Evolution of the growth hormone, prolactin, prolactin 2 and somatolactin family

Daniel **Ocampo Daza*** and **Dan Larhammar**

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

* Corresponding author at Subdepartment of Evolution and Development, Department of Organismal Biology, Uppsala University, Norbyvägen 18A, SE-752 36 Uppsala, Sweden.

E-mail address: daniel.ocampo.daza@ebc.uu.se

General and Comparative Endocrinology 264 (2018) 94-112

https://doi.org/10.1016/j.ygcen.2018.01.007

Received 8 August 2017

Revised 29 December 2017

Accepted 11 January 2018

Available online 12 January 2018

Figshare

https://doi.org/10.6084/m9.figshare.6494879

Highlights

- *PRL* and *PRL2* genes arose in the basal vertebrate tetraploidizations (1R and 2R).
- *SL* duplicated in teleost tetraploidization (3R) producing *SLa* and *SLb*.
- *GH*, *PRL/PRL2* and *SL* genes were present already before the vertebrate radiation.
- Genomic evidence suggests origin of family through local duplications before 1R/2R.
- *GH*, *PRL*, *PRL2*, *SL* emergence much earlier than diversification of *GHR*/*PRLR* receptors.

Abstract

Growth hormone (GH), prolactin (PRL), prolactin 2 (PRL2) and somatolactin (SL) belong to the same hormone family and have a wide repertoire of effects including development, osmoregulation, metabolism and stimulation of growth. Both the hormone and the receptor family have been proposed to have expanded by gene duplications in early vertebrate evolution. A key question is how hormone-receptor preferences have arisen among the duplicates. The first step to address this is to determine the time window for these duplications. Specifically, we aimed to see if duplications resulted from the two basal vertebrate tetraploidizations (1R and 2R). *GH* family genes from a broad range of vertebrate genomes were investigated using a combination of sequence-based phylogenetic analyses and comparisons of synteny. We conclude that the *PRL* and *PRL2* genes arose from a common ancestor in 1R/2R, as shown by neighboring gene families. No other gene duplicates were preserved from these tetraploidization events. The ancestral genes that would give rise to *GH* and *PRL/PRL2* arose from an earlier duplication; most likely a local gene duplication as they are syntenic in several species. Likewise, some evidence suggests that *SL* arose from a local duplication of an ancestral *GH/SL* gene in the same time window, explaining the lack of similarity in chromosomal neighbors to *GH*, *PRL* or *PRL2*. Thus, the basic triplet of ancestral *GH*, *PRL/PRL2* and *SL* genes appear to be unexpectedly ancient. Following 1R/2R, only *SL* was duplicated in the teleost-specific tetraploidization 3R, resulting in *SLa* and *SLb*. These time windows contrast with our recent report that the corresponding receptor genes GHR and PRLR arose through a local duplication in jawed vertebrates and that both receptor genes duplicated further in 3R, which reveals a surprising asynchrony in hormone and receptor gene duplications.

Keywords

Growth hormone; Somatolactin; Prolactin; Prolactin 2; Gene duplication; Molecular evolution; Phylogeny

1. Introduction

The growth hormone (*GH*), prolactin (*PRL*), prolactin 2 (*PRL2)* and somatolactin (*SL*) genes in vertebrates constitute a gene family within the large superfamily of class-1 helical cytokines (Huising et al., 2006). The four members share a common exon organization and the proteins display a characteristic conserved amino-acid framework with four cysteine residues forming two disulfide bridges that stabilize their tertiary four-helical structure (Pérez-Sánchez et al., 2002, Power, 2005, Whittington and Wilson, 2013). The four proteins share only about 25–30% amino acid sequence identity, and each one can display considerable variation in evolutionary rate across taxa, both among mammals (Wallis, 2014, Wallis, 2001, Wallis, 2000, Wallis et al., 2005) and teleost fishes (Ryynänen and Primmer, 2006). Nevertheless, the conserved cysteine framework, the exon organization of the genes and the binding properties define them as a monophyletic family within the cytokinesuperfamily (Huising et al., 2006). Despite sequence information from a vast number of species, it has remained unclear how the gene family was formed by gene duplications, and different proposals have been discussed extensively over the years (Bole-Feysot et al., 1998, Forsyth and Wallis, 2002, Kawauchi and Sower, 2006). The identification of only one family member, a GH-like sequence, in the jawless sea lamprey (*Petromyzon marinus*) (Kawauchi et al., 2002) was seen as evidence that the gene duplications that gave rise to *GH, PRL* and *SL* took place in an ancestor of jawed vertebrates (*Gnathostomata*), before their radiation (Kawauchi et al., 2002, Kawauchi and Sower, 2006). Subsequently lineage-specific gains and losses of genes occurred in both the ray-finned fish *(Actinopterygii*) and lobe-finned fish (*Sarcopterygii*) lineages. This includes the duplications and diversification of *GH* in primates (Papper et al., 2009, Wallis and Wallis, 2006), and *PRL* in rodents (Alam et al., 2006, Wiemers et al., 2003) and ruminants (Larson et al., 2006, Ushizawa et al., 2007b), producing the placental lactogens (PL) and prolactin-related proteins (Forsyth and Wallis, 2002). After the first identification of somatolactin in the olive flounder (*Paralichthys olivaceus*) (Ono et al., 1990) and Atlantic cod (*Gadus morhua*) (Rand-Weaver et al., 1991), this member of the hormone family has been identified in a multitude of teleost species (Benedet et al., 2008) as well as in the white sturgeon (*Acipenser transmontanus*) and the African lungfish (*Protopterus annectens*) (Amemiya et al., 1999, May et al., 1999). The latter findings demonstrated that all three hormones, including somatolactin, were present before the divergence of lobe-finned fishes and ray-finned fishes.

More recent discoveries complicate this view of the family's evolution. We have previously reported the existence of *PRL2* in teleost genomes and we noted that it likely emerged through the duplication of a large chromosome block (Ocampo Daza et al., 2009). The *PRL2*gene had already been discovered in the green spotted pufferfish (*Tetraodon nigroviridis*) genome and named *PRLb* as part of the analysis of the draft genome sequence (Huising et al., 2006). Subsequently *PRL2* was identified in a variety of vertebrate classes and characterized biologically with regard to tissue expression and receptor activation in several independent studies: The *PRL2* gene was identified in elephant shark (*Callorhinchus milii*), Russian sturgeon (*Acipenser gueldenstaedti*) (Huang et al., 2009) and chicken (*Gallus gallus*) (Wang et al., 2010), showing that both *PRL* and *PRL2* must have been present in a jawed vertebrate ancestor. This implies that the divergences between *GH*, the ancestral *PRL/PRL2*, and *SL* took place even earlier.

These findings lead to several important questions regarding the origin and diversification of this family of hormones. The origin of the first member is of special interest. Recent studies identified a putative GH-like homolog in the lancelet *Branchiostoma japonicum* and claimed to demonstrate its growth-promoting and osmoregulatory abilities (Li et al., 2017, Li et al., 2014). This was the first identification of a putative GH family gene and protein hormone in an invertebrate species, and led the authors to suggest that *GH* is the ancestral member, with *PRL* and *SL* emerging in the vertebrate lineage. Additionally, it is not clear whether the *GH* gene identified in the sea lamprey also represents the ancestor of all four genes in jawed vertebrates, or if there have been losses in this lineage.

Furthermore, the two types of *SL*, called *SLa* and *SLb*, first identified in cyprinid and salmonid fishes, have been suggested to have resulted from the basal teleost tetraploidization (3R) (Jaillon et al., 2004, Meyer and van de Peer, 2005). However, published analyses of sequence-based phylogeny (Benedet et al., 2008, Zhu et al., 2004) and conserved synteny (Fukamachi and Meyer, 2007) have been inconclusive.

In order to resolve these key questions, it becomes essential to determine whether the gene duplications that gave rise to *GH*, *PRL*, *PRL2* and *SL* genes resulted from the two rounds of basal vertebrate tetraploidization (1R and 2R) (Dehal and Boore, 2005, Holland et al., 2008, Nakatani et al., 2007, Putnam et al., 2008). This may also shed light on the relationships of lamprey *GH* regardless of whether cyclostomes have undergone both 1R and 2R, or only 1R (Lagman et al., 2013, Mehta et al., 2013, Werner, 2013). The involvement of 1R and 2R in the emergence of the growth hormone family has not yet been thoroughly investigated, although it has been previously proposed (Forsyth and Wallis, 2002) and discussed in some detail (Huang et al., 2009, Ocampo Daza et al., 2009, Wang et al., 2010). The time points of the origin of the family and the gene duplications will be essential to understand the origin and evolution of many neuroendocrine functions mediated by the pituitary in vertebrates. These time points may subsequently be correlated with the time points for the corresponding receptor gene duplications to provide much needed context for comparative and evolutionary studies of hormone-receptor binding preferences. We recently reported that the growth hormone receptor and prolactin receptor genes *GHR* and *PRLR* arose as a result of a local gene duplication in a jawed vertebrate ancestor, and subsequently both genes were duplicated in the basal teleost tetraploidization (Ocampo Daza and Larhammar, 2017).

In the present study, we have combined comprehensive phylogenetic and chromosomal synteny analyses in order to shed light on the evolution of the growth hormone, prolactin, prolactin 2, and somatolactin family of hormones. This has posed a series of challenges, not only due to the low sequence identity and differential evolutionary rates between the family members as well as between species or lineages, but also because of differential losses in some lineages and, as we have come to realize during the course of this work, the dynamic evolutionary history of the chromosome regions harboring the *GH*, *PRL*, *PRL2* and *SL* genes.

2. Materials and methods

2.1. Sequence identification in vertebrate genome databases

GH, *PRL*, *PRL2* and *SL* gene predictions were identified in the Ensembl genome browser (http://www.ensembl.org) and the National Center for Biotechnology Information (NCBI) genome resource (https://www.ncbi.nlm.nih.gov/genome/) by using gene orthology prediction and protein family functions (Fernández-Suárez and Schuster, 2010) as well as extensive TBLASTN searches (Altschul et al., 1990). For all such searches, hits were considered informative if expect values (E-values) were lower than $1e^{-30}$. Amino acid sequences were collected for all identified *GH*-family gene predictions, and all genomic locations and database identifiers were noted. All database identifiers and location data, as well as detailed genome assembly information, is included in Supplementary Data 1. All sequence identifiers and locations from the Ensembl genome browser have been verified against Ensembl version 91 (December 2017), and all locations and sequence identifiers from NCBI are up to date as of July 2017.

The following species were investigated through Ensembl: *Homo sapiens* (human), *Mus musculus* (mouse), *Monodelphis domestica* (grey short-tailed opossum), *Gallus gallus*(chicken), *Anas platyrhynchos* (mallard duck), *Taeniopygia guttata* (zebra finch), *Anolis carolinensis* (Carolina anole lizard), *Pelodiscus sinensis* (Chinese softshell turtle), *Latimeria chalumnae* (coelacanth), *Lepisosteus oculatus* (spotted gar), *Danio rerio* (zebrafish), *Oryzias latipes* (medaka), *Gasterosteus aculeatus* (three-spined stickleback), *Tetraodon nigroviridis* (spotted green pufferfish), *Ciona intestinalis* (vase tunicate) and *Drosophila* melanogaster (fruit fly). The following species were investigated through the NCBI genome resource: *Alligator mississipiensis* (American alligator), *Python* (*molurus*) *bivittatus* (Burmese python), *Xenopus (Silurana) tropicalis* (Western clawed frog), *Salmo salar* (Atlantic salmon), *Dicentrarchus labrax* (European sea bass), *Takifugu rubripes* (Japanese pufferfish, fugu), *Oreochromis niloticus* (Nile tilapia), *Rhincodon typus* (whale shark), *Callorhinchus milii* (elephant shark), *Petromyzon marinus* (sea lamprey), *Lethenteron camtschaticum* (Arctic lamprey), *Branchiostoma floridae* (Florida lancelet) and *Saccoglossus kowalevskii* (acorn worm). A smaller selection of species was used for the phylogenetic analyses of neighboring gene families.

In addition, *GH*-family gene predictions were identified in the genome of *Xenopus laevis* (African clawed frog), available through http://www.xenbase.org, through TBLASTN searches using *Xenopus tropicalis* sequences as queries. Similarly, *GH*-family cDNA sequences were sought in transcriptome assemblies from the salamander species *Ambystoma mexicanum* (axolotl), available from http://www.ambystoma.org, and *Notophthalmus viridescens* (red spotted newt), available from http://sandberg.cmb.ki.se/redspottednewt (Abdullayev et al., 2013).

For some neighboring families, additional invertebrate sequences were sought using profile-Hidden Markov Model searches (Finn et al., 2011) of the EnsemblMetazoa database (http://metazoa.ensembl.org).

2.2. Identification of additional somatolactin sequences

Additional *SL* sequences were identified in the NCBI nucleotide collection (nr/nt) using TBLASTN searches with the zebrafish *smtla* and *smtlb* sequences as queries. All hits annotated as somatolactin sequences were collected (translated amino acid sequences) and used to make an *SL*-specific sequence alignment and phylogenetic tree. Additional *SL* sequences were sought in the following genome assemblies through the NCBI genome resource: *Scleropages formosus* (Asian arowana), *Anguilla anguilla* (European eel), *Anguilla japonica* (Japanese eel), *Clupea harengus* (Atlantic herring), *Cyprinus carpio* (common carp), *Ictalurus punctatus* (channel catfish), *Esox lucius* (Northern pike), *Oncorhynchus kisutch* (coho salmon) and *Oncorhynchus mykiss* (rainbow trout). Complete information about the identified *SL* sequences, including database identifiers, location data and genome assembly versions, is provided in Supplementary Data 1.

2.3. Sequence analysis, alignment and phylogenetic tree construction

Any short, fragmented or otherwise faulty gene predictions, a common occurrence in genome databases (Prosdocimi et al., 2012), were curated by manual inspection of the genomic sequence, following consensus for gene initiation, splice donor and acceptor sites, as well as sequence homology to other family members. Amino acid sequences were aligned using the MUSCLE algorithm (Edgar, 2004) with default settings in AliView 1.18 (Larsson, 2014). The alignments were inspected and edited in AliView in order to identify and correct faulty predictions as described above, and to adjust poorly aligned sequence stretches with respect to exon boundaries and conserved sequence motifs.

Maximum likelihood phylogenies were constructed using IQ-TREE (Nguyen et al., 2015) with the following settings: the LG model of amino acid substitution (Le and Gascuel, 2008) was assumed; the proportion of invariable sites (+I), Gamma shape parameter (+G), and amino acid frequencies (+F) were estimated from the alignments; the number of substitution rate categories was increased from 4 to 8; the proportion of invariable sites p_invar was optimized; node supports were calculated using IQ-TREE's UltraFast Bootstrap (UFBoot) method (Minh et al., 2013) as well as the approximate Likelihood Ratio Test (aLRT) with SH-like supports (Anisimova and Gascuel, 2006). Phylogenetic trees were rooted with invertebrate family members using SeaView 4.6.1 (Gouy et al., 2010). Where no invertebrate out-group could be identified, trees are displayed as midpoint-rooted trees (Hess and De Moraes Russo, 2007). Phylogenetic trees were inspected and tree midpoints were detected using FigTree v1.4.3 (available from http://tree.bio.ed.ac.uk/software/figtree/). For neighboring gene family phylogenies, a smaller selection of species was used (details in Supplementary Data File Supplementary Figures S2–S32). Vase tunicate or Florida lancelet sequences were used as relative dating points for the time window of the basal vertebrate tetraploidizations (1R and 2R), while spotted gar sequences were used as a relative dating point for the basal teleost tetraploidization (3R).

2.4. Identification of neighboring gene families and conserved synteny blocks

Gene lists corresponding to the genomic regions 10 MB in each direction of *GH*, *PRL*, *PRL2*and *SL* genes in the chicken (assembly WASHUC2), spotted gar (assembly LepOcu1) and zebrafish (assembly Zv9) genomes were mined using Ensembl's BioMart function. These

gene lists were combined and sorted according to the corresponding Ensembl protein family predictions (Fernández-Suárez and Schuster, 2010). By this sorting, families with members neighboring several of the *GH*-family genes across species could be identified. Gene families with a large multitude of members and/or high degree of sequence conservation, such as the *zinc finger* family of transcription factors, were excluded from the analyses. For the additional neighboring gene analysis of the zebrafish *gh1* and *prl* loci on chromosome 3, smaller regions were used in order to avoid overlap. Gene lists for 5 MB in each direction of the *gh1* and *prl* genes were compared to the spotted gar and chicken regions described above. All data has been verified against Ensembl version 91 (December 2017) and all neighboring gene family gene locations and database identifiers were updated to account for more recent genome assemblies: human genome assembly GRCh38, chicken genome assembly Gallus_gallus-5.0, zebrafish genome assembly Zv10 and Japanese pufferfish genome assembly FUGU5.

The spotted gar and chicken were chosen for the identification of conserved synteny blocks because they have undergone relatively few large-scale genome rearrangements (Amores et al., 2011, Nakatani et al., 2007, Voss et al., 2011), and because they preserve *GH*-family genes that have been lost in related lineages. Notably, the spotted gar preserves all four ancestral *GH*-family member genes and the chicken preserves *PRL2*. The zebrafish genome was added to the analysis because it preserves both *SLa* and *SLb* genes.

2. Results

3.1. Nomenclature

For the human, mouse, chicken, *Xenopus*, and zebrafish genes, the approved gene symbols have been used, following the appropriate gene nomenclature guidelines for each species. Thus, we have applied the approved stem symbols *GH* for growth hormone genes, *SL* for somatolactin genes, *PRL* for "classical" prolactin genes, and *PRL2* for the recently described *PRL* paralog genes. The numeral "1" was added to the *PRL* stem symbol only where duplicates of *PRL* could be confused for *PRL2* genes, e.g. *PRL1.2* and *PRL2*. Lineage-specific duplicates are indicated by the Roman numerals –I and –II, unless other nomenclature systems are already in use. The *Xenopus* gene symbols were also applied to other amphibian species.

