

OPEN

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

Ingo Braasch^{1,34}, Andrew R Gehrke², Jeramiah J Smith³, Kazuhiko Kawasaki⁴, Tereza Manousaki⁵, Jeremy Pasquier⁶, Angel Amores¹, Thomas Desvignes¹, Peter Batzel¹, Julian Catchen⁷, Aaron M Berlin⁸, Michael S Campbell^{9,34}, Daniel Barrell^{10,11}, Kyle J Martin^{12,34}, John F Mulley¹³, Vydianathan Ravi¹⁴, Alison P Lee¹⁴, Tetsuya Nakamura², Domitille Chalopin^{15,34}, Shaohua Fan^{16,34}, Dustin Weisel^{17,18}, Cristian Cañestro^{19,20}, Jason Sydes¹, Felix E G Beaudry²¹, Yi Sun^{22,23}, Jana Hertel^{24,34}, Michael J Beam¹, Mario Fasold^{24,34}, Mikio Ishiyama²⁵, Jeremy Johnson⁸, Steffi Kehr²⁴, Marcia Lara⁸, John H Letaw¹, Gary W Litman²⁶, Ronda T Litman²⁶, Masato Mikami²⁷, Tatsuya Ota²⁸, Nil Ratan Saha²⁹, Louise Williams⁸, Peter F Stadler²⁴, Han Wang^{22,23}, John S Taylor²¹, Quenton Fontenot³⁰, Allyse Ferrara³⁰, Stephen M J Searle¹⁰, Bronwen Aken^{10,11}, Mark Yandell⁹, Igor Schneider³¹, Jeffrey A Yoder^{17,18}, Jean-Nicolas Volff¹⁵, Axel Meyer^{16,32}, Chris T Amemiya²⁹, Byrappa Venkatesh¹⁴, Peter W H Holland¹², Yann Guiguen⁶, Julien Bobe⁶, Neil H Shubin², Federica Di Palma^{8,34}, Jessica Alföldi⁸, Kerstin Lindblad-Toh^{8,33} & John H Postlethwait¹

To connect human biology to fish biomedical models, we sequenced the genome of spotted gar (*Lepisosteus oculatus*), whose lineage diverged from teleosts before teleost genome duplication (TGD). The slowly evolving gar genome has 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 (mediated, for example, by Hox, ParaHox and microRNA genes). Numerous conserved noncoding elements (CNEs; often *cis* regulatory) undetectable in direct human-teleost comparisons become apparent using gar: functional studies uncovered conserved roles for such cryptic CNEs, facilitating annotation of sequences identified in human genome-wide association studies. Transcriptomic analyses showed that the sums of expression domains and expression levels for duplicated teleost genes often approximate the patterns and levels of expression for 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.

Teleost fishes represent about half of all living vertebrate species¹ and provide important models for human disease (for example, zebrafish and medaka)^{2–9}. Connecting teleost genes and gene functions to human biology (Fig. 1a) can be challenging given (i) the two rounds of early vertebrate genome duplication (VGD1 and VGD2 (ref. 10), but see ref. 11) followed by reciprocal loss of some ohnologs (gene duplicates derived from genome duplication¹²) in teleosts and tetrapods, including humans^{13,14}; (ii) the TGD, which resulted in duplicates of many human genes^{15,16}; and (iii) rapid teleost sequence evolution^{17,18}, 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 (*L. oculatus*, henceforth ‘gar’; Supplementary Fig. 1 and Supplementary Note) because its lineage represents the unduplicated sister group of teleosts^{19,20} (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 microRNA (miRNA) gene families. Surprisingly, many entire gar chromosomes have been conserved with some tetrapods for 450 million years. Notably, gar facilitates the identification of CNEs, which are often regulatory, that teleosts and humans share but that are not detected by direct sequence comparisons. Global gene expression analyses show that expression domains and levels for TGD-generated duplicates usually sum to those for the corresponding gar gene, as expected if ancestral regulatory elements were 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 was sequenced to 90× coverage using Illumina technology. The ALLPATHS-LG²¹ draft assembly covers 945 Mb with quality metrics comparable

A full list of affiliations appears at the end of the paper.

Received 12 September 2015; accepted 12 February 2016; published online 7 March 2016; doi:10.1038/ng.3526

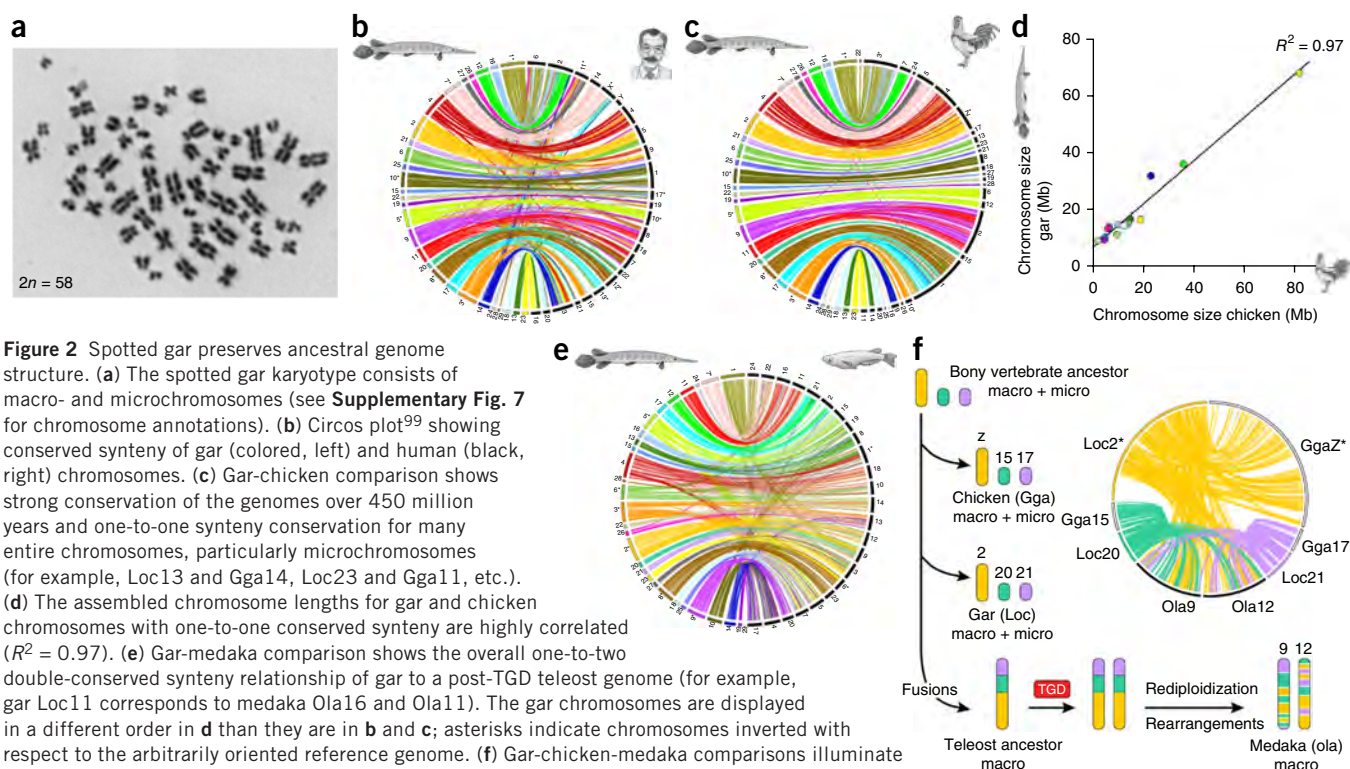


Figure 2 Spotted gar preserves ancestral genome structure. (a) The spotted gar karyotype consists of macro- and microchromosomes (see **Supplementary Fig. 7** for chromosome annotations). (b) Circos plot⁹⁹ showing conserved synteny of gar (colored, left) and human (black, right) chromosomes. (c) Gar-chicken comparison shows strong conservation of the genomes over 450 million years and one-to-one synteny conservation for many entire chromosomes, particularly microchromosomes (for example, Loc13 and Gga14, Loc23 and Gga11, etc.). (d) The assembled chromosome lengths for gar and chicken chromosomes with one-to-one conserved synteny are highly correlated ($R^2 = 0.97$). (e) Gar-medaka comparison shows the overall one-to-two double-conserved synteny relationship of gar to a post-TGD teleost genome (for example, gar Loc11 corresponds to medaka Ola16 and Ola11). The gar chromosomes are displayed in a different order in **d** than they are in **b** and **c**; asterisks indicate chromosomes inverted with respect to the arbitrarily oriented reference genome. (f) Gar-chicken-medaka comparisons illuminate the karyotype evolution leading to modern teleosts. The genome of the bony vertebrate ancestor contained both macro- and microchromosomes, some of which remain largely conserved in chicken and gar, for example, 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 the **Supplementary Note** for details.

Gar informs the evolution of bony vertebrate karyotypes

Gar represents the first chromonome²² 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 Fig. 7** and **Supplementary Note**). Aligning gar chromosomes to those of human, chicken and teleosts highlighted distinct conservation of orthologous segments in all species (**Fig. 2b–e**, **Supplementary Figs. 8** and **9**, and **Supplementary Note**). Strikingly, gar-chicken comparisons showed conservation of many entire chromosomes (**Fig. 2c**). The chicken and gar karyotypes differed 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** and **Supplementary Note**). This similarity in chromosome size and gene content is strong evidence that the karyotype of the common bony vertebrate ancestor of gar and chicken possessed both macro- and microchromosomes as Ohno³⁵ hypothesized, consistent with microchromosomes in coelacanth³⁶ and cartilaginous fishes³⁵, for which no chromosomes are yet available.

