RESEARCH ARTICLE   **Open Access**

CrossMark

Structural differences and differential expression among rhabdomeric opsins reveal functional change after gene duplication in the bay scallop, *Argopecten irradians* (Pectinidae)

Anita J. Porath-Krause[1†], Autum N. Pairett[1†], Davide Faggionato[1], Bhagyashree S. Birla[2,4], Kannan Sankar[3,4] and Jeanne M. Serb[1*]

Abstract

Background: Opsins are the only class of proteins used for light perception in image-forming eyes. Gene duplication and subsequent functional divergence of opsins have played an important role in expanding photoreceptive capabilities of organisms by altering what wavelengths of light are absorbed by photoreceptors (spectral tuning). However, new opsin copies may also acquire novel function or subdivide ancestral functions through changes to temporal, spatial or the level of gene expression. Here, we test how opsin gene copies diversify in function and evolutionary fate by characterizing four rhabdomeric (G_q-protein coupled) opsins in the scallop, *Argopecten irradians*, identified from tissue-specific transcriptomes.

Results: Under a phylogenetic analysis, we recovered a pattern consistent with two rounds of duplication that generated the genetic diversity of scallop G_q-opsins. We found strong support for differential expression of paralogous G_q-opsins across ocular and extra-ocular photosensitive tissues, suggesting that scallop G_q-opsins are used in different biological contexts due to molecular alternations outside and within the protein-coding regions. Finally, we used available protein models to predict which amino acid residues interact with the light-absorbing chromophore. Variation in these residues suggests that the four G_q-opsin paralogs absorb different wavelengths of light.

Conclusions: Our results uncover novel genetic and functional diversity in the light-sensing structures of the scallop, demonstrating the complicated nature of G_q-opsin diversification after gene duplication. Our results highlight a change in the nearly ubiquitous shadow response in molluscs to a narrowed functional specificity for visual processes in the eyed scallop. Our findings provide a starting point to study how gene duplication may coincide with eye evolution, and more specifically, different ways neofunctionalization of G_q-opsins may occur.

Keywords: Rhabdomeric photoreceptor, R-opsin, Gene duplication, Melanopsin, Rhodopsin, Vision

* Correspondence: serb@iastate.edu
†Equal contributors
[1]Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames 50011, IA, USA
Full list of author information is available at the end of the article

Porath-Krause *et al. BMC Evolutionary Biology* (2016) 16:250

Page 2 of 15

Background

Organisms detect environmental stimuli using an array of sensory receptors. Changes to the genetic basis of these sensory receptors has been shown to allow organisms to exploit new ecological niches [1] or alter signaling between conspecifics [2], which can affect individual fitness and, ultimately, have evolutionary consequences for the species. Duplication of the genes that code for the sensory receptor proteins is thought to play an important role in expanding the diversity of sensory systems by providing new genetic material for novel phenotypes [3–6]. If gene duplicates are retained, they can follow one of three evolutionary fates (first outlined by [7]; see also expanded models reviewed by [8–10]). First, if both paralogs have the exact same function or suite of functions, the existence of a second copy can increase production levels of encoded protein ("gene conservation" [11]). Under this scenario, the second copy provides functional redundancy that can buffer against neutral loss-of-function mutations over evolutionary time. However, more dramatic functional divergence may occur following the duplication event. In the second scenario, if the original gene managed a suite of functions, such as enzymatic activity and signal transduction, the duplicated copies could subdivide these tasks ("subfunctionalization" [12]). Subfunctionalization of paralogs may include changes in spatial or temporal expression patterns [13] and may release one gene copy from adaptive constraint ("escape from adaptive conflict" model [14]) so that both copies can be optimized for particular tasks [15]. Finally, one copy of the duplicated gene can acquire a novel function while the other copy retains the original, pre-duplication function ("neofunctionalization" [7]).

In photosensory systems, the ability of an animal to become sensitive to a broader range of wavelengths is most often mediated by an increase in the number of opsins [16–22]. Opsins encode a class of G-protein coupled receptors (GPCRs), proteins with seven alpha-helical domains that transverse the cell membrane (helix, H1-7) interspaced by loops that extend into the cytoplasm (cytoplasmic loops, CL1-3) and outside of the photoreceptive cell (extracellular loops, EC1-3). Opsins covalently bind a light-absorbing vitamin-A derived chromophore, such as 11-*cis*-retinal, using a lysine residue in H7. Together, the opsin protein and chromophore molecule form a photopigment sensitive to a specific portion of the light spectrum. Photopigments are often characterized by the wavelength at which the absorbance of light is the greatest (λ_{max}). When 11-*cis* retinal absorbs a light photon, it isomerizes to an all-*trans* state. As a result, the opsin undergoes a conformational change and releases a complex of heterotrimeric guanine nucleotide-binding proteins (G-proteins), which

are specific to that opsin (reviewed in [23]). The dissociated alpha-subunit of the G-protein activates the phototransduction cascade through second messenger molecules. Depending on the particular transduction pathway initiated by opsin, the photoreceptor cell may either hyperpolarize (e.g., G_t-protein coupled opsins in ciliary cells) or depolarize (e.g., G_q-protein coupled opsins in rhabdomeric cells) [24]. Opsin specificity to its G-protein partner is regulated by G-protein binding sites [25] and is associated with particular amino acid motifs in the fourth cytoplasmic loop [26]. Phylogenetically, opsins group into clades based, in part, by the G-protein partner and to a lesser extent by photoreceptor type (rhabdomeric versus ciliary cells) [27, 28].

Because a photopigment can only absorb a portion of the light spectrum, increasing the number and diversity of opsins through gene duplication and divergence allows an expansion of the photoresponse to new wavelengths of light. This may lead to color discrimination, if the photopigments have different light sensitivities. Under this neofunctionalization model, changes in the amino acid residues at positions that interact with the chromophore (e.g., "spectral tuning sites") shift the wavelength at which absorbance is the greatest (λ_{max}) of the duplicated visual pigment. Thus, the potential advantages for organisms with multiple and genetically diverse photopigments include extending the range of spectral perception, new functionality under different light conditions, generation of wavelength-specific behaviors, or providing the molecular substrate in the retina for color vision (reviewed in [29]). Any of these phenotypes may allow an animal to occupy new or more heterogeneous photic niches [30, 31].

While it is well-documented that duplicated opsin genes most often attain a new λ_{max} by neofunctionalization [32–40] it is less understood what other phenotypic outcomes may follow the duplication of opsin genes (but see [21]). Photoreceptors in invertebrates occur in multiple tissue types and in different life stages, and can function as both ocular and extra-ocular sensory receptors [41–46]. Thus, in invertebrates, neofunctionalization of opsins may include co-option between tissues, organs, or life stages after a gene duplication event. In order to distinguish among different evolutionary outcomes of opsin duplication and what effect gene duplication may have in the evolution of the photoreceptive cells and organs in a given system [47], it is necessary to first identify and then characterize the diversity of opsin proteins that are present.

Here, we assess the evolutionary history of G_q-opsins in scallop to examine the role of gene duplication in producing extant diversity. The molecular basis of photoreception in the scallop is complex. The mirror-type eyes of scallops contain at least two different

Porath-Krause *et al. BMC Evolutionary Biology* (2016) 16:250

Page 3 of 15

phototransduction systems based on opsins that presumably couple with G_o- and G_q-proteins [48]. Previously, we identified a duplication event of scallop G_q-protein coupled opsins that occurred over 230 Mya [49]. Because gene copies with identical gene function are unlikely to be maintained in the genome unless the new duplicate is advantageous [50], the long-term retention of these opsin duplicates in the scallop lineage suggests a fitness cost if the copies are not maintained. For these duplicates to persist over evolutionary time, opsin copies must have diverged phenotypically under one or more of the evolutionary fate models described above. To test this hypothesis, we determined the evolutionary fates of these duplicated scallop opsins. We first captured the genetic diversity of G_q-protein coupled opsin genes (herein *opnGq* for the gene or the coding region, and OPNGq for the protein) by generating transcriptomes of photosensitive tissues from adult animals and placed the genetic diversity of scallop G_q-opsins into an evolutionary framework by employing a phylogenetic analysis. We next asked how might these scallop OPNGq proteins interact with a chromophore. To do so, we capitalized on the x-ray crystallography data from the squid OPNGq ("squid rhodopsin") [51, 52] to model the tertiary structure of the scallop OPNGqs. Then, we examined if the protein characteristics of each paralog differ. As a first approximation to identify differences in λ_{max} among scallop G_q-opsins, we leveraged existing computational models that estimate electrostatic interactions between the amino acids and the chromophore of squid OPNGq and applied them to the scallop data. Finally, we examined differences in gene expression of *opnGq* paralogs across both ocular and extra-ocular photoreceptive organs. From these lines of evidence, we show that scallop G_q-opsin paralogs differ in 1) the biochemical properties of amino acid residues interacting with the chromophore; 2) expression levels of the gene; and 3) spatial expression of the gene among light-sensitive tissues in the adult organisms.

Methods

Transcriptome assembly and gene analyses

Thirty-six adult individuals of the bay scallop, *Argopecten irradians* (Pectinidae), were collected from the Gulf of Mexico near Sanibel, Florida during July, 2012. The adults were kept in recirculating saltwater tanks under a light regime of 13 h of light and 11 h of dark per 24-h cycle. To maximize the likelihood of capturing all G_q-opsin transcripts expressed, we collected tissues under both light and dark treatments (nine hours of light vs. nine hours of dark), with the expectation that the highest level of opsin expression would occur nine hours after sunrise [53, 54]. The tissues from dark-treated scallops were dissected under red-light. All eyes from the

left and right mantles were collected and pooled for each animal (~60 eyes/individual). Small sections of mantle tissue were sampled along the anterior-posterior axis from both left and right valves and pooled for each individual. A portion of adductor muscle equivalent in volume to the dissected eye tissue was collected from each individual. RNA was extracted from the three tissue types using the Ambion RiboPure RNA extraction kit (Life Technologies). RNA samples from the tissues of one light-treated and one dark-treated individual were sent to the Iowa State University DNA Facility for library creation and transcriptome sequencing on an Illumina HiSeq2000. Nearly 1.5 trillion 100 base pair (bp) paired-end reads were generated from six libraries: light/dark eyes, light/dark mantle, and light/dark adductor. A *de novo* assembly of a reference transcriptome from all six libraries was created in the Trinity sequence assembly and analysis pipeline [55] by first normalizing the raw reads to remove redundancy with the Trimmomatic script, then assembling the quality trimmed reads. This assembly resulted in 231,391 transcripts with a contig N50 of 2078 and an average contig length of 971 bp. The assembled transcriptome data was given the reference name of "AirradFL." Opsin sequences from two other scallop species [56] were used as queries to identify Gq-opsin sequences in the AirradFL reference transcriptome using BLAST. Putative opsin sequences from the AirradFL reference transcriptomes were blasted back to the NCBI nonredundant (nr) database to further confirm the sequence identities. Gene and protein nomenclature follows the general guidelines in invertebrate model organisms (e.g., http://www.wormbase.org), where gene and transcript names (italicized) are composed of a three-letter species prefix, followed by a hyphen, the class (homolog) of the gene, and a number (e.g., *Air-opnGq1*). The number provides the order of gene discovery of paralogs within a species or lineage. Proteins use the gene name, with the gene abbreviation without italics and in all uppercase (e.g., Air-OPNGq1).

