J Mol Evol (2017) 84:12–28
DOI 10.1007/s00239-016-9777-1

ORIGINAL ARTICLE

Ancient and Recent Duplications Support Functional Diversity of *Daphnia* Opsins

Christopher S. Brandon[1] · Matthew J. Greenwold[1] · Jeffry L. Dudycha[1]

Received: 19 September 2016 / Accepted: 4 December 2016 / Published online: 21 December 2016

Abstract *Daphnia pulex* has the largest known family of opsins, genes critical for photoreception and vision in animals. This diversity may be functionally redundant, arising from recent processes, or ancient duplications may have been preserved due to distinct functions and independent contributions to fitness. We analyzed opsins in *D. pulex* and its distant congener *Daphnia magna*. We identified 48 opsins in the *D. pulex* genome and 32 in *D. magna*. We inferred the complement of opsins in the last common ancestor of all *Daphnia* and evaluated the history of opsin duplication and loss. We further analyzed sequence variation to assess possible functional diversification among *Daphnia* opsins. Much of the opsin expansion occurred before the *D. pulex*-*D. magna* split more than 145 Mya, and both *Daphnia* lineages preserved most ancient opsins. More recent expansion occurred in pteropsins and long-wavelength visual opsins in both species, particularly *D. pulex*. Recent duplications were not random: the same ancestral genes duplicated independently in each modern species. Most ancient and some recent duplications involved differentiation at residues known to influence spectral tuning of visual opsins. Arthropsins show evidence of gene conversion between tandemly arrayed paralogs in functionally important domains. Intron–exon gene structure was generally conserved within clades inferred from sequences, although pteropsins showed substantial intron size variation. Overall, our analyses support the hypotheses that diverse opsins are maintained due to diverse functional roles in photoreception and vision, that functional diversification is both ancient and recent, and that multiple evolutionary processes have influenced different types of opsins.

Keywords Opsin · Gene family · Vision · Photoreception · Gene duplication · Pteropsin · Arthropsin · Neuropsin · Spectral tuning

Introduction

The sequenced genome of the freshwater microcrustacean *Daphnia pulex* revealed the largest family of opsins—genes critical for vision—of any known species (Colbourne et al. 2011). One explanation is that the large family size is a consequence of widespread duplication events that were specific to *D. pulex*. However, opsin expansion may instead pre-date the origin of *D. pulex* as a species. Thus, an alternate explanation is that ancient duplications and preferential retention of opsins with differentiated functions may have led to expansion of the gene family. Most *D. pulex* opsin genes code for complete proteins and are supported by expression evidence, implying that they continue to play functional roles in *Daphnia* vision and photoreception (Colbourne et al. 2011). Recently, the *Daphnia* Genomics Consortium sequenced the *Daphnia magna* genome. *D. pulex* and *D. magna* are members of separate subgenera with an estimated divergence time of 200 million years; they represent the deepest possible split within the genus (Colbourne and Hebert 1996), which has ~ 100 described species globally (Benzie 2005). This estimate is reinforced by fossils showing the subgenera had

Electronic supplementary material The online version of this article (doi:10.1007/s00239-016-9777-1) contains supplementary material, which is available to authorized users.

✉ Jeffry L. Dudycha
dudycha@biol.sc.edu

[1] Department of Biological Sciences, University of South Carolina, Columbia, SC 29208, USA

already diverged in the Mesozoic, more than 145 Mya (Kotov and Taylor 2011). By comparing the opsin families found in these two genomes, we can better understand the history of gene family expansion and functional diversification within the family.

Opsins are members of a large and diverse class of G protein-coupled receptors that include many proteins involved in sensory reception (reviewed in Terakita 2005). Opsins function in photoreception, and are a necessary component for vision. Visual pigments consist of an opsin covalently bound to a chromophore, typically an 11-*cis* vitamin A_1 derivative, that absorbs photons. Photon absorption changes the conformation of the visual pigment, initiating a signal transduction cascade (Nathans 1987; Rosenbaum et al. 2009). Opsins feature seven highly conserved transmembrane (TM) motifs, an N-terminus in the extracellular region, and a C-terminus located in the cytosol (Palczewski et al. 2000). Functionally photoreceptive opsins contain a lysine residue in the seventh transmembrane domain that binds the retinal chromophore through Schiff-base linkage (Lewis et al. 1978).

Across animals, opsins can be divided into three categories. The ciliary and rhabdomeric opsins are monophyletic clades of opsins that separated prior to the protostome–deuterostome split (Terakita 2005; Shichida and Matsuyama 2009; Hering and Mayer 2014). The remaining opsins are a heterogeneous group that includes the photoisomerases and neuropsins; some authors have termed this group the "Group 4" opsins (Porter et al. 2012). The *Daphnia* genome contains members of each of these groups (Colbourne et al. 2011; Hering and Mayer 2014).

Rhabdomeric opsins are responsible for vision in arthropods (Terakita et al.1993; Lee et al.1994). The molecular structure of an opsin defines which wavelength(s) of light the pigment is most efficient at absorbing. We therefore refer to visual opsins and their inferred color of maximum absorption on the basis of sequence similarity. Colbourne et al. (2011) used comparative sequence analysis to conclude that *D. pulex* had four distinct visual opsin subgroups maximally sensitive to separate wavelengths of light: ultraviolet, blue, green, and red. This is consistent with evidence from a low-resolution electrophysiological study in *D. magna* that found four peak wavelengths of spectral sensitivity (Smith and Macagno 1990). Colbourne et al. (2011) reported that the *D. pulex* genome contains 25 long-wavelength opsins, the largest subgroup of its opsin gene family, which they divided into a long-wavelength A clade (LOPA, predicted to be maximally sensitive to green light) and a long-wavelength B clade (LOPB, predicted to be maximally sensitive to red light). Additional individual opsins are specific to blue and ultraviolet light, respectively. This is the largest number of visual opsins yet known.

Opsin diversity in *Daphnia pulex* is not limited to visual opsins. Colbourne et al. (2011) described two other large clades of opsins. One was a novel type of rhabdomeric opsins the authors dubbed arthropsins. Little is known about their function, but evidence from a spider and a velvet worm has shown expression in central nervous tissue (Eriksson et al. 2013). The other clade contains nine pteropsins, which is the largest ciliary opsin group known in invertebrates (Hering and Mayer 2014). Pteropsins have been proposed to mediate circadian rhythm in some capacity (Velarde et al. 2005; Tierney et al. 2015), but no empirical studies have yet tested their biological role.

In this study, we conduct a comparative evolutionary analysis of the opsin gene family of two species of *Daphnia*, a genus that has provided insights on visual ecology (Smith and Baylor 1953; Stearns 1975; Young et al.1984; De Meester 1991, 1993; Storz and Paul 1998; Hamza and Ruggio 2000) and contemporary processes of eye evolution (Brandon and Dudycha 2014; Brandon et al.2015). We had two major aims. First, we sought to determine what the ancestral *Daphnia* opsin complement was, and thereby determine which opsins are the product of ancient versus recent duplications. Second, we sought to evaluate the temporal distribution of functional diversification within the gene family. In particular, we are interested in testing the hypothesis that there were four ancient visual opsins in the earliest *Daphnia*, each ancestral to one of the four clades inferred to represent different spectral classes in modern *Daphnia*.

Methods

Daphnia magna Opsin Gene Discovery

We built an initial list of *D. magna* opsins from a genome-wide assembly of predicted genes. We downloaded the 2012 *D. magna* predicted gene database, trall7set9rbest, from wfleasbase.org, and used the NCBI standalone BLAST+ version 2.2.29+ software to conduct the search (Camacho et al. 2009). We obtained amino acid sequences for the 44 *D. pulex* opsins described in Colbourne et al. (2011) from NCBI GenBank or the Joint Genome Institute's *Daphnia* database (http://genome.jgi-psf.org/Dappu1). Two genes that were curated as small fragments without protein translations were excluded. We used the *D. pulex* sequences as our query to search the *D. magna* database using the default settings of BLASTp. We retained sequences from our search with an e-value of 5×10^{-4} or lower.

We then used the initial list of *D. magna* opsin transcript sequences as the basis for a gene-by-gene search in the *D. magna* genome assembly 2.4 using the BLAST function available on wfleabase.org. For each *D. magna* sequence, we searched the genome using tBLASTn with default settings and recorded the location of the sequence with the lowest e-value. If it was already recorded from a previous query, we additionally recorded the sequence with the next lowest e-value. After querying with each of the initial *D. magna* opsin transcript sequences, we searched the *D. magna* 2.4 assembly once more with the output of the transcript queries. We also searched the *D. magna* genome using the amino acid sequences of *D. pulex* opsins as our queries to ensure that we had identified as many potential opsin sequences as possible. Finally, to confirm that the *D. magna* gene sequences were opsins, we performed a reciprocal BLAST search of the NCBI non-redundant protein sequence database using BLASTp. To ensure that sequences were complete, in some cases we did additional targeted searches for exons that were overlooked by gene model prediction algorithms.

After conducting the searches described above, we learned of a new opsin in *D. pulex*, a neuropsin/opsin-5, recently identified by Hering and Mayer (2014). Our searches had already produced a homolog of this opsin-5, but to ensure we did not miss another copy, we used the *D. pulex* protein sequence for opsin-5 for an additional tBLASTn search. This search uncovered an additional, novel ciliary opsin.

Phylogenetic Analyses

Ciliary and rhabdomeric opsins are known to have diverged prior to the protostome–deuterostome split (Hering and Mayer 2014; Feuda et al.2014). Because we are interested in the history of the *Daphnia* genus, we therefore analyzed ciliary and rhabdomeric opsins separately. One *Daphnia* opsin belonged to neither of these groups; we included it with the ciliary opsins because it diverged from them following the protostome-deuterostome split (Hering and Mayer 2014). To confirm that this gene was indeed an opsin and determine where it should be placed, we used the phylogenetically informed annotation tool developed by Speiser et al. (2014). This tool uses a maximum likelihood approach to place new sequences within an existing phylogenetic tree, providing a better estimate of orthologies than reciprocal BLAST.