The gar chromonome also tests the hypothesis that an increase in the number of interchromosomal rearrangements occurred in teleosts after, and possibly as a result of, the TGD²⁰. For each gar chromosome segment, teleosts usually have two ohnologous segments, verifying gar-teleost divergence before the TGD²⁰. Each TGD-derived pair in teleosts usually shows conserved synteny with more than one gar chromosome, indicating rearrangements before the TGD (**Fig. 2e**, **Supplementary Figs. 8** and **9**, and **Supplementary Note**). Gar shares

many whole chromosomes with chicken (**Fig. 2c**) but few with teleosts (**Fig. 2e**). These results indicate that chromosome fusions thought to have occurred in the ray-finned lineage after divergence from the lobe-finned lineage³⁷ actually occurred in the teleost lineage after divergence from gar but before the TGD (**Fig. 2f** and **Supplementary Fig. 10**). This finding explains how spotted gar has more chromosomes ($n = 29$; **Fig. 2a**) than typical teleosts ($n \sim 24$ or 25 ; ref. 38) without experiencing the TGD. Comparisons taking the TGD into account further found an average fission and translocation rate in percomorphs (stickleback, medaka and pufferfish) relative to gar that is similar to that in the chicken lineage. Zebrafish had a higher rearrangement rate, even after accounting for the TGD (**Supplementary Fig. 11** and **Supplementary Note**). 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 the identification of true orthologs^{22,39} and frustrates the translation of knowledge from teleost biomedical models to human biology¹³. Gar is uniquely informative because its lineage did not experience the TGD and often retains ancestral VGD1 and 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**).

Analyses of developmental gene families showed stability in the gar gene repertoire, including for Hox gene clusters (**Supplementary Note**). Gar has 43 Hox genes organized into four clusters, as expected for an unduplicated ray-finned fish (**Supplementary Fig. 12**). No

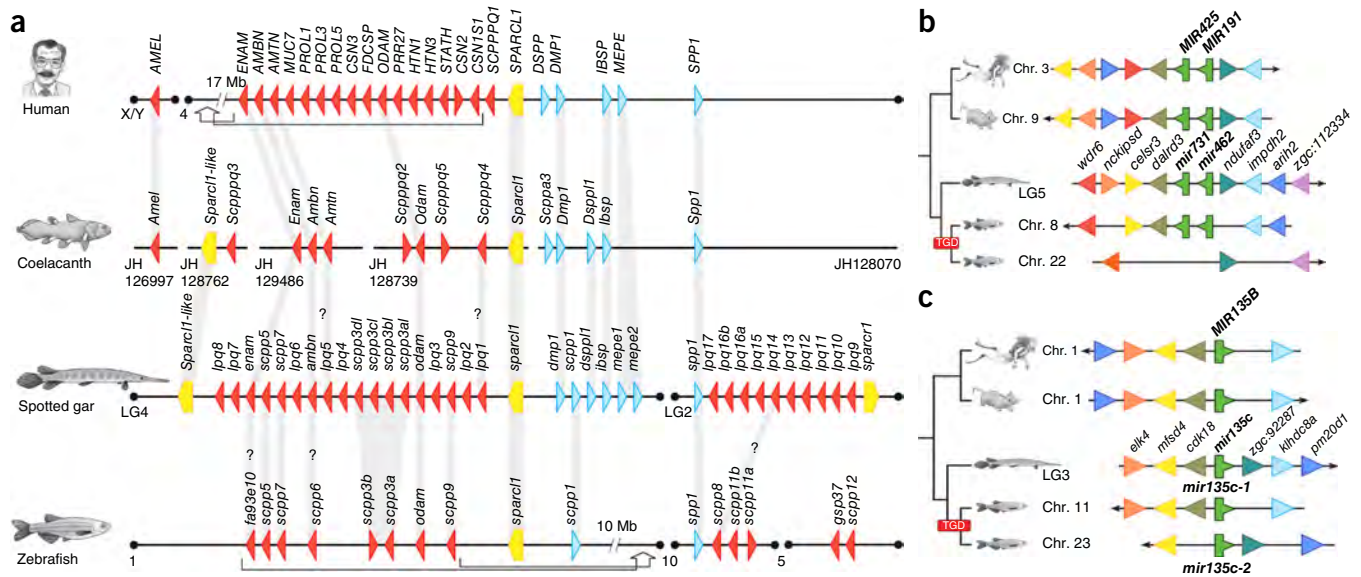


Figure 3 Gar helps connect vertebrate protein-coding and miRNA genes. (a) Scpp gene arrangements in human, coelacanth, gar and zebrafish including P/Q-rich (red) and acidic (blue) Scpp genes and Sparc-like genes (yellow) (**Supplementary Note**; ref. 68). Orthologies (gray vertical bars) among lobe-finned vertebrates (for example, human and coelacanth) and teleosts (for example, zebrafish) had previously been limited to Odam and Spp1 genes. Gar connects lineages through orthologs of genes previously known only from either teleosts (*scpp1*, *scpp3*, *scpp5*, *scpp7* and *scpp9*) or lobe-finned vertebrates (*enam*, *ambn*, *dmp1*, *dsppl1*, *ibsp* and *mepe*). Further putative orthologies supported by only short stretches of sequence similarity (indicated by a question mark) connect gar *enam*, *ambn* and *lpq14* genes with zebrafish *fa93e10*, *scpp6* and *scpp8* genes, respectively; gar *lpq1* and coelacanth *Scpppq4*; and gar *lpq5* with *Amtn* genes in lobe-finned vertebrates. Arrows in human and zebrafish indicate intrachromosomal rearrangements separating originally clustered genes into distant chromosomal locations (distance in Mb). Analysis of conserved synteny for 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 Fig. 26** and **Supplementary Note**). (b) The gar 'conserved synteny bridge' (**Supplementary Note**) 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 (highlighted in bold). (c) Gar newly connects through synteny zebrafish TGD-derived ohnologs *mir135c-1* and *mir135c-2* with mammalian *Mir135B* genes (highlighted in bold).

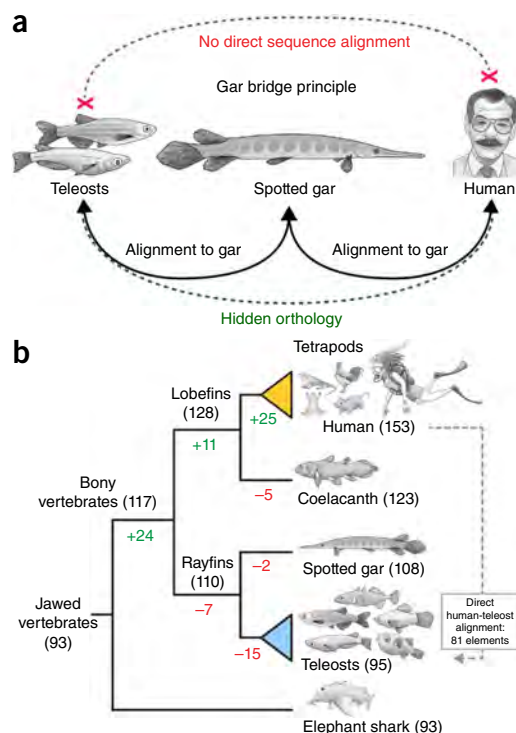
Hox gene has been completely lost in gar since divergence from the last common ray-finned ancestor. The *hoxd14* gene, missing from teleosts but present in paddlefish⁴⁰, is recognizable as a pseudogene in gar (**Supplementary Fig. 13**). In contrast, teleosts have far fewer Hox cluster genes than the 82 expected after genome duplication (for example, zebrafish has 49 genes and stickleback has 46 genes), demonstrating massive Hox gene loss after the TGD. Teleosts lack orthologs of *hoxa6* and *hoxd2*, zebrafish lacks all HoxDb cluster protein-coding genes¹⁵ and percomorphs lack the HoxCb cluster⁴¹, but gar lacks just one Hox cluster gene from the last common bony vertebrate ancestor (*hoxa14*), fewer than tetrapods (for example, human has three losses) and coelacanth (two losses) (**Supplementary Fig. 12**). Gar ParaHox clusters (**Supplementary Table 5** and **Supplementary Note**) are also more complete than those in teleosts and tetrapods, with four clusters containing seven genes. Gar retained *cdx2*, which highlights a VGD1/VGD2 ohnolog 'gone missing' from teleosts (**Supplementary Fig. 14**). Gar possesses the VGD1/VGD2 ohnolog *pdx2*, previously found only in cartilaginous fishes and coelacanth⁴², indicating that *pdx2* was lost independently teleosts and tetrapods (**Supplementary Figs. 14** and **15**). Retinoic acid regulates Hox cluster gene expression⁴³, but retinoic acid-synthesizing Aldh enzymes (**Supplementary Note**) vary in number among vertebrates⁴⁴: tetrapods have three genes (*Aldh1a1*, *Aldh1a2* and *Aldh1a3*), zebrafish has two genes (*aldh1a2* and *aldh1a3*) and medaka has just one (*aldh1a2*)⁴⁵. Finding all three genes in gar rules out the hypothesis⁴⁵ 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 Fig. 17**, **Supplementary Table 6** and **Supplementary Note**)⁴⁸ and opsin (**Supplementary Fig. 18**, **Supplementary Table 7** and **Supplementary Note**)⁴⁹ genes link the gene repertoires of teleosts and tetrapods; for example, gar clarifies which circadian genes originated in VGD events and which originated in the TGD event. Gar has pinopsin, present in tetrapods but absent from teleosts, along with exo-rhodopsin, previously thought to compensate for the lack of pinopsin in teleosts⁵⁰.

