Three amphioxus reference genomes reveal gene and chromosome evolution of chordates

Zhen Huang[1,*], Luohao Xu[2,*,†], Cheng Cai[3,*], Yitao Zhou[4,5,*], Jing Liu[2], Zexian Zhu[3], Wen Kang[3], Duo Chen[4,5], Surui Pei[6], Ting Xue[4,7], Wan Cen[4], Chenggang Shi[8], Xiaotong Wu[8], Yongji Huang[9], Chaohua Xu[1], Yanan Yan[1], Ying Yang[1], Wenjing He[1], Xuefeng Hu[4], Yanding Zhang[4], Youqiang Chen[4,7], Changwei Bi[10], Chunpeng He[10], Lingzhan Xue[11], Shijun Xiao[12], Zhicao Yue[13], Yu Jiang[6], Jr-Kai Yu[14,15], Erich D. Jarvis[16,17], Guang Li[8], Gang Lin[4,7,†], Qiujin Zhang[4,7,†], Qi Zhou[3,2,18,†]

1. Fujian Key Laboratory of Special Marine Bio-resources Sustainable Utilization, College of Life Sciences, Fujian Normal University, Fuzhou 350117, Fujian

2. Department of Neuroscience and Developmental Biology, University of Vienna, Vienna 1090, Vienna

3. The MOE Key Laboratory of Biosystems Homeostasis & Protection and Zhejiang Provincial Key Laboratory for Cancer Molecular Cell Biology, Life Sciences Institute, Zhejiang University, Hangzhou 310058, Zhejiang

4. The Public Service Platform for Industrialization Development Technology of Marine Biological Medicine and Product of State Oceanic Administration, College of Life Sciences, Fujian Normal University, Fuzhou 350117, Fujian

5. Fujian Key Laboratory of Developmental and Neurobiology, Fujian Normal University, Fuzhou 350117, Fujian

6. Annoroad Gene Technology Co., Ltd, Beijing

7. Center of Engineering Technology Research for Microalgae Germplasm Improvement of Fujian, Southern Institute of Oceanography, Fujian Normal University, Fuzhou 350117, Fujian

8. State Key Laboratory of Cellular Stress Biology, School of Life Sciences, Xiamen University, Xiamen, Fujian 361102, Fujian

9. Institute of Oceanography, Minjiang University, Fuzhou 350108, Fujian

10. State Key Laboratory of Bioelectronics, School of Biological Science and Medical Engineering, Southeast University, Nanjing 210096, Jiangsu

11. Aquaculture and Genetic breeding laboratory, Freshwater Fisheries Research Institute of Fujian, Fuzhou 350002, Fujian

12. College of Plant Protection, Jilin Agricultural University, Changchun, Jilin

13. Department of Cell Biology and Medical Genetics, Carson International Cancer Center, and Guangdong Key Laboratory for Genome Stability and Disease Prevention, Shenzhen University School of Medicine, Guangdong

14. Institute of Cellular and Organismic Biology, Academia Sinica, Taipei 11529, Taipei

15. Marine Research Station, Institute of Cellular and Organismic Biology, Academia Sinica, Yilan 26242, Taipei

16. Laboratory of Neurogenetics of Language, The Rockefeller University, New York 10065, New York

17. Howard Hughes Medical Institute, Chevy Chase, Maryland 20815, Maryland

18. Center for Reproductive Medicine, The 2nd Affiliated Hospital, School of Medicine, Hangzhou 310052, Zhejiang

[*] These authors contributed equally to the work

[†] Correspondence to: Luohao Xu: luohaox@gmail.com, Gang Lin: lgffz@fjnu.edu.cn, Qiujing Zhang: qjzhang@fjnu.edu.cn and Qi Zhou: zhouqi1982@zju.edu.cn

Abstract

The slow-evolving invertebrate amphioxus has an irreplaceable role in advancing our understanding into the vertebrate origin and innovations. Here we resolve the nearly complete chromosomal genomes of three amphioxus species, one of which best recapitulates the 17 chordate ancestor linkage groups. We reconstruct the fusions, retention or rearrangements between descendants of whole genome duplications (WGDs), which gave rise to the extant microchromosomes likely existed in the vertebrate ancestor. Similar to vertebrates, the amphioxus genome gradually establishes its 3D chromatin architecture at the onset of zygotic activation, and forms two topologically associated domains at the *Hox* gene cluster. We find that all three amphioxus species have ZW sex chromosomes with little sequence differentiation, and their putative sex-determining regions are nonhomologous to each other. Our results illuminate the unappreciated interspecific diversity and developmental dynamics of amphioxus genomes, and provide high-quality references for understanding the mechanisms of chordate functional genome evolution.

Introduction

Although first described in 1774, the lesser-known marine invertebrate amphioxus (or lancelets) only became recognized for its unparalleled value in elucidating the vertebrate origin and innovations until characterization of its *Hox* genes in 1990s (*1*). It was later established that amphioxus diverged from the ancestor of two other chordate subphyla, urochordates (tunicates) and vertebrates about 550 million years ago (MYA) (*2, 3*). Amphioxus has a vertebrate-like but simpler body plan, and underwent much less lineage-specific changes of chromosomes and genomic sequences than urochordates (*4*). Therefore, it represents the best known living proxy for the chordate ancestor (*5, 6*). Amphioxus has one, and the largest reported *Hox* gene cluster with 15 genes (*7*), which was found to form one structural and regulatory unit of topologically associated domain (TAD). By contrast, vertebrates have at least 4 *Hox* gene clusters and up to 13 genes per cluster, with the mouse *HoxA* and *HoxD* clusters each forming two TADs (*8*). Such a 4-fold difference of *Hox* gene cluster numbers provided early evidence for Ohno's hypothesis of two rounds of WGDs (the 2R hypothesis) (*9, 10*) that shaped the genome evolution and regulation of vertebrates since they diverged from other chordates.

Broader understanding beyond individual genes into the scenario and functional consequences of vertebrate WGDs, whose times and timing recently became a subject of debate (*11*), necessitate high-quality sequence assembly and annotation of genes and *cis*-regulatory elements of amphioxus (*12*), as a pre-WGD outgroup. The first draft genome of Florida amphioxus *Branchiostoma floridae* (Bf) was published over a decade ago, and has been frequently used to reconstruct the ancestral vertebrate protokaryotype, with however different estimates of ancestral linkage group number between studies (*11, 13-15*). A recent work improved the Bf genome into the chromosome-level and proposed a refined 2R hypothesis with 17 ancestral chordate linkage groups: the first WGD occurred in the ancestor of all vertebrates, and the second WGD only occurred in the lineage of jawed vertebrates (*16*). The duplicated gene products of WGDs in vertebrates ('ohnologues') seem to have generally a higher number of and more specialized regulatory elements and gene expression between copies, relative to their single-copy orthologs of amphioxus (*12*). Besides results at the gene-level, to address how vertebrates evolved globally more complex regulatory circuits after WGDs requires knowledge of higher-order chromatin organization of amphioxus.

An often-overlooked factor among previous studies using only one species' genome is amphioxus' largely unexplored interspecific genomic diversity. It is known that different amphioxus species have different chromosome numbers, and exhibit frequent disruptions of gene synteny which may confound the inference of vertebrate ancestral state (*17*). Moreover, the available amphioxus genome assemblies are either incomplete or fragmented because of the high intraspecific polymorphisms associated with their large effective population size (*4*). To elucidate the evolution of genes, genomes and chromatin landscapes of different amphioxus species compared to vertebrates, we resolved here the nearly complete haploid genomes of three *Branchiostoma* amphioxus species Chinese amphioxus (*B. belcheri*, Bb), Japanese amphioxus (*B. japonicum*, Bj) and Bf.

Results

Haploid chromosomal genomes of three amphioxus species

We estimated the genome-wide heterozygosity levels of three amphioxus species to range from 3.2% to 4.2%, among the highest in animal species (*18*) (**Supplementary Fig. S1**). To overcome this great challenge for genome assembly, we devised an interspecific trio-sequencing strategy and produced respectively more than 100-fold short and long sequencing reads for the F1 hybrids derived from Bf-Bb or Bf-Bj crosses (**Fig. 1a**, **Supplementary Fig. S2**). Given at least 50 MYs' species divergence time (**Supplementary Fig. S3**), the hybrids contain two haploid parental genomes that have become too diverged in sequences to form cross-species chimeric assembly (**Supplementary Fig. S1**). By mapping short-reads derived from the respective parental species, we were able to attribute each assembled contig into one of the four haploid (Bb, Bj and two Bf) genomes (**Fig. 1b-c**). The new haploid amphioxus genomes have an assembled size ranging from 382 to 491 Mb, and an over 200-fold improvement in contig N50 length (between 6.4 to 14.2Mb) compared to the published genomes(*4, 12, 17*), an over 97% genome completeness (measured by BUSCO) and a reduced level of false duplications (**Supplementary Table S1, Supplementary Fig. S4a-b**). Using Hi-C data, we anchored more than 98.6% of the contig sequences into chromosomes, with a much lower gap number (on average only 3.8 gaps) per chromosome than those of major vertebrate reference genomes and that of a recently improved Bf genome(*16*) (**Fig. 1d**, **Supplementary Fig. S4c**).

With some exceptions, all chromosomal sequences of the three species have been assembled from the telomere at one end to the centromere at the other (**Fig. 1e, Supplementary Fig. S5-6**). This is consistent with the reported predominantly telocentric karyotype of amphioxus (*19-21*), the low levels of recombination rate and nucleotide diversity at centromeric and pericentromeric regions (**Supplementary Fig. S7-8**); and is also verified by our fluorescent *in situ* hybridization (FISH) experiment for Bf (**Supplementary Fig. S9**). The telomeres contain conserved telomeric motifs $(TTAGGG)_n$ (*22*) with an average length of 3.6 kb, and they account for the majority of G-quadruplex content in the genome (**Supplementary Fig. S10**). Our cytogenetic and genomic investigations also confirmed the presence of interstitial telomeric sequences in a few amphioxus chromosomes (**Supplementary Fig. S5, S9**). The putative centromeric regions consist of species-specific satellite monomers of different sequences and lengths, with inverted repeat structures (**Supplementary Fig. S11**). Besides the telomeres and centromeres, our new genomes contain newly resolved complex repeat regions, including satellite DNA or rDNA arrays (**Supplementary Fig. S12**), that are partial or absent in the previous amphioxus genomes.

Our phylogenomic analyses using whole-genome alignments of amphioxus against other chordates and one invertebrate outgroup confirmed amphioxus as the most basal chordate lineage, with a relatively lower genome-wide substitution rate (**Fig. 1f**). Based on 3,653 single-copy orthologous genes, we estimated that different chordate lineages diverged about 552.0 MYA, and three amphioxus species diverged about 86.6 MYA (**Supplementary Fig. S3**). Over 73% vertebrate orthologous gene groups are present in amphioxus genomes (**Fig. 1g**). The vertebrate specific genes are enriched for various gene ontology (GO) categories including signalling pathway regulation and muscle functions, while the amphioxus specific genes are enriched for GOs of tissue regeneration (*23*) and apoptosis, among many others (**Supplementary Table S2**). We also identified 27,032 conserved sequence elements between vertebrates and amphioxus, and majorities of them (26,955) are located in protein-coding regions. Finally, the amphioxus genomes were found to have a moderate repeat content of about 30% (**Fig. 1h**), but they contain abundant MITEs (miniature inverted-repeat transposable elements) that are nearly absent in vertebrates. These MITEs seem to have propagated more recently in amphioxus species, relative to other DNA transposons (**Supplementary Fig. S13**).

Ancestral karyotypes of amphioxus, chordates and vertebrates

The assembled chromosome number of Bj, Bf and Bb is respectively 18, 19 and 20, consistent with their reported karyotypes (*22, 24*). Based on their whole-genome alignments, we inferred that similar to the karyotype of Bb, the *Branchiostoma* amphioxus ancestor had 20 linkage groups, which then underwent two chromosome fusions in Bj, and one fusion in Bf after their species divergence (**Fig. 2a, Supplementary Fig. S14**).

Genomic comparison between Bb vs. chicken allows us to reconstruct the karyotype of chordate ancestor. We chose chicken because it is one of the vertebrates that exhibit the lowest rates of lineage-specific chromosomal evolution (*15, 16, 25, 26*) and gene duplications (*27, 28*). Consistent with two rounds of WGDs followed by gene loss, one single-copy amphioxus gene typically has between one to four homologs in vertebrates (**Supplementary Fig. S15**). Moreover, genes from one Bb chromosome are more frequently found to have homologs distributed on four different chromosomes in chicken (**Supplementary Fig. S16**), compared to spotted gar or human (**Supplementary Fig. S17**), confirming that chicken better preserve the ancestral vertebrate karyotype with less interchromosomal rearrangements. We also found several Bb chromosomes share their combination of homologous chicken chromosomes. For instance, Bb chr13, chr14 and chr17 all have their homologous genes located on the chicken chr2 (GGA2), GGA7, GGA27 and GGA33 (**Fig. 2b**). This suggested that these three Bb chromosomes were likely derived from one single chordate ancestral linkage group (CLG) (**Fig. 2c**). Similarly, Bb chr1 shares its homologous chicken chromosomes exclusively with either Bb chr19 or chr20 (**Supplementary Fig. S16, Fig. 2c**), suggesting Bb chr1 originated from a translocation between two CLGs. Moreover, we inferred that Bb chr2 and chr16 fused at the vertebrate ancestor prior to whole genome duplication, while Bb chr3 was split into two (**Supplementary Fig. S16, Fig. 2c**). Taken together, we inferred that there was a total of 17 CLGs (**Fig. 2c**), consistent with previous results (*4, 13, 16*).

To reconstruct the evolutionary trajectories of how CLGs gave rise to the representative extant vertebrate karyotypes, we mapped the homologs of Bb genes assigned to 17 CLGs (**Fig. 2c**) across the chromosomes of chicken or spotted gar. Most chicken and gar microchromosomes have homologous Bb genes predominantly derived from one single CLG (**Fig. 2d, Supplementary Fig. S18**). Such striking evolutionary stability of microchromosomes spanning the entire chordate evolution supports the hypothesis that they were likely present at the ancestor of bony vertebrates (*16, 29-31*). Some chicken microchromosomes (e.g., GGA28 and GGA30), like most macrochromosomes, nevertheless are homologous to two or more CLGs (**Fig. 2d**). When the same combination of CLGs were found for two different homologous GGAs, e.g., GGA28 and GGAZ (homologous to CLG2 and CLG15), we inferred a fusion or translocation likely occurred between 1R and 2R, as illustrated in **Fig. 2e**. We identified a total of 5 such putative post-1R chromosome fusions or translocations (**Supplementary Fig. S19**), whose 2R descendant genes are predicted to be grouped together (**Fig. 2f**, e.g., GGA28 and GGAZ genes) apart from other ohnologs (GGA10 and GGA25 genes) of the same CLG origin but without undergoing post-1R fusions or translocations. This was broadly supported by the phylogenetic trees (**Fig. 2f**, **Supplementary Fig. S19**) constructed from chicken ohnolog gene groups (**Supplementary Table S3, Supplementary Fig. S20**). Extending our phylogenetic reconstructions to 243 chicken paralog groups with at least three ohnologs available, we found among the 9 CLGs that gave rise to ohnologs distributed on 4 GGAs (we termed genes of each of these 4 GGAs as 'ohno linkage group', ohno-A, B, C, D), 7 CLGs' ohnolog trees exhibited a phylogenetic structure that strongly supported the 2R hypothesis (**Supplementary Fig. S21**). That is, ohnologs from two GGAs of the same post-1R origin (ohno-A/B or C/D) were grouped together in their phylogenetic trees. When such ohno linkage groups involve microchromosomes, we revealed that microchromosomes always contain much less ohnologs than the other macrochromosomes of the same post-1R origin (**Supplementary Fig. S22**). This led to our hypothesis that microchromosomes possibly originated by asymmetric sequence loss after the 2R in the vertebrate ancestor.

