Evolution of the vertebrate phototransduction cascade activation steps

Trevor D. Lamb[1] and David M. Hunt[2,3]

[1]Eccles Institute of Neuroscience, John Curtin School of Medical Research, The Australian National University, ACT 2600, Australia, and [2]The Lions Eye Institute and [3]School of Biological Sciences, The University of Western Australia, WA 6009, Australia.

Correspondence: Professor Trevor Lamb <Trevor.Lamb@anu.edu.au>

Keywords: Evolution; Phototransduction; Transducin; Phosphodiesterase; Cyclic nucleotide-gated channel; Opsin

ABSTRACT

We examine the molecular phylogeny of the proteins underlying the activation steps of vertebrate phototransduction, for both agnathan and jawed vertebrate taxa. We expand the number of taxa analysed and we update the alignment and tree building methodology from a previous analysis. For each of the four primary components (the G-protein transducin alpha subunit, $G\alpha_T$, the cyclic GMP phosphodiesterase, PDE6, and the alpha and beta subunits of the cGMP-gated ion channel, CNGC), the phylogenies appear consistent with expansion from an ancestral proto-vertebrate cascade during two rounds of whole-genome duplication followed by divergence of the agnathan and jawed vertebrate lineages. In each case, we consider possible scenarios for the underlying gene duplications and losses, and we apply relevant constraints to the tree construction. From tests of the topology of the resulting trees, we obtain a scenario for the expansion of each component during 2R that accurately fits the observations. Similar analysis of the visual opsins indicates that the only expansion to have occurred during 2R was the formation of Rh1 and Rh2. Finally, we propose a hypothetical scenario for the conversion of an ancestral chordate cascade into the proto-vertebrate phototransduction cascade, prior to whole-genome duplication. Together, our models provide a plausible account for the origin and expansion of the vertebrate phototransduction cascade.

INTRODUCTION

The molecular mechanism of sensory reception is especially well understood in the case of phototransduction in vertebrate cones and rods. However, much less is known about the evolution of this cascade or its development in embryonic photoreceptor cells. One intriguing feature of vertebrate phototransduction is that the cone and rod photoreceptors (which operate under daytime and night-time light levels, respectively) in most cases utilise distinct isoforms of the proteins mediating activation, recovery, and adaptation of the cascade.

For the activation steps of the cascade, distinct genes encode the cone and rod isoforms for each of the four principal protein players: namely, the G-protein transducin alpha subunit ($G\alpha_T$), the cyclic GMP phosphodiesterase (PDE6), and the alpha and beta subunits of the cGMP-gated ion channel (CNGC). Examination of synteny has provided powerful evidence that, for each of these components, the separate cone and rod isoforms arose during two rounds of genome duplication (Nordström et al, 2004; Larhammar et al, 2009; Lagman et al, 2012, 2016).

According to a concept originally proposed by Ohno (1970), it is now widely accepted that many vertebrate genes diversified through two rounds (2R) of whole-genome duplication that occurred prior to the radiation of jawed vertebrates. However, there is debate as to whether both duplications actually involved the whole genome, and whether the second duplication occurred before or after the divergence of jawed and agnathan (jawless) vertebrates, as has been discussed for example by Kuraku et al (2009), Smith et al (2013) and Smith & Keinath (2015). For our analysis in this paper, we adopt the conventional 2R assumption, that both genome duplications occurred before the ancestors of jawed vertebrates diverged from the ancestors of extant agnathan vertebrates.

For phototransduction in agnathan vertebrates, the genes encoding $G\alpha_T$ and PDE6 were examined for a northern hemisphere lamprey by Muradov et al (2007, 2008). Recently, the sequences for all four activation components were obtained from one species of hagfish and two species of southern hemisphere lamprey, as well as several basal species of fish, including sharks, rays, gar and bowfin (Lamb et al, 2016). By constructing molecular phylogenies for each gene, the authors provided a tentative scenario for the origin of the different isoforms during 2R.

Here, we expand those results by incorporating the protein sequences from a larger number of jawed vertebrates. We then construct revised molecular phylogenies, and use these to update the description of the expansion of these genes that occurred during 2R and the subsequent divergence of agnathan and jawed vertebrate lineages. Finally we speculate on how the proto-vertebrate phototransduction cascade might have evolved from a simpler cascade in an ancestral chordate organism.

MATERIALS AND METHODS

Transcriptome data. The methods for obtaining the eye transcriptomes from basal vertebrate species were described in Lamb et al (2016), and here we use transcripts from that work. Sequences were available for each of the following species obtained from Australian waters: *Eptatretus cirrhatus*, broad-gilled hagfish; *Geotria australis*, pouched lamprey; *Mordacia mordax*, short-headed lamprey; *Aptychotrema vincentiana*, western shovelnose ray; *Aptychotrema rostrata*, eastern shovelnose ray; *Neotrygon kuhlii* (*N. australiae*), blue spot maskray; *Chiloscyllium punctatum*, bamboo shark; and *Carcharhinus amblyrhynchos*, grey reef shark. Descriptions of each of these species are available from Bray & Gomon (2016,

http://fishesofaustralia.net.au). Sequences were also obtained from bowfin, *Amia calva*, and Florida gar, *Lepisosteus platyrhincus*. In addition to the sequences that we reported previously (that were assigned Genbank protein accession numbers in the range ANV21075 – ANV21149), we now report 39 new sequences, which have been submitted to Genbank and assigned protein accession numbers [to follow].

Sequence selection. For the increased number of jawed vertebrate sequences presented here, we tried to use as uniform a set of taxa as possible. We aimed to select: human, plus one additional placental mammal; two marsupials; three birds; three reptiles; two amphibians; coelacanth; bowfin and gar; two sharks, two rays, and elephant shark (a chimaera). For our two shovelnose ray species, the orthologous sequences were nearly identical when we had both, and in those cases, we used only the Western shovelnose ray sequence. Likewise, for the two species of gar, we used only the Florida gar sequence when we had nearly-identical orthologs. For agnathan vertebrates, we used every available sequence, except for those partial sequences that we deemed to be too short (e.g. less than about half the expected length). For several partial sequences from agnathan species, we noticed a deterioration of the alignment near the end of the sequence. In these cases, we removed the poorly-aligned terminal residues; these sequences are listed as '-Trimmed' in the Figures. For outgroups, we searched for closely similar sequences from tunicates (*Ciona intenstinalis* and *C. savigni*), lancelets (*Branchiostoma floridae* and *B. belcheri*), and from two other more basal deuterostomes (*Strongylocentrotus purpuratus*, an echinoderm, and *Saccoglossus kowalevskii*, a hemichordate), as well as from the fruit fly, *Drosophila melanogaster*.

Sequence alignment. We performed multiple sequence alignment of protein sequences using SATé-II (version 2.2.7, Liu et al, 2012) with the following default settings: aligner, MAFFT; merger, MUSCLE; tree estimator, FASTTREE, model WAG+G20; decomposition, centroid. The maximum sub-problem size was set to 12 for PDE6 and the opsins, and to 20 for GNAT/GNAI and the CNGCs.

Tree estimation. We constructed unconstrained maximum likelihood (ML) phylogenetic trees using IQ-Tree (Windows multicore version 1.5.2, Nguyen et al, 2015) with 10,000 bootstrap replicates, using the ultrafast bootstrap approximation (Minh et al, 2013). We tested each of the basic protein substitution models JTT, WAG, and LG; the resulting phylogenies were usually very similar. Where possible, we used the more recent LG model (Lee & Gascuel, 2008), but in a few cases, use of this model led to 'fragmentation' of a well-established jawed vertebrate clade (in our hands, for CNGA3 and Rh2). Therefore, for the CNGC and opsin phylogenies, we used the WAG model (Whelan & Goldman, 2001). We also tested rate heterogeneity additions (e.g. '+R4', etc.) but found that these often generated trees containing highly implausible clades, and so we restricted analysis to the basic substitution models. To improve the chances of finding the tree with the best log likelihood, we increased to 200 the number of iterations (-numstop) performed, after an improvement had been found, before stopping. All other parameters were set to their default values.

Tree estimation constrained by models of 2R and speciation. Based on the unconstrained ML tree obtained for each gene, we constructed a model (or models) of the duplications, losses, and divergence that could plausibly have been consistent with the observed phylogeny. Using such models for guidance, we then constructed constrained trees, using the '-g' constraint option in IQ-Tree. In specifying the constraints, we chose to use the minimum set of sequences that would constrain the tree as we intended. Typically, we used just a single sequence (human where possible) representative of the relevant isoform, and we relied on the tightness of clading to constrain the other orthologs in the same manner. Each constraint tree that we used is shown as an inset next to the constrained tree. For each constrained tree obtained, we conducted tree

topology tests using the '-z' option in IQ-Tree, in order to test whether or not the constrained tree needed to be rejected in comparison with the unconstrained ML tree. The tests applied were *bp*-RELL, *c*-ELW and *p*-AU, representing respectively: the Bootstrap Proportion test using the RELL method (Kishino et al, 1990), the Expected Likelihood Weight test (Strimmer & Rambaut, 2002), and the Approximately Unbiased test (Shimodaira, 2002). Only those trees that passed all tests at the 95% confidence level (i.e. $p \geq 0.05$) were considered further. (For the AU test, the value returned was the p-value, which we report, but for the other two tests the numbers had a different meaning, and for those tests we only report pass/fail at the 95% confidence level.) Figures in the main text are shown in collapsed format; fully expanded trees are presented in the Supplementary Information. Numbers at each node represent percentage bootstrap support.

RESULTS

For each of the four cascade components (GNAT, PDE6, CNGA and CNGB), we noticed the following general features of the ML phylogenetic trees we constructed when no constraints were applied. Firstly, there was very high internal bootstrap support for each jawed vertebrate clade (e.g. GNAT1, GNAT2, GNAT3), and secondly there was very high support for each major group of presumed 2R clades (e.g. the set of GNATs). However, the position of the root within that major group was not robust, and could vary according to the taxa selected or the parameters of tree construction. In fact, behaviour of this kind might be expected if the time between 1R and 2R had been relatively short, in which case an approximation to quadruplication (a multi-furcation) may have occurred. Because of this lack of robustness, it is necessary to exercise caution in interpreting the unconstrained ML trees. Our approach has been to consider the unconstrained phylogeny in relation to plausible models for the manner in which gene duplications and losses might account for the observations. We then constructed ML trees constrained in light of those models, and thereafter applied tree topology tests to determine whether or not each constrained tree needed to be rejected, or could be considered further. This approach permitted us to determine a scenario that appeared highly plausible for the evolution of the multiple isoforms. We first analyse the phylogenies for GNAT, PDE6 and the CNGCs, and thereafter we consider the visual opsins, where the interpretations are somewhat different.

GTP-binding protein alpha subunits, GNATs

Jawed vertebrates possess three *GNAT* genes, and analysis of synteny has shown that these arose during 2R (Nordström et al, 2004; Larhammar et al, 2009; Lagman et al, 2012). *GNAT1* (encoding rod transducin) is utilised in rod photoreceptors, *GNAT2* (cone transducin) in cone photoreceptors, and *GNAT3* (gustducin) in taste receptor cells. Each of these genes is located in close proximity to one of three *GNAI* genes (in human, as well as in several other species), consistent with the concept that *GNAI* and *GNAT* in jawed vertebrates arose by tandem duplication of an ancestral gene prior to 2R.

Before investigating the case when agnathan vertebrate sequences are included, it is instructive first to consider the simpler case of the phylogeny obtained for jawed vertebrate sequences, as presented in Fig. 1. In this unconstrained ML phylogeny, bootstrap support for the GNAT sub-tree is unanimous, and support for each of the six jawed vertebrate clades ranges from 92% to unanimous. When we examine the pattern of gene duplication and loss most likely to have given rise to this observed phylogeny, there is a crucial constraint we can apply, based on synteny. We make the assumption that, for both *GNAT* and *GNAI*, the extant isoforms arose through 2R, by two successive duplications. As only three of the four isoforms remain, one pair must have diverged during the second round, and therefore be sisters, whereas the third isoform

remains from 1R and is sister to the other pair. Next we note the syntenic pairings of *GNAT* and *GNAI* genes, as described by Larhammar et al (2009), comprising: *GNAT1–GNAI2*, *GNAT2–GNAI3*, and *GNAT3–GNAI1*. Because of these pairings, there are only three possible patterns of divergence that need to be examined for the six genes in Fig. 1. For example, if *GNAT1* diverged at 1R, then it should be sister to the other pair (*GNAT2*, *GNAT3*), but crucially *GNAI2* should likewise be sister to (*GNAI3*, *GNAI1*); and *mutatis mutandis* for the other two cases.

Figure 1 near here

In Fig. 1, it is noteworthy that one *GNAI* pair (*GNAI3*, *GNAI1*) is unanimously supported as sister. It is the syntenic relationship of *GNAT2* with *GNAI3* and *GNAT3* with *GNAI1* that determines the only reasonable branching pattern for the *GNAT* gene family. Thus, in descriptive terms, the two potential phylogenies that disrupt this pairing of *GNAI*s are highly unlikely. Instead, by far the most likely scenario is obtained simply by shifting the root within the *GNAT* sub-tree to the position indicated by the arrow in Fig. 1.