Evolution of vertebrate immunity becomes clearer using gar (**Supplementary Note**). Major histocompatibility complex (MHC) class I and class II genes (**Supplementary Figs. 19–21**) are tightly linked in tetrapods and cartilaginous fishes but are unlinked in teleosts^{51,52}. In gar, at least one pair of class I and class II genes is 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-like, L-like and U/S-like, for example^{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 the MHC class II or class III region on Hsa6, and some are adjacent to orthologs of teleost MHC class I genes (**Supplementary Table 8**). The human MHC class III region on Hsa6 has syntenic segments on Hsa1, Hsa9 and Hsa19; these four ohnologs likely arose in VGD1 and VGD2 (ref. 57), as supported by the gar genome (**Supplementary Table 8**).

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 uncovered 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 shows 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 using gar as a bridge as in a. See **Supplementary Figure 37**, **Supplementary Table 22** and the **Supplementary Note** for details.



Gar immunoglobulin 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 immunity⁶⁰, 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 tropicalis*⁶¹, are downstream of V_H and J_H segments. Phylogenetic analyses of Toll-like receptor (TLR) genes (**Supplementary Fig. 24**) in tetrapods, teleosts and gar showed that the 16 identifiable gar TLRs encompass all six major TLR families⁶². Gar TLRs appear to share evolutionary histories with the TLRs from 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 the 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^{18,67,68}. Gar clarifies understanding of these genes and their evolution because it retains ancient characteristics both in its ganoid scales, which contain ganoin, hypothesized to be a type of enamel⁶⁹, and in its teeth, which are covered by both enameloid and enamel⁷⁰ (**Supplementary Note**). Mammalian genomes were thought to contain the largest number of Scpp genes (human, 23 genes; coelacanth, 14 genes; zebrafish, 15 genes), and only 2 genes (*Spp1* and *Odam*) seemed to be common to lobe-finned vertebrates and teleosts⁶⁸ (**Fig. 3a**). We identified 35 Scpp genes in gar in two clusters on LG2 and LG4 (**Fig. 3a**, **Supplementary Fig. 26**, **Supplementary Table 9** and **Supplementary Note**), which contain *spp1* and *odam*, respectively. Notably, gar includes orthologs of five Scpp genes previously found only in teleosts and six Scpp genes 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 Table 9** and **Supplementary Note**).

The enamel matrix protein genes encoding 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 ortholog for *Amel*) in the gar genome and transcriptomes. The gar *ambn* and *enam* genes show sequence similarity to zebrafish *scpp6* and *fa93e10*, respectively, suggesting that teleosts may have divergent orthologs, a hypothesis supported by conserved gene orders in the gar and zebrafish clusters (**Fig. 3a**).

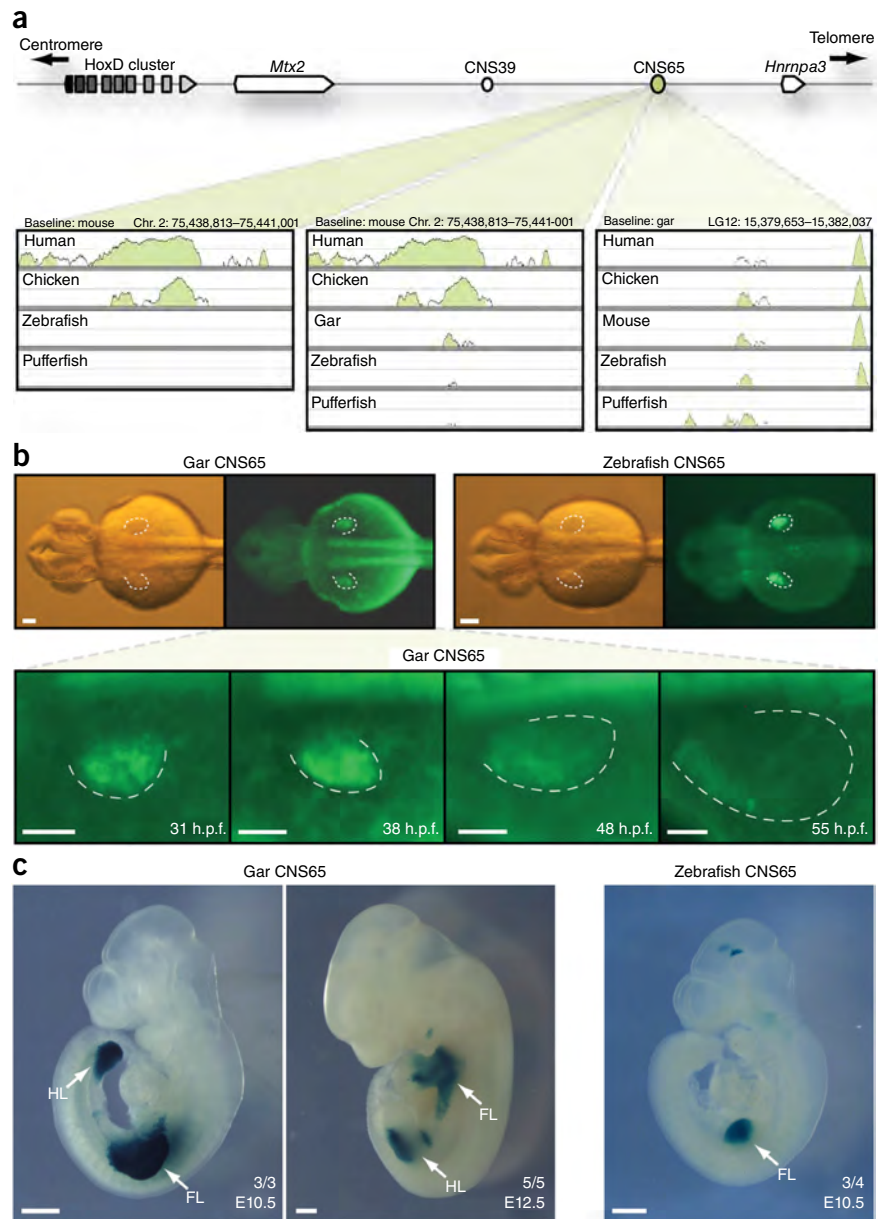
RT-PCR and our gar skin transcriptome analysis identified expression of *ambn* and *enam* in enamel-containing gar teeth and in gar skin that includes scales with ganoin (**Supplementary Table 9** and **Supplementary Note**), suggesting that strong expression of *ambn* and *enam* is limited to enamel 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 vertebrate enamel and in gar enamel and ganoin supports homology of these tissues. Analysis of gnathostome fossils suggested that ganoin is plesiomorphic for crown osteichthyans and arose before enamel⁷¹; thus, enamel-bearing teeth likely evolved by coopting enamel matrix genes originally used in ganoid scales. The *Amel* gene may have evolved subsequently to encode the principal organic component of the ‘true enamel’ that appears to have originated in lobe-finned vertebrates⁶⁸.

Gar expresses 12 additional Scpp genes (including the *odam* and *scpp9* hypermineralization genes⁶⁶) in both teeth and scales and another 4 genes in bone (**Supplementary Table 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 genes⁶⁵, gar has retained characteristics of both lineages.

Gar connects vertebrate microRNAomes

miRNA genes could become teleost or tetrapod specific^{18,72} by their loss in one lineage or gain in the other. We studied gar miRNAs computationally (**Supplementary Fig. 27**, **Supplementary Table 10** and **Supplementary Note**) and annotated them using a sequence-based approach (**Supplementary Note**). Small RNA-seq data for four tissues identified 302 mature miRNAs derived from 233 genes, of which 229 belong to 107 families and 4 lack a known family (**Supplementary Fig. 28** and **Supplementary Table 11**). Gar-zebrafish^{73,74} comparisons showed that four families and four individual miRNA genes emerged in teleosts. Of the 22 families thought to have been lost

Figure 5 Identification and functional analysis of the gar and teleost early-phase HoxD enhancer CNS65. **(a)** Top, schematic of the mouse HoxD telomeric gene desert, which contains the CNS39 and CNS65 enhancers that drive early-phase HoxD expression in limbs. Using mouse as the baseline, VISTA alignments of the HoxD gene desert show sequence conservation with human and chicken for CNS65 but not with teleosts (zebrafish and pufferfish) (bottom left). An alignment including gar, however, shows a peak of conservation in the gar sequence (middle). Using the identified gar CNS65 as the baseline identified 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 (h.p.f.) (top). Gar CNS65 has pectoral fin activity beginning at 31 h.p.f., which drives GFP expression throughout the fin, and becomes deactivated around 48 h.p.f. (bottom). Dashed lines indicate the distal portion of the pectoral fins. **(c)** Gar CNS65 drives expression throughout the early mouse forelimbs 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 is absent in developing digits (middle). Zebrafish CNS65 drives reporter expression in developing mouse limbs at E10.5 but only in forelimbs (right). The number of LacZ-positive embryos showing limb signal is indicated at the bottom right of each image; FL, forelimb, HL, hindlimb. Scale bars, 50 μ m **(b)** and 500 μ m **(c)**. See also the **Supplementary Note**.



in teleosts¹⁸, 2 actually belong to the same family and orthologs of 4 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 teleosts⁷⁴ but absent from zebrafish. A single family present in teleosts and lobe-finned fishes (*miR150*) was not found in gar. Notably, no miRNA family loss was specific to teleosts, suggesting that the TGD did not accelerate family loss.

