

SCIENTIFIC REPORTS



OPEN

Overcoming the loss of blue sensitivity through opsin duplication in the largest animal group, beetles

Camilla R. Sharkey¹, M. Stanley Fujimoto², Nathan P. Lord³, Seungwan Shin⁴, Duane D. McKenna⁴, Anton Suvorov¹, Gavin J. Martin¹ & Seth M. Bybee¹

Opsin proteins are fundamental components of animal vision whose structure largely determines the sensitivity of visual pigments to different wavelengths of light. Surprisingly little is known about opsin evolution in beetles, even though they are the most species rich animal group on Earth and exhibit considerable variation in visual system sensitivities. We reveal the patterns of opsin evolution across 62 beetle species and relatives. Our results show that the major insect opsin class (SW) that typically confers sensitivity to “blue” wavelengths was lost ~300 million years ago, before the origin of modern beetles. We propose that UV and LW opsin gene duplications have restored the potential for trichromacy (three separate channels for colour vision) in beetles up to 12 times and more specifically, duplications within the UV opsin class have likely led to the restoration of “blue” sensitivity up to 10 times. This finding reveals unexpected plasticity within the insect visual system and highlights its remarkable ability to evolve and adapt to the available light and visual cues present in the environment.

At the molecular level, the wavelength sensitivity of an animal photoreceptor is determined by the photopigment, comprising an opsin protein bound to a light-absorbing chromophore. Insects commonly possess three opsin proteins (UV, SW and LW) that form photopigments maximally sensitive to ultraviolet (~350 nm), blue (~440 nm) and green (~530 nm) wavelengths, respectively. As insect opsin genes form distinct phylogenetic clades according to their spectral class (UV, SW or LW) the sensitivity ranges of an insect visual system can usually be estimated by the complement of opsin genes present. In some insects photopigment sensitivity has extended outside of this range into the violet (~420 nm) and red (>600 nm) region of the light spectrum through duplications of the SW^{1,2} and LW opsins^{2,3}, respectively.

Gene duplications occur at a proposed rate of 1 per 100 MY⁴ and for the majority of duplications, gene copies are lost within ~2 MY, through the accumulation of deleterious mutations⁴. However, gene copies may be retained if they acquire a novel function through sequence mutation (subfunctionalization)⁴ that leads to increased fitness. In the case of opsins, beneficial mutations in duplicates may lead to changes in photopigment properties, such as spectral sensitivity. Such duplications alongside losses of opsins have shaped the major animal opsin classes we observe today. Duplications of long-wavelength sensitive opsins are widespread across Arthropoda and are numerous in some lineages (e.g., up to 21 LW opsins in aeshnid dragonflies², 25 in the genome of *Daphnia pulex*⁵ and seven in mosquitoes⁶). UV duplications by contrast are rare and lineage specific, with only single duplication events reported in *Heliconius* butterflies⁷, planthoppers (Delphacidae)⁸, and two lineages of beetles^{9,10}.

Molecular evidence suggests that the SW opsin class has been lost from a number of beetle lineages: fireflies (Lampyridae^{11,12}), diving beetles (*Thermonectus marmoratus*)^{9,13}, jewel beetles (Buprestidae)¹⁰ and darkling beetles (e.g., *Tribolium castaneum*)¹⁴. It has therefore been proposed that all beetles may lack the SW opsin class, which typically underpins visual sensitivity to blue wavelengths in insects. However, physiological evidence has revealed that a number of beetles do have blue sensitive photoreceptors: a ladybird (*Coccinella*

¹Department of Biology, Brigham Young University, 4102 LSB, Provo, UT, 84602, USA. ²Computer Science Department, Brigham Young University, Provo, Utah, 84602, USA. ³Department of Biological and Environmental Sciences, Georgia College & State University, Campus Box 081, Milledgeville, GA, 31061, USA. ⁴Department of Biological Sciences, University of Memphis, 3700 Walker Avenue, Memphis, TN, 38152, USA. Correspondence and requests for materials should be addressed to C.R.S. (email: camilla.r.sharkey@gmail.com)

septempunctata)^{15,16}, a leaf beetle (*Leptinotarsa decemlineata*)¹⁷, two ground beetles (*Carabus* spp.)¹⁸ and a jewel beetle (*Agrilus planipennis*)¹⁰.

Two evolutionary scenarios are therefore possible. Firstly, the SW opsin class was lost independently in numerous beetle lineages, or secondly, this opsin class was lost prior to or during the early evolutionary history of beetles and blue light sensitivity was regained multiple times, independently of the ancestral SW opsin. Due to the paucity of studies on opsins from beetles and related groups, the timings of the SW opsin losses remain unclear. To address this, we examined the opsin repertoire of a broad diversity of beetles representing most major lineages and included beetles with a wide range of life histories (e.g., diurnal, nocturnal, predaceous and pollinating beetles). Additionally, exemplars from the closely related orders Strepsiptera (twisted-wing parasites), Raphidioptera (snake flies), Megaloptera (alderflies, fishflies and dobsonflies) and Neuroptera (lacewings, antlions and mantidflies), comprising the other major lineages of the clade Neuropteroidea¹⁹ were also investigated.

Results

Loss of the SW opsin class. For this study, over two billion RNA-seq reads were assembled into more than six million gene transcripts (Table S1). In total, 204 opsins were recovered, of which 73% encode full-length proteins (Tables S2 and S3). In a homology search of 74 coleopteran transcriptomes across 29 families and 19 of 21 superfamilies, the SW opsin was not recovered. Furthermore, this opsin class was not detected in the 12 transcriptome assemblies from the other lineages of Neuropteroidea. While the majority of samples were derived from adult tissue (Table S1), larval tissue was also sequenced for 11 species. The SW opsin class was also not recovered from these samples, indicating that it has been lost in both major life stages. Similarly, both sexes were sequenced for 15 species, indicating that the SW opsin loss is not sex specific. Furthermore, we did not recover the SW opsin class from the genomes of four beetle species (Table S1). In all cases, there was no evidence for a SW opsin pseudogene but the absence of the SW gene in our analyses indicates that it has indeed been lost. We are confident that the phylogenetic breadth and quality of samples (Figure S1) used in this study was sufficient to rule out the possibility that this opsin class was present within Coleoptera but not recovered during our analysis. Bees, wasps and ants (Hymenoptera) are thought to be the closest relative of all remaining holometabolon insects, including the true flies (Diptera), and moths and butterflies (Lepidoptera)²⁰. The SW opsin has been retained in these orders^{21–23}, implying that the loss of the SW opsin occurred in the lineage leading to Neuropteroidea, after this clade diverged from the rest of Holometabola, approximately 300 million years ago^{19,24}.

Opsin duplications. As the insect SW opsin class was lost prior to the radiation of Coleoptera, blue light sensitivity must have secondarily arisen in Coleoptera through an alternative mechanism to that of other insects. Recent evidence suggests that blue light sensitivity in jewel beetles (Buprestidae) has been achieved through duplication and subfunctionalization of the UV and/or LW opsin genes¹⁰. We performed homology searches for the presence of opsin gene duplications on 89 assembled neuropteroid transcriptomes, totalling 72 species. Phylogenetic analyses were performed to reconstruct species and opsin gene trees, which we then used to interpret the evolutionary history of opsin gene duplication events.

