

## Article

# The Gluopsins: Opsins without the Retinal Binding Lysine

Martin Gühmann <sup>1,\*</sup> , Megan L. Porter <sup>2</sup> and Michael J. Bok <sup>3</sup><sup>1</sup> School of Biological Sciences, University of Bristol, Bristol BS8 1TQ, UK<sup>2</sup> Department of Biology, University of Hawai'i at Mānoa, Honolulu, HI 96822, USA<sup>3</sup> Lund Vision Group, Department of Biology, University of Lund, 223 62 Lund, Sweden

\* Correspondence: martin.guehmann@bristol.ac.uk

**Abstract:** Opsins allow us to see. They are G-protein-coupled receptors and bind as ligand retinal, which is bound covalently to a lysine in the seventh transmembrane domain. This makes opsins light-sensitive. The lysine is so conserved that it is used to define a sequence as an opsin and thus phylogenetic opsin reconstructions discard any sequence without it. However, recently, opsins were found that function not only as photoreceptors but also as chemoreceptors. For chemoreception, the lysine is not needed. Therefore, we wondered: Do opsins exist that have lost this lysine during evolution? To find such opsins, we built an automatic pipeline for reconstructing a large-scale opsin phylogeny. The pipeline compiles and aligns sequences from public sources, reconstructs the phylogeny, prunes rogue sequences, and visualizes the resulting tree. Our final opsin phylogeny is the largest to date with 4956 opsins. Among them is a clade of 33 opsins that have the lysine replaced by glutamic acid. Thus, we call them gluopsins. The gluopsins are mainly dragonfly and butterfly opsins, closely related to the RGR-opsins and the retinochromes. Like those, they have a derived NPxxY motif. However, what their particular function is, remains to be seen.

**Keywords:** opsin; evolution; photoisomerase; phylogeny; G-protein-coupled receptor; chemoreceptor; data mining; retinal binding site



**Citation:** Gühmann, M.; Porter, M.L.; Bok, M.J. The Gluopsins: Opsins without the Retinal Binding Lysine. *Cells* **2022**, *11*, 2441. <https://doi.org/10.3390/cells11152441>

Academic Editors: Annamaria Locascio and Maria I. Arnone

Received: 9 June 2022

Accepted: 28 July 2022

Published: 6 August 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Opsins are the molecules that allow us to see. They are G-protein-coupled receptors (GPCRs) [1,2], which are chemoreceptors and have seven transmembrane domains forming a binding pocket for a ligand [3,4]. The ligand for opsins is 11-cis-retinal [5–9], which is covalently bound to a lysine residue [10] in the seventh transmembrane domain [11–13]. However, 11-cis-retinal only blocks the binding pocket and does not activate the opsin. The opsin is only activated when 11-cis-retinal absorbs a photon of light and isomerizes to all-trans-retinal [14,15], the receptor activating form [16,17]. Thus, a chemoreceptor is converted to a light or photo(n)receptor.

Opsins and other GPCRs have a number of conserved sequence motifs and residues that are functionally important. All GPCRs have a highly conserved NPxxY<sup>7.53</sup> sequence motif in their seventh transmembrane domain. Here, we use the common GPCR numbering scheme for residues from Ballesteros and Weinstein [18]. The number before the period is the number of the transmembrane domain. The number after the period is set arbitrarily to 50 for the most conserved residue in that transmembrane domain among GPCRs known in 1995. For the seventh transmembrane domain, the proline in the NPxxY<sup>7.53</sup> motif is P<sup>7.50</sup>, the asparagine before is then N<sup>7.49</sup>, and the tyrosine three residues after is then Y<sup>7.53</sup>. The NPxxY<sup>7.53</sup> motif is important for G-protein activation. For instance, if proline<sup>7.50</sup> is replaced by alanine<sup>7.50</sup> then cattle rhodopsin has 141% of wild type activity [19], but this depends on the receptor. Other GPCRs have less activity [20] or have no activity at all [21]. However, a receptor could also have activity with alanine<sup>7.50</sup> and lose it with proline<sup>7.50</sup> [22]. The same is true for asparagine<sup>7.49</sup> [19,23] and tyrosine<sup>7.53</sup> [19,24].

The lysine in the seventh transmembrane domain of cattle rhodopsin (*Bos taurus*) is the 296th amino acid [12,25] and thus is named lysine 296<sup>7.43</sup> (here we also include the Ballesteros and Weinstein numbering). Cattle rhodopsin was the first opsin whose amino acid sequence was determined [25]. Since lysine 296<sup>7.43</sup> binds retinal, opsins without it are not photosensitive [26]. Other opsins may have more or fewer amino acids than cattle rhodopsin and the corresponding lysine may be at another position. However, the corresponding lysine can be easily identified by aligning those opsins to cattle rhodopsin. For simplicity, we call such homologous lysines also lysine 296<sup>7.43</sup> in accordance with the opsin literature.

Lysine 296<sup>7.43</sup> is well conserved among opsins, so well conserved that sequences without it are not even considered opsins and thus excluded from large scale phylogenetic reconstructions [27,28]. Feuda et al. [29] even reconstructed a group of *Trichoplax* GPCRs without lysine 296<sup>7.43</sup> that they found closely related to the opsins and thus called placopsins.

Beside light detection, some opsins are also involved in thermosensation [30], mechanoreception such as hearing [31] and other functions [32,33]. Recently, opsins have even been identified that can act as aristolochic acid chemoreceptors, even if light sensitivity is abolished by replacing lysine 296<sup>7.43</sup> by another amino acid [26]. These studies suggest a functional flexibility in opsins to facilitate tasks beyond photoreception. Therefore, we wondered: Do opsins exist that have lost lysine 296<sup>7.43</sup> during evolution? Such opsins would be nested within other groups of opsins that still have lysine 296<sup>7.43</sup>. Here, we built a new custom pipeline for reconstructing phylogenies and reconstructed the largest opsin phylogeny to date. In this phylogeny, we found a clade of opsins that have lost lysine 296<sup>7.43</sup>, which we call gluopsins.

## 2. Material and Methods

### 2.1. Protein Sequence Collection

To collect GPCR-protein sequences, we searched with BLAST [34] in the uniprot databases SPROT and TREMBL locally and in the NCBI databases nr, refseq\_protein, swissprot, and tsr\_nr remotely for sequences similar to opsins. As opsin bait sequences, we used 84 sequences from the data set of Ramirez et al. [35]. The bait sequences were spread over their phylogeny to cover a broad range of opsins. The chosen sequences are in Files S1 and S2; some bait sequences turned out to be rogue, which are in File S2 (see Section 2.2). We used a liberal e-value cutoff of  $1 \times 10^{-5}$  and collected the first 100,000 hits. Among the bait sequences, we used also a placopsin sequence. We restricted the number of queries to the servers of NCBI to one sequence per database at the same time so that we would not overuse this common and public resource. Additionally, we added sequences from Lowe et al. [36] and D'Aniello et al. [37], which were also used by Ramirez et al., but were not in one of the databases we searched.

We also added transcriptome sequences from the marine worm *Platynereis dumerilii* [38], which are available at <https://jekelylab.ex.ac.uk/blastdbs/index.html> (accessed on 1 June 2022) as assembly version 2. These sequences were tentatively annotated as opsins via BLAST. We also added sequences from fan worm transcriptomes that we identified as opsin related with our own version of Phylogenetically Informed Annotation (PIA, [39]). (The new sequences and all other sequences of our final tree are included in File S3). Our version of PIA is derived from that of Pérez-Moreno et al. [40] and is available at <https://github.com/MartinGuehmann/PIA2> (accessed on 1 June 2022).

### 2.2. Sequence Pruning

Since we collected sequences from different databases, we collected duplicates, which we removed with SeqKit [41]. However, the dataset still contained many very similar sequences. Therefore, we grouped the sequences that were 90% or more similar to each other into clusters and chose from each a representative with CD-Hit [42,43]. This data set of representatives contained 89,996 sequences. However, since our sequence search

was very permissive, many of those were non-opsin sequences. Using so many sequences to reconstruct a phylogenetic tree consumed more time and memory than our cluster computer could provide.

To purge non-opsin sequences, we split the dataset randomly into 128 subsets of about 900 sequences with SeqKit and added our opsin bait sequences to each set and a non-opsin sequence of an olfactory receptor. Then, we aligned each subset with PASTA [44] and reconstructed for each subset a phylogenetic tree with IQ-Tree 2 [45]. Each tree was rooted with *nw\_reroot* at the olfactory receptor so that we could extract the subtree spanned by our bait sequences with *nw\_clade* from the newick utilities [46]. For this, we had to remove seven sequences from the set of bait sequences including the placopsin and Go-opsin2 of *Terebratalia transversa*, since those were placed in some trees outside the opsins and thus would have given us non-opsin sequences.

From the trees, we extracted 8483 potential opsin sequences. We added back 1000 sequences that were randomly chosen by SeqKit to include a diverse non-opsin outgroup. Since we were interested in the placopsins too, which were removed because of the placement problem of the placopsin bait sequence, we also added back all sequences from *Trichoplax*. This way, our final dataset contained 9694 sequences before rogue removal.

### 2.3. Rogue and Long Branch Removal

Rogue sequences introduce instability to a phylogeny by jumping around from one place to another and may change the relationships and branch supports within the phylogeny unpredictably [47,48]. To remove such rogue sequences as well as long branches, we split the dataset randomly into subsets of about 900 sequences, aligned each subset with PASTA and built a phylogenetic tree and bootstrap trees with IQ-Tree 2. The bootstrap trees were passed to RogueNaRok [47,49] to identify rogue sequences and the main consensus tree was passed to TreeShrink [50] to identify long branches. The rogue sequences and the sequences of the long branches were removed from the main dataset with SeqKit. However, since what a rogue sequence is, depends also on the other sequences in the dataset [48], we repeated this splitting and removal procedure for 20 iterations. Once the 20 iterations were complete, we used iteration 10 as a base and built from there trees containing all sequences up to iteration 20. In each iteration, we removed the sequences identified as rogue sequences from the full tree and from the previous split trees at that iteration to build the next full tree. Our final data set at iteration 20 contained 6040 sequences. The sequences are available in File S3.

### 2.4. Phylogenetic Reconstruction

For phylogenetic reconstructions, we aligned the sequences with PASTA [44] with the default settings. From the alignment, we removed columns that contained more than 90% of sequences with a gap with TrimAl [51]. The trimmed alignment was then passed to IQ-Tree 2. IQ-Tree 2 selected the best substitution model (JTT + F + G4 for our final tree) for inferring a maximum likelihood tree and generated three kinds of support values: Shimodaira–Hasegawa-like approximate likelihood ratio test (SH-aLRT) values [52], aBayes [53], and ultra-fast bootstrap (UFBoot) values [54,55]. For both the SH-aLRT and the UFBoot values, we used 1000 replicates.

We used more than one kind of support value, because one may support a wrong branch while the other may not so that they may “compensate for each other’s failures” [52]. We rejected a branch if its SH-aLRT value was below 0.1 irrespective of whatever the other support values were [56] and accepted it if all three support values were above or equal to the following thresholds: 80 for SH-aLRT [52], 0.95 for aBayes [53], and 95 for UFBoot [54].

For alignment trimming, we used TrimAl with a moderate trimming value. In principle, removing data such as N- and C-terminal sequences of opsins, as done previously [28], can also remove phylogenetically informative sites and thus reduce the resolution of the phylogenetic reconstruction [57]. In fact, TrimAl did not improve the averages of support values as we could see, it may even have slightly worsened them. This agrees with rigorous

benchmark tests with real and simulated data [58–60]. However, since the aligner introduced many gaps with columns almost empty, TrimAl reduced the alignment file size: For instance, the alignment with all sequences originally had a size of 1 GB and after trimming had a size of 11 MB. This significantly reduced the time IQ-Tree needed to reconstruct the phylogeny. The final gap-reduced alignment is available in File S4, a version of the alignment sorted by the order of sequences in the final tree is available in File S5, and the final tree is available in newick format in File S6.

### 2.5. Tree Visualization

The trees were visualized with ETE 3 [61], which is a python package for programmatically visualizing phylogenetic trees. For that, we used a CSV file to define each clade with a name, a color to be used within a tree, and a sequence ID. The sequence ID represents a clade such as the peropsins and points to a leaf in the tree.