By concatenating chicken ohnologs from the same ohno linkage group (A, B, C or D), together with their orthologs of human, mouse and gar, we constructed their phylogenetic trees and dated the timing of 1R and 2R (**Fig. 2g**). The 1R was estimated to occur 547 MY ago, in less than 10 MY since the divergence of chordate common ancestor (**Fig. 2g**). In addition, we estimated that jawed vertebrates experienced 2R about 517 MY ago (**Fig. 2g**), 10 MY after their divergence from jawless vertebrates (**Supplementary Fig. S3**).

Amphioxus specific gene duplications

Although without undergoing WGDs, the three amphioxus species have a comparable number of protein-coding genes (between 22,733 to 26,497) to that of vertebrates (**Supplementary Fig.**

S23a). By phylogenetic reconstruction of 8,464 orthologous gene groups whose members are present in both amphioxus and vertebrates, we estimated that the amphioxus ancestor had acquired 4,855 genes (**Fig. 3a**), some of which may also result from gene loss in the vertebrate ancestor. Interestingly, genes that retained at least two paralogs in vertebrates are more likely to have undergone duplications in amphioxus (P < 1.71e-13, Fisher's exact test, **Supplementary Table S4**), suggesting convergent gene gains in vertebrates and amphioxus. For example, among the orthologous gene groups that have multi-copy genes in Bb, 74% have multi-copy homologs in chicken, but only 33% of the orthologous gene groups with single-copy Bb genes have homologs in chicken (**Fig. 3b**). We also found cases of recurrent duplication in amphioxus (**Supplementary Fig. S23b**) as demonstrated by a recent study for *MRF* genes (*32*). For instance, there are 3 ohnologs of the *Slc27a* gene family derived from a single chordate ancestral gene which was independently duplicated multiple times at the ancestor of amphioxus (**Fig. 3c**).

The other prominent case of convergent gene acquisition in amphioxus and vertebrates is demonstrated by certain members of *Hox* genes. Amphioxus has one prototypical *Hox* gene cluster (*AmphiHox*), whose posterior *Hox* genes have an ambiguous orthologous relationship with the vertebrate *Hox* paralog groups (HPGs), leaving the *Hox* gene number of chordate ancestor still controversial (*33-35*). Our phylogenetic analysis confirmed one-to-one homologous relationships of some *Hox* (1-5, 9, 15) genes between amphioxus and vertebrates, dating their likely existence to the chordate ancestor (**Fig. 3d**). Other *Hox* genes likely have undergone gain and loss events independently in the ancestors of the two clades' (**Fig. 3e**). For instance, the amphioxus *Hox6-8* and the vertebrate *HPG8* seem to be acquired after the two chordate clades diverged from each other. The posterior amphioxus *Hox* genes *Hox10-12* and *Hox13-14* are respectively grouped with the vertebrate *HPG9* and *HPG11-13*, suggesting amphioxus-specific duplications from an ancestral chordate *Hox* gene that might have subsequently become lost in the vertebrate ancestor. Similar to *HPGs*, amphioxus *Hox* genes exhibit a temporal colinearity of expression pattern, with the anterior genes expressed in earlier developmental stages than the posterior genes (**Supplementary Fig. S24**).

One major molecular mechanism that contributed to the gene acquisition of amphioxus is segmental duplications, which tend to be of more recent origin and often species-specific (**Supplementary Fig. S25**). Segmental duplications accounted for a higher percentage of the genome in amphioxus vs. vertebrates (9% vs. 3.5%, **Fig. 3f**); they are on average 7.8 kb long, but can be up to 300 kb (**Fig. 3g, Supplementary Fig. S26**). These duplicated segments encompass genes that are enriched for GO categories of G-protein coupled receptor activities, protein tyrosine kinase activities or nucleic acid binding functions (**Supplementary Table S5**). These genes are also frequently enriched for multi-copy ohnologs in vertebrates (*36-38*). Transcriptional factors or genes involved in early development that are often retained after vertebrate WGDs (*39, 40*), however, are not enriched in amphioxus segmental duplicates.

Developmental dynamics of amphioxus chromatin architecture

Eukaryotic genomes are folded into (active/A or inactive/B) chromatin compartments and to a finer scale of TADs. Such hierarchical three-dimensional (3D) chromatin architectures were previously shown in *Drosophila*, teleosts and mammals to be gradually established or reprogrammed during embryonic development (*41-43*).

To examine whether this is a broadly conserved feature between invertebrates and vertebrates, we collected time-series population Hi-C data of Bf spanning six developmental stages of 1-cell zygote, 32-cell, 64-cell embryos, gastrula, larvae, and adult muscle tissues (**Supplementary Table S9**). Both the percentage of actively transcribed genes (**Fig. 4a**) and the total number of TAD boundaries (TAB) (**Fig. 4b, Supplementary Fig. S27-28**) display a significant ($P < 0.01$, Wilcoxon test) increase after zygotic genome activation (ZGA) around the 64-cell stage (*44*). The strength of TABs measured by insulation scores also becomes generally intensified during development particularly in those strong TABs (**Fig. 4c**). These patterns are similar to those found in Drosophila and mammals (*42, 45*) where major TAD structures of zygote genomes emerge after, although do not necessarily depend on ZGA. In contrast to mammals and Drosophila, the amphioxus genome is highly compartmentalized before ZGA. The A/B compartment strength further becomes significantly (P<0.05**, Supplementary Fig. S29**) increased after embryonic stages, but becomes decreased, i.e., possibly reprogrammed on some chromosomes at the gastrula stage (**Fig. 4d-e, Supplementary Fig. S30**).

To explore the formation mechanisms of TADs in amphioxus, we examined the TABs and found that they are enriched for putative binding motifs of chromatin architectural protein CTCF (**Supplementary Fig. S31**), whose transcription level is also specifically increased at ZGA (**Supplementary Fig. S32**). There are disproportionately more (>52%) CTCF-binding site pairs present with convergent forward and reverse orientations at the two TABs of the same

TAD (**Supplementary Fig. S33**) (*46*). These results together suggested that similar to vertebrates, loop extrusion facilitated by CTCF protein might play an important role during the formation of TADs upon ZGA of amphioxus. Another mechanism of TAD formation, i.e., self-organization likely mediated by heterochromatin interactions, could also play a role, however it requires chromatin profiling data of different embryonic stages before and after ZGA to be tested in future.

Once established at 64-cell stage, 26.82% of the TABs are overlapped with those in all the later developmental stages, with about 16.88% to 22.95% of TABs only present in one certain stage or tissue (**Fig. 4b**). This indicates that similar to Drosophila and mammals, substantial numbers of, but not all TADs become stabilized and conserved across stages after ZGA, with many others showing dynamic changes during development. To further illustrate this process, we scrutinized the *Hox* cluster of Bf, which is encompassed in one single TAD from 1- to 64-cell stages but becomes segregated into two TADs (**Supplementary Fig. S28**) since the gastrula stage during later development (**Fig. 4f**). The TAB within the *Hox* cluster is weak at gastrula and larvae stages, but becomes clearer in the adult tissue (**Fig. 4f**). This is in contrast to the previous result that characterized the *Hox* cluster of European amphioxus (*B. lanceolatum*) as one TAD, with pooled samples of different embryonic stages and 4C technique (*8*). The *Hox* TABs in adult muscles seem to be conserved across different amphioxus species around *Hox7*. Interestingly, the entire *Hox* cluster of Bb (together with three neighbouring genes) is included in a large genomic inversion (**Fig. 4g**) that occurred after its divergence from Bj in the last 50 MY, with its functional impact on the Bb genome remained to be elucidated in future.

Evolutionary turnovers of sex determining regions between amphioxus species

The sex-determination (SD) mechanisms of amphioxus remain largely enigmatic, with no cytogenetic evidence for the existence of differentiated sex chromosome pair in Bf and Bb (*19, 47*). A recent genetic study suggested that Bf has a female heterogametic sex chromosome system (male ZZ, female ZW) (*48*). We confirmed this by generating a heterozygous female mutant strain of *Pitx* with transcription activator-like effector nucleases (TALENs), whose mutant alleles are only carried by their daughters. While the mutant alleles can be found in both sons and daughters of male heterozygous mutant strain (**Supplementary Fig. S34**). Using whole-genome re-sequencing data of between 10 to 48 individuals per sex per species

(**Supplementary Table S6**), we identified the sexually differentiated regions (SDR) that harbor female-associated SNPs, i.e. excessive female heterozygotes, and are not shared between the three amphioxus species (**Fig. 5a-c**). In particular, the SDR of Bf is located on Chr16 and harbors 189 genes (**Supplementary Table S7**); and those of Bj and Bb are located at two different genomic loci of Chr3, harboring 35 genes and one gene respectively (**Supplementary Table S8**). These SDRs consistently exhibit the highest levels of population differentiation (measured by F_{st}) between sexes throughout the genome (**Supplementary Fig. S35**), but do not exhibit sexually differentiated patterns of mapped read coverage. These results together indicated that all three amphioxus species have non-homologous female heterogametic sex chromosomes that have not become differentiated in their genomic sequences.

The homomorphic sex chromosomes of amphioxus are similar to those of many fish and frog species, sharing the feature of rapid evolutionary turnovers between species (*49*). This is in contrast to the relatively stable and highly differentiated sex chromosomes of most birds and mammals and may be explained by the 'fountain-of-youth' hypothesis. It postulates that occasional sex reversal may induce rare recombination between sex chromosomes and prevent them from becoming differentiated (*50*). Supporting this, we found between 10 to 40% of the phenotypic female or male individuals of the three species exhibit a genotype of the opposite sex in their SDRs (**Supplementary Fig. S36-38**).

With the advantage of fully assembled sequences of ChrZ of Bb and Bj, and particularly those of both ChrZ and ChrW of Bf (**Fig. 1**), we further reconstructed the evolutionary history of these species' SDRs. The SDR of Bf can be divided into two regions which likely have suppressed or reduced homologous recombination between ChrZ/W at different time points (termed 'evolutionary strata' (*51*)). The older stratum spans 4.1 Mb sequence at one end of Bf ChrW chromosome and exhibits uniformly much higher levels of ChrZ/W pairwise sequence divergence and intersexual F_{st} than the rest SDR (**Fig. 5d, Supplementary Fig. S36**). The boundary of this stratum aligns with that of chromosomal inversion between ChrZ/W of Bf (**Fig. 5e**), which probably accounted for the recombination suppression in this stratum. In contrast, the F_{st} values and ChrZ/W sequence divergence levels are not uniform in the rest SDR of Bf (4.1Mb-11.5Mb), suggesting homologous recombination may have been gradually reduced without involving chromosomal inversions (**Supplementary Fig. S36**). The SDRs of Bj and Bb do not exhibit a pattern of 'evolutionary strata' and seem to have gradually reduced recombination, suggested by their F_{st} patterns (**Fig. 5f, Supplementary Fig. S37-38**).

The SDR of each amphioxus species is expected to harbor respective upstream sex-determining genes, which may constitute the sex-determining pathways together with genes on the other chromosomes. We examined the orthologs of 10 reported vertebrate sex-determining genes, and found none of them are present in SDRs of amphioxus. Three upstream SD genes of some vertebrates, *Dmrt1* (**Supplementary Fig. S39**), *Amh* and *Rspo1* do not have an ortholog in the amphioxus genome (**Fig. 5g**); among the rest, only *Sf1* and *Foxl2* exhibit a testis- or ovary-biased expression pattern in amphioxus (**Fig. 5h**). Among the amphioxus SDR genes, we identified a candidate Bj SD gene that is absent in Bf and Bb (**Supplementary Fig. S40**), and a candidate Bb SD gene that are present in Bf and Bj, both of which have specific or biased expression in the gonads, and do not have a vertebrate homolog (**Fig. 5i-j**). These results together indicated that amphioxus and vertebrates independently evolved their SD pathways.

Conclusions

With three reference-quality genomes of amphioxus, we uncovered their interspecific diversities of genes and chromosomes to an unprecedented resolution. This enabled more direct and accurate reconstruction of ancestral status of the ancestors of both amphioxus and chordates, which was previously based on the draft genome of one amphioxus species. We inferred that there were 20 ancestral linkage groups in the ancestor of *Branchiostoma* amphioxus, best approximated by the Bb genome; and confirmed there were 17 ancestral linkage groups in the chordate ancestor (*13, 16*). Phylogenetic analyses of vertebrate ohnologs and their amphioxus orthologs dated the timing of WGDs, and further characterised the rearrangements and asymmetric loss/retention among the duplicated descendants of CLGs that gave rise to the vertebrate ancestral karyotype. These evolutionarily distant comparisons between amphioxus and vertebrates can be attributed to the slow-evolving genomes of the former relative to those of urochordates.

Our analyses also revealed shared or independently evolved genomic features of amphioxus and vertebrates. For example, both clades seem to establish their major TAD architecture after ZGA, and form two TADs within the *Hox* gene cluster, suggesting these patterns probably originated in their chordate ancestor. In the absence of WGDs, amphioxus species expanded their gene repertoire by segmental duplications or individual gene duplications; and independently evolved their sex-determination pathways from each other, and from vertebrates. By the development of rich genomic resources from this and previous works (*12, 16, 17*), as well as that of gene knockout techniques (*52*), we expect the resurgence of interest into this classic evo-devo model organism, with more functional insights into its genes to be uncovered in future.

Methods

Genome sequencing and assembly

Bb and Bj were collected from Xiamen Rare Marine Creature Conservation Areas (Fujian, China) and Bf was introduced from Dr. Jr-Kai Yu's laboratory (Institute of Cellular and Organismic Biology, Academia Sinica, Taiwan) (**Supplementary Fig. S1**). All of them were cultured as previously described (*52, 53*) . Interspecific hybrids were produced by pooling the sperm of one species, and the eggs of another species except that Bj and Bb cannot be crossed with each other. We extracted high molecular weight genomic DNAs from the muscle tissues of a single individual (male Bj/Bf F1 offspring, Bb/Bf F1 offspring with unidentified sex) using the DNeasy Blood & Tissue Kit (QIAGEN, Valencia, CA), and inspected the DNA quality by Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA) and 2100 Agilent Bioanalyzer (Agilent). We prepared the 20 kb SMRTbell^{TM} PacBio libraries and generated sequencing data of ~50G for the two hybrids (Bb/Bf and Bj/Bf). We estimated the heterozygosity levels of three species' genomes using Illumina reads of the three species using GenomeScope (*54*). For the hybrids, the estimated genome size was equivalent to the sum of the haploid genome sizes of the parental species (**Supplementary Fig. S1**).

