

Review



**Cite this article:** Liénard MA, Valencia-Montoya WA, Pierce NE. 2022 Molecular advances to study the function, evolution and spectral tuning of arthropod visual opsins. *Phil. Trans. R. Soc. B* **377**: 20210279. <https://doi.org/10.1098/rstb.2021.0279>

Received: 9 December 2021  
Accepted: 21 February 2022

One contribution of 17 to a theme issue ‘Understanding colour vision: molecular, physiological, neuronal and behavioural studies in arthropods’.

**Subject Areas:**  
ecology, evolution, molecular biology

**Keywords:**  
light-sensitive opsins, visual pigments, heterologous expression, comparative sequence analysis, protein modelling

**Author for correspondence:**  
Marjorie A. Liénard  
e-mail: [marjorie.lienard@biol.lu.se](mailto:marjorie.lienard@biol.lu.se)

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.6098624>.

# Molecular advances to study the function, evolution and spectral tuning of arthropod visual opsins

Marjorie A. Liénard<sup>1,2</sup>, Wendy A. Valencia-Montoya<sup>2</sup> and Naomi E. Pierce<sup>2</sup>

<sup>1</sup>Department of Biology, Lund University, 22362 Lund, Sweden

<sup>2</sup>Department of Organismic and Evolutionary Biology and Museum of Comparative Zoology, Harvard University, Cambridge, MA 02138, USA

MAL, 0000-0003-3193-3666; WAW, 0000-0001-9246-2330; NEP, 0000-0003-3366-1625

Visual opsins of vertebrates and invertebrates diversified independently and converged to detect ultraviolet to long wavelengths (LW) of green or red light. In both groups, colour vision largely derives from opsin number, expression patterns and changes in amino acids interacting with the chromophore. Functional insights regarding invertebrate opsin evolution have lagged behind those for vertebrates because of the disparity in genomic resources and the lack of robust *in vitro* systems to characterize spectral sensitivities. Here, we review bioinformatic approaches to identify and model functional variation in opsins as well as recently developed assays to measure spectral phenotypes. In particular, we discuss how transgenic lines, cAMP-spectroscopy and sensitive heterologous expression platforms are starting to decouple genotype–phenotype relationships of LW opsins to complement the classical physiological-behavioural-phylogenetic toolbox of invertebrate visual sensory studies. We illustrate the use of one heterologous method by characterizing novel LW Gq opsins from 10 species, including diurnal and nocturnal Lepidoptera, a terrestrial dragonfly and an aquatic crustacean, expressing them in HEK293T cells, and showing that their maximum absorbance spectra ( $\lambda_{max}$ ) range from 518 to 611 nm. We discuss the advantages of molecular approaches for arthropods with complications such as restricted availability, lateral filters, specialized photochemistry and/or electrophysiological constraints.

This article is part of the theme issue ‘Understanding colour vision: molecular, physiological, neuronal and behavioural studies in arthropods’.

## 1. Opsins: spectral light-sensitive receptors

Animal visual opsins are a group of proteins that belong to the superclass of G-protein-coupled membrane receptors (GPCRs) and form functionally diverse light-sensitive photopigments [1,2]. Opsins are unique among GPCRs because each one permanently binds an inactive form of its specific ligand, rendering almost an instantaneous response upon photo activation [3,4]. Photosensitivity is enabled by the bound molecule or chromophore, *cis*-retinal (or a relative) typically derived from vitamin A [3]. The free retinal chromophore maximally absorbs photons of light in the ultraviolet (UV) around 380 nm but its sensitivity is shifted to react to higher energy, shorter wavelengths, when bound to a UV short-wavelength (SW) opsin, and to lower energy, longer wavelengths when anchored to a blue middle-wavelength (MW) or long-wavelength (LW) opsin [4,5].

The chromophore forms a covalent bond with a lysine residue (K) in the ligand-binding site of the opsin via a protonated Schiff base. This bond is conserved across all animal opsins and is stabilized by a nearby negatively charged residue, the counterion [4,5]. When the inactive *cis* form of the chromophore

absorbs a photon of light, it changes to a new all-*trans*-retinal conformation. This, in turn, triggers a conformational change in the opsin coupled to a Gt protein in vertebrates or Gq protein in invertebrates, thus activating distinct downstream signalling cascades leading to the perception of light [3,6,7]. Since the chromophore-opsin complex (also called rhodopsin) is tuned to specific wavelengths of light, maximum opsin absorbance ( $\lambda_{\max}$ ) represents a directly measurable phenotype [8,9]. Although recent studies have uncovered non-visual roles of traditional visual opsins [10], such as temperature discrimination [11] and gustatory functions [12], the strong connection between genotype and spectral sensitivity phenotype makes opsins a powerful model to understand phenotypic adaptation.

Opsins originated early in the evolution of metazoans and duplicated to give rise to three major gene groups: c-opsins and r-opsins, the main visual pigments of vertebrates and arthropods, respectively, and the less known retinal G-coupled receptors/Go opsins [1]. Whereas vertebrate visual pigments are Gt-coupled proteins that activate a transducin protein (Gt) cascade leading to a hyperpolarization response, invertebrate opsins are coupled to Gq proteins that initiate the phosphoinositol cascade, resulting in a depolarization response by the photoreceptor cells [1,7,13]. In addition to activating different phototransduction pathways, each group contains distinct opsin genes that give rise to visual pigments that evolved from a single ancestor via idiosyncratic gene duplications and absorb maximally at different wavelengths [1,7,13]. The ancestral repertoire of vertebrate visual c-opsins includes four gene families, the short-wavelength-sensitive SWS1 (UV,  $\lambda_{\max}$  344–445 nm) and SWS2 (blue,  $\lambda_{\max}$  400–470 nm), and the middle and long-wavelength-sensitive opsins, MWS (green,  $\lambda_{\max}$  480–530 nm) and LWS (red,  $\lambda_{\max}$  500–570 nm), respectively [14–16]. The arthropod visual r-opsins comprise five well-characterized families, including short-wavelength sensitive (SW), middle-wavelength sensitive (MW1 and MW2), and long-wavelength sensitive (LW1 and LW2) [6,17]. Ancestral insect lineages very likely harboured three opsin types, an SW (or UV), an MW (or blue) and an LW (or green) opsin, with gene expansions and/or losses accounting for the remarkable opsin diversity observed across extant insect groups [17,18]. In summary, vertebrate and invertebrate opsin families are functionally analogous, yet phylogenetically distantly related, having converged independently to absorb similar ranges of the light spectrum.

The primary factor determining the spectral sensitivity of a photoreceptor is the absorption spectrum of the expressed visual pigment [1,3]. Optical and electrophysiological factors such as screening and filtering pigments, as well as neural processing, can further shift the peak sensitivity of a photoreceptor to a particular wavelength of light [3,17]. In addition to lateral filtering tuning effects, the absorption maxima of visual pigments can be shifted by changing the chromophore type from the widespread retinal (A1) or hydroxyretinal (A3) to dehydroretinal (A2), which prompts the visual pigment to absorb longer wavelengths [3,13]. Another mechanism involves changing opsin amino acid residues found within or adjacent to the chromophore-binding pocket, shifting the chromophore's ability to absorb specific wavelengths of light [3,9,19]. This mechanism accounts for the vast diversity of spectral sensitivities useful for colour vision and results from the dynamic history of diversification and molecular evolution of opsins [1,3].

The striking phenotypic convergence in spectral sensitivities across major vertebrate and invertebrate opsin groups does not necessarily imply convergence of tuning mechanisms at the molecular level. Since amino acid substitutions of the key residues surrounding the chromophore have occurred during the evolutionary history of each lineage, distinct tuning mechanisms are likely to have modulated spectral sensitivity of different opsin groups [20,21]. Mechanisms underlying spectral tuning in vertebrate visual opsins have been characterized in detail, notably by means of spectroscopic analysis with recombinant proteins of wild-type and mutant opsins [22–25]. By contrast, the tuning mechanisms of invertebrate opsins are still largely unknown owing to the difficulty of obtaining sufficient yields of purified recombinant pigments for spectroscopic analyses, particularly for certain groups of opsins [20,26,27]. Nevertheless, recent years have seen remarkable progress in the availability of invertebrate genomic resources, in the improvement of bioinformatics and modelling tools, and principally, in the development of robust molecular assays to study spectral sensitivities of invertebrate visual pigments [20,27–30].

Here, we review these advances and summarize a framework for identifying and testing molecular variation relevant for spectral tuning in invertebrate opsins. We capitalize on some of these techniques and show their promise by characterizing sensitivities of LW opsins across diverse taxa, including a crustacean, an early divergent insect (a dragonfly), and several nocturnal lepidopteran species. Finally, we discuss the role of these methods in advancing knowledge on the mechanistic basis of functional convergence between vertebrate and invertebrate visual pigments and the evolutionary pathways modulating spectral tuning of colour vision genes.

## 2. Comparative sequence analysis of opsins

### (a) Mining, alignment and variation of opsin sequences

The lag in our understanding of tuning mechanisms of arthropod light-sensitive pigments compared to those of vertebrates is partly owing to the relative scarcity of genomic and transcriptomic data. However, the dramatic drop in sequencing costs has led to large-scale initiatives to sequence genomes and transcriptomes of thousands of arthropod species in the last decade or so [31–37]. Currently, more than 2700 assembled arthropod genomes are available via GenBank, with approximately 60% released in the last year. Pilot projects such as the Earth Biogenome project, the Tree of Life project by the Wellcome Sanger Institute, the i5k initiative to sequence 5000 arthropod genomes, as well as the 1KITE to sequence more than 1000 insect transcriptomes have greatly improved data availability [31,32,36–38]. This abundance of new genomic and transcriptomic information has opened the door for studies of functional genomics of non-model organisms and comparative sequence analyses of opsins across a larger taxonomic breadth.

Once the genomic or transcriptomic resources have been found, three main approaches are routinely used to mine opsin sequences to identify orthologues and duplicated genes. These include sequence similarity, protein prediction through hidden Markov models (HMM), and phylogenetic inference [39–44] (table 1). The first two homology methods, sequence similarity and HMM profiles, can fall short in

**Table 1.** Summary of steps for comparative analysis of opsin sequences, including selected programs and resources available for the different steps.

step	aim	suggested programs or pipelines
<b>mining and annotation</b>	retrieve opsin sequences	<b>programs:</b> BLAST [39], HMMER [40,45], GMAP [41], BLAT [42], AUGUSTUS [46] <b>pipelines:</b> BRAKER [47], MAKER [48], phylogenetically informed annotation (PIA) [8,49].
<b>identifying orthologues and paralogues</b>	discriminate duplicated genes	in principle, the resources used for mining and annotation can aid in finding duplicates, but some specific tools to discriminate orthologues from paralogues include ORTHOFINDER [50], i-ADHoRe [51], MCSANX [52], InPARANOID [53].
<b>alignment and variant discovery</b>	characterize variation	<b>multiple sequence alignment:</b> MAFFT [54], MUSCLE [55], CLUSTAL OMEGA [56]. <b>mapping to reference sequence:</b> BWA [57], BOWTIE [58], STAR [59]. <b>variant calling:</b> GATK [60], bcftools [61], FREEBAYES [62].
<b>ancestral sequence reconstruction</b>	reconstruct likely sequences at specific nodes	<b>tree inference:</b> IQ-TREE [63], RAXML-NG [64], PHYML [65]. <b>ancestral reconstruction:</b> ANCESCON [66], PHYLOBOT [67], PROTASR2 [68], FIREPROT <sup>ASR</sup> [69], FASTML [70], PAML [71].
<b>selection analyses</b>	find residues under selection	<b>population level: outlier analyses (OA):</b> ARLEQUIN [72], FDIST [73], BAYESCAN [74]. <b>environmental (EEA):</b> BAYENV/BAYENV2 [75], LFMM [76], RDA [77]. <b>for diverged sequences: dS/dN ratio:</b> PAML [71], HyPHY [78]. <b>McDonald and Kreitman test:</b> iMKT [79]
<b>homology modelling</b>	identify candidate tuning sites through structure prediction	<b>programs:</b> MODELLER [80], SWISS-MODEL [81], PHYRE2 [82], ALPHA FOLD [83], <b>visualization:</b> PyMOL [84], UCFS CHIMERA-A [85].

identifying opsins in distantly related species given the low similarity of some opsins (less than 50%) in different families and the high computational costs when analysing large datasets [49]. Conversely, phylogenetic approaches for functional annotation of opsins have been shown to overcome the shortcomings of homology-based methods [49,86]. In fact, the robust and efficient phylogenetically informed annotation (PIA) pipeline, primarily developed to mine for opsins, has recently been updated and used to explore patterns of opsin evolution across diverse arthropods groups such as amphipods, crustaceans, and butterflies and moths [49,86]. Alternatively, non-opsin light-sensitive GPCR proteins such as in the bay scallop, *Argopecten irradians*, have been recovered by combining RNA-sequence analysis and protein modelling workflows that target chromophore-binding lysine residues in positions that differ from the canonical lysine on the seventh transmembrane domain in opsins [87].

