

Development

Evolution of the Muscarinic Acetylcholine Receptors in Vertebrates

Julia E. Pedersen,* Christina A. Bergqvist,* and Dan Larhammar

<https://doi.org/10.1523/ENEURO.0340-18.2018>

Department of Neuroscience, Unit of Pharmacology, Science for Life Laboratory, Uppsala University, Uppsala SE-751 24, Sweden

Abstract

The family of muscarinic acetylcholine receptors (mAChRs) consists of five members in mammals, encoded by the *CHRM1-5* genes. The mAChRs are G-protein-coupled receptors, which can be divided into the following two subfamilies: M2 and M4 receptors coupling to $G_{i/o}$; and M1, M3, and M5 receptors coupling to $G_{q/11}$. However, despite the fundamental roles played by these receptors, their evolution in vertebrates has not yet been fully described. We have combined sequence-based phylogenetic analyses with comparisons of exon–intron organization and conserved synteny in order to deduce the evolution of the mAChR receptors. Our analyses verify the existence of two ancestral genes prior to the two vertebrate tetraploidizations (1R and 2R). After these events, one gene had duplicated, resulting in *CHRM2* and *CHRM4*; and the other had triplicated, forming the *CHRM1*, *CHRM3*, and *CHRM5* subfamily. All five genes are still present in all vertebrate groups investigated except the *CHRM1* gene, which has not been identified in some of the teleosts or in chicken or any other birds. Interestingly, the third tetraploidization (3R) that took place in the teleost predecessor resulted in duplicates of all five mAChR genes of which all 10 are present in zebrafish. One of the copies of the *CHRM2* and *CHRM3* genes and both *CHRM4* copies have gained introns in teleosts. Not a single separate (nontetraploidization) duplicate has been identified in any vertebrate species. These results clarify the evolution of the vertebrate mAChR family and reveal a doubled repertoire in zebrafish, inviting studies of gene neofunctionalization and subfunctionalization.

Key words: acetylcholine; G-protein-coupled receptor; gene duplication; muscarinic; tetraploidization; zebrafish

Significance Statement

Despite their pivotal physiologic role, the evolution of the muscarinic acetylcholine receptors (mAChRs) has not yet been resolved. By investigating the genomes of a broad selection of vertebrate species and combining three different types of data, namely sequence-based phylogeny, conserved synteny, and intron organization, we have deduced the evolution of the mAChR genes in relation to the major vertebrate tetraploidizations (1R, 2R, and 3R). Our analyses show that all vertebrate mAChR gene duplications resulted from the tetraploidizations. Interestingly, following 3R, zebrafish doubled its gene number, resulting in the 10 mAChR genes present. By knowing how and when the mAChR genes arose, studies of receptor subtype specialization and possible neofunctionalization or subfunctionalization can follow.

Introduction

The muscarinic acetylcholine receptors (mAChRs) are G-protein-coupled receptors (GPCRs) involved in a variety of CNS processes such as cognition, learning, and

memory. They are also present in the peripheral nervous system and smooth muscle tissue. The mAChR family consists of five different receptor subtypes named M1–M5, which are encoded by the *CHRM1-5* genes. The

Received August 27, 2018; accepted October 17, 2018; First published October 24, 2018.

The authors declare no competing financial interests.

Author contributions: J.E.P., C.A.B., and D.L. designed the research. J.E.P. and C.A.B. performed research; J.E.P., C.A.B., and D.L. analyzed data; J.E.P. and D.L. wrote the paper.

structures of the muscarinic receptors follow the typical GPCR structure with the extracellular N terminus followed by seven transmembrane (TM) domains (TM domains 1–7), which are separated by three intracellular loops (ILs; 1–3), three extracellular loops (ELs; 1–3), and finally the intracellular C terminus. The orthosteric binding site for acetylcholine consists of a hydrophobic pocket formed by the side chains of TM domains 3–7. The crystal structures of the M2 and M3 receptors have been reported (Haga et al., 2012; Kruse et al., 2012), showing that the binding pocket contains identical amino acid residues in the M2 and M3 receptors (Haga et al., 2012; Kruse et al., 2012; Tautermann et al., 2013). In the study by Haga et al. (2012), 14 amino acid residues were found to form the antagonist binding sites and, following modeling of acetylcholine into the antagonist-binding pocket, 6 of these residues were suggested to bind acetylcholine. These six proposed acetylcholine-binding residues have also been reported to be conserved in *Drosophila melanogaster* (Collin et al., 2013). In the EL regions, the amino acid residues are less conserved, hence these have been targets for the design of drugs working as allosteric modulators (Christopoulos, 2002; Kruse et al., 2013, 2014). The M1, M3, and M5 receptors form one subfamily, coupling to $G_{q/11}$, and the M2 and M4 receptors form one subfamily, coupling to $G_{i/o}$. Hence, acetylcholine may give rise to different responses depending on which receptor subtype is present to initiate the signal transduction.

The mAChRs are widely expressed in the nervous system, the cardiovascular system, and the gastrointestinal tract, as well as elsewhere. In the peripheral nervous system, the mAChRs play a major role in the parasympathetic system stimulating smooth muscle contraction and glandular secretion as well as slowing the heart rate (Eglen, 2005). In the CNS of primates and rodents, the M1, M2, and M4 receptors are the most highly expressed mAChRs in the brain, but M3 and M5 are also present (Thiele, 2013; Lebois et al., 2018). Regarding mechanisms behind gene expression and similarities or dissimilarities among vertebrate species, little is known (Lebois et al., 2018). Each gene may have multiple promoters as has been demonstrated for CHRM2 (Krejci et al., 2004), and while some promoters are conserved across mammals, others differ and presumably contribute to anatomic or temporal differences in expression between species.

Although the muscarinic receptors have prominent roles in various nervous system functions, the evolution of the mAChR gene family has not yet been fully resolved. It

is important to deduce evolutionary relationships to distinguish orthologs (species homologs), paralogs (gene duplicates), and ohnologs (gene duplicates resulting specifically from tetraploidization events), especially when studying species that belong to evolutionarily distant groups, for instance the commonly used experimental animals mouse/rat, chicken, and zebrafish. Furthermore, the time points of the gene duplication events are important for studies of evolutionary change between orthologs and paralogs as well as ohnologs. It is now well established that the vertebrate predecessor underwent two rounds of whole-genome duplication (i.e., tetraploidizations) before the radiation of jawed vertebrates (Nakatani et al., 2007; Putnam et al., 2008). These two events are usually referred to as 1R and 2R. In addition, the ancestor of the teleosts went through a third tetraploidization (3R) after the divergence from the most basal lineages of ray-finned fishes (Jaillon et al., 2004).

As the availability of high-quality genome assemblies is continuously increasing, it is now possible to perform a more extensive analysis of the evolution of the mAChR family. We have used an approach that combines amino acid sequence-based phylogeny and analyses of chromosomal locations for comparison of synteny and duplicated chromosome regions. We report here that the 1R and 2R genome-doubling events duplicated the two ancestral mAChR genes to the five genes that are present today in all tetrapods investigated except birds, where *CHRM1* has not been found. Furthermore, the teleost 3R event doubled the repertoire once more, resulting in the 10 genes present today in zebrafish, albeit some teleosts seem to lack both copies of *CHRM1*. This long-lived multiplicity invites further studies of the roles of each of the subtypes.

