Evolution of vertebrate rod and cone phototransduction genes

Dan Larhammar1,*, Karin Nordström2 and Tomas A. Larsson1,†

1Department of Neuroscience, Unit of Pharmacology, Uppsala University, PO Box 593, SE-751 24 Uppsala, Sweden
2School of Molecular and Biomedical Science, The University of Adelaide, Adelaide SA 5005, Australia

Vertebrate cones and rods in several cases use separate but related components for their signal transduction (opsins, G-proteins, ion channels, etc.). Some of these proteins are also used differentially in other cell types in the retina. Because cones, rods and other retinal cell types originated in early vertebrate evolution, it is of interest to see if their specific genes arose in the extensive gene duplications that took place in the ancestor of the jawed vertebrates (gnathostomes) by two tetraploidizations (genome doublings). The ancestor of teleost fishes subsequently underwent a third tetraploidization. Our previously reported analyses showed that several gene families in the vertebrate visual phototransduction cascade received new members in the basal tetraploidizations. We here expand these data with studies of additional gene families and vertebrate species. We conclude that no less than 10 of the 13 studied phototransduction gene families received additional members in the two basal vertebrate tetraploidizations. Also the remaining three families seem to have undergone duplications during the same time period but it is unclear if this happened as a result of the tetraploidizations. The implications of the many early vertebrate gene duplications for functional specialization of specific retinal cell types, particularly cones and rods, are discussed.

Keywords: gene duplication; tetraploidization; eye; retina; phototransduction; opsin

1. INTRODUCTION

(a) Vertebrate phototransduction

Phototransduction is the process whereby light is converted to electrical signals in the eye. The signal transduction cascade of vertebrate rods and cones has been worked out in great detail and has been described in several excellent reviews (Fu & Yau 2007; Wensel 2008). The principal components of a vertebrate cone phototransduction machinery are shown schematically in figure 1. The cascade starts with the absorption of photons by the photoreceptor pigments, the opsins, which are G-protein-coupled receptors. The photon isomerizes the receptor's covalently bound 11-cis-retinal to all-trans-retinal which induces a structural change that activates the opsin. The signal is relayed by activation of transducin (a trimeric G protein) that stimulates the activity of phosphodiesterase (PDE), which reduces the level of the second messenger cGMP. This leads to closure of a cyclic nucleotide-gated (CNG) ion channel leading to hyperpolarization of the cell and reduced synaptic release of the neurotransmitter glutamate onto the bipolar and horizontal cells.

Signalling is reduced and terminated by phosphorylation of the opsin by the G-protein-coupled receptor kinase (GRK) followed by binding of arrestin. The system is restored by synthesis of cGMP by guanylyl cyclase (GC) stimulated by GC-activating protein (abbreviated GUCA or GCAP). Recoverin reduces the activity of GRK. A separate complicated machinery involving non-photoreceptor cell types returns the all-trans-retinal to 11-cis-retinal for re-insertion into opsin.

To some extent the proteins are the same in rods and cones, but there are also several cases where cones use one variant of a protein and rods use a different but closely related member of the same protein family expressed from a separate gene. For instance, the transducins, like all other trimeric G proteins, consist of α, β and γ subunits, and rods and cones have been found to express distinct variants of all three subunits: rods express the α subunit gene GNAT1, whereas cones use GNAT2, rods express the β subunit gene GNB1, while cones use GNB3, and rods express the γ subunit gene GNGT1, whereas cones use GNGT2 (see Nordström et al. 2004 for references).

Some of the phototransduction genes are also expressed differentially in other retinal cell types. One example is the transducin β subunit just mentioned, where the GNB1 gene has been found to be expressed in amacrine cells, whereas the GNB3 gene product was detected in some bipolar cells in macaque (cynomolgus monkey, Macaca fascicularis) as well as in rat, rabbit and cattle (Peng et al. 1992). The GNB3 gene is expressed in ON-bipolar cells in mouse (Huang et al. 2003). This shared use of specific gene family members by distinct retinal cell types may help shed light on the evolutionary and developmental relationships of the various cell types.
of the retina. Although several aspects of retinal development have been described (Dyer & Cepko 2001; Marquardt & Gruss 2002), many details remain to be resolved regarding the roles and the ontogeny of the many subtypes of bipolar, amacrine and ganglion cells.

The major cell types of the vertebrate retina, i.e. rods, cones, horizontal cells, bipolar cells, amacrine cells and ganglion cells, are thought to have originated in early vertebrate evolution (Lamb et al. 2007). Even subtypes of some of these cells arose in early vertebrates. Not only rods and cones are known to be present in lampreys, but also subtypes of bipolar cells and ganglion cells. Indeed, different categories of ganglion cells were first identified in a frog (Hartline 1938) and subtypes of bipolar cells were originally observed in a salamander (Dowling & Werblin 1969). Subsequently, such subtypes of cells have been found to exist in mammals, birds and teleost fishes. Therefore, it is of great interest to see if the genes that are differentially expressed in these cell types arose in the massive gene duplications that took place before the radiation of the jawed vertebrates, the gnathostomes (Miyata & Suga 2001), resulting from the recently confirmed tetraploidizations (Putnam et al. 2008).

We previously reported that many human phototransduction gene families are located in chromosomal regions that seem to have been duplicated in early vertebrate evolution, as based on sequence-based phylogenetic trees, supporting origin in the tetraploidizations (Nordström et al. 2004). We investigated nine gene families (opsins, transducins α, β and γ, PDE, CNG, GRK, arrestin and recoverin) and concluded that all nine of these received additional family members before the divergence of tetrapods and ray-finned fishes (although the recoverin family duplicated after the divergence of tetrapods and ray-finned fishes), and both of these lineages have at least four Hox clusters as concluded from the genomes of the two pufferfishes (Osteichthyes) and both of these lineages have at least four Hox clusters as concluded from the genomes of the two pufferfishes (Osteichthyes) and both of these lineages have at least four Hox clusters as concluded from the genomes of the two pufferfishes (Osteichthyes) and both of these lineages have at least four Hox clusters as concluded from the genomes of the two pufferfishes (Osteichthyes).