We focused our phylogenetic analyses on protein-coding DNA sequences rather than amino acid sequences because DNA provides better resolution for recently duplicated genes. We determined protein-coding nucleotide sequences by aligning the predicted amino acid sequences to genomic DNA using GeneWise (Birney et al. 2004). We used nucleotide sequences from cDNA and not gDNA for *D. magna* Arthropsin2 and 3, because large regions of their gDNA were missing from the genome assembly. We aligned codon sequences with an open gap penalty of −2.9 using MUSCLE as available in the MEGA6 software package (Tamura et al. 2013). We ran phylogenetic analyses using maximum likelihood in RAxML v. 8.1 (Stamatakis 2014) using a general time reversal (GTR) substitution matrix and GAMMA plus proportion of invariable sites estimate. We set RAxML to terminate bootstrap replication automatically, which terminated at 360 replicates for the ciliary opsins and 156 replicates for the rhabdomeric. We estimated the phylogeny without setting an outgroup to avoid constraining tree construction. To then root the phylogeny of *Daphnia* rhabdomeric opsins, we used a set of six vertebrate ciliary opsins from *Danio rerio* and *Bos taurus*. For our ciliary opsin analysis, we rooted the tree using four vertebrate melanopsins (non-visual rhabdomeric opsins). Additionally, we included a number of non-*Daphnia* invertebrate opsin sequences in both analyses. We obtained mRNA and amino acid sequences for the additional sequences from NCBI (accession numbers listed in Supplementary Table 1). Sequence alignments for tree building are in Supplemental Files 1 and 2.

Daphnia Opsin Amino Acid Conservation

We aligned the inferred sequences of the *Daphnia* proteins for each major opsin clade using ClustalW2 (Larkin et al. 2007). To identify functional domains and key residues, we then mapped the alignment to the bovine rhodopsin (GenBank accession: NP_001014890) and imported them into *MEGA*6 (Tamura et al. 2013). We identified transmembrane (TM) domains, cytoplasmic loops, and extracellular loops using the model of bovine rhodopsin secondary structure (Terakita 2005). We grouped alpha-helix VIII of bovine rhodopsin in the C-terminus domain. Alignments can be found in Supplementary File 3.

To examine the sites that may influence spectral tuning in *Daphnia*, we performed an amino acid alignment of *Daphnia* visual opsins and the bovine rhodopsin using MUSCLE as available in *MEGA6*. Amino acid sites that influence spectral tuning have been identified in several invertebrates through phylogenetic and selection analyses (Briscoe 2001, 2002; Porter et al. 2007, 2009) or experimental manipulations (Salcedo et al. 2003, 2009; Wakakuwa et al. 2010). We used this work as a guide to identify sites where variation among *Daphnia* opsin sequences may indicate differences in spectral tuning.

Gene Conversion and Recombination Analysis

The evolution of tandemly arrayed multigene families can be shaped by multiple processes such as diversification by mutation or homogenization through concerted evolution. Gene conversion, a mechanism of concerted evolution, is the non-reciprocal transfer of genetic material that leads to the homogenization of paralogous genes within a genome (Ohta 2010). Gene conversion can reduce differentiation between paralogs, limiting the potential for diversification, and is known to occur at relatively high rates in *Daphnia* (Omilian et al. 2006; Keith et al. 2016). We therefore sought to identify the frequency and distribution of gene conversion among opsin paralogs. Nucleotide sequences were grouped by clade, aligned in ClustalW2 (Larkin et al. 2007), and analyzed in GENECONV v.1.81a (Sawyer 1989) to detect statistically significant sequence homogenization between paralogs using Bonferroni-corrected Karlin-Altschul *p*-values with a cutoff of 0.05.

Opsin Gene Structures

We determined the intron–exon structures of *Daphnia* opsins to provide more clarity on the evolutionary relationships within each opsin subgroup and evaluate conservation of gene structure. We retrieved curated gene structures for *D. pulex* from JGI (http://genome.jgi-psf.org/Dappu1). We obtained structures for *D. magna* genes by pairwise alignment of amino acid sequences to genomic DNA using default parameters in GeneWise. Gaps in the genome assembly prevented us from determining structures for two *D. magna* opsins.

D. magna Opsin Gene Nomenclature

We named *D. magna* opsin genes following the system described in Colbourne et al. (2011). For opsin subgroups with multiple genes, we numbered the gene according to its most similar homolog in *D. pulex*. We numbered opsins with a decimal number if there was no clear gene-to-gene homology between *D. pulex* and *D. magna* sequences, reflecting duplications specific to *D. magna*.

Results

Opsin Gene Number and Discovery

The *D. pulex* genome contains 48 opsin genes, whereas the *D. magna* genome contains 32 (Table 1). The *D. magna* genome contains orthologs for representatives of all categories of opsins and all clades of visual opsins identified in *D. pulex* (Table 1). Lineage-specific expansion of long-wavelength (red and green) visual opsins occurred in both species, but was greater in *D. pulex* (Fig. 1). In both species, one pteropsin appears to have undergone independent expansions (Fig. 2). Other opsin groups showed no gains after divergence between *D. pulex* and *D. magna*. Individual gene losses occurred in the pteropsins, arthropsins, and long-wavelength visual opsins (Figs. 1, 2).

Table 1 Number of opsin genes in *Daphnia pulex* and *D. magna*

Opsin category	*D. pulex*	*D. magna*
Ultraviolet	1	1
Blue	1	1
Unknown wavelength, possibly visual	2	2
Long-wavelength A (green-like)	10	4
Long-wavelength B (red-like)	15	8
Arthropsin	8	7
Pteropsin	9	7
Opsin-5	1	1
Other ciliary opsin	1	1
Total	48	32

Within the rhabdomeric opsins (Fig. 1), *D. magna* and *D. pulex* have matched orthologous pairs of genes for both ultraviolet and blue visual opsins. Two other pairs of orthologs cluster with the visual opsins, but are separate and therefore specific wavelength sensitivity cannot be predicted. The two genomes have a similar complement of arthropsins, with most having an ortholog in each genome. In contrast, long-wavelength visual opsins differ substantially between the genomes, with *D. pulex* having roughly twice as many as *D. magna*. There are ten long-wavelength A (LOPA, putatively green-sensitive) opsins, but only four in *D. magna*. Fifteen long-wavelength B (LOPB, putatively red-sensitive) genes are in *D. pulex*, but only eight in *D. magna*. These differences are not simply due to expansion in *D. pulex*; both species harbor independent duplications in both clades of long-wavelength opsins.

Within the ciliary opsins (Fig. 2), each species has a similar number of pteropsins, but only half are present as orthologous pairs. We also found an orthologous pair of previously undescribed ciliary opsins. This pair is clearly distinct from the pteropsins (Fig. 2), and we therefore named it *c-opsin1*.

One orthologous pair of opsin genes is neither rhabdomeric nor ciliary. This opsin clusters with a group of vertebrate neuropsin genes known as opsin-5 genes (Fig. 2; Hering and Mayer 2014). We therefore named it *opsin5*.

A complete list of opsin genes for each species, including names, identifiers, and genomic location, is given in Supplementary Tables 2 (*D. pulex*) and 3 (*D. magna*).

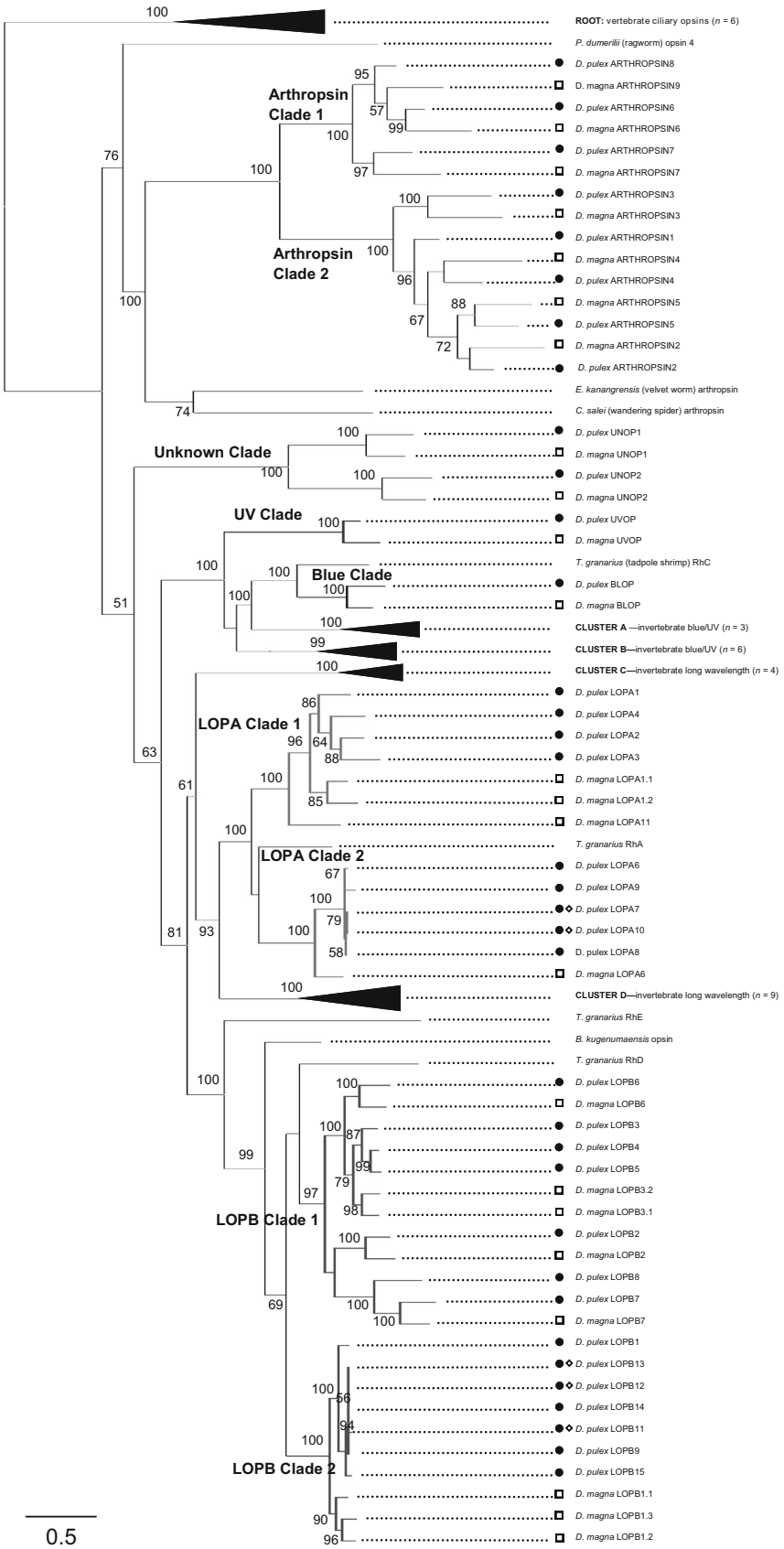

Fig. 1 Phylogeny of *D. pulex*, *D. magna*, and other species rhabdomeric opsins estimated from protein-coding nucleotide sequences using maximum likelihood in RAxML. The tree is rooted by vertebrate ciliary opsins. Bootstrap support values >50% are shown. *Scale bar* shows substitutions/site. *Black closed circles* identify *D. pulex* and *open boxes* identify *D. magna* sequences. *Diamonds* indicate genes for which only partial sequence data are available. Sequence identifiers for non-*Daphnia* clusters are given in Supplementary Table 1