To quantify these considerations, we calculated ML trees constrained according to the three possible branching patterns, and then applied tests of topology to the resulting trees. For the first case, in which GNAT1 and GNAI2 remain from 1R, we only needed to constrain the GNAT sub-tree because the GNAI sub-tree was already as required; this constraint led to a very small change in log likelihood, of ΔLogL $\approx$ 1.3, and the tree readily passed the three tests of topology (with p-AU > 0.5). For the second case, in which GNAT2 and GNAI3 remain from 1R, one of the two constraints we applied was that GNAI3 be sister to (GNAI1, GNAI2). While the constrained tree did indeed meet this condition, it placed the three clades as a trifurcation, demonstrating what we referred to above in qualitative terms as the difficulty in separating GNAI1 and GNAI3. There was a large change in log likelihood, of ΔLogL $\approx$ 18.7, but the tree did not actually fail the tests of topology (with p-AU $\approx$ 0.16). The existence of the trifurcation meant that the tree did not conform to a sensible interpretation of the postulated branching, and when taken in conjunction with the large ΔLogL, we consider it very unlikely that this second scenario could underlie the observed data. For the third case, with GNAT3 and GNAI1 remaining from 1R, we only needed to constrain the GNAI sub-tree, because the GNAT sub-tree was already as required (and it remained so); the resulting tree exhibited an intermediate change in log likelihood, of ΔLogL $\approx$ 14.8, but it was rejected by all three tests of topology (with p-AU < 0.02). Therefore, we consider the only realistic possibility to be the first case, in which the survivors from 1R are GNAT1 and GNAI2, as indicated diagrammatically by the alternative root position marked by the arrow in Fig. 1.

With this information about the likely origin of the jawed vertebrate clades, we next applied a corresponding approach with the agnathan sequences included. The phylogeny that we obtained, subject to the constraint that GNAT1 is sister to GNAT2 and GNAT3, is presented in Fig. 2. The unconstrained tree (Supplementary Fig. S1) was very similar; the main difference was that the root within its GNAT sub-tree occurred at the position indicated by the arrow in Fig. 2. The constraint we applied is specified by the inset below the GNAT sub-tree in Fig. 2. It utilised only the three human GNAT sequences plus an outgroup sequence (see Methods) and, because of the tight support within each clade, this was all that was needed to constrain the other sequences as well. As in the case for the jawed vertebrate tree in Fig. 1, application of this constraint caused only a small change in log likelihood, of ΔLogL $\approx$ 2.2, and the tree readily passed all the tests of topology (with p-AU > 0.6). Thus, the tree in Fig. 2 not only conforms to the requirements of synteny, but it is also acceptable on the basis of the topology tests.

Figure 2 near here

The phylogenetic tree in Fig. 2 can be represented accurately by the set of duplications, losses, and divergences illustrated in Fig. 3. The only features not accounted for precisely in this schematic are the relative positions of the agnathan GNAI sequences in Fig. 2, and we suggest that this may be a result of the paucity of sequence data for these species, in conjunction with the relatively high level of conservation within each of the GNAI clades.

Figure 3 near here

As for the jawed vertebrate case described above, we also tested the other two possible scenarios for the genes that could have survived from 1R, and both were rejected by the tree topology tests. With the survivors from 1R being GNAT2 and GNAI3, the change in log likelihood was large, at ΔLogL $\approx$ 33.6, and the tree failed the *bp*-REL and *c*-ELW tests but passed the third test with *p*-AU $\approx$ 0.09. Similarly, with (GNAT3 + GNATX) and GNAI1 surviving from 1R, the change in log likelihood was even larger, at ΔLogL $\approx$ 36.1, and the tree again failed the *bp*-REL and *c*-ELW tests but passed the third test with a very similar *p*-AU $\approx$ 0.09. In both these cases, the constrained GNAI tree was a trifurcation, in which GNAI3 and GNAI1 were not actually separated (data not shown). Accordingly, we reject these two alternate scenarios, and conclude that the only plausible scenario is that shown in Fig. 3.

A consequence of this interpretation is that GNATX is not just sister to GNAT3, but is in fact the agnathan ortholog of jawed vertebrate GNAT3. Accordingly, in Fig. 2, we have labelled this clade as GNAT3/X. (However, in order to avoid confusion in the naming of sequences, we have retained the contig names with 'X' from Lamb et al, 2016, in the expanded tree of Supplementary Fig. S2.)

Phosphodiesterase, PDE6

Jawed vertebrates possess three *PDE6* genes encoding catalytic subunits, and analysis of synteny has again shown that these arose during 2R (Nordström et al, 2004; Larhammar et al, 2009; Lagman et al, 2016). *PDE6A* and *PDE6B* encode α and β subunits, that together form a heterodimeric catalytic unit in rod photoreceptors, whereas *PDE6C* encodes the α' subunit that forms a homodimeric catalytic unit in cone photoreceptors. The ancestral vertebrate *PDE6* gene is thought to have originated by duplication of a more ancient *PDE5/6/11* gene. No gene that clades with vertebrate *PDE6*s has been found outside of vertebrates.

Figure 4A presents the ML phylogenetic tree calculated under a constraint described below. In the unconstrained ML tree (presented in Supplementary Fig. S3), we had seen that PDE6X was positioned as sister to all the other PDE6 clades (the arrow in Fig. 4A marks the root of the unconstrained tree), suggesting that PDE6X had diverged at 1R. However, if that were the case, we could see no feasible scheme whereby a second duplication in the other branch could have formed the additional three clades, PDE6A, PDE6B and PDE6C. Therefore we tested the other possible positions for the root within the PDE6 sub-tree. When we applied the constraint indicated in the inset, we obtained the tree shown. The change in log likelihood was ΔLogL $\approx$ 3.8, and the tree passed all three topology tests (with *p*-AU > 0.3). For the other three possible positions of the root, the constrained trees failed all three tests of topology. With the root constrained to be adjacent to the PDE6C clade, the p-value was *p*-AU < 0.01, and when adjacent to either PDE6B or PDE6A the value was *p*-AU < 10^{-4}.

Figure 4 near here

Thus, the only two acceptable positions for the root are, firstly, as illustrated by the phylogeny in Fig. 4A, and, secondly, as indicated by the arrow. Of these two potential configurations, the only one for which we can envisage a plausible 2R scenario is the former. That scenario of duplications, losses and speciation is shown schematically in Fig. 4B. On this basis, we conclude that a common ancestor of (PDE6C, PDE6X) most likely separated from a common ancestor of (PDE6A, PDE6B) at 1R, and that the second round of genome duplications completed the quadruplication.

Cyclic nucleotide-gated channels, CNGCs

The cyclic nucleotide-gated ion channel of jawed vertebrate photoreceptors and olfactory receptor neurons is a heterotetramer comprising α and β subunits encoded by *CNGA* and *CNGB* genes, both sets of which appear to have expanded during 2R (Nordström et al, 2004). Rod channels comprise three α1 subunits and one β1 subunit, encoded respectively by *CNGA1* and *CNGB1* genes, whereas cone channels comprise two α3 and two β3 subunits, encoded by *CNGA3* and *CNGB3*. The channels of canonical olfactory receptor neurons comprise two α2, one α4, and one β1 subunit.

In calculating the phylogenetic tree for CNGCs, where we had a large number of sequences, we again began with only jawed vertebrates included. With the LG substitution model, we noticed that support for the CNGA3 clade was not very high (82%) and furthermore, when the agnathan sequences were subsequently included, the jawed vertebrate group no longer remained as a distinct clade. On the other hand, the WAG substitution model performed better, with support for the CNGA3 clade of 95% (with only jawed vertebrate sequences) and 85% (with all sequences included). Therefore we chose the WAG model for the CNGCs. The tree obtained when only jawed vertebrate sequences were used is presented as Supplementary Fig. S5.

With the agnathan sequences included, Fig. 5 presents the ML phylogenetic tree obtained when a constraint (described below) was placed on the position of the root within the group of α sub-units that expanded during 2R. The unconstrained tree is presented in Supplementary Fig. S6, and the fully expanded constrained tree in Supplementary Fig. S7. Clearly, the duplication that gave rise to the α and β branches was ancient, and presumably occurred prior to the protostome-deuterostome split. Previously, Kaupp & Seifert (2002) observed that *CNGA4* has three additional introns in its C-terminal region, and concluded that it falls into a different subfamily from the other genes of the jawed vertebrate α branch that is expressed only in olfactory receptor neurons. Additional support for this idea comes from inspection of the syntenic arrangement for the *CNGA* genes reported by Nordström et al (2004) in their Fig. 6b. And here, in Fig. 5 and Supplementary Figs S5–S7, the basal branching position of α4 lends further support to the notion that its divergence was ancient, quite possibly predating the protostome/deuterostome split. Subsequently, during 2R, the other α branch (here denoted αQ) and the β branch both expanded.

Figure 5 near here

For the β division, the analysis of synteny by Nordström et al (2004) in their Fig. 6c provided compelling evidence that jawed vertebrate clades B1 and B3 arose during 2R WGD. Here, the unconstrained tree (Supplementary Fig. S6) extends this result by showing that the isoforms we previously denoted as BY and BX, respectively, clade tightly with B1 and B3. Accordingly, in Fig. 5 they are coloured blue and red, and are renamed as CNGB1/Y and CNGB3/X. (As above, we have retained the former names for the contigs in Supplementary Fig. S7.)

One difference between these new trees (constrained and unconstrained) and our previous tree (Fig. 6 of Lamb et al, 2016) is that here we have detected an additional agnathan clade, CNGAY. Previously, we had overlooked one *G. australis* partial sequence, GEOAU_57616-1-2, which we find has 81% identity and 91% similarity to *M. mordax* partial sequence MORMR_41352-3-3, with no gaps over 328 residues; in the unconstrained tree (Supplementary Fig. S6) these two sequences were found to clade together with 82% bootstrap support.

For the CNAQ (αQ) division, the bootstrap support was unanimous (not only in the unconstrained case, but also under all constraints that we examined). However, the position of the root within this sub-tree appeared insecure, in light of the low support of 46% and 65% shown at the AX and AY nodes in the unconstrained case (Supplementary Fig. S6), and so we examined other possibilities. In evaluating the phylogeny, we need to take into consideration any restrictions imposed by the 2R duplication events assumed to have generated the set of sequences. To assist in these considerations, we present the scenario shown schematically in Fig. 6; this represents our favoured scheme, but also serves as a framework for exploring alternatives.

Figure 6 near here

The first restriction is that, for the jawed vertebrate clades, we require A1 to have diverged from (A2 + A3) at the first duplication (1R); this is required because there was unanimous bootstrap support for the A2 and A3 clades as sisters, not only in the tree restricted to jawed vertebrates (Supplementary Fig. S5) but also in the tree with agnathan taxa included. However, when we constrained either of the agnathan clades AX or AY to sit as sister to either A2 or A3, the resulting trees all exhibited ΔLogL > 27 and failed each of the three tree topology tests at the 95% confidence level (with the borderline exception that in one of the three tests one tree had a score of *p*-AU = 0.051); this is presumably the result of the strong support for A2 and A3 clades as sisters. With AX and AY both constrained to sit within the (A2 + A3) sub-tree, the resulting trees had ΔLogL > 33 and failed all tests. Accordingly, we rejected the possibility that either AX or AY, or both, could occupy positions sister to A2 or A3. On the assumption that the speciation event (A-J) occurred after the second round of duplications, we considered that it was inappropriate to try to constrain either of those agnathan clades to lie between the root and (A2 + A3).

Instead, we reached the conclusion that both AX and AY need to be placed on the branch containing A1. For the two possible arrangements of this kind, the constrained trees passed all the tests of topology (with *p*-AU > 0.3), but it is unclear which of the two possibilities is correct. In fact, the likelihood score for the arrangement illustrated in Figs 5 and 6 was marginally poorer (ΔLogL = 8.5 versus 7.7) than with AX and AY interchanged. However, we consider the illustrated arrangement to be more plausible because it places the AX clade, containing the 'Rh1-only' species *E. cirrhatus*, as sister to the A1 clade that is expressed in jawed vertebrate rods. This would be expected if the same isoform of α subunit were expressed in the scotopic photoreceptors of both agnathan and jawed vertebrates. However, this argument is not strong, and so we have put a question mark in Fig. 6 to indicate that these two clades might be interchanged.

For the β division, it seems clear that the two agnathan isoforms diverged from the corresponding jawed vertebrate isoforms only through speciation rather than duplication, and they are therefore orthologs (which we have denoted B1/Y and B3/X). The lower section of Fig. 6 represents the most parsimonious of the branching patterns that could have generated the β isoforms, and has just a single gene loss. We cannot rule out two other possibilities, in which B1 and B3 diverged at the first round, but these appear less parsimonious as they would have

involved the loss of either two gene after 2R or of four genes occurred after speciation. However, in each of these cases, the two agnathan β isoforms would still be orthologs of the two jawed vertebrate β isoforms.

For the αQ division, we conclude that the jawed vertebrate isoform A1 has an ortholog in agnathans, and we suggest that it is more likely that this ortholog is AX rather than AY. On the other hand, we conclude that vertebrate A3 does not have an agnathan ortholog.

Visual opsins

In 1992, Okano et al analysed the molecular phylogeny of the vertebrate visual opsins that were then available, and showed that rod opsin (Rh1) appeared to have evolved after the cone opsin families had already been established. The cone opsin families that they reported to have predated Rh1 were L, S, M_1 and M_2, which in today's terminology correspond to LWS, SWS1, SWS2 and Rh (=Rh1+Rh2), respectively. Subsequently, when knowledge of the chromosomal locations of these genes was taken into account (Nordström et al, 2004; Larhammar et al, 2009; Lagman et al, 2013), it was instead proposed that the four shorter-wavelength-sensitive opsins (SWS1, SWS2, Rh1 and Rh2) had arisen through 2R quadruplication of an ancestral SWS gene and that the corresponding expansion of an ancestral LWS gene had presumably been followed by loss of all but one copy.

Here, we re-examine the molecular phylogeny of vertebrate visual opsins, using a set of 116 sequences. Vertebrate opsins are particularly well represented in sequence databases, and all the sequences used here were obtained from NCBI. Because two of the visual opsin classes (SWS2 and Rh2) were lost during early mammalian evolution, we chose to exclude all mammalian sequences, in order to lessen the chances of introducing bias between the classes. Figure 7A presents the unconstrained ML phylogeny obtained using the WAG substitution model; it is shown in collapsed format, and the expanded tree is given in Supplementary Fig. S8. With the exception of *G. australis* RhB/Rh2, each of the other eight agnathan visual opsin sequences (including the four Rh1s) fell within a conventional visual opsin clade, each of which exhibited bootstrap support of at least 99%. The location shown for RhB/Rh2 in Fig. 7, as sister to (Rh1 + Rh2), represents one of the two potential positions reported by Pisani et al (2006). Figure 7B is an extract showing the support levels for these two clades when the tree was recalculated with *G. australis* RhB/Rh2 constrained as sister to Rh2. For the constrained tree, ΔLogL was just 2.2 and the tree passed all three tests of topology (with p-AU > 0.4).