(b) Tetraploidizations in early vertebrate evolution (2R and 3R)

For several years the notion of tetraploidizations at the dawn of vertebrate evolution was highly controversial and widely discussed in the literature (Panopoulou & Pousta 2005). The concept of tetraploidizations was proposed already in 1970 by Susumu Ohno (Ohno 1970), but the time points for the events were rather vaguely discussed. The first hard evidence for extensive chromosome duplications was reported by Lars G. Lundin in 1993 who described four sets of chromosomes in the human and mouse genomes with similarities among the members of each set (Lundin 1993). Two of the sets consisted of four chromosomes indicating that the genome had been quadrupled through two tetraploidizations. The double tetraploidization scenario was later named 2R for two rounds of genome doubling and this abbreviation will be used in this article.

Detailed molecular genetic evidence for one quadrupled chromosome region was presented when the lancelet Branchiostoma floridae, also called amphioxus, a cephalochordate, was found to have a single gene cluster with Hox-type homeobox genes (Garcia-Fernández & Holland 1994), whereas tetrapods have four Hox clusters. But it was not until the long-awaited genome assembly of amphioxus was published in June 2008 (Putnam et al. 2008) that it was confirmed that many parts of its genome correspond to four paralogous regions in the genomes of Homo sapiens and other tetrapods. Also several previous studies had accumulated strong evidence for quadruplication of many large chromosome regions (Abi-Rached et al. 2002; Larhammar et al. 2002; Lundin et al. 2003; Vienne et al. 2003; Dehal & Boore 2005; Nakatani et al. 2007) in line with the double tetraploidization hypothesis. Related chromosome regions within a genome are said to be paralogous, and together they form a paralogon (Coulier et al. 2000), i.e. a set of related regions, usually a quartet.

Both of the basal vertebrate tetraploidizations have been confirmed to have taken place before the radiation of the gnathostomes (Vandepoele et al. 2004; figure 2). The first divergence among these led to Chondrichthyes (cartilaginous fishes) and Osteichthyes (ray-finned and lobe-finned fishes), and both of these lineages have at least four Hox clusters as concluded from the genome sequence of the elephant shark, Callorhinus milii (Venkatesh et al. 2007) and all studied representatives for Osteichthyes (Hoegg & Meyer 2005). Recently, we reported that the elephant shark also possesses members in a gene family located on all four chromosomes of a different quartet in tetrapods, namely the neuropeptide Y (NPY) receptor family (Larsson et al. 2008), thereby supporting its shared genome quadruplication with Osteichthyes (Venkatesh et al. 2007). Studies in ray-finned fishes (Actinopterygii) have shown that euteleost fishes (Euteleostei) have undergone a third tetraploidization named 3R. This was suggested already by the zebrafish genome (Postlethwait et al. 2000) and persuasively concluded from the genomes of the two pufferfishes Takifugu rubripes (Christoffers et al. 2004) and Tetraodon nigroviridis (Jaillon et al. 2004).

It is still unclear when the two basal tetraploidizations occurred relative to the divergencies of the two cyclostome groups, the lampreys and the hagfishes (figure 2). It has been proposed that one tetraploidization took place...
before the cyclostomes branched off, whereupon the second tetraploidization took place in the lineage leading to the gnathostomes (Holland et al. 1994; Furlong & Holland 2002). Recently, it was suggested that both of the tetraploidizations happened before the divergence of cyclostomes (Kuraku et al. 2009), implying perhaps that more gene losses may have taken place in the cyclostome lineages. Another issue pertinent to this discussion is whether the cyclostome lineages are monophyletic or paraphyletic (Stadler et al. 2004; Lamb et al. 2007).

Traditionally, the hagfishes have been considered to be the sister group of the vertebrates, and together with the cephalochordates (lancelets, amphioxus) they comprise the craniates (animals with a skull). However, several studies suggest that the hagfishes and lampreys are monophyletic, together forming the sister group of the gnathostomes (Takezaki et al. 2003; Kuraku et al. 2009). Nevertheless, the confirmation of the 2R scenario means that a huge number of gene duplicates was generated early in vertebrate evolution. A large proportion of the duplicates was presumably lost at an early stage after the tetraploidizations, and duplicates have continued to disappear in various lineages. However, it is likely that the extensive duplications resulting from the two more or less complete genome doublings gave rise to many new genes that facilitated the appearance of new functions (neo-functionalization) or more specific functions such as restricted gene expression (sub-functionalization), thereby providing raw material for emergence of new cell types with specialized functions (Arendt 2008).

Important for resolving the evolution of vertebrate genes is to have appropriate outgroups for comparison. For a long time the cephalochordates (lancelets, amphioxus) have been considered to be the sister group of craniates or vertebrates. The cephalochordates are comprised of more than 30 species of which the majority belongs to the genus Branchiostoma. Extensive sequence analyses have suggested that the cephalochordates are actually slightly more distantly related to the vertebrates than the urochordates (Delsuc et al. 2006). This was recently confirmed by the genome sequence of the Florida lancelet Branchiostoma floridae (Putnam et al. 2008).