To annotate the tree, we rooted the master tree at the outgroup leaf, which is defined as the last entry in the CSV file. Then, from each clade leaf, we traversed to the root and counted for each visited node how many clade leaves were descendants. The descendant count is the highest at the root and stays that high when traversed to any clade defining leaf back as long as all these leaves are decedents. This way, we could determine the last common ancestor of all ingroup clades and used that to reroot the tree. Similarly, we could define the root node of each clade, which is the last node from the defining leaf node with the descendant count of one. From that node we could color the clade or collapse the clade. We saved each clade as an independent tree.

This way, we did not only visualize the complete tree, but also the trees from the partial datasets we used for the initial rogue removal so that we could inspect them better. However, in these reduced trees, the sequence we used to define the clade leaf may not be included. Therefore, we used the clade subtrees from the main tree to find an alternative clade leaf, which we then could use to collapse or color the clades accordingly.

With ETE 3, we produced for each tree a pdf file with all branches, and a tree collapsed at the clade root nodes. We also printed the support values onto in the order SH-aLRT/aBayes/UFBoot the branches. In the collapsed tree, each value is represented by a pie chart. The pie chart is black for the following values SH-aLRT  $\geq$  80%, aBayes  $\geq$  0.95, and UFBoot  $\geq$  95%; otherwise, it is gray.

### 2.6. Finding Position 296<sup>7.43</sup> in Other Opsins and GPCRs

To find opsins that have lost lysine 296<sup>7.43</sup>, we used the gap-reduced alignment to determine the amino acid position in the other opsins and GPCRs that corresponds to lysine 296<sup>7.43</sup> in cattle rhodopsin. Since this alignment was gap reduced, we aligned the reference sequence of cattle rhodopsin to the gap-reduced cattle rhodopsin in the alignment. This allowed us to map lysine 296<sup>7.43</sup> to its actual position in the gap reduced alignment, and we could get the corresponding amino acid at the corresponding position for all sequences. This information was then applied to the trees generated by ETE 3. For the collapsed trees, this is simply the percentage of each amino acid of the sequences at that position in the collapsed subtree.

We did not consider single opsins without lysine 296<sup>7.43</sup> isolated within a clade as real, because we could not exclude that those were missequenced, misassembled, or pseudogenes.

### 2.7. Annotate the Sequences with Higher Taxa

We annotated the sequences in the tree with information about the corresponding higher taxa. For that, we used the NCBI taxonomy database, which our pipeline downloaded automatically and extracted all the information about the known genera. The pipeline checked the sequence IDs in the tree for whether they contained a genus string from the database. If it was in the database, the corresponding higher taxon was assigned to that sequence and if it was in our list of interesting taxa. We started with the taxon list of

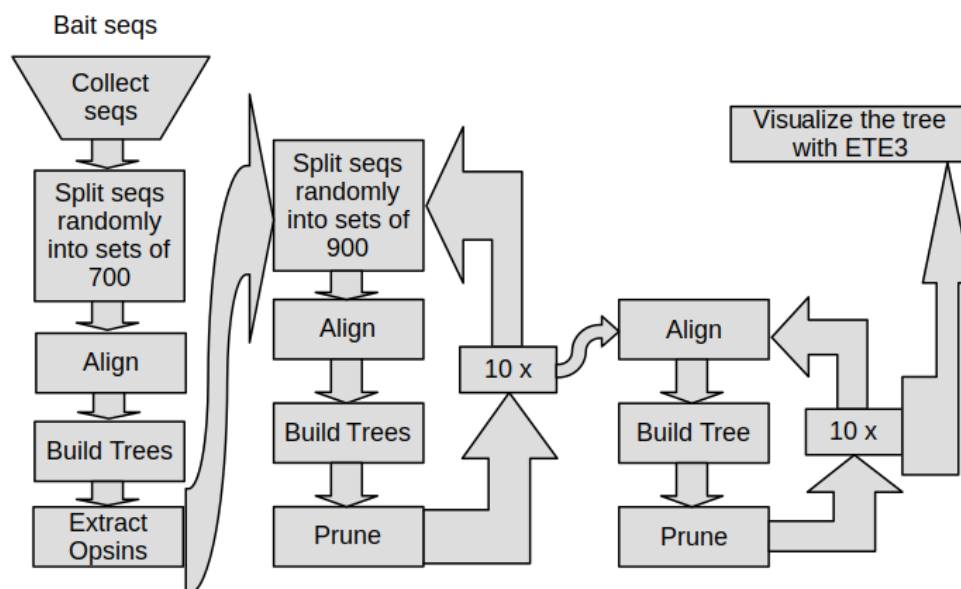
Porter et al. [28], and checked unidentified sequences, whether they were from a genus that was not within a higher taxon on our list of interesting taxa. In that case, we added that higher taxon. Some of these unidentified sequences had a sequence ID that did not contain a genus name, and thus we could not annotate it with a higher taxon. However, since this was only the case for 26 of 6040 sequences, we do not consider this as a problem.

### 2.8. Sequence Logo

For each defined clade, we generated a sequence logo around lysine 296<sup>7.43</sup> with Logomaker [62] and the library matplotlib [63]. The logo starts at residue 287<sup>7.34</sup>, ends at residue 324<sup>7.71</sup>, and spans 37 residues. This region does not contain any gaps in most gluopsins and contains conserved residues. Beside residue 296<sup>7.43</sup>, we highlighted the residues of the NPxxY<sup>7.53</sup> motif, and the residues 292<sup>7.39</sup> and 314<sup>7.64</sup>.

### 2.9. The Phylogeny Pipeline

Finally, we combined all the steps together with standard Linux tools including Bash. This way we built a pipeline with the aim of putting raw data in and getting publication ready figures out, while everything in between was processed, automatically (Figure 1).



**Figure 1.** Flowchart of the pipeline used for reconstructing our opsin phylogeny. The pipeline starts with the bait sequences and uses them to collect similar sequences with BLAST from public sequences databases. The sequences are filtered for opsins by building phylogenetic trees and extracting the opsin clade from them. It adds back outgroup sequences (not shown in the chart). With this set, it then reconstructs small trees ten times for fast rogue pruning, followed by trees with the full sequence set for rogue pruning ten times, since what a rogue is, is determined by all other sequences in the set. After rogue pruning, the final tree is visualized with ETE3. The last step is manual and can be applied to all trees.

Since this is a resource-intensive pipeline, we built it for the cluster computer BluePebble at the University of Bristol. BluePebble had computer nodes with up to 187 GB memory and 24 processor cores, and a maximum wall time of 72 h. Its scheduler system was first PBS-Pro, which was later replaced by Slurm. Therefore, the pipeline can be used with either PBS-Pro or Slurm, and if necessary other schedulers might be easily added. The code was modularized such that jobs could execute tasks in parallel or in sequence when the task depended on the result of a previous task.

The pipeline has two input types, the bait sequences, and the clade definitions. The bait sequences are used to find similar sequences in the public databases with BLAST. They are also used to filter for opsin sequences. However, for that, they need to be indeed



opsins. However, some bait sequences turned out not to be opsins or to be rogue sequences. Therefore, the pipeline can be paused at the filter step and these sequences can be marked as additional bait sequences to keep them for a total rerun, but not for filtering. We checked in the filter step by checking the trees generated by that step whether the bait sequences defined a clade that only contained opsins. For that, we used Dendroscope 3 [64].

The clade definitions were used for tree visualization and need also be updated after the first run. One reason was that CD-Hit clusters the sequences into groups of 90% similarity, and chooses one sequence as representative, but which one was beyond our control. The other reason was that the sequence that was supposed to define the clade was removed. This was the final and least computational step, which we ran on a Linux laptop after copying the data from the cluster computer, because ETE3 [61] required for tree visualization QT5, which required a running x-server, which is not available on a cluster computer in automatic non-interactive mode.

The code for the phylogeny pipeline is available at <https://github.com/MartinGuehmann/PhylogenyPipeline> (accessed on 1 June 2022) The data and the files to run it are available at <https://github.com/MartinGuehmann/Opsins> (accessed on 1 June 2022).

### 2.10. Opsin Nomenclature

During phylogenetic reconstruction, we encountered both new and previously unnamed clades of opsins. We aimed to give them names that strike a balance between descriptive content, uniqueness, and brevity. Our names typically consist of an opsin suffix and a presyllable derived from either a name from a species or higher taxon within the clade, or from a shared property of the opsins in the clade.

## 3. Results

### 3.1. The Phylogeny Pipeline

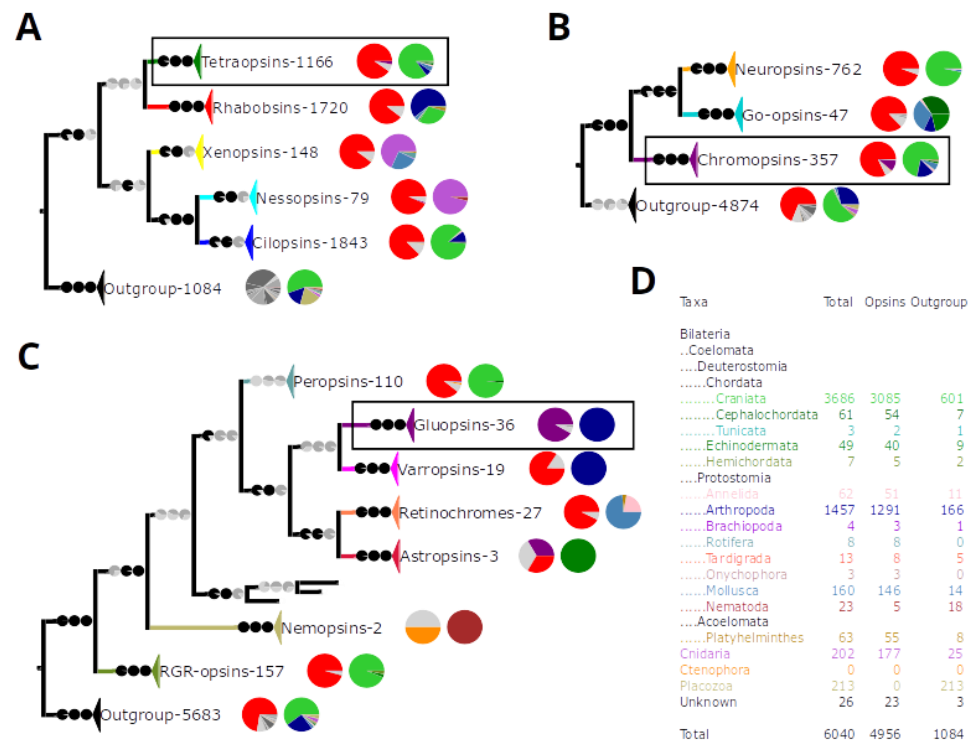
We reconstructed the relationship of the opsins to find a clade of opsins that has lost lysine 296<sup>7.43</sup>. For that, we used our custom phylogeny pipeline, which automatically executes almost all steps from sequence collection to phylogenetic reconstruction. However, when we filtered the collected sequences for opsins, a few of the original bait sequences turned out to be rogue sequences, grouping sometimes with non-opsin sequences. Furthermore, we could not recover the placopsins as a sister group of the opsins as we expected from the results of Feuda et al. [29]. Therefore, we removed the placopsins and the rogue sequences in the opsin filtering step.

The whole pipeline collects the sequences from the databases; makes the sequences unique within 90% of sequence identity; removes non-opsin sequences, except (in our case) the *Trichoplax* sequences and adds 1000 randomly chosen outgroup sequences; removes rogue sequences; and builds the final opsin tree, which then only needs to be visualized on a computer running an X-server (Figure 1).

Thus, the entire phylogenetic reconstruction can be reproduced easily. The pipeline can also be used for other GPCRs or proteins, by inputting different bait sequences and clade definitions.