Figure 7 near here

In considering the possible role that 2R whole genome duplication might have played in generating the four short-wave-sensitive visual opsins, an important question is whether or not the SWS1 and SWS2 clades could plausibly be sisters. This relationship is required on the assumption that the pairs (Rh1, Rh2) and (SWS1, SWS2) were generated by the second round of duplication. To examine this question, we re-calculated the tree with SWS1 and SWS2 constrained to be sisters (Fig. 7C, and Supplementary Fig. S9). The tree constrained in this way was rejected at the 95% confidence level by all three tests of topology (with p-AU < 0.03). Furthermore, the phylogenetic distance between the SWS1 and SWS2 clades (visualised roughly as sum of their branch lengths, and tabulated in Table 1 of Lamb et al, 2016) is much greater than the corresponding distance between Rh1 and Rh2. We interpret the combination of these two results to show clearly that SWS1 and SWS2 could not have diverged during 2R, but instead must have diverged considerably earlier. Therefore, we revert to the pattern of duplications originally proposed by Okano et al (1992), with the addition that the split between Rh1 and Rh2 occurred during 2R.

Once we assume that the LWS, SWS1, SWS2 and Rh classes already existed prior to 2R, we need to explain how the extant paralogon arrangement came about. First, it is self-evident that for each of the first three of these (LWS, SWS1, SWS2), only a single copy survived the two rounds of duplication, and the simplest explanation is that one copy of each of those genes was lost after 1R, and one copy was again lost after 2R. It is then a matter of allocating which combination of genes was lost from each daughter block (chromosome) after the two duplications. The factor that narrows down the potential set of losses to just a single option is the known synteny of opsins with GNATs and GNAIs. Thus, because of the syntenic arrangement Rh1–GNAT1–GNAI2, Rh2–GNAT2–GNAI3, and SWS1–GNAT3–GNAI1, the duplications and losses that in jawed vertebrates gave rise to the three GNATs and three GNAIs must correspondingly have given rise to Rh1, Rh2 and SWS1.

Figure 8 near here

Figure 8 shows the simplest combination of gene losses (following two block duplications) that we could find, that satisfied the dual requirements of (1) appropriate opsins on chromosomes, and (2) correct correspondence with the formation of GNAT and GNAI isoforms as shown in Fig. 2. Alternative variants of this scheme are possible, in which a given opsin gene was retained after 1R, and then three copies were lost after 2R, but these are less parsimonious. Although, the arrangement we propose here was predicated primarily on correspondence between the positions of Rh1, Rh2 and SWS1 and the positions of the GNATs and GNAIs, it also shows by elimination the pathway by which LWS and SWS2 adopted their current positions.

DISCUSSION

One important message from our phylogenetic analysis is that caution is needed in interpreting molecular phylogenies constructed for gene families that expanded during 2R. Thus, in the ML trees constructed conventionally (i.e. without any constraints) for each of GNAT (Fig. 1), PDE6 (Fig. 4A) and CNGA (Fig. 5), we think that the root of the 2R family was incorrectly assigned. Most likely, this indicates that the position of the root is poorly defined, as a result of the relatively small extent to which the post-2R isoforms had diverged from each other prior to the vertebrate radiation, in conjunction with the paucity of data from agnathan species. For each of these three gene families, the topological arrangement of clades *within* the 2R family in our final constrained tree was almost identical to that in the unconstrained tree. All that differed between the pairs of trees was the placement of the root (and an interchange of the positions of the agnathan CNGAX and CNGAY clades for our preferred arrangement in Fig. 5). Thus, in comparing the constrained and unconstrained trees, the root moved: by two nodes for GNAT (Fig. 2 versus Supplementary Fig. S1); by one node for PDE6 (Fig. 4A versus Supplementary Fig. S3); and by two nodes for CNGA (Fig. 5 versus Supplementary Fig. S6).

Since the root of any other 2R gene family (comprising more than two extant isoforms) may likewise be poorly specified, we suggest that it will be important to test the molecular phylogeny for any 2R family against plausible models of duplications, losses, and speciation, using suitable constraints on tree construction; this can readily be accomplished using the constraint option incorporated in recent releases of IQ-Tree (http://www.iqtree.org/). Under ideal circumstances, such constraints might involve multiple gene families from paralogous regions of chromosomes.

Expansion of the set of phototransduction protein isoforms during two rounds of whole-genome duplication

In our analysis of the molecular phylogenies of the phototransduction components in jawed and agnathan vertebrates, we have adopted the standard model that a common ancestor of agnathan and jawed vertebrates underwent two rounds (2R) of whole genome duplication. This accords with the conclusion by Smith et al (2013), in analysing the first assembly of the genome of *P. marinus*, "that genome-wide patterns of duplication are indicative of a shared history of two rounds of genome-wide duplication before lamprey-gnathostome divergence". Their results are also "consistent with the divergence of the two lineages shortly after the last whole-genome duplication event" (Smith et al, 2013); here, we have referred to that divergence using the shorthand 'A-J split'. It is further generally assumed that the interval between the first and second rounds of duplication was relatively short, and our finding of lack of robustness in the position of the root within each 2R sub-tree appears consistent with that assumption; however, there does not currently appear to be a great deal of hard evidence about the timing of that interval. On the other hand, it is clear that the A-J split occurred at least 500 My ago and, subsequently, within the stem jawed vertebrate lineage, there was at least a further 80 My of evolution before the radiation of jawed vertebrate species occurred, around 420 My ago (e.g. Erwin et al, 2011).

The scenarios that we developed in Figs 3, 4B and 6 for GNATs, PDE6s and CNGCs can be combined with the model of opsin duplications/losses in Fig. 8 to provide a comprehensive description of the expansion of isoforms underlying the activation steps of vertebrate phototransduction, and this is presented in Fig. 9. Here we have omitted agnathan vertebrates, in order to concentrate on the rods and cones of jawed vertebrates.

Figure 9 near here

In the opsin panel of Fig. 9 (top), the upward and downward branches at 1R denote the upper and lower blocks, respectively, in the middle panel of Fig. 8. Likewise, the upward and downward branches at 2R denote the upper and lower blocks, respectively, for the two pairs of blocks in the bottom panel of Fig. 8. Furthermore, for the GNATs, the upward and downward branches in Fig. 9 (second panel) correspond not only to the branches in Fig. 3, but also to the branches for Rh1, SWS2 and Rh2 in the opsin panel of Fig. 9 (top).

In our view, the hypothesised scenario in Fig. 9 provides a plausible description of all the sets of gene duplications and losses, relevant to the activation steps in phototransduction, that occurred during early vertebrate evolution, following the divergence of our ancestors from the ancestors of tunicates, and prior to their divergence from the ancestors of extant agnathan vertebrates. Although we have concentrated in Fig. 9 on jawed vertebrates, in order to avoid excessive complexity in the diagram, the scenario can readily be extended to agnathan vertebrates using the information in Figs 3, 4B and 6.

Isoforms utilised for phototransduction activation steps in agnathan cone-like and rod-like cells

For the proteins mediating the activation steps of vertebrate phototransduction (GNAT, PDE6, CNGA, CNGB), the isoforms utilised by the cones and rods of jawed vertebrates are well documented; next we investigate the complement of isoforms in the analogous cone-like and rod-like photoreceptors of agnathan vertebrates. Table 1 summarises the quantitative levels of transcripts measured in the eyes of our three agnathan species (from Supplementary Table 2 of Lamb et al, 2016); those measurements are now considered alongside the knowledge that these

three species have radically different complements of photoreceptors, ranging from cone-like only to rod-like only. However, we stress that these interpretations are made in the absence of other expression data, such as single-cell RT-PCR or *in situ* hybridisation, and therefore should be viewed with caution until they receive direct support.

Table 1 near here

The short-headed lamprey, *Mordacia mordax*, possesses only a single class of photoreceptor that exhibits a spectral peak at 526 nm, corresponding to the long-wave sensitive (LWS) opsin (Collin et al, 2004; N. Hart, personal communication). Furthermore, the visual opsin transcripts reported in *M. mordax* by Lamb et al (2016) were almost exclusively (99.7%) the *LWS* isoform. Accordingly, the photoreceptors of *M. mordax* are most likely analogous to the long-wavelength-sensitive cones of jawed vertebrates. Importantly, the second column of Table 1 shows that the eyes of this species contain transcripts for only one set of activation proteins: GNAT3/X, PDE6C, CNGAY and CNGB3/X.

In contrast to this, hagfish are known to possess a single class of photoreceptor (Locket & Jorgensen, 1998; Collin & Lamb, 2015) with a spectral peak at ~500 nm (Kobayashi, 1964; N Hart et al, personal communication), and the only visual opsin transcripts in the eyes of the broadgilled hagfish, *Eptatretus cirrhatus*, are for Rh1 (Lamb et al, 2016). The final column of Table 1 shows that in this species there is a reversal, in comparison with *M. mordax*, in the set of activation components expressed, with transcripts only for the proteins GNAT1, PDE6X, CNGA1/X and CNGB1/Y found in the eyes of *E. cirrhatus*.

The third species we studied, the pouched lamprey *Geotria australis*, is interesting for two reasons. First, it possesses five classes of visual opsin (Rh1, RhB/Rh2, SWS1, SWS2, LWS) which are individually expressed in five different morphological classes of photoreceptor (Collin et al, 2003a, 2003b; Warrington, 2016). Secondly, the two migratory stages of this species have very different compositions of their photoreceptor classes (Collin et al, 1999, 2003a; Davies et al, 2007) that correlate with their different lifestyles. The young, small, downstream migrants are active diurnally, under photopic conditions, and their photoreceptors express predominantly three of the cone-type opsins, LWS, SWS1 and SWS2, whereas the older, larger, upstream migrants are active nocturnally, under scotopic conditions, and the opsins expressed in their photoreceptors are predominantly Rh1 and RhB/Rh2 (Davies et al, 2007). The first of these was originally named RhA in *G. australis*, but we shall use the name Rh1 as we consider it to be the ortholog of the pigment present in rod photoreceptors of jawed vertebrates. The second was named RhB in *G. australis*, and we shall refer to it here as RhB/Rh2 because, although it has close similarity to jawed vertebrate Rh2, there is a possibility that it diverged prior to the divergence of agnathan and jawed vertebrates.

Table 1 shows that, while both sets of cascade activation components listed above are found in both migratory phases of *G. australis*, the set that we found in *M. mordax* dominates in the downstream (diurnally-active) migrants. Conversely, in the upstream (nocturnally-active) migrants, the dominant G-protein is GNAT1, as in hagfish, though the levels of the other three components are equivocal. From this analysis we conclude that in these agnathan species three of the cone-type opsins (SWS1, SWS2, LWS) couple to GNAT3/X, PDE6C, CNGAY and CNGB3/X, whereas the rod-type opsin (Rh1) couples to GNAT1, PDE6X, CNGA1/X and CNGB1/Y. This deduced arrangement is illustrated in Fig. 10, where the dashed vertical lines denote the phototransduction activation pathway for the two classes of agnathan photoreceptor cell. In addition, the corresponding pathways utilised in the classical cone and rod cascades of jawed vertebrates are denoted by the solid vertical lines.

For the scotopic activation pathway, the only difference between rods and the rod-like cells of agnathans involves the isoform of PDE6 utilised. Jawed vertebrate rods use the heterodimeric PDE6A/PDE6B, whereas agnathan rod-like cells use the homodimeric PDE6X. We speculate that the evolution of the heterodimeric PDE6 in the jawed vertebrate lineage may have been have been highly significant, because it has recently been established that the asymmetric nature of the PDE6A/PDE6B provides greatly increased immunity against noise resulting from spontaneous PDE activation elicited by baseline transducin activity (Qureshi et al, 2015). As a result, the 'continuous' electrical noise in darkness would be expected to be lower in rods, that express the heterodimeric PDE6A/PDE6B, than in agnathan rod-like cells, that express the homodimeric PDE6X (or in cones or cone-like cells, that express the homodimeric PDE6C).

We have not attempted to depict agnathan RhB/Rh2 in Fig. 10, because there is insufficient evidence upon which to base an assignment of the components to which it couples. However, it is interesting that, for *G. australis* in Table 1, the sum of transcript levels for *Rh1 + RhB/Rh2* yields approximately the expected ratio to *GNAT1* levels as would be anticipated if both Rh1 and RhB/Rh2 linked to GNAT1. Thus, we suggest that, in lamprey photoreceptors, RhB/Rh2 may couple to the rod-like cascade. This contrasts with the case for jawed vertebrates, where Rh2 is found in cones and is therefore presumed to couple to the cone cascade. For example, in the nocturnal gecko, Rh1 opsin and the rod phototransduction cascade isoforms are absent, and the photoreceptors express only cone opsins (including Rh2) and cone transduction components (Zhang et al, 2006). In relation to RhB/Rh2, two additional points may be relevant. First, the only agnathan species in which this isoform has been detected is *G. australis* (Collin et al, 2003b) and, secondly, the phylogenetic position of RhB/Rh2 remains enigmatic (e.g. Fig. 7, and Pisani et al, 2006).

The evolution of distinct classes of vertebrate photoreceptor

As we have discussed above, the five main classes of vertebrate visual opsin are (with just a few exceptions) expressed individually in distinct morphological classes of photoreceptor cell, not only in jawed vertebrates but also in the lamprey *G. australis*. According to our interpretation of the visual opsin phylogeny (Fig. 7), and in accord with Okano et al (1992), four visual opsins already existed prior to 1R. In jawed vertebrates, three of these (LWS, SWS1, SWS2) are cone opsins and are expressed in distinct classes of cone; likewise, in the lamprey, *G. australis*, the three orthologs are expressed in separate classes of cone-like cell. Therefore it seems highly likely that at least three classes of cone-like cell already existed in the proto-vertebrate ancestor, prior to 1R. It is also highly likely that there was a fourth class of photoreceptor cell, again with cone-like morphology, that expressed the forerunner of Rh1/Rh2.

