

Evolution of the receptors for growth hormone, prolactin, erythropoietin and thrombopoietin in relation to the vertebrate tetraploidizations

Daniel **Ocampo Daza**^{1,*} and Dan **Larhammar**

Department of Neuroscience, Science for Life Laboratory, Uppsala University, Box 593, SE-75124 Uppsala, Sweden.

* Corresponding author. E-mail address: Daniel.Ocampo.Daza@ebc.uu.se

¹ Present address: Subdepartment of Evolution and Development, Department of Organismal Biology, Uppsala University, Norbyvägen 18A, SE-752 36 Uppsala, Sweden.

General and Comparative Endocrinology 257 (2018) 143-160
<https://doi.org/10.1016/j.ygcen.2017.06.021>

Received 29 January 2017

Revised 16 June 2017

Accepted 22 June 2017

Available online 23 June 2017

Figshare

<http://dx.doi.org/10.6084/m9.figshare.5151157>

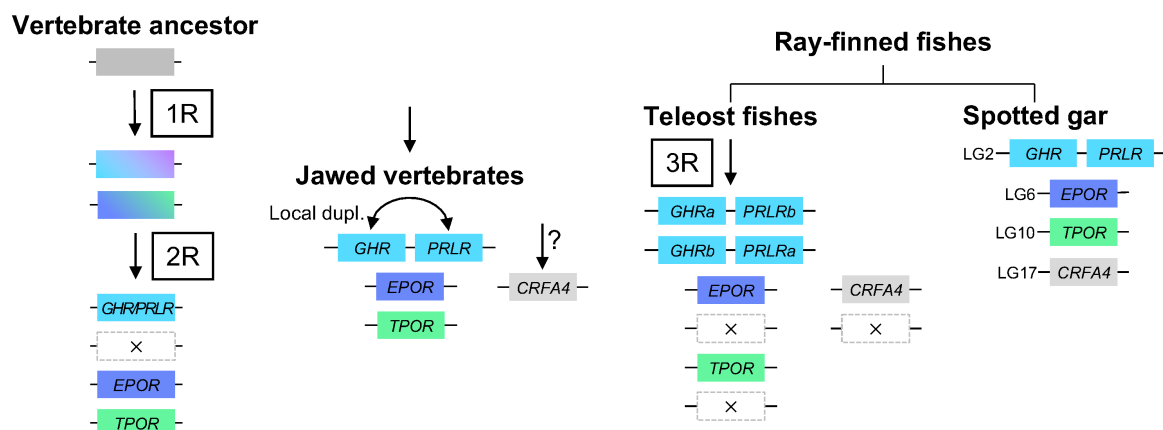
This document corresponds to the final published version of this article. You are free to download, print and distribute this document for any purposes under a Creative Commons Attribution 4.0 Unported License (<http://creativecommons.org/licenses/by/4.0/>), provided the original work is cited as specified.

Cite this article as: Ocampo Daza D, Larhammar D. Evolution of the receptors for growth hormone, prolactin, erythropoietin and thrombopoietin in relation to the vertebrate tetraploidizations, *General and Comparative Endocrinology* 257 (2018) 143-160. doi: <http://dx.doi.org/10.1016/j.ygcen.2017.06.021>

Abstract

The receptors for the pituitary hormones growth hormone (GH), prolactin (PRL) and somatolactin (SL), and the hematopoietic hormones erythropoietin (EPO) and thrombopoietin (TPO), comprise a structurally related family in the superfamily of cytokine class-I receptors. GH, PRL and SL receptors have a wide variety of effects in development, osmoregulation, metabolism and stimulation of growth, while EPO and TPO receptors guide the production and differentiation of erythrocytes and thrombocytes, respectively. The evolution of the receptors for GH, PRL and SL has been partially investigated by previous reports suggesting different time points for the hormone and receptor gene duplications. This raises questions about how hormone-receptor partnerships have emerged and evolved. Therefore, we have investigated in detail the expansion of this receptor family, especially in relation to the basal vertebrate (1R, 2R) and teleost (3R) tetraploidizations. Receptor family genes were identified in a broad range of vertebrate genomes and investigated using a combination of sequence-based phylogenetic analyses and comparative genomic analyses of synteny. We found that 1R most likely generated *EPOR/TPOR* and *GHR/PRLR* ancestors; following this, 2R resulted in *EPOR* and *TPOR* genes. No *GHR/PRLR* duplicate seems to have survived after 2R. Instead the single *GHR/PRLR* underwent a local duplication sometime after 2R, generating separate syntenic genes for GHR and PRLR. Subsequently, 3R duplicated the gene pair in teleosts, resulting in two GHR and two PRLR genes, but no *EPOR* or *TPOR* duplicates. These analyses help illuminate the evolution of the regulatory mechanisms for somatic growth, metabolism, osmoregulation and hematopoiesis in vertebrates.

Graphical abstract



Keywords

Growth hormone receptor, Prolactin receptor, Somatolactin receptor, Gene duplication, Molecular Evolution, Phylogeny

1. Introduction

The pituitary hormones that belong to the growth hormone family display an extraordinary diversity of physiological roles. Growth hormone itself (GH) is primarily known for its anabolic effects, especially growth stimulation of muscle, bone and cartilage, but it also influences numerous other metabolic processes. Prolactin (PRL) was named due to its stimulation of milk production in mammals, but the hormone arose in evolution long before mammary glands and has important roles in processes involving growth, electrolyte balance and seasonal regulation of reproduction, to name only a few (Grattan and Kokay, 2008). Somatolactin (SL) was discovered in teleost fish (Ono et al., 1990; Rand-Weaver et al., 1991) and has also been found in a lobe-finned fish, a lungfish (Amemiya et al., 1999), however it seems that the hormone was subsequently lost in the tetrapod lineage. Its functions in teleost fish are only partially known and include pigmentation (Fukamachi et al., 2009), electrolyte balance (Uchida et al., 2009) and lipid metabolism (Sasano et al., 2012).

The cell-surface receptors that respond to the growth hormone family of hormones belong to the superfamily of class I cytokine receptors (Huising et al., 2006). Together with the receptors for the hematopoietic hormones erythropoietin and thrombopoietin they constitute the subgroup of single-chain class I cytokine receptors that form functional receptor complexes through homodimerization (Boulay et al., 2003; Liongue and Ward, 2007), although heterodimers may also form in some species (Forsyth and Wallis, 2002). The primary mechanism of signal transduction is the activation of Janus kinases (JAKs) that leads to stimulation of STAT mechanisms (signal transducers and activators of transcription), but other enzymes may also be stimulated, such as the MAPK/ERK pathway (Freeman et al., 2000). The receptors that respond to the three pituitary hormones, growth hormone, prolactin and somatolactin, have primarily been characterized in mammals, chicken and teleost fishes. They are more or less selective for each of the three hormones and have been classified as either growth hormone receptors (GHR) or prolactin receptors (PRLR) based on preferential ligand binding and phylogenetic analyses (Fukamachi et al., 2005). Receptor duplicates have been reported in various species of teleost fishes for GHR (Calduch-Giner et al., 2001; Fukamachi et al., 2005; Gao et al., 2011; Jiao et al., 2006; Li et al., 2007; Ma et al., 2012; Pierce et al., 2012; Saera-Vila et al., 2005; Tse et al., 2003; Very et al., 2005) and PRLR (Fiol et al., 2009; Huang et al., 2007; Noh et al., 2012). The duplicate genes for GHR are likely a result of the teleost-specific tetraploidization, also known as 3R (Fukamachi and Meyer, 2007). A receptor with preferential binding of somatolactin was first reported from the masu salmon (*Oncorhynchus masou*) (Fukada et al., 2005). Subsequently homologous receptors were described from several other teleost species (Benedet et al., 2008; Fukamachi et al., 2005) and found to be orthologous to one of the GHR duplicates (Fukamachi and Meyer, 2007), suggesting that the specific SL-SLR hormone-receptor pairing originated in teleosts. However, somatolactin exists in both ray-finned fishes (*Actinopterygii*) and lobe-finned fishes (*Sarcopterygii*) (Amemiya et al., 1999), which suggests an earlier hormone-receptor pairing. Furthermore, the functional distinction between GHR and SLR is blurred in teleosts due to differing ligand selectivities between species for seemingly orthologous receptors (Chen et al., 2011).

We investigate here the evolution of the single-chain group of class I cytokine receptors by combining phylogenetic analyses and chromosomal studies of conserved synteny in a wide

selection of vertebrate species. Many receptor gene families have expanded through the two rounds of tetraploidization that took place in early vertebrate evolution (1R and 2R) (Nakatani et al., 2007; Putnam et al., 2008) as well as the teleost-specific tetraploidization (3R) (Glasauer and Neuhauss, 2014). While previous phylogenetic analyses point to a clear distinction between PRLR on the one hand and GHR on the other (Fukamachi et al., 2005), the time point for this basal divergence and the possible involvement of ancient vertebrate tetraploidizations have not been fully investigated. Furthermore, the unclear origin of SLR and the confusing nomenclature of the duplicate GHRs in teleost fishes has been debated several times in the literature (Benedet et al., 2008; Ellens et al., 2013; Fukamachi and Meyer, 2007; Walock et al., 2014). During the course of this work it became apparent that the closely related erythropoietin and thrombopoietin receptors (EPOR and TPOR) also needed to be taken into consideration in order to grasp the entire evolutionary scenario of this receptor family. The studies of conserved synteny also yielded insights into the evolution of several salient neighboring gene families, including the fibroblast growth factors 3, 7, 10 and 22, and the family of prostaglandin E₂ receptor 4, where we describe novel members.

The conclusions presented herein have important implications for studies of the functional specializations, hormone-receptor interactions and hormone selectivities of growth hormone, prolactin, somatolactin, erythropoietin and thrombopoietin receptors, as well as for our understanding of the origin and evolution of the fundamental functions that these receptors carry out in vertebrates.

2. Materials and methods

2.1. Database searches and sequence annotation

Amino acid sequences of *GHR*, *PRLR*, *CRFA4*, *EPOR* and *TPOR (MPL)*, as well as neighboring gene family gene predictions, were identified in the Ensembl genome browser (<http://www.ensembl.org>) using gene orthology and protein family predictions (Fernández-Suárez and Schuster, 2010). The corresponding amino acid sequences were collected and the genomic locations and database identifiers of all sequences were noted. Any gene predictions that were conspicuously missing from Ensembl's automatic orthology predictions were sought using TBLASTN (Altschul et al., 1990). All sequences and locations have been verified against Ensembl release 86 (October 2016) (Yates et al., 2016). For the human, chicken and zebrafish gene predictions, sequence identifiers and locations refer to previous genome assemblies GRCh37, Galgal4 and Zv9 respectively, accessed through <http://grch37.ensembl.org>. Complete information about the identified sequences, including database identifiers, location data and genome assembly versions is provided in Supplementary data 1 and 2. The following species were investigated through the Ensembl genome browser (common names in parenthesis): *Homo sapiens* (human), *Mus musculus* (mouse), *Monodelphis domestica* (Grey short-tailed opossum), *Gallus gallus* (chicken), *Anolis carolinensis* (Carolina anole lizard), *Xenopus (Silurana) tropicalis* (Western clawed frog), *Latimeria chalumnae* (coelacanth), *Lepisosteus oculatus* (spotted gar), *Danio rerio* (zebrafish), *Astyanax mexicanus* (Mexican cave tetra), *Gadus morhua* (Atlantic cod), *Oryzias latipes* (medaka), *Xiphophorus maculatus* (Southern platyfish), *Gasterosteus aculeatus*

(three-spined stickleback), *Tetraodon nigroviridis* (green spotted pufferfish), *Takifugu rubripes* (Japanese pufferfish, fugu), *Petromyzon marinus* (Sea lamprey), *Ciona intestinalis* (vase tunicate) and *Drosophila melanogaster* (fruit fly). For the phylogenetic analyses of neighboring gene families, a smaller selection of species was used (Supplementary data 2).

Additional *GHR*, *PRLR*, *CRFA4*, *EPOR* and *TPOR* gene predictions from the genomes of *Clupea harengus* (Atlantic herring), *Salmo salar* (Atlantic salmon), *Callorhynchus milii* (elephant shark), *Rhincodon typus* (whale shark), *Lethenteron camtschaticum* (Arctic lamprey), *Saccoglossus kowalevskii* (acorn worm) and *Branchiostoma floridae* (Florida lancelet) were sought through the National Center for Biotechnology Information (NCBI) genome resource (<https://www.ncbi.nlm.nih.gov/genome>). These gene predictions were identified through TBLASTN searches using human, coelacanth or zebrafish amino acid sequences as queries. For all such searches, hits were considered informative if expect values (E-values) were lower than $1e^{-30}$. cDNA sequences from a *Leucoraja erinacea* (little skate) transcriptome (NCBI BioProject: 361222) (Bellono et al., 2017) were sought in the same way using elephant shark and coelacanth sequences as queries.

Additional invertebrate sequences from Florida lancelet and *Strongylocentrotus purpuratus* (purple sea urchin) were sought in reference proteome databases through Hidden Markov Model searches (HMMER) using the HMMER web server (<http://hmmer.janelia.org>) (Finn et al., 2011).

Short, fragmented or otherwise faulty predictions in the Ensembl databases (Prosdocimi et al., 2012) as well as the NCBI Genome resource were curated manually from the genomic sequence following consensus for gene initiation and splice donor and acceptor sites as well as sequence homology to other family members.

2.2. Sequence alignments and phylogenetic analyses

Alignments were made using the MUSCLE algorithm (Edgar, 2004) applied through AliView 1.18 with default settings (Larsson, 2014). Alignments were inspected and edited manually in AliView in order to curate wrongly predicted sequences as described above and adjust poorly aligned sequence stretches. The main alignment of single-chain cytokine class I receptor sequences was edited to remove duplicated ligand-binding domain in some *PRLR* and *TPOR* sequences (see Section 3.3 in Results).

Phylogenetic analyses were carried out using the likelihood-based Phylogenetic Maximum Likelihood method (PhyML) (Guindon et al., 2010). PhyML trees were made using the PhyML 3.0 algorithm through Seaview 4.5.0 (Gouy et al., 2010) with the following settings: The LG model of amino acid substitution (Le and Gascuel, 2008) was assumed. Approximate Likelihood Ratio Tests (aLRT) with SH-like supports (Anisimova et al., 2011; Anisimova and Gascuel, 2006) were chosen for statistical support of the tree topologies. Amino acid (equilibrium) frequencies were estimated from the alignments (empirical), and the proportion of invariable sites and across-site substitution rate variation (gamma-shape) parameters were optimized. For the latter parameter, the number of substitution rate categories was increased from 4 to 8. The starting tree was estimated using BIONJ with optimized tree topology. This option optimizes the tree topology as well as branch lengths

of the starting tree. Both NNI and SPR tree improvement methods were considered to estimate the best tree topology. When possible, trees were rooted with the identified invertebrate family members as out-groups. Vase tunicate or lancelet sequences were used to provide a relative dating point for the time window of the basal vertebrate tetraploidizations (1R and 2R), and spotted gar sequences were used as a relative dating point for the basal teleost tetraploidization (3R). Where no out-group could be found, trees are displayed as midpoint-rooted phylograms (Hess and De Moraes Russo, 2007).

2.3. Identification of neighboring gene families and conserved syntenic blocks

In our first selection, neighboring gene families were defined as Ensembl protein family predictions (Fernández-Suárez and Schuster, 2010) that have members closer than 5 MB to *at least two* of the *GHR*, *PRLR*, *EPOR* and *MPL* (*TPOR*) genes in the human genome (assembly GRCh37). These chromosome regions were investigated by downloading and comparing lists of gene predictions from the Ensembl genome browser using the BioMart tool (Supplementary data 3). Out of necessity, additional neighboring gene families were identified in the medaka genome to better investigate the involvement of the basal teleost tetraploidization (3R). The medaka genome is an appropriate starting point for the analysis of conserved syntenic blocks in teleost fishes since it has preserved more of the ancestral teleost genome organization (Kasahara et al., 2007). In this second selection, Ensembl protein family predictions with members closer than 5 MB to *at least two* of the medaka *GHRa*, *GHRb*, *PRLRa* and *PRLRb* genes, as well as the human *GHR* and *PRLR* genes were considered (Supplementary data 3).

3. Results

3.1. Nomenclature

In human, mouse, chicken and zebrafish, thrombopoietin receptor genes have received the name *MPL* for “*myeloproliferative leukemia proto-oncogene*”. When referring to these genes in particular their approved gene symbols will be used; when referring to thrombopoietin genes in general the symbol *TPOR* will be used. For teleost genes, excluding the zebrafish *mpl* specifically, the symbols *GHRa*, *GHRb*, *PRLRa*, *PRLRb*, *EPOR* and *TPOR* will be used. In several teleost fishes, some GHR-type receptors have been named somatolactin receptors (SLR), including in the Atlantic salmon and medaka; however, we will use the symbols *GHRa1* and *GHRa* respectively (see Section 3.2 below).

3.2. Single-chain cytokine class I receptor sequences in vertebrate genomes

Our phylogeny of the single-chain cytokine class I receptors is presented in Fig. 1. We could identify gene predictions for growth hormone receptor (*GHR*), prolactin receptor (*PRLR*), erythropoietin receptor (*EPOR*) and thrombopoietin receptor (*TPOR*) in all ray-finned fishes (*Actinopterygii*), including spotted gar and teleost fishes, and all lobe-finned fishes (*Sarcopterygii*), including coelacanth and tetrapods, that were analyzed. In cartilaginous fishes (*Chondrichthyes*), we could only identify *GHR*, *EPOR* and *TPOR* sequences, which suggests that *PRLR* was lost in this lineage. As a whole our phylogeny shows that *GHR*, *PRLR*,

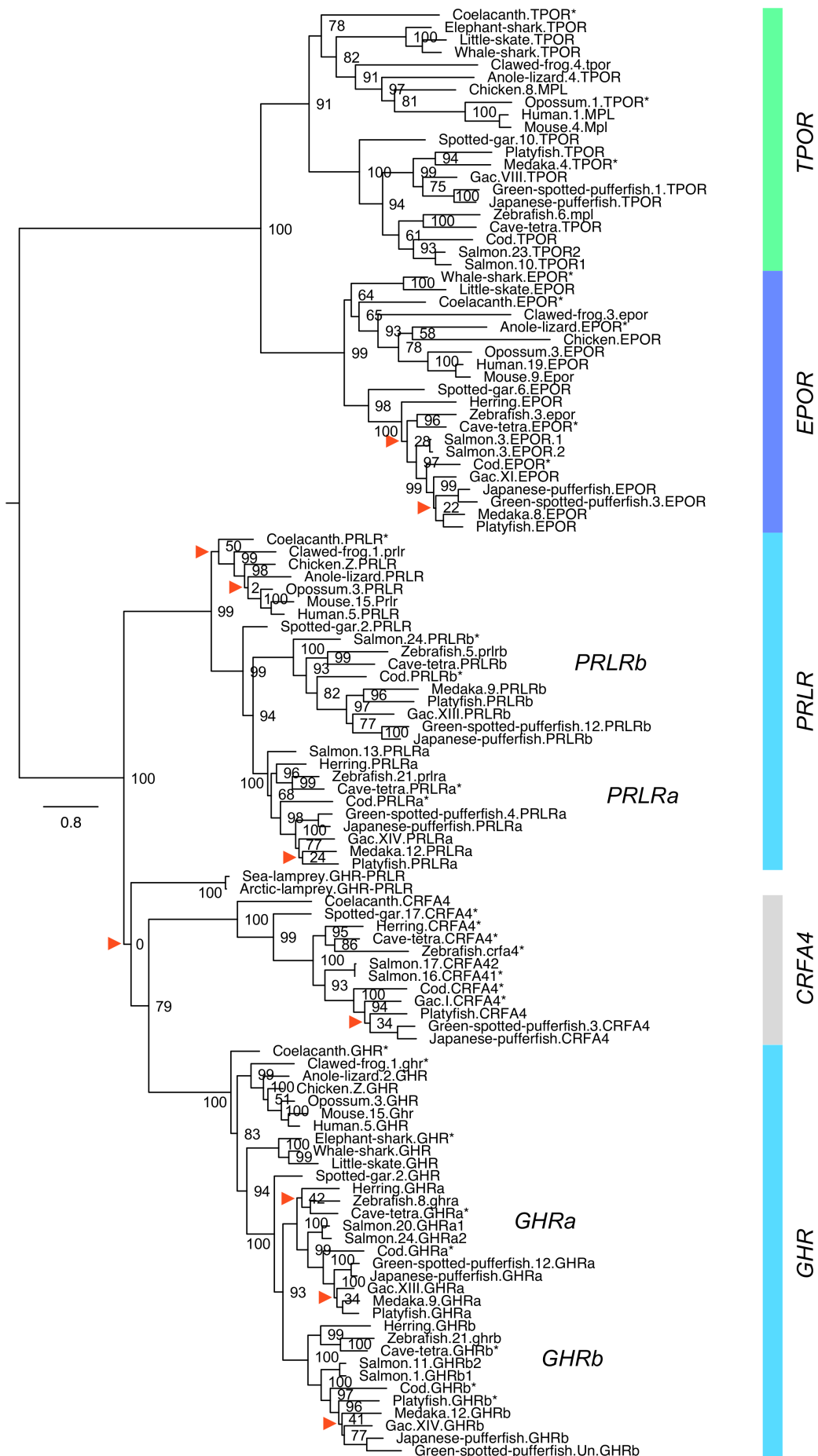


Fig. 1. Phylogenetic maximum likelihood tree of the single chain class I cytokine receptors. Approved gene symbols and nomenclature guidelines for human, mouse, chicken, Western clawed frog and zebrafish genes were used. Some species names are abbreviated for clarity, see Section 2.1 for a complete species list. Numbers following species names refer to chromosome or linkage group assignments for mapped genes. Asterisks denote partial sequences that do not span the full length of the alignment. The tree is supported by an approximate Likelihood Ratio Test (aLRT) with SH-like node supports (shown in percent). Nodes with support values $\leq 50\%$ were considered uninformative and are marked with arrowheads. For clarity, some support values for shallow nodes are not shown. The tree is displayed as a midpoint-rooted phylogram.

EPOR and *TPOR* genes were present at the base of jawed vertebrates (*Gnathostomata*). These sequences form four well-supported clades in our phylogeny, with *EPOR* and *TPOR* forming one cluster, and *GHR*, *PRLR* and a clade of *GHR/PRLR*-like sequences called *CRFA4* forming a second cluster. The *CRFA4* sequences were an unexpected finding, identified during complementary BLAST searches for missing *GHR* and *PRLR* sequences. Due to their relatively high sequence identity and similar gene structure to *GHR* and *PRLR*, they were also included in our analyses. In our phylogeny, the *CRFA4* sequences cluster with the *GHR* clade, with good support (Fig. 1). We could identify *CRFA4* sequences in all ray-finned fishes as well as the coelacanth, but not in any tetrapod nor cartilaginous fish species. Our phylogeny also includes the truncated *GHR/PRLR*-like sequence that has been described previously in the sea lamprey (Ellens et al., 2013). We could identify this sequence in the genomes of both sea lamprey and Arctic lamprey. These lamprey sequences cluster at the base of the larger *GHR*, *PRLR* and *CRFA4* clade, seemingly supporting the suggestion that they represent a common ancestor of both *GHR* and *PRLR* (Ellens et al., 2013). However, the lack of support for this branch in our phylogeny precludes any clear conclusion in this regard.

For the most part, our phylogeny follows the accepted vertebrate taxonomy. Due to the low overall sequence identity and uneven evolutionary rates within this family, especially for the intracellular domain (see Section 3.3 below), some inconsistencies are to be expected. Notably, the coelacanth sequences occupy basal positions within their respective clades, which could be an indication of their relatively slow evolutionary rate. The cartilaginous fish branches also diverge markedly from the accepted taxonomy within each clade.

All teleost species were found to have duplicate *GHR* and *PRLR* genes, and all but the Atlantic salmon were found to have single *EPOR* and *TPOR* genes. For both *GHR* and *PRLR*, the duplicate teleost sequences form well-supported distinct clades diverging after the respective spotted gar ortholog branches (Fig. 1). This is consistent with gene duplications through the basal teleost tetraploidization (3R). Among the species we have analyzed, the Atlantic salmon *GHRa1* and medaka *GHRa* sequences have been called somatolactin receptors (SLR) in the literature (Benedet et al., 2008; Fukamachi et al., 2005). We concur with more recent arguments against a separating *GHR* and *SLR* nomenclature (Ellens et al., 2013; Walock et al., 2014), because no clear-cut ligand selectivity has been demonstrated for most putative somatolactin receptors (Chen et al., 2011). We have named the teleost duplicate clades *GHRa* and *GHRb* based on the accepted zebrafish gene names and the naming convention for 3R-generated gene duplicates ("ZFIN Zebrafish Nomenclature Guidelines," 2016). In the Atlantic salmon genome, we could identify duplicate sequences for *GHRa*, *GHRb*, *CRFA4*, *EPOR* and *TPOR* (Fig. 1). The chromosomal locations of these duplicate sequences are consistent with paralogous chromosome blocks generated in the salmonid tetraploidization (4R) (see Fig. 2 in Lien et al., 2016), except for the two *EPOR*

sequences which are located on the same chromosome approximately 3.6 MB apart. Several of these genes are known by other names in Atlantic salmon and/or other salmonid species (see for instance Fig. 6 in Walock et al., 2014). We show a comparison of previous gene names and our suggested nomenclature in Supplementary data 1. For the salmonid duplicates, we have used numbers after the “a” and “b” designations, following a precedent set by the naming of duplicate salmonid insulin-like growth factor binding proteins (IGFBPs) (Macqueen et al., 2013). The local duplicates of *EPOR* have received the suffixes “.1” and “.2”, following the above cited nomenclature guidelines.

In most genomes with sufficiently high coverage to allow the assemblage of larger contigs, *GHR* and *PRLR* genes were found to be syntenic, i.e. located on the same chromosomes or linkage groups, while the *EPOR* and *TPOR* genes were found on separate chromosomes or linkage groups (Fig. 1, Supplementary data 1). In all mapped teleost genomes, except for the zebrafish and to an extent the Atlantic salmon, the duplicate *GHR* and *PRLR* genes are organized as syntenic pairs, *GHRa-PRLRb* and *GHRb-PRLRa*, on two different chromosomes or linkage groups (Fig. 1, Supplementary data 1). In the zebrafish genome, the *ghrb* and *prlra* genes are both located on chromosome 21 whereas the *ghra* and *prlrb* genes are located on separate chromosomes – 8 and 5 respectively. In the Atlantic salmon genome, the only preserved gene pair is *GHRa2* and *PRLRb* on chromosome 24. However, the pattern of 4R-generated paralogous chromosome blocks supports the ancestral linkage of the *GHRb2* locus on chromosome 11 and the *PRLRa* locus on chromosome 13 (see Fig. 2 in Lien et al. (2016)).