2. PHOTOTRANSDUCTION GENE FAMILIES
Below we describe in detail each of the gene families involved in the phototransduction cascade. We primarily describe gene locations in the human genome. Chromosomes are abbreviated as Hsa for Homo sapiens and Gga for Gallus gallus, i.e. chicken. For phylogenetic trees we usually refer to our previous study (Nordström et al. 2004) or to trees available in the

![Figure 1. Schematic outline of the phototransduction cascade in a vertebrate cone. All of the components are shown within or near the same membrane, the cell-surface membrane, as is assumed to have been the case in the ancestor of cones and rods (rods subsequently evolved an internalization mechanism for parts of the membrane to form intracellular discs whose membranes harbour some of the phototransduction components including rhodopsin). Upon absorption by a photon, 11-cis-retinal within the opsin is transformed to all-trans-retinal. This activates the opsin, which in turn activates the transducin (a G protein) consisting of three subunits (α, β, and γ). Transducin in turn activates phosphodiesterase 6 (PDE), which forms a dimer with α and β (or two α subunits), inhibited by two γ subunits. Active PDE hydrolyzes cGMP to GMP. cGMP keeps the cyclic nucleotide ion channel (CNG, a tetramer with α and β subunits) open. New cGMP is generated by guanylyl cyclase (GC), which is regulated by a GC-activating protein (GUCA or GCAP). Rhodopsin kinase (GRK) deactivates the opsin by phosphorylation, and is itself regulated by recoverin. Arrestin binds to the phosphorylated opsin to reduce its signalling to the G protein. GRK is anchored to the membrane and arrestin undergoes translocation to the membrane on activation (not shown in the figure).](http://rstb.royalsocietypublishing.org/)

Phil. Trans. R. Soc. B (2009)
Evolution of vertebrate phototransduction

TreeFam database v. 6.0 (www.treefam.org) (Li et al. 2006; Ruan et al. 2008). The trees have been accessed between 15 December 2008 and 15 January 2009.

(a) Opsins
Several recent reviews have focused on the evolution of opsins both within the vertebrates themselves (Lamb et al. 2007; Bowmaker 2008) and across the animal kingdom (Oakley 2003; Terakita 2005). As is well known, rhodopsin (RHO) is expressed in rods and the colour vision opsins are typically specific for distinct types of cones. However, some overlap of expression may occur; mice express the two cone opsins called S (ancestral UV, OPN1SW) and L (OPN1LW) in the same cones in a large part of the retina (Applebury et al. 2000). Note that the long-wavelength opsin was called M in the cited reference. Rabbits and guinea pigs also co-express two opsins (S and L) in their cones, but not as extensively as shown for the mouse. Co-expression of these opsins is not observed in adult rats and gerbils, but transiently during development (Szel et al. 2000).

Our initial analyses indicated that human RHO (RH1), OPN1SW (blue = ancestral ultraviolet), and OPN1LW (red) occupy corresponding positions on the three chromosomes Hsa3, Hsa7 and HsaX, suggesting that they might have arisen from a common ancestral opsin by chromosome duplication (Nordström et al. 2004). However, this interpretation was too simplistic because analyses from a broad range of vertebrates have shown that the ancestral gnathostome had five visual opsin genes: RHO, RH2 (ancestral ultraviolet), OPN1SW2 (blue) and OPN1LW (red) (Bowmaker 2008). Furthermore, chromosomal data in teleost fishes show that OPN1SW2 and OPN1LW are syntenic (i.e. they are located in the same chromosome region). Because phylogenetic sequence analyses show that the earliest branching in the visual opsin tree involves these two genes (Bowmaker 2008), the ancestral chromosome probably contained both of these genes before the chromosome quadruplication took place. If so, all duplicates of OPN1LW must have been lost as shown in the hypothetical chromosome duplication scheme shown in figure 3. However, the chromosomal locations of the visual opsin genes in teleost fishes seem to add further complexity to their evolutionary history, implying that even the duplication scheme proposed in figure 3 is too simplistic.

Additionally, some of the adjacent gene families in these chromosome regions need to be revised. As recently noted by Kuraku et al. (2009), the PLXN gene family is not a single series of paralogues, but in fact involves an ancestral gene pair with the ancestor of the PLXNA family and the ancestor of the PLXNB family. After considering this as well as additional information, the chromosome duplication scenario actually receives even more support from adjacent gene families as shown in figure 3 than it did in our previous study (Nordström et al. 2004).

The evolution of vertebrate visual opsins is further complicated by the fact that even though the cephalochordate amphioxus has five ‘ciliary’ opsins, all these are related to the encephalopsin/TMT-opsin subfamily, and no orthologues of the vertebrate visual opsins (including RHO) were identified (Holland et al. 2008) in a recent analysis of its genome. Opsins similar to the vertebrate ancient and RPE-retinal G-protein receptor (RGR) opsins have been identified in Ciona intestinalis (Kusakabe et al. 2001; Horie et al. 2008). To resolve this, chromosomal information is needed for tunicates, amphioxus and the cyclostome groups of hagfishes and lampreys, as well as the earliest gnathostome branch comprised of the cartilaginous fishes.

(b) Transducins α, β and γ
Transducins are trimeric G proteins found in photoreceptor cells. Each G protein consists of unrelated
\[\alpha, \beta\] and \[\gamma\] subunits and each of these has multiple members in vertebrate genomes. In mammals 16 \[\alpha\] (GNA) subunit genes (Birnbaumer 2007) have been identified that are sorted into subfamilies based on sequence phylogeny (Nordström et al. 2004) and functional coupling (Birnbaumer 2007). Two transducin \[\alpha\] genes (GNAT) are expressed in the retina of mammals where they stimulate the activity of phosphodiesterase, with GNAT1 expressed in rods and GNAT2 in cones (Downes & Gautam 1999). A third member, GNAT3 (gustducin), is expressed in taste cells. The two transducins and gustducin are almost equally closely related to each other (Downes & Gautam 1999). Interestingly, the gustducin gene has been found to be involved in the phototransduction cascade of the lizard parietal eye, whereas the transducins were not (Su et al. 2006). Each of the three genes is located close to a Gi\[a\] (G-protein inhibitory \[a\] subunit; GNAI) subunit gene and these are also equally distantly related to each other (Downes & Gautam 1999). The GNAT and GNAI genes have the same exon–intron organization (Peng et al. 1992). This led us to propose that an ancestral gene pair consisting of one GNAT gene and one GNAI gene was duplicated twice to give the three pairs on Hsa1p, 3p and 7q (see Nordström et al. 2004). These chromosome arms, together with Hsa12p, form a paralogon containing numerous other gene families.