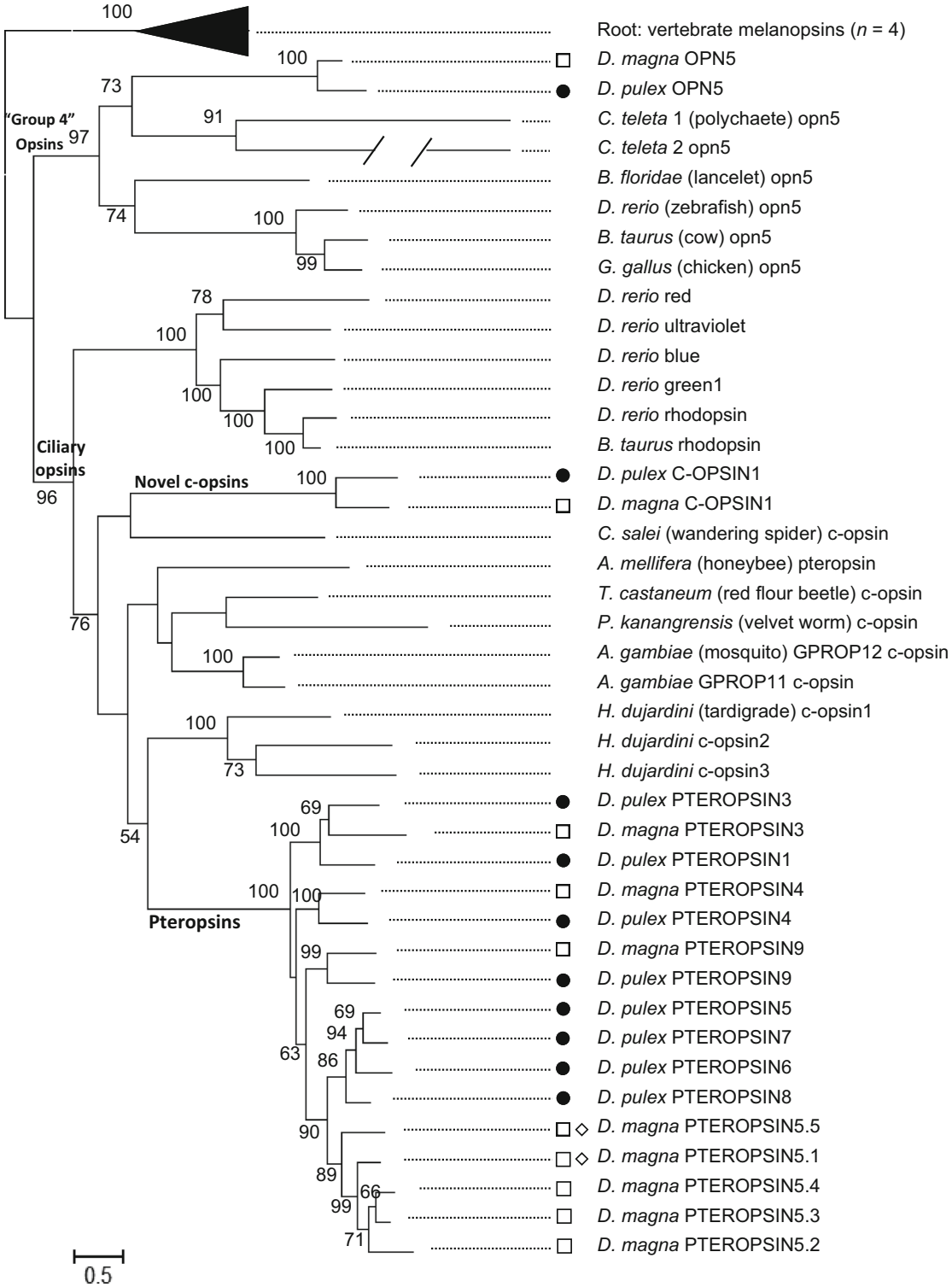

Fig. 2 Phylogeny of *D. pulex, D. magna*, and other species ciliary opsins estimated from protein-coding nucleotide sequences using maximum likelihood in RAxML. The tree is rooted by vertebrate melanopsins (a type of rhabdomeric opsin). Bootstrap support values >50% are shown. *Scale bar* shows substitutions/site. *Black closed circles* identify *D. pulex* and *open boxes* identify *D. magna* sequences. Sequence identifiers for the root cluster are given in Supplementary Table 1

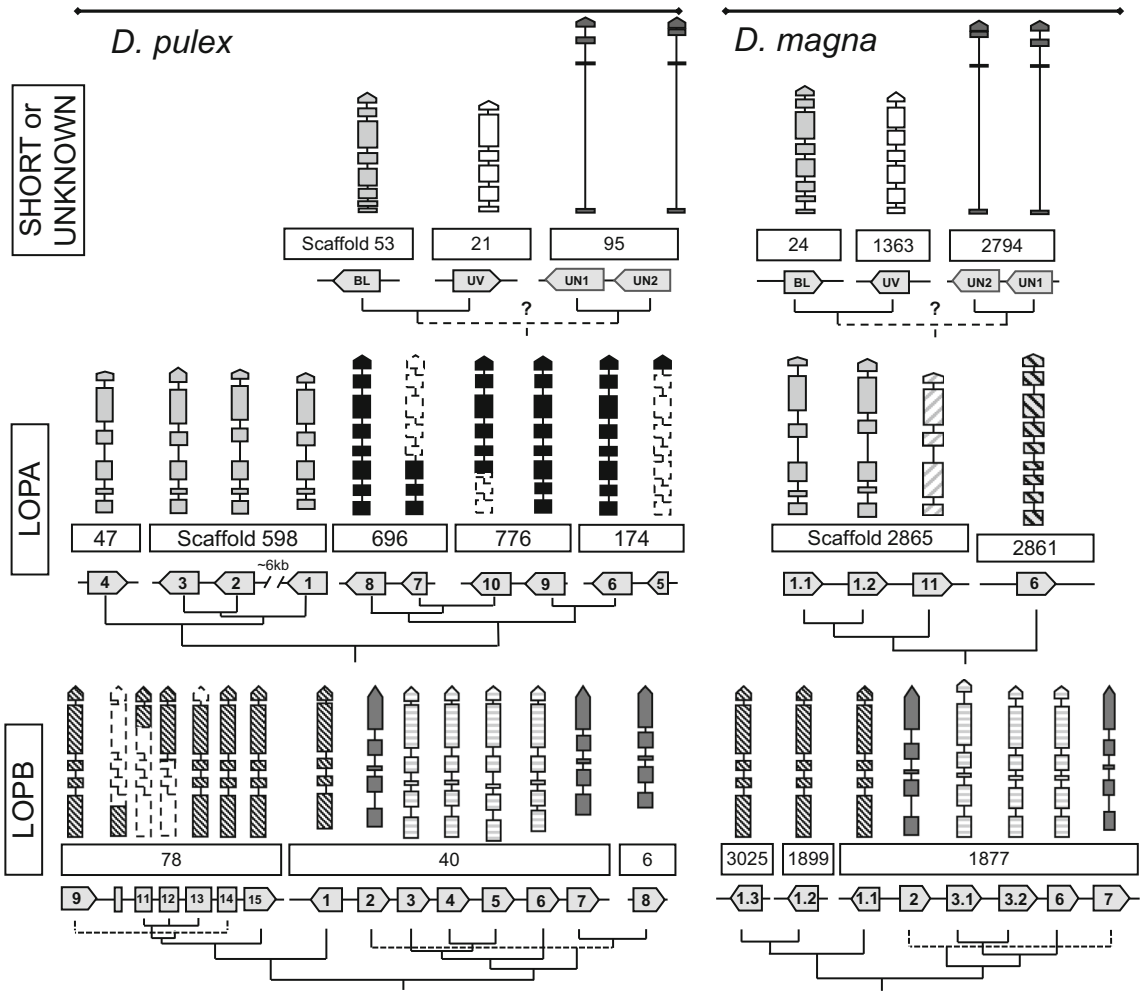

Fig. 3 Gene structure, genome location, and intra-species phylogenetic relationships of *Daphnia* visual opsins. *D. pulex* opsins are illustrated on the left and *D. magna* opsins on the right. *Diagrams* indicate exons with boxes and introns with lines. *Dashed exon* boxes show exons predicted to occur in regions of missing data. Physical scale is approximate, and differs across rows. Differential shading highlights similar gene structures within clades. Scaffold numbers are identified below gene structures. Genes are identified in *shaded boxes* that illustrate physical arrangement within scaffolds. Phylogenetic topology of each clade is mapped onto the genomic locations. *Dashed lines* illustrate branches where inferred evolutionary relationships conflict with spatial relationships

Short-Wavelength Opsins

D. magna and *D. pulex* have orthologous pairs of both the blue- (BLOP) and ultraviolet-sensitive (UVOP) opsins. The intron structure of blue-sensitive opsins (BLOPs) is conserved in *Daphnia* (Fig. 3). Both orthologs share eight exons, and the intron–exon lengths are similar between orthologs. The structure of ultraviolet opsins (UVOPs) is similar between species, but *D. magna* has one intron that is absent from *D. pulex*.

Unknown Opsins

Two pairs of orthologous rhabdomeric opsins with unknown wavelength-sensitivity cluster with the visual opsins (UNOP1 and UNOP2; Fig. 1). In a previous study that examined opsins more broadly in the Arthropoda (Hering and Mayer 2014), these genes clustered with BLOP and UVOP. In our analysis, they are instead sister to all other visual opsins. Both analyses show only moderate support for the connecting node, and thus it is not clear whether these are short-wavelength visual opsins or not. Both pairs of UNOP orthologs share a conserved intron–exon structure, although one intron is notably larger in UNOP1 than in UNOP2. The two UNOPs occur in tandem in each genome.

