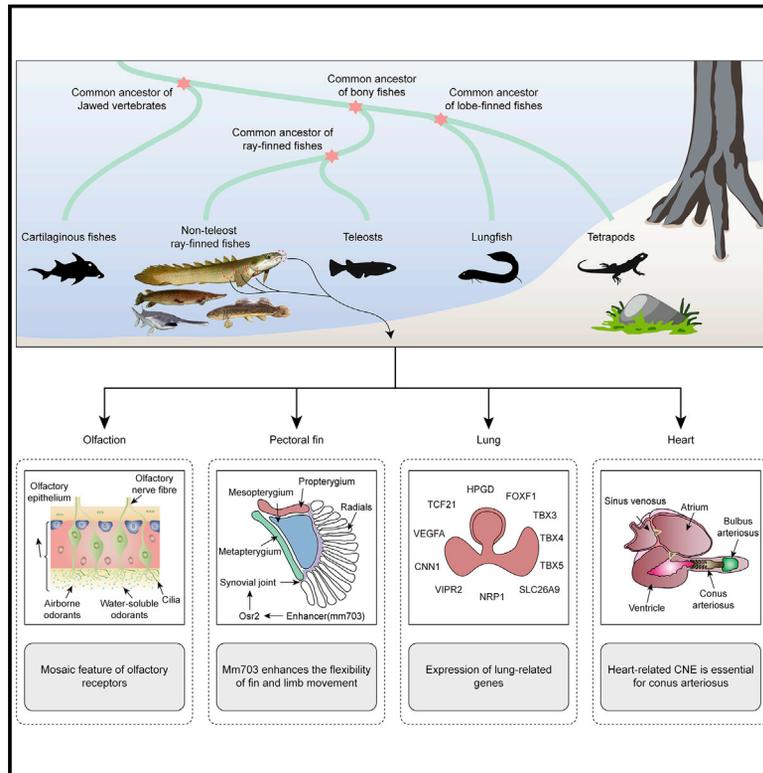


# Tracing the genetic footprints of vertebrate landing in non-teleost ray-finned fishes

## Graphical abstract



## Authors

Xupeng Bi, Kun Wang, Liandong Yang, ...,  
Min Zhu, Shunping He, Guojie Zhang

## Correspondence

clad@ihb.ac.cn (S.H.),  
guojie.zhang@bio.ku.dk (G.Z.),  
wwang@mail.kiz.ac.cn (W.W.),  
zhumin@ivpp.ac.cn (M.Z.)

## In Brief

Comparative analyses of divergent lineages of ray-finned fishes reveal that that these species exhibit mosaic genomic features that have facilitated the adaptive evolution of phenotypes that contributed to the water-to-land transition.

## Highlights

- *De novo* reference genome assemblies of four non-teleost ray-finned fishes
- Basal ray-finned fishes have key limb development regulatory elements
- Lung-related genes in early ray-finned fishes hold the key for the lung origin
- Cardio-respiratory systems' co-evolution during early air breathing evolution

Resource

# Tracing the genetic footprints of vertebrate landing in non-teleost ray-finned fishes

Xupeng Bi,<sup>1,2,17</sup> Kun Wang,<sup>3,17</sup> Liandong Yang,<sup>1,17</sup> Hailin Pan,<sup>2,17</sup> Haifeng Jiang,<sup>1,17</sup> Qiwei Wei,<sup>4,17</sup> Miaoquan Fang,<sup>2</sup> Hao Yu,<sup>2</sup> Chenglong Zhu,<sup>3</sup> Yiran Cai,<sup>2</sup> Yuming He,<sup>2</sup> Xiaoni Gan,<sup>1</sup> Honghui Zeng,<sup>1</sup> Daqi Yu,<sup>5,6</sup> Youan Zhu,<sup>7,8</sup> Huifeng Jiang,<sup>9</sup> Qiang Qiu,<sup>3</sup> Huanming Yang,<sup>2,10,11</sup> Yong E. Zhang,<sup>5,6,12</sup> Wen Wang,<sup>3,12,13,\*</sup> Min Zhu,<sup>7,8,14,\*</sup> Shunping He,<sup>1,12,15,\*</sup> and Guojie Zhang<sup>2,12,13,16,18,\*</sup>

<sup>1</sup>State Key Laboratory of Freshwater Ecology and Biotechnology, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, China

<sup>2</sup>BGI-Shenzhen, Shenzhen 518083, China

<sup>3</sup>School of Ecology and Environment, Northwestern Polytechnical University, Xi'an 710072, China

<sup>4</sup>Key Laboratory of Freshwater Biodiversity Conservation, Ministry of Agriculture and Rural Affairs, Yangtze River Fisheries Research Institute, Chinese Academy of Fishery Sciences, Wuhan, China

<sup>5</sup>Key Laboratory of Zoological Systematics and Evolution and State Key Laboratory of Integrated Management of Pest Insects and Rodents, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China

<sup>6</sup>University of Chinese Academy of Sciences, Beijing 100049, China

<sup>7</sup>Key Laboratory of Vertebrate Evolution and Human Origins, Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, 142 Xi-zhi-men-wai Street, Beijing 100044, China

<sup>8</sup>CAS Center for Excellence in Life and Paleoenvironment, Beijing 100044, China

<sup>9</sup>Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin, China

<sup>10</sup>James D. Watson Institute of Genome Sciences, Hangzhou, China

<sup>11</sup>Guangdong Provincial Academician Workstation of BGI Synthetic Genomics, BGI-Shenzhen, Shenzhen 518120, China

<sup>12</sup>Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, 32 Jiaochang Donglu, Kunming 650223, China

<sup>13</sup>State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China

<sup>14</sup>College of Earth and Planetary Sciences, University of Chinese Academy of Sciences, Beijing 100049, China

<sup>15</sup>Institute of Deep-Sea Science and Engineering, Chinese Academy of Sciences, Sanya 572000, China

<sup>16</sup>Villum Center for Biodiversity Genomics, Section for Ecology and Evolution, Department of Biology, University of Copenhagen, Copenhagen, Denmark

<sup>17</sup>These authors contributed equally

<sup>18</sup>Lead contact

\*Correspondence: [wwang@mail.kiz.ac.cn](mailto:wwang@mail.kiz.ac.cn) (W.W.), [zhumin@ivpp.ac.cn](mailto:zhumin@ivpp.ac.cn) (M.Z.), [clad@ihb.ac.cn](mailto:clad@ihb.ac.cn) (S.H.), [guojie.zhang@bio.ku.dk](mailto:guojie.zhang@bio.ku.dk) (G.Z.) <https://doi.org/10.1016/j.cell.2021.01.046>

## SUMMARY

Rich fossil evidence suggests that many traits and functions related to terrestrial evolution were present long before the ancestor of lobe- and ray-finned fishes. Here, we present genome sequences of the bichir, paddlefish, bowfin, and alligator gar, covering all major early divergent lineages of ray-finned fishes. Our analyses show that these species exhibit many mosaic genomic features of lobe- and ray-finned fishes. In particular, many regulatory elements for limb development are present in these fishes, supporting the hypothesis that the relevant ancestral regulation networks emerged before the origin of tetrapods. Transcriptome analyses confirm the homology between the lung and swim bladder and reveal the presence of functional lung-related genes in early ray-finned fishes. Furthermore, we functionally validate the essential role of a jawed vertebrate highly conserved element for cardiovascular development. Our results imply the ancestors of jawed vertebrates already had the potential gene networks for cardio-respiratory systems supporting air breathing.

## INTRODUCTION

The water-to-land transition during the Devonian is one of the most prominent events in vertebrate evolution. During this transition, various organs underwent adaptive changes, particularly those related to locomotion and respiration systems (Hinchliffe, 1994; Long and Gordon, 2004). Study of the genomic changes

that occurred during the origin of tetrapods is challenging because the animals involved in the transitional events are extinct. However, it has long been postulated that many of these innovations evolved in the ancestors of bony vertebrates (ray-finned fishes, lobe-finned fishes, and tetrapods) and thus still are present in some living basal groups of actinopterygians (ray-finned fishes) (Graham et al., 2014; Takeuchi et al., 2009;

Zaccone et al., 2009). For instance, the lung respiratory system that tetrapods rely on for survival also exists in several basal lobe-finned fishes, such as coelacanth, lungfish, and Polypteriformes, one of the early divergent ray-finned fish groups. Anatomically, the lungs are distributed in pairs on the ventral side and originate from the pharynx posterior to the gills. While the lungs disappeared and were replaced by the swim bladder in most ray-finned fishes, the presence of the lung or its homologous organs in the stem jawed vertebrates has been proposed for decades (Denison, 1941; Goujet, 2011; Perry et al., 2001). It is clear that the lungs should have originated by at least the last common ancestor (LCA) of ray-finned fishes and lobe-finned fishes, which is at least in Late Silurian, not very far away from the LCA of cartilaginous and bony fishes (presumably Upper Ordovician) (Andreev et al., 2015). Thus, the genetic basis of the traits that evolved during this major transition may be accessible through comparison of existing developmental pathways that could mirror the developmental plasticity in the ancestral lineage (Amemiya et al., 2013; Gibert, 2017; Muschick et al., 2011; West-Eberhard, 2003). If so, investigation on the genomes of extant species from these early divergent lineages of ray-finned fishes may illuminate the genetic regulatory processes underlying the trait evolution during the terrestrial transition (Qu et al., 2015).

The ray-finned fishes (Actinopterygii) are an extremely diverse group of vertebrates with over 32,000 species, thus representing about half of all extant vertebrate species (Friedman, 2015; Salan, 2014). Most Actinopterygii (ca. 95%) are teleosts, which initially diversified through adaptive radiation in the Mesozoic and underwent further dramatic diversification in the Cenozoic (Glasauer and Neuhauss, 2014). The other extant ray-finned fishes are distributed in three deeply divergent actinopterygian groups (Cladistia, Chondrostei, and Holostei), which are often regarded as “living fossils” due to their ancient evolutionary origins and possession of traits ascribed to a common ancestor of ray-finned fishes and lobe-finned fishes (Hurley et al., 2007; Noack et al., 1996). Many species in these groups possess fascinating characteristics that have been associated with air-breathing during vertebrate terrestrialization, such as ventral paired lungs (Farmer, 1999; Perry and Sander, 2004). Living cladistians (*Polypterus* and *Erpetoichthys*) also bear morphological similarities to the lobe-finned fishes, such as elongated body forms, large paired spiracles dorsally positioned on the head, and developed lobate portion of the pectoral fins. These similarities either evolved through convergent evolution (Friedman, 2015; Giles et al., 2017) or were inherited from the primitive forms of gnathostomes (Zhu et al., 2013; Zhu and Yu, 2009). These mosaic features place the non-teleost actinopterygians in a phylogenetic position that is crucial for comparative evolutionary and developmental studies that aim to illuminate some of the key transitions that occurred during the evolution of terrestrial vertebrates. Thus, to elucidate the genetic basis of these features in early diverging ray-finned fishes, we produced a chromosome level assembly of the bichir (*Polypterus senegalus*) and draft genome assemblies for the American paddlefish (*Polyodon spathula*), the bowfin (*Amia calva*), and the alligator gar (*Atractosteus spatula*), which together cover all three teleost outgroup lineages. Our analyses showed that these species exhibit many mosaic genomic

features of lobe- and ray-finned fishes, particularly relating to regulatory elements for limb development. We then undertook transcriptomic-based exploration to confirm the homology between the lung and swim bladder and revealed the presence of functionally lung-related genes in early ray-finned fishes. Finally, we performed functional experiments to validate the essential role of an element that is highly conserved in jawed vertebrate (gnathostome) ancestors in cardiovascular development.

## RESULTS

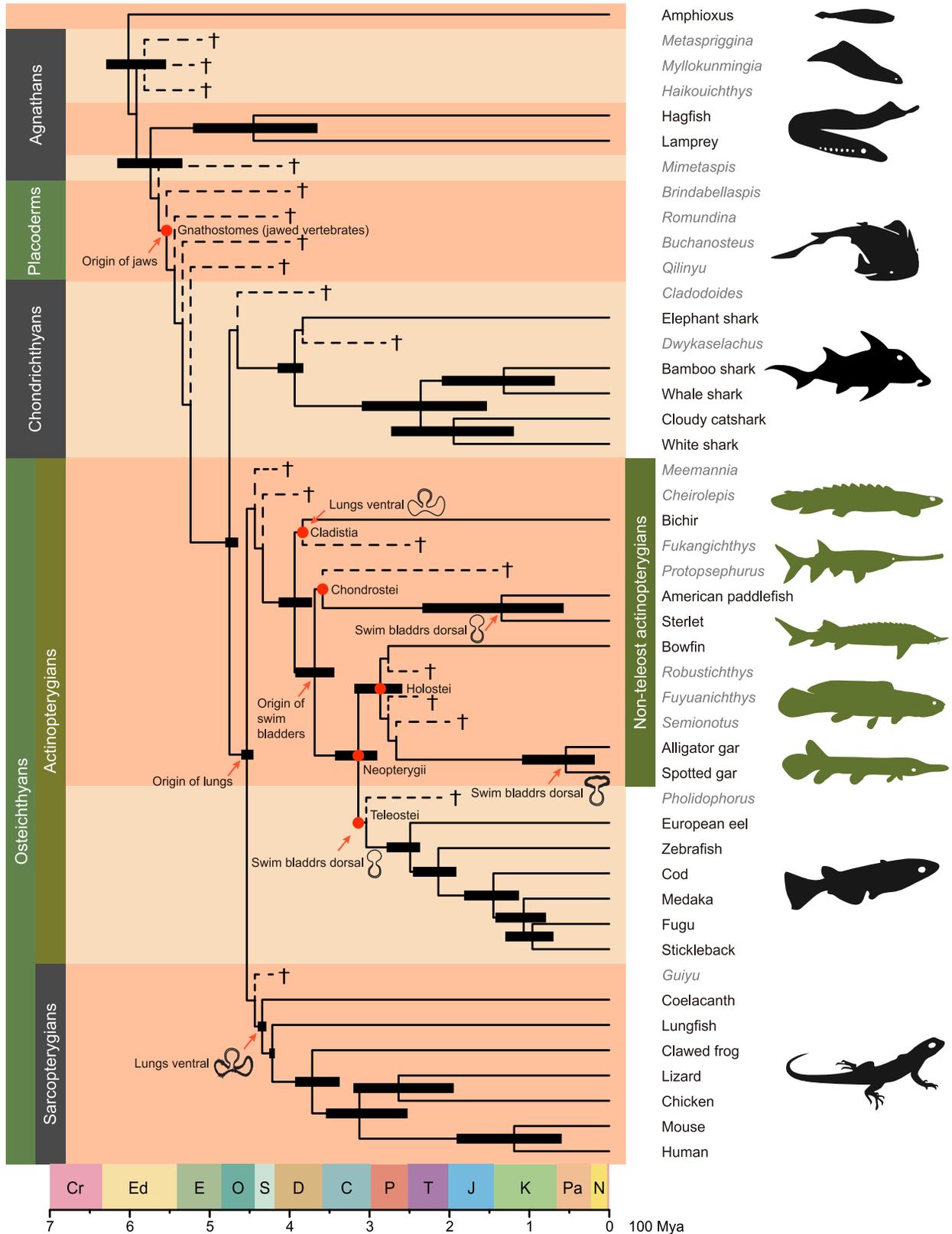
### Genome assembly and annotation

We generated from a bichir 250 Gb of long sequence reads using Oxford Nanopore Technology and used the data to produce a draft genome assembly. The assembly was polished using Illumina short reads and then further improved into a chromosome-level assembly with Hi-C data. The final assembly of the bichir genome has contig N50 size of 4.53 Mb and scaffold N50 size of 189.69 Mb, with 94.3% of the sequences anchored onto 18 chromosomes. The genomes of the American paddlefish, bowfin, and alligator gar were sequenced and assembled at 135–242x coverage with Illumina short reads from libraries with inserted sizes ranging from 250 bp to 40 kb (Table S1). From 89%–96% of the 2,586 vertebrate core sets of Benchmarking Universal Single-Copy Orthologs (BUSCO) are completely annotated in these four species (Simão et al., 2015) (Table S1). We also produced transcriptome data from nine tissues for each species to assist the annotation, which resulted in 18,839–23,374 protein coding genes for these genomes (Table S1).

The sizes of the non-teleost actinopterygians genomes studied to date vary considerably, from 0.95 to 3.67 Gb, which implies dramatic changes happened to their genomic structures during the early divergence of ray-finned fishes. Some of this size variation is due to differences in their proportions of repeat content, which range from 14% in the gar to 44% in the bichir (Table S1). Interestingly, TcMar-Tc1 seems to be the most abundant type of transposon in all basal ray-finned fishes (Figure S1A), in accordance with previous findings that TcMar-Tc1 transposons are active in freshwater species (Yuan et al., 2018). In addition, similar to the genomes of lobe-finned and cartilaginous fishes, these genomes contain abundant LINE/CR1 repeats, which might have originated from their common ancestor (Figure S1B).

### Basic interrelationships of ray-finned fishes

Despite extensive analyses on both extinct and extant species since the late nineteenth century (Allen, 1876; Cope, 1878), the early evolutionary history of actinopterygians remains controversial. The major debate regarding extant lineages concerns the relationship between Chondrostei and Holostei. While some studies have argued they are a monophyletic clade (Inoue et al., 2003; Venkatesh et al., 2001), others have indicated that Acipenseriformes (sturgeons and paddlefishes), an order of the Chondrostei subclass, should be treated as a sister group of the Neopterygii that is composed of Holostei and Teleostei (Kikugawa et al., 2004). We identified 359 one-to-one orthologs across 27 chordates (allowing up to one missing) that derived from the same ancestral chromosomes before the two rounds of whole-genome duplications (WGDs) that occurred at the



(legend on next page)

base of vertebrates (Sacerdot et al., 2018) (Figure S1C). Phylogenetic analyses with these orthologous genes confirmed that the Polypteriformes (bichirs) form a sister lineage of all other ray-finned fishes and resolved that Acipenseriformes are the sister lineage of the Neopterygii (Figure 1). This topology is also supported by morphological evidence (Gardiner et al., 2005) and further confirmed by the phylogenetic trees generated with 23 Mb of genome-wide orthologous sequences and 319 single copy genes of 14 selected vertebrate species (Figures S1D and S1E). Molecular dating using the 359 orthologs and calibrated using nine fossil records (Table S1) suggests that actinopterygians split from sarcopterygians about 453 million years ago (Mya, 95% confidence interval: 445.8–459.8 Mya), which is close to a previous estimate (422 Mya) based on mitochondrial genomes (Lalu et al., 2010).