We used Falcon (*55*) to assemble the PacBio subreads of two hybrid samples, after discarding raw subreads and corrected reads (preads) shorter than 8kb. We used the following parameters to avoid collapse of reads derived from different parental species: pa_HPCdaligner_option = -v -dal128 -t8 -e0.75 -M24 -l3200 -k18 -h480 -w8 -s100, ovlp_HPCdaligner_option = -v -dal128 -M24 -k24 -h1024 -e.96 -l2500 -s100. We used the arrow (from the Falcon assembler) algorithm to polish the contigs twice, followed by another two-round polishing with the Illumina reads derived from the same hybrid individual, using pilon (1.22) (*56*). To assign the contig sequences of hybrids to each parental species, we aligned the

Illumina reads of either parental species to the contigs by bwa-mem with default parameters, and only kept the alignments with a mapping quality higher than 60. For each contig, we calculated the proportion of nucleotide sequences that were mapped by each species' reads (coverage), without considering the contigs shorter than 20kb. We assigned a contig to either parental species if the sequencing coverage was larger than 10% for one parental species, while the sequencing coverage for the other species was below 1% (**Supplementary Fig. S2**). We then used minimap2 (2.15-r905) (*57*) to align the PacBio reads of hybrids to the assembly, with the option '--secondary=no', and partitioned the species-specific haploid reads. These partitioned reads were used for assembling the four haploid assemblies (one Bb, one Bj and two Bf) by Canu (1.6) (*58*) ('corOutCoverage=200 correctedErrorRate=0.15') and Falcon ('pa_daligner_option= -k18 -e0.7 -l2000 -h480 -w8 -s100, ovlp_daligner_option=-k24 -e.93 -l2000 -h600 -s100'). Since the read length of Bb/Bf was longer, we increased the '-l' parameter from 2000 to 2500 in 'pa_daligner_option' and from 2000 to 3000 in 'ovlp_daligner_option'. The polishing steps were similar to those for the diploid assembly of hybrids. Then contigs of two pipelines were merged: we aligned the Canu contigs against the falcon contigs using the nucmer aligner (MUMmer 3.0) (*59*) with the option -b 400. When one Falcon contig spanned the boundaries of two Canu contigs, we linked the Canu contigs with a gap of 200 Ns.

Finally, we used the Juicer (1.7.6) (*60*) pipeline and 3D-DNA (180922) (*61*) to connect the contigs into chromosome-level scaffolds. To reduce the false-positives of contig splitting, we used the following parameters: --editor-coarse-resolution 500000 --editor-coarse-region 1000000 --editor-saturation-centile 1 -r 0 --editor-repeat-coverage 1 --editor-coarse-stringency 70. We manually curated the chromosome assembly by editing the Hi-C contact map using Juicebox (1.90) (*62*). After that, we updated the assembly using the 'review' module of 3D-DNA. The unanchored scaffolds are highly repetitive, with repeat content as high as 79.0%, 63.5% and 81.7% for Bb,Bj and Bf respectively.

Genome annotation

To annotate genes, we generated Iso-seq and RNA-seq data from whole-body adult male and female individuals of the three species. We used IsoSeq3 (3.1.0) (*63*) and Trimmomatic (0.36) (*64*) for pre-processing the raw reads. Then we generated reference-guided and *de novo* assembled transcript sequences using Cupcake (5.8) with Iso-Seq reads, and StringTie

(1.3.3b)(*65*) (-m 300 -j 5 -c 8) and Cufflinks (2.2.1) (*66*) (–multi-read-correct –max-intron-length 30000) and Trinity (2.6.6) (*67*) (--min_glue 10 --path_reinforcement_distance 30 --min_contig_length 400 --jaccard_clip) with RNA-seq reads. We then used the Mikado (1.2.2) (*68*) to integrate all transcript sequences. We used RepeatModeler (1.0.10) (*69*), Tandem Repeat Finder (409) (*70*) ('2 7 7 80 10 50 500 -d -l 6') and MITE_Hunter (*71*) ('-I 86 -n 8 -c 8') for annotating and classifying the repeat families.

To produce a consensus gene model, we ran MAKER (2.31.10) (*72*), after masking the annotated repeats. We used the query protein sequences from NCBI RefSeq database (Bb: GCA_001625405.1 and Bf: GCA_000003815.1), and the transcriptome annotations produced by Mikado. The MAKER gene annotation was then used to train SNAP (2013-11-29) (*73*) (maker2zff -c 0.99 -e 0.99 -o 0.99 -l 800 -x 0.01) and AUGUSTUS (3.3) (*74*) for *ab initio* predictions. Gene evidence from protein alignment, StringTie transcripts, ISO-seq transcripts, SNAP and AUGUSTUS predictions, were combined by EVidenceModeler (EVM) (1.1.1) (*75*), with the highest weight on the protein alignment and StringTie transcripts (10), intermediate weight on ISO-seq transcripts (5) and lowest weight on the *ab initio* predictions. We used the PASApipeline (v2.3.3) (*76*) to polish the gene models. We used InterProScan (5.35-74.0) (*77*) to annotate gene ontology (GO) for the predicted coding genes.

To annotate putative centromeres, we counted the copy number and total length for each satellite repeat based on the RepeatMasker results and inferred the most abundant and longest satellite sequences to be associated with centromeres. The identified centromeric monomer of Bf is consistent with the reported result (*78*). The recombination rates were estimated with ReLERNN program, using the individually sequenced data (**Supplementary Table S8**). The nucleotide diversity was estimated in 100 kb windows using VCFtools (0.1.16) (*79 80*). To verify the centromere with the fluorescence in situ hybridization (FISH) technique, we prepared the probe of candidate centromeric monomer and the slides of Bf chromosomes. The details of the FISH experiment were described in Lie et al. (2002) (*79*). To annotate telomeres, we searched for clusters of (AACCCT)n repeats throughout the genomes using RepeatMasker. We only kept those with a total length of 200 bp (33.3 consecutive AACCCT repeats) to reduce false positives. We used the R package Quadron (*81*) to predict the G-quadruplexes (G4) throughout the genome with default settings. Then we calculated the length of G4 elements over 20kb sliding windows along the chromosomes using bedtools coverage.

Comparative genomic analyses

We included three amphioxus species and four vertebrate species to infer the orthologous gene groups. The Refseq annotations of human (GCF_000001405.39), mouse (GCF_000001635.26), zebrafish (GCF_000002035.6) and chicken (GCF_000002315.6) were downloaded from NCBI. When multiple isoforms were present, we selected the longest one for each gene. We ran OrthoFinder (2.2.7) (*82*) to group the orthologous genes, using diamond (0.9.21) for protein alignment. We used Last (1042) (*83*) to align genomes of mouse (GRCm38.p4), chicken (GRCg6a), zebrafish (GRCz11), Bb, Bj and Bf against the human reference genome (GRCh38.p12), with -uMAM4 for mouse alignment, and more sensitive -uMAM8 for the other species. The one-to-one best alignments were retained and merged by Multiz (v11.2) (*84*).

For the reconstructing the chordate phylogeny, we added *Ciona intestinalis* (GCA_009617815.1) (*85*) and scallop (*Mizuhopecten yessoensis*, ASM211388v2) (*86*), with the latter set as an outgroup. We excluded the alignments in which the sequences were aligned to non-homologous chromosomes among amphioxus, because they likely represent alignment errors. The filtered alignments contained 5,074 loci, with a total size of 276,373 bp. We used IQ-TREE (2.0-rc1) (*87*), with the substitution model (TVMe+R3), to construct the phylogenomic tree, and ran bootstrapping for 100 times.

We used the PhastCons from the PHAST package (1.5) (*88)* to annotate the conserved non-coding elements across the genomes. First we used (msa_view --4d) all fourfold degenerate (4d) sites for estimating a nonconserved phylogenetic model by phyloFit (a PHAST program), with the phylogenetic tree as ((((human,mouse),chicken),zebrafish),(bf,(bb,bj))). Then we ran the PhastCons program with the alignments and the nonconserved model to estimate the rho value and the conserved model (the nonconserved model remained the same as the 4d model). Finally, we ran the PhastCons program again but added --most-conserved option to identify the conserved elements. Then we compared the conserved elements with annotated features of the human genome (RefSeq GCF_000001405.39 and Ensembl annotation) using BEDTools (2.29.0) (*89*). For each conserved element, we assigned it to one feature if overlapped. If multiple features were overlapped with a single element, they were assigned under the following priority: protein-coding region > pseudogene > non-coding RNAs > lncRNA > UTR > intron > intergenic.

Ancestral karyotype reconstruction

We generated whole genome alignments between amphioxus species by minimap2 (2.15-r905) (*58*) (-x asm20) and visualized the alignments by D-Genies online tool (*90*) (**Supplementary Fig. S14**). We selected 7269 orthologous gene groups (orthogroups) in which Bb genes are located within the same chromosome. 1799 orthogroups contained more than one gene in chicken which were informative for reconstructing the chordate ancestral karyotype. For each Bb chromosome (*i*), we asked which chicken chromosome (*j*) its homologous genes belong to, and counted the gene number for each chicken chromosome (CK*ij*). Then we calculated the relative abundance of genes of a chicken chromosome for a given Bb chromosome (nCK*ij*):

$$\mathrm{nCK}_{ij} = \frac{\mathrm{CK}_{ij}}{\sum_{j=1}^{33} \mathrm{CK}_{ij}}$$

We included 33 chicken chromosomes, and retained a chicken chromosome when the nCK*ij* value was larger than 4%, for a given Bb chromosome (*i*). Then we visualised the nCK*ij* values for every Bb chromosome with a network-style graph (**Fig. 2b, Supplementary fig. 16**), using the igraph R package. We used 244 orthogroups that retained three or four chicken ohnologs and performed coding sequence alignments using MAFFT (v7.294b) (*91*). Then we constructed the phylogenetic tree using concatenated sequence alignments of the same CLG using IQ-TREE, with 1000 times bootstrapping. Based on the phylogenetic relationships, we assigned the four ohno-chromosomes derived from a single CLG as ohno-A, ohno-B, ohno-C and ohno-D. For each ohno-chromosome group, we further included the orthologous genes of human, mouse and spotted gar of the chicken gene in that group (**Supplementary Table S3** and **Supplementary Fig. S20**) Then all the coding sequences of three amphioxus species and vertebrates were aligned with MAFFT (7.427) (*91*) and GUIDANCE2 (2.02) (*92*) pipeline, producing concatenated alignments with 409,659 nucleotide sites. We then used BASEML (4.9j) (*93*) to estimate the overall mutation rate with the time calibration on the root node (550 MY for the vertebrate and amphioxus split (*94*)). The topology "((bf,(bb,bj)),((((chicken-Ohn_A,(human-Ohn_A,mouse-Ohn_A)),gar-Ohn_A),((chicken-Ohn_B,(human-Ohn_B,mouse-Ohn_B)),gar-Ohn_B)),(((chicken-Ohn_C,(human-Ohn_C,mouse-Ohn_C)),gar-Ohn_C),((chicken-Ohn_D,(human-Ohn_D,mouse-Ohn_D)),gar-Ohn_D))));" was used. General reversible substitution model and discrete gamma rates were estimated by maximum likelihood approach under the strict clock. The divergence time was then estimated using MCMCtree (4.9j) (*95*), with three soft-bound calibration time points: 534–566 MY for the vertebrate and Branchiostoma species split, and 62-101 MY for the human and mouse split, and, 306-332 MY for the chicken and mammal split, 416-422 for the teleost and tetropad split (*96*).

Gene evolution

We used SDquest (0.1) (*97*) to identify segmental duplications (SDs) in all amphioxus species and four vertebrates including human (hg38), mouse (mm10), chicken (galGal6) and zebrafish (danRer11). We excluded the sex chromosomes of human, mouse and chicken, and alternate-loci scaffolds of zebrafish as these sequences may confound the identification of SD. We only kept SDs that are longer than 1000 kb, and show a sequence similarity level of at least 70%. For studying gene gain and loss, we selected 8,464 orthologous gene groups that contain at least one vertebrate species and one amphioxus species as the input for Notung (2.9.1) (*98*) gene family reconstruction. We identified 200 orthogroups that had more than one gene copy in all amphioxus species, but had single-copy genes in vertebrates. The mean copy number of the expanded gene families were 3.6, 3.8 and 4.8 for Bb, Bj and Bf respectively.To elucidate the evolution of the *Hox* genes across chordate species, protein and CDS sequences of chicken, mouse, human and zebrafish *Hox* genes were downloaded from NCBI, and aligned to those of amphioxus species by MAFFT (v7.407), with alignment polishing by trimAl (v1.4.rev15) (*99)*. We used IQ-TREE to infer the phylogeny, and the AVX+FMA model was selected automatically by IQ-TREE. We used EvolView online tool (https://www.evolgenius.info/evolview) to visualize our phylogenetic tree. RNA-seq data of multiple Bf developmental stages were downloaded from NCBI SRA (PRJDB3785) for estimating the *Hox* gene expression level using HISAT2 (2.0.4) and featureCounts (v1.5.2).

3D genome analyses

in situ Hi-C libraries were constructed from the muscle and embryonic tissues of amphioxus as described before (*100*). Hi-C data were mapped to the genomes using bwa-mem (0.7.17-r1188) with parameters '-A 1 -B 4 -E 50 -L 0'. The quality control including valid pairs and cis/trans ratio of Hi-C data was finished by using pairtools(0.3.0)

(https://pairtools.readthedocs.io/en/latest/) and the estimated resolution was calculated by HiCRes(2.0) (*101*)

Then the mapped read-pairs were used to generate raw Hi-C contact matrix at 5kb, 15kb and 30kb resolution using hicBuildMatrix of the HiCExplorer (2.2.1) suite (*102*). We used the ICE method implemented in hicCorrectMatrix to remove the bins with extremely low or high numbers of reads, and visualized the matrix with hicPlotMatrix**.** We used hicFindTADs to generate the coordinates of TADs and the TAD insulation score of each bin (--thresholdComparisons=0.01, --delta=0.01). To investigate the overlaps of TAD boundaries between different developmental stages, we combined the TAD boundaries of all development stages into one set, and extended each boundary for 5kb of both sides to form 15kb windows and merged adjacent windows when their distance was not longer than 10kb. This generated a set of boundaries that existed in at least one developmental stage. We then compared the boundaries of each stage to this common set, and defined conservation of boundaries as an overlap of at least 15 kb in size. We used cooltools (0.3.2)(https://cooltools.readthedocs.io/en/latest/) and its call-compartments function to obtain the first eigenvector values (PC1) of each chromosome of the Hi-C matrices with a 250kb resolution. Regions with positive PC1 values are assigned as A (active) compartments and those with negative PC1 values are assigned as B (inactive) compartments, adjusted by the gene density of the region. Compartment strength was calculated as $AA \times BB/AB^2$ for each chromosome. Saddle plot was also obtained by cooltools. In brief, enrichment contact maps at the 250kb resolution were normalized by genomic distance into a 50 by 50 bin matrix to calculate the observed/expected (O/E) values as contact enrichment. Bins in the matrix were sorted by PC1 values and all contacts with similar PC1 values were aggregated to obtain compartmentalization saddle plots with B-B interactions in the upper left corner and A-A interactions in the lower right corner. The numbers in saddle plots indicate the strength of the top 20% of A-A interactions (over A-B interaction) and the bottom 20% of B-B interactions (over B-A interactions). We used FIMO (*103*) to search for human CTCT motif (MA0139.1) in the amphioxus genomes and identified 62,987 putative CTCF motifs. To test whether the CTCF motif was enriched in the TAD boundary, we used bedtools intersect to identify the CTCF motifs located in the 15kb TAD boundaries (5kb boundary extended by 5kb of both sides) of the six developmental stages. In addition, we also checked whether the TAD boundaries contain more CTCF motifs than by chance, we randomly selected 15 kb windows across the genome and calculated the proportion of windows that contain CTCF motifs (**Supplementary Fig. S41**). The pairings of convergent CTCF sites at domain boundaries is considered as a hallmark of the conserved role of CTCF/cohesion in TAD formation (*104)*. The enrichment pattern of putative CTCF binding sites (**Supplementary Fig. S32**) and the distribution pattern of convergent CTCF site pairs (**Supplementary Fig. S34**) were similar for TAD results derived from different TAD-calling bin sizes.

