The genome of the bowfin (*Amia calva*) illuminates the developmental evolution of ray-finned fishes

Andrew Thompson

Michigan State University

Michael Hawkins

Harvard University <https://orcid.org/0000-0002-5164-2893>

Elise Parey

Institut de Biologie de l'ENS (IBENS), Département de Biologie, École Normale Supérieure, CNRS, INSERM, Université PSL

Dustin Wcisel

North Carolina State University

Tatsuya Ota

SOKENDAI <https://orcid.org/0000-0002-1115-5169>

Kazuhiko Kawasaki

Pennsylvania State University <https://orcid.org/0000-0003-1090-5340>

Emily Funk

Cornell University

Mauricio Losilla

Michigan State University <https://orcid.org/0000-0003-0695-3120>

Olivia Fitch

Michigan State University

Qiaowei Pan

University of Lausanne <https://orcid.org/0000-0002-8501-0405>

Romain Feron

University of Lausanne <https://orcid.org/0000-0001-5893-6184>

Marine Milhes

INRAE, GeT-PlaGe

Brett Racicot

Michigan State University

Kevin Childs

Michigan State University

Quenton Fontenot

Nicholls State University

Allyse Ferrara

Nicholls State University

Solomon David

Nicholls State University

Amy McCune

Cornell University

Alex Dornburg

North Carolina State University

Jeffrey Yoder

North Carolina State University

Yann Guiguen

National Research Institute for Agriculture, Food and Environment <https://orcid.org/0000-0001-5464-6219>

Hugues Roest Crollius

IBENS <https://orcid.org/0000-0002-8209-173X>

Camille Berthelot

Institut de Biologie de l'ENS (IBENS), Département de Biologie, École Normale Supérieure, CNRS, INSERM, Université PSL

Matthew Harris

Boston Children's Hospital <https://orcid.org/0000-0002-7201-4693>

Ingo Braasch (✉ braasch@msu.edu)

Michigan State University <https://orcid.org/0000-0003-4766-611X>

Article

Keywords: bowfin (*Amia calva*) genome, developmental evolution of ray-finned fishes, neopterygian fishes

Posted Date: October 27th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-92055/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Nature Genetics on August 30th, 2021. See the published version at <https://doi.org/10.1038/s41588-021-00914-y>.

Abstract

The bowfin fish (*Amia calva*) diverged before the genome duplication in teleost fishes, and its archetypical body plan and slow rate of molecular evolution make it a key species for genomic exploration as a basal representative of the neopterygian fishes. To investigate the evolution and development of ray-finned fishes, we generated a chromosome-level genome assembly for bowfin that enables gene-order analyses which settle its long-debated, phylogenetic relationship with gars. We analyze the genomic underpinnings of the bowfin's unique combination of derived and ancestral phenotypes involving the immune system as well as scale, respiratory organ, and skeletal development. By detailing chromatin accessibility and gene expression through bowfin development, we connect developmental gene regulatory loci across vertebrates. We illustrate the utility of these genomic resources to connect developmental evolution across bony fishes, showing the importance of bowfin in understanding vertebrate biology and diversity.

Introduction

The monotypic bowfin, *Amia calva* (Linnaeus, 1776), is an iconic textbook example in comparative anatomy for its prototypical fish body plan and key phylogenetic position^{1,2}. Bowfin biology thus sheds light on the evolution and development of ray-finned fishes and bony vertebrates in general. Ray-finned fishes constitute the most diverse vertebrate lineage with >30,000 living species, of which >96% belong to the teleost fishes (Teleostei)³. Bowfin (Amiiformes) and seven gar species (Lepisosteiformes) represent the extant Holostei, the sister lineage of teleosts fishes, together comprising the Neopterygii⁴⁻⁹. These eight holosteans, however, capture just a minor fraction of this once speciose lineage. The fossil record shows that the biodiversity of holosteans is highly underappreciated, as they were much more abundant in the past and as species-rich as stem teleosts¹⁰.

With a multitude of teleosts and the spotted gar (*Lepisosteus oculatus*) having been sequenced^{7,11}, bowfin represents the last major neopterygian lineage remaining for genomic exploration. Comparing bowfin and gar covers the maximum holostean genomic diversity available to study, and – due to their early divergence during holostean evolution >250 million years ago⁸ – also spans the vast majority of holosteans to ever have existed¹⁰. Importantly, within Neopterygii the relations of bowfin, gars, and teleosts have been matter of a long-standing, controversial debate (e.g.¹²⁻¹⁴) with two main proposed alternative scenarios. Bowfin has been grouped with gars in the Holostei clade with strong phylogenomic support (e.g.⁴⁻⁹), but due to similarities in morphology, bowfin has also been historically grouped together with teleosts as Halecostomi (e.g.¹²⁻¹⁵). This Halecostomi scenario could further reconcile similarities of bowfin and teleost karyotypes¹⁶. The bowfin genome sequence is essential to test these two scenarios.

Bowfin is an indispensable outgroup to investigate teleost evolution. The common ancestor of extant teleosts underwent the teleost whole genome duplication (TGD), resulting in a sizeable fraction of often functionally divergent gene duplicates in living teleosts (e.g.¹⁷⁻²⁰). Genomic comparisons of teleosts with

tetrapods like humans are thus challenging, for example in biomedical studies leveraging teleost models like zebrafish, medaka, killifish, or cavefish. Bowfin and gar diverged from teleosts before the TGD and thus maintained more one-to-one gene relations with tetrapods⁷ and also have slower rates of molecular sequence evolution than teleosts^{7,21}. Thus, the genome of the gar has been leveraged as an intermediate stepping stone or ‘bridge’ for identifying hidden orthologies of genetic elements between teleosts and tetrapods^{7,22,23}. Adding the bowfin genome, the ‘holostean bridge’ will capture more genomic diversity across the ray-finned tree of life and will help resolve critical questions about vertebrate evolution in which analyzing only one holostean representative is insufficient (e.g.^{7,24}).

Bowfin features a unique suite of developmental, immunological, morphological, and behavioral phenotypes² ripe for study with genomic sequence information, including, e.g. a derived type of dermal scales²⁵, air breathing with a gas bladder^{26,27}, and a largely ancestral skeletal configuration^{1,15}. Genomic resources for the bowfin are thus poised to provide key insights into vertebrate developmental evolution. Here, we report a chromosome-level genome assembly that we use to unravel bowfin’s phylogenetic position and chromosomal evolution and investigate key gene families, gene regulatory regions, and developmental processes in bowfin to explore vertebrate biology. We increase the utility of holostean genomic resources by characterizing bowfin developmental chromatin dynamics and analysis of gene expression in development, enabling comparisons of non-coding regulatory regions across model and non-model fishes and tetrapods. The inclusion of the bowfin genomic landscape permits translation of genetic and genomic changes underlying vertebrate evolution and its regulation.

Results

Bowfin genome assembly and annotation

The bowfin genome was sequenced from a single adult male with a *de novo* genome assembly constructed using Meraculous²⁸ and further scaffolded with Chicago²⁹ and Hi-C approaches³⁰ using the HiRise software pipeline²⁹. The final assembly (N50 = 41.2 Mb) consists of 1,958 scaffolds including 23 superscaffolds that contain 99% of the assembly (Supplementary Fig. 1) and match bowfin’s chromosome number^{16,31}, consistent with a chromosome-level genome assembly. The final assembly (AmiCal1, NCBI accession PESF00000000; Supplementary Table 1) is largely complete as reported by CEGMA³² (98.3%: 244/248 eukaryotic CEGs) and BUSCO³³ (100%: 303/303 eukaryotic BUSCOs).

The bowfin genome consists of ~22% repeats (Supplementary Table 2) with a transposable element profile overall similar to that of spotted gar (repeat content ~21%⁷). Using transcriptomic evidence from ten adult tissues^{7,34}, we generated a MAKER³⁵ genome annotation reporting 21,948 protein-coding genes, very similar to spotted gar (21,443)⁷. OrthoFinder³⁶ predicted orthologies for 86.6% of these genes to eleven other vertebrates (Supplementary Fig. 2, Supplementary Tables 3-4). See Supplementary Notes 1-2 for genome assembly and annotation details.

Sex determination in bowfin

Unlike all other living non-teleost ray-finned fishes, bowfin shows prominent sexual dimorphism with male-specific color patterns (Fig. 1A) and complex reproductive behavior of paternal care with nest building and guarding of larvae². Such phenotypic differences among sexes could be based on differentiated sex chromosomes (e.g.³⁷), but cytogenetic sex differences have not been reported for bowfin^{16,31}, and its sex determination system remains unknown. Following a pool-sequencing strategy (Pool-Seq) contrasting 30 males vs. 30 females and using both reference genome-based and -free approaches, we did not find a genomic region exhibiting sex differentiation (Supplementary Fig. 3, Supplementary Table 5, Supplementary Note 3). Thus, sex determination in bowfin is either polyfactorial, too small to be detected using Pool-Seq, and/or based on environmental effects. Notably, no genomic sex differentiation loci have been found using similar approaches for both spotted gar³⁸ and the more distantly related sterlet sturgeon³⁹. Genetic sex determination mechanisms thus remain elusive for non-teleost ray-finned fishes, if they exist.

Genome organization and gene order support the monophyly of holostean fishes

Cytogenetic comparisons suggest that the bowfin karyotype is more similar to those of teleosts than gars¹⁶. Teleosts and bowfin both lack micro-chromosomes in contrast to gar and chicken, which retained them from the bony vertebrate ancestor^{7,40}. This shared absence of micro-chromosomes could imply a closer relationship of bowfin and teleosts (Halecostomi) opposed to the monophyly of bowfin and gars (Holostei). We have previously shown that the teleost ancestor experienced micro-chromosome fusions after its divergence from gar, but before the TGD⁷. We thus investigated if these teleost micro-chromosome fusions are shared with bowfin. By orthology to gar micro-chromosomes conserved from the bony vertebrate ancestor⁷, we identified bowfin and teleost chromosomes corresponding to fused ancestral micro-chromosomes. We find 3 such micro-chromosome fusions for bowfin and 5 for the teleost medaka, with none shared between them (Fig. 1C-D). Bowfin and teleosts thus independently fused different sets of micro-chromosomes and their karyotypic similarities are the result of convergent evolution and do not support the Halecostomi scenario.

We identified 4 chromosome pairs with one-to-one orthology between bowfin and chicken (Supplementary Fig. 4), while there are 14 one-to-one pairs between gar and chicken⁷. Thus, the bowfin karyotype is more derived than that of gar. However, in support of holostean monophyly, chicken chromosomes 13 and 23 correspond to a fusion chromosome in both bowfin and gar, a rearrangement not found in teleosts (Supplementary Fig. 4).

We investigated gene order evolution as a new line of evidence to illuminate bowfin's phylogenetic position. Genomic rearrangements (inversions, transpositions, translocations) introduce gene order differences across species over evolutionary time, but are much rarer than nucleotide substitutions and therefore useful for reconstructing phylogenies⁴¹. Using rearrangement distances, we reconstructed a phylogeny of bowfin, gar, and 10 teleosts, plus two tetrapod outgroups. We leveraged 3,223 marker genes

present in all genomes, estimated pairwise evolutionary distances between species using a normalized breakpoint distance, and constructed a rearrangement-based Neighbor-Joining species tree that supports holostean monophyly with high confidence (Fig. 1E). Focusing on all gene adjacencies acquired since the neopterygian ancestor (i.e. gene adjacencies found in gar, bowfin and/or teleosts, but not in outgroups), we positioned gains and losses of these adjacencies on the two alternative phylogenies using the Dollo algorithm⁴² assuming that a gene adjacency is gained only once, but can be lost numerous times independently⁴³. The Holostei scenario required a significantly lower number of evolutionary changes (760 vs. 1,024, KHT test p-value=0.028358), rejecting the Halecostomi scenario (Fig. 1F).