### 3.2. Five Basal Types of Opsins

To find opsins that have lost lysine 296<sup>7.43</sup>, we collected as many sequences as possible. We ended up with 89,996 unique GPCR sequences. Using so many sequences to reconstruct a phylogeny was computationally not feasible for us. Therefore, we automatically removed all non-opsin sequences and added back 1000 randomly chosen sequences plus 211 *Trichoplax* GPCRs as an outgroup. This gave us a reduced dataset of 9694 sequences. Additionally, we increased the quality of the phylogenetic reconstruction by repeatedly removing rogue sequences with RogueNaRok [47,49] and long branches with TreeShrink [50]. After rogue and long branch removal, our final phylogeny contained 6040 GPCRs including 4956 opsins (Figure 2D).



**Figure 2.** Phylogenetic reconstruction of the opsins. (A) The groups of the opsins and the outgroup are collapsed. The outgroup contains all other GPCRs. The frame highlights the tetraopsins, which are expanded in (B). (B) The groups of the tetraopsins are shown. The outgroup contains all other opsins and the non-opsin GPCRs. The frame highlights the chromopsins, which are expanded in (C). (C) The groups of the chromopsins are shown. The frame highlights the gluopsins. The outgroup contains all other opsins and the non-opsin GPCRs. (A–C) Next to each clade is the number of sequences within that clade shown. The first pie chart shows the percentage of a certain amino acid at lysine 296<sup>7.43</sup>. Red stands for lysine (K), purple stands for glutamic acid (E), the other amino acids are alternatively colored dark or mid-gray so that two adjacent amino acids have different shades of gray in the pie chart. Light gray stands for a gap at this position. This comes from sequences that are incomplete and do not contain the seventh transmembrane domain. The second pie chart gives the taxon composition for each clade, the colors correspond to the list of taxa in (D) The taxon composition is also given in a numerical format in Table S1. The support values are given as pie charts. They are from right to left SH-aLRT/aBayes/UFBboot. Splits are considered supported when SH-aLRT ≥ 80%, aBayes ≥ 0.95, and UFBboot ≥ 95%. If a support value is above its threshold the pie chart is black otherwise gray. (D) The list of higher taxa represented by the sequences. A few sequences are unidentified, since their sequence identifiers do not contain a genus name.

With this, we not only found the gluopsins, the opsins that have lost lysine 296<sup>7.43</sup> during evolution, but also reconstructed the largest yet opsin phylogeny with 4956 opsin sequences. This phylogeny is more than five times bigger than previous opsin phylogenies [28,35]. This 5k opsin phylogeny is itself interesting, but out of scope here. Therefore, we will describe it elsewhere so that we can focus on the gluopsins, here.

In addition to the 4956 opsins, we recovered an outgroup of 1084 sequences, which apart from two sequences do not have lysine 296<sup>7.43</sup>, but a variety of other residues (Figures 2A and S1). In contrast, most opsins sequences that are full length have lysine 296<sup>7.43</sup> (Figure 2A–C). This means if our sample of non-opsin GPCRs is representative, then a lysine 296<sup>7.43</sup> still indicates that a GPCR is most likely an opsin.

We did not recover any bathyopsins or chaopsins as reconstructed by Ramirez et al. [35]. Their bathyopsins and chaopsins have only four and seven sequences, respectively, which are not in our final data set and thus were removed by RogueNaRok. The same happened

to the ctenopsins, which is expected as they are known to behave like rogue sequences; in phylogenies, they jump around depending on the outgroup used [65].

Our phylogeny recovered five primary opsin clades: the ciliary opsins (cilopsins), the rhabdomeric opsins (rhabopsins), the tetraopsins, the xenopsins, which were first reconstructed by Ramirez et al. [35], and the nessopsins (Figure 2A). The cilopsins and the rhabopsins do not contain sequences from cnidarians (Figure 2A) but contain the visual opsins of vertebrates and arthropods, respectively. The xenopsins are absent from deuterostomes and the only clade with cnidarian opsins beside the nessopsins (Figure 2A).

The nessopsins contain mainly cnidarian opsins (Figure 2A). They are identical to an unnamed group that fell sister to the cilopsins, the rhabopsins, and the bathyopsins in the phylogenetic tree of Ramirez et al. [35]. They are also identical with the anthozoan opsins II of Quiroga Artigas et al. [66] and the cnidarian opsins of Rawlinson et al. [67]. We identified these groups as nessopsins, since they share the sequence XP\_015773304 of *Acropora digitifera* with our nessopsins. Since the nessopsins have had so far no established name, we call them nessopsins after the German word “Nesseltiere”, which means cnidarians.

The tetraopsins are the main group that contain the gluopsins. They are also known as RGR/Go-opsins or Group 4 opsins [28,29,35]. The tetraopsins, like the cilopsins and the rhabopsins, do not have cnidarian sequences (Figure 2A) and are subdivided into three clades: the neuropsins, the Go-opsins, and the chromopsins (Figure 2B).

### 3.3. The Chromopsins

Classically, the chromopsins contain the peropsins, RGR-opsins, and retinochromes (Figure 2C). We derived the name chromopsin from the retinochromes, because they were discovered first [68,69], before the RGR-opsins [70,71] and the peropsins [72]. Additionally, we reconstructed four more clades: the varropsins, astropsins, nemopsins, and gluopsins. How these clades relate to each other is unclear due to low support values (Figure 2C). Two chromopsin orthologs exist in deuterostomes, and one in protostomes.

The peropsins contain sequences from craniates and cephalochordates (Figure 2C). They exclude protostome sequences that were previously described as peropsins [73]. Instead, we reconstructed these sequences, either as retinochromes, gluopsins, or varropsins.

The varropsins only exist in *Limulus* and the arachnids, which are both arthropods (Figures 2C, S1 and S2). We named them after *Varroa destructor*, a mite with a varropsin. Although, varropsins have been phylogenetically described as peropsins, their relationships to vertebrate peropsins are unclear due to low support values [73–75]. Henze and Oakley [73] actually distinguish between two peropsin clades: the insect and non-insect arthropod peropsins, which are our gluopsins and varropsins, respectively. Their peropsin clades are even interspersed by a sequence from the marine ragworm *Platynereis dumerilii*, which was originally described as a peropsin [76] and has been reclassified as a retinochrome [35]. We do not have enough support to place the varropsins confidently either. Therefore, they could simply be arthropod peropsins or be indeed a different clade.

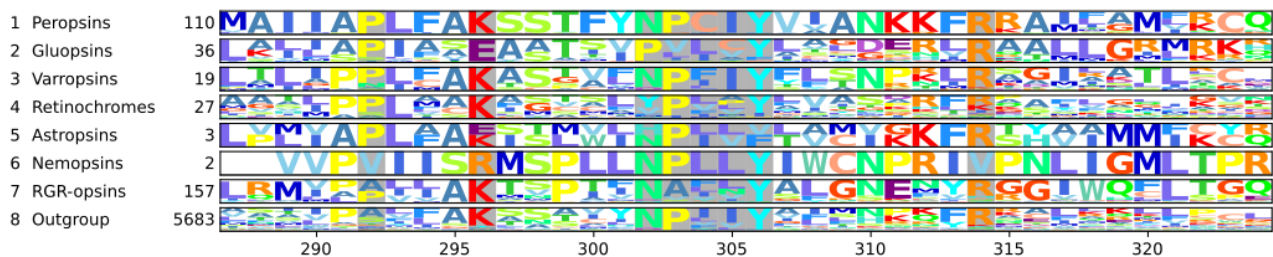
The RGR-opsins (short for Retinal G protein coupled receptors) have sequences from craniates, hemichordates, and echinoderms (Figures 2C, S1 and S2), while the retinochromes have sequences from mollusks, platyhelminths, and annelids (Figure 2C). The annelid sequences come from our transcriptomes except one sequence from *Platynereis dumerilii*, which was originally describes as a peropsin [76], and later reclassified as a retinochrome [35] in agreement with our phylogeny.

The astropsins are echinoderm specific chromopsins (Figure 2C). We named them after *Asterias rubens*, a sea star with an astropsin. Only three astropsins remain in our rogue-pruned final tree, only two cover the seventh transmembrane domain, and one has lysine 296<sup>7.43</sup> replaced by a glutamic acid (Figure 2C). With only two sequences, it is hard to draw conclusions. However, in total three astropsins with glutamic acid 296<sup>7.43</sup> have been reported [37]. We checked our alignments from previous iterations of pruning and found three sequences with glutamic acid 296<sup>7.43</sup> (data not shown). These sequences are all from sea urchins, while sea stars and sea cucumbers have astropsins with lysine 296<sup>7.43</sup>.



### 3.4. The Nemopsins Have Arginine at Position 296

The nemopsins are nematode chromopsins. Only two nemopsins remain in our final tree, and only one covers the seventh transmembrane domain where lysine 296<sup>7.43</sup> is replaced by an arginine (Figures 2C, 3, S1 and S2). Among the removed sequences are sequences from *Caenorhabditis elegans* (NP\_001364737.1) and *Pristionchus pacificus* (PDM61246.1); both species are model systems and both sequences have the arginine. The sequence from *C. elegans* has been previously described as an opsin like GPCR with arginine 296<sup>7.43</sup> and a conserved NPxxY<sup>7.53</sup> motif. It was named sro-1, short for serpentine receptor class o 1 and is expressed in chemosensory cells [77]. The nemopsins have not been included in a previous opsin phylogeny, and we are not aware of anything else known about them.



**Figure 3.** Consensus sequences of the different chromopsins. The first column contains a number for each chromopsin group for easy reference. The second column shows the names for each group. The groups are in the same order as in the tree in Figure 2C. The third contains the number of sequences in each group. And the fourth column contains the sequence logo, the height of the letters indicates the percentage of that amino acid given at that position. The x-axis gives the position of the amino acid corresponding to cattle rhodopsin. Positions 292<sup>7.39</sup> and 314<sup>7.64</sup> are highlighted in gray. Lysine (K) 296<sup>7.43</sup> is highlighted with a gray background, which is replaced in the nemopsins by arginine (R) and in the gluopsins by glutamic acid (E). The NPxxY<sup>7.53</sup> motif is highlighted with a gray background.

### 3.5. The Gluopsins Have Glutamic Acid at Position 296

The gluopsins are arthropod chromopsins mainly specific to butterflies and dragonflies (Figure 2C, Figures S1 and S2). Their lysine 296<sup>7.43</sup> is replaced by a glutamic acid residue (Figures 2C and 3). Therefore, we call them gluopsins, where glu is the three-letter abbreviation of glutamic acid. Our gluopsins form a clade of 36 opsins with 33 gluopsins having glutamic acid 296<sup>7.43</sup>, two are fragments without the seventh transmembrane domain or parts of it, and one is misassembled (File S5: Sorted alignment). However, since we have 33 gluopsins with glutamic acid and since we have them from different higher insect taxa, we can exclude that those are missequenced, misassembled or pseudogenes. With glutamic acid 296<sup>7.43</sup>, the gluopsins (and the astropsins and nemopsins) are special since all other opsins have lysine at this position. Interestingly, gluopsins and astropsins have apparently evolved glutamic acid 296<sup>7.43</sup> independently (Figure 2C).

In contrast to the astropsins, the gluopsins are better studied. Gluopsins in the dragonflies *Sympetrum frequens* (in our tree: BAQ54696.1, Figures S1 and S2) and *Orthetrum albistylum* are expressed sparsely in visual organs of the larva and the adult. These gluopsins have also been experimentally verified by reverse transcriptase PCR and sequencing. Despite this, the original study did not mention glutamic acid 296<sup>7.43</sup> [78]. Our dataset also contains sequences from butterflies and moths such as the common silk moth (*Bombyx mori*), which is of commercial interest, and the tobacco hawk moth (*Manduca sexta*) (XP\_021206870, XP\_030031533, respectively). Both species are model systems.

Previous opsin phylogenies did not include the gluopsins or notice their glutamic acid 296<sup>7.43</sup>. The gluopsins are neither in the datasets of Porter et al. [28] nor of Ramirez et al. [35], because any sequences without lysine 296<sup>7.43</sup> (apart from some outgroup sequences) were excluded. The gluopsins were previously reconstructed as insect peropsins by Henze and Oakley [73], who however did mention glutamic acid 296<sup>7.43</sup>. Böhm et al. [79] described

two putative gluopsins as peropsins, which came from head transcriptomes of scorpionflies. They did not mention glutamic acid 296<sup>7.43</sup>, either. Even so one sequence has it in their alignment while the other is a fragment without the seventh transmembrane domain. These sequences seem not to have been submitted to a sequence database and thus are not in our phylogeny. Böhm et al. [79] also mentioned a sequence from the jewel beetle *Agrilus planipennis* (XP\_025829857), which also has glutamic acid 296<sup>7.43</sup>. This sequence is probably a gluopsin too, even so it was removed from our dataset as a rogue sequence with other sequences from planthoppers, whiteflies, and termites, which all have glutamic acid 296<sup>7.43</sup> and a derived NPxxY<sup>7.53</sup> motif.