We recovered UV duplications from three of the four coleopteran suborders: from the single species sampled from the suborder Myxophaga (*Lepicerus* sp.), three of five species from the suborder Adephaga and 18 of 54 species from the suborder Polyphaga (Fig. 1; Table S2). UV duplications were not recovered in either species from suborder Archostemata (*Priacma serrata* and *Micromalthus debilis*) (Table S2). In three cases, the phylogenetic placement of both UV copies into separate monophyletic clades in the gene tree reconstructed from DNA sequences (Figs 2 and 3) suggests that opsin gene duplications occurred prior to the crown diversification of leaf beetles (Chrysomelidae), ladybirds (Coccinellidae) and jewel beetles (Buprestidae) (Fig. 2). Ancestral states of opsin duplication were reconstructed on the species phylogeny, to infer likely patterns of duplication across Coleoptera. Parsimony and maximum likelihood reconstructions (Figure S3), alongside the phylogenetic placement of opsin duplicates along independent lineages, suggest that many separate opsin gene duplication events occurred within Coleoptera. In total, including the previously described *Thermonectus marmoratus* and Buprestidae duplications, our results are consistent with the existence of 12 independent UV opsin duplications, two of which occurred in one species, the pollen beetle *Brassicoglyphus aeneus* (Nitidulidae) (Fig. 1). UV opsins are absent in three beetle species, *Dastarcus helophoroides*, *Monochamus alternatus* and *Tenebrio molitor*, suggesting low UV opsin gene expression or a lack of UV opsins. Physiological evidence also supports a lack of UV sensitivity in *T. molitor* (Fig. 4)²⁵. In the remaining neuropteroid orders, UV opsin duplications were observed in snakeflies (Raphidioptera) and one in twisted-wing parasites (Strepsiptera) (Figs 1 and 2).

Novel LW opsin gene duplications were discovered in five coleopteran species across five families: the flower weevil *Larinus minutus* (Curculionidae), the whirligig beetle *Gyrinus marinus* (Gyrinidae), the pollen beetle *Brassicoglyphus aeneus* (Nitidulidae), the monkey beetle *Heterochelus* sp. (Scarabaeidae) and the ground beetle *Carabus granulatus* (Carabidae). The DNA sequences of *Carabus granulatus* LW opsins were identical at the protein level, however, as the full length LW2 was not recovered in this species (137 amino acids; Table S3), it remains to be examined how distinct the full-length copies of these two opsins are. One additional buprestid (*Aphanisticus* sp.) was also found to have two LW opsin copies, confirming previous findings from six other buprestid species¹⁰. LW duplications are also present in two of three Strepsiptera species with *Xenos vesparum* possessing five unique LW opsin proteins.

Sites under selection. To explore potential amino acid sites responsible for spectral shifts in beetle UV opsins with putative blue light sensitivity, we tested branches in the opsin gene tree (Fig. 2) and individual amino acid sites for positive selection (PS). This was carried out in the three major clades of beetle UV opsins duplicates that have complementary physiological evidence of blue-sensitive photoreceptors (Buprestidae, Coccinellidae and Chrysomelidae). Site selection analyses reveal only four sites under PS adjacent to the chromophore-binding

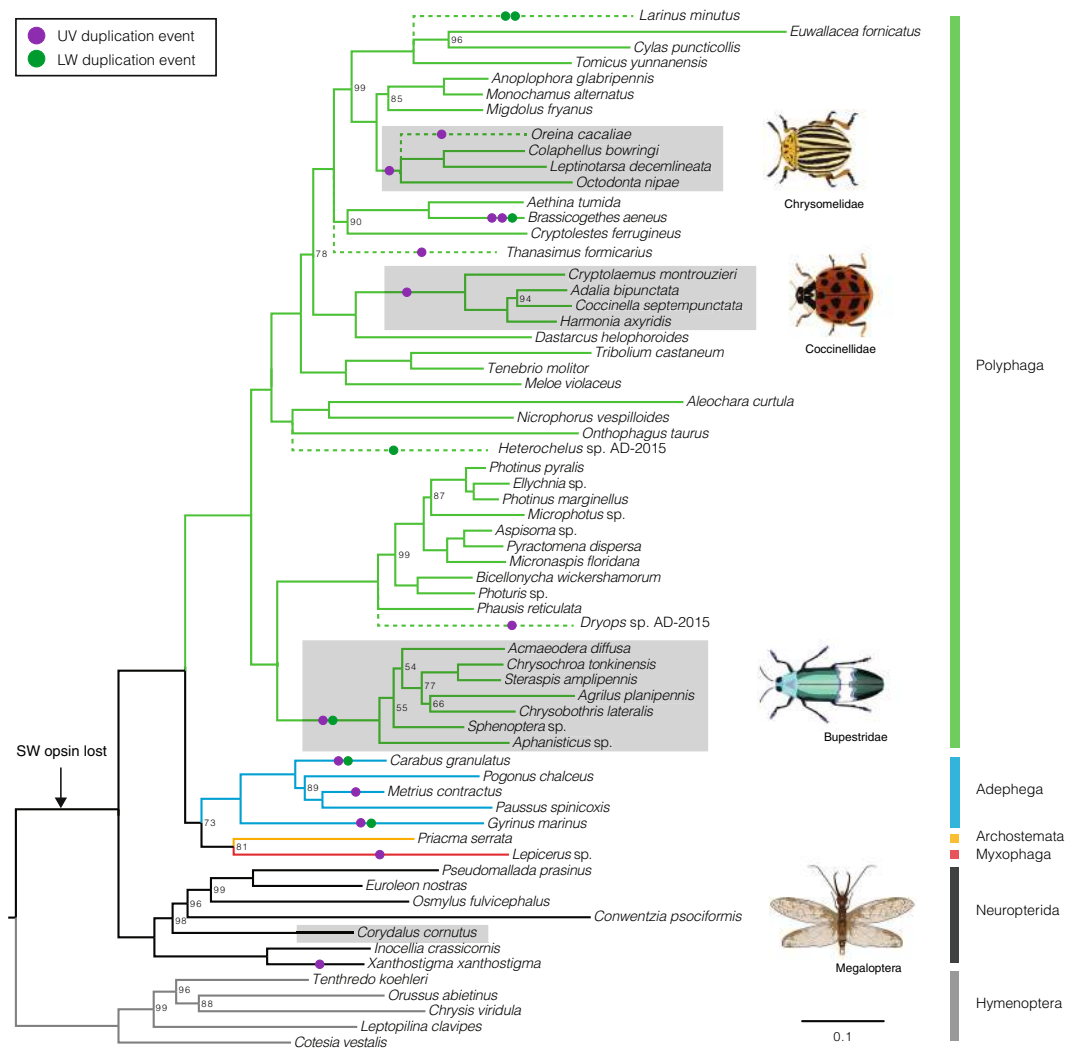


Figure 1. Opsin duplications within Coleoptera. Species phylogeny using 358 gene clusters from translated transcriptome assemblies, constructed using maximum likelihood. Node values indicate UFboot supports and only values <100 are shown. Strepsipterans are not included due to poor support and placement of these species. For the full topology see Figure S2. Three major coleopteran UV opsin clades are highlighted in grey. Additional species with opsin duplications that were not included in the original species phylogeny are included at proposed positions (dashed lines). Solid circles denote putative UV (purple) and LW (green) duplications. The proposed timing of the SW opsin class loss is indicated (black arrow).