Thus gene order, in agreement with sequence-based phylogenomic studies⁴⁻⁹ (Supplementary Fig. 2), strongly supports holostean monophyly regardless of methodology. While the bowfin karyotype is more derived than that of gar at a gross chromosomal level, this is not reflected in local gene order, where bowfin and gar present similar levels of species-specific rearrangements (branch length differences are non-significant; see Methods).

Unraveling the holostean immunogenome

Our previous analyses of spotted gar immune genes revealed shared characteristics with those of both teleosts and tetrapods, but left important aspects of the ray-finned fish immunogenome unresolved^{7,44}. For a comprehensive understanding of holostean immune genes, we surveyed the bowfin genome and a newly generated bowfin immune tissue transcriptome (Supplementary Note 4).

The human Major Histocompatibility Complex (MHC) is a cluster of >200 genes, broadly characterized as class I, II, and III genes that play various roles in antigen processing, antigen presentation, and inflammation^{45,46}. Class I and class II genes are tightly linked on one chromosome in cartilaginous fishes and tetrapods (e.g. human chromosome 6, Fig. 2)⁴⁷. In contrast, teleost class I and class II genes are not linked, and class III genes are scattered throughout teleost genomes⁴⁸⁻⁵⁰. The ancestral organization of the ray-finned MHC has remained unresolved because the gar MHC is highly fragmented in the reference assembly⁷. Bowfin, in contrast, has a cluster on superscaffold 14 that contains the majority of class I, II, and III genes (Fig. 2, Supplementary Fig. 5, Supplementary Tables 6-8, Supplementary Note 4). Thus, the overall organization of MHC regions in the holostean and ray-finned fish ancestors was similar to that in tetrapods and cartilaginous fish. Furthermore, the loss of linkage of teleost class I, II, and III genes therefore occurred after their divergence from holosteans and is associated with differential gene loss from TGD-duplicated chromosomes such as seen in the MHC regions on zebrafish chromosomes 19 and 16 (Fig. 2).

Antigen recognition receptors of the adaptive immune system, i.e. immunoglobulin (Ig) and T-cell receptor (TCR), have been identified in all jawed vertebrate lineages, but can differ in their genomic structure and organization^{51,52}. The bowfin genome contains all three canonical TCR loci, and remarkably, encodes not only antibody classes IgM and IgD, but also IgT (IgZ in zebrafish) (Supplementary Figs. 6-9, Supplementary Tables 9-11, Supplementary Note 4). IgT thus far has been considered teleost-specific⁵³⁻⁵⁶

with an important role for teleost mucosal immunity⁵⁷⁻⁵⁹. In contrast to our previous analysis⁷, using the bowfin IgT sequence we confirm an IgT ortholog in spotted gar. Thus, IgT-like antibodies date back to at least the neopterygian ancestor (Supplementary Note 4).

Toll-like Receptors (TLRs) provide initial immune responses to infection in animals⁶⁰. Bowfin possesses 20 TLR genes from all six families, twice the number of human TLRs⁶¹, more than the 16 functional spotted gar TLRs^{7,44}, and almost on par with the more than 20 teleost TLR genes^{62,63} (Supplementary Fig. 10, Supplementary Table 12, Supplementary Note 4). TLR complexity is therefore a more general pattern among neopterygians and the TGD is not the sole evolutionary mechanism leading to large TLR gene numbers in ray-finned fishes. In summary, the holostean immunogenome shows an overall complexity and diversity comparable to teleosts, with a chromosomal organization resembling tetrapods and more distant vertebrate lineages.

Biomineralization genes and the formation of ray-finned fish scales

While bowfin and gar both possess enamel-covered teeth, they prominently differ in scale biomineralization. Gars have ganoid scales, representing an ancestral bony vertebrate scale type covered with a thick layer of hypermineralized ganoin. Bowfin, in contrast, has thin and flexible elasmoid scales that secondarily lost the thick bony plate and ganoin^{12,25}. Teleosts have neither tooth enamel nor scale ganoin. It is thus expected that genes involved in scale formation differ among neopterygians.

The secretory calcium-binding phosphoprotein (SCPP) gene family, generated by complex successive gene duplications during bony vertebrate evolution, encodes proteins involved in biomineralization^{64,65}. We identified 22 SCPP genes in bowfin, 21 of which form two large genomic clusters arranged similarly to gar (Fig. 3, Supplementary File 1, Supplementary Note 5). Like gar^{7,66}, bowfin possesses the enamel-forming SCPP genes *enamelin* (*enam*) and *ameloblastin* (*ambn*) (Fig. 3). Their involvement in the formation of both tooth enamel and scale ganoin in gar supports that these two mineralized structures evolved from a common genetic program^{7,66}. Since bowfin lacks ganoin, this suggests that gene regulatory shifts in *enam* and *ambn* have occurred in the bowfin lineage that silenced their expression during scale development but not dental enamel formation.

The total number of bowfin SCPP genes (22) is considerably smaller than that of gar, which has the largest known SCPP gene repertoire among vertebrates (38)^{7,64}. All 22 bowfin SCPP genes have a gar ortholog. In contrast, orthologs of 16 gar SCPP genes could not be found in bowfin, 9 of which are located in a cluster on gar chromosome LG2 syntenic to a reduced gene cluster on bowfin superscaffold 9 (Fig. 3). Expression of most gar LG2 SCPP genes was weak or undetectable in tooth germs but strong in scale-forming skin⁷. This implies that the SCPP clusters on gar LG2 and bowfin 9 are particularly involved in scale formation, a hypothesis further supported by teleost SCPP genes (Supplementary Fig. 11, Supplementary Note 5). The orthologous zebrafish SCPP genes are located on TGD-duplicate clusters on chromosomes 5 and 10⁷ and expressed during zebrafish skin development (Supplementary Table 13, Supplementary Note 5). Furthermore, these clusters have been reduced to only one gene (*scpp8*) in the

scale-less channel catfish (Fig. 3). These multiple lines of evidence suggest that SSCP genes with major roles in scale formation are clustered in a genomic region that is expanded in the ganoid gar on LG2 and likely secondarily reduced in bowfin on superscaffold 9, corresponding to the formation of their modified elasmoid scales. The reduced biomineralization of the bowfin scale is thus attributed to both changes in gene regulation and the loss of SSCP genes.

Chromatin profiling through bowfin development connects gene regulation across vertebrate morphologies

To increase the power and utility of the holostean genome for comparative studies on vertebrate gene regulation, we used ATAC-Seq⁶⁷ to generate an atlas of open chromatin regions (OCRs) from six developmental stages of wild-caught bowfin embryos and larvae, from the conserved phylotypic stage onwards towards the end of the described larval development⁶⁸ (Supplementary Figs. 12-13, Supplementary Note 6). We identify a total of 163,771 OCRs, of which 80% (132,119) are non-coding OCRs (ncOCRs), and annotate their genomic feature location and nearest gene with HOMER⁶⁹ (Supplementary Tables 14-16). Adding RNA-Seq transcriptomes from the same stages, our dataset provides a rich resource for exploring the correlation between chromatin accessibility of putative gene regulatory elements and gene expression exemplified by bowfin ncOCRs overlapping key enhancers active during development of a variety of vertebrate tissues such as brain⁷⁰, heart²³, anterior and posterior fin/limb⁷¹⁻⁷³, and lungs^{73,74} (Fig. 4).

Conserved Non-coding elements (CNEs) and Ultra Conserved Elements (UCEs) are used to identify candidate enhancers and genetic loci for phylogenetic inference^{7,9}. We find that 33% of both gar-centric vertebrate CNEs⁷ (21,279/64,608) and bowfin UCEs⁹ (122/366) intersect with bowfin ncOCRs (Supplementary Table 17), thus many CNEs/UCEs are likely gene regulatory during bowfin development. To further investigate the comparative efficacy of our ATAC-Seq dataset, we identified the orthologous position of experimentally confirmed mammalian enhancers⁷⁰ in the bowfin genome using sequence conservation (Supplementary Tables 18-19, Supplementary Note 6). Orthologs of 61% of human enhancers (600/989) were found in bowfin and over half (52%; 314/600) overlap with bowfin ncOCRs. In contrast, orthology of only 45% (449/989) of human enhancers could be established in the zebrafish genome (Supplementary Table 18), illustrating the usefulness of the slowly evolving and 'unduplicated' bowfin genome to connect accessible, non-coding, regulatory regions from human to fish. Overlap with >2,000 ncOCRs from a mouse developmental single-nuclei ATAC-Seq atlas⁷⁵ established bowfin OCRs that are likely functioning in a cell/tissue type-specific manner, e.g. in the forebrain, neural crest cells, cardiomyocytes, and many other tissues (Supplementary Tables 20-21).

Currently, beyond zebrafish and medaka (e.g.^{76,77}), there are limited gene regulatory resources for fishes, making comparative analyses across species or whole genome duplications difficult. Our bowfin ATAC-Seq atlas is the first for any non-teleost and provides a critical platform to further connect gene regulatory homology across vertebrates while strengthening the utility of non-teleost fishes for examining genome

function in the context of long-standing questions about vertebrate evolution and development. For example, ever since Owen and Darwin, the homology of vertebrate air-filled organs has been debated⁷⁸. Using our bowfin ATAC-Seq atlas, we identified an intronic ncOCR within the *tbx4* gene (Fig. 4E) that is orthologous to both the previously identified mammalian *tbx4* 'lung enhancer'^{73,74} as well as a previously uncharacterized, conserved intronic region in teleosts (Fig. 4F). Specifically, the 'holostean bridge' establishes orthology of this important developmental enhancer from tetrapods via bowfin and gar to teleosts (Fig. 4F, Supplementary Fig. 14, Supplementary Note 7). Additionally, we show that this ncOCR activity correlates with critical timing of *tbx4* expression during bowfin gas bladder development²⁶ (Fig. 4E, Supplementary Note 7). Very intriguingly, these results support the deep homology among bony vertebrate air-filled organs, i.e. terrestrial lungs, holostean respiratory gas bladders, and buoyancy-controlling swim bladders in teleosts (Supplementary Note 7).

Evolutionary stasis of holostean Hox clusters and the enigmatic *hoxD14* gene

Hox cluster genes play essential roles in patterning the primary body axis during animal development as well the proximo-distal and anterior-posterior axes of vertebrate appendages. As result of the two early rounds of vertebrate genome duplication, non-teleost ray-finned fishes without additional genome duplication such as gar contain four Hox clusters A-D⁷. Highlighting the evolutionary stasis of the holostean lineage, the bowfin genome harbors the same repertoire of 43 bona fide Hox cluster genes as gar (Fig. 5A, Supplementary File 2, Supplementary Note 8), with only one gene lost since the bony vertebrate ancestor and none since the ray-finned and holostean fish ancestors^{7,39}.