Phylogenetic analysis

To determine the phylogenetic placement of putative scallop G_q-opsins, we compiled G_q-opsin sequences from genomes, transcriptomes or single genes from public databases at Genbank (http://www.ncbi.nlm.nih.gov/genbank/) and assembled data from Porter et al. [27] (Additional file 1: Table S2). We queried all five publically-available molluscan genomes for additional G_q-opsins: pearl oyster, *Pinctada fucata* (June, 2013); Pacific oyster, *Crassostrea gigas* (June, 2013); freshwater snail, *Biomphalara glabrata* (June, 2013); owl limpet, *Lottia gigantea* (June, 2013); and sea hare, *Aplysia californica* (June, 2013). G_q-opsin sequences were found by blasting scallop opsins against predicted gene models

Porath-Krause *et al. BMC Evolutionary Biology* (2016) 16:250

Page 4 of 15

from each molluscan genome using tblastx and an E-value cutoff of 1e-3. When gene models were not available, the genome contigs/scaffolds were used. The putative G_q-opsins identified through BLAST were then reciprocally blasted back to the NCBI nonredundant (nr) database and subjected to phylogenetic analyses with known metazoan G_q-opsins to confirm their identity.

Amino acid sequences of the 96 opsins from 42 taxa, including four annelids, 38 arthropods, 21 molluscs, and six platyhelminthes, (Additional file 1: Table S2) were aligned using MAFFT v 7.017 [57] as implemented in Geneious (v5.6.7). (http://www.geneious.com). This dataset included opsins from the G_i- and G_o-opsin families to test the monophyly of the G_q-opsin clade. The G_o-opsin from *Argopecten irradians* was used to root the phylogeny. The aligned dataset was then manually trimmed to remove long C- and N-terminus sequences and remove a single large (>50 aa) gap around position 258 in the H6. The trimmed, aligned dataset contained 355 amino acids. The best-fit model of protein evolution for this dataset was determined using ProtTest [58], which found the LG + G + I + F model [59] to have the lowest Akaike Information Criteria score (AIC). A maximum likelihood (ML) phylogeny of the aligned dataset was constructed using Randomized Axelerated Maximum Likelihood (RAxML) v 8 [60]. Node support was calculated using 1000 rapid bootstrap replications as implemented in RAxML. Using the same model of protein evolution, we also analyzed the data under Bayesian inference using MrBayes v3.2.6 [61] on the XSEDE tool available through the CIPRES Science Gateway [62]. We used the Metropolis Coupled Markov Chain Monte Carlo method with one cold and three hot chains for 3.1 million generations with a burnin of 1000 for two independent runs. Convergence was determined when the potential scale reduction factor (PSRF) approached 1.

PCR confirmation of scallop opsin transcripts

All opsin transcripts were confirmed to be single genes by PCR amplification of the complete coding region with UTR-specific primers from both cDNA and genomic DNA (Qiagen DNeasy Blood and Tissue kit) (Additional file 2: Table S1). PCR products were size-screened using agarose gel electrophoresis, bands of expected size were gel extracted (Qiagen Qiaquick Gel Extraction kit) and cloned using chemically competent *E. coli* cells (TOPO TA Cloning Kit with pCR2.1-TOPO). Positive colonies from blue-white screening were Sanger sequenced using an ABI 3730 Capillary Electrophoresis Genetic Analyzer at the Iowa State University DNA Sequencing Facility. The resulting sequences were translated and compared against contigs from the transcriptome. Using the same approach, we confirmed that a large contig sequence containing two G_q-opsin transcripts (*Air-opnGq3* and

Air-opnGq4) and an intergenic region of ~1690 bp was present in the genome. Because repetitive motifs can indicate gene duplication due to transposable elements [63], we searched for repetitive motifs in this intergenic region. To do so, the nucleotide sequence of the whole contig was screened with the RepeatMasker Web server v open-4.0.5 (http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker) using the cross_match search engine on slow speed/sensitivity and the bivalves *Crassostrea gigas*, *Pinctada fucata*, and *Mizuhopecten yessoensis* as DNA sources.

Homology modeling of scallop G_q-opsins

To identify amino acid changes that may result in functional differences among scallop G_q-opsins, we compare the Air-OPNGqs to the only molluscan opsin with a resolved crystal structure, the *Todarodes pacificus* "rhodopsin" (Tpa-OPSGq1; Genbank accession X70498) [51] We followed the amino acid numbering system of the squid where the first amino acid position in our alignment begins with the start codon (Met) of Tpa-OPNGq1. To examine the degree of resemblance among protein sequences, we calculated pairwise percent similarity of the scallop and squid amino acid sequences in the BLASTP 3.2.1 [64, 65] at NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins).

We also used the protein alignment to identify amino acid residues that may interact with the chromophore. We applied a quantum mechanics/molecular mechanics model based on the crystal structure of Tpa-OPNGq1 [66], which predicts the involvement of 38 sites in spectral tuning of G_q-opsins. We examined differences in the Air-OPNGq and Tpa-OPNGq1 sequences at these sites and noted changes in the biochemical properties of the residues.

Next we employed bioinformatic homology modeling to predict the tertiary structure of the four scallop G_q-opsin proteins. These models were based on the template of the only available crystal structure for a G_q-opsin, the rhodopsin from squid *Todarodes pacificus* 2ZIY [52]. The tertiary structure models of four scallop opsins (Air-OPNGq1, Air-OPNGq2, Air-OPNGq3 and Air-OPNGq4) were predicted using the Iterative Threading ASSEmbly Refinement (I-TASSER) server [67, 68]. The squid 2ZIY template was used to retrieve model proteins of similar folds from the Protein Data Bank (PDB) library using a locally installed meta-threading library. The continuous fragments excised from PDB templates were re-assembled into full-length models by replica- exchange Monte Carlo simulations and the unaligned regions were built by *ab-initio* modeling. The structure was then further refined with a second fragment assembly simulation. No restraints such as inter-residue contacts or inter-residue distances were specified

Porath-Krause *et al. BMC Evolutionary Biology* (2016) 16:250

Page 5 of 15

for the modeling. For each G_q-opsin, the top five predicted structures from I-TASSER were used for further quality assessment.

Assessing the quality of the modeled tertiary structures

The quality of the modeled structures was assessed using the Ramachandran plot and the confidence score (C-score) (Additional file 3: Table S3) from the I-TASSER server. The Ramachandran plot is a graph of the backbone dihedral angles ψ against ϕ of the amino acid residues in the structure. Good quality models have more than 90% of the residues in allowed regions (i.e. most favored and additionally allowed regions) of the Ramachandran plot. The Ramachandran plot of the modeled structures was obtained using PROCHECK [69] which has been implemented as part of the PDBSum Server [70].

The C-score (from I-TASSER server) is a scoring function to rank models based on their quality and is defined using the significance of threading template alignments and the convergence parameters of the structure assembly simulations (for more details see [67]). C-scores are typically between −5 and 2 with higher values representing better models. However it has been observed that the C-score is particularly low (and negative) for membrane proteins. The "best" models of the four G_q-opsin sequences were selected based on the highest C-score and maximum percentage of residues in the most favored and generously allowed regions according to the Ramachandran plots.

To quantify the overall shape differences among Gq-opsin tertiary structures, we performed a whole-molecule comparison between the predicted tertiary models calculating the Root-Mean-Square Deviation (RMSD) of the atomic positions of the alpha carbons between one opsin against each other. RMSD provided a quantitative computation of the average distance between the backbone atoms of two superimposed proteins. Variation in Air-OPNGq sequence length did not impact the RMSD values because a small portion of the N- and most of the C-termini were truncated from each sequence so the comparison occurs only between superimposed atoms. For RMSD comparison, only common one-to-one aligned residues, were included (V19 to K342). The values between each pair of structures were calculated using the standard 'align' program in PyMOL (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC). Lower RMSD values indicate a higher similarity between structures.

Scallop gene expression data

Paired-end RNA-seq data for three scallop tissues (eye, mantle, and adductor muscle) from the light treatment were aligned against the AirradFL assembly (nonredundant set of 231,391 transcripts grouped into 176,417 "genes") using Bowtie v. 1.0.1 [71] followed by read abundance estimation with RSEM v. 1.2.9 [72] through the Trinity sequence and assembly pipeline v. 2013_2-25 [55]. Relative levels of expression in Fragments Per Kilobase per Million fragments mapped (FPKM) for a given transcript were calculated using the Trinity toolkit v. 2013_2-25 [55]. We accepted expression levels for a given transcript when the FPKM value was equal to or greater than one as a conservative approach to compare levels of relative expression among tissue types. Because tissues under the light treatment had the greatest levels of G_q-opsin expression, only the results from light-treated tissues are reported here.

Oyster gene expression data

To compare interspecific differences in G_q-opsin expression patterns between bivalve taxa, opsin gene expression data for the Pacific oyster, *Crassostrea gigas*, were mined from the oyster genome database (OysterBase, http://www.oysterdb.com). We identified *opnGqs* from oyster by blasting our scallop G_q-opsins against the database using the OysterBase blast tools with default settings. Gene expression data in RPKM (Reads Per Kilobase per Million) of the oyster G_q-opsins (*Cgi-opnGqs*) were curated for each adult tissue type (digestive gland, gills, gonad, hemolymph, labial palp, mantle, and pallial mantle) and larval life stages (trochophore, D-shape larva, umbo larva, and pediveliger) from the website (OysterBase, http://www.oysterdb.com) and supplementary data tables (Table S12, S14) in Zhang [73]. However, comparing gene expression changes between the oyster (in RPKM) and scallop (in FPKM) tissues could only be described in relative terms.

