

Submission is intended as Research Article

Analysis of the Opsin Repertoire in the Tardigrade *Hypsibius dujardini* Provides Insights into the Evolution of Opsin Genes in Panarthropoda

Lars Hering^{1,*}, Georg Mayer¹

¹Animal Evolution and Development, Institute of Biology, University of Leipzig,
Talstraße 33, D-04103 Leipzig, Germany

*Author for Correspondence: Lars Hering, Animal Evolution and Development,
Institute of Biology, University of Leipzig, Talstraße 33, D-04103 Leipzig, Germany,
+493419736766, lars.hering@uni-leipzig.de

Abstract

Screening of a deeply sequenced transcriptome using Illumina sequencing as well as the genome of the tardigrade *Hypsibius dujardini* revealed a set of five opsin genes. To clarify the phylogenetic position of these genes and to elucidate the evolutionary history of opsins in Panarthropoda (Onychophora + Tardigrada + Arthropoda), we reconstructed the phylogeny of broadly sampled metazoan opsin genes using maximum likelihood and Bayesian inference methods in conjunction with carefully selected substitution models. According to our findings, the opsin repertoire of *H. dujardini* comprises representatives of all three major bilaterian opsin clades, including one r-opsin, three c-opsins, and a Group 4 opsin (neuropsin/opsin-5). The identification of the tardigrade ortholog of neuropsin/opsin-5 is the first record of this opsin type in a protostome, but our screening of available metazoan genomes revealed that it is also present in other protostomes. Our opsin phylogeny further suggests that two r-opsins, including an “arthropsin”, were present in the last common ancestor of Panarthropoda. While both r-opsin lineages were retained in Onychophora and Arthropoda, the “arthropsin” was lost in Tardigrada. The single (most likely visual) r-opsin found in *H. dujardini* supports the hypothesis of monochromatic vision in the panarthropod ancestor, whereas two duplications of the ancestral panarthropod c-opsin have led to three c-opsins in tardigrades. Although the early-branching nodes are unstable within the metazoans, our findings suggest that the last common ancestor of Bilateria possessed six opsins: two r-opsins, one c-opsin, and three Group 4 opsins, one of which (Go opsin) was lost in the ecdysozoan lineage.

Keywords

opsin evolution, vision, transcriptomics, Tardigrada, Panarthropoda

Introduction

Opsins are light-sensitive proteins used for photoreception. These proteins function as G protein-coupled receptors (GPCRs) that trigger phototransduction cascades associated with animal vision and circadian clocks (e.g. Arendt et al. 2004; review Fain et al. 2010; review Hankins et al. 2008; Rubin et al. 2006; review Shichida and Matsuyama 2009; Velarde et al. 2005). Previous studies had unveiled that the last common ancestor of Bilateria possessed representatives of three major opsin clades, including ciliary [=c-opsins], rhabdomeric [=r-opsins], and Group 4 opsins (review Porter et al. 2012). Among these three clades, only the c-opsins and r-opsins have been known to be involved in vision (e.g. Koyanagi et al. 2008; Koyanagi and Terakita 2013; Land and Nilsson 2012). Independent diversification of c-opsins in vertebrates and r-opsins in arthropods has led to convergent evolution of color vision in these animals (review Shichida and Imai 1998).

Within Panarthropoda (Onychophora + Tardigrada + Arthropoda), color vision has been confirmed only in arthropods, whereas onychophorans most likely show monochromatic vision due to the presence of a single visual r-opsin, dubbed *onychopsin* (Hering et al. 2012). Moreover, while all three major bilaterian opsin clades are represented in the arthropod lineage, Group 4 opsins are absent from onychophorans (Eriksson et al. 2013). However, since the corresponding information is completely missing from tardigrades, the opsin repertoire in the last common ancestor of Panarthropoda remains unknown.

Tardigrades typically possess a pair of simple, ocellus-like eyes (Dewel et al. 1993; Greven 2007; Kristensen 1983) – a situation which is similar to that in onychophorans but different from that in arthropods, which show two types of visual organs: the median ocelli, and the compound eyes (fig. 1A–C). While the onychophoran eyes have been homologized with the median ocelli of arthropods (Mayer 2006), the homology of the tardigrade eyes remains obscure (Greven 2007). Experimental data revealed that tardigrades clearly respond

to light (Baumann 1961; Beasley 2001; Marcus 1929; Ramazzotti and Maucci 1983), but beyond this nothing is known about the photoreceptive system in these animals.

The major objective of our study is therefore to analyze the opsin repertoire in a tardigrade to clarify the following questions: (i) Do tardigrades show only a single r-opsin as in onychophorans (Hering et al. 2012), or is there evidence for multiple visual pigments as in arthropods (e.g. review Briscoe and Chittka 2001; Henze et al. 2012; review Marshall et al. 2007)? (ii) Did the last common ancestor of Onychophora, Tardigrada and Arthropoda have monochromatic vision? (iii) How many orthologs of c-opsins and Group 4 opsins have been retained in the tardigrade lineage? (iv) Did losses and duplications of opsin genes occur in tardigrades and, if so, how many? (v) What was the opsin composition in the last common ancestor of Panarthropoda and Bilateria?

To answer these questions, we sequenced and analyzed the transcriptome of the eutardigrade *Hypsibius dujardini* using an Illumina-based sequencing approach. In addition, we screened various metazoan genomes, including the recently released genome of *H. dujardini*, and reconstructed the phylogeny of broadly sampled metazoan opsin genes, which allowed us to firmly place the tardigrade sequences in the bilaterian opsin tree.

Material and Methods

Specimens, library preparation, sequencing and assembly

Specimens of *Hypsibius dujardini* (Doyère, 1840) (Eutardigrada, Hypsibiidae) were obtained commercially from Sciento (Manchester, UK). Several hundred specimens were used to extract total RNA using TRIzol[®] reagent (Invitrogen, Carlsbad, CA) and RNeasy MinElute Cleanup Kit (Qiagen, Hilden, Germany) according to the manufacturers' protocols. Library preparation for double indexing (Kircher et al. 2012; Meyer and Kircher 2010), 76 cycles paired-end sequencing on an Illumina Genome Analyzer IIx (San Diego, CA), and post-sequencing processing (adapter trimming, removal of reads with falsely paired indices and filtering of reads at three different levels of stringency) were performed as described by Hering et al. (2012). Each of the three obtained datasets (Filter15, Filter25, and Filter30) was then assembled *de novo* using two different software packages to assess the occurrence of opsin transcripts in a broader methodological framework: CLC Genomics Workbench v5.1 (CLC bio, Århus, Denmark), and IDBA-Tran v1.1.0 (Peng et al. 2013). The IDBA-Tran assemblies were done twice, allowing for the retention of one or three isoforms of a transcript, respectively, using the *-num_isoform* option (additional assembly parameters and statistics; supplementary table S1, Supplementary Material online).