Sex chromosome analyses

Pitx mutants were generated and detected using the TALEN method as described before (*105*). The TALEN pair used for mutant generation are Fw3: 5'-GCAACCGTTCGACGAC-3' and Rv3 5'-TGTAGGCCGGCGAGTA-3' which are from the third coding exon of the gene. A *Tat* Ⅰ restriction site was included in the target site for genotyping and primer pair used for genotyping are Pitx-TALEN-PCR-F2 (5'-AGGTCTGGTTCAAGAACCG-3') and Pitx-TALEN-PCR-R4 (5'-TCACGGTAAGCGTAAGGCTG-3'). Two different mutant stains were generated. The founder of stain 1 is a female, which was crossed with a wild type male to generate F1 offspring, from which a female heterozygote was further crossed with a wild type to generate F2 descendants. In contrast, the founder of strain 2 was a male and an F1 heterozygous male was used to generate its F2 descendants.

We generated re-sequencing Illumina data of multiple individuals of both male and female (on average 25 individuals of each sex) at a coverage larger than 20X (**Supplementary Table S6**). The raw reads were mapped to the reference genome using bwa-mem (0.7.16a), with default parameters. After sorting the alignments with samtools (1.9) sort, we marked the duplications of reads using the MarkDuplicates function of the picardtools package (2.14.0). We then used the GATK (3.8) (*106*) pipeline to call variants. To do so, we ran HaplotypeCaller to generate GVCF output for each sample. This was done separately for each chromosome with the interval (-l) option, and then the GVCF outputs were combined with the GatherVcfs function of the Picard toolkits (*107*). Then we genotyped the variants with the GVCF files as inputs of all samples (joint calling) using GenotypeGVCFs. We selected single-nucleotide variants (SNPs) for further analysis and filtered the SNPs with the following criteria: QD < 2.0 || FS > 60.0 || MQRankSum < -12.5 || RedPosRankSum < -8.0 || SOR > 3.0 || MQ < 40.0. We used the biallelic SNPs (filtered by bcftools -m2 -M2) to screen for sex-linked variants. We further excluded the variants that have minor allele frequency less than 0.05 and missing rate larger than 10%. We used Beagle (28Sep18.793) (*108*) to do the imputation for the variants and obtain an initial set of phased genotype calls for all variants. SHAPEIT (v2.r904) (*109*) was then used to produce a more accurate set of phased genotypes on the variants. A total of 4,954,852, 7,213,889, and 12,016,687 high-quality phased SNPs in Bj, Bb and Bf respectively, were used to perform whole-genome association analysis for the sex trait (male or female) with EMMAX (efficient mixed-model association expedited, version 8.22) (*110*). Population stratification and the hidden relatedness were modeled with a kinship (K) matrix in the emmax-kin-intel package of EMMAX. The genome-wide significance thresholds of all tested traits were evaluated with the formula P=0.05/n (where n is the effective number of independent SNPs). Apart from identifying sex-associated regions, we screened for differentiated regions between the sexes. We calculated the F_{ST} values between male and female populations using VCFtools (0.1.13) (*80*). SNPs with more than two alleles were removed. The F_{ST} values were estimated in a 10 kb sliding window with an overlapping size of 5 kb. For Bb whose sex-determining region is much smaller, we used 5 kb windows instead of 10 kb. We defined the non-recombining regions of the sex chromosomes by the sex-linked SNPs identified through the whole-genome association tests. We evaluated the extend of sex chromosome differentiation with two measures: 1) F_{ST} and 2) the difference between male and female SNP density.

We collected transcriptomes of immature (identifiable but not functionally mature) and mature gonads for studying the candidate sex-determining genes of amphioxus. We mapped the RNA-seq reads against the genomes using HISAT2 (2.1.0) with the parameters '-k 4 --max-intronlen 50000 --min-intronlen 30'. The alignments with mapping quality score lower than 10 were removed (samtools view -q 10). Then we used featureCounts (1.6.2) (*111*) to count the reads mapped to the annotated transcripts. We used the TPM (transcripts per million) method to quantify and normalize the expression levels.

We chose 10 conserved vertebrate SD pathway genes: *Wnt4*, *Sf1*, *β-catenin*, *Rspo1*, *Sox9*, *Amh*, *Foxl2*, *Fst*, *Cyp19a1* and *Dmrt1* to check their presence or absence in the amphioxus genomes. We first checked the orthogroups that contain those SD genes and whether amphioxus is present in these orthogroups. If amphioxus is absent in the orthogroups, we searched the coding sequences of the SD genes against the amphioxus genomes by BLAST **(Supplementary Fig. S40)**. The absence of *Dmrt1* in amphioxus is consistent with a recent study (*112*).

Acknowledgements

Funding: Zhen Huang is supported by the 13th Five-Year Plan for the Marine Innovation and Economic Development Demonstration Projects (FZHJ14) from Fujian Provincial Fund, the scientific research innovation program "Xiyuanjiang River Scholarship" from the College of Life Sciences, Fujian Normal University. Qi Zhou is supported by the National Natural Science Foundation of China (31722050, 31671319 and 32061130208), Natural Science Foundation of Zhejiang Province (LD19C190001), European Research Council Starting Grant (grant agreement 677696) and start-up funds from Zhejiang University. Luohao Xu is supported by the Erwin Schrödinger Fellowship (J4477-B) from the Austrian Science Fund. Wan Cen is supported by the Research Foundation of the Education Bureau of Fujian Province (JT180104). Qiujin Zhang is supported by the Natural Science Foundation of Fujian Province (2019J01277). Jr-Kai Yu is supported by intramural funding from the Cellular and Organismic Biology, Academia Sinica, and grants from the Ministry of Science and Technology, Taiwan (105-2628-B-001-003-MY3; 108-2311-B-001-035-MY3). **Author contributions:** Q.Z., L.X., Q. J. Z., G. L. conceived the project; Z. H., Y. Z., D. C., S. P., T. X., W. C., C. S., X. W., Y. H., C. X., Y. N.Y., Y. Y., W. H., X. H., Y. Z., Y. C., C. B., C. H., L. X., S. X., Z. Y., Y. J. acquired the data; Z. H., L. X., C. C., J. L., Z. Z., W. K., Q. Z. performed the analyses; Z. H., L. X., C. C., J. K.Y., E.D.J., G. L., G. L., Q. J. Z., Q. Z. wrote the paper. **Competing interests:** The authors declare that they have no competing interests. **Data and code availability:** The genome assemblies have been deposited at GenBank under the accession PRJNA603158, PRJNA603159, PRJNA647830. All sequencing data has been deposited at PRJNA602496. A full list of accessions is available in the **Supplementary Table S8**. The scripts used in the study have been deposited at Github (https://github.com/lurebgi/amphioxusGenome)

References

1. N. D. Holland, L. Z. Holland, The ups and downs of amphioxus biology: a history. *Int J Dev Biol.* **61**, 575-583 (2017).
2. S. J. Bourlat, T. Juliusdottir, C. J. Lowe, R. Freeman, J. Aronowicz, M. Kirschner, E. S. Lander, M. Thorndyke, H. Nakano, A. B. Kohn, A. Heyland, L. L. Moroz, R. R. Copley, M. J. Telford, Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. *Nature.* **444**, 7115 (2006).
3. F. Delsuc, H. Brinkmann, D. Chourrout, H. Philippe, Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature.* **439**, 7079 (2006).
4. N. H. Putnam, T. Butts, D. E. Ferrier, R. F. Furlong, U. Hellsten, T. Kawashima, M. Robinson-Rechavi, E. Shoguchi, A. Terry, J. K. Yu, E. L. Benito-Gutierrez, I. Dubchak, J. Garcia-Fernandez, J. J. Gibson-Brown, I. V. Grigoriev, A. C. Horton, P. J. de Jong, J. Jurka, V. V. Kapitonov, Y. Kohara, Y. Kuroki, E. Lindquist, S. Lucas, K. Osoegawa, L. A. Pennacchio, A. A. Salamov, Y. Satou, T. Sauka-Spengler, J. Schmutz, I. T. Shin, A. Toyoda, M. Bronner-Fraser, A. Fujiyama, L. Z. Holland, P. W. Holland, N. Satoh, D. S. Rokhsar, The amphioxus genome and the evolution of the chordate karyotype. *Nature.* **453**, 7198 (2008).
5. S. Bertrand, H. Escriva, Evolutionary crossroads in developmental biology: amphioxus. *Development.* **138**, 22 (2011).
6. P. Holland, The dawn of amphioxus molecular biology - a personal perspective. *The International Journal of Developmental Biology.* **61**, 10-11-12 (2017).
7. J. Garcia-Fern, P. W. Holland, Archetypal organization of the amphioxus Hox gene cluster. *Nature.* **370**, (1994).
8. R. D. Acemel, J. J. Tena, I. Irastorza-Azcarate, F. Marletaz, C. Gomez-Marin, E. de la Calle-Mustienes, S. Bertrand, S. G. Diaz, D. Aldea, J. M. Aury, S. Mangenot, P. W. Holland, D. P. Devos, I. Maeso, H. Escriva, J. L. Gomez-Skarmeta, A single three-dimensional chromatin compartment in amphioxus indicates a stepwise evolution of vertebrate Hox bimodal regulation. *Nat Genet.* **48**, 3 (2016).
9. S. Ohno, Evolution by Gene Duplication. (1970).
10. P. W. Holland, J. Garcia-Fernàndez, N. A. Williams, A. Sidow, Gene duplications and the origins of vertebrate development. *Dev. Suppl.* 125-133 (1994).
11. J. J. Smith, M. C. Keinath, The sea lamprey meiotic map improves resolution of ancient vertebrate genome duplications. *Genome Res.* **25**, 8 (2015).
12. F. Marletaz, P. N. Firbas, I. Maeso, J. J. Tena, O. Bogdanovic, M. Perry, C. D. R. Wyatt, E. de la Calle-Mustienes, S. Bertrand, D. Burguera, R. D. Acemel, S. J. van Heeringen, S. Naranjo, C. Herrera-Ubeda, K. Skvortsova, S. Jimenez-Gancedo, D. Aldea, Y. Marquez, L. Buono, I. Kozmikova, J. Permanyer, A. Louis, B. Albuixech-Crespo, Y. Le Petillon, A. Leon, L. Subirana, P. J. Balwierz, P. E. Duckett, E. Farahani, J. M. Aury, S. Mangenot, P. Wincker, R.

Albalat, E. Benito-Gutierrez, C. Canestro, F. Castro, S. D'Aniello, D. E. K. Ferrier, S. Huang, V. Laudet, G. A. B. Marais, P. Pontarotti, M. Schubert, H. Seitz, I. Somorjai, T. Takahashi, O. Mirabeau, A. Xu, J. K. Yu, P. Carninci, J. R. Martinez-Morales, H. R. Crollius, Z. Kozmik, M. T. Weirauch, J. Garcia-Fernandez, R. Lister, B. Lenhard, P. W. H. Holland, H. Escriva, J. L. Gomez-Skarmeta, M. Irimia, Amphioxus functional genomics and the origins of vertebrate gene regulation. *Nature.* **564**, 7734 (2018).

13. C. Sacerdot, A. Louis, C. Bon, C. Berthelot, H. Roest Crollius, Chromosome evolution at the origin of the ancestral vertebrate genome. *Genome Biology.* **19**, 1 (2018).

14. M. Kohn, J. Hogel, W. Vogel, P. Minich, H. Kehrersawatzki, J. Graves, H. Hameister, Reconstruction of a 450-My-old ancestral vertebrate protokaryotype. *Trends in Genetics.* **22**, 4 (2006).

15. J. J. Smith, N. Timoshevskaya, C. Ye, C. Holt, M. C. Keinath, H. J. Parker, M. E. Cook, J. E. Hess, S. R. Narum, F. Lamanna, H. Kaessmann, V. A. Timoshevskiy, C. K. M. Waterbury, C. Saraceno, L. M. Wiedemann, S. M. C. Robb, C. Baker, E. E. Eichler, D. Hockman, T. Sauka-Spengler, M. Yandell, R. Krumlauf, G. Elgar, C. T. Amemiya, The sea lamprey germline genome provides insights into programmed genome rearrangement and vertebrate evolution. *Nat Genet.* **50**, 2 (2018).

16. O. Simakov, F. Marletaz, J. X. Yue, B. O'Connell, J. Jenkins, A. Brandt, R. Calef, C. H. Tung, T. K. Huang, J. Schmutz, N. Satoh, J. K. Yu, N. H. Putnam, R. E. Green, D. S. Rokhsar, Deeply conserved synteny resolves early events in vertebrate evolution. *Nat Ecol Evol.* **4**, 6 (2020).

17. S. Huang, Z. Chen, X. Yan, T. Yu, G. Huang, Q. Yan, P. A. Pontarotti, H. Zhao, J. Li, P. Yang, R. Wang, R. Li, X. Tao, T. Deng, Y. Wang, G. Li, Q. Zhang, S. Zhou, L. You, S. Yuan, Y. Fu, F. Wu, M. Dong, S. Chen, A. Xu, Decelerated genome evolution in modern vertebrates revealed by analysis of multiple lancelet genomes. *Nat Commun.* **5**, (2014).

18. G. Zhang, X. Fang, X. Guo, L. Li, R. Luo, F. Xu, P. Yang, L. Zhang, X. Wang, H. Qi, Z. Xiong, H. Que, Y. Xie, P. W. Holland, J. Paps, Y. Zhu, F. Wu, Y. Chen, J. Wang, C. Peng, J. Meng, L. Yang, J. Liu, B. Wen, N. Zhang, Z. Huang, Q. Zhu, Y. Feng, A. Mount, D. Hedgecock, Z. Xu, Y. Liu, T. Domazet-Loso, Y. Du, X. Sun, S. Zhang, B. Liu, P. Cheng, X. Jiang, J. Li, D. Fan, W. Wang, W. Fu, T. Wang, B. Wang, J. Zhang, Z. Peng, Y. Li, N. Li, J. Wang, M. Chen, Y. He, F. Tan, X. Song, Q. Zheng, R. Huang, H. Yang, X. Du, L. Chen, M. Yang, P. M. Gaffney, S. Wang, L. Luo, Z. She, Y. Ming, W. Huang, S. Zhang, B. Huang, Y. Zhang, T. Qu, P. Ni, G. Miao, J. Wang, Q. Wang, C. E. Steinberg, H. Wang, N. Li, L. Qian, G. Zhang, Y. Li, H. Yang, X. Liu, J. Wang, Y. Yin, J. Wang, The oyster genome reveals stress adaptation and complexity of shell formation. *Nature.* **490**, 7418 (2012).

19. K. Saotome, Y. Ojima, Chromosomes of the Lancelet Branchiostoma belcheri Gray. *Zoological Science.* **18**, 5 (2001).

20. C. Wang, S. Zhang, Y. Zhang, The karyotype of amphioxus Branchiostoma belcheri tsingtauense (Cephalochordata). *Journal of the Marine Biological Association of the United Kingdom.* **83**, 1 (2003).

21. D. Colombera, Male chromosomes in two populations of Branchiostoma lanceolatum. *Experientia.* **30**, 353-355 (1974).

22. L. F. Castro, P. W. Holland, Fluorescent in situ hybridisation to amphioxus chromosomes. *Zoolog Sci.* **19**, 12 (2002).