Results

Transcriptomic and phylogenetic analyses reveal four G_q-opsin genes in scallop

To determine the number of G_q-opsin genes in scallop, we performed deep transcriptome sequencing of tissue-specific libraries derived from dissected eyes, mantle tissue, and adductor muscle of *Argopecten irradians*. From light and dark treated animals, four transcripts were identified as putative *opnGqs* using a similarity-based analysis pipeline described in Pairett and Serb [56], which we named *Air-opnGq1*, *Air-opnGq2*, *Air-opnGq3*, and *Air-opnGq4* with ascending numbering according to the history of discovery (GenBank accession numbers KT426908, KT426909 KT426910, and KT426911). Visual inspection of the back mapped reads to each identified G_q-opsin sequence did not show any obvious misassembled regions or mismatches. The proteins varied in amino acid percent similarity (the ratio of residues with similar physio-chemical properties shared between two sequences), which were the greatest between Air-

Porath-Krause *et al. BMC Evolutionary Biology* (2016) 16:250

Page 6 of 15

OPNGq2 and Air-OPNSGq3 at 80.9%, and lowest between Air-OPNGq1 and Air-OPNGq4 (72.9%) (Table 1). Amino acid percent similarity was more conserved between the aligned Helix 1 (H1) through H7, and ranged from 92.6% (Air-OPNGq2 versus Air-OPNGq3) to 76.9% (Air-OPNGq1 versus Air-OPNGq4) (Table 1). Transcripts also differed in the sequence length from the first Met codon to the beginning of H1 (35–49 amino acids) and between the end of H7 and the stop codon (135–184 amino acids) (Fig. 1; Table 2).

To determine how Air-OPNGqs were evolutionarily related to other G_q-opsins, we conducted a phylogenetic analysis of their translated amino acid sequences with 96 metazoan opsins (Additional file 1: Table S2). Under both maximum likelihood and Bayesian inference, all four scallop sequences belonged to a clade that included G_q-opsins from four other bivalve species: two oysters (*Pinctada fucata*, *Crassostrea gigas*) and two additional scallops (*Placopecten magellanicus*, *Mizuhopecten yessoensis*) (Fig. 2, green box). Within this clade, there was one difference between the ML and BI topologies, where ML placed the two oyster OPNGq1s as the sister group to the scallop G_q-opsins 2–4, and the BI topology placed all bivalve OPNGq1s in a single clade (grey box in Additional file 4: Figure S1). However, values supporting these relationships were low (47% bootstrap support; 54 posterior probability). The bivalve-specific G_q-opsin clade (OPNGq1-4) was the sister group to a clade of opsins from cephalopod and gastropod molluscs, and part of a larger clade of well-characterized vertebrate (e.g., melanopsin) and arthropod (e.g., *Drosophila* rhodopsin) G_q-opsins (Fig. 2). A second molluscan G_q-opsin clade was also recovered which contained oyster and gastropod opsins, but no scallop opsins (Fig. 2, red box). A complete, uncollapsed ML phylogram is available as a supplemental document (Additional file 5: Figure S2).

We then asked whether the four scallop G_q-opsins possess the specific amino acid residues and sequence motifs required for photosensitivity. In addition to the seven transmembrane α-helices, it has been experimentally demonstrated that G_q-opsin proteins require certain sequence motifs to maintain structural integrity and bind to the chromophore [74]. These include: 1) two Cys residues in the TM3 and EC2 domains that are involved in disulfide bond formation, 2) a Glu180 in the EC2 that functions as a counter ion to the positive charge of the protonated Schiff base [75], 3) a E/DRY motif near the TM3/CL2 boundary that helps stabilize the inactive-state conformation [76], 4) Asn87 and Tyr111 residues that are hydrogen binding partners for the protonated Schiff base [52], 5) a lysine residue in TM7 that is covalently linked to the chromophore, and 6) a conserved NPxxY motif in the TM7 [74]. We found that all four scallop proteins were invariant for the expected amino acid residues and motifs needed for correct conformation with the exception of the E/DRY motif (Table 2). This motif was variable among the scallop opsins, where Y134C in Air-OPNGq2 and Y134F in Air-OPNGq3 and Air-OPNGq4. In addition, we examined a motif (positions 319–321) in the fourth cytoplasmic loop, which has been experimentally demonstrated to be important for opsin-G_t-protein interactions (positions 310–312 in bovine rhodopsin) [25]. Three of the four scallop opsins contain a HPK motif, an evolutionary conserved sequence that appears to be specific to G_q-protein binding [77] (Table 2). Air-OPNGq4 had a HPR motif, but R has similar biochemical properties to K. Based on these data, we conclude that the four transcripts are indeed OPNGqs possessing the amino acid residues required for molecular stabilization, chromophore binding, and G-protein interaction and thus likely form photopigments.

G_q-opsin transcripts are not the result of alternative splicing

To determine whether the four different *opnGq* transcripts were the result of alternative splicing of the same gene, we developed target-specific primers (Additional file 2: Table S1) from the flanking UTR sequences for each *Air-opnGq*. We then compared these sequences derived from genomic DNA (gDNA)
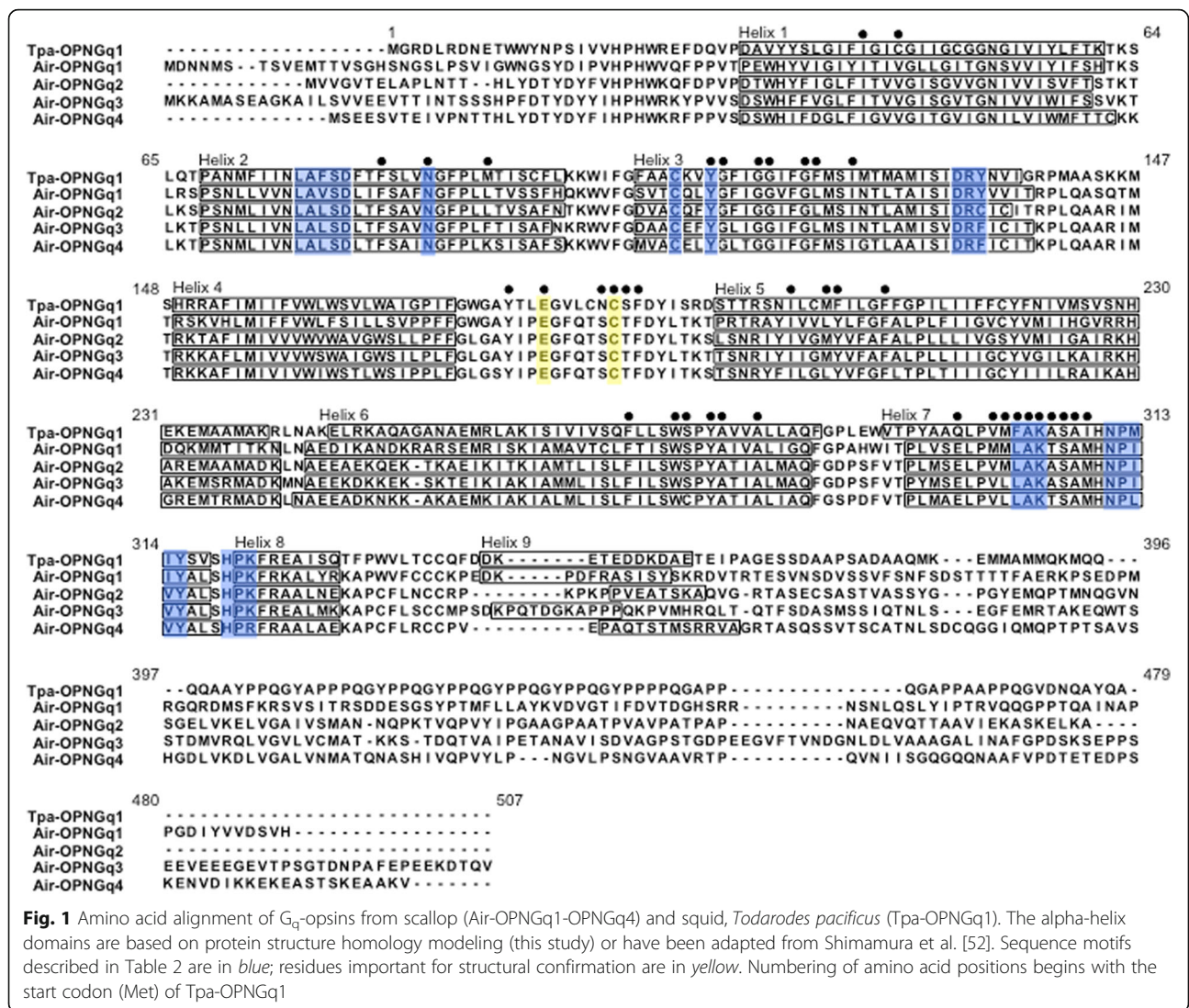
Table 1 Percent similarity (below diagonal) and RMSD (above diagonal) of scallop (Air) and squid (Tpa) proteins

	Air-OPNGq1	Air-OPNGq2	Air-OPNGq3	Air-OPNGq4	Tpa-OPNGq1
Air-OPNGq1	-	0.378[b]	**0.354**	0.489	0.589
Air-OPNGq2	74.7 (78.9) [a]	-	0.408	0.603	0.503
Air-OPNGq3	74.7 (77.9)	**80.9 (92.6)** [c]	-	0.699	0.601
Air-OPNGq4	72.9 (76.9)	76.9 (85.6)	74.6 (88.1)	-	0.549
Tpa-OPNGq1	**71.0 (72.4)**	73.4 (73.8)	75.2 (75.8)	73.3 (74.7)	-

[a] Percent similarity of amino acid sequence alignments from first methionine to stop codons; values in parentheses are percent identity from Helix 1 through Helix 7
[b] Atomic values in angstroms, where the lower RMSD value, the higher is the similarity between structures
[c] Numbers in bold indicate minimum and maximum values

Fig. 1 Amino acid alignment of G$_q$-opsins from scallop (Air-OPNGq1-OPNGq4) and squid, *Todarodes pacificus* (Tpa-OPNGq1). The alpha-helix domains are based on protein structure homology modeling (this study) or have been adapted from Shimamura et al. [52]. Sequence motifs described in Table 2 are in *blue*; residues important for structural confirmation are in *yellow*. Numbering of amino acid positions begins with the start codon (Met) of Tpa-OPNGq1

to transcripts derived from the transcriptomes. Alignments of 5′- and 3′-UTR DNA sequences and coding regions were identical between the transcripts and gDNA templates (data not shown). The flanking UTR sequences were not conserved and could not be

unambiguously aligned across the four *Air-opnGqs* (Additional file 6: Figure S3).

While three of the four *Air-opnGq* sequences lacked introns, we identified a 393 bp intron within the region coding of H3 that was unique to *Air-opnGq1*.