The evolution of the vertebrate Hox14 paralog group exhibits unique patterns of conservation, loss, and pseudogenization. Members of this paralogy group are absent from tetrapods and teleosts, but present in lamprey, cartilaginous fishes, paddlefish, sturgeon, and gar^{7,39,79-81}. Gar *hoxd14* is a pseudogene⁷ and we find sequence conservation with bowfin within the intergenic region between *evx2* and *hoxd13*. A large number of bowfin pectoral fin bud RNA-Seq reads (see below) align to this region, which also shows open chromatin status (Supplementary Fig. 15). We cloned two different *hoxd14* transcripts from bowfin pectoral fin bud cDNA (Figure 5B). RNA *in situ* hybridization revealed that bowfin *hoxd14* is expressed in stage 26 pectoral fin buds in a posterior mesenchymal domain (Figure 5C) similar to paddlefish⁸¹. Expression was also observed in the vent, similar to shark and lamprey⁷⁹, and in the tailbud (Figure 5C). Bowfin *hoxd14* is likely a pseudogene, with numerous stop codons and nonsynonymous changes throughout the coding region and homeodomain (Figure 5D). The hallmark Hox14 WFQNQR motif is preserved, however, as is the split homeodomain junction⁸⁰.

Initially, *hox14* genes were thought not to be subject to the ancestral regulatory mechanisms that control *hox1-hox13*^{79,82}. However, paddlefish fin buds express *hoxd14* posteriorly, consistent with spatially collinear Hox regulation⁸¹. Bowfin now supports temporal collinearity of *hoxd14* expression (see below) and its tailbud expression reveals another domain exhibiting global colinear regulation. The bowfin *hoxd14* pseudogene is still transcribed at high levels, spliced, and expressed in highly specific, ancestral

patterns. This discovery raises the possibility that, even though bowfin *hoxd14* likely no longer makes a functional protein, it plays some function in the regulation of *hox* expression and fin patterning and/or other to be identified processes.

Unexpected gene expression dynamics during holostean pectoral fin development

Given their predominance among neopterygians, teleosts have long been used as models for paired fin development and as outgroups for tetrapod limb development. However, the teleost pectoral fin endoskeleton has undergone severe reduction since the ancestral bony vertebrate, having lost the metapterygium and proximal-distal long bone articulation. Bowfin is the closest living teleost relative that retains the metapterygium and proximal-distal elaboration of the fin endoskeleton (Supplementary Fig. 16) and thus serves as an essential node connecting appendicular skeletal diversity of ray-finned fishes and those of lobe-finned vertebrates, including tetrapods.

We performed RNA-Seq across four early patterning stages of bowfin pectoral fin bud development (Fig. 1B, Fig. 6A, Supplementary Fig. 16, Supplementary Note 9), revealing dynamic expression of many known limb and fin patterning genes. The most dramatic expression change in bowfin is in the transcription of structural genes of actinodin and collagen encoding components of the actinotrichia, elastinoid fibrils supporting the early fin fold⁸³⁻⁸⁶ (Fig. 6B). The expression of actinodin genes spike at stage 26, as do type III and type IX collagen components (Fig. 6C). In contrast, *col1a1*, *col1a2* and *col2a1* exhibit relatively stable expression across the bowfin stages examined (Figure 6C), but exhibit dynamic expression and comprise the main structural component of actinotrichia during zebrafish fin development^{87,88}. Type III collagens have not been previously implicated in actinotrichia development and, as they are absent from teleosts genomes⁸⁷, may underlie differences between bowfin and teleost fin morphology.

Given the distinct variation in patterning of the endochondral component between holosteans and teleosts, we focused on changes observed in early patterning genes during bowfin development. Temporal and spatial collinearity of Hox cluster gene expression play critical roles in the patterning of fin and limb skeletons. We find collinearity in bowfin *hox* gene expression levels across early fin bud outgrowth. Peak expression of posterior *hox* genes track with their position within the Hox cluster, including that of *hoxd14* with its highest expression levels observed at stages 25-26 (Figure 6D). Other gene expression patterns are consistent with mechanisms described in limbs. Termination of tetrapod limb bud outgrowth is caused by Twist2, which negatively regulates *grem1* expression⁸⁹. Similarly, bowfin *twist2* expression peaks at stage 26, while *grem1* peaks at stage 23 and is then downregulated ~3-fold by stage 26 (Figure 6D). Other correlated expression profiles in bowfin are consistent with known genetic interactions, such as negative regulation of *meis* genes by Cyp26b1⁹⁰ (Figure 6D).

Surprisingly, we noticed the lack of expression of the critical limb development signaling ligand gene *fgf8* from the bowfin pectoral fin transcriptome at all stages. In contrast, other fin/limb apical ectodermal ridge (AER) markers such as *sp8*, which encodes a transcription factor that directly positively regulates *fgf8*

expression in the AER⁹¹, are highly expressed across all stages. Using RNA *in situ* hybridization, we found expected *sp8* and *fgf8* expression in the central nervous system and tailbud in both bowfin and gar embryos (Figure 6E). However, in embryos of both holosteans, *sp8* is highly expressed in the AER throughout early fin bud development while *fgf8* is not detected (Figure 6E). These findings are shocking as *fgf8* is expressed in every fin and limb bud in which it has been assessed across gnathostome lineages (e.g.⁹²⁻⁹⁶) and is critical for the outgrowth and patterning of these appendages. Even in species that lack a morphological AER, *fgf8* is still expressed⁹⁷⁻⁹⁹. Thus, the lack of *fgf8* from the holostean fin bud is entirely unexpected. Given that *sp8* is expressed in the AER at high levels (RPKM ~20 in all stages), we hypothesize that changes have occurred in the regulation of *fgf8* that remove *sp8* responsiveness in the fin bud, a unique holostean feature. Genomic comparisons (Supplementary Fig. 17, Supplementary Note 8) show that holosteans possess a complex set of known *fgf8* enhancers^{100,101}, some of which are marked as ncOCRs by ATAC-Seq (Supplementary Table 22). We did not detect an obvious enhancer near the *fgf8* locus not utilized or deleted in the bowfin (Supplementary Fig. 17). Since the epistasis in regulation of *fgf8* expression is likely to be complex as Fgf8 signaling at large still likely plays critical roles in development apart from the fin (Fig. 6D), shared enhancer use may be common. Our results reveal an unexpected drift in the genetic control of holostean fin development and represent unique case of robustness of appendage development in the absence of early *fgf8* regulation. As these changes are defining for holosteans and not seen in other fishes or tetrapods, these findings demonstrate the importance of holostean genomic information that challenges long-standing assumptions of core developmental mechanisms and informs new models of appendage patterning.

Discussion

Its phylogenetic position and unique suite of ancestral and derived phenotypes make bowfin an essential component for understanding vertebrate evolution^{1,15}. Our analyses using the chromosome-level genome assembly for this “living fossil”¹⁰² show that, despite its derived karyotype with convergent similarities to those of teleosts, its chromosomal organization nevertheless indicates a closer phylogenetic relationship to gar in the holostean clade. The slow-evolving, ancestrally unduplicated genome and the developmental epigenome of the bowfin will be a critical resource for comparative vertebrate genomics and evolutionary-developmental biology. After genome sequencing, many challenges remain in determining and functionally assaying non-coding gene regulatory loci in both model and non-model species. Here, we have identified putative enhancers for this non-model species and illustrate that our ATAC-Seq dataset can be easily connected to orthologous enhancers in fish and more distant vertebrate lineages. Bowfin not only informs genome evolution in holostean fishes, but also offers valuable insights into the genomic basis of its archetypical anatomy and the many developmental and physiological phenotypes in ray-finned fishes. While this species represents a once large taxonomic group that is now mostly extinct and thus lost for studying genomic diversity, the bowfin genome adds a major sequenced branch for genetic and developmental exploration to the fish tree of life.

Bowfin and gars have persisted for millions of years. Fortunately, the negative perception of holosteans as “trash fishes”¹⁰³ is changing to include greater appreciation for their evolutionary, ecological, and cultural importance¹⁰⁴ and, as shown here, their significance for understanding vertebrate genome biology. As the *Amia* genus may not be monotypic^{2,105}, our genome assembly will be of utmost importance to evaluate bowfin’s species status and conservation.

Methods

Animal work. Bowfin work was approved under Institutional Animal Care and Use Committee (IACUC) protocols from Nicholls State University (#IA046/IA053), Michigan State University (#10/16-179-00), and Cornell University (#2006-0013). Bowfins for genome sequencing, sex determination analysis, and immune transcriptomics were sampled in Louisiana (USA) as detailed see below. Bowfin embryos and larvae for developmental and fin bud transcriptomics, ATAC-Seq, and RNA *in situ* hybridization were collected from nests in Oneida Lake (New York, USA), then raised in the laboratory as described²⁷ (see also Supplementary Note 9) until sampling at the desired developmental stages⁶⁸. Spotted gar embryos were obtained, raised, and fixed as described¹⁰⁶ and approved by the Nicholls State University (#IA053) and the Michigan State University (#10/16-179-00) IACUC.

Genome sequencing and assembly. DNA was extracted from blood of a single wild male, collected from the Atchafalaya River Basin, Louisiana, USA (Supplementary Note 1). A *de novo* assembly was constructed using a paired-end sequencing library (mean assembly-based insert size ~410 bp) with Meraculous v2.2.2.5²⁸ (kmer size = 55). The input data consisted of 433.5 million read pairs sequenced from the paired-end library, totaling 122 Gbp after trimming for quality, sequencing adapters, and mate pair adapters using Trimmomatic v0.38¹⁰⁷.

A Chicago library was prepared as described²⁹. Briefly, ~500ng of HMW gDNA (mean fragment length = 50Kb) was reconstituted into chromatin *in vitro* and fixed with formaldehyde. Fixed chromatin was digested with DpnII, the 5' overhangs filled in with biotinylated nucleotides, and free blunt ends were ligated. After ligation, crosslinks were reversed, and DNA purified from protein. Purified DNA was treated to remove biotin that was not internal to ligated fragments. DNA was then sheared to ~350 bp mean fragment size and sequencing libraries were generated using NEBNext Ultra enzymes and Illumina-compatible adapters. Biotin-containing fragments were isolated using streptavidin beads before PCR enrichment of each library. The libraries were sequenced on an Illumina HiSeq platform to produce 177 million 2x101bp paired end reads, which provided 89.9x physical coverage of the genome (1-50kb pairs).

A HiC library was prepared as described³⁰. For each library, chromatin was fixed in place with formaldehyde in the nucleus and then extracted. Fixed chromatin was digested with DpnII, the 5' overhangs filled in with biotinylated nucleotides, and then free blunt ends were ligated. After ligation, crosslinks were reversed, and the DNA purified from protein. Purified DNA was treated to remove biotin that was not internal to ligated fragments. The DNA was sheared to ~350 bp mean fragment size and

sequencing libraries were generated using NEBNext Ultra enzymes and Illumina-compatible adapters. Biotin-containing fragments were isolated using streptavidin beads before PCR enrichment of each library. The libraries were sequenced on an Illumina HiSeq platform to produce 527 million 2x151 bp paired end reads, which provided 22,887x physical coverage of the genome (1-50kb pairs).

The input *de novo* assembly, shotgun, Chicago library, and HiC library reads were used as input data for HiRise²⁹ to scaffold genome assemblies with an iterative analysis. First, Shotgun and Chicago library sequences were aligned to the draft input assembly using a modified SNAP read mapper [<http://snap.cs.berkeley.edu>]. The separations of Chicago read pairs mapped within draft scaffolds were analyzed by HiRise to produce a likelihood model for genomic distance between read pairs. The model was used to identify and break putative misjoins, to score prospective joins, and make joins above a likelihood threshold. After aligning and scaffolding Chicago data, Dovetail HiC library sequences were aligned and scaffolded following the same method. After scaffolding, shotgun sequences were used to close gaps between contigs to generate the final genome assembly (AmiCal1).