After retrieving candidate coding sequences, the second step is to spot genetic variation such as single nucleotide polymorphisms (SNPs), deletions, or insertions possibly linked to functional divergence. Although the direct association between phylogenetic groups and the capacity to absorb within a specific range of the spectrum make opsins amenable to functional classification solely based on similarity (UV, blue, LW), even moderate variation can translate into significant shifts in peak absorbance within groups [22]. To characterize variation among orthologues or duplicated genes, alignments are routinely performed using various algorithms (table 1). Beyond outlining variation between extant sequences, studies of vertebrate opsins have advocated for the use of ancestral sequence reconstruction (ASR) to infer the most plausible sequence at the root of the common ancestor between opsins of interest [88–90] (table 1). Identifying variants between ancestral and extant sequences has helped to reconstruct plausible scenarios for

the historical accumulation of mutations and the order in which they appeared, as well as the evolutionary trajectories of spectral tuning of SWS1 visual pigments in vertebrates [23,24,91,92]. This approach, together with selection analyses, has been used to formulate hypotheses about opsin evolution in invertebrates, mostly Lepidoptera, Coleoptera and Hemiptera [93–96]. Nonetheless, candidate spectral tuning sites have seldom been validated, in sharp contrast to vertebrate opsins [27,29,30].

Most of the sequence variants identified in the previous step are likely to be synonymous changes or fall outside the binding pocket, thus playing a moderate part in generating opsin spectral diversity. The third step, therefore, is to narrow down the pool of variants to those that are more likely to be responsible for phenotypic changes in absorption maxima or protein stability, and thus might have been driven by or maintained by selection. Across species, rates of synonymous (dS) and non-synonymous substitutions (dN) and their ratio (dS/dN) are widely used to infer positive, negative or neutral selective pressures acting on opsin genes [97–99]. Using this method, studies have shown faster evolution of opsins in diurnal taxa, or more specifically, adaptive evolution in UV-sensitive opsins in day-flying insects and LWS opsins of day-flying Lepidoptera [6,96,100,101]. In addition, site-specific tests of dipteran, lepidopteran and stomatopod crustacean opsins have suggested positive selection at residues outside the chromophore-binding pocket; thus, hinting at adaptive roles potentially decoupled from spectral phenotypes [100,102,103]. At the population level, studies detecting genetic variants of opsins under selection have provided insights into the genetic basis of local adaptation to light environments and speciation [30,90,104,105]. The most widely used approaches to detect putatively adaptive SNPs at the population level are  $F_{ST}$  outlier analysis (OA), which identifies variants with higher genetic differentiation among

populations than expected under a neutral model, and environmental association analysis [106] (table 1). These approaches have been commonly used in vertebrates such as cichlids, guppies, sticklebacks, flycatchers and New World monkeys [107–111]. By contrast, only a handful of studies have characterized intraspecific variation of opsin genes in arthropods [30,112]. Consequently, little is known about the relationship between intraspecific variation and diversity of spectral sensitivities, and for the few examples known in arthropods, opsin diversification within species appears to be more consistent with natural selection unrelated to spectral tuning [30,112].

### (b) Protein modelling

Since visual opsins are also known to have adaptive extraocular functions [113], amino acid sites under selection may not be reliable predictors of variation in absorption maxima. Changes in the topographical distributions and orientation of the amino acid side chains surrounding the chromophore are primarily responsible for determining  $\lambda_{\max}$  values, and thus building three-dimensional models of the visual pigments is essential [4,5,114]. In the absence of experimentally determined protein three-dimensional structures for the vast majority of opsins, homology modelling plays a cost-effective role [115]. The process begins by choosing the best template three-dimensional structure on which the target sequence can be successfully threaded [115]. For three-dimensional modelling of arthropod opsins, the squid rhodopsin template is preferred compared to the bovine rhodopsin because unlike the monostable and Gt-coupled visual opsins of vertebrates, squid and arthropod opsins are Gq-coupled and bistable, undergoing a 2-photon bidirectional photoreaction, with the retinal bound throughout the cycle [4,19,24]. The first crystal structure of a Gq-coupled opsin, jumping spider rhodopsin-1, was also published in 2019, providing additional insights into the molecular architecture of Gq-opsins and constituting a second available template for modelling arthropod visual pigments [4].

A series of alignments and alignment correction rounds are performed based on the template, followed by the generation of the backbone three-dimensional structure [115]. In opsin studies, the most frequently used tools have been the SWISS-MODEL and the MODELLER program, which are knowledge-based evaluation methods, relying on scores representing energies obtained statistically within the context of all known experimental three-dimensional structures in the database [115] (table 1). For other photopigments with unknown homologous structures, machine learning-based methods such as ALPHA FOLD2 [83] are expected to shed light on the molecular architecture of the visual pigments of non-model organisms or elusive families of opsins such as the non-visual groups. Finally, the predicted three-dimensional models and the chromophore are visualized to identify the amino acids lining in the binding pocket, which may be responsible for changes in  $\lambda_{\max}$  values. As a rule of thumb, variants located less than or within 5 Å from the chromophore are spectral tuning sites candidates [116,117]. This is because 5 Å is the maximal length of weak hydrogen bonds, and the longest bonds between the visual pigment and the retinal that can alter the shape of the electrostatic environment within the binding pocket [116,117]. Although the variants surrounding the

chromophore are expected to influence opsin tuning, ultimately, their role in modulating absorbance can only be tested experimentally.

### (c) Modelling approaches to predict peak spectral sensitivities

Site-directed mutagenesis of spectral tuning site candidates and *in vitro* experiments have convincingly linked amino acid changes to shifts in  $\lambda_{\max}$  between opsin orthologues and duplicates, particularly in vertebrates [22,24,25]. Since the effects of some amino acid substitutions appear consistent across different vertebrate groups, these known spectral tuning sites have been used to infer  $\lambda_{\max}$  based on sequence data [23,25,118]. For instance, the ‘five-sites’ rule posits that the identities of five critical sites within the binding pocket of some mammalian LWS can predict peak spectra [118]. Similarly, the A/B ratio, a ratio between the hydrogen-bond areas formed by five key amino acids of the SWS1 opsin in vertebrates, is shown to relate to variation in UV-violet perception [23]. More recently, through molecular dynamic simulations, Patel *et al.* [119] developed a statistical model able to predict  $\lambda_{\max}$  values accurately for the blue opsin (Rh2) across teleost fishes. This study constituted an important leap forward in functionally predictive opsin modelling because it does not rely on data from laborious site-directed mutagenesis experiments, which have to account for additive as well as epistatic interactions. However, it is not clear whether this model extends beyond Rh2 opsins or to those from other animal taxa. In addition, although not requiring information about tuning residues, absorption maxima of the visual pigments are still necessary for model training. Thus, despite the potential for application of these methods to the prediction of  $\lambda_{\max}$  of invertebrate opsins, comprehensive baseline information about experimentally characterized  $\lambda_{\max}$  for purified invertebrate opsins remains largely limited. In the next section, we outline recent developments to characterize absorption spectra of Gq opsins *in vitro*, whose application to diverse organisms can fuel the development of predictive models tuned to arthropod visual pigments.

## 3. Molecular assays to elucidate absorbance maxima and tuning mechanisms

Molecular mechanisms that regulate invertebrate opsin absorption are still poorly understood [20,26]. However, in vertebrates, functional investigations of visual pigments have revealed critical mutations and amino acid substitutions responsible for adaptive shifts in absorption spectra [23,88,89,120]. This is predominantly because for the last 25 years or so, virtually any vertebrate opsin could be expressed and spectrally characterized *in vitro* [120]. In addition, vertebrate cells spontaneously bind 11-*cis*-retinal from solution, whereas for invertebrate opsins, it was not clear when this occurred or whether the retinal was eventually released [121]. Early attempts to evaluate the suitability of cell culture systems used in vertebrates to express invertebrate opsins, including mammalian COS1 cells, insect Sf9 cells, and amphibian *Xenopus* oocytes, failed to recover opsins capable of binding retinal [121]. At the time, these findings suggested that additional protein cofactors or chaperones for

invertebrate photoreceptors might be necessary for the functional expression of visual pigments [121]. However, expressing opsin complementary DNAs (cDNAs) of other invertebrate species in transgenic *Drosophila* succeeded in producing visual pigments that bound retinal, showed normal spectral properties, and coupled to the endogenous phototransduction cascade [121–123]. Consequently, *in vivo* assays using *Drosophila* mutants became among the first methods available to study tuning mechanisms of arthropod opsins [21,30,121–123].

### (a) Transgenic opsin rescue lines using *Drosophila*

Early studies of opsin absorbance using *Drosophila* took advantage of the 'blind' *ninaE*<sup>17</sup> mutant flies, which lack Rh1, the gene encoding the visual pigment expressed in the major class of photoreceptor cells [122]. Accordingly, the expression of functional opsins in these mutants rescued light response fully [122]. Thus, the opsin of interest, or modified versions by site-directed mutagenesis could be subcloned into an expression cassette containing the native *Drosophila* Rh1 promoter [21,122,123]. This construct is introduced into the genome of *ninaE*<sup>17</sup> by *P*-element mediated germline transformation, and spectral sensitivities are then measured by electrophysiological analyses with electroretinogram recordings of live flies exposed to different wavelengths of light [21,122,123] (figure 1a).

Unsurprisingly, the spectral properties of different *Drosophila* visual pigments were among the first to be characterized in arthropods, followed by opsins of the honeybee and *Limulus* horseshoe crabs [121,122,124]. The first study that mutated insect visual pigments was conducted using this method. Salcedo *et al.* [123] studied spectral tuning of *Drosophila* blue and UV receptors (Rh1 and Rh3, respectively) (2003) [123], and later the LWS (Rh6) (2009) receptor [21]. They identified the variants UV: lysine versus blue: asparagine or glutamate (G90 in bovine rhodopsin system) that shift the absorption of the blue (Rh1) pigment into the UV, as well as the serine/alanine substitution (Ala-292 in bovine rhodopsin) residues implicated in changing the absorption of LWS (Rh6) to shorter wavelengths [21,123]. Interestingly, this site (Ala-292, in bovine rhodopsin) is also responsible for regulating LWS sensitivity in birds and mammals, thus providing the first example of a convergent tuning mechanism between vertebrate and invertebrate visual pigments [21]. More recently, the *in vivo* spectral sensitivity of *Drosophila* photoreceptors was re-examined by selectively restoring photoreceptor activity in flies with otherwise no receptor activity (*norpA* mutants) [125]. Sharkey and colleagues engineered single-opsin rescue fruit fly lines expressing phospholipase C (PLC, encoded by *norpA*) under the control of a specific opsin promoter, which restored sensitivity of each pigment in its native photoreceptor cell surrounded by its native lateral pigment [125]. This approach makes it possible to express and record electroretinogram (ERG) activity from mutant opsins in a natural photoreceptor environment.