23. I. M. Somorjai, R. L. Somorjai, J. Garcia-Fernandez, H. Escriva, Vertebrate-like regeneration in the invertebrate chordate amphioxus. *Proc Natl Acad Sci U S A.* **109**, 2 (2012).

24. Q.-j. Zhang, G. Li, Y. Sun, Y.-q. Wang, Chromosome Preparation and Preliminary Observation of Two Amphioxus Species in Xiamen. *Zoological Research.* **30**, 2 (2009).

25. I. Braasch, A. R. Gehrke, J. J. Smith, K. Kawasaki, T. Manousaki, J. Pasquier, A. Amores, T. Desvignes, P. Batzel, J. Catchen, A. M. Berlin, M. S. Campbell, D. Barrell, K. J. Martin, J. F. Mulley, V. Ravi, A. P. Lee, T. Nakamura, D. Chalopin, S. Fan, D. Wcisel, C. Canestro, J. Sydes, F. E. Beaudry, Y. Sun, J. Hertel, M. J. Beam, M. Fasold, M. Ishiyama, J. Johnson, S. Kehr, M. Lara, J. H. Letaw, G. W. Litman, R. T. Litman, M. Mikami, T. Ota, N. R. Saha, L. Williams, P. F. Stadler, H. Wang, J. S. Taylor, Q. Fontenot, A. Ferrara, S. M. Searle, B. Aken, M. Yandell, I. Schneider, J. A. Yoder, J. N. Volff, A. Meyer, C. T. Amemiya, B. Venkatesh, P. W. Holland, Y. Guiguen, J. Bobe, N. H. Shubin, F. Di Palma, J. Alfoldi, K. Lindblad-Toh, J. H. Postlethwait, The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. *Nat Genet.* **48**, 4 (2016).

26. Y. Uno, C. Nishida, H. Tarui, S. Ishishita, C. Takagi, O. Nishimura, J. Ishijima, H. Ota, A. Kosaka, K. Matsubara, Y. Murakami, S. Kuratani, N. Ueno, K. Agata, Y. Matsuda, Inference of the protokaryotypes of amniotes and tetrapods and the evolutionary processes of microchromosomes from comparative gene mapping. *PLoS One.* **7**, 12 (2012).

27. T. Blomme, K. Vandepoele, S. De Bodt, C. Simillion, S. Maere, Y. Van de Peer, The gain and loss of genes during 600 million years of vertebrate evolution. *Genome Biol.* **7**, 5 (2006).

28. C. L. G. Zhang, Q. Li, B. Li, D. M. Larkin, C. Lee, J. F. Storz, A. Antunes, M. J. Greenwold, R. W. Meredith, A. Ödeen, J. Cui, Q. Zhou, L. Xu, H. Pan, Z. Wang, L. Jin, P. Zhang, H. Hu, W. Yang, J. Hu, J. Xiao, Z. Yang, Y. Liu, Q. Xie, H. Yu, J. Lian, P. Wen, F. Zhang, H. Li, Y. Zeng, Z. Xiong, S. Liu, L. Zhou, Z. Huang, N. An, J. Wang, Q. Zheng, Y. Xiong, G. Wang, B. Wang, J. Wang, Y. Fan, R. R. da Fonseca, A. Alfaro-Núñez, M. Schubert, L. Orlando, T. Mourier, J. T. Howard, G. Ganapathy, A. Pfenning, O. Whitney, M. V. Rivas, E. Hara, J. Smith, M. Farré, J. Narayan, G. Slavov, M. N. Romanov, R. Borges, J. P. Machado, I. Khan, M. S. Springer, J. Gatesy, F. G. Hoffmann, J. C. Opazo, O. Håstad, R. H. Sawyer, H. Kim, K.-W. Kim, H. J. Kim, S. Cho, N. Li, Y. Huang, M. W. Bruford, X. Zhan, A. Dixon, M. F. Bertelsen, E. Derryberry, W. Warren, R. K. Wilson, S. Li, D. A. Ray, R. E. Green, S. J. O'Brien, D. Griffin, W. E. Johnson, D. Haussler, O. A. Ryder, E. Willerslev, G. R. Graves, P. Alström, J. Fjeldså, D. P. Mindell, S. V. Edwards, E. L. Braun, C. Rahbek, D. W. Burt, P. Houde, Y. Zhang, H. Yang, J. Wang, C. Avian Genome, E. D. Jarvis, M. T. P. Gilbert, J. Wang, Comparative genomics reveals insights into avian genome evolution and adaptation. *Science.* **346**, 1311-1320 (2014).

29. D. W. Burt, Origin and evolution of avian microchromosomes. *CGR.* **96**, 97-112 (2002).

30. R. E. O'Connor, L. Kiazim, B. Skinner, G. Fonseka, S. Joseph, R. Jennings, D. M. Larkin, D. K. Griffin, Patterns of microchromosome organization remain highly conserved throughout avian evolution. *Chromosoma.* **128**, 1 (2019).

31. J. M. S. Ohno, C. Stenius, L. Christian, W. A. Kittrell, N. B. Atkin, Microchromosomes in holocephalian, chondrostean and holostean fishes. *Chromosoma.* **26,**, 35-40 (1969).

32. M. E. Aase-Remedios, C. Coll-Llado, D. E. K. Ferrier, More Than One-to-Four via 2R: Evidence of an Independent Amphioxus Expansion and Two-Gene Ancestral Vertebrate State for MyoD-Related Myogenic Regulatory Factors (MRFs). *Mol Biol Evol.* **37**, 10 (2020).

33. S. D. A. J. Pascual-Anaya, S. Kuratani, J. Garcia-Fernàndez, Evolution of Hox gene clusters in deuterostomes. *BMC Developmental Biology.* **13**, 26 (2013).

34. S. J. P. C. T. Amemiya, A. Hill-Force, A. Cook, J. Wasserscheid, D. E. K. Ferrier, J. Pascual-Anaya, J. Garcia-Fernàndez, K. Dewar, P. F. Stadler, The amphioxus Hox cluster: characterization, comparative genomics, and evolution. *J. Exp. Zool. B Mol. Dev. Evol.* **310**, 465-477 (2008).

35. L. Z. Holland, R. Albalat, K. Azumi, E. Benito-Gutierrez, M. J. Blow, M. Bronner-Fraser, F. Brunet, T. Butts, S. Candiani, L. J. Dishaw, D. E. K. Ferrier, J. Garcia-Fernandez, J. J. Gibson-Brown, C. Gissi, A. Godzik, F. Hallbook, D. Hirose, K. Hosomichi, T. Ikuta, H. Inoko, M. Kasahara, J. Kasamatsu, T. Kawashima, A. Kimura, M. Kobayashi, Z. Kozmik, K. Kubokawa, V. Laudet, G. W. Litman, A. C. McHardy, D. Meulemans, M. Nonaka, R. P. Olinski, Z. Pancer, L. A. Pennacchio, M. Pestarino, J. P. Rast, I. Rigoutsos, M. Robinson-Rechavi, G. Roch, H. Saiga, Y. Sasakura, M. Satake, Y. Satou, M. Schubert, N. Sherwood, T. Shiina, N. Takatori, J. Tello, P. Vopalensky, S. Wada, A. Xu, Y. Ye, K. Yoshida, F. Yoshizaki, J. K. Yu, Q. Zhang, C. M. Zmasek, P. J. de Jong, K. Osoegawa, N. H. Putnam, D. S. Rokhsar, N. Satoh, P. W. H. Holland, The amphioxus genome illuminates vertebrate origins and cephalochordate biology. *Genome Research.* **18**, 7 (2008).

36. J. Semyonov, J. I. Park, C. L. Chang, S. Y. Hsu, GPCR genes are preferentially retained after whole genome duplication. *PLoS One.* **3**, 4 (2008).

37. F. G. Brunet, J. N. Volff, M. Schartl, Whole Genome Duplications Shaped the Receptor Tyrosine Kinase Repertoire of Jawed Vertebrates. *Genome Biol Evol.* **8**, 5 (2016).

38. J. T. Li, G. Y. Hou, X. F. Kong, C. Y. Li, J. M. Zeng, H. D. Li, G. B. Xiao, X. M. Li, X. W. Sun, The fate of recent duplicated genes following a fourth-round whole genome duplication in a tetraploid fish, common carp (Cyprinus carpio). *Sci Rep.* **5**, (2015).

39. P. P. Singh, J. Arora, H. Isambert, Identification of Ohnolog Genes Originating from Whole Genome Duplication in Early Vertebrates, Based on Synteny Comparison across Multiple Genomes. *PLoS Comput Biol.* **11**, 7 (2015).

40. C. L. McGrath, J. F. Gout, P. Johri, T. G. Doak, M. Lynch, Differential retention and divergent resolution of duplicate genes following whole-genome duplication. *Genome Res.* **24**, 10 (2014).

41. Y. Ke, Y. Xu, X. Chen, S. Feng, Z. Liu, Y. Sun, X. Yao, F. Li, W. Zhu, L. Gao, H. Chen, Z. Du, W. Xie, X. Xu, X. Huang, J. Liu, 3D Chromatin Structures of Mature Gametes and Structural Reprogramming during Mammalian Embryogenesis. *Cell.* **170**, 2 (2017).

42. C. B. Hug, A. G. Grimaldi, K. Kruse, J. M. Vaquerizas, Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. *Cell.* **169**, 2 (2017).

43. X. Chen, Y. Ke, K. Wu, H. Zhao, Y. Sun, L. Gao, Z. Liu, J. Zhang, W. Tao, Z. Hou, H. Liu, J. Liu, Z. J. Chen, Key role for CTCF in establishing chromatin structure in human embryos. *Nature.* **576**, 7786 (2019).

44. K. Y. Yang, Y. Chen, Z. Zhang, P. K. Ng, W. J. Zhou, Y. Zhang, M. Liu, J. Chen, B. Mao, S. K. Tsui, Transcriptome analysis of different developmental stages of amphioxus reveals dynamic changes of distinct classes of genes during development. *Sci Rep.* **6**, (2016).

45. J. R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, B. Ren, Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature.* **485**, 7398 (2012).

46. S. S. Rao, M. H. Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov, J. T. Robinson, A. L. Sanborn, I. Machol, A. D. Omer, E. S. Lander, E. L. Aiden, A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell.* **159**, 7 (2014).

47. S. Z. C. Wang, J. Chu, G-banding patterns of the chromosomes of amphioxus Branchiostoma
belcheri tsingtauense. *Hereditas.* **141**, 2-7 (2004).

48. C. Shi, X. Wu, L. Su, C. Shang, X. Li, Y. Wang, G. Li, A ZZ/ZW Sex Chromosome System in Cephalochordate Amphioxus. *Genetics.* **214**, 3 (2020).

49. B. Vicoso, Molecular and evolutionary dynamics of animal sex-chromosome turnover. *Nat Ecol Evol.* **3**, 12 (2019).

50. N. Perrin, Sex reversal: a fountain of youth for sex chromosomes? *Evolution.* **63**, 12 (2009).

51. B. T. Lahn, D. C. Page, Four Evolutionary Strata on the Human X Chromosome. *Science.* **286**, 964-967 (1999).

52. G. Li, X. Liu, C. Xing, H. Zhang, S. M. Shimeld, Y. Wang, Cerberus-Nodal-Lefty-Pitx signaling cascade controls left-right asymmetry in amphioxus. *Proc Natl Acad Sci U S A.* **114**, 14 (2017).

53. Q.-J. Zhang, Y. Sun, J. Zhong, G. Li, X.-M. Lü, Y.-Q. Wang, Continuous culture of two lancelets and production of the second filial generations in the laboratory. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution.* **308B**, 4 (2007).

54. T. R. Ranallo-Benavidez, K. S. Jaron, M. C. Schatz, GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun.* **11**, 1 (2020).

55. C. S. Chin, P. Peluso, F. J. Sedlazeck, M. Nattestad, G. T. Concepcion, A. Clum, C. Dunn, R. O'Malley, R. Figueroa-Balderas, A. Morales-Cruz, G. R. Cramer, M. Delledonne, C. Luo, J. R. Ecker, D. Cantu, D. R. Rank, M. C. Schatz, Phased diploid genome assembly with single-molecule real-time sequencing. *Nat Methods.* **13**, 12 (2016).

56. J. Wang, B. J. Walker, T. Abeel, T. Shea, M. Priest, A. Abouelliel, S. Sakthikumar, C. A. Cuomo, Q. Zeng, J. Wortman, S. K. Young, A. M. Earl, Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLoS One.* **9**, 11 (2014).

57. H. Li, Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics.* **34**, 18 (2018).

58. S. Koren, B. P. Walenz, K. Berlin, J. R. Miller, N. H. Bergman, A. M. Phillippy, Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**, 5 (2017).

59. A. P. S. Kurtz, A. L. Delcher, M. Smoot, M. Shumway, C. Antonescu, S. L. Salzberg, Versatile and open software for comparing large genomes. *Genome Biol.* **5**, R12 (2004).

60. N. C. Durand, M. S. Shamim, I. Machol, S. S. Rao, M. H. Huntley, E. S. Lander, E. L. Aiden, Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst.* **3**, 1 (2016).

61. S. S. B. O. Dudchenko, A. D. Omer, S. K. Nyquist, M. Hoeger, N. C. Durand, M. S. Shamim, I. Machol, E. S. Lander, A. P. Aiden, E. L. Aiden, De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science.* **356**, 92-95 (2017).

62. N. C. Durand, J. T. Robinson, M. S. Shamim, I. Machol, J. P. Mesirov, E. S. Lander, E. L. Aiden, Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Syst.* **3**, 1 (2016).

63. D. Zheng, S. P. Gordon, E. Tseng, A. Salamov, J. Zhang, X. Meng, Z. Zhao, D. Kang, J. Underwood, I. V. Grigoriev, M. Figueroa, J. S. Schilling, F. Chen, Z. Wang, Widespread Polycistronic Transcripts in Fungi Revealed by Single-Molecule mRNA Sequencing. *PLoS One.* **10**, 7 (2015).

64. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* **30**, 15 (2014).

65. M. Pertea, G. M. Pertea, C. M. Antonescu, T. C. Chang, J. T. Mendell, S. L. Salzberg, StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol.* **33**, 3 (2015).

66. C. Trapnell, A. Roberts, L. Goff, G. Pertea, D. Kim, D. R. Kelley, H. Pimentel, S. L. Salzberg, J. L. Rinn, L. Pachter, Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc.* **7**, 3 (2012).

67. B. J. Haas, A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M. B. Couger, D. Eccles, B. Li, M. Lieber, M. D. MacManes, M. Ott, J. Orvis, N. Pochet, F. Strozzi, N. Weeks, R. Westerman, T. William, C. N. Dewey, R. Henschel, R. D. LeDuc, N. Friedman, A. Regev, De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc.* **8**, 8 (2013).

68. L. Venturini, S. Caim, G. G. Kaithakottil, D. L. Mapleson, D. Swarbreck, Leveraging multiple transcriptome assembly methods for improved gene structure annotation. *Gigascience.* **7**, 8 (2018).

69. J. M. Flynn, R. Hubley, C. Goubert, J. Rosen, A. G. Clark, C. Feschotte, A. F. Smit, RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci U S A.* **117**, 17 (2020).

70. G. Benson, Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573-580 (1999).

71. Y. Han, S. R. Wessler, MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. *Nucleic Acids Res.* **38**, 22 (2010).

72. B. L. Cantarel, I. Korf, S. M. C. Robb, G. Parra, E. Ross, B. Moore, C. Holt, A. Sanchez Alvarado, M. Yandell, MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research.* **18**, 1 (2007).