3.3. Structures of jawed vertebrate single-chain cytokine class I receptor genes

The structures of the human *GHR*, *PRLR*, *EPOR* and *MPL* (*TPOR*) genes are known and, excluding 5' and 3' untranslated regions, reflect the general structures of the single-chain cytokine class I receptor genes (Liongue and Ward, 2007): The first coding exon encodes a signal peptide (SP), this is followed by four exons encoding the extracellular ligand-binding domain (LBD), one exon encoding the transmembrane region (TM), and finally two exons encoding the intracellular domain (Fig. 2). This structure is also shared with all *CRFA4* genes we have analyzed. In the extracellular domain there are two conserved motifs: The TSXW motif, which is part of the hormone-binding site 1 of *GHR* and *PRLR* (see Ellens et al., 2013, and references therein), and the WSXWS motif (Y/FGE/DFS in *GHR*) that is involved in the structural changes of receptor activation (Dagil et al., 2012). The intracellular domain contains the so called “box 1” and “box 2” motifs, which are involved in Jak docking and receptor internalization. The eutherian mammal *GHR* genes include an additional exon (red in Fig. 2) that is specific to this lineage (Menzies et al., 2008). The *GHRa* genes in Atlantic cod, medaka, Southern platyfish, three-spined stickleback, Japanese pufferfish (fugu) and green spotted pufferfish have an additional 3' intron (blue in Supplementary Fig. S1), which suggests an intron insertion in a spiny-rayed fish (*Acanthomorpha*) ancestor.

The chicken *PRLR* gene is known to have two sets of LBD-encoding exons (Bu et al., 2013), which produces a “double antenna” structure in the mature receptor (Tanaka et al., 1992). We could identify the duplicated LBD-encoding unit also in the anole lizard *PRLR* sequence (Fig. 2). Several alternative transcript variants of the chicken *PRLR* gene are known, including

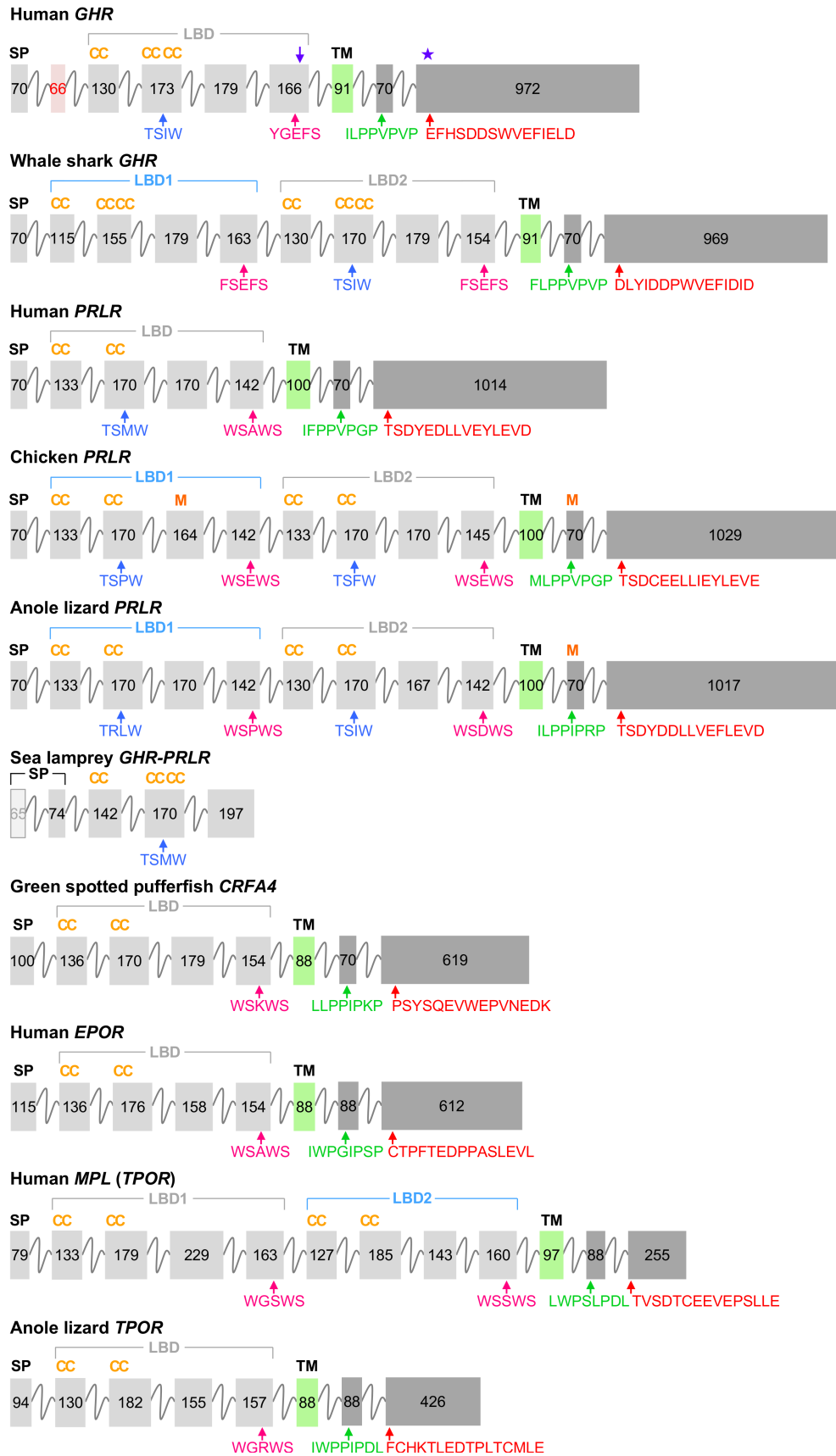


Fig. 2. Coding exon organizations of representative *GHR*, *PRLR*, *CRFA4*, *EPOR* and *TPOR* (*MPL*) genes. Exons are drawn to scale. Exon lengths are given in base pairs. Abbreviations: LBD, ligand binding domain; SP, signal peptide; TM, transmembrane region. Light grey boxes indicate extracellular domain encoding exons, dark grey boxes indicate intracellular domain-encoding exons. LBD sequences marked blue were excluded from the

alignment used to construct the phylogenetic tree in Fig. 1. The second exon of the human GHR (in red) is specific to eutherian mammals. The first exon of the lamprey GHR/PRLR gene has only been found in this lineage. Upper case Cs indicate the positions of conserved cysteine residue pairs. Upper case Ms indicate the positions of alternative start codons. The purple star over the human GHR indicates a premature stop codon in a different reading frame found in some alternative transcripts. The cleavage site for the formation of growth hormone binding proteins (GHBP) is indicated by a purple arrow. Conserved domains are indicated by coloured arrows: TSXW hormone-binding motif (blue), WSXWS motif (magenta), “box 1” (green), and “box 2” (red). The exon organizations of representative teleost GHRa, GHRb, PRLRa and PRLRb are shown in Supplementary Fig. S1.

those with alternative exons not shown in Fig. 2 (Bu et al., 2013; Tanaka et al., 2000). Two alternative start codons are shown for the chicken *PRLR* in Fig. 2. The anole lizard *PRLR* gene has a putative alternative start of translation that matches the second of these. In chicken this is the start of translation of a testis-specific 5'-truncated *PRLR* transcript (Tanaka et al., 2000). *TPOR* genes are also known to have duplicate sets of LBD-encoding exons (Liongue and Ward, 2007). We could identify the duplicate LBD-encoding exons in all investigated *TPOR* gene predictions except that of anole lizard (Fig. 2). In the chicken and mammalian *TPOR* genes, the third exon of LBD1 is notably longer than in other species. Surprisingly, we could also identify two sets of LBD-encoding exons in the cartilaginous fish *GHR* genes (Fig. 2), which has not been reported previously. Our phylogeny of only the ligand-binding domains (Supplementary Fig. S2) shows the *TPOR* LBD1 clustering with the single *EPOR* LBD with high statistical support, while the *TPOR* LBD2 sequences occupy a more basal position. This suggests that the duplication of the ligand-binding domain preceded the emergence of *EPOR* and *TPOR* from an ancestral gene, in which case *EPOR* has secondarily lost LBD2. In order to avoid large alignment gaps and phylogenetic artifacts due to the varying ligand-binding domain composition within this family of receptors, the LBD1 of the chicken and anole lizard *PRLR* sequences, and of the cartilaginous fish *GHR* sequences, as well as the LBD2 of the *TPOR* sequences, were excluded from the alignment used to calculate the phylogeny presented in Fig. 1. We have provided the final alignment and phylogenetic tree files alongside this article (see Section 3.15 below).

3.4. Lamprey *GHR/PRLR*-like genes

The lamprey *GHR/PRLR*-like genes have an additional 5' signal peptide exon and only three of the four LBD exons: the coding sequences end on a premature stop codon compared with full-length single-chain cytokine class I receptor sequences. Notably they lack the exon which contains the conserved WSXWS motif, as well as the exons encoding the transmembrane region and intracellular domain (Fig. 2). Despite detailed searches in both sea lamprey and Arctic lamprey genomes, we could not identify exon predictions with the conserved WSXWS motif, TM region, or intracellular “box 1” and “box 2” motifs. In the sea lamprey genome, the *GHR/PRLR*-like gene is flanked on both sides by large gaps in the assembly, precluding the retrieval of a putative full-length receptor sequence. Our findings mirror the sea lamprey *GHR/PRLR* cDNA described by Ellens et al. (2013), which lacks a transmembrane region as well as an intracellular domain. This suggests that the lamprey *GHR/PRLR*-like gene encodes a soluble binding protein rather than a membrane-bound receptor. Such soluble binding protein variants have been described from a number of species (Björnsson et al., 2002; Calduch-Giner et al., 2003; Ross et al., 1997; Saunders et al., 2002; Sohm et al., 1998). It is possible that this truncated form of the receptor gene is a derived feature in at least Northern hemisphere lampreys.

3.5. Putative invertebrate family members

Putative single-chain class I cytokine receptor sequences were sought through BLAST in the genomes of the tunicate *Ciona intestinalis*, the cephalochordate Florida lancelet, the hemichordate acorn worm and the echinoderm purple sea urchin, however no putative family member could be identified using this approach. Using a Hidden Markov model search approach, we could identify a gene prediction from the Florida lancelet. This sequence matches the lancelet sequence that was reported by Li et al. (2014) as *GH/PRL1BP*, for “growth hormone/prolactin binding protein”. It contains an extracellular ligand-binding domain typical of single-chain class I cytokine receptors, including a WSXWS motif, however the exon structure of the gene prediction differs from the vertebrate *GHR*, *PRLR*, *EPOR* and *TPOR* genes (Supplementary Fig. S3). Despite our efforts to identify any TM region and intracellular domain-encoding exons in the Florida lancelet genome, a full-length receptor sequence could not be found. In addition, the sequence has an extended N-terminal, partly encoded by additional 5' exons, which vertebrate single-chain class I cytokine receptor sequences lack. Owing to these discrepancies, which may cause phylogenetic artifacts, this putative family member in the Florida lancelet was not included in our main alignment and the phylogenetic analysis presented in Fig. 1. Instead, we carried out a phylogenetic analysis based on the extracellular ligand-binding domains only. The resulting unrooted phylogeny is presented in Supplementary Fig. S3. The lancelet *GH/PRL1BP* sequence clusters closer to the vertebrate *GHR*, *CRFA4* and *PRLR* clade, however the lack of a reliable outgroup to root this analysis makes its position within the wider class I cytokine receptor family uncertain. Further studies are underway to resolve this question, including searches in other invertebrate lineages and conserved synteny analyses.

3.6. Sequence completion and data quality

In some cases, full-length sequences could not be obtained despite repeated searches and manual curation. Out of 120 single-chain cytokine class I receptor gene predictions that we identified, 58 needed some level of manual correction and 32 could not be predicted in full length, mostly due to gaps in the genome assemblies. Of the latter, the putative Atlantic herring *TPOR* and medaka *CRFA4* sequences, as well as an *EPOR*-like sequence in Atlantic salmon, were so short they could not be used in our analyses. Two additional partial sequences from the Atlantic salmon genome are likely the result of assembly errors and were not investigated further (see comments in Supplementary data 1). Partial predictions are indicated by an asterisk next to the sequence name in Fig. 1. The first exon, which encodes only the signal peptide, was commonly missing from the Ensembl gene predictions, and in several cases the full intracellular domain sequences had not been identified. Details about the quality of each individual sequence are described in Supplementary data 1. These observations are indicative of the degree of errors found in genome databases (Prosdocimi et al., 2012), as well as the high degree of sequence divergence that has been observed for the intracellular domains in several species (Iso-Touru et al., 2009). In our inspection of the identified sequences we could also observe a high degree of variation in this domain, both with regard to sequence length, some of which is shown in Fig. 2, and amino acid composition.

3.7. Conserved synteny analyses

We identified 19 gene families in the vicinity of *GHR*, *PRLR*, *EPOR* and *TPOR* genes in the human and medaka genomes (Supplementary data 3). Out of these, we analyzed the chromosomal locations and phylogeny of 18 neighboring gene families. These are summarized in Table 1. The large “Zinc finger” protein family prediction was discarded due to its large number of members and the high degree of sequence conservation among them. In addition to the gene families identified in this study, the paralemmin (*PALM*) family was also analyzed as we could determine that it had members in the relevant chromosome blocks in several genomes. We have previously studied the evolution of this family (Hultqvist et al., 2012) and determined that it likely arose and diversified through the vertebrate tetraploidizations. The identified blocks of conserved synteny are presented for the genomes of human, chicken, spotted gar (Fig. 3), medaka, three-spined stickleback (Fig. 4) and zebrafish (Supplementary Fig. S4). By carrying out phylogenetic analyses on these neighboring gene families, we established the orthology (between species) and paralogy (within species) relationships between the investigated chromosomal regions. Importantly, these phylogenetic analyses also allowed us to investigate whether or not the chromosome blocks bearing *GHR*, *PRLR*, *EPOR* and *TPOR* genes diverged in the time windows of the two basal vertebrate tetraploidizations (1R and 2R), as well as the basal teleost tetraploidization (3R). These relative dating points were determined by the inclusion of vase tunicate or lancelet (*Branchiostoma ssp.*) family members for 1R and 2R, and spotted gar family members for 3R, in the phylogenies of the neighboring gene families.

Out of the 18 neighboring gene families that we analyzed in total, 12 support or are consistent with divergences in 1R and 2R: ADAMTS6/10 (Supplementary Fig. S5), BEST (Supplementary Fig. S6), C6-9 (Supplementary Fig. S7), CYP4 (Supplementary Fig. S8), FGF3/7/10/22 (Supplementary Fig. S9), MAST (Supplementary Fig. S12), NIM1K (Supplementary Fig. S13), PRDX (Supplementary Fig. S17), PTGER4 (Supplementary Fig. S18), SLC1A (Supplementary Fig. S19), ZFR (Supplementary Fig. S21) and ZSWIM (Supplementary Fig. S22). An additional four families are consistent with divergences in 1R and 2R, but lack out-groups and tunicate sequences to relatively date these divergences: GFL (Supplementary Fig. S10), LIFR/OSMR (Supplementary Fig. S11), PALM (Supplementary Fig. S16) and STRBP/ILF3 (Supplementary Fig. S20). The remaining two families, NIPBL (Supplementary Fig. S14) and OXCT (Supplementary Fig. S15), do not have multiple branches diverging in early vertebrate evolution.

With respect to 3R, 8 neighboring gene families support or are consistent with the divergence of duplicate teleost branches with members on the GHRa-PRLRb and GHRb-PRLRa-bearing chromosome blocks: FGF3/7/10/22 (Supplementary Fig. S9), GFL (Supplementary Fig. S10), LIFR/OSMR (Supplementary Fig. S11), NIM1K (Supplementary Fig. S13), NIPBL (Supplementary Fig. S14), OXCT (Supplementary Fig. S15), PTGER4 (Supplementary Fig. S18) and SLC1A (Supplementary Fig. S19). An additional 3 families have duplicate teleost genes in the vicinity of the GHRa-PRLRb and GHRb-PRLRa gene pairs, however the topologies of the phylogenetic trees are not entirely clear with respect to the concurrence with 3R: C6-9 (Supplementary Fig. S7), STRBP/ILF3 (Supplementary Fig. S20) and ZSWIM (Supplementary Fig. S22). Descriptions and analyses of each neighboring gene

Table 1. Neighboring gene families.

Symbol	Descriptions	Phylogenetic tree figure	Outgroup
ADAMTS6/10	ADAM metallopeptidase with thrombospondin type 1 motif 6 and 10	Suppl. Fig. S5	Purple sea urchin
BEST	Bestrophins	Suppl. Fig. S6	Fruit fly
C6-9	Complement components C6 - C9	Suppl. Fig. S7	Lancelet
CYP4	Cytochrome P450, family 4	Suppl. Fig. S8	Purple sea urchin
FGF3/7/10/22	Fibroblast growth factors 3, 7, 10 and 22	Suppl. Fig. S9	Lancelet
GFL	Glial cell-line neurotrophic factor (GDNF) family ligands	Suppl. Fig. S10	Unrooted
LIFR/OSMR	Oncostatin M receptor and Leukemia inhibitory factor receptor	Suppl. Fig. S11	Unrooted
MAST	Microtubule associated serine/threonine kinases	Suppl. Fig. S12	Fruit fly
NIM1K	NIM1 serine/threonine kinases	Suppl. Fig. S13	Purple sea urchin
NIPBL	Nipped-B homolog	Suppl. Fig. S14	Purple sea urchin
OXCT	3-oxoacid CoA-transferases	Suppl. Fig. S15	Fruit fly
PALM	Paralemmins	Suppl. Fig. S16	Unrooted
PRDX	Peroxiredoxins	Suppl. Fig. S17	Fruit fly
PTGER4	Prostaglandin E receptors EP ₄	Suppl. Fig. S18	Purple sea urchin
SLC1A	Solute carrier family 1	Suppl. Fig. S19	Fruit fly
STRBP/ILF3	Spermatid perinuclear RNA binding protein and Interleukin enhancer binding factor 3	Suppl. Fig. S20	Unrooted
ZFR	Zinc finger RNA binding proteins	Suppl. Fig. S21	Fruit fly
ZSWIM4/5/6	Zinc finger SWIM-type containing 4, 5 and 6	Suppl. Fig. S22	Vase tunicate

Gene family symbols and descriptions are derived from Human Gene Nomenclature Committee gene symbols and descriptions, except GFL which is an acronym of GDNF Family Ligands. Phylogenetic trees were rooted with identified sequences from fruit fly (*Drosophila melanogaster*), purple sea urchin (*Strongylocentrotus purpuratus*), lancelet (*Branchiostoma floridae*, *B. lanceolatum* and/or *B. belcheri*) or vase tunicate (*Ciona intestinalis*). Unrooted trees are shown as midpoint-rooted trees. Vase tunicate branches were used as relative dating points for the basal vertebrate tetraploidizations (1R/2R) and spotted gar (*Lepisosteus oculatus*) branches were used as relative dating points for the basal teleost tetraploidization (3R).

family phylogeny are provided in the corresponding supplementary figure captions. However, some gene families are worth noting here; either because they provided surprising secondary findings or because their phylogenies are particularly complex.

3.8. ADAM metallopeptidases with thrombospondin type 1 motif

The secreted metallopeptidases of the ADAMTS superfamily are crucial for the assembly and remodeling of the extracellular matrix (Brocker et al., 2009). In the human genome, the ADAMTS family is comprised of 19 genes, with an additional 4 ADAMTS-like genes also being known (Huxley-Jones et al., 2007). The Ensembl protein family prediction used in this study contained 8 ADAMTS genes as well as the ADAMTS-like papilin (*PLPN*) genes. However, only ADAMTS6 and ADAMTS10 were found to be located in the vicinity of single-chain class I cytokine receptor genes. The above-cited study by Huxley-Jones et al. (2007), indicates that these two genes constitute a distinct subfamily in vertebrates, with one clearly defined vase tunicate ortholog. Therefore, we proceeded to analyze only the ADAMTS6 and ADAMTS10 sequences, and our phylogeny supports the divergence of ADAMTS6 and ADAMTS10 clades in the time window of the 1R and 2R tetraploidizations (Supplementary Fig. S5).

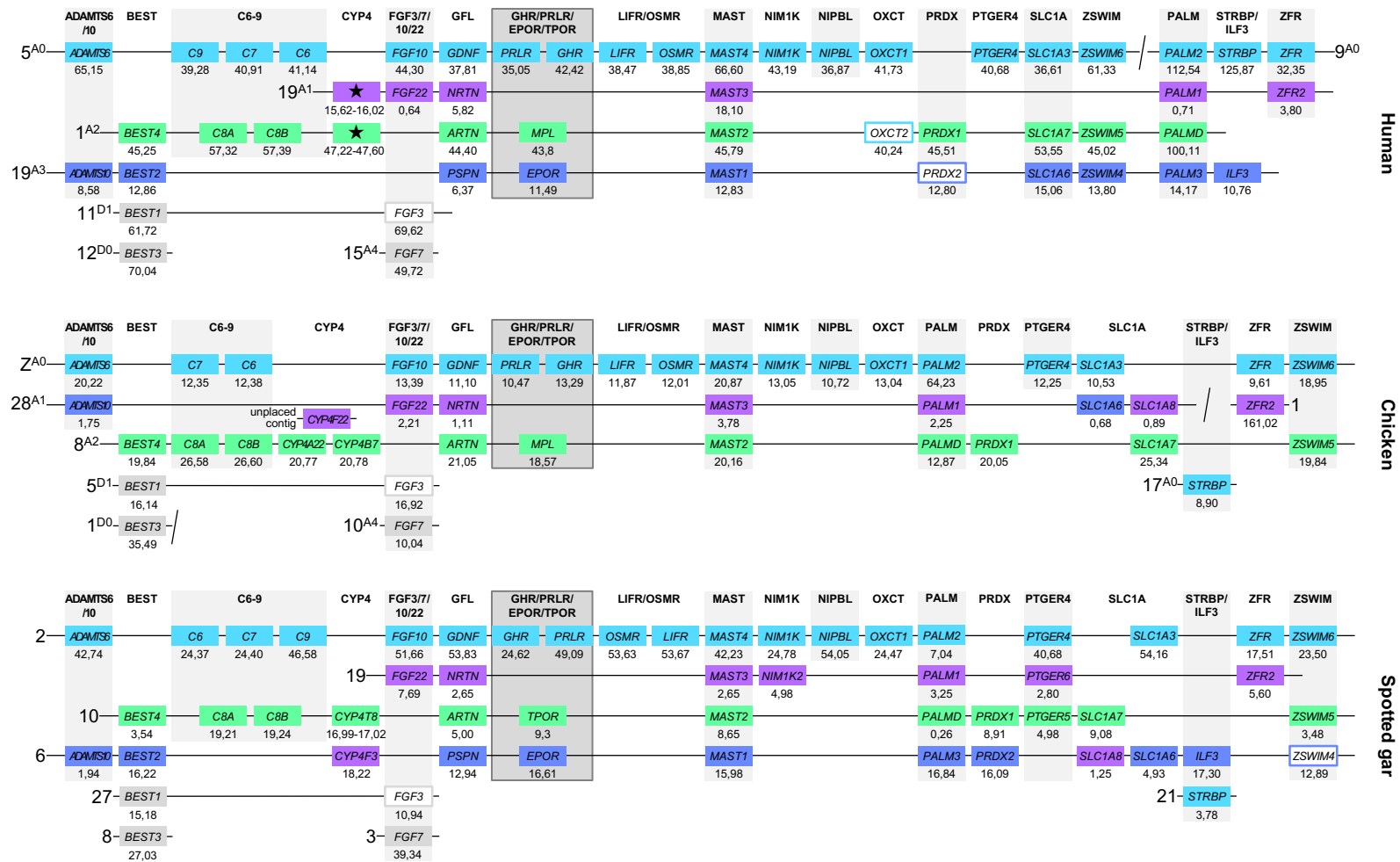


Fig. 3. Conserved synteny blocks in the human, chicken and spotted gar genomes. Gene families are organized alphabetically. Locations are given in MB unless specified. Colours are applied following the human and spotted gar chromosomes in order to show conserved synteny as well as sequence homology between species. Genes represented by white bordered boxes have unclear positions in the phylogenetic analyses. Stars correspond to clusters of CYP4 genes (see Section 3.9 in the results). A0-A3 and D0-D1 designations next to human and chicken (and medaka in Fig. 4) chromosome numbers correspond to vertebrate paralogous blocks identified by Nakatani et al. (2007), to have arisen through the 1R and 2R basal vertebrate tetraploidizations. Such designations within parentheses indicate our interpretation of the supplementary data provided by Nakatani et al. (2007).

3.9. Bestrophins

The bestrophins are a family of Ca^{2+} -activated monovalent anion channels that are expressed in a wide variety of tissues and participate in a wide gamut of functions. Mutations in the *BEST1* gene are associated with retinopathies (Dickson et al., 2014), and there is evidence for bestrophin function in the regulation of arterial pressure (Matchkov et al., 2015) and smooth muscle contractions (Bulley and Jaggar, 2014). Our phylogeny of the BEST family supports the emergence of *BEST1*, -2, -3 and -4 in the time window of the 1R and 2R tetraploidizations. However, the locations of the *BEST1* and *BEST3* genes suggest they are part of a separate set of paralogous chromosome blocks from that of *BEST2* and *BEST4* (Figs. 3 and 4). One interpretation of this is that a *BEST1/BEST3* ancestral gene and a *BEST2/BEST4* ancestral gene arose before the 1R tetraploidization and were located on different chromosomes, whereupon either 1R or 2R generated the four subtypes. The other is that 1R generated the *BEST1/BEST3* and *BEST2/BEST4* ancestral genes, whereupon chromosome rearrangements translocated one of the ancestral genes to a different chromosome before 2R generated the four subtypes. In either case, the phylogeny of the *BEST2* and *BEST4* clades, as well as their chromosomal locations, support the emergence of *EPOR* and *TPOR* in the 2R tetraploidization.

3.10. Complement components C6–C9

Together with the complement component 5b protein (C5b), the proteins encoded by this gene family form the terminal Membrane Attack Complex (MAC) of the innate immunity complement system. Our phylogeny (Supplementary Fig. S7), in combination with the chromosomal location data, indicates that there likely have been local duplications in early vertebrate evolution both preceding and following the 1R and 2R tetraploidizations; the latter generating only *C9* and an ancestral *C8* gene. Subsequently this *C8* gene gave rise to *C8A* and *C8B* through a local duplication some time before the divergence of ray-finned fishes and lobe-finned fishes (including tetrapods). Although the common ancestry of the MAC protein genes and their emergence through gene duplications has been discussed previously in the literature, the suggested time windows for these events have been unclear (Holland and Lambris, 2002). We could identify 9 MAC-like sequences in the vase tunicate, mirroring a previous study (Azumi et al., 2003), and at least 5 MAC-like sequence in the Florida lancelet (Supplementary data 2). Our results thus indicate that this gene family has had independent expansions in the three chordate lineages: vertebrates, tunicates and cephalochordates. With respect to the 3R tetraploidization, there are teleost duplicates of *C7* named *C7a* and *C7b*. However, in our phylogeny the spotted gar *C7* sequence clusters with the tetrapod *C7* sequences rather than basal to the teleost *C7a* and *C7b* branches, which makes the relative dating of the duplication somewhat uncertain.