The 'gar bridge' helps to identify miRNA orthologies. For example, the mammalian *Mir425* and *Mir191* genes, thought to be lost in teleosts¹⁸, are orthologs of teleost *mir731* and *mir462*, respectively (**Fig. 3b**). Additionally, mammalian *Mir135B* is orthologous to *mir135c* in gar and the zebrafish TGD-derived 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 retention rate for protein-coding genes (20–24%; ref. 75), consistent with the hypothesis that miRNA genes are likely to be retained after a duplication owing to their incorporation into multiple gene regulatory networks^{76–79}.

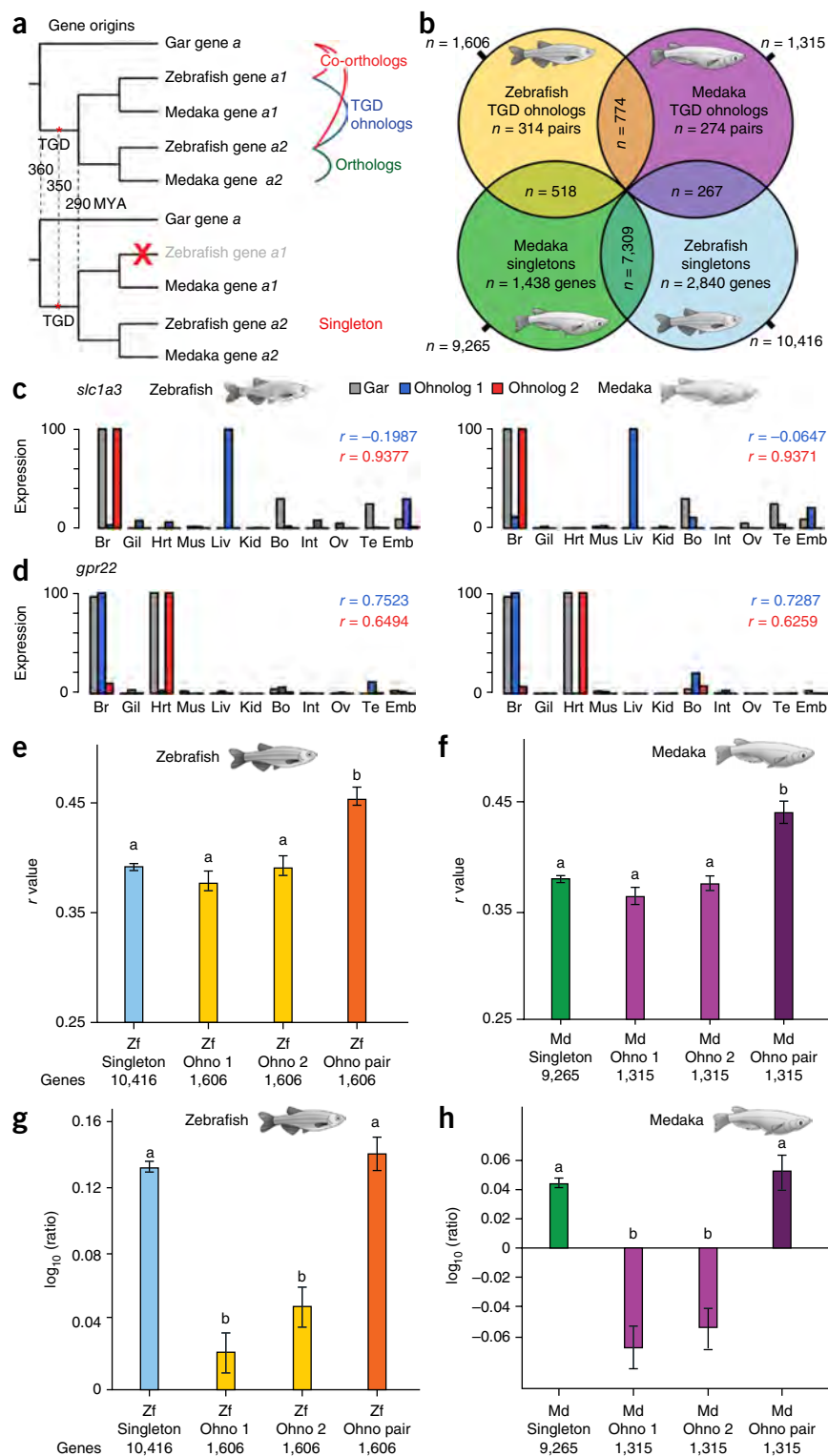
Gar highlights hidden orthology of cis-regulatory elements

CNEs often function as *cis*-acting regulators^{80,81}, but many appear to be absent in teleosts, presumably because of rapid teleost sequence evolution (**Fig. 1b** and **Supplementary Note**); ancestral CNEs identified in tetrapods, however, might be detected in ray-finned fish using the slowly evolving gar.

CNE analyses near developmental gene loci (Hox and ParaHox clusters, Pax6 and 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 (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 Figs. 14, 15 and 29–35, Supplementary Tables 12–19 and Supplementary Note**).

Gar elucidates the origins of tetrapod limb enhancers, evidenced by whole-genome alignments for 13 vertebrates (including gar, five teleosts, coelacanth, five tetrapods and elephant shark; **Supplementary Fig. 36, Supplementary Tables 20 and 21, and Supplementary Note**). Of 153 known human limb enhancers^{33,82–84}, human-centric alignments identified 71% (108) in gar, but only 53% (81) were identified through direct human-teleost alignments. Of the 72 human limb enhancers not detected by human-teleost alignment, 40% (29) aligned to gar, confirming their presence in the bony vertebrate

Figure 6 Gar illuminates gene expression evolution after the TGD. (a,b) The origin (a) and distribution (b) of gar and teleost singletons and TGD-derived ohnologs (Supplementary Table 23 and Supplementary Note). (c) Neofunctionalized ohnologs for *slc1a3* showing new expression in liver. (d) Subfunctionalized TGD orthologs of *gpr22* with one expressed in brain as in gar and the other expressed in heart as in gar. In c and d, the r values denote the correlation of the expression profile of each ohnolog with the gar pattern. The Supplementary Note lists neofunctionalization and subfunctionalization criteria. (e-h) Expression conservation for ohnologs and singletons in zebrafish (Zf; e,g) and medaka (Md; f,h) (Supplementary Note). (e,f) Mean correlation between the expression patterns of gar genes and teleost ortholog(s). The correlation between average expression levels for ohnolog pairs and gar genes was greater than that for ohnologs alone and than that for singletons, indicating sharing of ancestral subfunctions by the ohnolog pair (multiple Wilcoxon Mann-Whitney tests with Bonferroni correction, $\alpha = 0.05$ for significance). (g,h) Mean \log_{10} -transformed ratios of expression levels for gar genes and teleost ortholog(s). In comparison to gar genes, individual ohnologs were expressed at significantly lower levels than singletons; ohnolog pair/gar ratios were not statistically different from singleton/gar ratios, suggesting that the aggregate expression level of ohnolog pairs approaches the expression level of the preduplication gene (multiple two-sided Student's t test with Bonferroni correction, $\alpha = 0.05$ for significance). Error bars in e-h, s.e.m. Br, brain; Gil, gill; Hrt, heart; Mus, muscle; Liv, liver; Kid, kidney; Bo, bone; Int, intestine; Ov, ovary; Te, testis; Emb, embryo.



ancestor and loss or considerable divergence in teleosts. Of these 29 enhancers, 15 also aligned to elephant shark, highlighting their existence in the gnathostome ancestor. Fourteen occurred in gar but not in teleosts and would have been incorrectly characterized as lobe-finned vertebrate innovations without gar data (Supplementary Table 22 and Supplementary Note).

Using the gar bridge (Fig. 4a), we tested whether the 29 human enhancers not directly identified in teleosts might represent rapid divergence rather than definitive loss. Inspection of human-centric and then gar-centric alignments showed 48% (14/29) aligning to at least one teleost (Supplementary Table 22). Gar thus substantially improves understanding of the evolutionary origin of vertebrate limb enhancers and their fate in teleosts (Fig. 4b, Supplementary Fig. 37 and Supplementary Table 22). Strikingly, despite using the gar bridge, we found that teleosts lost substantially more limb enhancers (15) than gar (2) (Fig. 4b and Supplementary Fig. 37), suggesting that gar might be a better model than teleosts for investigating the fin-to-limb transition⁸⁵.

Functional studies of a HoxD limb enhancer tested the usefulness of a 'gar CNE bridge'. HoxD and HoxA clusters pattern proximal and distal mammalian limbs by 'early' and 'late' phases of gene expression, respectively⁸⁶. Early-phase HoxD expression in fins and limbs shows several features that are presumed to be homologous⁸⁷ and may derive from shared but cryptic regulatory elements. The CNS39 and CNS65 elements drive early-phase HoxD activation in mammals⁸⁸ (Fig. 5a). Human-centric (Supplementary Table 22) and local mouse-centric

(Fig. 5a) alignments failed to detect CNS39 in ray-finned fish but identified CNS65 in gar. Notably, CNS65 was identified in teleosts only by using the gar bridge (Fig. 5a and Supplementary Table 22).

To test whether cryptic CNE orthologs preserve enhancer function, we used CNS65-driven reporter constructs to generate transgenic zebrafish and mice (Supplementary Note). CNS65 from either gar or zebrafish drove early expression in the developing zebrafish pectoral fin (Fig. 5b). Gar CNS65 drove expression in the forelimbs and hindlimbs of embryonic day (E) 10.5 mice (Fig. 5c) that was indistinguishable from the activity of mouse CNS65 (ref. 88). Zebrafish CNS65 activated forelimb expression somewhat more weakly than gar CNS65 (Fig. 5c). At E12.5, gar CNS65 activated proximal but not distal limb expression (Fig. 5c), mimicking the endogenous mouse enhancer⁸⁸. These functional experiments suggest 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.