Table 2 Sequence and structural motifs in scallop (Air) and squid (Tpa) G$_q$-opsins

Motifs	Air-OPNGq1 493 aa	Air-OPNGq2 456 aa	Air-OPNGq3 519 aa	Air-OPNGq4 481 aa	Tpa-OPNGq1 448 aa
LxxxD TMII (pos 76–80)	LAVSD	LALSD	LALSD	LALSD	LAFSD
Disulfide bond	C108, C186	C108, C186	C108, C186	C108, C186	C108, C186
Hydrogen bond with Schiff base	N87, Y111	N87, Y111	N87, Y111	N87, Y111	N87, Y111
E/DRY TMIII (pos 132–134)	DRY	DRC	DRF	DRF	DRY
Counterion	E180	E180	E180	E180	E180
LAK TMVII (pos 305–307)	LAK	LAK	LAK	LAK	FAK
NPxxY TMVII (pos 311–315)	NPIIY	NPIVY	NPIVY	NPLVY	NPMIY
G-protein binding (pos 319–321)	HPK	HPK	HPK	HPR	HPK

The amino acid numbering system follows the amino acid position (pos) of squid rhodopsin

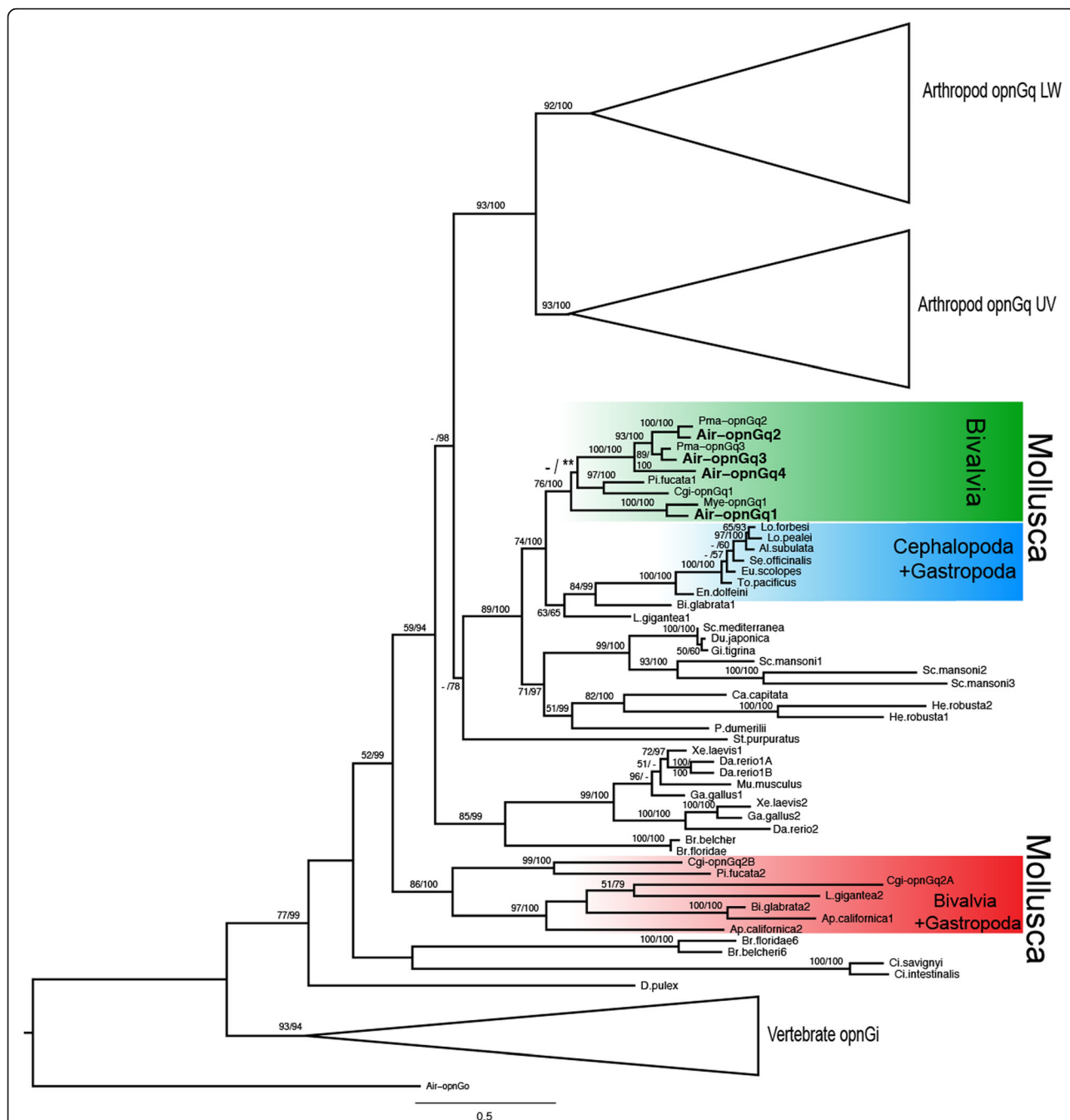Porath-Krause *et al. BMC Evolutionary Biology* (2016) 16:250

Page 8 of 15

Fig. 2 Maximum likelihood (ML) topology of G$_q$-opsins. The phylogenetic tree is based on aligned amino acid sequences with scallop G$_o$-opsin as the outgroup. Support values (>50%) of nodes were generated by 1000 bootstrap replicates in RAxML. Support values after the '/' are posterior probabilities from a Bayesian phylogenetic analysis (BI). Support values <50% are indicated by a '-'. The single difference between the ML and BI topologies occurs within the bivalve G$_q$-opsin clade (*green*) and is highlighted with an asterisk '*'. *Argopecten irradians* G$_q$-opsins (Air-OPNGqs) from this study are in bold. Two molluscan G$_q$-opsin clades were recovered, but only one clade (*green*) contained scallop G$_q$-opsins from *Argopecten irradians* (Air-OPNGqs), *Mizuhopecten yessoensis* (Mye-OPNGqs), or *Placopecten magellanicus* (Pma-OPNGqs). Two large clades of arthropod opsins that represent UV and long-wavelength (LW) opsins and a vertebrate G$_i$–opsin clade were collapsed for space

Additionally, gDNA sequencing determined that *Air-opnGq3* and *Air-opnGq4* were located in tandem, but in reverse orientation, with a 1690 bp intergenic region between the two coding regions. No repeat regions or putative transposable elements were identified in the intergenic region (data not shown). Variation in intron pattern and UTR sequences among the G$_q$-opsins indicates that these four genes are most likely located on

different physical places in the genome and are four separate loci.

Predicted tertiary structure and chromophore-associated residues differ among scallop G_q-opsins
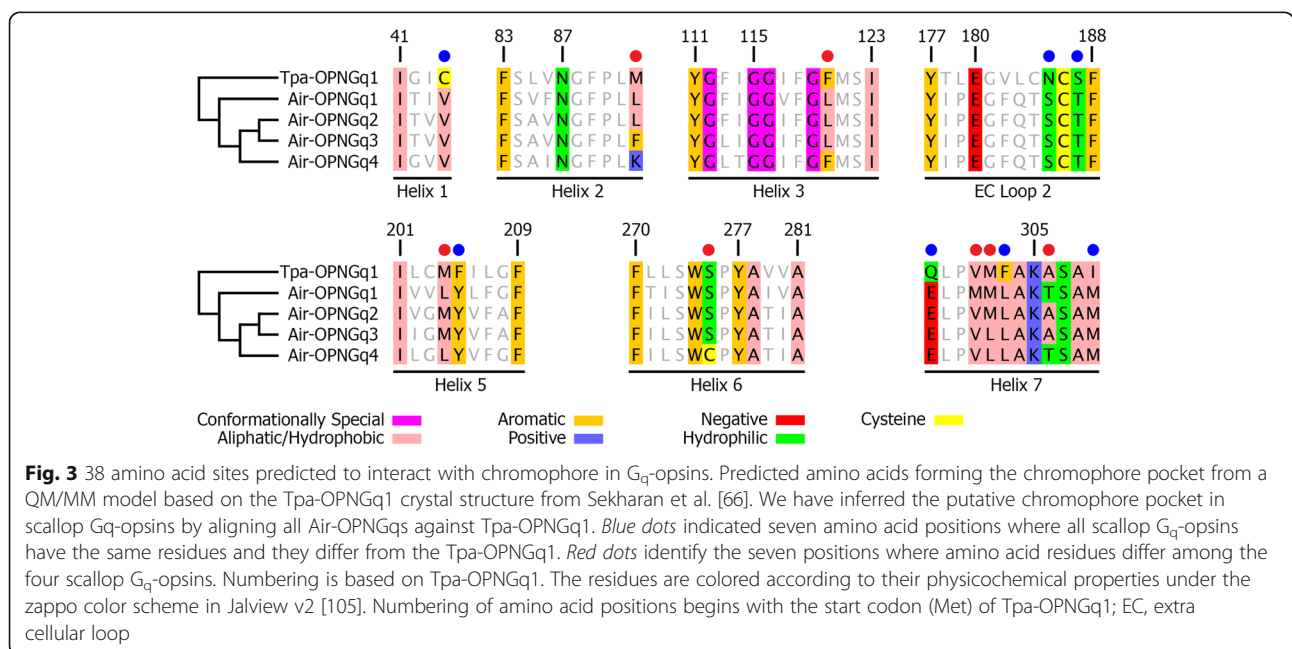
We generated three-dimensional models for each Air-OPNGq using crystallography data from the squid "rhodopsin" [52] as a template for homology models. This allowed us to examine differences in the tertiary structure among the four G_q-opsin sequences. The best model for each Air-OPNGq was selected based on the highest C-score and maximum percentage of residues in the most favored and generously allowed regions according to the Ramachandran plots (Additional file 3: Table S3). To quantify the overall shape differences among G_q-opsin tertiary structures, we performed a whole-molecule comparison between the predicted tertiary models calculating the Root-Mean-Square Deviation (RMSD) of the atomic positions of the alpha carbons between one opsin against each other. Based on the RMSD of atomic values, tertiary structures differed from 0.354 to 0.699 Å, where lower RMSD values indicate higher similarity between structures (Table 1). Predicted tertiary structures were the most similar among Air-OPNGq1, Air-OPNGq2, and Air-OPNGq3 proteins (RMSD ranged between 0.354 and 0.408), while Air-OPNGq3 was most different from Air-OPNGq4 (RMSD = 0.699) (Table 1). Air-OPNGq3 and Air-OPNGq4 are more different in tertiary structure from each other than either are to squid rhodopsin (RMSD = 0.503 and 0.601).