From our scenarios for the expansion of transduction cascade component isoforms in Fig 3, 4A and 6 (summarised for jawed vertebrates in Fig. 9), there is no suggestion of any scotopic *versus* photopic specialisation prior to 1R. Hence, the activation cascade in all four classes of cell would have been comparable, with the implication that the transduction cascade in the cell type expressing the Rh1/Rh2 precursor would also have been cone-like, without any rod-like specialisation. A major question to be answered, then, is how rapidly specialisation for the scotopic intensity range arose in a descendant cell, after the 1R duplication.

For the G-protein transducin, inspection of the branch lengths in Fig. 2 suggests that the ancestral GNAT underwent extensive evolution (~0.25 residue substitutions per position) after its split from GNAI but prior to 1R. On the other hand, there appears to have been relatively

minor change in GNAT1 after 1R but prior to the divergence of agnathans (as judged by the shortness of the blue branch below the '98' in Fig. 2).

For the PDE6 (Fig. 4A), there was an even greater change (of ~0.5 residue substitutions per position), from the time it split from a tunicate PDE until 1R. Then, in the precursor of the PDE6A and PDE6B isoforms that are expressed in the rods of jawed vertebrates, there was some further change (of ~0.1 substitutions per position), that according to the scheme in Fig. 4B would have occurred between 1R and 2R. But in the other branch, that includes the cone PDE6C and the agnathan scotopic PDE6X isoform, rather less change occurred between 1R and 2R. On this basis, we deduce that the agnathan PDE6X underwent relatively little change prior to the A-J split.

Comparable arguments can be made for the CNGC α subunit (Fig. 5). But for the β subunit, we have insufficient information, because there was only a single gene duplication (at either 1R or 2R). For the opsins, examination of the short branch length at the base of the Rh1 clade (~0.07 substitutions; Supplementary Fig. S8), suggests that relatively little change occurred in Rh1 prior to the A-J split.

When these different considerations are combined, we can conclude that, for each of the scotopic isoforms Rh1, GNAT1, PDE6A/B/X and CNGA1/X, only a relatively small degree of specialisation appears to have occurred between its genesis and the time of divergence of the agnathan lineage. As a result, we suggest that serious consideration must be given to the possibility that, at the time of the A-J split, the precursor of the rod cell and the precursors of cone cells differed to only a minor extent. If this were the case, then the subsequent specialisation to generate distinct rod- and cone-specific transduction pathways in jawed vertebrates, and corresponding rod-like and cone-like cascades in agnathans, may have proceeded substantially independently in the two lineages, from a common toolkit and under similar evolutionary pressures.

Origin of the proto-vertebrate phototransduction cascade

Finally, and somewhat more speculatively, we now ask how the proto-vertebrate phototransduction cascade could have arisen, prior to its expansion (addressed above) during 2R. Thus, we wish to investigate the transition from an ancient chordate cascade into the proto-vertebrate photoreceptor cascade that utilised GNAT and PDE6. First we describe our proposal for the nature of the ancestral chordate G_I-coupled phototransduction cascade, comparable to the cascade that is presumed to exist in extant *Branchiostoma* and *Ciona*. Then we propose a series of modifications to that cascade, each of which would have been beneficial to the organism, and which together could have culminated in the proto-vertebrate transducin-coupled cascade.

Ancestral chordate G_I-coupled phototransduction cascade

In order to formulate a hypothesis for the nature of the ancestral chordate phototransduction cascade, in the last common ancestor of cephalochordates and vertebrates, we draw upon the limited evidence that exists on transduction in the ciliary photoreceptors of extant lancelets and tunicates. Electrical recordings from the ciliary photoreceptors of larval tunicates have, to our knowledge, been reported in only a single study (Gorman et al, 1971), where they were described as being 'exceedingly difficult to record'. In the cells that were recorded, the light response was hyperpolarising, and was accompanied by a decrease in membrane conductance (channel closure), just as occurs for vertebrate photoreceptors. In *Branchiostoma*, the presumptive photoreceptor cells of the frontal eye express the inhibitory G-protein, GNAI, along with a ciliary opsin (Vopalensky et al, 2012). Furthermore, as discussed earlier, it is clear that

vertebrate transducin (GNAT) arose through duplication of an ancestral *GNAI* gene, and so it is natural to expect that the ancestral chordate phototransduction cascade would have utilised GNAI. Downstream, the obvious target for GNAI would be adenylyl cyclase (AC), which synthesises cAMP. The light-induced inhibition of AC by GNAI would be predicted to lower the cAMP concentration, and thereby to close cAMP-gated ion channels.

Based on these ideas, our proposal for the ancestral chordate phototransduction cascade is illustrated in Fig. 11A, which has close parallels with the transduction cascade in vertebrate olfactory receptor neurons (ORNs). We propose that this cascade utilised the cytoplasmic messenger cAMP, acting on cAMP-gated ion channels that comprised α and β subunits. cAMP was synthesised by adenylyl cyclase (AC) and hydrolysed by a phosphodiesterase (PDE). The most obvious candidate for the PDE in extant *Branchiostoma* would be the one illustrated in the outgroup of Fig. 5, namely XP_002608983, which bears a close resemblance to mammalian PDE5/11. PDE11 is a dual cAMP/cGMP phosphodiesterase, and we postulate that the ancestral chordate G_I-coupled phototransduction cascade employed a dual-substrate PDE of this kind. For the cyclase, we note that the genome of *Branchiostoma* includes three members of the class III adenylyl cyclases that resemble mammalian members ADCY3, ADCY5/6, and ADCY8, respectively. The second of these (XP_002586828) would be the obvious candidate for the AC in *Branchiostoma* phototransduction, because ADCY5 and ADCY6 are clearly inhibited by GNAIs, whereas ADCY3 and ADCY8 are not. Conventionally, it has been thought that G_I exerts its actions on AC primarily by inhibition of receptor-mediated stimulation (e.g. by G_S), but recently it has been shown that G_I is in fact able to inhibit the basal activity of ADCY5 independent of receptor activation (Melsom et al, 2014), as would be required in this system. Finally, we note that ADCY5 and ADCY6 are negatively regulated by micromolar concentrations of Ca^{2+}, and that a comparable effect in the ancestral AC would have provided a modicum of negative feedback regulation in the case of excessive channel opening.

Figure 11 near here

Accordingly, we propose that the light response in the ancestral chordate phototransduction cascade would have functioned as follows (Fig. 11A, black). Activated visual opsin (R*) catalysed the activation of an inhibitory G-protein (GNAI), to create $G\alpha_I$*. The activated $G\alpha_I$* bound to and inhibited an adenylyl cyclase (AC) that resembled ADCY5/6. Inhibition of this AC, in conjunction with continued activity of the PDE (that resembled PDE11), led to a decline in cytoplasmic cAMP concentration, thereby closing cAMP-gated ion channels.

The cascade proposed in Fig. 11A is almost identical to that utilised in the canonical transduction cascade of vertebrate olfactory receptor neurons (ORNs). Apart from the receptor protein, the main difference is that the illustrated cascade uses an inhibitory G_I rather than an excitatory G_S, to lower rather than to raise the level of cAMP, so that it would correspond to the cascade of a hypothetical inhibitory ORN. On this basis, we have suggested that the ancestral CNGCs may have comprised α_Q, α_4, and β subunits, as for the channels of canonical ORNs.

Transition to the proto-vertebrate phototransduction cascade

We hypothesise that the transition from the proposed ancestral chordate cascade to the proto-vertebrate cascade occurred via the following sequence of steps, where we can identify a positive advantage to the organism in each transition.

1. Advent of a soluble factor that inhibited the PDE. We propose that the first step in the transition from a G_I-mediated cascade to a G_T-mediated cascade was the advent of a soluble protein that provided a degree of inhibition of the phosphodiesterase (Fig. 11A, red). Zhang &

Artemyev (2010) concluded that the capacity of the ancestral PDE to bind the inhibitory γ subunit predated the emergence of the inhibitory subunit itself. Thus, by reconstructing the sequence of the ancestral PDE5/6/11 enzyme, they found that it already contained the signature motif 'Ile-Pro-Met' (IPM) at which binding of the C-terminus of modern PDEγ occurs. We suggest that a short protein, similar to the C-terminus of the modern PDEγ, appeared, and provided some degree of inhibition of hydrolytic activity. How might this have been advantageous? One possibility is that it may have permitted the photoreceptor to function at higher ambient intensities, when the AC was substantially inhibited by GNAI; thus, inhibition of the PDE by this soluble factor could have partly counteracted the saturation brought about by light-induced inhibition of the cyclase, thereby extending the organism's operating range to somewhat higher intensities. It might even have been the case that the expression level of this soluble protein was modulated diurnally, in just the way that occurs for PDE6H in extant zebrafish (Lagman et al, 2016), so as to adjust the cell's performance to day and night intensity levels. Modulation of expression level over a diurnal timescale is likely to have been easier to accomplish for a small soluble molecule than for the catalytic units of PDE, which form a large dimeric membrane protein.

2. Development of an interaction between the G-protein and the soluble inhibitory factor. Next, we propose that mutations led to a propensity for the G-protein to interact with (i.e. bind to) this soluble inhibitory factor. At this stage of evolution (Fig. 11B, red), we have denoted the G-protein as G_{IT}, to indicate a transitional form between G_I and G_T. The consequence of binding of G_{IT} to the soluble factor is likely to have been a loosening of the latter's binding to the PDE, and hence a partial relief of its inhibition, thus causing a light-activated increase in the hydrolysis of cAMP. In other words, this new interaction would have reinforced the effect of the ancestral mechanism (cyclase inhibition), and thereby have been beneficial to the organism. Indeed, it is possible that this new pathway could rapidly have provided much improved light sensitivity. If, for example, the K_D for inhibition by G_I^* of basal AC activity had been relatively weak, whereas the K_D for binding of $G\alpha_{IT}^*$ to the new PDEγ had been tighter, then the new pathway could have had higher efficacy. In this case, it would not be surprising if further mutations reduced the effectiveness of the G-protein in inhibiting the cyclase (Fig. 11B, blue). On the other hand, it is interesting to note that many extant transducin $G\alpha_T$ sequences, including those in human, mouse, chicken, *Xenopus*, anole and zebrafish, still retain a conserved motif (I184, K209, I212, H213 and E216) that is reported to be an AC interaction site (see NCBI Conserved Features for G-alpha, at https://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=cd00066).

3. Transition from cAMP to cGMP as the cytoplasmic messenger. Once the AC no longer contributed directly to the light response, we propose that a guanylyl cyclase (GC) was expressed in the outer segment (Fig. 11C, red); additionally (not shown) we presume that its regulatory GCAP (Guanylyl Cyclase Activating Protein) would also have been expressed. On our previous assumption that the PDE exhibited dual substrate activity, the cGMP level would have been modulated by light in a similar way to the modulation of cAMP. The major advantage to the organism of this change would have been the powerful Ca^{2+}-mediated negative feedback loop engendered by GCAP modulation of GC activity. We further suggest that, with cGMP as the messenger, the channels would no longer have needed the α4 subunit, which may no longer have been expressed, so that the channels would then simply have comprised $α_Q$ and β subunits, as is the case for the cGMP-gated ion channels of vertebrate photoreceptors. Once cGMP had 'taken over' the role previously mediated by cAMP, it is likely that AC-mediated formation of cAMP would no longer have contributed directly to modulation of the photoreceptor current (Fig. 11C, blue). Nevertheless, there is clear evidence an AC continues to be expressed in extant vertebrate photoreceptors. Type 1 adenylyl cyclase (AC1), which is closely related to AC5 and AC6, has been shown to be expressed in rods (Jackson et al, 2009),

and treatment with forskolin, which activates AC, raises the cytoplasmic concentration of cAMP in rod photoreceptors, and (by unknown mechanisms) causes the cells to appear more dark-adapted (Astakhova et al, 2012). Therefore, it is possible that the role of cAMP in phototransduction evolved from an ancestral function of directly controlling ion channel opening to a modern function involving regulation of the cells' adaptational state.

Although we stress that the scenario outlined above is entirely hypothetical, we put it forward as a plausible route by which the proto-vertebrate phototransduction cascade could have evolved. Any alternative proposal for such a transition needs to begin from a plausible ancestral chordate cascade, and end with a cascade closely resembling that in extant vertebrate photoreceptors.

Future directions

Our phylogenetic analysis and our hypothesised scenarios lead us to propose several avenues for future research into the evolution of vertebrate phototransduction, which we now outline in dot-point format:

- Apply analysis of synteny to the phototransduction genes of agnathan species. Currently this is not feasible. The draft genome of *Petromyzon marinus* is not sufficiently complete, and furthermore *P. marinus* is not ideal for this purpose because, with only two classes of retinal photoreceptor, it has lost many of the isoforms of interest. The genome of *G. australis* would be especially valuable.
- Obtain the sequences for the phototransduction activation components from additional species of agnathan vertebrate, especially from any species of lamprey that expresses multiple cone-like opsins. Such an extension would permit improved characterisation of the agnathan clades, which at present contain very few taxa.
- Identify the components of the phototransduction cascade in the ciliary photoreceptors of *Branchiostoma* and *Ciona*, so as to better understand the nature of the ancestral chordate cascade. In particular, it would be important to test whether those cascades rely on a reduction in cAMP concentration mediated by adenylyl cyclase.
- Apply comparable phylogenetic analysis to the proteins underlying response shut-off and Ca^{2+}-mediated light adaptation, in order to understand more completely the evolution of vertebrate phototransduction.
- Apply recently available single-cell transcriptome analysis to retinal cells, to confirm the linkage of opsin class to downstream phototransduction cascade components, in *G. australis* and other species of interest.