Across the gar genome, we identified approximately 28% of human-centric CNEs (39,964/143,525), more than in any of five aligned teleost genomes. Around 19,000 human-centric CNEs aligned to gar but not to any teleost (Supplementary Table 21 and Supplementary Note). 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) establishes hidden orthology from human to gar to zebrafish for many of these human-centric CNEs (30–36%, depending on overlap; Supplementary Table 21 and Supplementary Note). These approximately 6,500 newly connected human CNEs contain around 1,000 SNPs linked to human conditions in genome-wide association studies (GWAS), thereby connecting otherwise undetected disease-associated haplotypes to genomic locations in zebrafish (Supplementary Table 21). The gar bridge thus helps identify biomedically relevant candidate regions in model teleosts for functional testing, potentially enhancing teleost models for biomedical research.

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 partitioning of ancestral functions^{89–92}. Because the contribution of various fates has not yet been studied using a closely related TGD outgroup, we generated a list of gar genes and their orthologous TGD-derived ohnologs or singletons in zebrafish and medaka using phylogenetic⁹³ and conserved synteny⁹⁴ analyses (Fig. 6a,b, Supplementary Table 23 and Supplementary Note).

To compare tissue-specific gene expression patterns, we conducted RNA-seq analysis for ten adult organs and stage-matched embryos for gar, zebrafish and medaka and then normalized reads across tissues for each gene in each species (Supplementary Note). 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). New expression domains like this are expected if one ohnolog maintained ancestral patterns while the other evolved new functions⁹⁵ before the teleost radiation. In contrast, gar expressed *gpr22* mostly in brain and heart, but both teleosts expressed one ohnolog in brain and the other in heart (Fig. 6d), as expected from partitioning of ancestral regulatory subfunctions⁸⁹.

To characterize the effects of the TGD on evolution of gene expression, we plotted tissue-specific expression levels in gar versus (i) expression of orthologous teleost singletons, (ii) expression of each

TGD-derived ohnolog when both were retained and (iii) the averaged expression level of both retained ohnologs ('ohnolog pair'), and we then calculated correlation coefficients. Our results showed that the correlation between the expression patterns of gar genes and those of their teleost singleton orthologs was not significantly different from the correlation of expression patterns between gar genes and those of either copy of their teleost TGD-derived co-orthologs (Fig. 6e,f). Thus, when 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-derived duplicates correlated significantly more closely with gar than with either ohnolog taken alone and correlated more closely with gar than with singletons (Fig. 6e,f); thus, ancestral gene subfunctions tended to be partitioned between TGD-derived ohnologs, which maintained ancestral functions as a gene pair, as predicted by the subfunctionalization model⁸⁹.

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 were each expressed at significantly lower levels than singletons as compared to gar orthologs (Fig. 6g,h). The ohnolog pair/gar expression ratios, however, showed no statistical difference from the singleton/gar expression ratios (Fig. 6g,h). This finding suggests that the aggregate expression level for ohnolog pairs tends to evolve to approximately the expression level of the preduplication gene, as expected by quantitative subfunctionalization^{89,90,96}.

Taken together, our analyses indicate that, after the 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 expression for preduplication genes.

DISCUSSION

Gar is the first ray-finned fish genome sequence not affected by the TGD. Because of gar's 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. Our analyses show that gar bridges teleosts to tetrapods in genome arrangement, allowing the identification of orthologous genes by possessing ancient VGD ohnologs lost reciprocally in teleosts and tetrapods and elucidating the 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 genomic analyses show that sequences formerly considered unique to teleosts or tetrapods are often shared by ray-finned and lobe-finned vertebrates, including human. Notably, the gar bridge helps identify potential gene regulatory elements that are shared by teleosts and humans but are elusive in direct teleost-tetrapod comparisons. The availability of gar embryos and the ease of raising eggs to adults in the laboratory²² (Supplementary Fig. 1) make 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.

URLs. 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.siganae.org/index.html>; RepeatMasker, <http://www.repeatmasker.org/>.

METHODS

Methods and any associated references are available in the [online version of the paper](#).

Accession codes. The spotted gar genome assembly is available from GenBank under accession [GCA_000242695.1](#). RNA-seq data are available from the Sequence Read Archive (SRA) under accessions [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 from GenBank under accessions [KU189274–KU189300](#).

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank the Broad Institute Genomics Platform for constructing and sequencing gar DNA and RNA libraries and J. Turner-Maier for the gar transcriptome assembly. We thank the teams of the Bayosphere Research Laboratory (Nicholls State University) and the University of Oregon Fish Facility for gar work and husbandry. We thank J. Westlund for the design of the species illustrations. The generation of gar sequences and assemblies by the Broad Institute of MIT and Harvard University was supported by US National Institutes of Health (NIH)/National Human Genome Research Institute grant U54 HG03067. This work was further supported by US NIH grants R01 OD011116 (alias R01 RR020833) and R24 OD0119004 (J.H.P.); a Feodor Lynen Fellowship from the Alexander von Humboldt Foundation and the Volkswagen Foundation Initiative Evolutionary Biology, grant I/84 815 (I.B.); US NIH grant T32 HD055164 and National Science Foundation (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 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.); Agence Nationale de la Recherche (ANR) grant ANR-10-GENM-017 (PhyloFish; J.B.); the Wellcome Trust (grants WT095908 and WT098051) and the European Molecular Biology Laboratory (D.B., S.M.J.S. and B.A.); the Biomedical Research Council of the Agency for Science, Technology and Research (A*STAR), Singapore (B.V.); European Research Council grant 268513 (P.W.H.H. and K.J.M.); the Ministerio de Ciencia e Innovación (BFU2010-14875 and BFU2015-71340) and the Generalitat de Catalunya, AGAUR (SGR2014-290) (C.C.); US NIH grant R01 AI057559 (G.W.L. and J.A.Y.); US NIH grants R24 OD010922 and R01 GM079492 (C.T.A.); and the National Basic Research Program of China (973 Program) (2012CB947600) and the National Natural Science Foundation of China (NSFC) (31030062) (H.W.).

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, and F.D.P. was 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, and 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 TEs. 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. analyzed ParaHox genes; C.C. analyzed Aldh genes; Y.S. and H.W. analyzed circadian clock genes; F.E.G.B. and J.S.T. analyzed opsin genes; and D.W., G.W.L., R.T.L., T.O., N.R.S., C.T.A., J.H.P. and J.A.Y. analyzed immune genes. K.K., M.I., M.M., P.B. and I.B. carried out the annotation and expression analysis of mineralization-related genes. T.D., M.J.B., P.B., J.S. and J.H.P. annotated and analyzed miRNA genes on the basis of small RNA-seq data generated by T.D. (main text and **Supplementary Note**). J.H., M.F., S.K. and P.F.S. studied miRNAs *in silico* (**Supplementary Note**). 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 manuscript with input from other authors.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/3.0/>.

- Nelson, J.S. *Fishes of the World* 4th edn (John Wiley, 2006).
- Kettleborough, R.N. *et al.* A systematic genome-wide analysis of zebrafish protein-coding gene function. *Nature* **496**, 494–497 (2013).
- Patton, E.E., Mathers, M.E. & Scharl, M. Generating and analyzing fish models of melanoma. *Methods Cell Biol.* **105**, 339–366 (2011).
- Reitzel, A.M. *et al.* Genetic variation at aryl hydrocarbon receptor (*AHR*) loci in populations of Atlantic killifish (*Fundulus heteroclitus*) inhabiting polluted and reference habitats. *BMC Evol. Biol.* **14**, 6 (2014).
- Lee, O., Green, J.M. & Tyler, C.R. Transgenic fish systems and their application in ecotoxicology. *Crit. Rev. Toxicol.* **45**, 124–141 (2015).
- Albertson, R.C., Cresko, W., Detrich, H.W. III & Postlethwait, J.H. Evolutionary mutant models for human disease. *Trends Genet.* **25**, 74–81 (2009).
- Harel, I. *et al.* A platform for rapid exploration of aging and diseases in a naturally short-lived vertebrate. *Cell* **160**, 1013–1026 (2015).
- Pagán, A.J. & Ramakrishnan, L. Immunity and immunopathology in the tuberculous granuloma. *Cold Spring Harb. Perspect. Med.* **5**, a018499 (2015).
- Hagedorn, E.J., Durand, E.M., Fast, E.M. & Zon, L.I. Getting more for your marrow: boosting hematopoietic stem cell numbers with PGE2. *Exp. Cell Res.* **329**, 220–226 (2014).
- Dehal, P. & Boore, J.L. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol.* **3**, e314 (2005).
- Smith, J.J. & Keinath, M.C. The sea lamprey meiotic map improves resolution of ancient vertebrate genome duplications. *Genome Res.* **25**, 1081–1090 (2015).
- Wolfe, K. Robustness—it's not where you think it is. *Nat. Genet.* **25**, 3–4 (2000).
- Frankenberg, S.R. *et al.* The POU-er of gene nomenclature. *Development* **141**, 2921–2923 (2014).
- Braasch, I. *et al.* Connectivity of vertebrate genomes: paired-related homeobox (*Prrx*) genes in spotted gar, basal teleosts, and tetrapods. *Comp. Biochem. Physiol. C Toxicol. Pharmacol.* **163**, 24–36 (2014).
- Amores, A. *et al.* Zebrafish *hox* clusters and vertebrate genome evolution. *Science* **282**, 1711–1714 (1998).
- Taylor, J.S., Braasch, I., Frickey, T., Meyer, A. & Van de Peer, Y. Genome duplication, a trait shared by 22000 species of ray-finned fish. *Genome Res.* **13**, 382–390 (2003).
- Amemiya, C.T. *et al.* The African coelacanth genome provides insights into tetrapod evolution. *Nature* **496**, 311–316 (2013).
- Venkatesh, B. *et al.* Elephant shark genome provides unique insights into gnathostome evolution. *Nature* **505**, 174–179 (2014).
- Hoegg, S., Brinkmann, H., Taylor, J.S. & Meyer, A. Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish. *J. Mol. Evol.* **59**, 190–203 (2004).
- Amores, A., Catchen, J., Ferrara, A., Fontenot, Q. & Postlethwait, J.H. Genome evolution and meiotic maps by massively parallel DNA sequencing: spotted gar, an outgroup for the teleost genome duplication. *Genetics* **188**, 799–808 (2011).
- Gnerre, S. *et al.* High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc. Natl. Acad. Sci. USA* **108**, 1513–1518 (2011).
- 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.* **324**, 316–341 (2015).
- Holt, C. & Yandell, M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**, 491 (2011).
- Chalopin, D., Naville, M., Plard, F., Galiana, D. & Volff, J.N. Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates. *Genome Biol. Evol.* **7**, 567–580 (2015).
- Near, T.J. *et al.* Resolution of ray-finned fish phylogeny and timing of diversification. *Proc. Natl. Acad. Sci. USA* **109**, 13698–13703 (2012).
- Betancur-R, R. *et al.* The tree of life and a new classification of bony fishes. *PLoS Curr.* doi:10.1371/currents.tol.53ba26640df0cace75bb165c8c26288 (18 April 2013).



