Published in final edited form as:
Nat Genet. 2016 April ; 48(4): 427–437. doi:10.1038/ng.3526.

The spotted gar genome illuminates vertebrate evolution and facilitates human-to-teleost comparisons

A full list of authors and affiliations appears at the end of the article.

Abstract

To connect human biology to fish biomedical models, we sequenced the genome of spotted gar (*Lepisosteus oculatus*), whose lineage diverged from teleosts before the teleost genome duplication (TGD). The slowly evolving gar genome conserved in content and size many entire chromosomes from bony vertebrate ancestors. Gar bridges teleosts to tetrapods by illuminating the evolution of immunity, mineralization, and development (e.g., Hox, ParaHox, and miRNA genes). Numerous conserved non-coding elements (CNEs, often *cis*-regulatory) undetectable in direct human-teleost comparisons become apparent using gar: functional studies uncovered conserved roles of such cryptic CNEs, facilitating annotation of sequences identified in human genome-wide association studies. Transcriptomic analyses revealed that the sum of expression domains and levels from duplicated teleost genes often approximate patterns and levels of gar genes, consistent with subfunctionalization. The gar genome provides a resource for understanding evolution after genome duplication, the origin of vertebrate genomes, and the function of human regulatory sequences.

Keywords

GWAS; comparative medicine; polyploidy; zebrafish; medaka; neofunctionalization

Teleost fish represent about half of all living vertebrate species[1] and provide important models for human disease (e.g. zebrafish and medaka)[2-9]. Connecting teleost genes and gene functions to human biology (Fig. 1a) can be challenging, however, due to 1) two rounds of

Correspondence should be addressed to I.B. (braasch@msu.edu) or J.H.P. (jpostle@uoneuro.uoregon.edu).
[34]Present addresses: Department of Integrative Biology, Michigan State University, East Lansing, Michigan, USA (I.B.); Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, USA (M.S.C.); Department of Animal and Plant Sciences, University of Sheffield, Sheffield, United Kingdom (K.J.M.); Department of Genetics, University of Georgia, Athens, Georgia, USA (D.C.); Department of Genetics, University of Pennsylvania, Philadelphia, Pennsylvania, USA (S.F.); Young Investigators Group Bioinformatics and Transcriptomics, Department of Proteomics, Helmholtz Centre for Environmental Research – UFZ, Leipzig, Germany (J.H.); ecSeq Bioinformatics, Leipzig, Germany (M.F.); Vertebrate and Health Genomics, The Genome Analysis Center, Norwich, United Kingdom (F.D.P.).

Accession codes. GenBank accession number for the spotted gar genome: GCA_000242695.1. RNA-seq data are available under SRA accession numbers SRP042013 (Broad Institute gar transcriptome), SRP044781-SRP044784 (PhyloFish transcriptomes of zebrafish, gar, bowfin, and medaka), and SRP063942 (gar small RNA-seq for miRNA annotation). Gar *scpp* gene sequences are available under GenBank accession numbers KU189274-KU189300.

Note: Supplementary information is available in the online version of the paper.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

Europe PMC Funders Author Manuscripts

early vertebrate genome duplication (VGD1 and VGD2[10], but see[11]) followed by reciprocal loss of some ohnologs (gene duplicates derived from genome duplication[38]) in teleosts and tetrapods, including humans (e.g.,[12,13]); 2) the teleost genome duplication (TGD), which resulted in duplicates of many human genes[14,15]; and 3) rapid teleost sequence evolution[16,17], often due to asymmetric rates of ohnolog evolution that frustrates ortholog identification. To help connect teleost biomedicine to human biology, we sequenced the genome of spotted gar (*Lepisosteus oculatus*, henceforth 'gar'; see also Supplementary Note 1, Supplementary Fig. 1), because its lineage represents the unduplicated sister group of teleosts[18,19] (Fig. 1a).

Gar informs the evolution of vertebrate genomes and gene functions after genome duplication and illuminates evolutionary mechanisms leading to teleost biodiversity. The gar genome evolved comparatively slowly and clarifies the evolution and orthology of problematic teleost protein-coding and miRNA gene families. Surprisingly, many entire gar chromosomes have been conserved with some tetrapods for 450 million years. Importantly, gar reveals conserved non-coding elements (CNEs), which are often regulatory, that teleosts and humans share but that direct sequence comparisons do not detect. Global gene expression analyses show that expression domains and levels of TGD duplicates usually sum to those in gar, as expected if ancestral regulatory elements partitioned after the TGD. By illuminating the legacy of genome duplication, the gar genome bridges teleost biology to human health, disease, development, physiology, and evolution.

RESULTS

Genome assembly and annotation

The genome of a single adult gar female collected in Louisiana (USA) was Illumina sequenced to 90X coverage. The ALLPATHS-LG[20] draft assembly covers 945 Mb with quality metrics comparable to other vertebrate Illumina assemblies[20]. To generate a 'chromonome' (chromosome-level genome assembly[21]), we anchored scaffolds to a meiotic map[19] capturing 94% of assembled bases in 29 linkage groups (LGs) (Supplementary Note 2). Transcriptomes from adult tissues and developmental stages (Supplementary Note 3) facilitated a MAKER[22]-annotated gene set of 21,443 high confidence protein-coding genes, while ENSEMBL annotation identified 18,328 protein-coding genes (mostly a subset of MAKER annotations), 42 pseudogenes, and 2,595 ncRNAs (Supplementary Note 4), compared to human (20,296 protein coding genes) and zebrafish (25,642). About 20% of the gar genome is repetitive, including transposable elements (TEs) representing most lobe-finned and teleost TE superfamilies and a TE profile similar to that of coelacanth[23], thus clarifying TE phylogenetic origins (Supplementary Note 5, Supplementary Tabs.1-3, Supplementary Figs. 2-5).

The gar lineage evolved slowly

Phylogenies of 243 one-to-one orthologs in 25 jawed vertebrates[16], including gar and our transcriptome of bowfin *Amia calva* (Supplementary Notes 3,4, Supplementary File 1), strongly support the monophyly of Holostei (gar+bowfin) as sister group to teleosts (Fig. 1b,

Supplementary Note 6, Supplementary Fig. 6)[24-27], suggesting that morphologies shared by bowfin and teleosts[28,29] may be convergent or ancestral traits altered in the gar lineage.

Darwin applied his term 'living fossil' to 'ganoid fishes', including gars[30]; indeed, gars show low rates of speciation and phenotypic evolution[31]. Evolutionary rate analyses using cartilaginous fish outgroups show that gar and bowfin proteins evolved significantly slower than teleost sequences. Holostei have a significantly shorter branch length to the cartilaginous outgroup than most other bony vertebrates except coelacanth, the slowest evolving bony vertebrate[16,32] (Fig. 1b, Supplementary Note 7, Supplementary Tab. 4). Our results support the hypothesis that the TGD could have facilitated the high rate of teleost sequence evolution [16,17,33]. Gar TEs also show a low turnover rate compared to teleosts, mammals, and even coelacanth[23] (Supplementary Note 5, Supplementary Fig. 5).

Gar informs the evolution of bony vertebrate karyotypes

Gar represents the first chromonome[21] of a non-tetrapod, non-teleost jawed vertebrate, allowing for the first time long-range gene order analyses without the confounding effects of the TGD. The gar karyotype (2N=58) contains both macro- and microchromosomes (Fig. 2a, Supplementary Note 8.1, Supplementary Fig. 7). Aligning gar chromosomes to those of human, chicken, and teleosts revealed distinct conservation of orthologous segments in all species (Fig. 2b-e, Supplementary Note 8.2, Supplementary Figs. 8,9). Strikingly, gar-chicken comparisons revealed conservation of many entire chromosomes (Fig. 2c). Chicken and gar karyotypes differ only by about 17 large fissions, fusions, or translocations. Almost half of the gar karyotype (14/29 chromosomes) showed a nearly one-to-one relationship in gar-chicken comparisons, including macro- and microchromosomes with highly correlated chromosome assembly lengths (Fig. 2d, Supplementary Note 8.2). Similarity in chromosome size and gene content is strong evidence that the karyotype of the common bony vertebrate ancestor possessed both macro-and microchromosomes as Ohno (1969)[34] hypothesized, consistent with microchromosomes in coelacanth[35] and cartilaginous fish[34], for which no chromonomes are yet available.

The gar chromonome also tests the hypothesis that an increase in interchromosomal rearrangements occurred in teleosts after, and possibly due to, the TGD[19]. For each gar chromosome segment, teleosts usually have two ohnologous segments, verifying a pre-TGD gar-teleost divergence[19]. Each TGD pair in teleosts usually shares conserved synteny with more than one gar chromosome, indicating rearrangements before the TGD (Fig. 2e, Supplementary Note 8.2, Supplementary Figs. 8,9). Gar shares many whole chromosomes with chicken (Fig. 2c) but few with teleosts (Fig. 2e). These results show that chromosome fusions thought to have occurred in the ray-finned lineage after divergence from the lobe-finned lineage[36] actually occurred in the teleost lineage after divergence from gar but before the TGD (Fig. 2f, Supplementary Fig. 10). This finding explains how spotted gar has more chromosomes (N=29, Fig. 2a) than typical teleosts (N~24-25[37]) without experiencing the TGD. Comparisons taking the TGD into account further revealed an average fission/translocation rate in percomorphs (stickleback, medaka, pufferfish) relative to gar similar to that in the chicken lineage. Zebrafish has a higher rearrangement rate, however, even after accounting for the TGD (Supplementary Note 8.2, Supplementary Fig. 11). These

comparisons indicate that the TGD might not fully account for high teleost rearrangement rates.

Gar clarifies vertebrate gene family evolution

Lineage-specific loss of ohnologs often followed VGD1, VGD2, and the TGD (Fig. 1a), which complicates identification of true orthologs[21,39] and frustrates translating knowledge from teleosts biomedical models to human biology, e.g.,[12]. Gar is uniquely informative because its lineage did not experience the TGD and often retained ancestral VGD1/VGD2 ohnologs that were reciprocally lost in teleosts and tetrapods, thus clarifying the evolution of gene families involved in vertebrate development, physiology, and immunity (Supplementary Note 9).

Developmental gene family analyses revealed stability in the gar gene repertoire, including *Hox* clusters (Supplementary Note 9.1). Gar has 43 *hox* genes organized in four clusters expected for an unduplicated ray-finned fish (Supplementary Fig. 12). No *hox* gene has been completely lost in gar since divergence from the last common ray-finned ancestor. The *hoxD14* gene, missing from teleosts but present in paddlefish[40], is recognizable as a pseudogene in gar (Supplementary Fig. 13). In contrast, teleosts have far fewer *hox* cluster genes than the 82 expected after genome duplication (e.g., zebrafish, 49 genes; stickleback, 46), demonstrating massive *hox* cluster gene loss after the TGD. Teleosts lack orthologs of *hoxA6* and *hoxD2*, zebrafish lacks all *hoxDb* cluster protein-coding genes[14], and percomorphs lack the *hoxCb* cluster[41], but gar lacks just one *hox* cluster gene from the last common bony vertebrate ancestor (*hoxA14*), fewer than tetrapods (e.g., human: three losses) and coelacanth (two) (Supplementary Fig. 12). Gar *ParaHox* clusters (Supplementary Note 9.2, Supplementary Tab. 5) are also more complete than those in teleosts and tetrapods, with four clusters containing seven genes. Gar retained *cdx2*, revealing a VGD1/VGD2 ohnolog 'gone missing' from teleosts (Supplementary Fig. 14). Gar possesses the VGD1/VGD2 ohnolog *pdx2*, previously found only in cartilaginous fish and coelacanth[42], showing that *pdx2* was lost independently in teleosts and tetrapods (Supplementary Figs. 14,15). Retinoic acid regulates *Hox* cluster gene expression[43] but retinoic acid-synthesizing Aldh enzymes (Supplementary Note 9.3) vary in number among vertebrates[44]: tetrapods have three genes (*Aldh1a1*, *Aldh1a2*, *Aldh1a3)*, zebrafish has two (*aldh1a2*, *aldh1a3*), medaka just one (*aldh1a2*)[45]. Finding all three genes in gar rules out the hypothesis[45] that *Aldh1a1* was a lobe-finned innovation (Supplementary Fig. 16).

Physiological mechanisms are shared among vertebrates, including light control of circadian rhythms, despite important gene repertoire differences between teleosts and tetrapods [46,47]. Analyses of gar circadian clock (Supplementary Note 9.4, Supplementary Tab. 6, Supplementary Fig. 17)[48] and opsin genes (Supplementary Note 9.5, Supplementary Tab. 7, Supplementary Fig. 18)[49] link gene repertoires of teleosts and tetrapods: e.g., gar clarifies circadian gene origins in VGD vs. TGD events. Gar has *pinopsin*, present in tetrapods but absent from teleosts, along with *exo-rhodopsin*, previously thought to compensate the lack of *pinopsin* in teleosts[50].