### 3.6. The NPxxY<sup>7.53</sup> Motif Is Derived in Some Chromopsins

The NPxxY<sup>7.53</sup> motif in GPCRs is important for G-protein binding and signaling, and thus is conserved. However, the motif has mutations in RGR-opsins and retinochromes. Thus, these RGR-opsins and retinochromes have been claimed not to signal but to work as photoisomerases instead [80,81]. Therefore, we checked whether the varropsins, the astropsins, and the gluopsins also have mutations in their NPxxY<sup>7.53</sup> motif, as this could give us some clues about their function. To answer this question and to check for other conserved residues, we generated a sequence logo for each clade of the chromopsins (Figure 3).

The chromopsins have additional conserved residues such as proline 292<sup>7.39</sup> and arginine 314<sup>7.64</sup>, which is even shared with the other opsins. Besides that, the chromopsins fall into two groups: One group has the peropsins, varropsins, and nemopsins with a well conserved NPxxY<sup>7.53</sup> motif and the other group has the other chromopsins clades. Interestingly, each clade has its own mutations in the NPxxY<sup>7.53</sup> motif: The RGR-opsins have NAxxY<sup>7.53</sup>, while the retinochromes VPxxY<sup>7.53</sup> for annelids or YPxxY<sup>7.53</sup> for mollusks (Figure 3, File S5: Sorted alignment). The astropsins also have mutations within the motif, but since the sequence logo only contains two sequences, it is hard to say what the consensus is. Finally, the gluopsins have the most derived motif of all chromopsins. Most have either PVxxY<sup>7.53</sup> or PLxxY<sup>7.53</sup> (Figure 3). The sequences from Böhm et al. [79] and the jewel beetle (XP\_025829857) also have PVxxY<sup>7.53</sup>.

Whether these two groups are also phylogenetic groups, is unclear due to low support values. All chromopsins with a derived NPxxY<sup>7.53</sup> motif have different mutations, which means they could have evolved independently. However, this could still hint to shared functional requirements, which include also relaxed requirements if these chromopsins do not signal.

## 4. Discussion

### 4.1. The Number of Chromopsins in the Urbilaterian

The retinochromes, like the RGR-opsins and the gluopsins of the chromopsins, have a derived NPxxY<sup>7.53</sup> motif, while the motif of the nemopsins, the peropsins, and the varropsins is conserved (Figure 3). Therefore, we could assume that also the RGR-opsins, retinochromes, and the gluopsins form a group. However, our reconstruction does not recover that relationship with certainty, and the mutations, which are different for the three groups in the motif, suggest independent evolution. In principle based on our final pruned dataset, the urbilaterian could have had one paralogue of the chromopsins that was then duplicated in the craniate lineage. The protostomes have only one paralogue of either retinochromes, varropsins for arachnids, or gluopsins for beetles, scorpionflies, butterflies, and dragonflies. Possibly, the gluopsins are specific to insects, and have been lost in some clades.

For the gluopsins, we could find more insect taxa in our original dataset such as planthoppers, whiteflies, and termites. These sequences have glutamic acid 296<sup>7.43</sup> and a derived NPxxY<sup>7.53</sup> motif, which we can use as a diagnostic feature, to identify them as gluopsins. We also built a tree from the original dataset, where some sequences of annelids and mollusks were reconstructed as true peropsins (data not shown). However, we should

be careful here, because the peropsins do not have such clear features as the gluopsins and the point of pruning rogue sequences was to reconstruct a more reliable phylogeny. In the end, we cannot exclude that the urbilaterian had as many chromopsin paralogues as the seven we have reconstructed.

Postulating so many chromopsins that have been gained in the urbilaterian and then have been lost in the decedents is an unparsimonious hypothesis. In fact, the general assumption is that gains and losses of genes are rare, and therefore reconciliations of gene trees with species trees try to reduce such gains and losses [82,83].

However, opsin gains and losses are common among vertebrates [84] and are also known in protostomes. The water flea [85], the pineal shrimp [86], dragon and damselflies [78,87], *Limulus* [88], and the mantis shrimp [89] have all gained opsins. In contrast, *Drosophila melanogaster* has only seven opsins, which are all rhabdopsins [90]. Therefore, it must have lost the xenopsins, cilopsins, and the three paralogues of the tetraopsins since it evolved from the urbilaterian, as those are present in protostomes and deuterostomes. These are at least five opsin classes we show expanded in our tree (Figure 2). Furthermore, Ramirez et al. [35] concluded that the last common ancestor of deuterostomes and the protostomes had at least 9 opsins, they found that none of the lineages that evolved from the urbilaterian retained all 9 opsins.

Since gains and losses are common among opsins, we did not reconcile our phylogeny with a species tree, as Ramirez et al. [35] did for their phylogeny. Since they are common, the number of chromopsin paralogues may indeed range from one to seven in the urbilaterian.

#### 4.2. The Function of the Chromopsins

The chromopsins are an interesting group of opsins that have diverse and poorly understood functions. Some chromopsins have a derived NPxxY<sup>7.53</sup> motif (Figure 3), which may change their G-protein interaction or activation. Furthermore, many chromopsins bind all-trans-retinal in the dark, including peropsins [91], varropsins [74,92], RGR-opsins [93], and retinochromes [69]. This is unusual as most opsins bind 11-cis-retinal in the dark and convert it to all-trans-retinal when illuminated. Among the chromopsins, the gluopsins lack the well conserved lysine 296<sup>7.43</sup> (Figures 2C and 3). This may prevent them from binding retinal and raises fascinating questions about their evolution and function. The gluopsins could function like other opsins as thermoreceptors [30] or be involved in mechanoreception such as hearing [31].

However, since the gluopsins are more closely related to the other chromopsins, these may give us some clues about their function. The best-studied chromopsins are retinochromes and RGR-opsins. They have, like the gluopsins, a derived NPxxY<sup>7.53</sup> motif (Figure 3), and thus are claimed not to signal [80], but to produce 11-cis-retinal as photoisomerases. This view is considered established in the literature [33,74,80,94,95]. Therefore, it would be reasonable to assume that the gluopsins might be photoisomerases as well. Then, the missing lysine would be an optimization in a high throughput system as covalent binding might cost time. However, we disagree with the literature that the retinochromes and RGR-opsins are established as photoisomerases.

Only if RGR-opsins and retinochromes are indeed photoisomerases, it is reasonable to assume that the gluopsins are photoisomerases. Therefore, we discuss what a photoisomerase is, how something could be experimentally shown to be a photoisomerase, and how a separate photoisomerase could be useful. Then, we evaluate what is known about the different chromopsins beyond their existence, and finally we discuss possible other functions.

#### 4.3. Photoisomerases

A photoisomerase in the general sense is an enzyme that converts a molecule from one isomer to another with the energy of a photon. For that, it binds the molecule before the reaction and releases it afterwards as a photoproduct. Whether the photoproduct is just a byproduct, or the main product used for something else does not matter in this

definition. However, under this definition, cattle rhodopsin, and likely most cilopsins, are photoisomerases, since they release all-trans-retinal once they have converted it from 11-cis-retinal [17]. Here, all-trans-retinal is considered as a byproduct, while the main function is the phototransduction cascade, which is activated by the photoisomerization. In contrast, retinochromes and RGR-opsins have been claimed to produce 11-cis-retinal for other opsins only, and not to activate phototransduction cascades at all [80,81].

For retinal, photoisomerases change the absorption spectrum: 11-cis-retinal absorbs maximally at 380 nm [7] and all-trans-retinal at 387 nm [17]. In contrast, 11-cis-retinal covalently bound to cattle rhodopsin absorbs maximally at 498 nm [96]; and all-trans-retinal covalently bound to RGR-opsin pH-dependently at 469 nm or 370 nm [97], and to retinochrome at 492 nm [68]. Furthermore, binding all-trans-retinal to a lipid in the plasma membrane moves the absorption maximum to 450 nm, and thus this system acts as a protein free photoisomerase [98]. Another kind of photoisomerase exists in the honeybee. It is water-soluble and thus not a transmembrane domain protein such as opsins [99,100].

To show that something is a photoisomerase in the general sense is relatively easy. To show that the photoproduct is used and needed for something else is more difficult. Ideally, if the photoproduct is missing, a phenotype should result. Even if the photoisomerization only supplies a small fraction (e.g., 10%) of what is needed, it still should result in a phenotype, if missing, so that a function can be established.

To make photoisomerases work effectively, they should supply a significant amount of retinal to their target opsin. To do that they need to be present in a comparable amount or in a much higher amount as their target opsins. Additionally, they also should be placed next to their target opsins so that the distance for exchanging retinal is short. This requires the photoisomerase to use as much space as the target opsin. This might be a problem in rods, where the membranes are stacked to hold as many target opsin molecules as possible to catch every photon under low light, scotopic conditions.

These photons can either be used for vision or for the photoisomerase, unless the photoisomerase contributes to visual excitation, as well. This could be either achieved by binding to a different G-protein or by modifying the binding pocket of the photoisomerase, so that the pocket then binds all-trans-retinal in the dark and converts it to 11-cis-retinal for signaling. Basically, the target opsin would also become a photoisomerase in a two-opsin system, where both opsins are photoisomerases for each other. In this system, both opsins can be tuned spectrally independently of each other.

This might be not possible in a system with bistable opsins. Bistable opsins can convert all-trans-retinal back to 11-cis-retinal by absorbing another photon of a different wavelength without releasing retinal [80,101]. These bistable opsins would integrate the function of a photoreceptor, converting 11-cis-retinal to all-trans-retinal to activate the receptor and to convert it back to supply the receptor with 11-cis-retinal. However, the wavelength sensitivities of both photoreactions may depend on each other because they depend on the same molecule.

Having two opsins functioning as photoisomerase for each other or using bistable opsins might solve the space problem. However, the space problem is only crucial for high performance tasks such as vision or the UV-avoidance response of the larva of *Platynereis dumerilii*, which is also fast [102] and uses photoreceptors with stacked membranes [103]. In tasks that do not require the detection of many photons during a short period, photoisomerases might be more useful, but even here the photoisomerase and its target opsin should be next to each other.

#### 4.4. The Varropsins

The varropsins also bind all-trans-retinal in the dark state and isomerase it to 11-cis-retinal upon light exposure [74,92]. This way they are dark-active opsins that are deactivated by light [92]. Varropsins are expressed in the eyes of the spiders *Hasarius adansonii* [74] and *Cupiennius salei* [104]. In *Limulus*, a varropsin is expressed in glia and pigment cells of the eyes next to photoreceptor cells and in the central nervous system [75,88].

#### 4.5. The Peropsins

In mice, a peropsin is localized to the apical microvilli of the retinal pigment epithelium (RPE) [72]. There, it regulates storage or the movement of vitamin A from the retina to the RPE [105]. A peropsin is also expressed in keratinocytes of the human skin. In keratinocyte cell culture, it reacts to UV light if retinal is supplied [106]. In chicken, a peropsin and an RGR-opsin are expressed in the pineal gland and the retina [107]. In amphioxus, a peropsin binds in the dark-state all-trans-retinal instead of 11-cis-retinal [91]. Despite peropsins having been discovered 25 years ago, in 1997 [72], not much more is known about them. This might be, because the human peropsin could not be linked to an eye disease [108,109], which contrasts with human RGR-opsin, which could be linked to retinitis pigmentosa [110].

#### 4.6. The RGR-Opsins

The RGR-opsins have an N<sub>Axx</sub>Y<sup>7.53</sup> motif, instead of the well-conserved NP<sub>xx</sub>Y<sup>7.53</sup> motif (Figure 3, outgroup). This motif is important for G-protein binding. For instance, if it is mutated to N<sub>Axx</sub>Y<sup>7.53</sup> in the rat m3 muscarinic receptor, the receptor can still be activated but less efficiently [20]. Therefore, RGR-opsins are thought to neither signal nor activate a phototransduction cascade [80,81].