### (b) Towards the development of heterologous systems

Despite the significant contributions of transgenic *Drosophila* methods to the early understanding of spectral tuning of invertebrate opsins, these *in vivo* assays made it difficult to disentangle the contributions of filtering and screening

pigments to the spectral properties of photoreceptors from that of pure opsins, although the use of white-eye (*w*) mutants tends to reduce these effects. To understand the molecular evolution of spectral tuning sites, direct analysis of purified visual proteins is indispensable [20,26]. However, studies exploring the molecular bases of spectral tuning in invertebrate opsins remained limited by the lack of an optimized *in vitro* expression system that would yield sufficient active rhodopsin complexes comparable to those recovered for vertebrate opsins [20,26].

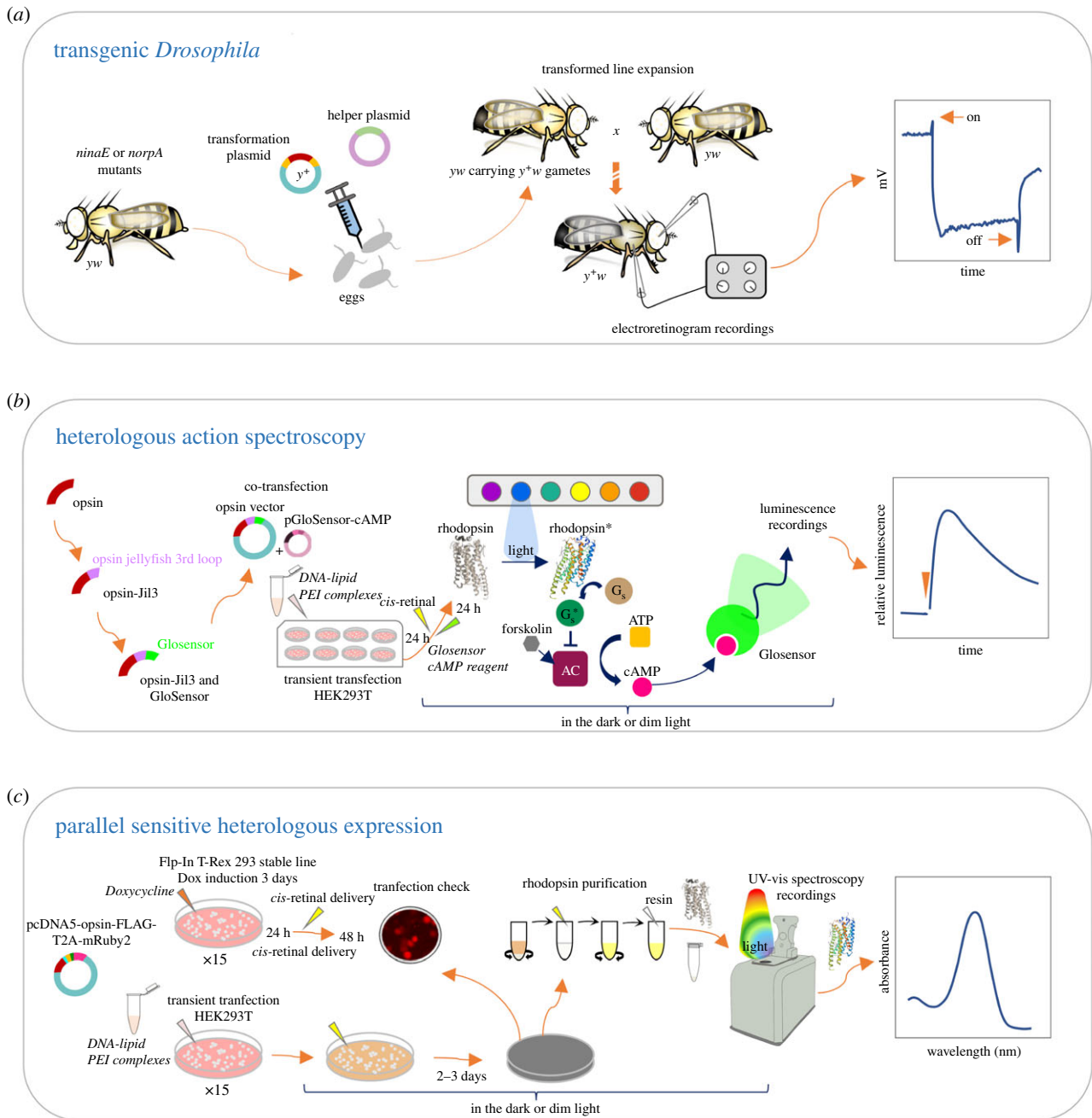
### (i) Short-wavelength opsin expression systems

Terakita and colleagues in 2008 were the first to describe heterologous expression of a Gq-protein-coupled opsin in cultured cells [26]. Unlike previous attempts, they used the human embryonic kidney (HEK) 293 cell system, which was a cultured tissue not surveyed before by the Knox *et al.* study [121]. Terakita *et al.* [26] successfully obtained protein yields of UV, blue and LWS opsins of the Japanese honeybee, the first arthropod opsins ever expressed in cultured cells [26]. However, although the expression of the three proteins was detected, pigment formation was not observed for the LWS opsin, suggesting that folding of this protein did not occur [26]. Similarly, they were unable to obtain active proteins for other Gq-coupled opsins such as those for molluscs and crustaceans, leading them to suggest that higher thermal stability or the need for other co-expressed chaperones might be required to obtain functional LWS opsins and other unsuccessfully expressed pigments [26].

Building upon the Terakita *et al.* [26] HEK cell assay, which proved to be efficient for studying opsins absorbing in shorter wavelengths [26], Wakawuka and colleagues [27] subsequently studied blue duplicated opsins of the cabbage white butterfly *Pieris rapae* [27]. They reconstituted the blue and violet receptors *in vitro* using HEK293 cultured cells and used spectroscopy to characterize absorption maxima [27]. Through site-directed mutagenesis, they found two amino acid substitutions between the duplicate copies crucial for shifting spectral sensitivity and whose positions suggested a tuning mechanism specific to invertebrates [27]. Frentiu *et al.* [30] also used HEK293 cultured cells to study SW opsins but focused on the role of intraspecific variation in contrast to variation between duplicated opsins [30]. They developed a first stable cell culture system for the expression of insect opsins to test the effect of clinal variation on spectral tuning of the blue opsin of *Limenitis arthemis* populations across a latitudinal gradient [30]. Stable cell lines can carry a genetic modification resulting in constant and modulable expression levels, which, if the opsin folds properly, yield higher protein levels for accurate measurement of  $\lambda_{max}$  [30]. In summary, until recently *in vitro* studies of arthropod opsins primarily allowed reconstituting visual pigments for SW and MW opsins from the UV and blue opsin families.

### (ii) Long-wavelength opsin expression systems

LW pigments were believed to be difficult to express in cultured cells on a large scale until optimized assays emerged to circumvent this problem [20,29]. To the best of our knowledge, two methods have been successful at consistently characterizing absorption spectra of arthropod LW opsins



**Figure 1.** Overview of molecular methods used to analyse spectral sensitivities of invertebrate opsins. (a) *Transgenic Drosophila*. An expression plasmid carrying an opsin of interest under control of the Rh1, *norpA* or other promoter, is co-injected with a helper plasmid to control recombination and gene cassette insertion in the embryonic germ line pole. Injected developing flies are mated and their  $F_1$  progeny screened for the presence of positive transformants, typically via  $y^+w$  phenotypes as schematized, or via mini-white  $w^+$  eye marker phenotypes. The spectral sensitivity of fruit fly photoreceptors from final expression lines bearing opsin genes and markers of interest can be phenotyped by directly analysing pigment absorption of crude eye extracts [30] or via electrophysiology [21,122–125]. See text for details. (b) *Heterologous action spectroscopy (HeAS)*.  $G_s$ -coupled opsins are engineered by replacing their third cytoplasmic loop with that of the  $G_s$ -couple jellyfish opsin. The opsin expression vectors are co-transfected with the pGloSensor cAMP plasmid. Following addition of 11, *cis*-retinal and pGloSensor reagents, and a 2-day opsin expression time course, light-induced changes in luminescence triggered by increased concentration of cAMP are measured by irradiation with lights of different spectral peaks. Modified based on [19,20,28]. See main text for details. (c) *Parallel sensitive heterologous expression (PaSHE)*. Opsin open reading frames amplified from eye cDNA or synthesized for codon-optimization are individually ligated into a derivative of the pcDNA5 expression vector fused to a C-terminal epitope flanked by a peptide linker, a 21-amino acid long T2A cleavage site and a cytoplasmic fluorescent marker for visualizing cell transfection efficiency. Constructs can be transfected in the Flp-In Trex 293 cell line alongside a co-helper plasmid pOG44 to induce recombination and integration in the genome. Stable transformants are generated upon antibiotic marker selection (Hygromycin and Blastidicin), and opsin integration is validated through Sanger sequencing. A time course expression induction is then conducted in the presence of doxycycline, followed by quantitative polymerase chain reaction and western blot analyses (see the electronic supplementary material, methods). The stable validated line is expanded and Dox-induced in the presence of 11, *cis*-retinal prior to purifying the reconstituted visual pigments (see the electronic supplementary material, appendix). Stable lines can also be generated using alternative expression vectors, cell lines and/or epitopes (see [30,54]). Alternatively, the same expression vector can be used to transiently transfect HEK293T cells (24  $\mu$ g DNA/plate at  $4.10^6$  cells), prior to 11, *cis*-retinal delivery under dim light. Opsin expression and trafficking to the plasma membrane is typically detected within 2–3 days, which should be verified by fluorescent microscopy and western blot analyses prior to conducting large-scale purification. In both method variants, cells expressing opsins (induced or transient) are harvested, rhodopsin complexes are natively solubilized with additional 11, *cis*-retinal and then solubilized from plasma membranes prior to purification by resin affinity, concentration and UV-vis spectroscopy recordings to obtain the visual pigment dark absorbance spectrum. Modified based on [29].

*in vitro*: heterologous action spectroscopy (HeAS), and parallel sensitive heterologous expression (PaSHE) [20,29].

**Heterologous action spectroscopy.** This first method was initially implemented to study non-visual opsins, such as Opn3, with unknown signalling cascades [19,28]. The first arthropod opsin whose spectral sensitivity curve was accurately estimated using heterologous action spectroscopy was the Gq-coupled jumping spider Rh1 (UV) [19,28]. More recently, Saito *et al.* [20] used HeAS to pinpoint the main helices and residues therein responsible for spectral tuning of duplicated LW opsins of the butterfly *Papilio xuthus*, thus showing that this method is not only functional for UV visual pigments but also for studying LW opsins [20].

Heterologous action spectroscopy is based on the quantification of light-dependent changes in the cyclic second messenger of cultured cells expressing opsins of interest [20,28] (figure 1*b*). Since arthropod Gq-coupled visual opsins drive phosphatidylinositol-related cascades instead of cyclic nucleotides cascades like vertebrate Gt-coupled, Gs, Gi and Go-coupled cnidarian opsins, which use the cyclic nucleotides cGMP and cAMP as second messengers in the phototransduction cascade [7,19], the first step to enable heterologous action spectroscopy is to engineer Gs-coupled versions of the Gq-coupled opsins by replacing the third cytoplasmic loop with that of the Gs-coupled jellyfish opsin by polymerase chain reaction [20,28]. The third intracellular loop is the one replaced because it is known to be a major determinant of G protein selectivity in opsins [19]. By changing its third loop, an opsin can be expressed *in vitro* even if the G-protein with which it couples is unknown [28]. The chimeric Gs-version of the protein of interest is transfected into HEK293 cells with the GloSensor plasmid, which codes for a cAMP-sensitive luciferase [20,28] (figure 1*b*). Upon activation with light stimuli, changes in cAMP levels are measured based on the luminescence derived from the GloSensor luciferase protein [20,28] (figure 1*b*). Light sources with different spectral emission peaks (from 410 nm up to 630 nm) are used for the estimation of wavelength-dependent changes in intracellular cAMP in response to each light stimulus [20,28] (figure 1*b*). To calculate relative sensitivities, the amplitude of these wavelength-dependent responses are extrapolated using a light intensity dose–response curve generated by quantifying responses to a single light stimulus (green 500 nm, orange 600 nm) while varying the intensity [20,28]. Finally, the absorption spectrum is estimated by fitting these relative sensitivities to a rhodopsin template [20,28].