Evolution of vertebrate immunity becomes clearer using gar (Supplementary Note 9.6). Major-histocompatibility complex (MHC) class I and class II genes (Supplementary Figs.

19-21) are tightly linked in tetrapods and cartilaginous fish but are unlinked in teleosts[51,52]. In gar, at least one pair of class I and class II genes are linked as in tetrapods[53,54], suggesting that gar retains the ancestral configuration although most gar MHC genes remain on unassembled scaffolds (Supplementary Fig. 21). Gar has some class I genes thought to be teleost-specific (Z/P-, L-, and U/S-like, e.g.[54-56]; Supplementary Fig. 19) and some class II genes similar to, and some distinct from, teleost DA/DB and DE lineages (Supplementary Fig. 20). Several gar MHC region genes are on unassembled scaffolds linked to genes whose human orthologs are encoded in MHC class II or MHC class III regions on Hsa6 and some are adjacent to orthologs of teleost MHC class I genes (Supplementary Tab. 8). The human MHC class III region on Hsa6 has syntenic segments on Hsa1, Hsa9, and Hsa19; these four ohnologons likely arose in VGD1 and VGD2[57] as supported by the gar genome (Supplementary Tab. 8).

Gar immunoglobulin (Ig) genes (Supplementary Fig. 22) and transcripts generally resemble those of teleosts. Unexpectedly, gar has a second, distinct IgM locus but lacks IgT (IgZ)[58,59], thought to provide mucosal immunity[60], suggesting that IgT is teleost-specific and that gar ganoid scales may suffice for exterior surface protection. Gar T-cell receptor genes (Supplementary Fig. 23) are tightly linked as in mammals, but unlike in *Xenopus*[61], they are downstream of V_H and J_H segments. Phylogenetic analyses of Toll-like receptor (TLR) genes (Supplementary Fig. 24) from tetrapods, teleosts, and gar revealed that the 16 identifiable gar TLRs embrace all six major TLR families[62]. Gar TLRs appear to share evolutionary histories with teleosts and/or tetrapods. Gar encodes NITR (novel immune-type receptor) genes (Supplementary Fig. 25), which function in allorecognition and were thought to be teleost-specific[63,64]. The 17 gar *nitr* genes form 15 families, suggesting few recent tandem duplications or rapid divergence after gene duplication. In sum, the gar immunogenome bridges teleosts to tetrapods.

Gar uncovers evolution of vertebrate mineralized tissues

Bony vertebrates share mineralized tissues (bone, dentin, enameloid, and enamel), yet gene repertoires for the secretory calcium-binding phosphoproteins (Scpp) that form these tissues[65,66] differ substantially between teleosts and tetrapods and their evolution remains controversial[17,67,68]. Gar clarifies understanding because it retains ancient characteristics both in its ganoid scales, which contain ganoin, hypothesized to be a type of enamel[69], and in its teeth, which are covered by both enameloid and enamel[70] (Supplementary Note 10). Mammalian genomes were thought to contain the largest number of *Scpp* genes (human, 23 genes; coelacanth, 14; zebrafish, 15) and only two (*Spp1* and *Odam*) seemed common between lobe-finned vertebrates and teleosts[68] (Fig. 3a). We identified 35 *scpp* genes in gar in two clusters on LG2 and LG4 (Fig. 3a, Supplementary Note 10, Supplementary Tab. 9), which contain *spp1* and *odam*, respectively. Importantly, gar includes orthologs of five *scpp* genes previously found only in teleosts and six known only from lobe-finned vertebrates. Another 18 gar *scpp* genes have no identified ortholog in either lobe-finned vertebrates or teleosts (Fig. 3a, Supplementary Note 10, Supplementary Tab. 9).

Enamel matrix protein genes *Ameloblastin* (*Ambn*), *Enamelin* (*Enam*), and *Amelogenin* (*Amel*) are found in lobe-finned vertebrates with enamel-bearing teeth, but not in teleosts,

which lack enamel-bearing teeth[66,68]. For the first time in a ray-finned vertebrate, we identified *ambn* and *enam* genes (but no *Amel* ortholog) in gar genome and transcriptomes. Gar *ambn* and *enam* genes show sequence similarities to zebrafish *scpp6* and *fa93e10*, respectively, suggesting that teleosts may have divergent orthologs, supported by conserved gene orders in gar and zebrafish clusters (Fig. 3a).

RT-PCR and our gar skin transcriptome revealed expression of *ambn* and *enam* in enamel-containing gar teeth and in gar skin that includes scales with ganoin (Supplementary Note 10, Supplementary Tab. 9), suggesting that strong expression of *ambn* and *enam* is limited to enamel and ganoin. Thus, enamel in teeth and ganoin in ganoid scales likely represent the same tissue and common expression of *Ambn* and *Enam* in lobe-finned enamel and in gar enamel/ganoin supports homology of these tissues. Analysis of gnathostome fossils suggested that ganoin is plesiomorphic for crown osteichthyans and arose before enamel[71]; thus, enamel-bearing teeth likely evolved by co-opting enamel matrix genes originally used in ganoid scales. *Amel* may have evolved subsequently to encode the principal organic component of “true enamel” that appears to have originated in lobe-finned vertebrates[68].

Gar expressed twelve additional *scpp* genes (including *odam* and *scpp9* hypermineralization genes[66]) in both teeth and scales and another four genes in bone (Supplementary Tab. 9), strongly suggesting that the common ancestor of extant bony vertebrates had a rich repertoire of *Scpp* genes, many of which were expressed in mineralized tissues, and that although teleosts and lobe-finned vertebrates independently lost subsets of ancient *Scpp* genes[65], gar retained characteristics of both lineages.

Gar connects vertebrate microRNAomes

MicroRNA genes could become teleost- or tetrapod-specific[17,72] by loss in one lineage or gain in the other. We studied gar miRNAs computationally (Supplementary Note 11.1, Supplementary Tab. 10, Supplementary Fig. 27) and annotated them using a sequence-based approach (Supplementary Note 11.2). Small RNA sequencing from four tissues identified 302 mature miRNAs from 233 genes, 229 belonging to 107 families and four without a known family (Supplementary Tab. 11, Supplementary Fig. 28). Gar-zebrafish[73,74] comparisons showed that four families and four individual miRNA genes emerged in teleosts. Of 22 families thought to be teleost losses[17], two actually belong to the same family and orthologs of four gar miRNA genes were previously overlooked in teleosts. Fourteen families are absent from both gar and teleosts, and three are present in gar and many teleosts[74] but absent from zebrafish. A single family present in teleosts and lobe-finned fish (*mir150*) was not found in gar. Notably, no miRNA family loss was teleost-specific, suggesting that the TGD did not accelerate family loss.

The ‘gar bridge’ helps identify miRNA orthologies. For example, mammalian *Mir425* and *Mir191*, thought to be lost in teleosts[17], are orthologs of teleost *mir731* and *mir462*, respectively (Fig. 3b). Additionally, mammalian *Mir135B* is orthologous to gar *mir135c* and zebrafish TGD ohnologs *mir135c-1* and *mir135c-2* (Fig. 3c). The post-TGD retention rate for zebrafish miRNA ohnologs is 39% (81/208 analyzable cases), considerably higher than the rate for protein-coding genes (20-24%[75]), consistent with the hypothesis that miRNA

genes are likely to be retained after duplication due to their incorporation into multiple gene regulatory networks[76-79].

Gar reveals hidden orthology of *cis*-regulatory elements

Conserved non-coding elements (CNEs) often function as *cis*-acting regulators[80,81], but many are not visible in teleosts, presumably due to rapid teleost sequence evolution (Fig. 1b, Supplementary Note 7); ancient CNEs identified in tetrapods, however, might be detected in ray-finned fish using the slowly evolving gar.

CNE analyses near developmental gene loci (*Hox*/*ParaHox* clusters, *Pax6*, *IrxB*) showed that gar contains more gnathostome CNEs (conserved between bony vertebrates and elephant shark) than teleosts. Analyses incorporating gar identified many bony vertebrate CNEs (i.e., absent from elephant shark) that were not predicted by direct human-teleost comparisons; furthermore, gar-based alignments identified CNEs recruited in the common ancestor of ray-finned fishes (Supplementary Notes 9.2, 12.1, Supplementary Tabs. 12-19, Supplementary Figs. 14-15,29-35).

Gar unravels the origins of tetrapod limb enhancers, evidenced by whole-genome alignments for 13 vertebrates (gar, five teleosts, coelacanth, five tetrapods, elephant shark, Supplementary Note 12.2, Supplementary Tabs. 20-21, Supplementary Fig. 36). For 153 known human limb enhancers [32, 82-84], human-centric alignments identified 71% (108) in gar but only 53% (81) in direct human-teleost alignments. Of 72 limb enhancers lacking a human-teleost alignment, 40% (29/72) aligned to gar, confirming their presence in the bony vertebrate ancestor and loss or considerable divergence in teleosts. Of these 29 enhancers, 15 also aligned to elephant shark, revealing their existence in the gnathostome ancestor. Fourteen occurred in gar but not teleosts and would have been incorrectly characterized as lobe-finned innovations without gar (Supplementary Note 12.3, Supplementary Tab. 22).

Using the ‘gar bridge’ (Fig. 4a), we tested whether these 29 enhancers not directly identified in teleosts might represent rapid divergence rather than definitive loss. Inspection of human-centric and gar-centric alignments revealed 48% (14/29) aligning to at least one teleost (Supplementary Tab. 22). Gar thus substantially improves understanding of the evolutionary origin of vertebrate limb enhancers and their fate in teleosts (Fig. 4b, Supplementary Tab. 22, Supplementary Fig. 37). Strikingly, despite using the ‘gar bridge’, we found that teleosts lost substantially more limb enhancers (15) than gar (two) (Fig. 4b, Supplementary Fig. 37), suggesting gar as a better model than teleosts for investigating the fin-to-limb transition[85].

Functional studies of a *HoxD* limb enhancer tested the utility of a ‘gar CNE bridge’. *HoxD* and *HoxA* clusters pattern proximal and distal mammalian limbs by ‘early’ and ‘late’ phases of gene expression, respectively[86]. Early phase *HoxD* expression in fins and limbs shows several presumed homologous features[87] and may derive from shared but cryptic regulatory elements. Elements CNS39 and CNS65 drive early phase *HoxD* activation in mammals[88] (Figure 5a). Human-centric (Supplementary Tab. 22) and local mouse-centric alignments (Figure 5a) failed to detect CNS39 in ray-finned fish, but identified CNS65 in gar. Significantly, CNS65 appeared in teleosts only using the ‘gar bridge’ (Figure 5a, Supplementary Tab. 22).

To test if cryptic CNE orthologs preserve enhancer function, we used CNS65-driven reporter constructs to generate transgenic zebrafish and mice (Supplementary Note 12.4). CNS65 from either gar or zebrafish drove early expression in the developing zebrafish pectoral fin (Figure 5b). Gar CNS65 drove expression in fore- and hindlimbs of stage e10.5 mice (Figure 5c) indistinguishable from murine CNS65[88]. Zebrafish CNS65 activated forelimb expression somewhat weaker than gar CNS65 (Figure 5c). At e12.5, gar CNS65 activated proximal but not distal limb expression (Figure 5c), mimicking the endogenous murine enhancer[88]. These functional experiments demonstrate that regulation of *HoxD* early phase expression in limbs and fins is an ancestral, conserved feature of bony vertebrates and that gar connects otherwise cryptic teleost regulatory mechanisms to mammalian developmental biology.

Gar bridges human CNEs to teleost biomedical models—Genome-wide, we identified approximately 28% of human-centric CNEs (39,964/143,525) in gar, more than in any of five aligned teleost genomes. Around 19,000 human-centric CNEs aligned to gar but not to any teleost (Supplementary Note 12.2, Supplementary Tab. 21). Without gar, one would have erroneously concluded that these elements originated in lobe-finned vertebrates or were lost in teleosts. The ‘gar bridge’ (Fig. 4a) established hidden orthology from human to gar to zebrafish for many of these human-centric CNEs (30-36%, depending on overlap; Supplementary Note 12.2, Supplementary Tab. 21). These approximately 6,500 newly connected human CNEs contain around a thousand SNPs linked to human conditions in genome-wide association studies (GWAS), thereby connecting otherwise undetected disease-associated haplotypes to genomic locations in zebrafish (Supplementary Tab. 21). The gar bridge thus helps identify biomedically relevant candidate regions in model teleosts for functional testing, thereby enhancing teleost models for personalized medicine.