Early work suggested that the \[\alpha\] subunit was responsible for the signalling of each G protein, but recently the roles of the \[\beta\] and \[\gamma\] subunits have received intense interest (for a recent review, see Birnbaumer 2007). The G-protein \[\beta\] subunit gene family consists of seven human members (Downes & Gautam 1999). Members GNB1 to GNB4 are approximately equally closely related to each other (figure 4), whereas GNB5 is more distantly related and GNB1L and GNB2L1 even more remote (Nordström et al. 2004). GNB1–4 show broad tissue distribution with differential expression in retinal cell types. Rods and amacrine cells express GNB1 while cones and bipolar cells express GNB3. GNB1–4 diverged from each other after the vertebrate lineage split from the urochordate lineage represented by two tunicates in figure 4 (see also Nordström et al. 2004). GNB1–4 (figure 5) are located on Hsa 1p, 3q, 7q and 12p and are flanked by many other gene families in this paralogon. Also GNB5 is expressed in rods as well as in the brain (Watson et al. 1996). Its role in phototransduction is little known but it has been reported to contribute to rod adaptation because knock-out mice show reduced G-protein deactivation (Krispel et al. 2003). GNB5 seems to have no close relatives resulting from 2R.

The G-protein \[\gamma\] subunit family consists of 12 mammalian members (Serb & Oakley 2005; Birnbaumer 2007) and many of these are expressed in the brain. Distinct \[\gamma\] subunit genes are used in rods and cones: GNGT1 and GNGT2. Whereas mammalian GNGT1 is expressed exclusively in the eye, the zebrafish orthologue is expressed both in the pineal gland and retina, but driven by different transcription factors (Ferguson et al. 1998). The human GNGT1 and GNGT2 genes are located on human chromosomes 7q and 17q, respectively, thereby belonging to the extended Hox paralogon (Nordström et al. 2004; Sundström et al. 2008). GNGT1 shows a low level of identity (<80%) compared to that observed for other \[\gamma\] subtypes that typically show greater than 90 per cent amino acid identity between teleosts and mammals (Leung et al. 2006). This lower level of identity may reflect adaptation to different environmental conditions. Alternatively, more rapid divergence of mammalian GNGT1 may have been facilitated by a duplication that generated a gene adjacenty on the same chromosome (Hsa7q) called GNG11 (Kaupp & Seifert 2002). GNG11 probably arose in a mammalian ancestor and is highly expressed in a variety of tissues such as the pineal gland, the heart, lung and skeletal muscle (Balcueva et al. 2000; Chen et al. 2007).

(c) Phosphodiesterases

The phosphodiesterases (PDEs) form a superfamily with 21 members in mammals that are classified into 11 families (Omori & Kotera 2007). Some of these enzymes hydrolyze cGMP to GMP, whereas others hydrolyze cAMP to AMP. The retina expresses the three members of a PDE family called PDE6\(\alpha\)–\(C\), all of which act on cGMP. Many of the PDEs function as dimers, including the PDE6 family. The PDE6\(\alpha\) and PDE6\(\beta\) genes encode the subunits PDE6\(\alpha\) and PDE6\(\beta\), respectively, that form a PDE\(\alpha\)\(\beta\) heterodimer in rods. The PDE6\(\alpha\) gene gives rise to the PDE6\(\alpha\)\' subunit which is expressed in cones and forms a homodimer, PDE6\(\alpha\)\(\alpha\)' (see Cote 2004 for review).

We previously reported (Nordström et al. 2004) that the three PDE6 protein sequences are more closely related to each other than to all other members of the PDE superfamily (see also Omori & Kotera 2007). The duplications that gave rise to the three copies took place in early vertebrate evolution as supported by sequence-based phylogeny, taxonomic distribution in vertebrate classes, and chromosomal positions of the three genes in the large paralogon described (see fig. 3 in Lundin et al. 2003), consisting largely of human chromosomes 4, 5, 2/8 and 10. This corresponds to paralogon C in Nakatani et al. (2007) and part of this paralogon has been described in detail in several species in connection with the NPY receptor family (Larsson et al. 2008).

The enzymatically active PDE6 dimer is inhibited by two copies of a \(\gamma\) subunit. Activation of PDE6 takes place when two transducin \(\alpha\) subunits bind and remove the \(\gamma\) subunits. The rod \(\gamma\) subunit is encoded by a gene called PDE6G, whereas cones have a distinct \(\gamma\) subunit encoded by the PDE6H gene (Cote 2004). Sequence-based phylogeny suggests that the two genes resulted from a duplication in early vertebrate evolution. The genes are located on Hsa12 and Hsa17, the same chromosomes as two of the Hox clusters, albeit some distance away from the Hox clusters. However, the PDE6 \(\gamma\) genes are closer to the Hox-cluster genes in the dog genome (data not shown), suggesting that they may be part of the same chromosome duplication events, but more detailed studies involving additional species are necessary to confirm this.

In addition to cones and rods, PDE has been detected in bipolar cells, namely PDE1C in ON bipolar (Dhingra et al. 2008). However, because PDE
Figure 3. Chromosomal locations of vertebrate visual opsin gene family members in the human genome (Hsa for *Homo sapiens*) and postulated positions for the lost ancestral blue and green opsins. The proposed position for the ancient green opsin (RHO2) is confirmed by the present location of the orthologue in the chicken genome in a region with conserved synteny with Hsa1. Because phylogenetic trees for vertebrate visual opsins have OPN1LW and OPN1SW as the most basal divergence (Bowmaker 2008), we postulate that there was an ancestral gene pair before the chromosome quadruplication. However, other scenarios are possible. This hypothetical chromosome duplication scenario needs to be investigated further by studies in genomes of representatives from additional vertebrate classes. The transducin $\alpha$ subunit (GNAT) involved in phototransduction is located in the same paralogon, with GNAT1 expressed in rods, and GNAT2 in cones.