Conclusions

We conclude that the scenarios that we present in Figs 10 and 11 together provide a credible description for the origin of the vertebrate phototransduction cascade, and thereby afford a framework for future research. Thus, Fig. 11 presents a plausible account for the transition from an ancestral chordate phototransduction cascade, utilising an inhibitory G-protein acting on adenylyl cyclase, into a proto-vertebrate cascade utilising transducin to activate PDE6. And Fig. 10 provides a parsimonious account for the origin in 2R of the duplex system of cone and rod isoforms in jawed vertebrates, while the earlier Figs 3, 4B, and 6 provide the additional information needed to explain the broadly parallel system in agnathan vertebrates. Prior to the A-J divergence, the each of the activation proteins appears to have undergone fairly minor changes in the scotopic branch relative to the photopic branch. This suggests that little of the specialisation that now distinguishes rod and cone transduction was present when agnathans diverged, and that instead much of the 'duplex' rod/cone specialisation evolved subsequently.

ACKNOWLEDGEMENT

REFERENCES

Astakhova LA, Samoiliuk EV, Govardovskii VI, Firsov ML. 2012. cAMP controls rod photoreceptor sensitivity via multiple targets in the phototransduction cascade. *J. Gen. Physiol.* **140**:421-433.

Bray DJ & Gomon MF. 2016. Fishes of Australia. http://fishesofaustralia.net.au

Collin SP, Hart NS, Shand J, Potter IC. 2003*a*. Morphology and spectral absorption characteristics of retinal photoreceptors in the southern hemisphere lamprey (*Geotria australis*). *Visual Neurosci.* **20**:119-30.

Collin SP, Hart NS, Wallace KM, Shand J, Potter IC. 2004. Vision in the southern hemisphere lamprey *Mordacia mordax*: Spatial distribution, spectral absorption characteristics, and optical sensitivity of a single class of retinal photoreceptor. *Visual Neurosci.* **21**:765-773.

Collin SP, Knight MA, Davies WL, Potter IC, Hunt DM, Trezise AEO. 2003*b*. Ancient colour vision: multiple opsin genes in the ancestral vertebrates. *Curr. Biol.* **13**:R864-R865.

Collin SP, Lamb TD. 2015. Photoreception in hagfishes: insights into the evolution of vision. In *Hagfish Biology* (Eds Edwards SL & Goss GG). Chap. 5, pp 129-148. Boca Raton, CRC Press.

Collin SP, Potter IC, Braekevelt CR. 1999. The ocular morphology of the southern hemisphere lamprey *Geotria australis* Gray, with special reference to optical specializations and the characterization and phylogeny of photoreceptor types. *Brain Behav. Evol*. **54**:96-118.

Davies WL, Cowing JA, Carvalho LS, Potter IC, Trezise AEO, Hunt DM, Collin SP. 2007. Functional characterization, tuning, and regulation of visual pigment gene expression in an anadromous lamprey. *FASEB Journal* **21**:2713-2724.

Erwin DH, Laflamme M, Tweedt SM, Sperling EA, Pisani D, Peterson KJ. 2011. The Cambrian conundrum: early divergence and later ecological success in the early history of animals. *Science* **334**:1091-1097.

Gorman ALF, McReynolds JS & Barnes SN. 1971. Photoreceptors in primitive chordates: Fine structure, hyperpolarizing receptor potentials, and evolution. *Science* **172**:1052-1054. doi:10.1126/science.172.3987.1052

Jackson CR, Chaurasia SS, Zhou H, Haque R, Storm DR, Iuvone PM. 2009. Essential roles of dopamine D4 receptors and the type 1 adenylyl cyclase in photic control of cyclic AMP in photoreceptor cells. *J. Neurochem.* **109**:148-157.

Kaupp UB, Seifert R. 2002. Cyclic nucleotide-gated ion channels. *Physiol Reviews* **82**:769-824.

Kishino H, Miyata T, Hasegawa M. 1990. Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. *J. Mol. Evol.* **31**:151-160.

Kobayashi H. 1964. On the photo-perceptive function in the eye of the hagfish, *Myxine garmani* Jordan et Snyder. *Journal of National Fisheries University* **13**:67-83. ISSN 0370-9361.

Kuraku S, Meyer A, Kuratani S. 2009. Timing of genome duplications relative to the origin of the vertebrates: did cyclostomes diverge before or after? *Mol. Biol. Evol.* **26**:47-59. doi:10.1093/molbev/msn222

Lagman D, Franzén IE, Eggert J, Larhammar D, Abalo XM. 2016. Evolution and expression of the phosphodiesterase 6 genes unveils vertebrate novelty to control photosensitivity. *BMC Evol Biol.* **16**:124.

Lagman D, Ocampo Daza D, Widmark J, Abalo XM, Sundström G, Larhammar D. 2013. The vertebrate ancestral repertoire of visual opsins, transducin alpha subunits and oxytocin/vasopressin receptors was established by duplication of their shared genomic region in the two rounds of early vertebrate genome duplications. *BMC Evol Biol.* **13**:238.

Lagman D, Sundström G, Ocampo Daza D, Abalo XM, Larhammar D. 2012. Expansion of transducin subunit gene families in early vertebrate tetraploidizations. *Genomics* **100**:203-211.

Lamb TD, Patel H, Chuah A, Natoli RC, Davies WIL, Hart NS, Collin SP & Hunt DM. 2016. Evolution of vertebrate phototransduction: cascade activation. *Mol. Biol. Evol.* **33**, 2064-87. doi:10.1093/molbev/msw095

Larhammar D, Nordström K, Larsson TA. 2009. Evolution of vertebrate rod and cone phototransduction genes. *Phil Trans Roy Soc.* B **364**:2867-2880.

Le SQ, Gascuel O (2008). An improved general amino acid replacement matrix. *Mol. Biol. Evol.* **25**:1307–1320.

Liu K, Warnow TJ, Holder MT, Nelesen S, Yu J, Stamatakis A and Linder CR. SATé-II: Very fast and accurate simultaneous estimation of multiple sequence alignments and phylogenetic trees. *System. Biol.* **61**:90-106.

Locket NA, Jorgensen JM. 1998. The eyes of hagfishes. In *The Biology of Hagfishes* (Eds Jorgensen JM, Lomholt JP, Weber RE & Malte H). Chap. 34, pp 541-546. London, Chapman and Hall.

Melsom CB, Ørstavik Ø, Osnes J-B, Skomedal T, Levy FO, Krobert KA. 2014. G_i proteins regulate adenylyl cyclase activity independent of receptor activation. *PLoS ONE* **9**:e106608. doi:10.1371/journal.pone.0106608

Minh BQ, Nguyen MAT and von Haeseler A. 2013. Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* **30**:1188-1195. doi:10.1093/molbev/mst024

Muradov H, Boyd KK, Kerov V, Artemyev NO. 2007. PDE6 in lamprey *Petromyzon marinus*: Implications for the evolution of the visual effector in vertebrates. *Biochemistry* **46**:9992-10000.

Muradov H, Kerov V, Boyd KK, Artemyev NO. 2008. Unique transducins expressed in long and short photoreceptors of lamprey *Petromyzon marinus*. *Vision Res.* **48**:2302-2308.

Nguyen L-T, Schmidt HA, von Haeseler A and Minh BQ. 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Mol. Biol. Evol.* **32**:268-274. doi:10.1093/molbev/msu300

Nordström K, Larsson TA, Larhammar D. 2004. Extensive duplications of phototransduction genes in early vertebrate evolution correlate with block (chromosome) duplications. *Genomics* **83**:852-72.

Ohno S. 1970. *Evolution by Gene Duplication*. London: Allen and Unwin, ISBN 0-04-575015-7.

Okano T, Kojima D, Fukada Y, Shichida Y, Yoshizawa T. 1992. Primary structures of chicken cone visual pigments: Vertebrate rhodopsins have evolved out of cone visual pigments. *Proc. Natl Acad. Sci. USA* **89**, 5932-5936.

Pisani D, Mohun SM, Harris SR, McInerney JO, Wilkinson M. 2006. Molecular evidence for dim-light vision in the last common ancestor of the vertebrates. *Curr Biol.* **16**:R318-R319.

Qureshi BM, Behrmann E, Schöneberg J, Loerke J, Bürger J, Mielke T, Giesebrecht J, Noé F, Hofmann KP, Spahn CMT, Heck M. 2015. Asymmetric properties of rod cGMP Phosphodiesterase 6 (PDE6): structural and functional analysis. *BMC Pharmacol Toxicol*. **16** (Suppl 1):A76.

Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. *Syst. Biol.* **51**:492-508.

Smith JJ, Keinath MC. 2015. The sea lamprey meiotic map improves resolution of ancient vertebrate genome duplications. *Genome Res*. **25**:1081-1090.

Smith JJ, Kuraku S, Holt C, Sauka-Spengler T, Jiang N, Campbell MS, Yandell MD, Manousaki T, Meyer A, Bloom OE, et al. 2013. Sequencing of the sea lamprey (*Petromyzon marinus*) genome provides insights into vertebrate evolution. *Nat Genet* **45**:415-421.

Strimmer K, Rambaut A. 2002. Inferring confidence sets of possibly misspecified gene trees. *Proc. R. Soc. Lond. B* **269**:137-142.

Vopalensky P, Pergnera J, Liegertova M, Benito-Gutierrez E, Arendt D, Kozmik Z. 2012. Molecular analysis of the amphioxus frontal eye unravels the evolutionary origin of the retina and pigment cells of the vertebrate eye. *Proc Natl Acad Sci USA* **109**:15383-15388.

Warrington R. 2016. Retinal photoreception in southern hemisphere lampreys. PhD Thesis, University of Western Australia. http://research-repository.uwa.edu.au/files/9252561/Warrington_Rachael_2016.pdf

Whelan S, Goldman N. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol.* **18**:691-699.

Zhang X, Wensel TG, Yuan C. 2006. Tokay gecko photoreceptors achieve rod-like physiology with cone-like proteins. *Photochem Photobiol.* **82**:1452–1460.

Zhang Z, Artemyev NO. 2010. Determinants for phosphodiesterase 6 inhibition by its γ-subunit. *Biochemistry* **49**:3862-3867.

Table 1. Transcript levels of phototransduction activation components in agnathans

RPKM-CDS[a] for:	*Mordacia mordax*	*Geotria australis* Downstream	Upstream	*Eptatretus cirrhatus*
SWS1+SWS2+LWS[b]	1890.6	2988.7	663.2	–
GNAT3/X	944.3	496.3	175.3	–
PDE6C	56.9	110.2	47.8	–
CNGAY	50.6	66.5	31.0	–
CNGB3/X	18.8	50.2	37.1	–
RhB/Rh2[c]	–	730.7	571.6	–
Rh1	(5.1)[d]	209.8	485.7	976.1
GNAT1	–	139.0	367.4	212.1
PDE6X	–	15.0	15.8	27.3
CNGA1/X	–	11.8	26.4	142.4
CNGB1/Y	–	6.7	11.3	7.9

Notes

a. Transcript levels are given in units of reads per kilobase per million mapped reads, calculated over only the coding region of the transcript, and referred to here as RPKM-CDS.

b. For the visual opsins, the transcript levels for the three cone-type isoforms (*SWS1*, *SWS2* and *LWS*) have been summed; of these isoforms, only *LWS* was present in *M. mordax*.

c. The lamprey opsin *RhB/Rh2* has been listed separately, as there is a possibility that it may link to a rod-type transduction cascade (see Text).

d. The *Rh1* transcript for *M. mordax* was only about half the expected length, and was present at only trace level. Hence it may represent a pseudogene, and it is therefore shown in brackets.

Figure 1. Molecular phylogeny for jawed vertebrate G-protein α subunits (GNAI and GNAT)

Unconstrained ML phylogenetic tree for GNAI and GNAT protein sequences from jawed vertebrates only (i.e. excluding agnathan vertebrate sequences). The sub-trees for each clade have been condensed. The arrow shows the alternative position for the root of the GNAT sub-tree analysed in the text. SATé alignment: maximum sub-problem size, 20; model WAG+G20. IQ-Tree construction: protein substitution model, LG.

Figure 2. Molecular phylogeny for vertebrate G-protein α subunits (GNAI and GNAT)

ML phylogenetic tree for GNAI and GNAT protein sequences, subject to the constraints that GNAT1 is sister to (GNAT2, GNAT3) and that GNAI2 is sister to (GNAI1, GNAI3); see Text. The unconstrained tree is presented as Supplementary Fig. S1, and the fully-expanded constrained tree is presented as Supplementary Fig. S2. Here, and in subsequent Figures, blue and red denote jawed vertebrate rod- and cone-specific isoforms, respectively; green denotes non-rod/non-cone vertebrate isoforms, and magenta is used for agnathan photoreceptor isoforms that cannot be identified as orthologous to a vertebrate isoform. SATé alignment: maximum sub-problem size, 20; model WAG+G20. IQ-Tree construction: protein substitution model, LG.

Figure 3. Scenario for the duplications, losses, and divergences underlying the expansion of G-protein α subunits (GNAI and GNAT) in jawed and agnathan vertebrates.

Dashed vertical lines denote hypothesised timing of genome duplications and speciation: 1R (first) and 2R (second) rounds of whole genome duplication; A-J, divergence of the ancestors of extant agnathan vertebrates (A) and jawed vertebrates (J). Branch marked with open square (□) denotes tandem gene duplication; branches marked with open circles (O) indicate genome duplication; unmarked branches denote speciation; incomplete lines indicate gene loss.

Figure 4. Molecular phylogeny and hypothesised origin for phosphodiesterase catalytic subunits (PDE6).