pocket, and the recovered positively selected sites were not congruent between lineages (Fig. 3; Table S4). It is therefore difficult to hypothesise about the potential specific shifting mechanisms behind photopigment spectral diversity in these groups. Interestingly, however, all coccinellid, two chrysomelid and one buprestid UV opsin duplicates have hallmarks of insect SW opsins (Tables S4 and S5). This includes a neutral or negatively charged amino acid rather than positively charged lysine (K) at bovine site 90 and phenylalanine (F) replacing tyrosine (Y) at bovine site 113²⁶. Replacing K with a neutral amino acid at site 90 in the *Drosophila* Rh3 UV opsin, shifts sensitivity by 73 nm into the blue wavelengths²⁶. This single amino acid substitution may be sufficient to render a number of beetle UV photopigments found in this study (*Harmonia axyridis* UV2, *Leptinotarsa decemlineata* UV2, *Octodonta nipae*, UV2 *Oreina cacaliae* UV2, *Steraspis amplipennis* UV1) functionally blue sensitive. Little is known about the spectral shifting mechanisms across Coleoptera^{10,12,27} but our findings, coupled with physiological evidence for photoreceptor diversity, highlight beetles as a highly attractive group for future studies.

Discussion

Opsin losses have occurred in other animals, including the American cockroach (*Periplaneta americana*)²⁸, deep sea fish²⁹, fossorial snakes³⁰, caecilians³¹ and both nocturnal and aquatic mammals^{32–34}. Such losses are typically associated with low-light or spectrally-attenuated environments. It has been proposed that the presumptive loss of tri- or di-chromatic colour vision under these conditions has little impact on fitness³⁵ and due to the high cost of maintaining retinal tissue³⁶, selection favours a reduction in visual system complexity (e.g., during the evolution of mammals)³⁵. Furthermore, under these scotopic conditions, the saliency of wavelength specific cues

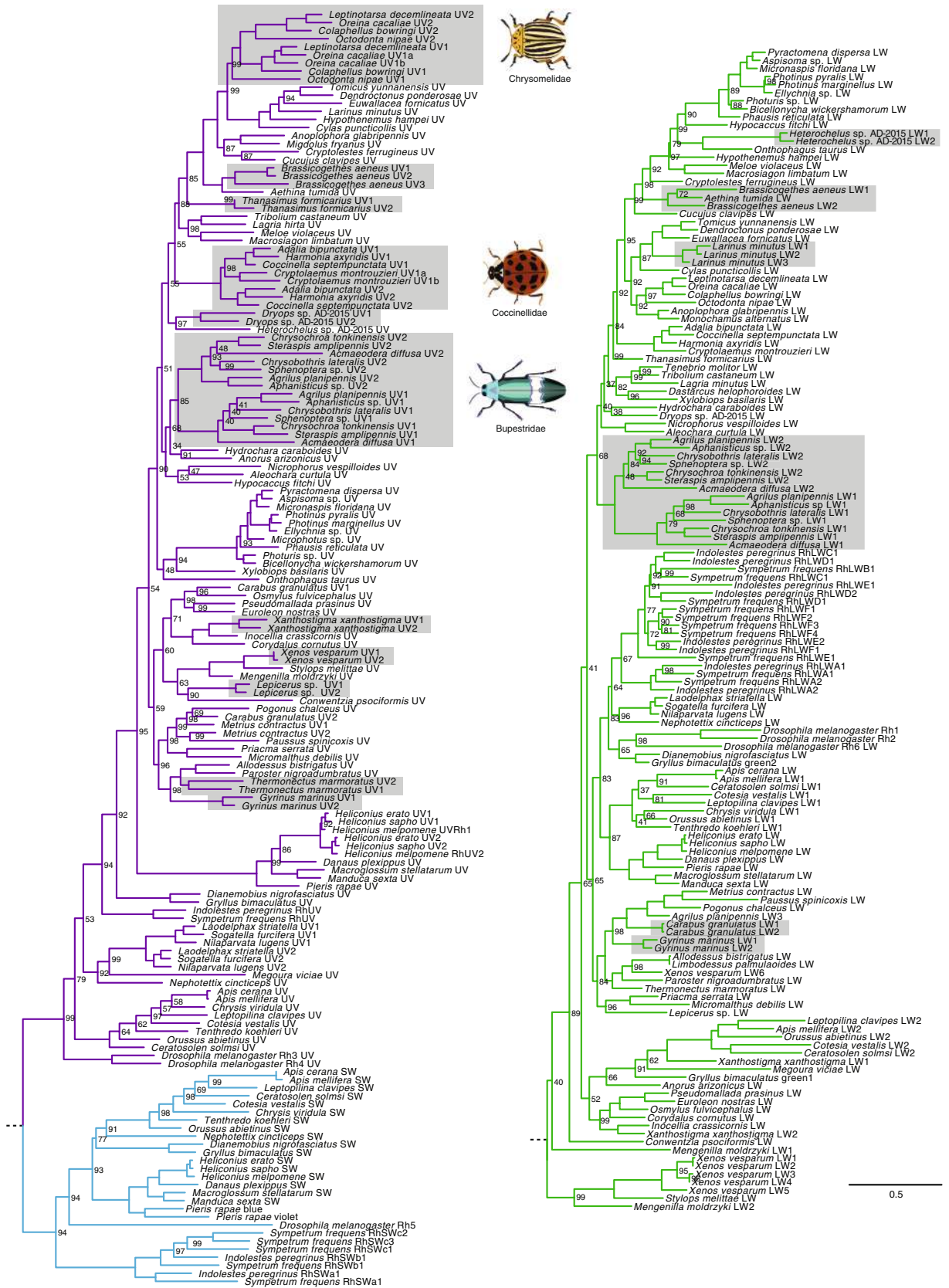


Figure 2. Beetle opsin gene phylogeny. Maximum likelihood DNA phylogeny of all visual opsin genes from this study and other insect UV (purple), SW (blue) and LW (green) opsins rooted to cephalopod opsins (not shown). Node values (UFBoot support values) are based on 10,000 replicates. Values of 100 are not shown. Duplicated neuropteroid opsins are highlighted in grey.