Figure 4. Phylogenetic trees for the G-protein $\beta$ subunit family GNB1–4. The species included are two mammals (human and mouse), one bird (chicken) and a pufferfish (*Takifugu rubripes*). As outgroup the GNB gene of the tunicates *Ciona intestinalis* and *Ciona savignyi* were used. The trees were calculated using the neighbour-joining method as implemented in MEGA 4.0 (Tamura et al. 2007) with default settings. Panel (a) shows the obtained tree collapsed for nodes with less than 50 per cent bootstrap support. Panel (b) shows the same tree with branch lengths representing evolutionary distance. The two *T. rubripes* genes GNB3a and GNB3b have evolved at a much higher rate than all of the other sequences and their common branch has been shortened as shown by the double dash in panel (b).
inhibitors do not affect the response of ON bipolars (Nawy 1999), it has been suggested that PDE has a modulatory role in these cells and that their glutamate receptor mGluR6 activates Gao to regulate ion channels (Dhingra et al. 2008).

(d) Cyclic nucleotide-gated ion channels
There are six members of the CNG gene family in mammalian genomes. Four of these encode α subunits (CNGA1–CNGA4) and two encode β subunits (CNGB1 and CNGB3) (Kaupp & Seifert 2002). Functional ion channels are heterotetramers, in rods and cones consisting of three CNGA subunits and one CNGB subunit (Zheng et al. 2002). Rods and cones use distinct subunits with rods having CNGA1 and CNGB1, whereas cone channels consist of CNGA3 and CNGB3 (Paillart et al. 2006). Olfactory neurons in contrast use three distinct subunits: CNGA2, CNGA4 and CNGB1b with two copies of the first-mentioned (Zheng et al. 2002). The CNG channels are non-selective cation channels, thus allowing influx of both Na⁺ and Ca²⁺.

Our previous analyses showed that the duplication that led to the α and β lineage took place before the protostome–deuterostome divergence (Nordström et al. 2004). The phylogenetic trees showed expansion of both lineages in early vertebrate evolution and this receives further support from data from additional vertebrate genomes in the TreeFam database (see also Paillart et al. 2006). Although quadruplicate resemblance could be detected for the four CNGA genes, they are not located in any of the large paralogs described in detail (Nordström et al. 2004). Our subsequent analyses suggest that some of the CNGA genes are located close to a few other gene families that seem to have been translocated after the chromosome duplications (data not shown), thereby obscuring the paralogon relationship. The two CNGB genes are the only survivors of the chromosome quadruplications that involved segments that are located on human chromosomes 8, 16, 18 and 20. Again, translocations make analyses complicated but it seems like these chromosome regions share an evolutionary history with the paralogon containing the four opioid receptor genes and several other gene families on Hsa1 and 20 (the CNGB genes are located close to these clusters), also involving Hsa1 and 6 (Dreborg et al. 2008), presumably due to chromosomal rearrangements.

Some teleost fishes have duplicates of both CNGA3 and CNGB3 (see TreeFam) and it will be interesting to see if these duplicates have become specialized by restriction of their expression to distinct cone types in the teleosts that have more colour opsins than mammals.

(e) Guanylyl cyclases
Many cell types possess the ability to produce cGMP via GC, either by soluble cytosolic GCs or by membrane GCs. The membrane-penetrating GC proteins (mGC) can be further subdivided into two major categories, those that serve as receptors for extracellular ligands and those that are activated by intracellular proteins (Sharma 2002). The last-mentioned type
G-protein-coupled receptor kinases

GRKs are Ser/Thr kinases that phosphorylate cytoplasmic parts of G-protein-coupled receptors and reduce their signalling to G proteins. The human genome contains seven GRK genes with GRK1 (previously called rhodopsin kinase, RHOK) predominantly expressed in rods and GRK7 (previously GPRK7) in cones, albeit with some interesting species differences. GRK1 is expressed mainly in rods but also in cones in humans (Zhao et al. 1998), as in monkey (Sears et al. 2000; Weiss et al. 2001) and mouse (Lubarsky et al. 2000) retinas, but not in the cones of dog and pig retinas (Weiss et al. 2001). GRK1 is expressed exclusively in rods of the medaka fish (Hisatomi et al. 1998). In contrast, GRK7 is concentrated in cones in both cone- and rod-dominant mammals (Weiss et al. 2001) and in fish (Hisatomi et al. 1998). An intriguing exception is that the expression of GRK7 is not detectable in mouse and rat retinas, nor is the GRK7 sequence present in mouse or rat cDNA or mouse genomic libraries and deposited genomic sequences (Weiss et al. 2001), although it is present in the guinea pig and rabbit genomes. GRK1 and GRK7 are attached to the membrane by an isoprenyl moiety (Inglese et al. 1992; Zhang et al. 2004).

In our earlier phylogenetic analyses, these two retinal GRKs did not branch together although their chromosomal positions suggested that they arose from a common ancestral gene in 2R (Nordström et al. 2004). More extensive data from various vertebrate genomes now add further support for their origin in early vertebrate evolution, as shown by their presence in both teleosts and tetrapods (Wada et al. 2006). Furthermore, their chromosomal locations can now be assigned to agree with paralogon F (Nakatani et al. 2007) with extensive regions of similarity (i.e. representatives from the same gene families) on Hsa3 (GRK7) and Hsa13 (GRK1) as well as HsaX (Nordström et al. 2004) and Hsa 1, 2 and 11 (Nakatani et al. 2007; see figure 5). These chromosome regions form a paralogue that also contains the GNb and arrestin gene families (figure 5). Teleost fishes have duplicates of both GRK1 and GRK7 that seem to agree with 3R. Furthermore, there is a third member forming a basal branch for this GRK family in teleost fishes (Wada et al. 2006; Imanishi et al. 2007) that may have been lost in mammals but appears to be present in chicken (Imanishi et al. 2007) and which appears not to be duplicated in teleost-specific tetraploidization 3R. More detailed analyses of chromosomal location in different teleost genomes and in the chicken genome are required to determine if this gene is a third product of 2R in addition to GRK1 and GRK7. If so, the names of some of the members of this gene family will probably require revision to comply with orthology.