3.11. Cytochrome P450 family 4

The cytochrome P450 (CYP) proteins form a large superfamily of enzymes found across all kingdoms of life. They mediate the biosynthesis and catabolism of small endogenous molecules, such as steroids, retinoids, fatty acids, and their derivatives, as well as the catabolism of exogenous substrates, such as environmental toxins and drugs (Thomas, 2007). The protein family prediction identified in this study comprises the CYP4A, -4B, -4F, -

4T, -4X and -4Z subtypes. However, the CYP4 family is larger and includes additional invertebrate-specific subtypes as well as the related *CYP4V* genes (Kirischian and Wilson, 2012). Our phylogeny of the CYP4 family shows two branches diverging early in vertebrate evolution; one *CYP4F* branch and one branch containing the *CYP4A*, -4B, -4T, -4X and -4Z (Supplementary Fig. S8). Two vase tunicate sequences cluster with the *CYP4F* branch, however the low support for this node (19%) indicates that they likely diverged at the base of the tree, which supports our suggested scenario. Our phylogeny, combined with chromosomal location data, suggests that there have been multiple waves of local gene duplications in tetrapods. Some of these are likely ancestral to at least amniotes, however most local gene duplications seem to be lineage specific. Including the known clusters of CYP4 genes in the human genome, on chromosomes 1 and 19 (Fig. 3), and the mouse genome, on chromosomes 4 and 17 respectively (Nelson et al., 2004).

3.12. Fibroblast growth factors 3, 7, 10 and 22

The fibroblast growth factors (FGF) are integral components in the signaling pathways required for embryonic development and morphogenesis, as well as various homeostatic and endocrine functions postnatally (Itoh and Ornitz, 2011). Our analysis of conserved synteny identified only *FGF10* and *FGF22* genes. We added *FGF3* and *FGF7* to our analyses based on previous studies grouping these 4 genes together (Itoh and Ornitz, 2011). We also included putative family members from vase tunicate, lancelet and purple sea urchin based on previous studies (Bertrand et al., 2011; Oulion et al., 2012; Satou et al., 2002). We present two phylogenies of this family (Supplementary Fig. S9) owing to the uncertain relationship between the *Branchiostoma* *FGFA* and *FGFB* genes and the vertebrate genes (see phylogeny in Satou et al. (2002), as well as phylogenies and synteny analyses in Bertrand et al. (2011), and Oulion et al. (2012)). Both our phylogenies suggest that *FGF3* constitutes a separate subfamily, and that *FGF7*, *FGF10* and *FGF22* arose early in vertebrate evolution. However, only the second phylogeny shows the putative tunicate family member clustering basal to the vertebrate *FGF7*, *FGF10* and *FGF22* clades.

3.13. NIM1 serine/threonine kinases

The human NIM1 kinase (NIM1K) is a poorly understood member of the AMP-activated protein kinase (AMPK)-related kinase family (Bright et al., 2009; Jaleel et al., 2005). Our phylogeny of the gene family that includes NIM1K (Supplementary Fig. S13) shows two branches diverging in the time window of the 1R and 2R tetraploidizations: one containing orthologs of the known human *NIM1K*, and one novel branch we have called *NIM1K2*. Both subtypes include duplicate teleost branches, however the topologies are somewhat unclear with respect to the 3R tetraploidization. The zebrafish *NIM1Ka* sequence diverges basal to both *NIM1Ka* and *NIM1Kb* sequences. This is likely due to a phylogenetic artifact caused by the uneven evolutionary rates between the two teleost subtypes (see branch lengths in our phylogeny). The positions of the spotted gar *NIM1K* and zebrafish *nim1kb* sequences nonetheless makes duplication through the 3R tetraploidization the most likely scenario. Regarding the *NIM1K2* branch, the spotted gar *NIM1K2* sequence clusters with the *NIM1K2a* branch rather than basal to both teleost duplicate branches. This is arguably a phylogenetic artifact caused by the lack of tetrapod *NIM1K2* sequences, as indicated by the low node support (17%) of the spotted gar branch.

3.14. Prostaglandin E receptors EP₄

The EP₄ receptor, encoded by the *PTGER4* gene, belongs to the family of “relaxant” prostanoid receptors together with the closely related EP₂ receptor (*PTGER2*) as well as the prostacyclin receptor IP (PTGIR) and the prostaglandin D₂ receptors DP₁ (*PTGDR1*) and DP₂ (*PTGDR2*) (Narumiya et al., 1999). Their common properties include the ability to elicit smooth muscle relaxation as well as signaling through cAMP-dependent intracellular signaling pathways. The EP₁ (*PTGER1*) and EP₃ (*PTGER3*) receptors belong to separate prostanoid receptor families.

In addition to the known *PTGER4* sequences, we describe here two hitherto undescribed *PTGER4*-related receptor subtypes found only in the coelacanth, spotted gar and teleost fish genomes. Our phylogeny of this family (Supplementary Fig. S18), together with the chromosomal locations, suggest that the novel subtype sequences constitute two ancestral vertebrate clades. Therefore, we propose the nomenclature *PTGER5* and *PTGER6* for these genes. The phylogenetic tree was rooted with a sequence from the purple sea urchin. However, since the evolutionary relationship between the sea urchin sequence and other members of the “relaxant” prostanoid receptors is unclear, we also included sequences identified in the genome of the Florida lancelet. Our phylogeny, together with the chromosomal data, supports the emergence of *PTGER4*, *PTGER5* and *PTGER6* in the 1R and 2R tetraploidizations, and the duplication of *PTGER4* in the 3R tetraploidization, giving rise to *PTGER4a* and *PTGER4b*.

The zebrafish *PTGER5* gene has received the name *ptger4c*, based on a limited phylogenetic analysis and physiological similarities to the zebrafish EP₄ receptors (Tsuge et al., 2013). Notably, these physiological similarities include the ability to elicit a cAMP surge in response to prostaglandin E₂ as well as an EP₄-selective agonist. However, the name *ptger4c* is misleading since it implies that the gene emerged in the same time window as the teleost 3R-generated *PTGER4a* and *PTGER4b* genes, rather than early in vertebrate evolution.

3.15. Description of supporting data and additional files

In addition to the supplementary figures and data files attached to the online version of this article (see Appendix A below), we have deposited a file set with a stable digital object identifier at <https://doi.org/10.6084/m9.figshare.4531364>. This file set includes all alignment and phylogenetic tree files described in this study, including neighboring gene families as well as files submitted before peer-review. It also includes an unaligned sequence file with all curated *GHR*, *PRLR*, *CRFA4*, *EPOR* and *TPOR* sequences identified in this study.

4. Discussion

4.1. *GHR, PRLR, EPOR and TPOR (MPL)-bearing chromosome blocks were duplicated in vertebrate tetraploidizations*

We have compared the chromosomal locations of 18 gene families with members in the vicinity of receptor genes for growth hormone (*GHR*), prolactin (*PRLR*), erythropoietin (*EPOR*) and thrombopoietin (*TPOR*) in several vertebrate genomes. This was combined with phylogenetic analyses of the gene families in order to 1) infer orthology and paralogy relationships between the identified chromosome blocks, and 2) determine the time window of the duplications that gave rise to them. Taken together, our chromosomal data and phylogenetic analyses indicate that the vertebrate *GHR*, *PRLR*, *EPOR* and *TPOR* genes are located in paralogous chromosome blocks that arose in the time window of the basal vertebrate tetraploidizations, called 1R and 2R. Subsequently, these chromosome blocks were further duplicated in the basal teleost tetraploidization, called 3R. This scenario is summarized in Fig. 5.

Following our analyses, we compared our results with the reconstruction of the ancestral vertebrate genome before and after 1R and 2R by Nakatani et al. (2007): The paralogous chromosome regions we have identified correspond to the ancestral vertebrate proto-chromosome “A” proposed by Nakatani et al. (2007), and the paralogous chromosome blocks “A0” to “A3” (Fig. 5) that likely arose in 1R and 2R. This lends further support to our conclusions. We also identified a few genes on seemingly unrelated chromosome blocks, namely *BEST1*, *BEST3*, *FGF3* and *FGF7* (indicated by grey lines on the chromosomes in Fig. 5). Of these, the *FGF7* genes are located in regions identified as part of the ancestral paralogous block “A4” (Figs. 3 and 4), which Nakatani et al. (2007), suggest arose through a fission from the ancestral “A2/A3” block between 1R and 2R. Furthermore, the paralogous chromosome regions we have identified correspond well with the 3R-generated chromosome blocks and ancestral teleost proto-chromosomes “i”, “m” and “e” proposed by Kasahara et al. (2007), and further ratified by Bian et al. (2016). In the teleost lineage, 3R was likely preceded by the partial or complete fusion of the “A1” and “A2” blocks to form part of teleost proto-chromosome “m” (Fig. 5). This was also suggested by Nakatani et al. (see Fig. 4 in that publication). After 3R, the ancestral teleost genome underwent a series of complex rearrangements (Bian et al., 2016; Kasahara et al., 2007). We could detect the fission of one of the duplicated “m” blocks sometime between 3R and the divergence of the lineage leading to zebrafish. Other chromosomal rearrangements prove more difficult to date, owing to the small number of translocated genes we could detect and contradicting data between different species (Figs. 3 and 4, Supplementary Fig. S4). Nonetheless, we have indicated such rearrangements in Fig. 5.

These paralogous chromosome regions have attracted attention previously, not least because they harbor the insulin/relaxin hormone and receptor gene families (Olinski et al., 2006; Yegorov and Good, 2012), and the idea that they arose through early vertebrate tetraploidizations goes back to the mid 1990's (Katsanis et al., 1996). Our conclusions and the scenario presented in Fig. 5 are in complete agreement with the latest and most comprehensive of these studies, by Yegorov and Good (2012) (see Fig. 2 in that publication), including suggested fusion and fission events, and with the overall conclusions reached by

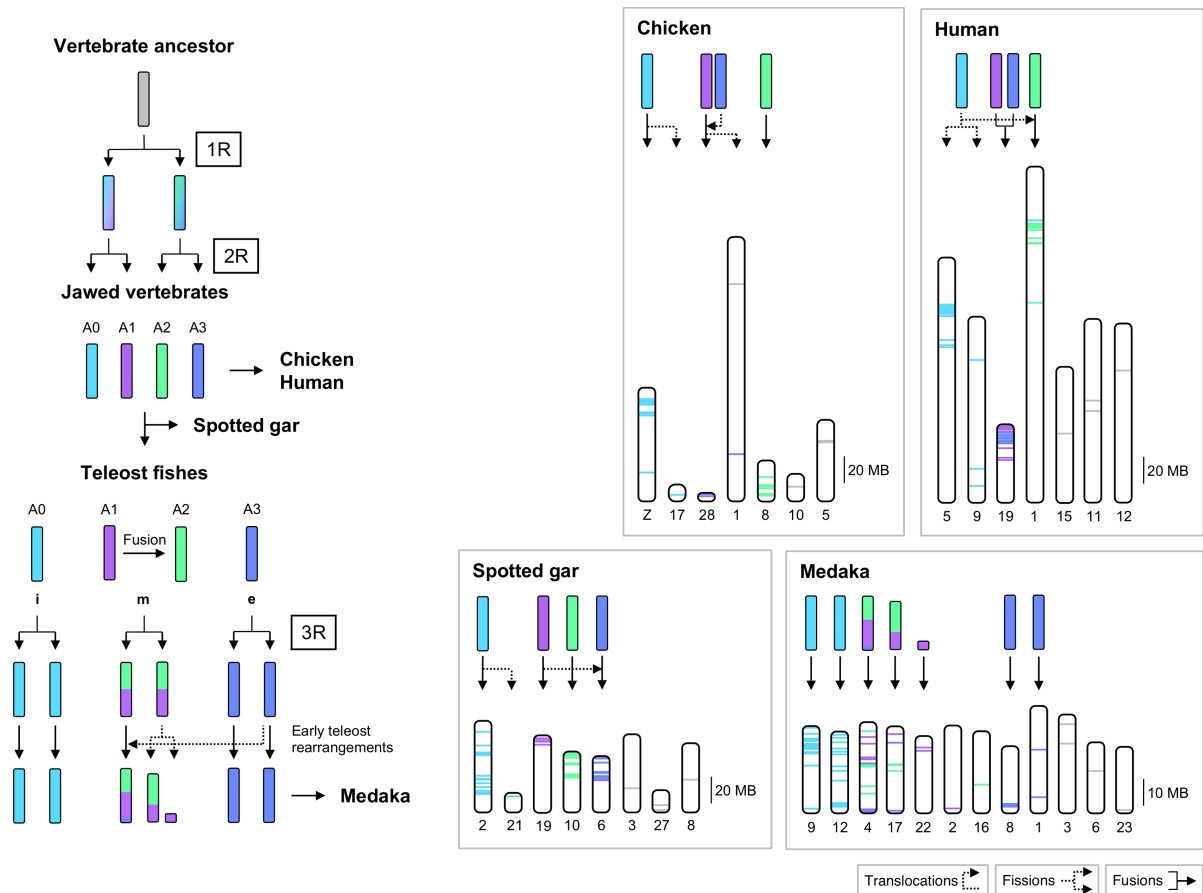


Figure 5. Evolution of paralogous chromosome blocks bearing *GHR*, *PRLR*, *EPOR* and *TPOR* genes in vertebrates. This scenario is based on the chromosomal locations and phylogenies of the *GHR*, *PRLR*, *EPOR* and *TPOR* genes, as well as 18 neighboring gene families, across a wide selection of vertebrate species. Here we show the corresponding chromosome regions in chicken, human, spotted gar and medaka. Ancestral chromosome block “A0” – “A3” designations after Nakatani et al., 2007; and “i”, “m”, “e” designations after Kasahara et al., 2007. The colour-coding of chromosome blocks follows Figs. 3 and 4.

Olinski et al. (2006). However, importantly our analyses improve upon these previous studies by including carefully constructed phylogenies of the syntenic gene families. These results are also consistent with our previously reported analysis of seven gene families, of which we have included the paralemmin (PALM) family in this study (Hultqvist et al., 2012).

4.2. Evolution of the single-chain class I cytokine receptors

As a whole, the chromosomal locations and phylogenies of the single-chain class I cytokine receptors, as well as 18 neighboring gene families, support the following scenario, summarized in Fig. 6: A vertebrate ancestral cytokine class I receptor gene was duplicated in 1R and 2R. The 1R tetraploidization gave rise to an ancestral *GHR/PRLR* gene and the ancestor of *EPOR* and *TPOR* genes. Following this, the *EPOR/TPOR* ancestor was duplicated in the 2R tetraploidization, resulting in the *EPOR* and *TPOR* genes. The 2R-generated duplicate of the ancestral *GHR/PRLR* gene was not preserved. Subsequently, the *GHR/PRLR* gene was duplicated locally and diverged giving rise to *GHR* and *PRLR*. The timing of this event is discussed in Section 4.3 below. With this local duplication event, the basic jawed vertebrate (*Gnathostomata*) setup of single-chain class I cytokine receptors was finally established: a *GHR* and a *PRLR* gene located on the same chromosome, as well as *EPOR* and *TPOR* genes located on separate related chromosomes. As for the *CRFA4* gene, which was a surprise addition to our analyses, we cannot provide a conclusive explanation of its origin.

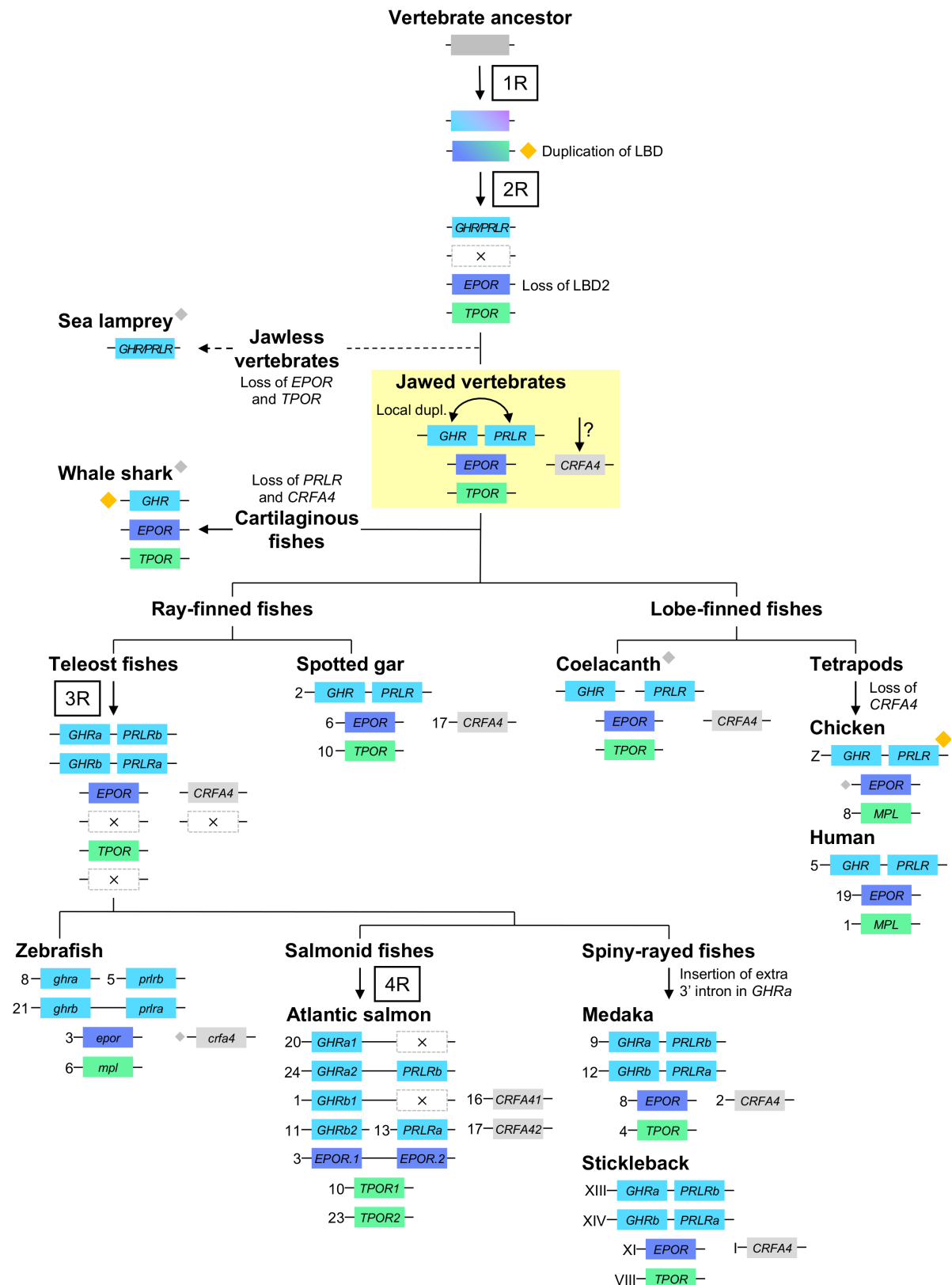


Figure 6. Evolutionary scenario of single-chain cytokine class I receptor genes. Based on chromosomal location data and phylogenetic analysis of 116 *GHR*, *PRLR*, *EPOR*, *TPOR* (*MPL*) and *CRFA4* gene predictions from 21 vertebrate species, 10 of which are shown in the figure. The uncertain divergence of the sea lamprey *GHR/PRLR*-like gene relative to 1R, 2R and the local duplication that gave rise to *GHR* and *PRLR* is indicated by dashed lines. Chromosomal/linkage group location data is not available for the sea lamprey, elephant shark and coelacanth genomes, nor the chicken *EPOR* and zebrafish *crfa4* gene predictions. Gene losses after the tetraploidization events are represented by crossed boxes. Yellow diamonds represent the suggested duplications of ligand binding domain (LBD)-encoding exons.

A *CRFA4* sequence was first identified in the green spotted pufferfish genome (Jaillon et al., 2004) and named tentatively as “*cytokine receptor family, class I receptor 4*”. Subsequently a *CRFA4* sequence was described in zebrafish (Liongue and Ward, 2007). However, their cognate hormone ligand has not been identified. Our analyses of these genes indicate that they indeed encode single-chain cytokine class I receptors, based on their primary amino acid sequence and exon structure (Fig. 2). Our phylogenetic analysis suggests a common origin with *GHR* genes, however no conserved synteny could be detected between *CRFA4*-bearing chromosome blocks and *GHR/PRLR*-bearing chromosome blocks. One possibility is that the *CRFA4* gene is the fourth family member generated by 2R, and that it was subsequently translocated to a non-paralogous chromosome block. However, as such translocations are more common between homologous chromosome regions through recombination, we make the more parsimonious interpretation that it is an ancestral family member. In such case, no 1R/2R-generated duplicates of the *CRFA4* ancestor gene survived. Although the time window for its origin is uncertain, we can date the presence of a *CRFA4* gene to a jawed vertebrate ancestor together with the *GHR*, *PRLR*, *EPOR* and *TPOR* genes.

After the basic setup of single-chain class I cytokine receptors was established in early vertebrate evolution, several lineage-specific events shaped the continued evolution of the gene family. Crucially, most jawed vertebrate species we have analyzed preserve the basic setup of genes, including the spotted gar and coelacanth, which represent basal lineages of ray-finned fishes (*Actinopterygii*) and lobe-finned fishes (*Sarcopterygii*), respectively (Fig. 6). Nevertheless, some gene losses seem to have occurred: 1) *CRFA4* was likely lost independently from the tetrapod and cartilaginous fish (*Chondrichthyes*) lineages, and 2) *PRLR* was likely lost early in cartilaginous fish evolution. We could not identify *PRLR* or *CRFA4* sequences in any of the cartilaginous fish species that were investigated, which represent all three orders of this lineage. Additionally, we could not identify an *EPOR* sequence in the genome of the elephant shark, however this is arguably a consequence of the incomplete status of this genome assembly. We could not find an *EPOR* sequence in the chicken genome nor in any available avian genome assemblies. However, a recent report identified the chicken *EPOR* sequence in a raw sequence read archive (Hron et al., 2015), and we included this sequence in our phylogenetic analysis. The conspicuous lack of some genes from avian genome databases seems to be related to high GC-content in both gene and intergenic sequences (Hron et al., 2015), which suggests that *EPOR* genes are located on microchromosomes in birds (Han et al., 2008; McQueen et al., 1996).

In the teleost fish lineage, the 3R tetraploidization duplicated the *GHR-PRLR* gene pair giving rise to *GHRa-PRLRb* and *GHRb-PRLRa* gene pairs (Fig. 6) on related chromosomes. This is supported by the chromosomal locations and phylogenies of 8 of the 18 neighboring gene families we analyzed. It is also supported by previous large-scale genomic analyses: In the analysis of the first draft medaka genome (Kasahara et al., 2007), the *GHR* and *PRLR*-bearing chromosomes 9 and 12, as well as the green spotted pufferfish chromosomes 4 and 12, correspond to an ancestral vertebrate linkage group that was duplicated in early teleost evolution, consistent with 3R. The disruption of linkage between the zebrafish *ghra* and *prlb* genes, located on chromosomes 8 and 5 respectively, post-3R is also supported by this analysis (see Fig. 4 in Kasahara et al. (2007)). Our scenario is also supported by a previous study of the growth hormone receptors by Fukamachi and Meyer (2007), which was the first to conclude that *GHR* genes had duplicated in the 3R tetraploidization. However, this study

did not include the syntenic *PRLR* genes nor phylogenetic analyses of neighboring gene families (Fukamachi and Meyer, 2007).

4.3. Divergence of *GHR* and *PRLR* genes

The *GHR* and *PRLR* genes most likely did not arise as a result of the 1R and 2R tetraploidizations. Their syntenic location in the human, chicken, spotted gar and several teleost genomes (Figs. 3 and 4, Supplementary Fig. S4) implies that they were located on the same chromosome at least before the split between lobe-finned fishes (including tetrapods) and ray-finned fishes approximately 440 million years ago (Amores et al., 2011), likely as the result of a local gene duplication. The possibility of a duplication through 2R followed by an early translocation of one of the genes, although less parsimonious, cannot be discarded. Cartilaginous fishes lack *PRLR*, however the position of the cartilaginous fish *GHR* sequences well within the *GHR* clade in our phylogeny (Fig. 1), suggests that the *GHR*-*PRLR* gene pair was present in a jawed vertebrate ancestor. A more precise time estimation for the emergence of this gene pair is made difficult due to the particular characteristics of the lamprey *GHR/PRLR* genes and the lack of a reliable invertebrate outgroup (see Section 3.5 above). The position of the lamprey *GHR/PRLR* sequences in our phylogeny suggests that they represent an ancestor of both *GHR* and *PRLR*. Indeed, the analyses by Ellens et al. (2013), which identified the sea lamprey *GHR/PRLR*-like cDNA, also suggest that this is the case, although their phylogeny included only a small number of *PRLR* sequences. This would place the local duplication that gave rise to *GHR* and *PRLR* after the divergence between jawless vertebrates (*Agnatha*) and jawed vertebrates, irrespective of when the jawless vertebrates diverged in relation to the basal vertebrate tetraploidizations (Mehta et al., 2013; Smith et al., 2013). We investigated the sea lamprey genome as well as the more complete genome of the Arctic lamprey (Mehta et al., 2013) and in both species, the predicted genes consist of only a portion of the extracellular ligand-binding domain (Fig. 2). It was not possible to identify a sequence in either lamprey species which spanned the full extent of a single-chain class I cytokine receptor sequence, including the transmembrane and intracellular domains with their characteristic conserved sequence motifs. The shortness of the sequences opens up for phylogenetic artifacts caused by a reduction in phylogenetic signal. Indeed, the lack of support for the lamprey *GHR/PRLR* branch in our phylogeny indicates the uncertainty in this regard, and we cannot completely discard the possibility that these sequences represent derived members of either the *GHR* or the *PRLR* clade. Thus, the local gene duplication that gave rise to *GHR* and *PRLR* might have taken place as early as before the divergence between jawless and jawed vertebrates.

4.4. Duplications of ligand-binding domain-encoding exons

The chicken prolactin receptor (Tanaka et al., 1992), as well as the thrombopoietin receptors across several species (Liongue and Ward, 2007), are known to have an extracellular “double antenna” structure. This is the result of duplicate sets of ligand-binding domain (LBD)-encoding exons in the corresponding genes. For the chicken prolactin receptor, only the second LBD seems to be required for prolactin binding (Bu et al., 2013), while for thrombopoietin receptors both LBDs seem to be required for thrombopoietin binding (Sabath et al., 1999). In addition to the receptor genes already mentioned, the anole lizard *PRLR* gene and the cartilaginous fish *GHR* genes were found to have duplicate sets of

LBD-encoding exons. We conclude that there have been at least three independent duplications of the ligand-binding domains during the evolution of this gene family (Fig. 6): In the ancestral *EPOR/TPOR* gene, preceding 2R, thus *EPOR* has secondarily lost LBD2; in the *PRLR* gene of a diapsid ancestor of birds and *Squamata* (lizards, snakes and *Amphisbaena*), at the latest; and in the *GHR* gene during early cartilaginous fish evolution.