27. Broughton, R.E., Betancur-R, R., Li, C., Arratia, G. & Ortí, G. Multi-locus phylogenetic analysis reveals the pattern and tempo of bony fish evolution. *PLoS Curr.* doi:10.1371/currents.tol.2ca8041495ffafd0c92756e75247483e (16 April 2013).
28. Faircloth, B.C., Sorenson, L., Santini, F. & Alfaro, M.E. A phylogenomic perspective on the radiation of ray-finned fishes based upon targeted sequencing of ultraconserved elements (UCEs). *PLoS One* **8**, e65923 (2013).
29. Grande, L. An empirical synthetic pattern study of gars (Lepisosteiformes) and closely related species, based mostly on skeletal anatomy. The resurrection of Holosteii. *Copeia* **10** (supplementary issue 2A), 1–863 (2010).
30. Sallan, L.C. Major issues in the origins of ray-finned fish (Actinopterygii) biodiversity. *Biol. Rev. Camb. Philos. Soc.* **89**, 950–971 (2014).
31. 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).
32. Rabosky, D.L. *et al.* Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. *Nat. Commun.* **4**, 1958 (2013).
33. Nikaido, M. *et al.* Coelacanth genomes reveal signatures for evolutionary transition from water to land. *Genome Res.* **23**, 1740–1748 (2013).
34. Ravi, V. & Venkatesh, B. Rapidly evolving fish genomes and teleost diversity. *Curr. Opin. Genet. Dev.* **18**, 544–550 (2008).
35. Ohno, S. *et al.* Microchromosomes in holoccephalian, chondrosteian and holosteian fishes. *Chromosoma* **26**, 35–40 (1969).
36. Bogart, J.P., Balon, E.K. & Bruton, M.N. The chromosomes of the living coelacanth and their remarkable similarity to those of one of the most ancient frogs. *J. Hered.* **85**, 322–325 (1994).
37. Nakatani, Y., Takeda, H., Kohara, Y. & Morishita, S. Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates. *Genome Res.* **17**, 1254–1265 (2007).
38. Naruse, K. *et al.* A medaka gene map: the trace of ancestral vertebrate proto-chromosomes revealed by comparative gene mapping. *Genome Res.* **14**, 820–828 (2004).
39. Postlethwait, J.H. The zebrafish genome in context: ohnologs gone missing. *J. Exp. Zool. B Mol. Dev. Evol.* **308**, 563–577 (2007).
40. Crow, K.D., Smith, C.D., Cheng, J.F., Wagner, G.P. & Amemiya, C.T. 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.* **4**, 937–953 (2012).
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* **370**, 75–82 (2006).
42. Mulley, J.F. & Holland, P.W. Parallel retention of *Pdx2* genes in cartilaginous fish and coelacanths. *Mol. Biol. Evol.* **27**, 2386–2391 (2010).
43. Duboule, D. Vertebrate *hox* gene regulation: clustering and/or colinearity? *Curr. Opin. Genet. Dev.* **8**, 514–518 (1998).
44. Duester, G. Retinoid signaling in control of progenitor cell differentiation during mouse development. *Semin. Cell Dev. Biol.* **24**, 694–700 (2013).
45. Cañestro, C., Catchen, J.M., Rodríguez-Marí, A., Yokoi, H. & Postlethwait, J.H. Consequences of lineage-specific gene loss on functional evolution of surviving paralogs: ALDH1A and retinoic acid signaling in vertebrate genomes. *PLoS Genet.* **5**, e1000496 (2009).
46. Wang, H. Comparative analysis of period genes in teleost fish genomes. *J. Mol. Evol.* **67**, 29–40 (2008).
47. Rennison, D.J., Owens, G.L. & Taylor, J.S. Opsin gene duplication and divergence in ray-finned fish. *Mol. Phylogenet. Evol.* **62**, 986–1008 (2012).
48. Toloza-Villalobos, J., Arroyo, J.I. & Opazo, J.C. The circadian clock of teleost fish: a comparative analysis reveals distinct fates for duplicated genes. *J. Mol. Evol.* **80**, 57–64 (2015).
49. Lagman, D. *et al.* The vertebrate ancestral repertoire of visual opsins, transducin α 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 (2013).
50. Mano, H., Kojima, D. & Fukada, Y. Exo-rhodopsin: a novel rhodopsin expressed in the zebrafish pineal gland. *Brain Res. Mol. Brain Res.* **73**, 110–118 (1999).
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* **46**, 129–134 (1997).
52. Sato, A. *et al.* Nonlinkage of major histocompatibility complex class I and class II loci in bony fishes. *Immunogenetics* **51**, 108–116 (2000).
53. Dijkstra, J.M., Grimholt, U., Leong, J., Koop, B.F. & 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 Evol. Biol.* **13**, 260 (2013).
54. Grimholt, U. *et al.* A comprehensive analysis of teleost MHC class I sequences. *BMC Evol. Biol.* **15**, 32 (2015).
55. Dirscherl, H., McConnell, S.C., Yoder, J.A. & de Jong, J.L. The MHC class I genes of zebrafish. *Dev. Comp. Immunol.* **46**, 11–23 (2014).
56. Dirscherl, H. & Yoder, J.A. Characterization of the Z lineage major histocompatibility complex class I genes in zebrafish. *Immunogenetics* **66**, 185–198 (2014).
57. Flajnik, M.F. & Kasahara, M. Origin and evolution of the adaptive immune system: genetic events and selective pressures. *Nat. Rev. Genet.* **11**, 47–59 (2010).
58. Danilova, N., Busmann, J., Jekosch, K. & Steiner, L.A. The immunoglobulin heavy-chain locus in zebrafish: identification and expression of a previously unknown isotype, immunoglobulin Z. *Nat. Immunol.* **6**, 295–302 (2005).
59. Hansen, J.D., Landis, E.D. & Phillips, R.B. Discovery of a unique Ig heavy-chain isotype (IgT) in rainbow trout: implications for a distinctive B cell developmental pathway in teleost fish. *Proc. Natl. Acad. Sci. USA* **102**, 6919–6924 (2005).
60. Zhang, Y.A. *et al.* IgT, a primitive immunoglobulin class specialized in mucosal immunity. *Nat. Immunol.* **11**, 827–835 (2010).
61. Parra, Z.E., Ohta, Y., Criscitiello, M.F., Flajnik, M.F. & Miller, R.D. The dynamic TCR δ : TCR δ chains in the amphibian *Xenopus tropicalis* utilize antibody-like V genes. *Eur. J. Immunol.* **40**, 2319–2329 (2010).
62. Roach, J.C. *et al.* The evolution of vertebrate Toll-like receptors. *Proc. Natl. Acad. Sci. USA* **102**, 9577–9582 (2005).
63. Cannon, J.P. *et al.* A bony fish immunological receptor of the NITR multigene family mediates allogeneic recognition. *Immunity* **29**, 228–237 (2008).
64. Yoder, J.A. Form, function and phylogenetics of NITRs in bony fish. *Dev. Comp. Immunol.* **33**, 135–144 (2009).
65. Kawasaki, K., Suzuki, T. & Weiss, K.M. Phenogenetic drift in evolution: the changing genetic basis of vertebrate teeth. *Proc. Natl. Acad. Sci. USA* **102**, 18063–18068 (2005).
66. Kawasaki, K. The SCPP gene repertoire in bony vertebrates and graded differences in mineralized tissues. *Dev. Genes Evol.* **219**, 147–157 (2009).
67. Ryll, B., Sanchez, S., Haitina, T., Tafforeau, P. & Ahlberg, P.E. The genome of *Callorhynchus* and the fossil record: a new perspective on SCPP gene evolution in gnathostomes. *Evol. Dev.* **16**, 123–124 (2014).
68. Kawasaki, K. & Amemiya, C.T. SCPP genes in the coelacanth: tissue mineralization genes shared by sarcopterygians. *J. Exp. Zool. B Mol. Dev. Evol.* **322**, 390–402 (2014).
69. Sire, J.Y. Light and TEM study of nonregenerated and experimentally regenerated scales of *Lepisosteus oculatus* (Holostei) with particular attention to ganoine formation. *Anat. Rec.* **240**, 189–207 (1994).
70. Sasagawa, I., Ishiyama, M., Yokosuka, H. & Mikami, M. Fine structure and development of the collar enamel in gars, *Lepisosteus oculatus*, Actinopterygii. *Front. Mater. Sci. China* **2**, 134–142 (2008).
71. Zhu, M. *et al.* The oldest articulated osteichthyan reveals mosaic gnathostome characters. *Nature* **458**, 469–474 (2009).
72. Hertel, J. & Stadler, P.F. The expansion of animal microRNA families revisited. *Life (Basel)* **5**, 905–920 (2015).
73. Desvignes, T., Beam, M.J., Batzel, P., Sydes, J. & Postlethwait, J.H. Expanding the annotation of zebrafish microRNAs based on small RNA sequencing. *Gene* **546**, 386–389 (2014).
74. Kozomara, A. & Griffiths-Jones, S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.* **39**, D152–D157 (2011).
75. Braasch, I. & Postlethwait, J.H. in *Polyploidy and Genome Evolution* (eds. Soltis, P.S. & Soltis, D.E.) Ch. 17, 341–383 (Springer, 2012).
76. Loh, Y.H., Yi, S.V. & Strelman, J.T. Evolution of microRNAs and the diversification of species. *Genome Biol. Evol.* **3**, 55–65 (2011).
77. Grimson, A. *et al.* Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature* **455**, 1193–1197 (2008).
78. Berezikov, E. *et al.* Deep annotation of *Drosophila melanogaster* microRNAs yields insights into their processing, modification, and emergence. *Genome Res.* **21**, 203–215 (2011).
79. Wheeler, B.M. *et al.* The deep evolution of metazoan microRNAs. *Evol. Dev.* **11**, 50–68 (2009).
80. Woolfe, A. *et al.* Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol.* **3**, e7 (2005).
81. Pennacchio, L.A. *et al.* In vivo enhancer analysis of human conserved non-coding sequences. *Nature* **444**, 499–502 (2006).
82. Montavon, T. *et al.* A regulatory archipelago controls *Hox* genes transcription in digits. *Cell* **147**, 1132–1145 (2011).
83. Berlivet, S. *et al.* Clustering of tissue-specific sub-TADs accompanies the regulation of *HoxA* genes in developing limbs. *PLoS Genet.* **9**, e1004018 (2013).
84. Visel, A., Minovitsky, S., Dubchak, I. & Pennacchio, L.A. VISTA Enhancer Browser—a database of tissue-specific human enhancers. *Nucleic Acids Res.* **35**, D88–D92 (2007).
85. Gehrke, A.R. *et al.* Deep conservation of wrist and digit enhancers in fish. *Proc. Natl. Acad. Sci. USA* **112**, 803–808 (2015).
86. Zakany, J. & Duboule, D. The role of *Hox* genes during vertebrate limb development. *Curr. Opin. Genet. Dev.* **17**, 359–366 (2007).
87. Schneider, I. & Shubin, N.H. The origin of the tetrapod limb: from expeditions to enhancers. *Trends Genet.* **29**, 419–426 (2013).
88. Andrey, G. *et al.* A switch between topological domains underlies *HoxD* genes collinearity in mouse limbs. *Science* **340**, 1234167 (2013).
89. Force, A. *et al.* Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**, 1531–1545 (1999).
90. Stoltzfus, A. On the possibility of constructive neutral evolution. *J. Mol. Evol.* **49**, 169–181 (1999).
91. Postlethwait, J., Amores, A., Cresko, W., Singer, A. & Yan, Y.L. Subfunction partitioning, the teleost radiation and the annotation of the human genome. *Trends Genet.* **20**, 481–490 (2004).