Obtaining opsin sequences from the transcriptome of *H. dujardini* and publicly available metazoan genomes

To obtain the sequences of putative opsin genes from the transcriptome of *H. dujardini*, BLAST v2.2.27+ (Altschul et al. 1997) and HMMER v3.1b1 (<http://hmmer.org/>; Eddy 1998) were used in custom Perl scripts on nine assemblies in total as described by Hering et al. (2012) with the following modifications: for the tBLASTn/BLASTP searches, 16 opsin sequences from all major opsin groups were used as bait sequences with an E value of 1e-5 as

a threshold (for accession numbers of all query sequences, see supplementary table S2, Supplementary Material online). For the HMMER search, the same value was used and the search was performed by applying previously built HMMER profiles (Hering et al. 2012). In total, 1,634 non-redundant contigs were obtained as putative opsin genes and used as candidates in further analyses. For reciprocal BLAST searches against the nr database of GenBank, 530 non-redundant contigs from a BLASTP search with the E value $1e-10$ and a HMMER search with the E value $1e-20$ were used and every best hit was stored (101 hits after the removal of redundant sequences). Furthermore, we mined the publicly available genomes of the annelids *Capitella teleta* and *Helobdella robusta*, and the mollusks *Lottia gigantea* (<http://www.jgi.doe.gov/>) and *Crassostrea gigas* (<http://gigadb.org/>) to enrich our metazoan opsin dataset by using BLAST searches. In addition, we screened the genomes of the aphid *Acyrtosiphon pisum* (<http://www.aphidbase.com/>) and the water flea *Daphnia pulex* (<http://www.jgi.doe.gov/>) to identify putative orthologs of the new *opsin-5* gene (vertebrate *opsin-5/neuropsin*-like gene) from *H. dujardini*. All identified putative opsin genes from these genomes were checked for the presence of lysine at the retinal-binding site corresponding to the K296 position of bovine rhodopsin (Palczewski et al. 2000). Due to the uncertain placement of one of the opsins of the ctenophore *Mnemiopsis leidyi* (MleiOpsin3; see Schnitzler et al. 2012) and an unusual insertion downstream of the predicted retinal-binding site corresponding to the K296 position of bovine rhodopsin, it is unclear whether or not *MleiOpsin3* is a functional opsin gene. We therefore decided to exclude MleiOpsin3 prior to our analyses (supplementary figure S8, Supplementary Material online).

Computational pre-analyses

To decide whether or not a contig identified by BLAST/HMMER is a potential opsin gene, several maximum likelihood analyses (ML) were performed to sort out non-opsin contigs.

Therefore, we aligned all query sequences for the BLAST searches and the opsin sequences used to build the HMMER profiles to an opsin master alignment using the web server of MAFFT version 7 (Kato and Standley 2013). Thereafter, we extended this alignment with the 101 unique best hits obtained by the reciprocal BLAST search using the *-add* option in MAFFT. In the next step, the initially obtained 1,634 candidate contigs were split into batches of 150 sequences and each batch was also added to the master alignment using MAFFT (*-add* option) resulting in 11 separate alignments. Each of the 11 alignments was masked using ALISCORE v2.2 (Kück et al. 2010; Misof and Misof 2009) and ALICUT v2.3 (<http://www.utilities.zfmk.de>) to exclude randomly aligned sections prior to tree reconstruction (sliding window size 64, comparing 10,000 random pairs). After running 11 independent ML analyses to obtain the best tree for each dataset using RAxML v7.5.8 (Stamatakis 2006) with the PROTGAMMAAUTO option for automatic selection of the best-fitting substitution model (10x LG+G, 1x MTZOA+G), a total of 43 in-group opsin transcripts from all assemblies were curated manually using BioEdit v7.0.9 (Hall 1999) and CLC Main Workbench v6.8.4 (CLC bio, Århus, Denmark). These transcripts yielded five different opsin genes (*Hd-r-opsin*, *Hd-c-opsin1*, *Hd-c-opsin2*, *Hd-c-opsin3* and *Hd-neuroopsin*) for the tardigrade *H. dujardini*, named after the clade in which they occur.

Cloning and Rapid Amplification of cDNA Ends (RACE)

Rapid amplification of cDNA ends was performed using SMARTer[®] RACE cDNA Amplification Kit (Clontech Laboratories, Inc., Mountain View, CA) according to the manufacturer's protocol for a putative c-opsin (*Hd-c-opsin3*) and putatively new *Hd-neuroopsin* of *H. dujardini*, due to the short length of the corresponding contigs obtained from our transcriptome data. The fragments of all five identified opsin genes of *H. dujardini* were cloned from cDNA using the pGEM[®]-T Vector System (Promega, Madison, WI) and verified

by Sanger sequencing (GATC Biotech, Konstanz, Germany) using a standard M13 amplification. The sequences of the tardigrade opsin genes (*Hd-r-opsin*, *Hd-c-opsin1*, *Hd-c-opsin2*, *Hd-c-opsin3* and *Hd-neuroopsin*) were submitted to GenBank under the accession numbers KM086335–KM086339.

Final sequence alignment and masking

The final dataset comprised 401 metazoan opsin and other closely related G protein-coupled receptors (GPCRs), such as receptors for somatostatin, allatostatin, dopamine, octopamine and melatonin as outgroups, selected according to the reconstructed trees from our pre-analyses. Prior to the final alignment, a Pfam v27.0 domain search (Punta et al. 2012) was performed and all sequences were trimmed manually at ± 20 amino acids up- and downstream from the predicted seven-transmembrane domain (7tm_1; PF00001). This was done to allow for a more accurate identification of homologous positions during the alignment step by removing poorly alignable regions *a priori*, such as the unconserved regions flanking the actual domain. The alignment was done using MAFFT version 7 with the most accurate option L-INS-i and default parameters. The software Noisy rel. 1.5.12 (Dress et al. 2008) was used to mask the alignment by removing homoplastic and random-like positions (*-cutoff=0.8*, *-seqtype=P*, *-shuffles=20,000*).

Model choice and cross-validation

Several analyses were performed to obtain the best-fitting substitution model for the phylogenetic analyses of the final dataset. First, ProtTest v3.3 (Darriba et al. 2011) revealed LG+G+F as the best-fitting model according to the Akaike Information Criterion (AIC; Akaike 1974), Bayesian Information Criterion (BIC; Schwarz 1978), corrected Akaike Information Criterion (AICc; Hurvich and Tsai 1989; Sugiura 1978) and Decision Theory Criterion (DT; Minin et al. 2003). Second, a 10-fold cross-validation with 10 replicates for

each chosen model (LG, GTR, CAT-LG, CAT-GTR, C20-LG, C20-GTR, C30-LG, C30-GTR, C40-LG, C40-GTR, C50-LG, C50-GTR, C60-LG, C60-GTR, WLSR5-LG, WLSR5-GTR) was performed for comparison using the multicore version of PhyloBayes MPI v1.4f (Lartillot et al. 2009). Therefore, the original alignment was split randomly 10 times into a learning set (9/10 of the initial dataset) and a test set (1/10 of the initial dataset). Markov chains were run for 1,500 generations on each learning set (160 chains in total) of the models for comparison. For each of the replicates, the cross-validated likelihoods were calculated under the test set, averaged over the posterior distribution of the learning set, discarding the first 500 sampled points as burn-in and using the remaining 1,000 generations. Finally, the cross-validation log-likelihood scores per model were averaged over the 10 replicates and used to rank the fit to the initial dataset. According to this, the site-heterogeneous CAT-GTR model fitted the data best in 8 out of the 10 replicates. As the best non-site-heterogeneous model, the dataset-specific GTR model fitted the data better than the empirical LG model (supplementary table S3, Supplementary Material online). We therefore decided to use the GTR model for the final ML analyses using RAxML (due to the lack of site-heterogeneous models) and CAT-GTR to conduct Bayesian inference analyses with PhyloBayes.

Maximum likelihood and Bayesian inference analyses

The best maximum likelihood tree was obtained running 100 independent inferences on the final dataset with GAMMA correction of the final tree using the Pthreads-AVX version of RAxML v7.7.8 (fig. 2; supplementary figure S1, Supplementary Material online). A dataset-specific GTR substitution matrix was estimated during a prior single run and then provided as substitution matrix for the above-mentioned run. Bootstrap support values (BS) were calculated using the rapid bootstrapping algorithm implemented in RAxML from 1,000 pseudoreplicates on the original alignment. To check whether or not a sufficient number of

replicates has been generated (Pattengale et al. 2009), bootstrap convergence was assessed *a posteriori* according to the weighted Robinson-Foulds (WRF) distance criterion (WC) using the *-I autoMRE* option of RAxML (supplementary figure S1, Supplementary Material online). Bayesian inference analysis was performed using the multicore version of PhyloBayes MPI v1.4f under the site-heterogeneous CAT-GTR model, which is the best-fitting model according to the cross validation test. Two Markov chains were run independently for 60,000 generations each. Convergence of the chains was assessed using *bpcomp* and *tracecomp* statistics of PhyloBayes and Tracer v1.5.0 (Rambaut and Drummond 2009). Therefore, *bpcomp* and *tracecomp* were run multiple times. The burn-in was increased by 1,000 every iteration (sampling every 10th tree), beginning with a burn-in of 1,000 and finishing with a burn-in of 59,000. The obtained statistics and the log-likelihood traces of the runs were summarized and used for reliable assessment of chain convergence as described in the PhyloBayes manual, in dependency on the burn-in (summary statistics and parameter for chain convergence; supplementary figure S9, Supplementary Material online). Accordingly, the first 21,000 trees of each chain were discarded and every second tree thereafter was used to calculate the 50% majority rule consensus tree with posterior probability support values (supplementary figure S2, Supplementary Material online).