Long-Wavelength A Opsins—Predicted Green

The long-wavelength A (LOPA; predicted to be maximally sensitive to green light) opsins include two distinct clades that each cluster with 100% bootstrap support (Fig. 1). *D.*

pulex LOPA1-4 and *D. magna* LOPA1.1 and 1.2 form homologous groups derived from the same ancient opsin that expanded independently within their respective lineages. In *D. pulex*, LOPA 1-3 are located in tandem at the end of scaffold 598, and LOPA4 is at the end of scaffold 47, so it is possible that all four are in tandem (Fig. 3). A second ancestral LOPA Clade 1 gene appears to have been lost in *D. pulex*, where there is no direct ortholog for the *D. magna* LOPA11. A third ancient LOPA was the ancestor of LOPA Clade 2, which expanded in *D. pulex* but not *D. magna*.

Intron–exon gene structures provide further evidence supporting two distinct clades within the LOPA genes (Fig. 3). In Clade 1, a six-exon gene structure has been largely conserved in both species. *D. magna* LOPA11 lacks the intron separating exons 2 and 3 in the other members of that clade. In Clade 2, an eight-exon structure is conserved across the *D. pulex* genes with complete sequence information. The single-orthologous opsin in *D. magna* is similar, but has an additional intron. Two *D. pulex* genes that belong to these subclades, LOPA7 and LOPA10, are annotated in the JGI *Daphnia pulex* database as partial duplicates with in-phase start and stop codons. EST libraries in the *Daphnia* genomic database, wfleabase.org, suggest that both of these genes are expressed. The final exon of a third partial gene, LOPA5, appears in the genome sequence, but no specific information on its expression is available. Closer examination of the genome sequence indicated that these three may actually be complete opsin genes, but were annotated as partial duplicates due to missing data.

Long-Wavelength B Opsins—Predicted Red

The long-wavelength B (LOPB; predicted to be maximally sensitive to red light) genes also cluster into two distinct clades, each with >97% bootstrap support (Fig. 1). Many of the LOPB opsins have orthologous pairs between the two species, indicating duplication that predates the divergence of the species. Duplications also occurred in each lineage after their split.

Two smaller clusters comprise LOPB Clade 1, which are also supported by gene structural information (Fig. 3). *D. pulex* LOPB3-6 share a conserved six-exon structure with *D. magna* LOPB6 and LOPB3.1 and 3.2. *D. pulex* and *D. magna* LOPB2, 7, and 8 all share a five-exon structure. All the *D. magna* LOPB Clade 1 opsins are located on scaffold 1877, arrayed in tandem (Fig. 3). Most *D. pulex* LOPB Clade 1 opsins are also arrayed in tandem on scaffold 40, but *D. pulex* LOPB8 is located on scaffold 6 (Fig. 3). LOPB8 is positioned within the scaffold such that it could not be in tandem with the others.

In LOPB Clade 2, expansions in both *D. pulex* and *D. magna* arose from a single common LOPB ancestor (Fig. 1). In *D. pulex,* LOPB11-15 are likely a lineage-specific expansion (Fig. 1; LOPB10 was excluded from phylogenetic analyses due to limited sequence information). LOPB10-14 were all curated in the JGI database as tandem partial duplicates, most of which had in-phase start and stop codons. Closer examination showed that the missing exons were all associated with sections of missing data, and alternate intron–exon boundaries would allow each gene to continue into these sections. For our analyses, we retained sequences for LOPB10-13 as previously curated in the JGI *Daphnia* genome database. For LOPB14, we were able to identify the complete gene sequence, and we recurated the JGI database. All these genes can be found in *Daphnia pulex* EST libraries at wfleabase.org. Despite the partial nature of sequence data for some genes, LOPB Clade 2 is also supported by gene structures, with all complete gene sequences sharing five exons and partial sequences showing matching structure (Fig. 3). Both *D. magna* and *D. pulex* LOPB Clade 2 opsins are located across separate scaffolds in an arrangement that precludes tandem organization in the genome.

Arthropsins

The arthropsins group sister to the visual rhabdomeric opsins and have six orthologous pairs in both species, plus three genes lacking conspecific orthologs (Figs. 1, 4), indicating a total of nine ancestral arthropsins. The arthropsins are clustered into two distinct clades, each with 100% bootstrap support. The clades are mirrored by scaffold locations, with each clade tandemly arrayed in each species (Fig. 4).

A single scaffold contains arthropsins 1–5 in *D. pulex*, where a large intergenic region splits the arthropsins into a tandem pair and a tandem triplet (Fig. 4). These genes, comprising Arthropsin Clade 2, likely derived from five arthropsins already present in the last common ancestor of *Daphnia*. Orthologs of four of these genes are in *D. magna* on scaffold 1036, which has several regions of missing data. One region is positioned where the third member of the tandem triplet would occur, and thus may contain the missing ortholog of Arthropsin1. Although RNAseq information confirms that Arthropsins 2–5 are complete in the *D. magna* genome, sequence gaps in scaffold 1036 render the structures of Arthropsin 2 and 3 uncertain. Part of Arthropsin 3 is on scaffold 1167, which may fit into a gap on 1036, and parts of Arthropsin 2 are on unassembled contigs (23047 and 29514) that may fit into other gaps.

In Arthropsin Clade 1, *D. pulex* and *D. magna* Arthropsin7 and Arthropsin6 form orthologous pairs. The remaining two arthropsins in this clade do not appear to be orthologs. Rather, the analysis suggests reciprocal losses of

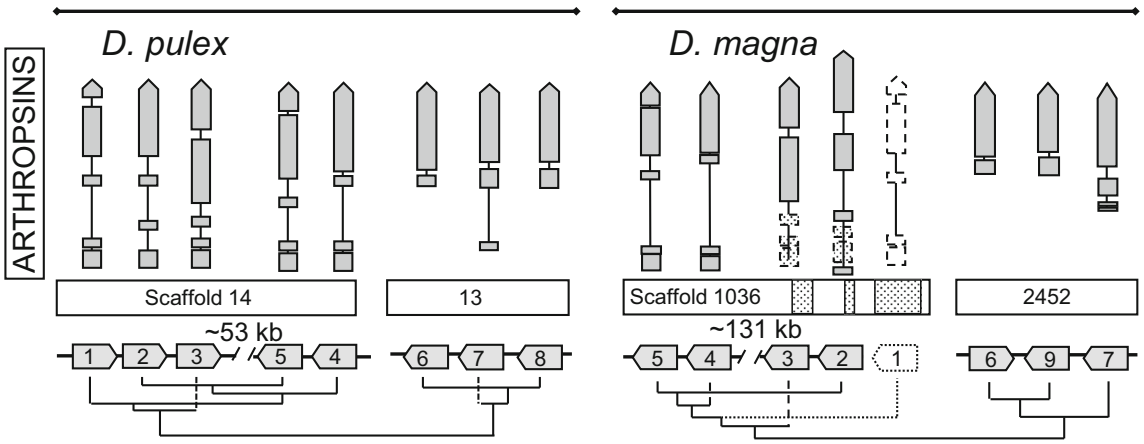

Fig. 4 Gene structure, genome location, and intra-species phylogenetic relationships of *Daphnia* arthropsins. Illustration features are the same as described for Fig. 3. In *D. magna*, stippling shows regions of Scaffold 1036 that are missing data (see text for further explanation). The location of *D. magna* Arthropsin1 is predicted to be in a gap in the genome assembly, and is illustrated with a *dashed-line box*

Arthropsin8 and Arthropsin9 in the two species, implying four members of this clade in the ancestral *Daphnia*. However, the node separating these genes is weak, and data from other species are needed to clarify orthology.

Pteropsins

The *Daphnia* pteropsins form a monophyletic clade among other ciliary opsins (Fig. 2). Phylogenetic analysis shows that the pteropsin sub-family expanded before the *D. pulex*-*D. magna* split and subsequently expanded in each lineage. *D. pulex* pteropsins 5–8 are a sister group to *D. magna* pteropsins 5.1–5.4, but without gene-to-gene orthology (Fig. 2). Incomplete assembly of the *D. magna* genome made it impossible to determine the size of one intron for pteropsin5.1, but the parts of the gene found on different scaffolds (first half on scaffold 1253, second half on the small scaffold 1097) combine to match a single-RNAseq transcript sequence. In *D. magna*, the ortholog of Pteropsin1 appears to have been lost.

Pteropsin2 is a pseudogene in *D. pulex*, and we were unable to reconstruct the gene sequence from available information (Colbourne et al. 2011). We therefore excluded it from analyses. Colbourne et al. (2011) also identified Pteropsin5 as a pseudogene. However, alternate intron–exon boundaries would produce a complete protein with all conserved components of an opsin. We therefore recurated Pteropsin5 and suggest that it is functional.

Most pteropsins share a consistent seven-exon structure (Fig. 5). Pteropsins 7 and 8 in *D. pulex* have an additional intron that divides what would be the sixth exon in two. Unlike the other groups of opsins, pteropsins show great variation in intron length, leading gene prediction algorithims to incorrectly model many of the *D. magna* pteropsins as partial duplicates.

A Novel Ciliary Opsin

We found an additional non-pteropsin ciliary opsin in both *Daphnia* genomes. It diverged deeply from the *Daphnia* pteropsins, but occurs phylogenetically within other invertebrate ciliary opsins (Fig. 2). Intron–exon structure is conserved between the *D. pulex* and *D. magna* orthologs (Fig. 5). To the best of our knowledge, no direct ortholog has been described in the published literature, and we thus have named the gene *c-opsin1*.

Neuropsin

Neuropsins are neither ciliary nor rhabdomeric, but are part of the so-called "Group 4" opsins (Porter et al. 2012). Both *Daphnia* contain a copy of the neuropsin *opsin5*, which together form an orthologous pair that groups with other invertebrate and vertebrate opsin-5 genes (Fig. 2). Both genes consist of several exons separated by large introns (Fig. 5). Until recently, opsin-5 opsins were known only in vertebrates, but sequences have since been described in a few invertebrates (Hering and Mayer 2014). In vertebrates, opsin-5 expression responds to ultraviolet light and is expressed in deep brain and outer ear tissue (Kojima et al. 2011; Nakane et al. 2014), as well as in the neural retina (Yamashita et al. 2010).