3.2. Growth hormone family genes in vertebrate genomes

All identified full-length sequences contain the conserved GH family framework of four cysteine residues (Huising et al., 2006) and share a common organization of five protein-coding exons (Bachelot and Binart, 2007) encoding prehormone sequences of approximately 200–230 amino acids (Supplementary Data 2). Our maximum likelihood phylogeny of the GH family is presented in Fig. 1. It supports four main clusters – *GH*, *PRL*, *PRL2* and *SL* – diverging early in vertebrate evolution, each including sequences from lobe-finned fishes (*Sarcopterygii*), including tetrapods and coelacanth, as well as ray-finned fishes (*Actinopterygii*), including teleost fishes and spotted gar. This phylogeny is in agreement with the phylogeny previously published by Yamaguchi et al. in this journal (Yamaguchi et al.,

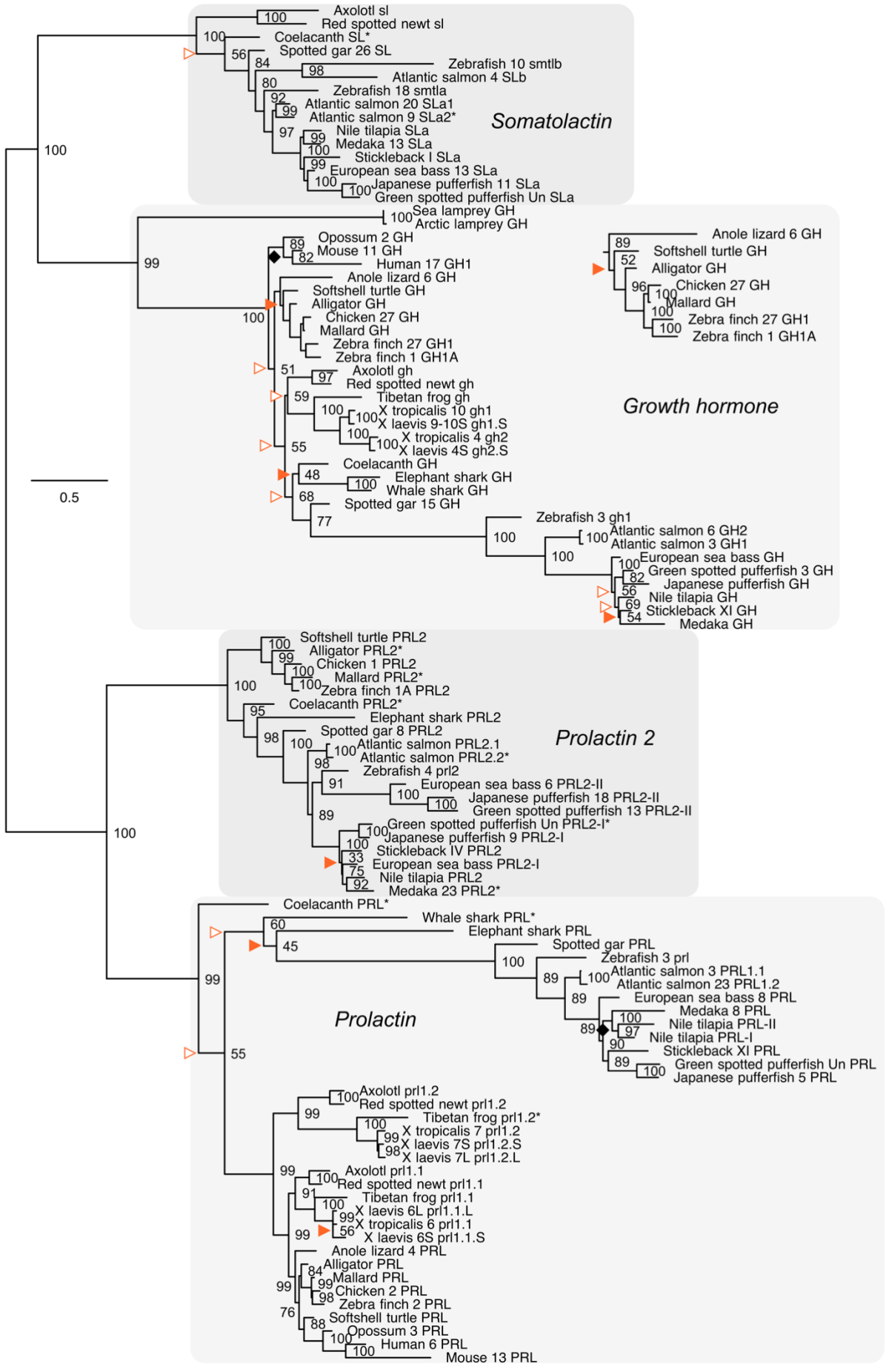

Somatolactin
Axolotl sl
Red spotted newt sl
Coelacanth SL*
Spotted gar 26 SL
Zebrafish 10 smtlb
Atlantic salmon 4 SLb
Zebrafish 18 smtla
Atlantic salmon 20 SLa1
Atlantic salmon 9 SLa2*
Nile tilapia SLa
Medaka 13 SLa
Stickleback I SLa
European sea bass 13 SLa
Japanese pufferfish 11 SLa
Green spotted pufferfish Un SLa
Growth hormone
Sea lamprey GH
Arctic lamprey GH
Opossum 2 GH
Mouse 11 GH
Human 17 GH1
Anole lizard 6 GH
Softshell turtle GH
Alligator GH
Chicken 27 GH
Mallard GH
Zebra finch 27 GH1
Zebra finch 1 GH1A
Axolotl gh
Red spotted newt gh
Tibetan frog gh
X tropicalis 10 gh1
X laevis 9-10S gh1.S
X tropicalis 4 gh2
X laevis 4S gh2.S
Coelacanth GH
Elephant shark GH
Whale shark GH
Spotted gar 15 GH
Zebrafish 3 gh1
Atlantic salmon 6 GH2
Atlantic salmon 3 GH1
European sea bass GH
Green spotted pufferfish 3 GH
Japanese pufferfish GH
Nile tilapia GH
Stickleback XI GH
Medaka GH
0.5
Prolactin 2
Softshell turtle PRL2
Alligator PRL2*
Chicken 1 PRL2
Mallard PRL2*
Zebra finch 1A PRL2
Coelacanth PRL2*
Elephant shark PRL2
Spotted gar 8 PRL2
Atlantic salmon PRL2.1
Atlantic salmon PRL2.2*
Zebrafish 4 prl2
European sea bass 6 PRL2-II
Japanese pufferfish 18 PRL2-II
Green spotted pufferfish 13 PRL2-II
Green spotted pufferfish Un PRL2-I*
Japanese pufferfish 9 PRL2-I
Stickleback IV PRL2
European sea bass PRL2-I
Nile tilapia PRL2
Medaka 23 PRL2*
Prolactin
Coelacanth PRL*
Whale shark PRL*
Elephant shark PRL
Spotted gar PRL
Zebrafish 3 prl
Atlantic salmon 3 PRL1.1
Atlantic salmon 23 PRL1.2
European sea bass 8 PRL
Medaka 8 PRL
Nile tilapia PRL-II
Nile tilapia PRL-I
Stickleback XI PRL
Green spotted pufferfish Un PRL
Japanese pufferfish 5 PRL
Axolotl prl1.2
Red spotted newt prl1.2
Tibetan frog prl1.2*
X tropicalis 7 prl1.2
X laevis 7S prl1.2.S
X laevis 7L prl1.2.L
Axolotl prl1.1
Red spotted newt prl1.1
Tibetan frog prl1.1
X laevis 6L prl1.1.L
X tropicalis 6 prl1.1
X laevis 6S prl1.1.S
Anole lizard 4 PRL
Alligator PRL
Mallard PRL
Chicken 2 PRL
Zebra finch 2 PRL
Softshell turtle PRL
Opossum 3 PRL
Human 6 PRL
Mouse 13 PRL

Fig. 1. Maximum likelihood phylogeny of the *GH*, *PRL*, *PRL2* and *SL* gene family. The tree topology is supported by non-parametric Ultra-Fast Bootstrap (UFBoot) and approximate Likelihood-Ratio Test (aLRT) analyses with 1000 replicates. Only UFBoot values are shown. Filled arrowheads indicate nodes with low support in both UFBoot (≤70%) and aLRT (≤50%) tests. Outlined arrowheads indicate nodes with low UFBoot support only, and diamonds indicate nodes with low aLRT support only. For clarity, some support values for shallow nodes are not shown. The tree is displayed as a midpoint-rooted phylogram. Approved gene symbols and nomenclature guidelines for human, mouse, chicken, *Xenopus tropicalis* 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. Download full-size image: https://doi.org/10.6084/m9.figshare.6494879.

2015), including the presence of all four family members in the coelacanth and spotted gar, as well as of *GH*, *PRL* and *PRL2* in the cartilaginous fish (*Chondrichthyes*) elephant shark. These species are important because they place the origin of at least the *GH*, *PRL* and *PRL2* genes to before the divergence of cartilaginous fishes and bony fishes (*Osteichthyes*). We could also identify *GH* and *PRL* sequences in the genome of the whale shark (Fig. 1), as well as a *GH* sequence in the draft genome of the little skate (Supplementary Data 1). The lack of *PRL2* in the whale shark, and of both *PRL* and *PRL2* in the little skate, are likely the result of the incomplete status of these genome assemblies rather than due to lineage-specific gene losses. In any case, an *SL* sequence could not be found in any of the investigated cartilaginous fish species. Fragments of the known sea lamprey *GH* sequence first identified by Kawauchi et al. 2002 and analyzed further by Moriyama et al. 2006 could be found in the current genome assembly. A full-length *GH* sequence could also be predicted from the genome assembly of the closely related Arctic lamprey. These sequences cluster in the basal-most position of the *GH* branch with reliable support (Fig. 1). No putative orthologs of *PRL*, *PRL2* or *SL* could be identified in either of the cyclostome species.

In mammalian genomes, only the known *GH* and *PRL* sequences could be found, including the many known mammalian lineage-specific duplicates (Forsyth and Wallis, 2002, Wallis, 2014) such as the human *GH2*, *CSH1*, *CSH2* and *CSHL* (Papper et al., 2009) and the duplicate genes of the greatly expanded *PRL* locus in the mouse genome (Wiemers et al., 2003). These relatively recent lineage-specific duplicates were not used in the phylogenetic analysis as the focus of this study was primarily on gene duplications in the early stages of vertebrate and teleost evolution.

In avian and non-avian reptile species, we could identify *PRL2* in addition to *GH* and *PRL* in all genomes that were investigated (Fig. 1). The short anole lizard *PRL2* sequence fragment first reported by Huang et al. 2009 was found in the genome database mapped to chromosome 5 (Supplementary Data 1), but was not included in the phylogenetic analyses due to its shortness. Additionally, the zebra finch has duplicate *GH* genes called *GH1* and *GH1A* (Xie et al., 2010) on chromosomes 27 and 1 respectively. These likely resulted from a gene duplication in a common ancestor of passerine birds (Yuri et al., 2008).

In amphibians, we could identify several *GH* and *PRL* sequences in the genome assemblies of the clawed frogs *Xenopus (Silurana) tropicalis* and *Xenopus laevis*, the Tibetan frog *Nanorana parkeri*, as well as, surprisingly, *SL* sequences alongside *GH* and *PRL* sequences in transcriptome assemblies from the salamander species axolotl and red spotted newt (Fig. 1). This is, to our knowledge, the first identification of *SL* in tetrapod species. *PRL2*sequences could not be found in any of the amphibian species. The duplicate *GH*

sequences found in both *Xenopus* species (Fig. 1) correspond to the known duplicate genes named *gh1* and *gh2*, first identified in *Xenopus laevis* (Huang and Brown, 2000, Martens et al., 1989). These could be found in both *X. tropicalis* and *X. laevis*, thus showing that the duplication took place in a common ancestor rather than in the tetraploidization event that occurred in the lineage leading to *X. laevis* (Hughes and Hughes, 1993). Only one *GH* sequence could be found in the *Nanorana parkeri* genome and the axolotl and red spotted newt transcriptome data. The scenario for the multiple *PRL* sequences that exist in amphibians is more ambiguous. We can identify two main *PRL* clades in all amphibian species which we have called *prl1.1* and *prl1.2*, following established gene nomenclature guidelines for *Xenopus* species (http://www.xenbase.org/gene/static/geneNomenclature.jsp). These correspond to the *X. tropicalis* sequences identified as *"PRL1A"* and *"PRL1B"* by Yamaguchi et al. 2015, and their proposed duplication scenario will be discussed below. There are additional copies of both *prl1.1* and *prl1.2* in *X. laevis*, and their chromosomal locations suggest that they arose through the tetraploidization event in this species (Fig. 1).

In teleost fishes, we could find all four family members (Fig. 1), with interesting additional duplicates in some lineages. The known duplicate *SLa* and *SLb* sequences (Zhu et al., 2004) were only found in the zebrafish and Atlantic salmon in this set of species; therefore, additional SL sequences were collected and used to construct a phylogeny of the *SL* branch (described below). There is also evidence of a duplication of *PRL2* in the green spotted pufferfish and Japanese pufferfish (*Tetraodontiformes*) as previously reported by Wang et al. 2010, as well as in the European sea bass, as previously reported by Tine et al. 2014. We have called these duplicates *PRL2-I* and *PRL2-II* in Fig. 1. The phylogeny is not completely resolved with respect to these duplicates, however the most parsimonious scenario is that the *PRL2* gene was duplicated in an eupercarian ancestor of both pufferfishes and the European sea bream. We could also identify the tandem duplicates of *PRL* in the Nile tilapia genome (*PRL-I* and *PRL-II* in Fig. 1) previously reported as *tiPRL-I* and *tiPRL-II* by Rentier-Delrue et al. 1989. In the Atlantic salmon genome, we could identify the known *GH1* and *GH2* genes (von Schalburg et al., 2008) as well as previously unknown duplicates of *PRL, PRL2* and *SLa* genes, which we have named *PRL1.1*, *PRL1.2*, *PRL2.1*, *PRL2.2*, *SLa1* and *SLa2* in Fig. 1. The chromosomal locations of these duplicate genes support an origin through the salmonid tetraploidization – see Fig. 2 in Lien et al. 2016. However, the *PRL2.1*and *PRL2.2* genes are located on unplaced scaffolds, so no clear inference as to their origin can be made. The Atlantic salmon *GH1* and *GH2* genes have likely had an intron insertion in the last coding exon, producing 6 rather than 5 coding exons (von Schalburg et al., 2008).

It has been previously observed that sequences of the *GH* family often do not cluster in accordance with the accepted species phylogeny in phylogenetic trees – See for instance Forsyth and Wallis, 2002, Huising et al., 2006). This is true also for our phylogeny (Fig. 1), where the positions of several sequences within each of the four main clusters deviate from the species phylogeny. This is especially notable for the elephant shark, whale shark, coelacanth and spotted gar sequences. The statistical support for several nodes within each cluster is also low, especially within the *GH* cluster. This is likely due to a combination of causes, including the relatively low sequence conservation between the gene family members as well as the varying evolutionary rates for these sequences in some lineages, in particular for *GH* and *PRL* (Forsyth and Wallis, 2002, Ryynänen and Primmer, 2006, Wallis,

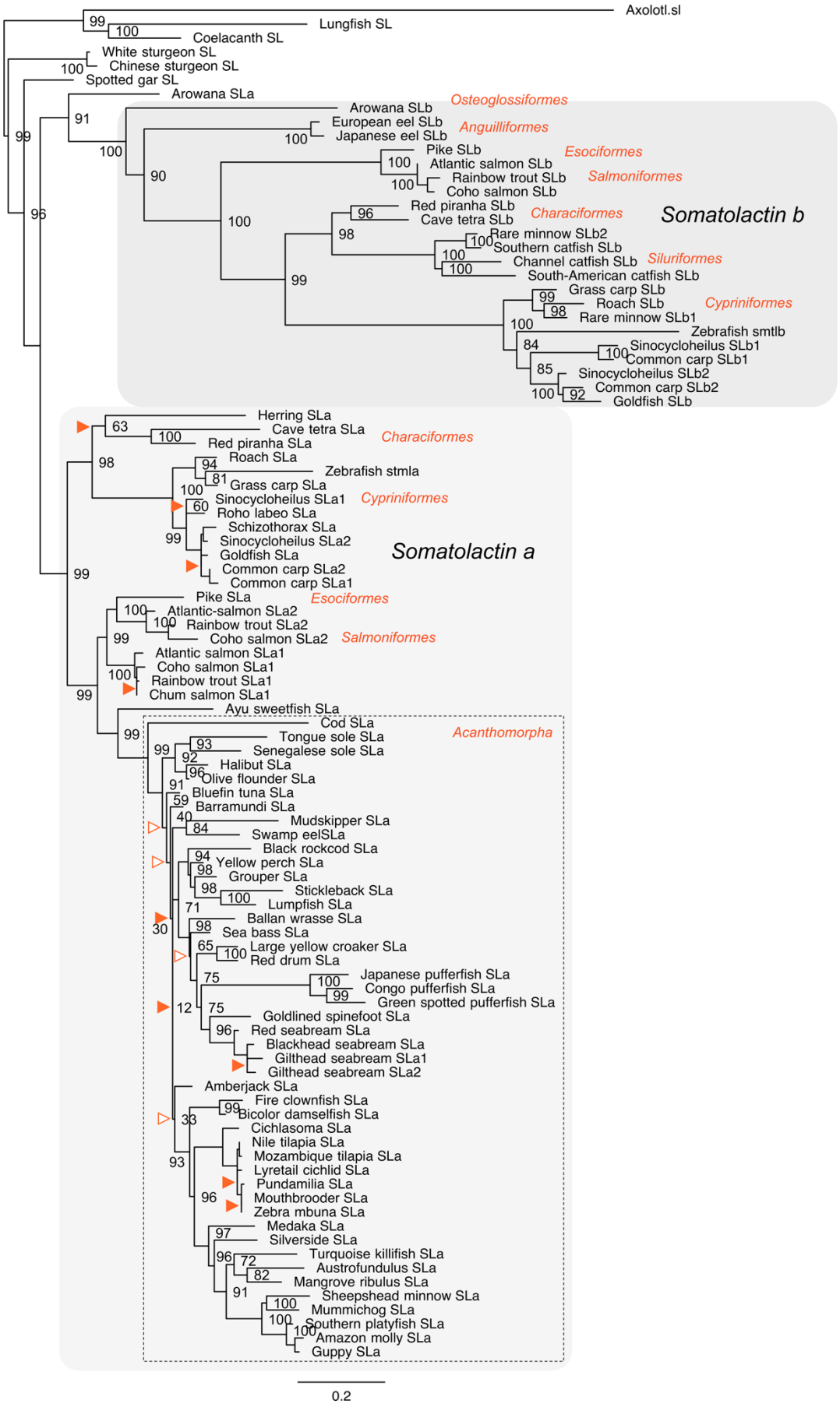

Axolotl.sl
Lungfish SL
Coelacanth SL
White sturgeon SL
Chinese sturgeon SL
Spotted gar SL
Arowana SLa
Osteoglossiformes
Arowana SLb
European eel SLb
Anguilliformes
Japanese eel SLb
Pike SLb
Esociformes
Atlantic salmon SLb
Rainbow trout SLb
Salmoniformes
Coho salmon SLb
Red piranha SLb
Cave tetra SLb
Characiformes
Somatolactin b
Rare minnow SLb2
Southern catfish SLb
Channel catfish SLb
Siluriformes
South-American catfish SLb
Grass carp SLb
Roach SLb
Cypriniformes
Rare minnow SLb1
Zebrafish smtlb
Sinocycloheilus SLb1
Common carp SLb1
Sinocycloheilus SLb2
Common carp SLb2
Goldfish SLb
Herring SLa
Cave tetra SLa
Characiformes
Red piranha SLa
Roach SLa
Zebrafish stmla
Grass carp SLa
Sinocycloheilus SLa1
Cypriniformes
Roho labeo SLa
Schizothorax SLa
Sinocycloheilus SLa2
Somatolactin a
Goldfish SLa
Common carp SLa2
Common carp SLa1
Pike SLa
Esociformes
Atlantic-salmon SLa2
Rainbow trout SLa2
Coho salmon SLa2
Salmoniformes
Atlantic salmon SLa1
Coho salmon SLa1
Rainbow trout SLa1
Chum salmon SLa1
Ayu sweetfish SLa
Acanthomorpha
Cod SLa
Tongue sole SLa
Senegalese sole SLa
Halibut SLa
Olive flounder SLa
Bluefin tuna SLa
Barramundi SLa
Mudskipper SLa
Swamp eelSLa
Black rockcod SLa
Yellow perch SLa
Grouper SLa
Stickleback SLa
Lumpfish SLa
Ballan wrasse SLa
Sea bass SLa
Large yellow croaker SLa
Red drum SLa
Japanese pufferfish SLa
Congo pufferfish SLa
Green spotted pufferfish SLa
Goldlined spinefoot SLa
Red seabream SLa
Blackhead seabream SLa
Gilthead seabream SLa1
Gilthead seabream SLa2
Amberjack SLa
Fire clownfish SLa
Bicolor damselfish SLa
Cichlasoma SLa
Nile tilapia SLa
Mozambique tilapia SLa
Lyretail cichlid SLa
Pundamilia SLa
Mouthbrooder SLa
Zebra mbuna SLa
Medaka SLa
Silverside SLa
Turquoise killifish SLa
Austrofundulus SLa
Mangrove ribulus SLa
Sheepshead minnow SLa
Mummichog SLa
Southern platyfish SLa
Amazon molly SLa
Guppy SLa
0.2

Fig. 2. Maximum likelihood phylogeny of somatolactin sequences. Node supports as in Fig. 1. For clarity, some support values for shallow nodes are not shown. The tree was rooted with the lobe-finned fish *SL* branch. Download full-size image: https://doi.org/10.6084/m9.figshare.6494879.

2014, Wallis, 2001, Wallis, 2000). Based on the relative branch lengths in our phylogenetic tree, several lineages seem to have had a relatively faster evolutionary rate – these include the teleost *GH* branch, the ray-finned fish *PRL* branch, the amphibian *prl1.2* branch, the *SLb* branch in teleost fishes, and the *PRL2-II* sequences found in pufferfishes and European sea bass (Fig. 1). In addition, some of the newly identified sequences could not be predicted in full length (indicated with asterisks in Fig. 1), which introduces gaps into the alignment and could affect the phylogenetic signal negatively. The reported low rate of sequence evolution in protein-coding genes for coelacanth (Amemiya et al., 2013), spotted gar (Braasch et al., 2016) and elephant shark (Venkatesh et al., 2014) could also contribute to the deviations from the species phylogeny. In our tree, this low evolutionary rate can be seen for the coelacanth *GH*, *PRL2* and *SL* sequences, and notably also for the spotted gar orthologs.

3.3. Putative GH-family members in invertebrates

Putative invertebrate homologs of GH-family sequences were sought using BLAST (Altschul et al., 1990), Pattern-Hit Initiated (PHI) BLAST (Sayers et al., 2010) and profile-Hidden Markov protein sequence similarity (Finn et al., 2011) search strategies in several genome and reference sequence databases of invertebrate species. It was not possible to identify any putative invertebrate homologs with the relevant sequence similarity, even with suboptimal expect values. A detailed description of our search methods and results is available at https://doi.org/10.6084/m9.figshare.750470.