**Parallel sensitive heterologous expression.** PaSHE is an *in vitro* assay that allows efficient purification of arthropod visual opsins for spectroscopic analysis [29]. PaSHE is essentially a heterologous expression system that builds upon previous protocols used to express opsins in HEK293T [26,27,30,87], but through a series of optimized steps is able to consistently recover active rhodopsin complexes for the different Gq opsin families, including the elusive LWS [29]. Synthetically codon-optimized opsin sequences are effectively expressed, although constructs using native sequences work comparably well depending on the species investigated. Briefly, this simple system uses an expression cassette under a strong cytomegalovirus promoter, engineered to contain a mammalian transcription sequence upstream of the opsin start codon, and fused in 3' with a Flag epitope [29]. The Flag epitope is directly flanked by a T2A peptide and a cytoplasmic fluorescent mRuby2 marker (figure 1*c*). The T2A peptide contains

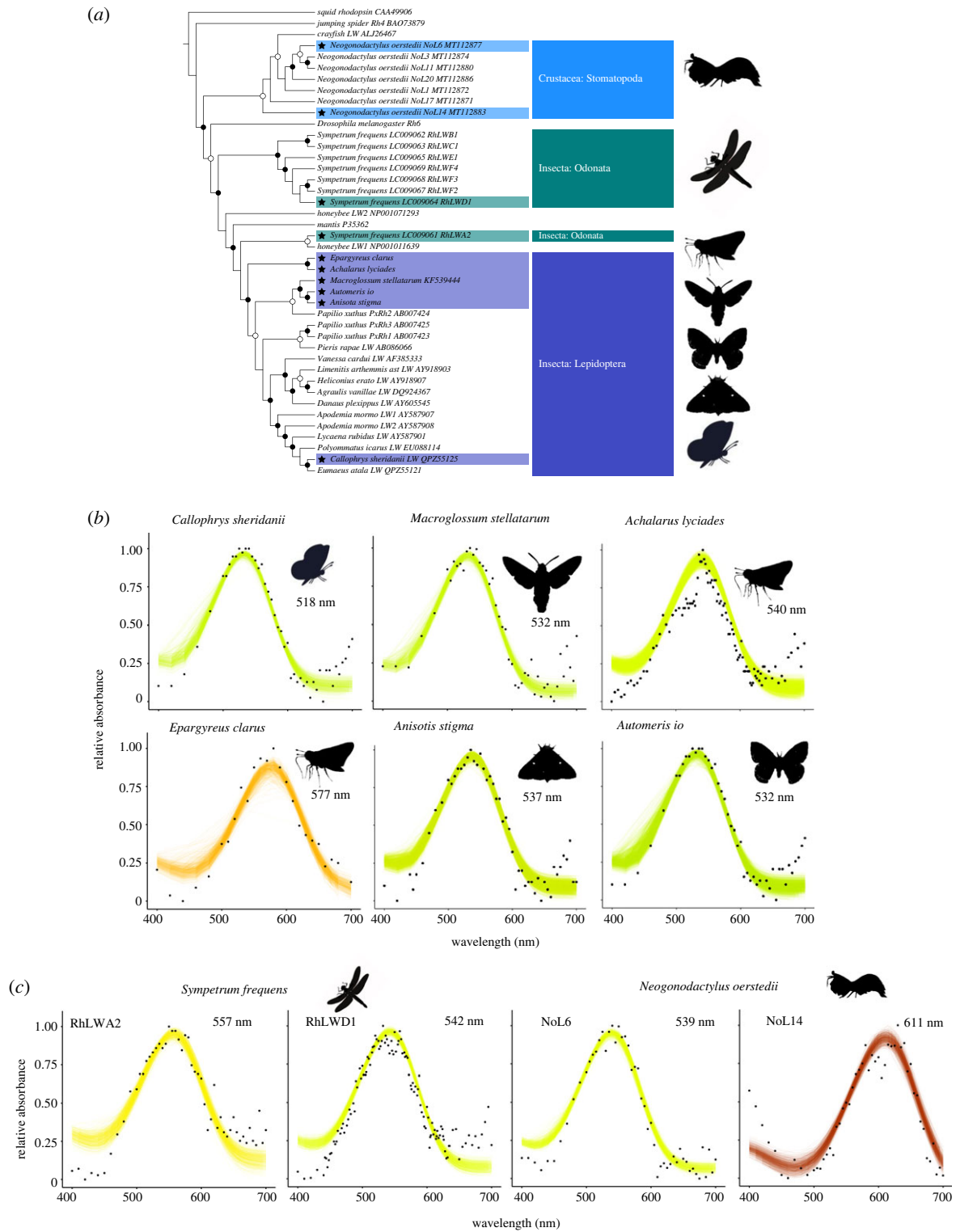
a Proline residue that mediates cleaving of the co-translated proteins [126]. Small-scale transfections and western blot are performed to select high expression plasmids, followed by large-scale purification. Transfected plates are supplemented with *cis*-retinal after 6 h and incubated in the dark for at least two nights to maximize membrane protein expression. Cells are harvested, and opsin-FLAG proteins are solubilized from plasma membranes then separated from the crude extract using FLAG resin and a subsequent column purification prior to competitive binding with FLAG-peptide. Finally, rhodopsin complexes are eluted and concentrated by centrifugation. Dark absorbance spectra of concentrated opsins are measured using UV–Vis spectroscopy, followed by estimation of  $\lambda_{\max}$  by fitting the raw data to a visual template (figures 1*c* and 2*b,c*).

The methods described above to characterize  $\lambda_{\max}$  *in vitro* open the door to dissect LW opsin tuning mechanisms and to start answering long-standing questions about the genetic basis of functional convergence between invertebrate and vertebrate LW opsins. For example, in a series of elegant chimeric and site-directed mutant experiments, Saito *et al.* [20] used HeAS to successfully narrow down two single mutations in helix III, A116G (+6 nm) and F120Y (+3 nm), that synergistically contribute to a +15 nm shift between duplicated PxRh1 (540 nm) and PxRh3 (570 nm) *Papilio* opsins [20]. By contrast, two sites in helices IV and VI in LW duplicated opsins in primates are the main contributors to their 30 nm difference in  $\lambda_{\max}$ ; thus suggesting different mechanisms underlying similar shifts in absorbance in primates and insects [20]. However, it is not clear how general the LW tuning mechanism observed in *Papilio xuthus* is across butterflies and more generally, across arthropods. A first step towards investigating the potential diversity of spectral sensitivities and tuning mechanisms of LW across arthropods is to increase the sampling and characterization of LW opsins, especially for independently duplicated opsins. In the next section, we present new data for LW absorption spectra across Lepidoptera and other arthropod groups to show the range of spectral sensitivities spanned by this gene family and as a baseline for studying independently evolved LW duplications in invertebrates.

## 4. Functional insights into arthropod long-wavelength opsin sensitivities

To expand the range of PaSHE applications to study Gq opsins of more distantly related arthropod groups and gain insights into spectral sensitivities of invertebrate LW opsins, we analysed 10 additional LW opsins across two insect orders and a crustacean. In this study, we aimed to encompass phylogenetic and ecological diversity by comparing: (i) diurnal and nocturnal lepidopteran species, (ii) species with terrestrial and/or aquatic lifestyles, and (iii) species with a large number of duplicated LW opsins (figure 2*a*). For each LW opsin, we used PaSHE to obtain dark absorbance spectra from purified LW Gq opsins (figure 2*b,c*; electronic supplementary material, dataset S1, SI appendix, figures S1 and S2), including six lepidopteran species from four families (Lycaenidae, Hesperidae, Sphingidae and Saturniidae) (figure 2*b*), a dragonfly, *Sympetrum frequens* (Odonata) and a stomatopod crustacean (*Neogonodactylus oerstedii*) (figure 2*c*).

Insect LW photoreceptors are known to harbour a large amount of spectral variation, accounting for rhabdomeric



**Figure 2.** Functional expression of insect and crustacean long-wavelength (LW) opsins. (a) Maximum-likelihood phylogeny of LW opsins in exemplar taxa from Insecta, Arachnida and Crustacea. Branch names comprise the full species name, the GenBank accession number and sequence acronym. LW opsins functionally characterized in this study are labelled with a black star. (b,c) Dark absorbance spectra of long-wavelength rhodopsin visual pigments reconstituted and purified from cell cultures in the presence of 11-*cis*-retinal using the PaSHE workflow. (b) Dark absorbance spectra of selected long-wavelength lepidopteran insect opsins. (c) Dark absorbance spectra of selected long-wavelength odonate insects and stomatopod crustacean opsins. The black dots represent mean absorbances. Absorbance at 380 nm when present, is owing to residual unbound *cis*-retinal. Relative absorbance data are fitted to a visual template [127] with polynomial function analyses computed in R to obtain the best estimates of  $\lambda_{max}$  for each opsin. Number of protein eluate aliquot measurements ( $n$ ): *Callophrys sherdanii* ( $n = 4$ ), *Macroglossum stellatarum* ( $n = 6$ ), *Automeris io* ( $n = 2$ ), *Anisotis stigma* ( $n = 3$ ), *Achalarus lyciades* ( $n = 4$ ), *Epargyreus clarus* ( $n = 2$ ), *Sympetrum frequens* ( $n_{RhLWA2} = 3$ ,  $n_{RhLWD1} = 9$ ), *Neogonodactylus oerstedii* ( $n_{NoL6} = 9$ ,  $n_{NoL14} = 14$ ).

filtering pigments, functional variation in the expressed visual pigments and visual pigment co-expression, altogether leading to LW spectral sensitivity peaks ranging from 500 to

above 600 nm [17,128]. We show that variation at the photoreceptor level is mirrored by maximal peak sensitivities of individual rhodopsins (figure 2b,c; electronic supplementary