92. He, X. & Zhang, J. Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics* **169**, 1157–1164 (2005).
93. Vilella, A.J. *et al.* EnsemblCompara GeneTrees: complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res.* **19**, 327–335 (2009).
94. Catchen, J.M., Conery, J.S. & Postlethwait, J.H. Automated identification of conserved synteny after whole-genome duplication. *Genome Res.* **19**, 1497–1505 (2009).
95. Ohno, S. *Evolution by Gene Duplication* (Springer-Verlag, 1970).
96. Scannell, D.R. & Wolfe, K.H. A burst of protein sequence evolution and a prolonged period of asymmetric evolution follow gene duplication in yeast. *Genome Res.* **18**, 137–147 (2008).
97. Chang, N. *et al.* Genome editing with RNA-guided Cas9 nuclease in zebrafish embryos. *Cell Res.* **23**, 465–472 (2013).
98. Hwang, W.Y. *et al.* Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat. Biotechnol.* **31**, 227–229 (2013).
99. Krzywinski, M. *et al.* Circo: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).

¹Institute of Neuroscience, University of Oregon, Eugene, Oregon, USA. ²Department of Organismal Biology and Anatomy, University of Chicago, Chicago, Illinois, USA. ³Department of Biology, University of Kentucky, Lexington, Kentucky, USA. ⁴Department of Anthropology, Pennsylvania State University, University Park, Pennsylvania, USA. ⁵Institute of Marine Biology, Biotechnology and Aquaculture, Hellenic Centre for Marine Research, Heraklion, Greece. ⁶Institut National de la Recherche Agronomique (INRA), UR1037 Laboratoire de Physiologie et Génomique des Poissons (LPGP), Campus de Beaulieu, Rennes, France. ⁷Department of Animal Biology, University of Illinois, Urbana-Champaign, Illinois, USA. ⁸Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. ⁹Eccles Institute of Human Genetics, University of Utah, Salt Lake City, Utah, USA. ¹⁰Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, UK. ¹¹European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, UK. ¹²Department of Zoology, University of Oxford, Oxford, UK. ¹³School of Biological Sciences, Bangor University, Bangor, UK. ¹⁴Comparative Genomics Laboratory, Institute of Molecular and Cell Biology, Agency for Science, Technology and Research (A*STAR), Singapore. ¹⁵Institut de Génomique Fonctionnelle de Lyon, Ecole Normale Supérieure de Lyon, Lyon, France. ¹⁶Department of Biology, University of Konstanz, Konstanz, Germany. ¹⁷Department of Molecular Biomedical Sciences, North Carolina State University, Raleigh, North Carolina, USA. ¹⁸Center for Comparative Medicine and Translational Research, North Carolina State University, Raleigh, North Carolina, USA. ¹⁹Departament de Genètica, Universitat de Barcelona, Barcelona, Spain. ²⁰Institut de Recerca de la Biodiversitat, Universitat de Barcelona, Barcelona, Spain. ²¹Department of Biology, University of Victoria, Victoria, British Columbia, Canada. ²²Center for Circadian Clocks, Soochow University, Suzhou, China. ²³School of Biology and Basic Medical Sciences, Medical College, Soochow University, Suzhou, China. ²⁴Bioinformatics Group, Department of Computer Science, Universität Leipzig, Leipzig, Germany. ²⁵Department of Dental Hygiene, Nippon Dental University College at Niigata, Niigata, Japan. ²⁶Department of Pediatrics, University of South Florida Morsani College of Medicine, St. Petersburg, Florida, USA. ²⁷Department of Microbiology, Nippon Dental University School of Life Dentistry at Niigata, Niigata, Japan. ²⁸Department of Evolutionary Studies of Biosystems, SOKENDAI (Graduate University for Advanced Studies), Hayama, Japan. ²⁹Molecular Genetics Program, Benaroya Research Institute, Seattle, Washington, USA. ³⁰Department of Biological Sciences, Nicholls State University, Thibodaux, Louisiana, USA. ³¹Instituto de Ciências Biológicas, Universidade Federal do Pará, Belem, Brazil. ³²International Max Planck Research School for Organismal Biology, University of Konstanz, Konstanz, Germany. ³³Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden. ³⁴Present 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, UK (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.) and Vertebrate and Health Genomics, Genome Analysis Center, Norwich, UK (F.D.P.). Correspondence should be addressed to I.B. (braasch@msu.edu) or J.H.P. (jpostle@uoneuro.uoregon.edu).

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, Louisiana (**Supplementary Note**). It was sequenced using Illumina sequencing technology and jumping libraries to 90× coverage and assembled into LepOcu1 (GenBank accession [AHAT00000000.1](#)) using ALLPATHS-LG²¹. 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 those of other vertebrate Illumina genome assemblies²¹. A total of 209 scaffolds were anchored in 29 linkage groups using 2,153 of 8,406 meiotic map restriction site-associated DNA (RAD) tag markers²⁰, thus capturing 891 Mb of sequence or 94.2% of bases in the chromosome assembly (**Supplementary Note**).

RNA-seq transcriptomes. The Broad Institute gar RNA-seq transcriptome (**Supplementary Note**) was generated from ten tissues (stage 28 embryo¹⁰⁰, 8-day larvae, eye, liver, heart, skin, muscle, kidney, brain and testis) and assembled using Trinity¹⁰¹. PhyloFish RNA-seq transcriptomes of gar, bowfin, zebrafish and medaka (**Supplementary Note**) were generated from ten adult tissues (ovary, testis, brain, gills, heart, muscle, liver, kidney, bone and intestine) and one embryonic stage ('pigmented eye' stage of gar, zebrafish and medaka) and assembled using the Velvet/Oases package¹⁰².

