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

Ingo Braasch1,34, Andrew R Gehrke2, Jeramiah J Smith3, Kazuhiko Kawasaki4, Tereza Manousakis5, Jeremy Pasquier6, Angel Amores1, Thomas Desvignes1, Peter Batzel1, Julian Catchen7, Aaron M Berlin8, Michael S Campbell9,34, Daniel Barrell10,11, Kyle J Martin12,13, John F Mulley12, Vyvianathan Ravi14, Alison P Lee14, Tetsuya Nakamura2, Domitille Chalopin15,34, Shaohua Fan16,34, Dustin Wcisel17,18, Cristian Cañestro19,20, Jason Sydes1, Felix E G Beaudry21, Yi Sun22,23, Jana Hertel24,34, Michael J Beam1, Mario Fasold24,34, 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 Wang22,23, John S Taylor21, Quenton Fontenot30, Allyse Ferrara30, Stephen M J Searle10, Bronwen Aken10,11, Mark Yandell9, Igor Schneider31, Jeffrey A Yoder17,18, Jean-Nicolas Volff15, Axel Meyer16,32, Chris T Amemiya29, Byrappa Venkatesh14, Peter W H Holland12, Yann Guiguens6, Julien Bobe6, Neil H Shubin2, Federica Di Palma8,34, Jessica Alföldi8, Kerstin Lindblad-Toh8,33 & 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 gene families 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 humans13,14; (ii) the TGD, which resulted in duplicates of many human genes15,16; and (iii) rapid teleost sequence evolution17,18, often due to asymmetric rates of ohnolog evolution, that frustrates ortholog identification. To help connect teleost biomecience 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 teleosts19,20 (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 duplicated gar gene families 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.

RESULTS

Genome assembly and annotation

The genome of a single adult gar female collected in Louisiana was sequenced to 90x 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; corrected after print 25 April 2016; doi:10.1038/ng.3526
to those for other vertebrate Illumina assemblies\(^ {23} \). To generate a ‘chromosome’ (chromosome-level genome assembly\(^ {22} \)) we anchored scaffolds to a meiotic map\(^ {20} \), capturing 94% of assembled bases in 29 linkage groups (LGs) (Supplementary Note). Transcriptomes from adult tissues and developmental stages (Supplementary Note) facilitated the construction of a gene set annotated by MAKER\(^ {23} \) of 21,443 high-confidence protein-coding genes and Ensembl annotation identified 18,328 protein-coding genes (mostly a subset of the MAKER annotations), 42 pseudogenes and 2,595 noncoding RNAs (Supplementary Note), in comparison to human (20,296 protein-coding genes) and zebrafish (25,642 protein-coding genes). 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\(^ {24} \), thus clarifying TE phylogenetic origins (Supplementary Figs. 2–5, Supplementary Tables 1–3 and Supplementary Note).

The gar lineage evolved slowly

Phylogenies of 243 one-to-one orthologs in 25 jawed vertebrates\(^ {17} \), including the gar genome and our transcriptome of the bowfin Amia calva (Supplementary Note and Supplementary Data Set), strongly supported the monophyly of Holostei (gar and bowfin) as the sister group to teleosts (Fig. 1b, Supplementary Fig. 6 and Supplementary Note)\(^ {25–28} \), suggesting that morphologies shared by bowfin and teleosts\(^ {29,30} \) may be convergent or may be ancestral traits that were altered in the gar lineage.

Darwin applied his term ‘living fossil’ to ‘ganoid fishes’, including gars\(^ {31} \); indeed, gars show low rates of speciation and phenotypic evolution\(^ {32} \). Evolutionary rate analyses using cartilaginous fish outgroups showed that gar and bowfin proteins have evolved significantly slower than teleost sequences. Holostei had a substantially shorter branch length to the cartilaginous outgroup than most other bony vertebrates except coelacanth, the slowest evolving bony vertebrate\(^ {17,33} \) (Fig. 1b, Supplementary Table 4 and Supplementary Note). Our results support the hypothesis that the TGD could have facilitated the high rate of teleost sequence evolution\(^ {17,18,34} \). Gar TEs also showed a low turnover rate as compared to TEs in teleosts, mammals and even coelacanth\(^ {24} \) (Supplementary Fig. 5 and Supplementary Note).
Gar informs the evolution of bony vertebrate karyotypes