73. I. Korf, Gene finding in novel genomes. *BMC Bioinformatics.* **5**, 59 (2004).

74. M. Stanke, O. Schoffmann, B. Morgenstern, S. Waack, Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics.* **7**, (2006).

75. B. J. Haas, S. L. Salzberg, W. Zhu, M. Pertea, J. E. Allen, J. Orvis, O. White, C. R. Buell, J. R. Wortman, Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**, 1 (2008).

76. B. J. Haas, A. L. Delcher, S. M. Mount, J. R. Wortman, R. K. Smith, Jr., L. I. Hannick, R. Maiti, C. M. Ronning, D. B. Rusch, C. D. Town, S. L. Salzberg, O. White, Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* **31**, 19 (2003).

77. P. Jones, D. Binns, H. Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell, G. Nuka, S. Pesseat, A. F. Quinn, A. Sangrador-Vegas, M. Scheremetjew, S. Y. Yong, R. Lopez, S. Hunter, InterProScan 5: genome-scale protein function classification. *Bioinformatics.* **30**, 9 (2014).

78. K. R. B. D. P. Melters, H. A. Young, N. Telis, M. R. May, J. G. Ruby, R. Sebra, P. Peluso, J. Eid, D. Rank, J. F. Garcia, J. L. DeRisi, T. Smith, C. Tobias, J. Ross-Ibarra, I. Korf, S. W. L. Chan, Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biol.* **14**, R10 (2013).

79. J. D. Liu, M. S. Yi, G. Zhao, F. Zhou, D. Q. Wang, Q. X. Yu, Sex chromosomes in the spiny eel <i>(Mastacembelus aculeatus)</i> revealed by mitotic and meiotic analysis. *Cytogenetic and Genome Research.* **98**, 4 (2002).

80. P. Danecek, A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin, The variant call format and VCFtools. *Bioinformatics.* **27**, 15 (2011).

81. A. B. Sahakyan, V. S. Chambers, G. Marsico, T. Santner, M. Di Antonio, S. Balasubramanian, Machine learning model for sequence-driven DNA G-quadruplex formation. *Scientific Reports.* **7**, 1 (2017).

82. D. M. Emms, S. Kelly, OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology.* **20**, 1 (2019).

83. S. M. Kielbasa, R. Wan, K. Sato, P. Horton, M. C. Frith, Adaptive seeds tame genomic sequence comparison. *Genome Res.* **21**, 3 (2011).

84. W. J. K. M. Blanchette, C. Riemer, L. Elnitski, A. F. A. Smit, K. M. Roskin, R. Baertsch, K. Rosenbloom, H. Clawson, E. D. Green, D. Haussler, W. Miller, Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res.* **14**, 708-715 (2004).

85. Y. Satou, R. Nakamura, D. Yu, R. Yoshida, M. Hamada, M. Fujie, K. Hisata, H. Takeda, N. Satoh, R. O'Neill, A Nearly Complete Genome of Ciona intestinalis Type A (C. robusta) Reveals the Contribution of Inversion to Chromosomal Evolution in the Genus Ciona. *Genome Biology and Evolution.* **11**, 11 (2019).

86. S. Wang, J. Zhang, W. Jiao, J. Li, X. Xun, Y. Sun, X. Guo, P. Huan, B. Dong, L. Zhang, X. Hu, X. Sun, J. Wang, C. Zhao, Y. Wang, D. Wang, X. Huang, R. Wang, J. Lv, Y. Li, Z. Zhang, B. Liu, W. Lu, Y. Hui, J. Liang, Z. Zhou, R. Hou, X. Li, Y. Liu, H. Li, X. Ning, Y. Lin, L. Zhao, Q. Xing, J. Dou, Y. Li, J. Mao, H. Guo, H. Dou, T. Li, C. Mu, W. Jiang, Q. Fu, X. Fu, Y. Miao, J. Liu, Q. Yu, R. Li, H. Liao, X. Li, Y. Kong, Z. Jiang, D. Chourrout, R. Li, Z. Bao, Scallop genome provides insights into evolution of bilaterian karyotype and development. *Nature Ecology & Evolution.* **1**, 5 (2017).

87. H. A. S. B. Q. Minh, O. Chernomor, D. Schrempf, M. D. Woodhams, A. von Haeseler, IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution.* **37**, 5 (2020).

88. G. B. A. Siepel, J. S. Pedersen, A. S. Hinrichs, M. Hou, K. Rosenbloom, H. Clawson, J. Spieth, L. W. Hillier, S. Richards, G. M. Weinstock, R. K. Wilson, R. A. Gibbs, W. J. Kent, W. Miller, D. Haussler, Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034-1050 (2005).

89. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* **26**, 6 (2010).

90. F. Cabanettes, C. Klopp, D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ.* **6**, (2018).

91. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* **30**, 4 (2013).

92. I. Sela, H. Ashkenazy, K. Katoh, T. Pupko, GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. *Nucleic Acids Research.* **43**, W1 (2015).

93. Z. Yang, PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* **24**, 8 (2007).

94. P. C. J. D. M. J. Benton, R. J. Asher, "Calibrating and constraining molecular clocks" in The timetree of Life. *Hedges, S. B, S. Kumar, Eds. (Oxford University Press, United Kingdom).* (2009).

95. Z. Yang, B. Rannala, Bayesian Estimation of Species Divergence Times Under a Molecular Clock Using Multiple Fossil Calibrations with Soft Bounds. *Molecular Biology and Evolution.* **23**, 1 (2006).

96. M. J. Benton, P. C. Donoghue, Paleontological evidence to date the tree of life. *Mol Biol Evol.* **24**, 1 (2007).

97. L. Pu, Y. Lin, P. A. Pevzner, Detection and analysis of ancient segmental duplications in mammalian genomes. *Genome Res.* **28**, 6 (2018).

98. D. D. K. Chen, M. Farach-Colton, NOTUNG: a program for dating gene duplications and optimizing gene family trees. *J. Comput. Biol.* **7**, 429-447 (2000).

99. S. Capella-Gutierrez, J. M. Silla-Martinez, T. Gabaldon, trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics.* **25**, 15 (2009).

100. J. Shi, X. Ma, J. Zhang, Y. Zhou, M. Liu, L. Huang, S. Sun, X. Zhang, X. Gao, W. Zhan, P. Li, L. Wang, P. Lu, H. Zhao, W. Song, J. Lai, Chromosome conformation capture resolved near complete genome assembly of broomcorn millet. *Nature Communications.* **10**, 1 (2019).

101. C. Marchal, N. Singh, X. Corso-Díaz, A. Swaroop, HiCRes: a computational method to estimate and predict the resolution of HiC libraries 2. *bioRxiv.* (2020).

102. F. Ramírez, V. Bhardwaj, L. Arrigoni, K. C. Lam, B. A. Grüning, J. Villaveces, B. Habermann, A. Akhtar, T. Manke, High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nature Communications.* **9**, 1 (2018).

103. C. E. Grant, T. L. Bailey, W. S. Noble, FIMO: scanning for occurrences of a given motif. *Bioinformatics.* **27**, 7 (2011).

104. M. J. Rowley, M. H. Nichols, X. Lyu, M. Ando-Kuri, I. S. M. Rivera, K. Hermetz, P. Wang, Y. Ruan, V. G. Corces, Evolutionarily Conserved Principles Predict 3D Chromatin Organization. *Mol Cell.* **67**, 5 (2017).

105. G. Li, J. Feng, Y. Lei, J. Wang, H. Wang, L. K. Shang, D. T. Liu, H. Zhao, Y. Zhu, Y. Q. Wang, Mutagenesis at specific genomic loci of amphioxus Branchiostoma belcheri using TALEN method. *J Genet Genomics.* **41**, 4 (2014).

106. M. A. DePristo, E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, G. del Angel, M. A. Rivas, M. Hanna, A. McKenna, T. J. Fennell, A. M. Kernytsky, A. Y. Sivachenko, K. Cibulskis, S. B. Gabriel, D. Altshuler, M. J. Daly, A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* **43**, 5 (2011).

107. “Picard Toolkit.”. *Broad Institute, GitHub Repository. http://broadinstitute.github.io/picard/.* (2019).

108. S. R. Browning, B. L. Browning, Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet.* **81**, 5 (2007).

109. O. Delaneau, J.-F. Zagury, J. Marchini, Improved whole-chromosome phasing for disease and population genetic studies. *Nature Methods.* **10**, 1 (2013).

110. H. M. Kang, J. H. Sul, S. K. Service, N. A. Zaitlen, S.-y. Kong, N. B. Freimer, C. Sabatti, E. Eskin, Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics.* **42**, 4 (2010).

111. Y. Liao, G. K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics.* **30**, 7 (2013).

112. S. Mawaribuchi, Y. Ito, M. Ito, Independent evolution for sex determination and differentiation in the DMRT family in animals. *Biol Open.* **8**, 8 (2019).

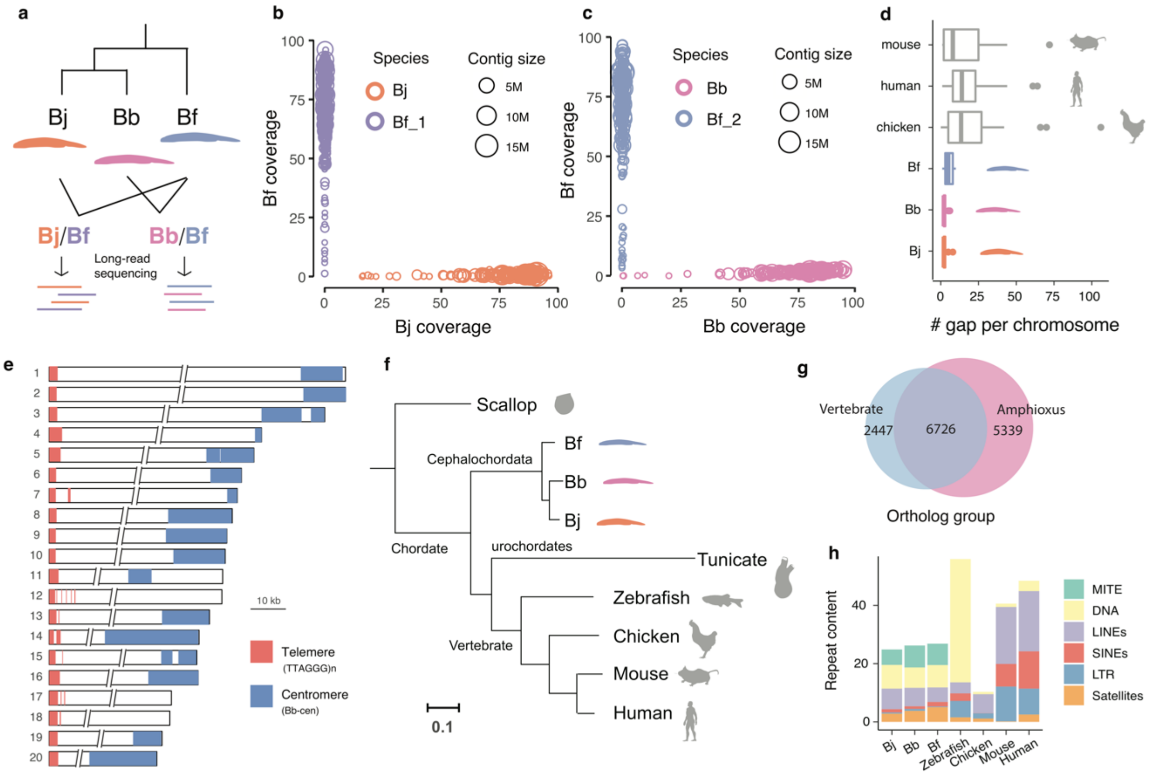


Figure 1 Three haploid genomes of amphioxus species. a) We performed long-read sequencing of interspecific hybrids between the three amphioxus species and assembled their haploid genomes. Bj: *B. japonicum*, Bb: *B. belcheri*, Bf: *B. floridae*. **b-c)** Contig sequences of the hybrids were assigned to the haploid genome of each parental species, according to their coverage (proportion of mapped sequences) mapped by the short-reads of parental species. **d)** The amphioxus haploid genomes have a lower gap content (numbers of gaps per chromosome) compared to other vertebrate reference genomes. **e)** Most amphioxus chromosomes are telocentric. The 10kb scale applies to the two tips of the chromosomes only, and the two slash lines represent the gaps between the two chromosomal tips. **f)** Phylogenomic tree based on whole-genome alignments of amphioxus vs. other chordate species. **g)** A large number of orthologous gene groups (6726) is shared between amphioxus and vertebrates, but amphioxus species have 5,339 specific gene groups. **h)** MITEs (green) comprise ~6.7% of the amphioxus genomes but are largely absent in vertebrates. In the DNA transposon category (yellow) MITE was excluded.

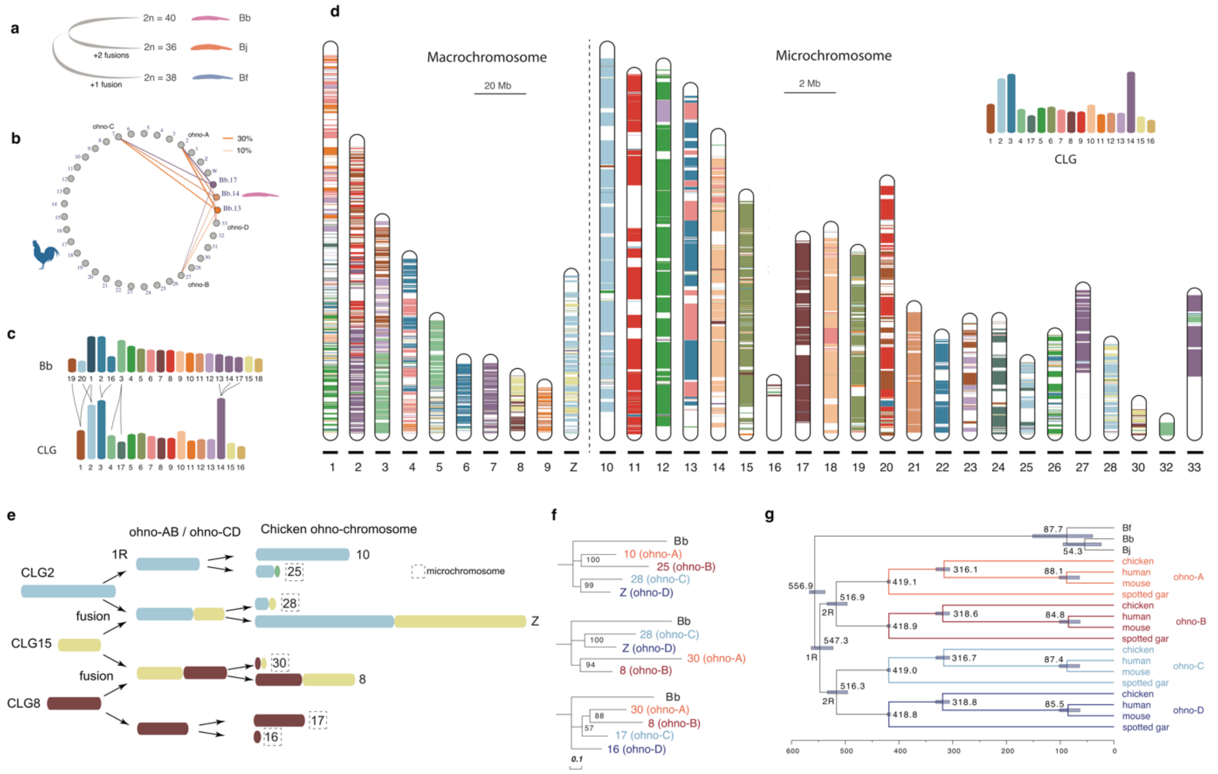


Figure 2 Ancestral karyotypes of amphioxus, chordates and vertebrates.

a) Bb probably best recapitulates the ancestral karyotype of *Branchiostoma* amphioxus, with Bj and Bf having undergone chromosomal fusions. **b)** Genes on chr13, chr14 and chr17 of Bb have their homologous genes located on the same set of chicken chromosomes. Each line connecting chromosomes of Bb and chicken chromosomes is scaled to the proportion of Bb genes that are homologous to the genes of one chicken chromosome. **c)** The inferred relationship between Bb chromosomes and CLG. **d)** Composition of chicken chromosome by CLG homologous sequences. The colored bands represent the Bb-chicken synteny blocks. A different scale for macrochromosomes (20 Mb) and microchromosomes (2 Mb) was used. **e)** Reconstructed 1R and 2R of three CLGs. One color represents one CLG, and when one chromosome is composed with more than one CLG, two or more CLG blocks are linked together. **f)** The ohnolog genes were used to construct the phylogeny of ohno-chromosomes (ohno-A, B, C, D), which refer to gene groups derived from WGDs. Bb homologs were used as the outgroup. Bootstrapping values shown placed at the internal nodes. **g)** 244 ohnolog gene groups were used to date 1R and 2R. Fossil calibration for the mouse-human nodes: 62-101 MY, bird-mammal nodes: 306-332 MY.