More recently, a putative *GH*-like family member was identified in the Japanese lancelet (*Branchiostoma japonicum*) and the Florida lancelet (*Branchiostoma floridae*) (Li et al., 2014). Notably, these lancelet sequences lack the conserved framework of four cysteine residues that characterize the *GH* family hormones. We conducted BLAST searches of the *Branchiostoma floridae* genome using the sequences reported in Li et al. (2014) as queries and identified two additional sequences (Supplementary Data 1). We constructed a phylogeny including the identified *Branchiostoma* sequences, which shows the *Branchiostoma* branch diverging from a basal node between the *GH*/*SL* and *PRL*/*PRL2*clades (Fig. S1). This is consistent with the suggestion that the *Branchiostoma* sequences represent the common ancestor of all four basal vertebrate *GH* family members. However, this analysis also indicates that the *Branchiostoma* sequences are highly derived and have had an accelerated evolutionary rate in comparison with the vertebrate *GH* family sequences, which may introduce artefacts into the phylogeny. The lack of a reliable outgroup to root this analysis also makes its position within the wider class I cytokine family uncertain. The phylogenetic analysis presented by Li et al. (2014), which includes a wide selection of class I cytokines, also shows the putative *Branchiostoma GH*-like sequence clustering basal to the vertebrate *GH, PRL, PRL2,* and *SL* sequences. However, it does not include a node support value for the *Branchiostoma* branch, which makes the interpretation of the phylogeny inconclusive. The lack of a reliable outgroup also means that the *Branchiostoma* sequences can't be used to determine the relative timing of the earliest divergences in the *GH*-family

tree with certainty. Further studies are underway to resolve these questions, including synteny analyses in the *Branchiostoma floridae* genome.

3.4. Phylogeny of the SL branch in teleost fishes

In addition to the SL sequences included in Fig. 1, we identified 92 more from 63 teleost fish species, two sturgeon species as well as the West African lungfish (Supplementary Data 1). Our phylogenetic tree of the SL branch (Fig. 2) supports the duplication of an ancestral *SL* gene after the divergence of the spotted gar, likely giving rise to *SLa* and *SLb* early in teleost evolution within the time window of the basal teleost tetraploidization (3R). Most of the identified *SL* sequences belong to the *SLa* branch. The smaller *SLb* cluster in this tree includes known sequences from several cyprinid fishes (*Cypriniformes*) – goldfish (Cheng et al., 1997), zebrafish (Zhu et al., 2004), grass carp (Jiang et al., 2008) and roach (Kroupová et al., 2012); a silurid fish (*Siluriformes*) – the South American catfish (Baldisserotto et al., 2014); salmonid fishes (*Salmoniformes*) – Atlantic salmon (Benedet et al., 2008) and rainbow trout (Yang and Chen, 2003); as well as from the European eel (May et al., 1997) (*Elapomorpha*). In addition to these, we found *SLb* sequences from the Asian arowana (*Osteoglossomorpha*), Japanese eel (*Elopomorpha*), Northern pike (*Esociformes*), coho salmon (*Salmoniformes*), Mexican cave tetra and red piranha (*Characiformes*), Southern catfish and channel catfish (*Siluriformes*), rare minnow, common carp and the cave barbel *Sinocycloheilus anshuensis* (*Cypriniformes*). For most these species, *SLa* is also known or could be found. Notably, *SLa* could not be found from the European and Japanese eels, the rare minnow and most catfish species. However, a likely *SLa* pseudogene could be identified in the channel catfish genome (Supplementary Data 1). The Asian arowana *SLa* sequence clusters basal to the *SLb* branch, rather than at the basal-most position of the *SLa* branch (Fig. 2). This is likely due to its relatively slow evolutionary rate (as demonstrated by its short branch length). Several diagnostic amino acid positions within the alignment nonetheless indicate that it is an *SLa* sequence (Supplementary Data 3).

The species where *SLb* could be found represent basal lineages that diverged after the teleost whole genome doubling (3R), but before the large teleost radiations within spiny-rayed fishes (*Acanthomorpha*) (Near et al., 2013, Near et al., 2012). The Atlantic herring also represents a basal teleost lineage, *Clupeiformes*; however, in this species only an *SLa* sequence could be found, despite repeated searches of its genome with *SLb* sequences from closely related species. Similarly, only an *SLa* sequence could be identified from the ayu sweetfish, which belongs to the basal teleost order *Osmeriformes*. Thus, this suggests at least two, possibly three, independent losses of *SLb*.

We could identify duplicates of both *SLa* and *SLb* in the genomes of the cyprinid species common carp and *Sinocycloheilus anshuiensis*. These are named *SLa1*, *SLa2*, *SLb1* and *SLb2* in Fig. 2. However, the duplicates may have arisen independently in these two lineages because of several independent tetraploidizations within *Cypriniformes* (Glasauer and Neuhauss, 2014), including in the common carp and *Sinocycloheilus* (Wang et al., 2012, Xu et al., 2014, Yang et al., 2016). To complicate this scenario further, the chromosomal locations of the common carp *SLa1*, *SLa2*, *SLb1* and *SLb2* genes (Supplementary Data 1) do not match the identified patterns of paralogy after tetraploidization – see Figs. 1 and 2 in Xu et al. 2014. Thus, further analyses in cyprinid species are needed to ascertain the origin of these *SLa* and *SLb* duplicates as well as their orthology relationships between species. It has

been suggested that the tetraploidization event in common carp is shared with the closely related goldfish (Glasauer and Neuhauss, 2014), however we could only identify one *SLa* and one *SLb* sequence from goldfish (Fig. 2), and the lack of a genome sequence from this species precludes the identification of any putative duplicates.

As for the salmonid tetraploidization, we could identify orthologs of the Atlantic salmon *SLa2* gene in the genomes of the coho salmon and rainbow trout (Fig. 2). The known *SLa* sequences from these two species, as well as from the chum salmon, are orthologs of the Atlantic salmon *SLa1*. In our phylogeny of the *SL* branch (Fig. 2), the salmonid *SLa2* branch clusters with the Northern pike *SLa* sequence, rather than with its sister *SLa1* branch. This is likely due to the shortness of the partial *SLa* sequence from the Northern pike. The most parsimonious scenario is that the salmonid *SLa1* and *SLa2* genes arose in the ancestral salmonid tetraploidization event, as described in Section 3.1 above.

3.5. Patterns of conserved cysteine residues

In addition to the conserved framework of four cysteine residues that is characteristic for the growth hormone family (Huising et al., 2006), the tetrapod prolactin proteins have two additional cysteine residues in the amino terminal region. It has been proposed that this cysteine pair is ancestral, and has been lost in the teleost prolactins, possibly through the partial deletion of the second exon (Manzon, 2002, Whittington and Wilson, 2013). We could identify this cysteine residue pair in all PRL, PRL2 and SL amino acid sequences, with the exception of the teleost and spotted gar PRL sequences, the elephant shark PRL and PRL2 sequences (but the cysteine pair is present in the whale shark PRL), and the green spotted pufferfish SLa sequence, where the first position has been substituted. The alignment file that our analyses are based on has been included as Supplementary Data 2 (see also Section 3.8 below). The elephant shark PRL sequence has a single cysteine in this region, but it does not seem to correspond to any of the two conserved amino terminal cysteines. This amino terminal cysteine pair is also present in both PRL and PRL2 proteins in sturgeon (Huang et al., 2009, Noso et al., 1993) These findings could suggest a closer relationship between SL and PRL/PRL2. However, it is also conceivable that it is an ancestral feature that has been lost in GH, ray-finned fish PRL (except sturgeon) and elephant shark PRL and PRL2 independently, but this explanation seems less parsimonious.

Most identified SL sequences have a third conserved cysteine residue encoded by the second exon, including the axolotl, red spotted newt, coelacanth and spotted gar SL sequences (Supplementary data 3). This cysteine residue is substituted in most SLb proteins (Benedet et al., 2008, Zhu et al., 2004), with the exception of the eel SLb sequences, as well as in the SLa sequences of few species (Fig. S2). The role of this cysteine is unclear.

3.6. Conserved synteny analyses

Three subsets of neighboring gene families were selected from the overall list of gene predictions in the vicinity of *GH*, *PRL*, *PRL2* and *SL*-genes across the chicken, spotted gar and zebrafish genomes. In the first subset, we selected gene families neighboring *PRL* and *PRL2* genes to test the hypothesis that these two genes arose through the duplication of an ancestral *PRL/PRL2* gene in the basal vertebrate tetraploidizations (1R and 2R). This resulted in 14 gene families neighboring *PRL* and *PRL2*: BCAT, CALU, KCNA, LIN7, LRRC4, MYBPC,

PAX4/6/10, PPFIA, PRMT1/8, SLC16A1, SYT1/2/5/8, SYT3/6/9/10, TNNT and TSPAN4/6/25 (Table 1). In the second subset of neighboring gene families, we selected those with members in the vicinity of *smtla* (*SLa*) and *smtlb* (*SLb*) genes in the zebrafish genome, *as well as* the *SL* gene in the spotted gar genome, to test the hypothesis that the *SLa* and *SLb* genes arose through the duplication of an ancestral *SL* gene in the basal teleost tetraploidization (3R). This resulted in six gene families: CLMP, FCHSD, GRAMD1, KCNJ1/10/15, KIRREL and RSF1 (Table 2). Lastly, we investigated the subset of gene families neighboring *at least three* of the *GH*, *PRL*, *PRL2*, and *SL* genes regardless of the species. This was done to test all alternative hypotheses regarding the origin of the *GH*-family genes, including whether *GH* and *SL* also arose in the time window of 1R and 2R. Eleven additional gene families were thus selected: ABCC8/9, CRY, D2, KCNJ3/5/6/9, MAZ, QSER/PRR12, RAP, SETD1, SHANK, SLC1A and YPEL (Table 3). The locations of the identified member genes in these families were recorded for 11 vertebrate species with mapped genome assemblies. Phylogenetic analyses of all neighboring gene families were carried out to determine the relative time windows of their expansions through gene duplications, and to determine orthology (between species) and paralogy (within species) relationships (Supplementary Figs. S2 – S32). Complete information about the neighboring gene families, including database identifiers, location data as well as sequence annotation and prediction notes, is provided in Supplementary Data 4. All corresponding alignments and phylogenetic trees described below have been provided as a citable file set with a stable identifier – See Section 3.8 below. The gene lists that underlie our conserved synteny analyses and the selection of neighboring gene families are included in Supplementary Data 5.

The two basal vertebrate tetraploidizations resulted in a large number of quartets of related chromosome regions, each such quartet is called a *paralogon* and related chromosome regions are said to be *paralogous*. From the studies of conserved synteny described above we could determine two sets of paralogous chromosome regions. The first set is centered around the chromosomal regions of *GH*, *PRL* and *PRL2*, and the second is centered around the *SL*-bearing chromosome regions. Notably, the D2 family has members in both regions.

Table 1. Neighboring gene families in the vicinity of *PRL* and *PRL2* genes

Symbol	Description	Suppl. Fig.	Outgroup
BCAT	Branched chain amino-acid transaminase	Fig. S2	Fruit fly
CALU	Reticulocalbin and calumenin	Fig. S3	Fruit fly
KCNA	Voltage-gated potassium channel subfamily A	Fig. S4	Fruit fly
LIN7	*lin-7* homolog	Fig. S5	Fruit fly
LRRC4	Leucine-rich repeat containing 4	Fig. S6	Florida lancelet
MYBPC	Myosin-binding protein C	Fig. S7	Vase tunicate
PAX4/6/10	Paired box 4, -6 and -10	Fig. S8	Fruit fly
PPFIA	PTPRF-interacting protein alpha	Fig. S9	Fruit fly
PRMT1/8	Protein arginine methyltransferase 1 and -8	Fig. S10	Fruit fly
SLC16A1	Solute carrier family 16, subfamily 1	Fig. S11	Unrooted
SYT1/2/5/8	Synaptotagmin 1, -2, -5 and -8	Fig. S12	Fruit fly
SYT3/6/9/10	Synaptotagmin 3, -6, -9 and -10	Fig. S13	Honey bee
TNNT	Troponin T	Fig. S14	Florida lancelet
TSPAN4/6/25	Tetraspanin 4, -9 and -25 (CD53)	Fig. S15	Purple sea urchin

Table 2. Neighboring gene families in the vicinity of *SLa* and *SLb* genes.

Symbol	Description	Suppl. Fig.	Outgroup
CLMP	Coxsackie- and adenovirus receptor-like membrane	Fig. S16	Elephant shark
FCHSD	FCH and double SH3 domain protein	Fig. S17	Fruit fly
GRAMD1	GRAM domain-containing protein 1	Fig. S18	Red flour beetle
KCNJ1/10/15	Voltage-gated potassium channel J, member 1, -10 and	Fig. S19	Unrooted
KIRREL	Kin of irregular chiasm-like protein	Fig. S20	Fruit fly
RSF1	Remodeling and spacing factor 1	Fig. S21	Elephant shark

Table 3. Neighboring gene families in the vicinity of several *GH*-family members.

Symbol	Description	Suppl. Fig.	Outgroup
ABCC8/9	ATP-binding cassette subfamily C, member 8 and -9	Fig. S22	Fruit fly
CRY	Cryptochrome	Fig. S23	Fruit fly
D2	Dopamine receptor D2 family	Fig. S24	Fruit fly
KCNJ3/5/6/9	Voltage-gated potassium channel J, member 3, -5, -6	Fig. S25	Fruit fly
MAZ	MYC-associated zinc finger transcription factors	Fig. S26	Unrooted
QSER/PRR12	Glutamine and serine rich 1 protein and proline rich 12	Fig. S27	Unrooted
RAP	RAS oncogene family	Fig. S28	Fruit fly
SETD1	SET domain-containing 1	Fig. S29	Fruit fly
SHANK	SH3 and multiple ankyrin repeat domains	Fig. S30	Fruit fly
SLC1A	Solute carrier family 1	Fig. S31	Fruit fly
YPEL	*yippee* like	Fig. S32	Fruit fly

In total, 24 neighboring families, from subsets 1 and 3 combined, were found to have member genes within the *PRL* and *PRL2*-bearing paralogous regions. In the human genome, these regions correspond mainly to segments of chromosomes 1, 11, 12, 22 and 7, and 19 (Fig. 3). Note the absence of *PRL2* as well as the locations of the *GH1* and *PRL* genes outside of the paralogy group in the human genome. In the chicken genome, there is a *PRL2* gene on chromosome 1 but the *GH* and *PRL* genes are located outside of the identified paralogy group, while in the *Xenopus tropicalis* genome, the *PRL2* gene is absent but there are *GH* and *PRL* genes both within the paralogy group (*gh2* and *prl1.2*) and outside of it (*gh1* and *prl1.1*) (Fig. 3). In the chicken genome, the corresponding genes that make up the paralogous region on human chromosome 19 could not be identified, possibly because they are located on a GC-rich microchromosome (Bornelöv et al., 2017, Hron et al., 2015, McQueen et al., 1996). These paralogous regions are also apparent in ray-finned fishes (*Actinopterygii*), such as the spotted gar and the teleost fishes medaka and zebrafish (Fig. 4). In teleost genomes the *PRL* genes are located within the identified paralogy group, on chromosome 8 in the medaka and chromosome 3 in the zebrafish (Fig. 4). In several teleost genomes, the *GH* and *PRL* genes are syntenic. Note that the location of the spotted gar *PRL* gene is unknown, and many of the *PRL*-neighboring genes could not be found in this genome.

Most of the identified neighboring gene families in this set of paralogous chromosome regions, 13 out of 24, support the duplication of an ancestral *PRL*/*PRL2* gene,

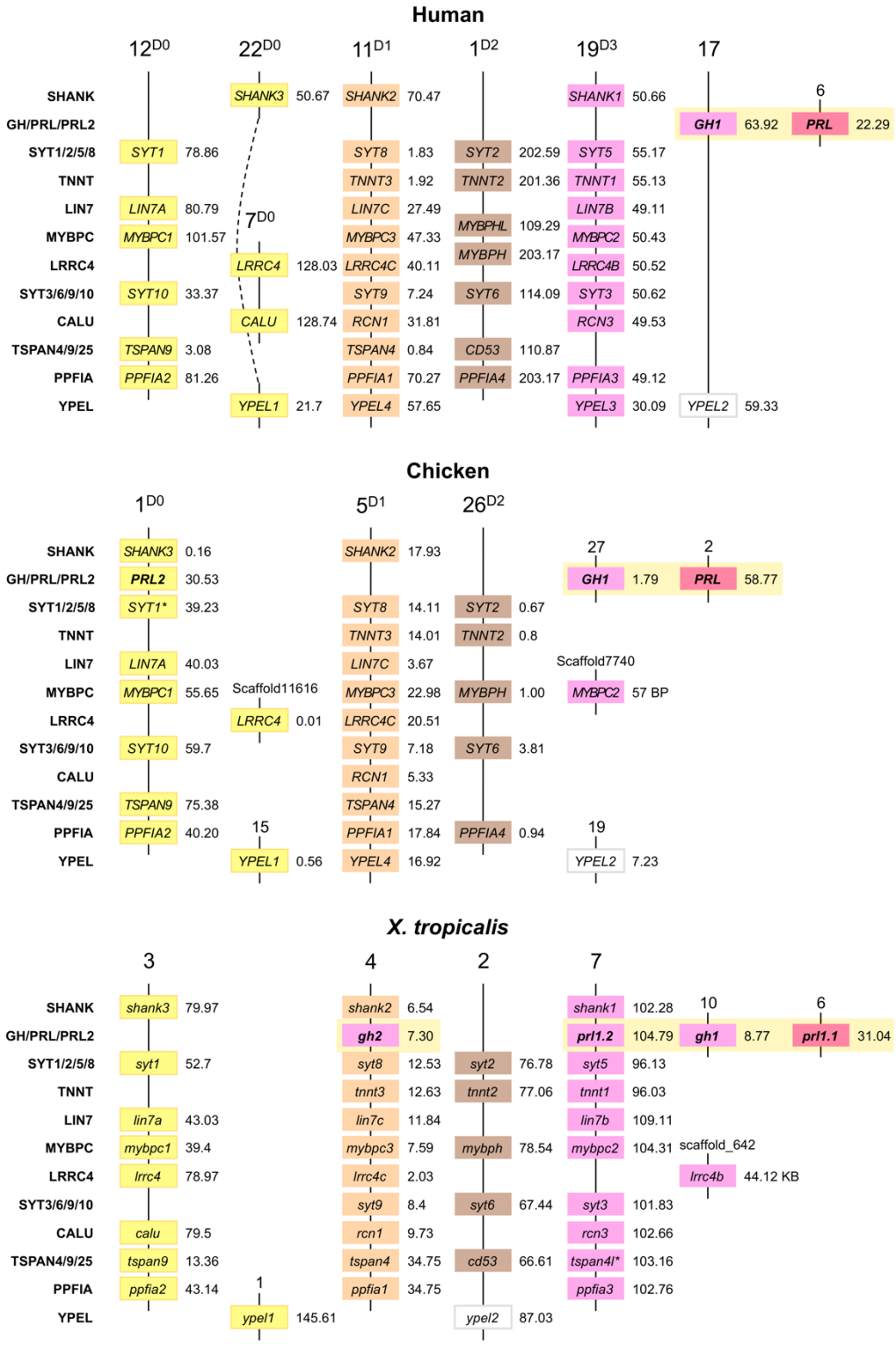


Fig. 3. Conserved synteny blocks in the vicinity of GH, PRL, and PRL2 genes in the human, chicken and *Xenopus tropicalis* genomes. 11 out of 24 neighboring gene families are shown. For the full representation of gene families see Figs. S2, S3 and S4. Locations are given in MB unless indicated otherwise. Gene order follows chicken chromosome 1. New suggested gene symbols are indicated by asterisks. Colors are applied following human chromosomes in order to show conserved synteny as well as gene orthology between species. GH genes are also colored magenta to indicate the ancestral co-localization of GH and PRL. A deeper color is used to indicate tetrapod PRL genes located in non-paralogous chromosome regions. The *YPEL2* branch is not supported in the phylogeny of this family (white bordered boxes). Grey color indicates uncertain synteny

relationships to the identified paralogous chromosome regions (*YPEL2*, *TNNT5* in Fig. 4). D0 – D3 designations next to the human, chicken and medaka (Fig. 4) chromosome numbers correspond to vertebrate paralogous blocks identified by Nakatani et al., 2007 to have arisen through 1R and 2R. Download full-size image: https://doi.org/10.6084/m9.figshare.6494879.

generating *PRL* and *PRL2*, in the same time-window as the basal vertebrate tetraploidizations (1R/2R), i.e. after the divergence of invertebrate chordates and vertebrates, but before the divergence of lobe-finned fishes (*Sarcopterygii*) and ray-finned fishes: BCAT (Fig. S2), CALU (Fig. S3), KCNA (Fig. S4), LIN7 (Fig. S5), LRRC4 (Fig. S6), PPFIA (Fig. S9), PRMT1/8 (Fig. S10), SYT1/2/5/8 (Fig. S12), SYT3/6/9/10 (Fig. S13), TNNT (Fig. S14), TSPAN4/6/25 (Fig. S15), SETD1B (Fig. S29), and SHANK (Fig. S30). Five more families; MAZ (Fig. S26), MYBPC (Fig. S7), PAX4/6/10 (Fig. S8), SLC16A1 (Fig. S11), and YPEL (Fig. S32) have members in the *PRL* and *PRL2*-bearing chromosome blocks, and are consistent with expansions through 1R and 2R. However, their phylogenies are either inconclusive with regard to the time window of 1R/2R, or have branches that have low support or are unresolved. See the corresponding supplementary figure captions for details. An additional six families from subset 3 support duplications in 1R and 2R – ABCC8/9 (Fig. S22), CRY (Fig. S23) and SLC1A (Fig. S31) – or are consistent with this scenario – D2 (Fig. S23), QSER/PRR12 (Fig. S27) and RAP (Fig. S28). However, their member genes are located in chromosome blocks other than those bearing *PRL* and *PRL2*. None of these neighboring gene families directly contradict gene duplications in the time window of 1R/2R. Several of these neighboring gene families also have duplicate teleost branches that support duplications in the time-window of the basal teleost tetraploidization (3R), and some have branches that diverged before 1R/2R – The large SLC1A family is made up of nine genes in vertebrates (Fig. S31), of which only the *SLC1A2* and *SLC1A9* genes are located in this paralogon. In addition, the *CRY4* and *CRY5* branches of the CRY family seem to pre-date vertebrates (Fig. S23), although the genes are located in the vicinity of *CRY2* and *CRY1B* respectively. See the corresponding supplementary figure captions for details on the phylogeny of each gene family.