A. ML phylogenetic tree for vertebrate PDE6, calculated subject to the constraint that the combined members of the PDE6X and PDE6C groups are sister to the combined members of the PDE6A and PDE6B groups. The arrow shows the position of the root in the unconstrained tree (see Supplementary Fig. S3). The constraint tree applied is shown as the inset at the top left, and the fully expanded constrained tree is given as Supplementary Fig. S4. The limb leading to PDE6 is quite long, indicating that substantial changes occurred during the transition from the ancestral phosphodiesterase to vertebrate PDE6. SATé alignment: maximum sub-problem size, 12; model WAG+G20. IQ-Tree construction: protein substitution model, LG.
B. Scenario for the duplications, losses, and divergences underlying the expansion of PDE6 catalytic subunits in jawed and agnathan vertebrates. Conventions are as for Fig. 3.

Figure 5. Constrained molecular phylogeny for cyclic nucleotide-gated channel α and β subunits (CNGA and CNGB).

Molecular phylogeny of CNGA and CNGB proteins, subject to a constraint on the position of the root within the CNGAQ group specified by the constraint tree in the inset at the top left; see

Text. The CNGA sub-tree has an additional clade, CNGAY, that includes an agnathan sequence that we previously overlooked (see Text). The unconstrained tree is presented as Supplementary Fig. S6, and the expanded constrained tree is presented as Supplementary Fig. S7. SATé alignment: maximum sub-problem size, 20; model WAG+G20. IQ-Tree construction: protein substitution model, WAG.

Figure 6. Scenario for the duplications, losses, and divergences underlying the expansion of cyclic nucleotide-gated channel α and β subunits (CNGA and CNGB) in jawed and agnathan vertebrates.

Conventions are as for Fig. 3.

Figure 7. Unconstrained ML molecular phylogeny for vertebrate visual opsins.

A. Unconstrained phylogeny shown in collapsed format; the fully expanded tree is shown in Supplementary Fig. S8. Here, the RhB/Rh2 sequence for *G. australis* did not fall within the Rh2 clade, though with some alternative selections of taxa it did (data not shown). B. Extract from the tree obtained when this sequence was constrained to clade with Rh2; the change in log likelihood was ΔLogL $\approx$ 2.2 and the tree passed all topology tests (with p-AU > 0.4). C. Extract from the tree obtained subject to the constraint that SWS1 and SWS2 were sister; the constraint tree used is shown below. In this case, the tree failed all three tests of topology (with p-AU < 0.03). The complete tree is presented as Supplementary Fig. S9. In B and C, the asterisks denote constrained clades. SATé alignment: maximum sub-problem size, 12; model WAG+G20. IQ-Tree construction: protein substitution model, WAG.

Figure 8. Hypothesised 2R gene duplications and losses that led to the chromosomal arrangement of vertebrate visual opsin and G-protein genes.

The lowermost panel shows the chromosomal locations, in the human genome, of the three visual opsins and relevant G-proteins (Rh1–GNAT1–GNAI2 on 3q, GNAT2–GNAI3 on 1p, SWS1–GNAT3–GNAI1 on 7q, and LWS on Xq), together with the presumed locations of the two visual opsins that were lost in an ancestral mammal (SWS2 on Xq, and Rh2 on 1p); see Larhammar et al, 2009, and Text. The possible equivalent position of the agnathan RhB gene is discussed in the Text. The top panel shows the presumed arrangement in the proto-vertebrate organism, prior to 2R, with four ancestral visual opsin genes plus GNAT and GNAI arranged in tandem; note, though, that their order is arbitrary. The middle panel shows how the upper and lower arrangements could have been linked, following the first round (1R) of whole-genome duplication. The crosses in the middle and lower panels show where a gene was lost. Note that, in this hypothetical scenario, the only opsin gene that underwent expansion during 2R was the ancestral Rh gene, that gave rise to Rh1, Rh2 and (if it is separate) RhB.

Figure 9. Gene duplications and losses hypothesised to have occurred during 2R.

The first dashed vertical line indicates the time at which the tunicate lineage diverged from the stem vertebrate lineage. The subsequent two dashed vertical lines indicate the timing of the two whole-genome duplications, labelled 1R and 2R, for the first and second rounds. The time-scale is arbitrary. Open squares (□) denote individual (tandem) gene duplications, whereas open

circles (○) denote genome duplications. For the visual opsins and the GNATs, the labels at the right indicate the human chromosome on which the gene is located (or, in the case of SWS2 and Rh2, on which the gene would have been located before its loss in the mammalian lineage, shown in [ ]); the braces emphasise the syntenic pairings: Rh1–GNAT1, Rh2–GNAT2, and SWS1–GNAT3. For the visual opsins, the colour coding is indicative of the spectral sensitivities.

Figure 10. Expression of activation proteins in different classes of photoreceptor.

Left: Presumed expression pattern in the ancestral proto-vertebrate organism prior to 2R. Right: Expression pattern in extant vertebrates. The isoforms employed in cones and rods of jawed vertebrates are shown in red and blue, respectively, with solid red and blue lines indicating the corresponding activation pathways. The isoforms determined from Table 1 to be utilised in *M. mordax* and *E. cirrhatus* are shown in the columns so labelled. The dotted red and blue lines indicate the deduced pathway for phototransduction in the exclusively cone-like and exclusively rod-like photoreceptors of these two species. Magenta indicates that PDE6X, and probably CNGAY, are distinct from any jawed vertebrate isoforms. The remaining agnathan isoforms are orthologous to jawed vertebrate isoforms: GNAT1, CNGA1 and CNGB1 (blue); PDE6C and CNGB3 (red); and GNAT3 (green). *G. australis* possesses five opsin classes, and from the values in Table 1 it seems probable that its cone-like photoreceptors (SWS1, SWS2, LWS) and its rod-like photoreceptors (Rh1) utilise the corresponding pathways. There is insufficient information to determine the pathway for the RhB-expressing *G. australis* photoreceptors, though it is possible (see Text) that these cells use a rod-type pathway, in contrast to the case for Rh2 in jawed vertebrates where the cone pathway is used.

Figure 11. Proposed origin of the proto-vertebrate phototransduction cascade.

Open arrows indicate catalysis; dotted lines indicate binding; $\oplus$ indicates potentiation; $\otimes$ indicates inhibition; $\downarrow$ indicates reduced concentration; red denotes new or modified component; blue denotes loss (or potential loss) of component.
A. Hypothesised ancestral chordate phototransduction cascade (black). Activated photopigment (R*) catalysed the activation of the inhibitory G-protein alpha subunit (Gα_I*), which bound to and inhibited adenylyl cyclase (AC). The decreased production of cAMP, in conjunction with continued hydrolysis by a phosphodiesterase (PDE) led to a decrease in cytoplasmic concentration of cAMP (cA), decreasing the fraction of open cyclic nucleotide-gated ion channels (CNGC). As the first step in the transition to the proto-vertebrate cascade, a small soluble protein, similar to the C-terminus of modern PDEγ, is proposed to have emerged (red γ); this would have provided some degree of inhibition of the PDE, thereby potentially extending the cell's operating range to higher light intensities.
B. Development of an interaction between Gα and the new inhibitory subunit. We hypothesise that mutations in both *GNAI* and the inhibitory factor permitted the G-protein (now denoted Gα_{IT}*) to bind to γ (red dotted line), thereby weakening the interaction between γ and the PDE, and hence potentiating the PDE activity (red $\oplus$). This would have augmented the reduction in cAMP elicited by R*. Once this new pathway had become dominant, it is possible that the interaction between Gα_{IT}* and AC would have weakened and eventually disappeared (blue cross); however, it may have been retained, because many extant transducin molecules possess an interaction motif for AC. The ancestral *CNGA* duplicated to form *CNGA4* and *CNGAQ* (see Figs 5 and 6), and we hypothesise that the ion channel then included both forms of alpha subunit (red α4, αQ), as in the channels of 'canonical' vertebrate olfactory receptor neurons.