### Genome evolution of early ray-finned fishes

The split of actinopterygians and sarcopterygians was a major evolutionary step during the transition toward current vertebrate diversity (Evans, 2008; Soltis and Soltis, 2012; Volff, 2005). Using the chromosome assemblies of bichir, sterlet, and spotted gar (Braasch et al., 2016; Du et al., 2020), we reconstructed the ancestral karyotype of ray-finned fishes and obtained 38 proto-chromosomes for Actinopterygii (Figure 2A; see Figure S2A for a full description of karyotype reconstruction). Remarkably, the non-teleost actinopterygians have a higher level of genome synteny with chicken than teleosts, whose genomes have undergone more fusions, fissions, and rearrangements following teleost-specific genome duplication (Figures 2B and S2B). For example, we found that 51% of the bichir genome has conserved synteny with the chicken genome. When expanding the analyses to include more jawed vertebrates, we found that non-teleost actinopterygians share 9%–36% synteny with humans, thus higher than both the proportions shared by any of the teleosts (6%–7%) examined and humans (Table S2). These findings clearly indicate that the early actinopterygians largely preserved the ancestral genome structure of bony vertebrates before the teleost-specific genome duplication. Despite the overall conservation with the ancestral genome, 22 chromosome fusion events appear to have occurred in the bichir lineage, in which ancestral micro-chromosomes merged with other macro-chromosomes. These rearrangement processes were accompanied with an elevated level of transposon element activity, which increases genome size in these species.

Recent analyses have shown that a third round of WGD (3R WGD) occurred in the sterlet (*Acipenser ruthenus*) (Du et al., 2020), a species closely related to the paddlefish. We detected seven Hox clusters in the paddlefish genome, indicating that a

3R WGD also occurred in paddlefish (Figure S2C). To investigate whether the 3R WGD event occurred in the common ancestor of paddlefish and sterlet, or independently in the two lineages, we analyzed the distributions of synonymous substitution rates (Ks) and 4-fold degenerative third-codon transversions (4dTv) among the 3,743 paddlefish and 8,297 sterlet paralogs. We found that the Ks age distributions of the two species had different ranges, suggesting that independent WGD events occurred in the two lineages (Figure 2C). This pattern was also confirmed by both the distribution of 4dTv values, and gene phylogeny analyses (Figures S2D and S2E). Based on the divergence time (135 Mya) between paddlefish and sterlet (Figure 1), we estimate that the 3R WGD events in paddlefish and sterlet occurred at 121 Mya and 51 Mya, respectively. Although our estimated date for the 3R WGD in sterlet is much more recent than that estimated in a previous study, that prior estimate based on a sterlet-Atlantic sturgeon divergence time of 166 Mya, which is much earlier than the divergence time (74 Mya) estimated in other studies (Rabosky et al., 2013).

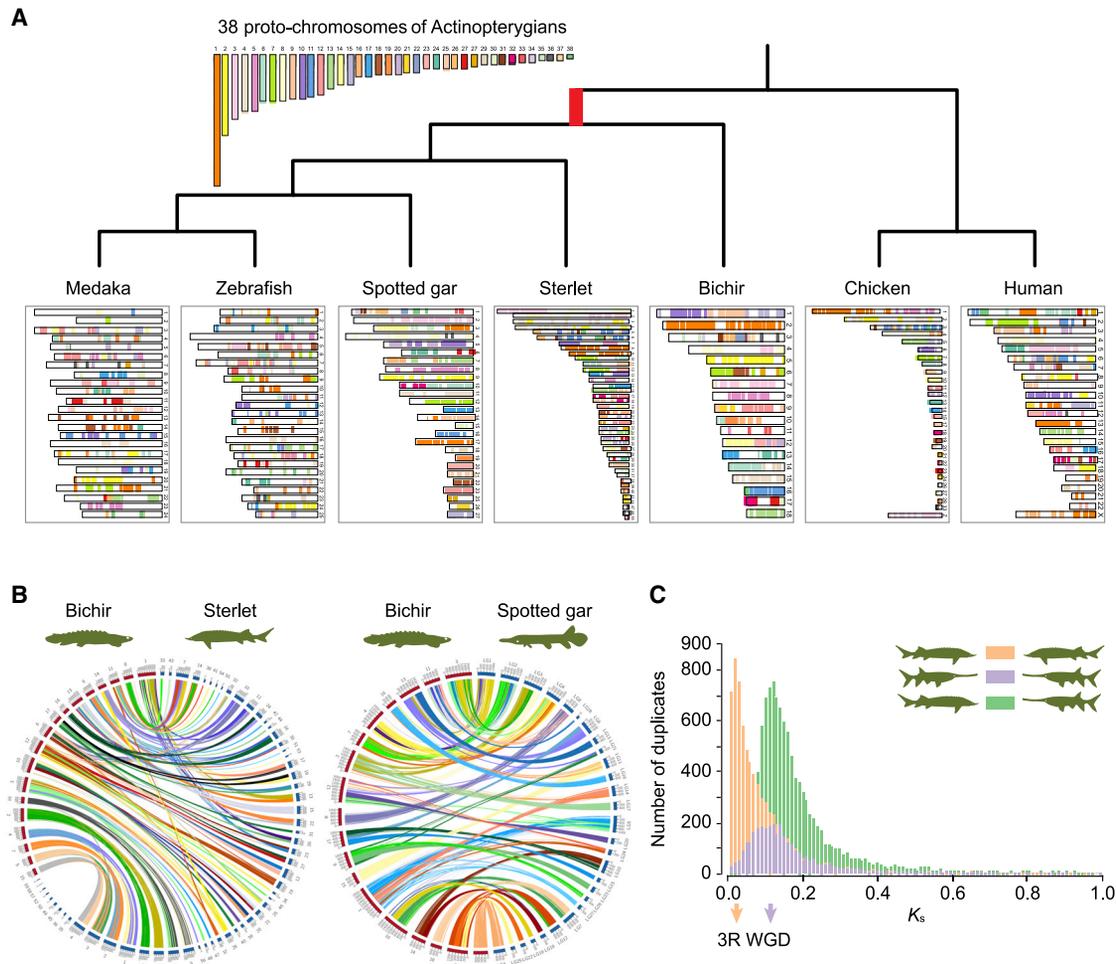
The independent duplication of the whole genome in these two closely related lineages offers an excellent model for studying the evolutionary fate of duplicate genes after WGD. Unlike the sterlet, which retained eight Hox clusters, the paddlefish genome has lost one HoxC cluster since its 3R WGD. Interestingly, convergent loss of this HoxC cluster after 3R WGD also occurred in some teleost fishes, such as fugu (*Takifugu rubripes*) and medaka (*Oryzias latipes*) (Figure S2C). At the global genome level, based on 8,377 singletons and 3,743 paralogous pairs detected in the paddlefish genome, we estimated a duplicate gene retention rate after the paddlefish 3R WGD of around 31%, which is considerably lower than the retention rate in sterlet (70%), in accordance with the indications that the 3R WGD was more recent in the sterlet than in the paddlefish. However, the rate in the paddlefish is higher than the corresponding rate (15%–20%) following the teleost WGD (Hrbek et al., 2007). Notably, duplicated genes involved in protein turnover, signal transduction, cell proliferation, and development were preferably maintained in the sterlet (Du et al., 2020), but duplicated genes in the paddlefish are enriched in functions related to biosynthetic process, gene expression, and metabolic process (Table S2). These findings suggest that substantially different selection pressures acted on the two lineages.

### Genetic regulation related to limb flexibility

The evolution of terrestrial locomotion, one of the major transitions in vertebrate evolution, involved the appearance of derived supporting appendicular structures (Carroll and Holmes, 2007;

### Figure 1. Interrelationship and divergence times of a broad selection of living and extinct chordates

Amphioxus was used as an outgroup, and a species tree was generated using RAxML and ASTRAL on the basis of 359 orthologs. The divergence times of extant species were estimated with the MCMCTree program under a correlated molecular clock and general time reversible (GTR) substitution model. Extinct species representing taxa in each of the major clades were added to the phylogenetic tree using information in previous literature (Coates et al., 2017; Morris, 2008; Dupret et al., 2014; Giles et al., 2015; 2017; Grande et al., 2002; Maisey, 2005; Olsen and McCune, 1991; Shu et al., 1999; 2003; Stensiö, 1927; Xu et al., 2018; Young, 1980; Zhu et al., 2013; 2009). The black rectangles at the nodes represent 95% confidence intervals of the corresponding estimated divergence times. The “†” symbols indicate branches leading to extinct species with fossil records (dated to times indicated by the ends of the respective branches). First appearances of the jaw, lung, and swim bladder are marked on the tree, and geological periods are indicated at the bottom (from left to right: Cryogenian, Ediacaran, Cambrian, Ordovician, Silurian, Devonian, Carboniferous, Permian, Triassic, Jurassic, Cretaceous, Paleogene, Neogene, and Quaternary).



**Figure 2. Reconstruction of proto-chromosomes for the common ancestor of ray-finned fishes**

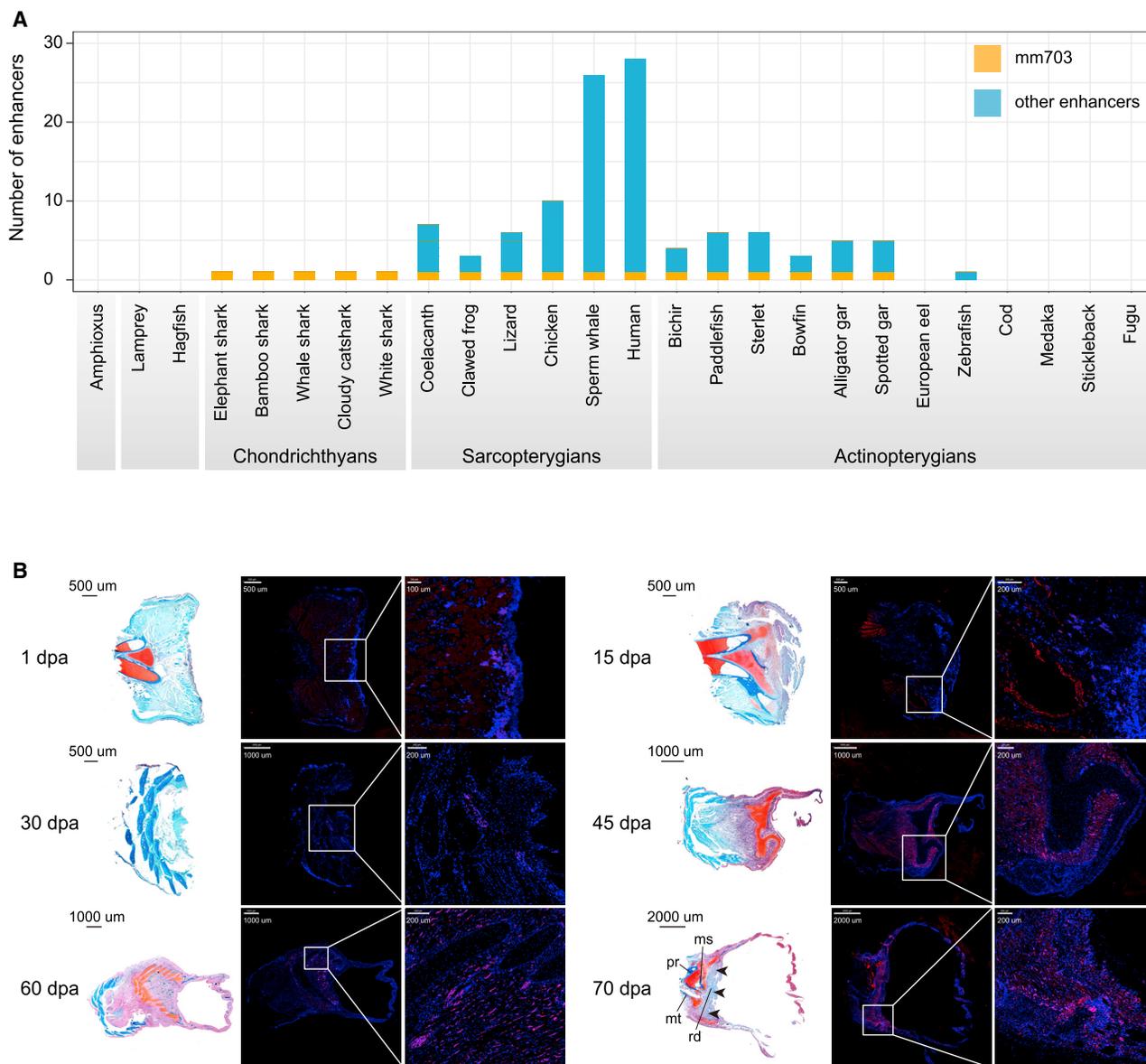
(A) 38 Actinopterygii proto-chromosomes (color-coded) were constructed using chicken as an outgroup.

(B) Circos plots showing conserved synteny between bichir and sterlet (left) and bichir and spotted gar (right).

(C)  $K_s$  distributions of 3,743 paralogous genes in paddlefish (purple), 8,297 paralogous genes in sterlet (orange), and 10,729 orthologs between paddlefish and sterlet (green). Times of putative 3R WGD events are indicated by the x axis.

Long and Gordon, 2004; Neyt et al., 2000). Some of these anatomical features are also present in the basal ray-finned fishes (Clack, 2009). For example, the humerus is a proximal basal element of the sarcopterygian forelimb, which evolved from the metapterygium and putatively appeared in the common ancestor of cartilaginous and bony fishes but was completely lost in the teleost pectoral fin (Davis et al., 2004; Tanaka, 2016; Wagner and Chiu, 2001; Woltering et al., 2020). A previous study on the fleshy pectoral fins of coelacanth showed that the flexible and movable elbow joint connects the humerus and ulna with two crisscrossed ligaments (Miyake et al., 2016). Meanwhile, *Polypterus* possess preaxial and postaxial muscles in both pectoral and pelvic fins that resemble those in the lobe-finned fishes *Latimeria* and *Neoceratodus* (Molnar et al., 2017). We reason that an ancestral developmental potential for the flexible movement of pectoral fins through the joint connections might have been present in the crown jawed vertebrates but was lost in teleosts.

To test this hypothesis, we first searched for the presences of homologs for 264 genes related to limb morphogenesis development, limb bud formation, and limb joint morphogenesis in all jawed vertebrates. However, we did not detect much difference in the copy numbers of these genes when comparing teleosts with other vertebrates, nor did we detect specific gene loss in teleosts, suggesting the loss of movement flexibility in teleost might not be through the significant change over the protein-coding sequences (Table S3). Therefore, we next screened for the presence of conserved enhancer sequences that have been functionally validated in mice. Among 14 tissues investigated, teleosts showed the lowest numbers of conserved enhancers with mice in all tissue types, which might partially be due to the higher genomic turnover rates after the 3R WGD in this group. Interestingly, we found there was a much higher number of conserved enhancers that specifically function in limbs, in non-teleost actinopterygians, lobe-finned fishes, and cartilaginous fishes, than in teleosts. Specifically,



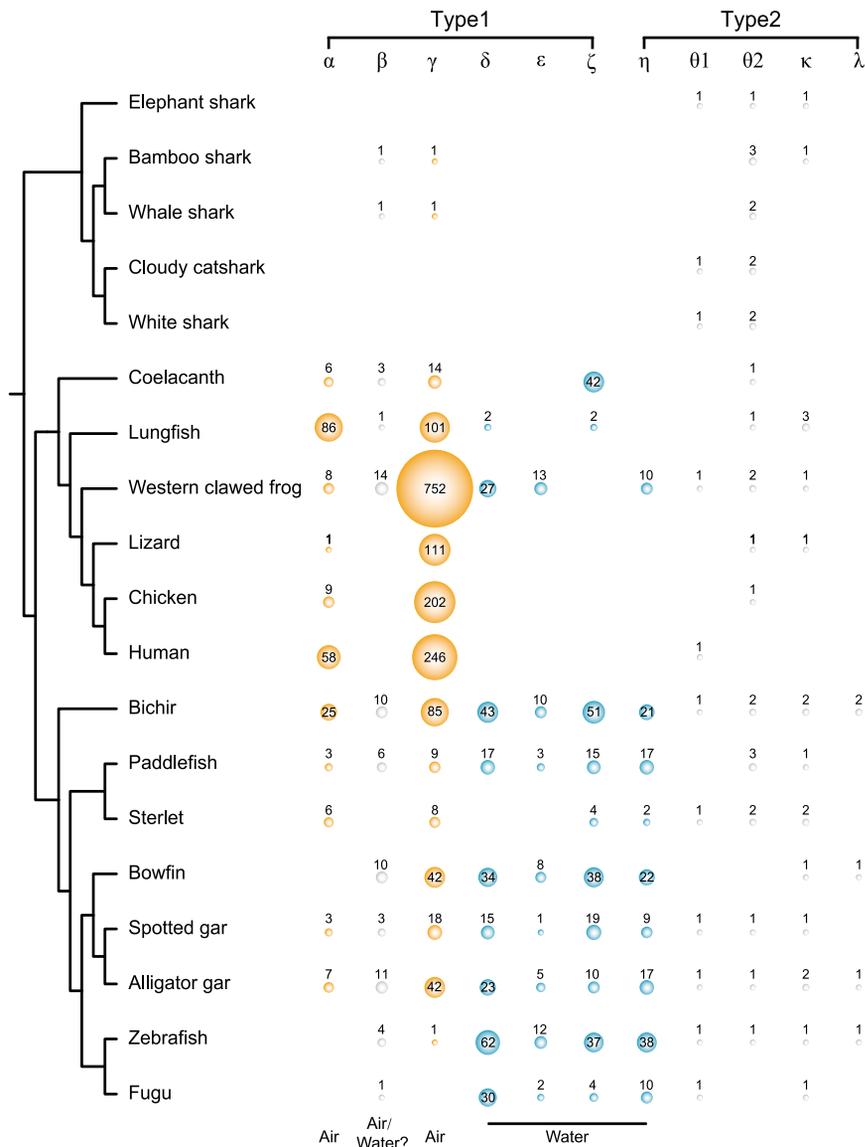
**Figure 3. Early limb development and *in situ* expression of *Osr2* in the pectoral fin of bichir**

(A) Annotation results for the 40 limb-related enhancers in 25 chordate species. The mm703 enhancer originated after emergence of crown gnathostomes and is preserved in lobe-finned fishes, tetrapods, and basal ray-finned fishes (as indicated by yellow bars).