The remaining five GRKs in the human genome belong to two other paralogons. GRKs 4, 5 and 6 (Nordström et al. 2004) are located in the NPY receptor paralogue involving Hsa4, 5, 2/8 and 10 (Larsson et al. 2008), i.e. paralogue C in Nakatani et al. (2007). GRK4 was previously called GPRK2L. GRKs 2 and 3 (formerly called β-adrenergic receptor kinases 1 and 2, abbreviated ADRBK1 and 2) are located on Hsa11 and 22, respectively, and may belong to paralogue H (Nakatani et al. 2007). Both GRK2 and GRK3 are present in the genomes of mammals and chicken. However, more extensive analyses are required to confirm their duplication before the radiation of gnathostomes.

(f) G-protein-coupled receptor kinases

GRKs are Ser/Thr kinases that phosphorylate cytoplasmic parts of G-protein-coupled receptors and reduce their signalling to G proteins. The human genome contains seven GRK genes with GRK1 (previously called rhodopsin kinase, RHOK) predominantly expressed in rods and GRK7 (previously GPRK7) in cones, albeit with some interesting species differences. GRK1 is expressed mainly in rods but also in cones in humans (Zhao et al. 1998), as in monkey (Sears et al. 2000; Weiss et al. 2001) and mouse (Lubarsky et al. 2000) retinas, but not in the cones of dog and pig retinas (Weiss et al. 2001). GRK1 is expressed exclusively in rods of the medaka fish (Hisatomi et al. 1998). In contrast, GRK7 is concentrated in cones in both cone- and rod-dominant mammals (Weiss et al. 2001) and in fish (Hisatomi et al. 1998). An intriguing exception is that the expression of GRK7 is not detectable in mouse and rat retinas, nor is the GRK7 sequence present in mouse or rat cDNA or mouse genomic libraries and deposited genomic sequences (Weiss et al. 2001), although it is present in the guinea pig and rabbit genomes. GRK1 and GRK7 are attached to the membrane by an isoprenyl moiety (Inglese et al. 1992; Zhang et al. 2004).

In our earlier phylogenetic analyses, these two retinal GRKs did not branch together although their chromosomal positions suggested that they arose from a common ancestral gene in 2R (Nordström et al. 2004). More extensive data from various vertebrate genomes now add further support for their origin in early vertebrate evolution, as shown by their presence in both teleosts and tetrapods (Wada et al. 2006). Furthermore, their chromosomal locations can now be assigned to agree with paralogon F (Nakatani et al. 2007) with extensive regions of similarity (i.e. representatives from the same gene families) on Hsa3 (GRK7) and Hsa13 (GRK1) as well as HsaX (Nordström et al. 2004) and Hsa 1, 2 and 11 (Nakatani et al. 2007; see figure 5). These chromosome regions form a paralogue that also contains the GNb and arrestin gene families (figure 5). Teleost fishes have duplicates of both GRK1 and GRK7 that seem to agree with 3R. Furthermore, there is a third member forming a basal branch for this GRK family in teleost fishes (Wada et al. 2006; Imanishi et al. 2007) that may have been lost in mammals but appears to be present in chicken (Imanishi et al. 2007) and which appears not to be duplicated in teleost-specific tetraploidization 3R. More detailed analyses of chromosomal location in different teleost genomes and in the chicken genome are required to determine if this gene is a third product of 2R in addition to GRK1 and GRK7. If so, the names of some of the members of this gene family will probably require revision to comply with orthology.

The remaining five GRKs in the human genome belong to two other paralogons. GRKs 4, 5 and 6 (Nordström et al. 2004) are located in the NPY receptor paralogue involving Hsa4, 5, 2/8 and 10 (Larsson et al. 2008), i.e. paralogue C in Nakatani et al. (2007). GRK4 was previously called GPRK2L. GRKs 2 and 3 (formerly called β-adrenergic receptor kinases 1 and 2, abbreviated ADRBK1 and 2) are located on Hsa11 and 22, respectively, and may belong to paralogue H (Nakatani et al. 2007). Both GRK2 and GRK3 are present in the genomes of mammals and chicken. However, more extensive analyses are required to confirm their duplication before the radiation of gnathostomes.

(g) Arrestins (ARR)

Arrestins bind to phosphorylated G-protein-coupled receptors to reduce or arrest their signalling and may also signal internalization of cell-surface receptors as well as induce other signal transduction pathways than the G-protein-mediated signalling (Lefkowitz &
(h) Recoverin (RCVRN)

Recoverin is a negative regulator of GRK1, i.e. RHOK (Chen 2002; Makino et al. 2004). It contains a Ca\(^{2+}\)-binding domain with similarity to calmodulin (Ames & Ikura 2002) and is attached to membrane by a myristoyl group upon Ca\(^{2+}\) binding (Desmeules et al. 2002). The number of members of the RCVRN family depends on what criteria are used to define a family. One classification encompasses recoverin itself plus the five human hippocalcin (HPCA)-like proteins, one of which is neurocalcin delta (NCALD), see the TreeFam database. Recoverin is expressed only in the retina (De Raad et al. 1995). Also NCALD is expressed in the retina in amacrine and ganglion cells, but not in rods and cones (Krishnan et al. 2004). NCALD is also expressed in many other organs (Krishnan et al. 2004). NCALD is also activated by a Ca\(^{2+}\)-myristoyl switch (Krishnan et al. 2004) that allows it to stimulate GC.