4.5. Emergence of somatolactin receptors: Implications for growth hormone receptor differentiation and hormone-receptor interactions

The gene family of cognate ligands which includes growth hormone, prolactin and somatolactin all seem to have been present very early in vertebrate evolution (Ocampo Daza and Larhammar, manuscript in preparation). In contrast, all somatolactin receptors (SLR) that have been identified belong to the *GHRa* clade of receptors, which arose much later in the 3R tetraploidization. Thus, somatolactin and growth hormone must have shared GHR for a long period of time before an SLR arose. This would appear to be an example of sub-functionalization, i.e., the teleost *GHRa* and *GHRb* receptors partitioned the functions of the ancestral GHR receptor, evolving differential ligand binding preferences for either growth hormone or somatolactin after 3R. However, it is not at all clear that this is what has occurred in teleost fishes. Out of the receptors that have been reported as SLRs, only the masu salmon receptor has shown preferred binding of somatolactin (Fukada et al., 2005). Subsequent studies in black seabream (*Acanthopagrus schlegelii*) (Jiao et al., 2006) and zebrafish (Chen et al., 2011) show no clear differentiation between *GHRa* and *GHRb* in terms of ligand selectivity; and in addition, the Japanese eel (*Anguilla japonica*) *GHRa* receptor preferentially binds the native growth hormone, not somatolactin, in a competition assay (Ozaki et al., 2006). The rainbow trout (*Oncorhynchus mykiss*) receptors named Ghr1 and Ghr2 correspond to the salmonid 4R-generated duplicates we have called *GHRb1* and *GHRb2*, and have both been shown to preferentially bind the native rainbow trout growth hormone, albeit eliciting somewhat different physiological responses (Reindl et al., 2009). Taken together, these reports suggest a far more complex functional differentiation between *GHRa* and *GHRb* than the sub-functionalization of growth hormone and somatolactin binding. There is evidence to suggest that differing tissue expression patterns, differing responses to physiological states such as nutritional restriction and osmotic challenge, and the engagement of different intracellular signaling pathways, all contribute more to GHR differentiation (see Walock et al. (2014), and references therein, as well as Breves et al. (2010), and Breves et al. (2011)). The binding of somatolactin to one or both of the teleost GHRs is likely more dependent on inter-species differences between the hormones used experimentally, or factors that are hitherto unknown, rather than different selectivities between the native hormones and receptor subtypes in each species. To complicate things further, some teleost fishes have duplicates of somatolactin, called *SL α* and *SL β* , whose receptor preferences have not been fully investigated.

Many questions remain to be answered regarding the emergence of somatolactin-receptor interactions and the differentiation of GHR and PRLR subtypes in teleost fishes. We show here that there was a substantial time lag between the origin of growth hormone, prolactin and somatolactin (Ocampo Daza and Larhammar, manuscript in preparation) and the duplications resulting in their receptors, first through a local duplication after 2R in a jawed vertebrate ancestor, and later in teleost 3R and 4R. This means that the hormones have

shared the same receptor during an extensive period early in vertebrate evolution, and the functional distinction between growth hormone receptors and somatolactin receptors with regard to preferences for growth hormone and somatolactin is still blurred in teleost fishes. This lack of clear-cut selectivity is possibly also found in those lobe-finned fishes that preserve somatolactin.

4.6. Implications for the origin and evolution of vertebrate hematopoiesis

We were surprised to uncover the common ancestry of *EPOR* and *TPOR* genes with *GHR* and *PRLR* genes in an early vertebrate ancestor. This finding is relevant for the study of the origin and early evolution of hematopoiesis in vertebrates, in particular the differentiation and proliferation of erythrocytes as well as megakaryocytes and thrombocytes. An analysis of the colonial tunicate *Botryllus schlosseri* genome (Voskoboinik et al., 2013) found homologs of genes associated with hematopoietic stem cells in humans, including genes that define myeloid cell populations i.e. those that give rise to erythrocytes and megakaryocytes/thrombocytes. In an abstract presented to the Journal of Immunology, several of the same authors argue that the common ancestor of tunicates and vertebrates had a true hematopoietic myeloid lineage (Rosental et al., 2016). However, invertebrates (including tunicates) do not have erythrocytes or thrombocytes. Thus, while it seems that much of the genetic program required for erythrocyte and thrombocyte differentiation was present in an early chordate ancestor, the separate EPO/EPOR and TPO/TPOR hormone systems most likely emerged within the vertebrate lineage thanks to the basal vertebrate tetraploidizations 1R and 2R. We were not able to find any single-chain cytokine class I receptor sequences in the genome of the vase tunicate, nor that of *Botryllus schlosseri* (accessed through <http://botryllus.stanford.edu/botryllusgenome>).

4.7. Conclusions

The vertebrate growth hormone, prolactin, erythropoietin and thrombopoietin receptor genes *GHR*, *PRLR*, *EPOR* and *TPOR* arose in early in vertebrate evolution through the following mechanisms: The first round of basal vertebrate tetraploidization (1R) gave rise to *EPOR/TPOR* and *GHR/PRLR* ancestors, whereupon the second tetraploidization (2R) gave rise to the *EPOR* and *TPOR* genes. In contrast, the *GHR* and *PRLR* genes likely arose through a local duplication after 2R. Thus, surprisingly, the erythropoietin and thrombopoietin receptors arose from a common ancestor with growth hormone and prolactin receptors in early vertebrate evolution. The orphan receptor CRFA4 shares primary sequence and gene structure similarity with growth hormone and prolactin receptors, but no conserved synteny with *GHR*, *PRLR*, *EPOR* or *TPOR* genes, thereby making its origin highly enigmatic. In the teleost lineage, the basal teleost tetraploidization (3R) generated duplicates of *GHR* and *PRLR*; *GHRa*, *GHRb*, *PRLRa* and *PRLRb*. The salmonid tetraploidization (4R) contributed further duplicates of *GHRa*, *GHRb*, *CRFA4* and *TPOR*. The somatolactin receptors that have been identified thus far belong to the *GHRa* subtype, however somatolactin-receptor interactions are not clear-cut and likely predate the split between ray-finned fishes and lobe-finned fishes.

These studies raise fundamental questions about how functional specializations and hormone-receptor interactions have emerged and evolved in growth hormone, prolactin

and somatolactin systems. Our results are also relevant for comparative and evolutionary studies of the physiological processes influenced by growth hormone, prolactin and somatolactin, as well as erythropoietin and thrombopoietin, including the regulation of growth and metabolism, osmoregulation and hematopoiesis.

Acknowledgments

We thank Christina A. Bergqvist for valuable help with preliminary data gathering for neighboring gene families as well as proofreading. This work was supported by grants from the Swedish Research Council and Carl Trygger's Foundation.

References

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. *J. Mol. Biol.* 215, 403–10. <http://dx.doi.org/10.1006/jmbi.1990.9999>
- Amemiya, Y., Sogabe, Y., Nozaki, M., Takahashi, A., Kawauchi, H., 1999. Somatolactin in the white sturgeon and African lungfish and its evolutionary significance. *Gen. Comp. Endocrinol.* 114, 181–90. <http://dx.doi.org/10.1006/gcen.1998.7250>
- Amores, A., Catchen, J.M., Ferrara, A., Fontenot, Q., Postlethwait, J.H., 2011. Genome evolution and meiotic maps by massively parallel DNA sequencing: spotted gar, an outgroup for the teleost genome duplication. *Genetics* 188, 799–808. <http://dx.doi.org/10.1534/genetics.111.127324>
- Anisimova, M., Gascuel, O., 2006. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst. Biol.* 55, 539–52. <http://dx.doi.org/10.1080/10635150600755453>
- Anisimova, M., Gil, M., Dufayard, J.-F., Dessimoz, C., Gascuel, O., 2011. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Syst. Biol.* 60, 685–99. <http://dx.doi.org/10.1093/sysbio/syr041>
- Azumi, K., De Santis, R., De Tomaso, A., Rigoutsos, I., Yoshizaki, F., Pinto, M.R., Marino, R., Shida, K., Ikeda, M., Ikeda, M., Arai, M., Inoue, Y., Shimizu, T., Satoh, N., Rokhsar, D.S., Du Pasquier, L., Kasahara, M., Satake, M., Nonaka, M., 2003. Genomic analysis of immunity in a Urochordate and the emergence of the vertebrate immune system: “waiting for Godot.” *Immunogenetics* 55, 570–581. <http://dx.doi.org/10.1007/s00251-003-0606-5>
- Bellono, N.W., Leitch, D.B., Julius, D., 2017. Molecular basis of ancestral vertebrate electroreception. *Nature*. <http://dx.doi.org/10.1038/nature21401>
- Benedet, S., Björnsson, B.T., Taranger, G.L., Andersson, E., 2008. Cloning of somatolactin alpha, beta forms and the somatolactin receptor in Atlantic salmon: seasonal expression profile in pituitary and ovary of maturing female broodstock. *Reprod. Biol. Endocrinol.* 6, 42. <http://dx.doi.org/10.1186/1477-7827-6-42>
- Bertrand, S., Camasses, A., Somorjai, I., Belgacem, M.R., Chabrol, O., Escande, M.-L., Pontarotti, P., Escriva, H., 2011. Amphioxus FGF signaling predicts the acquisition of vertebrate morphological traits. *Proc. Natl. Acad. Sci. U. S. A.* 108, 9160–5. <http://dx.doi.org/10.1073/pnas.1014235108>
- Bian, C., Hu, Y., Ravi, V., Kuznetsova, I.S., Shen, X., Mu, X., Sun, Y., You, X., Li, J., Li, X., Qiu, Y.,

- Tay, B.-H., Thevasagayam, N.M., Komissarov, A.S., Trifonov, V., Kabilov, M., Tupikin, A., Luo, J., Liu, Y., Song, H., Liu, C., Wang, X., Gu, D., Yang, Y., Li, W., Polgar, G., Fan, G., Zeng, P., Zhang, H., Xiong, Z., Tang, Z., Peng, C., Ruan, Z., Yu, H., Chen, J., Fan, M., Huang, Y., Wang, M., Zhao, X., Hu, G., Yang, H., Wang, J., Wang, J., Xu, X., Song, L., Xu, G., Xu, P., Xu, J., O'Brien, S.J., Orbán, L., Venkatesh, B., Shi, Q., 2016. The Asian arowana (*Scleropages formosus*) genome provides new insights into the evolution of an early lineage of teleosts. *Sci. Rep.* 6, 24501. <http://dx.doi.org/10.1038/srep24501>
- Björnsson, B.T., Johansson, V., Benedet, S., Einarsdóttir, I.E., Hildahl, J., Agustsson, T., Jönsson, E., 2002. Growth hormone endocrinology of salmonids: Regulatory mechanisms and mode of action. *Fish Physiol. Biochem.* 27, 227–242. <http://dx.doi.org/10.1023/B:FISH.0000032728.91152.10>
- Boulay, J.-L., O'Shea, J.J., Paul, W.E., 2003. Molecular phylogeny within type I cytokines and their cognate receptors. *Immunity* 19, 159–63. [http://dx.doi.org/10.1016/S1074-7613\(03\)00211-5](http://dx.doi.org/10.1016/S1074-7613(03)00211-5)
- Breves, J.P., Fox, B.K., Pierce, A.L., Hirano, T., Grau, E.G., 2010. Gene expression of growth hormone family and glucocorticoid receptors, osmosensors, and ion transporters in the gill during seawater acclimation of mozambique tilapia, *Oreochromis mossambicus*. *J. Exp. Zool. Part A Ecol. Genet. Physiol.* 313 A, 432–441. <http://dx.doi.org/10.1002/jez.613>
- Breves, J.P., Seale, A.P., Helms, R.E., Tipsmark, C.K., Hirano, T., Grau, E.G., 2011. Dynamic gene expression of GH/PRL-family hormone receptors in gill and kidney during freshwater-acclimation of Mozambique tilapia. *Comp. Biochem. Physiol. - A Mol. Integr. Physiol.* 158, 194–200. <http://dx.doi.org/10.1016/j.cbpa.2010.10.030>
- Bright, N.J., Thornton, C., Carling, D., 2009. The regulation and function of mammalian AMPK-related kinases. *Acta Physiol.* 196, 15–26. <http://dx.doi.org/10.1111/j.1748-1716.2009.01971.x>
- Brocker, C.N., Vasiliou, V., Nebert, D.W., 2009. Evolutionary divergence and functions of the ADAM and ADAMTS gene families. *Hum. Genomics* 4, 43–55. <http://dx.doi.org/10.1186/1479-7364-4-1-43>
- Bu, G., Wang, C.Y., Cai, G., Leung, F.C., Xu, M., Wang, H., Huang, G., Li, J., Wang, Y., 2013. Molecular characterization of prolactin receptor (cPRLR) gene in chickens: gene structure, tissue expression, promoter analysis, and its interaction with chicken prolactin (cPRL) and prolactin-like protein (cPRL-L). *Mol. Cell. Endocrinol.* 370, 149–62. <http://dx.doi.org/10.1016/j.mce.2013.03.001>
- Bulley, S., Jaggar, J.H., 2014. Cl⁻ channels in smooth muscle cells. *Pflugers Arch. Eur. J. Physiol.* 466, 861–872. <http://dx.doi.org/10.1007/s00424-013-1357-2>
- Calduch-Giner, J.A., Duval, H., Chesnel, F., Boeuf, G., Pérez-Sánchez, J., Boujard, D., 2001. Fish growth hormone receptor: molecular characterization of two membrane-anchored forms. *Endocrinology* 142, 3269–73. <http://dx.doi.org/10.1210/endo.142.7.8407>
- Calduch-Giner, J.A., Mingarro, M., Vega-Rubín de Celis, S., Boujard, D., Pérez-Sánchez, J., 2003. Molecular cloning and characterization of gilthead sea bream (*Sparus aurata*) growth hormone receptor (GHR). Assessment of alternative splicing. *Comp. Biochem. Physiol. B. Biochem. Mol. Biol.* 136, 1–13. [http://dx.doi.org/10.1016/S1096-4959\(03\)00150-7](http://dx.doi.org/10.1016/S1096-4959(03)00150-7)
- Chen, M., Huang, X., Yuen, D.S.H., Cheng, C.H.K., 2011. A study on the functional interaction between the GH/PRL family of polypeptides with their receptors in zebrafish: Evidence against GHR1 being the receptor for somatolactin. *Mol. Cell. Endocrinol.* 337, 114–21.

- <http://dx.doi.org/10.1016/j.mce.2011.02.006>
- Dagil, R., Knudsen, M.J., Olsen, J.G., O'Shea, C., Franzmann, M., Goffin, V., Teilum, K., Breinholt, J., Kragelund, B.B., 2012. The WSXWS motif in cytokine receptors is a molecular switch involved in receptor activation: Insight from structures of the prolactin receptor. *Structure* 20, 270–282. <http://dx.doi.org/10.1016/j.str.2011.12.010>
- Dickson, V.K., Pedi, L., Long, S.B., 2014. Structure and insights into the function of a Ca²⁺-activated Cl⁻ channel. *Nature* 516, 213–218. <http://dx.doi.org/10.1038/nature13913>
- Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–7. <http://dx.doi.org/10.1093/nar/gkh340>
- Ellens, E.R., Kittilson, J.D., Hall, J.A., Sower, S.A., Sheridan, M.A., 2013. Evolutionary origin and divergence of the growth hormone receptor family: insight from studies on sea lamprey. *Gen. Comp. Endocrinol.* 192, 222–36. <http://dx.doi.org/10.1016/j.ygcen.2013.05.008>
- Fernández-Suárez, X.M., Schuster, M.K., 2010. Using the ensembl genome server to browse genomic sequence data. *Curr. Protoc. Bioinforma.* Chapter 1, Unit1.15. <http://dx.doi.org/10.1002/0471250953.bi0115s30>
- Finn, R.D., Clements, J., Eddy, S.R., 2011. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 39 Suppl 2, W29–37. <http://dx.doi.org/10.1093/nar/gkr367>
- Fiol, D.F., Sanmarti, E., Sacchi, R., Kültz, D., 2009. A novel tilapia prolactin receptor is functionally distinct from its paralog. *J. Exp. Biol.* 212, 2007–2015. <http://dx.doi.org/10.1242/jeb.025601>
- Forsyth, I.A., Wallis, M., 2002. Growth hormone and prolactin--molecular and functional evolution. *J. Mammary Gland Biol. Neoplasia* 7, 291–312. <http://dx.doi.org/10.1023/A:1022804817104>
- Freeman, M.E., Kanyicska, B., Lerant, A., Nagy, G., 2000. Prolactin: structure, function, and regulation of secretion. *Physiol. Rev.* 80, 1523–631. <http://physrev.physiology.org/content/80/4/1523>
- Fukada, H., Ozaki, Y., Pierce, A.L., Adachi, S., Yamauchi, K., Hara, A., Swanson, P., Dickhoff, W.W., 2005. Identification of the salmon somatolactin receptor, a new member of the cytokine receptor family. *Endocrinology* 146, 2354–61. <http://dx.doi.org/10.1210/en.2004-1578>
- Fukamachi, S., Meyer, A., 2007. Evolution of receptors for growth hormone and somatolactin in fish and land vertebrates: lessons from the lungfish and sturgeon orthologues. *J. Mol. Evol.* 65, 359–72. <http://dx.doi.org/10.1007/s00239-007-9035-7>
- Fukamachi, S., Yada, T., Meyer, A., Kinoshita, M., 2009. Effects of constitutive expression of somatolactin alpha on skin pigmentation in medaka. *Gene* 442, 81–7. <http://dx.doi.org/10.1016/j.gene.2009.04.010>
- Fukamachi, S., Yada, T., Mitani, H., 2005. Medaka receptors for somatolactin and growth hormone: phylogenetic paradox among fish growth hormone receptors. *Genetics* 171, 1875–83. <http://dx.doi.org/10.1534/genetics.105.048819>
- Gao, F.-Y., Lu, M.-X., Ye, X., Huang, Z.-H., Wang, H., Zhu, H.-P., Yang, L.-P., 2011. Identification and expression analysis of two growth hormone receptors in zanzibar tilapia (*Oreochromis hornorum*). *Fish Physiol. Biochem.* 37, 553–65. <http://dx.doi.org/10.1007/s10695-010-9457-9>
- Glasauer, S.M.K., Neuhauss, S.C.F., 2014. Whole-genome duplication in teleost fishes and its evolutionary consequences. *Mol. Genet. Genomics* 289, 1045–1060.

- <http://dx.doi.org/10.1007/s00438-014-0889-2>
- Gouy, M., Guindon, S., Gascuel, O., 2010. SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. *Mol. Biol. Evol.* 27, 221–224. <http://dx.doi.org/10.1093/molbev/msp259>
- Grattan, D.R., Kokay, I.C., 2008. Prolactin: a pleiotropic neuroendocrine hormone. *J. Neuroendocrinol.* 20, 752–63. <http://dx.doi.org/10.1111/j.1365-2826.2008.01736.x>
- 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, 307–21. <http://dx.doi.org/10.1093/sysbio/syq010>
- Han, L., Su, B., Li, W.-H., Zhao, Z., 2008. CpG island density and its correlations with genomic features in mammalian genomes. *Genome Biol.* 9, R79. <http://dx.doi.org/10.1186/gb-2008-9-5-r79>
- Hätinen T., Holm L. and Airaksinen M.S., 2007. Loss of neurturin in frog-Comparative genomics study of GDNF family ligand-receptor pairs, *Mol. Cell. Neurosci.* 34 (2), 155–167. <http://dx.doi.org/10.1016/j.mcn.2006.10.009>
- Hess, P.N., De Moraes Russo, C.A., 2007. An empirical test of the midpoint rooting method. *Biol. J. Linn. Soc. Lond.* 92, 669–674. <http://dx.doi.org/10.1111/j.1095-8312.2007.00864.x>
- Holland, M.C.H., Lambris, J.D., 2002. The complement system in teleosts. *Fish Shellfish Immunol.* 12, 399–420. <http://dx.doi.org/10.1006/fsim.2001.0408>
- Hron, T., Pajer, P., Pačes, J., Bartůněk, P., Elleder, D., 2015. Hidden genes in birds. *Genome Biol.* 16, 164. <http://dx.doi.org/10.1186/s13059-015-0724-z>
- Huang, X., Jiao, B., Fung, C.K., Zhang, Y., Ho, W.K.K., Chan, C.B., Lin, H., Wang, D., Cheng, C.H.K., 2007. The presence of two distinct prolactin receptors in seabream with different tissue distribution patterns, signal transduction pathways and regulation of gene expression by steroid hormones. *J. Endocrinol.* 194, 373–92. <http://dx.doi.org/10.1677/JOE-07-0076>
- Huising, M.O., Kruiswijk, C.P., Flik, G., 2006. Phylogeny and evolution of class-I helical cytokines. *J. Endocrinol.* 189, 1–25. <http://dx.doi.org/10.1677/joe.1.06591>
- Hultqvist, G., Ocampo Daza, D., Larhammar, D., Kilimann, M.W., 2012. Evolution of the vertebrate paralemmin gene family: ancient origin of gene duplicates suggests distinct functions. *PLoS One* 7, e41850. <http://dx.doi.org/10.1371/journal.pone.0041850>
- Huxley-Jones, J., Clarke, T.-K., Beck, C., Toubaris, G., Robertson, D.L., Boot-Handford, R.P., 2007. The evolution of the vertebrate metzincins; insights from *Ciona intestinalis* and *Danio rerio*. *BMC Evol. Biol.* 7, 63. <http://dx.doi.org/10.1186/1471-2148-7-63>
- Iso-Touru, T., Kantanen, J., Li, M.-H., Gizejewski, Z., Vilkkii, J., 2009. Divergent evolution in the cytoplasmic domains of PRLR and GHR genes in Artiodactyla. *BMC Evol. Biol.* 9, 172. <http://dx.doi.org/10.1186/1471-2148-9-172>
- Itoh, N., Ornitz, D.M., 2011. Fibroblast growth factors: From molecular evolution to roles in development, metabolism and disease. *J. Biochem.* 149, 121–130. <http://dx.doi.org/10.1093/jb/mvq121>
- Jaillon, O., Aury, J.-M., Brunet, F.G., Petit, J.-L., Stange-Thomann, N., Mauceli, E., Bouneau, L., Fischer, C., Ozouf-Costaz, C., Bernot, A., Nicaud, S., Jaffe, D.B., Fisher, S., Lutfalla, G., Dossat, C., Segurens, B., Dasilva, C., Salanoubat, M., Levy, M., Boudet, N., Castellano, S., Anthouard, V., Jubin, C., Castelli, V., Katinka, M., Vacherie, B., Biémont, C., Skalli, Z., Cattolico, L., Poulain, J., De Berardinis, V., Cruaud, C., Duprat, S., Brottier, P.,

- Coutanceau, J.-P., Gouzy, J., Parra, G., Lardier, G., Chapple, C., McKernan, K.J., McEwan, P., Bosak, S., Kellis, M., Volff, J.-N., Guigó, R., Zody, M.C., Mesirov, J., Lindblad-Toh, K., Birren, B., Nusbaum, C., Kahn, D., Robinson-Rechavi, M., Laudet, V., Schachter, V., Quétier, F., Saurin, W., Scarpelli, C., Wincker, P., Lander, E.S., Weissenbach, J., Roest Crollius, H., 2004. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* 431, 946–57.
<http://dx.doi.org/10.1038/nature03025>
- Jaleel, M., McBride, A., Lizcano, J.M., Deak, M., Toth, R., Morrice, N.A., Alessi, D.R., 2005. Identification of the sucrose non-fermenting related kinase SNRK, as a novel LKB1 substrate. *FEBS Lett.* 579, 1417–1423. <http://dx.doi.org/10.1016/j.febslet.2005.01.042>
- Jiao, B., Huang, X., Chan, C.B., Zhang, L., Wang, D., Cheng, C.H.K., 2006. The co-existence of two growth hormone receptors in teleost fish and their differential signal transduction, tissue distribution and hormonal regulation of expression in seabream. *J. Mol. Endocrinol.* 36, 23–40. <http://dx.doi.org/10.1677/jme.1.01945>
- Kasahara, M., Naruse, K., Sasaki, S., Nakatani, Y., Qu, W., Ahsan, B., Yamada, T., Nagayasu, Y., Doi, K., Kasai, Y., Jindo, T., Kobayashi, D., Shimada, A., Toyoda, A., Kuroki, Y., Fujiyama, A., Sasaki, T., Shimizu, A., Asakawa, S., Shimizu, N., Hashimoto, S.-I., Yang, J., Lee, Y., Matsushima, K., Sugano, S., Sakaizumi, M., Narita, T., Ohishi, K., Haga, S., Ohta, F., Nomoto, H., Nogata, K., Morishita, T., Endo, T., Shin-I, T., Takeda, H., Morishita, S., Kohara, Y., 2007. The medaka draft genome and insights into vertebrate genome evolution. *Nature* 447, 714–9. <http://dx.doi.org/10.1038/nature05846>
- Katsanis, N., Fitzgibbon, J., Fisher, E.M.C., 1996. Paralogy Mapping: Identification of a Region in the Human MHC Triplicated onto Human Chromosomes 1 and 9 Allows the Prediction and Isolation of Novel PBX and NOTCH Loci. *Genomics* 35, 101–108.
<http://dx.doi.org/10.1006/geno.1996.0328>
- Kirischian, N.L., Wilson, J.Y., 2012. Phylogenetic and functional analyses of the cytochrome P450 family 4. *Mol. Phylogenet. Evol.* 62, 458–471.
<http://dx.doi.org/10.1016/j.ympev.2011.10.016>
- Larsson, A., 2014. AliView: a fast and lightweight alignment viewer and editor for large data sets. *Bioinformatics* 30, btu531-. <http://dx.doi.org/10.1093/bioinformatics/btu531>
- Le, S.Q., Gascuel, O., 2008. An improved general amino acid replacement matrix. *Mol. Biol. Evol.* 25, 1307–20. <http://dx.doi.org/10.1093/molbev/msn067>
- Li, M., Gao, Z., Ji, D., Zhang, S., 2014. Functional characterization of GH-like homolog in amphioxus reveals an ancient origin of GH/GH receptor system. *Endocrinology* 155, 4818–30. <http://dx.doi.org/10.1210/en.2014-1377>
- Li, Y., Liu, X., Zhang, Y., Zhu, P., Lin, H., 2007. Molecular cloning, characterization and distribution of two types of growth hormone receptor in orange-spotted grouper (*Epinephelus coioides*). *Gen. Comp. Endocrinol.* 152, 111–22.
<http://dx.doi.org/10.1016/j.ygcen.2007.01.033>
- Lien, S., Koop, B.F., Sandve, S.R., Miller, J.R., Kent, M.P., Nome, T., Hvidsten, T.R., Leong, J.S., Minkley, D.R., Zimin, A., Grammes, F., Grove, H., Gjuvsland, A., Walenz, B., Hermansen, R.A., von Schalburg, K., Rondeau, E.B., Di Genova, A., Samy, J.K.A., Olav Vik, J., Vigeland, M.D., Caler, L., Grimholt, U., Jentoft, S., Inge Våge, D., de Jong, P., Moen, T., Baranski, M., Palti, Y., Smith, D.R., Yorke, J.A., Nederbragt, A.J., Tooming-Klunderud, A., Jakobsen, K.S., Jiang, X., Fan, D., Hu, Y., Liberles, D.A., Vidal, R., Iturra, P., Jones, S.J.M., Jonassen, I., Maass, A., Omholt, S.W., Davidson, W.S., 2016. The Atlantic salmon genome provides insights into rediploidization. *Nature* 533, 200–205.