(B) Pectoral fin and endoskeleton regeneration in bichir. Panels to the left show Safranin O/Fast green staining at indicated days after amputation (see also Figure S3B), with cartilage and bone stained red and blue, respectively. The left side in these images is near the base of the pectoral fin. Middle and right panels show results of *in situ* hybridization of the *Osr2* gene at different magnifications, revealing that *Osr2* is weakly expressed in blastema at 1 dpa (days post amputation) and highly expressed from 30 dpa. The blue dot is DAPI staining, showing the morphology and location of the nucleus. Pink around the blue dot indicates that the gene is expressed here. Abbreviations: ms, mesopterygium; mt, metapterygium; pr, propterygium; rd, radials (rd). Black arrows in the 70 dpa image indicate synovial joints in the pectoral fin.

across the six teleost genomes we investigated, only one species (zebrafish) contains one limb-specific enhancer. In contrast, we identified 3–6 in non-teleost actinopterygian species and 3–40 in sarcopterygians (Figure 3A). Notably, all of the seven limb enhancers detected in all non-teleost actinopterygian species are also present in the coelacanth genome (Table S3).

Strikingly, we detected a highly conserved limb enhancer (mm703) in all five cartilaginous fishes and all bony vertebrates except teleosts. This enhancer is located in the downstream region of *Osr2*. A previous study in mice has revealed that mm703 is a cis-regulatory module of the *Lmx1b* transcription factor that could regulate the expression of the *Osr2* gene (Haro et al., 2017). Study of the *Osr2* transcription factor in mice has shown



**Figure 4. Properties of olfactory receptors in jawed vertebrates**

“Air” (yellow circles) and “water” (blue circles) refer to the detection of airborne and water-soluble odorants, respectively. Sizes of the circles indicate numbers of intact OR genes, which possess a full-length OR protein coding sequence that is at least 250 amino acids long.

the regeneration process (Figures 3B and S3B). Though further functional experiments would be valuable to reveal the detail mechanism of how mm703 and *Osr2* mediate joint modification, our findings suggest an ancient role of *Osr2* in synovial joints, which might have evolved in early vertebrates.

#### Early origination of air olfactory receptors

Vertebrate genomes contain numerous olfactory receptor (OR) genes, predominantly expressed in the olfactory epithelium of the nasal cavity where they play key roles in detection of odorous molecules in the environment (Zhang and Firestein, 2002). We characterized OR genes and found cartilaginous fishes have much smaller (~3) numbers of intact OR genes, while the numbers have been expanded to ~100 and ~200 in ray-finned fishes and lobe-finned fishes, respectively. Group  $\alpha$  and  $\gamma$  OR genes are required to detect airborne odorants and are the dominant OR type in tetrapods but absent in teleost fishes, which instead have a large number of OR genes that participate in detection of water-soluble odorants (Niimura, 2009). Interestingly, most non-teleost actinopterygians have

that this gene is essential for formation of synovial joints (Gao et al., 2011), which greatly enhances the flexibility of limb movement and thus has a presumed association with vertebrate terrestrialization (Askary et al., 2016). Our Hi-C data also revealed a close interaction between this enhancer and *Osr2* in bichir, and the pattern was conserved with that in mice, implying a conserved regulatory function might also be present in the bichir (Figure S3A). Consistent with this, it has been shown that the bichir can regenerate pectoral fins with remarkable accuracy, resembling the limb regeneration process found in amphibians (Cuervo et al., 2012). To further infer the potential role of *Osr2* in the bichir pectoral fin, we monitored the expression profile of *Osr2* during regeneration of the bichir pectoral fin (as detailed in the STAR methods section), a homologous structure to the front limb in tetrapods. We found that *Osr2* was mainly expressed at joints in the pectoral fin, especially between the metapterygium and radials, and its expression level increased during

mosaic OR features with a mixture of both types of OR genes, in accordance with the air-breathing characteristics of some species in these lineages (Figure 4; Table S4). Among them, the bichir has the highest number of both  $\alpha$  and  $\gamma$  ORs, which is consistent with previous finding that spiracle-mediated aspiration accounts for up to 93% of all air breaths in *Polypterus* (Graham et al., 2014). It should be noted that both  $\alpha$  and  $\gamma$  ORs had already appeared in bony vertebrates and expanded separately in different lineages in a later period (Figure S4A). Nevertheless, there is no evidence so far that Acipenseriformes can breathe in the air. Accordingly, there are fewer OR genes in this group than in other non-teleost actinopterygians. Notably, previous studies have detected no OR gene related to air-breathing in lampreys or elephant sharks (Niimura, 2009), but we found a  $\gamma$  gene in the bamboo shark (*Chiloscyllium punctatum*) and whale shark (*Rhincodon typus*) (Figure 4). The presence of  $\alpha$  and  $\gamma$  in some of the fishes that cannot breathe air remains mysterious. We



speculate that these ORs might primitively have been functionally associated with water-soluble odorants but later specialized toward interactions with airborne odors in cooperation with aerial olfaction during the fish-tetrapod transition. However, further functional experiments are needed to verify this hypothesis.

In addition to the main olfactory system, many vertebrates also use the vomeronasal system as an accessory olfactory system for pheromone detection. Most terrestrial vertebrates have a higher copy number of vomeronasal type-1 receptors (V1Rs) than type-2 receptors (V2Rs), which respectively detect small volatile molecules diffusing in air and molecules that are soluble in aquatic environments (Silva and Antunes, 2017). Although basal ray-finned fishes have more V2Rs than V1Rs like teleosts, we found that the bichir has significantly higher copy numbers of V1Rs than other teleost fishes (Figure S4B). The exact roles of V1Rs in this aquatic fish thus would be interesting for further experimental study.

### Transcriptional profiles reveal the evolution of vertebrate lungs

To address the evolution of lungs and their relationship with the swim bladders of fish, a question that has intrigued researchers for centuries, we sequenced transcriptomes from 355 samples representing 10 tissues of eight vertebrate species (three lobe-finned fishes, three basal ray-finned fishes, and two teleosts) with five biological replications per tissue per species (Table S5). An expression level matrix based on 5,046 1:1 orthologs of the eight species was generated and standardized with quantile normalization. The principal-component analysis (PCA) clustered the lung and swim bladder expression profiles, something also supported by the neighbor-joining (NJ) tree of all tissues based on gene expression levels after normalization (Figure 5A). This result is in accordance with Darwin's hypothesis that they are homologous organs (Darwin, 1859) and the previous anatomic analysis of pulmonary arteries in non-teleost actinopterygians that also showed lungs and swim bladders are homologous organs (Longo et al., 2013). Additionally, the topology of the tree generated with transcriptome data strongly resembles the species phylogeny, demonstrating the high correlation between genomic and transcriptomic evolution (Figure S5A).

We also assessed the phylogenetic relationships between lung and swim bladder samples based on transcriptome distances that were evaluated as 1 Spearman's correlation coefficient. We found the swim bladders were nested within the lung samples, which supports the hypothesis that the swim bladder evolved from the lung, with a locational transition from the ventral to the dorsal side of the esophagus (Sagai et al., 2017) after divergence of Polypteriformes and other ray-finned lineages (Figures 5B and S5B). Moreover, the swim bladder of the alligator gar clustered with the bichir lung rather than the swim bladder of paddlefish, indicating that the alligator gar's swim bladder retains the function of the ancestral lungs that evolved in the common ancestor of ray-finned fishes and lobe-finned fishes (Figure 5B). This is consistent with the anatomical structures of the species' lungs and swim bladders, as surfaces of the bichir's lung and the alligator gar's swim bladder are highly vascularized, unlike the surface of the paddlefish swim bladder (Figures 5C and S5C).

A Kolmogorov-Smirnov test showed that expression levels of 461 genes (after quantile normalization) were significantly higher ( $p < 0.01$  after correction) in the swim bladders than in all lung samples. We found that some of the most strongly expressed genes in swim bladder (e.g., *Ruvbl1*, *Amot*, and *Mmp2*) are associated with cell migration (Figure 5B), which might have been involved in the ventral to dorsal transition during swim bladder development. We also identified 466 genes that are more strongly expressed in air-breathing than in non-air-breathing vertebrates. These genes are significantly enriched in functions related to development of alveoli and alveolar lamellar bodies (Table S5). Among them we found 24 genes related to angiogenesis, which is crucial for formation of blood vessels at the surface of alveoli (Carmeliet and Jain, 2011), and 15 of these were among the top 50 most significantly differentially expressed genes (Figures 5B and 5D). These findings support the above-described differences in anatomical structure between lungs and swim bladders.

We further compared the lung expression profiles with those of all other tissues in the bichir and the three lobe-finned species and identified 11 genes exhibiting lung-specific expression (Figure 5E). Interestingly, all of them reportedly participate in lung development and function (Anagnostopoulou et al., 2012; Arora

### Figure 5. Analyses of basal ray-finned species illuminating the evolution of lung-related gene expression

(A) Score plot obtained from PCA of 355 transcriptome samples based on expression profiling of 5,046 orthologs. The NJ tree in the top right corner is based on expression data and indicates a homologous relationship between lung and swim bladder. The red dotted circle represents the clustering of lungs and swim bladders.

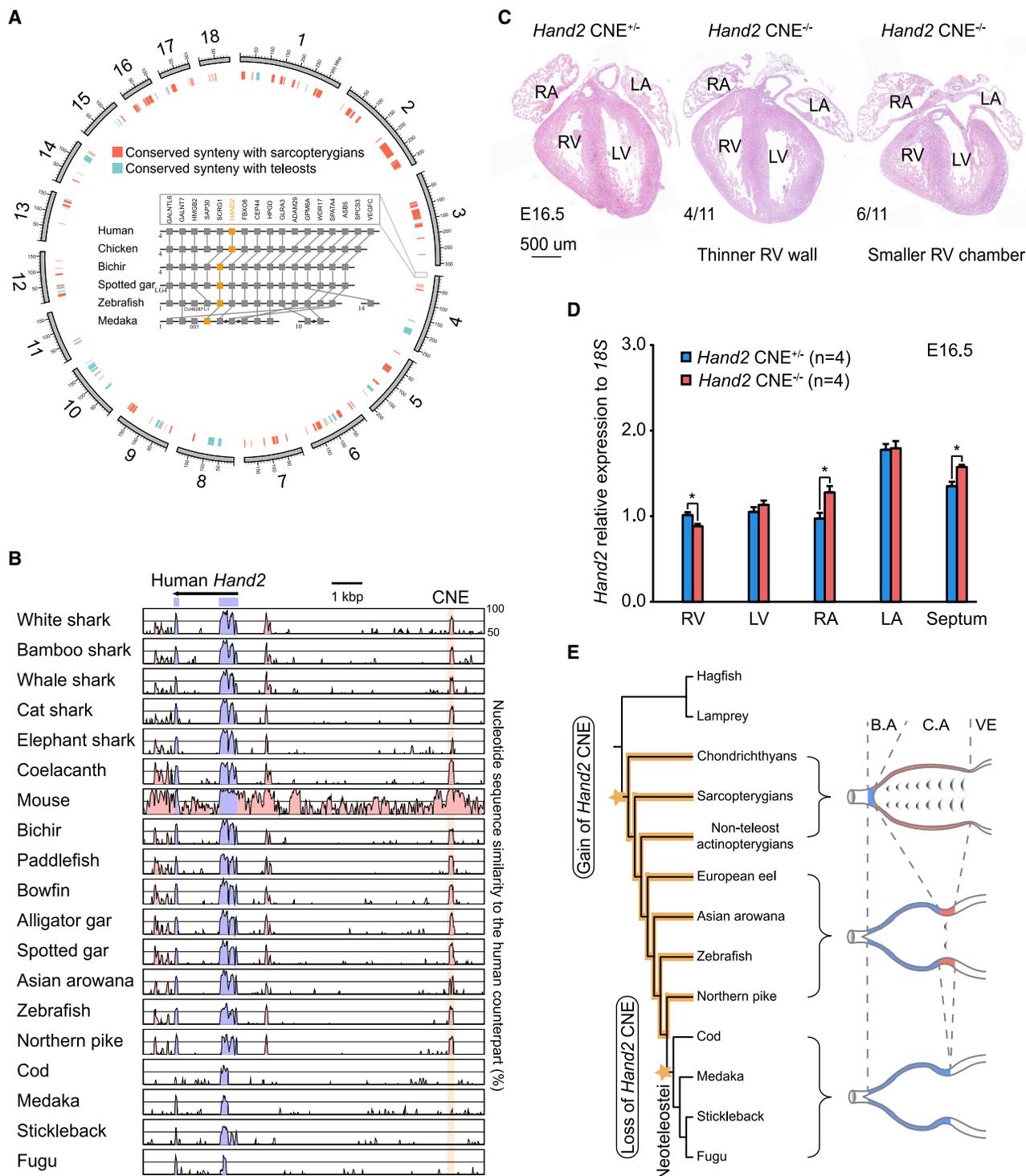
(B) NJ tree of lung and swim bladder samples based on transcriptome distances, clearly separating actinopterygians and sarcopterygians. Schematic morphological diagrams of lungs and swim bladders of each species are displayed under the tree. The heatmap below shows expression profiles of 21 of the most significantly highly expressed genes in the air-breathing and swim bladder groups.

(C) Anatomic illustration of the lungs of the bichir and swim bladders of the American paddlefish and alligator gar.

(D) The angiogenesis process involves three consecutive steps. First an endothelial cell tip cell is selected to ensure blood vessel branching formation. Second, tip cells navigate in response to guidance signals and adhere to the extracellular matrix to migrate, while stalk cells proliferate and elongate. Finally, neighboring branches fuse, and a new vessel forms. Eleven highly expressed genes in the air-breathing group participate during these three processes. Genes in the 50 most highly expressed sets are marked in blue.

(E) The presence and absence of lung-related genes in representative vertebrates.

(F) Results of *in situ* hybridization analysis of five representative genes in lungs of the bichir and the mouse and swim bladder of the zebrafish, showing that these genes are expressed in the region of the alveolar cells (indicated by the arrowhead). Zebrafish have two copies of *TBX3*, *NRP1*, and *VEGFA*, one copy of a *VIPR2* gene, but completely lost *SLC26A9*. Zebrafish-a and -b show the expression pattern of the paralogous copies when the paralogous genes exist. The red signals show the expression of the targeted gene.



**Figure 6. Genomic sequences specifically shared by non-teleost actinopterygians and lobe-finned fishes**

(A) Distributions of conserved synteny blocks in the bichir genome shared with sarcopterygians or teleosts. The gene alignment map in the middle shows the syntenic map for the *Hand2* gene.

(B) VISTA plot showing the presence of a heart-related CNE upstream of the *Hand2* gene across all jawed vertebrates except the Neoteleostei. Peaks (blue, exons; red, non-coding regions) indicate regions with conserved sequences to the human counterpart. The *Hand2* CNE is highlighted in pale yellow.

(C) Hematoxylin and eosin-stained transverse sections of E16.5 hearts. *Hand2* CNE<sup>-/-</sup> mutant hearts showed slightly reduced right ventricular compact myocardium of right ventricle body wall (4 out of 11) and significantly decreased chamber volume (6 out of 11). Numbers in the bottom indicate the hearts with

(legend continued on next page)

et al., 2012; Hong et al., 2007; Morrisey and Hogan, 2010; Quaggin et al., 1999; Vaidya et al., 2017). Most of the 11 genes are also expressed in swim bladders of teleosts, suggesting that they still play functional roles in both swim bladders and lungs. One of these genes (*Tcf21*) has been apparently present since the common ancestor of chordates, and two (*Foxf1* and *Hpgd*) apparently originated with the common ancestor of vertebrates. The others appeared after the split of jawed and jawless vertebrates and are present in all jawed vertebrates. It is striking to observe such a high proportion of lung-specific expression genes in the bichir. These include two well-known lung development genes (*Tbx4* and *Tbx5*) that were apparently present in the common ancestor of cartilaginous and bony fishes. It is also consistent with a previous report on an important lung-related regulatory enhancer (lung mesenchyme-specific enhancer, LME) of *Tbx4*, which is present in cartilaginous fishes and also exists in the bichir, bowfin, and spotted gar (Hara et al., 2018; Tatsumi et al., 2016) (Figure S5D). *In situ* hybridization experiments with five of these genes confirmed that they are all highly active in alveolar cells of both the mouse and the bichir but weakly expressed in the swim bladder surface of the zebrafish (Figure 5F). Previous analysis of fossils has shown that the pharynx has been drastically rebuilt during the origin of gnathostomes, and the lungs were derivatives of the pharynx posterior to the gills (Perry et al., 2001). Considering that some of these lung-related genes (such as *Tbx3*) are also expressed in the pharyngeal region (Mesbah et al., 2008), we hypothesize that they might have appeared in the common ancestor of cartilaginous and bony fishes and play a role in forming the distinctive pharynx in gnathostome (Dupret et al., 2014; Kuratani and Ahlberg, 2018).