We then examined if the positions predicted to interact with the chromophore differ in their residues among the four scallop G_q-opsins. We employed results from a quantum mechanics/molecular mechanics (QM/MM) model based on the Tpa-OPNGq1 crystal structure [66]. This model predicts 38 amino acid sites that may play a role in spectral tuning of G_q-opsins. The scallop G_q-opsins differed from the Tpa-OPNGq1 at seven of the 38 positions, but only three of these had residues with another biochemical property (Fig. 3, blue dots). Among the four scallop G_q-opsins, seven of the 38 positions varied (Fig. 3, red dots). At four positions, at least one of the scallop opsins had an amino acid residue with a different biochemical property. Position 92 was the most divergent among Air-OPNGq proteins and included nonpolar aliphatic/hydrophobic (Air-OPNGq1 and Air-OPNGq2) and aromatic residues (Air-OPNGq3), while Air-OPNGq4 had a positive polar residue (Lys) at this position. At position 275, a conserved serine was substituted by cysteine in Air-OPNGq4, and at position 306, adjacent to the lysine forming the Schiff base, Air-OPNGq1 and Air-OPNGq4 have an hydrophilic residue instead of an hydrophobic/aliphatic residue (Fig. 3).

G_q-opsins are differentially expressed across the eye, mantle and adductor muscle tissues

To determine whether the expression patterns from the four G_q-opsins in *A. irradians* differ spatially, we compared the relative expression level of each G_q-opsin among the six tissue-specific transcriptomes from adult animals collected after a nine-hour light treatment or a nine-hour dark treatment. We found that spatial expression of the four G_q-opsins was consistent in the light and dark adapted animals (data not shown); however,

Fig. 3 38 amino acid sites predicted to interact with chromophore in G_q-opsins. Predicted amino acids forming the chromophore pocket from a QM/MM model based on the Tpa-OPNGq1 crystal structure from Sekharan et al. [66]. We have inferred the putative chromophore pocket in scallop Gq-opsins by aligning all Air-OPNGqs against Tpa-OPNGq1. *Blue dots* indicated seven amino acid positions where all scallop G_q-opsins have the same residues and they differ from the Tpa-OPNGq1. *Red dots* identify the seven positions where amino acid residues differ among the four scallop G_q-opsins. Numbering is based on Tpa-OPNGq1. The residues are colored according to their physicochemical properties under the zappo color scheme in Jalview v2 [105]. Numbering of amino acid positions begins with the start codon (Met) of Tpa-OPNGq1; EC, extra cellular loop

Porath-Krause *et al. BMC Evolutionary Biology* (2016) 16:250

Page 10 of 15

tissues under the light treatment had the greatest levels of G_q-opsin expression and we only the report these results here.

We found all four scallop G_q-opsins were expressed in the eye. Outside of the eye, both *Air-opnGq1* and *Air-opnGq2* were expressed in the mantle, but only *Air-opnGq2* was expressed in the adductor muscle at levels above our expression threshold (≥1.0 FPKM; Fig. 4). As a general pattern across all tissue types, *Air-opnGq2* had the highest expression levels, while *Air-opnGq4* was expressed at the lowest level or not at all. When comparing relative expression levels in the eye, *Air-opnGq2* and *Air-opnGq3* had the highest relative expression levels with *Air-opnGq2* expression (10,001.27 FPKM) at ~38 times higher than *Air-opnGq3* (260.64 FPKM), 275-times higher compared to *Air-opnGq1* (36.46 FPKM), and over 5800-times higher *Air-opnGq4* (1.72 FPKM) (Fig. 4).

We then examined relative levels of gene expression in the Pacific oyster (*Crassostrea gigas*). Since this species is eyeless as an adult, we anticipated that its genome would contain a limited number of G_q-opsins. However, our analyses identified three different G_q-opsins in the *C. gigas* genome (*Cgi-opnGq1, Cgi-opnGq2A,* and *Cgi-opnGq2B*) that showed a degree of differential expression across tissues and life stages. *Cgi-opnGq1*, the oyster G_q-opsin most closely related to the scallop opsins identified here (Fig. 2, green box), was found to have low (<1.0 RPKM) expression levels across the adult oyster tissues, but relatively higher expression in the larval umbo (2.508 RPKM) and pediveliger (21.355 RPKM) stages. *Cgi-opnGq2A* and *Cgi-opnGq2B* belonged to a second clade of gastropod and bivalve G_q-opsins (Fig. 2, red box). *Cgi-opnGq2A* was most highly expressed in the adult tissues, with the labial palp (organs that move food to the mouth for ingestion) and pallial mantle (the tissue most similar to the scallop eye-containing mantle edge) showing the greatest *Cgi-opnGq2A* expression (2.290 RPKM and 4.080 RPKM, respectively). *Cgi-opnGq2B* showed the lowest expression across all tissues and life stages (<1.0 RPKM).

Discussion

The duplication of opsin genes is considered to be an important mechanism for the expansion of light-sensing capabilities of photosensory systems by either enhancing wavelength discrimination or increasing the spatial expression. While some of the best studied examples of photosensitivity expansion are the separate origins of
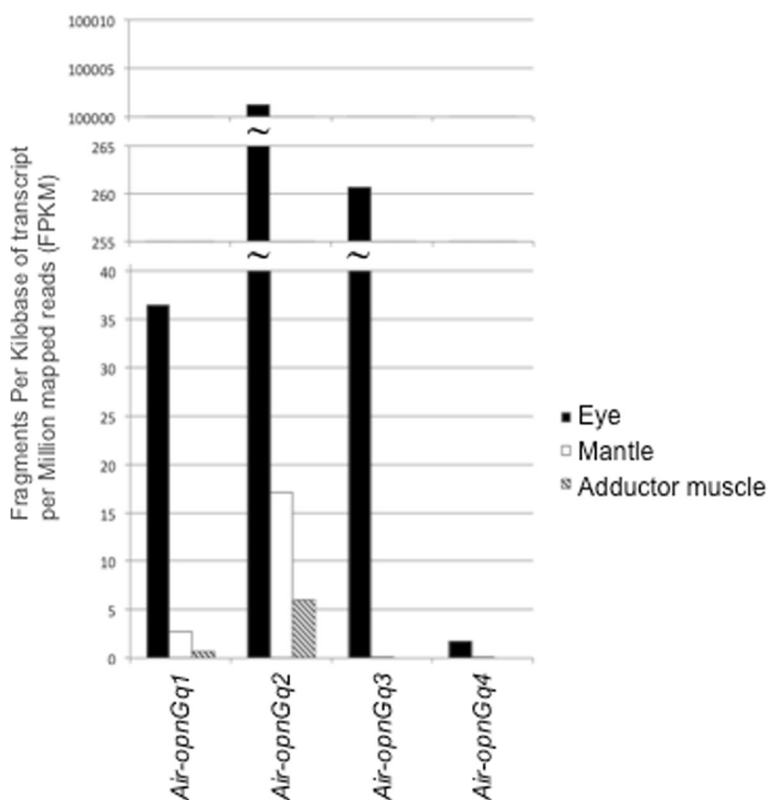

Fig. 4 Expression profiles of scallop G_q-opsin genes across three tissues from a single light-treated animal. Gene-specific mRNA levels were quantified using RNA-seq of tissue-specific libraries: eye (*black*), mantle (*white*), and adductor muscle (*striped*). Expression levels are reported in Fragments Per Kilobase of transcript per Million mapped reads (FPKM)

Porath-Krause *et al. BMC Evolutionary Biology* (2016) 16:250

Page 11 of 15

color vision in insects [22, 42, 78] and vertebrates [17, 79, 80], where shifts in absorbance spectra are attributed to nonsynonymous substitutions to the coding region of one opsin copy, post-duplication fates of opsins need not be limited to changes in the coding region. Functional divergence of opsin copies can also be driven by changes to the untranslated regions of the gene, which contain regulatory elements influencing gene expression and translation. This latter phenomenon has been less studied in post-duplicated opsins (but see [81]). While we did not directly investigate regulation of scallop G_q-opsin, our discovery of tissue-specific expression of G_q-opsin paralogs in the scallop, *Argopecten irradians*, not only provides circumstantial evidence that there may be differences in regulatory regions, but offers an opportunity to investigate how these gene copies diversified in function and evolutionary fates. One-to-one matches between transcript and genomic amplicons strongly support the presence of at least four G_q-opsin paralogs in the *A. irradians* genome. All four genes were identified as G_q-opsins by both sequence similarity and phylogenetic analysis, and are most likely the result of duplication events in a lineage that includes the orders Pectinoida and Limoida [49], either through whole genome duplication events [82] or duplication of small segments of the genome [83]. The specific timing of these events will require denser taxonomic sampling within the subclass Pteriomorphia, but if the phylogenetic pattern from our study holds, it would appear that *opnGq1* and *opnGq4* are derived from the first round of gene or genome duplication. Subsequently, *opnGq4* may have undergone a tandem duplication, and the paralog underwent a second round of duplication to create *opnGq2* and *opnGq3* (Fig. 5).

We present evidence that all four *Air-opnGqs* products, when reconstituted with the proper chromophore, could form photopigments. Each scallop Gq-opsin has the sequence motifs necessary for protein conformation and chromophore binding (Table 2). Tertiary structural models developed for each Air-OPNGq contain the expected protein domains and loops for a functional opsin protein. Interestingly, all four scallop protein models predict eighth and ninth cytoplasmic α-helices (Fig. 1), features unique to G_q-opsins [51]. In the Tpa-OPNGq1 crystal structure, the C-terminus of H9 interacts with the cytoplasmic extension of H6, that together with H5 form a rigid column projecting 25 Å from the membrane surface; however the rotational freedom of H9 is restricted by its interactions with H8. Thus, others have predicted that this four-domain cytoplasmic feature, in conjunction with the HKP motif in H8 [26], functions as the recognition mechanism for specific G-protein partners [51]. In summary, our bioinformatic analyses support that all four scallop G_q-opsins form photopigments
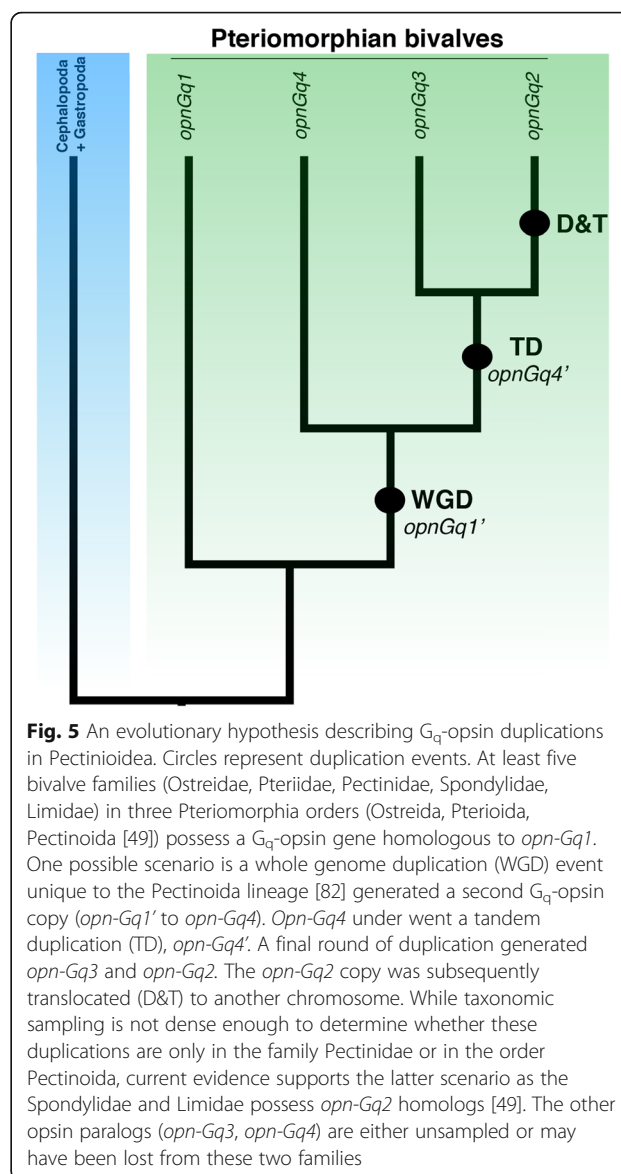