In total, eight neighboring gene families from subsets 2 and 3 were found in the set of paralogous chromosome regions bearing the *SL* genes. The identified paralogous chromosome blocks in human, chicken, *Xenopus tropicalis,* spotted gar, zebrafish and medaka are summarized in Fig. 5. Of the eight *SL*-neighboring gene families, five support the duplication of an ancestral *SL* gene giving rise to *SLa* and *SLb* in the time window of 3R – CLMP (Fig. S16), GRAMD1 (Fig. S18), KCNJ1/10/15 (Fig. S19), KIRREL (Fig. S20), and RSF1 (Fig. S21). For a sixth family, FCHSD, the *FCHSD2a* and *-2b* branches do not diverge in the time-window of 3R, likely due to uneven evolutionary rates between the two duplicates (Fig. S17). However, their chromosomal locations are compatible with duplications in 3R (Fig. 5). In addition, the D2 (Fig. S24) and KCNJ3/5/6/9 (Fig. S25) families have members in the same paralogous chromosome blocks as *SLa*, but not *SLb*, in teleost fishes. Taken together, four of these gene families also support gene duplications in the 1R/2R tetraploidizations (GRAMD1, FCHSD, KIRREL and KCNJ3/5/6/9), and two more are consistent with this scenario, at the very least (D2 and KCNJ1/10/15). The remaining two gene families (CLMP and RSF1) do not have additional members diverging early in vertebrate evolution. See the corresponding supplementary figure captions for details.

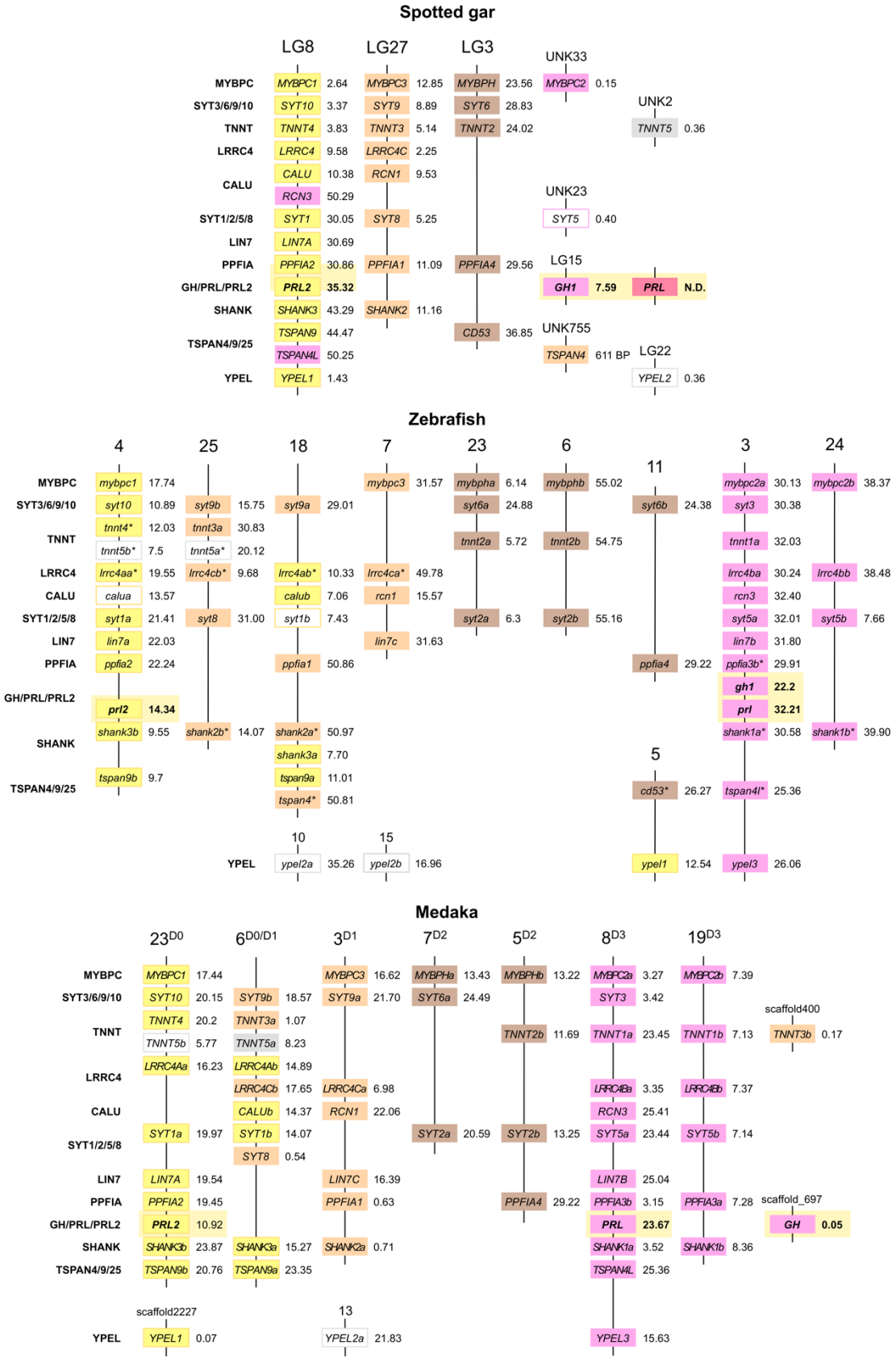


Fig. 4. Conserved synteny blocks in the vicinity of GH, PRL, and PRL2 genes in the spotted gar, zebrafish and medaka genomes. See Fig. 3 caption for details. Gene order follows spotted gar linkage group 8 (except RCN3 and TSPAN4L for clarity). The spotted gar SYT5 gene could not be used in the phylogeny of this family. White

bordered boxes in zebrafish and medaka can indicate unclear divergence times relative to 3R as well as 1R/2R. The chromosomal location of the spotted gar PRL gene is currently unknown. Download full-size image: https://doi.org/10.6084/m9.figshare.6494879.

The *PRL*/*PRL2*-bearing paralogous chromosome regions and the *SL*-bearing paralogous chromosome regions coincide on some chromosomes, for instance on human chromosomes 1, 11 and 19, chicken chromosome 1, and *Xenopus tropicalis* chromosome 7 (Fig. 3, Fig. 5), as well as zebrafish chromosomes 7, 10, 18 and 24 (Fig. 4, Fig. 5). However, the overlap between the two sets of paralogous chromosome blocks is not complete or perfect. For example, in some species, such as the spotted gar and medaka, the chromosomes/linkage groups do not coincide at all. In any case, our results suggest that the *SL*-bearing chromosome blocks are not part of the same fourfold paralogy as the *PRL/PRL2*-bearing paralogy group, but rather that the two paralogy groups partially overlap. This indicates a complex evolutionary history for these chromosomal regions.

3.7. Analysis of GH and PRL gene co-localization

In the zebrafish genome, as well as several other mapped teleost genomes, the *GH* and *PRL* genes are co-located on the same chromosomes, whereas in tetrapod species they are not. Because we could not identify neighboring genes in the vicinity of *GH* in the conserved synteny analysis described above, and to elucidate the ancestral organization of *GH* genes relative to *PRL* genes, we carried out an additional comparative analysis of neighboring genes. We selected neighboring genes shared between at least three of the *GH* and *PRL*-bearing chromosome regions in the zebrafish, spotted gar and chicken genomes. Subsequently, the locations of the gene orthologs in the human and *Xenopus tropicalis* genomes were added to the analysis (Supplementary Data 6). These results are summarized in Supplementary Fig. S36. We found a remarkable lack of conserved synteny between the *PRL*-bearing chromosome block in zebrafish and the *PRL*-bearing blocks in human and chicken. Our synteny analysis described in Section 3.6 showed conserved synteny between the *gh1* and *prl*-bearing chromosome blocks on zebrafish chromosome 3 and the *Xenopus tropicalis prl1.2* locus, but not the *prl1.1* locus (Fig. 3). The same complex pattern of conserved synteny was shown by Yamaguchi et al. (2015), and our results corroborate their proposed scenario for this loss of conserved synteny (discussed below). Conversely, there is a clear pattern of conserved synteny between the *gh1* and *prl*-bearing chromosome blocks on zebrafish chromosome 3 and the *GH*-bearing chromosome blocks in the spotted gar, chicken and human genomes (Fig. S36). In *Xenopus tropicalis*, which has two *GH* genes, *gh1* and *gh2*, conserved synteny was found with only the *gh1* gene. These results suggest that the co-location of *GH* and *PRL* genes in teleost fish genomes likely represents the ancestral organization, at least just before the divergence between lobe-finned fishes and ray-finned fishes. Subsequently, the chromosomal regions carrying *GH* and *PRL* broke apart in the tetrapod lineage, likely due to chromosome fissions and rearrangements. Aside from *GH* and *PRL,* we found no paralogous genes shared between the two chromosome blocks.

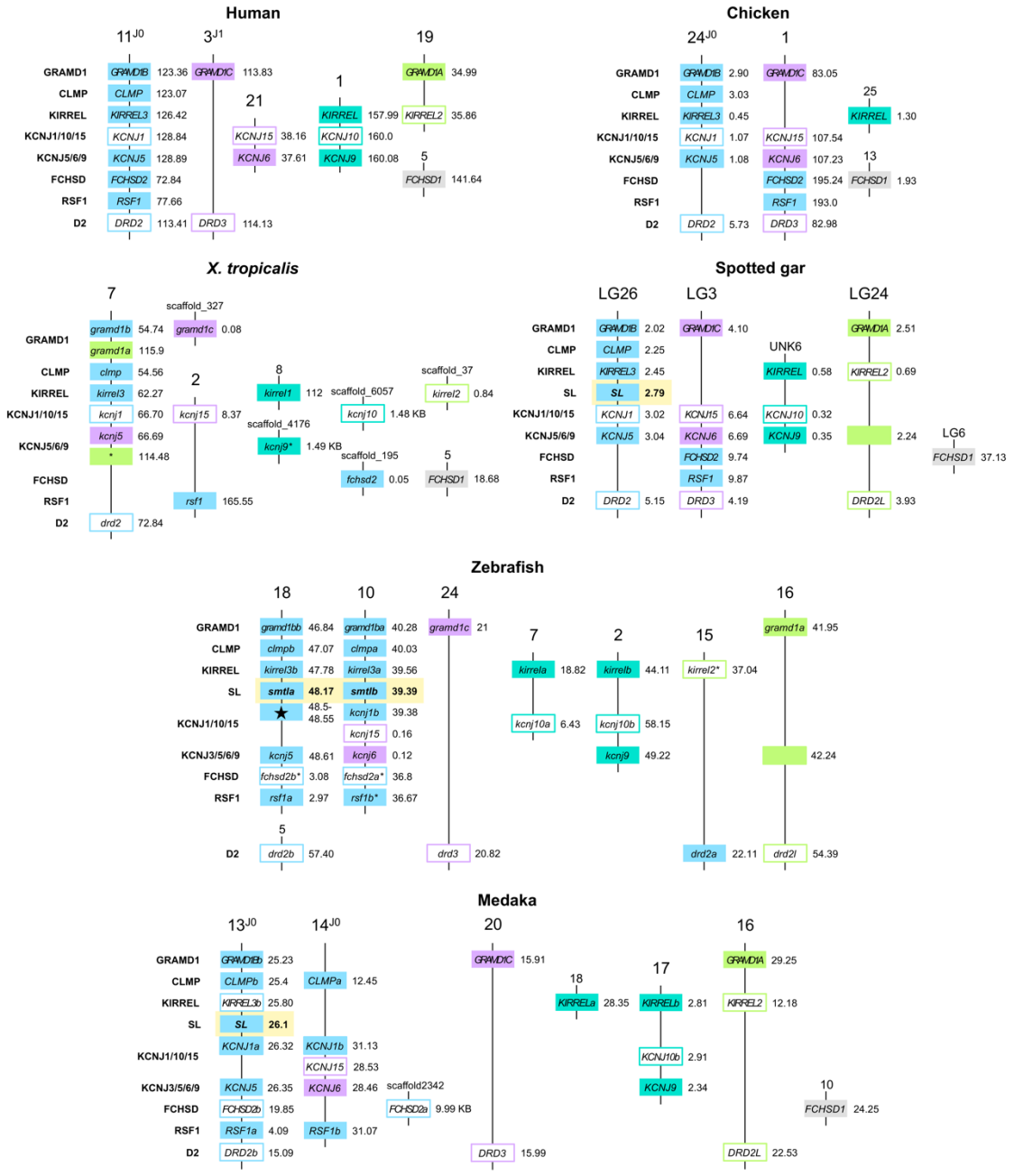


Fig. 5. Conserved synteny blocks in the vicinity of SL genes in the human, chicken, Xenopus tropicalis, spotted gar, medaka and zebrafish genomes. Locations are given in MB unless indicated otherwise. Gene order follows spotted gar linkage group 26 and colors are applied following the human chromosomes. New suggested gene symbols are indicated by asterisks. White bordered boxes indicate genes with unclear divergence times relative to 1R/2R; due to the lack of outgroups for relative dating (D2 and KCNJ1/10/15 families) or low support in the phylogenetic analyses (KIRREL2) It also indicates genes with unclear divergence times relative to 3R (FCHSD2a and -2b, and KCNJ10a and -10b). Grey color indicates uncertain synteny relationships to the identified paralogous chromosome regions (FCHSD1). The unnamed KCNJ gene on Xenopus tropicalis chromosome 7 (marked with an asterisk) has wrongly been identified as kcnj9. In fact, we could find the correct kcnj9 gene on scaffold_7147. Putative orthologs of this gene in the spotted gar and zebrafish are shown as empty boxes. J0 – J1 designations next to human, chicken and medaka chromosome numbers correspond to vertebrate paralogous blocks identified by Nakatani et al., 2007 to have arisen through 1R and 2R. For the D2 family, only the genes located within this paralogon (excluding DRD4 and DRD4-RS) are shown. Download full-size image: https://doi.org/10.6084/m9.figshare.6494879.

3.8. Description of supporting data 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.5284144. This file set includes all alignment and phylogenetic tree files described in this study, including neighboring gene families. It also includes an unaligned sequence file with all curated *GH*, *PRL*, *PRL2* and *SL* sequences identified in this study.

4. Discussion

4.1. Evolution of the GH, PRL, PRL2 and SL family

The evolution of the growth hormone family of genes has been extensively investigated over the years, including discussions of varying evolutionary rates in different vertebrate lineages (Ryynänen and Primmer, 2006, Wallis, 2001); the duplication and diversification of GH and PRL genes in some species (Soares, 2004, Wallis, 2014); as well as evolutionary aspects of receptor interactions (Chen et al., 2011, Ellens et al., 2013, Fukamachi and Meyer, 2007). In this study, we have investigated the first appearance, duplications and divergence of the four family members – growth hormone (*GH)*, prolactin (*PRL*), prolactin 2 (*PRL2*) and somatolactin (*SL*) – in relation to the two rounds of basal vertebrate tetraploidization, called 1R and 2R, as well as the basal teleost tetraploidization 3R. Our approach is broader than in previously published studies: In addition to comprehensive phylogenetic analyses of the growth hormone family itself, we have also analyzed the genomic regions of the *GH* family genes for patterns of conserved synteny and reconstructed the phylogenies of 31 neighboring gene families. This allowed us to infer the orthology and paralogy relationships between the *GH* family gene-bearing chromosome regions and determine the time window of the duplications that gave rise to them.

Our analyses support the following complex scenario, summarized in Fig. 6: Ancestral *GH*, *PRL/PRL2*, and possibly *SL*, genes were present early in vertebrate evolution, before the first basal tetraploidization 1R. The ancestral *GH* and *PRL/PRL2* genes were likely located on the same chromosome. Subsequently, one of the basal vertebrate tetraploidizations, either 1R or 2R, gave rise to the *PRL* and *PRL2* genes. We present both possible scenarios in Fig. 6. No other gene duplicates were preserved from these tetraploidization events. Subsequently, there were gene losses in the jawless vertebrate (*Agnatha*) lineage (*PRL, PRL2* and *SL*), as well as in the cartilaginous fish (*Chondrichthyes*) lineage (*SL*) within the jawed vertebrates (*Gnathostomata*). All four member genes are present in the coelacanth and the spotted gar, confirming that all four genes were present in a bony fish (*Osteichthyes*) ancestor, before the divergence of lobe-finned fishes (*Sarcopterygii*) and ray-finned fishes (*Actinopterygii*). Our scenario contradicts an earlier interpretation that the *GH* in lampreys represents the ancestral hormone of the family in vertebrates (Kawauchi and Sower, 2006). Our results also contradict several points in previously proposed models for the emergence and evolution of the *GH* family in vertebrates (Huang et al., 2009, Wang et al., 2010), where *GH* and *SL* on the one hand, and *PRL* and *PRL2* on the other, were proposed to constitute a 1R/2R-generated quartet of genes.

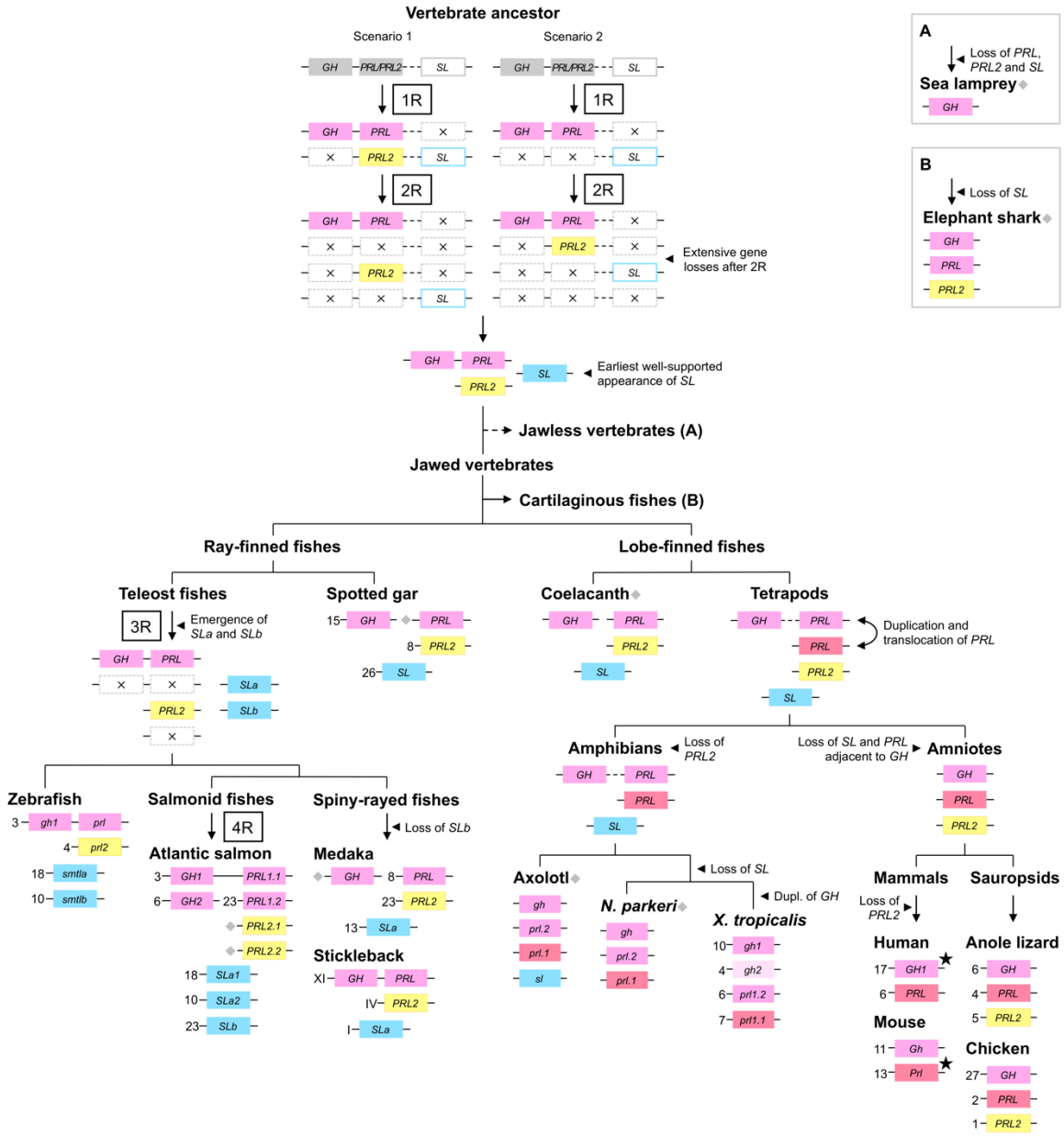


Fig. 6. Evolutionary scenario of the growth hormone family. Based on chromosomal location data and phylogenetic analysis of 105 GH, PRL, PRL2 and SL genes from 28 vertebrate species, 15 of which are shown here. Two scenarios for the emergence of PRL and PRL2 in either 1R or 2R are presented. The uncertain divergence of the sea lamprey GH gene relative to 1R/2R is indicated by a dashed line. Chromosomal/linkage group location data is not available for the sea lamprey, elephant shark, coelacanth, axolotl and Nanorana parkeri, nor for the spotted gar PRL and medaka GH genes, which is indicated by grey diamonds. Gene losses after 1R/2R and 3R are represented by crossed boxes. Asterisks indicate local expansions of the GH and PRL loci in the human and mouse genomes, respectively. Download full-size image: https://doi.org/10.6084/m9.figshare.6494879.