### Concomitant changes in circulatory systems

Vertebrates have evolved sophisticated cardio-respiratory systems. The respiratory system provides an adequate supply of oxygen to maintain proper cardiac function (under conditions that are not excessively stressful), while the circulatory system efficiently transfers the oxygen to other organs. During the evolution of air-breathing respiration, selection pressures may have driven co-adaptive changes in the circulatory system. In the lung circulatory system, the oxygen-depleted blood is pumped away from the heart to lungs through the pulmonary artery. After oxygenation, the blood flows back to heart through the pulmonary veins. However, most air-breathing fish did not evolve the specialized chamber to separate the two streams of blood and thus mix deoxygenated and oxygenated blood in their circulatory system (Ishimatsu, 2012). It has been hypothesized that dominance of the conus arteriosus was an ancestral characteristic of the vertebrate cardiac outflow tract (Icardo, 2006; Lorenzale et al., 2018). The conus arteriosus emerged in the ancestor of jawed vertebrates; is preserved in cartilaginous

fishes, non-teleost actinopterygians, basal teleosts, and amphibians; and has been presumably absorbed into the right ventricle (RV) in amniotic animals and completely lost in the more advanced teleosts (Icardo, 2006; Maldanis et al., 2016). This tissue contains compact myocardium and functions as an extra ventricle or accessory pumping chamber that minimizes pressure fluctuations and regulates the outflow of blood from the ventricle to the ventral aorta (Icardo et al., 2002a; 2002b; Lorenzale et al., 2018; Satchell and Jones, 1967).

Through genome synteny comparisons, we found that over 15% of the bichir genome has conserved genomic synteny with the lobe-finned fish genomes but only 6% with teleosts (Figure 6A). Interestingly, gene ontology (GO) enrichment analysis for genes shared the same synteny between non-teleost actinopterygians, and lobe-finned fishes showed significantly enriched in the cardiac development-related GO terms (Figure S6A), supporting the different cardiac functions displayed between the two groups. It also pinpointed that these highly conserved synteny blocks over long evolutionary history might be favored by natural selection and possibly maintain the similar function through conserved enhancer regulation that present in the same genomic regulatory blocks (Kikuta et al., 2007).

From the specifically conserved synteny between non-teleost actinopterygians and sarcopterygians, we found a highly conserved non-coding element (CNE) in all jawed vertebrates, except the Neoteleostei, located near the *Hand2* gene (Figure 6B). *Hand2* is essential for cardiac morphogenesis, especially formation of the RV. Its expression is also positively correlated with cardiomyocyte proliferation (Han et al., 2019). To understand the function of this CNE, we knocked out the element in mice, which resulted in perinatal lethality for homozygous mutated animals (Figures S6B and S6C). The homozygous knockout mice developed a thin, small RV and hence congenital heart defects (Figures 6C and S6D). We also detected weaker expression of *Hand2* in the RV of homozygous knockout than in the heterozygous knockouts (Figure 6D). Thus, our knockout experiment suggested that conserved CNE has a regulatory role in *Hand2* expression patterns. We also detected this CNE in all selected non-Neoteleostei teleosts (Elopomorpha, Osteoglossomorpha, Otocephala, and Protacanthopterygii) but not other crown teleost groups (Figure 6E). This is consistent with the anatomical observation that a conus arteriosus-like structure is present in the cardiac outflow tracts of basal teleost lineages but has been completely lost in advanced teleosts, which have a single chamber (bulbus arteriosus) in these tracts (Icardo, 2006). Thus, we speculate that loss of this CNE in the Neoteleostei might explain the complete absence of conus arteriosus in these fishes, which have developed a diversified cardiovascular system that co-appeared with the gas bladder respiratory

defect out of the total hearts analyzed. The cell nuclei are stained blue with hematoxylin, and the extracellular matrix and cytoplasm pink with eosin. LA, left atrium; LV, left ventricle; RA, right atrium; RV, right ventricle; C.A, conus arteriosus; B.A, bulbus arteriosus; VE, ventricle.

(D) Results of qRT-PCR analysis of *Hand2* gene expression in cardiomyocytes dissected from E16.5 embryonic mouse hearts. Data are mean  $\pm$  SEM (\* $p$  < 0.05, t test).

(E) Hypothetical transitions of cardiac outflow anatomic structure. Dominance of the conus arteriosus in the cardiac outflow tract evolved in the common ancestor of jawed vertebrates and was preserved in chondrichthyans, sarcopterygians, and non-teleost actinopterygians. A conus arteriosus-like structure can still be detected in some basal teleosts but has been completely lost in other teleosts.

system in adaptation to aquatic environments with low oxygen tensions (Harter and Brauner, 2017).

## DISCUSSION

The high-quality genome assemblies of four basal ray-finned fishes obtained in this study have filled gaps in knowledge of early evolutionary processes of bony fish genomes. We find these fishes possess many ancestral genomic features of the bony fishes. Particularly, it is remarkable that many limb- and lung-related genes and associated regulatory elements are present in cartilaginous and bony fishes, as validated by a number of experiments, suggesting that the ancestral regulatory networks that latterly used to participate in the terrestrial locomotion and air-breathing functions have evolved in early jawed vertebrates. Our study suggests some of these ancestral developmental regulation networks have been preserved in non-teleost actinopterygians, which might explain their mosaic biological features between lobe-finned fishes and ray-finned fishes. These ancestral developmental potentials have also facilitated the adaptive evolution of many phenotypes that contributed to the vertebrate landing.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Source organism
- METHOD DETAILS
  - Genome sequencing, assembly and annotation
  - Phylogeny reconstruction
  - Reconstruction of ancestral karyotypes
  - Synteny analysis
  - Whole genome duplication events analysis
  - Limb enhancer identification
  - Fin regeneration and *Osr2* *in situ* expression
  - Identification of OR and VR genes
  - Transcriptome analysis
  - *In situ* hybridization of lung-specific genes
  - Knockout of the *Hand2* CNE
- QUANTIFICATION AND STATISTICAL ANALYSIS

## Supplemental information

Supplemental Information can be found online at <https://doi.org/10.1016/j.cell.2021.01.046>.

## ACKNOWLEDGMENTS

We thank Qiwu Xu, Guangyi Fan, and Xiaoli Chen for Hi-C library construction and sequencing, and also thank the support from the Fish10K project and Wuhan Frasergen Bioinformatics Co., Ltd. We thank Tom Gilbert for the helpful suggestions. This research was supported by grants from the State Key Lab-

oratory of Freshwater Ecology and Biotechnology, Institute of Hydrobiology, Chinese Academy of Sciences and from the National Natural Science Foundation of China (41876179 and 31972866) to S.H. and funded by the Strategic Priority Research Program of the Chinese Academy of Sciences grants (no. XDB13000000 and no. XDB31020000) and a Villum Investigator grant (no. 25900) from The Villum Foundation to G.Z. It was also supported by the Guangdong Provincial Academician Workstation of BGI Synthetic Genomics (no. 2017B090904014) to H. Yang and Youth Innovation Promotion Association, Chinese Academy of Sciences to L.Y. (<http://www.yicas.cn>).

## AUTHOR CONTRIBUTIONS

G.Z., S.H., M.Z., and W.W. conceived the study. L.Y. and Haifeng Jiang collected the materials. L.Y., Haifeng Jiang, X.G., and H.Z. performed the morphological laboratory work. X.B. and H. Yu performed the genome assembly and genome annotation. H.P. and X.B. designed evolutionary analyses. X.B., H.P., K.W., M.F., C.Z., Y.C., and Y.H. performed evolutionary analyses. S.H., M.Z., W.W., H.P., Q.W., Y.E.Z., H. Yang, Q.Q., Huifeng Jiang, Y.Z., and D.Y. participated in discussions and provided suggestions. G.Z., X.B., and K.W. wrote the manuscript with the input from all co-authors.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 15, 2020

Revised: November 11, 2020

Accepted: January 27, 2021

Published: February 4, 2021

## REFERENCES

- Allen, J.A. (1876). *The American bison, living and extinct* Vol. 10 (University Press).
- Amemiya, C.T., Alföldi, J., Lee, A.P., Fan, S., Philippe, H., Maccallum, I., Braasch, I., Manousaki, T., Schneider, I., Rohner, N., et al. (2013). The African coelacanth genome provides insights into tetrapod evolution. *Nature* 496, 311–316.
- Anagnostopoulou, P., Riederer, B., Duerr, J., Michel, S., Binia, A., Agrawal, R., Liu, X., Kalitzki, K., Xiao, F., Chen, M., et al. (2012). SLC26A9-mediated chloride secretion prevents mucus obstruction in airway inflammation. *J. Clin. Invest.* 122, 3629–3634.
- Andreev, P.S., Coates, M.I., Shelton, R.M., Cooper, P.R., Smith, M.P., and Sansom, I.J. (2015). Upper Ordovician chondrichthyan-like scales from North America. *Palaeontology* 58, 691–704.
- Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Birney, E., Biswas, M., Bucher, P., Cerutti, L., Corpet, F., Croning, M.D., et al. (2001). The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Res.* 29, 37–40.
- Arora, R., Metzger, R.J., and Papaioannou, V.E. (2012). Multiple roles and interactions of Tbx4 and Tbx5 in development of the respiratory system. *PLoS Genet.* 8, e1002866.
- Askary, A., Smeeton, J., Paul, S., Schindler, S., Braasch, I., Ellis, N.A., Postlethwait, J., Miller, C.T., and Crump, J.G. (2016). Ancient origin of lubricated joints in bony vertebrates. *eLife* 5, e16415.
- Birney, E., Clamp, M., and Durbin, R. (2004). GeneWise and Genomewise. *Genome Res.* 14, 988–995.
- Braasch, I., Gehrke, A.R., Smith, J.J., Kawasaki, K., Manousaki, T., Pasquier, J., Amores, A., Desvignes, T., Batzel, P., Catchen, J., et al. (2016). The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. *Nat. Genet.* 48, 427–437.
- Carmeliet, P., and Jain, R.K. (2011). Molecular mechanisms and clinical applications of angiogenesis. *Nature* 473, 298–307.

- Carroll, R.L., and Holmes, R. (2007). Evolution of the appendicular skeleton of amphibians. In *Fins into Limbs: Evolution, Development, and Transformation*, B.K. Hall, ed. (The University of Chicago Press), pp. 185–224.
- Chauve, C., and Tannier, E. (2008). A methodological framework for the reconstruction of contiguous regions of ancestral genomes and its application to mammalian genomes. *PLoS Comput Biol.* 4, e1000234.
- Chen, Y., Chen, Y., Shi, C., Huang, Z., Zhang, Y., Li, S., Li, Y., Ye, J., Yu, C., Li, Z., et al. (2018). SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *Gigascience* 7, 1–6.
- Clack, J.A. (2009). The fin to limb transition: new data, interpretations, and hypotheses from paleontology and developmental biology. *Annual Review of Earth and Planetary Sciences* 37, 163–179.
- Coates, M.I., Gess, R.W., Finarelli, J.A., Criswell, K.E., and Tietjen, K. (2017). A symmoriform chondrichthyan braincase and the origin of chimaeroid fishes. *Nature* 541, 208–211.
- Morris, S.C. (2008). A redescription of a rare chordate, *metaspriggina walcotti* Simonetta and Insom, from the Burgess shale (middle Cambrian), British Columbia, Canada. *Journal of Paleontology* 82, 424–430.
- Cope, E.D. (1878). *Descriptions of New Extinct Vertebrata from the Upper Tertiary and Dakota Formations.* (Harvard University).
- Cuervo, R., Hernández-Martínez, R., Chimal-Monroy, J., Merchant-Larios, H., and Covarrubias, L. (2012). Full regeneration of the tribasal Polypterus fin. *Proc. Natl. Acad. Sci. USA* 109, 3838–3843.
- Darwin, C. (1859). *On the Origin of Species.* (Harvard University Press).
- Davis, M.C., Shubin, N.H., and Force, A. (2004). Pectoral fin and girdle development in the basal actinopterygians *Polyodon spathula* and *Acipenser transmontanus*. *J. Morphol.* 262, 608–628.
- Denison, R.H. (1941). The soft anatomy of *Bothriolepis*. *Journal of Paleontology* 15, 553–561.
- Du, K., Stöck, M., Kneitz, S., Klopp, C., Woltering, J.M., Adolphi, M.C., Feron, R., Prokopov, D., Makunin, A., Kichigin, I., et al. (2020). The sterlet sturgeon genome sequence and the mechanisms of segmental rediploidization. *Nat. Ecol. Evol.* 4, 841–852.
- Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C., Shamim, M.S., Machol, I., Lander, E.S., Aiden, A.P., and Aiden, E.L. (2017). De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* 356, 92–95.
- Dupret, V., Sanchez, S., Goujet, D., Tafforeau, P., and Ahlberg, P.E. (2014). A primitive placoderm sheds light on the origin of the jawed vertebrate face. *Nature* 507, 500–503.
- Edgar, R.C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5, 113.
- Evans, B.J. (2008). Genome evolution and speciation genetics of clawed frogs (*Xenopus* and *Silurana*). *Front. Biosci.* 13, 4687–4706.
- Farmer, C.G. (1999). Evolution of the vertebrate cardio-pulmonary system. *Annu. Rev. Physiol.* 61, 573–592.
- Frazer, K.A., Pachter, L., Poliakov, A., Rubin, E.M., and Dubchak, I. (2004). VISTA: computational tools for comparative genomics. *Nucleic Acids Res.* 32, W273–9.
- Friedman, M. (2015). The early evolution of ray-finned fishes. *Palaeontology* 58, 213–228.
- Gao, Y., Lan, Y., Liu, H., and Jiang, R. (2011). The zinc finger transcription factors *Osr1* and *Osr2* control synovial joint formation. *Dev. Biol.* 352, 83–91.
- Gardiner, B.G., Schaeffer, B., and Masserie, J.A. (2005). A review of the lower actinopterygian phylogeny. *Zoological Journal of the Linnean Society* 144, 511–525.
- Gibert, J.M. (2017). The flexible stem hypothesis: evidence from genetic data. *Dev. Genes Evol.* 227, 297–307.
- Giles, S., Coates, M.I., Garwood, R.J., Brazeau, M.D., Atwood, R., Johanson, Z., and Friedman, M. (2015). Endoskeletal structure in *Cheirolepis* (Osteichthyes, Actinopterygii), An early ray-finned fish. *Palaeontology* 58, 849–870.
- Giles, S., Xu, G.-H., Near, T.J., and Friedman, M. (2017). Early members of ‘living fossil’ lineage imply later origin of modern ray-finned fishes. *Nature* 549, 265–268.
- Glasauer, S.M., and Neuhauss, S.C. (2014). Whole-genome duplication in teleost fishes and its evolutionary consequences. *Mol. Genet. Genomics* 289, 1045–1060.
- Goldman, N., and Yang, Z. (1994). A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.* 11, 725–736.
- Goujet, D. (2011). “Lungs” in Placoderms, a persistent palaeobiological myth related to environmental preconceived interpretations. *Comptes Rendus Palevol* 10, 323–329.
- Graham, J.B., Wegner, N.C., Miller, L.A., Jew, C.J., Lai, N.C., Berquist, R.M., Frank, L.R., and Long, J.A. (2014). Spiracular air breathing in polypterid fishes and its implications for aerial respiration in stem tetrapods. *Nat. Commun.* 5, 3022.
- Grande, L., Jin, F., Yabumoto, Y., and Bemis, W.E. (2002). *Protosephurus liui*, a well-preserved primitive paddlefish (Acipenseriformes: Polyodontidae) from the Lower Cretaceous of China. *Journal of Vertebrate Paleontology* 22, 209–237.
- Han, X., Zhang, J., Liu, Y., Fan, X., Ai, S., Luo, Y., Li, X., Jin, H., Luo, S., Zheng, H., et al. (2019). The lncRNA *Hand2os1/Uph* locus orchestrates heart development through regulation of precise expression of *Hand2*. *Development* 146, dev176198.
- Hara, Y., Yamaguchi, K., Onimaru, K., Kadota, M., Koyanagi, M., Keeley, S.D., Tatsumi, K., Tanaka, K., Motone, F., Kageyama, Y., et al. (2018). Shark genomes provide insights into elasmobranch evolution and the origin of vertebrates. *Nat. Ecol. Evol.* 2, 1761–1771.
- Haro, E., Watson, B.A., Feenstra, J.M., Tegeler, L., Pira, C.U., Mohan, S., and Oberg, K.C. (2017). Lmx1b-targeted cis-regulatory modules involved in limb dorsalization. *Development* 144, 2009–2020.
- Harris, R.S. (2007). Improved pairwise Alignment of genomic DNA. PhD thesis (Pennsylvania State University).
- Harter, T.S., and Brauner, C.J. (2017). The O<sub>2</sub> and CO<sub>2</sub> Transport System in Teleosts and the Specialized Mechanisms That Enhance Hb–O<sub>2</sub> Unloading to Tissues. In *Fish Physiology*, A.K. Gamperl, T.E. Gillis, A.P. Farrell, and C.J. Brauner, eds. (Elsevier), pp. 1–106.
- Hinchliffe, J.R. (1994). Evolutionary developmental biology of the tetrapod limb. *Dev. Suppl.* 1994, 163–168.
- Hong, T.-M., Chen, Y.-L., Wu, Y.-Y., Yuan, A., Chao, Y.-C., Chung, Y.-C., Wu, M.-H., Yang, S.-C., Pan, S.-H., and Shih, J.-Y. (2007). Targeting neuropilin 1 as an antitumor strategy in lung cancer. *Clin. Cancer Res.* 13, 4759–4768.
- Hrbek, T., Seckinger, J., and Meyer, A. (2007). A phylogenetic and biogeographic perspective on the evolution of poeciliid fishes. *Molecular Phylogenetics and Evolution* 43, 986–998.
- Hurley, I.A., Mueller, R.L., Dunn, K.A., Schmidt, E.J., Friedman, M., Ho, R.K., Prince, V.E., Yang, Z., Thomas, M.G., and Coates, M.I. (2007). A new time-scale for ray-finned fish evolution. *Proc. Biol. Sci.* 274, 489–498.
- Icardo, J.M. (2006). *Conus arteriosus* of the teleost heart: dismissed, but not missed. *Anat. Rec. A Discov. Mol. Cell. Evol. Biol.* 288, 900–908.
- Icardo, J.M., Colvee, E., Cerra, M.C., and Tota, B. (2002a). Structure of the conus arteriosus of the sturgeon (*Acipenser naccarii*) heart. I: the conus valves and the subendocardium. *Anat. Rec.* 267, 17–27.
- Icardo, J.M., Colvee, E., Cerra, M.C., and Tota, B. (2002b). The structure of the conus arteriosus of the sturgeon (*Acipenser naccarii*) heart: II. The myocardium, the subepicardium, and the conus-aorta transition. *Anat. Rec.* 268, 388–398.
- Inoue, J.G., Miya, M., Tsukamoto, K., and Nishida, M. (2003). Basal actinopterygian relationships: a mitogenomic perspective on the phylogeny of the “ancient fish”. *Mol. Phylogenet. Evol.* 26, 110–120.
- Ishimatsu, A. (2012). Evolution of the cardiorespiratory system in air-breathing fishes. *Aqua-BioScience Monographs* 5, 1–28.