Sensitivity analyses using leaf stability

To identify “unstable” taxa in the reconstructed phylogeny of opsins, especially among the “early-branching” clades, we calculated the leaf stability indices (LS; Thorley and Wilkinson 1999) for each branch using Phyutility v2.2.6 (Smith and Dunn 2008), with 1,000 trees derived from RAxML bootstrapping as input, and mapped these LS indices on the corresponding ML tree using iTOL v2.1 (Letunic and Bork 2011) (LS indices; supplementary figure S1, Supplementary Material online). Following this, we excluded those taxa from the

original dataset with LS indices ≤ 0.50 and ≤ 0.60 , respectively, except for the ctenophore opsins (LS=0.57) due to their importance for our study. These pruned datasets were aligned, masked and reanalyzed separately using RAxML v7.7.8 (GTR+G) with 1,000 bootstrap pseudoreplicates, including the *a posteriori* Bootstrap convergence assessment as described above for the full dataset (supplementary figure S3+S5, Supplementary Material online). In contrast to the Bayesian inference on the full dataset, two Markov chains for each pruned dataset were run for 30,000 instead of 60,000 generations using PhyloBayes MPI v1.4f (CAT-GTR). After the assessment of chain convergence, as described for the full dataset (supplementary figure S10+S11, Supplementary Material online), the first 18,000 trees (LS ≤ 0.50 dataset) and 16,000 trees, respectively (LS ≤ 0.60 dataset), were discarded as burn-in and every tree thereafter was used to calculate the 50% majority rule consensus tree with posterior probability support values (supplementary figure S4+S6, Supplementary Material online).

Results

Identification of opsin genes in the transcriptome of the tardigrade *Hypsibius dujardini*

The sequencing of the *H. dujardini* transcriptome library with an Illumina Genome Analyzer IIx yielded 68,214,238 filtered paired-end reads. Our screening pipeline to search for putative opsin candidates included three different filtering stringencies of the raw data (Filter15, Filter25, Filter30), two different assemblers (CLC, IDBA), and two different search algorithms (BLAST, HMMER) (see the Materials and Methods section for more details). Sorting out the false positives resulted in a total of 43 contigs belonging to five different opsin genes. The sequences of two of them had to be extended by RACE. We successfully cloned partial sequences of all opsin transcripts, which we named after the clade in which they occur in our phylogenetic analyses (*Hd-r-opsin*: 703 nt, *Hd-c-opsin1*: 664 nt, *Hd-c-opsin2*: 704 nt, *Hd-c-opsin3*: 702 nt, *Hd-neuropsin*: 605 nt).

To verify our results and to check for additional opsin genes, we also screened the recently released genome of *H. dujardini* (http://badger.bio.ed.ac.uk/H_dujardini/). With the exception of a second, slightly different predicted transcript from the genome (nHd.2.3.1.g15325), which is most similar to our *Hd-neuropsin* and therefore perhaps an isoform of it, we found no evidence for any additional opsin genes apart from those already obtained from our transcriptomic data.

Phylogenetic analyses of metazoan opsin genes and the placement of *H. dujardini* opsins

In our maximum likelihood (ML) analysis of the full metazoan opsin dataset, all three major bilaterian opsin subgroups were recovered, including r-opsins, c-opsins, and Group 4 opsins *sensu* Porter et al. (2012) (fig. 2; supplementary figure S1, Supplementary Material online).

The best ML tree shows a monophyletic clade of the bilaterian c-type opsins with a weak

bootstrap support (BS<50), containing vertebrate visual pigments, chordate brain opsins (pinopsins, parapinopsins, vertebrate ancient opsins, teleost multiple tissue opsins, and encephalopsins), arthropod brain opsins (pteropsins), a brain opsin from the annelid *Platynereis dumerilii* (see Arendt et al. 2004), and a c-type opsin from the onychophoran *Euperipatoides kanangrensis* (see Eriksson et al. 2013) (fig. 2C; supplementary figure S1, Supplementary Material online). Notably, none of the opsin sequences obtained from the trochozoan genomes, including the annelids *Capitella teleta* and *Helobdella robusta*, and the mollusks *Lottia gigantea* and *Crassostrea gigas*, fall into this clade.

Within the bilaterian c-opsin clade, three of the five obtained opsin genes from the tardigrade *H. dujardini* (*Hd-c-opsin1*, *Hd-c-opsin2* and *Hd-c-opsin3*) emerge as the sister group to all known c-opsins/pteropsins from arthropods. The tardigrade and arthropod c-opsins/pteropsins in turn cluster with the onychophoran c-opsin, altogether forming a well-supported monophyletic clade (BS=85) of panarthropod c-opsins/pteropsins (fig. 2C; supplementary figure S1, Supplementary Material online). Our analysis further revealed a highly supported clade (BS=99) consisting exclusively of the anthozoan opsin genes from *Nematostella vectensis* and *Acropora digitifera* (hereafter referred to as “Anthozoa II”) as the sister group to the bilaterian c-opsins (supplementary figure S1, Supplementary Material online). However, this relationship exhibits a low bootstrap support value (BS<50) and our leaf stability analysis shows that the position of Anthozoa II is unstable in the metazoan opsin tree (LS=0.47).

A group of several uncharacterized opsins from the genomes of *Lottia gigantea* and *Crassostrea gigas*, the sequence of *Sp-opsin2* from the sea urchin *Strongylocentrotus purpuratus* (see Raible et al. 2006), and the c-type opsin from the brachiopod *Terebratalia transversa* (see Passamaneck et al. 2011) was recovered as sister to the clade of the “cnidopsins” *sensu* Plachetzki et al. (2007; 2010) (hereafter used for all cnidarian opsins to

the exclusion of the anthozoan opsins I and II; fig 2; supplementary figure S1, Supplementary Material online) and the ctenophoran opsins (hereafter referred to as “ctenopsins”), also with low nodal support (BS<50). This clade in turn occurs as the sister group to the clade [Anthozoa II + c-opsins], although this relationship shows only weak bootstrap and leaf stability values (BS<50; LS≤0.59). Interestingly, the two ctenopsins from both ctenophoran species included in the analyses (*Mnemiopsis leidyi* and *Pleurobrachia pileus*) are monophyletic with maximum support value (BS=100), suggesting that they are ctenophoran in-paralogs.

The second major clade of the bilaterian opsins – Group 4 opsins *sensu* Porter et al. (2012) – includes peropsins, RGR opsins, Go-opsins, and an assemblage of neuropsins and “opsin-5” genes. The sister group relationship of Group 4 opsins to all above-mentioned c-type opsins, cnidopsins, Anthozoa II opsins, ctenopsins, and related opsins is weakly supported in our analyses (BS<50; see fig. 2; supplementary figure S1, Supplementary Material online). Surprisingly, one of our obtained putative opsin genes from the tardigrade *H. dujardini* appears together with other protostome taxa as the sister group to the deuterostome neuropsins/opsin-5 group (Tarttelin et al. 2003), both forming a monophyletic clade (BS=52) within the Group 4 opsins (fig. 2B). Besides the tardigrade sequence, this protostome neuropsin clade consists of single putative neuropsin homologs of *Daphnia pulex* and *Crassostrea gigas*, two homologs of *Lottia gigantea* and six homologs of *Capitella teleta* based on their screened genomes.