However, the human MT2 melatonin receptor signals via a G-protein and has an N<sub>Axx</sub>Y<sup>7.53</sup> motif natively. If this motif is mutated to NP<sub>xx</sub>Y<sup>7.53</sup>, the receptor cannot be activated, but can be rescued partially if the motif is mutated to NV<sub>xx</sub>Y<sup>7.53</sup> [22]. Furthermore, when the motif is mutated to N<sub>Axx</sub>Y<sup>7.53</sup> in cattle rhodopsin, the mutant has 141% of wild type activity [19]. This evidence shows that a GPCR does not need a standard NP<sub>xx</sub>Y<sup>7.53</sup> motif to signal.

RGR-opsins bind all-trans-retinal instead of 11-cis-retinal in the dark [93] and are involved in retinal regeneration [111]. Therefore, RGR-opsins were thought to be photoisomerases [81]. However, in the retinal pigment epithelium (RPE) cells, they are located in the smooth endoplasmic reticulum [112] and regulate retinoid traffic and production [113,114]. In particular, they speed up the production of 11-cis-retinol (an 11-cis-retinal precursor) from all-trans-retinyl-esters, light-independently [115]. The all-trans-retinyl-esters, however, are made available light-dependently by the same RGR-opsins for the Rpe65-isomerase in the RPE. Therefore, RGR-opsins signal, but it is unclear whether they signal via a G-protein or some other mechanism [116]. This is contrasted by the results of Zhang et al. [117], who found that the Rpe65-isomerase activity does not depend on light. However, they used a cell-free RPE-microsome system for their experiments. Microsomes are generated from cell fragments, and thus lack internal lipid storage where substrate for the Rpe65-isomerase could come from. Basically, this only shows that RGR-opsins are photoisomerases in the general sense, it does not show whether they also produce all-trans-retinal for other opsins.

Although, RGR-opsins are present in a relatively high amount compared to the total amount of protein in RPE-cells [70], RPE-cells do not stack their membranes as densely as rods and cones do, so the amount of RGR-opsin in RPE-cells should be relatively low compared to visual opsins in rod and cones. Additionally, RGR-opsins have in vitro only 34% of the photosensitivity of cattle rhodopsin, and they do not readily release 11-cis-retinal, which can be displaced by all-trans-retinal [93]. Finally, RGR-opsins are located in the RPE behind the rod and cones, which take out a significant fraction of photons. In sum, all these properties do not support RGR-opsins as photoisomerases for visual opsins.

Therefore, RGR-opsins apparently function in the RPE primarily as photoreceptors. They may just work in reverse: In the dark, they are active and are inactivated by light, like chicken Opn5L1 (Opn8), which is inactivated by light by isomerizing all-trans-retinal to 11-cis-retinal [118].

In principle, RGR-opsins could still contribute to chromophore production in the RPE as photoisomerases. However, this is difficult to determine, since they increase the substrate



supply for another isomerase [116], and an RGR-opsin knockout thus removes both the 11-cis-retinal produced by the RGR-opsin itself and that from the isomerase.

The cone outer segments also contain RGR-d, an RGR-opsin splice variant [119]. However, RGR-d lacks most of transmembrane domain 6 and thus is inactive [117], but it could still indicate RGR-opsin expression [119]. This could fulfill the need for a highly efficient and abundant photoisomerase in the cone disks next to their target opsins. Additionally, *in vitro* experiments suggest that RGR-opsins may serve as photoisomerases in Müller cells [120] where they are located in the endoplasmic reticulum [112]. However, whether RGR-opsins serve as photoisomerases in cones or whether they contribute a significant amount to chromophore production on their own in the RPE, remains to be seen. It rather seems that the RGR-opsins are provided with all-trans-retinal by the visual opsins and thus the visual opsins serve them as photoisomerases.

Furthermore, if RGR-opsin was present in a high amount then it would reflect so much light of specific wavelengths to the eyes of an observer that it would act as a pigment and give color to its host cells, like rhodopsin, which stains the rods purple. This would be visible to curious researchers; however, the first RGR-opsins were not discovered visually but rather via molecular techniques [70,71]. In fact, 28 years before the first RGR-opsin was found, the first chromopsin was discovered by eye [68]. It was subsequently named retinochrome [69].

#### 4.7. The Retinochromes

Like RGR-opsins, retinochromes are also thought to function as photoisomerases. They bind in the dark state all-trans-retinal [121], which is isomerized by light to 11-cis-retinal and released immediately, so that the retinochrome can readily take up another molecule of all-trans-retinal. Squid retinochrome is so abundant that it could be discovered by eye [68]. In an *in vitro* solution, it can supply cattle rhodopsin with 11-cis-retinal, regenerates with all-trans-retinal several tens of times faster than cattle rhodopsin with 11-cis-retinal [122], and it is more light-sensitive than the visual squid opsin [121]. Therefore, retinochrome would serve as an effective photoisomerase. However, in the squid eye, it is mostly located in the inner segments of the photoreceptor cells, while the visual opsin is located in the outer segments [123]. The inner and outer segments are separated by a dense screen of pigment, so that retinochrome could only use light that comes through body tissues and not from the eyes [68]. Some retinochrome is also found in the outer segments, but in lower amount and not next to the visual opsins in the rhabdoms [124]. However, the visual opsins and the retinochromes could exchange retinal via a shuttle protein [125].

In contrast to the squid eye, the visual opsin is co-expressed with retinochrome in extraocular tissue, such as the longitudinal bundles of central fin muscle, the arm ganglia, the sucker peduncle nerves, the epidermal hair cells, and the parolfactory vesicles [126]. In the parolfactory vesicles, the amount of visual opsin and retinochrome is roughly equal [127], and thus retinochrome could serve there as a photoisomerase for the visual opsin.

This evidence shows that retinochromes can function as photoisomerases. However, this comes from biochemical *in vitro* and *in situ* experiments, which are highly artificial systems, and thus cannot tell whether retinochromes also function as photoisomerases *in vivo*.

#### 4.8. The Gluopsins: Opsins without Lysine 296<sup>7,43</sup>

Here, we have covered everything that is known about the chromopsins and their functions. From that, we conclude that the evidence is not enough to determine if RGR-opsins and retinochromes are photoisomerases. This influences the plausibility whether gluopsins are photoisomerases. Therefore, we should also think beyond the photoisomerase hypothesis.

Recently, the *Drosophila* rhabdoms Rh1, Rh4, and Rh7 were reported to function not only as photoreceptors, but also as chemoreceptors for aristolochic acid. These opsins still

have lysine 296<sup>7.43</sup> like other opsins. However, if this lysine is replaced by an arginine in Rh1, then Rh1 loses light sensitivity but still responds to aristolochic acid. Thus, lysine 296<sup>7.43</sup> is not needed for Rh1 to function as chemoreceptor [26]. Therefore, we wondered whether any opsins existed that had lost lysine 296<sup>7.43</sup> during evolution. Such opsins would form a clade embedded within the other opsins.

Indeed, we found such a clade: the gluopsins. The gluopsins have glutamic acid instead of lysine 296<sup>7.43</sup> and form a strong clade of 36 member sequences (Figures 2C and 3). We also reconstructed the astropsins with one member that also has glutamic acid 296<sup>7.43</sup> (Figures 2C and 3). Previously, three members of the astropsins have been reported to have glutamic acid 296<sup>7.43</sup>. These members were mined from transcriptome and genome databases [37]. However, we could trace two astropsin sequences with glutamic acid 296<sup>7.43</sup> to two independent genome projects from sea urchins [128,129]. Beyond their sequences, however, nothing is known about them. Beside the gluopsins and astropsins, the nemopsins have lysine 296<sup>7.43</sup> replaced, however with arginine (Figures 2C and 3), they also have a conserved NPxxY<sup>7.53</sup> motif. A nemopsin is expressed in chemosensory cells in *C. elegans*. Therefore, the nemopsins are thought to be chemoreceptors [77].

All other opsins without lysine 296<sup>7.43</sup> are isolated single sequences, so that we can assume they have been missequenced, misassembled, or are pseudogenes. For example, in an encephalopsin of the clawed frog, *Xenopus tropicalis* (XM\_002935666), lysine 296<sup>7.43</sup> is replaced by isoleucine and thus is regarded by Kato et al. [95] as a pseudogene. It has been subsequently removed from the NCBI database. Therefore, we do not consider single opsins without lysine 296<sup>7.43</sup> within a clade as real, unless these opsins are at least sequence verified and ideally characterized functionally or by expression.

Gluopsins have been previously described in butterflies, dragonflies, scorpionflies, and in beetles. They have been confirmed in dragonflies by expression data and their sequences have been verified by Sanger sequencing of their cDNA [78]. They were also found in head transcriptomes of scorpionflies [79]. They have been reconstructed as a clade by Henze and Oakley [73], who however did not mention glutamic acid 296<sup>7.43</sup>. Beyond that, nothing is known about them, especially about their function.

The gluopsins share with RGR-opsins and the retinochromes a derived NPxxY<sup>7.53</sup> motif. The motif in gluopsins is PVxxY<sup>7.53</sup> or PLxxY<sup>7.53</sup> (Figure 3). This has two mutations compared to the NAXxY<sup>7.53</sup> motif of the RGR-opsins. Even with this derived motif, we should not exclude that gluopsins signal unless shown otherwise experimentally, since the whole receptor could have acquired compensatory mutations.

Indeed, gluopsins should signal, otherwise they could not function as chemoreceptors, which is a possibility. They could sense chemicals like the *Drosophila* opsins Rh1, Rh4, and Rh7, which sense aristolochic acid without lysine 296<sup>7.43</sup> [26]. However, in cattle rhodopsin, the retinal binding lysine can be shifted from position 296 to other positions, even into other transmembrane domains, without changing the activity [130]. This way the gluopsins could serve as photoreceptors or even as photoisomerases. Even so, it has not been conclusively shown that RGR-opsins and retinochromes are photoisomerases as we have discussed above.

From inspecting our alignment manually, we could not find an alternative position for the retinal binding lysine that is conserved across all gluopsins. However, different gluopsins may have switched the retinal binding lysine to different positions so that all the gluopsins could serve as photoreceptors. Beside light and chemicals, the gluopsins could, like other opsins, also detect heat, phospholipids, mechanical stimuli, or other stimuli [32,33]. In the end however, whether the gluopsins are photoreceptors, chemoreceptors, or something else remains to be determined experimentally.

## 5. Summary

To answer our question, whether opsins exist that have lost lysine 296<sup>7.43</sup> during evolution, we built an automatic phylogeny pipeline that can be easily adjusted for reconstructing the phylogeny of other GPCRs and other proteins. We reconstructed the first 5k

opsin phylogeny, which contains more than 5 times the number of opsin sequences compared to previous large-scale phylogenies [28,35]. The full description of this phylogeny will be published elsewhere. Finally, we answered our question: Opsins that lost lysine 296<sup>7.43</sup> during evolution do exist. In these opsins, lysine 296<sup>7.43</sup> is replaced by glutamic acid 296<sup>7.43</sup>, and thus we call them gluopsins. The gluopsins are found in insects such as beetles, scorpionflies, dragonflies, and butterflies including the silk moth, which is of commercial interest. The gluopsins are an exciting target to study the fascinating functional flexibility of opsins, especially as more opsins with functions beyond light sensitivity are discovered. However, what the function of the gluopsins is, is unknown and so it is to be answered by future research.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/cells11152441/s1>, Figure S1: Phylogenetic reconstruction of the opsins (full tree). Figure S2: Phylogenetic reconstruction of the opsins (chromopsins only). Table S1: The taxon composition in percentage in a table from the pie charts in Figure 2. File S1: Bait sequences. File S2: Rogue bait sequences. File S3: Final sequences of iteration 20. File S4: Gap reduced alignment of iteration 20. File S5: Gap reduced sorted alignment of iteration 20. File S6: Final tree of iteration 20 in newick format.