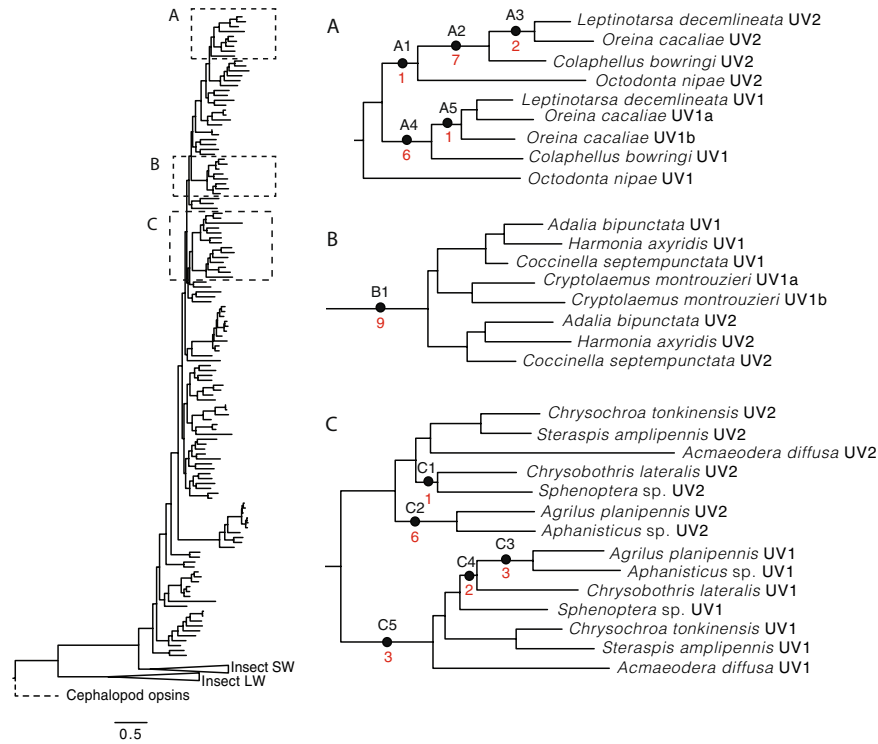


Figure 3. Signals of selection within beetle UV duplication clades. Branches (closed circles) and numbers of sites (red values) under positive selection in the three major coleopteran UV opsin clades. Positions of these clades, (A) (Chrysomelidae), (B) (Coccinellidae) and (C) (Buprestidae), on the full opsin DNA phylogeny are indicated (dashed boxes). For further positive selection statistics and amino acid sites see Tables S4 and S5.

Related taxa in this study	Species with physiological data	Spectral sensitivities					Opsins				
		UV	Blue	Blue	Green	Red	UV	UV	UV	LW	LW
<i>Tenebrio molitor</i>	<i>Tenebrio molitor</i>				Green						
<i>Dendroctonus ponderosae</i>	<i>Dendroctonus pseudotsugae</i>		Blue*		Green						
<i>Onthophagus taurus</i>	<i>Lethrus apterus</i>	UV			Green						
Lampyridae	Lampyridae	UV			Green						
<i>Onthophagus taurus</i>	<i>Onitis alexis</i>	UV			Green						
<i>Coccinella septempunctata</i>	<i>Coccinella septempunctata</i>	UV	Blue		Green						
<i>Leptinotarsa decemlineata</i>	<i>Leptinotarsa decemlineata</i>	UV	Blue		Green						
<i>Carabus granulatus</i>	<i>Carabus</i> spp.	UV	Blue		Green	Red					
<i>Agrilus planipennis</i>	<i>Agrilus planipennis</i>	UV	Blue		Green	Red					
<i>Brassicogethes aeneus</i>	<i>Brassicogethes aeneus</i>	UV	Blue		Green						

Figure 4. Comparison of spectral sensitivities and opsins found in this study for matching or similar species. *Measurements from *Dendroctonus pseudotsugae* were made in the absence of UV stimuli. The “blue-sensitive” photoreceptors discovered in this species most likely peak in the UV.

is diminished³⁷ and a system using two rather than three photoreceptor classes may be advantageous by increasing overall sensitivity to light³⁸. Similar to what has been observed in other animal visual systems with reduced opsin diversity, it seems probable that the ancestor of Neuropteroidea was nocturnal or functioned at low light levels. Support for this hypothesis is also given by the preponderance of nocturnality (Neuroptera, Megaloptera, Raphidioptera) and evidence for visual function associated with nocturnality (Strepsiptera) in the other orders of Neuropteroidea^{39,40}.

Opsin gene duplication and divergence is the primary mechanism by which novel photopigment sensitivities arise⁴¹. Throughout animal evolution, many instances of opsin gene duplications and losses have occurred, shaping the diverse repertoire of animal visual system sensitivities we observe today^{41,42}. A well-known example is the duplication of the LWS opsin gene in Old World monkeys that has led to increased discrimination between long wavelengths and therefore an increased ability to detect ripe fruits against a foliage background^{43,44}. The

distribution of opsin duplications across Coleoptera recovered in our analyses clearly show that opsin duplicates are not derived from an opsin-diverse common ancestor but, instead, appeared secondarily and independently in numerous lineages of beetles. The prevalence of highly divergent UV opsin duplicates across Coleoptera (Table S3) and the rarity of such duplicates in other animal visual systems suggests that there may be a greater selective advantage for additional UV opsins in beetles than other animals. It is highly unlikely that duplications reflect adult and larval opsin copies, as no difference in opsin copy number was found between the available adult and larval samples and the majority of samples used were derived from adult tissue (Table S1).

Spectral sensitivity data was available for six species included in this study, from two closely related *Carabus* species and one *Dendroctonus* species (Fig. 4)^{16–18,25,27,45–49}. Comparing the number of opsin copies with photoreceptor sensitivities reveals that all instances of blue sensitivity co-occur with UV duplication (*Coccinella septempunctata*, *Leptinotarsa decemlineata*, *Brassicogethes aeneus*, *Agrilus planipennis* and *Carabus* spp.; Fig. 4). We propose that coleopteran UV duplications and subsequent amino acid changes serve to extend the sensitivity of the beetle visual system into the short or blue wavelengths, overcoming the ancestral loss of sensitivity to this region of the light spectrum and its presumed limitations on spectral sensitivity and discrimination in diurnal and other comparatively high-light environments.

We found that beetle species with additional photoreceptor sensitivities had extra opsin copies (Fig. 4) suggesting duplication and subfunctionalization as a route for the evolution of novel photopigment sensitivities. If the evolution of beetle opsin duplicates has been shaped by visual ecology, one might expect to find a clear link between opsin diversity and the use of visual cues or visual environment. Indeed, duplications are widespread amongst known diurnal species and those with eye morphologies that are indicative of activity in high light environments (e.g., large compound eyes, high density of corneal pigments, apposition-type eye structure)⁵⁰. Opsin duplications were predominantly found in species with behaviours often guided by visual cues, and in many of these species (see references), vision has been shown to be the primary cue for such behaviours: flower visitation (*Larinus minutus*, *Brassicogethes aeneus*⁵¹ and *Heterocheilus* sp.^{52,53}), predation (coccinellids⁵⁴, gyrids^{55,56}, *Thanasimus formicarius*, *Carabus granulatus*, *Metrius contractus*), host plant detection (chrysomelids^{57,58}, coccinellids⁵⁹), and mate recognition (buprestids⁶⁰). Duplications were notably absent in nocturnal species, with the exception of the nocturnal active predator, *Carabus granulatus*, which has been shown to possess a number of spectrally distinct photoreceptors¹⁸.