- Jiang, Y., Loh, Y.E., Rajarajan, P., Hirayama, T., Liao, W., Kassim, B.S., Javidfar, B., Hartley, B.J., Kleofas, L., Park, R.B., et al. (2017). The methyltransferase SETDB1 regulates a large neuron-specific topological chromatin domain. *Nat. Genet.* **49**, 1239–1250.
- Katoh, M. (2013). Functional and cancer genomics of ASXL family members. *Br. J. Cancer* **109**, 299–306.
- Kikugawa, K., Katoh, K., Kuraku, S., Sakurai, H., Ishida, O., Iwabe, N., and Miyata, T. (2004). Basal jawed vertebrate phylogeny inferred from multiple nuclear DNA-coded genes. *BMC Biol.* **2**, 3.
- Kikuta, H., Laplante, M., Navratilova, P., Komisarczuk, A.Z., Engström, P.G., Fredman, D., Akalin, A., Caccamo, M., Sealy, I., Howe, K., et al. (2007). Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates. *Genome Res.* **17**, 545–555.
- Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360.
- Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59.
- Kuratani, S., and Ahlberg, P.E. (2018). Evolution of the vertebrate neurocranium: problems of the premandibular domain and the origin of the trabecula. *Zoological Lett.* **4**, 1.
- Lalu, X.C., Kosen, J.D., Tjakrawidjaja, A.H., Kusumah, R.V., Sadhotomo, B., Pouyaud, L., Slembrouck, J., and Paradis, E. (2010). Mitochondrial genomic divergence in coelacanth (Latimeria): slow rate of evolution or recent speciation? *Mar. Biol.* **157**, 2253–2262.
- Li, L., Stoeckert, C.J., Jr., and Roos, D.S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178–2189.
- Livak, K.J., and Schmittgen, T.D.J.m. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>-</sup>(Delta Delta C(T)) method. *Methods* **25**, 402–408.
- Long, J.A., and Gordon, M.S. (2004). The greatest step in vertebrate history: a paleobiological review of the fish-tetrapod transition. *Physiol. Biochem. Zool.* **77**, 700–719.
- Longo, S., Riccio, M., and McCune, A.R. (2013). Homology of lungs and gas bladders: insights from arterial vasculature. *Journal of Morphology* **274**, 687–703.
- Lorenzale, M., López-Unzu, M.A., Rodríguez, C., Fernández, B., Durán, A.C., and Sans-Coma, V. (2018). The anatomical components of the cardiac outflow tract of chondrichthyans and actinopterygians. *Biol. Rev. Camb. Philos. Soc.* **93**, 1604–1619.
- Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., et al. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* **1**, 18.
- Maisey, J.G. (2005). Braincase of the Upper Devonian shark *Cladodoides wildungensis* (Chondrichthyes, Elasmobranchii), with observations on the braincase in early chondrichthyans. *Bulletin of the American Museum of Natural History* **2005**, 1–103.
- Majoros, W.H., Pertea, M., and Salzberg, S.L. (2004). TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics* **20**, 2878–2879.
- Maldanis, L., Carvalho, M., Almeida, M.R., Freitas, F.I., de Andrade, J.A.F.G., Nunes, R.S., Rochitte, C.E., Poppi, R.J., Freitas, R.O., Rodrigues, F., et al. (2016). Heart fossilization is possible and informs the evolution of cardiac outflow tract in vertebrates. *eLife* **5**, e14698.
- Matsui, A., Go, Y., and Niimura, Y. (2010). Degeneration of olfactory receptor gene repertoires in primates: no direct link to full trichromatic vision. *Mol. Biol. Evol.* **27**, 1192–1200.
- Mesbah, K., Harrelson, Z., Théveniau-Ruissy, M., Papaioannou, V.E., and Kelly, R.G. (2008). Tbx3 is required for outflow tract development. *Circ. Res.* **103**, 743–750.
- Mirarab, S., Reaz, R., Bayzid, M.S., Zimmermann, T., Swenson, M.S., and Warnow, T. (2014). ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics* **30**, i541–i548.
- Miyake, T., Kumamoto, M., Iwata, M., Sato, R., Okabe, M., Koie, H., Kumai, N., Fujii, K., Matsuzaki, K., Nakamura, C., et al. (2016). The pectoral fin muscles of the coelacanth *Latimeria chalumnae*: Functional and evolutionary implications for the fin-to-limb transition and subsequent evolution of tetrapods. *Anat. Rec. (Hoboken)* **299**, 1203–1223.
- Molnar, J.L., Johnston, P.S., Esteve-Altava, B., and Diogo, R. (2017). Musculoskeletal anatomy of the pelvic fin of *Polypterus*: implications for phylogenetic distribution and homology of pre- and postaxial pelvic appendicular muscles. *J. Anat.* **230**, 532–541.
- Morrisey, E.E., and Hogan, B.L. (2010). Preparing for the first breath: genetic and cellular mechanisms in lung development. *Dev. Cell* **18**, 8–23.
- Muschick, M., Barluenga, M., Salzburger, W., and Meyer, A. (2011). Adaptive phenotypic plasticity in the Midas cichlid fish pharyngeal jaw and its relevance in adaptive radiation. *BMC Evol. Biol.* **11**, 116.
- Neyt, C., Jagla, K., Thisse, C., Thisse, B., Haines, L., and Currie, P.D. (2000). Evolutionary origins of vertebrate appendicular muscle. *Nature* **408**, 82–86.
- Niimura, Y. (2009). On the origin and evolution of vertebrate olfactory receptor genes: comparative genome analysis among 23 chordate species. *Genome Biol. Evol.* **1**, 34–44.
- Noack, K., Zardoya, R., and Meyer, A. (1996). The complete mitochondrial DNA sequence of the bichir (*Polypterus ornatipinnis*), a basal ray-finned fish: ancient establishment of the consensus vertebrate gene order. *Genetics* **144**, 1165–1180.
- Olsen, P.E., and McCune, A.R. (1991). Morphology of the *Semionotus elegans* species group from the Early Jurassic part of the Newark Supergroup of Eastern North America with comments on the family Semionotidae (Neopterygii). *Journal of Vertebrate Paleontology* **11**, 269–292.
- Perry, S.F., and Sander, M. (2004). Reconstructing the evolution of the respiratory apparatus in tetrapods. *Respir. Physiol. Neurobiol.* **144**, 125–139.
- Perry, S.F., Wilson, R.J., Straus, C., Harris, M.B., and Remmers, J.E. (2001). Which came first, the lung or the breath? *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* **129**, 37–47.
- Qu, Q., Haitina, T., Zhu, M., and Ahlberg, P.E. (2015). New genomic and fossil data illuminate the origin of enamel. *Nature* **526**, 108–111.
- Quaggin, S.E., Schwartz, L., Cui, S., Igarashi, P., Deimling, J., Post, M., and Rossant, J. (1999). The basic-helix-loop-helix protein pod1 is critically important for kidney and lung organogenesis. *Development* **126**, 5771–5783.
- Rabosky, D.L., Santini, F., Eastman, J., Smith, S.A., Sidlauskas, B., Chang, J., and Alfaro, M.E. (2013). Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. *Nat. Commun.* **4**, 1958.
- Ruan, J., and Li, H. (2020). Fast and accurate long-read assembly with wtdbg2. *Nat. Methods* **17**, 155–158.
- Sacerdot, C., Louis, A., Bon, C., Berthelot, C., and Crollius, H. (2018). Chromosome evolution at the origin of the ancestral vertebrate genome. *Genome Biol.* **19**, 1–15.
- Sagai, T., Amano, T., Maeno, A., Kimura, T., Nakamoto, M., Takehana, Y., Naruse, K., Okada, N., Kiyonari, H., and Shiroishi, T. (2017). Evolution of Shh endoderm enhancers during morphological transition from ventral lungs to dorsal gas bladder. *Nat. Commun.* **8**, 14300.
- Sallan, L.C. (2014). Major issues in the origins of ray-finned fish (Actinopterygii) biodiversity. *Biol. Rev. Camb. Philos. Soc.* **89**, 950–971.
- Satchell, G.H., and Jones, M.P. (1967). The function of the conus arteriosus in the Port Jackson shark, *Heterodontus portusjacksoni*. *J. Exp. Biol.* **46**, 373–382.
- Servant, N., Varoquaux, N., Lajoie, B.R., Viara, E., Chen, C.J., Vert, J.P., Heard, E., Dekker, J., and Barillot, E. (2015). HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259.
- Shu, D.G., Luo, H.-L., Morris, S.C., Zhang, X.L., Hu, S.-X., Chen, L., Han, J., Zhu, M., Li, Y., and Chen, L.-Z. (1999). Lower Cambrian vertebrates from south China. *Nature* **402**, 42–46.

- Shu, D.G., Morris, S.C., Han, J., Zhang, Z.F., Yasui, K., Janvier, P., Chen, L., Zhang, X.L., Liu, J.N., Li, Y., and Liu, H.Q. (2003). Head and backbone of the Early Cambrian vertebrate *Haikouichthys*. *Nature* **421**, 526–529.
- Sievers, F., and Higgins, D.G. (2014). Clustal Omega, accurate alignment of very large numbers of sequences. *Methods Mol. Biol.* **1079**, 105–116.
- Silva, L., and Antunes, A. (2017). Vomeronasal Receptors in Vertebrates and the Evolution of Pheromone Detection. *Annu. Rev. Anim. Biosci.* **5**, 353–370.
- Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212.
- Soltis, P.S., and Soltis, D.E. (2012). Polyploidy and genome evolution *Volume 665* (Springer).
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313.
- Stanke, M., Diekhans, M., Baertsch, R., and Haussler, D. (2008). Using native and syntetically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**, 637–644.
- Stensiö, E.A. (1927). The Downtonian and Devonian vertebrates of Spitsbergen: Family Cephalaspidae 1 (I kommission hos J. Dybwad).
- Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**, W609–12.
- Takeuchi, M., Okabe, M., and Aizawa, S. (2009). The genus *Polypterus* (bichirs): a fish group diverged at the stem of ray-finned fishes (Actinopterygii). *Cold Spring Harb. Protoc.* **2009**, pdb.emo117.
- Tanaka, M. (2016). Fins into limbs: Autopod acquisition and anterior elements reduction by modifying gene networks involving 5'Hox, Gli3, and Shh. *Dev. Biol.* **413**, 1–7.
- Tatsumi, N., Kobayashi, R., Yano, T., Noda, M., Fujimura, K., Okada, N., and Okabe, M. (2016). Molecular developmental mechanism in polypterid fish provides insight into the origin of vertebrate lungs. *Sci. Rep.* **6**, 30580.
- Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111.
- Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515.
- Vaidya, R., Zambrano, R., Hummler, J.K., Luo, S., Duncan, M.R., Young, K., Lau, L.F., and Wu, S. (2017). Recombinant CCN1 prevents hyperoxia-induced lung injury in neonatal rats. *Pediatr. Res.* **82**, 863–871.
- Vaser, R., Sović, I., Nagarajan, N., and Šikić, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* **27**, 737–746.
- Venkatesh, B., Erdmann, M.V., and Brenner, S. (2001). Molecular synapomorphies resolve evolutionary relationships of extant jawed vertebrates. *Proc. Natl. Acad. Sci. USA* **98**, 11382–11387.
- Volff, J.N. (2005). Genome evolution and biodiversity in teleost fish. *Heredity* **94**, 280–294.
- Wagner, G.P., and Chiu, C.H. (2001). The tetrapod limb: a hypothesis on its origin. *J. Exp. Zool.* **291**, 226–240.
- Wang, Y., Tang, H., Debarry, J.D., Tan, X., Li, J., Wang, X., Lee, T.H., Jin, H., Marler, B., Guo, H., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**, e49.
- Wang, K., Wang, J., Zhu, C., Yang, L., Ren, Y., Ruan, J., Fan, G., Hu, J., Xu, W., Bi, X., et al. (2021). The African lungfish genome sheds light on the vertebrate water-to-land transition. *Cell* **184**, this issue, ■■■■-■■■■.
- West-Eberhard, M.J. (2003). *Developmental plasticity and evolution*. (Oxford University Press).
- Wolff, J., Bhardwaj, V., Nothjunge, S., Richard, G., Renschler, G., Gilsbach, R., Manke, T., Backofen, R., Ramírez, F., and Grüning, B.A. (2018). Galaxy HiCExplorer: a web server for reproducible Hi-C data analysis, quality control and visualization. *Nucleic Acids Res.* **46** (W1), W11–W16.
- Woltering, J.M., Irisarri, I., Ericsson, R., Joss, J.M.P., Sordino, P., and Meyer, A. (2020). Sarcopterygian fin ontogeny elucidates the origin of hands with digits. *Sci. Adv.* **6**, eabc3510.
- Xu, G.-H., Ma, X.-Y., and Ren, Y. (2018). *Fuyuanichthys wangi* gen. et sp. nov. from the Middle Triassic (Ladinian) of China highlights the early diversification of ginglymodian fishes. *PeerJ* **6**, e6054.
- Yanai, I., Benjamin, H., Shmoish, M., Chalifa-Caspi, V., Shklar, M., Ophir, R., Bar-Even, A., Horn-Saban, S., Safran, M., Domany, E., et al. (2005). Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* **21**, 650–659.
- Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591.
- Yang, Z., and Rannala, B. (2006). Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Mol. Biol. Evol.* **23**, 212–226.
- Young, G.C. (1980). A new Early Devonian placoderm from New South Wales, Australia, with a discussion of placoderm phylogeny. *Palaeontographica Abteilung A* **167**, 10–76.
- Yuan, Z., Liu, S., Zhou, T., Tian, C., Bao, L., Dunham, R., and Liu, Z. (2018). Comparative genome analysis of 52 fish species suggests differential associations of repetitive elements with their living aquatic environments. *BMC Genomics* **19**, 141.
- Zaccone, G., Mauceri, A., Maisano, M., and Fasulo, S. (2009). Innervation of lung and heart in the ray-finned fish, bichirs. *Acta Histochem.* **111**, 217–229.
- Zhang, X., and Firestein, S. (2002). The olfactory receptor gene superfamily of the mouse. *Nat. Neurosci.* **5**, 124–133.
- Zhu, M., and Yu, X. (2009). Stem sarcopterygians have primitive polybasal fin articulation. *Palaeontology* **5**, 372–375.
- Zhu, M., Zhao, W., Jia, L., Lu, J., Qiao, T., and Qu, Q. (2009). The oldest articulated osteichthyan reveals mosaic gnathostome characters. *Nature* **458**, 469–474.
- Zhu, M., Yu, X., Ahlberg, P.E., Choo, B., Lu, J., Qiao, T., Qu, Q., Zhao, W., Jia, L., Blom, H., and Zhu, Y. (2013). A Silurian placoderm with osteichthyan-like marginal jaw bones. *Nature* **502**, 188–193.