**Author Contributions:** M.G. designed the study and programmed the pipeline. M.L.P. and M.J.B. generated the fan-worm transcriptome data. M.G. drafted the manuscript. All authors reviewed, edited, and contributed ideas to the final manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research and the APC were funded by the European Union through the Marie Skłodowska-Curie Actions of Horizon 2020 (PhoToBe 846655).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The code of the phylogeny pipeline is available at <https://github.com/MartinGuehmann/PhylogenyPipeline> (accessed on 1 June 2022). The final sequences, the alignments, and the newick tree for the final iteration are included in the Supplemental Files S3 to S6. The complete data generated by the pipeline is available at <https://github.com/MartinGuehmann/Opsins> (accessed on 1 June 2022). Our version of PIA is available at <https://github.com/MartinGuehmann/PIA2> (accessed on 1 June 2022).

**Acknowledgments:** This work was carried out using the computational facilities of the Advanced Computing Research Centre at the University of Bristol: <http://www.bristol.ac.uk/acrc/> (accessed on 1 June 2022).

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analysis, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References

1. Casey, P.J.; Gilman, A.G. G protein involvement in receptor-effector coupling. *J. Biol. Chem.* **1988**, *263*, 2577–2580. [[CrossRef](#)]
2. Attwood, T.K.; Findlay, J.B.C. Fingerprinting G-protein-coupled receptors. *Protein Eng. Des. Sel.* **1994**, *7*, 195–203. [[CrossRef](#)] [[PubMed](#)]
3. Dixon, R.A.F.; Kobilka, B.K.; Strader, D.J.; Benovic, J.L.; Dohlman, H.G.; Frielle, T.; Bolanowski, M.A.; Bennett, C.D.; Rands, E.; Diehl, R.E.; et al. Cloning of the gene and cDNA for mammalian  $\beta$ -adrenergic receptor and homology with rhodopsin. *Nature* **1986**, *321*, 75–79. [[CrossRef](#)] [[PubMed](#)]
4. Dixon, R.A.F.; Sigal, I.S.; Rands, E.; Register, R.B.; Candelore, M.R.; Blake, A.D.; Strader, C.D. Ligand binding to the  $\beta$ -adrenergic receptor involves its rhodopsin-like core. *Nature* **1987**, *326*, 73–77. [[CrossRef](#)]
5. Wald, G. Carotenoids and the Vitamin A Cycle in Vision. *Nature* **1934**, *134*, 65. [[CrossRef](#)]
6. Wald, G.; Brown, P.K.; Hubbard, R.; Oroshnik, W.; Brown, P.K.; Habburd, R.; Wald, G.; Hubbard, R.; Wald, G.; Brown, P.K.; et al. Hindered Cis Isomers of Vitamin A and Retinene: The Structure of the Neo-B Isomer. *Proc. Natl. Acad. Sci. USA* **1955**, *41*, 438–451. [[CrossRef](#)]
7. Brown, P.; Wald, G. The neo-b isomer of vitamin A and retinene. *J. Biol. Chem.* **1956**, *222*, 865–877. [[CrossRef](#)]

8. Oroshnik, W. The synthesis and configuration of neo-b vitamin A and neoretinene b. *J. Am. Chem. Soc.* **1956**, *78*, 2651–2652. [[CrossRef](#)]
9. Oroshnik, W.; Brown, P.K.; Hubbard, R.; Wald, G. Hindered cis isomers of vitamin A and retinene: The structure of the neo-b isomer. *Proc. Natl. Acad. Sci. USA* **1956**, *42*, 578–580. [[CrossRef](#)]
10. Bownds, D. Site of Attachment of Retinal in Rhodopsin. *Nature* **1967**, *216*, 1178–1181. [[CrossRef](#)]
11. Hargrave, P.A.; McDowell, J.H.; Curtis, D.R.; Wang, J.K.; Juszczak, E.; Fong, S.-L.; Rao, J.K.M.; Argos, P. The structure of bovine rhodopsin. *Biophys. Struct. Mech.* **1983**, *9*, 235–244. [[CrossRef](#)]
12. Palczewski, K.; Kumasaka, T.; Hori, T.; Behnke, C.A.; Motoshima, H.; Fox, B.A.; Le Trong, I.; Teller, D.C.; Okada, T.; Stenkamp, R.E.; et al. Crystal Structure of Rhodopsin: A G Protein-Coupled Receptor. *Science* **2000**, *289*, 739–745. [[CrossRef](#)]
13. Murakami, M.; Kouyama, T. Crystal structure of squid rhodopsin. *Nature* **2008**, *453*, 363–367. [[CrossRef](#)]
14. Hubbard, R.; Kropf, A. The action of light on rhodopsin. *Proc. Natl. Acad. Sci. USA* **1958**, *44*, 130–139. [[CrossRef](#)]
15. Kropf, A.; Hubbard, R. The mechanism of bleaching rhodopsin. *Ann. N. Y. Acad. Sci.* **1958**, *74*, 266–280. [[CrossRef](#)]
16. Choe, H.-W.; Kim, Y.J.; Park, J.H.; Morizumi, T.; Pai, E.F.; Krauß, N.; Hofmann, K.P.; Scheerer, P.; Ernst, O.P. Crystal structure of metarhodopsin II. *Nature* **2011**, *471*, 651–655. [[CrossRef](#)]
17. Wald, G. The Molecular Basis of Visual Excitation. *Nature* **1968**, *219*, 800–807. [[CrossRef](#)]
18. Ballesteros, J.A.; Weinstein, H. Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. *Methods Neurosci.* **1995**, *25*, 366–428.
19. Fritze, O.; Filipek, S.; Kuksa, V.; Palczewski, K.; Hofmann, K.P.; Ernst, O.P. Role of the conserved NPxxY(x)<sub>5,6</sub>F motif in the rhodopsin ground state and during activation. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 2290–2295. [[CrossRef](#)]
20. Wess, J.; Nanavati, S.; Vogel, Z.; Maggio, R. Functional role of proline and tryptophan residues highly conserved among G protein-coupled receptors studied by mutational analysis of the m3 muscarinic receptor. *EMBO J.* **1993**, *12*, 331–338. [[CrossRef](#)]
21. Galés, C.; Kowalski-Chauvel, A.; Dufour, M.N.; Seva, C.; Moroder, L.; Pradayrol, L.; Vaysse, N.; Fourmy, D.; Silvente-Poirot, S. Mutation of Asn-391 within the Conserved NPXXY Motif of the Cholecystokinin B Receptor Abolishes Gq Protein Activation without Affecting Its Association with the Receptor. *J. Biol. Chem.* **2000**, *275*, 17321–17327. [[CrossRef](#)] [[PubMed](#)]
22. Mazna, P.; Grycova, L.; Balik, A.; Zemkova, H.; Friedlova, E.; Obsilova, V.; Obsil, T.; Teisinger, J. The role of proline residues in the structure and function of human MT2 melatonin receptor. *J. Pineal Res.* **2008**, *45*, 361–372. [[CrossRef](#)] [[PubMed](#)]
23. Borroto-Escuela, D.O.; Romero-Fernandez, W.; García-Negredo, G.; Correia, P.A.; Garriga, P.; Fuxe, K.; Ciruela, F. Dissecting the Conserved NPxxY Motif of the M3 Muscarinic Acetylcholine Receptor: Critical Role of Asp-7.49 for Receptor Signaling and Multiprotein Complex Formation. *Cell. Physiol. Biochem.* **2011**, *28*, 1009–1022. [[CrossRef](#)] [[PubMed](#)]
24. Zhou, Q.; Yang, D.; Wu, M.; Guo, Y.; Guo, W.; Zhong, L.; Cai, X.; Dai, A.; Jang, W.; Shakhnovich, E.I.; et al. Common activation mechanism of class A GPCRs. *eLife* **2019**, *8*, e50279. [[CrossRef](#)]
25. Ovchinnikov, Y. Rhodopsin and bacteriorhodopsin: Structure-function relationships. *FEBS Lett.* **1982**, *148*, 179–191. [[CrossRef](#)]
26. Leung, N.Y.; Thakur, D.P.; Gurav, A.S.; Kim, S.H.; Di Pizio, A.; Niv, M.Y.; Montell, C. Functions of Opsins in Drosophila Taste. *Curr. Biol.* **2020**, *30*, 1367–1379.e6. [[CrossRef](#)]
27. Ramirez, M.D.; Oakley, T.H. Eye-independent, light-activated chromatophore expansion (LACE) and expression of phototransduction genes in the skin of *Octopus bimaculoides*. *J. Exp. Biol.* **2015**, *218*, 1513–1520. [[CrossRef](#)]
28. Porter, M.L.; Blasic, J.R.; Bok, M.J.; Cameron, E.G.; Pringle, T.; Cronin, T.W.; Robinson, P.R. Shedding new light on opsin evolution. *Proc. R. Soc. B Biol. Sci.* **2011**, *279*, 3–14. [[CrossRef](#)]
29. Feuda, R.; Hamilton, S.C.; McInerney, J.O.; Pisani, D. Metazoan opsin evolution reveals a simple route to animal vision. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 18868–18872. [[CrossRef](#)]
30. Shen, W.L.; Kwon, Y.; Adegbola, A.A.; Luo, J.; Chess, A.; Montell, C. Function of Rhodopsin in Temperature Discrimination in *Drosophila*. *Science* **2011**, *331*, 1333–1336. [[CrossRef](#)]
31. Senthilan, P.R.; Piepenbrock, D.; Ovezmyradov, G.; Nadrowski, B.; Bechstedt, S.; Pauls, S.; Winkler, M.; Möbius, W.; Howard, J.; Göpfert, M.C. Drosophila Auditory Organ Genes and Genetic Hearing Defects. *Cell* **2012**, *150*, 1042–1054. [[CrossRef](#)]
32. Feuda, R.; Menon, A.K.; Göpfert, M.C. Rethinking Opsins. *Mol. Biol. Evol.* **2022**, *39*, msac033. [[CrossRef](#)]
33. Leung, N.Y.; Montell, C. Unconventional Roles of Opsins. *Annu. Rev. Cell Dev. Biol.* **2017**, *33*, 241–264. [[CrossRef](#)]
34. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [[CrossRef](#)]
35. Ramirez, M.D.; Pairett, A.N.; Pankey, M.S.; Serb, J.; Speiser, D.I.; Swafford, A.J.; Oakley, T.H. The last common ancestor of most bilaterian animals possessed at least 9 opsins. *Genome Biol. Evol.* **2016**, *8*, 3640–3652. [[CrossRef](#)]
36. Lowe, E.K.; Garm, A.L.; Ullrich-Lüter, E.; Cuomo, C.; Arnone, M.I. The crowns have eyes: Multiple opsins found in the eyes of the crown-of-thorns starfish *Acanthaster planci*. *BMC Evol. Biol.* **2018**, *18*, 168. [[CrossRef](#)]
37. D’Aniello, S.; Delroisse, J.; Valero-Gracia, A.; Lowe, E.; Byrne, M.; Cannon, J.; Halanych, K.; Elphick, M.; Mallefet, J.; Kaul-Strehlow, S.; et al. Opsin evolution in the Ambulacraria. *Mar. Genom.* **2015**, *24*, 177–183. [[CrossRef](#)]
38. Conzelmann, M.; Williams, E.A.; Krug, K.; Franz-Wachtel, M.; Macek, B.; Jékely, G. The neuropeptide complement of the marine annelid *Platynereis dumerilii*. *BMC Genom.* **2013**, *14*, 906. [[CrossRef](#)]
39. Speiser, D.I.; Pankey, M.S.; Zaharoff, A.K.; Battelle, B.A.; Bracken-Grissom, H.D.; Breinholt, J.W.; Bybee, S.M.; Cronin, T.W.; Garm, A.; Lindgren, A.R.; et al. Using phylogenetically-informed annotation (PIA) to search for light-interacting genes in transcriptomes from non-model organisms. *BMC Bioinform.* **2014**, *15*, 350. [[CrossRef](#)]