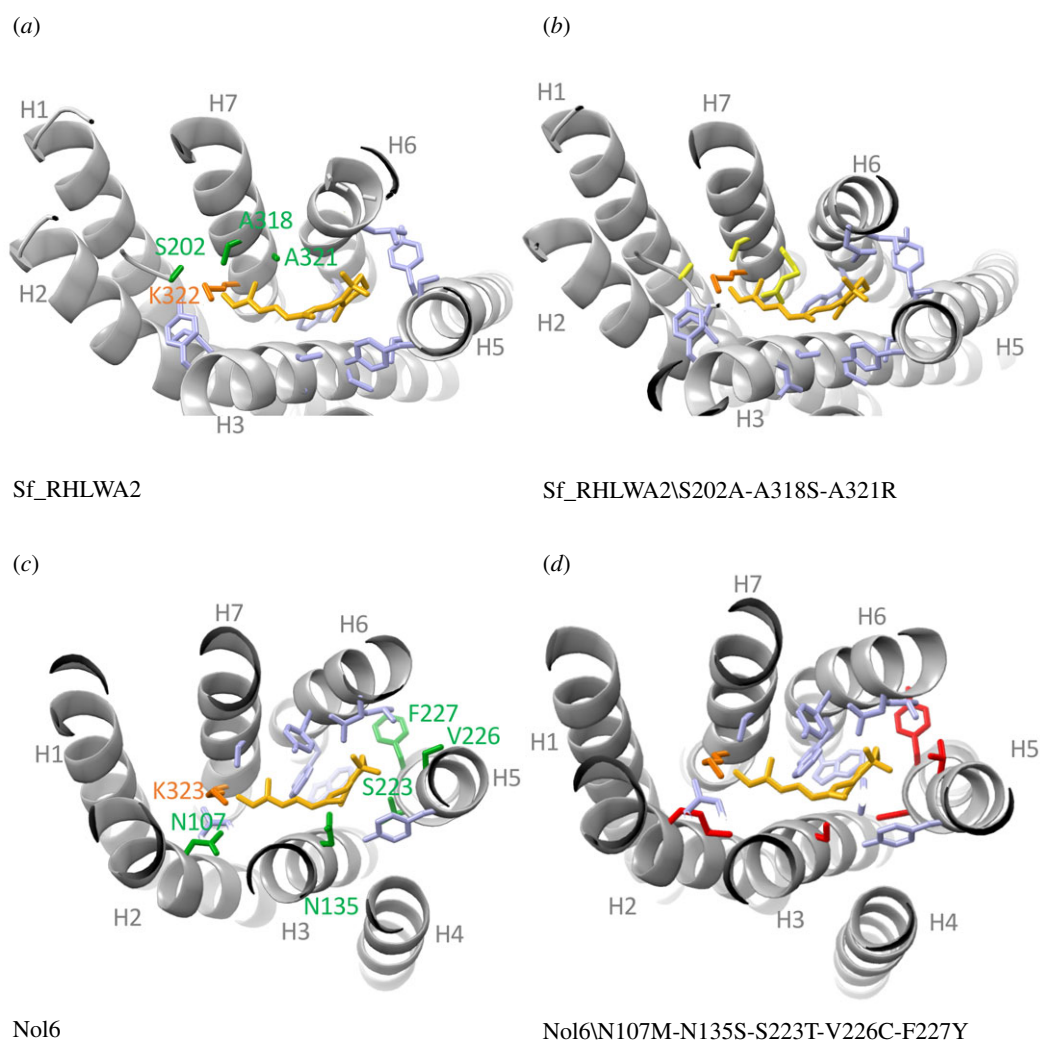
material, dataset S1). The reconstituted LW opsin from *Callophrys sheridanii* absorbed maximally at 517.8 nm. The LW opsin from the diurnal hawkmoth *Macroglossum stellatarum* absorbed at 531.7 nm, a peak absorbance a few nanometres above previous ERG recording estimates ( $521 \pm 3.6$  nm [127]), possibly owing to the optics of hawkmoth eyes and/or to using A1 chromophore instead of the native A3 retinal. LW opsins from diurnal and nocturnal saturniid moths, *Anisota stigma* and *Automeris io*, absorbed maximally at 537 nm and 532 nm, respectively. The two opsins share a high level of sequence similarity with only 11 variant residues, which our homology modelling predicts to fall outside candidate sites interacting with the chromophore. The close absorbance peaks are thus expected and support the reliability and reproducibility of the method in obtaining precise estimates of  $\lambda_{\max}$ .

Skippers (Hesperiidae) are unusual among butterflies for possessing eyes similar to the superposition eyes more typical of diurnal moths. While it remains a challenge to conduct *in vivo* studies of spectral sensitivities of visual pigments for superposition eyes, *in vitro* assays have the advantage of being effective regardless of eye morphology. We cloned an LW opsin *de novo* from the eye cDNA of a grass skipper *Epargyreus clarus* (Eudaminae), which has a single LW opsin gene, similarly to *Lerema accius* (Hesperiinae). Taking advantage of the high-quality genome of the Hoary edge, *Achalarus lyciades* (Hesperiidae, Eudaminae) [129], which inhabits open woodlands, forest edges and roadsides, we additionally mined its duplicated LW1 and LW2 opsin gene sequences (see methods). Of these, we chose to purify the paralogues *A. lyciades* LW2 and *E. clarus* LW opsins, which we found absorb maximally at 540 and 577 nm, respectively. These results illustrate the potential to explore visual gene functions from a genomic and functional angle, and to infer patterns regarding how ecological specializations might be correlated with adaptations in peripheral light-sensing genes.

For the LW opsin in the skipper butterfly *E. clarus* (figure 2b), we generated a Doxycycline-inducible stable line using the Flp-In TRex system (Invitrogen) as described in the electronic supplementary material, methods. The recombinant opsin expressed well from 3 days post-induction (electronic supplementary material, figure S2), a time point selected for purification. From spectroscopy measurements, we observed that the recombinant rhodopsin complexes have a two to threefold higher absolute absorbance (electronic supplementary material, dataset S1) consistent with earlier absorption measurement levels obtained with a non-inducible stable cell line expressing *Limnitis a. astyanax* BRh [30]. Establishing stable cell lines, however, takes time, requiring on average 30 days before obtaining stable polyclonal lines, plus additional time to validate the line and test optimal gene and protein expression time course. Alternatively, using a high-expression cytomegalovirus promoter together with native full-length transiently expressed opsins represents a time efficient, scalable method for parallel purifications, which is adaptable to all arthropod Gq opsins. Cloning opsin sequences sequentially *de novo* or reamplifying open reading frames from fresh biological material may also be replaced by synthetic codon-optimized sequences, providing a time and cost-effective option further maximizing high protein expression levels.

Most methodological advances to study arthropod LW opsins *in vitro* have focused on butterfly species. Thus, an outstanding question is whether paralogous LW opsin genes provide similar potential variability in their corresponding peak spectral sensitivities across more distantly related lineages. To explore this question, we experimentally measured peak spectral sensitivities of duplicated opsins in the autumn darter dragonfly, *Sympetrum frequens* (Sf) and the stomatopod crustacean, *Neogonodactylus oerstedii* (No), reconstituted with *cis*-retinal (figure 2c), a required step prior to inferring the ancestral sequences for extant LW opsins in these arthropod lineages, as well as their evolutionary trajectories and timing of spectral substitutions. We observed that Sf\_RhLWD1 and Sf\_RhLWA2, opsins expressed in the dorsal and ventral adult eyes across at least 12 surveyed dragonfly species [130] have peak absorbances for green wavelengths at 542 nm and 557 nm, respectively (figure 2c). Homology when compared with the jumping spider rhodopsin structure provided a reliable model, recovering all key conserved sites known to interact with the chromophore (lysine, counterion and glycines). Among 127 variant amino acid positions, three amino acids are within 5 Å of the vicinity of the chromophore (electronic supplementary material, dataset S3), including an Sf\_RhLWD1-Alanine to Sf\_RhLWA2-Serine substitution in EL2, which is known to cause shifts (towards longer wavelengths), and represents a robust candidate for *in vitro* site-mutagenesis experiments (figure 3a,b).

For the stomatopod, *N. oerstedii*, phylogenetic and *in situ* hybridization expression patterns previously showed that NoL6 and NoL14 belong to distinct phylogenetic opsin subgroups (A and F) known to be expressed in separate LWS photoreceptive units [102,131]. NoL6 is expressed in the mid-band eye region in ommatidial rows 5 and 6, also known to be tuned for polarization vision [132] and NoL14 is expressed across all seven (R1-R7) photoreceptive units in the dorsal and ventral eye hemispheres, but not in the mid band region [102]. Our proof-of-concept results show that both constructs formed active rhodopsin complexes in cultured cells in the presence of *cis*-retinal, with each purified visual pigment complex exhibiting maximal absorbance peaking at 539 nm (NoL6) and 611 nm (NoL14), respectively (figure 2c). The recorded peak sensitivities of stomatopod crustacean photoreceptors are notoriously variable resulting from rhabdoms divided into tiered layers, each expressing one or several distinct opsin visual pigments, alone or in addition to photostable filter pigments [132], which together produce up to 16 spectral receptor classes across the eye [102]. Opsin *in vitro* sensitivities are in line with the finding that NoL6 is co-expressed with a candidate middle-wave-sensitive (MWS) opsin, NoM10, resulting in photoreceptor sensitivity near 520 nm in mid-band rows 5 and 6 [102,132]. Although it is challenging to compare *in vitro* to microspectrophotometry estimates because of co-expression, our findings provide preliminary support for the functional distinctiveness of co-expressed LWS opsins. Although the two opsin proteins characterized here were selected in part because they are known to differ in expression patterns, phylogenetic distances and sequence (NoL6 and NoL14 share 69% identical residues and contain 155 amino acid variants), we identified a minimal number (five) of candidate spectral substitutions (electronic supplementary material, dataset S3) out of 30 previously predicted sites [102] that could underlie transitions from green to red-shifted absorbing opsins and can be readily tested *in vitro*.



**Figure 3.** Example of homology modelling to identify tuning sites among duplicate LW Gq opsins. (a) Structure of the dragonfly *Sympetrum frequens* Sf\_RhLWA2 opsin built against homology modelling with the jumping spider crystal structure (PDB 6i9k). Among 127 variant sites with Sf\_RhLWD1, three variant sites are within 5 Å of the *cis*-retinal chromophore. Helices are represented in light grey, the *cis*-retinal is represented in yellow, residues predicted to interact with the chromophore are in light blue (electronic supplementary material, dataset S3). Among these, the three variant candidate spectral sites are coloured in green. (b) Sf\_RhLWA2 structure with the three variant sites mutated to those of Sf\_RhLWD1 and coloured in yellow. (c) Structure of the green-sensitive NoL6 LW Gq opsin of the stomatopod *Neogonodactylus oerstedii* modelled against 6i9k. Among 155 variant sites with the red-shifted sensitive NoL14 opsin, five candidate spectral sites, coloured in green, are found in the vicinity of the chromophore. (d) NoL6 structure with mutated residues corresponding to NoL14 at each candidate spectral site. The templates and models were built in SWISSMODEL followed by visualization in PyMol.

## 5. Conclusion

In this essay, we compare and contrast available methods to gain insights into the diversity and function of invertebrate Gq rhodopsins. A combination of reproducible bioinformatic, modelling and heterologous methods are now available to characterize sequence variation and spectral sensitivity of virtually any invertebrate Gq opsin. These represent advances on multiple fronts: first, the growing availability of opsin sequences and bioinformatic tools allow for identification of residues under selection and systematic characterization prior to experimental testing of candidate tuning sites and/or duplication events. Second, *in vitro* methods such as HeAS and PaSHE make it possible to isolate the contribution of single pigments to visual sensitivity across invertebrate groups with complex visual systems and photoreceptor classes. As opsins show a strong connection between genotype and phenotype, these workflows pave the way to larger scale evolutionary and comparative studies to reveal the mechanistic basis of

visual adaptation across arthropods. These methods now allow us to retrace spectral tuning trajectories to elucidate the genetic basis of functional convergence and the role of epistasis, mutation and mutation-order accumulations in shaping spectral phenotypes.

This review focuses on methods to investigate visual opsins, as they are the most comprehensively studied to date. Nonetheless, these methods have the potential to be easily extended to examine the functions of extraocular opsins. In fact, HeAS was primarily developed to evaluate the light-dependency of vertebrate non-visual opsins expressed in the brain [28]. Similarly, heterologous systems expressing the non-visual Rh7 opsin were used to characterize their role as circadian photoreceptors in the brain [133,134]. Other non-canonical functions of opsins in thermosensation, hearing and taste are more challenging to research, although, in principle, these *in vitro* systems could be adapted with other recording assays to test a range of ligands and responses.

A remaining challenge to these methods is the speed and efficiency of screening for the presence and function of

different variants. The number of experiments grows exponentially with the number of candidate spectral sites and their potential interactions, placing constraints on the number of opsin variants that can be characterized simultaneously. Nevertheless, insights gained from being able to characterize LW opsins with spectral sensitivities absorbing far in the red have already brought promise for a diversity of new applications in optogenetics, which in turn can fuel the development of high throughput methods to screen arthropod visual opsins. Robust workflows to survey the diversity of opsins across arthropods provide the critical first step to uncovering the repertoire of GPCR opsins and thereby expanding our understanding of the full palette of wavelengths that are perceived and used by organisms across the Tree of Life.