As is the case in butterflies and moths (Lepidoptera)⁶¹ and dragon/damselflies (Odonata)^{2,62}, beetle LW opsin duplications may serve to extend sensitivity into the longer wavelengths. However, further physiological measurements from beetle species with LW duplications are necessary to confirm this. Interestingly, whilst LW duplications are less common than UV duplications within Coleoptera, they are ubiquitous amongst flower-visiting lineages: *Brassicogethes aeneus* (Nitidulidae: 3 UV and 2 LW), *Heterocheilus* sp. (Scarabaeidae: 1 UV and 2 LW) and the weevil, *Larinus minutus* (Curculionidae: 1 UV and 3 LW). This points towards a potential role for extended long wavelength sensitivity to locate flowers. A recent study found long wavelength-shifted photoreceptors (628 nm) in the flower-pollinating scarab, *Pygopleurus israelitus*, which is likely an adaptation to enhance the detection of red and orange flowers⁶³. Physiological data from the red palm weevil, *Rynchophorus ferrugineus*, also reveals additional long wavelength-shifted photoreceptors⁶⁴. The pollen beetle, *B. aeneus*, with 3 UV and 2 LW opsins, has the highest molecular complexity of opsins amongst all beetles studied thus far.

Our study reveals that some groups of beetles may have spectral sensitivity capabilities that match or even exceed those of pollinating bees and wasps (Hymenoptera), particularly in the long wavelengths. Traditionally it was assumed that trichromatic insect pollinators shaped the evolution of flower visual signals. Interestingly, our results suggest that beetle pollinators secondarily evolved the potential for trichromacy alongside the evolution of angiosperms. As beetles were presumably some of the first pollinating insects, this raises the interesting question of whether or not colour vision in beetles co-evolved with flowers or whether the beetle visual system adapted to detect flowers whose visual signals had already evolved alongside trichromatic pollinations, such as Hymenoptera.

Methods

Samples used. All available neuropteroid RNA sequence reads were obtained from the Sequence Read Archive (SRA) and raw RNA-seq data were included from refs 10 and 11 (Table S1). Data for an additional buprestid species (*Aphanisticus* sp.) were also generated (see ref. 10 for methods). Trimmomatic (v0.32)⁶⁵ was used to trim raw sequence reads of Illumina-specific adapters followed by gentle trimming of low-quality bases (Phred score <5) from the start and ends of reads according to refs 66 and 67. Finally, all reads with a resulting minimum length of <25 bases were removed. The resulting reads were assembled using Trinity (v2.1.1)⁶⁸ with default parameters. Thirteen additional unpublished assemblies were provided by 1KITE (see Tables S1 and S6). For these samples, RNA extraction and sequencing was carried out as described in ref. 24. Details regarding assembly and steps to remove cross-contamination are outlined in ref. 69. Transcriptome assemblies have been deposited into the NCBI TSA database (see Table S6 for accession, Bioproject IDs and assembly versions).

Assembly quality quantification. To ensure that all whole-body transcriptomes were of high quality for opsin detection, the completeness of each assembly was estimated using orthology detection (BUSCO v1.1b1)⁷⁰ using default settings to determine the presence of 2675 arthropod Benchmarking Universal Single-Copy Orthologs (BUSCOs; Benchmarking Universal Single-Copy Orthologs). The proportion of genes detected was used as an indicator of coverage for each transcriptome. Additionally, genes from the insect phototransduction pathway were extracted from each assembly, using a database of phototransduction gene proteins, obtained from the genomes of 10 insect species, including one coleopteran (*Tribolium castaneum*, *Acrthosiphon pisum*, *Pediculus humanus corporis*, *Nasonia vitripennis*, *Solenopsis invicta*, *Aedes aegypti*, *Culex quinquefasciatus*, *Plutella xylostella*, *Apis mellifera*, *Camponotus floridanus*; KEGG ID tca04745). Coding regions within the assemblies

were predicted using TransDecoder (<http://transdecoder.github.io>), which retains the longest open reading frame (ORF). To ensure that no putative phototransduction genes were discarded prematurely, all predicted ORFs were also searched against a BLAST database of the phototransduction genes using blastp (BLAST+ v2.2.31)⁷¹ with e-value threshold of 0.001. The resulting ORF dataset was then searched against an HMM database of the phototransduction genes using hmmscan implemented in HMMER (v3.1b2)⁷². Assembly statistics, including total contigs recovered and N50 values, were used alongside the number of phototransduction genes present and the quality of opsins recovered (see below) to assess transcriptome quality. After examining all criteria, assemblies with fewer than 10 of 15 phototransduction gene orthologs were removed prior to further analysis.

Opsin extraction. Opsins were extracted from all remaining assemblies (74 Coleoptera, four Neuroptera, five Strepsiptera, two Raphidioptera, one Megaloptera and five Hymenoptera; Table S1). Opsins were extracted using the same methods used for the detection of phototransduction genes (see above) but instead using a database of known arthropod opsins (orthoDB EOG8NKF98) plus full-length coleopteran opsin genes from Lampyridae and *Thermonectus marmoratus*. To ensure that all available opsin copies were extracted, assemblies were secondarily analysed using an additional pipeline (PIA)⁷³. All sequences were also BLASTed (<https://blast.ncbi.nlm.nih.gov/>) and subject to phylogenetic analysis to confirm that they were visual opsin proteins. Many of the samples contained tissue from multiple individuals, therefore, to exclude inter-specific variation in opsin gene copy, highly similar opsin duplicates (>99% sequence identity) were removed (CD-hit v4.6.4)^{74,75}. Sequence identity was approximated for the remaining opsin duplicates using CD-hit (v4.6.4) and BLAST. Opsin sequences have been deposited in GenBank with accession numbers KY368182–KY368379. Opsins were also obtained from four coleopterans using a tBLASTn search of coleopteran opsins against *Tribolium castaneum* (Tenebrionidae), *Anoplophora glabripennis* (Cerambycidae)⁷⁶, *Dendroctonus ponderosae* and *Hypothenemus hampei* (Curculionidae) genomes (Table S1).

Opsin phylogeny. Putative nucleotide opsin sequences were subject to a codon alignment using MAFFT (v7.273)⁷⁷ with 98 insect opsin sequences, and five outgroup cephalopod opsins (see Table S7 for accession numbers) specifying automatic alignment strategy selection. All coleopteran Rh7 opsins, of which the function is unknown, were excluded to reduce the number of alignment gaps. Potential contaminate opsin sequences and non-visual opsins (peropsins and RGRs) were also removed. Phylogenetic inference was performed on all final opsin nucleotide sequences using maximum likelihood (IQ-TREE v1.4.1)⁷⁸ with 10,000 UFBoot iterations. Estimation of the best fitting model was carried out using ModelFinder within IQ-TREE. LG + F + I + G4 was selected as the best-fit model. All trees were edited in FigTree (v1.4.2 <http://tree.bio.ed.ac.uk/software/figtree/>).