- <http://dx.doi.org/10.1038/nature17164>
- Liongue, C., Ward, A.C., 2007. Evolution of Class I cytokine receptors. *BMC Evol. Biol.* 7, 120. <http://dx.doi.org/10.1186/1471-2148-7-120>
- Ma, Q., Liu, S.-F., Zhuang, Z.-M., Sun, Z.-Z., Liu, C.-L., Tang, Q.-S., 2012. The co-existence of two growth hormone receptors and their differential expression profiles between female and male tongue sole (*Cynoglossus semilaevis*). *Gene* 511, 341–52. <http://dx.doi.org/10.1016/j.gene.2012.09.082>
- Macqueen, D.J., Garcia de la serrana, D., Johnston, I.A., 2013. Evolution of Ancient Functions in the Vertebrate Insulin-Like Growth Factor System Uncovered by Study of Duplicated Salmonid Fish Genomes. *Mol. Biol. Evol.* 30, 1060–1076. <http://dx.doi.org/10.1093/molbev/mst017>
- Matchkov, V. V., Boedtkjer, D.M., Aalkjaer, C., 2015. The role of Ca²⁺ activated Cl⁻ channels in blood pressure control. *Curr. Opin. Pharmacol.* 21, 127–137. <http://dx.doi.org/10.1016/j.coph.2015.02.003>
- McQueen, H.A., Fantes, J., Cross, S.H., Clark, V.H., Archibald, A.L., Bird, A.P., 1996. CpG islands of chicken are concentrated on microchromosomes. *Nat. Genet.* 12, 321–4. <http://dx.doi.org/10.1038/ng0396-321>
- Mehta, T.K., Ravi, V., Yamasaki, S., Lee, A.P., Lian, M.M., Tay, B.-H., Tohari, S., Yanai, S., Tay, A., Brenner, S., Venkatesh, B., 2013. Evidence for at least six Hox clusters in the Japanese lamprey (*Lethenteron japonicum*). *Proc. Natl. Acad. Sci. U. S. A.* 110, 16044–9. <http://dx.doi.org/10.1073/pnas.1315760110>
- Menzies, B.R., Shaw, G., Fletcher, T.P., Pask, A.J., Renfree, M.B., 2008. Exon 3 of the growth hormone receptor (GH-R) is specific to eutherian mammals. *Mol. Cell. Endocrinol.* 296, 64–8. <http://dx.doi.org/10.1016/j.mce.2008.07.010>
- 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, 1254–65. <http://dx.doi.org/10.1101/gr.6316407>
- Narumiya, S., Sugimoto, Y., Ushikubi, F., 1999. Prostanoid receptors: structures, properties, and functions. *Physiol. Rev.* 79, 1193–226. <http://physrev.physiology.org/content/79/4/1193>
- Nelson, D.R., Zeldin, D.C., Hoffman, S.M.G., Maltais, L.J., Wain, H.M., Nebert, D.W., 2004. Comparison of cytochrome P450 (CYP) genes from the mouse and human genomes, including nomenclature recommendations for genes, pseudogenes and alternative-splice variants. *Pharmacogenetics* 14, 1–18. <http://dx.doi.org/10.1097/01.fpc.0000054151.92680.31>
- Noh, G.E., Lim, H.K., Kim, J.-M., 2012. Characterization of genes encoding prolactin and prolactin receptors in starry flounder *Platichthys stellatus* and their expression upon acclimation to freshwater. *Fish Physiol. Biochem.* <http://dx.doi.org/10.1007/s10695-012-9697-y>
- Olinski, R.P., Lundin, L.-G., Hallböök, F., 2006. Conserved synteny between the Ciona genome and human paralogs identifies large duplication events in the molecular evolution of the insulin-relaxin gene family. *Mol. Biol. Evol.* 23, 10–22. <http://dx.doi.org/10.1093/molbev/msj002>
- Ono, M., Takayama, Y., Rand-Weaver, M., Sakata, S., Yasunaga, T., Noso, T., Kawauchi, H., 1990. cDNA cloning of somatolactin, a pituitary protein related to growth hormone and prolactin. *Proc. Natl. Acad. Sci. U. S. A.* 87, 4330–4. <http://dx.doi.org/10.1073/pnas.87.11.4330>

- Oulion, S., Bertrand, S., Escriva, H., 2012. Evolution of the FGF Gene Family. *Int. J. Evol. Biol.* 2012, 1–12. <http://dx.doi.org/10.1155/2012/298147>
- Ozaki, Y., Fukada, H., Kazeto, Y., Adachi, S., Hara, A., Yamauchi, K., 2006. Molecular cloning and characterization of growth hormone receptor and its homologue in the Japanese eel (*Anguilla japonica*). *Comp. Biochem. Physiol. B. Biochem. Mol. Biol.* 143, 422–31. <http://dx.doi.org/10.1016/j.cbpb.2005.12.016>
- Pierce, A.L., Breves, J.P., Moriyama, S., Uchida, K., Grau, E.G., 2012. Regulation of growth hormone (GH) receptor (GHR1 and GHR2) mRNA level by GH and metabolic hormones in primary cultured tilapia hepatocytes. *Gen. Comp. Endocrinol.* 179, 22–9. <http://dx.doi.org/10.1016/j.ygcen.2012.07.010>
- Prosdocimi, F., Linard, B., Pontarotti, P., Poch, O., Thompson, J.D., 2012. Controversies in modern evolutionary biology: the imperative for error detection and quality control. *BMC Genomics* 13, 5. <http://dx.doi.org/10.1186/1471-2164-13-5>
- Putnam, N.H., Butts, T., Ferrier, D.E.K., Furlong, R.F., Hellsten, U., Kawashima, T., Robinson-Rechavi, M., Shoguchi, E., Terry, A., Yu, J.-K., Benito-Gutiérrez, È., Dubchak, I., Garcia Fernández, J., Gibson-Brown, J.J., Grigoriev, I. V., Horton, A.C., de Jong, P.J., Jurka, J., Kapitonov, V. V., Kohara, Y., Kuroki, Y., Lindquist, E., Lucas, S.M., Osoegawa, K., Pennacchio, L.A., Salamov, A.A., Satou, Y., Sauka-Spengler, T., Schmutz, J., Shin-I, T., Toyoda, A., Bronner-Fraser, M., Fujiyama, A., Holland, L.Z., Holland, P.W.H., Satoh, N., Rokhsar, D.S., 2008. The amphioxus genome and the evolution of the chordate karyotype. *Nature* 453, 1064–71. <http://dx.doi.org/10.1038/nature06967>
- Rand-Weaver, M., Noso, T., Muramoto, K., Kawauchi, H., 1991. Isolation and characterization of somatolactin, a new protein related to growth hormone and prolactin from Atlantic cod (*Gadus morhua*) pituitary glands. *Biochemistry* 30, 1509–15. <http://dx.doi.org/10.1021/bi00220a010>
- Reindl, K.M., Kittilson, J.D., Sheridan, M.A., 2009. Differential ligand binding and agonist-induced regulation characteristics of the two rainbow trout GH receptors, Ghr1 and Ghr2, in transfected cells. *J. Endocrinol.* 202, 463–71. <http://dx.doi.org/10.1677/JOE-09-0057>
- Rosental, B., Kowarsky, M.A., Corey, D.M., Ishizuka, K.J., Palmeri, K.J., Chen, S.-Y., Sinha, R., Voskoboinik, A., Weissman, I.L., 2016. Finding the evolutionary precursor of vertebrate hematopoietic lineage: Functional and molecular characterization of *B. schlosseri* immune system. *J. Immunol.* 196 (1 Supplement), 216.2. http://www.jimmunol.org/content/196/1_Supplement/216.2
- Ross, R.J., Esposito, N., Shen, X.Y., Von Laue, S., Chew, S.L., Dobson, P.R., Postel-Vinay, M.C., Finidori, J., 1997. A short isoform of the human growth hormone receptor functions as a dominant negative inhibitor of the full-length receptor and generates large amounts of binding protein. *Mol. Endocrinol.* 11, 265–73. <http://dx.doi.org/10.1210/mend.11.3.9901>
- Sabath, D.F., Kaushansky, K., Broudy, V.C., 1999. Deletion of the extracellular membrane-distal cytokine receptor homology module of Mpl results in constitutive cell growth and loss of thrombopoietin binding. *Blood* 94, 365–367. <http://www.bloodjournal.org/content/94/1/365>
- Saera-Vila, A., Caldusch-Giner, J.A., Pérez-Sánchez, J., 2005. Duplication of growth hormone receptor (GHR) in fish genome: gene organization and transcriptional regulation of GHR type I and II in gilthead sea bream (*Sparus aurata*). *Gen. Comp. Endocrinol.* 142, 193–203. <http://dx.doi.org/10.1016/j.ygcen.2004.11.005>

- Sasano, Y., Yoshimura, A., Fukamachi, S., 2012. Reassessment of the function of somatolactin alpha in lipid metabolism using medaka mutant and transgenic strains. *BMC Genet.* 13, 64. <http://dx.doi.org/10.1186/1471-2156-13-64>
- Satou, Y., Imai, K.S., Satoh, N., 2002. Fgf genes in the basal chordate *Ciona intestinalis*. *Dev. Genes Evol.* 212, 432–438. <http://dx.doi.org/10.1007/s00427-002-0266-8>
- Saunders, M.C., Gemmell, R.T., Waters, M.J., Curlew, J.D., 2002. Plasma growth hormone and growth hormone-binding protein during development in the marsupial brushtail possum (*Trichosurus vulpecula*). *J. Endocrinol.* 173, 507–15. <http://dx.doi.org/10.1677/joe.0.1730507>
- Smith, J.J., Kuraku, S., Holt, C., Sauka-Spengler, T., Jiang, N., Campbell, M.S., Yandell, M.D., Manousaki, T., Meyer, A., Bloom, O.E., Morgan, J.R., Buxbaum, J.D., Sachidanandam, R., Sims, C., Garruss, A.S., Cook, M., Krumlauf, R., Wiedemann, L.M., Sower, S.A., Decatur, W.A., Hall, J.A., Amemiya, C.T., Saha, N.R., Buckley, K.M., Rast, J.P., Das, S., Hirano, M., McCurley, N., Guo, P., Rohner, N., Tabin, C.J., Piccinelli, P., Elgar, G., Ruffier, M., Aken, B.L., Searle, S.M.J., Muffato, M., Pignatelli, M., Herrero, J., Jones, M., Brown, C.T., Chung-Davidson, Y.-W., Nanlohy, K.G., Libants, S. V, Yeh, C.-Y., McCauley, D.W., Langeland, J.A., Pancer, Z., Frittsch, B., de Jong, P.J., Zhu, B., Fulton, L., Theising, B., Flicek, P., Bronner, M.E., Warren, W.C., Clifton, S.W., Wilson, R.K., Li, W., 2013. Sequencing of the sea lamprey (*Petromyzon marinus*) genome provides insights into vertebrate evolution. *Nat. Genet.* 45, 415–21. <http://dx.doi.org/10.1038/ng.2568>
- Sohm, F., Manfroid, I., Pezet, A., Rentier-Delrue, F., Rand-Weaver, M., Kelly, P.A., Boeuf, G., Postel-Vinay, M.C., de Luze, A., Edery, M., 1998. Identification and modulation of a growth hormone-binding protein in rainbow trout (*Oncorhynchus mykiss*) plasma during seawater adaptation. *Gen. Comp. Endocrinol.* 111, 216–24. <http://dx.doi.org/10.1006/gcen.1998.7106>
- Tanaka, M., Maeda, K., Okubo, T., Nakashima, K., 1992. Double antenna structure of chicken prolactin receptor deduced from the cDNA sequence. *Biochem. Biophys. Res. Commun.* 188, 490–6. [http://dx.doi.org/10.1016/0006-291X\(92\)91082-2](http://dx.doi.org/10.1016/0006-291X(92)91082-2)
- Tanaka, M., Yamamoto, I., Hayashida, Y., Nakao, N., Ohkubo, T., Wakita, M., Nakashima, K., 2000. Two novel first exons in the prolactin receptor gene are transcribed in a tissue-specific and sexual maturation-dependent manner to encode multiple 5'-truncated transcripts in the testis of the chicken. *Biochim. Biophys. Acta* 1491, 279–284. [http://dx.doi.org/10.1016/S0167-4781\(00\)00019-1](http://dx.doi.org/10.1016/S0167-4781(00)00019-1)
- Thomas, J.H., 2007. Rapid birth-death evolution specific to xenobiotic cytochrome P450 genes in vertebrates. *PLoS Genet.* 3, 720–728. <http://dx.doi.org/10.1371/journal.pgen.0030067>
- Tse, D.L.Y., Tse, M.C.L., Chan, C.B., Deng, L., Zhang, W.M., Lin, H.R., Cheng, C.H.K., 2003. Seabream growth hormone receptor: molecular cloning and functional studies of the full-length cDNA, and tissue expression of two alternatively spliced forms. *Biochim. Biophys. Acta* 1625, 64–76. [http://dx.doi.org/10.1016/S0167-4781\(02\)00591-2](http://dx.doi.org/10.1016/S0167-4781(02)00591-2)
- Tsuge, K., Iwasaki, R., Morimoto, K., Inazumi, T., Kawahara, O., Kawahara, A., Tsuchiya, S., Sugimoto, Y., 2013. Molecular and pharmacological characterization of zebrafish “contractile” and “inhibitory” prostanoid receptors. *Biochem. Biophys. Res. Commun.* 438, 353–358. <http://dx.doi.org/10.1016/j.bbrc.2013.07.075>
- Uchida, K., Moriyama, S., Breves, J.P., Fox, B.K., Pierce, A.L., Borski, R.J., Hirano, T., Grau, E.G., 2009. cDNA cloning and isolation of somatolactin in Mozambique tilapia and effects of seawater acclimation, confinement stress, and fasting on its pituitary

- expression. *Gen. Comp. Endocrinol.* 161, 162–70.
<http://dx.doi.org/10.1016/j.ygcen.2008.11.031>
- Very, N.M., Kittilson, J.D., Norbeck, L.A., Sheridan, M.A., 2005. Isolation, characterization, and distribution of two cDNAs encoding for growth hormone receptor in rainbow trout (*Oncorhynchus mykiss*). *Comp. Biochem. Physiol. B. Biochem. Mol. Biol.* 140, 615–28.
<http://dx.doi.org/10.1016/j.cbpc.2004.12.008>
- Voskoboynik, A., Neff, N.F., Sahoo, D., Newman, A.M., Pushkarev, D., Koh, W., Passarelli, B., Fan, H.C., Mantalas, G.L., Palmeri, K.J., Ishizuka, K.J., Gissi, C., Griggio, F., Ben-Shlomo, R., Corey, D.M., Penland, L., White, R.A., Weissman, I.L., Quake, S.R., 2013. The genome sequence of the colonial chordate, *Botryllus schlosseri*. *Elife* 2013, 1–24.
<http://dx.doi.org/10.7554/eLife.00569>
- Walock, C.N., Kittilson, J.D., Sheridan, M.A., 2014. Characterization of a novel growth hormone receptor-encoding cDNA in rainbow trout and regulation of its expression by nutritional state. *Gene* 533, 286–294. <http://dx.doi.org/10.1016/j.gene.2013.09.046>
- Wolkowicz U.M. and Cook A.G., 2012. NF45 dimerizes with NF90, Zfr and SPNR via a conserved domain that has a nucleotidyltransferase fold, *Nucleic Acids Res.* 40 (18), 9356–9368. <http://dx.doi.org/10.1093/nar/gks696>
- Yates, A., Akanni, W., Amode, M.R., Barrell, D., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., Fitzgerald, S., Gil, L., Girón, C.G., Gordon, L., Hourlier, T., Hunt, S.E., Janacek, S.H., Johnson, N., Juettemann, T., Keenan, S., Lavidas, I., Martin, F.J., Maurel, T., McLaren, W., Murphy, D.N., Nag, R., Nuhn, M., Parker, A., Patricio, M., Pignatelli, M., Rahtz, M., Riat, H.S., Sheppard, D., Taylor, K., Thormann, A., Vullo, A., Wilder, S.P., Zadissa, A., Birney, E., Harrow, J., Muffato, M., Perry, E., Ruffier, M., Spudich, G., Trevanion, S.J., Cunningham, F., Aken, B.L., Zerbino, D.R., Flicek, P., 2016. Ensembl 2016. *Nucleic Acids Res.* 44, D710–D716. <http://dx.doi.org/10.1093/nar/gkv1157>
- Yegorov, S., Good, S., 2012. Using paleogenomics to study the evolution of gene families: Origin and duplication history of the relaxin family hormones and their receptors. *PLoS One* 7, 1–12. <http://dx.doi.org/10.1371/journal.pone.0032923>
- ZFIN Zebrafish Nomenclature Guidelines [WWW Document], 2016. . Zebrafish Model Org. Database. URL <https://wiki.zfin.org/x/d4BOAQ> (accessed 1.24.17).

Appendix A. Supplementary data

The following supplementary content has been shared to

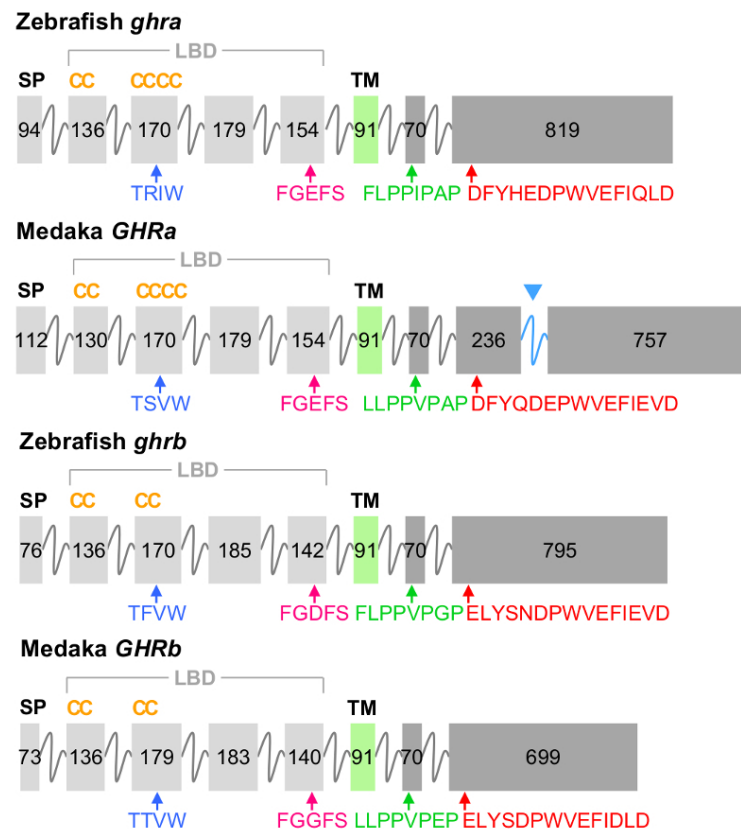
<https://doi.org/10.6084/m9.figshare.5151157>.

Supplementary data 1. Location data, sequence identifiers and prediction/annotation notes for all identified GHR, PRLR, CRFA4, EPOR and TPOR (MPL) sequences. The table also includes genome assembly and sequence quality information.

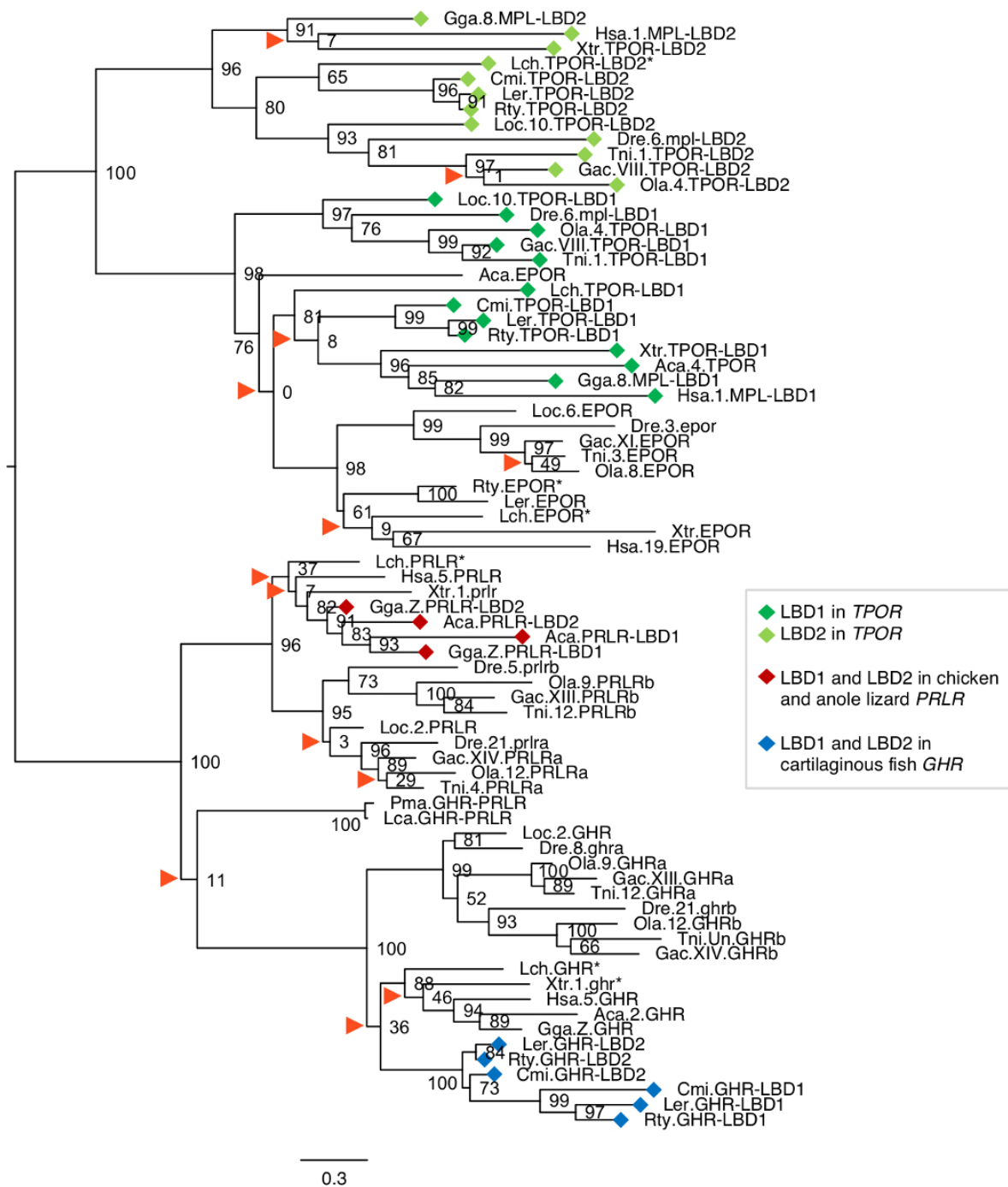
Supplementary data 2. Location data, sequence identifiers and prediction/annotation notes for all neighboring gene families. The included tables also detail sequence quality information and outgroup choice. Sequence names follow approved gene symbols and nomenclature guidelines for human, mouse, chicken, Western clawed frog and zebrafish. For gene predictions where no gene symbol could be assigned, the sequences are named for their assigned chromosomes, linkage groups or scaffolds only. Duplicates generated in the basal teleost tetraploidization (3R) are assigned the lower-case letters “a” and “b”, following zebrafish guidelines. Local duplicates in spotted gar and teleost fishes are assigned numeral suffixes, “.1”, “.2” et. c., also following zebrafish guidelines. For lineage-specific duplicates of unclear origin we have assigned the roman numerals, “-I” and “-II”. Species abbreviations: Human (Hsa), mouse (Mmu), chicken (Gga), Carolina anole lizard (Aca), Western clawed frog (Xtr), coelacanth (Lch), spotted gar (Loc), zebrafish (Dre), medaka (Ola), three-spined stickleback (Gac), green spotted pufferfish (Tni), Japanese pufferfish/fugu (Tru), vase tunicate (Cin), Florida lancelet (Bfl), common lancelet (Bla), Belcher’s lancelet (Bbe), purple sea urchin (Spu), fruit fly (Dme).

Supplementary data 3. Gene lists from GHR, PRLR, EPOR and TPOR gene-bearing chromosome blocks in the human and medaka genomes. Gene lists from each chromosome block were downloaded from Ensembl BioMart, colour-coded according to chromosome, and sorted by Ensembl protein family IDs and the number of chromosome blocks on which each protein family has member genes. In this way, we were able to identify gene families represented across several chromosome regions, i.e. conserved synten. “List 1” shows Ensembl protein families shared between at least two of the human chromosome regions. “List 2” shows Ensembl protein families shared between at least two of the medaka GHRA, GHRb, PRLRa and PRLRb gene-bearing regions as well as the human GHR and PRLR gene-bearing region. “Hsa” designates human chromosomes, “Ola” designates medaka chromosomes.