C. Transition from cAMP to cGMP as the cytoplasmic messenger. We presume that, like PDE11, the ancestral phosphodiesterase was a dual-substrate PDE, that could hydrolyse both cAMP and cGMP. Likewise, we presume that the ion channels had a propensity to bind either cAMP or cGMP. Accordingly, if a guanylyl cyclase had been expressed in the outer segment, as we hypothesise (red GC), then the same cascade would have operated with cGMP as its cytoplasmic messenger (red cG). Once the G-protein and the PDE had evolved to optimise this pathway, we denote them as $G\alpha_T$* and PDE6 (red). We suggest that enhanced selectivity for cGMP may have been permitted by eliminating the α4 subunit from the ion channel. Once the cascade had transitioned to using cGMP as its messenger, it is possible that AC would no longer have been expressed in the outer segment (blue cross). The resulting cascade represents what we presume to have been the activation steps in the phototransduction cascade of the stem proto-vertebrate organism that existed prior to the two rounds of whole-genome duplication.
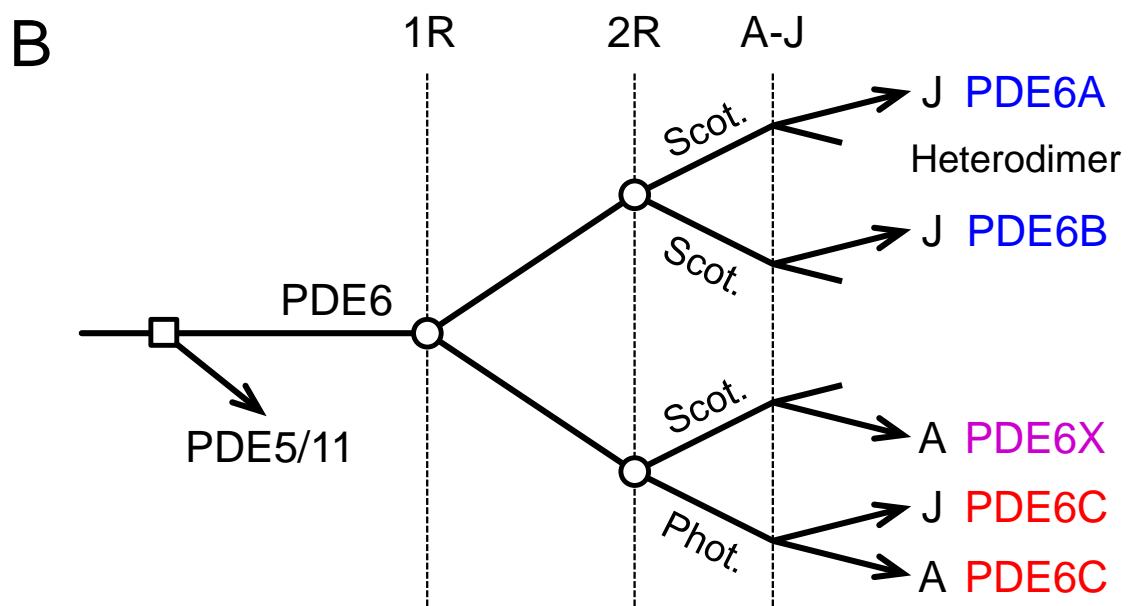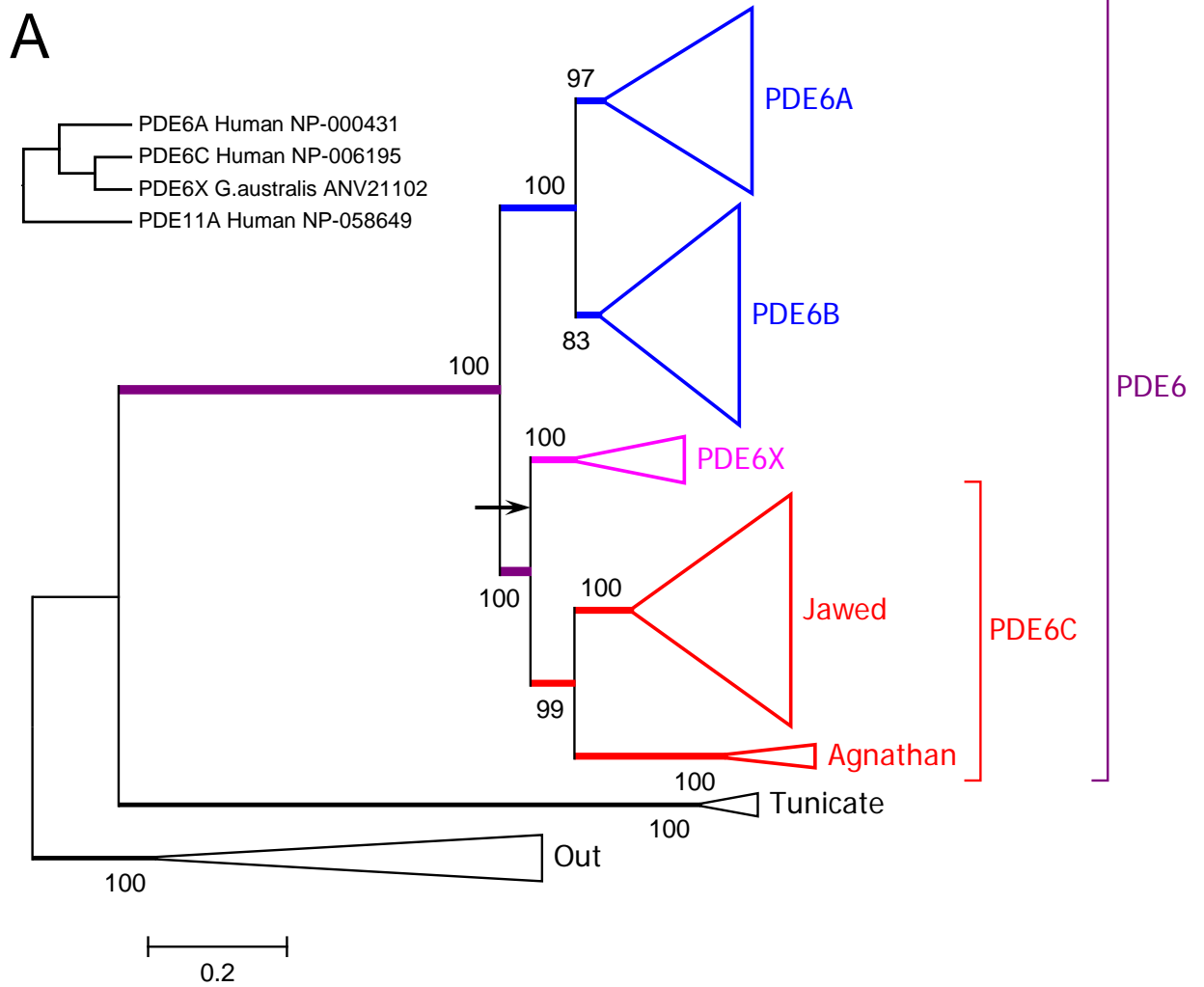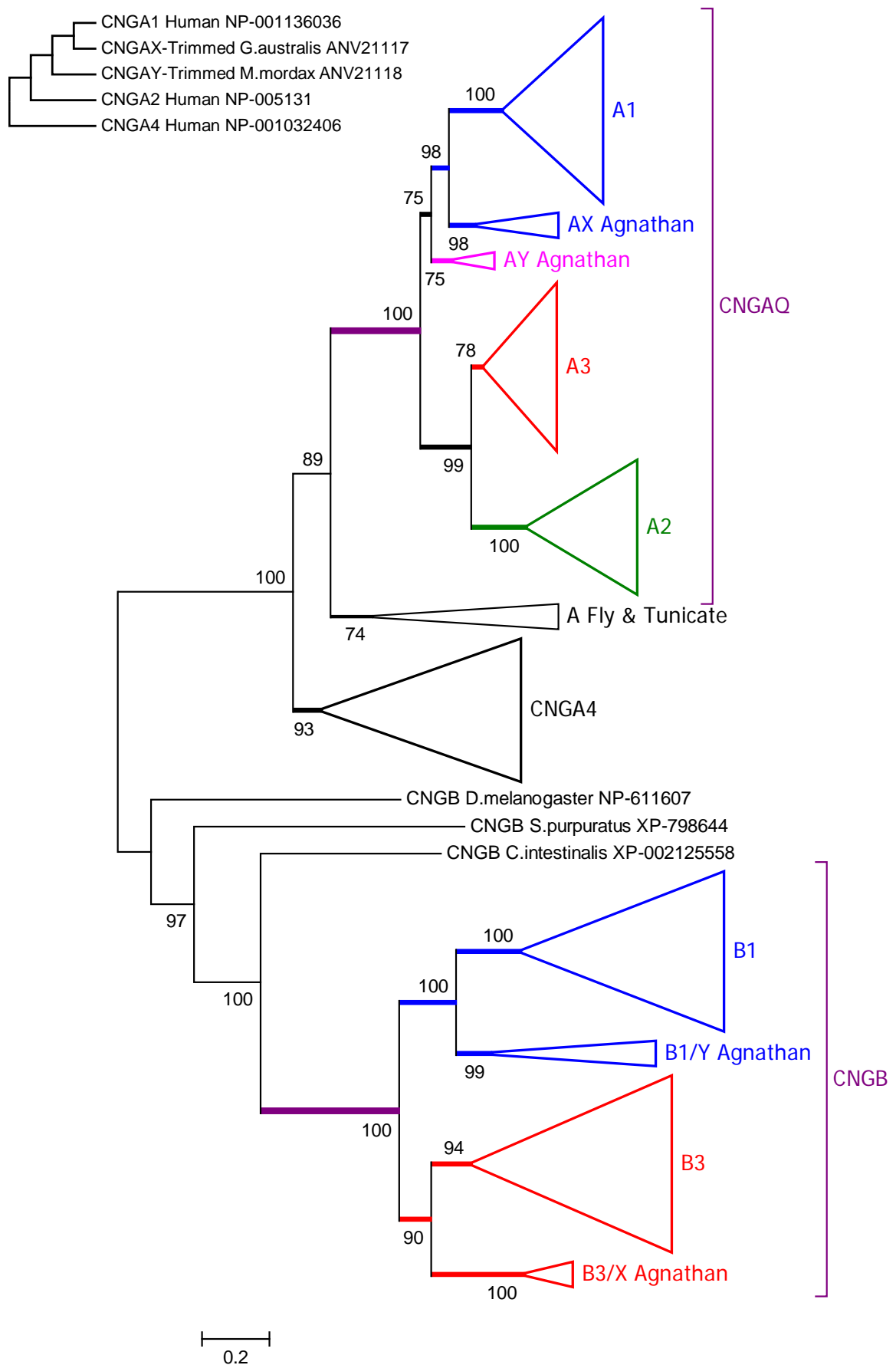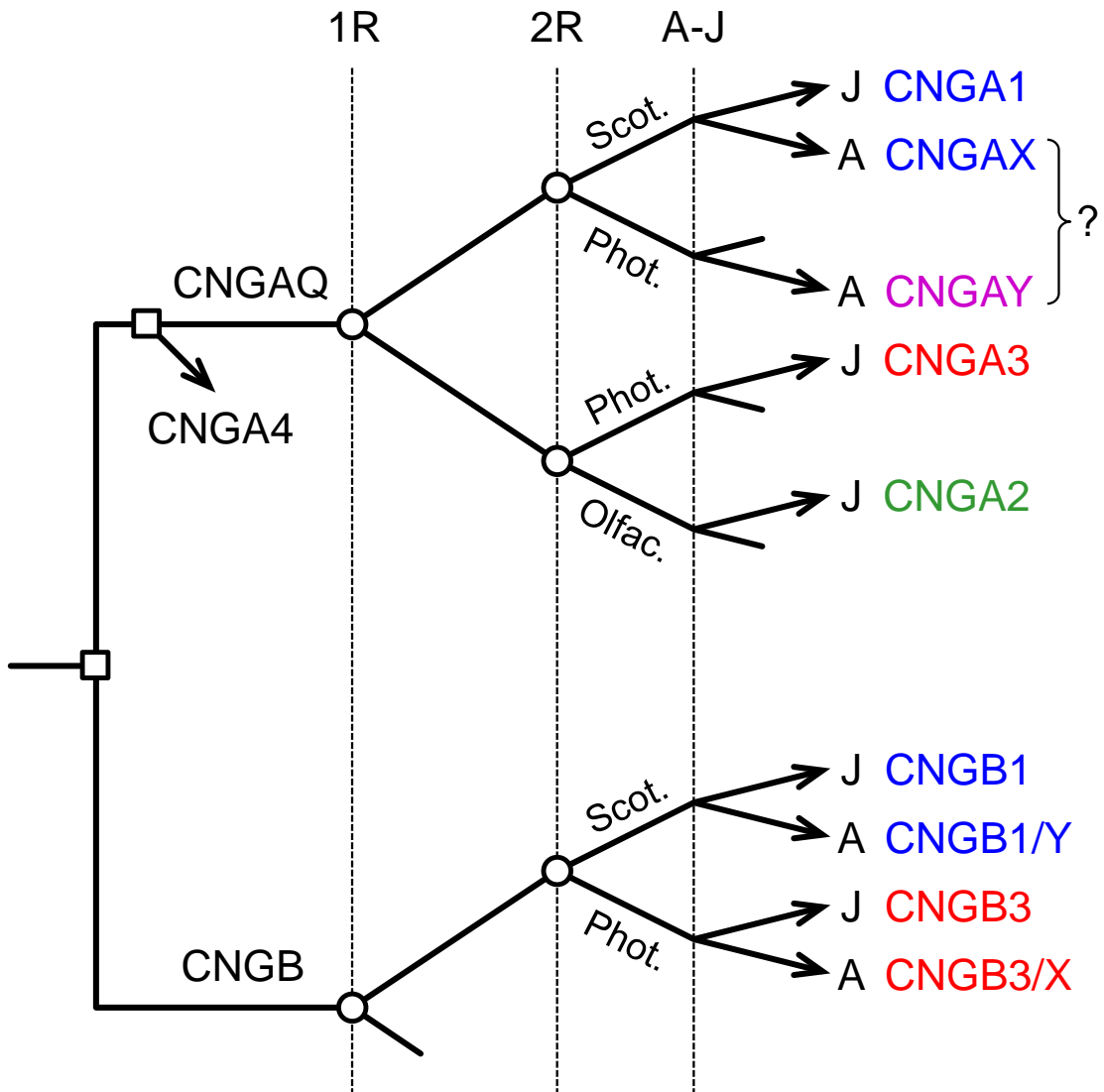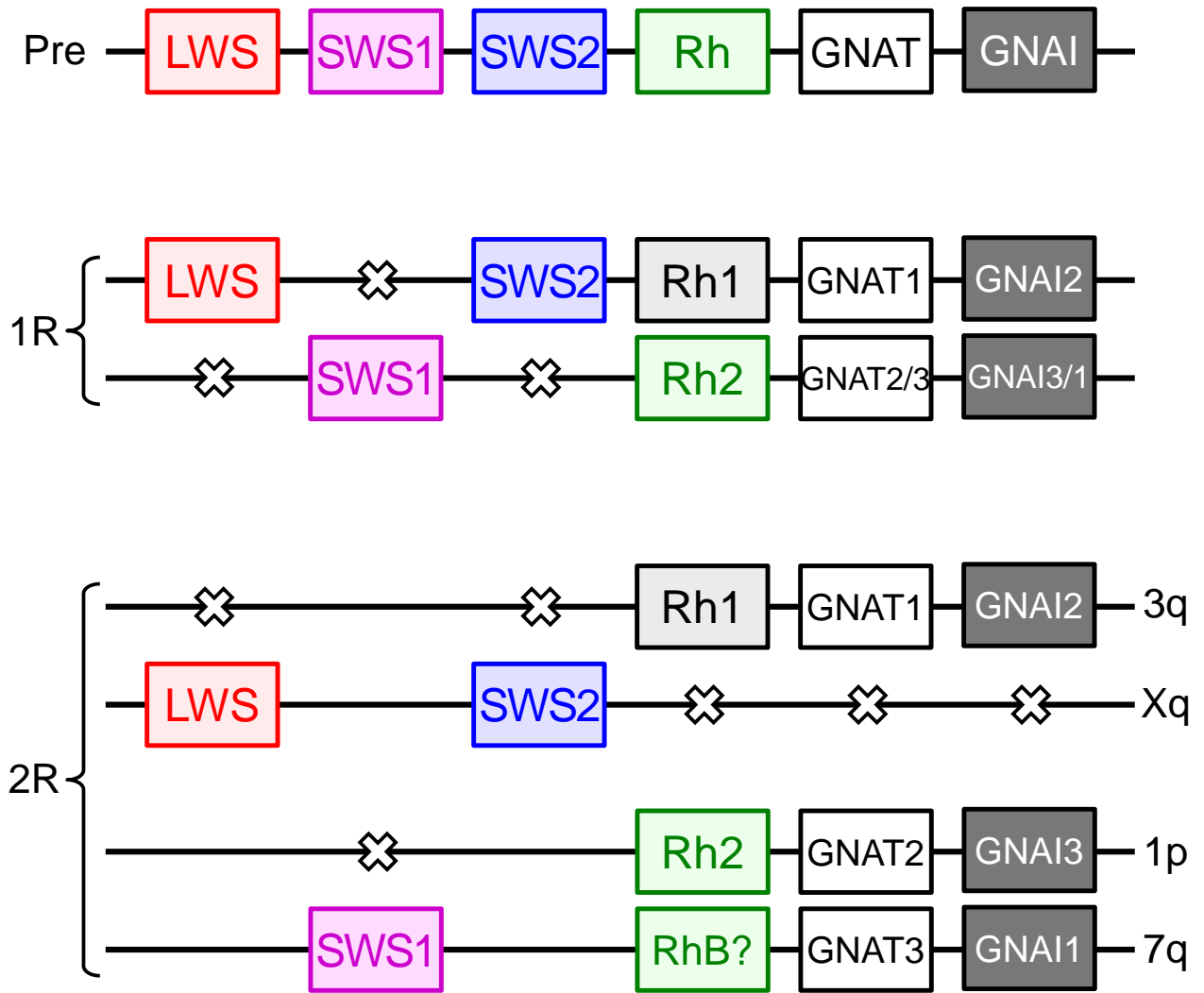
Figure 1
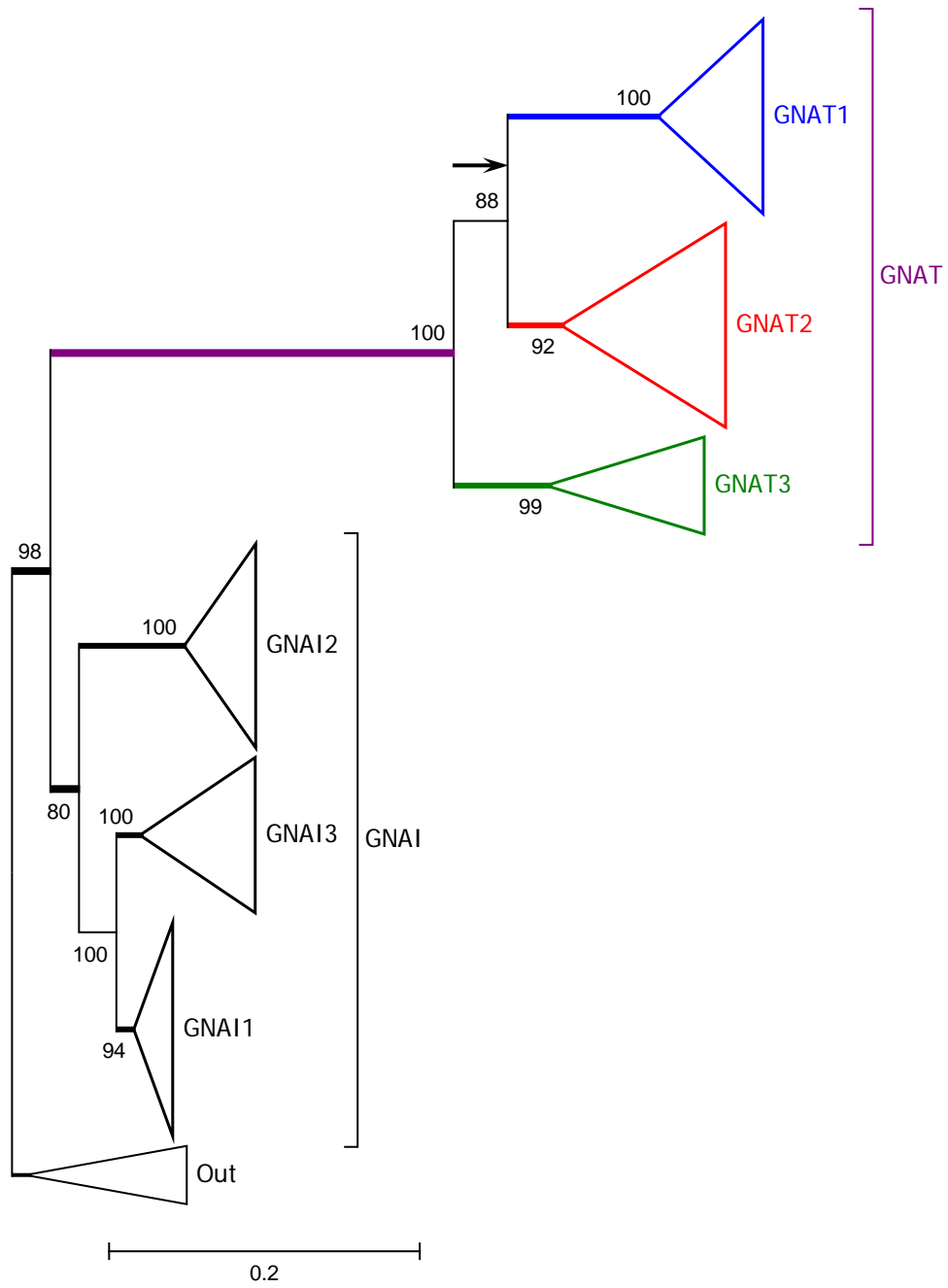
Figure 2
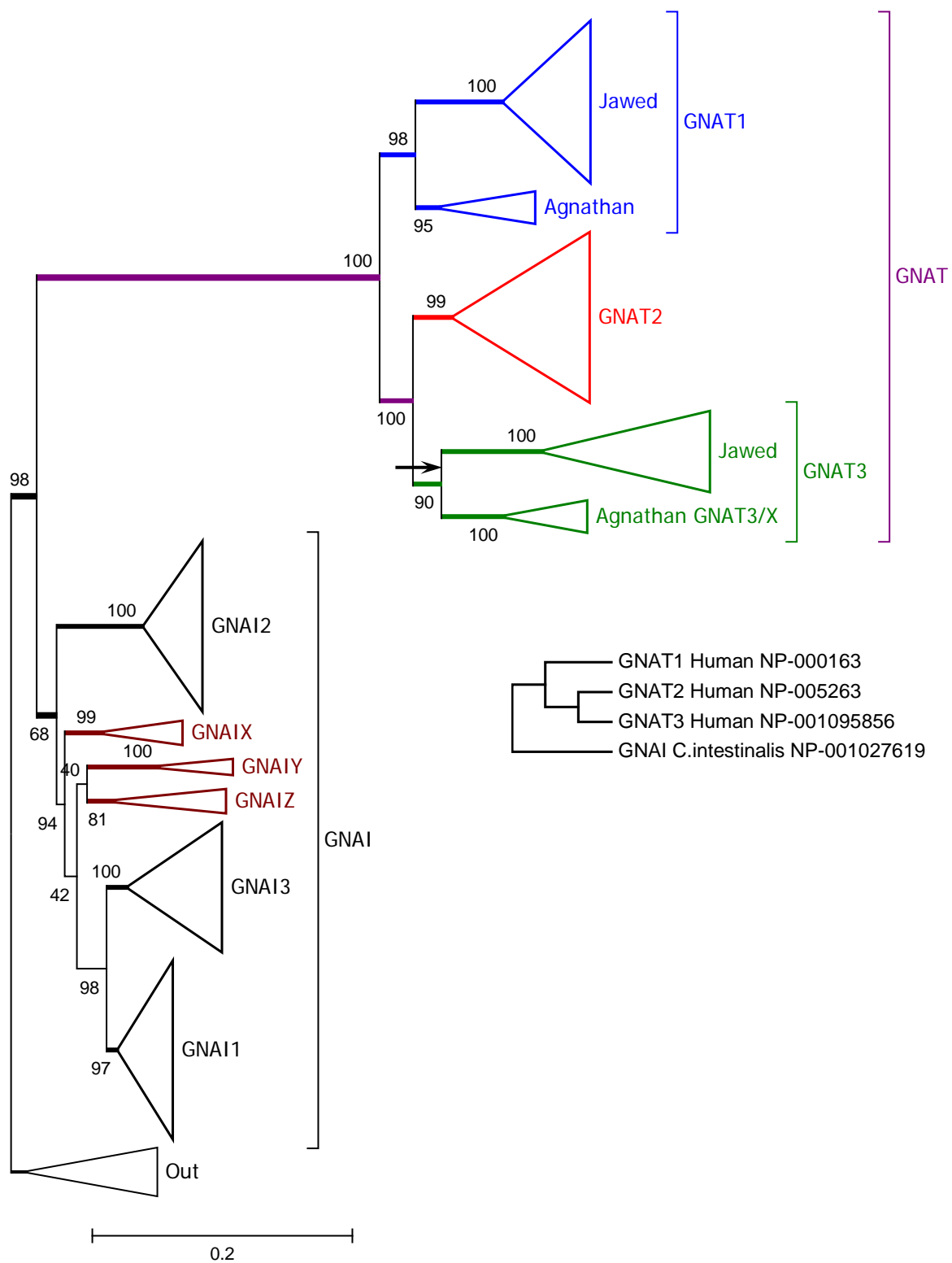
Figure 3
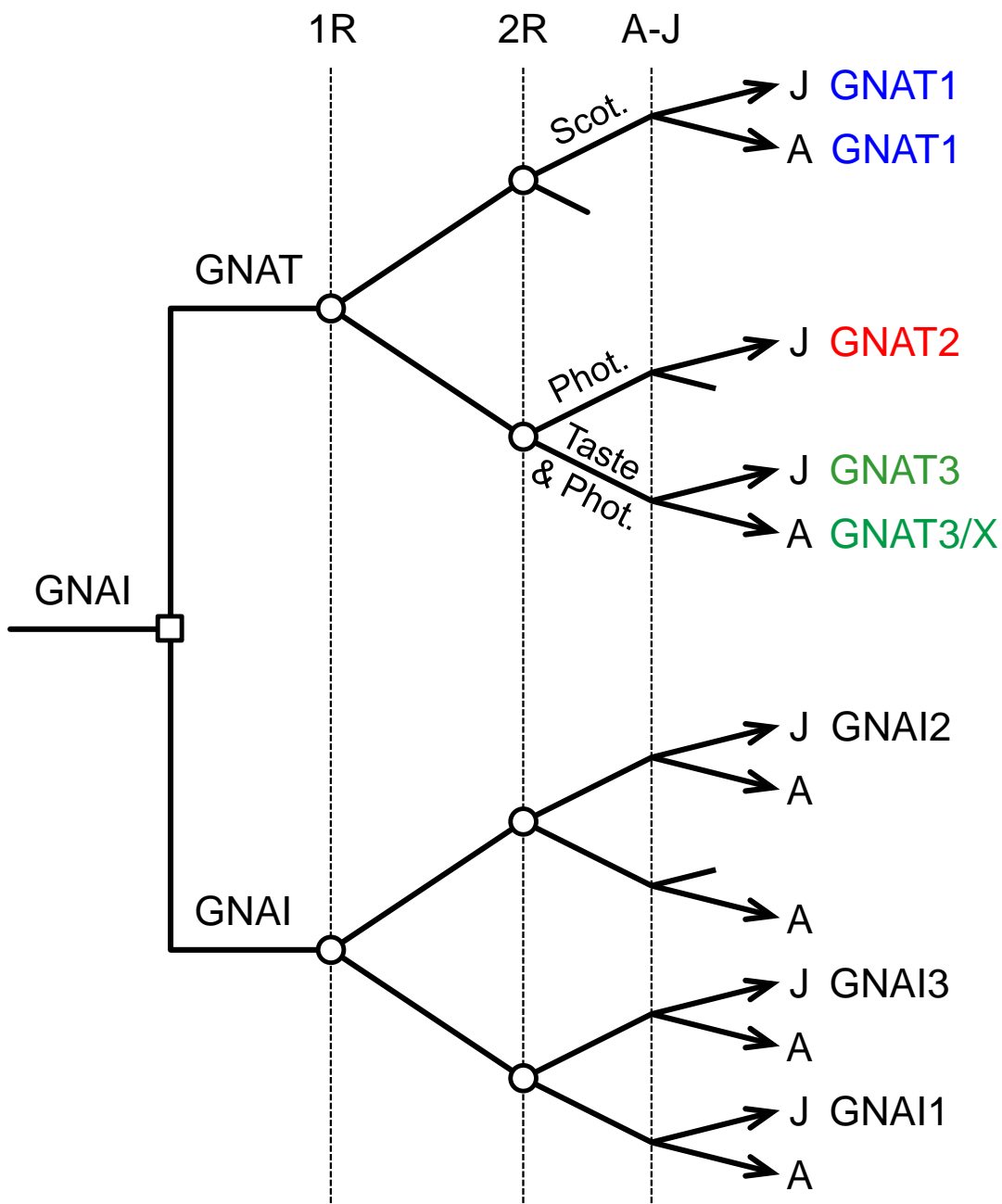
Figure 4