In teleost fishes, the *SL* gene was duplicated in the basal teleost tetraploidization (3R), giving rise to *SLa* and *SLb*, as has been previously proposed (Fukamachi and Meyer, 2007). Subsequently, the *SLb* gene was lost from the lineage leading to spiny-rayed fishes (*Acanthomorpha*), which diverged approximately 200–250 MYA (Near et al., 2012); thus, it was lost from the most diverse and species-rich group of teleost fishes. The fourth round

(4R) of tetraploidization at the base of salmonid fishes (Berthelot et al., 2014, Lien et al., 2016) duplicated all but the *SLb* gene, increasing the number of genes in this family from five to nine.

In the lobe-finned fish lineage, our results show that *SL* was lost much later than previously thought. We could identify *SL* sequences from two species of salamander, the axolotl and the red-spotted newt, which shows that *SL* was present in the tetrapod ancestor. This is to our knowledge the first identification of *SL* in any tetrapod species. Previously, the presence of *SL* in the African lungfish (May et al., 1999) was used to suggests that *SL* had been lost before the origin of tetrapods. Instead, *SL* seems to have been lost independently from an amniote ancestor and from the lineage leading to frogs (*Anura*) within amphibians (Fig. 6). *PRL2* was also lost twice independently in tetrapods: Once at the base of amphibians and once early in mammalian evolution. We could identify *PRL2* in all avian and non-avian reptile species that were investigated (Fig. 1). With respect to *PRL* in tetrapods, our results corroborate the scenario proposed by Yamaguchi et al. (2015) whereby the *PRL* gene was duplicated early on in tetrapod evolution, and one of the copies became translocated to an unrelated chromosome region, likely through the same non-homologous recombination event that produced the duplication. Amniotes have preserved only that *PRL* copy located in the unrelated chromosome region, while amphibians have preserved both copies – *prl1.1*and *prl1.2,* named "*PRL1A*" and "*PRL1B*" in Fig. 4 of Yamaguchi et al. (2015). It is interesting to note that the *GH* family experienced an expansion, and that *SL* is preserved, at this stage of tetrapod evolution, characterized by the transition to a more terrestrial life history. In teleost fishes, growth hormone and prolactin have osmoregulatory functions in the acclimation to sea water and freshwater, respectively (Breves et al., 2011, Manzon, 2002, Sakamoto and McCormick, 2006). Somatolactin also seems to have osmoregulatory-related functions, primarily related to the adaptive stress response to changes in water salinity (Khalil et al., 2012, Rhee et al., 2012) and acidity (Furukawa et al., 2010). The duplication of *PRL* and retention of *SL* during this stage of tetrapod evolution might be related to an increased adaptability to differing osmotic challenges and a more fine-tuned regulation of hydromineral homeostasis. This might also explain why only one of the *PRL* duplicates was retained and *SL* was lost from the amniote lineage.

Within amphibians, the *GH* gene was duplicated at the base of *Xenopus* (Huang and Brown, 2000, Martens et al., 1989). In the *Xenopus tropicalis* genome, the *gh2* gene is located on a chromosome block related by 2R to the location of the *gh1* gene (Fig. 4), however our phylogenetic analysis (Fig. 1) strongly indicates that this is due to a secondary translocation of the gene rather than an indication of an ancient duplication of *GH* in 2R. Additionally, all but the *gh2* gene were duplicated in the tetraploidization of the *Xenopus laevis* genome, increasing the number of *GH* family members to seven in this species.

Finally, numerous independent duplications have happened in the various vertebrate lineages. The *GH* locus has expanded at least twice independently in primates (González Alvarez et al., 2006, Papper et al., 2009, Wallis and Wallis, 2006), and duplicate *GH* genes can be found in passerine birds (Yuri et al., 2008) in addition to the duplicate *GH* genes in *Xenopus*. The *PRL* gene has duplicated independently in rodents (Alam et al., 2010, Alam et al., 2006, Wiemers et al., 2003) and ungulates (Larson et al., 2006, Schuler and Kessler, 1992, Ushizawa et al., 2007a, Ushizawa et al., 2007b), giving rise to expanded *PRL* loci that include the placental lactogen genes (*PL*) and the prolactin-related protein (*PRP*) or prolactin-like protein (*PLP*) genes. These expansions of both *GH* and *PRL* loci within eutherian mammals

have been related to the evolution of placental functions (Carter, 2012, Larson et al., 2006, Soares, 2004). The *PRL2* gene has duplicated once, in pufferfishes (*Tetraodontidae*) (Fig. 2) (Wang et al., 2010). We can also determine that the *PRL* duplicates in the Nile tilapia (Fig. 2) (Rentier-Delrue et al., 1989) likely represent a lineage-specific tandem duplication.

Our results are corroborated by 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 proposed ancestral vertebrate proto-chromosome "D". We have previously analyzed this ancestral paralogy group extensively in relation to the evolution of the visual opsin and oxytocin/vasopressin receptor gene families (Lagman et al., 2013). The *PRL* and *PRL2*-bearing chromosome regions correspond to the paralogous blocks named "D0" and "D3" by Nakatani et al. (2007). The relatively low number of neighboring gene families that we could identify in the vicinities of *SLa* and *SLb* likely result from the relatively short length of the *SL*-bearing linkage group (LG26) in spotted gar, approximately 13.7 MB, which was used to select the neighboring gene families. Nevertheless, the paralogous chromosome blocks we have identified are in agreement with reconstructions of the ancestral teleost genome before and after 3R – see Fig. 4 in Kasahara et al., 2007, Bian et al., 2016, and Fig. 3 in Nakatani and McLysaght (2017) – as well as to the ancestral vertebrate proto-chromosome "J" identified by Nakatani et al. (2007).

4.2. Genomic insights into GH, PRL/PRL2 and SL origins

Since only one family member, *GH*, has been identified in lampreys, and some concerns remain regarding the putative *GH* family members from lancelets (see Section 3.3 above), the relative timing of the early events that gave rise to the ancestral *GH*, *PRL/PRL2* and *SL* genes cannot be definitely deduced from this family's sequences alone, but the genomic regions in which the genes are located must also be considered. A summarized view of our conserved synteny analyses and the evolution of these genomic regions is included in Supplementary Fig. S37.

Our analyses of conserved synteny, together with a previously published study (Yamaguchi et al., 2015), support *GH* and *PRL* being located on the same chromosome at least before the divergence of lobe-finned fishes and ray-finned fishes. Together with phylogenetic analyses of the *GH* family, and the emergence of *PRL* and *PRL2* in 1R/2R, this suggests that an ancestral *GH*/*SL* gene and an ancestral *PRL*/*PRL2* gene arose through a local duplication at the latest between 1R and 2R, possibly earlier. There is some genomic evidence that supports this scenario. We found that the *GH* gene is located in the vicinity of *HOX* and *SCNA* genes in the human, chicken, spotted gar and zebrafish genomes (Fig. S36). The *HOX*-bearing chromosome regions have been studied extensively in the past, including by our research group, and it has been concluded that they constitute a paralogon that arose in the basal vertebrate tetraploidizations (Ocampo Daza et al., 2012a, Sundström et al., 2008, Widmark et al., 2011). In two separate reconstructions of the ancestral genome before 1R and 2R, by Nakatani et al., 2007, Putnam et al., 2008, the *HOX* (and *GH*)-bearing paralogous chromosome regions we found to coincide often on the same chromosomes as the *PRL*/*PRL2*- bearing paralogous chromosome regions in extant vertebrates. For example, on human chromosomes 3, 7, 10 and 12 – See Fig. 4 in Nakatani et al. (2007) and Fig. 3 in Putnam et al. (2008). The *HOX*-bearing chromosome regions correspond to ancestral vertebrate proto-chromosome "E" in Nakatani et al. and ancestral chordate linkage group 16

in Putnam et al. The *PRL*/*PRL2*-bearing chromosome regions correspond to "D" and 13, respectively. This pattern suggests that the ancestral chromosome regions coincided on the same chromosome in the vertebrate ancestor pre-1R/2R, thus preserving the association between the two chromosomal regions across the fourfold paralogy generated by 1R and 2R. The alternative hypothesis is that *GH* and *PRL* were located on separate chromosomes after 2R, which subsequently fused in the period of increased chromosome fusions that likely followed the tetraploidizations (Nakatani et al., 2007). However, considering the degree of association between the two ancestral regions across several chromosomes, this explanation seems less parsimonious.

The origin of *SL* is even less clear. The emergence of *PRL* and *PRL2* in 1R/2R, and our phylogeny of the *GH*-family, indicate that *GH* and *SL* likely emerged from an ancestral *GH/SL* gene at the latest between 1R and 2R, possibly earlier. However, there are also some indications that *SL* might be more closely related to *PRL* and *PRL2* genes rather than *GH* (see Section 3.2 above). The earliest time point by which we can confidently place the presence of an *SL* gene is before the divergence between jawless vertebrates and jawed vertebrates (Fig. 6). The *GH* sequences from the sea lamprey and Arctic lamprey provide a relative dating point for this timing: They cluster at the base of the *GH* branch with reliable support in our phylogenetic analysis (Fig. 1) as well as a previously published phylogeny (Yamaguchi et al., 2015). The lack of conserved synteny between either *GH* or *PRL/PRL2*genes and *SL* genes in our analyses, as well as a previously published study (Fukamachi and Meyer, 2007), at least indicates that an origin through either 1R or 2R is unlikely. We found overlap between the *SL*-bearing chromosome regions and one of the four synteny blocks in the *PRL/PRL2* paralogy group in some species – See human chromosome 11 and zebrafish chromosome 18 in in Fig. S37. This could be an indication that the ancestral *GH*, *PRL/PRL2* and *SL* genes were located in adjacent chromosome regions at one point in early vertebrate evolution or earlier. If so, the synteny of *SL* to *GH* and *PRL/PRL2* genes was broken due to differential gene losses immediately after 1R or 2R, as shown in Fig. 6. Thus, it is possible that the ancestral *GH* and *PRL*/*PRL2* genes arose through a local duplication at the latest before 2R, possibly before 1R, and that a subsequent second local duplication of either gene, within the same time window, gave rise to *SL*. However, it is also possible that the *SL*-bearing chromosome blocks were secondarily joined to one of the chromosome regions in the *PRL/PRL2*-bearing paralogy group by independent chromosome rearrangements. In most of the species we have analyzed, the *SL*-bearing chromosome blocks do not co-localize with any of the identified chromosome regions in the *PRL/PRL2*-bearing paralogy group (Fig. S37). The uncertainty regarding the origin of the *SL* gene is indicated by outlined *SL*-boxes in Fig. 6.

In summary, phylogenetic analyses of the *GH* family, conserved synteny analyses, as well as reconstructions of ancestral genomes before the basal vertebrate tetraploidizations (1R and 2R), suggest that the vertebrate ancestor had three family members: *GH*, an ancestral *PRL*/*PRL2*, and *SL*. The two former, and possibly also *SL*, were likely located on the same chromosome. Only one additional gene was added to the family through 1R and 2R, when the *PRL*/*PRL2* ancestor gave rise to *PRL* and *PRL2*. There are of course alternative hypotheses that are difficult to rule out. For instance, that the pattern of synteny between the *GH*-bearing, *PRL*/*PRL2*-bearing and *SL*-bearing paralogous chromosome regions is the result of chromosome fusions after 2R, rather than of ancestral co-localization. In any case, the three main branches of the *GH* family tree seem to have been present already before the vertebrate radiation.

4.3. Asynchrony in the evolution of GH family hormones and their corresponding receptors

Our proposed scenario for the evolution of the *GH* family stands in contrast to the evolution of the corresponding family of receptors, which acquired its basic setup of genes much later. We have recently resolved the evolution of the growth hormone and prolactin receptor genes and found that *GHR* and *PRLR* arose through a local duplication in a common ancestor of jawed vertebrates (Ocampo Daza and Larhammar, 2017). Subsequently both genes were duplicated in the basal teleost tetraploidization 3R. Thus, the different *GH* family hormones must have shared the same ancestral receptor during an extensive period early in vertebrate evolution. Together with the frequent lineage-specific duplications of the *GH* family genes, this indicates either a frequent remodeling of the hormone-receptor relationships during vertebrate evolution, or blurry functional distinctions between receptor subtypes with regard to hormone binding preferences. We have discussed some of the implications for GHR differentiation and the evolution of hormone-receptor interactions with GH and SL in the above cited article.

4.4. Other implications for functional studies of the GH family

The complex evolutionary trajectory of the growth hormone family differs strikingly from other gene families involved in the growth hormone pathway. The somatostatin receptors (SSTR), which inhibit GH release, and the insulin-like growth factor binding proteins (IGFBP) that bind to IGF1, which is released upon GH stimulation, expanded dramatically in the vertebrate whole genome doublings. We have previously shown that the SSTR family went from two ancestral genes to six in the basal vertebrate tetraploidizations (1R and 2R), and that the basal teleost tetraploidization (3R) generated three additional members (Ocampo Daza et al., 2012b). Likewise, IGFBP started with two ancestral vertebrate genes that the 1R/2R tetraploidizations brought to six, whereupon 3R resulted in 9–11 genes, depending on the species (Ocampo Daza et al., 2011). The fourth whole genome doubling in salmonids took this number to 19 (Macqueen et al., 2013). For the *GH* family on the other hand, 1R and 2R generated only a single additional gene, by giving rise to *PRL* and *PRL2*, and 3R only produced a single surviving duplicate, by giving rise to *SLa* and *SLb*. Notably, *SLb* was subsequently lost from the most diverse and species-rich group of teleost fishes. Whereas no single lineage-specific duplicate of either *SSTR* or *IGFBP* genes has been reported, the *GH* family abounds with such duplicates, sometimes resulting in quite extraordinary numbers of duplicates - like of *PRL* in rodents and ruminants, and *GH* in primates.

It has been proposed that tetraploidizations generally lead to a higher survival of duplicates compared with small-scale gene duplications because thereby the balance in whole cell-biological pathways (“modules”) may be maintained (Birchler et al., 2005, Birchler and Veitia, 2010, Freeling and Thomas, 2006, van de Peer et al., 2009). Furthermore, tetraploidizations preserve identical regulatory elements for the resulting gene duplicates. Individual gene duplicates, on the other hand, have been proposed to be preserved more frequently for the final components in cell-biological pathways, the last effector step, rather than to initiating or intermediary components. Additionally, local gene duplications, for instance by unequal crossing-over, will not necessarily transfer all of the regulatory regions, and transposon- or retrotransposition-generated duplicates will certainly not have identical regulatory elements. Therefore, these latter types of duplication likely affect the

concentration of the gene product less than the duplication of a large block or of whole chromosomes, as occurs in tetraploidizations. Except for in the unlikely case that the gene duplicate is put under the influence of a strong promotor or enhancer. Thus, duplicate genes for dosage-sensitive products will probably be lost rapidly after a tetraploidization. The gene duplicates that survive after a tetraploidization are probably less dosage-sensitive, and will therefore have a greater likelihood of neo- or subfunctionalization.

As most duplicates of *GH* family members were lost after the vertebrate tetraploidizations, it seems like they have generally not been tolerated in double gene dose with identical regulatory elements. Local, lineage-specific duplicates, on the other hand, have arisen on multiple occasions for both GH and PRL. In contrast, the SSTR and IGFBP gene families seem to have undergone no local duplications at all during vertebrate evolution. Both the SSTRs and IGFBPs have numerous roles distinct from their participation in the growth hormone hormonal axis (Duan and Xu, 2005, Tostivint et al., 2014), which has likely influenced the survival of their duplicates, in co-ordinance with extensive neo- and subfunctionalization. Another difference is that both the SSTRs (Ocampo Daza et al., 2012b) and the IGFBPs (Zhou et al., 2013) arose before the origin of the vertebrates, and thus presumably had functional roles before the *GH* family members entered the stage. None of the many members of the helical cytokines (Huising et al., 2006) seem to have unambiguous invertebrate homologs.

Not only do the GH-family genes have an unusual story of duplications, they also display an exceptionally jerky mode of sequence evolution (Wallis, 2000, Wallis, 1996, Wallis et al., 2005). This presumably reflects variable selection pressures during the course of evolution, probably both positive (Wallis, 2014) and negative. This resembles the situation for several other protein families, notably components of the immune system and the visual opsins. Such variable evolutionary rates compromise efforts by researchers to deduce phylogeny by sequences analyses alone. It is in such situations that information on chromosomal locations across several lineages becomes a powerful additional tool to resolve phylogeny, both with studies of conserved synteny and analyses of similarity between paralogous gene family members.

4.5. Conclusions

We present a comprehensive phylogenetic and chromosomal synteny analysis of the growth hormone, prolactin, prolactin 2, and somatolactin family of hormones. Previously published analyses have suggested that *GH* and *PRL* genes arose in the basal vertebrate tetraploidizations, 1R and 2R (Forsyth and Wallis, 2002, Wang et al., 2010), or that the whole quartet arose from a single ancestral GH gene after the divergence of lampreys (Kawauchi et al., 2002, Kawauchi and Sower, 2006). In contrast, and to our surprise, we found that the basic triplet of *GH* family genes – *GH*, an ancestral *PRL*/*PRL2*, and *SL* – seems to have been present already before the vertebrate radiation. We propose a scenario in which local gene duplications gave rise to this early triplet. Subsequently, the classical *PRL* gene and its paralog *PRL2* arose through the first of two rounds of basal vertebrate tetraploidization, increasing the number of family members to four. Subsequently, the third round of tetraploidization in teleost fishes duplicated the *SL* gene, giving rise to *SLa* and *SLb*– although the *SLb* gene has been lost from the most species-rich and diverse teleost lineage, the spiny-rayed fishes. In addition to these events, there have been several lineage-specific gene gains

and losses, in particular at the base of tetrapod evolution and in amphibians, possibly related to the adaptation to a more terrestrial life history. We also report for the first time the presence of *SL* in tetrapods, in two salamander species.

In summary, the GH family of hormones became a multi-member family earlier than previously thought, and the family only preserved one additional duplicate as a result of 1R and 2R, likely because of deleterious dosage effects. This is contrary to many other vertebrate peptide hormone families, which expanded largely through the vertebrate tetraploidizations.

Acknowledgments

This work was supported by grants from the Swedish Research Council and Carl Trygger's Foundation.

References

Abdullayev, I., Kirkham, M., Björklund, Å.K., Simon, A., Sandberg, R., 2013. A reference transcriptome and inferred proteome for the salamander Notophthalmus viridescens. Exp. Cell Res. 319, 1187–1197. https://doi.org/10.1016/j.yexcr.2013.02.013

Alam, S.M.K., Ain, R., Konno, T., Ho-Chen, J.K., Soares, M.J., 2006. The rat prolactin gene family locus: species-specific gene family expansion. Mamm. Genome 17, 858–77. https://doi.org/10.1007/s00335-006-0010-1

Alam, S.M.K., Konno, T., Rumi, M.A.K., Dong, Y., Weiner, C.P., Soares, M.J., 2010. Prolactin family of the guinea pig, Cavia porcellus. Endocrinology 151, 3918–28.