Repeat analysis. A custom bowfin repeat database was constructed using Repeat Modeler v1.0.8¹⁰⁸ and combined with custom repeat libraries for gar⁷, platyfish¹⁰⁹, coelacanth¹¹⁰, and all repeats from the Vertebrata RepBase¹¹¹. These repeat elements were input to RepeatMasker¹¹² within MAKER v2.31³⁵. Repeat proteins were identified by searching against a library packaged with MAKER via RepeatRunner¹¹³.

Gene annotation. Protein-coding genes were annotated with a bowfin reference transcriptome from multiple adult tissues^{7,34} as evidence. The EST2genome function in MAKER v2.31³⁵ was used identify putative bowfin genes based on BLAST¹¹⁴ and Exonerate¹¹⁵ transcript alignments. Best scoring genes with an Annotation Edit Distance of 0.2 or less were used to train Hidden Markov Models with SNAP¹¹⁶ and AUGUSTUS¹¹⁷ as described¹¹⁸. Ensembl and RefSeq protein sequences from other vertebrate species were used as additional evidence: gar (LepOcu1), coelacanth (LatCha1), mouse (GRCm38.p5), chicken (Gallus_gallus-5.0), human (GRCh38.p10), *Xenopus* (JGI 4.2), anole lizard (AnoCar2.0), zebrafish (GRCz10), medaka (HdrR), arowana (GCA_001624265.1 ASM162426v1), and elephant shark (GCA_000165045.2 Callorhinchus_milii-6.1.3). Proteins were aligned to the genome with BLAST and Exonerate, guiding gene predictions by HMMs (from SNAP and AUGUSTUS) in the MAKER workflow. MAKER output all predicted genes with and without transcript and protein evidence (MAKER-Max). Pfam protein domains¹¹⁹ were identified within the MAKER-Max gene set using HMMER v3.0b3 (hmmrscan e-value < 1e-5)¹²⁰. MAKER-Max genes with transcript or protein or Pfam domain support were retained as the final MAKER-Standard gene set¹²¹.

We ran OrthoFinder v1.1.3³⁶ (fastME distance method: -t 20 -M msa -A mafft -T FastTree) to identify orthologous genes between bowfin and other vertebrates using protein sequences of the longest isoforms from the same species set used for MAKER annotation.

Bowfin sex determination. Fin clips were sampled from 30 males and 30 females caught near Thibodaux, Louisiana, after gonadal observation. Individual genomic DNAs were extracted (Qiagen DNeasy Blood & Tissue kit), quantified (NanoDrop spectrophotometer), and mixed in equimolar quantities to produce a male pool and a female pool. Pool-Sequencing (Pool-Seq) libraries were prepared with the Illumina TruSeq Nano DNA HT Library Prep Kit, quality assessed (Advanced Analytical Fragment Analyzer), and quantified by QPCR using (Kapa Library Quantification Kit). Pool-Seq libraries (mean insert size: 419 bp female pool, 412 bp male pool) were sequenced on an Illumina NovaSeq6000 S4 lane (paired-end read length of 2x150 bp), resulting in 283.86 and 284.28 million reads for the female and male pool.

Pool-Seq analyses were carried out using a snakemake workflow [<https://github.com/SexGenomicsToolkit/PSASS-workflow>] and figures generated with SexGenomicsToolkit R package, “SGTR” v1.1.1 [10.5281/zenodo.3773063]. Reads were aligned to the reference genome with BWA mem v0.7.17¹²², and PCR duplicates removed using samtools rmdup. A file containing nucleotide counts for each genomic position was generated with the PSASS pileup command v3.0.1b [47310.5281/zenodo.3702337] and used to compute FST between males and females, numbers of male- and female-specific SNPs, and male and female read depth in a sliding window along the entire genome using the PSASS analyze command. Analyses were run using two sets of parameters: stringent (–window-size 50000, –output-resolution: 1000, –freq-het 0.5, –range-het 0.05, –freq-hom 1, –range-hom 0, and –group-snps set as true), and default (–window-size 50000, –output-resolution: 1000, –freq-het 0.5, –range-het 0.10, –freq-hom 1, –range-hom 0.05, and –group-snps set as true). As a reference-free approach, we searched for sex-biased k-mers in the Pool-Seq data to overcome potential failure to identify sex-specific regions due to a genome reference from the homogametic sex. First, 31-mers were identified and counted in the male and female Pool-Seq reads using the Jellyfish v2.2.10¹²³ (“count” command, option “-C” counting only canonical 31-mers and retaining only k-mers with occurrence 5-50,000). For each sex, 31-mer count tables were obtained with Jellyfish “dump”. 786,071,887 and 786,272,592 31-mers were identified in the female and male reads, respectively. 31-mer count tables were merged and then filtered with Kpool [<https://github.com/SexGenomicsToolkit/kpool>].

Genome structure and gene order analyses. Orthologous genes were extracted from reconciled gene trees built with the Ensembl Compara pipeline¹²⁴. The reconciled gene trees contain 78 species, including 55 fish, 18 other vertebrates, and 5 non-vertebrate outgroups. Briefly, starting with the set of predicted coding sequences, we performed an all-against-all BLAST¹¹⁴, followed by clustering with hcluster_sg¹²⁵ to define gene families, multiple alignment inference using M-Coffee¹²⁶, and phylogenetic tree construction with TreeBeST¹²⁴. From these trees, gene families were defined as groups of genes that derive from the same ancestral gene. Depending on the set of species used, we chose different ancestral species to define the families, each time taking the most recent common ancestor (see below).

We identified pairs of chromosomes between species sharing significantly more orthologs than expected after random gene shuffling ($p < 0.05$, Chi2 test; Yates correction for small sample size, Bonferroni correction for multiple testing). Karyotype comparisons with spotted gar were based on all orthologous

genes inherited from the Neopterygii ancestor (n=16,398 orthologous genes between bowfin and gar; n=14,374 between medaka and gar). Comparisons with chicken were based on all orthologous genes inherited from the Euteleostomi ancestor (n=8,219 orthologous genes between bowfin and chicken; n=7,835 between gar and chicken).

We used a distance-based method to reconstruct the phylogeny of 14 species using gene order data from high-confidence orthologs inherited from the common Euteleostomi ancestor, reducing all studied genomes to 3,223 marker genes: 1,527 were singletons in all species; 1,614 present in one copy in bowfin, gar, chicken, and *Xenopus* and one or two copies in post-TGD teleosts; and 82 present in one copy in all non-teleosts and two copies in all teleosts. Intervals between these marker genes cover >60% of the most fragmented genomes and >80% of all other genomes and are arranged into 16,064 gene adjacencies that exist in at least one study species. Differences in genome assembly qualities and gene duplications resulted in varying numbers of adjacent gene markers ranging from 3,028 (Amazon molly) to 3,800 (arowana).

Normalized breakpoint distances between pairs of genomes were computed as the number of gene adjacencies that exist in G1 but not in G2, normalized to the number of adjacencies in G1, where G1 is the genome with the smallest number of adjacencies, in line with previously reported adjustments^{127,128}. A Neighbor-Joining (NJ) species tree was reconstructed on the resulting distance matrix with bootstrap replicates generated by re-sampling 100 times with replacement from the 16,064 gene adjacencies. Bootstrapped NJ trees were used to evaluate the 95% confidence interval of the branch length difference between gar and bowfin, which was non-significant (95CI: [-0.009, 0.006]). To address the possibility that adjacency losses in teleosts result from loss of TGD duplicate genes instead of rearrangements, we explicitly ignored any adjacency loss for marker genes on alternative TGD-derived regions, which did not affect the results.

Comparing the Holostei vs. Halecostomi scenarios within a maximum parsimony framework, we treated each of the 566 Neopterygian-specific gene adjacencies as an independent character, either absent or present in each Neopterygian genome, and placed adjacency gains and losses on the two alternative phylogenies according to the assumptions of Dollo's parsimony⁴² as chances are very low that distinct rearrangements would (re)create the same gene adjacency. To assess if one scenario was significantly more parsimonious than the other, we used the KHT test (Phylip package¹²⁹, dollop program). We generated 100 bootstrapped trees by resampling with replacement to evaluate the 95% confidence interval of the branch length difference between gar and bowfin, which was non-significant (95CI: [-55.05, 4.05]).

Gene family analyses. Immune, hox, and SSCP gene were annotated and analyzed as detailed in Supplementary Notes 4, 5, and 8, respectively. Generally, spotted gar and/or other vertebrate sequences were used in BLAST searches¹¹⁴ against the bowfin genome assembly to generate manually curated annotations further supported by diverse bowfin transcriptomic evidence (described below and

published^{7,34}). Orthologies were assigned based on phylogenetic and synteny evidence as described in the respective Supplementary Notes.

Developmental transcriptomics. Total RNA was extracted (Qiagen RNeasy mini plus) from RNAlater-preserved (Ambion) embryos/larvae at stages 22-23, 23-24, 24-25, 26-27, 28-29, and 30-31⁶⁸. Stage 22-23 was defined as the bowfin phylotypic stage by similarity to the phylotypic stages of zebrafish and medaka⁷⁶. Five embryos were pooled for stages 22-23 and 24-25; one embryo/larvae for stages 23-24, 26-27, 28-29, and 30-31. Libraries were generated with an Illumina TruSeq Stranded mRNA Library Preparation Kit with IDT for Illumina Unique Dual Index and sequenced on an Illumina NextSeq v2.5 (Mid Output flow cell, 2x75bp paired end format, 150 cycle v2.5 NextSeq500 reagent cartridge). After base calling with Illumina Real Time Analysis v2.4.11, sequences were demultiplexed and converted to fastq format with Illumina Bcl2fastq v2.19.1. Read pairs were trimmed for quality and adaptor sequences with Trimmomatic v0.38¹⁰⁷. Remaining paired end reads were mapped to the repeat masked bowfin genome via Bowtie2¹³⁰ ("very-sensitive", other parameters set to default). Reads mapping to the mitochondrial genome were removed with removeChrom.py [<https://github.com/harvardinformatics/ATAC-seq/blob/master/atacseq/removeChrom.py>], reads from PCR duplicates were removed with PicardTools v 2.18.1 (mark duplicates function) [<https://broadinstitute.github.io/picard/>], and Samtools view (-b -q 10) was used to remove poorly mapped reads¹³¹. Reads were visualized in IGV¹³² after normalization with BPM (bin size =25) with the deepTools v3.1.1¹³³ bamCoverage tool.

Immune tissue transcriptomics. Immune tissues (gill, spleen, intestine) were dissected from a single adult female bowfin from Bayou Chevreuil, Louisiana. RNA was extracted from each individual tissue (Qiagen RNeasy kit), quantified (ThermoFisher NanoDrop and Agilent BioAnalyzer), and pooled in equal amounts prior to library barcoding. Sequencing reads from NovaSeq 6000 were quality-trimmed using Trimmomatic v0.38¹⁰⁷ and assembled into transcripts using genome-guided Trinity v2.8.5^{134,135}.