40. Pérez-Moreno, J.L.; DeLeo, D.M.; Palero, F.; Bracken-Grissom, H.D. Phylogenetic annotation and genomic architecture of opsin genes in Crustacea. *Hydrobiologia* **2018**, *825*, 159–175. [CrossRef]
41. Shen, W.; Le, S.; Li, Y.; Hu, F. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLoS ONE* **2016**, *11*, e0163962. [CrossRef]
42. Li, W.; Godzik, A. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **2006**, *22*, 1658–1659. [CrossRef]
43. Fu, L.; Niu, B.; Zhu, Z.; Wu, S.; Li, W. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* **2012**, *28*, 3150–3152. [CrossRef]
44. Mirarab, S.; Nguyen, N.-P.; Guo, S.; Wang, L.-S.; Kim, J.; Warnow, T. PASTA: Ultra-Large Multiple Sequence Alignment for Nucleotide and Amino-Acid Sequences. *J. Comput. Biol.* **2015**, *22*, 377–386. [CrossRef]
45. Minh, B.Q.; Schmidt, H.A.; Chernomor, O.; Schrempf, D.; Woodhams, M.D.; von Haeseler, A.; Lanfear, R.; Teeling, E. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **2020**, *37*, 1530–1534. [CrossRef]
46. Junier, T.; Zdobnov, E.M. The Newick utilities: High-throughput phylogenetic tree processing in the UNIX shell. *Bioinformatics* **2010**, *26*, 1669–1670. [CrossRef]
47. Aberer, A.J.; Krompass, D.; Stamatakis, A. Pruning Rogue Taxa Improves Phylogenetic Accuracy: An Efficient Algorithm and Web Service. *Syst. Biol.* **2013**, *62*, 162–166. [CrossRef]
48. Saunders, A.M.; Ashlock, D.; Graether, S.P. Testing the rogue taxa hypothesis for clustering instability. *J. Theor. Biol.* **2019**, *472*, 36–45. [CrossRef]
49. Aberer, A.J.; Krompaß, D.; Stamatakis, A. RogueNaRok: An Efficient and Exact Algorithm for Rogue Taxon Identification. Heidelberg Institute for Theoretical Studies: Exelixis-RRDR-2011–10. 2011. Available online: <https://cme.h-its.org/exelixis/pubs/Exelixis-RRDR-2011-10.pdf> (accessed on 1 June 2022).
50. Mai, U.; Mirarab, S. TreeShrink: Fast and accurate detection of outlier long branches in collections of phylogenetic trees. *BMC Genom.* **2018**, *19*, 272. [CrossRef]
51. Capella-Gutiérrez, S.; Silla-Martínez, J.M.; Gabaldón, T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **2009**, *25*, 1972–1973. [CrossRef]
52. Guindon, S.; Dufayard, J.-F.; Lefort, V.; Anisimova, M.; Hordijk, W.; Gascuel, O. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Syst. Biol.* **2010**, *59*, 307–321. [CrossRef] [PubMed]
53. Anisimova, M.; Gil, M.; Dufayard, J.-F.; Dessimoz, C.; Gascuel, O. Survey of Branch Support Methods Demonstrates Accuracy, Power, and Robustness of Fast Likelihood-based Approximation Schemes. *Syst. Biol.* **2011**, *60*, 685–699. [CrossRef] [PubMed]
54. Minh, B.Q.; Nguyen, M.A.T.; Von Haeseler, A. Ultrafast Approximation for Phylogenetic Bootstrap. *Mol. Biol. Evol.* **2013**, *30*, 1188–1195. [CrossRef] [PubMed]
55. Hoang, D.T.; Chernomor, O.; Von Haeseler, A.; Minh, B.Q.; Vinh, L.S. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* **2018**, *35*, 518–522. [CrossRef]
56. Simmons, M.P.; Norton, A.P. Divergent maximum-likelihood-branch-support values for polytomies. *Mol. Phylogenet. Evol.* **2014**, *73*, 87–96. [CrossRef]
57. Liu, K.; Raghavan, S.; Nelesen, S.; Linder, C.R.; Warnow, T. Rapid and Accurate Large-Scale Coestimation of Sequence Alignments and Phylogenetic Trees. *Science* **2009**, *324*, 1561–1564. [CrossRef]
58. Ali, R.H.; Bogusz, M.; Whelan, S. Identifying Clusters of High Confidence Homologies in Multiple Sequence Alignments. *Mol. Biol. Evol.* **2019**, *36*, 2340–2351. [CrossRef]
59. Chang, J.-M.; Di Tommaso, P.; Notredame, C. TCS: A New Multiple Sequence Alignment Reliability Measure to Estimate Alignment Accuracy and Improve Phylogenetic Tree Reconstruction. *Mol. Biol. Evol.* **2014**, *31*, 1625–1637. [CrossRef]
60. Tan, G.; Muffato, M.; Ledergerber, C.; Herrero, J.; Goldman, N.; Gil, M.; Dessimoz, C. Current Methods for Automated Filtering of Multiple Sequence Alignments Frequently Worsen Single-Gene Phylogenetic Inference. *Syst. Biol.* **2015**, *64*, 778–791. [CrossRef]
61. Huerta-Cepas, J.; Serra, F.; Bork, P. ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Mol. Biol. Evol.* **2016**, *33*, 1635–1638. [CrossRef]
62. Tareen, A.; Kinney, J.B. Logomaker: Beautiful sequence logos in Python. *Bioinformatics* **2020**, *36*, 2272–2274. [CrossRef]
63. Hunter, J.D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95. [CrossRef]
64. Huson, D.H.; Scornavacca, C. Dendroscope 3: An Interactive Tool for Rooted Phylogenetic Trees and Networks. *Syst. Biol.* **2012**, *61*, 1061–1067. [CrossRef]
65. Feuda, R.; Rota-Stabelli, O.; Oakley, T.; Pisani, D. The Comb Jelly Opsins and the Origins of Animal Phototransduction. *Genome Biol. Evol.* **2014**, *6*, 1964–1971. [CrossRef]
66. Artigas, G.Q.; Lapébie, P.; Leclère, L.; Takeda, N.; Deguchi, R.; Jékely, G.; Momose, T.; Houliston, E. A gonad-expressed opsin mediates light-induced spawning in the jellyfish *Clytia*. *eLife* **2018**, *7*, 1–12. [CrossRef]
67. Rawlinson, K.A.; Lapraz, F.; Ballister, E.R.; Terasaki, M.; Rodgers, J.; McDowell, R.J.; Girstmair, J.; Criswell, K.E.; Boldogkoi, M.; Simpson, F.; et al. Extraocular, rod-like photoreceptors in a flatworm express xenopsin photopigment. *eLife* **2019**, *8*, e45465. [CrossRef]
68. Hara, T.; Hara, R. New Photosensitive Pigment Found in the Retina of the Squid *Ommastrephes*. *Nature* **1965**, *206*, 1331–1334. [CrossRef]



69. Hara, T.; Hara, R.; Takeuchi, J. Vision in Octopus and Squid: Rhodopsin and Retinochrome in the Octopus Retina. *Nature* **1967**, *214*, 572–573. [[CrossRef](#)]
70. Jiang, M.; Pandey, S.; Fong, H.K. An opsin homologue in the retina and pigment epithelium. *Investig. Ophthalmol. Vis. Sci.* **1993**, *34*, 3669–3678.
71. Shen, D.; Jiang, M.; Hao, W.; Tao, L.; Salazar, M.; Fong, H.K.W. A Human Opsin-Related Gene That Encodes a Retinaldehyde-Binding Protein. *Biochemistry* **1994**, *33*, 13117–13125. [[CrossRef](#)]
72. Sun, H.; Gilbert, D.J.; Copeland, N.G.; Jenkins, N.A.; Nathans, J. Peropsin, a novel visual pigment-like protein located in the apical microvilli of the retinal pigment epithelium. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 9893–9898. [[CrossRef](#)]
73. Henze, M.J.; Oakley, T.H. The Dynamic Evolutionary History of Pancrustacean Eyes and Opsins. *Integr. Comp. Biol.* **2015**, *55*, 830–842. [[CrossRef](#)]
74. Nagata, T.; Koyanagi, M.; Tsukamoto, H.; Terakita, A. Identification and characterization of a protostome homologue of peropsin from a jumping spider. *J. Comp. Physiol. A Sens. Neural Behav. Physiol.* **2010**, *196*, 51–59. [[CrossRef](#)]
75. Battelle, B.-A.; Kempler, K.E.; Saraf, S.R.; Marten, C.E.; Dugger, D.R.; Speiser, D.; Oakley, T.H. Opsins in Limulus eyes: Characterization of three visible light-sensitive opsins unique to and co-expressed in median eye photoreceptors and a peropsin/RGR that is expressed in all eyes. *J. Exp. Biol.* **2014**, *218*, 466–479. [[CrossRef](#)]
76. Marlow, H.; Tosches, M.A.; Tomer, R.; Steinmetz, P.R.; Lauri, A.; Larsson, T.; Arendt, D. Larval body patterning and apical organs are conserved in animal evolution. *BMC Biol.* **2014**, *12*, 7. [[CrossRef](#)]
77. Troemel, E.R.; Chou, J.H.; Dwyer, N.D.; Colbert, H.A.; Bargmann, C.I. Divergent seven transmembrane receptors are candidate chemosensory receptors in *C. elegans*. *Cell* **1995**, *83*, 207–218. [[CrossRef](#)]
78. 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. USA* **2015**, *112*, E1247–E1256. [[CrossRef](#)]
79. Böhm, A.; Meusemann, K.; Misof, B.; Pass, G. Hypothesis on monochromatic vision in scorpionflies questioned by new transcriptomic data. *Sci. Rep.* **2018**, *8*, 9872. [[CrossRef](#)]
80. Tsukamoto, H.; Terakita, A. Diversity and functional properties of bistable pigments. *Photochem. Photobiol. Sci.* **2010**, *9*, 1435. [[CrossRef](#)]
81. Terakita, A. The opsins. *Genome Biol.* **2005**, *6*, 213. [[CrossRef](#)]
82. Chen, K.; Durand, D.; Farach-Colton, M. NOTUNG: A Program for Dating Gene Duplications and Optimizing Gene Family Trees. *J. Comput. Biol.* **2000**, *7*, 429–447. [[CrossRef](#)] [[PubMed](#)]
83. Szöllösi, G.J.; Tannier, E.; Daubin, V.; Boussau, B. The Inference of Gene Trees with Species Trees. *Syst. Biol.* **2014**, *64*, e42–e62. [[CrossRef](#)] [[PubMed](#)]
84. Musser, J.M.; Arendt, D. Loss and gain of cone types in vertebrate ciliary photoreceptor evolution. *Dev. Biol.* **2017**, *431*, 26–35. [[CrossRef](#)] [[PubMed](#)]
85. Colbourne, J.K.; Pfrender, M.E.; Gilbert, D.; Thomas, W.K.; Tucker, A.; Oakley, T.H.; Tokishita, S.; Aerts, A.; Arnold, G.J.; Basu, M.K.; et al. The Ecoresponsive Genome of *Daphnia pulex*. *Science* **2011**, *331*, 555–561. [[CrossRef](#)] [[PubMed](#)]
86. Zhang, X.; Yuan, J.; Sun, Y.; Li, S.; Gao, Y.; Yu, Y.; Liu, C.; Wang, Q.; Lv, X.; Zhang, X.; et al. Penaeid shrimp genome provides insights into benthic adaptation and frequent molting. *Nat. Commun.* **2019**, *10*, 356. [[CrossRef](#)]
87. Suvorov, A.; Jensen, N.O.; Sharkey, C.R.; Fujimoto, M.S.; Bodily, P.; Wightman, H.M.C.; Ogden, T.H.; Clement, M.J.; Bybee, S.M. Opsins have evolved under the permanent heterozygote model: Insights from phylotranscriptomics of Odonata. *Mol. Ecol.* **2017**, *26*, 1306–1322. [[CrossRef](#)]
88. Battelle, B.-A.; Ryan, J.F.; Kempler, K.E.; Saraf, S.R.; Marten, C.E.; Warren, W.C.; Minx, P.J.; Montague, M.J.; Green, P.J.; Schmidt, S.A.; et al. Opsin Repertoire and Expression Patterns in Horseshoe Crabs: Evidence from the Genome of *Limulus polyphemus* (Arthropoda: Chelicerata). *Genome Biol. Evol.* **2016**, *8*, 1571–1589. [[CrossRef](#)]
89. Porter, M.L.; Awata, H.; Bok, M.J.; Cronin, T.W. Exceptional diversity of opsin expression patterns in *Neogonodactylus oerstedii* (Stomatopoda) retinas. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 8948–8957. [[CrossRef](#)]
90. Feuda, R.; Goulty, M.; Zadra, N.; Gasparetti, T.; Rosato, E.; Pisani, D.; Rizzoli, A.; Segata, N.; Ometto, L.; Stabelli, O.R. Phylogenomics of Opsin Genes in Diptera Reveals Lineage-Specific Events and Contrasting Evolutionary Dynamics in *Anopheles* and *Drosophila*. *Genome Biol. Evol.* **2021**, *13*, evab170. [[CrossRef](#)]
91. Koyanagi, M.; Terakita, A.; Kubokawa, K.; Shichida, Y. Amphioxus homologs of Go-coupled rhodopsin and peropsin having 11-*cis*- and all-*trans*-retinals as their chromophores. *FEBS Lett.* **2002**, *531*, 525–528. [[CrossRef](#)]
92. Nagata, T.; Koyanagi, M.; Lucas, R.; Terakita, A. An all-*trans*-retinal-binding opsin peropsin as a potential dark-active and light-inactivated G protein-coupled receptor. *Sci. Rep.* **2018**, *8*, 3535. [[CrossRef](#)]
93. Hao, W.; Fong, H.K.W. The Endogenous Chromophore of Retinal G Protein-coupled Receptor Opsin from the Pigment Epithelium. *J. Biol. Chem.* **1999**, *274*, 6085–6090. [[CrossRef](#)]
94. Gehring, W.J. The evolution of vision. *Wiley Interdiscip. Rev. Dev. Biol.* **2014**, *3*, 1–40. [[CrossRef](#)]
95. Kato, M.; Sugiyama, T.; Sakai, K.; Yamashita, T.; Fujita, H.; Sato, K.; Tomonari, S.; Shichida, Y.; Ohuchi, H. Two Opsin 3-Related Proteins in the Chicken Retina and Brain: A TMT-Type Opsin 3 Is a Blue-Light Sensor in Retinal Horizontal Cells, Hypothalamus, and Cerebellum. *PLoS ONE* **2016**, *11*, e0163925. [[CrossRef](#)]
96. Wald, G. The Photochemical Basis of Rod Vision. *J. Opt. Soc. Am.* **1951**, *41*, 949–955. [[CrossRef](#)]