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. https://doi.org/10.1210/en.2010-0239

Amemiya, C.T., Alföldi, J., Lee, A.P., Fan, S., Philippe, H., MacCallum, I., Braasch, I., Manousaki, T., Schneider, I., Rohner, N., Organ, C.L., Chalopin, D., Smith, J.J., Robinson, M., Dorrington, R.A., Gerdol, M., Aken, B.L., Biscotti, M.A., Barucca, M., Baurain, D., Berlin, A.M., Blatch, G.L., Buonocore, F., Burmester, T., Campbell, M.S., Canapa, A., Cannon, J.P., Christoffels, A., De Moro, G., Edkins, A.L., Fan, L., Fausto, A.M., Feiner, N., Forconi, M., Gamieldien, J., Gnerre, S., Gnirke, A., Goldstone, J. V, Haerty, W., Hahn, M.E., Hesse, U., Hoffmann, S., Johnson, J., Karchner, S.I., Kuraku, S., Lara, M., Levin, J.Z., Litman, G.W., Mauceli, E., Miyake, T., Mueller, M.G., Nelson, D.R., Nitsche, A., Olmo, E., Ota, T., Pallavicini, A., Panji, S., Picone, B., Ponting, C.P., Prohaska, S.J., Przybylski, D., Saha, N.R., Ravi, V., Ribeiro, F.J., Sauka-Spengler, T., Scapigliati, G., Searle, S.M.J., Sharpe, T., Simakov, O., Stadler, P.F., Stegeman, J.J., Sumiyama, K., Tabbaa, D., Tafer, H., Turner-Maier, J., van Heusden, P., White, S., Williams, L., Yandell, M.D., Brinkmann, H., Volff, J.-N., Tabin, C.J., Shubin, N., Schartl, M., Jaffe, D.B., Postlethwait, J.H., Venkatesh, B., Di Palma, F., Lander, E.S., Meyer, A., Lindblad-Toh, K., 2013. The African coelacanth genome provides insights into tetrapod evolution. Nature 496, 311–316. https://doi.org/10.1038/nature12027

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. https://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. https://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. https://doi.org/10.1080/10635150600755453

Bachelot, A., Binart, N., 2007. Reproductive role of prolactin. Reproduction 133, 361–9. https://doi.org/10.1530/REP-06-0299

Baldisserotto, B., Martos-Sitcha, J.A., Menezes, C.C., Toni, C., Prati, R.L., Garcia, L. de O., Salbego, J., Mancera, J.M., Martínez-Rodríguez, G., 2014. The effects of ammonia and water hardness on the hormonal, osmoregulatory and metabolic responses of the freshwater silver catfish Rhamdia quelen. Aquat. Toxicol. 152, 341–352. https://doi.org/10.1016/j.aquatox.2014.04.023

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. https://doi.org/10.1186/1477-7827-6-42

Berthelot, C., Brunet, F.G., Chalopin, D., Juanchich, A., Bernard, M., Noël, B., Bento, P., da Silva, C., Labadie, K., Alberti, A., Aury, J.-M., Louis, A., Dehais, P., Bardou, P., Montfort, J., Klopp, C., Cabau, C., Gaspin, C., Thorgaard, G.H., Boussaha, M., Quillet, E., Guyomard, R., Galiana, D., Bobe, J., Volff, J.-N., Genêt, C., Wincker, P., Jaillon, O., Roest Crollius, H., Guiguen, Y., 2014. The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. Nat. Commun. 5, 3657. https://doi.org/10.1038/ncomms4657

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. https://doi.org/10.1038/srep24501

Birchler, J.A., Riddle, N.C., Auger, D.L., Veitia, R.A., 2005. Dosage balance in gene regulation: Biological implications. Trends Genet. 21, 219–226. https://doi.org/10.1016/j.tig.2005.02.010

Birchler, J.A., Veitia, R.A., 2010. The gene balance hypothesis: Implications for gene regulation, quantitative traits and evolution. New Phytol. 186, 54–62. https://doi.org/10.1111/j.1469-8137.2009.03087.x

Bole-Feysot, C., Goffin, V., Edery, M., Binart, N., Kelly, P.A., 1998. Prolactin (PRL) and its receptor: actions, signal transduction pathways and phenotypes observed in PRL receptor knockout mice. Endocr. Rev. 19, 225–68. https://doi.org/10.1210/edrv.19.3.0334

Bornelöv, S., Seroussi, E., Yosefi, S., Pendavis, K., Burgess, S.C., Grabherr, M., Friedman-Einat, M., Andersson, L., 2017. Correspondence on Lovell et al.: identification of chicken genes previously assumed to be evolutionarily lost. Genome Biol. 18, 112. https://doi.org/10.1186/s13059-017-1231-1

Braasch, I., Gehrke, A.R., Smith, J.J., Kawasaki, K., Manousaki, T., Pasquier, J., Amores, A., Desvignes, T., Batzel, P., Catchen, J., Berlin, A.M., Campbell, M.S., Barrell, D., Martin, K.J., Mulley, J.F., Ravi, V., Lee, A.P., Nakamura, T., Chalopin, D., Fan, S., Wcisel, D., Cañestro, C., Sydes, J., Beaudry, F.E.G., Sun, Y., Hertel, J., Beam, M.J., Fasold, M., Ishiyama, M., Johnson, J., Kehr, S., Lara, M., Letaw, J.H., Litman, G.W., Litman, R.T., Mikami, M., Ota, T., Saha, N.R., Williams, L., Stadler, P.F., Wang, H., Taylor, J.S., Fontenot, Q., Ferrara, A., Searle, S.M.J., Aken, B., Yandell, M., Schneider, I., Yoder, J.A., Volff, J.-N., Meyer, A., Amemiya, C.T., Venkatesh, B., Holland, P.W.H., Guiguen, Y., Bobe, J., Shubin, N.H., Di Palma, F., Alföldi, J., Lindblad-Toh, K., Postlethwait, J.H., 2016. The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. Nat. Genet. https://doi.org/10.1038/ng.3526

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. https://doi.org/10.1016/j.cbpa.2010.10.030

Carter, A.M., 2012. Evolution of Placental Function in Mammals: The Molecular Basis of Gas and Nutrient Transfer, Hormone Secretion, and Immune Responses. Physiol. Rev. 92, 1543–1576. https://doi.org/10.1152/physrev.00040.2011

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. https://doi.org/10.1016/j.mce.2011.02.006

Cheng, K.-W., Chan, Y.-H., Chen, Y.-D., Yu, K.-L., Chan, K.M., 1997. Sequence of a cDNA clone encoding a novel somatolactin in goldfish, Carassius auratus. Biochem. Biophys. Res. Commun. 232, 282–7. https://doi.org/10.1006/bbrc.1997.6271

Dehal, P., Boore, J.L., 2005. Two rounds of whole genome duplication in the ancestral vertebrate. PLOS Biol. 3, e314. https://doi.org/10.1371/journal.pbio.0030314

Duan, C., Xu, Q., 2005. Roles of insulin-like growth factor (IGF) binding proteins in regulating IGF actions. Gen. Comp. Endocrinol. 142, 44–52. https://doi.org/10.1016/j.ygcen.2004.12.022

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–7. https://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. https://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. https://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. https://doi.org/10.1093/nar/gkr367

Forsyth, I.A., Wallis, M., 2002. Growth hormone and prolactin--molecular and functional evolution. J. Mammary Gland Biol. Neoplasia 7, 291–312. https://doi.org/10.1023/A:1022804817104

Freeling, M., Thomas, B.C., 2006. Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. Genome Res. 16, 805–814. https://doi.org/10.1101/gr.3681406

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. https://doi.org/10.1007/s00239-007-9035-7

Furukawa, F., Watanabe, S., Kaneko, T., Uchida, K., 2010. Changes in gene expression levels of somatolactin in the pituitary and morphology of gill mitochondria-rich cells in Mozambique tilapia after transfer to acidic freshwater (pH 3.5). Gen. Comp. Endocrinol. 166, 549–555. https://doi.org/10.1016/j.ygcen.2010.01.025

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. https://doi.org/10.1007/s00438-014-0889-2

González Alvarez, R., Revol de Mendoza, A., Esquivel Escobedo, D., Corrales Félix, G., Rodríguez Sánchez, I., González, V., Dávila, G., Cao, Q., de Jong, P.J., Fu, Y.-X., Barrera Saldaña, H. a, 2006. Growth hormone locus expands and diverges after the separation of New and Old World Monkeys. Gene 380, 38–45. https://doi.org/10.1016/j.gene.2006.05.017

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. https://doi.org/10.1093/molbev/msp259

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. https://doi.org/10.1111/j.1095-8312.2007.00864.x

Holland, L.Z., Albalat, R., Azumi, K., Benito-Gutiérrez, È., Blow, M.J., Bronner-Fraser, M., Brunet, F.G., Butts, T., Candiani, S., Dishaw, L.J., Ferrier, D.E.K., Garcia Fernàndez, J., Gibson-Brown, J.J., Gissi, C., Godzik, A., Hallböök, F., Hirose, D., Hosomichi, K., Ikuta, T., Inoko, H., Kasahara, M., Kasamatsu, J., Kawashima, T., Kimura, A., Kobayashi, M., Kozmik, Z., Kubokawa, K., Laudet, V., Litman, G.W., McHardy, A.C., Meulemans, D., Nonaka, M., Olinski, R.P., Pancer, Z., Pennacchio, L.A., Pestarino, M., Rast, J.P., Rigoutsos, I., Robinson-Rechavi, M., Roch, G.J., Saiga, H., Sasakura, Y., Satake, M., Satou,

Y., Schubert, M., Sherwood, N.M., Shiina, T., Takatori, N., Tello, J.A., Vopalensky, P., Wada, S., Xu, A., Ye, Y., Yoshida, K., Yoshizaki, F., Yu, J.-K., Zhang, Q., Zmasek, C.M., de Jong, P.J., Osoegawa, K., Putnam, N.H., Rokhsar, D.S., Satoh, N., Holland, P.W.H., 2008. The amphioxus genome illuminates vertebrate origins and cephalochordate biology. Genome Res. 18, 1100–11. https://doi.org/10.1101/gr.073676.107

Hron, T., Pajer, P., Pačes, J., Bartůněk, P., Elleder, D., 2015. Hidden genes in birds. Genome Biol. 16, 164. https://doi.org/10.1186/s13059-015-0724-z

Huang, H., Brown, D.D., 2000. Overexpression of Xenopus laevis growth hormone stimulates growth of tadpoles and frogs. Proc. Natl. Acad. Sci. U. S. A. 97, 190–4. https://doi.org/10.1073/pnas.97.1.190

Huang, X., Hui, M.N.Y., Liu, Y., Yuen, D.S.H., Zhang, Y., Chan, W.Y., Lin, H.R., Cheng, S.H., Cheng, C.H.K., 2009. Discovery of a novel prolactin in non-mammalian vertebrates: evolutionary perspectives and its involvement in teleost retina development. PLoS One 4, e6163. https://doi.org/10.1371/journal.pone.0006163

Hughes, M.K., Hughes, A.L., 1993. Evolution of duplicate genes in a tetraploid animal, Xenopus laevis. Mol. Biol. Evol. 10, 1360–9. https://doi.org/10.1093/oxfordjournals.molbev.a040080

Huising, M.O., Kruiswijk, C.P., Flik, G., 2006. Phylogeny and evolution of class-I helical cytokines. J. Endocrinol. 189, 1–25. https://doi.org/10.1677/joe.1.06591

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. https://doi.org/10.1038/nature03025

Jiang, Q., He, M., Wang, X., Wong, A.O.L., 2008. Grass carp somatolactin: II. Pharmacological study on postreceptor signaling mechanisms for PACAP-induced somatolactin-alpha and -beta gene expression. Am. J. Physiol. Endocrinol. Metab. 295, E477-90. https://doi.org/10.1152/ajpendo.90386.2008

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. https://doi.org/10.1038/nature05846

Kawauchi, H., Sower, S.A., 2006. The dawn and evolution of hormones in the adenohypophysis. Gen. Comp. Endocrinol. 148, 3–14. https://doi.org/10.1016/j.ygcen.2005.10.011

Kawauchi, H., Suzuki, K., Yamazaki, T., Moriyama, S., Nozaki, M., Yamaguchi, K., Takahashi, A., Youson, J.H., Sower, S.A., 2002. Identification of growth hormone in the sea lamprey, an extant representative of a group of the most ancient vertebrates. Endocrinology 143, 4916–21. https://doi.org/10.1210/en.2002-220810

Khalil, N.A., Hashem, A.M., Ibrahim, A.A.E., Mousa, M.A., 2012. Effect of Stress During Handling, Seawater Acclimation, Confinement, and Induced Spawning on Plasma Ion Levels and Somatolactin-Expressing Cells in Mature Female Liza ramada. J. Exp. Zool. Part A Ecol. Genet. Physiol. 317, 410–424. https://doi.org/10.1002/jez.1734

Kroupová, H.K., Trubiroha, A., Wuertz, S., Frank, S.N., Sures, B., Kloas, W., 2012. Nutritional status and gene expression along the somatotropic axis in roach (Rutilus rutilus) infected with the tapeworm Ligula intestinalis. Gen. Comp. Endocrinol. 177, 270–7. https://doi.org/10.1016/j.ygcen.2012.04.007

Lagman, D., Ocampo Daza, D., Widmark, J., Abalo, X.M., Sundström, G., Larhammar, D., 2013. The vertebrate ancestral repertoire of visual opsins, transducin alpha subunits and oxytocin/vasopressin receptors was established by duplication of their shared genomic region in the two rounds of early vertebrate genome duplications. BMC Evol. Biol. 13, 238. https://doi.org/10.1186/1471-2148-13-238

Larson, J.H., Kumar, C.G., Everts, R.E., Green, C.A., Everts-van der Wind, A., Band, M.R., Lewin, H.A., 2006. Discovery of eight novel divergent homologs expressed in cattle placenta. Physiol. Genomics 25, 405–13. https://doi.org/10.1152/physiolgenomics.00307.2005

Larsson, A., 2014. AliView: a fast and lightweight alignment viewer and editor for large data sets. Bioinformatics 30, btu531-. https://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. https://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. https://doi.org/10.1210/en.2014-1377

Li, M., Jiang, C., Zhang, Y., Zhang, S., 2017. Activities of Amphioxus GH-Like Protein in Osmoregulation: Insight into Origin of Vertebrate GH Family. Int. J. Endocrinol. 2017, 1–13. https://doi.org/10.1016/j.ins.2017.03.015

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. https://doi.org/10.1038/nature17164

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. https://doi.org/10.1093/molbev/mst017

Manzon, L.A., 2002. The role of prolactin in fish osmoregulation: a review. Gen. Comp. Endocrinol. 125, 291–310. https://doi.org/10.1006/gcen.2001.7746

Martens, G.J., Groenen, P.J., Braks, A.A., Bussemakers, M.J.G., 1989. Expression of two growth hormone genes in the Xenopus pituitary gland. Nucleic Acids Res. 17, 3974. https://doi.org/10.1093/nar/17.10.3974

May, D., Alrubaian, J., Patel, S., Dores, R.M., Rand-Weaver, M., 1999. Studies on the GH/SL gene family: cloning of African lungfish (Protopterus annectens) growth hormone and somatolactin and toad (Bufo marinus) growth hormone. Gen. Comp. Endocrinol. 113, 121–35. https://doi.org/10.1006/gcen.1998.7185

May, D., Todd, C.M., Rand-Weaver, M., 1997. cDNA cloning of eel (Anguilla anguilla) somatolactin. Gene 188, 63–7. https://doi.org/10.1016/S0378-1119(96)00777-9

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. https://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. https://doi.org/10.1073/pnas.1315760110

Meyer, A., van de Peer, Y., 2005. From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). BioEssays 27, 937–45. https://doi.org/10.1002/bies.20293

Minh, B.Q., Nguyen, M.A.T., Von Haeseler, A., 2013. Ultrafast approximation for phylogenetic bootstrap. Mol. Biol. Evol. 30, 1188–1195. https://doi.org/10.1093/molbev/mst024

Moriyama, S., Oda, M., Takahashi, A., Sower, S.A., Kawauchi, H., 2006. Genomic structure of the sea lamprey growth hormone-encoding gene. Gen. Comp. Endocrinol. 148, 33–40. https://doi.org/10.1016/j.ygcen.2005.09.023

Nakatani, Y., McLysaght, A., 2017. Genomes as documents of evolutionary history: a probabilistic macrosynteny model for the reconstruction of ancestral genomes. Bioinformatics 33, i369–i378. https://doi.org/10.1093/bioinformatics/btx259

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. https://doi.org/10.1101/gr.6316407

Near, T.J., Dornburg, A., Eytan, R.I., Keck, B.P., Smith, W.L., Kuhn, K.L., Moore, J.A., Price, S.A., Burbrink, F.T., Friedman, M., Wainwright, P.C., 2013. Phylogeny and tempo of diversification in the superradiation of spiny-rayed fishes. Proc. Natl. Acad. Sci. U. S. A. 110, 12738–12743. https://doi.org/10.1073/pnas.1304661110

Near, T.J., Eytan, R.I., Dornburg, A., Kuhn, K.L., Moore, J.A., Davis, M.P., Wainwright, P.C., Friedman, M., Smith, W.L., 2012. Resolution of ray-finned fish phylogeny and timing of diversification. Proc. Natl. Acad. Sci. U. S. A. 109, 13698–703. https://doi.org/10.1073/pnas.1206625109

Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274. https://doi.org/10.1093/molbev/msu300

Noso, T., Nicoll, C.S., Polenov, A.L., Kawauchi, H., 1993. The primary structure of sturgeon prolactin: phylogenetic implication. Gen. Comp. Endocrinol. 91, 90–5. https://www.sciencedirect.com/science/article/pii/S0016648083711081

Ocampo Daza, D., Bergqvist, C.A., Larhammar, D., 2012a. Phylogenetic analyses of the insulin-like growth factor binding protein (IGFBP) family [WWW Document]. figshare. URL http://dx.doi.org/10.6084/m9.figshare.103144 (accessed 12.6.12).

Ocampo Daza, D., Larhammar, D., 2017. Evolution of the receptors for growth hormone, prolactin, erythropoietin and thrombopoietin in relation to the vertebrate tetraploidizations. Gen. Comp. Endocrinol. https://doi.org/10.1016/j.ygcen.2017.06.021

Ocampo Daza, D., Sundström, G., Bergqvist, C.A., Duan, C., Larhammar, D., 2011. Evolution of the insulin-like growth factor binding protein (IGFBP) family. Endocrinology 152, 2278–89. https://doi.org/10.1210/en.2011-0047

Ocampo Daza, D., Sundström, G., Bergqvist, C.A., Larhammar, D., 2012b. The evolution of vertebrate somatostatin receptors and their gene regions involves extensive chromosomal rearrangements. BMC Evol. Biol. 12, 231. https://doi.org/10.1186/1471-2148-12-231

Ocampo Daza, D., Sundström, G., Larsson, T.A., Larhammar, D., 2009. Evolution of the growth hormone-prolactin-somatolactin system in relation to vertebrate tetraploidizations. Ann. N. Y. Acad. Sci. 1163, 491–3. https://doi.org/10.1111/j.1749-6632.2008.03671.x

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. https://doi.org/10.1073/pnas.87.11.4330

Papper, Z., Jameson, N.M., Romero, R., Weckle, A.L., Mittal, P., Benirschke, K., Santolaya-Forgas, J., Uddin, M., Haig, D., Goodman, M., Wildman, D.E., 2009. Ancient origin of placental expression in the growth hormone genes of anthropoid primates. Proc. Natl. Acad. Sci. U. S. A. 106, 17083–8. https://doi.org/10.1073/pnas.0908377106

Pérez-Sánchez, J., Calduch-Giner, J.A., Mingarro, M., Vega-Rubín de Celis, S., Gómez-Requeni, P., Saera-Vila, A., Astola, A., Valdivia, M.M., 2002. Overview of Fish Growth Hormone Family. New Insights in Genomic Organization and Heterogeneity of Growth Hormone Receptors. Fish Physiol. Biochem. 27, 243–258. https://doi.org/10.1023/B:FISH.0000032729.72746.c8

Power, D.M., 2005. Developmental ontogeny of prolactin and its receptor in fish. Gen. Comp. Endocrinol. 142, 25–33. https://doi.org/10.1016/j.ygcen.2004.10.003

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. https://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 Fernàndez, 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. https://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. https://doi.org/10.1021/bi00220a010

Rentier-Delrue, F., Swennen, D., Prunet, P., Lion, M., Martial, J.A., 1989. Tilapia prolactin: molecular cloning of two cDNAs and expression in Escherichia coli. DNA 8, 261–70. https://doi.org/10.1089/dna.1.1989.8.261

Rhee, J.S., Kim, B.M., Seo, J.S., Kim, I.C., Lee, Y.M., Lee, J.S., 2012. Cloning of growth hormone, somatolactin, and their receptor mRNAs, their expression in organs, during development, and on salinity stress in the hermaphroditic fish, Kryptolebias marmoratus. Comp. Biochem. Physiol. - A Mol. Integr. Physiol. 161, 436–442. https://doi.org/10.1016/j.cbpa.2012.01.004

Ryynänen, H.J., Primmer, C.R., 2006. Varying signals of the effects of natural selection during teleost growth hormone gene evolution. Genome 49, 42–53. https://doi.org/10.1139/g05-079