Fin bud expression analyses. Whole animals were incubated overnight in RNAlater at 4°C, followed by dissection of pectoral fin buds using tungsten needles with 3 replicates of 6 to 7 animals collected at the desired stages. Tissues were dissolved by incubation in RLT Plus. RNA was isolated from both whole animals (for cDNA cloning) and dissected fin buds using the RNeasy Plus Micro Kit (Qiagen). For standard cloning, cDNA libraries were produced from RNA extracted from stage 23 and stage 25 single whole embryos and from stage 26 fin buds using the SuperScript IV Reverse Transcriptase System (Invitrogen). Following poly-A selection, stranded mRNAseq libraries were produced using the PrepX RNA-Seq Library Kit for Illumina (IntegenX) from stage 23, 24, 25, and 26 fin bud RNA. For each developmental stage, two replicate libraries were sequenced with 75 bp single-end reads, and a third was sequenced with 150 bp paired-end reads on an Illumina NextSeq machine. Reads were filtered using Trimmomatic v0.38¹⁰⁷, and ribosomal RNA reads were removed using SortMeRNA¹³⁶. Filtered non-rRNA reads were aligned to the bowfin nuclear transcriptome using CLC Genomics Workbench (Qiagen). The transcriptome toolkit in CLC Genomics Workbench was used to calculate RPKM values, fold-change, and analysis of

differential gene expression. Heatmaps to visualize gene expression dynamics were created using the Pretty Heatmaps package in R [<https://rdr.io/cran/pheatmap/>].

ATAC-Seq analysis. ATAC-Seq library preparation and amplification with barcoding was performed using a modification of published protocols^{67,137,138}. Samples were taken from the same clutches and stages as for the developmental transcriptome series (see above). One whole embryo per stage was dissociated into a cell solution with 0.125% collagenase (Sigma C9891) at 37°C until completely dissociated, then strained (100uM filter) and counted on an improved Neubauer chamber. 100K cells were collected and lysed per sample for DNA transposition and PCR amplification. Agencourt AMPure XP Beads were used to clean up PCR products. ATAC-Seq libraries were individually barcoded, pooled, and sequenced with Illumina NextSeq v2 (High Output flow cell, 2x75bp paired end, 150 cycle v2 NextSeq 500 reagent cartridge). Base calling, demultiplexing, FastQ format conversion, read trimming, mapping to the bowfin genome, alignments filtering, and data visualization of ATAC-Seq reads in IGV¹³² were performed as described above for the developmental transcriptome RNA-Seq data.

To identify Open Chromatin Regions (OCRs), ATAC-Seq peaks were called with MACS2 v2.2.6^{139,140} (callpeak -f BAMPE -g 767196669 -B -q 0.05 -s 75 --call-summits). Bedtools¹⁴¹ merge function was used to collapse OCRs with overlapping coordinates within stages. Merged OCRs were annotated with HOMER⁶⁹ annotatePeaks.pl (-size 'given' -annStats). Transcriptional start sites (TSS) were defined from -1kb to +100bp; transcriptional termination site (TTS) were defined from -100 bp to +1kb. BedTools merge was also used to merge OCRs across stages for a global picture of open chromatin through development. BedTools subtract -A was used to subtract exon coordinates (UTRs and coding regions) from merged OCR coordinates in each stage to generate non-coding OCRs (ncOCRs), defined as OCRs with no overlap (0bp) with any MAKER annotated exon. BedTools merge was used to merge ncOCRs across stages to get a global picture of non-coding open chromatin. Merged OCR (both all OCRs and ncOCRs only) overlap and uniqueness between developmental stages were quantified with HOMER mergePeaks⁶⁹ (-d given, -matrix matrix, -venn venn). Pairwise Jaccard distances were calculated between OCR profiles of all six stages. Read pileup and merged OCRs from MACS2 were visualized in IGV¹³² after normalization with BPM (bin size =25) with the deepTools v3.1.1¹³³ bamCoverage tool. To explore overlap of bowfin OCRs with elements from diverse vertebrate data sets (bowfin Ultra Conserved Elements, UCEs⁹; VISTA Enhancer Browser⁷⁰; mouse embryonic single nucleus OCRs⁷⁵, gar-centric Conserved Non-Coding Elements (CNEs⁷), these elements were blasted (BLASTN $e < 10^{-5}$) against the repeat-masked bowfin genome. Overlap between bowfin OCRs and blast hits for these queries was extracted with BedTools (Intersect -wa; overlaps: 1bp+ and 33%+ length overlap for either element).

Genomic sequence conservation analyses. Repeat-masked genomic regions of bowfin and other vertebrates were aligned with mVISTA¹⁴² using SLAGAN¹⁴³ as detailed in Supplementary Notes 7 and 9.

RNA *in situ* hybridization. For RNA *in situ* hybridization and skeletal staining, bowfin embryos were fixed overnight at 4°C in 4% paraformaldehyde in phosphate buffered saline, washed twice for 5 minutes in

PBS + 0.1% Tween, washed once in 100% methanol for 5 minutes, and stored in 100% methanol at -20°C. Amplified bowfin and gar cDNAs (see Supplementary Note 9 for primer sets) were cloned into pGEM-Teasy (Promega). Antisense probes were synthesized as described¹⁴⁴ and whole-mount in situ hybridization was performed using a standard embryo protocol¹⁴⁵.

Data availability

The bowfin reference genome assembly (AmiCal1) is available at GenBank under the accession number PESF00000000. Transcriptomic and ATAC-Seq reads are available at the Sequence Read Archive under accession numbers SRP281665 and SRP252716; assembled transcripts are available under accession number GIOP00000000.

Code availability

Scripts and workflows for the Pool-Seq sex determination analysis is available at <https://github.com/RomainFeron/paper-sexdetermination-bowfin>. Other software was used as provided in the methods section.

Declarations

Acknowledgments

We are thankful for grant support from NIH R01OD011116 (I.B.), NSF DDIG DEB-1600920 (M.B.H.), NSF IOS-1755242 (A.D.), NSF IOS-1755330 (J.A.Y.). Parts of this work have been supported by the Agence Nationale de la Recherche, France (ANR GenoFish project, 2016-2021, ANR-16-CE12-0035) to H.R.C., C.B., E.P., and Y.G. E.P., C.B., and H.R.C. received support under the program « Investissements d’Avenir » launched by the French Government and implemented by ANR with the references ANR-10-IDEX-0001-02 PSL* Université Paris. We thank Francis Feng for access to Oneida Lake bowfin spawning habitats and the Bauer Core at Harvard University for sequencing fin transcriptomes. We are exceedingly grateful to the late José Luis Gómez-Skarmeta for his guidance on establishing ATAC-Seq in holosteans.

Author contributions

A.W.T. and I.B. conceived the study and wrote the article with input from co-authors; I. B. coordinated genome sequencing in collaboration with Dovetail Genomics; E.F., A.R.M., Q.F., A.F., S.R.D. provided bowfin samples for genomic, chromatin, and transcriptomic analyses; A.W.T., K.L.C., and I.B. performed genome annotation; S.R.D., A.W.T., B.L.R., and I.B. sampled bowfin for the sex determination analyses performed by Q.P., R.F., M.M., and Y.G.; E.P., H.R.C., C.B., and I.B. analyzed genome evolution; D.J.W., T.O., A.D., and J.A.Y. performed immunogenome analyses; A.W.T., O.F., and I.B. performed gene regulatory analyses; M.L., I.B., M.B.H., and M.P.H. analyzed Hox genes; M.B.H., M.P.H., and I.B. investigated fin development; K. K. analyzed SCPP genes.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information for this paper is available online at XXXX.

Correspondence and requests for materials should be addressed to I.B.

References

- 1 Jarvik, J. *Basic Structure and Evolution of Vertebrates*. Vol. 1 (Academic Press, 1980).
- 2 Burr, B. M. & Bennett, M. G. in *Freshwater Fishes of North America* Vol. 1 (eds M. L. Warren & M. G. Burr) (John Hopkins University Press, 2014).
- 3 Nelson, J. S. *Fishes of the world*. 4th edn, (John Wiley, 2006).
- 4 Near, T. J. *et al.* Resolution of ray-finned fish phylogeny and timing of diversification. *Proc Natl Acad Sci U S A* **109**, 13698-13703, doi:10.1073/pnas.1206625109 (2012).
- 5 Betancur, R. R. *et al.* The tree of life and a new classification of bony fishes. *PLoS currents* **5**, doi:10.1371/currents.tol.53ba26640df0ccaee75bb165c8c26288 (2013).
- 6 Broughton, R. E., Betancur, R. R., Li, C., Arratia, G. & Orti, G. Multi-locus phylogenetic analysis reveals the pattern and tempo of bony fish evolution. *PLoS currents* **5**, doi:10.1371/currents.tol.2ca8041495ffafd0c92756e75247483e (2013).
- 7 Braasch, I. *et al.* The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. *Nat Genet* **48**, 427-437, doi:10.1038/ng.3526 (2016).
- 8 Hughes, L. C. *et al.* Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data. *Proc Natl Acad Sci U S A* **115**, 6249-6254, doi:10.1073/pnas.1719358115 (2018).
- 9 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**, doi:10.1371/journal.pone.0065923 (2013).
- 10 Clarke, J. T., Lloyd, G. T. & Friedman, M. Little evidence for enhanced phenotypic evolution in early teleosts relative to their living fossil sister group. *Proc Natl Acad Sci U S A* **113**, 11531-11536, doi:10.1073/pnas.1607237113 (2016).
- 11 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*, doi:10.1002/jez.b.22589 (2014).

- 12 Patterson, C. in *Interrelationships of Fishes* Vol. Supplement 1 (eds P. H. Greenwood, R. S. Miles, & C. Patterson) 233-305 (Academic Press, 1973).
- 13 Grande, L. An Empirical Synthetic Pattern Study of Gars (Lepisosteiformes) and Closely Related Species, Based Mostly on Skeletal Anatomy. The Resurrection of Holostei. *Copeia*, 1-863 (2010).
- 14 Sallan, L. C. Major issues in the origins of ray-finned fish (Actinopterygii) biodiversity. *Biological reviews of the Cambridge Philosophical Society*, doi:10.1111/brv.12086 (2014).
- 15 Grande, L. & Bemis, W. E. A Comprehensive Phylogenetic Study of Amiid Fishes (Amiidae) Based on Comparative Skeletal Anatomy. an Empirical Search for Interconnected Patterns of Natural History. *Journal of Vertebrate Paleontology* **18**, 1-696, doi:10.1080/02724634.1998.10011114 (1998).
- 16 Majtanova, Z., Symonova, R., Arias-Rodriguez, L., Sallan, L. & Rab, P. "Holostei versus Halecostomi" Problem: Insight from Cytogenetics of Ancient Nonteleost Actinopterygian Fish, Bowfin *Amia calva*. *J Exp Zool B Mol Dev Evol* **328**, 620-628, doi:10.1002/jez.b.22720 (2017).
- 17 Amores, A. *et al.* Zebrafish hox clusters and vertebrate genome evolution. *Science* **282**, 1711-1714 (1998).
- 18 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, doi:10.1101/gr.640303 (2003).
- 19 Braasch, I. & Postlethwait, J. H. in *Polyploidy and Genome Evolution* (eds P. S. Soltis & D. E. Soltis) Ch. 17, 341-383 (Springer, 2012).
- 20 Jaillon, O. *et al.* Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* **431**, 946-957, doi:10.1038/nature03025 (2004).
- 21 Takezaki, N. Global Rate Variation in Bony Vertebrates. *Genome Biol Evol* **10**, 1803-1815, doi:10.1093/gbe/evy125 (2018).
- 22 Clement, Y., Torbey, P., Gilardi-Hebenstreit, P. & Crollius, H. R. Enhancer-gene maps in the human and zebrafish genomes using evolutionary linkage conservation. *Nucleic Acids Res* **48**, 2357-2371, doi:10.1093/nar/gkz1199 (2020).
- 23 Yuan, X. *et al.* Heart enhancers with deeply conserved regulatory activity are established early in zebrafish development. *Nat Commun* **9**, 4977, doi:10.1038/s41467-018-07451-z (2018).
- 24 Peskin, B. *et al.* Notochordal Signals Establish Phylogenetic Identity of the Teleost Spine. *Curr Biol* **30**, 2805-2814 e2803, doi:10.1016/j.cub.2020.05.037 (2020).
- 25 Sire, J. Y., Donoghue, P. C. & Vickaryous, M. K. Origin and evolution of the integumentary skeleton in non-tetrapod vertebrates. *J Anat* **214**, 409-440, doi:10.1111/j.1469-7580.2009.01046.x (2009).