Gar illuminates gene expression evolution following the TGD

Ohnologs experience several non-exclusive fates after genome duplication: loss of one copy, evolution of new expression domains or protein functions, and the partitioning of ancestral functions[89-92]. Because the contribution of various fates has not yet been studied using a close TGD outgroup, we generated a list of gar genes and their orthologous TGD ohnologs or singletons in zebrafish and medaka using phylogenetic[93] and conserved synteny analyses[94] (Fig. 6a,b, Supplementary Note 13.1, Supplementary Tab. 23).

To compare tissue-specific gene expression patterns, we conducted RNA-seq for ten adult organs and for stage-matched embryos for gar, zebrafish, and medaka, then normalized reads across tissues for each gene in each species (Supplementary Notes 3.2,13.2). For example, gar expressed *slc1a3* mainly in brain, bone, and testis, but both teleosts expressed one ohnolog primarily in brain and the other primarily in liver, a novel expression domain, with little expression in bone or testis (Fig. 6c). Novel expression domains like this are expected if one ohnolog maintained ancestral patterns while the other evolved new functions[95] before the teleost radiation. In contrast, gar expressed *gpr22* mostly in brain and heart but both teleosts expressed one ohnolog in brain, the other in heart (Fig. 6d), as expected by partitioning of ancestral regulatory subfunctions[89].

To characterize effects of the TGD on evolution of gene expression, we plotted tissue-specific expression levels in gar vs. 1) expression of orthologous teleost singletons, 2) expression of each TGD ohnolog when both were retained, and 3) the averaged expression level of both retained ohnologs ('ohnolog-pair'), and then calculated correlation coefficients. Results showed that the correlation of *expression patterns* of gar genes to their teleost singleton orthologs ('Singl' in Fig. 6e,f) was not significantly different from the correlation of expression patterns of gar genes to either copy of their teleost TGD co-orthologs ('Ohno1' and 'Ohno2' in Fig. 6e,f); thus, compared to ancestral single-copy genes as estimated from gar, teleost ohnologs binned at random do not appear to have evolved expression pattern differences significantly more rapidly than singletons. In contrast, the average tissue-specific patterns of both TGD duplicates ('OhnoPair' in Fig. 6e,f) correlated significantly more closely to gar than either ohnolog taken alone and more closely than singletons; thus, ancestral gene subfunctions tended to partition between TGD ohnologs and to maintain ancestral functions as a gene pair, as predicted by the subfunctionalization model[89].

We next calculated average *expression levels* for each gene over the 11 tissues and computed the ratio of each teleost gene to its gar ortholog. Comparisons showed that individual ohnologs (Fig. 6g,h) were each expressed at significantly lower levels than singletons (Fig. 6g,h) compared to their gar orthologs. The ratio of expression levels of ohnolog-pairs to gar expression levels, however, showed no statistical difference from singleton/gar expression ratios (Fig. 6g,h). This finding suggests that the aggregate expression level of ohnolog-pairs tends to evolve to approximate the level of the pre-duplication gene as expected by quantitative subfunctionalization[89,90,96].

Taken together, our analyses indicate that post-TGD, ohnolog-pairs evolved so that the sum of their expression domains and the sum of their expression levels usually approximated the patterns and levels of pre-duplication genes.

DISCUSSION

Gar is the first ray-finned fish genome sequence not impacted by the TGD. Due to its phylogenetic position, slow rate of sequence evolution, dense genetic map, and ease of laboratory culture, this resource provides a unique bridge between tetrapods and teleost biomedical models. Analysis revealed that gar bridges teleosts to tetrapods in genome arrangement, identifying orthologous genes, possessing ancient VGD ohnologs lost reciprocally in teleosts and tetrapods, understanding evolution of vertebrate-specific features including adaptive immunity and mineralized tissues, and the evolution of gene expression. Clarification of gene orthology and history is crucial for the design, analysis, and interpretation of teleost models of human disease, including those generated with CRISPR/Cas9-induced genome editing[97,98]. Gar analyses show that sequences formerly considered unique to teleosts or tetrapods are often shared between ray-finned and lobe-finned vertebrates including human. Importantly, the gar bridge helps identify potential gene regulatory elements that are shared by teleosts and humans but invisible in direct teleost-tetrapod comparisons. Availability of gar embryos and ease of raising eggs to adults in the laboratory[21] (Supplementary Fig. 1) makes gar a ray-finned species of choice when

analyzing many vertebrate developmental and physiological features. In conclusion, the gar bridge facilitates the connectivity of teleost medical models to human biology.

ONLINE METHODS

A full description of methods can be found in the Supplementary Note. Animal work was approved by the University of Oregon Institutional Animal Care and Use Committee (Animal Welfare Assurance Number A-3009-01, IACUC protocol 12-02RA).

Gar genome sequencing and assembly

The spotted gar genome was sequenced and assembled using DNA from a single adult female gar wild-caught in Bayou Chevreuil, St. James Parish, Louisiana, USA (Supplementary Note 1). It was sequenced by Illumina sequencing technology and jumping libraries to 90X coverage and assembled into LepOcu1 (accession number AHAT00000000.1) using ALLPATHS-LG[20]. The draft assembly is 945 Mb in size and is composed of 869 Mb of sequence plus gaps between contigs. The spotted gar genome assembly has a contig N50 size of 68.3 kb, a scaffold N50 size of 6.9 Mb, and quality metrics comparable to other vertebrate Illumina genome assemblies[20]. A total of 209 scaffolds were anchored in 29 linkage groups using 2,153 of 8,406 meiotic map RAD-tag markers[19], thus capturing 891 Mb of sequence or 94.2% of bases in the chromonome assembly (Supplementary Note 2).

RNA-seq transcriptomes

The Broad Institute gar RNA-seq transcriptome (Supplementary Note 3.1) was generated from 10 tissues (stage 28 embryo[99], 8 day larvae, eye, liver, heart, skin, muscle, kidney, testis) and assembled using Trinity[100]. PhyloFish RNA-seq transcriptomes of gar (Supplementary Note 3.2), bowfin (Supplementary Note 3.3), zebrafish, and medaka (Supplementary Note 13.2) were generated from 10 adult tissues (ovary, testis, brain, gills, heart, muscle, liver, kidney, bone, intestine) and one embryonic stage (‘pigmented eye’ stage of gar, zebrafish, medaka) and assembled using the Velvet/Oases package[101].

Genome annotation

Using evidence from the Broad and PhyloFish gar transcriptomes (Supplementary Note 3), all RefSeq teleost proteins, and all Uniprot/Swissprot proteins, MAKER2[22] annotated 25,645 protein-coding genes (Supplementary Note 4.1). Using the Broad transcriptome (Supplementary Note 3.1), the Ensembl gene annotation pipeline identified 18,328 protein-coding genes for 22,470 transcripts along with 42 pseudogenes and 2,595 ncRNAs (Supplementary Note 4.2). Annotations for 762 and 6,877 genes are specific to Ensembl and MAKER, respectively. The 21,443 high confidence gene set predicted by MAKER is likely close to the true number of gar protein-coding genes.

Annotation of transposable elements

Manual and automatic classification (using RepeatScout and RepeatModeler) of gar TEs was performed on the basis of Wicker’s nomenclature[102] and identified elements were combined into a single library (Supplementary Note 5), which was then used to mask the

Europe PMC Funders Author Manuscripts

genome with RepeatMasker. The TE age profile was determined using Kimura distances of individual TE copies to the corresponding TE consensus sequence (Supplementary Note 5).

Phylogenomic and evolutionary rate analyses

Phylogenetic analyses (Supplementary Note 6) were based on protein-coding sequence alignments described for the coelacanth genome analysis[16] but updated with orthologous sequences from gar and bowfin (Supplementary Notes 3,4) and from the slowly evolving Western painted turtle[103]. Phylogenetic reconstructions were carried out with RAxML[104] and PhyloBayes MPI[105]. Molecular rate analyses (Supplementary Note 7) were performed at the protein alignment level with Tajima's relative rate tests[106] and at the level of the reconstructed phylogenies with Two-Cluster tests[107].

Genome structure analyses

The spotted gar karyotype was determined from caudal fin fibroblast cell cultures established as described for zebrafish[108] (Supplementary Note 8.1). Conserved synteny analyses between gar, tetrapods (human, chicken) and teleosts (Supplementary Note 8.2) were performed with 1) Circos plots[109] based on orthology relations from Ensembl75 and as described in Supplementary Note 13.1); 2) the Synteny Database[94] after integration of the gar genome assembly (Ensembl74 version); and 3) comparative synteny maps derived as described in refs.[16,110].

Gene family analyses

Individual gene families were analyzed as described in Supplementary Notes 9. RT-PCR and sequencing was performed to annotate and to analyze gene expression of *scpp* mineralization genes using cDNA libraries from gar teeth, jaw, and scales (Supplementary Note 10).

miRNA annotation and analysis

Gar miRNAs were studied *in silico* (Supplementary Note 11.1) by blasting teleost and tetrapod miRNAs from miRBase[74,111-113] against the gar genome assembly and confirmed with RNAfold[114] (see also ref.[72]). miRNA annotation and analyses based on sequencing data of gar miRNAs (Supplementary Note 11.2) was performed as described for zebrafish[73] by utilizing small RNA-seq data from adult brain, heart, testis, and ovary tissue, which were processed and annotated with Prost![115] according to miRNA gene nomenclature guidelines[116]; miRNA orthologies based on conserved synteny were established using Ensembl[117], the Synteny Database[94] and Genomicus[118,119].

Analysis of conserved non-coding elements

Investigation of CNEs in developmental gene loci were performed using SLAGAN[120] in VISTA[121] (Supplementary Notes 9.2,12.1). Gar-, zebrafish-, and human-centric 13-way multi-genome alignments were generated with MultiZ[122] based on lastZ[123] pairwise whole genome alignments (WGAs). We used phyloFit[124] to generate a neutral model of 4d site evolution to identify conserved elements with phastCons[124]; genic elements and repetitive sequences were filtered out to obtain CNEs. Evolution of human limb enhancers[32,82-84] was

established using WGAs and conserved synteny curation. Genome-wide connectivity of CNEs and embedded GWAS-SNPs from human to zebrafish through gar was established from WGAs using liftOver[125] and BEDtools[126] (Supplementary Notes 12.2,12.3).

HoxD enhancer functional analysis

Gar and teleost orthologs of *HoxD* early enhancer CNS65 were identified with VISTA (LAGAN)[121]. Gar and zebrafish CNS65 were cloned into pXIG-cFos-eGFP and Gateway-Hsp68-LacZ vectors for zebrafish[127] and mouse transgenesis (Cyagen Biosciences), respectively (Supplementary Note 12.4).

Comparative gene expression analyses

Curated lists of TGD ohnologs and TGD singletons of zebrafish and medaka and their gar (co-)orthologs were generated by integrating phylogenetic information from EnsemblCompara GeneTrees[93] (Ensembl74) and conserved synteny data from the Synteny Database[94] (Supplementary Note 13.1). For all three species, RNA-seq reads from the PhyloFish transcriptomes (Supplementary Note 3.2,13.2) were mapped against the longest Ensembl reference coding sequence of each gene with BWA-Bowtie[128,129], counted with SAMtools[130] and normalized for each gene across the 11 tissues using DESeq[131]. The correlation of expression patterns and relative levels of expression between each zebrafish/medaka gene and its gar ortholog and of singletons, ohnolog 1, ohnolog 2, and 'ohnolog pairs' were determined using R[132]. See Supplementary Note 13.2 for additional information including definition of 'ohnolog pair' expression and criteria for detecting neo- and subfunctionalization detection.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Ingo Braasch[1,34], Andrew R. Gehrke[2], Jeramiah J. Smith[3], Kazuhiko Kawasaki[4], Tereza Manousaki[5], Jeremy Pasquier[6], Angel Amores[1], Thomas Desvignes[1], Peter Batzel[1], Julian Catchen[7], Aaron M. Berlin[8], Michael S. Campbell[9,34], Daniel Barrell[10,11], Kyle J. Martin[12,34], John F. Mulley[13], Vydianathan Ravi[14], Alison P. Lee[14], Tetsuya Nakamura[2], Domitille Chalopin[15], Shaohua Fan[16,34], Dustin Wcisel[17,18], Cristian Cañestro[19,20], Jason Sydes[1], Felix E. G. Beaudry[21], Yi Sun[22,23], Jana Hertel[24,34], Michael J. Beam[1], Mario Fasold[24,34], Mikio Ishiyama[25], Jeremy Johnson[8], Steffi Kehr[24], Marcia Lara[8], John H. Letaw[1], Gary W. Litman[26], Ronda T. Litman[26], Masato Mikami[27], Tatsuya Ota[28], Nil Ratan Saha[29], Louise Williams[8], Peter F. Stadler[24], Han Wang[22,23], John S. Taylor[21], Quenton Fontenot[30], Allyse Ferrara[30], Stephen M. J. Searle[10], Bronwen Aken[10,11], Mark Yandell[9], Igor Schneider[31], Jeffrey A. Yoder[17,18], Jean-Nicolas Volff[15], Axel Meyer[16,32], Chris T. Amemiya[29], Byrappa Venkatesh[14], Peter W. H. Holland[12], Yann Guiguen[6], Julien Bobe[6], Neil H. Shubin[2], Federica Di Palma[8,34], Jessica Alföldi[8], Kerstin Lindblad-Toh[8,33], and John H. Postlethwait[1]

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Affiliations

[1]Institute of Neuroscience, University of Oregon, Eugene, Oregon, USA.