Inferred Ancestral Opsins

We inferred the ancestral tree topology for opsins present prior to the *D. magna*-*D. pulex* split (Fig. 6) by concatenating one-to-one orthologs between the species and collapsing species-specific duplications. This shows that the ancestor of all *Daphnia* already had the full complement of arthropsins, five pteropsins, the neuropsin, and the newly identified ciliary opsin. The long-wavelength visual opsins

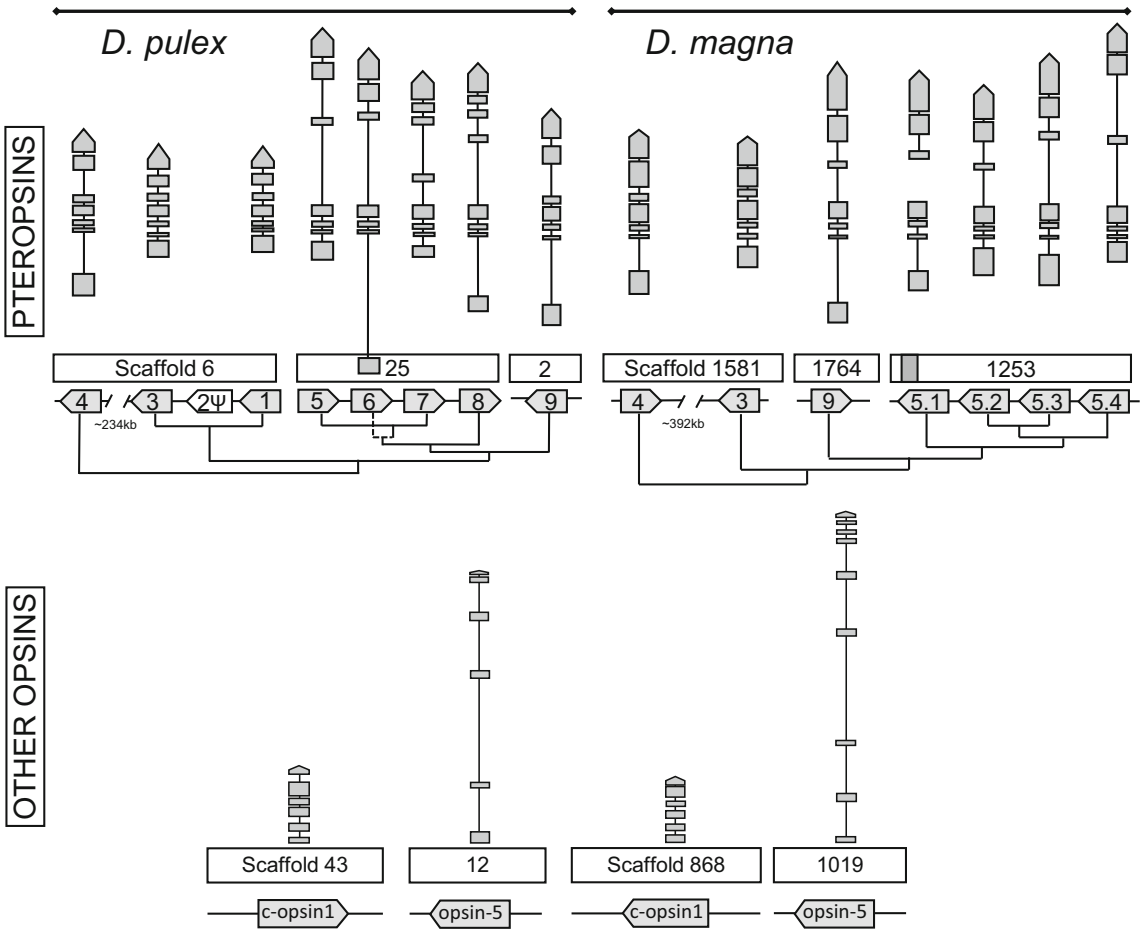

Fig. 5 Gene structure, genome location, and intra-species phylogenetic relationships of *Daphnia* pteropsins, the novel ciliary opsin, and the neuropsin *opsin5*, and their intra-species phylogenetic relationships. Illustration features are the same as described for Fig. 3. Size of the fourth intron of *D. magna* Pteropsin 5.1 could not be determined due to incomplete assembly of the genome. See text for details

had also differentiated substantially, with three genes in the LOPA clade, and six in the LOPB. Blue, UV, and both unknown opsins were also present.

Daphnia Opsin Amino Acid Conservation

We examined the conservation of functionally important amino acids by comparison to the bovine rhodopsin, the first opsin to have its structure determined (Palczewski et al. 2000) and the standard reference protein in opsin biology. The retinal-binding site amino acid, lysine (K296; Palczewski et al. 2000), of transmembrane domain VII (TM VII) is conserved across all phylogenetic clades (Supplementary File 3). It is absent only in genes for which there are missing sequence data covering the location in which it would occur.

We found that the counterion at position 113 (glutamic acid) of bovine rhodopsin TM III, which is conserved across vertebrates, was not conserved in *Daphnia*. Most sequences have tyrosine, phenylalanine, or histidine at that position, which is congruent with other non-vertebrate opsins (Terakita 2005). The counterion 181 (glutamic acid) found in the extracellular loop 2 (EC loop 2) domain is conserved across most *Daphnia* opsin clades (Supplementary File 3). The exception is the LOPB (long wavelength, red-sensitive) clade which contains an aspartic acid at that site. Furthermore, we found that within-clade amino acid conservation was very high in EC loop 2. Among clades, however, EC loop 2 varied in both length and amino acid conservation, with only site 187 (cysteine) conserved across all *Daphnia* opsins. EC loop 2 forms part of the chromophore binding pocket, and thus is directly involved in light capture and signal transduction (Yan et al. 2002).

Sites Involved in Spectral Tuning

Approximately three-fourths of the visual opsins we identified have unique haplotypes for amino acid residues known or inferred to influence spectral tuning, or share

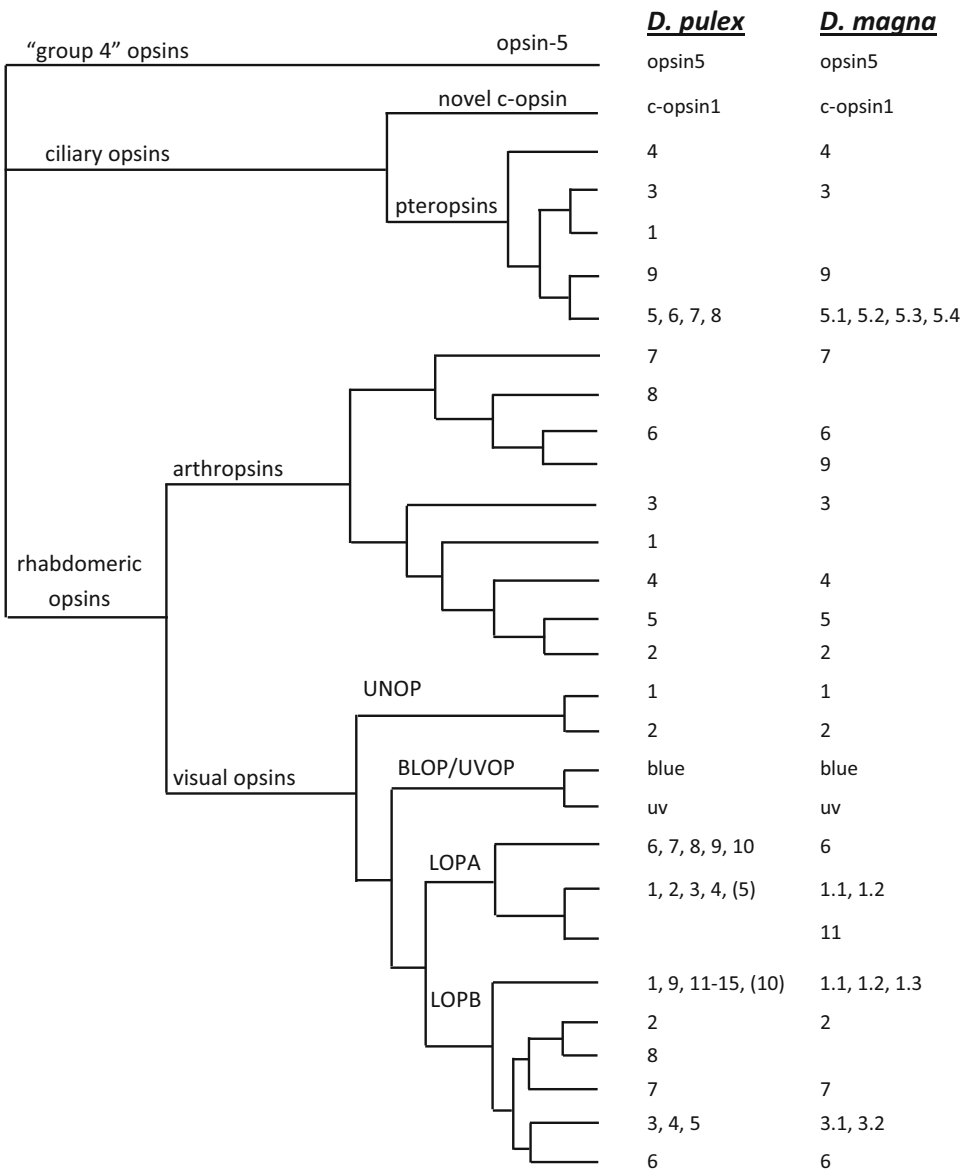

Fig. 6 Inferred phylogeny of the family of opsins present in the most recent common ancestor of all *Daphnia* species (approximately 200 mya; Colbourne and Hebert 1996). Numbers to the right identify which opsins in modern *Daphnia* are descended from each ancestral opsin

their haplotype with only their conspecific ortholog (Table 2). Eighteen different haplotypes were found in the *D. pulex* visual opsins; fourteen haplotypes were found in *D. magna*. Differences among spectral classes are sharp, but there is also substantial variation within wavelength classes, including variation among species-specific duplicates.

The strongest evidence for influence on spectral tuning in invertebrates comes from site-directed mutagenesis in the butterfly *Pieris rapae*, showing that residues at positions 118 and 178 act independently and synergistically to shift spectral tuning (Wakakuwa et al. 2010). Site 118 differs among the wavelength classes, but is largely conserved within them. The exceptions are LOPA9 in *D. pulex*, which has a glycine rather than a serine at site 118, and LOPA7, which has a methionine. Site 178 is a tyrosine in the majority of *Daphnia* visual opsins, but is a phenylalanine in both species' UV opsins, in the *D. magna* blue opsin and LOPA 2.1, and in the *D. pulex* LOPB8. Wakakuwa et al. (2010) showed that a change from tyrosine to phenylalanine shifts spectral tuning by at least 4 nm, depending on the residue at position 118.