Gar represents the first chromosomem 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 Ohno hypothesized, consistent with microchromosomes in coelacanth and cartilaginous fishes, for which no chromonomes 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 TGD. For each gar chromosome segment, teleosts usually have two or more homologous segments, verifying gar-teleost divergence before the TGD. 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 lineage 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 ~ 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 and frustrates the translation of knowledge from tetrapod biomedical models to human biology. 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
Hox gene has been completely lost in gar since divergence from the last common ray-finned ancestor. The *hoxa14* gene, missing from teleosts but present in paddlefish, 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 *hoxl2*, zebrafish lacks all HoxDb cluster protein-coding genes and percomorphs lack the HoxCb cluster, 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 gene missing from teleosts (Supplementary Fig. 14). Gar possesses the VGD1/VGD2 ohnolog *pdx2*, previously found only in cartilaginous fishes and coelacanth, indicating that *pdx2* was lost independently in tetrapods and teleosts (Supplementary Fig. 14 and 15). Retinoic acid regulates Hox cluster gene expression, but retinoic acid–synthesizing Aldh enzymes (Supplementary Table 5 and Supplementary Note) vary in number among vertebrates: tetrapods have three genes (*Aldh1a1, Aldh1a2* and *Aldh1a3*), zebrafish has two genes (*aldh1a2* and *aldh1a3*) and medaka has just one (*aldh1a2*). Finding all three genes in gar rules out the hypothesis 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. Analyses of gar circadian clock (Supplementary Fig. 17, Supplementary Table 6 and Supplementary Note) and *opsin* (Supplementary Fig. 18, Supplementary Table 7 and Supplementary Note) 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 teleosts.

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. In gar, at least one pair of class I and class II genes is linked as in tetrapods, 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, 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).
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 tropicalis61, are downstream of VH and JH 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 alloreognition and were thought to be teleost specific63,64. 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, enamelled and enamel), yet the gene repertoires for the secretory calcium-binding phosphoproteins (Scpp) that form these tissues65,66 differ substantially between teleosts and tetrapods and their evolution remains controversial18,67,68. 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 enamelin69, and in its teeth, which are covered by both enamelled 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 teeth66,68. We identified ambn and enam genes (but no ortholog for Amel) in the gar genome and transcriptomes, a conclusion drawn by others using our data released before publication133. The gar ambn and enam genes show sequence similarity to zebrafish scpp6 (ref. 133) 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 enamel71,133, 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 vertebrates68,133.

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 exant 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 specific18,72 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-zebrafish73,74 comparisons showed that four families and four individual miRNA genes emerged in teleosts. Of the 22 families thought to have been lost
in teleosts\textsuperscript{18}, 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 teleosts\textsuperscript{74} 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 teleosts\textsuperscript{18}, 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\textsuperscript{76–79}.

Gar highlights hidden orthology of cis-regulatory elements

CNEs often function as cis-acting regulators\textsuperscript{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 Irx8) 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\textsuperscript{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 orthologs (Supplementary Table 23 and Supplementary Note). (c) Neofunctionalized orthologs 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 ortholog with the gar pattern. The Supplementary Note lists neofunctionalization and subfunctionalization criteria. (e-h) Expression conservation for orthologs 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 ortholog pairs and gar genes was greater than that for orthologs alone and than that for singletons, indicating sharing of ancestral subfunctions by the ortholog pair (multiple Wilcoxon Mann-Whitney tests with Bonferroni correction, α = 0.05 for significance). (g,h) Mean log₁₀-transformed ratios of expression levels for gar genes and teleost ortholog(s). In comparison to gar genes, individual orthologs were expressed at significantly lower levels than singletons; ortholog pair/gar ratios were not statistically different from singleton/gar ratios, suggesting that the aggregate expression level of ortholog pairs approaches the expression level of the preduplication gene (multiple two-sided Student’s t test with Bonferroni correction, α = 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 transition.

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, respectively. Early-phase HoxD expression in fins and limbs shows several features that are presumed to be homologous and may derive from shared but cryptic regulatory elements. The CNS39 and CNS65 elements drive early-phase HoxD activation in mammals (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 functions89–92. 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 sub-functions 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 subfunctionalization89,90,96.

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 editing97,98. 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://syntenyd.db.uoregon.edu/synteny_db/; PhyloFish Portal, http://phylofish.sigenae.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_0000242695.1. RNA-seq data are available from the Sequence Read Archive (SRP) under accession SRP042013 (Broad Institute gar transcriptome), SRP044781–SRP044784 (PhyloFish transcriptomes of zebrafish, gar, bowfin and medaka) and SRP063942 (small RNA-seq for miRNA annotation). Gar Sccp gene sequences are available from GenBank under accession 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 BayouSphere 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 OD01119004 (J.H.P.); a Feodor Lynen Fellowship from the Alexander von Humboldt Foundation and the Volkswagen Foundation Initiative Evolutionary Biology, grant 1/84 815 (L.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 (L.S.); the Brimson Foundation and the University of Chicago Biological Sciences Division (N.H.S.); NSF grant BCS0725227 (K.K.); and ARISTEIA of the National Strategic Reference Framework 2007–2013 (SPARCMP36), Ministry of Education and Religious Affairs of Greece (T.M.); Agence Nationale de la Recherche (ANR) grant ANR–10–GENM–017 (PlyPhish; J.B.); the Wellcome Trust grants WT0959908 and WT098051; 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.T.A.); 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) (31060062) (H.W.).