- 26 Funk, E., Lencer, E. & McCune, A. Dorsoventral inversion of the air-filled organ (lungs, gas bladder) in vertebrates: RNAsequencing of laser capture microdissected embryonic tissue. *J Exp Zool B Mol Dev Evol*, doi:10.1002/jez.b.22998 (2020).
- 27 Funk, E. C., Breen, C., Sanketi, B. D., Kurpios, N. & McCune, A. Changes in Nkx2.1, Sox2, Bmp4 and Bmp16 expression underlying the lung-to-gas bladder evolutionary transition in ray-finned fishes. *Evolution and Development in press* (2020).
- 28 Chapman, J. A. *et al.* Meraculous: de novo genome assembly with short paired-end reads. *PLoS One* **6**, e23501, doi:10.1371/journal.pone.0023501 (2011).
- 29 Putnam, N. H. *et al.* Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Res* **26**, 342-350, doi:10.1101/gr.193474.115 (2016).
- 30 Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289-293, doi:10.1126/science.1181369 (2009).
- 31 Ohno, S. *et al.* Microchromosomes in holocephalian, chondrosteian and holostean fishes. *Chromosoma* **26**, 35-40, doi:10.1007/BF00319498 (1969).
- 32 Parra, G., Bradnam, K. & Korf, I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23**, 1061-1067, doi:10.1093/bioinformatics/btm071 (2007).
- 33 Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210-3212, doi:10.1093/bioinformatics/btv351 (2015).
- 34 Pasquier, J. *et al.* Gene evolution and gene expression after whole genome duplication in fish: the PhyloFish database. *BMC Genomics* **17**, 368, doi:10.1186/s12864-016-2709-z (2016).
- 35 Holt, C. & Yandell, M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**, 491, doi:10.1186/1471-2105-12-491 (2011).
- 36 Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* **20**, 238, doi:10.1186/s13059-019-1832-y (2019).
- 37 Dean, R. & Mank, J. E. The role of sex chromosomes in sexual dimorphism: discordance between molecular and phenotypic data. *J Evol Biol* **27**, 1443-1453, doi:10.1111/jeb.12345 (2014).
- 38 Feron, R. *et al.* RADSex: a computational workflow to study sex determination using Restriction Site-Associated DNA Sequencing data. *bioRxiv*, 2020.2004.2022.054866, doi:10.1101/2020.04.22.054866 (2020).

- 39 Du, K. *et al.* The sterlet sturgeon genome sequence and the mechanisms of segmental rediploidization. *Nat Ecol Evol* **4**, 841-852, doi:10.1038/s41559-020-1166-x (2020).
- 40 Sacerdot, C., Louis, A., Bon, C., Berthelot, C. & Roest Crolius, H. Chromosome evolution at the origin of the ancestral vertebrate genome. *Genome Biol* **19**, 166, doi:10.1186/s13059-018-1559-1 (2018).
- 41 Moret, B. M. E., Tang, J., Wang, L.-S. & Warnow, T. Steps toward accurate reconstructions of phylogenies from gene-order data. *Journal of Computer and System Sciences* **65**, 508-525, doi:[https://doi.org/10.1016/S0022-0000\(02\)00007-7](https://doi.org/10.1016/S0022-0000(02)00007-7) (2002).
- 42 Farris, J. S. Phylogenetic Analysis Under Dollo's Law. *Systematic Biology* **26**, 77-88, doi:10.1093/sysbio/26.1.77 (1977).
- 43 Lin, Y., Hu, F., Tang, J. & Moret, B. M. Maximum likelihood phylogenetic reconstruction from high-resolution whole-genome data and a tree of 68 eukaryotes. *Pac Symp Biocomput*, 285-296 (2013).
- 44 Wcisel, D. J., Ota, T., Litman, G. W. & Yoder, J. A. Spotted Gar and the Evolution of Innate Immune Receptors. *J Exp Zool B Mol Dev Evol* **328**, 666-684, doi:10.1002/jez.b.22738 (2017).
- 45 Trowsdale, J. The MHC, disease and selection. *Immunol Lett* **137**, 1-8, doi:10.1016/j.imlet.2011.01.002 (2011).
- 46 MHCSequencingConsortium. Complete sequence and gene map of a human major histocompatibility complex. The MHC sequencing consortium. *Nature* **401**, 921-923, doi:10.1038/44853 (1999).
- 47 Ohta, Y. *et al.* Primitive synteny of vertebrate major histocompatibility complex class I and class II genes. *Proc Natl Acad Sci U S A* **97**, 4712-4717, doi:10.1073/pnas.97.9.4712 (2000).
- 48 Flajnik, M. F., Ohta, Y., Namikawa-Yamada, C. & Nonaka, M. Insight into the primordial MHC from studies in ectothermic vertebrates. *Immunol Rev* **167**, 59-67, doi:10.1111/j.1600-065x.1999.tb01382.x (1999).
- 49 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, doi:10.1186/1471-2148-13-260 (2013).
- 50 Grimholt, U. *et al.* A comprehensive analysis of teleost MHC class I sequences. *BMC Evol Biol* **15**, 32, doi:10.1186/s12862-015-0309-1 (2015).
- 51 Flajnik, M. F. & Kasahara, M. Origin and evolution of the adaptive immune system: genetic events and selective pressures. *Nat Rev Genet* **11**, 47-59, doi:10.1038/nrg2703 (2010).

- 52 Flajnik, M. F. A cold-blooded view of adaptive immunity. *Nat Rev Immunol* **18**, 438-453, doi:10.1038/s41577-018-0003-9 (2018).
- 53 Salinas, I., Zhang, Y. A. & Sunyer, J. O. Mucosal immunoglobulins and B cells of teleost fish. *Dev Comp Immunol* **35**, 1346-1365, doi:10.1016/j.dci.2011.11.009 (2011).
- 54 Fillatreau, S. *et al.* The astonishing diversity of Ig classes and B cell repertoires in teleost fish. *Front Immunol* **4**, 28, doi:10.3389/fimmu.2013.00028 (2013).
- 55 Danilova, N., Bussmann, 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, doi:10.1038/ni1166 (2005).
- 56 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 U S A* **102**, 6919-6924, doi:10.1073/pnas.0500027102 (2005).
- 57 Zhang, Y. A. *et al.* IgT, a primitive immunoglobulin class specialized in mucosal immunity. *Nat Immunol* **11**, 827-835, doi:10.1038/ni.1913 (2010).
- 58 Salinas, I. The Mucosal Immune System of Teleost Fish. *Biology (Basel)* **4**, 525-539, doi:10.3390/biology4030525 (2015).
- 59 Piazzon, M. C. *et al.* Differential Modulation of IgT and IgM upon Parasitic, Bacterial, Viral, and Dietary Challenges in a Perciform Fish. *Front Immunol* **7**, 637, doi:10.3389/fimmu.2016.00637 (2016).
- 60 Aderem, A. & Ulevitch, R. J. Toll-like receptors in the induction of the innate immune response. *Nature* **406**, 782-787, doi:10.1038/35021228 (2000).
- 61 Fitzgerald, K. A. & Kagan, J. C. Toll-like Receptors and the Control of Immunity. *Cell* **180**, 1044-1066, doi:10.1016/j.cell.2020.02.041 (2020).
- 62 Aoki, T., Hikima, J., Hwang, S. D. & Jung, T. S. Innate immunity of finfish: primordial conservation and function of viral RNA sensors in teleosts. *Fish Shellfish Immunol* **35**, 1689-1702, doi:10.1016/j.fsi.2013.02.005 (2013).
- 63 Quiniou, S. M., Boudinot, P. & Bengten, E. Comprehensive survey and genomic characterization of Toll-like receptors (TLRs) in channel catfish, *Ictalurus punctatus*: identification of novel fish TLRs. *Immunogenetics* **65**, 511-530, doi:10.1007/s00251-013-0694-9 (2013).
- 64 Kawasaki, K. *et al.* SSCP Genes and Their Relatives in Gar: Rapid Expansion of Mineralization Genes in Osteichthyans. *J Exp Zool B Mol Dev Evol* **328**, 645-665, doi:10.1002/jez.b.22755 (2017).

- 65 Kawasaki, K. & Weiss, K. M. Mineralized tissue and vertebrate evolution: the secretory calcium-binding phosphoprotein gene cluster. *Proc Natl Acad Sci U S A* **100**, 4060-4065, doi:10.1073/pnas.0638023100 (2003).
- 66 Qu, Q., Haitina, T., Zhu, M. & Ahlberg, P. E. New genomic and fossil data illuminate the origin of enamel. *Nature* **526**, 108-111, doi:10.1038/nature15259 (2015).
- 67 Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**, 1213-1218, doi:10.1038/nmeth.2688 (2013).
- 68 Ballard, W. W. Stages and rates of normal development in the holostean fish, *Amia calva*. *Journal of Experimental Zoology* **238**, 337-354, doi:10.1002/jez.1402380308 (1986).
- 69 Heinz, S. *et al.* Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell* **38**, 576-589, doi:10.1016/j.molcel.2010.05.004 (2010).
- 70 Visel, A., Minovitsky, S., Dubchak, I. & Pennacchio, L. A. VISTA Enhancer Browser - A database of tissue-specific human enhancers. *Nucleic Acids Research* **35**, doi:10.1093/nar/gkl822 (2007).
- 71 Abbasi, A. A. *et al.* Human intronic enhancers control distinct sub-domains of Gli3 expression during mouse CNS and limb development. *BMC developmental biology* **10**, 44, doi:10.1186/1471-213X-10-44 (2010).
- 72 Adachi, N., Robinson, M., Goolsbee, A. & Shubin, N. H. Regulatory evolution of Tbx5 and the origin of paired appendages. *Proc Natl Acad Sci U S A* **113**, 10115-10120, doi:10.1073/pnas.1609997113 (2016).
- 73 Menke, D. B., Guenther, C. & Kingsley, D. M. Dual hindlimb control elements in the Tbx4 gene and region-specific control of bone size in vertebrate limbs. *Development* **135**, 2543-2553, doi:10.1242/dev.017384 (2008).
- 74 Zhang, W. *et al.* Spatial-temporal targeting of lung-specific mesenchyme by a Tbx4 enhancer. *BMC Biol* **11**, 111, doi:10.1186/1741-7007-11-111 (2013).
- 75 Pijuan-Sala, B. *et al.* Single-cell chromatin accessibility maps reveal regulatory programs driving early mouse organogenesis. *Nature Cell Biology* **22**, 487-497, doi:10.1038/s41556-020-0489-9 (2020).
- 76 Tena, J. J. *et al.* Comparative epigenomics in distantly related teleost species identifies conserved cis-regulatory nodes active during the vertebrate phylotypic period. *Genome Res* **24**, 1075-1085, doi:10.1101/gr.163915.113 (2014).