Gene Conversion

We found that the LOPA, LOPB, arthropsins, and pteropsins form separate clades and that some genes within each clade are tandem paralogs in each species. Therefore,

Table 2 Variation of residues associated with spectral sensitivity in invertebrate visual opsins

Residue position	90	113	117	118	120	121	122	123	178	186	187	189	207	265	274	292
Domain	TM2	TM3	TM3	TM3	TM3	TM3	TM3	TM3	ECL2	ECL2	ECL2	ECL2	TM4	TM6	TM6	TM7
Evidence Type	PSA	PSA	PSA	M	PSA	PSA	PSA	PSA	M	PSA	PSA	PSA	PSA	PSA	PSA	M
B· taurus rhodopsin	G	E	A	T	G	G	E	I	Y	S	C	I	M	W	Y	A
DpUVOP	T	F	G	A	V	·	P	C	F	·	·	F	L	·	I	·
DmUVOP	T	Y	G	A	V	·	P	C	F	·	·	F	L	·	I	·
DpBLOP	S	F	G	S	N	·	I	G	·	T	·	F	I	·	T	·
DmBLOP	T	F	G	S	N	·	I	G	F	T	·	F	I	·	T	·
DpUNOP1	M	Y	S	G	S	·	T	S	·	·	·	F	F	·	L	·
DmUNOP1	M	Y	S	G	S	·	T	S	·	·	·	F	F	·	L	·
DpUNOP2	M	Y	S	G	T	·	T	T	·	·	·	F	F	·	L	·
DmUNOP2	M	Y	S	G	T	·	T	T	·	·	·	F	F	·	L	·
DpLOPA1	Q	Y	G	S	F	·	L	C	·	·	·	F	L	·	R	·
DpLOPA2	Q	Y	G	S	F	·	L	C	·	·	·	F	L	·	Q	·
DpLOPA3	Q	Y	G	S	F	·	L	C	·	·	·	F	L	·	Q	·
DpLOPA4	Q	Y	G	S	F	·	L	C	·	·	·	F	L	·	Q	·
DmLOPA1.1	Q	Y	G	S	F	·	L	C	·	·	·	F	L	·	Q	·
DmLOPA1.2	Q	Y	G	S	F	·	L	C	·	·	·	F	L	·	Q	·
DmLOPA11	Q	Y	·	S	F	·	L	G	F	·	·	F	V	·	Q	·
DpLOPA6	Q	Y	G	S	F	·	A	T	·	·	·	F	L	·	K	·
DpLOPA7	Q	Y	G	M												
DpLOPA8	Q	Y	G	S	F	·	A	T	·	·	·	F	L	·	K	·
DpLOPA9	Q	Y	G	G	F	·	A	T	·	·	·	F	L	·	K	S
DpLOPA10	Q	Y	G	S	F	·	A	T	·	·	·	F	L	·	K	·
DmLOPA6	Q	Y	G	S	F	·	A	T	·	·	·	F	L	·	Q	·
DpLOPB1	L	H	G	A	F	·	Y	N	·	T	·	Y	A	·	M	V
DpLOPB9	L	H	G	A	F	·	Y	N	·	T	·	Y	A	·	M	V
DpLOPB11															M	V
DpLOPB12										T	·	Y	A	·	M	V
DpLOPB13	L	H	G	A	F	·	Y	N	·	T	·	Y	A	·	M	V
DpLOPB14	L	H	G	A	F	·	Y	N	·	T	·	Y	A	·	M	V
DpLOPB15	L	H	G	A	F	·	Y	N	·	T	·	Y	A	·	V	V
DmLOPB1.1	L	H	G	A	F	·	Y	N	·	T	·	Y	A	·	M	V
DmLOPB1.2	L	H	G	A	F	·	Y	N	·	T	·	Y	A	·	M	V
DmLOPB1.3	L	H	G	A	F	·	Y	N	·	T	·	Y	A	·	L	V
DpLOPB2	L	H	G	A	M	·	Y	N	·	T	·	F	A	·	L	I
DmLOPB2	L	H	G	A	F	·	Y	N	·	T	·	F	A	·	L	I
DpLOPB7	L	H	G	A	F	·	Y	S	·	T	·	F	S	V	V	I
DmLOPB7	L	H	G	A	F	·	Y	G	·	T	·	F	S	V	V	I
DpLOPB8	L	H	G	A	F	·	Y	S	F	T	·	F	S	·	I	I
DpLOPB3	L	H	G	A	C	·	Y	S	·	T	·	F	C	·	I	V
DpLOPB4	L	H	G	A	C	·	Y	S	·	T	·	F	T	·	I	V
DpLOPB5	L	H	G	A	C	·	Y	S	·	T	·	F	S	·	L	V
DmLOPB3.1	L	H	G	A	C	·	Y	S	·	T	·	F	C	·	L	V
DmLOPB3.2	L	H	G	A	C	·	Y	S	·	T	·	F	T	·	L	V
DpLOPB6	L	H	G	A	C	·	Y	S	·	T	·	F	T	C	L	V
DmLOPB6	L	H	G	A	C	·	Y	S	·	T	·	F	S	·	L	V

Residue position indicates the position in the bovine rhodopsin. Domain indicates whether the residue is found in an extracellular loop (ECL) or a transmembrane (TM) domain. Evidence type indicates whether the role in spectral sensitivity of the residue is inferred from phylogenetic selection analyses (PSA), or known from mutant analysis (M). See text for relevant citations. Amino acids are given in the standard single-letter code. Dots represent positions identical to the reference protein, blanks indicate missing data. Sequence names in bold indicate a unique haplotype within that species. Lines separate clades of genes descended from a distinct ancestral opsin in the most recent common ancestor of all *Daphnia*

we investigated the occurrence of gene conversion in these clades of opsins using GENECONV (Sawyer 1989) to determine the extent to which conversion could constrain functional differentiation. After correcting for multiple comparisons, we found 29 statistically significant segments of gene conversion in our dataset (Supplementary Table 4). All the gene pairs with a gene conversion fragment are also linked on the same scaffold. Gene conversion was not

distributed evenly across the clades of opsins; a large majority of the observed conversions occurred in *D. pulex* arthropsins, a few in LOPA genes, and the remaining opsin clades had none.

Twenty-three of the observed conversions were between pairs of arthropsins in *D. pulex*, most of which involved ~180 bp that code for the TM1/intracellular loop 1/TM2 domains, or ~150 bp that code for TM7 and part of the N-terminus. Additionally, 2 segments of gene conversion were found between arthropsin genes of *D. magna*. TM7 includes the retinal-binding site (K296) essential for light capture (Terakita 2005) and one of the residues known to influence spectral sensitivity in visual opsins (Table 2). Gene pairs with conversion fragments in *D. pulex* were always among Arthropsin 1–5 (on scaffold 14) or among Arthropsin 6–8 (on scaffold 13). Similarly, conversions in *D. magna* arthropsins occurred either within scaffold 1036 or within scaffold 2452. This further supports the inference that each species contains two unlinked clusters of tandemly arrayed arthropsins.

Four gene conversion segments were identified among the LOPA genes; one between two *D. magna* LOPA genes and three fragments between *D. pulex* genes.

Discussion

Our analyses revealed that the expansive suite of opsins present in *D. pulex* is also characteristic of *D. magna*, albeit to a lesser degree. We found fewer opsins in *D. magna* (32) than in *D. pulex* (48). Both *Daphnia* lineages have maintained many complete opsin genes in a variety of different clades of opsins, supporting the hypothesis that individual genes have differentiated functional roles in photoreception and vision. If functions were not differentiated, we would expect mutational erosion of functionally redundant gene copies (Lynch and Conery 2000) over the 200MY since *D. pulex* and *D. magna* diverged (Colbourne and Hebert 1996). Assuming a base substitution rate of 4×10^{-9} per site per generation in *Daphnia* (Keith et al. 2016), a conservative estimate of five generations per year, and that nonfunctionalization would require mutations to ~10% of the nucleotides in a gene, nonfunctionalization of a superfluous opsin would occur in only five million years. This is in line with the estimate of four million years for the average half-life of a eukaryotic gene duplicate (Lynch and Conery 2000).

In general, we found more evidence for ancient functional diversification than we anticipated—rather than one gene for each color of visual opsins, and one for each category of non-visual opsins, we found that the long-wavelength visual opsins, arthropsins, and pteropsins were already diverse in the last common ancestor of all *Daphnia*. Largely, this diversity has been preserved throughout prolonged evolution, with further expansion in the long-wavelength visual opsins and pteropsins. This diversity suggests that photoreception is involved in a variety of physiological processes beyond vision in *Daphnia* and potentially that spectral sensitivity is important to those processes. Given that *Daphnia* are transparent organisms, it is easy to imagine that any tissue could have a light-dependent aspect to its function.

Strikingly, recent duplications were not distributed randomly among the ancient opsins (Fig. 6). Only five ancient opsins led to recent, lineage-specific duplications. Four of these opsins duplicated in both *Daphnia* species, independently. Nodes supporting the separate expansion in each species are all strong (>85%), making it unlikely that the congruence between species is a consequence of phylogenetic error. The probability of this degree of congruence occurring randomly across the 29 opsins inferred to be present in the last common ancestor of *Daphnia* is 0.0002. One potential explanation is that parallel duplications occur because specific opsins are located in duplication hotspots. However, several sets of recently duplicated genes are located near other opsins that have not been recently duplicated. An alternative explanation is that duplication is widespread, but duplicates are retained for only specific genes where additional copies are advantageous or become advantageous through diversification.

Opsins and Diversification of *Daphnia* Vision

Both *Daphnia pulex* and *D. magna* are pond-dwelling species, which is likely the case for the ancestral *Daphnia* since most extant species inhabit ponds (Benzie 2005; Colbourne et al. 1997). Many ponds have high concentrations of colored dissolved organic matter (CDOM), which preferentially absorbs green–blue-UV light, thus creating a red–orange dominated light environment. While we are not advancing an ecological link per se between ponds and the expansion and maintenance of *Daphnia* long-wavelength opsins, the potential for an association stands out as something worth further investigation across other species of *Daphnia*.

Extracellular electrophysiological work has demonstrated that *D. magna* has four peaks of wavelength sensitivity in its compound eye (Smith and Macagno 1990), and the animals respond behaviorally to light with widely different wavelengths (Smith and Baylor 1953; Young et al. 1984). Barlow (1982) suggested that four classes of photoreceptors may provide enough information to decode color information in their environment, and this view has been emphasized recently (Marshall and Arikawa 2014, Marshall et al. 2015). However, four may be inadequate if

parts of the spectrum are unavailable in a particular habitat. Different opsins can be expressed within a single photoreceptor (Sakamoto et al. 1996); thus, one possibility is that multiple long-wavelength opsins with similar, but offset, spectral sensitivities broaden the spectral sensitivity of the photoreceptor (Arikawa 2003). Furthermore, most areas of lakes and ponds will have only dim light, and it may therefore be important to produce opsins that are finely tuned for capturing the available light spectra in order to discriminate color. A blue opsin is simply useless once all blue light has been absorbed; color discrimination may then depend on having several somewhat different green or red sensitivities. For example, gradations of green and red sensitivity may be important if *Daphnia* are using vision to sense micropatches of different types of algae, as we suggested elsewhere (Brandon et al. 2015).