Figure 5

Figure 6

Figure 7
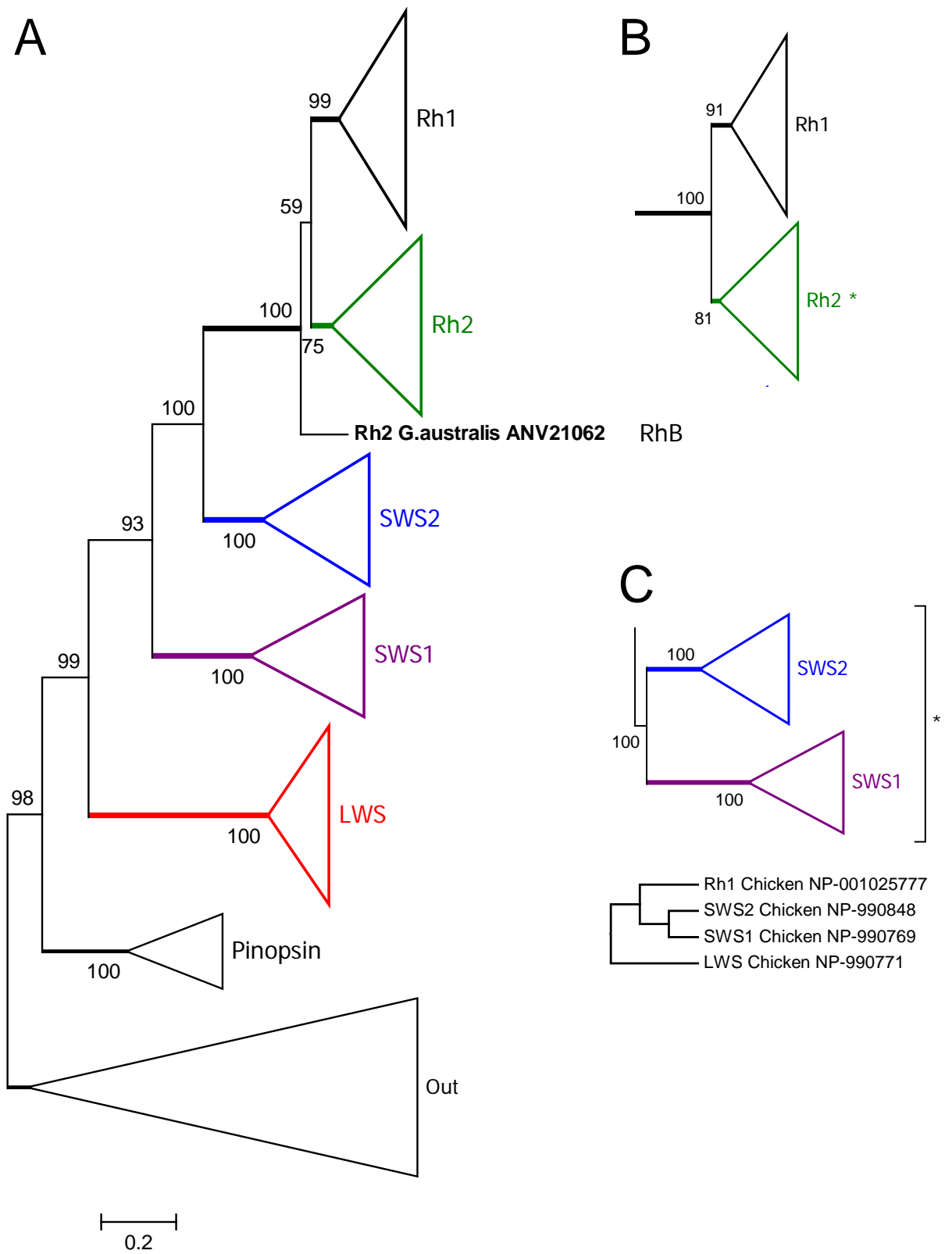
Figure 8

Figure 9

Figure 10

Proto-
vertebrate

Modern agnathan and jawed vertebrates

Photopic

Scotopic

4 Opsins

SWS1, SWS2, LWS

Rh1

GNAT

GNAT2   GNAT3 & Taste

GNAT1

PDE6

PDE6C

PDE6X   PDE6A
         PDE6B

Olfactory   Cone   *Mordacia mordax*   *Eptatretus cirrhatus*   Rod

CNGAQ
CNGB

CNGA2   CNGA3   CNGAY   CNGA1
         CNGB3   CNGB1

Geotria australis

Figure 11