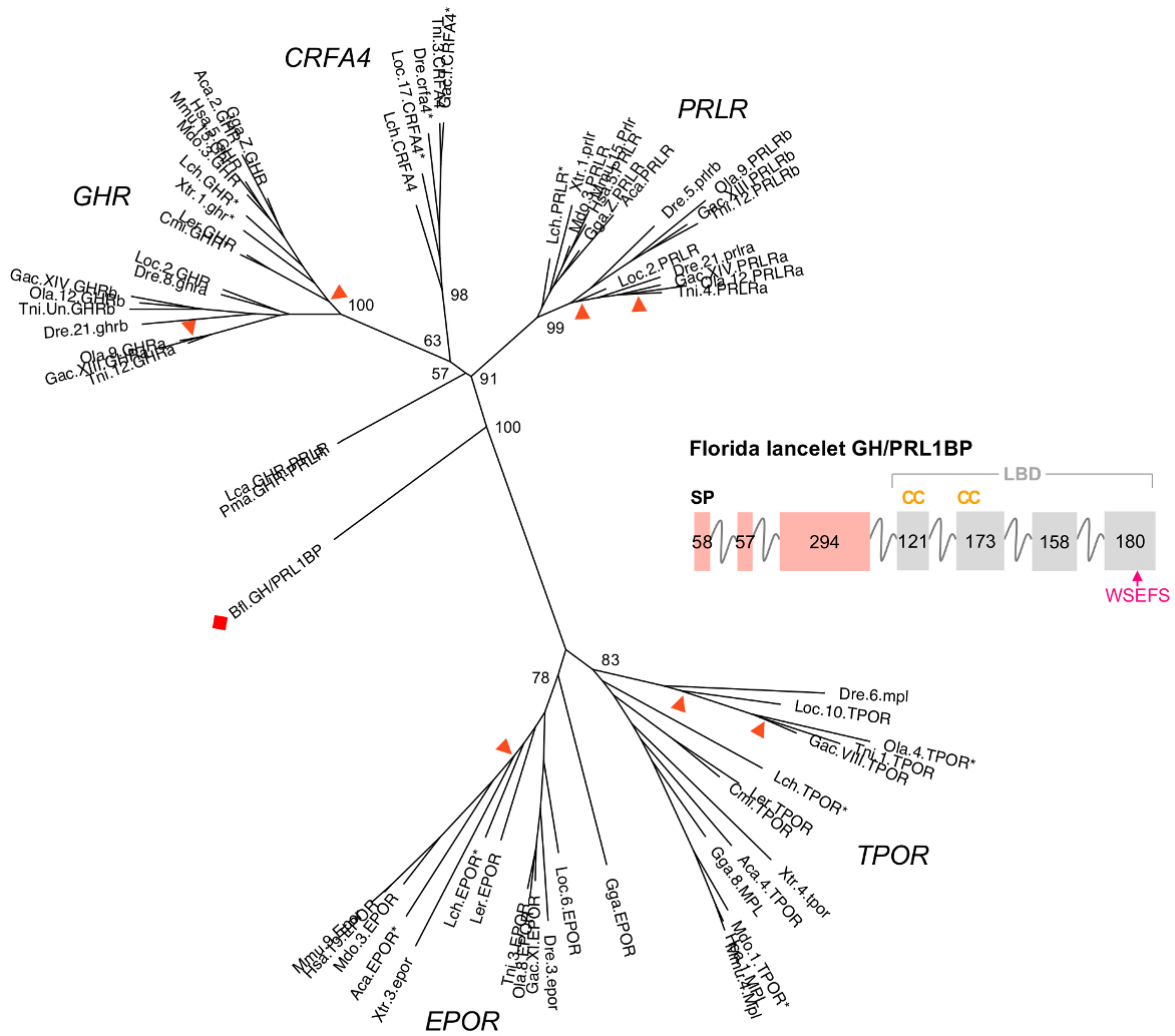
Supplementary figures



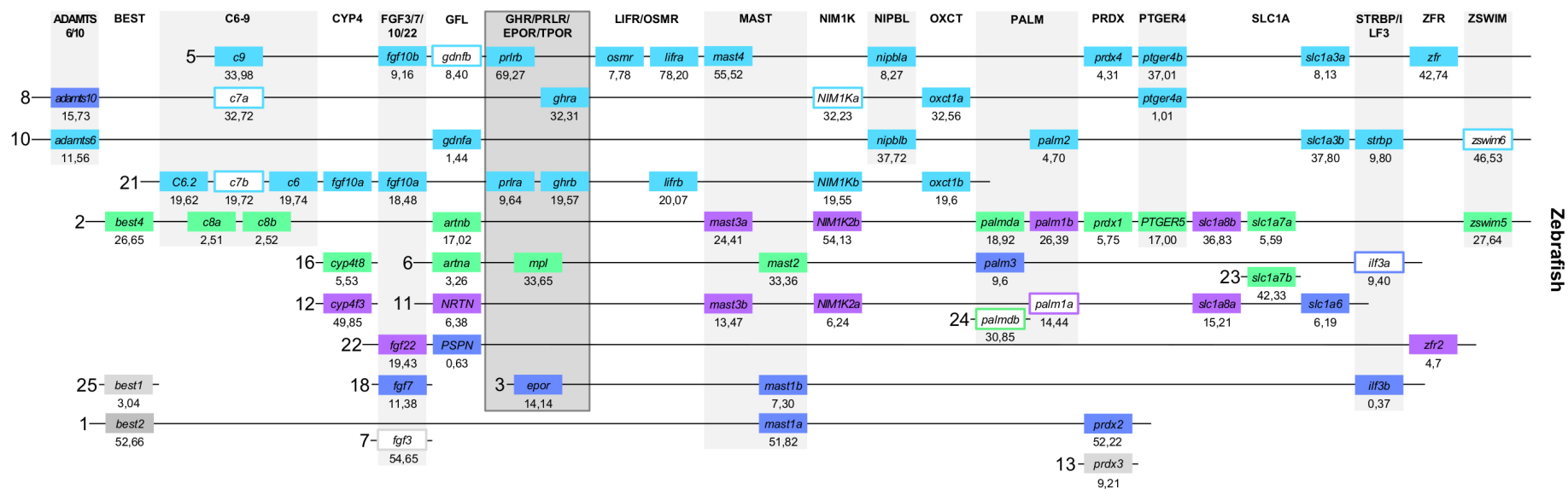
Supplementary Figure S1. Coding exon organizations of representative teleost *GHRa*, *GHRb*, *PRLRa* and *PRLRb* genes. The Atlantic cod, medaka, Southern platyfish, three-spined stickleback, Japanese pufferfish (fugu) and green spotted pufferfish *GHRa* genes have an additional 3' intron (in blue) that was not found in any other of the analyzed genes.



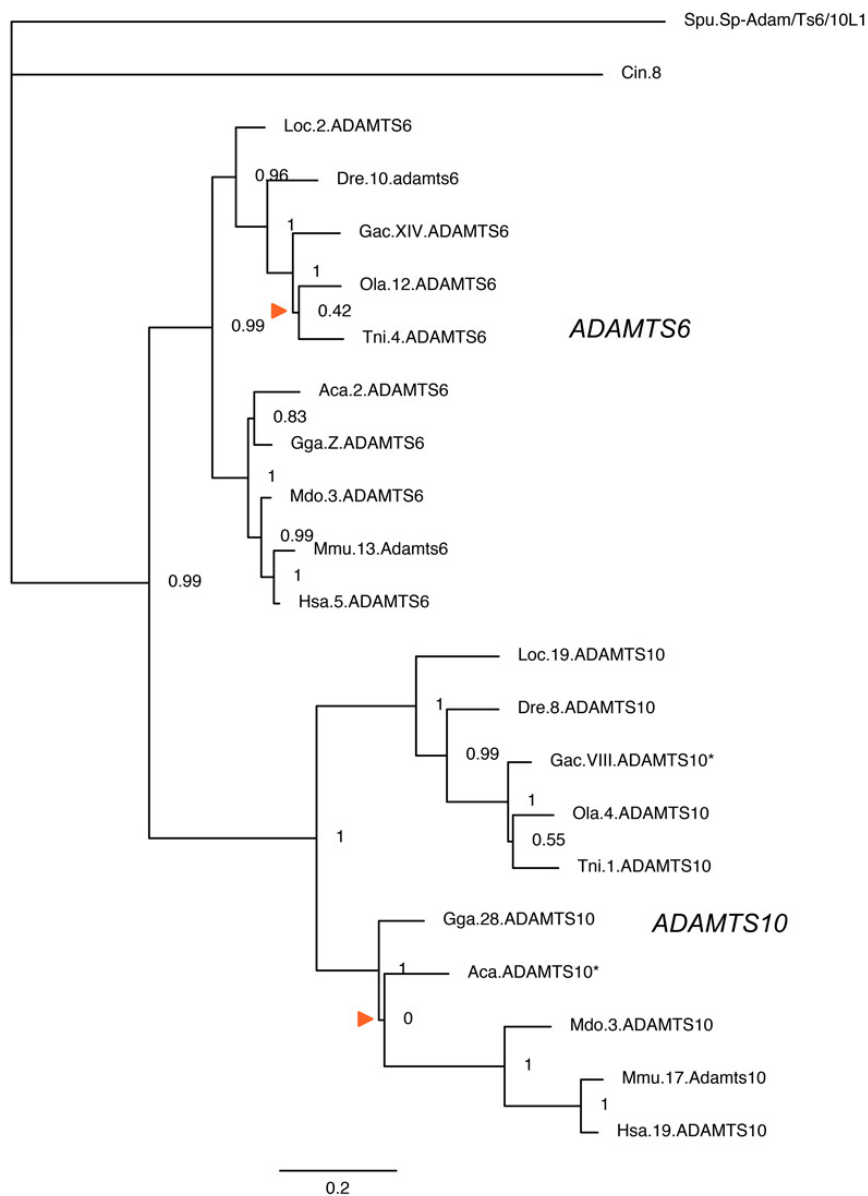
Supplementary Figure S2. Phylogenetic analysis of ligand binding domain (LBD) sequences. Uninformative nodes (support $\leq 50\%$) are marked with arrowheads. Species abbreviations: Human (Hsa), chicken (Gga), Carolina anole lizard (Aca), Western clawed frog (Xtr), coelacanth (Lch), spotted gar (Loc), zebrafish (Dre), medaka (Ola), three-spined stickleback (Gac), green spotted pufferfish (Tni), whale shark (Rty), little skate (Lca), elephant shark (Cmi), sea lamprey (Pma), Arctic lamprey (Lca). The tree is displayed as a midpoint-rooted phylogram.



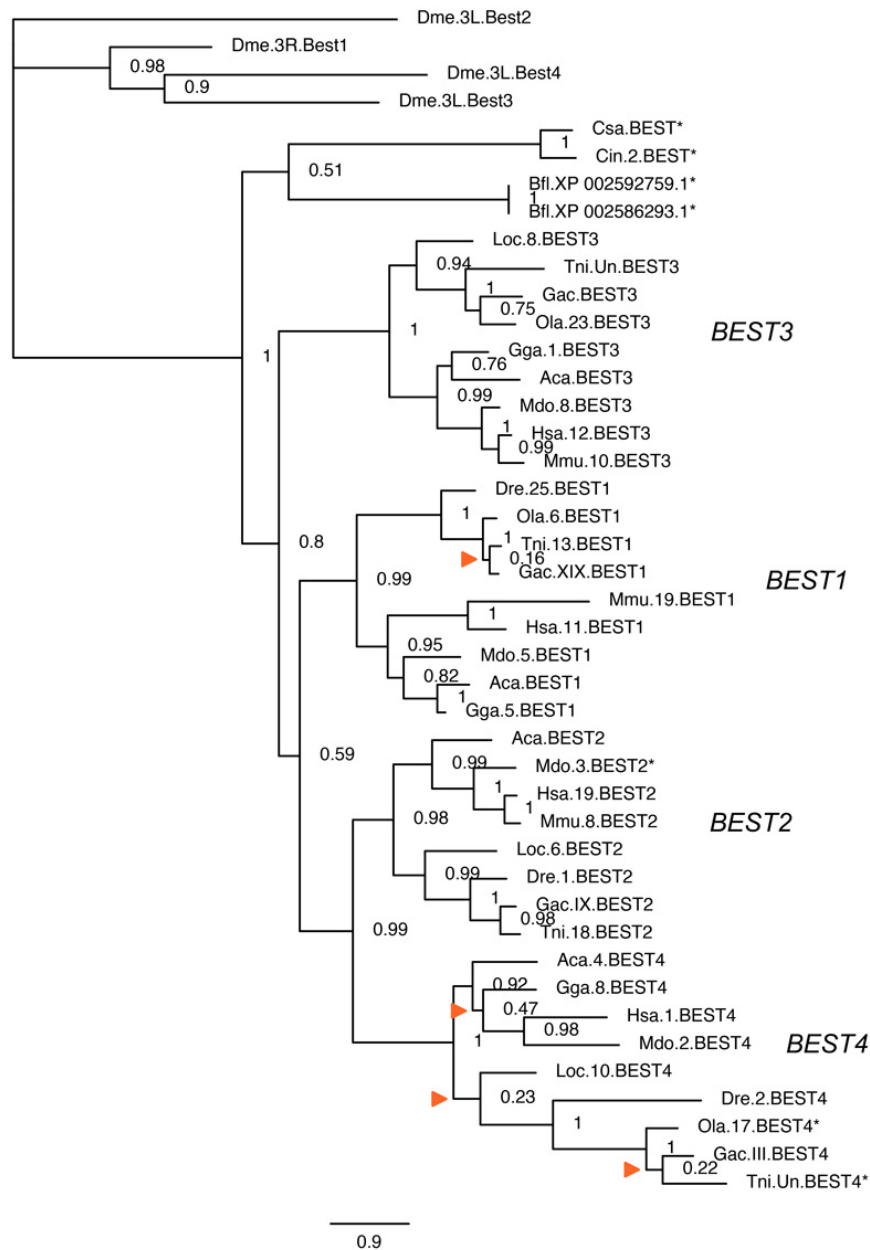
Supplementary Figure S3. Phylogenetic analysis and exon organization of the Florida lancelet GH/PRL1BP sequence. Uninformative nodes (support $\leq 50\%$) are marked with arrowheads. For clarity, shallow node support values are not shown. The Florida lancelet putative GH/PRL1BP sequence is marked by a red diamond. Species abbreviations as in Supplementary figure S2, except for Florida lancelet (Bfl). Coding exon lengths for the predicted GH/PRL1BP gene are given in base pairs. Abbreviations: LBD, ligand binding domain; SP, signal peptide. Red boxes indicate divergent exon organization compared with vertebrate GHR, PRLR, CRFA4, EPOR and TPOR genes. Upper case Cs indicate the positions of conserved cysteine residue pairs. The conserved WSXWS motif is indicated by a magenta arrow.



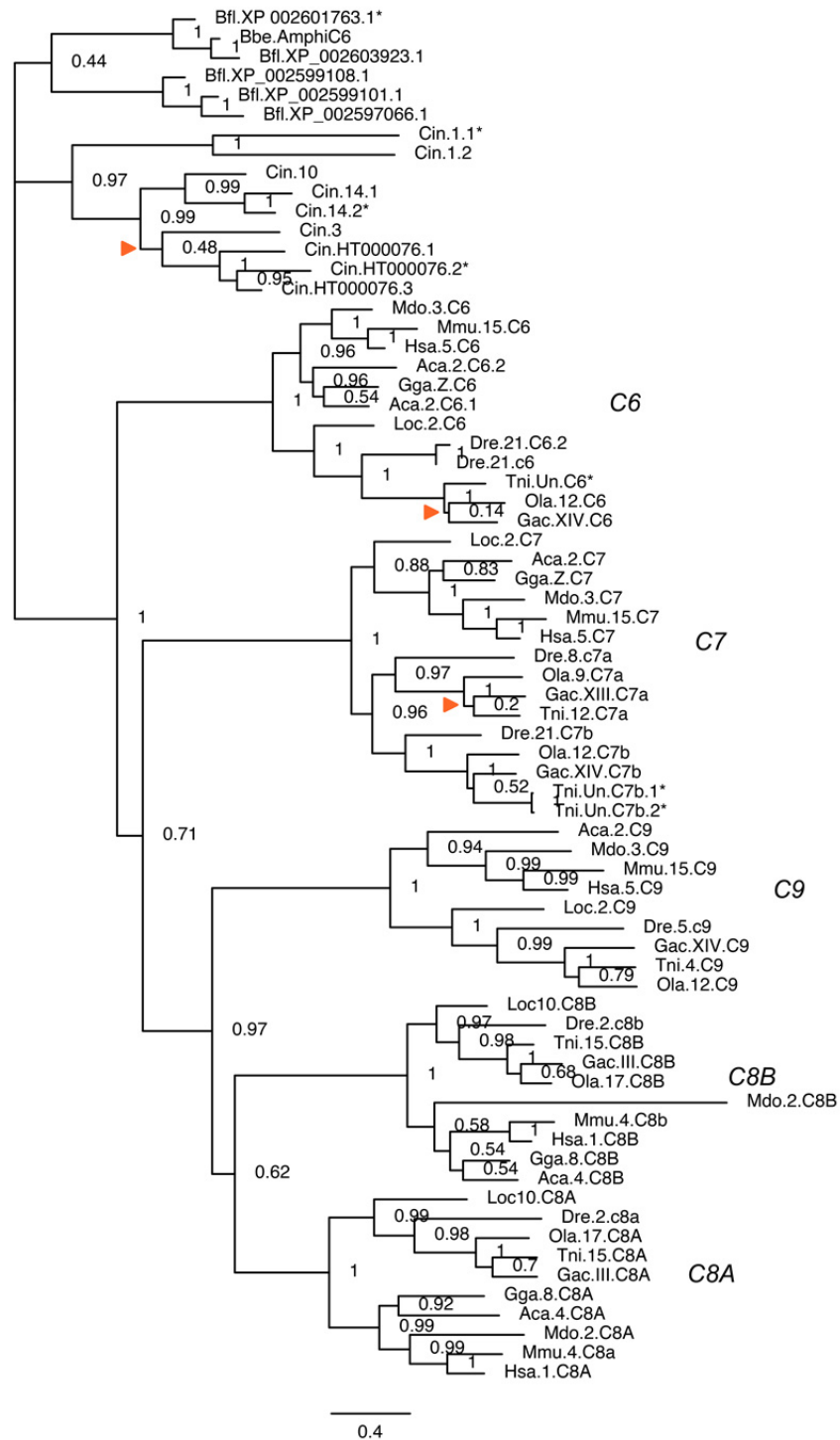
Supplementary Figure S4. Conserved synteny blocks in the zebrafish genome. Accepted zebrafish gene nomenclature is used (lower case), except in the case of novel suggested gene symbols, *NIM1K2* and *PTGER5*, which are written in upper case.



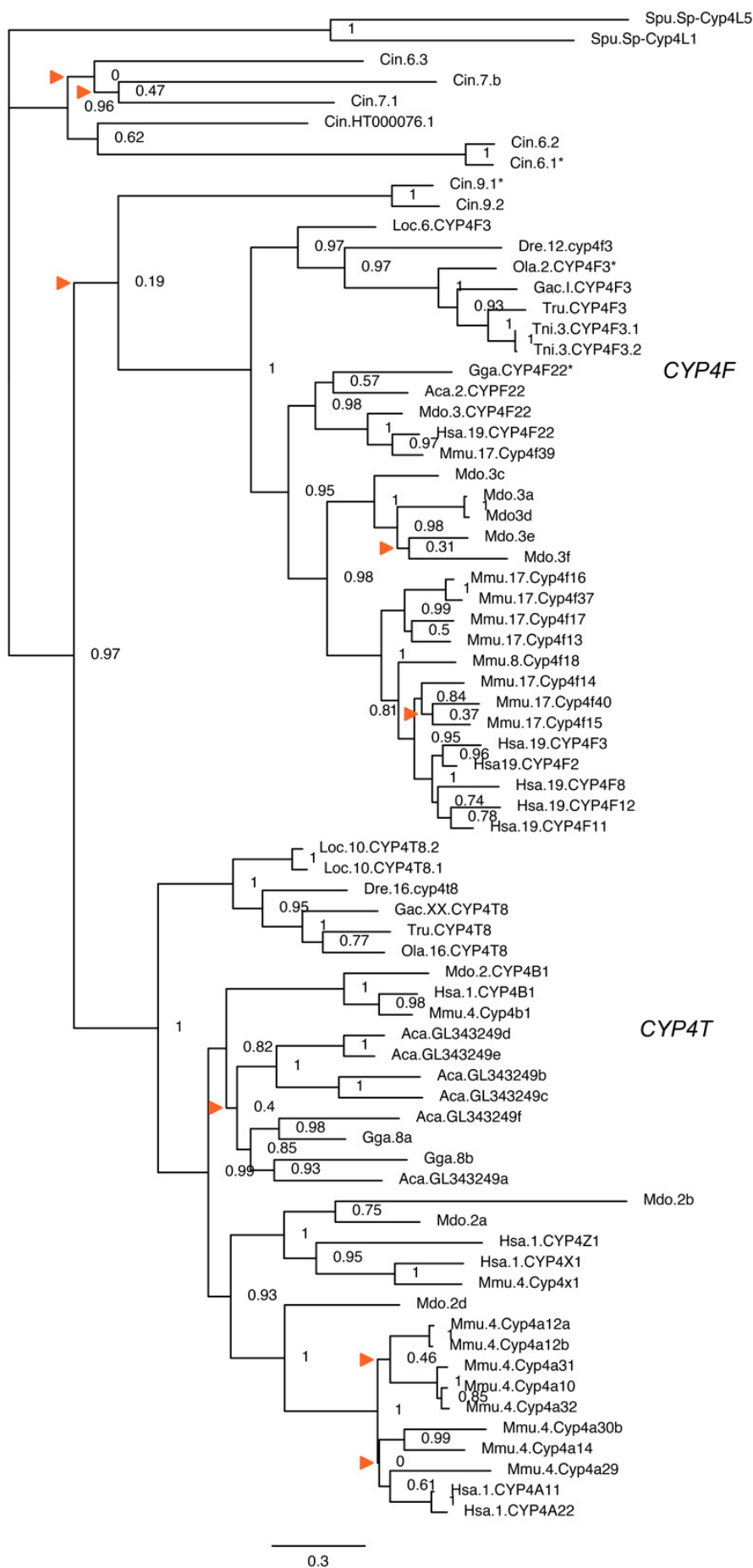
Supplementary Figure S5. Phylogeny of ADAMTS6/10. For species abbreviations and sequence naming details, see Supplementary data 2 caption. Uninformative nodes (support ≤50%) are marked with arrowheads. The phylogeny of *ADAMTS6* and *ADAMTS10* supports the divergence of the two clades in the time window of the 1R and 2R tetraploidizations.



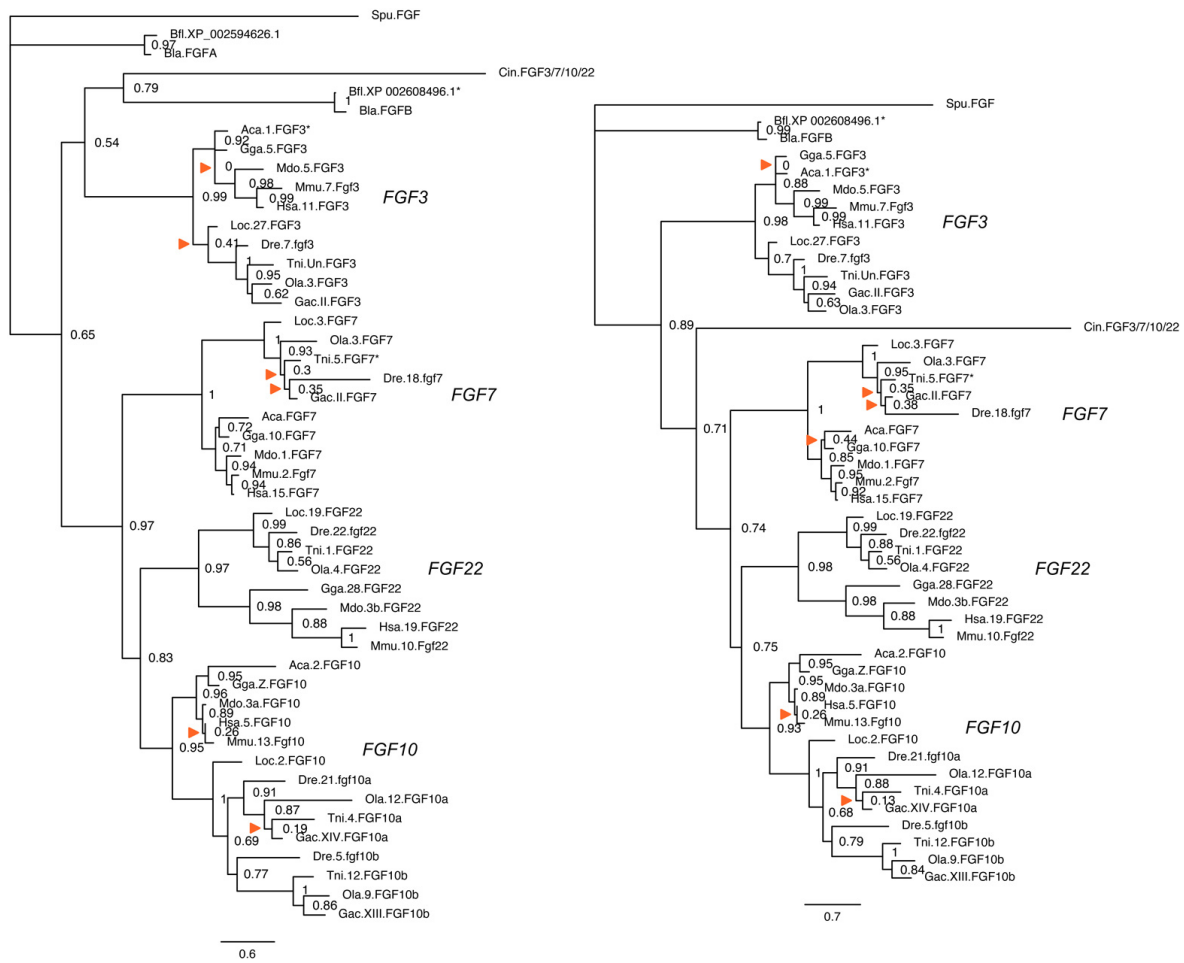
Supplementary Figure S6. Phylogeny of BEST. For species abbreviations and sequence naming details, see Supplementary data 2 caption. Uninformative nodes (support ≤ 50%) are marked with arrowheads. The phylogeny supports the divergence of BEST1, -2, -3 and -4 in the time window of the 1R and 2R tetraploidizations. However, the chromosomal locations of the BEST1 and BEST3 genes suggests that at least the duplication that gave rise to the BEST1/BEST3 ancestral gene and the BEST2/BEST4 ancestral gene occurred before 1R.



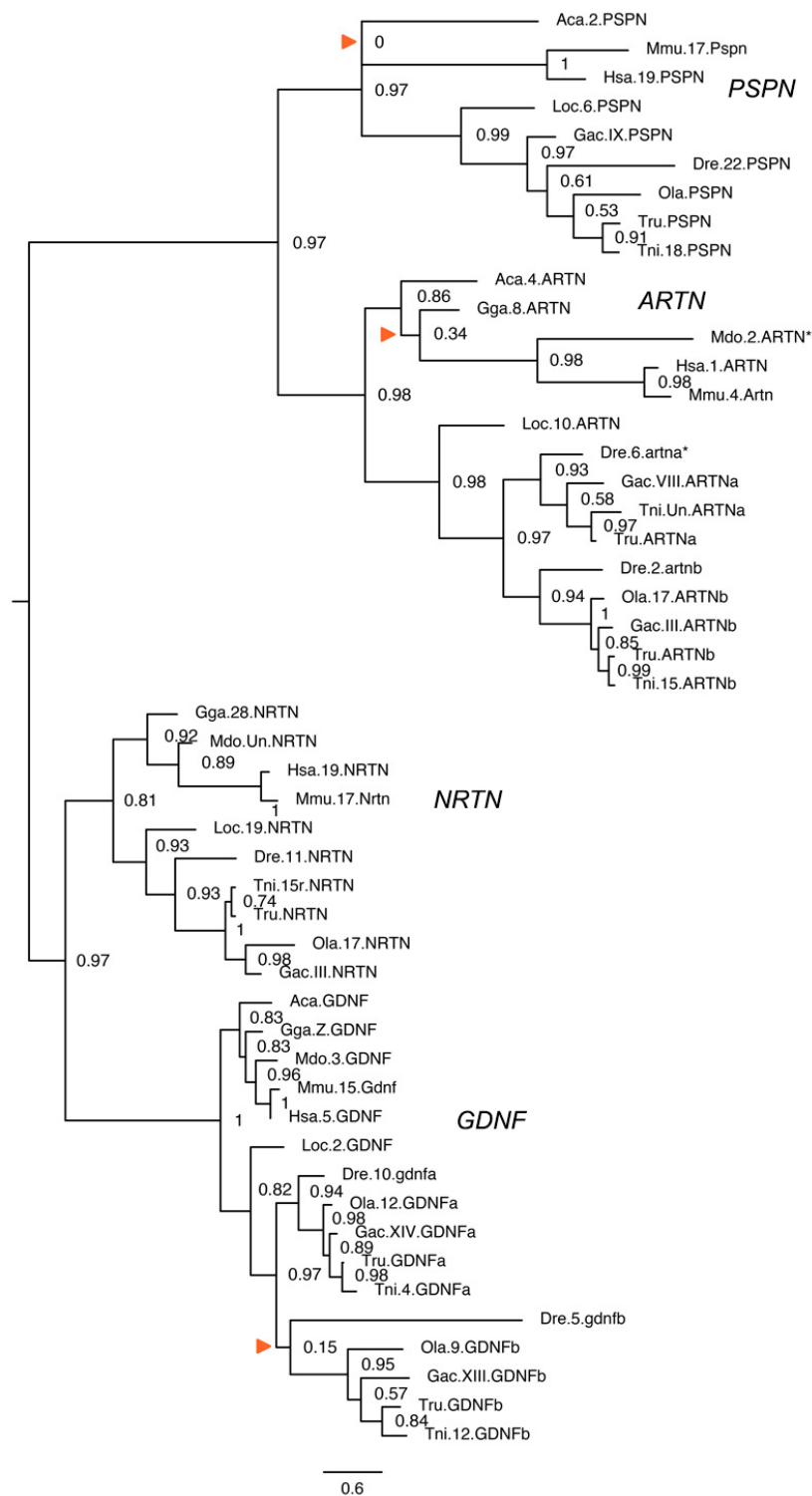
Supplementary Figure S7. Phylogeny of C6-9. For species abbreviations and sequence naming details, see Supplementary data 2 caption. Uninformative nodes (support ≤50%) are marked with arrowheads. See Section 3.9 in the results for a description of the C6-9 phylogeny.