Instances of opsin duplication have led to the evolution of different wavelength sensitivities in visual pigments (Frentiu et al. 2007; Hofmann and Carleton 2009), and the potential for differences in wavelength sensitivity of *Daphnia* visual pigments is supported by our evidence that residues important to spectral sensitivity differ among opsins within both the red and green spectral classes (Table 2). Within the LOPA (putatively green) visual opsins, the three gene copies that were present prior to the *D. pulex—D. magna* split have descendants with distinct haplotypes for residues known to influence spectral sensitivity. Similarly, the six ancient LOPB (putatively red) visual opsins had five, and possibly six, distinct haplotypes at these residues (Table 2; Fig. 6). Thus, the ancestor of all *Daphnia* had at least 11 visual opsins (six red, three green, one blue, and one UV), each potentially with different spectral sensitivities. Since UNOP1 and UNOP2 also differ at these residues in each species, if these are visual opsins, the ancient *Daphnia* would have had 13 spectrally different visual opsins.

Recent duplications are much more likely to produce opsins with identical spectral sensitivity haplotypes, but some lineage-specific duplications show differentiation. At least two recent duplications have led to new haplotypes within the green-like clade *D. pulex*. LOPA9 is distinct from LOPA6, 7, 8, and 10, and LOPA1 is distinct from LOPA2, 3, and 4. LOPA 7 may also be distinct, but that interpretation is subject to confirmation when the full gene sequence becomes known. In the red-like clade of visual opsins, recent duplications have produced unique haplotypes in LOPB15 and each of LOPB3, 4, and 5 for *D. pulex*, and betwen LOPB3.1 and 3.2 in *D. magna.* All told, for genes with complete data, 7 out of 18 lineage-specific duplications have led to differentiation at the residues in question (Table 2). Because the determinants of spectral sensitivity are incompletely known, our analysis may underestimate the potential for diversification of spectral sensitivity.

Opsin differentiation may also involve expression location or timing. Oakley and Huber (2004) showed that six of the eight opsins in the ostracod *Skogsberia lerneri* were expressed only in the compound eye; the other two only in the simple eye. Similarly, *Daphnia* possess both a compound eye and a simple eye. A comparable distinction may occur in *Daphnia*, with different opsins localized to one eye. The red-like, green-like, and short-wavelength opsins are each divided into two deeply split clades, and these divisions could reflect separation of compound-expressed and simple-expressed opsins. Timing of expression may also differ, either developmentally or in response to environmental cues such as diel or seasonal cycles.

The ancestral *Daphnia* species contained both a putative ultraviolet- and blue-sensitive opsin (Fig. 6). It is notable that these have not duplicated in either species given the substantial duplications of other opsins (Fig. 1). While short-wavelength light may be almost absent from high-CDOM habitats, these wavelengths are important to many *Daphnia*. Lakes around the world are dominated by blue light, so many species are found in primarily blue environments. These species may have undergone expansion in their short-wavelength opsins. UV light is the most thoroughly studied color of light with respect to *Daphnia* behavior and ecology, and light sensitivity in this range is critical for fitness in some *Daphnia* (Hessen et al. 1999; Rhode et al. 2001; Miner and Kerr 2011). An analysis of the history of opsin duplication across fish (Rennison et al. 2012) also found that long-wavelength opsins were more likely to duplicate or be retained than were short-wavelength opsins. Without information on the light environment in which the specific fish taxa live, we cannot know if these duplications are consistent with the hypothesis that opsin duplicates are generally retained on the basis of spectral quality of the light environment.

The unknown-wavelength rhabdomeric opsins duplicated before the *Daphnia* species radiation, and two paralogs have been maintained in each *Daphnia* lineage. In a previous report, these opsins clustered among other arthropod short-wavelength opsins (Hering and Mayer 2014). One possibility is that they provide diversity in sensitivity to UV/blue light or even shorter wavelengths. However, our analysis places them sister to all visual opsins (Fig. 1), and they differ from the blue and UV opsins at nearly all residues thought or known to influence spectral tuning in invertebrates (Table 2). Experimental work will be needed to determine their wavelength sensitivity, and if they are indeed expressed in visual photoreceptors.

Arthropsins

Arthropsins expanded early in *Daphnia* evolution, hinting that there may be multiple arthropsins in other crustaceans. Hering and Mayer (2014) identified several additional arthropsin sequences in other taxa that were not previously recognized as arthropsins (Koyanagi et al. 2005; Randel et al. 2013) and further exploration may find others.

Empirical information on arthropsin function is almost non-existent. Eriksson et al. (2013) found arthropsin expression in the central nervous tissue of a spider and in the brain of a velvet worm. The distinct locations suggest that different arthropsins may have divergent functions. Given that seven ancient arthropsins have persisted in the *Daphnia* genus, we predict that the different genes in *Daphnia* have distinct functions. With respect to amino acid residues known to influence spectral sensitivity, we found four different haplotypes in *D. pulex* and six in *D. magna* (Supplemental Table 5). However, we also found evidence for concerted evolution via gene conversion among arthropsins located tandemly on the same scaffold (Supplemental Table 4), suggesting strong conservation of at least some aspects of functionality across different arthropsins. Thus, we further predict that functional divergence will be greater among arthropsins that are not in tandem array. If opsins form dimers in native membranes (Liang et al. 2003), one explanation for preserving similarity but not identity may be that tandem *Daphnia* arthropsin genes have become necessary subunits of an arthropsin dimer.

Pteropsins

Pteropsins are ciliary opsins first described in the honeybee *Apis mellifera*, and may play a role in circadian rhythm entrainment (Velarde et al. 2005). Further work by Koyanagi et al. (2013) has determined that these proteins have peak wavelength sensitivities ranging from blue to green. Such spectral sensitivities would not be particularly useful in the red-dominated light environment of high-CDOM ponds. The *Daphnia* pteropsins form a monophyletic group within the genus, suggesting that the expansion of this clade occurred after crustaceans diverged from other arthropods (Fig. 2). Because the most recent common ancestor of *Daphnia* likely contained five pteropsins (Fig. 6), this clade may have expanded early in cladoceran evolutionary history and possibly during early crustacean evolution. More recent pteropsin duplications occurred in both *D. magna* and *D. pulex* lineages. All the ancient duplications led to variation at sites known to influence spectral sensitivity, while none of the recent duplications did so (Supplemental Table 5). Pteropsins stand out as having unusually large variation in the extracellular loop 2 domain, which forms the critical chromophore binding pocket. From a *Daphnia*—and indeed a broader zooplankton—perspective, the potential role of pteropsins in circadian rhythm mediation is worth investigating further because the ecologically important diel vertical migration behaviors of *Daphnia* are partially influenced by the circadian clock (reviewed in Cohen et al. 2009).

Mechanisms of Opsin Evolution

Tandem duplication appears to be the usual type of duplication, with at least 15 out of 22 recent duplication events and 13 out of 16 ancient duplication events (within functional categories) occuring in tandem. The majority of remaining duplications cannot be determined due to incomplete assemblies, but the origin of LOBP8 in *D. pulex* appears to have not been in tandem. This duplicate may have arisen through transposition, or been subjected to non-duplicative transposition following its origin.

Gene conversion did not constrain diversification outside of the arthropsins. This suggests, particularly for the visual opsins, that multiple copies of opsins are not maintained due to the need to rapidly upregulate transcript abundance of functionally equivalent opsins.

Intron–exon structures can be preserved for enormous time-spans—even across taxonomic kingdoms (Rogozin et al. 2003). Our data show that intron–exon evolution generally reinforces the phylogenetic topology inferred from sequence analysis, with single genes differing in presence/absence of an intron. The pteropsins show the greatest variation of intron size, with several introns ranging from fewer than 100 bp to much more than 1000 bp in different pteropsins. There is no apparent tendency toward larger or smaller introns in the two species. In most cases, we could not determine the evolutionary direction of intron presence–absence across genes, but these cases never involved conversion of a sequence between intron- and exon-status.

The prevalence of gene conversion in the arthropsins, intron expansion in the pteropsins, and recent tandem duplication in the long-wavelength visual opsins shows that different mechanisms of evolution have played important roles in diversification of different types of opsins. The incomplete assembly of the *Daphnia* genomes makes it difficult to suggest whether this mechanistic evolution is associated with physical geography or functional differences of the genes involved.

Acknowledgements We appreciate discussion during this project and commentary on earlier drafts of this manuscript from Dan Speiser, Todd Oakley, Sally Woodin, and Axios. The *Daphnia* Genomics Consortium provided access to the data on which this manuscript is based. CSB was supported by a U.S. Department of Education GAANN Fellowship. *D. magna* sequence data were produced by The Center for Genomics and

Bioinformatics at Indiana University and distributed via wFleaBase in collaboration with the *Daphnia* Genomics Consortium at http://daphnia.cgb.indiana.edu. *D. magna* sequencing was supported in part by National Institutes of Health award 5R24GM078274 "Daphnia Functional Genomics Resources."