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Biological samples</b>		
Bichir	This paper	N/A
American paddlefish	This paper	N/A
Bowfin	This paper	N/A
Alligator gar	This paper	N/A
<b>Critical commercial assays</b>		
EZNA Tissue DNA Kit	Omega Bio-Tek	N/A
Safranin O/Fast Green staining kit	Servicebio, China	N/A
TRIzol reagent	Life Technologies	N/A
<b>Deposited data</b>		
The sequence data and assemblies	This paper	NCBI BioProject database with accession number PRJNA599026. China National GeneBank with accession number CNP0000841
<b>Software and algorithms</b>		
WTDBG2 v 2.3	Ruan and Li, 2020	<a href="https://github.com/ruanjue/wtdbg2">https://github.com/ruanjue/wtdbg2</a>
RACON v1.3.2	Vaser et al., 2017	<a href="https://github.com/isovic/racon">https://github.com/isovic/racon</a>
HiC-Pro v2.8.0	Servant et al., 2015	<a href="https://github.com/nservant/HiC-Pro">https://github.com/nservant/HiC-Pro</a>
3D-DNA v170123	Dudchenko et al., 2017	<a href="https://github.com/aidenlab/3d-dna">https://github.com/aidenlab/3d-dna</a>
SOAPdenovo v2.0	Luo et al., 2012	<a href="https://github.com/aquaskyline/SOAPdenovo2">https://github.com/aquaskyline/SOAPdenovo2</a>
BUSCO v3.0.2	Simão et al., 2015	<a href="https://busco.ezlab.org">https://busco.ezlab.org</a>
Genewise v 2.4.1	Birney et al., 2004	<a href="https://www.ebi.ac.uk/~birney/wise2/">https://www.ebi.ac.uk/~birney/wise2/</a>
AUGUSTUS v3.2.1	Stanke et al., 2008	<a href="http://bioinf.uni-greifswald.de/augustus/">http://bioinf.uni-greifswald.de/augustus/</a>
GlimmerHMM v3.0.4	Majoros et al., 2004	<a href="https://ccb.jhu.edu/software/glimmerhmm/">https://ccb.jhu.edu/software/glimmerhmm/</a>
SNAP v2013-11-29	Korf, 2004	<a href="https://github.com/KorfLab/SNAP">https://github.com/KorfLab/SNAP</a>
Tophat v2.1.1	Trapnell et al., 2009	<a href="https://github.com/DaehwanKimLab/tophat">https://github.com/DaehwanKimLab/tophat</a>
Cufflinks v2.2.1	Trapnell et al., 2010	<a href="https://github.com/cole-trapnell-lab/cufflinks">https://github.com/cole-trapnell-lab/cufflinks</a>
MAFFT, v7.310	Katoh, 2013	<a href="https://mafft.cbrc.jp/alignment/software/linuxportable.html">https://mafft.cbrc.jp/alignment/software/linuxportable.html</a>
Clustal Omega, v1.2.4	Sievers and Higgins, 2014	<a href="http://www.clustal.org/omega/">http://www.clustal.org/omega/</a>
ASTRAL v5.5.6	Mirarab et al., 2014	<a href="https://github.com/smirarab/ASTRAL">https://github.com/smirarab/ASTRAL</a>
RaxML v8.2.11	Stamatakis, 2014	<a href="https://github.com/stamatak/standard-RAxML">https://github.com/stamatak/standard-RAxML</a>
MCMCtree v4.5	Yang, 2007	<a href="http://abacus.gene.ucl.ac.uk/software/paml.html#download">http://abacus.gene.ucl.ac.uk/software/paml.html#download</a>
LastZ v 1.04.00	Harris, 2007	<a href="https://github.com/lastz/lastz">https://github.com/lastz/lastz</a>
OrthoMCL v14-137	Li et al., 2003	<a href="https://orthomcl.org/orthomcl/">https://orthomcl.org/orthomcl/</a>
MUSCLE v3.8.31	Edgar, 2004	N/A
SOAPnuke v1.5.6	Chen et al., 2018	<a href="https://github.com/BGI-flexlab/SOAPnuke">https://github.com/BGI-flexlab/SOAPnuke</a>
VISTA v1.4.26	Frazer et al., 2004	<a href="http://genome.lbl.gov/vista/index.shtml">http://genome.lbl.gov/vista/index.shtml</a>
MCScaN v1.1	Wang et al., 2012	<a href="https://github.com/wyp1125/MCScaN">https://github.com/wyp1125/MCScaN</a>
ANGES v1.01	Chauve and Tannier, 2008	<a href="https://github.com/cchauve/ANGES">https://github.com/cchauve/ANGES</a>
pal2nal v14	Suyama et al., 2006	<a href="https://github.com/drostlab/orthologr/tree/master/inst/pal2nal/pal2nal.v14">https://github.com/drostlab/orthologr/tree/master/inst/pal2nal/pal2nal.v14</a>
HiCEXplorer v3.5	Wolff et al., 2018	<a href="https://github.com/deeptools/HiCEXplorer">https://github.com/deeptools/HiCEXplorer</a>
HISAT2 v2.0.4	Kim et al., 2015	<a href="https://daehwankimlab.github.io/hisat2/">https://daehwankimlab.github.io/hisat2/</a>

## RESOURCE AVAILABILITY

### Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Guojie Zhang ([guojie.zhang@bio.ku.dk](mailto:guojie.zhang@bio.ku.dk)).

### Materials availability

This study did not generate new unique reagents.

### Data and code availability

All the assemblies and raw sequencing data generated during this study are available in NCBI-SRA under the BioProject: PRJNA599026.

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

### Source organism

Bichir (*Polypterus senegalus*), bowfin (*Amia calva*) and alligator gar (*Atractosteus spatula*) specimens were bought from an ornamental fish market in Guangzhou, China. The paddlefish (*Polyodon spathula*) was provided by Yangtze River Fisheries Research Institute, Chinese Academy of Fisheries Sciences, China. The four species were taxonomically identified by both morphological observations and molecular methods of mitochondrial cytochrome C oxidase subunit 1 (mitochondrial loci, COI) and collected at about 2 years old. Genomic DNA was extracted from muscle tissue of the specimens using an EZNA Tissue DNA Kit (Omega Bio-Tek) following the manufacturer's instructions. The collection and sampling protocols were approved by the Institute of Hydrobiology, strictly following its Animal care and ethics policy. All animal experiments were approved by the Animal Research and Ethics Committee of the Institute of Hydrobiology, Chinese Academy of Sciences. For transcriptome and *in situ* hybridization, samples of bichir, paddlefish, alligator gar, African lungfish (*Protopterus annectens*), African clawed frog (*Xenopus laevis*) and snakehead (*Channa argus*) were collected at about 2 years old. Mouse (*Mus musculus*, strain: C57BL/6) and zebrafish (*Danio rerio*, strain: AB) were collected at three and ten months, respectively. We didn't identify the gender for these samples. Knockout of the *Hand2* CNE was performed in *Mus musculus* C57BL/6 strain mice. The following developmental ages were used in this study: embryonic day E16.5, E17.5 and E18.5 mice; 3 month old mice. Animals of both sexes were used in the analysis.

## METHOD DETAILS

### Genome sequencing, assembly and annotation

Genomic DNA of *P. senegalus* was extracted from muscle tissue for nanopore sequencing using a QIAGEN DNA purification kit. For each Nanopore library, gDNA was size-selected (> 20 kb) with a Blue Pippin (Sage Science, Beverly, MA) and processed using a Ligation sequencing 1D kit (SQK-LSK108, ONT, UK) according to the manufacturer's instructions. Fifteen libraries were constructed and sequenced in 15 flow cells using a GridION X5 sequencer (ONT, UK) for 48 h each by Nextomics Biosciences (Wuhan, China). Base calling was performed on fast5 files using ONT Albacore software v0.8.4, and "passed filter" reads with 22,519 bp average length (233.3 Gb of clean data, providing approximately 64-fold genome sequencing coverage) were subsequently used for genome assembly. We used WTDBG2 (Ruan and Li, 2020) (v 2.3) with default parameters for bichir genome *de novo* assembly. Subsequently, RACON (Vaser et al., 2017) (v1.3.2) with the nanopore reads as input was used for consensus calling for this assembly, then Illumina reads were used as input to the WTDBG-CNS tool for further polishing. To generate a chromosome-scale genome, we constructed two Hi-C libraries and obtained a total of approximately 436 Gb raw data. After HiC-Pro filtering (Servant et al., 2015) (v 2.8.0), the remaining valid base had 105 Gb data with a genome depth of about 28X. The 3D-DNA (Dudchenko et al., 2017) pipeline was used to create contact maps for the final genome assembly, yielding 18 long scaffolds corresponding to the 18 chromosomes in bichir, containing about 94.3% of the total sequences. The remaining 9,838 small scaffolds (maximum length, 624 kb) account for the remaining 5.7% of the genome. Genomic DNA of American paddlefish, bowfin and alligator gar was also extracted from muscle tissue and insert DNA libraries were constructed (with sizes ranging from approximately 250 bp to 40 kbp). All of the raw reads were generated using Illumina HiSeq 2000 and 4000 platforms with coverage depths of 245X, 208X, and 135X, respectively, and SOAPdenovo (Luo et al., 2012) (v2.0) was used to assemble the reads into scaffolds (Table S1). Vertebrate sets of Benchmarking Universal Single-Copy Orthologs are identified using BUSCO (Simão et al., 2015) (v3.0.2).

We integrated the results from homology analysis, *de novo* predictions and transcript analysis to annotate protein-coding genes of the four genomes. Whole protein sequences from *Danio rerio*, *Takifugu rubripes*, *Gasterosteus aculeatus*, *Gadus morhua*, *Oryzias latipes*, *Lepisosteus oculatus*, *Callorhinchus milii*, *Homo sapiens*, *Latimeria chalumnae*, and *Petromyzon marinus* were used to perform homologous gene structure predictions. The homology-based pipeline started with homology searching against a non-redundant collection of protein sequences using TBLASTN with an E-value cutoff of 1E-5, followed by selection of the most similar proteins for each region with homologous protein matching. Regions with homologous blocks shorter than 25% of query proteins were then excluded, and Genewise (Birney et al., 2004) (v2.4.1) was used to generate gene structures based on the homology

alignments. AUGUSTUS (Stanke et al., 2008) (v3.2.1), SNAP (Korf, 2004) (v2013-11-29) and GlimmerHMM (Majoros et al., 2004) (v3.0.4) were chosen for *de novo* predictions. Predicted gene models obtained by the two methods were integrated by GLEAN. RNA-seq reads were mapped to the genome by Tophat (Trapnell et al., 2009) (v2.1.1), then Cufflinks (Trapnell et al., 2010) (v2.2.1) was used to assemble transcripts, and the assembled transcripts were used to predict ORFs. Transcript-based gene models with intact ORFs that had no overlap with the GLEAN gene sets were added, and if a transcript-based gene model with an intact ORF covered more than one GLEAN gene, we replaced the GLEAN gene with the transcript-derived gene. Transcripts without intact ORFs were used to extend the incomplete GLEAN gene models to find start and stop codons. Functions were assigned to the genes based on the best alignments to entries in the Swiss-Prot database (v15.10) using BLASTP. Then we searched InterPro databases (Apweiler et al., 2001) (v29.0) including the Pfam, PRINTS, PROSITE, ProDom, Superfamily and SMART databases to identify motifs and domains of the genes. GO terms for each gene were obtained from the corresponding InterPro entries. All genes were aligned against the KEGG database using the KEGG Automatic Annotation Server to identify pathways in which the genes might be involved. We annotated about 20,000 protein-coding genes, most of which could be aligned to entries in Swiss-Prot, InterPro or KEGG databases (Table S1), and thus assigned putative functions.

We used known repeat element libraries in the Repbase database (v21.01) to predict repetitive elements in the four sequenced species and eight other species (spotted gar, medaka, fugu, stickleback, cod, eel, coelacanth, elephant shark) by Proteinmask, Repeatmasker and TRF. We also constructed a *de novo* repeat library with LTR finder and Repeatmodeler to identify repeat elements (Table S1). Repetitive elements of human and zebrafish were downloaded from the UCSC website (<https://genome.ucsc.edu/>).

### Phylogeny reconstruction

We used information from the spotted gar chromosomes corresponding to the 17 sets of ancestral vertebrate chromosomes before the two rounds of WGDs that occurred at the vertebrate base (Sacerdot et al., 2018) to produce orthologous groups composed of 27 living species of chordates (amphioxus, lamprey, hagfish, elephant shark, brownbanded bamboo shark, whale shark, cloudy catshark, white shark, spotted gar, alligator gar, bowfin, American paddlefish, sterlet, bichir, fugu, stickleback, medaka, cod, zebrafish, European eel, human, coelacanth, lungfish (Wang et al., 2021), mouse, chicken, lizard and Western clawed frog). We first used spotted gar as a reference to find orthologs with other species using the reciprocal best hit (RBH) method. The 17 sets of ancestral vertebrate chromosomes underwent two earlier rounds of WGD events, leading to up to four copies of genes in each descendent species, so we selected one copy with the largest number of orthologous genes from the four copies to find orthologs (Figure S1C). Finally, we obtained 359 one-to-one orthologs in all species, allowing up to one missing species. The protein sequences of orthologous genes were aligned using Clustal Omega (Sievers and Higgins, 2014) (v1.2.4) with default parameters. A species tree was reconstructed using ASTRAL (Mirarab et al., 2014) after inferring a set of gene trees by RaxML (Stamatakis, 2014) (v8.2.11) with the PROTGAMMAJTTF model and used to estimate divergence times. The correlated molecular clock and GTR substitution model were implemented in MCMCTree (Yang and Rannala, 2006) to estimate divergence times between species. The concatenated all codon sites of orthologous genes were used as inputs. Nine calibration time points based on fossil records were used as constraints in the MCMCTree estimation (Table S1). The MCMC process was run for 10,000,000 steps and sampled every 500 steps.

We used the bichir genome as reference, and aligned it with genomes of 13 other species (elephant shark, coelacanth, human, American paddlefish, bowfin, alligator gar, spotted gar, European eel, zebrafish, cod, medaka, stickleback and fugu) using LastZ (Harris, 2007) with the following parameters: H = 2000, Y = 9400 (3400 for 'distant alignments'), L = 3000 (6000), K = 3000 (2200). 'Distant alignments' were generated between ray-finned fishes and sarcopterygians. All genomes were masked softly before alignment. The resulting LastZ maf files were combined into a single multiple genome alignment by MultiZ, which we then obtained a total of 23,307,703 bp alignment across all species. About 70% (17,039,649) of the alignment included at least 13 species and 31% (7,188,055) included all 14 species. We used alignment blocks including at least 12 species to construct phylogenetic trees by Raxml using GTRGAMMA and GTRCAT modes separately with 100 bootstrap replicates. Results obtained with these two modes generated the same topology with high bootstrap support and similar branch lengths. The GTRGAMMA results were used for subsequent estimation of divergence times. Ranges of divergence time in each inner tree node were collected from the TimeTree website (<http://www.timetree.org/>) and used to calibrate the divergence time estimator by MCMCTree (Figure S1D).

We also used OrthoMCL (Li et al., 2003) (v14-137) to identify 319 single copy gene families among 19 vertebrates. Protein sequences of each single copy gene family were aligned, then we connected all single copy gene family alignments into a super alignment and constructed a phylogenetic tree using Raxml and estimated divergence times with MCMCTree (Figure S1E).

### Reconstruction of ancestral karyotypes

An Ancestral Actinopterygians Karyotype (AAK), Ancestral Teleosts Karyotype (ATK), and Ancestral Sarcopterygians Karyotype (ASK) were reconstructed using six chromosome-scale genome assemblies (Chicken was used as an outgroup and AAK was constructed using chicken and spotted gar; Spotted gar was used as an outgroup and ATK was constructed using spotted gar, zebrafish and medaka; Bichir was used as an outgroup and ASK was constructed using bichir, human and chicken). The ancestral genome reconstruction procedure is illustrated in Figure S2A. Briefly, we first used the RBH method with chicken and bichir as references to find orthologous genes in the reconstruction of AAK and ASK. Due to the effect of teleost genome duplication, we used orthologous genes identified in a previous study (Braasch et al., 2016) to reconstruct the ATK. We then used the chi-square test followed by False Discovery Rate (FDR) correction to identify homologous chromosomes between pairs of species, with an orthologous

gene number over 30 and FDR  $q$  value less than 0.05 as cutoffs. Next, we used MCScanX (Wang et al., 2012) (v1.1) with default parameters to find conserved synteny blocks between homologous chromosomes. Finally, ANGES (Chauve and Tannier, 2008) (v1.01) was applied for ancestral chromosome reconstruction. Synteny blocks between bichir and sterlet were also found using this method, and the karyotype region on the sterlet genome was deduced from the ancestral karyotype region of bichir. Circos plots of bichir-gar-chicken and gar-zebrafish-medaka comparisons were generated, based on the one-to-one orthology relations (Figure S2B).

We also used the bichir and spotted gar genomes as references to analyze the evolution of the ancestral karyotype of the ray-finned fishes in teleosts and lobe-finned fishes. Chromosome regions found to be conserved in all basal ray-finned fishes and lobe-finned fishes, but changed in some or all of the teleosts due to rearrangements and rediploidization, were defined as specific regions for the basal ray-finned fishes and lobe-finned fishes. Regions specific to the basal ray-finned fishes and teleosts were identified in the same manner. Gene ontology enrichments for specific regions of the basal ray-finned fishes and lobe-finned fishes are shown in Figure S6A.

### Synteny analysis

We extended the synteny analysis to 14 vertebrate species: elephant shark, human, mouse, chicken, Western clawed frog, coelacanth, bichir, bowfin, alligator gar, spotted gar, zebrafish, medaka, stickleback and fugu. We used BLASTP to identify orthologous genes in pairs of these species. Synteny blocks between each pair of species were searched by seeking seeds with at least five consecutive orthologous genes between the two species, then extending the seeds in both directions with gaps of no more than four genes. After that, the proportion synteny between pairs of species were calculated (Table S2).

### Whole genome duplication events analysis

To identify Hox clusters we used homolog-based methodology to obtain the evolutionary history of Hox genes in jawed vertebrates. Protein sequences of human and zebrafish Hox genes were used as homologous sequences. We first used TBLASTN to search for candidate gene regions, then predicted gene structures using Genewise (Birney et al., 2004) (v2.4.1). Finally, the annotation results were aligned back to the corresponding human and zebrafish Hox genes. If an alignment length accounted for more than half of the length of a homologous protein, it was considered to be an intact gene, and otherwise a partial gene. If there was a premature stop codon or frameshift mutation in the predicted gene structure, it was regarded as a pseudogene (Figure S2C).

To calculate the time of the 3R WGD for paddlefish and sterlet, we identified 3,743 paralogs in paddlefish and 8,297 paralogs in sterlet. Paralogous genes were identified by all-to-all protein alignment using BLAST v2.7.1, followed by clustering using MCL (Li et al., 2003) (v14-137). Multiple sequence alignment was performed between pairs of genes in each gene family using MUSCLE (Edgar, 2004) (v3.8.31), then pairs with alignment identity  $> 0.5$  and alignment rate  $> 0.6$  were regarded as paralogs. We also identified 10,729 orthologs between paddlefish and sterlet by the RBH method. Protein sequences from these paralogs and orthologs were aligned by Clustal Omega (Sievers and Higgins, 2014) (v1.2.4) and converted to CDS alignments by PAL2NAL (Suyama et al., 2006) (v14). 4-fold degenerative third-codon transversion (4dTv) values were calculated and corrected with the HKY model in the PAML package (Yang, 2007) (Figure S2D). Based on 4dTv distributions, the 3R WGDs in sterlet and paddlefish were estimated to have occurred about 42.2 Mya ( $[135 \text{ Mya}/0.048] * 0.015$ ) and 112.5 Mya ( $[135 \text{ Mya}/0.048] * 0.04$ ). The divergency time between paddlefish and sterlet at around 135 Mya was obtained from this article and illustrated in Figure 1.

Synonymous substitution rates (Ks values) were also calculated using the maximum likelihood method implemented in codeml of the PAML package under the F3x4 model (Goldman and Yang, 1994). Based on the Ks values of paddlefish-sterlet orthologs (with a median value of 0.1509) and estimated divergence time of the two species (135 Mya), we obtained an estimated number of substitutions per synonymous site per year ( $r$ ) of  $5.59\text{E}-10$  (divergence date =  $Ks/2r$ ). The same  $r$  value was applied to calculate dates of 3R WGD events based on median Ks values of their paralogs (sterlet, 0.0574; paddlefish, 0.1349). We also identified 1,935 genes of which one copy has been retained in spotted gar, but two copies are still present in sterlet and paddlefish. Construction of gene trees by the maximum likelihood method for each group showed that 34.3% of the gene trees supported independent 3R WGD events in sterlet and paddlefish, while only 16.6% supported the ancestral duplication hypothesis (Figure S2E).