- 77 Li, Y. *et al.* Dynamic transcriptional and chromatin accessibility landscape of medaka embryogenesis. *Genome Res* **30**, 924-937, doi:10.1101/gr.258871.119 (2020).
- 78 Graham, J. B. *Air-Breathing Fishes*. (Academic Press, 1997).
- 79 Kuraku, S. *et al.* Noncanonical role of Hox14 revealed by its expression patterns in lamprey and shark. *Proc Natl Acad Sci U S A* **105**, 6679-6683, doi:10.1073/pnas.0710947105 (2008).
- 80 Powers, T. P. & Amemiya, C. T. Evidence for a Hox14 paralog group in vertebrates. *Curr Biol* **14**, R183-184, doi:10.1016/j.cub.2004.02.015 (2004).
- 81 Tulenko, F. J. *et al.* HoxD expression in the fin-fold compartment of basal gnathostomes and implications for paired appendage evolution. *Sci Rep* **6**, 22720, doi:10.1038/srep22720 (2016).
- 82 Feiner, N., Ericsson, R., Meyer, A. & Kuraku, S. Revisiting the origin of the vertebrate Hox14 by including its relict sarcopterygian members. *J Exp Zool B Mol Dev Evol* **316**, 515-525, doi:10.1002/jez.b.21426 (2011).
- 83 Wood, A. Early pectoral fin development and morphogenesis of the apical ectodermal ridge in the killifish, *Aphyosemion scheeli*. *The Anatomical Record* **204**, 349-356, doi:10.1002/ar.1092040408 (1982).
- 84 Fisher, S., Jagadeeswaran, P. & Halpern, M. E. Radiographic analysis of zebrafish skeletal defects. *Dev Biol* **264**, 64-76, doi:10.1016/s0012-1606(03)00399-3 (2003).
- 85 Huang, C. C. *et al.* Collagen IX is required for the integrity of collagen II fibrils and the regulation of vascular plexus formation in zebrafish caudal fins. *Dev Biol* **332**, 360-370, doi:10.1016/j.ydbio.2009.06.003 (2009).
- 86 Zhang, J. *et al.* Loss of fish actinotrichia proteins and the fin-to-limb transition. *Nature* **466**, 234-237, doi:10.1038/nature09137 (2010).
- 87 Duran, I. *et al.* Collagen duplicate genes of bone and cartilage participate during regeneration of zebrafish fin skeleton. *Gene Expr Patterns* **19**, 60-69, doi:10.1016/j.gep.2015.07.004 (2015).
- 88 Duran, I., Mari-Beffa, M., Santamaria, J. A., Becerra, J. & Santos-Ruiz, L. Actinotrichia collagens and their role in fin formation. *Dev Biol* **354**, 160-172, doi:10.1016/j.ydbio.2011.03.014 (2011).
- 89 Wade, C., Brinas, I., Welfare, M., Wicking, C. & Farlie, P. G. Twist2 contributes to termination of limb bud outgrowth and patterning through direct regulation of Grem1. *Dev Biol* **370**, 145-153, doi:10.1016/j.ydbio.2012.07.025 (2012).
- 90 Yashiro, K. *et al.* Regulation of Retinoic Acid Distribution Is Required for Proximodistal Patterning and Outgrowth of the Developing Mouse Limb. *Developmental Cell* **6**, 411-422, doi:10.1016/S1534-5807(04)00062-0 (2004).

- 91 Kawakami, Y. *et al.* Sp8 and Sp9, two closely related buttonhead-like transcription factors, regulate Fgf8 expression and limb outgrowth in vertebrate embryos. *Development* **131**, 4763-4774, doi:10.1242/dev.01331 (2004).
- 92 Gillis, J. A., Dahn, R. D. & Shubin, N. H. Shared developmental mechanisms pattern the vertebrate gill arch and paired fin skeletons. *Proc Natl Acad Sci U S A* **106**, 5720-5724, doi:10.1073/pnas.0810959106 (2009).
- 93 Tulenko, F. J. *et al.* Fin-fold development in paddlefish and catshark and implications for the evolution of the autopod. *Proc Biol Sci* **284**, doi:10.1098/rspb.2016.2780 (2017).
- 94 Jovelin, R. *et al.* Duplication and divergence of fgf8 functions in teleost development and evolution. *J Exp Zool B Mol Dev Evol* **308**, 730-743, doi:10.1002/jez.b.21193 (2007).
- 95 Hodgkinson, V. S., Ericsson, R., Johanson, Z. & Joss, J. M. P. The apical ectodermal ridge in the pectoral fin of the Australian Lungfish (*Neoceratodus forsteri*): keeping the fin to limb transition in the fold. *Acta Zoologica* **90**, 253-263, doi:10.1111/j.1463-6395.2008.00349.x (2009).
- 96 Gehrke, A. R. & Shubin, N. H. Cis-regulatory programs in the development and evolution of vertebrate paired appendages. *Semin Cell Dev Biol* **57**, 31-39, doi:10.1016/j.semcdb.2016.01.015 (2016).
- 97 Gross, J. B., Kerney, R., Hanken, J. & Tabin, C. J. Molecular anatomy of the developing limb in the coqui frog, *Eleutherodactylus coqui*. *Evol Dev* **13**, 415-426, doi:10.1111/j.1525-142X.2011.00500.x (2011).
- 98 Doroba, C. K. & Sears, K. E. The divergent development of the apical ectodermal ridge in the marsupial *Monodelphis domestica*. *Anat Rec (Hoboken)* **293**, 1325-1332, doi:10.1002/ar.21183 (2010).
- 99 Purushothaman, S., Elewa, A. & Seifert, A. W. Fgf-signaling is compartmentalized within the mesenchyme and controls proliferation during salamander limb development. *Elife* **8**, doi:10.7554/eLife.48507 (2019).
- 100 Marinic, M., Aktas, T., Ruf, S. & Spitz, F. An integrated holo-enhancer unit defines tissue and gene specificity of the Fgf8 regulatory landscape. *Dev Cell* **24**, 530-542, doi:10.1016/j.devcel.2013.01.025 (2013).
- 101 Komisarczuk, A. Z., Kawakami, K. & Becker, T. S. Cis-regulation and chromosomal rearrangement of the fgf8 locus after the teleost/tetrapod split. *Dev Biol* **336**, 301-312, doi:10.1016/j.ydbio.2009.09.029 (2009).
- 102 Schultze, H.-P. & Wiley, E. O. in *Living Fossils* (eds N. Eldredge & S.M. Stanley) 153-159 (Springer, 1984).
- 103 Scarnecchia, D. L. A Reappraisal of Gars and Bowfins in Fishery Management. *Fisheries* **17**, 6-12, doi:10.1577/1548-8446(1992)017<0006:Arogab>2.0.Co;2 (1992).

- 104 Smith, N. G., Daugherty, D. J., Schlechte, J. W. & Buckmeier, D. L. Modeling the Responses of Alligator Gar Populations to Harvest under Various Length-Based Regulations: Implications for Conservation and Management. *Transactions of the American Fisheries Society* **147**, 665-673, doi:10.1002/tafs.10040 (2018).
- 105 Sinopoli, D. *Morphological Variation of Bowfin (Amiidae: Amia calva Linnaeus 1766) Populations from the Mississippi River Basin: Taxonomic and Conservation Implications*. Master of Science thesis, State University of New York, (2019).
- 106 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, doi:10.1016/j.cbpc.2014.01.005 (2014).
- 107 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120, doi:10.1093/bioinformatics/btu170 (2014).
- 108 Smit, A. F. A. & Hubley, R. *RepeatModeler Open-1.0*. <http://www.repeatmasker.org> (2008).
- 109 Schartl, M. *et al.* The genome of the platyfish, *Xiphophorus maculatus*, provides insights into evolutionary adaptation and several complex traits. *Nat Genet* **45**, 567-572, doi:10.1038/ng.2604 (2013).
- 110 Amemiya, C. T. *et al.* The African coelacanth genome provides insights into tetrapod evolution. *Nature* **496**, 311-316, doi:10.1038/nature12027 (2013).
- 111 Jurka, J. Repbase update: a database and an electronic journal of repetitive elements. *Trends Genet* **16**, 418-420, doi:10.1016/s0168-9525(00)02093-x (2000).
- 112 Smit, A. F. A., Hubley, R. & Green, P. *RepeatMasker Open-4.0*. <http://www.repeatmasker.org> (2013).
- 113 Smith, C. D. *et al.* Improved repeat identification and masking in Dipterans. *Gene* **389**, 1-9, doi:10.1016/j.gene.2006.09.011 (2007).
- 114 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *Journal of molecular biology* **215**, 403-410, doi:10.1016/S0022-2836(05)80360-2 (1990).
- 115 Slater, G. S. & Birney, E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**, 31, doi:10.1186/1471-2105-6-31 (2005).
- 116 Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59, doi:10.1186/1471-2105-5-59 (2004).
- 117 Stanke, M. & Waack, S. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19 Suppl 2**, ii215-225, doi:10.1093/bioinformatics/btg1080 (2003).

- 118 Bowman, M. J., Pulman, J. A., Liu, T. L. & Childs, K. L. A modified GC-specific MAKER gene annotation method reveals improved and novel gene predictions of high and low GC content in *Oryza sativa*. *BMC Bioinformatics* **18**, 522, doi:10.1186/s12859-017-1942-z (2017).
- 119 Bateman, A. *et al.* The Pfam Protein Families Database. <http://www.sanger.ac.uk/Software/Pfam/> (2000).
- 120 Eddy, S. R. Multiple alignment using hidden Markov models. *Proc Int Conf Intell Syst Mol Biol* **3**, 114-120 (1995).
- 121 Campbell, M. S., Holt, C., Moore, B. & Yandell, M. Genome Annotation and Curation Using MAKER and MAKER-P. *Curr Protoc Bioinformatics* **48**, 4 11 11-39, doi:10.1002/0471250953.bi0411s48 (2014).
- 122 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).
- 123 Marcais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764-770, doi:10.1093/bioinformatics/btr011 (2011).
- 124 Vilella, A. J. *et al.* EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res* **19**, 327-335, doi:10.1101/gr.073585.107 (2009).
- 125 Ruan, J. *et al.* TreeFam: 2008 Update. *Nucleic Acids Res* **36**, D735-740, doi:10.1093/nar/gkm1005 (2008).
- 126 Wallace, I. M., O'Sullivan, O., Higgins, D. G. & Notredame, C. M-Coffee: combining multiple sequence alignment methods with T-Coffee. *Nucleic Acids Res* **34**, 1692-1699, doi:10.1093/nar/gkl091 (2006).
- 127 Sankoff, D., Deneault, M., Bryant, D., Lemieux, C. & Turmel, M. in *Comparative Genomics. Computational Biology* Vol. 1 (eds D. Sankoff & J.H. Nadeau) (Springer, 2000).
- 128 Sawa, G., Dicks, J. & Roberts, I. N. Current approaches to whole genome phylogenetic analysis. *Brief Bioinform* **4**, 63-74, doi:10.1093/bib/4.1.63 (2003).
- 129 Felsenstein, J. PHYLIP-Phylogeny Inference Package (Ver. 3.2). *Cladistics* **5**, 164-166 (1989).
- 130 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).
- 131 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).
- 132 Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**, 178-192,

doi:10.1093/bib/bbs017 (2013).