Materials and Methods

Species included in analysis and amino acid sequence retrieval

Species sequences included in the analysis of the mAChR family were the human (*Homo sapiens*; Hsa), mouse (*Mus musculus*; Mmu), opossum (*Monodelphis domestica*; Mdo), chicken (*Gallus gallus*; Gga), anole lizard (*Anolis carolinensis*; Aca), frog (*Xenopus tropicalis*; Xtr), coelacanth (*Latimeria chalumnae*; Lch), spotted gar (*Lepisosteus oculatus*; Loc), Japanese eel (*Anguilla japonica*; Aja), European eel (*Anguilla anguilla*; Aan), zebrafish (*Danio rerio*; Dre), stickleback (*Gasterosteus aculeatus*; Gac), medaka (*Oryzias latipes*; Ola), tunicates (*Ciona intestinalis*; Cin and *Ciona savignyi*; Csa), and nematode (*Caenorhabditis elegans*; Cel). The amino acid sequences from the species listed were retrieved from the Ensembl genome browser (release 87; Zerbino et al., 2018; Ensembl Genome Browser, RRID:SCR_013367) or NCBI (NCBI; RRID:SCR_006472) databases. If sequences were not found in either of the databases, the sequence of a closely related species was used as a query sequence in a TBLASTN search (TBLASTN; RRID:SCR_011822). The Japanese and European eel genome assemblies are not annotated. Therefore, spotted gar mAChR gene sequences were used as templates to run a TBLASTN search and retrieve mAChR orthologs present in the Jap-

This project was supported by grants from the Carl Trygger Foundation and the FACIAS Foundation.

*J.E.P. and C.A.B. contributed equally.

We thank Jan-Erik Borg for preliminary analyses at the initial stage of this project.

Correspondence should be addressed to Dan Larhammar, Department of Neuroscience, Unit of Pharmacology, Science for Life Laboratory, Box 593, Uppsala University, SE-751 24 Uppsala, Sweden. E-mail: dan.larhammar@neuro.uu.se.

<https://doi.org/10.1523/ENEURO.0340-18.2018>

Copyright © 2018 Pedersen et al.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license, which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

anese eel; thereafter, Japanese eel sequences were used as templates in TBLASTN search for mAChR orthologs present in the European eel. Due to a high degree of sequence conservation between all mAChR genes in the Japanese eel and European eel, only the European eel sequences were included in sequence-based phylogenetic analysis as this species contained a more complete mAChR gene repertoire than the Japanese eel.

Sequence alignments and phylogenetic analyses

The retrieved amino acid sequences were aligned using Jalview 2.10.3b1, with Muscle default settings (Waterhouse et al., 2009; Jalview, RRID:SCR_006459). If the amino acid sequences were aligning poorly and the predictions appeared questionable, the genomic sequences were investigated and the sequences were manually edited, by comparing sequence homology and consensus donor and acceptor splice sites. Manual corrections were made where the alignment appeared shifted due to a low degree of conservation, such as the first part of the sequence or the IL3 loop region between TM5 and TM6. However, these adjustments were kept to a minimum. All sequence details are included in Fig. 1–4, and the full alignment is available in Fig. 1–1, and Fig. 1–2. A maximum likelihood analysis was performed using the PhyML 3.0 web server (PhyML; RRID:SCR_014629; Guindon et al., 2010). The optimal substitution model was selected by the “Automatic Model Selection by SMS” option, with the Akaike information criterion. Additional settings selected were as follows: BIONJ as starting tree, Subtree-pruning-regrafting (SPR) for tree improvement, no number of random starting tree, no fast likelihood methods, and finally perform bootstrapping with 100 replicates. The resulting tree was displayed in FigTree version 1.4.2 (FigTree, RRID:SCR_008515), rooted with *C. elegans*.

Conserved synteny and paralogon analysis

For synteny and paralogon analyses, the neighboring regions of the mAChR genes were investigated in human, chicken, and spotted gar. Lists of gene families in the genomic regions 10 Mb upstream and 10 Mb downstream of the *CHRM2/CHRM4* and *CHRM1/CHRM3/CHRM5* genes, respectively, in the spotted gar genome were retrieved using the Biomart function in Ensembl version 83 (Ensembl Genome Browser; RRID:SCR_013367). For the *CHRM2/CHRM4* genes, lists of gene families were also retrieved in a similar manner in the chicken genome, due to the small number of neighboring gene families in the genomic regions in the spotted gar. Thereafter, families containing two members or more were phylogenetically analyzed by retrieving the amino acid sequences for human, chicken, coelacanth, spotted gar, and zebrafish. The region surrounding the *CHRM1*, *CHRM3*, and *CHRM5* genes contained a higher number of families; therefore, families with at least three gene family members in the spotted gar were analyzed. As outgroups tunicates, amphioxus, drosophila, *C. elegans* and in some cases other human gene sequences were included. aLRT SH-like trees were constructed using the PhyML 3.0 web server (PhyML; RRID:SCR_014629; Guindon et al., 2010) to verify the sequence orthology. To apply the most optimal selection model the “Automatic model selection by SMS” model was selected, with Akaike

information criterion. SPR was used as tree improvement method. If the members of a family showed unclear topology and/or weak node support and/or if the family showed a high degree of conservation and/or lack of outgroups and/or massive expansions due to local duplications, it was excluded from the analysis. The relatively low number of gene families in the regions surrounding the *CHRM2* and *CHRM4* genes resulted in lack of information of the fourth chromosome and its gene family members in the spotted gar. Therefore, synteny figures of the current neighboring gene repertoire in the zebrafish were prepared and included, for the paralogon structure in the ray-finned fishes. Based on the results from the phylogeny and synteny analyses, some genes appeared incorrectly named or were lacking names; therefore, genes were renamed or named according to the “ZFIN Zebrafish Nomenclature Conventions” (ZFIN; RRID:SCR_002560), and the proposed gene names were submitted to ZFIN. The details of the gene family sequences included in the analysis are provided in Fig. 2–3, for the *CHRM2* and *CHRM4* paralogon and in Fig. 3–3, for the *CHRM1*, *CHRM3*, and *CHRM5* paralogon and the aLRT SH-like trees are included in Fig. 2–1, and 3–1, respectively.

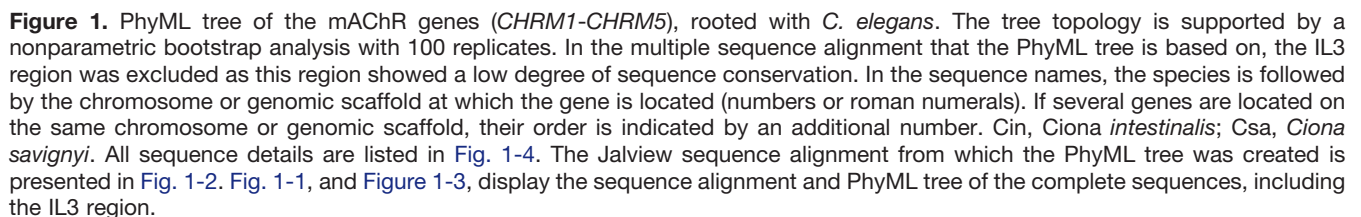
Intron position analysis in teleosts

To analyze specific intron gains in the *CHRM2b*, *CHRM3b*, *CHRM4a*, and *CHRM4b* teleost genes, additional teleost species included in the analysis were the Amazon molly (*Poecilia formosa*; Pfo) and fugu (*Takifugu rubripes*; Tru). Sequences were analyzed and aligned as described above. Intron positions were determined by manual investigation of amino acid and nucleotide sequences. Transmembrane domain regions were predicted by consulting the TMHMM Server version 2.0 (TMHMM Server; RRID:SCR_014935). For comparative analyses and confirmation of intron positions, the Japanese eel was also analyzed.