Species phylogeny. The 13 transcriptome assemblies provided by 1KITE were excluded from the species phylogeny according to 1KITE's data release policy. To construct a species phylogeny, protein ORFs were generated from the remaining transcriptomes, using only one per species, with male adult samples used when possible. Orthologous gene clusters were generated using OrthoMCL (v2.0.9)⁷⁹ and individually aligned using MAFFT (v7.273)⁷⁷ with automatic alignment strategy selection. Each cluster represented a single gene and for species with multiple isoforms per gene, only the longest isoform was retained. Clusters were then filtered using a machine-learning algorithm implemented in OGCleaner^{80,81} to remove low quality putative homology clusters. Alignment quality was assessed using Aliscore (v2.0)⁸², and Alicut (v2.0) was used to remove ambiguously aligned positions in the multiple sequence alignments^{82,83}. Only clusters with genes from at least 20% of species were used for further analysis. Lastly, a supermatrix was generated of 358 aligned gene clusters, which was used to infer a maximum-likelihood species phylogeny (IQ-TREE)⁷⁸ with 10,000 ultrafast bootstrap (UFBoot) iterations⁸⁴. UV and LW opsin duplication events were reconstructed on the species phylogeny, using both parsimony and maximum likelihood frameworks for ancestral state reconstruction in Mesquite (v3.04)⁸⁵. Likelihood was estimated using the Markov k-state 1 (Mk1) parameter model, whereby all character changes are equally probable. Two characters were specified according to the opsin gene data: single opsin gene or duplicated opsin. Multiple duplications within one species were assigned the latter category. Both ancestral state reconstruction and the topology of the opsin gene tree were used to infer the pattern of duplications across Coleoptera.

Positive selection analysis. Selection analysis was performed on the maximum likelihood topology using a full length DNA alignment of all opsins in this study, 98 additional insect opsins and five cephalopod opsins (Table S7). To test for signals of episodic positive selection (PS), we used CodeML within PAML (v4.9a)⁸⁶. The branch-site model A, which allows ω variation among sites as well as tree branches, was used to detect signatures of PS affecting certain lineages and sites⁸⁷. In order to determine significance of inferred selection patterns, we compared the branch-site null model A, which assumes fixed $\omega = 1$ (i.e. neutrality) against the branch-site model A. Then, the test statistic for likelihood ratio test (LRT) was calculated as twice the difference in log likelihood (ℓ) between the two models ($2\Delta\ell$). Significance was assessed using a chi-square distribution with one degree of freedom. To avoid local optima, starting values of 0.1, 1 and 2 ω were used and the result with the best ℓ retained. A Bayes empirical Bayes approach was taken to determine amino acid sites likely to be under PS according a posterior probability with significance at $\geq 95\%$. The position of the chromophore-binding pocket (i.e. the region of the protein that interacts with the chromophore) was identified using 3D protein modelling (I-TASSER online server)^{88,89} using squid rhodopsin as a template (PDB model 2Z73A)⁹⁰. Sites within the binding pocket were highlighted as potential candidates for spectral tuning and compared with known tuning sites in other taxa.

References

1. Wakakuwa, M. *et al.* Evolution and mechanism of spectral tuning of blue-absorbing visual pigments in butterflies. *PLoS One* **5**, e15015 (2010).
2. Futahashi, R. *et al.* Extraordinary diversity of visual opsin genes in dragonflies. *Proc. Natl. Acad. Sci.* **112**, E1247–E1256 (2015).