[2]Department of Organismal Biology and Anatomy, The University of Chicago, Chicago, Illinois, USA.

[3]Department of Biology, University of Kentucky, Lexington, Kentucky, USA.

[4]Department of Anthropology, Pennsylvania State University, University Park, Pennsylvania, USA.

[5]Institute of Marine Biology, Biotechnology and Aquaculture, Hellenic Centre for Marine Research, Heraklion, Greece.

[6]INRA, UR1037 LPGP Fish Physiology and Genomics, Campus de Beaulieu, Rennes, France.

[7]Department of Animal Biology, University of Illinois, Urbana-Champaign, Illinois, USA.

[8]Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA.

[9]Eccles Institute of Human Genetics, University of Utah, Salt Lake City, Utah, USA.

[10]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom.

[11]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom.

[12]Department of Zoology, University of Oxford, Oxford, United Kingdom.

[13]School of Biological Sciences, Bangor University, Bangor, Gwynedd, United Kingdom.

[14]Comparative Genomics Laboratory, Institute of Molecular and Cell Biology, A*STAR, Biopolis, Singapore.

[15]Institut de Génomique Fonctionnelle de Lyon, Ecole Normale Supérieure de Lyon, Lyon, France.

[16]Department of Biology, University of Konstanz, Konstanz, Germany.

[17]Department of Molecular Biomedical Sciences, North Carolina State University, Raleigh, North Carolina, USA.

[18]Center for Comparative Medicine and Translational Research, North Carolina State University, Raleigh, North Carolina, USA.

[19]Departament de Genètica, Universitat de Barcelona, Barcelona, Spain.

[20]Institut de Recerca de la Biodiversitat, Universitat de Barcelona, Barcelona, Spain.

[21]University of Victoria, Department of Biology, Victoria, British Columbia, Canada.

[22]Center for Circadian Clocks, Soochow University, Suzhou, Jiangsu, China.

[23]School of Biology & Basic Medical Sciences, Medical College, Soochow University, Suzhou, Jiangsu, China.

[24]Bioinformatics Group, Department of Computer Science, Universität Leipzig, Leipzig, Germany.

[25]Department of Dental Hygiene, The Nippon Dental University College at Niigata, Niigata, Japan.

[26]Department of Pediatrics, University of South Florida Morsani College of Medicine, St. Petersburg, Florida, USA.

[27]Department of Microbiology, The Nippon Dental University School of Life Dentistry at Niigata, Niigata, Japan.

[28]Department of Evolutionary Studies of Biosystems, SOKENDAI (The Graduate University for Advanced Studies), Hayama, Japan.

[29]Molecular Genetics Program, Benaroya Research Institute, Seattle, Washington, USA.

[30]Department of Biological Sciences, Nicholls State University, Thibodaux, Louisiana, USA.

[31]Instituto de Ciencias Biologicas, Universidade Federal do Para, Belem, Brazil.

[32]International Max-Planck Research School for Organismal Biology, University of Konstanz, Konstanz, Germany.

[33]Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden.

ACKNOWLEDGEMENTS

We thank the Broad Institute Genomics Platform for constructing and sequencing gar DNA and RNA libraries and Jason Turner-Maier for the gar transcriptome assembly. We thank the teams of the Bayousphere Research Lab (Nicholls State University) and the University of Oregon Fish Facility for gar work and husbandry. We thank John Westlund for design of species illustrations. Generation of gar sequences and assemblies by the Broad Institute of the Massachusetts Institute of Technology (MIT) and Harvard University was supported by NIH/NHGRI grant U54 HG03067. This work was further supported by NIH grants R01 OD011116 (alias R01 RR020833) and R24 OD01119004 (J.H.P.); Feodor Lynen Fellowship, Alexander von Humboldt Foundation and Volkswagen Foundation Initiative Evolutionary Biology, grant I/84 815 (I.B.); NIH Grant T32 HD055164 and NSF Doctoral Dissertation Improvement Grant 1311436 (A.R.G); Uehara Memorial Foundation Research Fellowship 2013, Japan Society for the Promotion of Science Postdoctoral Research Fellowship 2012-127, and the Marine Biological Laboratory Research Award 2014 (T.N.); Brazilian National Council for Scientific and Technological Development (CNPq) Grants 402754/2012-3 and 477658/2012-1 (I.S.); The Brinson Foundation and the University of Chicago Biological Sciences Division (N.H.S.); NSF grant BCS0725227 (K.K.); Call “ARISTEIA I” of the National Strategic Reference Framework 2007–2013 (SPARCOMP, #36), Ministry of Education and Religious Affairs of Greece (T.M.); ANR grant ANR-10-GENM-017 (PhyloFish; J.B.); Wellcome Trust (grants WT095908 and WT098051) and the European Molecular Biology Laboratory (D.B., S.M.J.S., B.A.); Biomedical Research Council of Agency for Science, Technology and Research (A*STAR), Singapore (B.V.); European Research Council grant 268513 (P.W.H.H., K.J.M.); Ministerio de Ciencia e Innovación: BFU2010-14875 and BFU2015-71340; Generalitat de Catalunya, AGAUR: SGR2014-290 (C.C.); NIH R01 AI057559 (G.W.L., J.A.Y.); NIH R24 OD010922 and R01 GM079492 (C.T.A.); and National Basic Research Program of China (973 Program) (# 2012CB947600) and the National Natural Science Foundation of China (NSFC) (#31030062) (H.W.).

Europe PMC Funders Author Manuscripts

AUTHOR CONTRIBUTIONS

J.H.P., I.B., J.A. and K.L.T. planned and oversaw the project. J.J. was in charge of genome sequencing management, F.D.P. in charge of overall project management and coordination. Q.F. and A.F. provided gar and bowfin samples for genome and transcriptome sequencing. M.L. prepared DNA for genome sequencing; L.W. prepared libraries for genome sequencing. A.M.B. performed genome assembly and anchoring. A.A. and J.C. developed the gar genetic map. A.A. prepared gar RNA for transcriptome sequencing and assembly by the Broad Institute Genomics Platform. J.P., Y.G., and J.B. generated PhyloFish RNA-seq transcriptomes of gar, bowfin, medaka, and zebrafish. M.S.C., M.Y., D.B., S.M.J.S., and B.A. annotated the genome. D.C., S.F., J.-N.V., and A.M. analyzed transposable elements. T.M. and A.M. performed phylogenomic and gene relative rate analyses. A.A. generated karyotype data. J.J.S., I.B., J.S., J.H.L., J.C., and J.H.P. analyzed conserved synteny data. K.J.M. and P.W.H.H. analyzed *Hox* genes, J.F.M. *ParaHox* genes; C.C. *Aldh* genes; Y.S. and H.W. clock genes; F.E.G.B. and J.S.T. *opsin* genes; D.W., G.W.L., R.T.L, T.O., N.R.S., C.T.A., J.H.P., and J.A.Y. immune genes. K.K., M.I., M.M., P.B., and I.B. carried out the annotation and expression analysis of mineralization genes. T.D., M.J.B., P.B., J.S., and J.H.P. annotated and analyzed miRNA genes based on small RNA-seq data generated by T.D. (main text and Supplementary Note 11.2). J.H., M.F., S.K., and P.F.S. studied miRNAs *in silico* (Supplementary Note 11.1). V.R., A.P.L., and B.V. carried out CNE analyses for developmental gene loci. I.B., P.B., J.S., and J.H.P. performed whole genome alignments and global CNE analyses. I.B. analyzed limb enhancer evolution. A.R.G., T.N., I.S., and N.H.S. analyzed *HoxD* enhancer functions. J.P., I.B., P.B., J.H.P., Y.G., and J.B. performed comparative gene expression analysis of gar, medaka, and zebrafish. I.B. and J.H.P wrote the paper with input from other authors.

URLs

GenBank spotted gar genome assembly GCA_000242695.1,http://www.ncbi.nlm.nih.gov/assembly/GCA_000242695.1/; spotted gar genome at Ensembl,http://www.ensembl.org/Lepisosteus_oculatus/Info/Index; Synteny Database,http://syntenydb.uoregon.edu/synteny_db/; PhyloFish Portal,http://phylofish.sigenae.org/index.html; Prost! (PRocessing Of Small Transcripts),http://dx.doi.org/10.5281/zenodo.35461; RepeatMasker, http://www.repeatmasker.org.

REFERENCES

1. Nelson , J S. Fishes of the world. 4. John Wiley; 2006.
2. Kettleborough RN, et al. A systematic genome-wide analysis of zebrafish protein-coding gene function. Nature. 2013; 496:494–497. [PubMed: 23594742]
3. Patton EE, Mathers ME, Schartl M. Generating and analyzing fish models of melanoma. Methods in cell biology. 2011; 105:339–366. [PubMed: 21951537]
4. Reitzel AM, et al. Genetic variation at aryl hydrocarbon receptor (AHR) loci in populations of Atlantic killifish (Fundulus heteroclitus) inhabiting polluted and reference habitats. BMC evolutionary biology. 2014; 14:6. [PubMed: 24422594]
5. Lee O, Green JM, Tyler CR. Transgenic fish systems and their application in ecotoxicology. Crit Rev Toxicol. 2015; 45:124–141. [PubMed: 25394772]

6. Albertson RC, Cresko W, Detrich HW 3rd, Postlethwait JH. Evolutionary mutant models for human disease. Trends Genet. 2009; 25:74–81. [PubMed: 19108930]
7. Harel I, et al. A platform for rapid exploration of aging and diseases in a naturally short-lived vertebrate. Cell. 2015; 160:1013–1026. [PubMed: 25684364]
8. Pagan AJ, Ramakrishnan L. Immunity and Immunopathology in the Tuberculous Granuloma. Cold Spring Harbor perspectives in medicine. 2014; 5:a018499. [PubMed: 25377142]
9. Hagedorn EJ, Durand EM, Fast EM, Zon LI. Getting more for your marrow: boosting hematopoietic stem cell numbers with PGE2. Experimental cell research. 2014; 329:220–226. [PubMed: 25094063]
10. Dehal P, Boore JL. Two rounds of whole genome duplication in the ancestral vertebrate. PLoS Biol. 2005; 3:e314. [PubMed: 16128622]
11. Smith JJ, Keinath MC. The sea lamprey meiotic map improves resolution of ancient vertebrate genome duplications. Genome research. 2015; 25:1081–1090. [PubMed: 26048246]
12. Frankenberg SR, et al. The POU-er of gene nomenclature. Development. 2014; 141:2921–2923. [PubMed: 25053425]
13. Braasch I, et al. Connectivity of vertebrate genomes: Paired-related homeobox (Prrx) genes in spotted gar, basal teleosts, and tetrapods. Comparative biochemistry and physiology. Toxicology & pharmacology : CBP. 2014; 163:24–36. [PubMed: 24486528]
14. Amores A, et al. Zebrafish hox clusters and vertebrate genome evolution. Science. 1998; 282:1711–1714. [PubMed: 9831563]
15. Taylor J, Braasch I, Frickey T, Meyer A, Van De Peer Y. Genome duplication, a trait shared by 22,000 species of ray-finned fish. Genome Res. 2003; 13:382–390. [PubMed: 12618368]
16. Amemiya CT, et al. The African coelacanth genome provides insights into tetrapod evolution. Nature. 2013; 496:311–316. [PubMed: 23598338]
17. Venkatesh B, et al. Elephant shark genome provides unique insights into gnathostome evolution. Nature. 2014; 505:174–179. [PubMed: 24402279]
18. Hoegg S, Brinkmann H, Taylor JS, Meyer A. Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish. J Mol Evol. 2004; 59:190–203. [PubMed: 15486693]
19. Amores A, Catchen J, Ferrara A, Fontenot Q, Postlethwait JH. Genome evolution and meiotic maps by massively parallel DNA sequencing: spotted gar, an outgroup for the teleost genome duplication. Genetics. 2011; 188:799–808. [PubMed: 21828280]
20. Gnerre S, et al. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proceedings of the National Academy of Sciences of the United States of America. 2011; 108:1513–1518. [PubMed: 21187386]
21. Braasch I, et al. A new model army: Emerging fish models to study the genomics of vertebrate Evo-Devo. J Exp Zool B Mol Dev Evol. 2015; 324:316–341. [PubMed: 25111899]
22. Holt C, Yandell M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics. 2011; 12:491. [PubMed: 22192575]
23. Chalopin D, Naville M, Plard F, Galiana D, Volff JN. Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates. Genome Biol Evol. 2015; 7:567–580. [PubMed: 25577199]
24. Near TJ, et al. Resolution of ray-finned fish phylogeny and timing of diversification. Proceedings of the National Academy of Sciences of the United States of America. 2012; 109:13698–13703. [PubMed: 22869754]
25. Betancur RR, et al. The tree of life and a new classification of bony fishes. PLoS Curr. 2013; 5doi: 10.1371/currents.tol.53ba26640df0ccaee75bb165c8c26288
26. Broughton RE, Betancur RR, Li C, Arratia G, Orti G. Multi-locus phylogenetic analysis reveals the pattern and tempo of bony fish evolution. PLoS Curr. 2013; 5doi: 10.1371/currents.tol.2ca8041495ffafd0c92756e75247483e
27. Faircloth BC, Sorenson L, Santini F, Alfaro ME. A Phylogenomic Perspective on the Radiation of Ray-Finned Fishes Based upon Targeted Sequencing of Ultraconserved Elements (UCEs). PLoS One. 2013; 8:e65923. [PubMed: 23824177]