In our previous analyses we proposed that there was a local gene duplication of the ancestral hippocalcin gene before 2R, whereupon the tetraploidizations gave rise to three additional copies (Nordström et al. 2004). Recoverin, on the other hand, had no duplicates in any tetrapod in our earlier studies. The phylogenetic tree of representative sequences (fig. 9 in Nordström et al. 2004) indicated a duplication before the actinopterygian–sarcopterygian divergence, but none of the tetrapods included in the analysis possessed two gene copies. Only the pufferfish gave rise to three additional copies (Nordström et al. 2004). The third member, a local gene duplication of the ancestral hippocalcin gene quite recently. In addition, the early recoverin duplicate RCV1 has been duplicated in the teleosteos, possibly as a result of 3R. Naturally, it will be highly interesting to see which retinal cells express the RCV1 gene(s) and if there has been sub- or neo-functionalization relative to the RCVRN gene.

(i) Guanylyl cyclase-activating proteins

More distant relatives of the recoverin superfamily are the GUCAa (also called GCAP), which can be considered a separate family. We used these remote relatives to root the phylogenetic tree in our previous analysis of the recoverin superfamily (Nordström et al. 2004). The GUCA family members possess Ca\(^{2+}\)-binding domains like other relatives of the calmodulin superfamily and perform the function that their name describes, i.e. they activate GCs. Three GUCA genes have been identified in mammalian genomes. GUCA1A (GCAP1) and GUCA1B (GCAP2) activate the two GCs present in rods and cones (Palczewski et al. 2004). The third member, GUCA1C (GCAP3), was discovered in human retina (Haeseleer et al. 1999) but is a pseudogene in mouse (Imanishi et al. 2002) although it is present in platypus, chicken, the frog *X. tropicalis* and in duplicate in teleosteos (see TreeFam). The GUCA1C gene is expressed in cones in human and zebrafish (Imanishi et al. 2002). The loss of a functional GUCA1C in mouse is compatible with their reduced dependence on cone-based (colour) vision. The GUCA1A and GUCA1B genes are located adjacent to each other on Hsa6 and the GUCA1C gene is on Hsa3 and is more closely related to GUCA1A according to phylogenetic analyses (Imanishi et al. 2004), suggesting that the local duplication on Hsa6 took place before GUCA1C arose from GUCA1A. Although the phylogenies suggest duplications in early vertebrate evolution, it is not apparent that these regions of human chromosomes 3 and 6 were duplicated in 2R.

Additional GUCA gene copies exist in teleost fishes with a total of six in zebrafish and seven and eight in the pufferfishes, *T. rubripes* and *Tetraodon nigroviridis*, respectively. Some of these duplicates are likely to have arisen in 3R, but more detailed analyses of their chromosomal locations as well as sequence phylogenies are required to resolve their evolutionary relationships.

3. DISCUSSION

The retina is thought to have undergone rapid evolution at the dawn of the vertebrates with dramatic increase in the number of visual ciliary opsins from one to five (Bowmaker 2008). Our previous simplistic model for opsin gene duplications by chromosome...
duplication is elaborated here to comply with data from several additional species representing the various vertebrate classes. The scheme proposed in figure 3 is speculative and involves several gene losses, but is compatible both with the phylogenetic tree for vertebrate opsins and the chromosomal locations of the opsin genes in mammals and chicken. Note also that the transducin α subunit genes (GNAT) are located in these chromosome regions together with the GNAI genes, suggesting that GNAI and GNAT were generated from a common ancestral GNA gene by a local duplication prior to the chromosome duplications. Because a GNA that inhibits adenylyl cyclase is present also in invertebrates, it seems that GNAI has retained the ancestral properties after the gene duplication, whereas GNAT evolved coupling to PDE.

For a gene family like the opsins, which is subjected to strong selective forces, it should be expected that the evolutionary rate may be somewhat uneven and therefore give rise to phylogenetic trees that do not necessarily reflect the actual order of events for the gene duplications. Chromosomal location may therefore be a useful complement to disentangle the evolutionary history. However, data from more species, particularly actinopterygian (ray-finned) fishes and cartilaginous fishes, as well as the various pre-gnathostomes, is required before more definitive conclusions can be drawn. Indeed, preliminary analyses of teleosts indicate that the scheme proposed in figure 3 perhaps does not give the whole story.

The duplications of the opsin genes seem to have been reasonably concomitant with the diversification of retinal cell types, not only the cones and rods which express the various opsins but also many other retinal cell types. Our analyses reported previously (Nordström et al. 2004) and expanded here show that a large number of gene families besides the opsins received additional copies in the two tetraploidizations preceding the origin and radiation of the gnathostomes.

Of the 13 gene families in the phototransduction cascade investigated here, no less than 10 seem to have expanded in 2R, namely the opsins, G-protein (transducin) α, β and γ subunits (GNAT, GNB and GNGT), PDE6 α/βα′ββ′ and γ subunits (PDE6α, B and C) and γ subunits (PDE6G and H), CNG α and β subunits, G-protein-coupled receptor kinases (GRK1 and 7) and arrestins (ARR3 and SAG). Note that the CNG α and β subunits can be considered as two separate gene families because both existed before the origin of the gnathostomes (in our previous study Nordström et al. 2004) we regarded them as a single family). The remaining three gene families, GCs (GUCY), GC-activating proteins (GUCA) and recoverin (RCVRN and RCV1) also received additional members during this phase of (pre)vertebrate evolution but the gene duplicates either do not seem to be located in any of the more well-characterized paralogons (GUCY and GUCA) or have not yet been investigated in detail (RCVRN–RCV1). However, it cannot be ruled out that also these gene families expanded as a result of 2R. As an extension of the present analysis, also the GNB5 gene may be considered although its role in cones and rods may be minor relative to its functions elsewhere as indicated by its wide distribution in the brain as well as in other organs in human and mouse (Jones et al. 1998). GNB5 is distantly related to GNB1–4 and does not seem to have any surviving 2R duplicates.