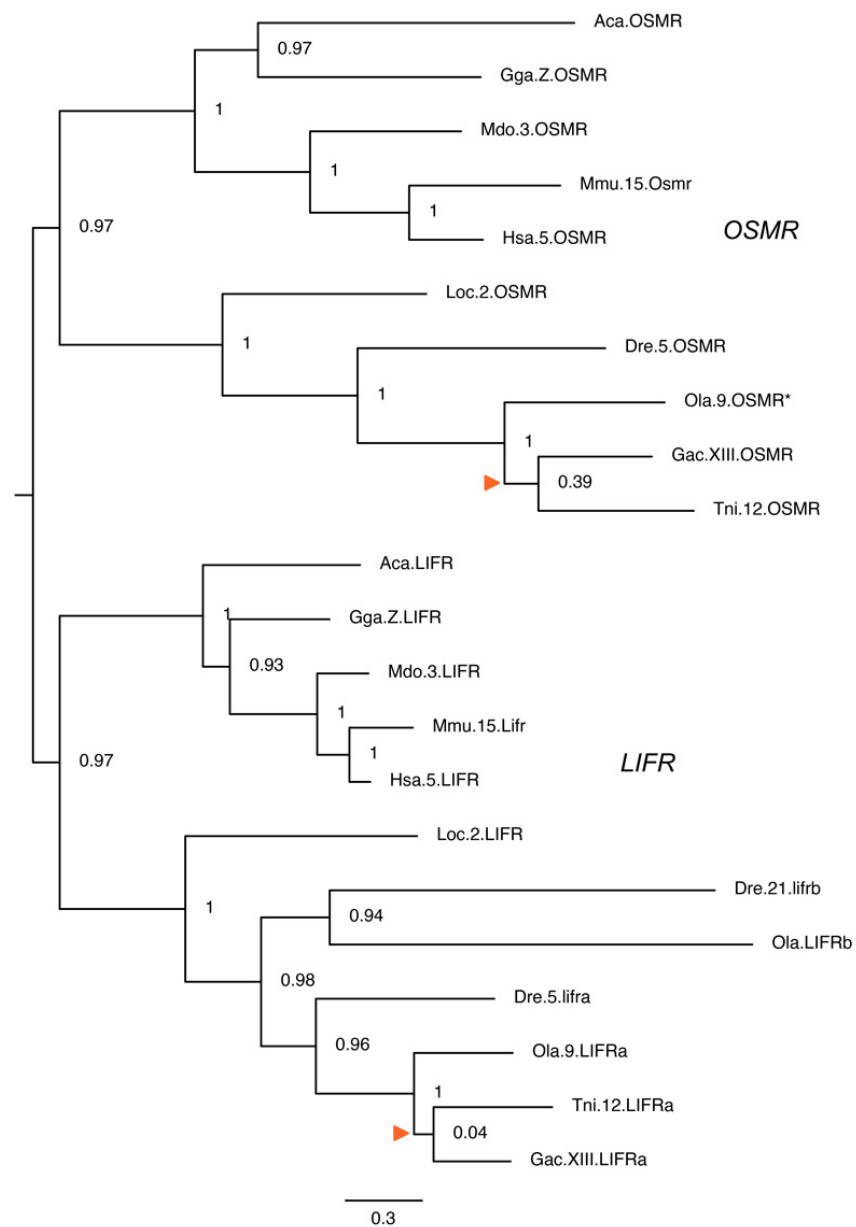
Supplementary Figure S8. Phylogeny of CYP4. For species abbreviations and sequence naming details, see Supplementary data 2 caption. Uninformative nodes (support $\leq 50\%$) are marked with arrowheads. See Section 3.10 in the results for a description of the CYP4 phylogeny.



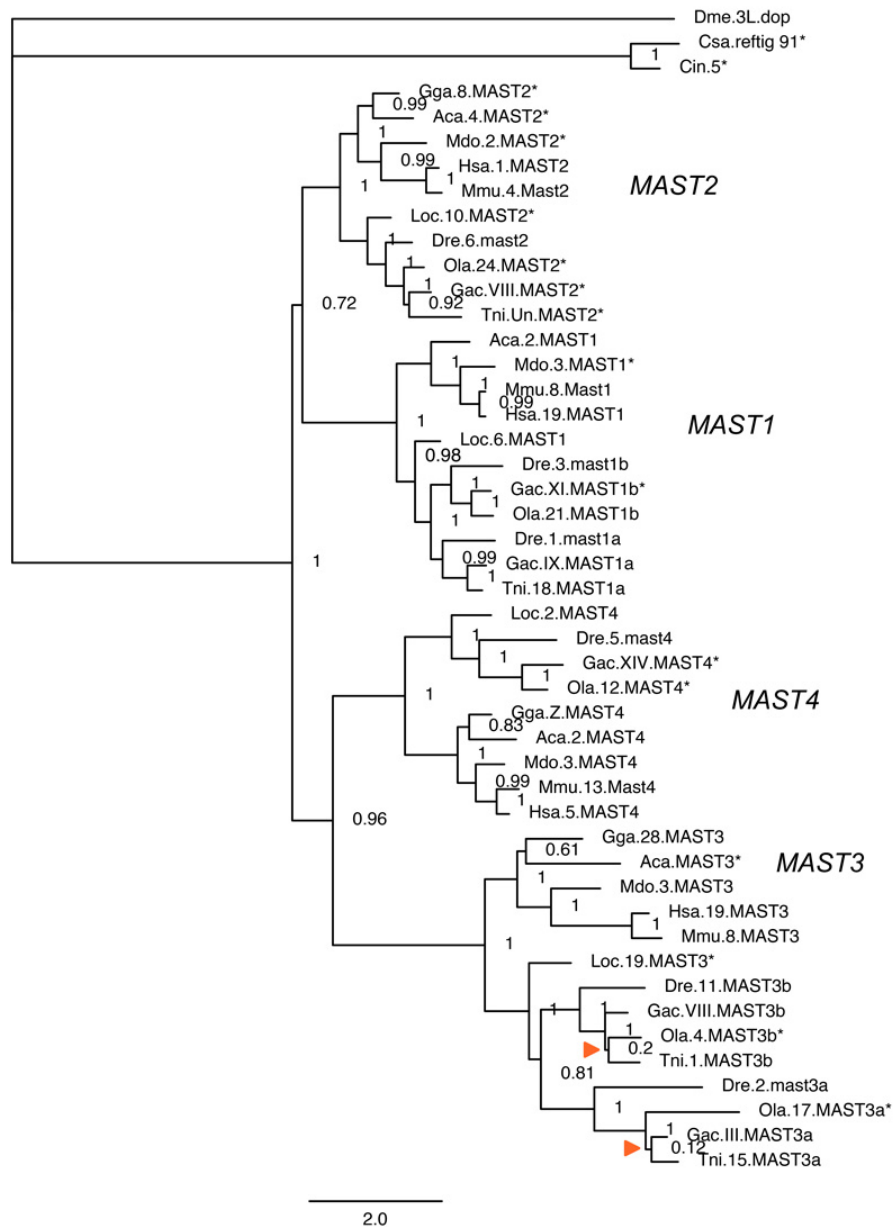
Supplementary Figure S9. Phylogeny of FGF3/7/10/22. For species abbreviations and sequence naming details, see Supplementary data 2 caption. Uninformative nodes (support ≤50%) are marked with arrowheads. See Section 3.11 in the results for a description of the two FGF3/7/10/22 phylogenies. Both phylogenies support the duplication of FGF10 in the basal teleost tetraploidization (3R), giving rise to FGF10a and FGF10b genes on the same chromosome blocks as the GHRb-PRLRa and GHRA-PRLRb gene pairs, respectively.



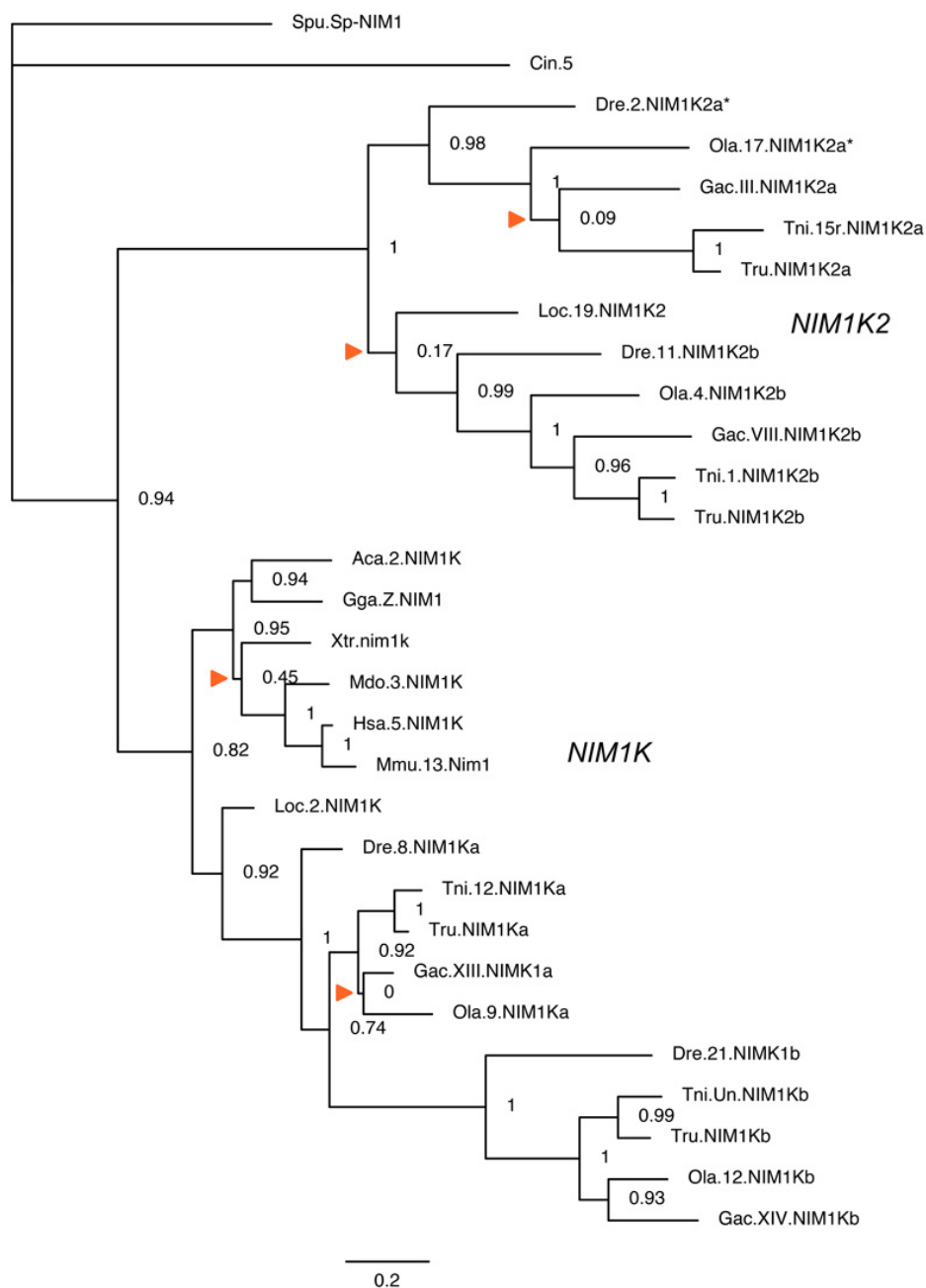
Supplementary Figure S10. Phylogeny of GFL. For species abbreviations and sequence naming details, see Supplementary data 2 caption. Uninformative nodes (support $\leq 50\%$) are marked with arrowheads. The protein family prediction used in this study included only the glial cell-derived neurotrophic factor (GDNF) genes. Artemin (ARTN), neurturin (NRTN) and persephin (PSPN) sequences were added to the dataset based on a previous study grouping these four neurotrophic factors together (Hätinen, Holm, & Airaksinen, 2007). The phylogeny is consistent with the emergence of an ARTN/PSPN ancestral gene and a GDNF/NRTN ancestral gene in the 1R tetraploidization, whereupon 2R generated the four subtype genes. However, we could not identify invertebrate family members to provide relative dating points for these divergences. The phylogeny supports the divergence of GDNFa and GDNFb in the time window of the basal teleost tetraploidization (3R), however the zebrafish gdnfb branch is not well-supported.



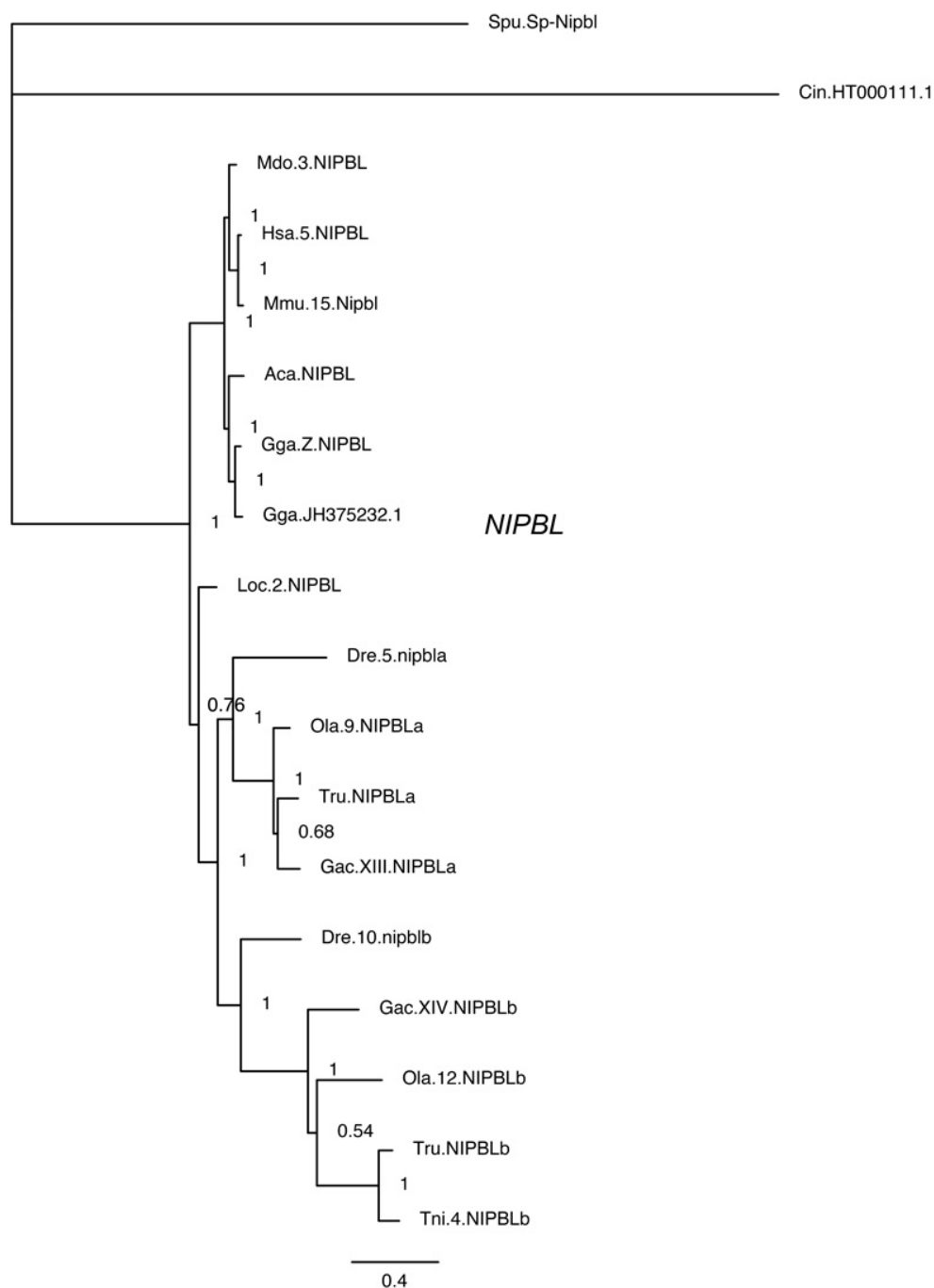
Supplementary Figure S11. Phylogeny of LIFR/OSMR. For species abbreviations and sequence naming details, see Supplementary data 2 caption. Uninformative nodes (support $\leq 50\%$) are marked with arrowheads. The phylogeny of the LIFR clade is consistent with the emergence of LIFRa and LIFRb in the basal teleost tetraploidization (3R). The chromosomal locations of LIFR and OSMR genes suggest they arose through an ancient local duplication.



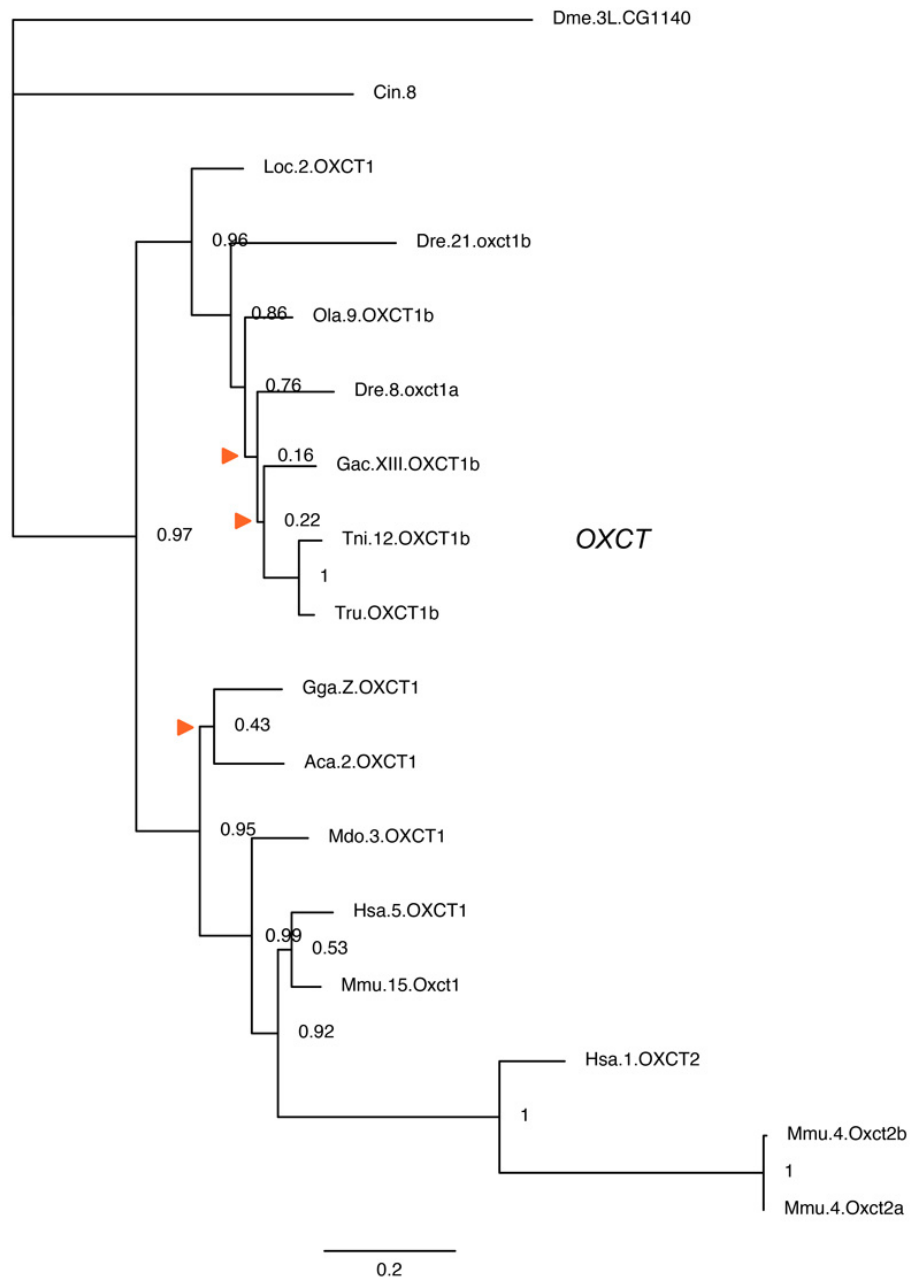
Supplementary Figure S12. Phylogeny of MAST. For species abbreviations and sequence naming details, see Supplementary data 2 caption. Uninformative nodes (support $\leq 50\%$) are marked with arrowheads. The phylogeny supports the divergence of MAST1, -2, -3 and -4 in the time window of the 1R and 2R tetraploidizations, and of MAST1a and -1b, as well as MAST3a and -3b, in the time window of the basal teleost tetraploidization (3R).



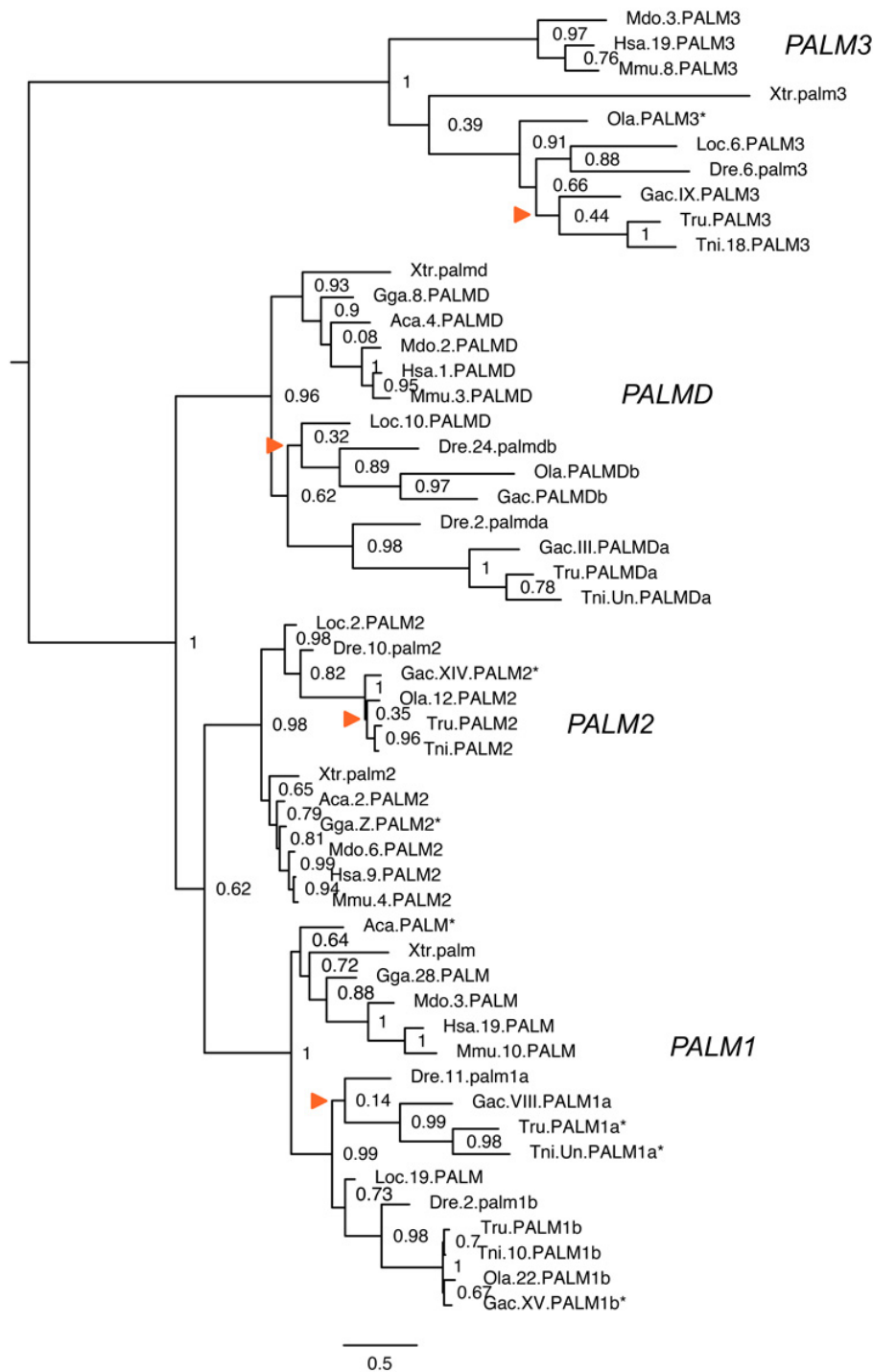
Supplementary Figure S13. Phylogeny of NIM1K. For species abbreviations and sequence naming details, see Supplementary data 2 caption. Uninformative nodes (support ≤50%) are marked with arrowheads. See Section 3.12 in the results for a description of the *NIM1K* and *NIM1K2* phylogeny.



