Genomics 100 (2012) 203–211

Contents lists available at SciVerse ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Expansion of transducin subunit gene families in early vertebrate tetraploidizations

David Lagman *, Görel Sundström [1], Daniel Ocampo Daza, Xesús M. Abalo, Dan Larhammar

Department of Neuroscience, Science for Life Laboratory, Uppsala University, Box 593, SE-751 24 Uppsala, Sweden

ARTICLE INFO

Article history:
Received 17 April 2012
Accepted 6 July 2012
Available online 17 July 2012

Keywords:
Tetraploidization
Gene duplication
Phototransduction
G-protein
Transducin
Teleost fish

ABSTRACT

Hundreds of gene families expanded in the early vertebrate tetraploidizations including many gene families in the phototransduction cascade. We have investigated the evolution of the heterotrimeric G-proteins of photoreceptors, the transducins, in relation to these events using both phylogenetic analyses and synteny comparisons. Three alpha subunit genes were identified in amniotes and the coelacanth, *GNAT1–3*; two of these were identified in amphibians and teleost fish, *GNAT1* and *GNAT2*. Most tetrapods have four beta genes, *GNB1–4*, and teleosts have additional duplicates. Finally, three gamma genes were identified in mammals, *GNGT1*, *GNG11* and *GNGT2*. Of these, *GNGT1* and *GNGT2* were found in the other vertebrates. In frog and zebrafish additional duplicates of *GNGT2* were identified. Our analyses show all three transducin families expanded during the early vertebrate tetraploidizations and the beta and gamma families gained additional copies in the teleost-specific genome duplication. This suggests that the tetraploidizations contributed to visual specialisations.

1. Introduction

Many gene families have retained copies from the two rounds of tetraploidization (1R and 2R) that occurred in the early vertebrate ancestors and the third round (3R) that occurred in the lineage leading to teleost fish. Not only studies of several gene families and their chromosomal regions give support to this conclusion, such as the endothelin system [1], *HOX* gene clusters [2], neuropeptide Y (*NPY*) system [3,4] and the opioid system [5,6] but also whole genome sequence analyses [7,8]. Each group of homologous chromosome regions is called a paralogon [9].

Preliminary analyses using sequence-based phylogenies combined with chromosomal data of the transducin subunit gene families have proposed that they also expanded in these tetraploidizations [10]. Transducins are the heterotrimeric G proteins (guanine nucleotide binding proteins) in the phototransduction cascade of vertebrate visual photoreceptor cells relaying the signal from the opsins to phosphodiesterase 6 (*PDE6*) [11]. The heterotrimeric G proteins consist of three subunits named Gα, Gβ and Gγ, and the genes that encode these three subunits, named *GNA*, *GNB* and *GNG* respectively, form three unrelated gene families. Each family has multiple members in vertebrate genomes. The *GNA* and *GNG* gene families in particular consist of several subfamilies out of which we have analysed those that include members expressed in rods and cones. These two cell types express different but related genes from all three transducin subunit families [12–14].

The human genome contains sixteen *GNA*, five *GNB* and thirteen *GNG* genes. The superfamily encompassing the sixteen *GNA* genes can be subdivided into four different classes based on sequence similarities and functional specialisations: *GNAI*, *GNAS*, *GNAQ* and *GNA12/13* [15,16]. An additional class has been identified in teleost fish as well as in invertebrates, *GNAV*, giving a total of five *GNA* gene classes [15]. The *GNAI* class consists of the four families *GNAI*, *GNAZ*, *GNAO* and *GNAT*, where the latter family includes the two genes encoding the visual transducin Gα subunits [16]: *GNAT1* is specifically expressed in rods and *GNAT2* specifically in cones [17]. In addition, there is a third member of this family, *GNAT3*, also known as gustducin, expressed mainly in taste receptor cells [18].

The three genes *GNAT1–3* are located on different chromosomes and each gene is located adjacent to a *GNAI* gene [10]. Specifically, *GNAT1* is located next to *GNAI2*, *GNAT2* is close to *GNAI3* and *GNAT3* is together with *GNAI1* [12]. Several studies suggest a tandem duplication of an ancestral *GNA* gene as the origin of the ancestral *GNAT* and *GNAI* pair followed by chromosomal duplications leading to three such pairs [10,19,20].