Interestingly, the only retinal gene that did not seem to have been duplicated in our previous study (Nordström et al. 2004), namely recoverin, has indeed also been duplicated, but the duplicate RCV1 has been lost in placental mammals. It would be interesting to try to deduce if the loss of RCV1 coincides with loss of any cell types in the mammalian retina.

Most, if not all, of these gene families have also expanded in the teleost fish lineage through the teleost-specific tetraploidization, 3R. Because many teleosts have retained more of the original retinal complexity of the gnathostome ancestor when compared with the degenerate mammals (that have lost two of the ancestral four colour vision opsins), they may have been able to benefit from the additional 3R duplicates by evolving additional functions or abilities by either neo- or sub-functionalization. It should perhaps be expected that all these additional genes might have led to great divergence of the retinas in the various teleost fish lineage. More detailed comparative studies of, for instance, cell type-specific gene expression using the various gene family members and multiple fish species will therefore be highly interesting.

An intriguing finding was that the rhodopsin gene expressed in the rods of ray-finned fishes lacks introns, whereas an intron-containing rhodopsin gene is expressed in the pineal gland of teleost fishes. It was proposed that the intron-less rhodopsin gene arose early in ray-finned fish evolution by retrotransposition (Bellingham et al. 2003). More recent results suggest that at least the zebrafish has one additional intron-containing rhodopsin gene whose expression pattern remains to be investigated.

Some of the gene families seem to have undergone no obvious subfunctionalization of the daughter genes. The two GCs GUCY2D and GUCY2F are expressed both in cones and rods. However, the level of expression differs considerably, with form D being expressed at 25-fold higher level than F in bovine rods (Helten et al. 2007). It will be interesting to see if other vertebrate taxa have different proportions between the two forms. Another pair of 2R daughter genes shows variable regulation across mammals, namely GRK1 and GRK7, as detailed above, indicating that changes in the regulation of expression have taken place long after the duplication.

In our previous study, we noted that some of the retinal gene families might have been located in the same paralogons. However, some of the gene families seemed to have been quite far apart in their respective chromosome regions making coordinated gene regulation unlikely. While some of these expanded paralogons may have been prematurely assigned due to lack of information from species other than human, our present analyses have identified another expanded paralogon harbouring three distinct retinal gene families, namely GNB, GRK and ARR as shown an intriguing finding was that the rhodopsin gene expressed in the rods of ray-finned fishes lacks introns, whereas an intron-containing rhodopsin gene is expressed in the pineal gland of teleost fishes. It was proposed that the intron-less rhodopsin gene arose early in ray-finned fish evolution by retrotransposition (Bellingham et al. 2003). More recent results suggest that at least the zebrafish has one additional intron-containing rhodopsin gene whose expression pattern remains to be investigated.

Some of the gene families seem to have undergone no obvious subfunctionalization of the daughter genes. The two GCs GUCY2D and GUCY2F are expressed both in cones and rods. However, the level of expression differs considerably, with form D being expressed at 25-fold higher level than F in bovine rods (Helten et al. 2007). It will be interesting to see if other vertebrate taxa have different proportions between the two forms. Another pair of 2R daughter genes shows variable regulation across mammals, namely GRK1 and GRK7, as detailed above, indicating that changes in the regulation of expression have taken place long after the duplication.

In our previous study, we noted that some of the retinal gene families might have been located in the same paralogons. However, some of the gene families seemed to have been quite far apart in their respective chromosome regions making coordinated gene regulation unlikely. While some of these expanded paralogons may have been prematurely assigned due to lack of information from species other than human, our present analyses have identified another expanded paralogon harbouring three distinct retinal gene families, namely GNB, GRK and ARR as shown
in figure 5. It still seems far-fetched to assume coordinated gene regulation, particularly in consideration of the numerous intra-chromosomal rearrangements that have scrambled the order of the gene families. On the other hand, duplication by tetraploidization has been proposed to allow equilibrium to be maintained in biochemical pathways and may subsequently be followed by more gradual changes in the levels, sites or time points of expression that may result in specialization.

One intriguing possibility for specialization involves the pineal gland. Several of the genes investigated are expressed both in the retina and the pineal gland, namely GNAT1, arrestin, recoverin, RHOK and GUCY2D (Takano et al. 2003). However, it is a bit surprising to see no example, to our knowledge, of a gene family with one 2R product exclusively expressed in the pineal gland and one or more specifically expressed in the retina. The pineal opsin expressed in the chicken pineal gland, pinopsin (Downes & Gautam 1999), is the result of a duplication preceding 2R. This is compatible with the idea that the origin of the pineal gland precedes 2R because amphioxus has a putative pineal homologue, the dorsal lamellar body (Lacalli 2008). The mammalian G-protein γ subunit GN11 is expressed in the pineal and elsewhere, but not in the retina (Balcueva et al. 2000), but this is a much more recent duplicate than 2R.

**4. CONCLUSIONS**

These extended analyses of the previously described phototransduction gene families corroborate their expansion as a result of the basal vertebrate tetraploidizations, 2R. Furthermore, we have described here investigations of additional phototransduction gene families that follow the same pattern. Thus, the 2R events seem to have paved the way for the evolution of elaborate vision in the vertebrates by offering an expanded gene repertoire that could be selected for novel or more highly specialized functions. Presumably, the 3R tetraploidization in the teleost fish ancestor contributed additional genetic raw material for functional specialization in this lineage. These observations suggest that the basal vertebrate tetraploidizations may have contributed to functional specialization also in other organs.

We thank Susanne Dreborg for tentative analyses of some of the gene families and Görel Sundström and Ola Gustafsson for analyses of the opsins and their paralogs. We are grateful to Lars G. Lundin for critical reading of the manuscript. This work was supported by grants from the Swedish Research Council and the Australian Research Council (DP0880983).

**REFERENCES**