Results

Two ancestral mAChR genes expanded to five following 1R and 2R

The multiple sequence alignment analysis of the mAChR genes confirmed a generally high degree of overall sequence identity both across receptor subtypes and across species. The degree of identity for the seven TM regions between one of the most slowly evolving vertebrate model species, spotted gar, and human is ~83% for *CHRM1*, 87% for *CHRM3*, and 90% for *CHRM5*. In the other subfamily, the identity is even higher with 96% for *CHRM2* and 95% for *CHRM4*, displayed by Jalview 2.10.3b pairwise alignment. Overall, *CHRM2* displays the highest degree of conservation, whereas the *CHRM1* displays the lowest. The *CHRM2/CHRM4* subfamily displays a higher degree of conservation than the *CHRM1/CHRM3/CHRM5* subfamily. This is also confirmed in human–chicken ortholog comparisons and human paralog comparisons. However, when including the complete amino acid sequences in the pairwise alignment, the percentage of identity drops considerably. For instance, the well conserved *CHRM2* subtype decreases from 96% to 75% identity when including the complete sequence. One



The phylogenetic maximum likelihood (PhyML) tree of the predicted mAChR proteins is displayed in [Figure 1](#). The tree is rooted with a sequence from the nematode *C. elegans*. After the split between protostomes and deuterostomes, the primordial gene was duplicated in the deuterostome lineage to form the two ancestral mAChR genes present in the vertebrate predecessor, later to form the two mAChR subfamilies. The closest relatives of these two vertebrate subfamilies are two groups of tunicate sequences. The ancestor of the *CHRM2/CHRM4* subfamily duplicated in the 1R-2R tetraploidizations, as shown by paralogon comparisons described below and also sup-

The *CHRM1/CHRM3/CHRM5* subfamily has a few tunicate sequences as its closest relatives (Fig. 1), whereas the *CHRM2/CHRM4* subfamily does not. Instead, there

are two groups of tunicate sequences present basally to both of the vertebrate subfamilies. However, the bootstrap values show that there is weak node support for the positioning of the tunicate sequences in the PhyML tree, and some of them have also evolved very fast, as shown by their long branches; hence, their positioning in the PhyML may not mirror the actual phylogeny.

The positioning of the tunicates basally to the *CHRM1/CHRM3/CHRM5* subfamily shows that the expansion of this subfamily occurred after the divergence of the vertebrates and the invertebrate chordates (here represented by tunicates), a period that coincides with the timing of the 1R and 2R events. This is further supported by the species distribution of these three subtypes. Although a tunicate group is missing for the *CHRM2/CHRM4* subfamily, the timing of the duplication events, as well as the species distribution of the two subtypes, coincides with the duplication events in the *CHRM1/CHRM3/CHRM5* subfamily, supporting the 1R and 2R expansion hypothesis also for this subfamily. Hence, from this phylogenetic analysis we conclude that the mAChR family most likely expanded from two ancestral members present before 1R and 2R, to five members present following the vertebrate tetraploidizations. Additionally, duplicates found in the group of teleosts coincide with the timing of the teleost-specific tetraploidization, hence suggesting that those gene duplicates are paralogs resulting from 3R and can thus be called ohnologs (see below). The PhyML tree resulting from the analysis of the complete multiple sequence alignment, also including the IL3 region, is shown in Fig. 1-3. Details about all sequences included in the analysis are listed in Fig. 1-4.

Analysis of synteny blocks confirms expansion of the mAChR family by 1R and 2R

To explore the hypothesis that two ancestral mAChR genes expanded to five as a result of the basal vertebrate tetraploidizations, analyses of the mAChR neighboring genomic regions were conducted. If the members of each of the two mAChR subfamilies are located in chromosomal regions that also contain representatives from several other gene families, this would strongly indicate that a large block of genes, or even a chromosomal region or an entire chromosome, had been duplicated. The related genes resulting from these events are named ohnologs, as described in the Introduction. On the other hand, if the members of an mAChR subfamily are in completely different chromosomal neighborhoods, this would indicate independent duplications of an mAChR gene and insertion into unrelated chromosomal regions. Related chromosomal regions that arose as a result of the 1R and 2R tetraploidizations (or any other tetraploidization event) are referred to as comprising a paralogon (Coulier et al., 2000). Thus, the vertebrate ancestral genome consisted of paralogons with quartets of related regions, whereas the teleost ancestor had paralogons with eight related members as a result of 3R. In extant species, the paralogons have often secondarily lost some of the ohnologs. Our phylogenetic analysis of the mAChR family showed that the expansion of the mAChR gene family coincides

with the time period of the tetraploidizations. We therefore analyzed the neighboring gene families to see whether these too expanded during this time period and also whether they have representatives in the other chromosomal regions of the same paralogon.

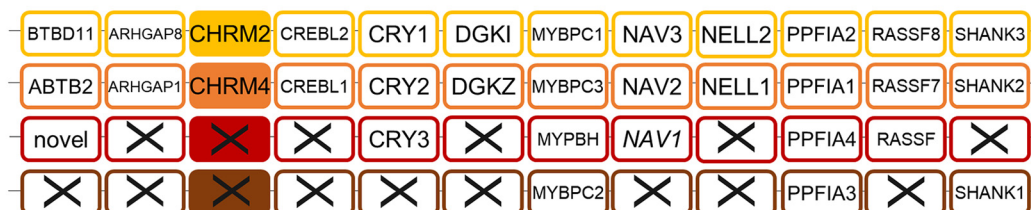
The genomic regions surrounding the *CHRM2/CHRM4* genes in the spotted gar were retrieved, and the gene families with at least two members present were analyzed. Following the exclusion criteria at the preliminary analysis stage stated in the Materials and Methods, five gene families were included in the final analysis, namely *ARHGAP*, *NAV*, *NELL*, *PPFIA*, and *SHANK*. However, due to the low number of gene families found in spotted gar, the genomic regions surrounding the *CHRM2/CHRM4* genes were investigated in the chicken and six additional gene families were identified. The reason why they were not found in the neighboring regions in spotted gar could be chromosomal rearrangements. The additional neighboring families are *ABTB2*, *CREBL*, *CRY* (for sequence details and phylogenetic analysis, see Haug et al., 2015), *DGK*, *MYBP*, and *RASSF*. Information about the neighboring gene families is included in Fig. 2-3, and their phylogenetic trees [aLRT (approximate likelihood-ratio test) SH (Shimodaira–Hasegawa)-like trees] are included in Fig. 2-1. From the phylogenetic analyses, orthologous and paralogous genes were determined, and the chromosomal locations of the neighboring gene families in human, chicken, and spotted gar are presented in Figure 2. In humans, the first member of this paralogon (Fig. 2, yellow) consists of regions located on three separate chromosomes (chromosomes 7, 22, and 12). In chicken and spotted gar, the orthologs are located on a single chromosome (chromosome 1 in the chicken and LG8 in spotted gar). This strongly suggests that this paralogon member in the human genome most likely was split by chromosomal translocations. The second paralogon member (Fig. 2, orange) is located on a single chromosome in human and chicken, and with one exception also in spotted gar. Also the third paralogon member is restricted to a single chromosome for these gene families in all three species. Note that the *MYBPHL* and *MYBPH* genes in human are a result of a more recent local duplication (Fig. 2, red). This paralogon member has undergone more ohnolog losses than the other two members. The fourth paralogon member (Fig. 2, brown) seems to have undergone even more ohnolog losses and is only present in the human genome, with members from the *MYBP*, *PPFIA*, and *SHANK* families (on chromosome 19).