3. Briscoe, A. D., Bernard, G. D., Szeto, A. S., Nagy, L. M. & White, R. H. Not all butterfly eyes are created equal: rhodopsin absorption spectra, molecular identification, and localization of ultraviolet-, blue-, and green-sensitive rhodopsin-encoding mRNAs in the retina of *Vanessa cardui*. *J. Comp. Neurol.* **458**, 334–349 (2003).
4. Lynch, M. & Conery, J. S. The evolutionary fate and consequences of duplicate genes. *Science* **290**, 1151–1155 (2000).
5. Colbourne, J. K. *et al.* The ecoresponsive genome of *Daphnia pulex*. *Science* **331**, 555–561 (2011).
6. Hill, C. A. *et al.* G protein coupled receptors in *Anopheles gambiae*. *Science* **298**, 176–178 (2002).
7. Briscoe, A. D. *et al.* 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 (2010).
8. Matsumoto, Y., Wakakuwa, M., Yukuhiro, F., Arikawa, K. & Noda, H. Attraction to Different Wavelength Light Emitting Diodes (LEDs), the Compound Eye Structure, and opsin Genes in *Nilaparvata lugens*. *Japanese J. Appl. Entomol. Zool.* **58**, 111–118 (2014).
9. Maksimovic, S., Cook, T. A. & Buschbeck, E. K. Spatial distribution of opsin-encoding mRNAs in the tiered larval retinas of the sunburst diving beetle *Thermonectus marmoratus* (Coleoptera: Dytiscidae). *J. Exp. Biol.* **212**, 3781–3794 (2009).
10. Lord, N. P. *et al.* A cure for the blues: opsin duplication and subfunctionalization for short-wavelength sensitivity in jewel beetles (Coleoptera: Buprestidae). *BMC Evol. Biol.* **16**, 107 (2016).
11. Martin, G. J., Lord, N. P., Branham, M. A. & Bybee, S. M. Review of the firefly visual system (Coleoptera: Lampyridae) and evolution of the opsin genes underlying color vision. *Org. Divers. Evol.* **15**, 513–526 (2015).
12. Sander, S. E. & Hall, D. W. Variation in opsin genes correlates with signalling ecology in North American fireflies. *Mol. Ecol.* **24**, 4679–4696 (2015).
13. Tierney, S. M. *et al.* Opsin transcripts of predatory diving beetles: a comparison of surface and subterranean photic niches. *R. Soc. open Sci.* **2**, 140386 (2015).
14. Jackowska, M. *et al.* Genomic and gene regulatory signatures of cryptozoic adaptation: Loss of blue sensitive photoreceptors through expansion of long wavelength-opsin expression in the red flour beetle *Tribolium castaneum*. *Front. Zool.* **4**, 24 (2007).
15. Lin, J. T. & Wu, C. Y. A comparative study on the color vision of four coleopterous insects. *Bull. Inst. Zool. Acad. Sin.* **31**, 81–88 (1992).
16. Lin, J. T. Identification of photoreceptor locations in the compound eye of *Coccinella septempunctata* Linnaeus (Coleoptera, Coccinellidae). *J. Insect Physiol.* **39**, 555–562 (1993).
17. Döring, T. F. & Skorupski, P. Host and non-host leaves in the colour space of the Colorado potato beetle (Coleoptera: Chrysomelidae). *Entomol. Gen.* **29**, 081–095 (2007).
18. Hasselmann, E. M. Über die relative spektrale Empfindlichkeit von Käfer- und Schmetterlingsaugen bei verschiedenen Helligkeiten. *Zool. Jahrbucher* **69**, 573–576 (1962).
19. McKenna, D. D. *et al.* The beetle tree of life reveals that Coleoptera survived end-Permian mass extinction to diversify during the Cretaceous terrestrial revolution. *Syst. Entomol.* **40**, 835–880 (2015).
20. Peters, R. S. *et al.* The evolutionary history of holometabolous insects inferred from transcriptome-based phylogeny and comprehensive morphological data. *BMC Evol. Biol.* **14**, 52 (2014).
21. Townson, S. M. *et al.* Honeybee blue- and ultraviolet-sensitive opsins: cloning, heterologous expression in *Drosophila*, and physiological characterization. *J. Neurosci.* **18**, 2412–2422 (1998).
22. Carulli, J. P., Chen, D.-M., Stark, W. S. & Hartl, D. L. Phylogeny and physiology of *Drosophila* opsins. *J. Mol. Evol.* **38**, 250–262 (1994).
23. Chase, M. R., Bennett, R. R. & White, R. H. Three opsin-encoding cDNAs from the compound eye of *Manduca sexta*. *J. Exp. Biol.* **200**, 2469–2478 (1997).
24. Misof, B. *et al.* Phylogenomics resolves the timing and pattern of insect evolution. *Science* **346**, 763–767 (2014).
25. Yinon, U. The visual mechanisms of *Tenebrio molitor*: some aspects of the spectral response. *J. Exp. Biol.* **53**, 221–229 (1970).
26. Salcedo, E., Zheng, L., Phistry, M., Bagg, E. E. & Britt, S. G. Molecular basis for ultraviolet vision in invertebrates. *J. Neurosci.* **23**, 10873–10878 (2003).
27. Lall, A. B., Strother, G. K., Cronin, T. W. & Seliger, H. H. Modification of spectral sensitivities by screening pigments in the compound eyes of twilight-active fireflies (Coleoptera: Lampyridae). *J. Comp. Physiol. A* **162**, 23–33 (1988).
28. French, A. S., Meisner, S., Liu, H., Weckström, M. & Torkkeli, P. H. Transcriptome analysis and RNA interference of cockroach phototransduction indicate three opsins and suggest a major role for TRPL channels. *Front. Physiol.* **6**, 1–10 (2015).
29. Hunt, D. M., Dulai, K. S., Partridge, J. C., Cottrell, P. & Bowmaker, J. K. The molecular basis for spectral tuning of rod visual pigments in deep-sea fish. *J. Exp. Biol.* **204**, 3333–3344 (2001).
30. Simões, B. F. *et al.* Visual system evolution and the nature of the ancestral snake. *J. Evol. Biol.* **28**, 1309–1320 (2015).
31. Mohun, S. M. *et al.* Identification and characterization of visual pigments in caecilians (Amphibia: Gymnophiona), an order of limbless vertebrates with rudimentary eyes. *J. Exp. Biol.* **213**, 3586–3592 (2010).
32. Carvalho, L. D. S., Cowing, J. A., Wilkie, S. E., Bowmaker, J. K. & Hunt, D. M. Shortwave visual sensitivity in tree and flying squirrels reflects changes in lifestyle. *Curr. Biol.* **16**, 81–83 (2006).
33. Newman, L. A. & Robinson, P. R. Cone visual pigments of aquatic mammals. *Vis. Neurosci.* **22**, 873–879 (2005).
34. Cowing, J. A., Arrese, C. A., Davies, W. L., Beazley, L. D. & Hunt, D. M. Cone visual pigments in two marsupial species: the fat-tailed dunnart (*Sminthopsis crassicaudata*) and the honey possum (*Tarsipes rostratus*). *Proc. R. Soc. B* **275**, 1491–1499 (2008).
35. Jacobs, G. H. Losses of functional opsin genes, short-wavelength cone photopigments, and color vision—A significant trend in the evolution of mammalian vision. *Vis. Neurosci.* **30**, 39–53 (2013).
36. Laughlin, S. B., de Ruyter van Steveninck, R. R. & Anderson, J. C. The metabolic cost of neural information. *Nat. Neurosci.* **1**, 36–41 (1998).
37. Kelber, A. & Lind, O. Limits of colour vision in dim light. *Ophthalmic Physiol. Opt.* **30**, 454–459 (2010).
38. Kelber, A., Balkenius, A. & Warrant, E. J. Scotopic colour vision in nocturnal hawkmoths. *Nature* **419**, 922–925 (2002).
39. Vas, J., Ábrahám, L. & Markó, V. Study of Nocturnal and Diurnal Activities of Lacewings (Neuropteroidea: Raphidioptera, Neuroptera) by Suction Trap. *Acta Phytopathol. Entomol. Hungarica* **34**, 149–152 (1999).
40. Buschbeck, E. K., Ehmer, B. & Hoy, R. R. The unusual visual system of the Strepsiptera: external eye and neuropils. *J. Comp. Physiol. A* **189**, 617–630 (2003).
41. Terakita, A. The opsins. *Genome Biol.* **6**, 213 (2005).
42. Hering, L. *et al.* Opsins in Onychophora (velvet worms) suggest a single origin and subsequent diversification of visual pigments in arthropods. *Mol. Biol. Evol.* **29**, 3451–3458 (2012).
43. Mollon, J. D. ‘Tho’ she kneel’d in that place where they grew...’ The uses and origins of primate colour vision. *J. exp. Biol.* **146**, 21–38 (1989).
44. Osorio, D. & Vorobyev, M. Colour vision as an adaptation to frugivory in primates. *Proc. R. Soc. B* **236**, 593–599 (1996).
45. Warrant, E. J. & McIntyre, P. D. Limitations to resolution in superposition eyes. *J. Comp. Physiol. A* **167**, 785–803 (1990).
46. Groberman, L. J. & Borden, J. H. Electrophysiological response of *Dendroctonus pseudotsugae* and *Ips paraconfusus* (Coleoptera: Scolytidae) to selected wavelength regions of the visible spectrum. *Can. J. Zool.* **60**, 2180–2189 (1982).
47. Gribakin, F. G. Automatic spectrosensitometry of photoreceptors in *Lethrus* (Coleoptera, Scarabaeidae). *J. Comp. Physiol. A* **142**, 95–102 (1981).
48. Crook, D. J. *et al.* Laboratory and field response of the emerald ash borer (Coleoptera: Buprestidae), to selected regions of the electromagnetic spectrum. *J. Econ. Entomol.* **102**, 2160–2169 (2009).