Fig. 5 An evolutionary hypothesis describing G_q-opsin duplications in Pectinioidea. Circles represent duplication events. At least five bivalve families (Ostreidae, Pteriidae, Pectinidae, Spondylidae, Limidae) in three Pteriomorphia orders (Ostreida, Pterioida, Pectinoida [49]) possess a G_q-opsin gene homologous to *opn-Gq1*. One possible scenario is a whole genome duplication (WGD) event unique to the Pectinoida lineage [82] generated a second G_q-opsin copy (*opn-Gq1'* to *opn-Gq4*). *Opn-Gq4* under went a tandem duplication (TD), *opn-Gq4'*. A final round of duplication generated *opn-Gq3* and *opn-Gq2*. The *opn-Gq2* copy was subsequently translocated (D&T) to another chromosome. While taxonomic sampling is not dense enough to determine whether these duplications are only in the family Pectinidae or in the order Pectinoida, current evidence supports the latter scenario as the Spondylidae and Limidae possess *opn-Gq2* homologs [49]. The other opsin paralogs (*opn-Gq3*, *opn-Gq4*) are either unsampled or may have been lost from these two families

that could be used to detect light. How might these gene copies have diverged after the duplication event? Molecular changes in paralogous scallop opsin genes appeared to have occurred both outside and within the protein-coding region.

We find differential gene expression across ocular and extra-ocular structures in the adult, suggesting there have been changes in the regulatory regions of scallop G_q-opsin paralogs. Specifically, while all *Air-opnGqs* are expressed in eyes, the level of expression is vastly different (ranging from a 38- to 5815-fold difference). In addition, only two of the four G_q-opsins, *Air-opnGq1* and *Air-opnGq2*, are significantly expressed outside of the eye, and presumably they are used in a nonvisual context such as the "shadow response" [84]. Taken together, these data suggest that scallop opsin paralogs are

Porath-Krause *et al. BMC Evolutionary Biology* (2016) 16:250

Page 12 of 15

used in different biological contexts. Some may preferentially be employed in eyes (*Air-opnGq3* and *Air-opnGq4*), while others (*Air-opnGq1* and *Air-opnGq2*) are used for both ocular and extra-ocular based functions.

Spatial patterning and expression level differences among the scallop G_q-opsin paralogs suggest they have undergone neofunctionalization since duplication. When we compare the scallop opsin expression data to the closest related bivalve with a sequenced genome, the Pacific oyster, *Crassostrea gigas* [73] we find a dramatic difference in the relative levels of gene expression and spatial patterning. From the oyster genome, we identified three G_q-opsins, but only one (*Cgi-opnGq1*) was phylogenetically similar to the scallop opsins (Fig. 2). This *Cgi-opnGq1* is broadly expressed at low levels across the adult non-ocular tissues (e.g., 0.10 RPKM in mantle tissue to 0.29 RPKM in gonad) [73]. In contrast, the adult scallop has high levels of expression (up to 10,001.27 FPKM) of different G_q-opsin gene copies in eyes, and low or no expression of these opsins in non-ocular tissues (Fig. 4). Could an increase in opsin expression level and/or greater number of gene copies be related to the origin of eyes? Currently available opsin sequences from bivalve species represent a very restricted taxonomic sampling. But based on the nearly ubiquitous shadow response in Bivalvia and Gastropoda, the few instances of eyes in bivalves [85], and the results from our study, we anticipate that the ancestral state for G_q-opsin spatial expression in bivalves is across multiple tissue types while the derived condition of spatial expression is narrowed (limited) to eyes and may indicate functional specificity for visual processes. If one or both of the scallop opsin duplication events were concurrent with the origin of eyes, it would support the notion of neofunctionalization of the new G_q-opsin copies.

Do the differential levels of gene expression indicate an even finer spatial partitioning of *Air-opnGqs*? We anticipate this to be the case. Depending on the scallop species, an adult animal can have between 35 to over 200 eyes along the mantle margins lining both valves (Serb, unpublished) that can vary in size [86, 87]. Visual fields from adjacent eyes overlap such that, as a conservative estimate, at least five eyes would convey similar information from a given point in the environment (estimated from a 30-eyed animal [88]). One way to reduce functional redundancy would be to distribute Air-OPNGq proteins of dissimilar absorbance spectra across non-adjacent eyes. However, due to the limitations of library construction, which required the pooling of all 60 eyes from one light- and 60 eyes from one dark-treated animal, we are unable to determine if a single eye expresses all or a just subset of *Air-opnGqs*. Furthermore, the expression pattern of *Air-opnGqs* at the level of single photoreceptors also needs to be elucidated. Since

Air-opnGqs are phylogenetically similar to the first reported scallop G_q-opsin in *Mizuhopecten yessoensis*, which is presumed to be co-expressed with G_q-protein in rhabdomeric photoreceptors of the proximal retina ("depolarizing layer") [48], we can predict that Air-OPNGqs will share a similar gross expression pattern. At a cellular level, it has been shown that more than one G_q-opsin can be expressed in a single photoreceptor cell [89–92] and this can lead to a broader spectral range for a given photoreceptor if opsins differ in λ_{max} values. Thus, to understand how spatial partitioning may have changed as gene copies diversified phenotypically in the scallop, future work will require the development of probes specific to each *Air-opnGq* gene or protein.

Spectral sensitivity may differ among the scallop G_q-opsin photopigments. We identified changes in amino acid sequence at seven sites that are predicted to influence spectral tuning of G_q-opsins [66]. The electrostatic contribution of individual residues at these sites has been modeled previously on Tpa-OPNGq1 [66, 75]. Among the scallop G_q-opsins, residues at position 92 had the most dissimilar biochemical properties (nonpolar aliphatic/hydrophobic in Air-OPNGq1 and Air-OPNGq2; aromatic in Air-OPNG3; positive polar in Air-OPNGq4). Position 306 is also of interest because there is a difference in charge and a presence/absence of a hydroxyl group. Air-OPNGq1 and Air-OPNGq4 have a polar, hydroxyl-bearing Thr306 while Air-OPNGq2 and Air-OPNGq3 contain a non-polar Ala306. Evidence from previous studies [93–95] suggests that shifts in λ_{max} values can be achieved via a change of charge (polar vs non-polar) or a gain/loss of a hydroxyl group that ultimately affects the electrostatic potential around the protonated Schiff base [66]. Based on our results, we hypothesize that the λ_{max} may differ among some or all of the Air-OPNGqs. This hypothesis contradicts results from previous studies where only a single λ_{max} value was measured for depolarizing rhabdomeric photoreceptors [96, 97]. While some of the earliest work on spectral sensitivity of scallops was based on behavior trials, and was unable to test specific visual pigments, photoreceptor cells, or account for extra-ocular photoreception (e.g., [98]), more sophisticated methods have been employed to record membrane potential changes of individual photoreceptor cells (e.g., [97, 99, 100]). Most recently, microspectrophotometry has been used on dark-adapted scallop retinas to measure λ_{max} directly [96]. For rhabdomeric photoreceptors of *A. irradians*, both intracellular recordings [97] and microspectrophotometry results [96] recover a single spectral curve with a λ_{max} value of ~500 nm. Though, with the limited number of photoreceptor cells examined (N = 4 versus N = 21 [96, 97]) and a 38- to 5815-fold higher expression level difference of *Air-opnGq2* to other *Air-opnGqs* (this study), it is unlikely that all

Porath-Krause et al. BMC Evolutionary Biology (2016) 16:250

Page 13 of 15

four G_q-opsins were sampled. An alternative approach will be needed to determine if there are any differences in λ_{max} by targeting individual Air-OPNGqs. One approach would be to directly test λ_{max} of each Air-OPNGq photopigment *in vitro*, but the well-known technical challenges of expressing G_q-opsin proteins in transient heterologous systems will need to be overcome [101, 102] or stable transfection of cell lines [103] or animals [104] will need to be employed.

Conclusions

Gene duplication and subsequent functional divergence of opsins have played an important role in expanding photoreceptive capabilities of organisms by altering what wavelengths of light are preferentially absorbed by photoreceptors (spectral tuning). However, new opsin copies may also acquire new or subdivide ancestral functions through changes to temporal, spatial or the level of gene expression. As the first molecular characterization of scallop G_q-opsins, our study highlights how opsin duplication and diversification may not only affect the evolution of the visual system, but also non-visual photoreception. Sequence variation among the scallop G_q-opsins suggests different biochemical properties of the proteins, which may translate into differences in light absorption and/or G protein affinity. Changes to spatial pattern and level of gene expression are illustrative of transitions between broad non-visual photoreception and eye-specific expression indicating neofunctionalization after opsin-duplication.

It is important to extend the taxonomic sampling of intraspecific opsin diversity in non-arthropod invertebrates in the future to understand diversification and plasticity of G_q-opsins. As such, molluscs are a rich system to study protein evolution, but have been underused due to a lack of basic information about their genic composition. Our work demonstrates the need for more studies looking at the visual evolution of molluscs to further their impact on the fields of molecular, sensory, and evolutionary biology.