To investigate whether this paralogon member is present in other species, the neighboring gene family repertoire and chromosomal locations were investigated in zebrafish, presented in Fig. 2-2. The fourth paralogon member was indeed found to be present in zebrafish, represented by *MYBP*, *PPFIA*, and *SHANK* ohnologs (on chromosomes 3 and 24). However, it appears that additional translocations have occurred in zebrafish, most likely following the 3R event. These genes are located in a paralogon that has been studied previously, where the focus was on a region containing numerous other gene families including the visual opsins, transducin alpha sub-

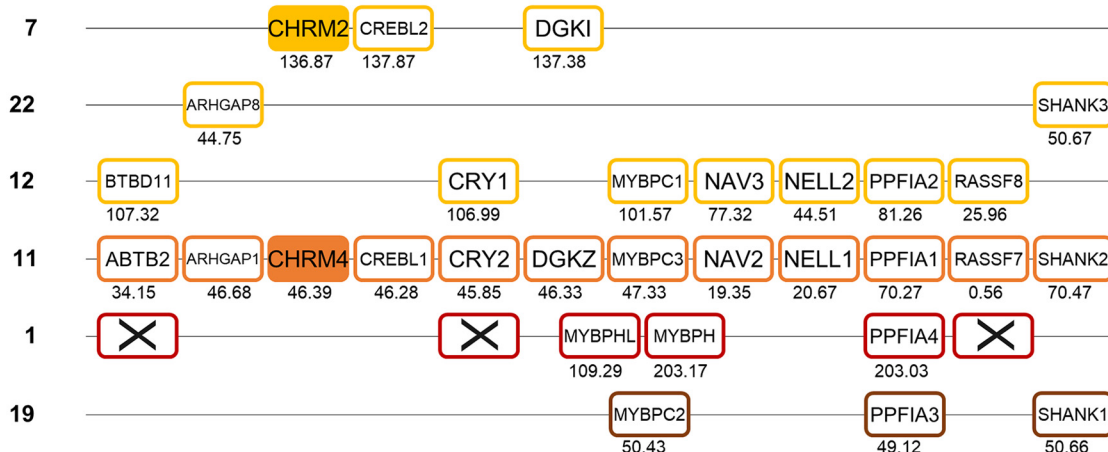
Vertebrate predecessor prior to 2R

ABTB2 ARHGAP CHRM CREBL CRY DGK MYBP NAV NELL PPFIA RASSF SHANK

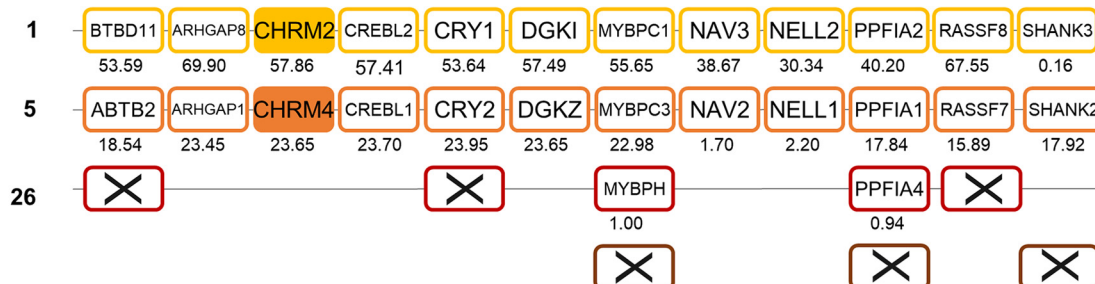
1R+2R ↓



Human



Chicken



Spotted gar

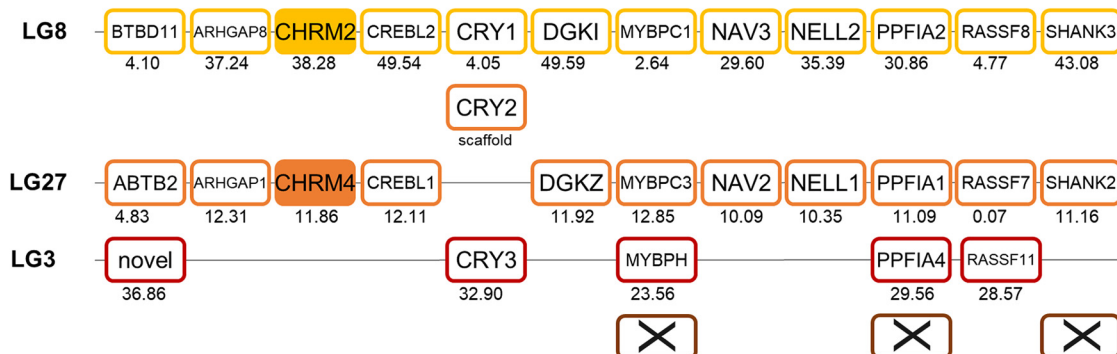


Figure 2. The evolutionary history and analysis of chromosomal regions and conserved synteny of the *CHRM2* and *CHRM4* genes and their

continued

neighboring gene families. The gene repertoire present in the vertebrate predecessor is displayed in the top panel, the duplication scheme further displays which orthologs were retained in the vertebrate ancestor following 1R and 2R, and finally the last three panels display the gene repertoire present in the human, chicken, and spotted gar. Crosses indicate gene loss or gene not (yet) identified. Dashed boxes represent incomplete sequences. Each paralogon member is presented in a separate color. Chicken and spotted gar illustrations are reused with permission from Daniel Ocampo Daza (source: www.egosumdaniel.se). The sequence details are listed in Fig. 2-3. The aLRT SH-like trees are displayed in Fig. 2-1, and the chromosomal regions and conserved synteny of the *CHRM2* and *CHRM4* genes and their neighboring gene families in zebrafish are displayed in Fig. 2-2.

units and oxytocin/vasopressin receptors (Lagman et al., 2013). Therefore, although the third and fourth members of this region have undergone several ohnolog losses, it is nevertheless clear that the *CHRM2* and *CHRM4* genes are located in a paralogon that arose in the basal vertebrate tetraploidizations.

The *CHRM1/CHRM3/CHRM5* subfamily arose from a separate ancestor gene. Our investigation of the paralogon hypothesis was initiated by retrieving the genomic regions surrounding the *CHRM1/CHRM3/CHRM5* genes in spotted gar. These contained a higher number of gene families than the regions surrounding the *CHRM2/CHRM4* genes and therefore the analysis was restricted to gene families with three or four members. Following the exclusion criteria stated in the Materials and Methods, 15 neighboring gene families were included in the final analysis, namely *ATL*, *EHD*, *FERMT*, *JAG*, *LTBP*, *MERTK*, *NRXN*, *PLD*, *PRKD*, *PROX*, *PRPH2*, *PYG*, *SLC24A*, *SPTB*, and *TGFB*. Information about the neighboring gene families is included in Fig. 3-3, and the phylogenetic trees (aLRT SH like) are included in Figure 3-1. Based on the phylogenetic analyses, the chromosomal locations of the gene family members are shown for human, chicken, and spotted gar in Figure 3.