28. Grande L. An Empirical Synthetic Pattern Study of Gars (Lepisosteiformes) and Closely Related Species, Based Mostly on Skeletal Anatomy. The Resurrection of Holostei. Copeia. 2010:1–863.
29. Sallan LC. Major issues in the origins of ray-finned fish (Actinopterygii) biodiversity. Biol Rev Camb Philos Soc. 2014; 89:950–971. [PubMed: 24612207]
30. Darwin, C. On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life. John Murray; 1859.
31. Rabosky DL, et al. Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. Nat Commun. 2013; 4:1958. [PubMed: 23739623]
32. Nikaido M, et al. Coelacanth genomes reveal signatures for evolutionary transition from water to land. Genome research. 2013; 23:1740–1748. [PubMed: 23878157]
33. Ravi V, Venkatesh B. Rapidly evolving fish genomes and teleost diversity. Curr Opin Genet Dev. 2008; 18:544–550. [PubMed: 19095434]
34. Ohno S, et al. Microchromosomes in holocephalian, chondrostean and holostean fishes. Chromosoma. 1969; 26:35–40. [PubMed: 5799423]
35. Bogart JP, Balon EK, Bruton MN. The chromosomes of the living coelacanth and their remarkable similarity to those of one of the most ancient frogs. J Hered. 1994; 85:322–325. [PubMed: 7930502]
36. Nakatani Y, Takeda H, Kohara Y, Morishita S. Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates. Genome research. 2007; 17:1254–1265. [PubMed: 17652425]
37. Naruse K, et al. A medaka gene map: the trace of ancestral vertebrate proto chromosomes revealed by comparative gene mapping. Genome research. 2004; 14:820–828. [PubMed: 15078856]
38. Wolfe K. Robustness--it's not where you think it is. Nat Genet. 2000; 25:3–4. [PubMed: 10802639]
39. Postlethwait JH. The zebrafish genome in context: ohnologs gone missing. J Exp Zoolog B Mol Dev Evol. 2007; 308:563–577.
40. Crow KD, Smith CD, Cheng JF, Wagner GP, Amemiya CT. An independent genome duplication inferred from Hox paralogs in the American paddlefish--a representative basal ray-finned fish and important comparative reference. Genome Biol Evol. 2012; 4:937–953. [PubMed: 22851613]
41. Kurosawa G, et al. Organization and structure of hox gene loci in medaka genome and comparison with those of pufferfish and zebrafish genomes. Gene. 2006; 370:75–82. [PubMed: 16472944]
42. Mulley JF, Holland PW. Parallel retention of Pdx2 genes in cartilaginous fish and coelacanths. Molecular biology and evolution. 2010; 27:2386–2391. [PubMed: 20463047]
43. Duboule D. Vertebrate hox gene regulation: clustering and/or colinearity? Curr Opin Genet Dev. 1998; 8:514–518. [PubMed: 9794816]
44. Duester G. Retinoid signaling in control of progenitor cell differentiation during mouse development. Semin Cell Dev Biol. 2013; 24:694–700. [PubMed: 23973941]
45. Canestro C, Catchen JM, Rodriguez-Mari A, Yokoi H, Postlethwait JH. Consequences of lineage-specific gene loss on functional evolution of surviving paralogs: ALDH1A and retinoic acid signaling in vertebrate genomes. PLoS Genet. 2009; 5:e1000496. [PubMed: 19478994]
46. Wang H. Comparative analysis of period genes in teleost fish genomes. J Mol Evol. 2008; 67:29–40. [PubMed: 18535754]
47. Rennison DJ, Owens GL, Taylor JS. Opsin gene duplication and divergence in ray-finned fish. Mol Phylogenet Evol. 2012; 62:986–1008. [PubMed: 22178363]
48. Toloza-Villalobos J, Arroyo JI, Opazo JC. The circadian clock of teleost fish: a comparative analysis reveals distinct fates for duplicated genes. J Mol Evol. 2015; 80:57–64. [PubMed: 25487517]
49. Lagman D, et al. 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 evolutionary biology. 2013; 13:238. [PubMed: 24180662]
50. Mano H, Kojima D, Fukada Y. Exo-rhodopsin: a novel rhodopsin expressed in the zebrafish pineal gland. Brain Res Mol Brain Res. 1999; 73:110–118. [PubMed: 10581404]

51. Bingulac-Popovic J, et al. Mapping of Mhc class I and class II regions to different linkage groups in the zebrafish, *Danio rerio*. Immunogenetics. 1997; 46:129–134. [PubMed: 9162099]
52. Sato A, et al. Nonlinkage of major histocompatibility complex class I and class II loci in bony fishes. Immunogenetics. 2000; 51:108–116. [PubMed: 10663573]
53. Dijkstra JM, Grimholt U, Leong J, Koop BF, Hashimoto K. Comprehensive analysis of MHC class II genes in teleost fish genomes reveals dispensability of the peptide-loading DM system in a large part of vertebrates. BMC evolutionary biology. 2013; 13:260. [PubMed: 24279922]
54. Grimholt U, et al. A comprehensive analysis of teleost MHC class I sequences. BMC evolutionary biology. 2015; 15:32. [PubMed: 25888517]
55. Dirscherl H, McConnell SC, Yoder JA, de Jong JL. The MHC class I genes of zebrafish. Developmental and comparative immunology. 2014; 46:11–23. [PubMed: 24631581]
56. Dirscherl H, Yoder JA. Characterization of the Z lineage Major histocompatability complex class I genes in zebrafish. Immunogenetics. 2014; 66:185–198. [PubMed: 24287892]
57. Flajnik MF, Kasahara M. Origin and evolution of the adaptive immune system: genetic events and selective pressures. Nat Rev Genet. 2010; 11:47–59. [PubMed: 19997068]
58. Danilova N, Bussmann J, Jekosch K, Steiner LA. The immunoglobulin heavy-chain locus in zebrafish: identification and expression of a previously unknown isotype, immunoglobulin Z. Nat Immunol. 2005; 6:295–302. [PubMed: 15685175]
59. Hansen JD, Landis ED, Phillips RB. Discovery of a unique Ig heavy-chain isotype (IgT) in rainbow trout: Implications for a distinctive B cell developmental pathway in teleost fish. Proceedings of the National Academy of Sciences of the United States of America. 2005; 102:6919–6924. [PubMed: 15863615]
60. Zhang YA, et al. IgT, a primitive immunoglobulin class specialized in mucosal immunity. Nat Immunol. 2010; 11:827–835. [PubMed: 20676094]
61. Parra ZE, Ohta Y, Criscitiello MF, Flajnik MF, Miller RD. The dynamic TCRdelta: TCRdelta chains in the amphibian Xenopus tropicalis utilize antibody-like V genes. Eur J Immunol. 2010; 40:2319–2329. [PubMed: 20486124]
62. Roach JC, et al. The evolution of vertebrate Toll-like receptors. Proceedings of the National Academy of Sciences of the United States of America. 2005; 102:9577–9582. [PubMed: 15976025]
63. Cannon JP, et al. A bony fish immunological receptor of the NITR multigene family mediates allogeneic recognition. Immunity. 2008; 29:228–237. [PubMed: 18674935]
64. Yoder JA. Form, function and phylogenetics of NITRs in bony fish. Developmental and comparative immunology. 2009; 33:135–144. [PubMed: 18840463]
65. Kawasaki K, Suzuki T, Weiss KM. Phenogenetic drift in evolution: the changing genetic basis of vertebrate teeth. Proceedings of the National Academy of Sciences of the United States of America. 2005; 102:18063–18068. [PubMed: 16332957]
66. Kawasaki K. The SCPP gene repertoire in bony vertebrates and graded differences in mineralized tissues. Dev Genes Evol. 2009; 219:147–157. [PubMed: 19255778]
67. Ryll B, Sanchez S, Haitina T, Tafforeau P, Ahlberg PE. The genome of Callorhinchus and the fossil record: a new perspective on SCPP gene evolution in gnathostomes. Evol Dev. 2014; 16:123–124. [PubMed: 24712871]
68. Kawasaki K, Amemiya CT. SCPP genes in the coelacanth: tissue mineralization genes shared by sarcopterygians. J Exp Zool B Mol Dev Evol. 2014; 322:390–402. [PubMed: 25243252]
69. Sire JY. Light and TEM study of nonregenerated and experimentally regenerated scales of Lepisosteus oculatus (Holostei) with particular attention to ganoine formation. Anat Rec. 1994; 240:189–207. [PubMed: 7992885]
70. Sasagawa I, Ishiyama M, Yokosuka H, Mikami M. Fine structure and development of the collar enamel in gars, Lepisosteus oculatus, Actinopterygii. Frontiers of Materials Science in China. 2008; 2:134–142.
71. Zhu M, et al. The oldest articulated osteichthyan reveals mosaic gnathostome characters. Nature. 2009; 458:469–474. [PubMed: 19325627]
72. Hertel J, Stadler PF. The Expansion of Animal MicroRNA Families Revisited. Life (Basel). 2015; 5:905–920. [PubMed: 25780960]