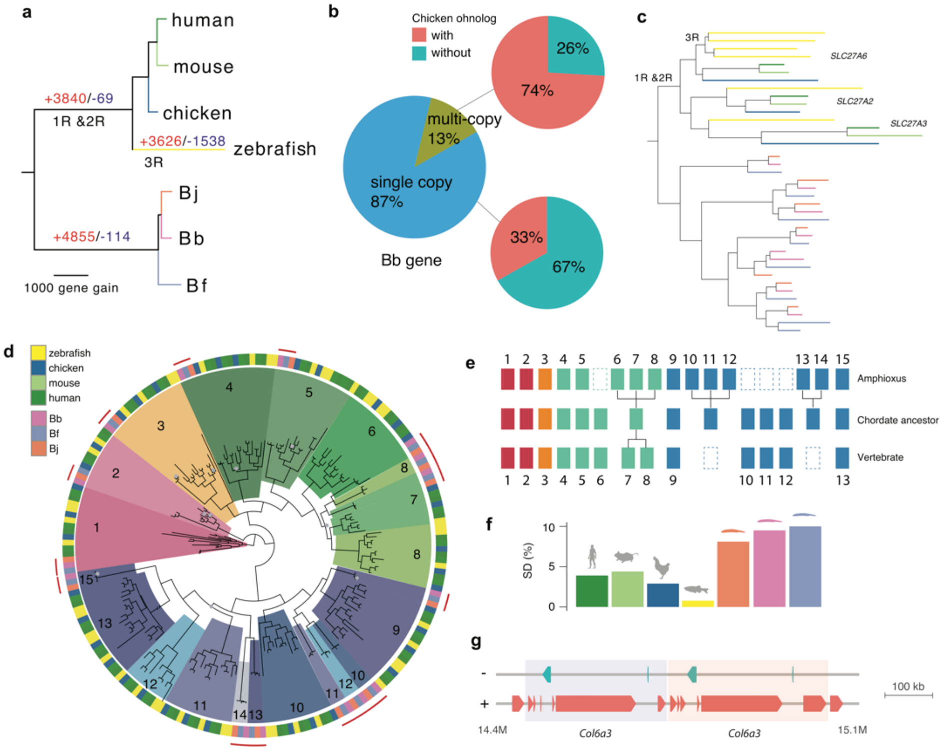


Figure 3 Gene expansion in amphioxus species

a) Reconstructed gene gains (red) and losses (blue) events during the chordate evolution based on the ortholog gene groups. The branch length is scaled to the number of gene gain. **b)** Using Bb as an example, we show duplicated genes in amphioxus more frequently have paralogs in vertebrates. A majority (74%) of the Bb multi-copy genes have chicken paralogs compared with only 33% of Bb single-copy genes. **c)** Independent expansion of *SLC27A* gene copies in vertebrates (due to WGD) and amphioxus (due to gene duplication). Each species (zebrafish and three amphioxus species) is marked with the same color as shown in a). **d)** Phylogenetic tree of *Hox* genes. The homologous *Hox* gene (denoted by the number) group of amphioxus and vertebrates was marked in the same color. The grey dots at the internal nodes indicate a bootstrapping value lower than 60. **e)** An inferred model of *Hox* gene evolution in chordates according to the results of d). Dashed boxes denote gene loss, each aligned column denotes homologous relationship, individual gene duplications are also shown for either amphioxus or vertebrates. **f)** Amphioxus has a higher portion of genome derived from segmental duplication compared to vertebrates **g)** One example of segmental duplication involving *Col6a3* in Bf. The two copies are next to each other highlighted in different background colors

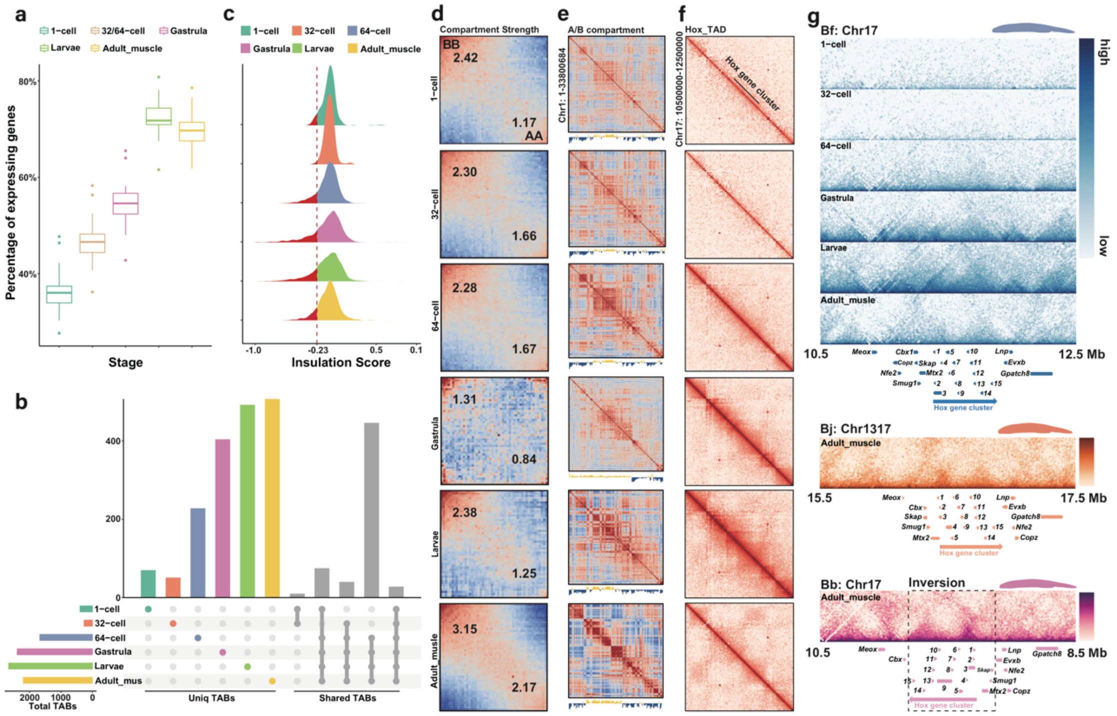


Figure 4 Developmental dynamics of amphioxus chromatin architecture

a) The percentage of actively transcribed genes (TPM>1) across five developmental stages of 1-cell zygote, 32/64-cell, gastrula, larvae, and adult muscle tissues of Bf. **b)** The number of TAD boundaries (TABs) at 5kb resolution across six developmental stages of Bf. The horizontal bars show the number of TABs of each stage. The vertical colored bars show the number of specific TABs of each stage, and the grey bars show the number of shared TABs among 6 stages. **c)** The distribution of insulation scores of TABs across different stages. The smaller the insulation score is, the higher strength the TAB has. **d)** Saddle plots of amphioxus Hi-C data binned at 250kb resolution at six different developmental stages. Bins are sorted by their PC1 value. B-B (inactive-inactive) interactions are in the upper left corner, and preferential A-A (active-active) interactions are in the lower right corner. Numbers in the corners show the strength of AA interactions as compared to AB interaction and BB interactions against BA interactions. **e)** Correlation matrix and eigenvector 1(PC1) values value tracks for amphioxus chromosome 1 at 250kb resolution at six different developmental stages. **f)** Distribution of interaction at 15kb resolution at the Bf *Hox* cluster. **g)** Distribution of TADs at the 15kb resolution in three different amphioxus *Hox* regions with the gene tracks.

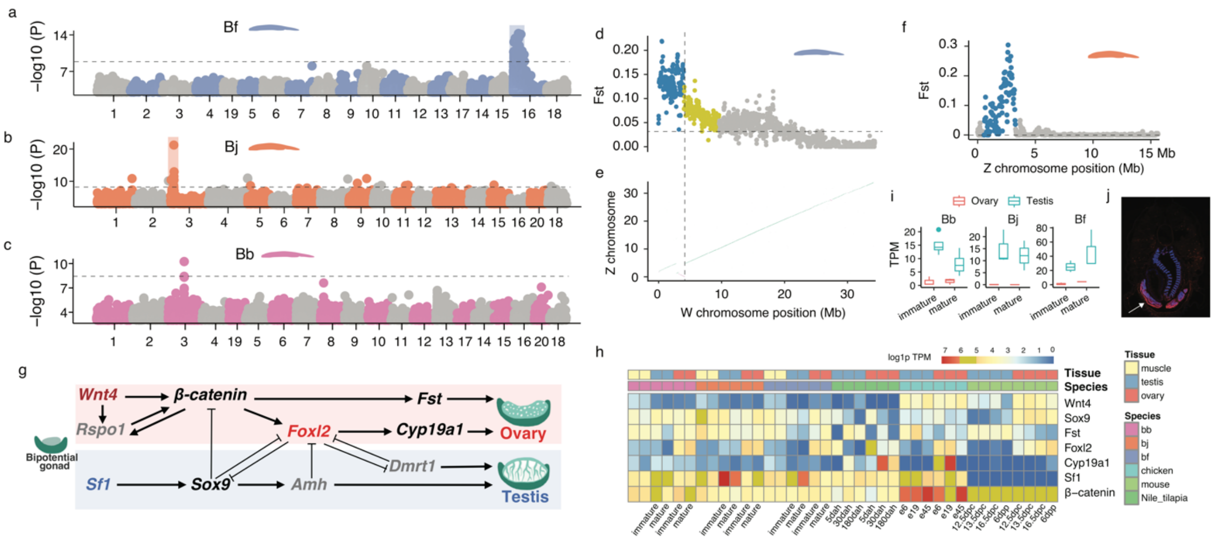
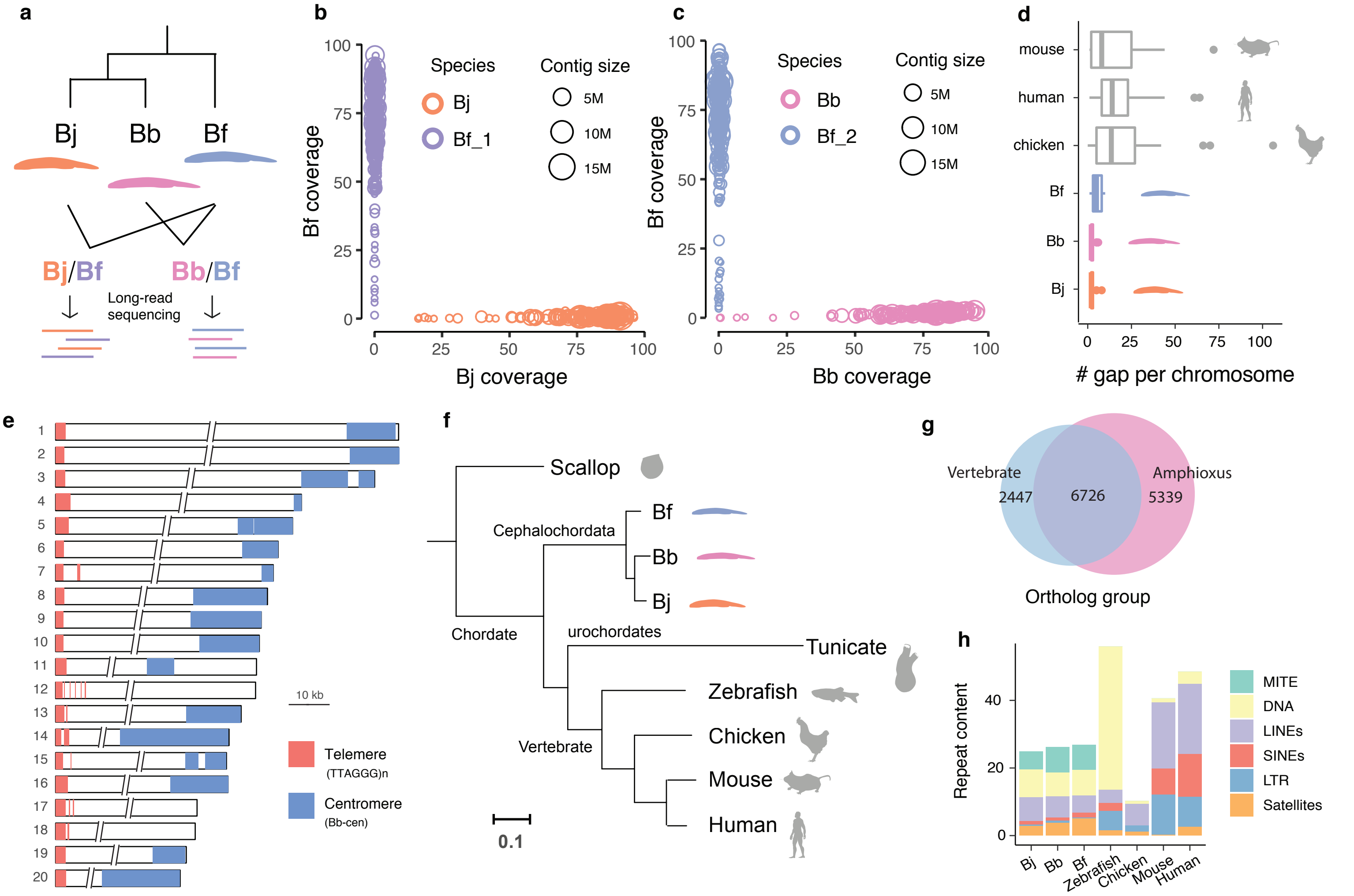
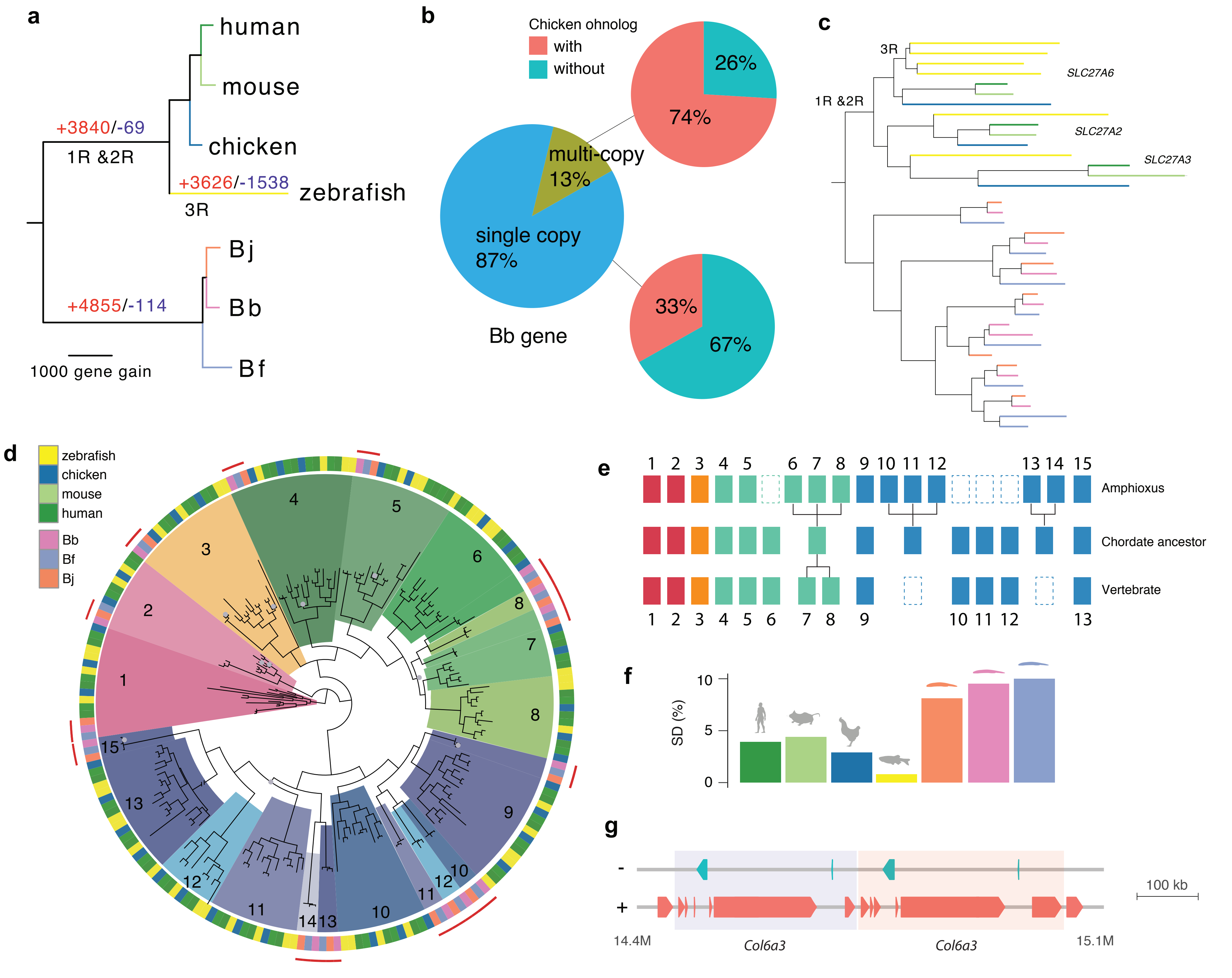