A monophyletic clade of r-type opsins, the third major bilaterian opsin group, was recovered as the sister group to the clade [c-opsins (including related opsins) + Anthozoa II + cnidopsins + ctenopsins + Group 4 opsins] with strong bootstrap support (BS=85). Compared to the c-opsin and Group 4 opsin clades, the r-opsins occur as the best-supported and the earliest-branching clade. Moreover, the leaf stabilities of the r-opsins are by far the highest

among all taxa included ($LS \geq 78$), exceeded only by the outgroup taxa (supplementary figure S1, Supplementary Material online). Within the r-opsin clade, the fifth opsin candidate from the tardigrade *H. dujardini* occurs as sister to the visual r-opsins of Onychophora (onychopsin genes sensu Hering et al. 2012), which together form a well-supported monophyletic clade ($BS=97$) (fig. 2A; supplementary figure S1, Supplementary Material online). The r-opsins of the Tardigrada + Onychophora clade emerge as the closest relatives to a highly diverse group of visual r-opsins of Arthropoda. These arthropod r-opsins are subdivided into two major subgroups that to some extent reflect their spectral sensitivity, i.e. UV and short-wavelength sensitive vs. medium- and long-wavelength sensitive opsins. The tardigrade r-opsin is unlikely to be a UV-sensitive visual pigment, as it has methionine (M) instead of lysine (K) at the position corresponding to G90 of the bovine rhodopsin (Palczewski et al. 2000), which has been shown to be responsible for ultraviolet tuning properties of arthropod opsins (Salcedo et al. 2003).

In addition to the panarthropod visual r-opsins, the rhabdomeric opsins are further subdivided into three monophyletic clades according to our analyses: (i) the lophotrochozoan (most likely visual) r-opsins, (ii) the chordate non-visual r-opsins (=melanopsins), and (iii) the arthropods (initially described by Colbourne et al. 2011) that form the earliest-branching subgroup. Most intriguingly, in addition to the eight arthropod paralogs described from the crustacean *Daphnia pulex* and the two putative arthropods from the onychophoran *Euperipatoides kanangrensis* and the spider *Cupiennius salei* (see Colbourne et al. 2011; Eriksson et al. 2013), we identified putative arthropod sequences in the genomes of the pea aphid *Acyrtosiphon pisum* as well as the annelid *Capitella teleta* and the mollusks *Lottia gigantea* and *Crassostrea gigas*. Moreover, even *r-opsin4* from the annelid *Platynereis dumerilii* (see Randel et al. 2013) and *Amphio6* from the lancelet *Branchiostoma belcheri* (see

Koyanagi et al. 2002) occur as members of the arthropsin clade in our cladograms (supplementary figure S1, Supplementary Material online).

The least stable taxa in our full dataset ML analysis ($LS \leq 0.46$) are the *Sp-opsin5* sequences from two sea urchins (Lesser et al. 2011; Raible et al. 2006) and the *Go-opsin2* sequence from the brachiopod *Terebratalia transversa* (see Passamanek and Martindale 2013) that branch off at the base of the tree as well as a second clade of anthozoan-specific opsins (hereafter referred to as “Anthozoa I” opsins) containing two paralogs from the genome of *Nematostella vectensis* and the *acropsin3* sequences (Mason et al. 2012) from two *Acropora* species. *Sp-opsin5* and its ortholog from *Strongylocentrotus droebachiensis* have been classified as rhabdomeric opsins (Lesser et al. 2011), but these opsins clearly cluster outside the r-opsin clade in our and Lesser et al.’s (2011) phylogenetic analyses. In contrast, according to our results, *Sp-opsin4* from *Strongylocentrotus purpuratus* is deeply nested within the r-opsin clade (supplementary figure S1, Supplementary Material online).

The 50% majority rule consensus tree of our Bayesian inference (BI) analysis of the full metazoan opsin dataset revealed a similar topology to our ML analysis, except that the Anthozoa II clade is deeply nested within the c-opsins clade. However, as mentioned above, the Anthozoa II opsins show low leaf stability indices ($LS = 0.47$) (supplementary figure S2, Supplementary Material online). The most conspicuous deviation from the ML tree is the unresolved topology at the base of the tree, indicated by polytomous branches of the particular opsin clades (bilaterian c-opsins, r-opsins, Group 4 opsins, cnidopsins, ctenopsins, and Anthozoa I). Nevertheless, the placement of the tardigrade opsins in the BI analysis corresponds to that in the ML tree. The same holds true for the ML and BI analyses, from which we excluded the putatively unstable taxa with leaf stability indices $LS \leq 0.50$ and $LS \leq 0.60$, respectively (supplementary figures S3-S6, Supplementary Material online).

Discussion

Opsin repertoire in the tardigrade *Hypsibius dujardini*

Our transcriptomic analyses of the opsin repertoire in the tardigrade *H. dujardini* revealed a set of five opsin genes, including one r-opsin (*Hd-r-opsin*), three c-opsins (*Hd-c-opsin1*, *Hd-c-opsin2*, and *Hd-c-opsin3*), and a neuropsin/opsin-5 (*Hd-neuropsin*). We found essentially the same set of genes in the publicly available genome of this species, with the exception of an additional neuropsin/opsin-5 isoform. However, this isoform is unlikely to be a functional gene, as its expression could not be confirmed by the transcriptomic analyses and gene cloning. Moreover, we have noticed some inconsistencies between our cloned transcripts on the one hand and the genomic sequences and predicted transcripts on the other hand, in which short fragments from putatively expressed regions are either absent or repeated (supplementary figure S7, Supplementary Material online). This suggests that either the genomic contigs and/or scaffolds were assembled incorrectly or, if the assembly was correct and the introns indeed contain such repeated exonic sequences, the actual splice sites were predicted incorrectly during the gene annotation process. It is therefore unclear whether the additional neuropsin/opsin-5 isoform is a real, albeit non-functional sequence, or an assembly artifact. Nevertheless, irrespective of these inconsistencies and minor differences in nucleotide composition (probably due to heterozygosity), the available genome sequences correspond well to our transcriptomic data, suggesting that *H. dujardini* has at least five functional opsin genes. According to our phylogenetic analyses, these genes cluster within the three major bilaterian opsin clades.

Evidence for monochromatic vision in the last common ancestor of Panarthropoda

Typically, either r-opsins (as in arthropods) or c-opsins (as in vertebrates) are involved in animal vision (e.g. Arendt 2003; Arendt and Wittbrodt 2001; Land and Nilsson 2012;

Vopalensky and Kozmik 2009). Given that Tardigrada is one of the closest arthropod relatives, one would expect that *H. dujardini* employs the single identified r-opsin as the sole visual pigment, suggesting monochromatic vision in this tardigrade species. However, in addition to numerous microvilli that may act as rhabdomeric photoreceptive structures, at least one ciliary cell has been reported from the eye of *Milnesium tardigradum* and *Halobiotus crispae* (see Dewel et al. 1993; Greven 2007; Kristensen 1983). Although the cilia described from *M. tardigradum* and *H. crispae* are unlikely to be involved in photoreception, a potential function of the three identified c-opsins in *H. dujardini* in visual photoreception and, hence, in color vision cannot be excluded as long as the corresponding gene expression data are unavailable. However, irrespective of the function of these genes, our phylogenetic analyses suggest that the three c-opsin genes have evolved by gene duplication in the tardigrade lineage (or a tardigrade subgroup) and that the last common ancestor of Panarthropoda possessed only one c-opsin gene (fig. 3A). This finding and the identification of a single r-opsin gene in *H. dujardini* are in line with the assumption of monochromatic vision in the last common ancestor of Onychophora, Tardigrada and Arthropoda (Hering et al. 2012).