73. Desvignes T, Beam MJ, Batzel P, Sydes J, Postlethwait JH. Expanding the annotation of zebrafish microRNAs based on small RNA sequencing. Gene. 2014; 546:386–389. [PubMed: 24835514]
74. Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res. 2011; 39:D152–157. [PubMed: 21037258]
75. Braasch, I.; Postlethwait, JH. Ch. 17. In: Soltis, PS.; Soltis, DE., editors. Polyploidy and Genome Evolution. Springer; 2012. p. 341-383.
76. Loh YH, Yi SV, Streelman JT. Evolution of microRNAs and the diversification of species. Genome Biol Evol. 2011; 3:55–65. [PubMed: 21169229]
77. Grimson A, et al. Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. Nature. 2008; 455:1193–1197. [PubMed: 18830242]
78. Berezikov E, et al. Deep annotation of Drosophila melanogaster microRNAs yields insights into their processing, modification, and emergence. Genome research. 2011; 21:203–215. [PubMed: 21177969]
79. Wheeler BM, et al. The deep evolution of metazoan microRNAs. Evol Dev. 2009; 11:50–68. [PubMed: 19196333]
80. Woolfe A, et al. Highly conserved non-coding sequences are associated with vertebrate development. PLoS Biol. 2005; 3:e7. [PubMed: 15630479]
81. Pennacchio LA, et al. In vivo enhancer analysis of human conserved non-coding sequences. Nature. 2006; 444:499–502. [PubMed: 17086198]
82. Montavon T, et al. A regulatory archipelago controls Hox genes transcription in digits. Cell. 2011; 147:1132–1145. [PubMed: 22118467]
83. Berlivet S, et al. Clustering of tissue-specific sub-TADs accompanies the regulation of HoxA genes in developing limbs. PLoS Genet. 2013; 9:e1004018. [PubMed: 24385922]
84. Visel A, Minovitsky S, Dubchak I, Pennacchio LA. VISTA Enhancer Browser--a database of tissue-specific human enhancers. Nucleic Acids Res. 2007; 35:D88–92. [PubMed: 17130149]
85. Gehrke AR, et al. Deep conservation of wrist and digit enhancers in fish. Proceedings of the National Academy of Sciences of the United States of America. 2015; 112:803–808. [PubMed: 25535365]
86. Zakany J, Duboule D. The role of Hox genes during vertebrate limb development. Curr Opin Genet Dev. 2007; 17:359–366. [PubMed: 17644373]
87. Schneider I, Shubin NH. The origin of the tetrapod limb: from expeditions to enhancers. Trends Genet. 2013; 29:419–426. [PubMed: 23434323]
88. Andrey G, et al. A switch between topological domains underlies HoxD genes collinearity in mouse limbs. Science. 2013; 340:1234167. [PubMed: 23744951]
89. Force A, et al. Preservation of duplicate genes by complementary, degenerative mutations. Genetics. 1999; 151:1531–1545. [PubMed: 10101175]
90. Stoltzfus A. On the possibility of constructive neutral evolution. J Mol Evol. 1999; 49:169–181. [PubMed: 10441669]
91. Postlethwait J, Amores A, Cresko W, Singer A, Yan YL. Subfunction partitioning, the teleost radiation and the annotation of the human genome. Trends Genet. 2004; 20:481–490. [PubMed: 15363902]
92. He X, Zhang J. Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. Genetics. 2005; 169:1157–1164. [PubMed: 15654095]
93. Vilella AJ, et al. EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. Genome research. 2009; 19:327–335. [PubMed: 19029536]
94. Catchen JM, Conery JS, Postlethwait JH. Automated identification of conserved synteny after whole-genome duplication. Genome research. 2009; 19:1497–1505. [PubMed: 19465509]
95. Ohno, S. Evolution by Gene Duplication. Springer-Verlag; 1970.
96. Scannell DR, Wolfe KH. A burst of protein sequence evolution and a prolonged period of asymmetric evolution follow gene duplication in yeast. Genome research. 2008; 18:137–147. [PubMed: 18025270]

97. Chang N, et al. Genome editing with RNA-guided Cas9 nuclease in zebrafish embryos. Cell Res. 2013; 23:465–472. [PubMed: 23528705]

98. Hwang WY, et al. Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat Biotechnol. 2013; 31:227–229. [PubMed: 23360964]

Methods-only references

99. Long WL, Ballard WW. Normal embryonic stages of the longnose gar, Lepisosteus osseus. BMC Dev Biol. 2001; 1:6. [PubMed: 11319037]

100. Grabherr MG, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011; 29:644–652. [PubMed: 21572440]

101. Schulz MH, Zerbino DR, Vingron M, Birney E. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics. 2012; 28:1086–1092. [PubMed: 22368243]

102. Wicker T, et al. A unified classification system for eukaryotic transposable elements. Nat Rev Genet. 2007; 8:973–982. [PubMed: 17984973]

103. Shaffer HB, et al. The western painted turtle genome, a model for the evolution of extreme physiological adaptations in a slowly evolving lineage. Genome biology. 2013; 14:R28. [PubMed: 23537068]

104. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014; 30:1312–1313. [PubMed: 24451623]

105. Lartillot N, Rodrigue N, Stubbs D, Richer J. PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. Syst Biol. 2013; 62:611–615. [PubMed: 23564032]

106. Tajima F. Simple methods for testing the molecular evolutionary clock hypothesis. Genetics. 1993; 135:599–607. [PubMed: 8244016]

107. Takezaki N, Rzhetsky A, Nei M. Phylogenetic test of the molecular clock and linearized trees. Molecular biology and evolution. 1995; 12:823–833. [PubMed: 7476128]

108. Amores A, Postlethwait JH. Banded chromosomes and the zebrafish karyotype. Methods in cell biology. 1999; 60:323–338. [PubMed: 9891345]

109. Krzywinski M, et al. Circos: an information aesthetic for comparative genomics. Genome research. 2009; 19:1639–1645. [PubMed: 19541911]

110. Smith JJ, et al. Sequencing of the sea lamprey (Petromyzon marinus) genome provides insights into vertebrate evolution. Nat Genet. 2013; 45:415–421. [PubMed: 23435085]

111. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. miRBase: tools for microRNA genomics. Nucleic Acids Res. 2008; 36:D154–158. [PubMed: 17991681]

112. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res. 2006; 34:D140–144. [PubMed: 16381832]

113. Griffiths-Jones S. The microRNA Registry. Nucleic Acids Res. 2004; 32:D109–111. [PubMed: 14681370]

114. Lorenz R, et al. ViennaRNA Package 2.0. Algorithms Mol Biol. 2011; 6:26. [PubMed: 22115189]

115. Batzel, P.; Desvignes, T.; Sydes, J.; Eames, BF.; Postlethwait, JH. Prost!, a tool for miRNA annotation and next generation smallRNA sequencing experiment analysis. 2015.

116. Desvignes T, et al. miRNA Nomenclature: A View Incorporating Genetic Origins, Biosynthetic Pathways, and Sequence Variants. Trends Genet. 2015; 31:613–626. [PubMed: 26453491]

117. Flicek P, et al. Ensembl 2013. Nucleic Acids Res. 2013; 41:D48–55. [PubMed: 23203987]

118. Muffato M, Louis A, Poisnel CE, Roest Crollius H. Genomicus: a database and a browser to study gene synteny in modern and ancestral genomes. Bioinformatics. 2010; 26:1119–1121. [PubMed: 20185404]

119. Louis A, Muffato M, Roest Crollius H. Genomicus: five genome browsers for comparative genomics in eukaryota. Nucleic Acids Res. 2013; 41:D700–705. [PubMed: 23193262]

120. Brudno M, et al. Glocal alignment: finding rearrangements during alignment. Bioinformatics. 2003; 19(Suppl 1):i54–62. [PubMed: 12855437]

121. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. Nucleic Acids Res. 2004; 32:W273–279. [PubMed: 15215394]

122. Blanchette M, et al. Aligning multiple genomic sequences with the threaded blockset aligner. Genome research. 2004; 14:708–715. [PubMed: 15060014]

123. Harris, RS. PhD thesis. Pennsylvania State University; 2007. Improved pairwise alignment of genomic DNA.

124. Siepel A, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome research. 2005; 15:1034–1050. [PubMed: 16024819]

125. Hinrichs AS, et al. The UCSC Genome Browser Database: update 2006. Nucleic Acids Res. 2006; 34:D590–598. [PubMed: 16381938]

126. Quinlan AR. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. Curr Protoc Bioinformatics. 2014; 47:11 12 11–11 12 34. [PubMed: 25199790]

127. Fisher S, et al. Evaluating the biological relevance of putative enhancers using Tol2 transposon-mediated transgenesis in zebrafish. Nat Protoc. 2006; 1:1297–1305. [PubMed: 17406414]

128. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25:1754–1760. [PubMed: 19451168]

129. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology. 2009; 10:R25. [PubMed: 19261174]

130. Li H, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25:2078–2079. [PubMed: 19505943]

131. Anders S, Huber W. Differential expression analysis for sequence count data. Genome biology. 2010; 11:R106. [PubMed: 20979621]

132. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; Vienna, Austria: 2015.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

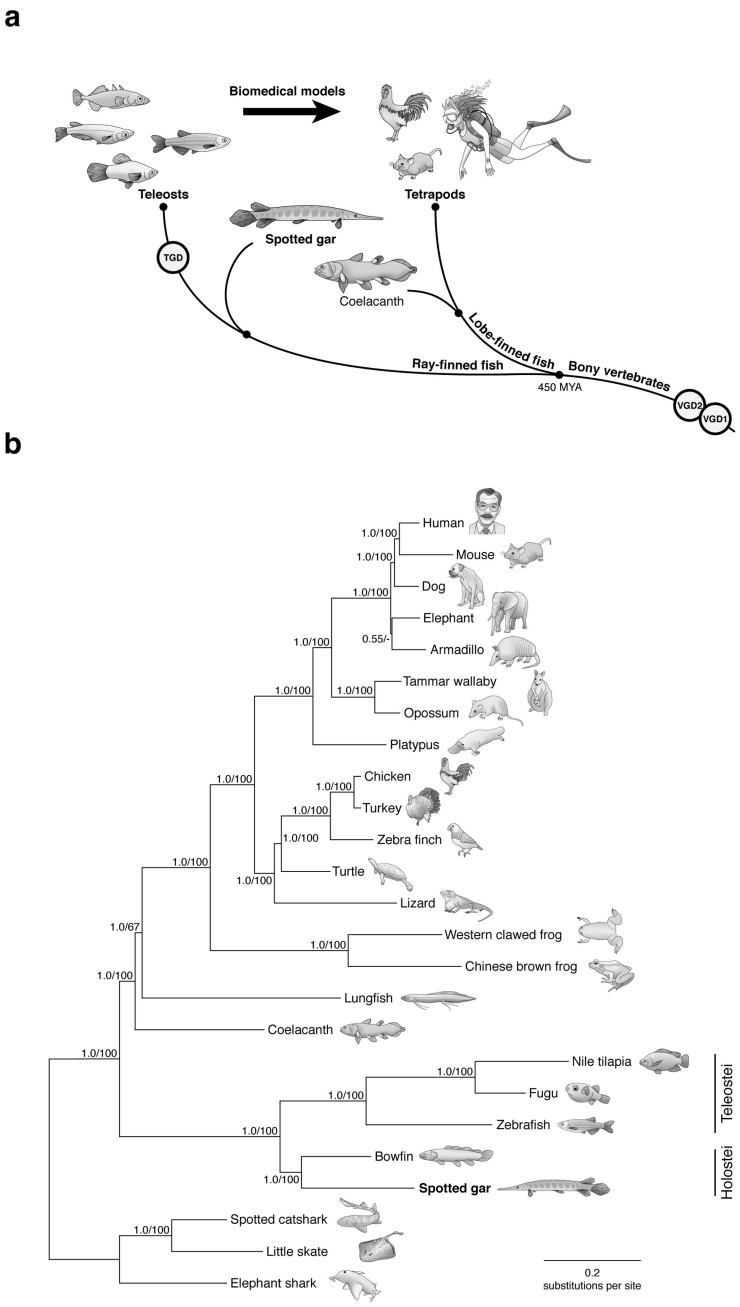

Figure 1. Spotted gar bridges vertebrate genomes

a) Spotted gar is a ray-finned fish that diverged from teleost fish, including the major biomedical models zebrafish, platyfish, medaka, and stickleback, before the teleost genome duplication (TGD). Gar connects teleosts to lobe-finned vertebrates, such as coelacanth and tetrapods, including human, by clarifying evolution after two earlier rounds of vertebrate genome duplication (VGD1, VGD2) that occurred before the divergence of ray-finned and lobe-finned fish 450 million years ago (MYA). b) Bayesian phylogeny inferred from an alignment of 97,794 amino acid site positions from 243 proteins with one-to-one orthology

ratio from 25 jawed (gnathostome) vertebrates using PhyloBayes under the CAT+GTR+Γ4 model and rooted on cartilaginous fish. Node support is shown as posterior probability and bootstrap support from maximum likelihood analysis (Supplementary Fig. 6). The tree shows the monophyly and slow evolution of Holostei (gar plus bowfin) compared to their sister lineage, the teleosts (Teleostei). See also Supplementary File 1 and Source Dataset 1.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