49. Döring, T. F., Skellern, M., Watts, N. & Cook, S. M. Colour choice behaviour in the pollen beetle *Meligethes aeneus* (Coleoptera: Nitidulidae). *Physiol. Entomol.* **37**, 360–378 (2012).
50. Caveney, S. The phylogenetic significance of ommatidium structure in the compound eyes of polyphagan beetles. *Can. J. Zool.* **64**, 1787–1819 (1986).
51. Jönsson, M., Rosdahl, K. & Anderson, P. Responses to olfactory and visual cues by over-wintered and summer generations of the pollen beetle, *Meligethes aeneus*. *Physiol. Entomol.* **32**, 188–193 (2007).
52. Van Kleunen, M., Nänni, I., Donaldson, J. S. & Manning, J. C. The role of beetle marks and flower colour on visitation by monkey beetles (Hopliini) in the Greater Cape Floral Region, South Africa. *Ann. Bot.* **100**, 1483–1489 (2007).
53. Johnson, S. D. & Midgley, J. J. Pollination by Monkey Beetles (Scarabaeidae: Hopliini): Do Color and Dark Centers of Flowers Influence Alighting Behavior? *Environ. Entomol.* **30**, 861–868 (2001).
54. Harmon, J. P., Losey, J. E. & Ives, A. R. The role of vision and color in the close proximity foraging behavior of four coccinellid species. *Oecologia* **115**, 287–292 (1998).
55. Kolmes, S. A. Ecological and sensory aspects of prey capture by the whirligig beetle *Dineutes discolor* (Coleoptera: Gyrinidae). *J. New York Entomol. Soc.* **91**, 405–412 (1983).
56. Lin, C. & Strausfeld, N. J. Visual inputs to the mushroom body calyces of the whirligig beetle *Dineutus sublineatus*: Modality switching in an insect. *J. Comp. Neurol.* **520**, 2562–2574 (2012).
57. Szentesi, A., Weber, D. C. & Jermy, T. Role of visual stimuli in host and mate location of the Colorado potato beetle. *Entomol. Exp. Appl.* **105**, 141–152 (2002).
58. Van der Ent, L. J. & Visser, J. H. The visual world of the Colorado potato beetle. *Exp. Appl. Entomol. Proc. Netherlands Entomol. Soc.* **2**, 80–85 (1991).
59. Mondor, E. B. & Warren, J. L. Unconditioned and conditioned responses to colour in the predatory coccinellid, *Harmonia axyridis* (Coleoptera: Coccinellidae). *Eur. J. Entomol.* **97**, 463–467 (2000).
60. Lelito, J. P. *et al.* Visually mediated 'paratrooper copulations' in the mating behavior of *Agrilus planipennis* (Coleoptera: Buprestidae), a highly destructive invasive pest of North American ash trees. *J. Insect Behav.* **20**, 537–552 (2007).
61. Frentiu, F. D., Bernard, G. D., Sison-Mangus, M. P., Brower, A. V. Z. & Briscoe, A. D. Gene duplication is an evolutionary mechanism for expanding spectral diversity in the long-wavelength photopigments of butterflies. *Mol. Biol. Evol.* **24**, 2016–2028 (2007).
62. Suvorov, A. *et al.* Opsins have evolved under the permanent heterozygote model: insights from phylotranscriptomics of Odonata. *Mol. Ecol.* doi:10.1111/mec.13884 (2016).
63. Martínez-Harms, J. *et al.* Evidence of red sensitive photoreceptors in *Pygopleurus israelitus* (Glaphyridae: Coleoptera) and its implications for beetle pollination in the southeast Mediterranean. *J. Comp. Physiol. A* **198**, 451–463 (2012).
64. Ilić, M., Pirić, P. & Belušić, G. Four photoreceptor classes in the open rhabdom eye of the red palm weevil, *Rynchophorus ferrugineus* Olivier. *J. Comp. Physiol. A* **202**, 203–213 (2016).
65. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
66. Mbandi, S. K., Hesse, U., Rees, D. J. G. & Christoffels, A. A glance at quality score: Implication for de novo transcriptome reconstruction of Illumina reads. *Front. Genet.* **5**, 1–5 (2014).
67. MacManes, M. D. On the optimal trimming of high-throughput mRNA sequence data. *Front. Genet.* **5**, 1–7 (2014).
68. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).
69. Mayer, C. *et al.* BaitFisher: A software package for multi-species target DNA enrichment probe design. *Mol. Biol. Evol.* **33**, 1875–1886 (2016).
70. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
71. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
72. Eddy, S. R. Accelerated profile HMM searches. *PLoS Comput. Biol.* **7**, e1002195 (2011).
73. Speiser, D. I. *et al.* Using phylogenetically-informed annotation (PIA) to search for light-interacting genes in transcriptomes from non-model organisms. *BMC Bioinformatics* **15**, 350 (2014).
74. Li, W. & Godzik, A. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
75. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
76. McKenna, D. D. *et al.* Genome of the Asian longhorned beetle (*Anoplophora glabripennis*), a globally significant invasive species, reveals key functional and evolutionary innovations at the beetle-plant interface. *Genome Biol.* **17**, 1–18 (2016).
77. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
78. Nguyen, L. T., Schmidt, H. A., Von Haeseler, A. & Minh, B. Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
79. Li, L., Stoeckert, C. J. J. & Roos, D. S. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178–2189 (2003).
80. Fujimoto, M. S., Suvorov, A., Jensen, N. O., Clement, M. J. & Bybee, S. M. Detecting false positive sequence homology: a machine learning approach. *BMC Bioinformatics* **17**, 101 (2016).
81. Fujimoto, M. S. *et al.* The OGCleaner: filtering false-positive homology clusters. *Bioinformatics* **33**, 125–127 (2016).
82. Misof, B. & Misof, K. A Monte Carlo Approach Successfully Identifies Randomness in Multiple Sequence Alignments: A More Objective Means of Data Exclusion. *Syst. Biol.* **58**, 21–34 (2009).
83. Kück, P. *et al.* Parametric and non-parametric masking of randomness in sequence alignments can be improved and leads to better resolved trees. *Front. Zool.* **7**, 10 (2010).
84. Minh, B. Q., Nguyen, M. A. T. & Von Haeseler, A. Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* **30**, 1188–1195 (2013).
85. Maddison, W. & Maddison, D. Mesquite: a modular system for evolutionary analysis. (2011).
86. Yang, Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
87. Zhang, J., Nielsen, R. & Yang, Z. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol. Biol. Evol.* **22**, 2472–2479 (2005).
88. Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**, 40 (2008).
89. Yang, J. *et al.* The I-TASSER Suite: Protein structure and function prediction. *Nat. Methods* **12**, 7–8 (2015).
90. Murakami, M. & Kouyama, T. Crystal structure of squid rhodopsin. *Nature* **453**, 363–367 (2008).

Acknowledgements

We thank all researchers whose published RNA-seq data were used in this study. We are grateful to Rebecca Plimpton for her assistance with opsin protein modelling. We thank Stephen Richards and the Insect 5000 Genomes Project (Baylor College of Medicine Human Genome Sequencing Center; <https://www.hgsc.bcm.edu>)

for access to data from the unpublished i5k genome for *Anoplophora glabripennis*. We also thank Dan O'Sullivan for creating the beetle illustrations. Further, we thank the following people for collecting and/or providing specimens: Eric Anton, Rolf Beutel, Dan Young, Margarethe Brummerman, Ralph Peters, Carola Greve, Sven Nekum, Alexander Blanke, Susanne Dobler, Cary Minter, Dirk Ahrens, Jonas Eberle and Silvia Fabrizi. Finally we thank Alexander Donath, Lars Posiadlowski, Xin Zhou, Karl Kjer, Bernhard Misof, Karen Meusemann, Adam Ślipiński and the 1KITE Coleoptera subproject for allowing access to unpublished transcriptome assemblies, providing sample information and assisting with 1KITE coordination. This work was supported by National Science Foundation Division of Environmental Biology grants [#1355169 DDM] and [#1265714 SMB].

Author Contributions

C.R.S. performed data analyses, interpreted data and wrote the manuscript. M.S.F. and A.S. analysed data. G.J.M., N.L. and S.M.B. assisted with data interpretation. D.D.M. provided specimens, coordinated with 1KITE, and is the analytical lead on the 1KITE beetle project. S.S. analysed 1KITE beetle project data and coordinated with 1KITE. G.J.M., N.L., S.M.B., D.D.M. and S.S. edited the manuscript.

Additional Information

Supplementary information accompanies this paper at doi:[10.1038/s41598-017-00061-7](https://doi.org/10.1038/s41598-017-00061-7)

Competing financial interests: The authors declare no competing financial interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017