Neuropsins were present in the last common bilaterian ancestor

Our phylogenetic analyses revealed that at least one opsin gene of *H. dujardini* belongs to the neuropsin/opsin-5 clade – a subgroup of the Group 4 opsins. To our knowledge, this is the first report of a neuropsin from a protostome, as this type of opsin so far has been known only from vertebrates (review Koyanagi and Terakita 2013). In addition to the tardigrade neuropsin, we identified up to six putative neuropsin homologs in the genomes of other protostomes, including the annelid *Capitella teleta*, the mollusks *Crassostrea gigas* and *Lottia gigantea*, and the crustacean *Daphnia pulex*. The latter is most astonishing because of its

well-characterized opsin repertoire of hitherto 46 identified opsin genes (Colbourne et al. 2011). The identification of neuropsin/opsin-5 members in representatives of Lophotrochozoa, Ecdysozoa and Deuterostomia suggests an origin of this opsin lineage prior to the split Deuterostomia + Protostomia (fig. 3B). While the function of neuropsin/opsin-5 homologs in protostomes is unknown, the vertebrate Opsin-5 (Opn5) homolog acts as a UV-sensitive G protein-coupled receptor (GPCR), which activates the Gi-type G proteins (Yamashita et al. 2010) and is expressed in various tissues, including the neural retina, deep brain, testis, and even the outer ear (Kojima et al. 2011; Nakane et al. 2010; Tarttelin et al. 2003; Yamashita et al. 2014). In light of our findings, it would be interesting to know whether the protostome neuropsin/opsin-5 homologs also have a function as UV-sensitive GPCRs or rather act as retinal photoisomerases, similar to the closely related peropsins (Koyanagi et al. 2002).

“Arthropins” are not restricted to arthropods

Arthropins were initially described from the genome of the crustacean *Daphnia pulex*, in which eight members of this putatively novel opsin group clustered as an early-branching clade within the r-opsins (Colbourne et al. 2011). Interestingly, another early-branching r-opsin, *Amphiop6* – which was identified previously in the chordate *Branchiostoma belcheri* and did not group with members of the melanopsin clade (Koyanagi et al. 2005) – was subsequently recovered as sister to the arthropins (Hering et al. 2012). Moreover, Randel et al. (2013) identified “another stable r-opsin clade (Clade I) with both mollusk and annelid sequences”, the phylogenetic relationship of which remained unclear, as neither the *Daphnia* arthropins nor *Amphiop6* were included in their study. The results of our phylogenetic analyses now show that all of these sequences, including Randel et al.’s (2013) “Clade I” and additional members from other bilaterian taxa, form a monophyletic group, which thus

includes representatives of Onychophora, Arthropoda, Lophotrochozoa, and Chordata. This implies that “arthropsins” are not restricted to arthropods but are the result of an ancient duplication of the r-opsin precursor in the last common bilaterian ancestor (fig. 3B). While additional duplications have occurred in some lineages (e.g. in *Daphnia pulex*; see Colbourne et al. 2011), an arthropsin homolog might have been lost in the tardigrade lineage, as it is not found in the transcriptomic or genomic data from *H. dujardini* (fig. 3A).

Conclusions

To determine the position of the tardigrade opsins, we performed an extensive phylogenetic analysis of broadly sampled metazoan opsins, including newly identified members from the available metazoan genomes. Our results suggest that the last common bilaterian ancestor possessed six opsins belonging to the three major opsin clades (fig. 3B): (i) two r-opsins (one of which was an “arthropsin”); (ii) one c-opsin; and (iii) three Group 4 opsins (including a neuropsin, a peropsin, and a Go opsin that was most likely lost in the ecdysozoan lineage). Since none of the opsins from the non-bilaterian taxa falls into any of these major clades, we suggest that a single duplication of the r-opsin precursor and two duplications within the Group 4 opsin clade have occurred within the bilaterian lineage (fig. 3B). Unfortunately, beyond this no unambiguous conclusion is possible on the origin and relationship of the three major bilaterian opsin clades, possibly due to the lack of sufficient phylogenetic signal to robustly resolve the deepest nodes of the opsin gene tree. This, in conjunction with the generally ambiguous placement of the non-bilaterian taxa, such as Cnidaria and Ctenophora within the metazoans (review Philippe et al. 2011), makes it a challenging task to draw any conclusions about pre-bilaterian opsin evolution.

Our analyses revealed three clades of cnidarian opsins: Anthozoa I, Anthozoa II, and “cnidopsins” (sensu Plachetzki et al. 2007). While the Anthozoa I and II clades exclusively contain anthozoan sequences, the cnidopsins comprise representatives of Anthozoa, Cubozoa and Hydrozoa, supporting the results of previous studies (Plachetzki et al. 2010; Porter et al. 2012; Suga et al. 2008). However, the placement of the cnidopsins differs among these studies, in which they form the sister group to various clades, including [r-opsins + Group 4 opsins] (Plachetzki et al. 2007), [r-opsins + c-opsins + Group 4 opsins] (Plachetzki et al. 2010), [Group 4 opsins] (Feuda et al. 2012), and [c-opsins] (Hering et al. 2012; Porter et al. 2012; Suga et al. 2008), depending on the underlying reconstruction method and the

substitution model used. In our analyses, the cnidopsins generally form the sister group of the ctenophoran opsins (=ctenopsins), but the placement of the entire cnidopsins/ctenopsins clade is ambiguous and depends on the analysis parameters used. The same applies to the Arthropsin I and II clades, the position of which is unstable.

These discrepancies between the studies and the methods used are not surprising, given the old age of the major metazoan lineages, dating back to ~700 million years ago, and the cladogenesis events that were highly compressed in time (e.g. Rokas et al. 2005). This might have led to a considerable accumulation of homoplasies and, hence, to an erosion of the phylogenetic signal in the molecular data. Therefore, currently it seems impossible to reconstruct with confidence the early-branching nodes and to reconcile a reliable scenario on the evolutionary history of the metazoan opsins based solely on opsin phylogeny. Nevertheless, our phylogenetic framework allows for the following conclusions on the evolution of opsins in panarthropods:

- (1) The last common ancestor of Panarthropoda most likely possessed a c-opsin, two r-opsins (an arthropsin and an additional [probably visual] r-opsin, which was not UV-sensitive), and two Group 4 opsins (a neuropsin and a peropsin) – a set that had been inherited from the last common ancestor of Ecdysozoa (fig. 3A+B).
- (2) This ancestral set of opsin genes was retained in the last common ancestor of Arthropoda (fig. 3A), although subsequent duplications and losses occurred in some arthropod lineages (e.g. Briscoe 2000; Colbourne et al. 2011; Porter et al. 2013).
- (3) The last common ancestor of Onychophora had retained a c-opsin and two r-opsins (arthropsin and onychopsin; see Eriksson et al. 2013; Hering et al. 2012) from the panarthropod ancestor, whereas the two Group 4 opsins (neuropsin and peropsin) were lost in the onychophoran lineage (fig. 3A). However, this hypothesis requires confirmation by genomic analyses.

- (4) The last common ancestor of Tardigrada (or the tardigrade subgroup containing *H. dujardini*) most likely possessed an r-opsin, three c-opsins, and a neuropsin, whereas the arthropsin and the peropsin were lost in the tardigrade lineage (fig. 3A).

Supplementary Material

Supplementary figures S1–S11 and tables S1–S3 are available at *Genome Biology and Evolution* online (<http://gbe.oxfordjournals.org/>).

Acknowledgments

We are thankful to Matthias Meyer, Birgit Nickel and Martin Kircher for assistance with library preparation and sequencing, Vladimir Gross for proofreading, and Sandy Richter and Christoph Bleidorn for comments on the manuscript. This work was supported by a grant from the German Research Foundation (DFG; grant Ma 4147/3-1) to G.M, who is a Research Group Leader supported by the Emmy Noether Programme of the DFG.