The first paralogon member (Fig. 3, green) is located on a single chromosome in human and spotted gar (chromosome 11 and LG28, respectively). Three of the genes located on LG28 (*LTBP3*, *PROX*, and *SPTBN2*; Fig. 3, dashed boxes) were incomplete in the genome database and contain <50% of the sequence. As this might impact the topology in the aLRT SH-like trees, trees were also generated where these sequences were excluded; nevertheless, the results remained the same (Fig. 3-1). None of the genes of the first paralogon member are present in chicken, except the *SPTB* family member *SPTBN2* (located on a scaffold). As mentioned previously, the *CHRM1* gene is absent in the chicken, as well as in other birds. The second and third paralogon members in human are located on three and two different chromosomes, respectively (Fig. 3, turquoise and light blue). The second member is located on two different chromosomes in spotted gar (LG1 and LG16). In chicken, both the second and the third paralogon members are located on a single chromosome. Finally, the fourth paralogon member is located on chromosome 19 in human and on LG2 in spotted gar (Fig. 3, dark blue). In chicken, the fourth member could not be identified for four of the gene families, and three ohnologs are located on scaffolds (*EHD2*, *PLD3*, and *SPTBN4*). However, one member of the *TGFB* family, the *TGFB1* gene, is located on chromosome 32, which is a very short chromosome in the chicken, consisting of only ~78 kb.

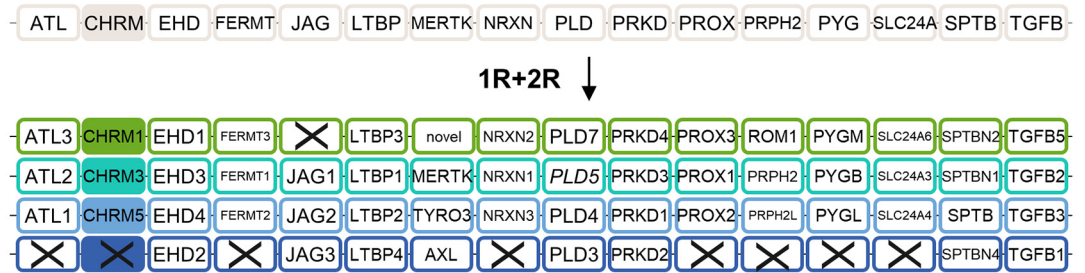
The synteny analysis in human, chicken, and spotted gar shows that these genomic regions have undergone a number of rearrangements such as translocations, and several ohnologs could not be identified. This is further seen when analyzing the *CHRM1/CHRM3/CHRM5* neighboring gene repertoire and chromosomal locations in zebrafish, presented in Fig. 3-2. It appears that a number of translocations have occurred in zebrafish, as for instance the first paralogon member (Fig. 3-2, green) is located on five different chromosomes, and the second member is located on no less than seven different chromosomes (Fig. 3-2, turquoise). This paralogon too has been studied in detail in a previous study focusing on somatostatin receptor genes including multiple neighboring gene families. These chromosomal regions were found to be related through the 1R, 2R, and 3R events and thereby comprise a paralogon (Ocampo Daza et al., 2012).

Thus, the analysis of conserved synteny and paralogons of the two mAChR subfamilies and their neighboring gene families confirms our hypothesis based on the phylogenetic analyses that the mAChR gene family expanded by the 1R and 2R tetraploidizations from two ancestral genes into the five members present today in most tetrapods and some basally diverging vertebrates, and 8-10 members in teleosts, forming two distinct subfamilies belonging to two separate paralogons.

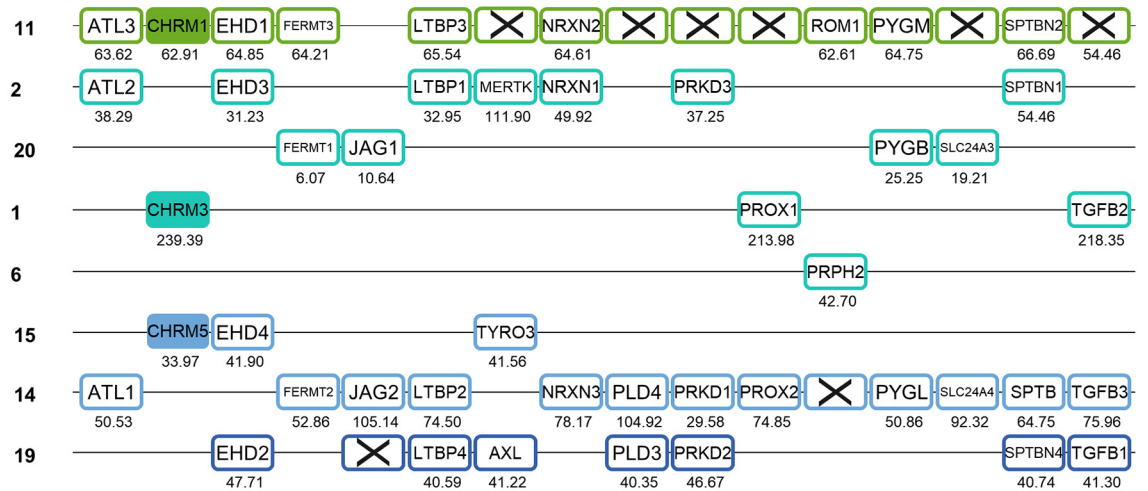
Teleost-specific intron gains in the *CHRM2b*, *CHRM3b*, *CHRM4a*, and *CHRM4b* genes

The amino acid sequence analyses and alignments revealed that some of the teleost sequences contained annotated introns in the genome assemblies, although the mAChR genes in general have been said to lack introns in the coding region (Bonner et al., 1987, 1988; Peralta et al., 1987; Seo et al., 2009). To investigate whether these introns were indeed teleost-specific gains, or whether they could be the results of gene annotation or sequencing difficulties, an extended sequence repertoire from teleosts was analyzed. In the sequence analyses, it was found that the *CHRM2b*, *CHRM3b*, *CHRM4a*, and *CHRM4b* genes have independently gained at least one intron in the proximity of the region encoding TM1 and at least one intron in the region encoding the IL between TM5 and TM6, in at least one of the teleost species investigated (Fig. 4). In *CHRM2b*, stickleback and medaka have gained one intron in the end of the TM1 domain (Fig. 4). This intron gain is supported by analysis of fugu and Amazon molly *CHRM2b* sequences, as they too contain this intron (Fig. 4-1). One intron is also present in the IL3 domain, located between TM5 and TM6, in zebrafish, stickleback, and medaka. However, it seems that this intron is not the same in the three teleost species.