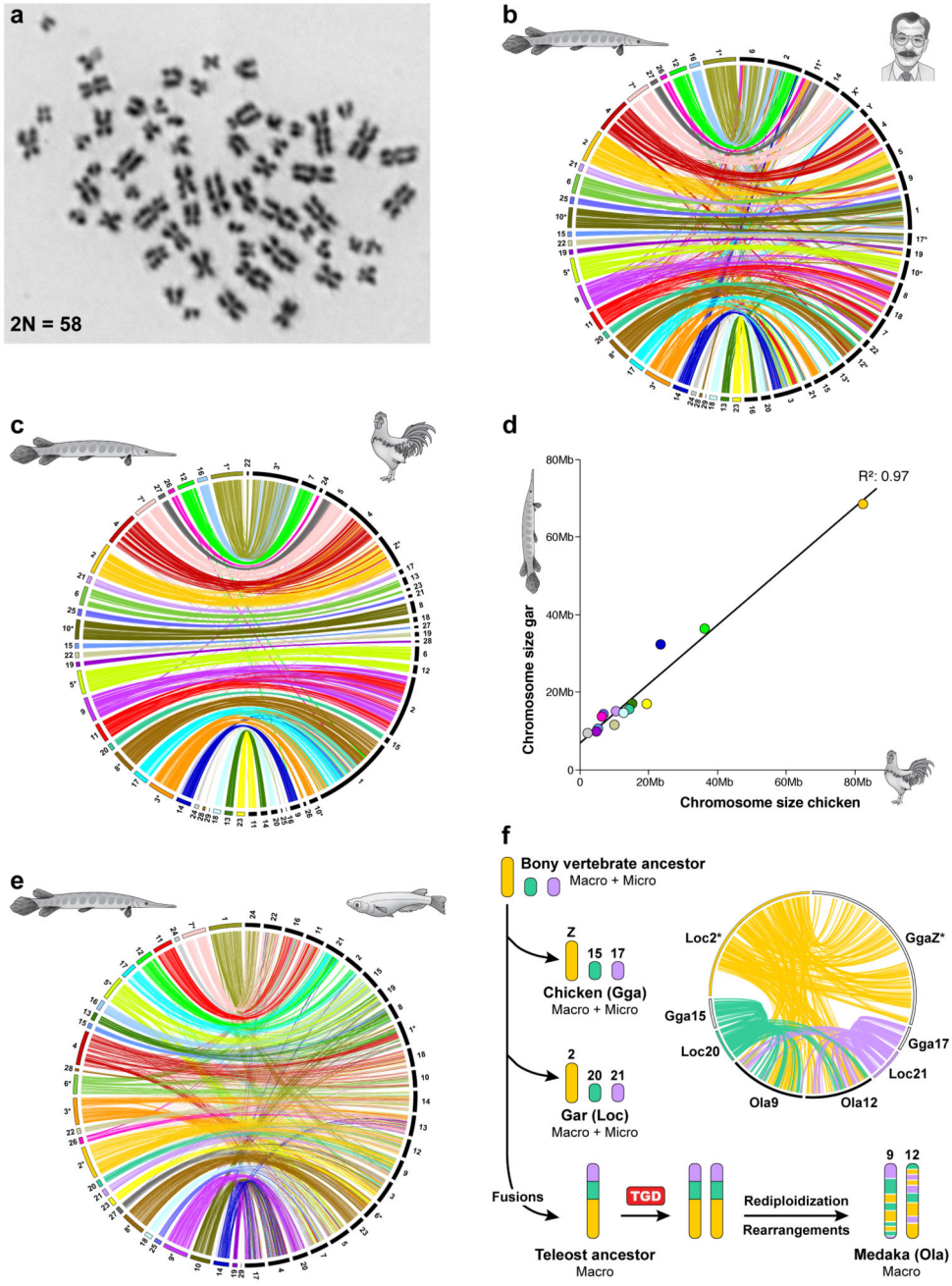

Figure 2. Spotted gar preserves ancestral genome structure

a) The spotted gar karyotype consists of macro- and microchromosomes (see Supplementary Fig. 7 for chromosome annotations). b) Circos plot[109] showing conserved synteny of gar (colored, left) vs. human (black, right) chromosomes. c) Gar vs. chicken shows strong conservation of genomes for 450 million years and one-to-one synteny conservation for many entire chromosomes, particularly microchromosomes (e.g., Loc13 and Gga14; Loc23 and Gga11, etc.). d) Assembled chromosome lengths (in megabases, Mb) for gar and chicken chromosomes with one-to-one conserved synteny are highly correlated ($R^2 = 0.97$).

e) Gar vs. medaka shows the overall one-to-two double-conserved synteny relationship of gar to a post-TGD teleost genome (e.g., gar Loc24 and Ola16/Ola11). Gar chromosomes are displayed in a different order in d compared to b/c; asterisks indicate chromosomes inverted with respect to the arbitrarily oriented reference genomes. f) Gar-chicken-medaka comparisons illuminate karyotype evolution leading to modern teleosts. The bony vertebrate ancestor genome contained both macro- and microchromosomes, some of which remain largely conserved in chicken and gar, e.g., macrochromosome Loc2/GgaZ and microchromosomes Loc20/Gga15 and Loc21/Gga17. All three chromosomes possess double conserved synteny with medaka chromosomes Ola9 and Ola12, which is explained by chromosome fusion in the lineage leading to teleosts after divergence from gar, followed by TGD duplication of the fusion chromosome and subsequent intrachromosomal rearrangements and rediploidization. Multiple examples of such pre-TGD chromosome fusions explain the absence of microchromosomes in teleosts. See Supplementary Note 8.2 and Source Dataset 2 for details.

Europe PMC Funders Author Manuscripts

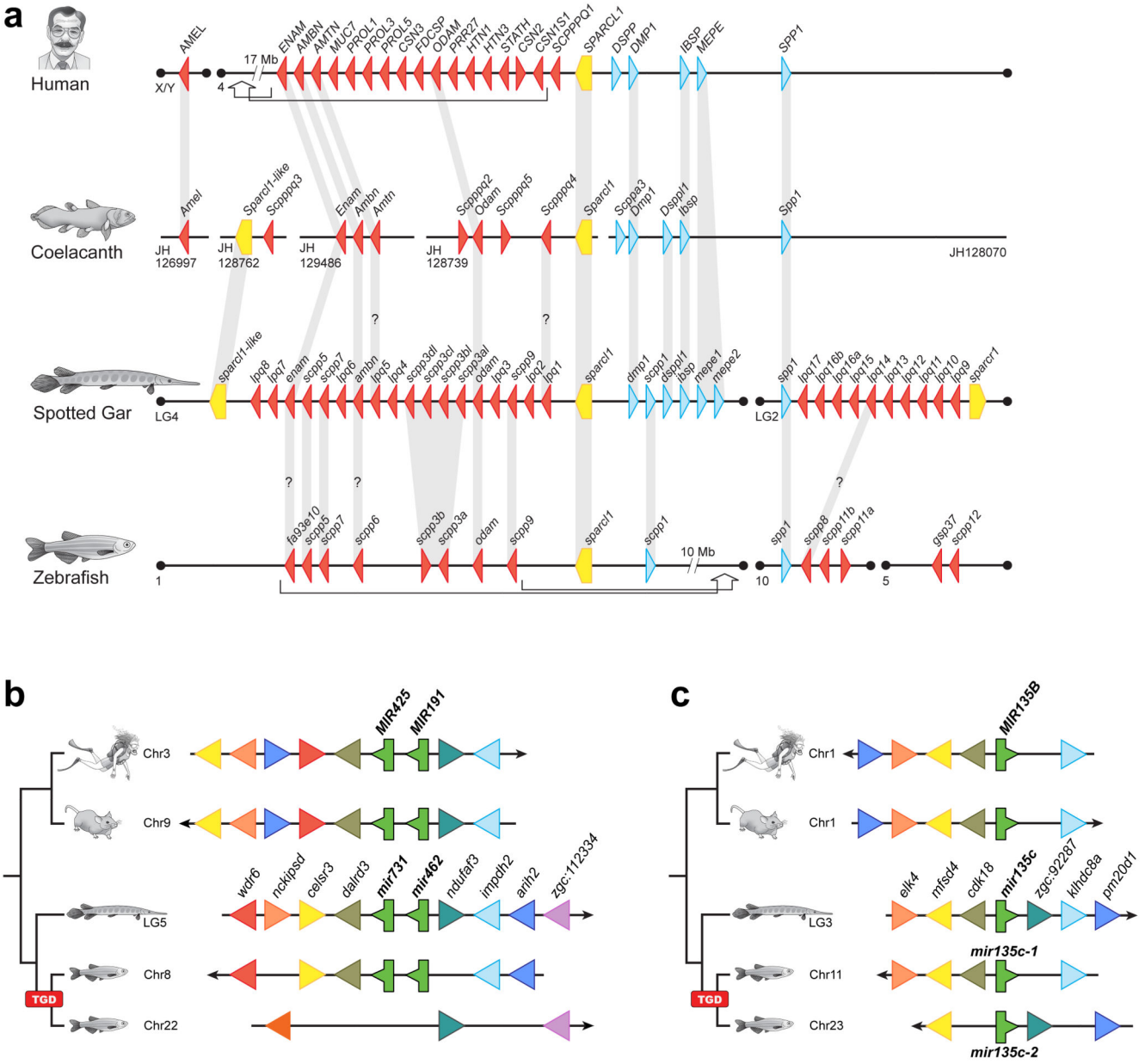

Figure 3. Gar helps connect vertebrate protein-coding and miRNA genes

a) *Scpp* gene arrangement in human, coelacanth, gar, and zebrafish including P/Q-rich (red) and acidic *Scpp* genes (blue) and *Sparc*-like genes (yellow) (Supplementary Note 10, ref.[68]). Orthologies (gray vertical bars) among lobe-finned vertebrates (e.g., human, coelacanth) and teleosts (e.g., zebrafish) had previously been limited to *Odam* and *Spp1*. Gar connects lineages through orthologs of genes previously known only from either teleosts (*scpp1*, *scpp3* genes, *scpp5*, *scpp7*, *scpp9*) or lobe-finned vertebrates (*enam*, *ambn*, *dmp1*, *dsppl1*, *ibsp*, *mepe*). Further putative orthologies supported by only short stretches of sequence similarity (‘?’) connect gar *enam*, *ambn*, and *lpq14* with zebrafish *fa93e10*, *scpp6*, and *scpp8*, respectively; gar *lpq1* and *Scpppq4* in coelacanth; and gar *lpq5* with *Amtn* in lobe-finned vertebrates. Arrows in human and zebrafish indicate intra-chromosomal

Europe PMC Funders Author Manuscripts

rearrangements separating originally clustered genes into distant chromosomal locations (distance in megabases, Mb). Conserved synteny analysis of the gar *scpp* gene cluster on LG2 suggests that the *scpp* gene regions on zebrafish chromosomes 10 and 5 are derived from the TGD (Supplementary Note 10, Supplementary Fig. 26). b) The gar 'conserved synteny bridge' (Supplementary Note 11.2) infers that the miRNA cluster of *mir731* and *mir462* on gar LG4 and zebrafish chromosome 8 and a miRNA-free region on zebrafish chromosome 2 are TGD ohnologous to the mammalian *Mir425-191* cluster. c) Gar newly connects through synteny zebrafish TGD ohnologs *mir135c-1* and *mir135c-2* with mammalian *Mir135B*. See Source Dataset 3 for genomic locations in a-c.