References

- Akaike H. 1974. A new look at the statistical model identification. *IEEE T Automat Control*. 19:716–723.
- Altschul SF, et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 25:3389–3402.
- Arendt D. 2003. Evolution of eyes and photoreceptor cell types. *Int J Dev Biol*. 47:563–571.
- Arendt D, Tessmar-Raible K, Snyman H, Dorresteijn AW, Wittbrodt J. 2004. Ciliary photoreceptors with a vertebrate-type opsin in an invertebrate brain. *Science*. 306:869–871.
- Arendt D, Wittbrodt J. 2001. Reconstructing the eyes of Urbilateria. *Philos Trans R Soc B*. 356:1545–1563.
- Baumann H. 1961. Der Lebenslauf von *Hypsibius (H.) convergens* Urbanowicz (Tardigrada). *Zool Anz*. 167:362–381.
- Beasley CW. 2001. Photokinesis of *Macrobiotus hufelandi* (Tardigrada, Eutardigrada). *Zool Anz*. 240:233–236.
- Briscoe AD. 2000. Six opsins from the butterfly *Papilio glaucus*: molecular phylogenetic evidence for paralogous origins of red-sensitive visual pigments in insects. *J Mol Evol*. 51:110–121.
- Briscoe AD, Chittka L. 2001. The evolution of color vision in insects. *Annu Rev Entomol*. 46:471–510.
- Colbourne JK, et al. 2011. The ecoresponsive genome of *Daphnia pulex*. *Science*. 331:555–561.
- Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics*. 27:1164–1165.
- Dewel RA, Nelson DR, Dewel WC. 1993. Tardigrada. In: Harrison FW and Rice ME, editors. *Microscopic Anatomy of Invertebrates: Onychophora, Chilopoda, and lesser Protostomata*. New York: Wiley-Liss. p. 143–183.
- Dress A, et al. 2008. Noisy: identification of problematic columns in multiple sequence alignments. *Alg Mol Biol*. 3:7.
- Eddy SR. 1998. Profile hidden Markov models. *Bioinformatics*. 14:755–763.
- 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.
- Fain GL, Hardie R, Laughlin SB. 2010. Phototransduction and the evolution of photoreceptors. *Curr Biol*. 20:R114–R124.
- Feuda R, Hamilton SC, McInerney JO, Pisani D. 2012. Metazoan opsin evolution reveals a simple route to animal vision. *Proc Natl Acad Sci USA*. 109:18868–18872.
- Greven H. 2007. Comments on the eyes of tardigrades. *Arthropod Struct Dev*. 36:401–407.
- Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser*. 41:95–98.
- Hankins MW, Peirson SN, Foster RG. 2008. Melanopsin: an exciting photopigment. *Trends Neurosci*. 31:27–36.
- Henze M, Dannenhauer K, Kohler M, Labhart T, Gesemann M. 2012. Opsin evolution and expression in arthropod compound eyes and ocelli: insights from the cricket *Gryllus bimaculatus*. *BMC Evol Biol*. 12:163.
- Hering L, et al. 2012. Opsins in Onychophora (velvet worms) suggest a single origin and subsequent diversification of visual pigments in arthropods. *Mol Biol Evol*. 29:3451–3458.

- Hurvich CM, Tsai C-L. 1989. Regression and time series model selection in small samples. *Biometrika*. 76:297–307.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 30:772–780.
- Kircher M, Sawyer S, Meyer M. 2012. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res*. 40:e3.
- Kojima D, et al. 2011. UV-sensitive photoreceptor protein OPN5 in humans and mice. *PLOS One*. 6:e26388.
- 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, Nagata T, Katoh K, Yamashita S, Tokunaga F. 2008. Molecular evolution of arthropod color vision deduced from multiple opsin genes of jumping spiders. *J Mol Evol*. 66:130–137.
- Koyanagi M, Terakita A. 2013. Diversity of animal opsin-based pigments and their optogenetic potential. *BBA-Bioenergetics*. 1837:710–716.
- Koyanagi M, Terakita A, Kubokawa K, Shichida Y. 2002. Amphioxus homologs of Go-coupled rhodopsin and peropsin having 11-*cis*- and all-*trans*-retinals as their chromophores. *FEBS Lett*. 531:525–528.
- Kristensen RM. 1983. The first record of cyclomorphosis in Tardigrada based on a new genus and species from Arctic meiobenthos. *J Zool Syst Evol Res*. 20:249–270.
- Kück P, et al. 2010. Parametric and non-parametric masking of randomness in sequence alignments can be improved and leads to better resolved trees. *Front Zool*. 7:10.
- Land MF, Nilsson D-E. 2012. *Animal eyes*. New York: Oxford University Press.
- Lartillot N, Lepage T, Blanquart S. 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics*. 25:2286–2288.
- Lesser MP, Carleton KL, Böttger SA, Barry TM, Walker CW. 2011. Sea urchin tube feet are photosensory organs that express a rhabdomeric-like opsin and PAX6. *Proc R Soc B*. 278:3371–3379.
- Letunic I, Bork P. 2011. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res*. 39:W475–W478.
- Marcus E. 1929. Tardigrada. In: Bronn HG, editor. *Klassen und Ordnungen des Tierreichs*. Leipzig: Akademische Verlagsgesellschaft m.b.H. p. 608.
- Marshall J, Cronin TW, Kleinlogel S. 2007. Stomatopod eye structure and function: a review. *Arthropod Struct Dev*. 36:420–448.
- Mason B, et al. 2012. Evidence for multiple phototransduction pathways in a reef-building coral. *PLOS One*. 7:e50371.
- Mayer G. 2006. Structure and development of onychophoran eyes: what is the ancestral visual organ in arthropods? *Arthropod Struct Dev*. 35:231–245.
- Mayer G, et al. 2013. Selective neuronal staining in tardigrades and onychophorans provides insights into the evolution of segmental ganglia in panarthropods. *BMC Evol Biol*. 13:230.
- Meyer M, Kircher M. 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc*. 2010:pdb.prot5448–.
- Minin V, Abdo Z, Joyce P, Sullivan J. 2003. Performance-based selection of likelihood models for phylogeny estimation. *Syst Biol*. 52:674–683.
- Misof B, Misof K. 2009. A Monte Carlo approach successfully identifies randomness in multiple sequence alignments: a more objective means of data exclusion. *Syst Biol*. 58:21–34.

- Nakane Y, et al. 2010. A mammalian neural tissue opsin (Opsin 5) is a deep brain photoreceptor in birds. *Proc Natl Acad Sci USA*. 107:15264–15268.
- Palczewski K, et al. 2000. Crystal structure of rhodopsin: a G protein-coupled receptor. *Science*. 289:739–745.
- Passamaneck Y, Furchheim N, Hejnol A, Martindale M, Luter C. 2011. Ciliary photoreceptors in the cerebral eyes of a protostome larva. *EvoDevo*. 2:6.
- Passamaneck YJ, Martindale MQ. 2013. Evidence for a phototransduction cascade in an early brachiopod embryo. *Integr Comp Biol*. 53:17–26.
- Pattengale ND, Alipour M, Bininda-Emonds ORP, Moret BME, Stamatakis A. 2009. How many bootstrap replicates are necessary? In: Batzoglou S, editor. *Research in Computational Molecular Biology*. Berlin, Heidelberg: Springer. p. 184–200.
- Peng Y, et al. 2013. IDBA-tran: a more robust de novo de Bruijn graph assembler for transcriptomes with uneven expression levels. *Bioinformatics*. 29:i326–i334.
- Philippe H, et al. 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. *PLOS Biol*. 9:e1000602.
- Plachetzki DC, Degnan BM, Oakley TH. 2007. The origins of novel protein interactions during animal opsin evolution. *PLOS One*. 2:e1054.
- Plachetzki DC, Fong CR, Oakley TH. 2010. The evolution of phototransduction from an ancestral cyclic nucleotide gated pathway. *Proc R Soc B*.
- Porter ML, et al. 2012. Shedding new light on opsin evolution. *Proc R Soc B*. 279:3–14.
- Porter ML, et al. 2013. The evolution of complexity in the visual systems of stomatopods: insights from transcriptomics. *Integr Comp Biol*. 53:39–49.
- Punta M, et al. 2012. The Pfam protein families database. *Nucleic Acids Res*. 40:D290–D301.
- Raible F, et al. 2006. Opsins and clusters of sensory G-protein-coupled receptors in the sea urchin genome. *Dev Biol*. 300:461–475.
- Ramazzotti G, Maucci W. 1983. *Il Phylum Tardigrada*. Verbania Pallanza: Memorie dell'Istituto Italiano di Idrobiologia Dott. Marco De Marchi.
- Rambaut A, Drummond AJ. 2009. Tracer v1.5. Last accessed 28.01.2014.
- 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.
- Rokas A, Krüger D, Carroll SB. 2005. Animal evolution and the molecular signature of radiations compressed in time. *Science*. 310:1933–1938.
- Rubin EB, et al. 2006. Molecular and phylogenetic analyses reveal mammalian-like clockwork in the honey bee (*Apis mellifera*) and shed new light on the molecular evolution of the circadian clock. *Genome Res*. 16:1352–1365.
- Salcedo E, Zheng L, Phistry M, Bagg EE, Britt SG. 2003. Molecular basis for ultraviolet vision in invertebrates. *J Neurosci*. 23:10873–10878.
- Schnitzler C, et al. 2012. Genomic organization, evolution, and expression of photoprotein and opsin genes in *Mnemiopsis leidyi*: a new view of ctenophore photocytes. *BMC Biol*. 10:107.
- Schwarz G. 1978. Estimating the dimension of a model. *Ann Stat*. 6:461–464.
- Shichida Y, Imai H. 1998. Visual pigment: G-protein-coupled receptor for light signals. *Cell Mol Life Sci*. 54:1299–1315.
- Shichida Y, Matsuyama T. 2009. Evolution of opsins and phototransduction. *Philos Trans R Soc B*. 364:2881–2895.
- Smith SA, Dunn CW. 2008. Phyutility: a phyloinformatics tool for trees, alignments and molecular data. *Bioinformatics*. 24:715–716.
- Stamatakis A. 2006. RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. 22:2688–2690.