### Limb enhancer identification

The transition from fin to limb is a crucial adaptation for terrestrialization. Thus, we downloaded 174 limb-related enhancer sequences in mice from VISTA Enhancer Browser (<https://enhancer.lbl.gov/>) and annotated corresponding enhancers in target genomes (amphioxus, lamprey, hagfish, elephant shark, brownbanded bamboo shark, whale shark, cloudy catshark, white shark, spotted gar, alligator gar, bowfin, American paddlefish, sterlet, bichir, fugu, stickleback, medaka, cod, zebrafish, European eel, human, coelacanth, sperm whale, chicken, lizard and Western clawed frog). However, some of these enhancers are not only related to limbs, but also activated in other tissues (such as the brain, heart, etc.). Here, we restricted our comparison on tissue-specific enhancers that are specifically functional in limbs to make sure the presence and absence patterns are only related to the limb functions (40/174; list in Table S3). The annotation method can be summarized as follows. First, we searched for limb regulatory elements in target genomes with BLASTN using VISTA enhancer sequences as queries. Then, to obtain candidate homologous hits of these enhancers, we selected the best hit from each alignment result, and extended it to both flanking regions until it was consistent with the length of the enhancer on the mouse genome. Finally, we aligned the selected genomic regions (blast best hit region plus flanking sequences) back to the mouse genome using BLASTN. If the alignment position of the candidate region overlapped with more than

100 nucleotides in the genomic region of the mouse limb enhancer, we considered the enhancer present in the target genome. The mm703 enhancer and *Osr2* are in the same topologically associating domain, which is supported by mice HiC data from PENNSTATE website (<http://3dgenome.fsm.northwestern.edu/view.php>, assembly: mm10, tissue: neuron, type: Jiang\_2017-raw, resolution: 40 kb) (Jiang et al., 2017). Bichir TADs are obtained using hicFindTADs tool in HiCExplorer program (Wolff et al., 2018) (v3.5) (Figure S3A).

### Fin regeneration and *Osr2* in situ expression

The animal experiment was approved by the Animal Care and Use Committee of Institute of Hydrobiology, Chinese Academy of Sciences (Approval ID: Y21304501). Adult bichirs were commercially obtained and maintained in glass tanks at 26°C, and anesthetized in 0.1% tricaine mesylate (MS-222) before sampling. The bichirs were then washed with pure water and dried with filter paper to clean their pectoral fins, which were amputated with small surgical scissors. Stumps were collected at 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60 and 70 days after amputation and stored in 4% paraformaldehyde solution for subsequent histological studies.

The above samples of pectoral fin regeneration were decalcified and the safranin O-fast green staining was performed by using the Safranin O/Fast Green staining kit (Servicebio, China). First, the slides were subjected to two 20-min incubations in xylene, two 5-min incubations in ethanol and one in 75% ethanol then washed with distilled water. The samples were then dyed with Fast Green for 1–5 min and washed with distilled water until the cartilage was transparent. Next, the samples were incubated in 1% hydrochloric acid ethanol for 10 s and washed with distilled water. The samples were subsequently stained with Safranin O for 1–5 s, then rapidly dehydrated with rinses (5, 2, and 10 s) in ethanol. The slides were then prepared for primary microscopic examination by treating with xylene for 5 min and sealing with neutral balsam. Finally, the Safranin O/Fast Green stained slides were photographed.

Tissues of regenerating pectoral fin were dehydrated in an ethanol gradient, dipped in paraffin then embedded in paraffin wax. Each resulting paraffin block was sectioned (5 μm), and mounted on slides. The slides were passed through xylene, ethanol and DEPC water sequences, then boiled in repairing liquid for 5 min and cooled naturally. The samples on the slides were subsequently circled by Pap Pen then digested with 20 μg/mL proteinase K at 37°C for 20 min. After washing with purified water, the slides were subjected to three 5-min rinses with PBS. To block endogenous peroxidase, 3% H<sub>2</sub>O<sub>2</sub> was added (in a methanolic stock solution) to each sample, which was next incubated for 15 min at room temperature in the dark then subjected to three 5-min washes with PBS. The samples were subsequently incubated in a prehybridization solution at 37°C for 1 h then in a hybridization solution containing a digoxigenin (DIG)-labeled RNA probe at 37°C for 12–16 h. The hybridized samples were then washed in a solution of 2 × SSC at 37°C for 10 min, 1 × SSC at 37°C for 5 min twice, and 0.5 × SSC at room temperature for 10 min. After blocking with BSA for 30 min at room temperature, the slides were incubated with anti-DIG-AP antibody (Jackson) at 37°C for 50 min then subjected to three 5-min washes with PBS. The FISH hybridization signals were visualized by staining with anti-DIG antibody-conjugated peroxidase and tyramide signal amplification (FITC-TSA)/ (DAPI). Images of the treated tissues were acquired with a fluorescence microscope (NIKON Eclipse ci, JAPAN) (Figure S3B).

### Identification of OR and VR genes

To identify members of the olfactory receptor (OR) gene family, we first searched for them in bichir, American paddlefish, sterlet, bowfin, alligator gar, spotted gar, coelacanth, lungfish, elephant shark, brownbanded bamboo shark, whale shark, cloudy catshark and white shark genome sequences using TBLASTN with known amino acid sequences from 20 species (amphioxus, chicken, chimpanzee, cow, dog, elephant shark, fugu, human, lizard, macaque, medaka, mouse, opossum, platypus, rat, sea lamprey, spotted green pufferfish, stickleback, Western clawed frog, zebrafish) with E-values ≤ 10. Sequences were obtained from a previous study (Matsui et al., 2010; Niimura, 2009). Non-overlapping blast hits were then extracted, and used with 20 kbp extensions in both directions as Genewise (Birney et al., 2004) (v2.4.1) inputs to predict gene structures of olfactory receptors. After extracting the protein sequences of the OR genes obtained from the homologous annotation, we double-checked the completeness of these proteins by BLASTP searches against the known database of OR genes. Only proteins with alignment longer than 50% of the queries and identity > 40% were picked. Finally, all OR genes encoding proteins at least 250 amino acids long were retained. We used the ‘einsi’ method in MAFFT (Katoh, 2013) (v7.310) to generate multiple alignments and neighbor joining (NJ) trees were constructed. Six non-OR G protein coupled receptors were also included to serve as outgroups. OR genes were subsequently classified as intact genes and pseudogenes. Sequences were regarded as “pseudogenes” if they had at least one premature stop codon and/or frameshift, while the others were considered as intact genes (Table S4 and Figure S4A).

We also searched for vomeronasal receptor (VR) genes in 25 chordates using Genewise with amino acid sequences of all human, mouse and zebrafish VR proteins as queries. We then aligned all homologous annotations back to all human, mouse and zebrafish genes by BLAST (v2.7.1). When the best alignment results were consistent with the homologous protein name used for annotation, and the amino acid length accounted for more than 30% of the homologous protein, the gene was considered to be a VR gene (Figure S4B).

### Transcriptome analysis

We collected RNA samples from 10 tissues (gill, lung, swim bladder, brain, heart, liver, spinal cord, muscle, jaw and skin) of eight species, including three sarcopterygians (mouse, lungfish and African clawed frog), three basal ray-finned fishes (bichir, American paddlefish and alligator gar) and two teleosts (snakehead and zebrafish), with at least five replicates for each tissue. In total, we collected 543 samples for RNA-seq. We constructed RNA-seq libraries with insert sizes of around 200–400 bp. We used 100 bp

paired end reads for transcriptome sequencing on the BGISEQ-500 platform and produced about 8.7 Gb data per sample. Low-quality reads and reads with adaptors and unknown bases were filtered using SOAPnuke (Chen et al., 2018) (v1.5.6).

Protein-coding genes for every other species were sought with BLASTP using mouse protein sequences as queries and selecting RBH pairs as orthologous genes. In total, we identified 5,046 1:1 orthologs among the eight species. After mapping clean reads to respective genomes using HISAT2 (Kim et al., 2015) (v2.0.4), gene expression levels were calculated in terms of RPKM (reads per kilobase of transcript per million mapped reads) for each gene in 543 samples. Pearson correlation coefficients were used to evaluate transcriptome similarity between different samples. For samples of the same tissue from the same species, five samples with the highest mutual correlation coefficient ( $> 0.7$ ) were chosen for the following transcriptional profile analysis. We selected 355 samples in total for downstream analyses with five biological replicates for each tissue from each species (Table S5). After combining all 355 samples, RPKM values were transformed to log<sub>2</sub> scale and one was added (log<sub>2</sub>(RPKM+1)). The data were subjected to quantile normalization using the R package preprocessCore, followed by Principal Component Analysis (Figure 5A). The neighbor-joining (NJ) tree for all species based on the expression matrix and anatomical structure of the lungs and swim bladder is shown in Figures S5A and S5C. Differentially expressed genes that were significantly upregulated or downregulated between the two groups were identified by the Kolmogorov-Smirnov test. The R function ks.test was used for this, with the threshold P value for rejecting the null hypothesis set to 0.01. GO term enrichments of 466 genes exhibiting higher expression levels in the air-breathing group are shown in Table S5.

To identify tissue-specific genes, we first calculated fold-differences in expression (RPKM) between the sample in which each gene was most strongly expressed and samples of the same tissue in each species. If three of the resulting ratios  $\geq 4$ , the sample in which the gene was most strongly expressed was excluded and RPKM values for that gene in the remaining samples were averaged, otherwise all the RPKM values for samples of that combination of tissue and species were averaged. A tissue specificity index ( $\tau$ ) was then calculated for each gene (Yanai et al., 2005), varying from 0 to 1, where 0 means uniformly expressed in all sampled tissues, and 1 indicates complete tissue specificity. Tissue-specific genes were identified that had  $\tau$  values  $\geq 0.8$ , were most strongly expressed in one of the 10 sampled tissues with RPKM  $\geq 1$ . In addition, if the RPKM for the tissue with the second highest expression level had to be  $\geq 5$ -fold higher than that of the tissue with the third highest expression level, it also regarded as tissue-specific genes. We finally identified 11 genes showing lung-specific expression pattern in at least three of the four species (mouse, African clawed frog, lungfish and bichir) (Figure 5E). A plot of *Tbx4* lung mesenchyme-specific enhancer (LME) expression in jawed vertebrates was generated by VISTA (Frazer et al., 2004) (v1.4.26) (Figure S5D).

### In situ hybridization of lung-specific genes

Lung from mouse and bichir, and swim bladder from zebrafish were sampled and fixed in 4% paraformaldehyde (PFA) at 4°C for 2–12 h. The tissues were then dehydrated in an ethanol gradient, dipped in paraffin and embedded in paraffin wax. Each resulting paraffin block was sectioned (5  $\mu$ m), and mounted on slides. The slides were then passed sequentially through xylene, ethanol and DEPC water, then boiled in repairing liquid for 5 min and cooled naturally. The samples on the slides were subsequently circled by Pap Pen then digested with 20  $\mu$ g/mL proteinase K at 37°C for 20 min. After a brief wash with purified water, the slides were subjected to three 5-min washes in PBS. To block endogenous peroxidase, 3% H<sub>2</sub>O<sub>2</sub> was added (in a methanolic stock solution) to each sample, which was next incubated for 15 min at room temperature in the dark then subjected to three 5-min washes with PBS. Subsequently, samples were incubated in prehybridization solution at 37°C for 1 h then in a hybridization solution containing a digoxigenin (DIG)-labeled RNA probe at 37°C for 12–16 h. The hybridized samples were then washed in a solution of 2  $\times$  SSC at 37°C for 10 min, 1  $\times$  SSC at 37°C for 5 min twice, and 0.5  $\times$  SSC at room temperature for 10 min. After blocking in BSA for 30 min at room temperature, the slides were incubated with anti-DIG-AP antibody (Jackson) at 37°C for 50 min then subjected to three 5-min washes in PBS. The FISH hybridization signals were visualized by staining with anti-DIG antibody-conjugated peroxidase and tyramide signal amplification (FITC-TSA)/ (DAPI). Images of the treated tissues were acquired with a fluorescence microscope (NIKON Eclipse ci, JAPAN) (Figure 5F).

### Knockout of the *Hand2* CNE

Mice with a genomic deletion of *Hand2* CNE (167 bp) (also called *Hand2* CNE<sup>-/-</sup> mice) were generated by CRISPR/Cas9-mediated genome editing (Cyagen Biosciences). Briefly, two single-guide RNAs (sgRNAs) flanking the mouse *Hand2* CNE were used to delete the element. gRNA and Cas9 mRNA were generated by *in vitro* transcription. An oligo donor with targeting sequence, flanked by homologous sequences on both sides was synthesized. The gRNA to mouse *Hand2* CNE, the donor vector, and Cas9 mRNA were co-injected into mouse zygotes. After confirmation of genetic deletion, the germline transmission of male pups was performed for two generations by mating with wild-type mice to generate F1 heterozygous mutants, and inter-cross heterozygous F1 mice to generate homozygous F2 mice. All mice we used had a C57BL/6 background, and embryos at stage E16.5 were generated from timed matings of heterozygous or homozygous mice. Sequences of all genotyping primers are provided in Table S6.

For histological observation, E16.5 embryonic hearts were fixed in 4% paraformaldehyde/PBS overnight at room temperature, dehydrated through a graded ethanol series (50%, 75%, 90%, 95%, 100%) and embedded in paraffin. Sections (7  $\mu$ m) were mounted onto Superfrost Plus microscope slides (Thermo scientific) then deparaffinized in xylene and rehydrated through a graded ethanol series (100%, 95%, 75%). Eosin and hematoxylin staining was performed according to standard procedures. The ratio of RV area was calculated as the RV chamber area divided by the whole ventricle area. The thickness of the RV body wall was calculated as

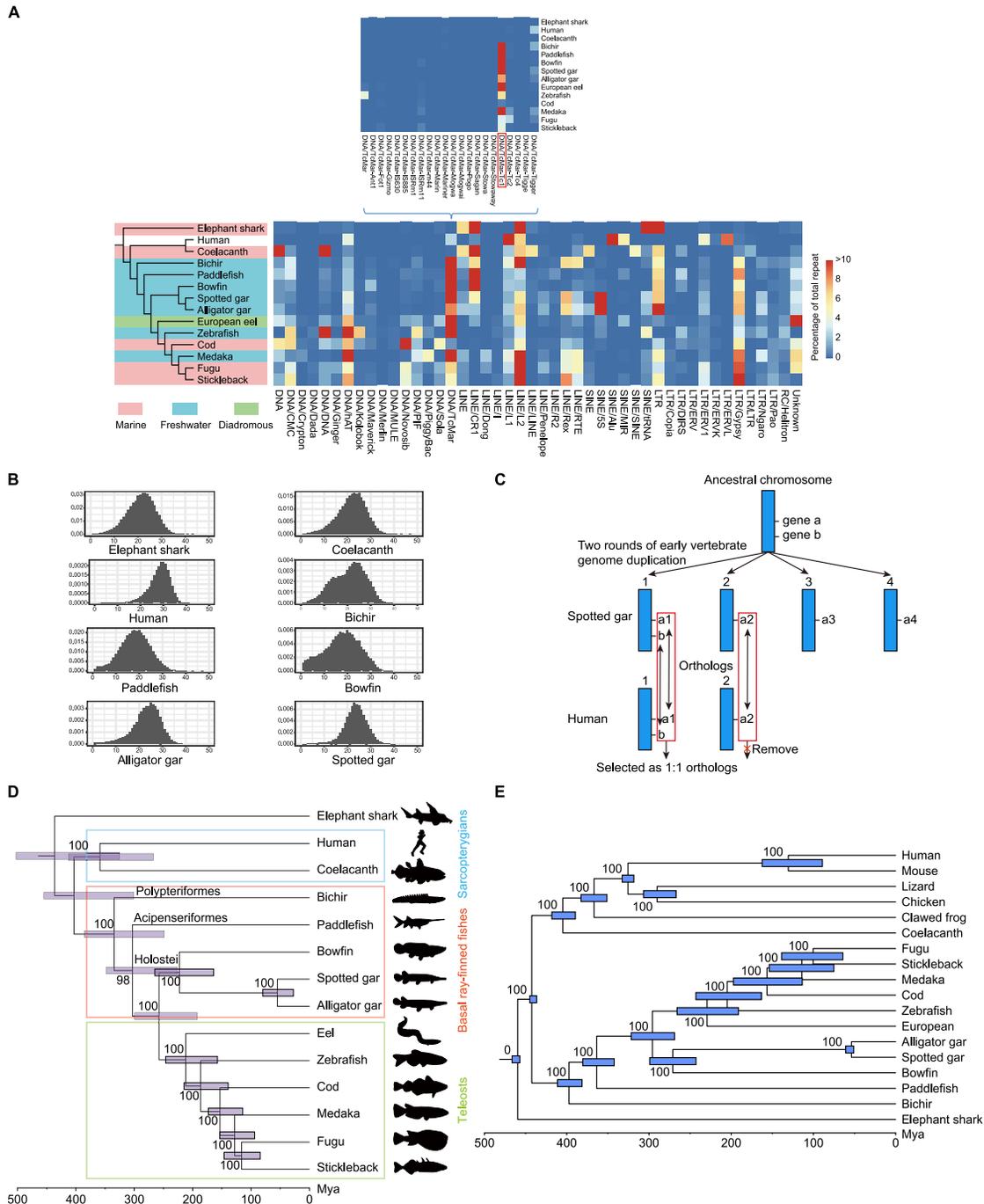
the ventricular compact myocardial area divided by its outer circumference. The relative thickness of the RV body wall was calculated as the thickness of the RV body wall divided by the whole ventricle circumference (Han et al., 2019). The variables were measured in CaseViewer after selecting image areas with myocardial color range (Figure S6D).