The five genes encoding Gβ in mammalian genomes are named *GNB1*, *GNB2*, *GNB3*, *GNB4* and *GNB5* and are located on five different chromosomes in the human genome. The mammalian proteins encoded by the *GNB1–4* genes share a relatively high protein sequence identity, 80–88%, while the *GNB5* protein only shares about 50% identity with the others [17]. The *GNB1–4* genes seem to belong to a paralogon that arose in 2R, based on their chromosomal locations [10,12]. However, their sequence-based phylogenies have not been totally congruent with this conclusion [10,12]. Out of these four genes, *GNB1* is expressed in rods and *GNB3* in cones and their protein products are thus used in the transducin heterotrimers [14].

* Corresponding author. Fax: +46 18511540.
E-mail address: david.lagman@neuro.uu.se (D. Lagman).
[1] Present address: Department of Medical Biochemistry and Microbiology, Uppsala University, Box 582, SE-751 23 Uppsala, Sweden.

0888-7543/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2012.07.005

There are thirteen *GNG* genes in the human genome and the proteins encoded by these genes, Gγ, are short, only about 70 amino acids and bind tightly to the Gβ subunits [21,22]. The *GNG* genes of particular interest with regard to photoreceptor cells are the closely related *GNGT1*, *GNGT2* and *GNG11* genes [22]. *GNGT1* and *GNGT2* are expressed in rods and cones, respectively [21]. Furthermore a study on the interactions between the five beta subunits and some of the gamma subunits has revealed that $G\beta_1$ has high affinity towards $G\gamma_{t1}$ [23]. *GNG11* is expressed in a variety of tissues but outside of the retina and is probably not involved in the vertebrate visual phototransduction cascade [22]. The *GNG11* gene has only been identified in mammals and is located in tandem with *GNGT1*, suggesting a local duplication from *GNGT1* in the mammalian lineage [10].

The *GNGT1*, *GNG11* and *GNGT2* genes are located in the same genomic regions as the developmentally important *HOX* gene clusters [2,10]. These gene clusters are well studied in relation to the early vertebrate genome doublings and the subsequent duplication in the teleost lineage. Most vertebrates have four clusters (post 2R) and teleost fish at least seven (post 3R) [24,25].

Here we investigate in detail the evolution of the three transducin gene families to see if their evolution correlates with duplications in the basal vertebrate tetraploidizations and with the origin of rod and cone photoreceptor cells. Since the transducin subunits are highly conserved and particularly the gamma subunits are quite small, they give unreliable sequence-based phylogenetic trees. Therefore we have also analysed several neighbouring gene families to get more accurate phylogenies for these chromosomal regions.

2. Results

2.1. Transducin alpha subunit genes (GNAT)

Three *GNAT* genes, *GNAT1–3*, were identified in amniotes and the sarcopterygian fish included in this study, while only *GNAT1* and *GNAT2* could be identified in the teleost and amphibian genomes (Fig. 1A). BLAST [26] searches to identify putative *GNAT* orthologs in invertebrate genomes were unsuccessful. Two *GNAT* genes, named short- (GαtS) and long (GαtL) photoreceptor cell transducin alpha subunit, have previously been reported and cloned in sea lamprey [27]. These were referred to as Pma.Short (GenBank accession no. **ACB69761.1**) and Pma.Long (GenBank accession no. **ACB69760.1**) respectively in the phylogenetic trees. The sea lamprey *GNAT* amino acid sequences were included in the analysis to provide relative dating for the transducin alpha subunit divergences.

The phylogenetic maximum likelihood (PhyML) and neighbour-joining (NJ) trees of the *GNAT* family were rooted using the midpoint-root method and show that the sequences form three distinct clusters with good statistical support (Figs. 1A and S1). The sea lamprey sequences are distantly related to all other *GNAT* sequences, with basal branching points in both phylogenetic analyses. However their orthology relationships with the gnathostome sequences are unclear. Both Pma.Long and Pma.Short branches basally to *GNAT1* in the PhyML analysis (Fig. 1A), while in the NJ analysis Pma.Long clusters basally to both *GNAT1* and *GNAT3* while Pma.Short clusters with the *GNAT1* sequences (Fig. S1).