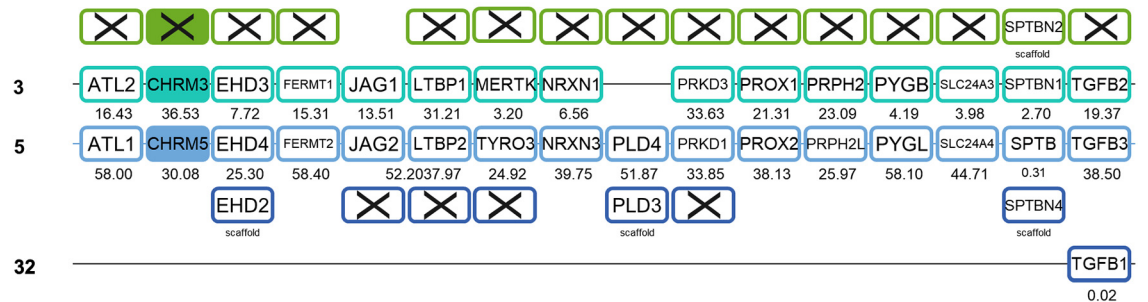
Vertebrate predecessor prior to 2R



Human



Chicken



Spotted gar

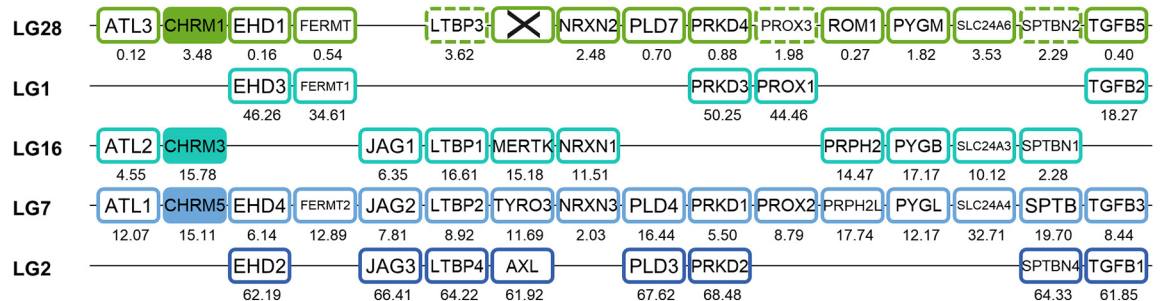


Figure 3. The evolutionary history and analysis of chromosomal regions and conserved synteny of the *CHRM1*, *CHRM3*, and *CHRM5* genes and their neighboring gene families. The gene repertoire present in the vertebrate predecessor is displayed in the top panel, the duplication scheme further displays which orthologs were retained in the vertebrate ancestor following 1R and 2R, and finally the last three panels display the gene repertoire present in the human, chicken, and spotted gar. Crosses indicate gene loss or gene not (yet) identified. Each paralogon member is presented in a separate color. Chicken and spotted gar illustrations are reused with

continued

permission from Daniel Ocampo Daza (source: www.egosumdaniel.se). The sequence details are listed in Fig. 3-3. The aLRT SH-like trees are displayed in Figure 3-1, and the chromosomal regions and conserved synteny of the *CHRM2* and *CHRM4* genes and their neighboring gene families in zebrafish are displayed in Figure 3-2.

Rather, one intron seems to have been gained in zebrafish and a separate intron was gained in the ancestor of stickleback and medaka (Fig. 4). No *CHRM2b* gene could be identified in European eel (or in the Japanese eel), and therefore it was not possible to determine the exact time point when this intron was gained in zebrafish. The intron present in stickleback and medaka was also found in fugu and Amazon molly (Extended Data Fig. 4-1).

The *CHRM3b* gene has gained one intron located in the beginning of the region encoding TM1 in medaka and stickleback and one additional intron in the N-terminal region in stickleback (Fig. 4). However, the first exon could not be identified in stickleback, although the presence of an intron at this position is supported by an identical

intron found in fugu, which is most likely present also in Amazon molly (Extended Data Fig. 4-2). No exons upstream of the intron in the beginning of TM1 in medaka *CHRM3b* could be identified, and therefore it was not possible to confirm the presence of additional introns in medaka (Extended Data Fig. 4-2).

The *CHRM4a* gene has gained the largest number of introns. It has gained one intron in the N-terminal region in the ancestor of medaka and stickleback (Fig. 4). The first exon could not be found in stickleback, but there is a suitable consensus splice acceptor site present at the position corresponding to the intron present in medaka. This possible splice site is also present in fugu and Amazon molly (Extended Data Fig. 4-3). Another intron is

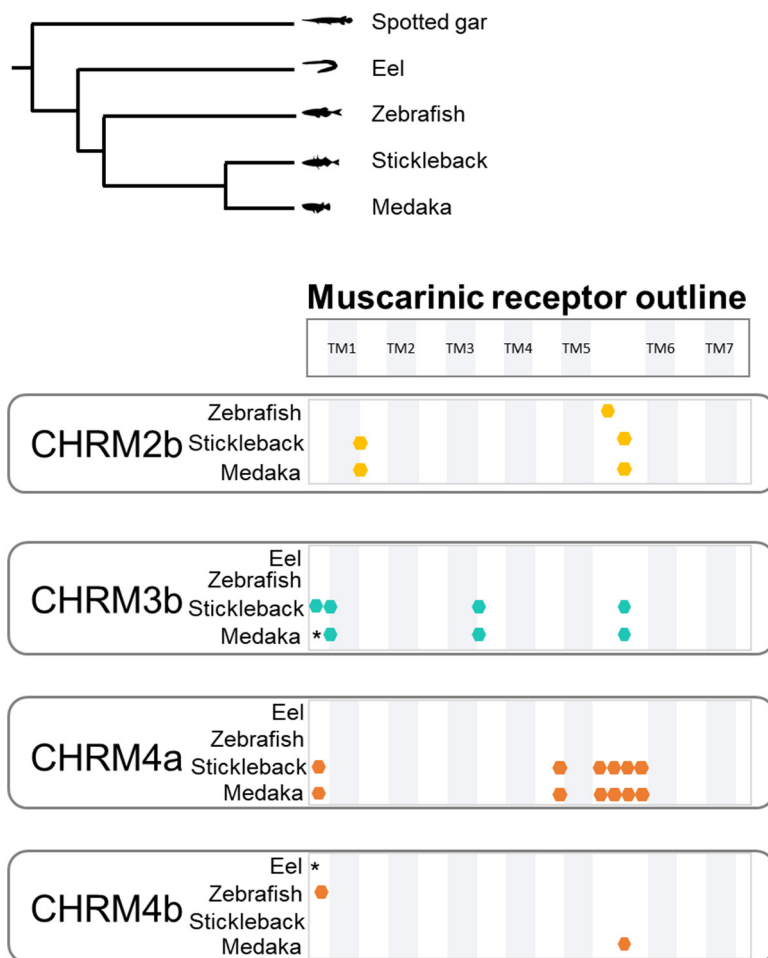


Figure 4. The localization of teleost-specific intron gains for the *CHRM2b*, *CHRM3b*, *CHRM4a*, and *CHRM4b* genes in the European eel, zebrafish, stickleback, and medaka. The top panel displays the relationship between the teleost species included in intron analysis, with the spotted gar as reference species followed by the mAChR outline and specific intron gains (indicated by colored hexagon) for the *CHRM2b*, *CHRM3b*, and *CHRM4a*. No *CHRM2b* sequence was identified in the European eel. Asterisk is present where an intron gain could not be confirmed. The sequence details are listed in Fig. 4-5. The Jalview sequence alignments of the teleost sequences analyzed are displayed for the *CHRM2b*, *CHRM3b*, *CHRM4a*, and *CHRM4b* genes in Fig. 4-1, Fig. 4-2, Fig. 4-3, and Fig. 4-4, respectively.

present in the region encoding EL2, just before TM5, in stickleback and medaka (Fig. 4), as well as in fugu and Amazon molly (Extended Data Fig. 4-3). Finally, four introns have been gained in the large region encoding IL3 of *CHRM4a* in the ancestor of stickleback and medaka (Fig. 4) and is also present in fugu and Amazon molly (Extended Data Fig. 4-3). Among these four genes, *CHRM4b* is the one that has gained the lowest number of introns. There is one intron present in the N-terminal region in zebrafish (Fig. 4). This intron is not found in any of the other teleosts analyzed. However, there is one possible intron present in European eel, but before this position there is also a methionine present that could act as translation initiator, meaning that this intron may not be present in European eel (Extended Data Fig. 4-4). However, the methionine is not present in Japanese eel, which implies that there should be an intron present at this position. Due to these inconsistencies between European eel and Japanese eel, it is not possible to conclude whether or not there is an intron present in the N-terminal region in these species. There is also one intron gained in IL3 in *CHRM4b* in medaka (Fig. 4), an intron that is also present in Amazon molly (Fig. 4). Information about the teleost sequences included in this analysis is provided in Extended Data Fig. 4-5.