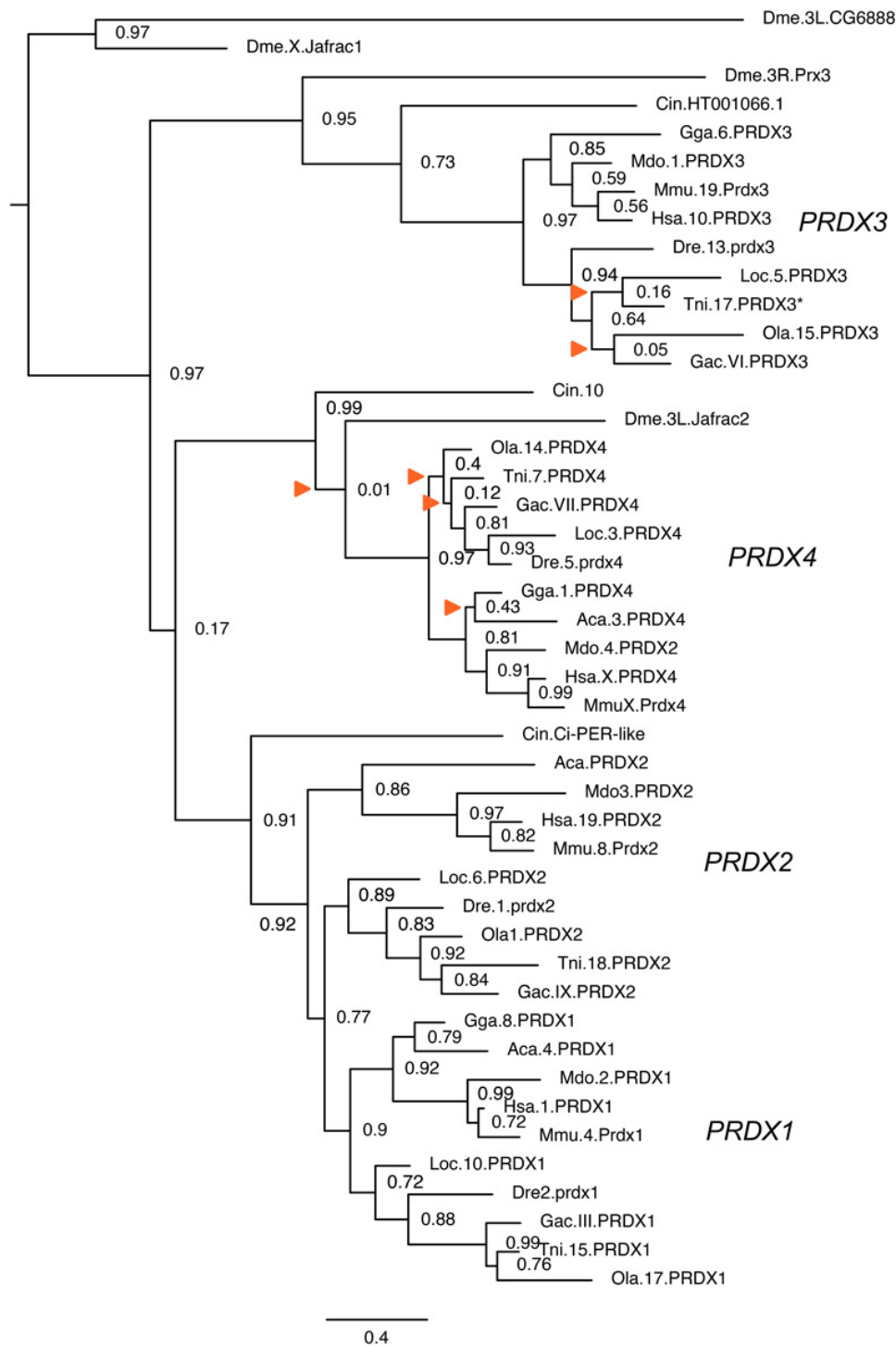
Supplementary Figure S14. Phylogeny of NIPBL. For species abbreviations and sequence naming details, see Supplementary data 2 caption. Uninformative nodes (support $\leq 50\%$) are marked with arrowheads. The phylogeny supports the divergence of the *NIPBLa* and *-b* clades in the time window of the basal teleost tetraploidization (3R).



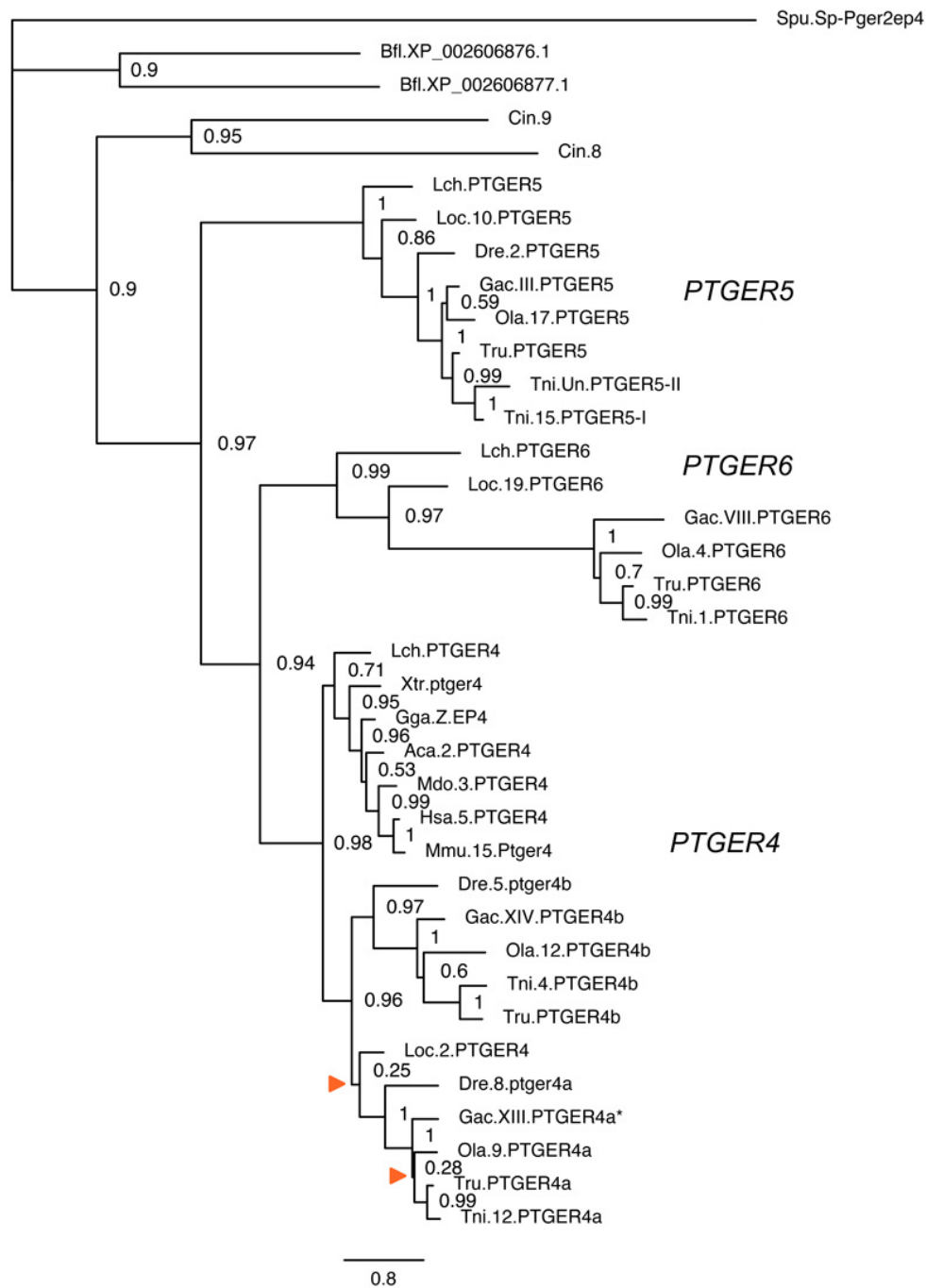
Supplementary Figure S15. Phylogeny of OXCT. For species abbreviations and sequence naming details, see Supplementary data 2 caption. Uninformative nodes (support $\leq 50\%$) are marked with arrowheads. The phylogeny is consistent with the emergence of *OXCT1a* and *1b* genes in the time window of the basal teleost tetraploidization (3R), however an *OXCT1b* sequence could only be found in the zebrafish. The *OXCT2* gene is a mammalian retrogene, based on its lack of introns and species representation.



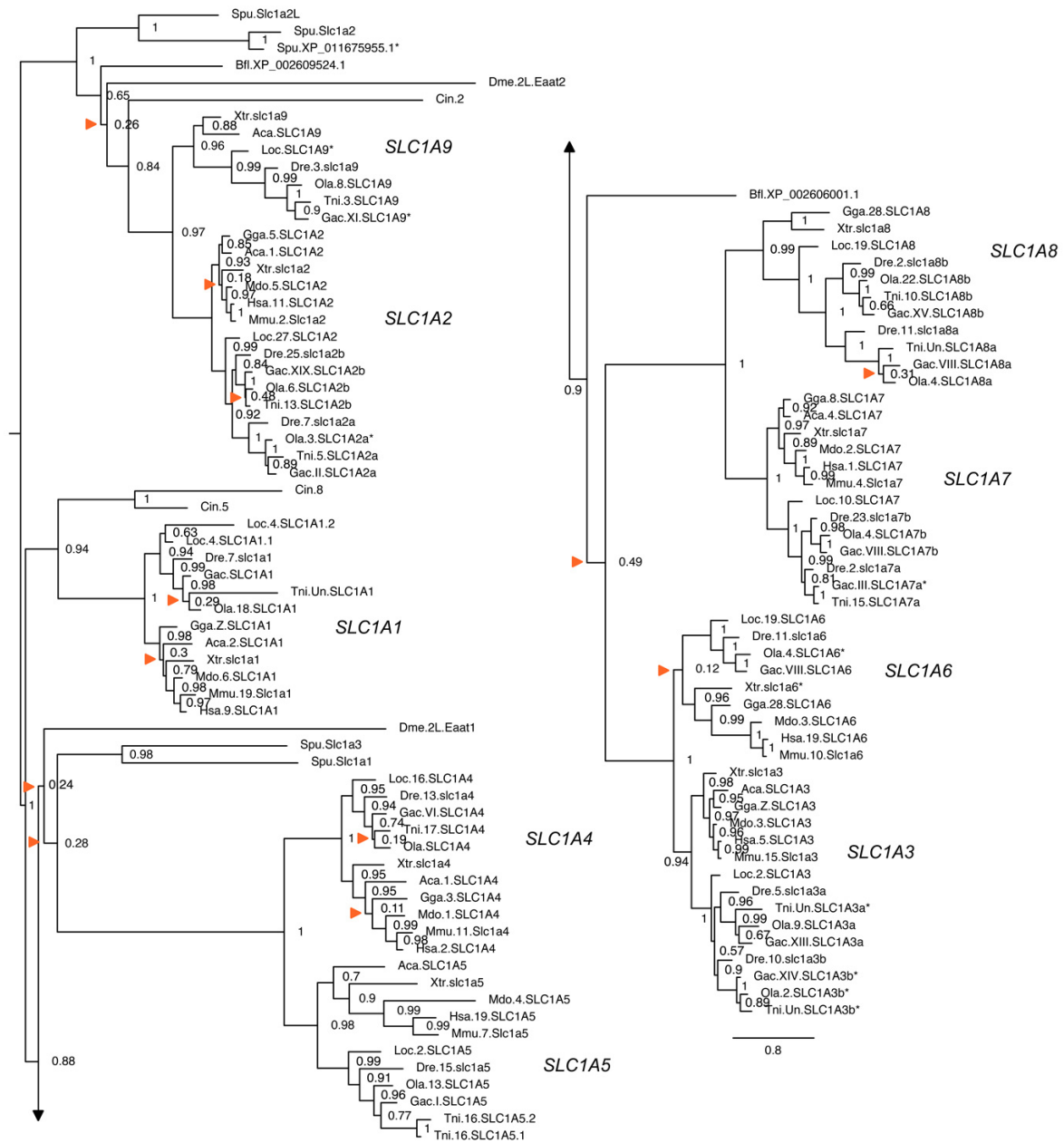
Supplementary Figure S16. Phylogeny of PALM. For species abbreviations and sequence naming details, see Supplementary data 2 caption. Uninformative nodes (support ≤50%) are marked with arrowheads. The phylogeny is consistent with duplications in the 1R and 2R tetraploidizations giving rise to *PALM*, *PALM2*, *PALM3* and *PALMD*. However, we could not identify invertebrate family members to provide relative dating points for these duplications. The phylogeny is somewhat unclear with regard to the emergence of *PALM1a* and *-1b* as well as *PALMDa* and *-Db* in the time window of the basal teleost tetraploidization (3R), however these inconsistencies are likely the result of the uneven evolutionary rates (represented as branch lengths) and low sequence similarity within this family.



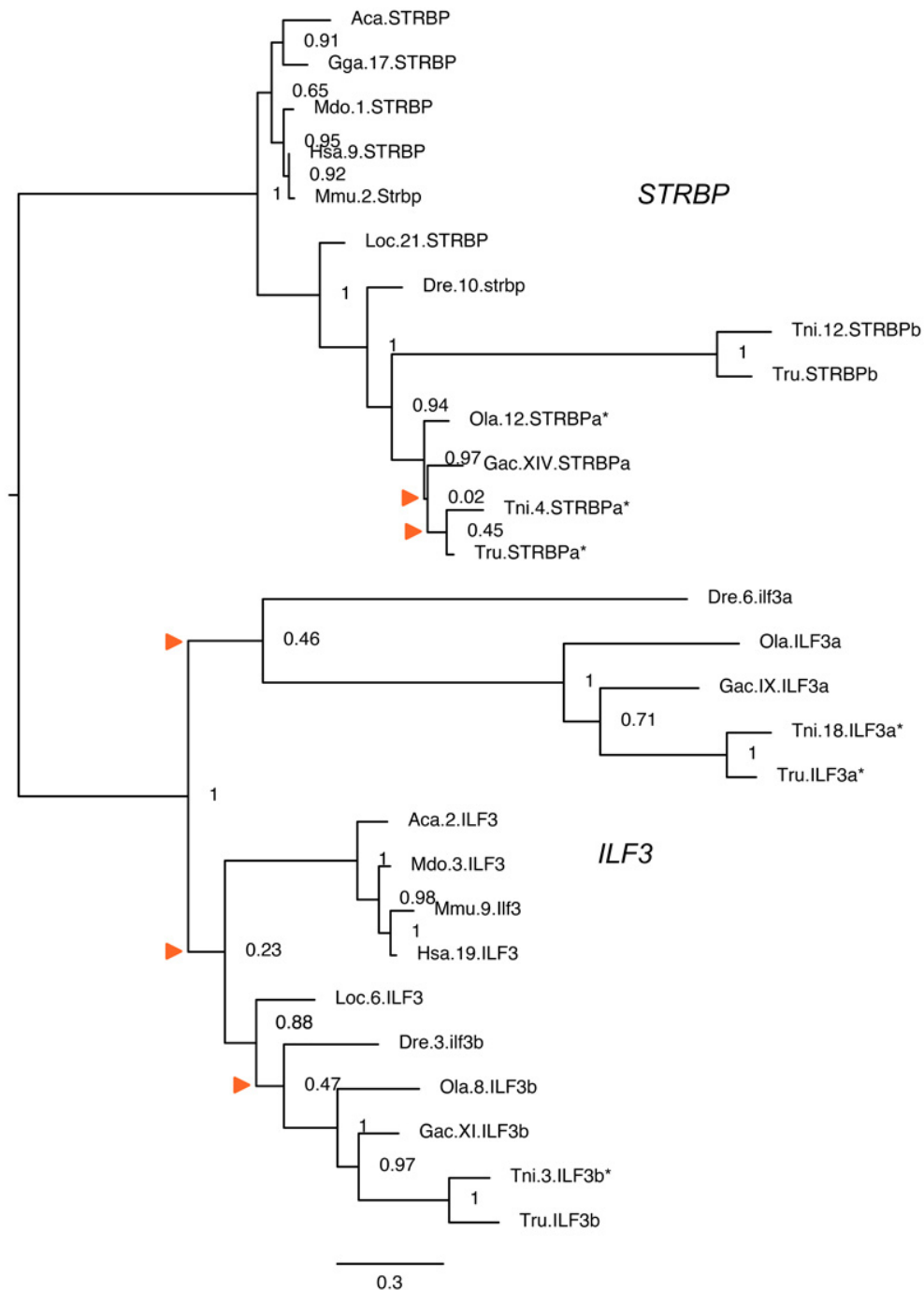
Supplementary Figure S17. Phylogeny of PRDX. For species abbreviations and sequence naming details, see Supplementary data 2 caption. Uninformative nodes (support $\leq 50\%$) are marked with arrowheads. The phylogeny of the PRDX1 and PRDX2 sequences is consistent with a divergence in the time window of the 1R and 2R tetraploidizations. However, the PRDX2 clade is not resolved. The chromosomal locations of the PRDX2 genes nonetheless make an emergence of PRDX1 and PRDX2 in 2R the likely scenario. The PRDX3 and PRDX4 genes likely emerged much earlier, and seemingly no 1R or 2R-generated duplicates were preserved.



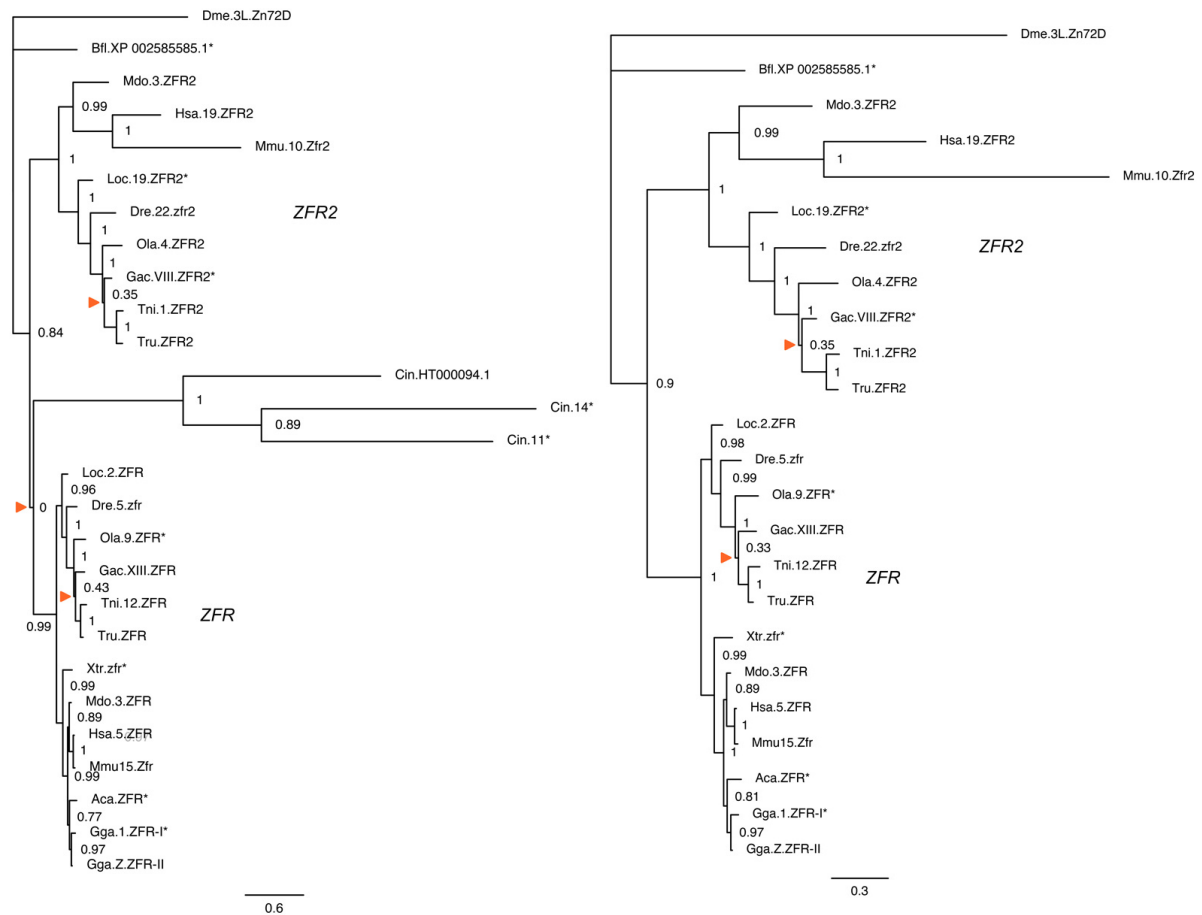
Supplementary Figure 18. Phylogeny of PTGER4. For species abbreviations and sequence naming details, see Supplementary data 2 caption. Uninformative nodes (support $\leq 50\%$) are marked with arrowheads. The phylogeny supports the divergence of PTGER4 as well as two previously undescribed clades we have called PTGER5 and PTGER6 in the time window of the 1R and 2R tetraploidizations. The phylogeny is also consistent with the divergence of PTGER4a and -4b clades in the time window of the basal teleost tetraploidization (3R). See Section 3.13 in the results for a more detailed description of the phylogeny.



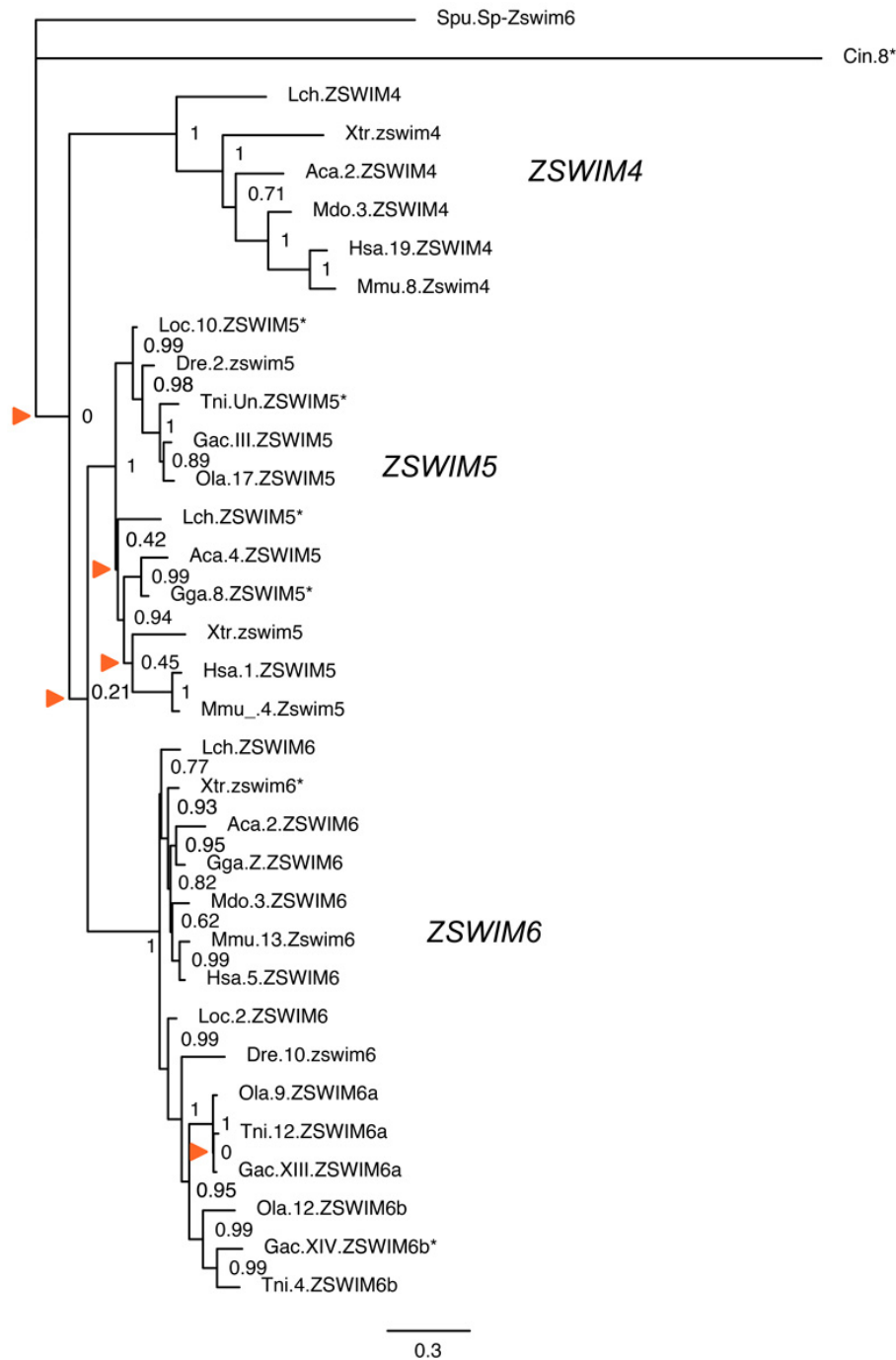
Supplementary Figure 19. Phylogeny of SLC1A. For species abbreviations and sequence naming details, see Supplementary data 2 caption. Uninformative nodes (support ≤50%) are marked with arrowheads. The phylogeny supports the divergence of the SLC1A3, SLC1A6, SLC1A7 and SLC1A8 clades in the time window of the 1R and 2R tetraploidizations, and of SLC1A3a and -3b, SLC1A7a and -7b, as well as SLC1A8a and -8b, in the basal teleost tetraploidization (3R). Furthermore, the phylogeny also supports the divergence of the SLC1A2 and SLC1A9 clades in the time window of 1R and 2R, and of SLC1A2a and -2b in the time window of 3R. However, these genes are not located in the paralogous chromosome regions that carry the GHR, PRLR, EPOR and TPOR genes. It is also likely that SLC1A4 and SLC1A5 clades emerged in 1R and 2R, however the lack of vase tunicate or lancelet sequences for this branch makes the relative dating of this duplication unclear.



Supplementary Figure 20. Phylogeny of STRBP/ILF3. For species abbreviations and sequence naming details, see Supplementary data 2 caption. Uninformative nodes (support $\leq 50\%$) are marked with arrowheads. The phylogeny and chromosomal locations of STRBP and ILF3 are consistent with an emergence in the 1R tetraploidization. However, we could not identify invertebrate family members to provide a relative dating point for this divergence. The phylogeny does not support the duplication of ILF3 in the basal teleost tetraploidization (3R) giving rise to ILF3a and -b. The chromosomal locations of the green spotted pufferfish STRBP_a and STRBP_b genes suggest an origin in 3R. The phylogeny is unclear in this regard; however, the inconsistency is likely brought about by the preservation of only one strbp gene in zebrafish.



Supplementary Figure 21. Phylogeny of ZFR. For species abbreviations and sequence naming details, see Supplementary data 2 caption. Uninformative nodes (support $\leq 50\%$) are marked with arrowheads. The ZFR and ZFR2 sequences were grouped with the STRBP and ILF3 sequences in the protein family prediction that was used in this study (identified as “BINDING” in Supplementary data 3). However, our subsequent phylogenetic analyses, as well as the domain compositions of the ZFR, ZFR2, STRBP and ILF3 sequences (Wolkowicz & Cook, 2012), do not support this. The vase tunicate sequences in this protein family prediction are highly divergent, and consequently have an unsupported branching point in our first phylogeny. For this reason, we also included a putative lancelet family member and constructed an additional phylogeny without the vase tunicate sequences. Both phylogenies of ZFR and ZFR2 are consistent with the divergence of the two clades in the time window of the 1R and 2R tetraploidizations.



Supplementary Figure 22. Phylogeny of ZSWIM. For species abbreviations and sequence naming details, see Supplementary data 2 caption. Uninformative nodes (support $\leq 50\%$) are marked with arrowheads. The phylogeny is consistent with the divergence of ZSWIM4, ZSWIM5 and ZSWIM6 clades in the time window of the 1R and 2R tetraploidizations. The low support for the basal nodes in the tree is likely caused by the high degree of sequence conservation within the family, which reduces the overall phylogenetic signal. A putative ZSWIM4 partial sequence from spotted gar was not used in this phylogeny due to its shortness and high degree of sequence divergence. The phylogeny is unclear with regard to the duplication of ZSWIM6 in the time window of the basal teleost tetraploidization (3R), generating ZSWIM6a and -6b. This inconsistency is likely brought about by the preservation of only one zswim6 gene in zebrafish.