Sakamoto, T., McCormick, S.D., 2006. Prolactin and growth hormone in fish osmoregulation. Gen. Comp. Endocrinol. 147, 24–30. https://doi.org/10.1016/j.ygcen.2005.10.008

Sayers, E.W., Barrett, T., Benson, D.A., Bolton, E., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., Dicuccio, M., Federhen, S., Feolo, M., Geer, L.Y., Helmberg, W., Kapustin, Y., Landsman, D., Lipman, D.J., Lu, Z., Madden, T.L., Madej, T., Maglott, D.R., Marchler-Bauer, A., Miller, V., Mizrachi, I., Ostell, J., Panchenko, A., Pruitt, K.D., Schuler, G.D., Sequeira, E., Sherry, S.T., Shumway, M., Sirotkin, K., Slotta, D., Souvorov, A., Starchenko, G., Tatusova, T.A., Wagner, L., Wang, Y., John Wilbur, W., Yaschenko, E., Ye, J., 2010. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 38, D5-16. https://doi.org/10.1093/nar/gkp967

Schuler, L.A., Kessler, M.A., 1992. Bovine placental prolactin-related hormones. Trends Endocrinol. Metab. 3, 334–338. https://doi.org/10.1016/1043-2760(92)90112-E

Soares, M.J., 2004. The prolactin and growth hormone families: pregnancy-specific hormones/cytokines at the maternal-fetal interface. Reprod. Biol. Endocrinol. 2, 51. https://doi.org/10.1186/1477-7827-2-51

Sundström, G., Larsson, T.A., Larhammar, D., 2008. Phylogenetic and chromosomal analyses of multiple gene families syntenic with vertebrate Hox clusters. BMC Evol. Biol. 8, 254. https://doi.org/10.1186/1471-2148-8-254

Tine, M., Kuhl, H., Gagnaire, P.-A., Louro, B., Desmarais, E., Martins, R.S.T., Hecht, J., Knaust, F., Belkhir, K., Klages, S., Dieterich, R., Stueber, K., Piferrer, F., Guinand, B., Bierne, N., Volckaert, F.A.M., Bargelloni, L., Power, D.M., Bonhomme, F., Canario, A.V.M., Reinhardt, R., 2014. European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. Nat. Commun. 5, 5770. https://doi.org/10.1038/ncomms6770

Tostivint, H., Ocampo Daza, D., Bergqvist, C. a, Quan, F.B., Bougerol, M., Lihrmann, I., Larhammar, D., 2014. Molecular evolution of GPCRs: Somatostatin/urotensin II receptors. J. Mol. Endocrinol. 52, T61–T86. https://doi.org/10.1530/JME-13-0274

Ushizawa, K., Takahashi, T., Hosoe, M., Kizaki, K., Abe, Y., Sasada, H., Sato, E., Hashizume, K., 2007a. Gene expression profiles of novel caprine placental prolactin-related proteins similar to bovine placental prolactin-related proteins. BMC Dev. Biol. 7, 16. https://doi.org/10.1186/1471-213X-7-16

Ushizawa, K., Takahashi, T., Hosoe, M., Ohkoshi, K., Hashizume, K., 2007b. Expression and characterization of novel ovine orthologs of bovine placental prolactin-related proteins. BMC Mol. Biol. 8, 95. https://doi.org/10.1186/1471-2199-8-95

van de Peer, Y., Maere, S., Meyer, A., 2009. The evolutionary significance of ancient genome duplications. Nat. Rev. Genet. 10, 725–32. https://doi.org/10.1038/nrg2600

Venkatesh, B., Lee, A.P., Ravi, V., Maurya, A.K., Lian, M.M., Swann, J.B., Ohta, Y., Flajnik, M.F., Sutoh, Y., Kasahara, M., Hoon, S., Gangu, V., Roy, S.W., Irimia, M., Korzh, V., Kondrychyn, I., Lim, Z.W., Tay, B.-H., Tohari, S., Kong, K.W., Ho, S., Lorente-Galdos, B., Quilez, J., Marques-Bonet, T., Raney, B.J., Ingham, P.W., Tay, A., Hillier, L.W., Minx, P., Boehm, T., Wilson, R.K., Brenner, S., Warren, W.C., 2014. Elephant shark genome provides unique insights into gnathostome evolution. Nature 505, 174–179. https://doi.org/10.1038/nature12826

von Schalburg, K.R., Yazawa, R., de Boer, J., Lubieniecki, K.P., Goh, B., Straub, C. a, Beetz-Sargent, M.R., Robb, A., Davidson, W.S., Devlin, R.H., Koop, B.F., 2008. Isolation, characterization and comparison of Atlantic and Chinook salmon growth hormone 1 and 2. BMC Genomics 9, 522. https://doi.org/10.1186/1471-2164-9-522

Voss, S.R., Kump, D.K., Putta, S., Pauly, N., Reynolds, A., Henry, R.J., Basa, S., Walker, J.A., Smith, J.J., 2011. Origin of amphibian and avian chromosomes by fission, fusion, and retention of ancestral chromosomes. Genome Res. 21, 1306–12. https://doi.org/10.1101/gr.116491.110

Wallis, M., 1996. The molecular evolution of vertebrate growth hormones: a pattern of near-stasis interrupted by sustained bursts of rapid change. J. Mol. Evol. 43, 93–100. https://doi.org/10.1007/BF02337353

Wallis, M., 2000. Episodic evolution of protein hormones: molecular evolution of pituitary prolactin. J. Mol. Evol. 50, 465–73. https://doi.org/10.1007/s002390010049

Wallis, M., 2001. Episodic evolution of protein hormones in mammals. J. Mol. Evol. 53, 10–8. https://doi.org/10.1007/s002390010187

Wallis, M., 2014. Molecular evolution of growth hormone. Biochem. (Lond). 36, 4–8. http://www.biochemist.org/bio/refer.htm?VOL=36&ISSUE=1&PAGE=4

Wallis, O.C., Mac-Kwashie, A.O., Makri, G., Wallis, M., 2005. Molecular evolution of prolactin in primates. J. Mol. Evol. 60, 606–14. https://doi.org/10.1007/s00239-004-0239-9

Wallis, O.C., Wallis, M., 2006. Evolution of growth hormone in primates: the GH gene clusters of the New World monkeys marmoset (Callithrix jacchus) and white-fronted capuchin (Cebus albifrons). J. Mol. Evol. 63, 591–601. https://doi.org/10.1007/s00239-006-0039-5

Wang, J.-T., Li, J.-T., Zhang, X.-F., Sun, X.-W., 2012. Transcriptome analysis reveals the time of the fourth round of genome duplication in common carp (Cyprinus carpio). BMC Genomics 13, 96. https://doi.org/10.1186/1471-2164-13-96

Wang, Y., Li, J., Yan Kwok, A.H., Ge, W., Leung, F.C., 2010. A novel prolactin-like protein (PRL-L) gene in chickens and zebrafish: cloning and characterization of its tissue expression. Gen. Comp. Endocrinol. 166, 200–10. https://doi.org/10.1016/j.ygcen.2009.10.007

Werner, H.B., 2013. Do we have to reconsider the evolutionary emergence of myelin? Front. Cell. Neurosci. 7, 415–21, 421–2. https://doi.org/10.1145/2445196.2445322

Whittington, C.M., Wilson, A.B., 2013. The role of prolactin in fish reproduction. Gen. Comp. Endocrinol. 191, 123–36. https://doi.org/10.1016/j.ygcen.2013.05.027

Widmark, J., Sundström, G., Ocampo Daza, D., Larhammar, D., 2011. Differential evolution of voltage-gated sodium channels in tetrapods and teleost fishes. Mol. Biol. Evol. 28, 859–71. https://doi.org/10.1093/molbev/msq257

Wiemers, D.O., Shao, L.-J., Ain, R., Dai, G., Soares, M.J., 2003. The mouse prolactin gene family locus. Endocrinology 144, 313–25. https://doi.org/10.1210/en.2002-220724

Xie, F., London, S.E., Southey, B.R., Annangudi, S.P., Amare, A., Rodriguez-Zas, S.L., Clayton, D.F., Sweedler, J. V, 2010. The zebra finch neuropeptidome: prediction, detection and expression. BMC Biol. 8:28. https://doi.org/10.1186/1741-7007-8-28

Xu, P., Zhang, X., Wang, X., Li, J., Liu, G., Kuang, Y., Xu, J., Zheng, X., Ren, L., Wang, G., Zhang, Y., Huo, L., Zhao, Z., Cao, D., Lu, C., Li, C., Zhou, Y., Liu, Z., Fan, Z., Shan, G., Li, X., Wu, S., Song, L., Hou, G., Jiang, Y., Jeney, Z., Yu, D., Wang, L., Shao, C., Song, L., Sun, J., Ji, P., Wang, J., Li, Q., Xu, L., Sun, F., Feng, J., Wang, C., Wang, S., Wang, B., Li, Y., Zhu, Y., Xue, W., Zhao, L., Wang, J., Gu, Y., Lv, W., Wu, K., Xiao, J., Wu, J., Zhang, Z., Yu, J., Sun, X., 2014. Genome sequence and genetic diversity of the common carp, Cyprinus carpio. Nat. Genet. 46(11): 1212-9. https://doi.org/10.1038/ng.3098

Yamaguchi, Y., Takagi, W., Kuraku, S., Moriyama, S., Bell, J.D., Seale, A.P., Lerner, D.T., Grau, E.G., Hyodo, S., 2015. Discovery of conventional prolactin from the holocephalan elephant fish, Callorhinchus milii. Gen. Comp. Endocrinol. 224, 216–227. https://doi.org/10.1016/j.ygcen.2015.08.020

Yang, B.-Y., Chen, T.T., 2003. Identification of a new growth hormone family protein, somatolactin-like protein, in the rainbow trout (Oncorhyncus mykiss) pituitary gland. Endocrinology 144, 850–7. https://doi.org/10.1210/en.2002-221005

Yang, J., Chen, X., Bai, J., Fang, D., Qiu, Y., Jiang, W., Yuan, H., Bian, C., Lu, J., He, S., Pan, X., Zhang, Y., Wang, X., You, X., Wang, Y., Sun, Y., Mao, D., Liu, Y., Fan, G., Zhang, H., Chen, X., Zhang, X., Zheng, L., Wang, J., Cheng, L., Chen, J., Ruan, Z., Li, J., Yu, H., Peng, C., Ma, X., Xu, J., He, Y., Xu, Z., Xu, P., Wang, J., Yang, H., Wang, J., Whitten, T., Xu, X., Shi, Q., 2016. The Sinocyclocheilus cavefish genome provides insights into cave adaptation. BMC Biol. 14, 1. https://doi.org/10.1186/s12915-015-0223-4

Yuri, T., Kimball, R.T., Braun, E.L., Braun, M.J., 2008. Duplication of accelerated evolution and growth hormone gene in passerine birds. Mol. Biol. Evol. 25, 352–61. https://doi.org/10.1093/molbev/msm260

Zhou, J., Xiang, J., Zhang, S., Duan, C., 2013. Structural and functional analysis of the amphioxus IGFBP gene uncovers ancient origin of IGF-independent functions. Endocrinology 1–11. https://doi.org/10.1155/2013/105930

Zhu, Y., Stiller, J.W., Shaner, M.P., Baldini, A., Scemama, J.-L., Capehart, A.A., 2004. Cloning of somatolactin alpha and beta cDNAs in zebrafish and phylogenetic analysis of two distinct somatolactin subtypes in fish. J. Endocrinol. 182, 509–18. https://doi.org/10.1677/joe.0.1820509

Appendix A. Supplementary data

The following supplementary data has been shared to https://doi.org/10.6084/m9.figshare.6494879.

Supplementary data 1. Location data, sequence identifiers and prediction/annotation notes for all identified *GH*, *PRL*, *PRL2* and *SL* sequences. The data file also includes genome assembly as well as sequence quality and annotation information.

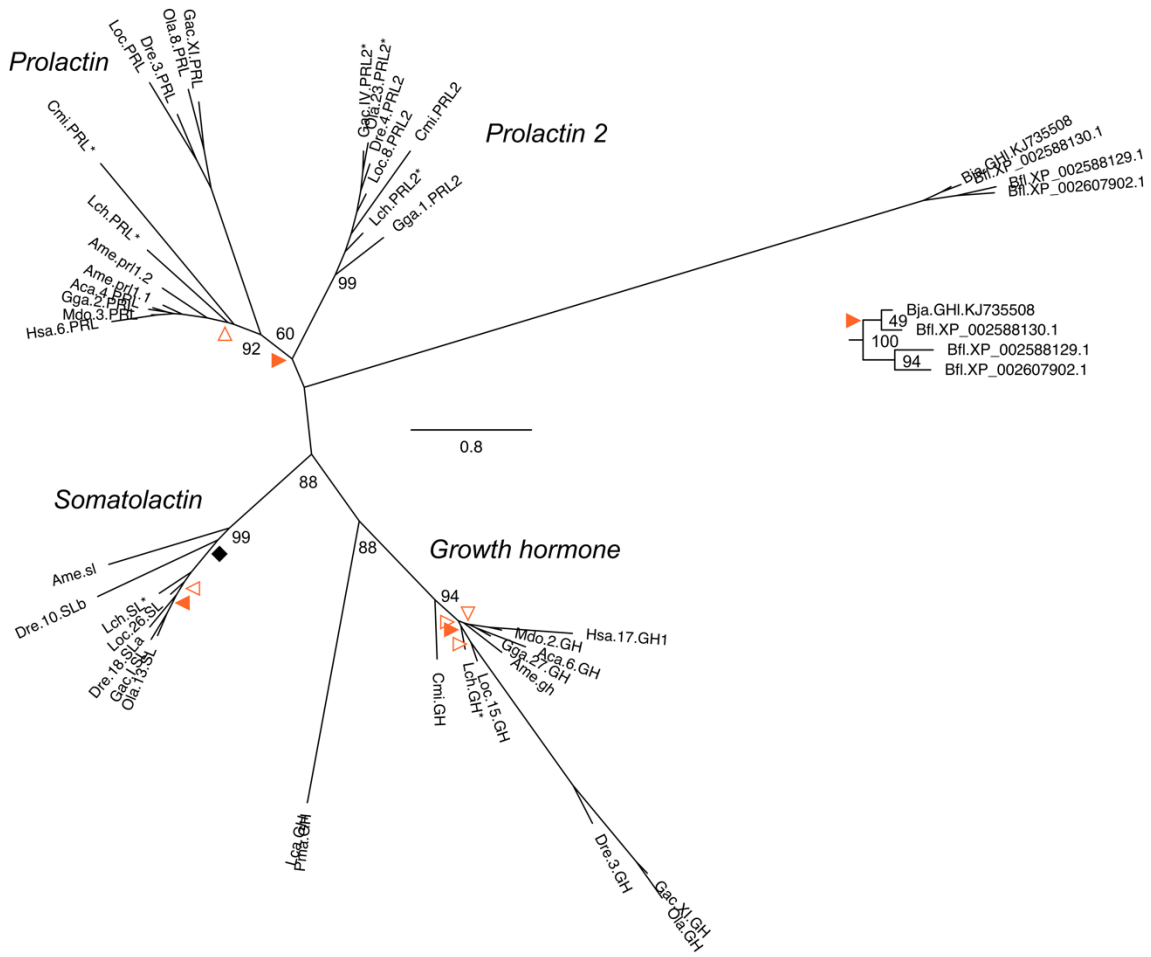
Supplementary data 2. Amino acid sequence alignment of *GH*, *PRL*, *PRL2* and *SL* sequences used to construct *GH* family phylogeny. Species abbreviations are explained in Supplementary data 1.

Supplementary data 3. Amino acid sequence alignment of *SLa* and *SLb* sequences used to construct the *SL* phylogeny. Species abbreviations are explained in Supplementary data 1.

Supplementary data 4. Up-to-date 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. New suggested gene symbols in these species are indicated by yellow highlights. For gene predictions where no gene symbol could be assigned, the sequences are named for their assigned chromosomes, linkage groups or scaffolds only. Duplicate genes generated in 3R are assigned the lower-case letters “a” and “b”, following zebrafish guidelines.

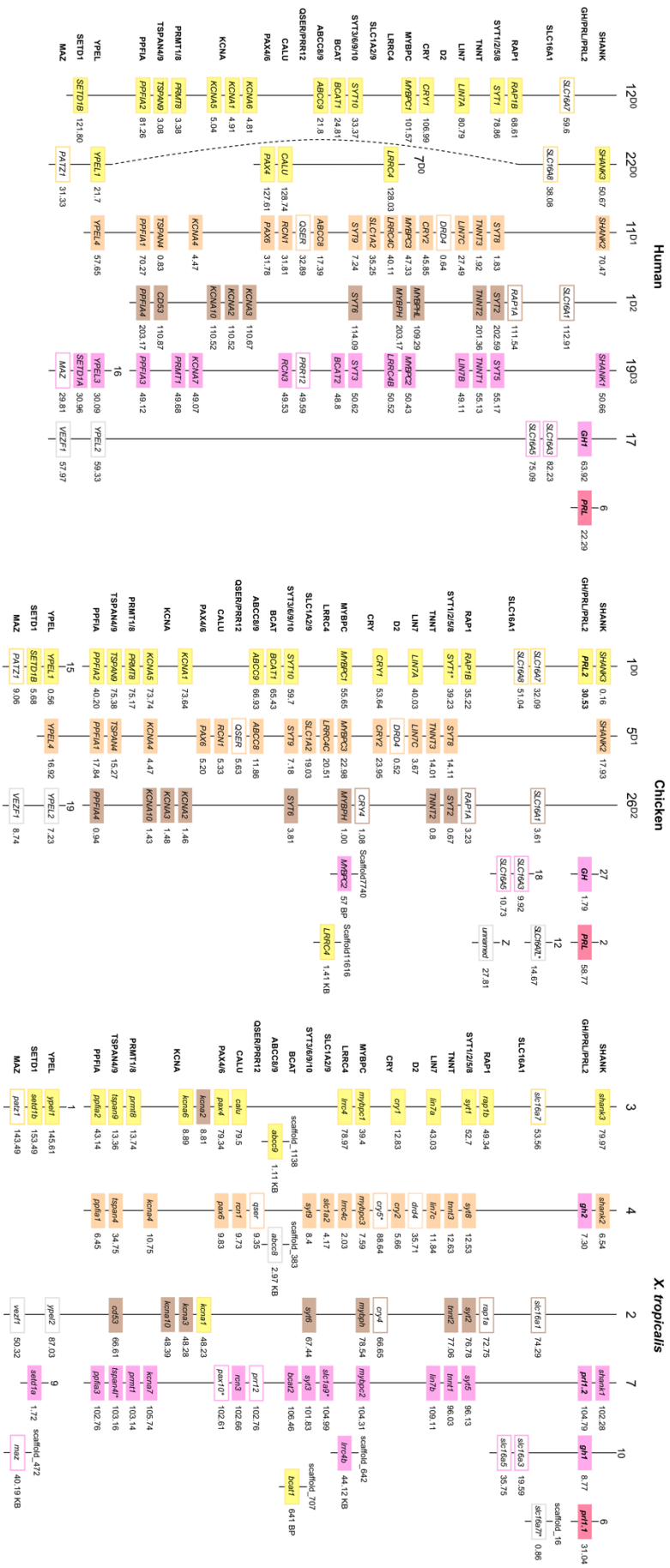
Supplementary data 5. Conserved synteny results for the GH, PRL, PRL2 and SL-bearing chromosome blocks in the chicken, spotted gar and zebrafish genomes. Gene lists from each chromosome block were downloaded from Ensembl BioMart, color-coded according to chromosome, and sorted by Ensembl protein family 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 synteny. The complete gene list includes all identified Ensembl protein families with members on at least two of the chromosome blocks. From this list, different selections were made to investigate different conserved synteny relationships. “Subset 1” includes families with members on at least two of the PRL and PRL2-bearing chromosome blocks. “Subset 2” includes families with members on at least two of the SL/SLa/SLb-bearing chromosome blocks. “Subset 3” includes families with members on at least two of the GH, PRL/PRL2 and SL/Sla/Slb-bearing chromosome blocks.

Supplementary data 6. Conserved synteny results for GH and PRL synteny analysis. A gene list for the gh1 and prl-bearing chromosome block in zebrafish was compared with gene lists for the GH and PRL-bearing blocks in spotted gar and chicken. Neighboring genes represented on at least three of these chromosome blocks were chosen for the analysis. The data file also includes up-to-date location data and sequence identifiers for all neighboring genes identified in the analysis, including for Xenopus tropicalis and human.