## 6. Methods

### (a) Phylogenetic analyses

Predicted LW opsin amino acid sequences from selected arthropod species were retrieved from GenBank, aligned using MAFFT [54] computed with IQ-TREE [135] with LG + F + G4 as the best-fit substitution model [136], using the SH-aLRT branch test [137] and visualized using EVOLVIEW [138]. The *Achalarus lyciades* LW opsin sequence was retrieved from genomic data [129] (MOOZ01000541, scaffold887\_len869015\_cov66) and annotated using AUGUSTUS [46]. LW opsin sequences from *Callophrys sheridanii*, *Anisota stigma*, *Automeris io* and *Epargyreus clarus* were characterized from eye messenger RNA following a protocol previously described for *Arhopala japonica* [29].

### (b) Heterologous expression and purification of insect and crustacean opsin genes

To express the *E. clarus* LW opsin, we generated a doxycycline-inducible stable line using the native cDNA sequence as detailed in the electronic supplementary material, methods, figure S1, dataset S2). For other LW opsins, native open reading frames were amplified from cDNA (for *C. sheridanii*, *A. io* and *A. stigma*), or retrieved from GenBank with accession numbers listed in the electronic supplementary material, dataset S1 and codon-optimized sequences were synthesized by Genscript,

with restriction sites suitable for subcloning in pCDNA5-FLAG-T2A-mRuby2, then overexpressed by transient transfection in HEK293T following the PaSHE procedure [29] followed by purification, spectroscopy and western blotting (electronic supplementary material, methods, figure S2). For each rhodopsin, we obtained estimates of  $\lambda_{\max}$  through nonlinear least-square fitting to the absorbance data according to the template formula proposed by Govardovskii *et al.* [139]. We performed 1000 bootstrap replications to calculate  $\lambda_{\max}$  predictions and confidence intervals in R v. 3.6.6 using the packages *rsample* and *tidymodels*.

**Data accessibility.** The opsin sequence data for *A. stigma*, *A. io* and *E. clarus* are available in GenBank under accession nos OK930068–OK930070. Source data and code underlying functional expression, qPCR analyses and homology modelling are available as electronic supplementary material files [140].

**Authors' contributions.** M.A.L.: conceptualization, formal analysis, funding acquisition, visualization, writing—original draft, writing—review and editing; W.A.V.-M.: conceptualization, funding acquisition, visualization, writing—original draft, writing—review and editing; N.E.P.: conceptualization, funding acquisition, resources, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

**Conflict of interest declaration.** We declare we have no competing interests.

**Funding.** This work was supported by the Swedish Research Council (VR 2020-0517) to M.A.L., a grant from the Mind, Brain and Behaviour (MBB) interfaculty initiative at Harvard University to W.A.V.-M and N.E.P., and grants from the National Science Foundation (PoLS 1411123 and DEB 1541560) to N.E.P.

**Acknowledgements.** The authors thank Sam Jaffe (<https://www.thecaterpillarlab.org>) for sharing his deep knowledge of lepidopteran life histories and for providing pupae of *Anisota stigma*, *Automeris io* and *Epargyreus clarus*. We also thank Gary D. Bernard for stimulating discussion, insights and wise advice about invertebrate vision, for providing tissues of *Callophrys sheridanii* and for his feedback on the paper. Rosalie Crouch provided the 11-*cis*-retinal, Jean-Marc Lاسance annotated the skipper opsin sequence using AUGUSTUS and provided helpful guidance with fits in R, Brendan Dean Zhi Min and Siliang Song were extremely helpful in the laboratory and assisted in amplifying the *A. stigma* and *A. io* LW opsin cDNA sequences. We also thank Feng Zhang, Nanfang Yu, Richard Childers, Sorcha Ashe, Cheng-Chia Tsai, Wei-Ping Chan, Avalon Owens and James Crall for lively debates about insect vision, 'wings and wavelengths' and other topics discussed in this chapter.

## References

1. Feuda R, Hamilton S, McInerney J, Pisani D. 2012 Metazoan opsin evolution reveals a simple route to animal vision. *Proc. Natl Acad. Sci. USA* **109**, 18 868–18 872. (doi:10.1073/pnas.1204609109)
2. Devine E, Oprian D, Theobald D. 2013 Relocating the active-site lysine in rhodopsin and implications for evolution of retinylidene proteins. *Proc. Natl Acad. Sci. USA* **110**, 13 351–13 355. (doi:10.1073/pnas.1306826110)
3. Cronin T, Johnsen S, Marshall J, Warrant E. 2014 *Visual ecology*. Princeton, NJ: Princeton University Press.
4. Varma N, Mutt E, Mühle J, Panneels V, Terakita A, Deupi X, Nogly P, Schertler G, Lesca E. 2019 Crystal structure of jumping spider rhodopsin-1 as a light sensitive GPCR. *Proc. Natl Acad. Sci. USA* **116**, 14 547–14 556. (doi:10.1073/pnas.1902192116)
5. Terakita A, Yamashita T, Shichida Y. 2000 Highly conserved glutamic acid in the extracellular IV–V loop in rhodopsins acts as the counterion in retinochrome, a member of the rhodopsin family. *Proc. Natl Acad. Sci. USA* **97**, 14 263–14 267. (doi:10.1073/pnas.260349597)
6. Feuda R, Marlétaz F, Bentley MA, Holland PWH. 2016 Conservation, duplication, and divergence of five opsin genes in insect evolution. *Gen. Biol. Evol.* **8**, 579–587. (doi:10.1093/gbe/evw015)
7. Terakita A, Kawano-Yamashita E, Koyanagi M. 2012 Evolution and diversity of opsins. *Wiley Interdiscip. Rev. Membr. Transport Signal.* **1**, 104–111. (doi:10.1002/wmts.6)
8. Smedley G, McElroy K, Serb J. 2021 Additive and epistatic effects influence spectral tuning in molluscan retinochrome opsin. *bioRxiv* 2021.05.26.445805. (doi:10.1101/2021.05.26.445805)
9. Terakita A. 2005 The opsins. *Protein Family Rev.* **6**, 213.
10. Leung N, Montell C. 2017 Unconventional roles of opsins. *Annu. Rev. Cell Dev. Biol.* **33**, 241–264. (doi:10.1146/annurev-cellbio-100616-060432)
11. Shen W, Kwon Y, Adegbola A, Luo J, Chess A, Montell C. 2011 Function of rhodopsin in temperature discrimination in *Drosophila*. *Science* **331**, 1333–1336. (doi:10.1126/science.1198904)

12. Pisani D, Rota-Stabelli O, Feuda R. 2020 Sensory neuroscience: a taste for light and the origin of animal vision. *Curr. Biol.* **30**, R773–R775. (doi:10.1016/j.cub.2020.05.009)
13. Matsui S, Seidou M, Uchiyama I, Sekiya N, Hiraki K, Yoshihara K, Kito Y. 1988 4-Hydroxyretinal, a new visual pigment chromophore found in the bioluminescent squid, *Watasenia scintillans*. *Biochim. Biophys. Acta* **966**, 370–374. (doi:10.1016/0304-4165(88)90087-6)
14. Kelber A. 2019 Bird colour vision - from cones to perception. *Curr. Opin. Behav. Sci.* **30**, 34–40. (doi:10.1016/j.cobeha.2019.05.003)
15. Jacobs GH. 2009 Evolution of colour vision in mammals. *Phil. Trans. R. Soc. B* **364**, 2957–2967. (doi:10.1098/rstb.2009.0039)
16. Hauser F, Chang B. 2017 Insights into visual pigment adaptation and diversity from model ecological and evolutionary systems. *Curr. Opin. Genet. Dev.* **47**, 110–120. (doi:10.1016/j.gde.2017.09.005)
17. van der Kooi C, Stavenga D, Arikawa K, Belusic G, Kelber A. 2021 Evolution of insect colour vision – from spectral sensitivity to visual ecology. *Annu. Rev. Entomol.* **66**, 435–461. (doi:10.1146/annurev-ento-061720-071644)
18. Briscoe AD, Chittka L. 2001 The evolution of color vision in insects. *Annu. Rev. Entomol.* **46**, 471–510. (doi:10.1146/annurev.ento.46.1.471)
19. Terakita A, Nagata T. 2014 Functional properties of opsins and their contribution to light sensing physiology. *Zool. Sci.* **31**, 653–659. (doi:10.2108/zs140094)
20. Saito T, Koyanagi M, Sugihara T, Nagata T, Arikawa K, Terakita A. 2019 Spectral tuning mediated by helix III in butterfly long wavelength-sensitive visual opsins revealed by heterologous action spectroscopy. *Zool. Lett.* **5**, 35. (doi:10.1186/s40851-019-0150-2)
21. Salcedo E, Farrell D, Zheng L, Phistry M, Bagg E, Britt S. 2009 The green-absorbing *Drosophila* rh6 visual pigment contains a blue-shifting amino acid substitution that is conserved in vertebrates. *J. Biol. Chem.* **284**, 5717–5722. (doi:10.1074/jbc.M807368200)
22. Yokoyama S, Xing J, Liu Y, Faggionato D, Altun A, Stamer WT. 2014 Epistatic adaptive evolution of human color vision. *PLoS Genet.* **10**, e1004884. (doi:10.1371/journal.pgen.1004884)
23. Yokoyama S, Tada T, Liu Y, Faggionato D, Altun A. 2016 A simple method for studying the molecular mechanisms of ultraviolet and violet reception in vertebrates. *BMC Evol. Biol.* **16**, 64. (doi:10.1186/s12862-016-0637-9)
24. Takahashi Y, Yokoyama S. 2005 Genetic basis of spectral tuning in the violet-sensitive visual pigment of African clawed frog, *Xenopus laevis*. *Genetics* **171**, 1153–1160. (doi:10.1534/genetics.105.045849)
25. Hauser FE, Iv H, Chang BSW. 2014 Spectral tuning in vertebrate short wavelength-sensitive 1 (SWS1) visual pigments: can wavelength sensitivity be inferred from sequence data? *J. Exp. Zool.* **322**, 529–539. (doi:10.1002/jez.b.22576)
26. Terakita A, Tsukamoto H, Koyanagi M, Sugahara M, Yamashita T, Shichida Y. 2008 Expression and comparative characterization of Gq-coupled invertebrate visual pigments and melanopsin. *J. Neurochem.* **105**, 883–890. (doi:10.1111/j.1471-4159.2007.05184.x)
27. Wakakuwa M, Terakita A, Koyanagi M, Stavenga DG, Shichida Y, Arikawa K. 2010 Evolution and mechanism of spectral tuning of blue-absorbing visual pigments in butterflies. *PLoS ONE* **5**, e15015. (doi:10.1371/journal.pone.0015015)
28. Sugihara T, Takashi N, Mason B, Koyanagi M, Terakita A. 2016 Absorption characteristics of vertebrate non-visual opsin, Opn3. *PLoS ONE* **17**, e0161215. (doi:10.1371/journal.pone.0161215)
29. Lienard M *et al.* 2021 The evolution of red colour vision is linked to coordinated rhodopsin tuning in lycaenid butterflies. *Proc. Natl Acad. Sci. USA* **118**, e2008986118. (doi:10.1073/pnas.2008986118)
30. Frentiu F, Yuan F, Savage W, Bernard GD, Mullen SP, Briscoe AD. 2014 Clines in butterflies suggest novel roles for insect photopigments. *Mol. Biol. Evol.* **32**, 368–379. (doi:10.1093/molbev/msu304)
31. Thomas G *et al.* 2020 Gene content evolution in the arthropods. *Genome Biol.* **21**, 15. (doi:10.1186/s13059-019-1925-7)
32. Misof B *et al.* 2014 Phylogenomics resolves the timing and pattern of insect evolution. *Science* **7**, 763–767. (doi:10.1126/science.1257570)
33. Kjer K, Misof B, Zhou X. 2012 The 1KITE insect phylogenetics initiative. See <https://1kite.cngb.org>.
34. Li F, Zhao X, Li M, He K, Huang C, Zhou Y, Li Z, Walters J. 2019 Insect genomes: progress and challenges. *Insect. Mol. Biol.* **28**, 739–758. (doi:10.1111/imb.12599)
35. Yin C *et al.* 2016 InsectBase: a resource for insect genomes and transcriptomes. *Nucleic Acids Res.* **44**, D801–D807. (doi:10.1093/nar/gkv1204)
36. Consortium IK. 2013 The i5 K initiative: advancing arthropod genomics for knowledge, human health, agriculture, and the environment. *J. Heredity* **104**, 595–600. (doi:10.1093/jhered/est050)
37. Levine R. 2011 i5 k: the 5,000 insect genome project. *Am. Entomol.* **57**, 110–113. (doi:10.1093/ae/57.2.110)
38. Lewin H *et al.* 2018 Earth BioGenome project: sequencing life for the future of life. *Proc. Natl Acad. Sci. USA* **115**, 4325–4333. (doi:10.1073/pnas.1720115115)
39. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990 Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410. (doi:10.1016/S0022-2836(05)80360-2)
40. Eddy S. 1998 Profile hidden Markov models. *Bioinformatics* **14**, 755–763. (doi:10.1093/bioinformatics/14.9.755)
41. Wu T, Watanabe C. 2005 GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21**, 1859–1875. (doi:10.1093/bioinformatics/bti310)
42. Bhagwat M, Young L, Robison R. 2012 Using BLAT to find sequence similarity in closely related genomes. *Curr. Protoc. Bioinform. Ch.* **10**, 10–8. (doi:10.1002/0471250953.bi1008s37)
43. States DJ, Gish W. 1994 Combined use of sequence similarity and codon bias for coding region identification. *J. Comput. Biol.* **1**, 39–50. (doi:10.1089/cmb.1994.1.39)
44. Finn R, Clements J, Eddy S. 2011 HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* **39**, W29–W37. (doi:10.1093/nar/gkr367)
45. Eddy S. 2011 Accelerated profile HMM searches. *PLoS Comput. Biol.* **7**, e1002195. (doi:10.1371/journal.pcbi.1002195)
46. Stanke M, Kollmar M, Keller O, Waack S. 2011 A novel hybrid gene prediction method employing protein multiple sequence alignments. *Bioinformatics* **27**, 757–763. (doi:10.1093/bioinformatics/btr010)
47. Hoff K, Lomsadze A, Borodovsky M, Stanke M. 2019 Whole-genome annotation with BRAKER. *Methods Mol. Biol.* **1962**, 65–95. (doi:10.1007/978-1-4939-9173-0\_5)
48. Cantarel B, Korf I, Robb S, Parra G, Ross E, Moore B, Holt C, Sánchez Alvarado A, Yandell M. 2008 MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* **18**, 188–196. (doi:10.1101/gr.6743907)
49. Speiser DI *et al.* 2014 Using phylogenetically-informed annotation (PIA) to search for light-interacting genes in transcriptomes from non-model organisms. *BMC Bioinf.* **15**, 350. (doi:10.1186/s12859-014-0350-x)
50. Emms D, Kelly S. 2019 OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238. (doi:10.1186/s13059-019-1832-y)
51. Proost S, Fostier J, De Witte D, Dhoedt B, Demeester P, Van de Peer Y, Vandepoel K. 2012 i-ADHoRe 3.0—fast and sensitive detection of genomic homology in extremely large data sets. *Nucleic Acids Res.* **40**, e11. (doi:10.1093/nar/gkr955)
52. Wang Y *et al.* 2012 MScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**, e49. (doi:10.1093/nar/gkr1293)
53. Sonnhammer E, Östlund G. 2015 InParanoid 8: orthology analysis between 273 proteomes, mostly eukaryotic. *Nucleic Acids Res.* **43**, D234–D239. (doi:10.1093/nar/gku1203)
54. Katoh K, Standley DM. 2013 MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780. (doi:10.1093/molbev/mst010)
55. Edgar RC. 2004 MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797. (doi:10.1093/nar/gkh340)
56. Sievers F, Higgins D. 2018 Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci.* **27**, 135–145. (doi:10.1002/pro.3290)
57. Li H, Durbin R. 2009 Fast and accurate short read alignment with Burrows–Wheeler transform.