97. Hao, W.; Fong, H.K.W. Blue and Ultraviolet Light-Absorbing Opsin from the Retinal Pigment Epithelium. *Biochemistry* **1996**, *35*, 6251–6256. [CrossRef]
98. Kaylor, J.J.; Xu, T.; Ingram, N.T.; Tsan, A.; Hakobyan, H.; Fain, G.; Travis, G.H. Blue light regenerates functional visual pigments in mammals through a retinyl-phospholipid intermediate. *Nat. Commun.* **2017**, *8*, 16. [CrossRef]
99. Pepe, I.M.; Cugnoli, C. New trends in photobiology. *J. Photochem. Photobiol. B Biol.* **1992**, *13*, 5–17. [CrossRef]
100. Goldsmith, T.H. Evolutionary tinkering with visual photoreception. *Vis. Neurosci.* **2012**, *30*, 21–37. [CrossRef]
101. Koyanagi, M.; Terakita, A. Diversity of animal opsin-based pigments and their optogenetic potential. *Biochim. Biophys. Acta* **2013**, *1837*, 710–716. [CrossRef]
102. Veraszto, C.; Gühmann, M.; Jia, H.; Rajan, V.B.V.; Bezares-Calderón, L.A.; Piñeiro-Lopez, C.; Randel, N.; Shahidi, R.; Michiels, N.K.; Yokoyama, S.; et al. Ciliary and rhabdomeric photoreceptor-cell circuits form a spectral depth gauge in marine zooplankton. *eLife* **2018**, *7*, e36440. [CrossRef]
103. Arendt, D.; Tessmar-Raible, K.; Snyman, H.; Dorresteijn, A.W.; Wittbrodt, J. Ciliary Photoreceptors with a Vertebrate-Type Opsin in an Invertebrate Brain. *Science* **2004**, *306*, 869–871. [CrossRef]
104. Eriksson, B.J.; Fredman, D.; Steiner, G.; Schmid, A. Characterisation and localisation of the opsin protein repertoire in the brain and retinas of a spider and an onychophoran. *BMC Evol. Biol.* **2013**, *13*, 186. [CrossRef]
105. Cook, J.D.; Ng, S.Y.; Lloyd, M.; Eddington, S.; Sun, H.; Nathans, J.; Bok, D.; Radu, R.A.; Travis, G.H. Peropsin modulates transit of vitamin A from retina to retinal pigment epithelium. *J. Biol. Chem.* **2017**, *292*, 21407–21416. [CrossRef]
106. Toh, P.P.C.; Yap, A.M.Y.; Sriram, G.; Bigliardi, P.; Bigliardi-Qi, M. Expression of peropsin in human skin is related to phototransduction of violet light in keratinocytes. *Exp. Dermatol.* **2016**, *25*, 1002–1005. [CrossRef]
107. Bailey, M.J.; Cassone, V.M. Opsin Photoisomerases in the Chick Retina and Pineal Gland: Characterization, Localization, and Circadian Regulation. *Investig. Ophthalmol. Vis. Sci.* **2004**, *45*, 769–775. [CrossRef]
108. Ksantini, M.; Sénéchal, A.; Humbert, G.; Arnaud, B.; Hamel, C.P. RRH, Encoding the RPE-Expressed Opsin-Like Peropsin, Is Not Mutated in Retinitis Pigmentosa and Allied Diseases. *Ophthalmic Genet.* **2007**, *28*, 31–37. [CrossRef]
109. Rivolta, C.; Berson, E.L.; Dryja, T.P. Mutation Screening of the Peropsin Gene, a Retinal Pigment Epithelium Specific Rhodopsin Homolog, in Patients with Retinitis Pigmentosa and Allied Diseases. *Mol. Vis.* **2006**, *12*, 1511–1515.
110. Morimura, H.; Saindelle-Ribeaudéau, F.; Berson, E.L.; Dryja, T.P. Mutations in RGR, encoding a light-sensitive opsin homologue, in patients with retinitis pigmentosa. *Nat. Genet.* **1999**, *23*, 393–394. [CrossRef]
111. Chen, P.; Hao, W.; Rife, L.; Wang, X.P.; Shen, D.; Chen, J.; Ogden, T.; Van Boemel, G.B.; Wu, L.; Yang, M.; et al. A photic visual cycle of rhodopsin regeneration is dependent on Rgr. *Nat. Genet.* **2001**, *28*, 256–260. [CrossRef]
112. Pandey, S.; Blanks, J.C.; Spee, C.; Jiang, M.; Fong, H.K. Cytoplasmic Retinal Localization of an Evolutionary Homolog of the Visual Pigments. *Exp. Eye Res.* **1994**, *58*, 605–613. [CrossRef] [PubMed]
113. Shichida, Y.; Matsuyama, T. Evolution of opsins and phototransduction. *Philos. Trans. R. Soc. B: Biol. Sci.* **2009**, *364*, 2881–2895. [CrossRef] [PubMed]
114. Nagata, T.; Koyanagi, M.; Terakita, A. Evolution and Functional Diversity of Opsin-Based Photopigments. Available online: <http://photobiology.info/Terakita.html> (accessed on 3 August 2021).
115. Wenzel, A.; Oberhauser, V.; Pugh, E.N.; Lamb, T.D.; Grimm, C.; Samardzija, M.; Fahl, E.; Seeliger, M.W.; Remé, C.E.; von Lintig, J. The Retinal G Protein-coupled Receptor (RGR) Enhances Isomerohydrolase Activity Independent of Light. *J. Biol. Chem.* **2005**, *280*, 29874–29884. [CrossRef] [PubMed]
116. Radu, R.A.; Hu, J.; Peng, J.; Bok, D.; Mata, N.L.; Travis, G.H. Retinal Pigment Epithelium-Retinal G Protein Receptor-Opsin Mediates Light-dependent Translocation of All-trans-retinyl Esters for Synthesis of Visual Chromophore in Retinal Pigment Epithelial Cells. *J. Biol. Chem.* **2008**, *283*, 19730–19738. [CrossRef]
117. Zhang, J.; Choi, E.H.; Tworak, A.; Salom, D.; Leinonen, H.; Sander, C.L.; Hoang, T.V.; Handa, J.T.; Blackshaw, S.; Palczewska, G.; et al. Photic generation of 11-cis-retinal in bovine retinal pigment epithelium. *J. Biol. Chem.* **2019**, *294*, 19137–19154. [CrossRef]
118. Sato, K.; Yamashita, T.; Ohuchi, H.; Takeuchi, A.; Gotoh, H.; Ono, K.; Mizuno, M.; Mizutani, Y.; Tomonari, S.; Sakai, K.; et al. Opm5L1 is a retinal receptor that behaves as a reverse and self-regenerating photoreceptor. *Nat. Commun.* **2018**, *9*, 1255. [CrossRef]
119. Zhang, Z.; Fong, H.K. Coexpression of nonvisual opsin, retinal G protein-coupled receptor, and visual pigments in human and bovine cone photoreceptors. *Mol. Vis.* **2018**, *24*, 434–442.
120. Morshedian, A.; Kaylor, J.J.; Ng, S.Y.; Tsan, A.; Frederiksen, R.; Xu, T.; Yuan, L.; Sampath, A.P.; Radu, R.A.; Fain, G.L.; et al. Light-Driven Regeneration of Cone Visual Pigments through a Mechanism Involving RGR Opsin in Müller Glial Cells. *Neuron* **2019**, *102*, 1172–1183.e5. [CrossRef]
121. Hara, T.; Hara, R. Vision in Octopus and Squid: Rhodopsin and Retinochrome in the Squid Retina. *Nature* **1967**, *214*, 573–575. [CrossRef]
122. Hara, T.; Hara, R. Regeneration of Squid Retinochrome. *Nature* **1968**, *219*, 450–454. [CrossRef]
123. Ozaki, K.; Hara, R.; Hara, T. Histochemical localization of retinochrome and rhodopsin studied by fluorescence microscopy. *Cell Tissue Res.* **1983**, *233*, 335–345. [CrossRef]
124. Hara, T.; Hara, R. Distribution of rhodopsin and retinochrome in the squid retina. *J. Gen. Physiol.* **1976**, *67*, 791–805. [CrossRef]
125. Terakita, A.; Hara, R.; Hara, T. Retinal-binding protein as a shuttle for retinal in the rhodopsin-retinochrome system of the squid visual cells. *Vis. Res.* **1989**, *29*, 639–652. [CrossRef]

126. Kingston, A.C.N.; Wardill, T.J.; Hanlon, R.T.; Cronin, T.W. An Unexpected Diversity of Photoreceptor Classes in the Longfin Squid, *Doryteuthis pealeii*. *PLoS ONE* **2015**, *10*, e0135381. [[CrossRef](#)]
127. Hara, T.; Hara, R. Retinochrome and rhodopsin in the extraocular photoreceptor of the squid, *Todarodes*. *J. Gen. Physiol.* **1980**, *75*, 1–19. [[CrossRef](#)]
128. Sea Urchin Genome Sequencing Consortium; Sodergren, E.; Weinstock, G.M.; Davidson, E.H.; Cameron, R.A.; Gibbs, R.A.; Angerer, R.C.; Angerer, L.M.; Arnone, M.I.; Burgess, D.R.; et al. The Genome of the Sea Urchin *Strongylocentrotus purpuratus*. *Science* **2006**, *314*, 941–952. [[CrossRef](#)]
129. Davidson, P.L.; Guo, H.; Wang, L.; Berrio, A.; Zhang, H.; Chang, Y.; Soborowski, A.L.; McClay, D.R.; Fan, G.; Wray, G.A. Chromosomal-Level Genome Assembly of the Sea Urchin *Lytechinus variegatus* Substantially Improves Functional Genomic Analyses. *Genome Biol. Evol.* **2020**, *12*, 1080–1086. [[CrossRef](#)]
130. Devine, E.L.; Oprian, D.D.; Theobald, D.L. Relocating the active-site lysine in rhodopsin and implications for evolution of retinylidene proteins. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 13351–13355. [[CrossRef](#)]