Discussion

Our analyses of the mAChR gene family are based on the following three types of information: sequence-based phylogenetic analysis; synteny and paralogon analysis; as well as analysis of teleost-specific intron gains. The combined results of these analyses show that the mAChR family expanded from two ancestral genes present in the vertebrate predecessor to five mAChR genes in an early vertebrate ancestor, as a result of the 1R and 2R tetraploidization events (Fig. 5). All five members could be identified in the vertebrate classes of mammals, birds, reptiles, amphibians, and bony fishes, with the exception of the *CHRM1* gene, which, surprisingly, has not been identified in chicken or any other bird. It remains possible that the gene exists and is located on a microchromosome, because it is well known that these are under-represented in the genome sequencing projects, probably partly due to their extremely high GC content (Burt, 2002; Zhang et al., 2014).

We also identified all five mAChR genes in a ray-finned fish, the spotted gar, which represents an early branch in the ray-finned fish tree. The teleosts, which constitute 99.9% of all ray-finned fishes, are descendants of a lineage that underwent a third tetraploidization, and for the mAChR family all 10 genes deriving from this event have been retained in zebrafish. The phylogenetic analysis shows that the teleost-specific tetraploidization 3R resulted in duplicates of all mAChR genes in zebrafish, resulting in a total of 10 mAChR genes (Fig. 5). A doubled mAChR repertoire in zebrafish has been reported before (Seo et al., 2009; Nuckels et al., 2011); however, neither of the previous studies tied it to the 3R tetraploidization. Here we can explain all of these duplications by a single genomic event (and previous gene duplications by the 1R/2R events). The European eel has retained nine of the

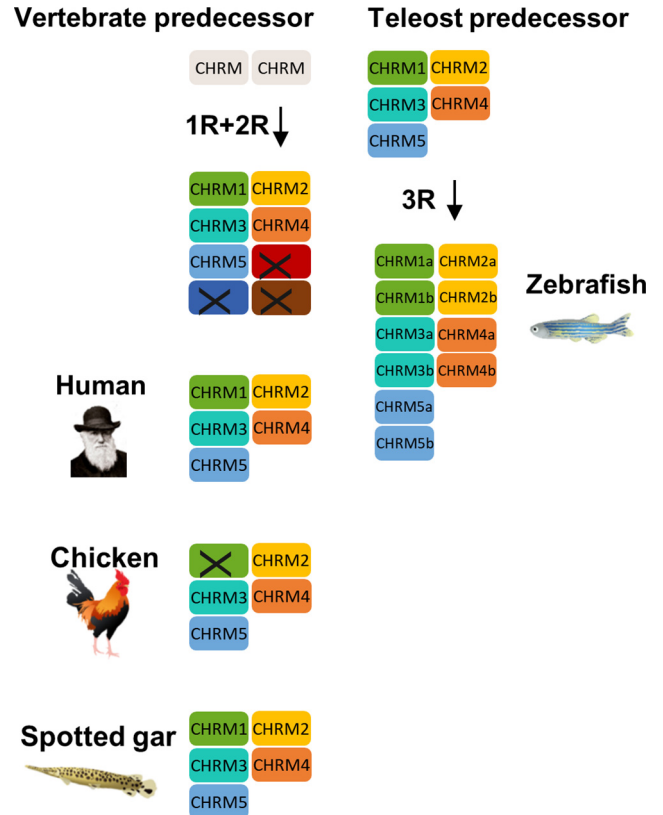


Figure 5. Duplication scheme of the mAChR genes following 1R, 2R, and 3R. Two mAChR genes present in the vertebrate predecessor expanded to five mAChR genes following 1R and 2R. All genes were retained in the human, chicken, and spotted gar except for the *CHRM1* gene in the chicken. In the teleost predecessor, 5 mAChR genes expanded to 10 genes following 3R, of which all duplicates are retained in the zebrafish. Chicken, spotted gar, and zebrafish illustrations are reused with permission from Daniel Ocampo Daza (source: www.egosumdaniel.se).

genes, lacking one of the *CHRM2* duplicates, whereas medaka and stickleback are lacking the *CHRM1* gene. Notably, no nontetraploidization duplicates of any of the mAChR genes was found in any of the vertebrate species analyzed.

The *CHRM2/CHRM4* subfamily contains two ohnologs resulting from the 1R-2R tetraploidizations. It is unclear whether these two genes arose in 1R, and both of their duplicates after 2R were lost, or whether one copy was lost after 1R and the other was duplicated in 2R. The *CHRM1/CHRM3/CHRM5* subfamily contains three of the ohnologs resulting from the two tetraploidizations. This means that the ancestral gene duplicated once in 1R, and then both copies duplicated once more in 2R, after which one ohnolog was lost. As the two tetraploidizations were probably very close in time to one another, it is difficult to say which two may be the results of the 2R tetraploidization.

Thus, the repertoire of mAChR genes is quite consistent across vertebrates. This is presumably a reflection of unique functional roles for each of the gene products. The only gene that deviates from this pattern is *CHRM1*. Not only does it seem to be missing in birds, it has also not been identified in a few teleosts, namely stickleback and

medaka. The pairwise alignments of the mAChR sequences showed that the *CHRM1* gene displayed the lowest degree of sequence identity; hence, it is more likely that this gene could be lost, or that the low degree of conservation has impeded its identification in the species where it has so far not been found. In fact, the whole paralogon member is missing in chicken, but there are still two possible reasons for this: either this whole chromosomal region was lost; or the whole region ended up on a microchromosome that is as yet unsequenced. In fact, one gene in this paralogon member (*SPTBN2*) has been identified, it is on a scaffold, perhaps indicating that additional members may be possible to identify.

Interestingly, a more thorough analysis of the amino acid sequences and especially the IL3 region, which has a low degree of sequence conservation, revealed that there has been a number of teleost specific introns gained in the coding regions of the *CHRM2b*, *CHRM3b*, *CHRM4a*, and *CHRM4b* genes. A previous study by Seo et al. (2009), which focused on a subset of mAChRs and smooth muscle contraction responses in Nile tilapia reported that all five mAChR genes present had retained paralogs in zebrafish, resulting in 10 mAChR genes present in total. The study by Seo et al. (2009) also reported that no introns were present in the mAChR genes studied. However, with the increased availability of data, and especially genome assemblies, our analyses have identified and verified a number of intron gains in several teleosts (Fig. 4, Fig. 4-1, Fig. 4-2, Fig. 4-3, and Fig. 4-4). Our findings are supported by those of a previous study reporting that the intron turnover in Actinopterygii is high, especially in the stickleback and zebrafish (Venkatesh et al., 2014; Ravi and Venkatesh, 2018). However, to our knowledge intron gains have not been previously reported for the mAChR genes in teleosts. The intron gains have taken place in teleost genes that evolve faster than their orthologs in other lineages, especially *CHRM2b*, *CHRM3b*, and *CHRM4b* (Fig. 1 and Fig. 1-3).

Duplication by chromosome doubling means by default that the two ohnologs deriving from the same mother gene must initially have had identical gene regulatory elements and hence identical expression patterns, anatomically, temporally, and quantitatively. This may initially result in an additive effect on the level of gene expression, unless compensatory mechanisms are at play. Subsequently, it is possible that either or both of the ohnologs may begin to accumulate mutations, either regulatory or structural, that will alter the functions of the gene. One possibility is that one of the ohnologs maintains the functions of the mother gene, leaving the other free to take on other roles (i.e., neofunctionalization). This was a possibility favored by Ohno (1970). Alternatively, the two ohnologs may lose regulatory elements such that they subdivide the functions of the mother gene between them in a process called subfunctionalization. As the mAChR genes were doubled in zebrafish and all duplicates are retained, it would be interesting to study possible neofunctionalizations or subfunctionalizations of these genes, initially by investigating gene expression patterns in anatomic mapping studies.