To compare *Hand2* expression in embryo heart between mutants and controls, mature four-chamber hearts of E16.5 embryos were dissected into distinct compartments in ice-cold PBS and harvested in TRIzol reagent (Life Technologies, #15596018). First-strand cDNA was synthesized from 500 ng of total RNA samples using M-MLV Reverse Transcriptase (Promega, Madison, WI, USA) and diluted 1:10 as an amplification template. Reverse transcription and quantitative PCR (qRT-PCR) was performed with 20  $\mu$ L mixtures using the Hieff® qPCR SYBR® Green Master Mix (Yeasen, China) and a LightCycler® 480 II Instrument (Roche, Switzerland). RNA levels were normalized to those of a housekeeping gene. Relative gene expression levels were determined using the comparative CT method (Livak and Schmittgen, 2001). The 18S gene was used as an internal reference and fold changes applied in the calculations are means obtained from analyses of four biological replicates. Primers are listed in Table S6. Results for qRT-PCR, relative thickness of RV body wall and ratio of RV area are shown as mean values with error bars representing the standard error (SEM). Statistical analysis was performed and P values were calculated using an unpaired two-tailed Student's t test and OriginPro software.

### QUANTIFICATION AND STATISTICAL ANALYSIS

All details of the applied statistics (e.g., for gene expression analysis in the 10 selected tissues) are provided together with details of the relevant analyses in the Methods Details section.

# Supplemental Figures



**Figure S1. Repeat annotation and phylogenetic tree construction, related to Figure 1**

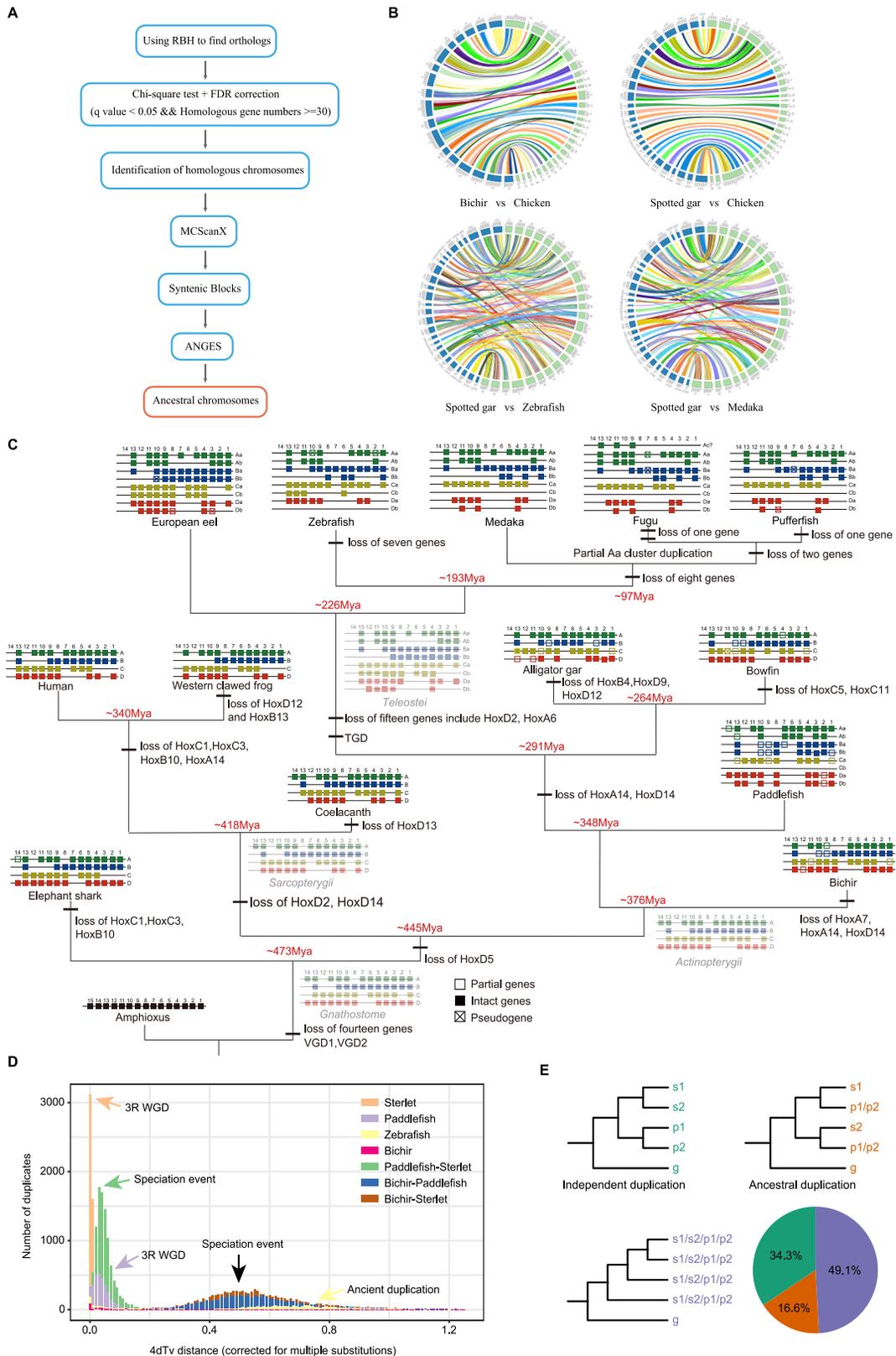
(A) Classification and distribution of repetitive elements in 14 species.

(B) The divergence profiles of LINE (CR1) in ray-finned fishes, lobe-finned fishes, and elephant shark, with the fraction of the genome shown on the y axis, and percent divergence on the x axis.

(C) Schematic diagram for selecting orthologous genes after two rounds of early vertebrate.

(D) Time tree based on whole genome alignments.

(E) Time tree based on 319 single copy genes.



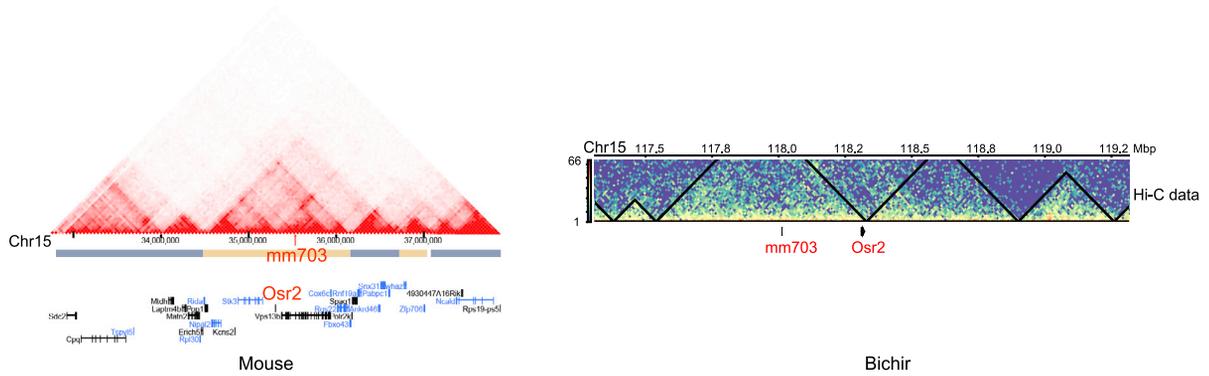
(legend on next page)

---

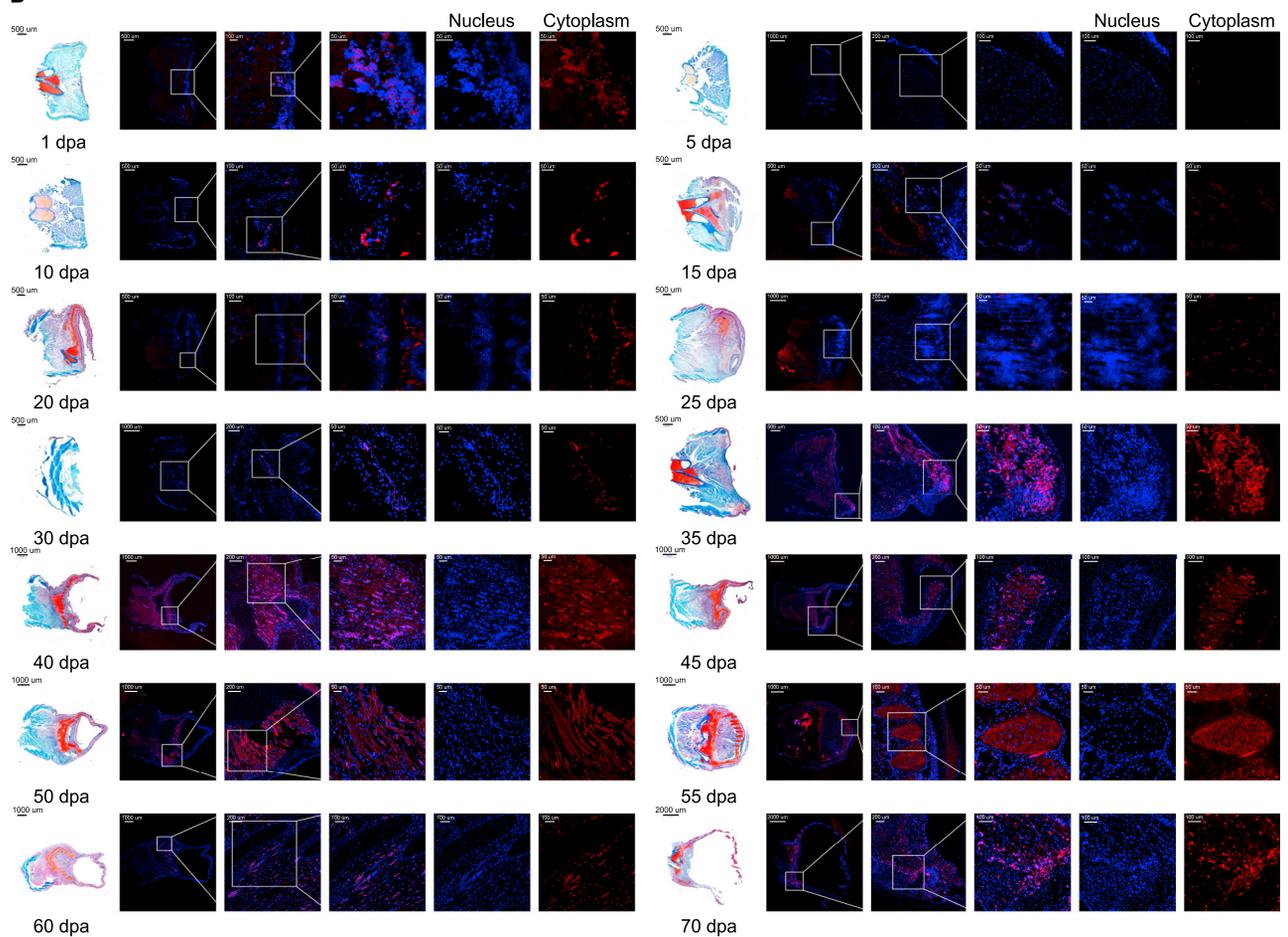
**Figure S2. The karyotype evolution and WGD analysis in vertebrates, related to Figure 2**

- (A) The ancestral genome reconstruction procedure.
- (B) Genome synteny of bichir-spotted gar-chicken and spotted gar-zebrafish-medaka.
- (C) Evolution of HOX clusters in vertebrates.
- (D) 4DTv comparison to determine 3R WGD time in paddlefish and sterlet.
- (E) Construction of 1935 gene trees among spotted gar (g), paddlefish (p) and sterlet (s). The green topology represents independent 3R WGD between paddlefish and sterlet. The orange topology represents ancestral 3R WGD. The purple topology cannot prove whether the 3R WGD occurred independently or not. The pie chart represents the proportion of genes for three topologies.

A



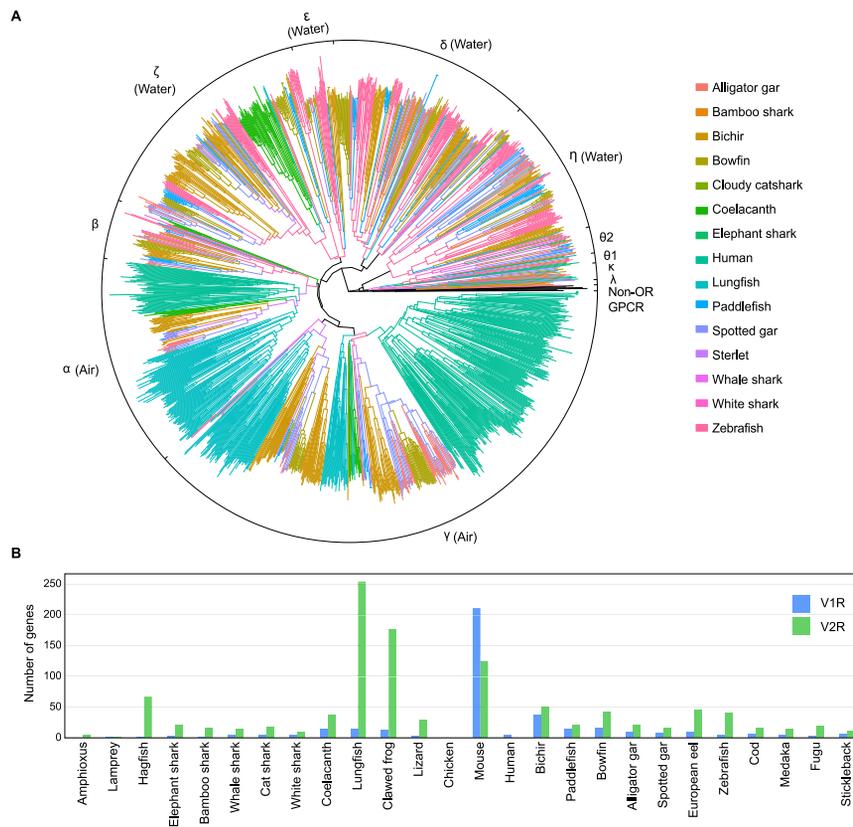
B



**Figure S3. mm703 is a potential regulatory sequence of Osr2, and pectoral fin regeneration in *P. senegalus*, related to Figure 3**

(A) The mm703 and Osr2 are in the same topologically associating domain supported by HiC data.

(B) FISH experiment of Osr2 during pectoral fin regeneration in *P. senegalus*.

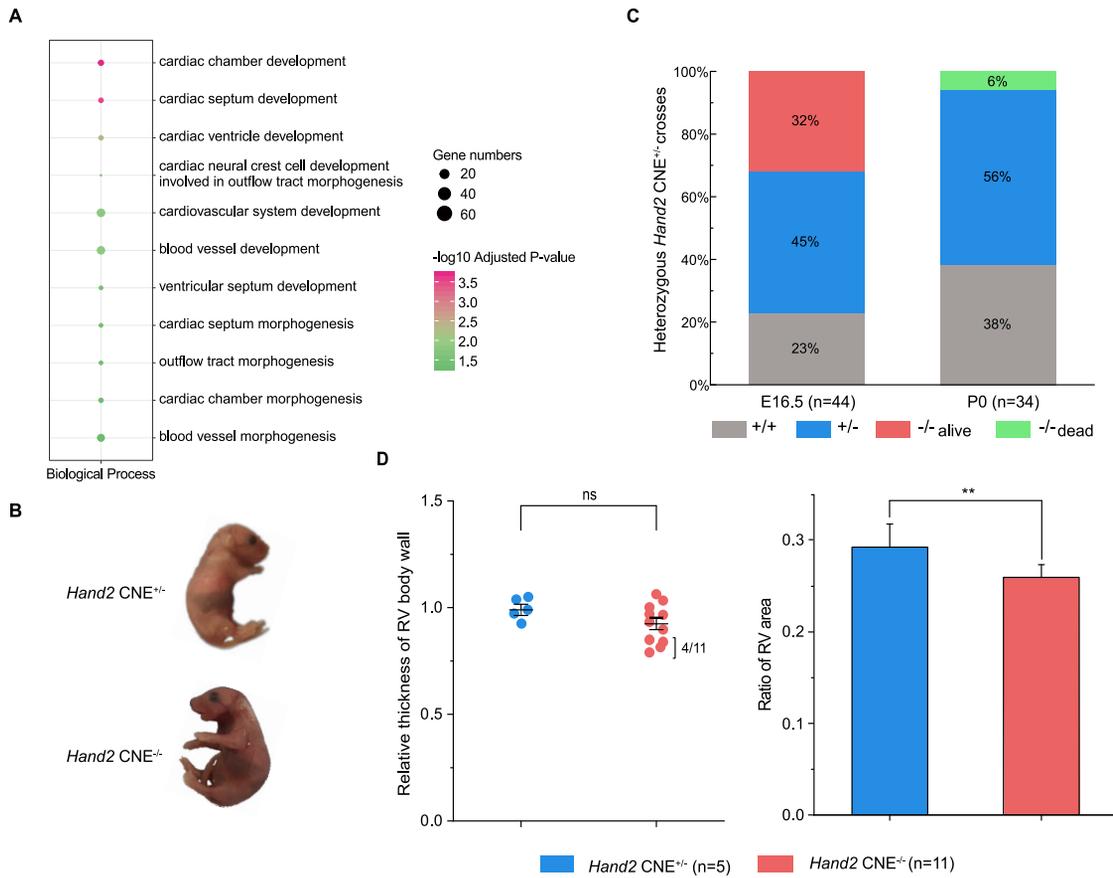


**Figure S4. Olfactory and vomeronasal receptor genes in chordates, related to Figure 4**

(A) NJ tree of annotated OR genes.

(B) Vomeronasal receptor gene repertoire in chordates.





**Figure S6. Deletion of the *Hand2* CNE causes congenital heart defects and perinatal lethality, related to Figure 6**

(A) Cardiac development-related GO terms.

(B) *Hand2* CNE<sup>+/-</sup> and *Hand2* CNE<sup>-/-</sup> newborns.

(C) Survival analysis of *Hand2* CNE<sup>+/-</sup> intercrosses.

(D) Morphometric analysis of the ratio of right ventricle area and relative thickness of right ventricle body wall.

Data are mean  $\pm$  SEM \*\*p < 0.01. ns (not significant).