AUTHOR CONTRIBUTIONS

J.H.P., L.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.P.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. M.S.C., M.Y., D.B., S.M.J.S. and B.A. carried out CNE analyses for developmental gene loci. I.B., P.B., J.H.P., Y.G. and J.B. performed comparative gene expression analysis of gar, medaka and zebrafish. L.B. and J.H.P. wrote the paper.

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 articles 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/.

1. Nelson, J.S. Fishes of the World 4th edn (John Wiley, 2006).
2. Kettleborough, R.N. et al. Scripted genome-wide analysis of zebrafish protein-coding gene function. Nature 496, 494–497 (2013).
3. Patton, E.E., Mathers, M.E. & Schartl, M. Generating and analyzing fish models of melanoma. Methods Cell Biol. 105, 359–366 (2011).
4. Revez, A.M. et al. Characterizing the expression variation of an hydrotrope receptor (APR) loci in populations of Atlantic killifish (Fundulus heteroclitus) inhabiting polluted and reference habitats. BMC Evol. Biol. 14, 6 (2014).
5. Lee, O., Green, J.M. & Tyler, C.R. Transgenic fish systems and their application in ecotoxicology. Crit. Rev. Toxicol. 45, 124–141 (2015).
6. Albertson, D.C., Cresko, W., Debich, H.W.I. & Postlethwait, J.H. Evolutionary mutant models for human disease. Trends Genet. 25, 74–81 (2009).
7. Harel, I. et al. A platform for rapid exploration of aging and diseases in a naturally short-lived vertebrate. Cell 160, 1013–1026 (2015).
8. Pagán, A.J. & Ramakrishnan, L. Immunity and immunopathology in the tuberculosis granuloma. Cold Spring Harb. Perspect. Med. 5, a018495 (2014).
9. 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, 252–259 (2014).
10. Dehail, P. & Boore, J.L. Two rounds of whole genome duplication in the ancestral vertebrate phylum. PLoS Biol. 3, e305 (2005).
11. Smith, J.J. & Keinath, M.C. The sea lamprey meiotic map improves resolution of ancient vertebrate genome duplications. Genome Res. 25, 1081–1090 (2015).
12. Wolfe, K. Robustness—it’s not where you think it is. Nat. Genet. 25, 3–4 (2000).
13. Frankenberg, S.R. et al. The POU-er of gene nomenclature. Development 141, 2921–2923 (2014).
14. Braasch, I. et al. Connectivity of vertebrate genomes: paired-related homeobox (Pox) genes in spotted gar, basal teleosts, and tetrapsods. Comp. Biochem. Physiol. C Toxicol. Pharmacol. 163, 24–36 (2014).
15. Anoors, A. et al. Zebrafish fox clusters and vertebrate genome evolution. Science 282, 1711–1714 (1998).
16. 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).
17. Amemiya, C.T. et al. The African cichlid genome provides insights into tetrapod evolution. Nature 496, 311–316 (2013).
18. Venkatase, B. et al. Elephant shark genome provides unique insights into gnathostome evolution. Nature 505, 174–179 (2014).
19. Hogg, 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).
20. Anoors, A., Catchen, J., Frickey, T., 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).
21. Gniece, S. et al. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc. Natl. Acad. Sci. USA 108, 1513–1518 (2011).
22. Braasch, I. et al. A new model army: emerging fish models to study the genomics of vertebrate Evo-Devo. J. Exp. Zool. 324, 316–341 (2015).
23. Holt, C. & Yandell, M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics 12, 491 (2011).
24. Chalopin, D., Naville, M., Pfard, F., Galiana, D. & Voltz, J.N. Comparative analysis of transposable elements highlights mobile diversity and evolution in vertebrates. Genome Biol. Evol. 7, 567–580 (2015).
25. Near, T.J. et al. Resolution of ray-finned fish phylogeny and timing of diversification. Proc. Natl. Acad. Sci. USA 109, 13698–13703 (2012).
26. Belancour, R. et al. The tree of life and a new classification of bony fishes. PLoS Curr. doi:10.1371-currents.toll.53ba26640df0caca7e790b165cb26288 (18 April 2013).
A switch between topological domains underlies a bony fish immunological receptor of the NITR

Clustering of tissue-specific sub-TADs accompanies the

Highly conserved non-coding sequences are associated with...
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. Kryzowski, M. et al. Circos: an information aesthetic for comparative genomics. Genome Res. 19, 1639–1645 (2009).
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 90x coverage and assembled into LepOcu1 (GenBank accession AHAT00000001) using ALLPATHS-LG51. 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 assemblies52. 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 markers59, 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 embryo120, 8-day larvae, eye, liver, heart, skin, muscle, kidney, brain and testis) and assembled using Trinity60. 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. The spotted gar genome assembly was used as anchor for MAKER2 (ref. 22) and gene-prediction analysis was performed in the Broad Institute transcriptome. 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).