References

- Bonner TI, Buckley NJ, Young AC, Brann MR (1987) Identification of a family of muscarinic acetylcholine receptor genes. *Science* 237:527–532. [Medline](#)
- Bonner TI, Young AC, Brann MR, Buckley NJ (1988) Cloning and expression of the human and rat m5 muscarinic acetylcholine receptor genes. *Neuron* 1:403–410. [Medline](#)
- Burt DW (2002) Origin and evolution of avian microchromosomes. *Cytogenet Genome Res* 96:97–112. [CrossRef Medline](#)
- Christopoulos A (2002) Allosteric binding sites on cell-surface receptors: novel targets for drug discovery. *Nat Rev Drug Discov* 1:198–210. [CrossRef](#)
- Collin C, Hauser F, de Valdivia EG, Li S, Reisenberger J, Carlsen EMM, Khan Z, Hansen NØ, Puhm F, Søndergaard L, Niemiec J, Heninger M, Ren GR, Grimmelikhuijzen CJP (2013) Two types of muscarinic acetylcholine receptors in *Drosophila* and other arthropods. *Cell Mol Life Sci* 70:3231–3242. [CrossRef](#)
- Coulier F, Popovici C, Villet R, Birnbaum D (2000) MetaHox gene clusters. *J Exp Zool* 288:345–351. [CrossRef Medline](#)
- Eglen RM (2005) Muscarinic receptor subtype pharmacology and physiology. In: *Progress in medicinal chemistry* (King FD, Lawton G, eds), pp 105–136. Amsterdam: Elsevier.
- Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59:307–321. [CrossRef Medline](#)
- Haga K, Kruse AC, Asada H, Yurugi-Kobayashi T, Shiroishi M, Zhang C, Weis WI, Okada T, Kobilka BK, Haga T, Kobayashi T (2012) Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist. *Nature* 482:547–551. [CrossRef](#)
- Haug MF, Gesemann M, Lazović V, Neuhauss SCF (2015) Eumetazoan cryptochrome phylogeny and evolution. *Genome Biol Evol* 7:601–619. [CrossRef Medline](#)
- Jaillon O, et al. (2004) Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* 431:946–957. [CrossRef Medline](#)
- Krejci A, Bruce AW, Dolezal V, Tucek S, Buckley NJ (2004) Multiple promoters drive tissue-specific expression of the human M muscarinic acetylcholine receptor gene. *J Neurochem* 91:88–98. [CrossRef Medline](#)
- Kruse AC, Hu J, Pan AC, Arlow DH, Rosenbaum DM, Rosemond E, Green HF, Liu T, Chae PS, Dror RO, Shaw DE, Weis WI, Wess J, Kobilka BK (2012) Structure and dynamics of the M3 muscarinic acetylcholine receptor. *Nature* 482:552–556. [CrossRef Medline](#)
- Kruse AC, Ring AM, Manglik A, Hu J, Hu K, Eitel K, Hübner H, Pardon E, Valant C, Sexton PM, Christopoulos A, Felder CC, Gmeiner P, Steyaert J, Weis WI, Garcia KC, Wess J, Kobilka BK (2013) Activation and allosteric modulation of a muscarinic acetylcholine receptor. *Nature* 504:101–106. [CrossRef Medline](#)
- Kruse AC, Kobilka BK, Gautam D, Sexton PM, Christopoulos A, Wess J (2014) Muscarinic acetylcholine receptors: novel opportunities for drug development. *Nat Rev Drug Discov* 13:549–560. [CrossRef Medline](#)
- 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. [CrossRef Medline](#)
- Lebois EP, Thorn C, Edgerton JR, Popielek M, Xi S (2018) Muscarinic receptor subtype distribution in the central nervous system and relevance to aging and Alzheimer's disease. *Neuropharmacology* 136:362–373. [CrossRef](#)
- Nakatani Y, Takeda H, Kohara Y, Morishita S (2007) Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates. *Genome Res* 17:1254–1265. [CrossRef Medline](#)

- Nuckels RJ, Forstner MRJ, Capalbo-Pitts EL, García DM (2011) Developmental expression of muscarinic receptors in the eyes of zebrafish. *Brain Res* 1405:85–94. [CrossRef](#)
- Ocampo Daza D, Sundström G, Bergqvist CA, Larhammar D (2012) The evolution of vertebrate somatostatin receptors and their gene regions involves extensive chromosomal rearrangements. *BMC Evol Biol* 12:231. [CrossRef](#) [Medline](#)
- Ohno S (1970) *Evolution by gene duplication*. Berlin, Heidelberg: Springer.
- Peralta EG, Ashkenazi A, Winslow JW, Smith DH, Ramachandran J, Capon DJ (1987) Distinct primary structures, ligand-binding properties and tissue-specific expression of four human muscarinic acetylcholine receptors. *EMBO J* 6:3923–3929. [Medline](#)
- Putnam NH, et al. (2008) The amphioxus genome and the evolution of the chordate karyotype. *Nature* 453:1064–1071. [CrossRef](#) [Medline](#)
- Ravi V, Venkatesh B (2018) The divergent genomes of teleosts. *Annu Rev Anim Biosci* 6:47–68. [CrossRef](#) [Medline](#)
- Seo JS, Kim M-S, Park EM, Ahn SJ, Kim NY, Jung SH, Kim JW, Lee HH, Chung JK (2009) Cloning and characterization of muscarinic receptor genes from the Nile tilapia (*Oreochromis niloticus*). *Mol Cells* 27:383–390. [CrossRef](#) [Medline](#)
- Tautermann CS, Kiechle T, Seeliger D, Diehl S, Wex E, Banholzer R, Gantner F, Pieper MP, Casarosa P (2013) Molecular basis for the long duration of action and kinetic selectivity of tiotropium for the muscarinic M3 receptor. *J Med Chem* 56:8746–8756. [CrossRef](#) [Medline](#)
- Thiele A (2013) Muscarinic signaling in the brain. *Annu Rev Neurosci* 36:271–294. [CrossRef](#) [Medline](#)
- Venkatesh B, Lee AP, Ravi V, Maurya AK, Lian MM, Swann JB, Ohta Y, Flajnik MF, Sutoh Y, Kasahara M, Hoon S, Gangu V, Roy SW, Irimia M, Korzh V, Kondrychyn I, Lim ZW, Tay BH, Tohari S, Kong KW, et al. (2014) Elephant shark genome provides unique insights into gnathostome evolution. *Nature* 505:174–179. [CrossRef](#) [Medline](#)
- Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ (2009) Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25:1189–1191. [CrossRef](#)
- Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, Billis K, Cummins C, Gall A, Girón CG, Gil L, Gordon L, Haggerty L, Haskell E, Hourlier T, Izuogu OG, Janacek SH, Juettemann T, To JK, Laird MR, et al. (2018) Ensembl 2018. *Nucleic Acids Res* 46:D754–D761. [CrossRef](#) [Medline](#)
- Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, Storz JF, Antunes A, Greenwold MJ, Meredith RW, Ödeen A, Cui J, Zhou Q, Xu L, Pan H, Wang Z, Jin L, Zhang P, Hu H, Yang W, et al. (2014) Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* 346:1311–1320. [CrossRef](#) [Medline](#)