- Suga H, Schmid V, Gehring WJ. 2008. Evolution and functional diversity of jellyfish opsins. *Curr Biol*. 18:51–55.
- Sugiura N. 1978. Further analysis of the data by Akaike's information criterion and the finite corrections. *Commun Stat Theory*. 7:13–26.
- Tarttelin EE, Bellingham J, Hankins MW, Foster RG, Lucas RJ. 2003. Neuropsin (Opn5): a novel opsin identified in mammalian neural tissue. *FEBS Lett*. 554:410–416.
- Thorley JL, Wilkinson M. 1999. Testing the phylogenetic stability of early tetrapods. *J Theor Biol*. 200:343–344.
- Velarde RA, Sauer CD, O. Walden KK, Fahrbach SE, Robertson HM. 2005. Pteropsin: a vertebrate-like non-visual opsin expressed in the honey bee brain. *Insect Biochem Molec*. 35:1367–1377.
- Vopalensky P, Kozmik Z. 2009. Eye evolution: common use and independent recruitment of genetic components. *Philos Trans R Soc B*. 364:2819–2832.
- Yamashita T, et al. 2010. Opn5 is a UV-sensitive bistable pigment that couples with Gi subtype of G protein. *Proc Natl Acad Sci USA*. 107:22084–22089.
- Yamashita T, et al. 2014. Evolution of mammalian Opn5 as a specialized UV-absorbing pigment by a single amino acid mutation. *J Biol Chem*. 289:3991–4000.

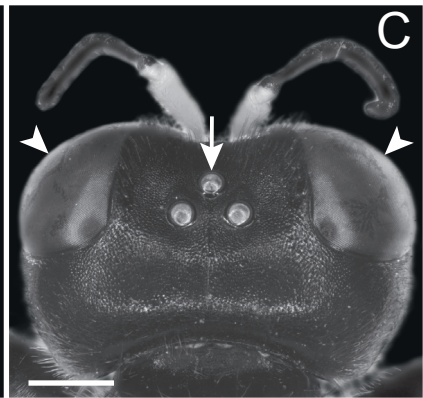
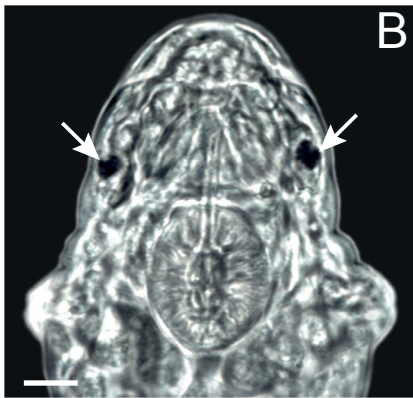
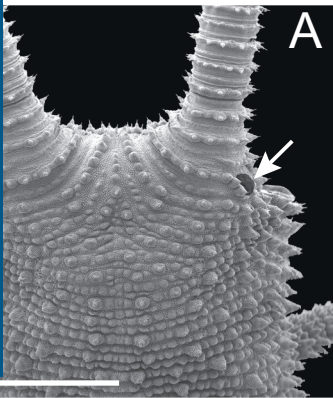
Figure Legends

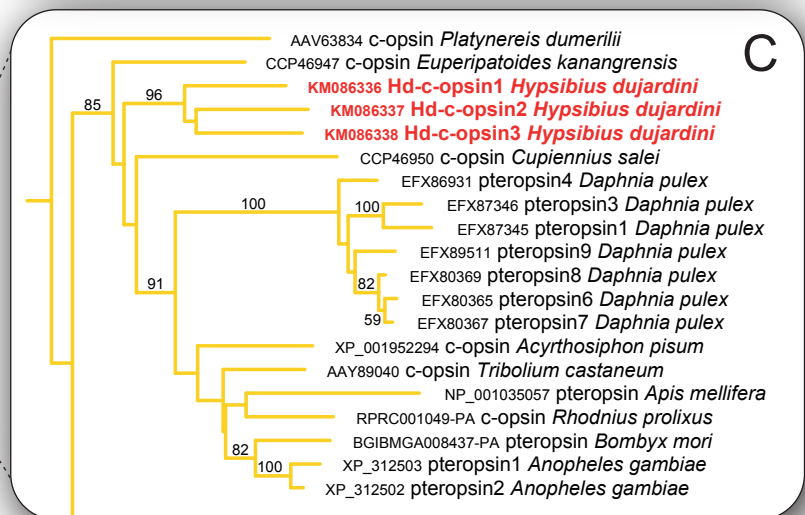
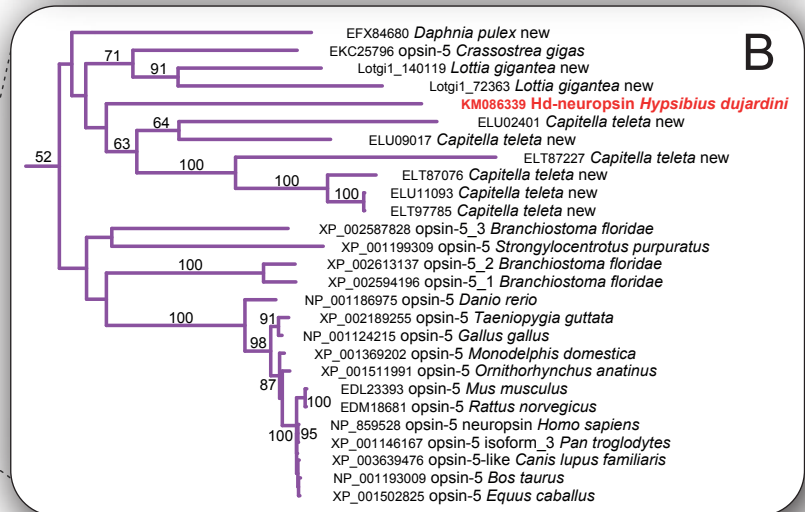
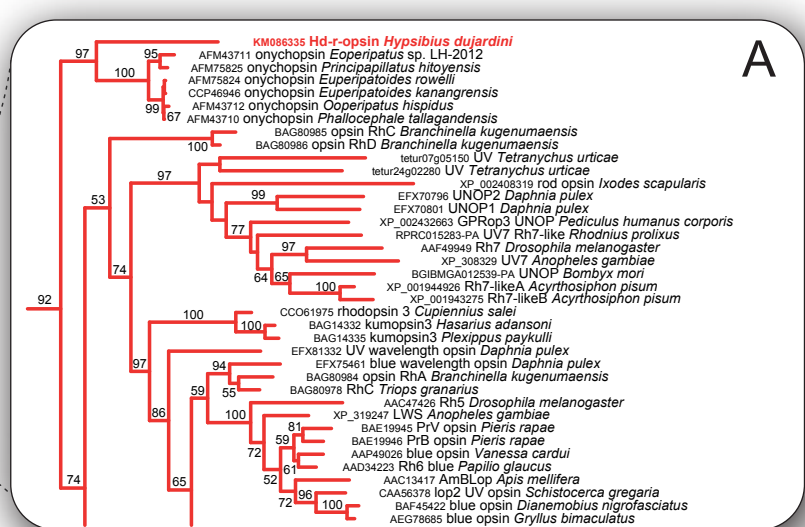
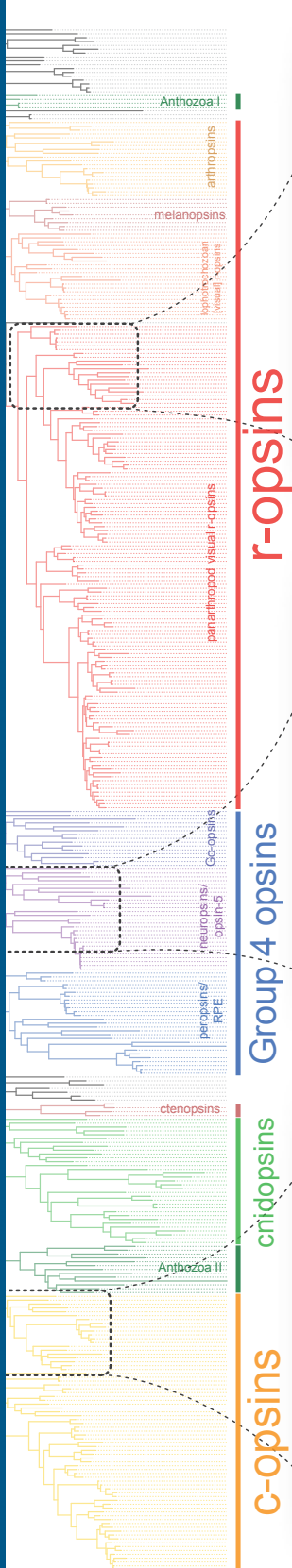
Fig. 1. Visual organs in representatives of Panarthropoda. While most onychophorans and tardigrades have a pair of simple, ocellus-like eyes (arrows in A and B), arthropods usually show two types of visual organs: the compound eyes (arrowheads in C), and the median ocelli (arrow in C). (A) Scanning electron micrograph of the head of the onychophoran *Principapillatus hitoyensis* in dorsal view. Scale bar: 1 mm. (B) Light micrograph of the head of the tardigrade *Hypsibius dujardini* in dorsal view. Scale bar: 10 μ m. (C) Stereomicrograph of the head of the hymenopteran *Ectemnius cavifrons* in dorsal view. Scale bar: 1 mm.