Analysis of conserved noncoding elements. Investigation of CNEs in developmental gene loci was performed using SLAGAN120 in VISTA121 (Supplementary Note). Gar, zebrasfish- and human-centric 13-way multi-genome alignments were generated with MulTI22 on the basis of lastZI23 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, 125. Genes and repetitive elements were filtered out to obtain CNEs. Evolution of human limb enhancers132–134 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 BEDTool126 (Supplementary Note).

HoxD enhancer functional analysis. Gar and teleost orthologs of the HoxD early enhancer CNS65 were identified with VISTA (LAGAN)122. Gar and zebrafish CNS65 elements were cloned into pXIG–Cfos–eGFP and Gateway-Hsps68–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 GeneTrees53 (Ensembl 74) and conserved synteny data from the Synteny Database34 (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- Bowtie128, 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.

miRNA annotation and analysis. Gar miRNAs were studied in silico (Supplementary Note) by BLAST comparison of teleost and tetrapod miRNAs from miRBase74–111–113 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 Genomics118, 119.

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 two pairwise whole-genome alignments without a reference genome. Nat. Biotechnol. 29, 644–652 (2011).

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).

Stamatakis, A. RAXML version B: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313 (2014).

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).

Tajima, F. Simple methods for testing the molecular evolutionary clock hypothesis. Genetics 135, 599–607 (1993).

Takezaki, N., Rzhetsky, A. & Nei, M. Phylogenetic test of the molecular clock and linearized trees. Mol. Biol. Evol. 12, 823–833 (1995).

Amores, A. & Postlethwait, J.H. Banded chromosomes and the zebrafish karyotype. Methods Cell Biol. 60, 323–338 (1999).

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 Sccp mineralization-related genes using cDNA libraries from gar teeth, jaw and scales (Supplementary Note).
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. Prost!, 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 Crollius, 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 Crollius, 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).

133. Qu, Q., Haitina, T., Zhu, M. & Ahlberg, P.E. New genomic and fossil data illuminate the origin of enamel. Nature 526, 108–111 (2015).
Corrigendum: The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons

Ingo Braasch, Andrew R Gehrke, Jeramiah J Smith, Kazuhiko Kawasaki, Tereza Manousaki, Jeremy Pasquier, Angel Amores, Thomas Desvignes, Peter Batzel, Julian Catchen, Aaron M Berlin, Michael S Campbell, Daniel Barrell, Kyle J Martin, John F Mulley, Vydiyathan Ravi, Alison P Lee, Tetsuya Nakamura, Domitille Chalopin, Shaohua Fan, Dustin Weisel, Cristian Cañestro, Jason Sydes, Felix E G Beaudry, Yi Sun, Jana Hertel, Michael J Beam, Mario Fasold, Mikio Ishiyama, Jeremy Johnson, Steffi Kehr, Marcia Lara, John H Letaw, Gary W Litman, Ronda T Litman, Masato Mikami, Tatsuya Ota, Nil Ratan Saha, Louise Williams, Peter F Stadler, Han Wang, John S Taylor, Quenton Fontenot, Allysse Ferrara, Stephen M J Searle, Bronwen Aken, Mark Yandell, Igor Schneider, Jeffrey A Yoder, Jean-Nicolas Volff, Axel Meyer, Chris T Amemiya, Byrappa Venkatesh, Peter W H Holland, Yann Guiguen, Julien Bobe, Neil H Shubin, Federica Di Palma, Jessica Alföldi, Kerstin Lindblad-Toh & John H Postlethwait

Nat. Genet. 48, 427–437 (2016); published online 7 March 2016; corrected after print 25 April 2016

As we intended, other researchers have been able to use the draft spotted gar genome sequence available from the Broad Institute website since December 2011, the assembly LepOcu1 publicly available from NCBI since 13 January 2012 under accession code GCA000242695.1, and the Ensembl gene annotation (version 74, December 2013; http://www.ensembl.org/Lepisosteus_oculatus/Info/Annotation) and recent annotation by NCBI on 15 May 2014 guided by RNA sequence data from ten tissues. While this article was in review, a paper (Nature 526, 108–111, 2015) was published that arrives at conclusions similar to some of our own analyses, and we wish to acknowledge that publication, which used our unpublished data and genome annotations, emphasizing the importance of the strategy of early release of sequence data. The correction has been made to the HTML and PDF versions of the article.