- Bioinformatics* **25**, 1754–1760. (doi:10.1093/bioinformatics/btp324)
58. Langmead B, Trapnell C, Pop M, Salzberg S. 2009 Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25. (doi:10.1186/gb-2009-10-3-r25)
  59. Dobin A, Davis C, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras T. 2013 STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21. (doi:10.1093/bioinformatics/bts635)
  60. DePristo M *et al.* 2011 A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498. (doi:10.1038/ng.806)
  61. Li H. 2011 A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993. (doi:10.1093/bioinformatics/btr509)
  62. Garrison E, Marth G. 2012 Haplotype-based variant detection from short-read sequencing. *arXiv*, 1207.3907 [q-bio].
  63. Minh B, Schmidt H, Chernomor O, Schrempf D, Woodhams M, von Haeseler A, Lanfear R. 2020 IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534. (doi:10.1093/molbev/msaa015)
  64. Kozlov A, Darriba D, Flouri T, Morel B, Stamatakis A. 2019 RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453–4455. (doi:10.1093/bioinformatics/btz305)
  65. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010 New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321. (doi:10.1093/sysbio/syq010)
  66. Cai W, Pei J, Grishin N. 2004 Reconstruction of ancestral protein sequences and its applications. *BMC Evol. Biol.* **4**, 33. (doi:10.1186/1471-2148-4-33)
  67. Hanson-Smith V, Johnson A. 2016 PhyloBot: a web portal for automated phylogenetics, ancestral sequence reconstruction, and exploration of mutational trajectories. *PLoS Comput. Biol.* **12**, e1004976. (doi:10.1371/journal.pcbi.1004976)
  68. Arenas M, Bastolla U. 2020 ProtASR2: ancestral reconstruction of protein sequences accounting for folding stability. *Methods Ecol. Evol.* **11**, 248–257. (doi:10.1111/2041-210X.13341)
  69. Musil M, Khan R, Beier A, Stourac J, Konegger H, Damborsky J, Bednar D. 2021 FireProtASR: a web server for fully automated ancestral sequence reconstruction. *Brief. Bioinform.* **22**, bbaa337. (doi:10.1093/bib/bbaa337)
  70. Ashkenazy H, Penn O, Doron-Faigenboim A, Cohen O, Cannarozzi G, Zomer O, Pupko T. 2012 FastML: a web server for probabilistic reconstruction of ancestral sequences. *Nucleic Acids Res.* **40**, W580–W584. (doi:10.1093/nar/gks498)
  71. Yang Z. 2007 PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591. (doi:10.1093/molbev/msm088)
  72. Excoffier L, Laval G, Schneider S. 2005 Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol. Bioinform. Online* **1**, 117693430500100000. (doi:10.1177/117693430500100003)
  73. Beaumont M, Nichols R. 1996 Evaluating loci for use in the genetic analysis of population structure. *Proc. R. Soc. Lond. B* **263**, 1619–1626. (doi:10.1098/rspb.1996.0237)
  74. Foll M, Gaggiotti O. 2008 A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* **180**, 977–993. (doi:10.1534/genetics.108.092221)
  75. Coop G, Witonsky D, Di Rienzo A, Pritchard J. 2010 Using environmental correlations to identify loci underlying local adaptation. *Genetics* **185**, 1411–1423. (doi:10.1534/genetics.110.114819)
  76. Frichot E, Schoville S, de Villemereuil P, Gaggiotti O, François O. 2015 Detecting adaptive evolution based on association with ecological gradients: orientation matters! *Heredity* **115**, 22–28. (doi:10.1038/hdy.2015.7)
  77. McGaughan A, Morgan K, Sommer R. 2014 Environmental variables explain genetic structure in a beetle-associated nematode. *PLoS ONE* **9**, e87317. (doi:10.1371/journal.pone.0087317)
  78. Pond S, Frost S, Muse S. 2005 HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21**, 676–679. (doi:10.1093/bioinformatics/bti079)
  79. Murga-Moreno J, Coronado-Zamora M, Hervas S, Casillas S, Barbadilla A. 2019 iMKT: the integrative McDonald and Kreitman test. *Nucleic Acids Res.* **47**, W283–W288. (doi:10.1093/nar/gkz372)
  80. Eswar N, Webb B, Marti-Renom M, Madhusudhan M, Eramian D, Shen M, Pieper U, Sali A. 2006 Comparative protein structure modeling using modeller. *Curr. Protoc. Bioinform.* **15**, Unit-5.6. (doi:10.1002/0471250953.bi0506s15)
  81. Waterhouse A *et al.* 2018 SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303. (doi:10.1093/nar/gky427)
  82. Kelley L, Mezulis S, Yates C, Wass M, Sternberg M. 2015 The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858. (doi:10.1038/nprot.2015.053)
  83. Jumper J *et al.* 2021 Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589. (doi:10.1038/s41586-021-03819-2)
  84. Schrödinger. 2016 The PyMOL molecular graphics system, version 1.8. Schrödinger, LLC. See <https://pymol.org/2/>.
  85. Pettersen E, Goddard T, Huang C, Couch G, Greenblatt D, Meng E, Ferrin T. 2004 UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612. (doi:10.1002/jcc.20084)
  86. Pérez-Moreno J, DeLeo D, Palero F, Bracken-Grissom H. 2018 Phylogenetic annotation and genomic architecture of opsin genes in Crustacea. *Hydrobiologia* **825**, 159–175. (doi:10.1007/s10750-018-3678-9)
  87. Faggionato D, Serb JM. 2017 Strategy to identify and test putative light sensitive non-opsin G-protein-coupled receptors: a case study. *Biol. Bull.* **233**, 70–82. (doi:10.1086/694842)
  88. Gerrard E *et al.* 2018 Convergent evolution of tertiary structure in rhodopsin visual proteins from vertebrates and box jellyfish. *Proc. Natl Acad. Sci. USA* **115**, 6201–6206. (doi:10.1073/pnas.1721333115)
  89. Chang B, Jönsson K, Kazmi M, Donoghue M, Sakmar T. 2002 Recreating a functional ancestral archosaur visual pigment. *Mol. Biol. Evol.* **19**, 1483–1489. (doi:10.1093/oxfordjournals.molbev.a004211)
  90. O’Quin K, Hofmann C, Hofmann H, Carleton K. 2010 Parallel evolution of opsin gene expression in African cichlid fishes. *Mol. Biol. Evol.* **27**, 2839–2854. (doi:10.1093/molbev/msq171)
  91. Shi Y, Radlwimmer F, Yokoyama S. 2001 Molecular genetics and the evolution of ultraviolet vision in vertebrates. *Proc. Natl Acad. Sci. USA* **98**, 11 731–11 736. (doi:10.1073/pnas.201257398)
  92. Yokoyama S. 2012 Synthesis of experimental molecular biology and evolutionary biology: an example from the world of vision. *BioScience* **62**, 939–948. (doi:10.1525/bio.2012.62.11.3)
  93. Lord NP, Plimpton RL, Sharkey CR, Suvorov A, Lelito JP, Willardson BM, Bybee SM. 2016 A cure for the blues: opsin duplication and subfunctionalization for short-wavelength sensitivity in jewel beetles (Coleoptera: Buprestidae). *BMC Evol. Biol.* **16**, 107. (doi:10.1186/s12862-016-0674-4)
  94. Sharkey CR, Fujimoto MS, Lord NP, Shin S, McKenna DD, Suvorov A, Martin GJ, Bybee SM. 2017 Overcoming the loss of blue sensitivity through opsin duplication in the largest animal group, beetles. *Sci. Rep.* **7**, 8. (doi:10.1038/s41598-017-00061-7)
  95. Frentiu F, Bernard G, Sison-Mangus M, Van Zandt Brower A, Briscoe A. 2007 Gene duplication is an evolutionary mechanism for expanding spectral diversity in the long-wavelength photopigments of butterflies. *Mol. Biol. Evol.* **24**, 2016–2028. (doi:10.1093/molbev/msm132)
  96. Briscoe AD, Bybee SM, Bernard GD, Yuan F, Sison-Mangus MP, Reed RD, Warren AD, Llorente-Bousquets J, Chiao C-C. 2010 Positive selection of a duplicated UV-sensitive visual pigment coincides with wing pigment evolution in *Heliconius* butterflies. *Proc. Natl Acad. Sci. USA* **107**, 3628–3633. (doi:10.1073/pnas.0910085107)
  97. Mugal C, Wolf J, Kaji I. 2014 Why time matters: codon evolution and the temporal dynamics of dN/dS. *Mol. Biol. Evol.* **31**, 212–231. (doi:10.1093/molbev/mst192)
  98. Messer P, Petrov D. 2013 Frequent adaptation and the McDonald–Kreitman test. *Proc. Natl Acad. Sci. USA* **110**, 8615–8620. (doi:10.1073/pnas.1220835110)
  99. Yang Z. 1998 Likelihood ratio tests for detecting positive selection and application to primate