- 133 Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A. & Manke, T. DeepTools: A flexible platform for exploring deep-sequencing data. *Nucleic Acids Research* **42**, W187-W191, doi:10.1093/nar/gku365 (2014).
- 134 Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* **29**, 644-652, doi:10.1038/nbt.1883 (2011).
- 135 Henschel, R. *et al.* in *Proceedings of the 1st Conference of the Extreme Science and Engineering Discovery Environment: Bridging from the eXtreme to the campus and beyond* Article 45 (Association for Computing Machinery, Chicago, Illinois, USA, 2012).
- 136 Kopylova, E., Noé, L. & Touzet, H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**, 3211-3217, doi:10.1093/bioinformatics/bts611 (2012).
- 137 Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol* **109**, 21 29 21-21 29 29, doi:10.1002/0471142727.mb2129s109 (2015).
- 138 Fernandez-Minan, A., Bessa, J., Tena, J. J. & Gomez-Skarmeta, J. L. Assay for transposase-accessible chromatin and circularized chromosome conformation capture, two methods to explore the regulatory landscapes of genes in zebrafish. *Methods Cell Biol* **135**, 413-430, doi:10.1016/bs.mcb.2016.02.008 (2016).
- 139 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biology* **9**, 1-9, doi:10.1186/gb-2008-9-9-r137 (2008).
- 140 Gaspar, J. M. Improved peak-calling with MACS2. *bioRxiv* **496521**, doi:10.1101/496521 (2018).
- 141 Quinlan, A. R. & Hall, I. M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).
- 142 Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M. & Dubchak, I. VISTA: computational tools for comparative genomics. *Nucleic Acids Res* **32**, W273-279, doi:10.1093/nar/gkh458 (2004).
- 143 Brudno, M. *et al.* Glocal alignment: finding rearrangements during alignment. *Bioinformatics* **19 Suppl 1**, i54-62, doi:10.1093/bioinformatics/btg1005 (2003).
- 144 Lufkin, T. In situ hybridization of whole-mount mouse embryos with RNA probes: preparation of embryos and probes. *CSH Protoc* **2007**, pdb prot4822, doi:10.1101/pdb.prot4822 (2007).
- 145 Lufkin, T. In situ hybridization of whole-mount mouse embryos with RNA probes: hybridization, washes, and histochemistry. *CSH Protoc* **2007**, pdb prot4823, doi:10.1101/pdb.prot4823 (2007).

- 146 Negrisol, E. *et al.* Different phylogenomic approaches to resolve the evolutionary relationships among model fish species. *Mol Biol Evol* **27**, 2757-2774, doi:10.1093/molbev/msq165 (2010).
- 147 McConnell, S. C. *et al.* Alternative haplotypes of antigen processing genes in zebrafish diverged early in vertebrate evolution. *Proc Natl Acad Sci U S A* **113**, E5014-5023, doi:10.1073/pnas.1607602113 (2016).
- 148 Traver, D. & Yoder, J. A. in *The Zebrafish in Biomedical Research* (eds S. Cartner *et al.*) 191–216 (Academic Press, 2020).
- 149 Nikaido, M. *et al.* Coelacanth genomes reveal signatures for evolutionary transition from water to land. *Genome Res* **23**, 1740-1748, doi:10.1101/gr.158105.113 (2013).
- 150 Long, W. L. & Ballard, W. W. Normal embryonic stages of the longnose gar, *Lepisosteus osseus*. *BMC developmental biology* **1**, 6 (2001).

Figures

Figure 1

Bowfin and the evolution of neopterygian genome organization. A) Adult male bowfin with green breeding color and caudal ocellus. B) Bowfin ontogenetic stages (st.) 23-2668 covering critical phases of pectoral fin and gas bladder development. Scale bar: 1mm. C) Orthologies between bowfin and spotted gar chromosomes and inferred micro-chromosome fusions for bowfin. D) Orthologies between the medaka and gar chromosomes and inferred micro-chromosome fusions for medaka. Circles represent the number of orthologous genes shared by bowfin and gar/medaka chromosomes (if in excess compared to random expectations). Ancestral micro-chromosomes are highlighted with colors, with micro-chromosome fusions highlighted by a dotted box. Numbers in colored boxes refer to the gar chromosome number. Micro-chromosome fusions differ between C and D: For instance, medaka chromosomes 9 and 12 are the result of a pre-TGD fusion of ancestral chromosomes orthologous to gar micro-chromosomes LG20 and LG21 and macro-chromosome LG2, followed by duplication of the fusion chromosome in the TGD (D).

Bowfin 15 in contrast is a fusion of ancestral chromosomes orthologous to gar micro-chromosomes LG13 and LG20; bowfin 7 in turn is a fusion of ancestral chromosomes orthologous to gar LG1 and micro-chromosomes LG21 and LG23 (C). E) Neighbor-joining phylogeny based on gene order divergence built using a normalized breakpoint distance and with bootstrap support shown in dark red. The Holostei clade is strongly supported (red box). Note that the location of the stickleback branch (gray) is in disagreement with the consensus phylogeny^{4,8} (but see¹⁴⁶). F) Dollo parsimony applied to gains and losses of local gene adjacencies in the Holostei (top) and Halecostomi (bottom) scenarios. Adjacencies shared by bowfin and gar only are shown in pink, those shared by bowfin and teleosts only in yellow, and by gar and teleosts only in blue.

Figure 2

Bowfin Major Histocompatibility Complex. (A) The bowfin MHC on superscaffold 14 contains class I, class II, and class III with orthologous relationships to the human MHC on chromosome 6 and zebrafish chromosomes 19, 16 and 15. Boxes represent genes or gene clusters of related sequences; gene placement is not to scale. Note that the zebrafish classical MHC class I “U” lineage genes on Dre19 display gene content variation147,148, but only one haplotype is presented here. Only those genes with

bowfin orthologs are shown in human and zebrafish. Pseudogenes of non-MHC genes, and RNA coding genes are not included. (B) Detailed lists of genes within each human HLA gene cluster indicated in panel A are provided. Asterisks indicate MHC or MHC-related pseudogenes.

Figure 3

Organization of biomineralization genes in neopterygian fishes. Genomic arrangements of SSCP genes and SPARC-family genes in the genomes of spotted gar, bowfin, and the teleosts zebrafish, Mexican tetra (cavefish), and channel catfish. The direction of triangles and pentagons indicates the transcriptional direction of each gene, encoding P/Q-rich SCPs (red), acidic SCPs (blue), and SPARC-family proteins (yellow). Orthologs are connected by dashed lines, unresolved orthologies indicated with a “?”. Intervening non-SCP genes are not shown. In the zebrafish genome, a cluster of sparcl1 and scpp1 are separated ~10 Mb from the other genes encoding P/Q-rich SCPs (arrow). Two cavefish SSCP gene clusters are located on unplaced scaffolds

Figure 4

Chromatin and gene expression profiles during bowfin development. Non-coding open chromatin regions (ncOCRs) overlapping with key developmental enhancers are indicated by black boxes. ATAC-Seq profiles are shown in blue, MACS2-called OCRs in green, and mRNA-Seq transcriptional profile in red. Nearest annotated genes are shown. Scale bars: 1Kb. A) Forebrain enhancer e12270 downstream of *arx* is conserved in bowfin (Supplementary Tab. 19) and open throughout bowfin developmental stages (st.). B) A pan-cardiac heart enhancer ZaCNE3, conserved from tetrapods to fish23 and upstream of *bmp4*, transitions from open to closed chromatin as development progresses. C) Intragenic *gli3* limb enhancer CNE1171, connected by sequence orthology from tetrapods to teleosts through *gar7* and bowfin (not shown), opens at st. 23 when pectoral fins are first visible68 and remains open onward through the development of pectoral fins (Supplemental Fig. 16). D) Pectoral fin enhancer CNS1272 is open during st. 22-27 when the pectoral fins develop (Supplemental Fig. 16). E) The *tbx4* hindlimb enhancer HLEB73 is MACS2-called as ncOCR at st. 28-30, shortly after the onset of pelvic fin development in st. 2768. Another ncOCR is located on the *tbx4* 'lung enhancer' critical for lung development in mouse73,74. Its chromatin is open throughout the bowfin gas bladder budding and outgrowth stages (st. 24-27) when *tbx4* is expressed in the bowfin gas bladder26. A putative swim bladder enhancer (SBE)149, conserved with teleosts (Supplementary Fig. 14), is marked by open chromatin as well. F) A VISTA conservation plot

(SLAGAN) with bowfin as reference species shows sequence conservation of the *tbx4* lung enhancer from tetrapods to lobe-finned fish (coelacanth) and teleosts (stickleback, pufferfish, pike), demonstrating the power of the ‘holostean bridge’ to reveal deep homology of gene regulation of key vertebrate morphologies. See Supplementary Fig. 14 and Supplementary Note 7 for detailed *tbx4* analyses.

Figure 5

Holostean Hox gene clusters and the bowfin *hoxd14* pseudogene. A) Hox cluster organization in bowfin and spotted gar genomes. B) Relative position and orientation of Hox14 paralogs in the Hox D cluster between *evx2* and *hoxd13* in zebrafish, bowfin, and horn shark. We identified two alternative splice

variants in bowfin. C) Whole mount RNA in situ hybridization of the *hoxd14* pseudogene in the tail bud (tb) and vent at stage 24 and in the posterior pectoral fin mesenchyme at stage 26 (white arrowhead). Scale bars are 125 μ m. D) Multiple sequence alignment of the HoxD14 homeodomain across species including bowfin (*Aca*), spotted gar (*Loc*), paddlefish (*Psp*, *Polyodon spathula*), elephant shark (*Cmi*, *Callorhinchus milii*), horn shark (*Hfr*, *Heterodontus francisci*), and small-spotted catshark (*Sca*, *Scyliorhinus canicula*). Black arrowhead indicates conserved intra-domain splicing position, and underline indicates conserved WFQNQR motif.

Figure 6

Gene expression analyses of developing pectoral fins in bowfin. A) Bowfin developmental stages (st.)68 used for RNA-Seq transcriptome analysis. The fin bud is first visible at stage 23 (black arrowhead, anterior to left). Isolated fin buds showing increasing size from st. 24 to 25 to 26 (anterior to top, distal to right). B) Actinotrichia fibrils first appear at st. 27. C) Transcriptional log2 fold-change in fin fold structural genes relative to st. 23. Between st. 23 and 26, *and1/2* and *and2/3* show 862- and 285-fold increases, respectively; *col3a1*, *col9a3-like a*, and *col9a3-like b* exhibit 133-, 417-, and 19-fold increases. Stable expression across stages is seen for *col1a1*, *col1a2*, and *col2a1*, each showing a fold-change range between 0.5 and 1.6. D) Transcriptional log2 fold-change in fin/limb patterning genes relative to st. 23 fin

buds. Indicative of their temporal collinearity, *hoxa9/10* and *hoxd9/10* peak at stage 23, while the more posterior *hoxa11/13*, and *hoxd11/12/13* peak at stage 25. Expression of *hoxa13* is highest at stage 26. Highest expression level of *hoxd14* is at stages 25-26. E) The transcription factor gene *sp8* is expressed in the apical ectodermal ridge (AER) of *gar* (st. 26/50) and *bowfin* (st. 23), while *fgf8* expression is not detected in the AER of both holosteans.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [BowfinSupplementsNG201013.pdf](#)
- [SupplFile1BowfinScppGenesv3.pdf](#)
- [SupplFile2BowfinHoxv3.txt](#)
- [SupplTab6BowfinMHCgenes.pdf](#)
- [SupplTab7GarMHCgenes.xlsx](#)
- [SupplTab812ImmuneGeneAccessions.pdf](#)
- [SupplTab19VISTAenhancersbowfinOCRs.xls](#)