References

Arikawa K (2003) Spectral organization of the eye of a butterfly, Papilio. J Comp Physiol A 189:791–800

Barlow HB (1982) What causes trichromacy? A theoretical analysis using comb-filtered spectra. Vis Res 22:635–643

Benzie JAH (2005) The Genus *Daphnia* (including *Daphniopsis*): *(Anomopoda, Daphniidae*). Kenobi Productions, Ghent

Birney E, Clamp M, Durbin R (2004) GeneWise and genomewise. Genome Res 14:988–995

Eriksson BJ, Fredman D, Steiner G, Schmid A (2013) Characterisation and localisation of the opsin protein repertoire in the brain and retinas of a spider and an onychophoran. BMC Evol Biol 13:186

Brandon CS, Dudycha JL (2014) Ecological constraints on sensory systems: compound eye size in *Daphnia* is reduced by resource limitation. J Comp Physiol A 200:749–758

Brandon CS, James T, Dudycha JL (2015) Selection on incremental variation of eye size in a wild population of *Daphnia*. J. Evol Biol 28:2112–2118

Briscoe AD (2001) Functional diversification of lepidopteran opsins following gene duplication. Mol Biol Evol 18:2270–2279

Briscoe AD (2002) Letter to the editor homology modeling suggests a functional role for parallel amino acid substitutions. Mol Biol Evol 19:983–986

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K et al (2009) BLAST+: architecture and applications. BMC Bioinform 10:421

Cohen JH, Richard B, Forward RB (2009) Zooplankton diel vertical migration—a review of proximate control. Oceanogr Mar Biol 47:77–110

Colbourne JK, Hebert PDN (1996) The systematics of North American *Daphnia* (Crustacea: Anomopoda): a molecular phylogenetic approach. Philos Trans R Soc Lond B 351:349–360

Colbourne JK, Hebert PDN, Taylor DJ (1997) Evolutionary origins of phenotypic diversity in *Daphnia*. In: Givnish TJ, Sytsma KJ (eds) Molecular evolution and adaptive radiation. Cambridge University Press, Cambridge, pp 163–188

Colbourne JK, Pfrender ME, Gilbert D, Thomas WK, Tucker A, Oakley TH et al (2011) The ecoresponsive genome of *Daphnia pulex*. Science 331:555–561

De Meester L (1991) An analysis of the phototactic behaviour of *Daphnia magna* clones and their sexual descendants. Hydrobiologia 225:217–227

De Meester L (1993) Genotype, fish-mediated chemical, and phototactic behavior in *Daphnia magna*. Ecology 74:1467–1474

Feuda R, Rota-Stabelli O, Oakley TH, Pisani D (2014) The comb jelly opsins and the origins of animal phototransduction. Genome Biol Evol 6(8):1964–1971

Frentiu FD, Bernard GD, Sison-Mangus MP, Van Zandt Brower A, Briscoe AD (2007) Gene duplication is an evolutionary mechanism for expanding spectral diversity in the long-wavelength photopigments of butterflies. Mol Biol Evol 24:2016–2028

Hamza W, Ruggiu D (2000) Swimming behaviour of *Daphnia galeata x hyalina* as a response to algal substances and to opaque colours. Int Rev Hydrobiol 85:157–166

Hering L, Mayer G (2014) Analysis of the opsin repertoire in the tardigrade Hypsibius dujardini provides insights into the evolution of opsin genes in panarthropoda. Genome Biol Evol 6:2380–2391

Hessen DO, Borgeraas J, Kessler K, Refseth UH (1999) UV-B susceptibility and photoprotection of Arctic *Daphnia* morphotypes. Polar Res 18:345–352

Hofmann CM, Carleton KL (2009) Gene duplication and differential gene expression play an important role in the diversification of visual pigments in fish. Integr Comp Biol 49:630–643

Keith N, Tucker AE, Jackson CE, Sung W, Lucas Lledo JI, Schrider DR, Schaack S, Dudycha JL, Shaw J, Lynch M (2016) High mutational rates of large-scale duplication and deletion in *Daphnia pulex*. Genome Res 26(1):60–69

Kojima D, Mori S, Torii M, Wada A, Morishita R, Fukada Y (2011) UV-sensitive photoreceptor protein OPN5 in humans and mice. PLOS ONE 6(10):e26388

Kotov AA, Taylor DJ (2011) Mesozoic fossils (>145 Mya) suggest the antiquity of the subgenera of *Daphnia* and their coevolution with chaoborid predators. BMC Evol Biol 11:129

Koyanagi M, Kubokawa K, Tsukamoto H, Shichida Y, Terakita A (2005) Cephalochordate melanopsin: evolutionary linkage between invertebrate visual cells and vertebrate photosensitive retinal ganglion cells. Curr Biol 15:1065–1069

Koyanagi M, Takada E, Nagata T, Tsukamoto H, Terakita A (2013) Homologs of vertebrate Opn3 potentially serve as a light sensor in nonphotoreceptive tissue. Proc Natl Acad Sci USA 110:4998–5003

Larkin MA, Blackshields G, Brown NP, Chenna R, Mcgettigan PA, McWilliam H et al (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–2948

Lee YJ, Shah S, Suzuki E, Zars T, O'Day PM, Hyde DR (1994) The *Drosophila* dgq gene encodes a G α protein that mediates phototransduction. Neuron 13(5):1143–1157

Lewis A, Marcus MA, Ehrenberg B, Crespi H (1978) Experimental evidence for secondary protein-chromophore interactions at the Schiff base linkage in bacteriorhodopsin: molecular mechanism for proton pumping. Proc Natl Acad Sci USA 75:4642–4646

Liang Y, Fotiadis D, Filipek S, Saperstein DA, Palczewski K, Engel A (2003) Organization of the G protein-coupled receptors rhodopsin and opsin in native membranes. J Biol Chem 278:21655–21662

Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science 290:1151–1155

Marshall J, Arikawa K (2014) Unconventional colour vision. Curr Biol 24:1150–1154

Marshall J, Carleton KL, Cronin T (2015) Colour vision in marine organisms. Curr Opin Neurobiol 34:86–94

Miner BE, Kerr B (2011) Adaptation to local ultraviolet radiation conditions among neighbouring *Daphnia* populations. Proc R Soc Lond B 278:1306–1313

Nakane Y, Shimmura T, Abe H, Yoshimura T (2014) Intrinsic photosensitivity of a deep brain photoreceptor. Curr Biol 24:R596–R597

Nathans J (1987) Molecular biology of visual pigments. Annu Rev Neurosci 10:163–194

Oakley TH, Huber DR (2004) Differential expression of duplicated opsin genes in two eyetypes of ostracod crustaceans. J Mol Evol 59:239–249

Ohta T (2010) Gene conversion and evolution of gene families: an overview. Genes (Basel) 1:349–356

Omilian A, Cristescu MEA, Dudycha JL, Lynch M (2006) Ameiotic recombination in asexual lineages of *Daphnia*. Proc Natl Acad Sci 103:18638–18643. PMC1693715. Erratum: PNAS 104: 2554

Palczewski K, Kumasaka T, Hori T, Behnke CA, Fox BA, Trong I et al (2000) Crystal structure of rhodopsin: a G protein-coupled receptor. Science 289:739–745

Porter ML, Cronin TW, McClellan DA, Crandall KA (2007) Molecular characterization of crustacean visual pigments and the evolution of pancrustacean opsins. Mol Biol Evol 24:253–268

Porter ML, Bok MJ, Robinson PR, Cronin TW (2009) Molecular diversity of visual pigments in Stomatopoda (Crustacea). Vis Neurosci 26:255–265

Porter ML, Blasic JR, Bok MJ, Cameron EG, Pringle T, Cronin TW et al (2012) Shedding new light on opsin evolution. Proc R Soc B 279:3–14

Randel N, Bezares-Calderón LA, Gühmann M, Shahidi R, Jékely G (2013) Expression dynamics and protein localization of rhabdomeric opsins in platynereis larvae. Integr Comp Biol 53:7–16

Rennison DJ, Owens GL, Taylor JS (2012) Opsin gene duplication and divergence in ray-finned fish. Mol Phyl Evol 62:986–1008

Rhode SC, Pawlowski M, Tollrian R (2001) The impact of ultraviolet radiation on the vertical distribution of zooplankton of the genus *Daphnia*. Nature 412:69–72

Rogozin IB, Wolf YI, Sorokin AV, Mirkin BG, Koonin EV (2003) remarkable interkingdom conservation of intron positions and massive, lineage-specific intron loss and gain in eukaryotic evolution. Curr Biol 13:1512–1517

Rosenbaum DM, Rasmussen SGF, Kobilka BK (2009) The structure and function of G-protein-coupled receptors. Nature 459:356–363

Sakamoto K, Hisatomi O, Tokunaga F, Eguchi E (1996) Two opsins from the compound eye of the crab Hemigrapsus sanguineus. J Exp Biol 199:441–450

Salcedo E, Zheng L, Phistry M, Bagg EE, Britt SG (2003) Molecular basis for ultraviolet vision in invertebrates. J Neurosci 23:10873–10878

Salcedo E, Farrell DM, Zheng L, Phistry M, Bagg EE, Britt SG (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

Sawyer S (1989) Statistical tests for detecting gene conversion. Mol Biol Evol 6:526–538

Shichida Y, Matsuyama T (2009) Evolution of opsins and phototransduction. Philos Trans R Soc Lond B 364:2881–2895

Smith FE, Baylor ER (1953) Color responses in the cladocera and their ecological significance. Am Nat 87:49–55

Smith KC, Macagno ER (1990) UV photoreceptors in the compound eye of Daphnia magna (Crustacea, Branchiopoda). A fourth spectral class in single ommatidia. J Comp Physiol A 166:587–606

Speiser DI, Pankey MS, Zaharoff AK, Battelle BA, Bracken-grissom HD, Breinholt JW et al (2014) Using phylogenetically-informed annotation (PIA) to search for light-interacting genes in transcriptomes from non-model organisms. BMC Bioinform 15:1–12

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313

Stearns SC (1975) Light responses of *Daphnia pulex*. Limnol Oceanogr 20:564–570

Storz UC, Paul RJ (1998) Phototaxis in water fleas (*Daphnia magna*) is differently influence by visible and UV light. J Comp Phys A 183:709–717

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–2729

Terakita A (2005) Protein family review the opsins. Genome Biol 6:1–9

Terakita A, Hariyama T, Tsukahara Y, Katsukura Y, Tashiro H (1993) Interaction of GTP-binding protein Gq with photoactivated rhodopsin in the photoreceptor membranes of crayfish. FEBS Lett 330(2):197–200

Tierney SM, Cooper SJB, Saint KM, Bertozzi T, Hyde J, Humphreys WF et al (2015) Opsin transcripts of predatory diving beetles: a comparison of surface and subterranean photic niches Subject Category: subject Areas: Authors for correspondence: R. Soc Open Sci 2:140386

Velarde RA, Sauer CD, Walden KK, Fahrbach SE, Robertson HM (2005) Pteropsin: a vertebrate-like non-visual opsin expressed in the honey bee brain. Insect Biochem Mol Biol 35(12):1367–1377

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

Yamashita T, Ohuchi H, Tomonari S, Ikeda K, Sakai K, Shichida Y (2010) Opn5 is a UV-sensitive bistable pigment that couples with Gi subtype of G protein. Proc Natl Acad Sci USA 107:22084–22089

Yan EC, Kazmi MA, De S, Chang BS, Seibert C, Marin EP,… Sakmar TP (2002) Function of extracellular loop 2 in rhodopsin: glutamic acid 181 modulates stability and absorption wavelength of metarhodopsin II. Biochemistry 41(11): 3620–3627

Young S, Taylor VA, Watts E (1984) Visual factors in *Daphnia* feeding. Limnol Oceanogr 29:1300–1308