Evolution of the vertebrate phototransduction cascade activation steps

Trevor D. Lamb[a],[*], David M. Hunt[b],[c]

[a] *Eccles Institute of Neuroscience, John Curtin School of Medical Research, The Australian National University, ACT 2600, Australia*
[b] *The Lions Eye Institute, The University of Western Australia, WA 6009, Australia*
[c] *School of Biological Sciences, The University of Western Australia, WA 6009, Australia*

ARTICLE INFO

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.

1. 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 and 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.

2. Materials and methods

2.1. 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 and Gomon (2016). 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 accession numbers in the range KT749668 – KT749760), we now report 40 new sequences, which have been submitted to GenBank and assigned accession numbers KY820586 – KY820625.

2.2. 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*.

2.3. 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.

2.4. 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 (Le and 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 and 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.

2.5. 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 the clades 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 and 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.

3. 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 multifurcation) 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