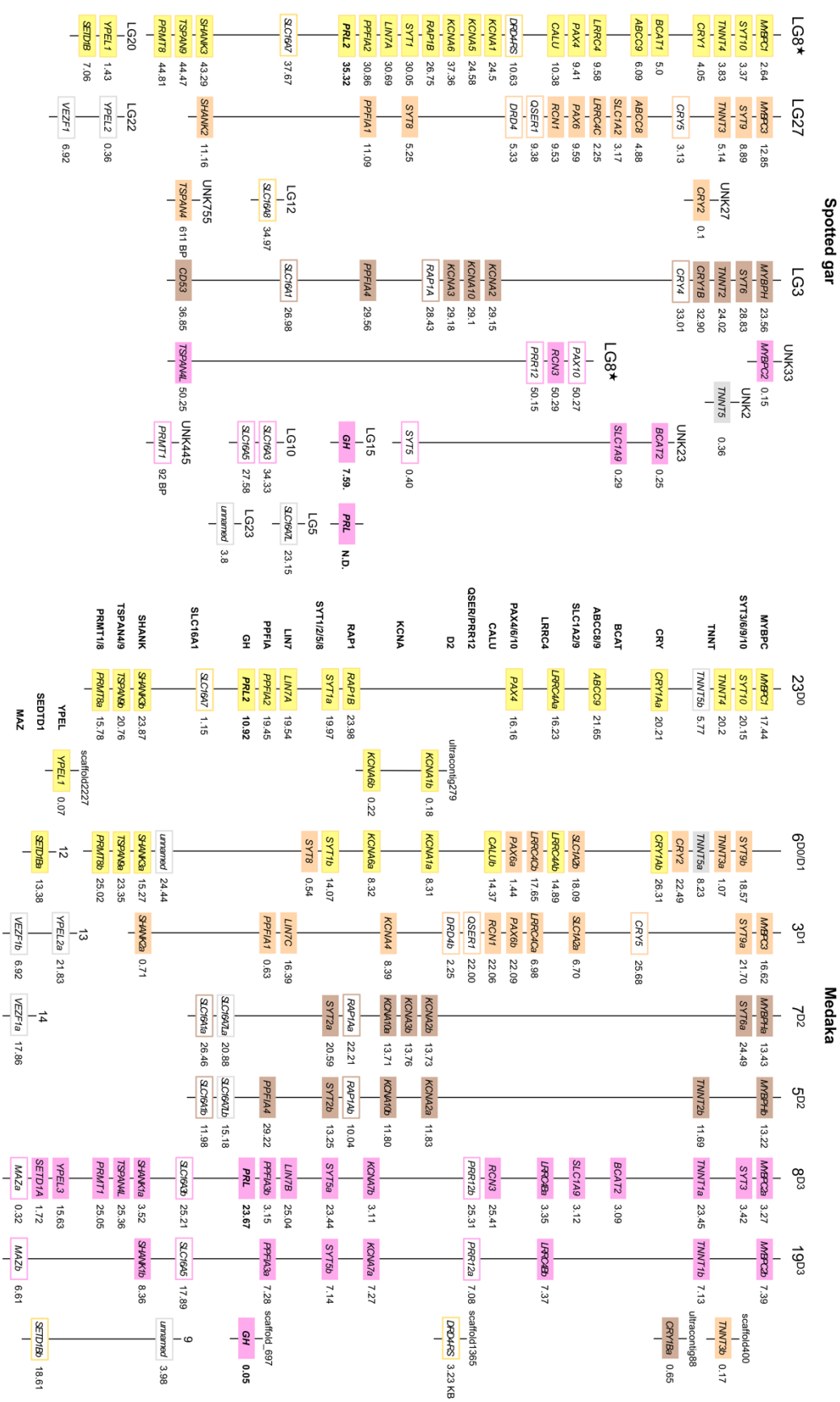
Supplementary figure S1. Phylogenetic analysis of putative *GH* sequences from *Branchiostoma belcheri* and *Branchiostoma floridae*. The maximum likelihood phylogeny is supported by non-parametric Utra-Fast Bootstrap (UFBoot) and approximate Likelihood Ratio Test (aLRT) analyses. Only UFBoot values values are shown. Filled arrowheads indicate nodes with low support in both UFBoot (≤70%) and aLRT (≤50%) tests, outlined arrowheads indicate nodes with low UFBoot support only, and diamonds indicate nodes with low aLRT support only. For clarity, some support values for shallow nodes are not shown. The tree is displayed as a radial unrooted phylogram. Abbreviations: Aca (Carolina anole lizard), Ame (axolotl), Bbe (*Branchiostoma belcheri*), Bfl (*Branchiostoma floridae*), Cmi (elephant shark), Dre (zebrafish), Gac (three-spined stickleback), Gga (chicken), Has (human), Lca (Arctic lamprey), Lch (coelacanth), Loc (spotted gar), Mdo (grey short-tailed opossum), Ola (medaka), Pma (sea lamprey). Download full-size image: https://doi.org/10.6084/m9.figshare.6494879.

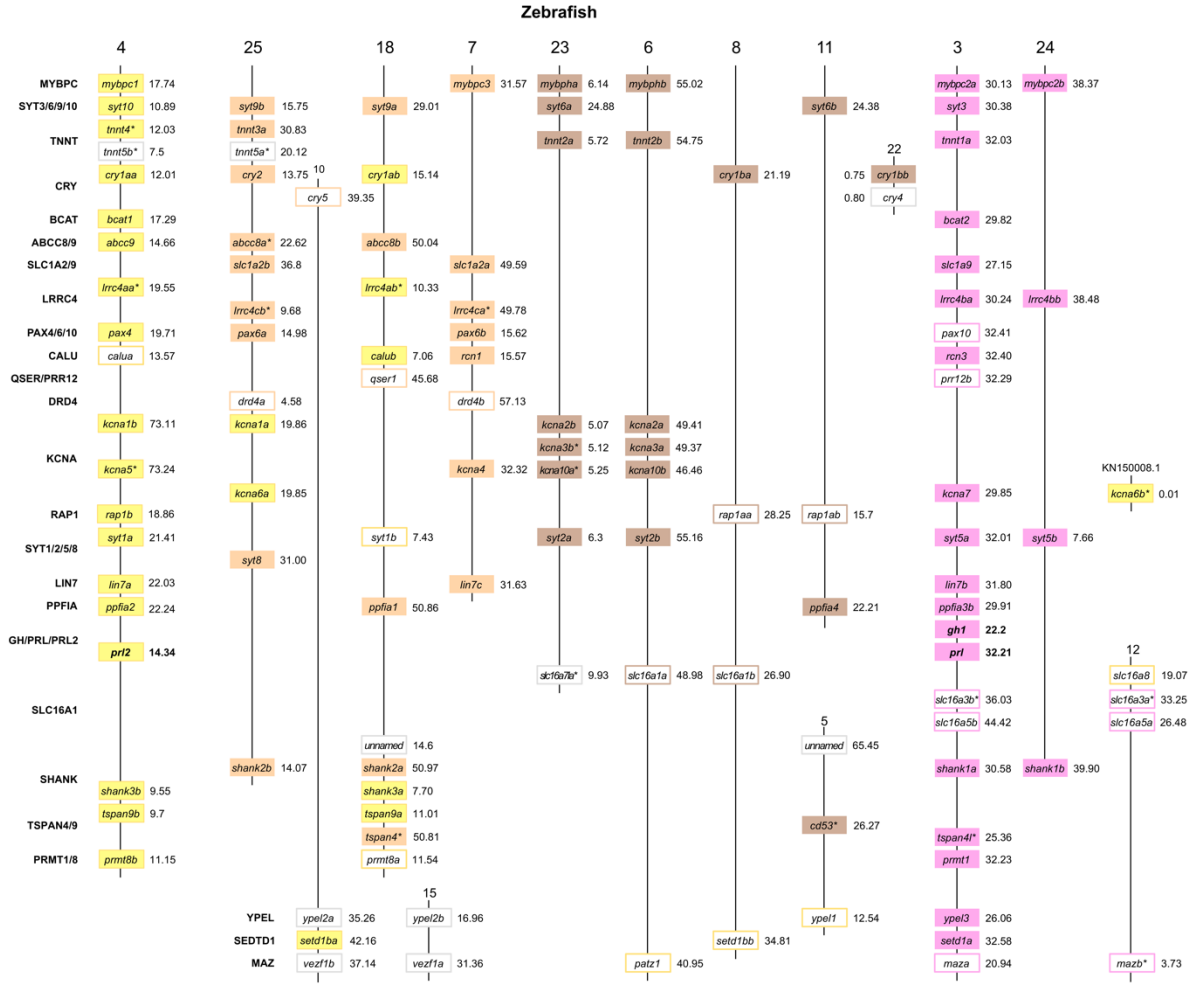
Supplementary Figures S2 – S32. See separate PDF document shared to https://doi.org/10.6084/m9.figshare.6494879.


Human
Chicken
X. tropicalis

Supplementary figure S33. Conserved synteny blocks in the vicinity of *GH*, *PRL*, and *PRL2* genes in the human, chicken and *Xenopus tropicalis* genomes. Locations are given in MB unless indicated otherwise. Gene order follows chicken chromosome 1. New suggested gene symbols are indicated by asterisks. Colors are applied following human chromosomes in order to show conserved synteny as well as gene orthology between species. *GH* and *GH*-neighboring genes are colored magenta to indicate the ancestral co-localization of *GH* and *PRL*. D0 – D3 designations next to the human, chicken and medaka (Fig. S3) chromosome numbers correspond to vertebrate paralogous blocks identified by Nakatani et al., 2007 to have arisen through 1R/2R. White bordered boxes indicate genes with unclear divergence times relative to 1R/2R; due to the lack of outgroups for relative dating (D2, MAZ, QSER/PRR12 and SLC16A1 families), low support in the phylogenetic analyses (*PAX10, RAP1A, SLC16A7L, MAZ* and *YPEL2*), or earlier divergence times (*CRY4* and *CRY5*). Grey color indicates uncertain synteny relationships to the identified paralogous chromosome regions (*TNNT5* and *YPEL2*). For the D2 and SLC1A families, only the genes located within this paralogon are shown – See Figs. S24 and S31 for the full families. Download full-size image: https://doi.org/10.6084/m9.figshare.6494879.

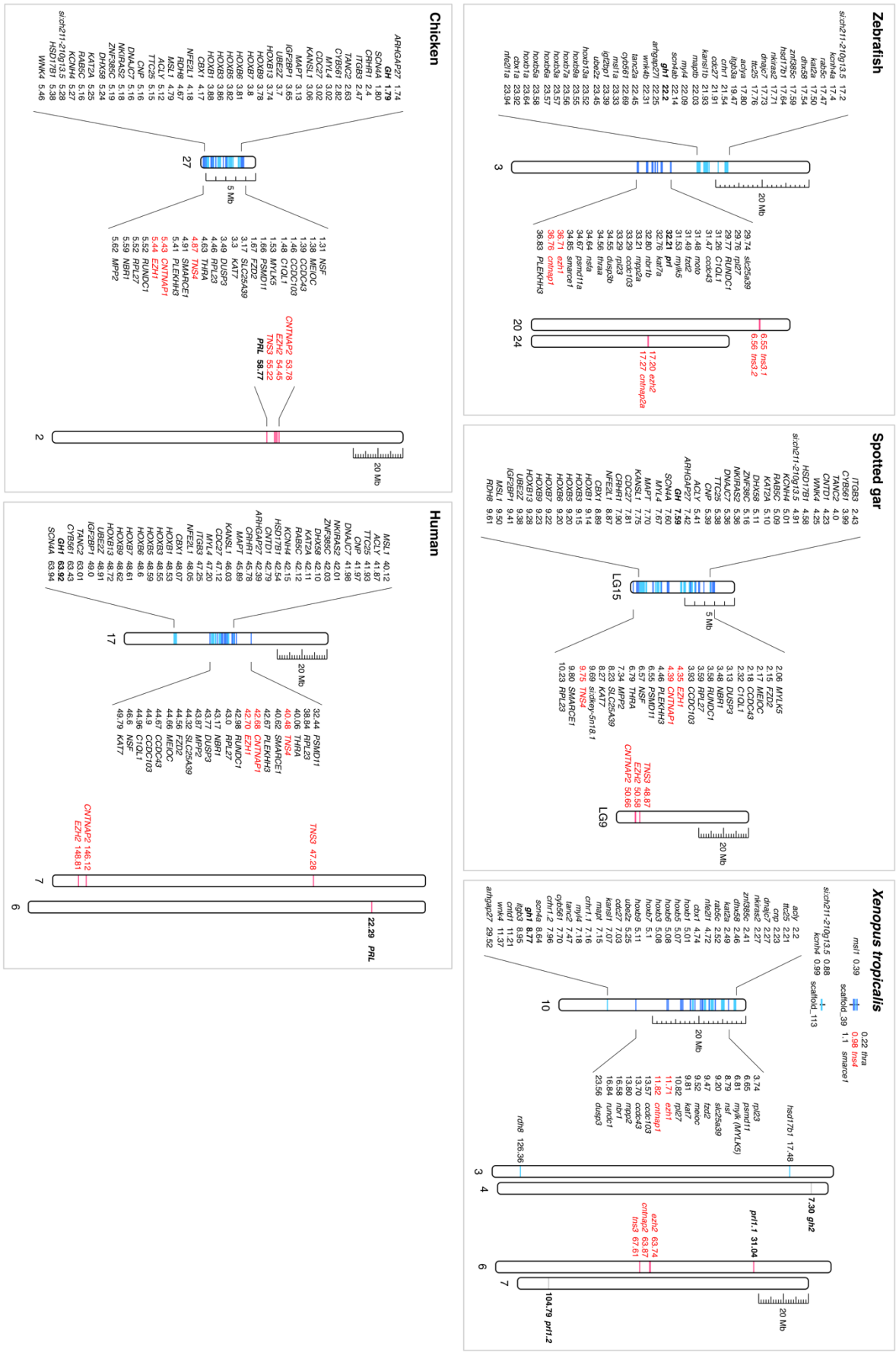
Supplementary figure S34 (next page). Conserved synteny blocks in the vicinity of *GH*, *PRL*, and *PRL2* genes in the spotted gar and medaka genomes. See Fig. S2 caption for details. Gene order follows spotted gar linkage group 8. Note that spotted gar linkage group 8 is represented twice (black star). The chromosomal location of the spotted gar *PRL* gene is currently unknown. The spotted gar *SYT5* gene could not be used in the phylogeny of this family. White bordered boxes in zebrafish and medaka can indicate unclear divergence times relative to 3R as well as 1R/2R. Download full-size image: https://doi.org/10.6084/m9.figshare.6494879.

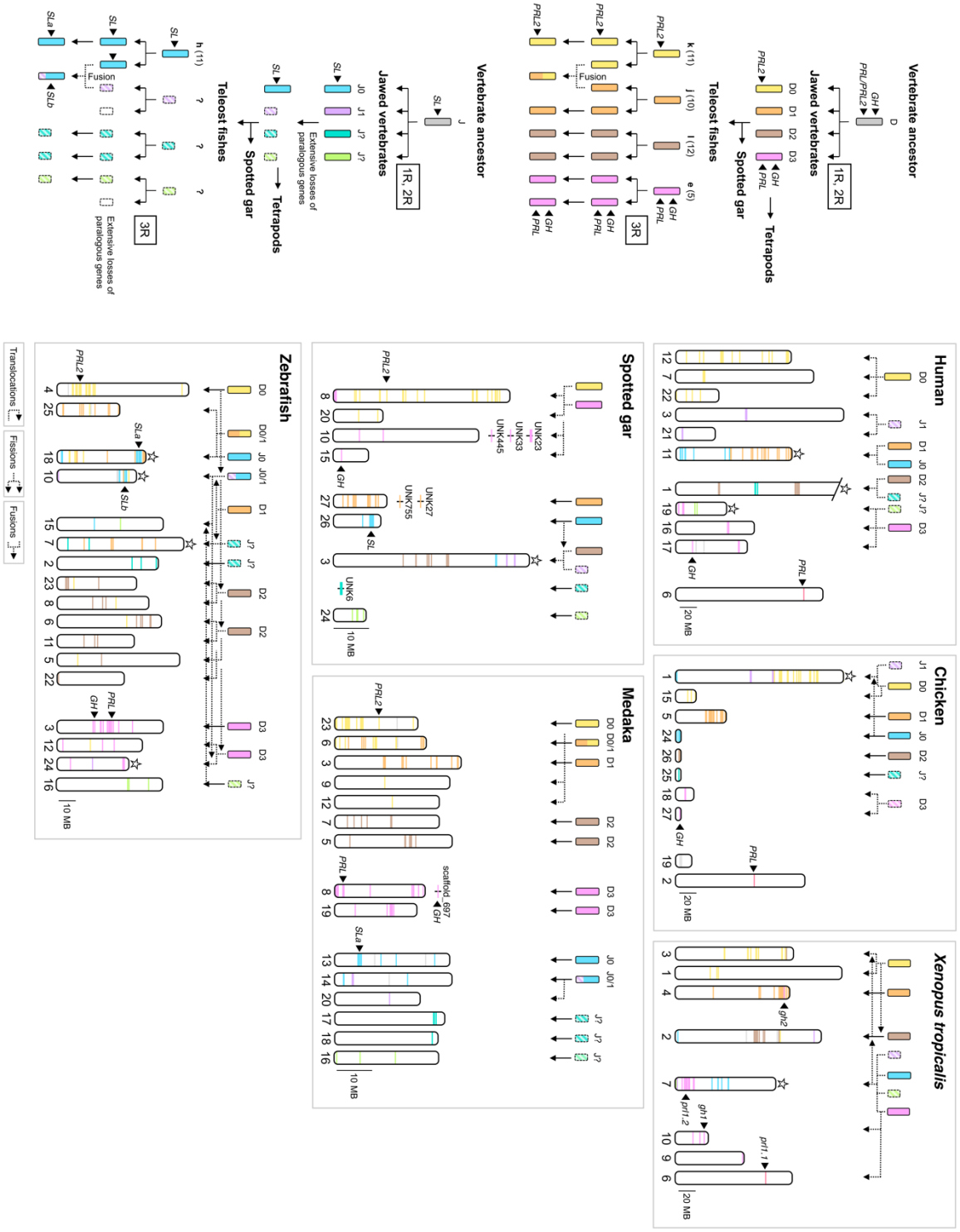

Spotted gar
Medaka
MYBPC
SYT3/6/9/10
TNNT
CRY
BCAT
ABCC8/9
SLC1A2/9
LRRC4
PAX4/6/10
CALU
QSER/PRR12
D2
KCNA
RAP1
SYT1/2/5/8
LIN7
PPFIA
GH
SHANK
TSPAN4/9
PRMT1/8
SLC16A1
YPEL
SEDTD1
MAZ
LG8★
LG27
LG20
LG22
LG3
UNK755
LG12
UNK27
UNK33
LG8★
UNK2
UNK23
UNK445
LG10
LG15
LG5
LG23
23D0
6D0/D1
3D1
7D2
5D2
8D3
19D3
12
13
14
9



Supplementary figure S35. Conserved synteny blocks in the vicinity of *GH*, *PRL*, and *PRL2* genes in the zebrafish genome. See Figs. S2 and S3 captions for details.Download full-size image: https://doi.org/10.6084/m9.figshare.6494879.

Supplementary figure S36 (next page). Conserved synteny between *GH* and *PRL*-bearing chromosome blocks. Colors are applied following the *gh1* (light blue) and *prl* (dark blue) blocks on zebrafish chromosome 3. Red color indicates paralogous genes shared between the *GH*-bearing block on chromosome 27 and the *PRL*-bearing block on chromosome 2 in the chicken genome. We identified 35 gene predictions within 5 MB in each direction of the zebrafish *gh1* gene, and 21 gene predictions within 5 MB in each direction of the *prl* gene, both located on chromosome 3, approximately 10 MB apart. These 56 genes have predicted orthologs located in the vicinity of *GH* genes, but not *PRL* genes, in the spotted gar, *Xenopus tropicalis*, chicken and human genomes. Three gene families have paralogous genes located on both *GH*-bearing and *PRL*-bearing chromosomes in the chicken and *Xenopus tropicalis* genomes: CNTNAP, EZH and TNS. These are highlighted in red. The location of the spotted gar *PRL* is as yet unknown. Other than *gh1* and *prl,* we found no paralogous genes shared between the two chromosome blocks on zebrafish chromosome 3. Download full-size image: https://doi.org/10.6084/m9.figshare.6494879.


Zebrafish
Spotted gar
Xenopus tropicalis
Chicken
Human



Supplementary figure S37. Conserved synteny blocks mapped onto the corresponding chromosomes in the human, chicken, *Xenopus tropicalis*, spotted gar, zebrafish and medaka genomes. "D0" – "D3" and "J0" – "J1" designations correspond to vertebrate paralogous blocks that arose in 1R and 2R, as identified by Nakatani et al., 2007. Lower case letter designations "k", "j" et c. correspond to teleost chromosome blocks pre-3R, as identified by Kasahara et al. 2007 and Bian et al. 2016. Numbers in parentheses correspond to teleost chromosome blocks pre-3R, as identified by Nakatani & McLysaght 2017. Conserved synteny blocks containing less than 5 genes are drawn smaller and with patterned backgrounds. Asterisks mark chromosomes where the "D" and "J" paralogy blocks co-localize. Download full-size image: https://doi.org/10.6084/m9.figshare.6494879.