Fig. 2. Simplified best obtained maximum likelihood tree of the full metazoan opsin gene dataset and the placement of the five opsin genes from the tardigrade *Hypsibius dujardini* (highlighted in red). See supplementary figure S1, Supplementary Material online, for the uncondensed tree. (A) Detail view of the bilaterian r-opsin clade, in which *Hd-r-opsin* of *H. dujardini* occurs as the sister group to the onychophoran *onychopsin* clade. (B) Detail view of the neuropsin/opsin-5 subgroup of Group 4 opsins, illustrating the placement of *Hd-neuropsin* of *H. dujardini* among the neuropsins from other protostomes. (C) Detail view of the bilaterian c-opsin clade, in which *Hd-c-opsin1*, *Hd-c-opsin2* and *Hd-c-opsin3* of *H. dujardini* are nested as a monophyletic group within the panarthropod c-opsin clade containing the arthropod pteropsins.

Fig. 3. Hypotheses on the evolutionary history of opsin genes in Panarthropoda and Bilateria based on our phylogenetic analyses of metazoan opsin sequences. (A) Scenario on opsin evolution in Panarthropoda. The trichotomy is due to the uncertain position of Tardigrada as the sister group of Arthropoda, Onychophora, or Onychophora plus Arthropoda (see Mayer et al. 2013 and references therein). According to this scenario, the last common ancestor of

Panarthropoda possessed five opsin genes: one c-opsin, two r-opsins, and two Group 4 opsins. The three c-opsin genes of *H. dujardini* might have evolved by gene duplications either in the tardigrade lineage, or a tardigrade subgroup. (B) Scenario on opsin evolution in Bilateria, according to which the last common bilaterian ancestor possessed six opsin genes. While the arthropsin and melanopsin/[visual] r-opsin lineages arose by a duplication of the r-opsin precursor, two additional duplications in the bilaterian Group 4 opsin clade gave rise to the neuropsin/opsin-5, peropsin, and Go-opsin lineages. Note that the Go-opsin gene was subsequently lost in the last common ancestor of Ecdysozoa.



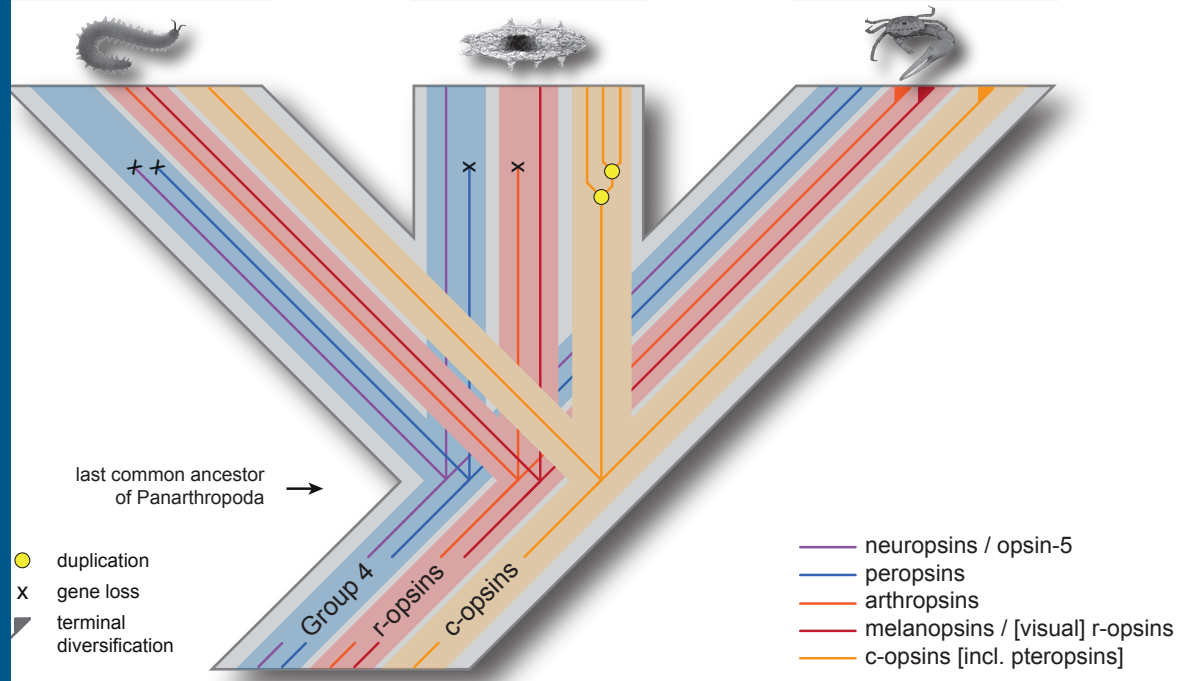


Panarthropoda

Onychophora

Tardigrada

Arthropoda



Bilateria

Deuterostomia

Lophotrochozoa

Ecdysozoa