Additional files

Additional file 1: Table S2. G_q-opsin sequences included in the phylogenetic analysis. Asterisks represent sequences obtained through Porter et al. [27]. For additional information regarding sequence acquisition not available on Genbank, see supplementary material in Porter et al. [27]. (DOCX 25 kb)

Additional file 2: Table S1. Primers used to amplify scallop G_q-opsins and intergenic region between *Air-opnGq3* and *Air-opnGq4*. (DOCX 13 kb)

Additional file 3: Table S3. Ramachandran plot values and C-scores for top G_q-opsin models. For each Air-OPNGq, the top five models reported by I-TASSER were analyzed for their quality using PROCHECK and the C-score. All the reported models have > 90% of their residues in allowed regions of the Ramachandran plot, indicating a good quality model. The C-scores for the best models was in the range of −3 to −2. While these values are lower than the suggested cutoff of −1.5, this is not unexpected

for GPCRs because there are relatively few solved GPCR protein structures and GPCRs often show high sequence diversity. The best model for each Air-OPNGq (highlighted) was selected as the structure having the highest C-score and highest percentage of residues in allowed regions of the Ramachandran plot. (DOCX 14 kb)

Additional file 4: Figure S1. Bayesian inference phylogram of G_q-opsins. The phylogenetic tree is based on 96 aligned amino acid sequences with scallop *Argopecten irradians* G_o-opsin as the outgroup. Support values at nodes are posterior probabilities >0.50. The grey box highlights a clade of bivalve *opnGq1* not recovered in the ML analysis. A black bar indicates the monophyletic G_q-opsin clade. (DOCX 94 kb)

Additional file 5: Figure S2. Maximum likelihood phylogram of G_q-opsins. The phylogenetic tree is based on 96 aligned amino acid sequences with scallop, *Argopecten irradians*, G_o-opsin as the outgroup. Support values (>50%) of nodes were generated by 1000 bootstrap replicates in RAxML. A black bar indicates the G_q-opsin clade. (DOCX 98 kb)

Additional file 6: Figure S3. Fifty base pair alignment of 5′- and 3′-UTRs from the four scallop G_q-opsins. Vertical lines represent the beginning and end of the coding region. (DOCX 49 kb)

Authors' contributions

JMS conceived the work; JMS, AJPK and ANP planned the research design; ANP collected samples; AJPK and ANP performed lab work; BSB, KS and DF performed protein modeling; ANP performed transcriptome assembly and phylogenetic analyses; all authors participated in manuscript preparation. All authors approve the final version of the manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Author details

[1]Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames 50011, IA, USA. [2]Department of Genetics, Development, and Cell Biology, Iowa State University, Ames 50011, IA, USA. [3]Department of Biochemistry, Biophysics, and Molecular Biology, Iowa State University, Ames 50011, IA, USA. [4]Interdepartmental Graduate Program in Bioinformatics and Computational Biology, Iowa State University, Ames 50011, IA, USA.

References

1. Briscoe AD, Macias-Muñoz A, Kozak KM, Walters JR, Yuan F, Jamie GA, Martin SH, Dasmahapatra KK, Ferguson LC, Mallet J, Jacquin-Joly E, Jiggins CD. Female behaviour drives expression and evolution of gustatory receptors in butterflies. PLoS Genet. 2013;9:e1003620.

Porath-Krause *et al. BMC Evolutionary Biology* (2016) 16:250

Page 14 of 15

2. Grus WE, Zhang J. Rapid turnover and species-specificity of vomeronasal pheromone receptor genes in mice and rats. Gene. 2004;340:303–12.

3. Yokoyama S. Molecular genetic basis of adaptive selection: examples from color vision in vertebrates. Annu Rev Genet. 1997;31:315–36.

4. Frentiu FD, Bernard GD, Sison-Mangus MP, Van Zandt Brower A, Briscoe AD. Gene duplication is an evolutionary mechanism for expanding spectral diversity in the long-wavelength photopigments of butterflies. Mol Biol Evol. 2007;24:2016–28.

5. Dong D, Jones G, Zhang S. Dynamic evolution of bitter taste receptor genes in vertebrates. BMC Evol Biol. 2009;9:12.

6. Niimura Y, Nei M. Extensive gains and losses of olfactory receptor genes in mammalian evolution. PLoS One. 2007;2:e708.

7. Ohno S. Evolution by Gene Duplication. Berlin: Springer; 1970.

8. Walsh B. Population-genetic models of the fates of duplicated genes. Genetica. 2003;118:279–94.

9. Hahn MW. Distinguishing among evolutionary models for the maintenance of gene duplicates. J Hered. 2009;100:605–17.

10. Innan H, Kondrashov F. The evolution of gene duplications: classifying and distinguishing between models. Nat Rev Genet. 2010;11:97–108.

11. Zhang J. Evolution by gene duplication: An update. Trends Ecol Evol. 2003; 18:292–8.

12. Force A, Lynch M, Postlethwait J. Preservation of duplicate genes by subfunctionalization. Am Zool. 1999;39:0.

13. Spady TC, Parry JWL, Robinson PR, Hunt DM, Bowmaker JK, Carleton KL. Evolution of the cichlid visual palette through ontogenetic subfunctionalization of the opsin gene arrays. Mol Biol Evol. 2006;23:1538–47.

14. Hittinger CT, Carroll SB. Gene duplication and the adaptive evolution of a classic genetic switch. Nature. 2007;449:677–81.

15. Piatigorsky J, Wistow G. The recruitment of crystallins: new functions precede gene duplication. Science. 1991;252:1078–9.

16. Nathans J, Thomas D, Hogness DS. Molecular genetics of human color vision: the genes encoding blue, green, and red pigments. Science. 1986; 232:193–202.

17. Yokoyama S. Molecular evolution of color vision in vertebrates. Gene. 2002; 300:69–78.

18. Porter ML, Bok MJ, Robinson PR, Cronin TW. Molecular diversity of visual pigments in Stomatopoda (Crustacea). Vis Neurosci. 2009;26:255–65.

19. Briscoe AD. Reconstructing the ancestral butterfly eye: focus on the opsins. J Exp Biol. 2008;211(Pt 11):1805–13.

20. O'Quin KE, Hofmann CM, Hofmann HA, Carleton KL. Parallel evolution of opsin gene expression in African cichlid fishes. Mol Biol Evol. 2010;27:2839–54.

21. Hofmann CM, Carleton KL. Gene duplication and differential gene expression play an important role in the diversification of visual pigments in fish. Integr Comp Biol. 2009;49:630–43.

22. Futahashi R, Kawahara-Miki R, Kinoshita M, Yoshitake K, Yajima S, Arikawa K, Fukatsu T. Extraordinary diversity of visual opsin genes in dragonflies. Proc Natl Acad Sci. 2015;112:E1247–E1256. doi:10.1073/pnas.1424670112.

23. Palczewski K. G protein-coupled receptor rhodopsin. Annu Rev Biochem. 2006;75:743–67.

24. Yarfitz S, Hurley JB. Transduction mechanisms of vertebrate and invertebrate photoreceptors. J Biol Chem. 1994;269:14329–32.

25. Marin EP. The amino terminus of the fourth cytoplasmic loop of rhodopsin modulates rhodopsin-transducin interaction. J Biol Chem. 2000;275:1930–6.

26. Plachetzki DC, Degnan BM, Oakley TH. The origins of novel protein interactions during animal opsin evolution. PLoS One. 2007;2:e1054.

27. Porter ML, Blasic JR, Bok MJ, Cameron EG, Pringle T, Cronin TW, Robinson PR. Shedding new light on opsin evolution. Proc Biol Sci. 2012;279:3–14.

28. Feuda R, Rota-Stebelli O, Oakley TH, Pisani D. The comb jelly opsins and the origins of animal phototransduction. Genome Biol Evol. 2014;6:1964–71.

29. Cronin TW, Porter ML. The evolution of invertebrate photopigments and photoreceptors. In: Hunt DM, Hankins MW, Collin SP, Marshall NJ, editors. Evol Vis Non-visual Pigment. New York: Springer International Publishing; 2014. p. 105–35.

30. Fuller RC, Carleton KL, Fadool JM, Spady TC, Travis J. Genetic and environmental variation in the visual properties of bluefin killifish, Lucania goodei. J Evol Biol. 2005;18:516–23.

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

32. Dulai KS, von Dornum M, Mollon JD, Hunt DM. The evolution of trichromatic color vision by opsin gene duplication in New World and Old World primates. Genome Res. 1999;9:629–38.

33. Briscoe AD. Functional diversification of lepidopteran opsins following gene duplication. Mol Biol Evol. 2001;18:2270–9.

34. Koyanagi M, Nagata T, Katoh K, Yamashita S, Tokunaga F. Molecular evolution of arthropod color vision deduced from multiple opsin genes of jumping spiders. J Mol Evol. 2008;66:130–7.

35. Chinen A, Hamaoka T, Yamada Y, Kawamura S. Gene duplication and spectral diversification of cone visual pigments of zebrafish. Genetics. 2003; 163:663–75.

36. Carulli JP, Chen DM, Stark WS, Hartl DL. Phylogeny and physiology of Drosophila opsins. J Mol Evol. 1994;38:250–62.

37. Porter ML, Cronin TW, McClellan DA, Crandall KA. Molecular characterization of crustacean visual pigments and the evolution of pancrustacean opsins. Mol Biol Evol. 2007;24:253–68.

38. Wang D, Oakley T, Mower J, Shimmin LC, Yim S, Honeycutt RL, Tsao H, Li WH. Molecular Evolution of Bat Color Vision Genes. Mol Biol Evol. 2004;21: 295–302.

39. Cortesi F, Musilová Z, Stieb SM, Hart NS, Siebeck UE, Malmstrøm M, Tørresen OK, Jentoft S, Cheney KL, Marshall NJ, Carleton KL, Salzburger W. Ancestral duplications and highly dynamic opsin gene evolution in percomorph fishes. Proc Natl Acad Sci. 2015;112:1493–8.

40. Yokoyama S. Gene duplications and evolution of the short wavelength-sensitive visual pigments in vertebrates. Mol Biol Evol. 1994;11:32–9.

41. Oakley TH, Huber DR. Differential expression of duplicated opsin genes in two eye types of ostracod crustaceans. J Mol Evol. 2004;58:1–11.

42. Spaethe J, Briscoe AD. Early duplication and function diversification of the opsin gene family in insects. Mol Biol Evol. 2004;21:1583–94.

43. Henze MJ, Dannenhauer K, Kohler M, Labhart T, Gesemann M. Opsin evolution and expression in Arthropod compound Eyes and Ocelli: Insights from the cricket Gryllus bimaculatus. BMC Evol Biol. 2012;12:163.

44. Pollock J, Benzer S. Transcript localization of four opsin genes in the three visual organs of Drosophila; RH2 is ocellus specific. Nature. 1988;333(6175): 779–82. doi:10.1038/333779a0.

45. Tong D, Rozas NS, Oakley TH, Mitchell J, Colley NJ, McFall-Ngai MJ. Evidence for light perception in a bioluminescent organ. Proc Natl Acad Sci U S A. 2009;106:9836–41.

46. Frank TM, Porter M, Cronin TW. Spectral sensitivity, visual pigments and screening pigments in two life history stages of the ontogenetic migrator Gnathophausia ingens. J Mar Biol Assoc U K. 2009;89:119–29.