- lysozyme evolution. *Mol. Biol. Evol.* **15**, 568–573. (doi:10.1093/oxfordjournals.molbev.a025957)
100. Sondhi Y, Ellis E, Bybee S, Theobald J, Kawahara A. 2021 Light environment drives evolution of color vision genes in butterflies and moths. *Commun. Biol.* **4**, 177. (doi:10.1038/s42003-021-01688-z)
101. Yuan F, Bernard G, Le J, Briscoe A. 2010 Contrasting modes of evolution of the visual pigments in *Heliconius* butterflies. *Mol. Biol. Evol.* **27**, 2392–2405. (doi:10.1093/molbev/msq124)
102. Porter ML, Awata H, Bok MJ, Cronin TW. 2020 Exceptional diversity of opsin expression patterns in *Neogonodactylus oerstedii* (Stomatopoda) retinas. *Proc. Natl Acad. Sci. USA* **117**, 8948. (doi:10.1073/pnas.1917303117)
103. Feuda R *et al.* 2021 Phylogenomics of opsin genes in diptera reveals lineage-specific events and contrasting evolutionary dynamics in *Anopheles* and *Drosophila*. *Genome Biol. Evol.* **13**, evab170. (doi:10.1093/gbe/evab170)
104. Schulte J, O'Brien C, Conte M, O'Quin K, Carleton K. 2014 Interspecific variation in Rx1 expression controls opsin expression and causes visual system diversity in African cichlid fishes. *Mol. Biol. Evol.* **31**, 2297–2308. (doi:10.1093/molbev/msu172)
105. Smith A, Carleton K. 2010 Allelic variation in Malawi cichlid opsins: a tale of two genera. *J. Mol. Evol.* **70**, 593–604. (doi:10.1007/s00239-010-9355-x)
106. Ahrens C, Rymer P, Stow A, Bragg J, Dillon S, Umbers K, Dudaniac R. 2018 The search for loci under selection: trends, biases and progress. *Mol. Ecol.* **27**, 1342–1356. (doi:10.1111/mec.14549)
107. Marques DA, Taylor JS, Jones FC, Di Palma F, Kingsley DM, Reimchen TE. 2017 Convergent evolution of SWS2 opsin facilitates adaptive radiation of threespine stickleback into different light environments. *PLoS Biol.* **15**, e2001627. (doi:10.1371/journal.pbio.2001627)
108. Hiwataishi T *et al.* 2010 An explicit signature of balancing selection for color-vision variation in New World Monkeys. *Mol. Biol. Evol.* **27**, 453–464. (doi:10.1093/molbev/msp262)
109. Tezuka A *et al.* 2014 Divergent selection for opsin gene variation in guppy (*Poecilia reticulata*) populations of Trinidad and Tobago. *Heredity* **113**, 381–389. (doi:10.1038/hdy.2014.35)
110. Terai Y *et al.* 2006 Divergent selection on opsins drives incipient speciation in Lake Victoria cichlids. *PLoS Biol.* **4**, e433. (doi:10.1371/journal.pbio.0040433)
111. Lehtonen P *et al.* 2012 Candidate genes for colour and vision exhibit signals of selection across the pied flycatcher (*Ficedula hypoleuca*) breeding range. *Heredity* **108**, 431–440. (doi:10.1038/hdy.2011.93)
112. Lower SE, Stanger-Hall KF, Hall DW. 2018 Molecular variation across populations of a widespread North American firefly, *Photinus pyralis*, reveals that coding changes do not underlie flash color variation or associated visual sensitivity. *BMC Evol. Biol.* **18**, 129. (doi:10.1186/s12862-018-1251-9)
113. Leung N, Thakur D, Gurav A, Kim S, Di Pizio A, Niv M, Montell C. 2020 Functions of opsins in *Drosophila* taste. *Curr. Biol.* **30**, 1367–1379.e1366. (doi:10.1016/j.cub.2020.01.068)
114. Briscoe A. 2002 Homology modeling suggests a functional role for parallel amino acid substitutions between bee and butterfly red- and green-sensitive opsins. *Mol. Biol. Evol.* **19**, 983–986. (doi:10.1093/oxfordjournals.molbev.a004158)
115. Haddad Y, Adam V, Heger Z. 2020 Ten quick tips for homology modeling of high-resolution protein 3D structures. *PLoS Comput. Biol.* **16**, e1007449. (doi:10.1371/journal.pcbi.1007449)
116. Sekharan S, Morokuma K. 2011 Why 11-*cis*-Retinal? Why not 7-*cis*, 9-*cis* or 13-*cis*-retinal in the eye? *J. Am. Chem. Soc.* **133**, 19 052–19 055. (doi:10.1021/ja208789h)
117. Jeffrey G. 1997 *An introduction to hydrogen bonding*. New York, NY: Oxford University Press.
118. Yokoyama S, Radlwimmer F. 1998 The 'five-sites' rule and the evolution of red and green color vision in mammals. *Mol. Biol. Evol.* **15**, 560–567. (doi:10.1093/oxfordjournals.molbev.a025956)
119. Patel J, Brown C, Ytreberg F, Stenkamp D. 2018 Predicting peak spectral sensitivities of vertebrate cone visual pigments using atomistic molecular simulations. *PLoS Comput. Biol.* **14**, e1005974. (doi:10.1371/journal.pcbi.1005974)
120. Yokoyama S. 2002 Molecular evolution of color vision in vertebrates. *Gene* **300**, 69–78. (doi:10.1016/s0378-1119(02)00845-4)
121. Knox B, Salcedo E, Mathiesz K, Schaefer J, Chou W-H, Chadwell L, Smith W, Britt S, Barlow R. 2003 Heterologous expression of *Limulus* rhodopsin. *J. Biol. Chem.* **278**, 40 493–40 502. (doi:10.1074/jbc.M304567200)
122. Salcedo E, Huber A, Henrich S, Chadwell L, Chou W, Paulsen R, Britt S. 1999 Blue- and green-absorbing visual pigments of *Drosophila*: ectopic expression and physiological characterization of the R8 photoreceptor cell-specific Rh5 and Rh6 rhodopsins. *J. Neurosci.* **19**, 10 716–10 726. (doi:10.1523/JNEUROSCI.19-24-10716.1999)
123. Salcedo E, Zheng L, Phistry M, Bagg E, Britt S. 2003 Molecular basis for ultraviolet vision in invertebrates. *J. Neurosci.* **23**, 10 873–10 878. (doi:10.1523/JNEUROSCI.23-34-10873.2003)
124. Townson S, Chang B, Salcedo E, Chadwell L, Pierce N, Britt S. 1998 Honeybee blue- and ultraviolet-sensitive opsins: cloning, heterologous expression in *Drosophila*, and physiological characterization. *J. Neurosci.* **18**, 2412–2422. (doi:10.1523/JNEUROSCI.18-07-02412.1998)
125. Sharkey C, Blanco J, Leibowitz M, Pinto-Benito D, Wardill T. 2020 The spectral sensitivity of *Drosophila* photoreceptors. *Sci. Rep.* **10**, 18242. (doi:10.1038/s41598-020-74742-1)
126. Szymczak-Workman AL, Vignali KM, Vignali DAA. 2012 Generation of 2A-linked multicistronic cassettes by recombinant PCR. *Cold Spring Harbor Protocols* **2012**, pdb.prot067884. (doi:10.1101/pdb.prot067884)
127. Telles FJ, Lind O, Henze MJ, Rodríguez-Gironés MA, Goyret J, Kelber A. 2014 Out of the blue: the spectral sensitivity of hummingbird hawkmoths. *J. Comp. Physiol. A* **200**, 537–546. (doi:10.1007/s00359-014-0888-0)
128. Arikawa K. 2017 The eyes and vision of butterflies. *J. Physiol.* **595**, 5457–5464. (doi:10.1113/JP273917)
129. Cong Q, Borek D, Otwinowski Z, Grishin NV. 2015 Skipper genome sheds light on unique phenotypic traits and phylogeny. *BMC Genomics* **16**, 639. (doi:10.1186/s12864-015-1846-0)
130. 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. USA* **112**, E1247. (doi:10.1073/pnas.1424670112)
131. Porter M, Bok M, Robinson P, Cronin T. 2009 Molecular diversity of visual pigments in Stomatopoda (Crustacea). *Vis. Neurosci.* **26**, 255–265. (doi:10.1017/S0952523809090129)
132. Cronin TW, Marshall NJ, Caldwell RL. 2000 Spectral tuning and the visual ecology of mantis shrimps. *Phil. Trans. R. Soc. Lond. B* **355**, 1263–1267. (doi:10.1098/rstb.2000.0680)
133. Ni JD, Baik LS, Holmes TC, Montell C. 2017 A rhodopsin in the brain functions in circadian photoentrainment in *Drosophila*. *Nature* **545**, 340. (doi:10.1038/nature22325)
134. Sakai T, Tsutsui K, Yamashita T, Iwabe N, Takahashi K, Wada A, Shichida Y. 2017 *Drosophila melanogaster* rhodopsin Rh7 is a UV-to-visible light sensor with an extraordinarily broad absorption spectrum. *Sci. Rep.* **7**, 7349. (doi:10.1038/s41598-017-07461-9)
135. Trifinopoulos J, Nguyen L-T, von Haeseler A, Minh B. 2016 W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* **44**, W232–W235. (doi:10.1093/nar/gkw256)
136. Kalyaanamoorthy S, Minh B, Wong T, von Haeseler A, Jermin L. 2017 ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589. (doi:10.1038/nmeth.4285)
137. Hoang D, Chernomor O, von Haeseler A, Minh B, Vinh L. 2018 UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522. (doi:10.1093/molbev/msx281)
138. Zhang H, Gao S, Lercher M, Hu S, Chen W-H. 2012 EvolView, an online tool for visualizing, annotating and managing phylogenetic trees. *Nucleic Acids Res.* **40**, W569–W572. (doi:10.1093/nar/gks576)
139. Govardovskii V, Fyhrquist N, Reuter T, Kuzmin D, Donner K. 2000 In search of the visual pigment template. *Vis. Neurosci.* **17**, 509–528. (doi:10.1017/s0952523800174036)
140. Liénard MA, Valencia-Montoya WA, Pierce NE. 2022 Molecular advances to study the function, evolution, and spectral tuning of arthropod visual opsins. Figshare. (doi:10.6084/m9.figshare.c.6098624)