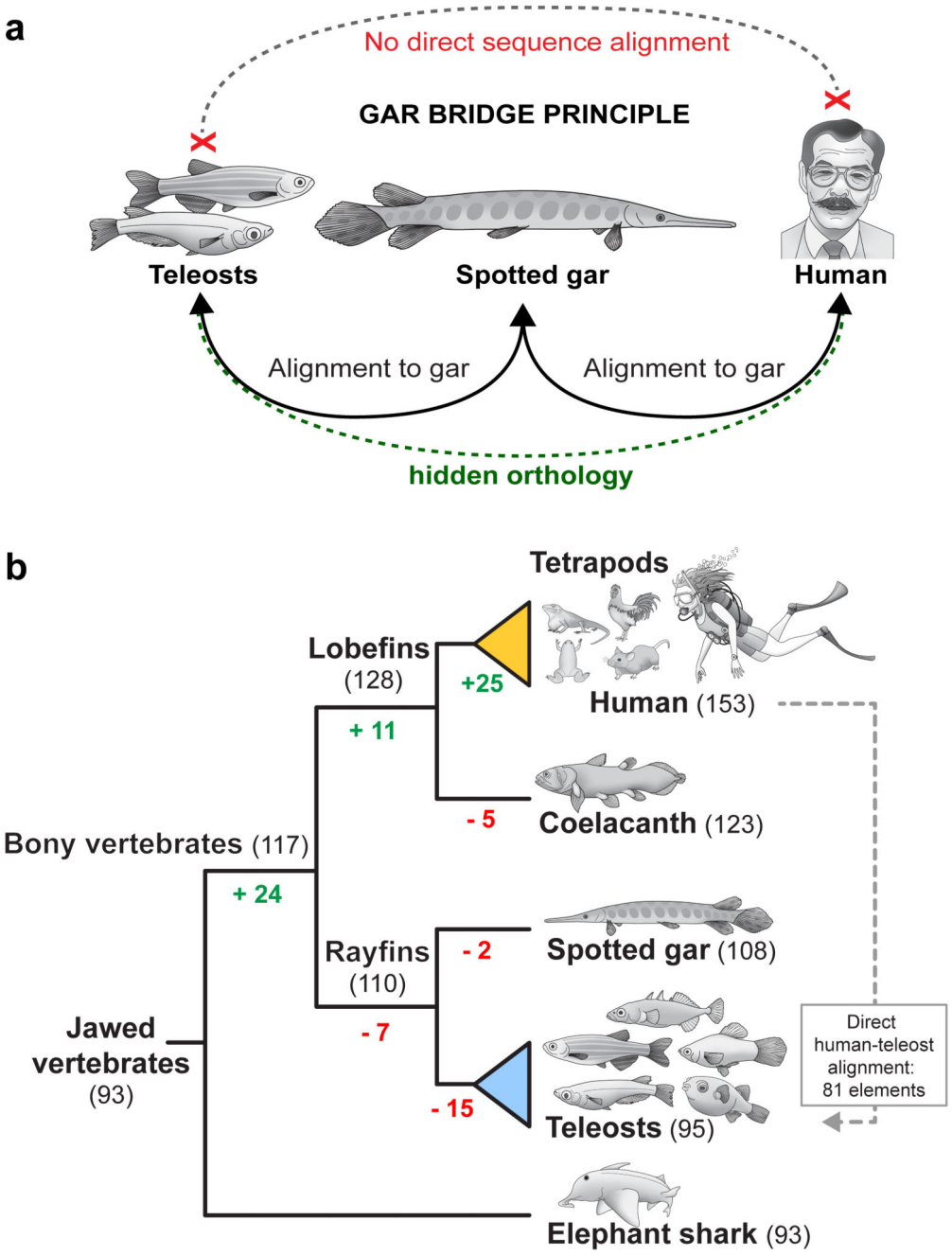

Figure 4. Gar provides connectivity of vertebrate regulatory elements

a) The ‘gar bridge principle’ of vertebrate CNE connectivity from human through gar to teleosts. Hidden orthology is revealed for elements that do not directly align between human and teleosts but become evident when first aligning tetrapod genomes to gar, and then aligning gar and teleost genomes. b) Connectivity analysis of 13-way whole-genome alignments reveals the evolutionary gain (green) and loss (red) of 153 human limb enhancers. Direct human-teleost orthology could only be established for 81 elements as

opposed to 95 when taking gar as bridge (a). See Supplementary Notes 12.2,12.3, Supplementary Tab. 22, and Supplementary Fig. 37 for details.

Europe PMC Funders Author Manuscripts

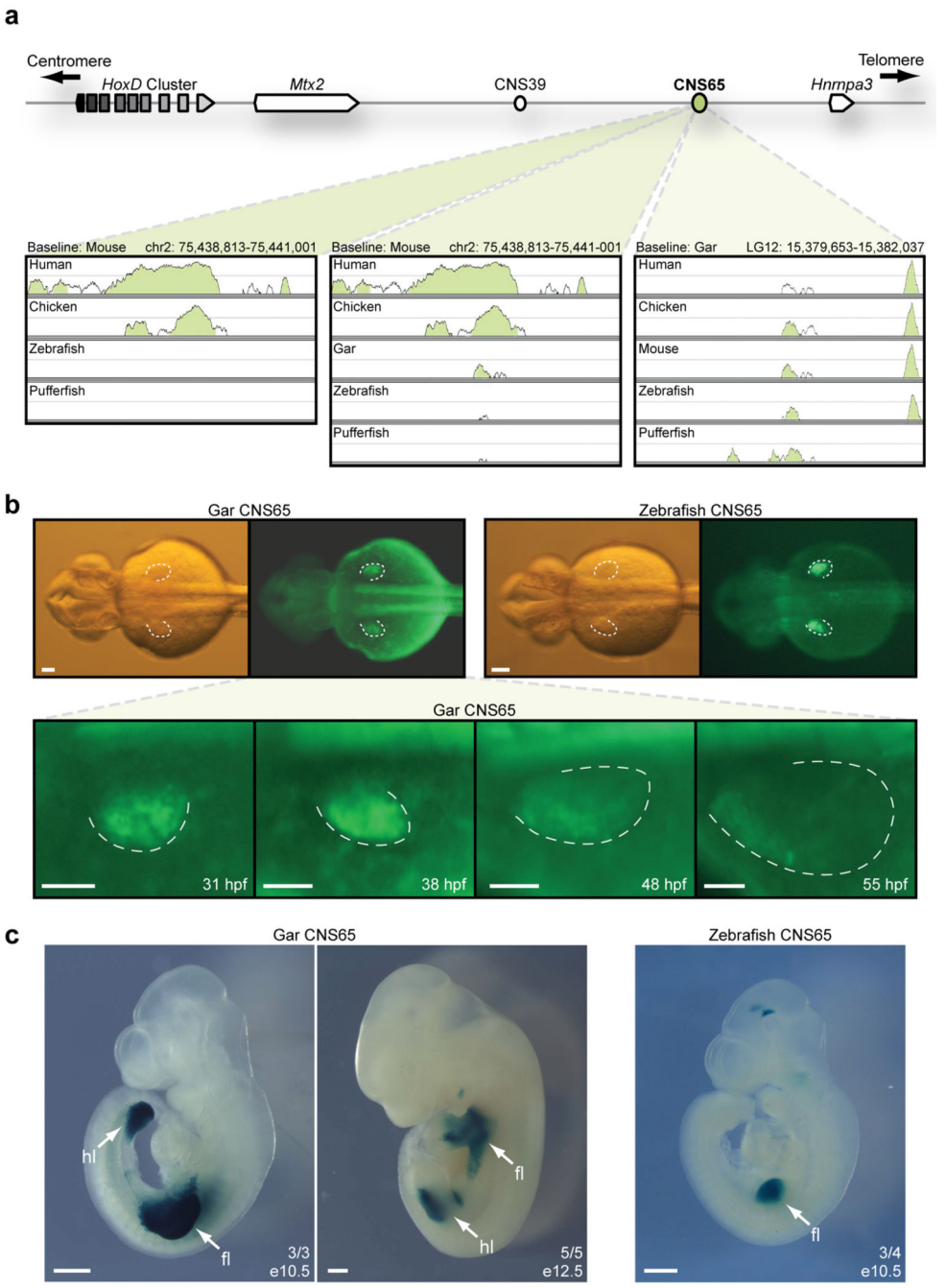

Figure 5. Identification and functional analysis of the gar and teleost early phase *HoxD* enhancer CNS65

a) Schematic of the mouse *HoxD* telomeric gene desert, which contains enhancers CNS39 and CNS65 that drive early phase *HoxD* expression in limbs (upper part). Using mouse as baseline, Vista alignments of the *HoxD* gene desert show sequence conservation with human and chicken for CNS65, but not with teleosts (zebrafish, pufferfish) (lower part, left). An alignment including gar, however, reveals a significant peak of conservation in the gar sequence (middle). Using the identified gar CNS65 as baseline revealed CNS65 orthologs in zebrafish and pufferfish (right). b) Gar (left) and zebrafish (right) CNS65 orthologs drive

robust and reproducible GFP expression in zebrafish pectoral fins at 36 hours post fertilization (hpf) (upper panel). Pectoral fin activity of gar CNS65 begins at 31 hpf, drives activity throughout the fin, and becomes deactivated around 48 hpf (lower panel). Dotted lines: distal portion of the pectoral fins. c) Gar CNS65 drives expression throughout the early mouse fore-and hindlimbs (arrows) at stage e10.5 (left). At later stages (e12.5), gar CNS65 activity is restricted to the proximal portion of the limb and absent in developing digits (middle). Zebrafish CNS65 drives reporter expression in developing mouse limbs at e10.5, but only in forelimbs (right). Number of LacZ-positive embryos showing limb signal is indicated at the bottom right; fl, forelimb, hl, hindlimb (c). Scale bars: 50 μm (b); 500 μm (c). See also Supplementary Note 12.4.

Europe PMC Funders Author Manuscripts

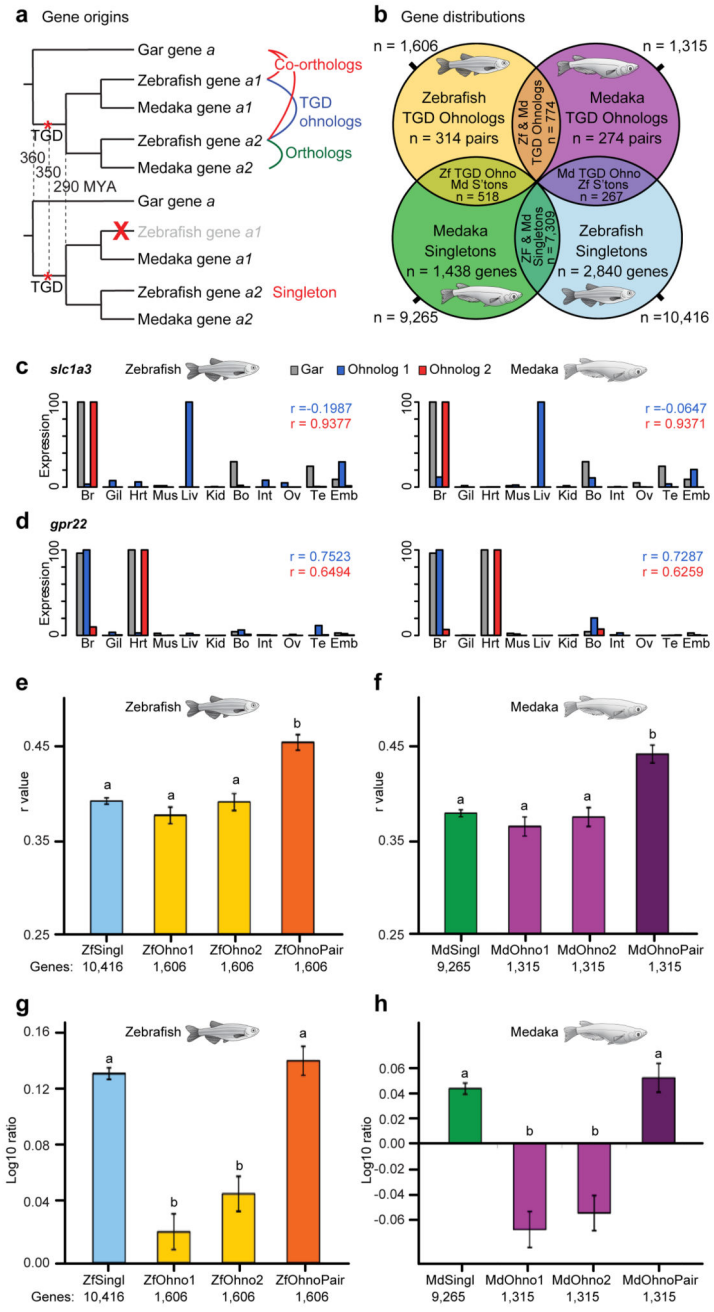

Figure 6. Gar illuminates gene expression evolution post-TGD

Origin (a) and distribution (b) of gar and teleost singletons or TGD ohnologs (Supplementary Note 13.1, Supplementary Tab. 23). c) Neofunctionalized ohnologs (*slc1a3*): novel expression in liver; d) Subfunctionalized ohnologs (*gpr22*): one is expressed in brain like in gar, the other in heart like in gar; r: correlation of expression profiles of each ohnolog vs. gar pattern. Supplementary Note 13.2 lists neo- and subfunctionalization criteria. e-h) Expression conservation for ohnologs or singletons in zebrafish (Zf; e, g) and medaka (Md; f, h) (Supplementary Note 13.2). e, f) Mean correlations (r values) between

expression patterns of gar genes and teleost ortholog(s). Correlations of average expression levels of ohnolog-pairs to gar were greater than ohnologs alone and than singletons, showing sharing of ancestral subfunctions between the ohnolog-pair (multiple Wilcoxon Mann-Whitney tests with Bonferroni correction; alpha value 0.05 for significance). g, h) Mean Log10 ratios between expression levels of gar genes and teleost ortholog(s). Individual ohnologs compared to gar were expressed at significantly lower levels than singletons, but ohnolog-pair/gar ratios were not statistically different from singleton/gar ratios, suggesting that expression levels of ohnolog-pairs approach pre-duplication genes (multiple two-sided Student t-test with Bonferroni correction; alpha value 0.05 for significance). Error bars: standard error of the mean (s.e.m.). 'OhnoPair': average expression of ohnolog-pair (Supplementary Note 13.2). Br, brain; Gil, gill; Hrt, heart; Mus, muscle; Liv, liver, Kid, kidney; Bo, bone; Int, intestine; Ov, ovary; Te, testis; Emb, embryo. Source Dataset 4 contains data for Fig. 6c-h.

Europe PMC Funders Author Manuscripts