47. Rivera AS, Pankey MS, Plachetzki DC, Villacorta C, Syme AE, Serb JM, Omilian AR, Oakley TH. Gene duplication and the orgins of morphological complexity in pancrustacean eyes, a genomic approach. BMC Evol Biol. 2010;10:123.

48. Kojima D, Terakita A, Ishikawa T, Tsukahara Y, Maeda A, Shichida Y. A novel Go-mediated phototransduction casade in scallop visual cells. J Biol Chem. 1997;272:22979–82.

49. Serb JM, Porath-Krause AJ, Pairett AN. Uncovering a gene duplication of the photoreceptive protein, opsin, in scallops (Bivalvia: Pectinidae). Integr Comp Biol. 2013;53:68–77.

50. Kimura M. The Neutral Theory of Molecular Evolution. Cambridge: Cambridge University Press: 1983.

51. Murakami M, Kouyama T. Crystal structure of squid rhodopsin. Nature. 2008; 453:363–7.

52. Shimamura T, Hiraki K, Takahashi N, Hori T, Ago H, Masuda K, Takio K, Ishiguro M, Miyano M. Crystal structure of squid rhodopsin with intracellularly extended cytoplasmic region. J Biol Chem. 2008;283:17753–6.

53. Dalal JS, Jinks RN, Cacciatore C, Greenberg RM, Battelle B-A. Limulus opsins: diurnal regulation of expression. Vis Neurosci. 2003;20:523–34.

54. Halstenberg S, Lindgre H, Samagh S, Nadal-Vicens M, Balt S, Fernald RD. Diurnal rhythm of cone opsin expression in the teleost fish Haplochromis burtoni. Vis Neurosci. 2005;22:135–41.

55. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, Macmanes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, Leduc RD, Friedman N, Regev A. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc. 2013;8:1494–512.

56. Pairett AN, Serb JM. De novo assembly and characterization of two transcriptomes reveal multiple light-mediated functions in the scallop eye (Bivalvia: Pectinidae). PLoS One. 2013;8:e69852.

57. Katoh K, Kuma K, Toh H, Miyata T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. 2005;33:511–8.

Porath-Krause et al. BMC Evolutionary Biology (2016) 16:250

Page 15 of 15

58. Abascal F, Zardoya R, Posada D. ProtTest: Selection of best-fit models of protein evolution. Bioinformatics. 2005;21:2104–5.
59. Le S, Gascuel O. An improved general amino acid replacement matrix. Mol Biol Evol. 2008;25:1307–20.
60. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30:1412–22.
61. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogeny. Bioinformatics. 2001;17:754–5.
62. Miller M, Pfeiffer W, Schwartz T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In. Proc Gatew Comput Environ Work; 2010. p. 1–8. http://www.phylo.org/sub_sections/portal/sc2010_paper.pdf.
63. Lonnig W-E, Saedler H. Chromosome rearrangements and transposable elements. Annu Rev Genet. 2002;36:389–410.
64. Altschul S, Madden T, Schaffer A, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein search programs. Nucleic Acids Res. 1997;25:3389–402.
65. Altschul S, Wooten J, Gertz E, Agarwala R, Morgulis A, Schaffer A, Yu Y-K. Protein database searches using compositionally adjusted substitution matrices. FEBS. 2005;272:5101–9.
66. Sekharan S, Wei JN, Batista VS. The active site of melanopsin: the biological clock photoreceptor. J Am Chem Soc. 2012;134:19536–9.
67. Zhang Y. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics. 2008;9:40.
68. Roy A, Kucikural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc. 2010;5:725–38.
69. Laskowski R, MacArthur M, Moss D, Thornton J. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr. 1993;26:283–91.
70. Laskowski R, Hutchinson E, Michie A, Wallace A, Jones M, Thornton J. PDBsum: a Web-based database of summaries and analyses of all PDB structures. Trends Biochem Sci. 1997;22:488–90.
71. Langmead B, Trapnell C, Pop M, Salzberg S. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10:R25.
72. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12:323.
73. Zhang G, Fang X, Guo X, Li L, Luo R, Xu F, Yang P, Zhang L, Wang X, Qi H, Xiong Z, Que H, Xie Y, Holland PWH, Paps J, Zhu Y, Wu F, Chen Y, Wang J, Peng C, Meng J, Yang L, Liu J, Wen B, Zhang N, Huang Z, Zhu Q, Feng Y, Mount A, Hedgecock D, et al. The oyster genome reveals stress adaptation and complexity of shell formation. Nature. 2012;490:49–54.
74. Rosenbaum DM, Rasmussen SGF, Kobilka BK. The structure and function of G-protein-coupled receptors. Nature. 2009;459:356–63.
75. Sekharan S, Altun A, Morokuma K. Photochemistry of visual pigment in a G (q) protein-coupled receptor (GPCR)–insights from structural and spectral tuning studies on squid rhodopsin. Chem Eur J. 2010;16:1744–9.
76. Vogel R, Mahalingam M, Lüdeke S, Huber T, Siebert F, Sakmar TP. Functional Role of the "Ionic Lock"-An Interhelical Hydrogen-Bond Network in Family A Heptahelical Receptors. J Mol Biol. 2008;380:648–55.
77. Plachetzki DC, Oakley TH. Key transitions during the evolution of animal phototransduction: novelty, "tree-thinking", co-option, and co-duplication. Integr Comp Biol. 2007;47:759–69.
78. Briscoe AD, Chittka L. The evolution of color vision in insects. Annu Rev Entomol. 2001;46:471–510.
79. Hart N, Hunt D. Avian Visual Pigments: Characteristics, Spectral Tuning, and Evolution. Am Nat. 2007;169:S7–S26.
80. Bowmaker JK. Evolution of vertebrate visual pigments. Vis Res. 2008;48:2022–41.
81. O'Quin KE, Smith AR, Sharma A, Carleton KL. New evidence for the role of heterochrony in the repeated evolution of cichlid opsin expression. Evol Dev. 2011;13:193–203.
82. Wang Y, Guo X. Chromosomal rearrangement in pectinidae revealed by rRNA loci and implications for bivalve evolution. Biol Bull. 2004;207:247–56.
83. Zhang L, Bao Z, Wang S, Huang X, Hu J. Chromosome rearrangements in Pectinidae (Bivalvia: Pteriomorphia) implied based on chromosomal localization of histone H3 gene in four scallops. Genetica. 2007;130:193–8.
84. Wilkens LA. Primary inhibition by light : A unique property of bivalve photoreceptors. Am Malacol Bull. 2008;26:101–9.
85. Morton B. The evolution of eyes in the Bivalvia. Oceanogr Mar Biol Annu Rev. 2001;39:165–205.
86. Speiser DI, Johnsen S. Comparative morphology of the concave mirror eyes of scallops (Pectinoidea). Am Malacol Bull. 2008;26:27–33.
87. Gutsell JS. Natural history of the bay scallop. Bull Bur Fish. 1930;46:569–632.
88. Wilkens LA. Neurobiology and behavior of the scallop. In: Shumway SE, Parsons GJ, editors. Scallops Biol Ecol Aquac. Amsterdam: Elsevier; 2006:317–356.
89. Mazzoni EO, Celik A, Wernet MF, Vasiliauskas D, Johnston RJ, Cook TA, Pichaud F, Desplan C. Iroquois complex genes induce co-expression of rhodopsins in Drosophila. PLoS Biol. 2008;6:825–35.
90. Arikawa K, Mizuno S, Kinoshita M, Stavenga DG. Coexpression of two visual pigments in a photoreceptor causes an abnormally broad spectral sensitivity in the eye of the butterfly Papilio xuthus. J Neurosci. 2003;23:4527–32.
91. Katti C, Kempler K, Porter ML, Legg A, Gonzalez R, Garcia-Rivera E, Dugger D, Battelle B. Opsin co-expression in Limulus photoreceptors: differential regulation by light and a circadian clock. J Exp Biol. 2010;213(Pt 15):2589–601.
92. Hu X, Leming MT, Whaley MA, O'Tousa JE. Rhodopsin coexpression in UV photoreceptors of Aedes aegypti and Anopheles gambiae mosquitoes. J Exp Biol. 2014;217:1003–8.
93. Asenjo AB, Rim J, Oprian DD. Molecular determinants of human red/green color discrimination. Neuron. 1994;12:1131–8.
94. Yokoyama S, Tada T, Zhang H, Britt L. Elucidation of phenotypic adaptations: Molecular analyses of dim-light vision proteins in vertebrates. Proc Natl Acad Sci U S A. 2008;105:13480–5.
95. Hauser FE, van Hazel I, Chang BSW. Spectral tuning in vertebrate short wavelength-sensitive 1 (SWS1) visual pigments: Can wavelength sensitivity be inferred from sequence data? J Exp Zool B Mol Dev Evol. 2014;3228:529–39.
96. Speiser DI, Loew ER, Johnsen S. Spectral sensitivity of the concave mirror eyes of scallops: potential influences of habitat, self-screening and longitudinal chromatic aberration. J Exp Biol. 2011;214(Pt 3):422–31.
97. McReynolds JS, Gorman ALF. Membrane conductances and spectral sensitivities of Pecten photoreceptors. J Gen Physiol. 1970;56:392–406.
98. Cronly-Dillon JR, Cronly-Dillion JR. Spectral sensitivity of the scallop Pecten maximus. Science. 1966;151:345–6.
99. Cornwall MC, Gorman ALF. The cation selectivity and voltage dependence of the light-activated potassium conductance in scallop distal photoreceptor. J Physiol. 1983;340:287–305.
100. Gomez MP, Nasi E. The light-sensitive conductance of hyperpolarizing invertebrate photoreceptors: a patch-clamp study. J Gen Physiol. 1994;103:939–56.
101. Terakita A, Tsukamoto H, Koyanagi M, Sugahara M, Yamashita T, Shichida Y. Expression and comparative characterization of Gq-coupled invertebrate visual pigments and melanopsin. J Neurochem. 2008;105:883–90.
102. Matsuyama T, Yamashita T, Imamoto Y, Shichida Y. Photochemical properties of mammalian melanopsin. Biochemistry. 2012;51:5454–62.
103. Frentiu FD, Yuan F, Savage WK, Bernard GD, Mullen SP, Briscoe AD. Opsin Clines in Butterflies Suggest Novel Roles for Insect Photopigments. Mol Biol Evol. 2014;32:368–79.
104. Knox BE, Salcedo E, Mathiesz K, Schaefer J, Chou W-H, Chadwell LV, Smith WC, Britt SG, Barlow RB. Heterologous expression of limulus rhodopsin. J Biol Chem. 2003;278:40493–502.
105. Waterhouse A, Procter J, Martin D, Clamp M, Barton GJ. Jalview Version 2–a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009;25:1189–91.