Genome annotation. Using evidence from the Broad Institute and PhyloFish gar transcriptomes (**Supplementary Note**), all RefSeq teleost proteins and all UniProt/SwissProt proteins, MAKER2 (ref. 23) annotated 25,645 protein-coding genes (**Supplementary Note**). Using the Broad Institute transcriptome, the Ensembl gene annotation pipeline identified 18,328 protein-coding genes for 22,470 transcripts along with 42 pseudogenes and 2,595 noncoding RNAs (**Supplementary Note**). Annotations for 762 and 6,877 genes are specific to Ensembl and MAKER, respectively. The gene set with 21,443 high-confidence genes predicted by MAKER likely has close to the true number of gar protein-coding genes.

Annotation of transposable elements. Manual and automated classification (using RepeatScout and RepeatModeler) of gar TEs was performed on the basis of Wicker's nomenclature¹⁰³, and identified elements were combined into a single library (**Supplementary Note**), which was then used to mask the genome with RepeatMasker. The TE age profile was determined using the Kimura distances of individual TE copies to the corresponding TE consensus sequence (**Supplementary Note**).

Phylogenomic and evolutionary rate analyses. Phylogenetic analyses (**Supplementary Note**) were based on protein-coding sequence alignments described for the coelacanth genome analysis¹⁷ but updated with orthologous sequences from gar and bowfin (**Supplementary Note**) and from the slowly evolving Western painted turtle¹⁰⁴. Phylogenetic reconstructions were carried out with RAxML¹⁰⁵ and PhyloBayes MPI¹⁰⁶. Molecular rate analyses (**Supplementary Note**) were performed at the protein alignment level with Tajima's relative rate tests¹⁰⁷ and at the level of the reconstructed phylogenies with two-cluster tests¹⁰⁸.

Genome structure analyses. The spotted gar karyotype was determined from caudal fin fibroblast cell cultures established as described for zebrafish¹⁰⁹ (**Supplementary Note**). Analyses of conserved synteny between gar, tetrapods (human and chicken) and teleosts (**Supplementary Note**) were performed with (i) Circos plots⁹⁹ on the basis of orthology relationships from Ensembl 75 and as described in the **Supplementary Note**; (ii) the Synteny Database⁹⁴ after integration of the gar genome assembly (Ensembl version 74); and (iii) comparative synteny maps derived as described in refs. 17,110.

Gene family analyses. Individual gene families were analyzed as described in the **Supplementary Note**. RT-PCR and sequencing was performed to annotate

and analyze gene expression of Scpp mineralization-related genes using cDNA libraries from gar teeth, jaw and scales (**Supplementary Note**).

miRNA annotation and analysis. Gar miRNAs were studied *in silico* (**Supplementary Note**) by BLAST comparison of teleost and tetrapod miRNAs from miRBase^{74,111–113} against the gar genome assembly and confirmed with RNAfold¹¹⁴ (see also ref. 72). miRNA annotation and analyses based on the sequencing data of gar miRNAs (**Supplementary Note**) were performed as described for zebrafish⁷³ by using small RNA-seq data from adult brain, heart, testis and ovary, which were processed and annotated with Prost! (ref. 115) according to miRNA gene nomenclature guidelines¹¹⁶; miRNA orthologies based on conserved synteny were established using Ensembl¹¹⁷, the Synteny Database⁹⁴ and Genomicus^{118,119}.

Analysis of conserved noncoding elements. Investigation of CNEs in developmental gene loci was performed using SLAGAN¹²⁰ in VISTA¹²¹ (**Supplementary Note**). Gar-, zebrafish- and human-centric 13-way multi-genome alignments were generated with MultiZ¹²² on the basis of lastZ¹²³ pairwise whole-genome alignments. We used phyloFit¹²⁴ to generate a neutral model of the evolution of fourfold-degenerate sites to identify conserved elements with phastCons¹²⁴; genic elements and repetitive sequences were filtered out to obtain CNEs. Evolution of human limb enhancers^{33,82–84} was established using whole-genome alignments and conserved synteny curation. Genome-wide connectivity of CNEs and embedded GWAS SNPs from human to zebrafish through to gar was established from whole-genome alignments using liftOver¹²⁵ and BEDtools¹²⁶ (**Supplementary Note**).

HoxD enhancer functional analysis. Gar and teleost orthologs of the HoxD early enhancer CNS65 were identified with VISTA (LAGAN)¹²¹. Gar and zebrafish CNS65 elements were cloned into pXIG-cFos-eGFP and Gateway-Hsp68-LacZ vectors for zebrafish¹²⁷ and mouse (Cyagen Biosciences) transgenesis, respectively (**Supplementary Note**).

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 Ensembl Compara GeneTrees⁹³ (Ensembl 74) and conserved synteny data from the Synteny Database⁹⁴ (**Supplementary Note**). For all three species, RNA-seq reads from the PhyloFish transcriptomes (**Supplementary Note**) were mapped against the longest Ensembl reference coding sequence of each gene with BWA-Bowtie^{128,129}, counted with SAMtools¹³⁰ and normalized for each gene across the 11 tissues using DESeq¹³¹. The correlation of expression patterns and relative levels of expression between each zebrafish or medaka gene and its gar ortholog and of singletons, ohnolog 1, ohnolog 2 and ohnolog pairs was determined using R (ref. 132). See the **Supplementary Note** for additional information, including the definition of ohnolog pair expression and criteria for the detection of neofunctionalization and subfunctionalization.

100. Long, W.L. & Ballard, W.W. Normal embryonic stages of the longnose gar, *Lepisosteus osseus*. *BMC Dev. Biol.* **1**, 6 (2001).
101. Grabherr, M.G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
102. Schulz, M.H., Zerbino, D.R., Vingron, M. & Birney, E. Oases: robust *de novo* RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* **28**, 1086–1092 (2012).
103. Wicker, T. *et al.* A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **8**, 973–982 (2007).
104. Shaffer, H.B. *et al.* The Western painted turtle genome, a model for the evolution of extreme physiological adaptations in a slowly evolving lineage. *Genome Biol.* **14**, R28 (2013).
105. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
106. Lartillot, N., Rodrigue, N., Stubbs, D. & Richer, J. PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Syst. Biol.* **62**, 611–615 (2013).
107. Tajima, F. Simple methods for testing the molecular evolutionary clock hypothesis. *Genetics* **135**, 599–607 (1993).
108. Takezaki, N., Rzhetsky, A. & Nei, M. Phylogenetic test of the molecular clock and linearized trees. *Mol. Biol. Evol.* **12**, 823–833 (1995).
109. Amores, A. & Postlethwait, J.H. Banded chromosomes and the zebrafish karyotype. *Methods Cell Biol.* **60**, 323–338 (1999).

110. Smith, J.J. *et al.* Sequencing of the sea lamprey (*Petromyzon marinus*) genome provides insights into vertebrate evolution. *Nat. Genet.* **45**, 415–421 (2013).
111. Griffiths-Jones, S., Saini, H.K., van Dongen, S. & Enright, A.J. miRBase: tools for microRNA genomics. *Nucleic Acids Res.* **36**, D154–D158 (2008).
112. Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A. & Enright, A.J. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.* **34**, D140–D144 (2006).
113. Griffiths-Jones, S. The microRNA Registry. *Nucleic Acids Res.* **32**, D109–D111 (2004).
114. Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **6**, 26 (2011).
115. Batzel, P., Desvignes, T., Sydes, J., Eames, B.F. & Postlethwait, J.H. Prosti!, a tool for miRNA annotation and next generation smallRNA sequencing experiment analysis. *Zenodo* doi:10.5281/zenodo.35461 (20 November 2015).
116. Desvignes, T. *et al.* miRNA nomenclature: a view incorporating genetic origins, biosynthetic pathways, and sequence variants. *Trends Genet.* **31**, 613–626 (2015).
117. Flicek, P. *et al.* Ensembl 2013. *Nucleic Acids Res.* **41**, D48–D55 (2013).
118. Muffato, M., Louis, A., Poisnel, C.E. & Roest Crolius, H. Genomicus: a database and a browser to study gene synteny in modern and ancestral genomes. *Bioinformatics* **26**, 1119–1121 (2010).
119. Louis, A., Muffato, M. & Roest Crolius, H. Genomicus: five genome browsers for comparative genomics in eukaryota. *Nucleic Acids Res.* **41**, D700–D705 (2013).
120. Brudno, M. *et al.* Glocal alignment: finding rearrangements during alignment. *Bioinformatics* **19** (suppl. 1), i54–i62 (2003).
121. Frazer, K.A., Pachter, L., Poliakov, A., Rubin, E.M. & Dubchak, I. VISTA: computational tools for comparative genomics. *Nucleic Acids Res.* **32**, W273–W279 (2004).
122. Blanchette, M. *et al.* Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res.* **14**, 708–715 (2004).
123. Harris, R.S. *Improved Pairwise Alignment of Genomic DNA*. PhD thesis, Penn. State Univ. (2007).
124. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).
125. Hinrichs, A.S. *et al.* The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.* **34**, D590–D598 (2006).
126. Quinlan, A.R. BEDTools: the Swiss-army tool for genome feature analysis. *Curr. Protoc. Bioinformatics* **47**, 11.12.11–11.12.34 (2014).
127. Fisher, S. *et al.* Evaluating the biological relevance of putative enhancers using Tol2 transposon-mediated transgenesis in zebrafish. *Nat. Protoc.* **1**, 1297–1305 (2006).
128. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
129. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
130. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
131. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
132. R Development Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2015).