The *GNAT* gene amino acid sequences are highly conserved in size and sequence. The *GNAT1* sequences consist of 350 amino acid residues in the investigated species; the tetrapod *GNAT2* and *GNAT3* sequences each consists of 354 residues; and the teleost *GNAT2* ortholog sequences consist of 350 residues. The latter sequences lack four amino acid residues close to the amino terminus at the same positions as all *GNAT1* sequences, positions 12–15 in the *GNAT* alignment. Human and zebrafish *GNAT1* amino acid sequences share 93% identity to each other and human and zebrafish *GNAT2* share 81% identity. The specific amino acid residues forming contacts with the beta–gamma dimer [28] are all conserved throughout the *GNAT* family in all vertebrate sequences included in this study. However we could observe individual *GNAT1*, *GNAT2* and *GNAT3* subtype specific amino acid differences within the region required for beta–gamma dimer binding, positions 1–29, which overlaps with the first opsin-binding region [28], positions 7–27. In this region the *GNAT3* sequences have serine at position seven while all other sequences have alanine: *GNAT1* has glutamic acid at position

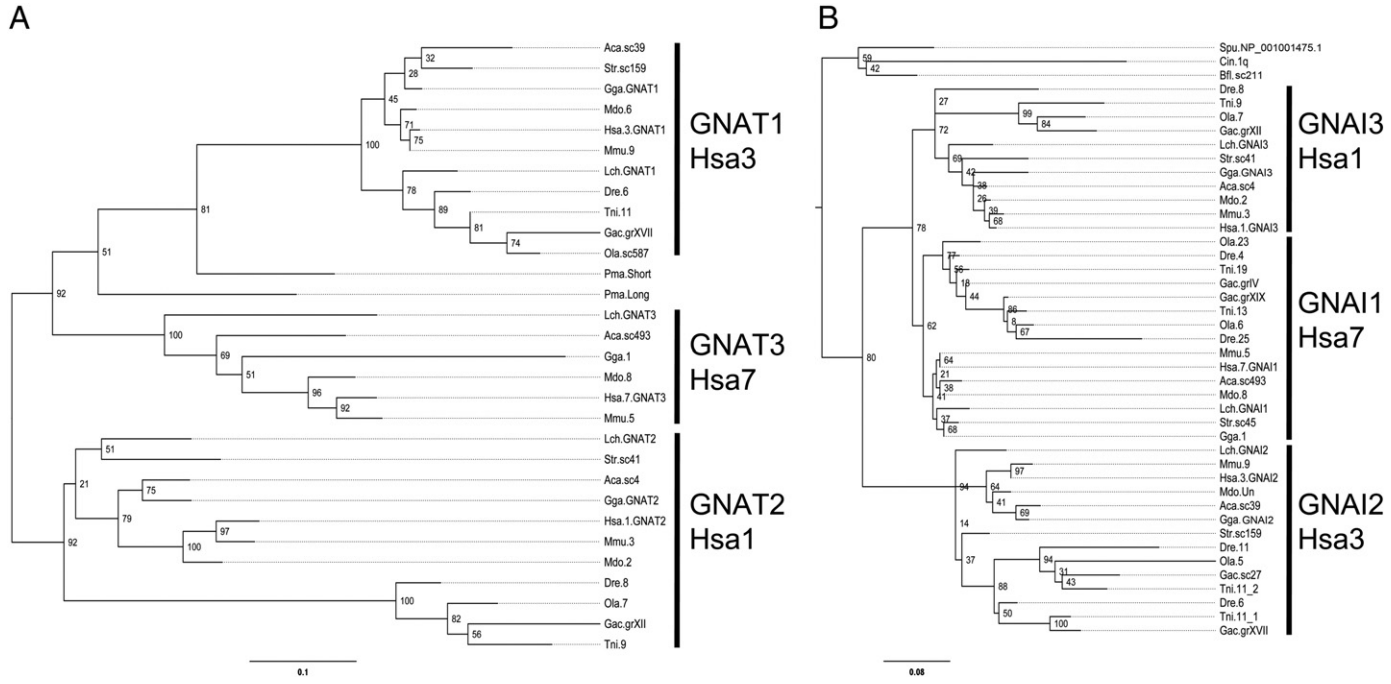

Fig. 1. Phylogenetic maximum likelihood trees of the *GNAT* and *GNAI* gene families. A) Midpoint-rooted PhyML tree with vertebrate *GNAT1*, *GNAT2* and *GNAT3* amino acid sequences. B) PhyML tree with the vertebrate *GNAI1*, *GNAI2* and *GNAI3* and invertebrate orthologous amino acid sequences. The tree is rooted with the identified fruit fly ortholog; the root is not displayed. The three letter abbreviations represent the species and the numbers or roman numerals represent the chromosome or genomic scaffold carrying the gene. Bootstrap values are shown at nodes.