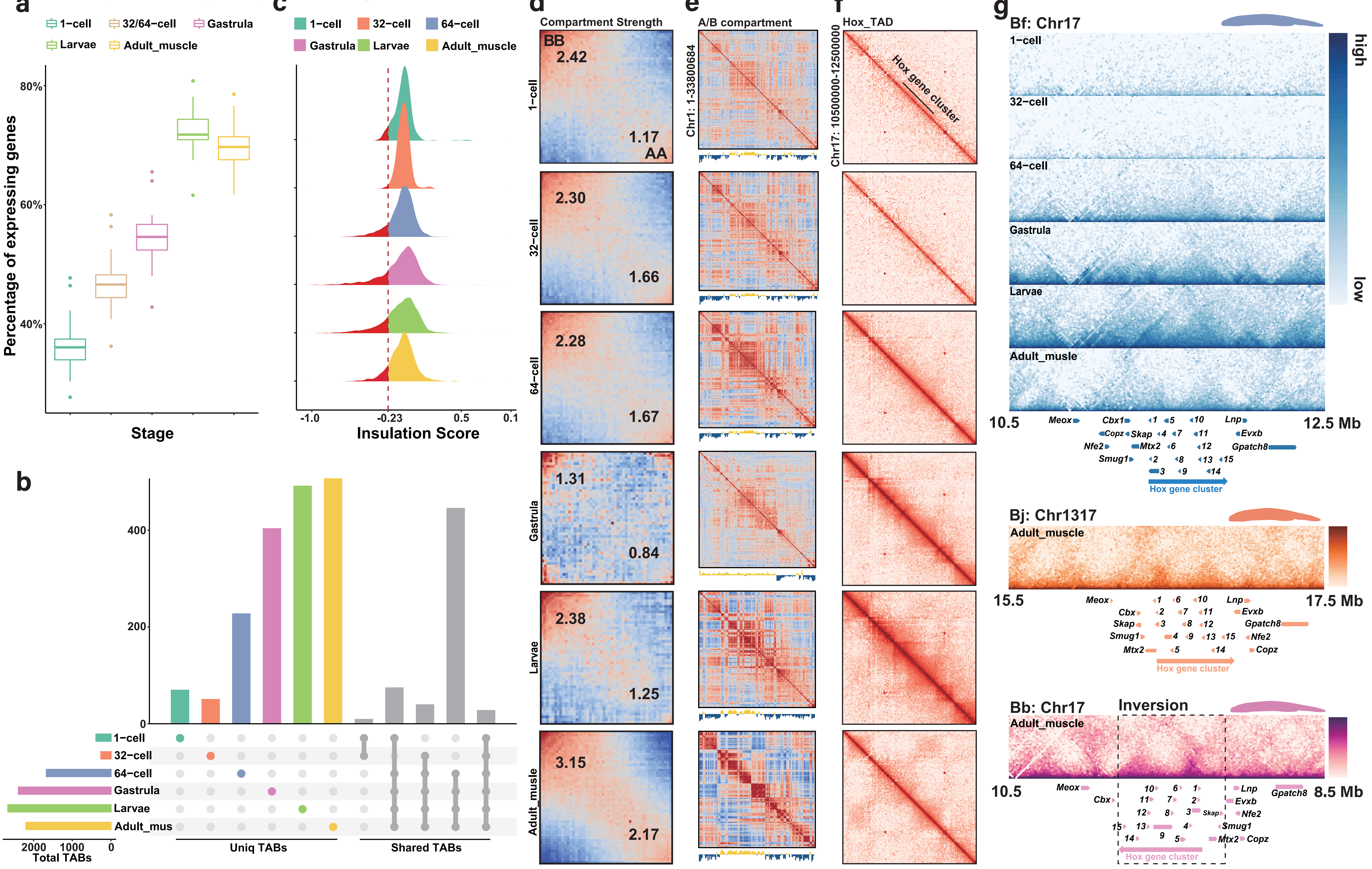
Figure 5 Turnovers of sexually differentiated regions between amphioxus species

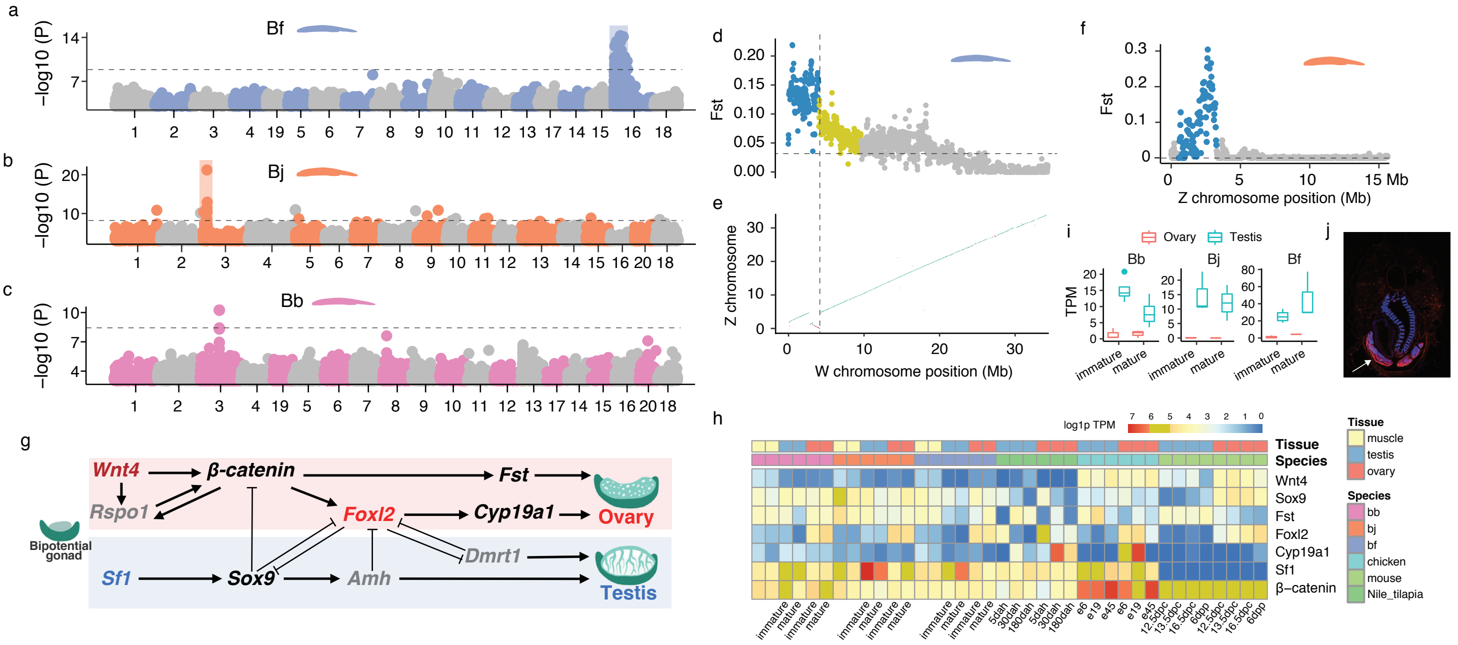
a-c) Genome-wide association study (GWAS) identified the sex-linked regions in amphioxus. The Y axis shows the log10 transformed p-value of GWAS. **d)** The F_{ST} statistics between male and female populations of Bf reveal the evolutionary strata. Each dot represents a 50 kb sliding window. The horizontal dashed lines show the genomic average levels. **e)** The synteny plot between the Z and W chromosomes of Bf. The purple lines represent reversed alignments. The vertical dashed line indicates the boundary of stratum 1 as well as the inversion. **f)** The F_{ST} statistics between male and female populations of Bj. **g)** The 10 conserved vertebrate SD pathway genes, genes in grey are absent in amphioxus. Only *Foxl2* and *Sf1* are sex-biased in amphioxus. **h)** The expression profiles of chordate SD-related genes over developing gonads. **i)** The candidate Bb SD gene has a conserved testis-biased expression. **j)** RNA fluorescence *in situ* hybridization shows the candidate Bb SD gene has a specific expression in testis


a
Bj
Bb
Bf
Bj/Bf
Bb/Bf
Long-read
sequencing
b
Bf coverage
Bj coverage
Species
Bj
Bf_1
Contig size
5M
10M
15M
0
25
50
75
100
c
Bf coverage
Bb coverage
Species
Bb
Bf_2
Contig size
5M
10M
15M
d
mouse
human
chicken
Bf
Bb
Bj
# gap per chromosome
0
25
50
75
100
e
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
10 kb
Telemere
(TTAGGG)n
Centromere
(Bb-cen)
f
Scallop
Cephalochordata
Bf
Bb
Bj
Chordate
urochordates
Tunicate
Zebrafish
Chicken
Mouse
Human
Vertebrate
0.1
g
Vertebrate
2447
6726
Amphioxus
5339
Ortholog group
h
Repeat content
0
20
40
Bj
Bb
Bf
Zebrafish
Chicken
Mouse
Human
MITE
DNA
LINEs
SINEs
LTR
Satellites

a
2n = 40
Bb
2n = 36
Bj
+2 fusions
2n = 38
Bf
+1 fusion
b
ohno-C
ohno-A
ohno-D
ohno-B
30%
10%
Bb.17
Bb.14
Bb.13
Z
W
c
Bb
19 20 1 2 16 3 4 5 6 7 8 9 10 11 12 13 14 17 15 18
CLG
1 2 3 4 17 5 6 7 8 9 10 11 12 13 14 15 16
d
Macrochromosome
20 Mb
Microchromosome
2 Mb
1 2 3 4 17 5 6 7 8 9 10 11 12 13 14 15 16
CLG
1 2 3 4 5 6 7 8 9 Z 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 30 32 33
e
ohno-AB / ohno-CD
Chicken ohno-chromosome
1R
CLG2
10
25
microchromosome
fusion
28
Z
CLG15
fusion
30
8
CLG8
17
16
f
Bb
100
10 (ohno-A)
25 (ohno-B)
99
28 (ohno-C)
Z (ohno-D)
Bb
100
28 (ohno-C)
Z (ohno-D)
30 (ohno-A)
94
8 (ohno-B)
Bb
88
30 (ohno-A)
57
8 (ohno-B)
17 (ohno-C)
16 (ohno-D)
0.1
g
87.7
Bf
Bb
Bj
54.3
556.9
316.1
88.1
chicken
human
mouse
spotted gar
ohno-A
419.1
516.9
2R
318.6
84.8
ohno-B
418.9
547.3
1R
316.7
87.4
ohno-C
419.0
516.3
2R
318.8
85.5
ohno-D
418.8
600
500
400
300
200
100
0


a
human
mouse
chicken
+3840/-69
1R &2R
+3626/-1538
zebrafish
3R
Bj
Bb
+4855/-114
Bf
1000 gene gain
b
Chicken ohnolog
with
without
26%
74%
multi-copy
13%
single copy
87%
Bb gene
33%
67%
c
3R
1R &2R
SLC27A6
SLC27A2
SLC27A3
d
zebrafish
chicken
mouse
human
Bb
Bf
Bj
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
e
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Amphioxus
Chordate ancestor
Vertebrate
1 2 3 4 5 6 7 8 9 10 11 12 13
f
SD (%)
0
5
10
g
-
+
100 kb
14.4M
Col6a3
Col6a3
15.1M


a
1−cell
32/64−cell
Gastrula
Larvae
Adult_muscle
Percentage of expressing genes
80%
60%
40%
Stage
c
1−cell
32−cell
64−cell
Gastrula
Larvae
Adult_muscle
-1.0
-0.23
0.5
0.1
Insulation Score
d
Compartment Strength
BB
AA
1−cell
2.42
1.17
32−cell
2.30
1.66
64−cell
2.28
1.67
Gastrula
1.31
0.84
Larvae
2.38
1.25
Adult_musle
3.15
2.17
e
A/B compartment
Chr1: 1-33800684
f
Hox_TAD
Chr17: 10500000-12500000
Hox gene cluster
g
Bf: Chr17
1−cell
32−cell
64−cell
Gastrula
Larvae
Adult_musle
high
low
10.5
12.5 Mb
Meox
Cbx1
1
5
10
Lnp
Copz
Skap
4
7
11
Evxb
Nfe2
Mtx2
6
12
Gpatch8
Smug1
2
8
13
15
3
9
14
Hox gene cluster
Bj: Chr1317
Adult_muscle
15.5
17.5 Mb
Meox
1
6
10
Lnp
Cbx
2
7
11
Evxb
Skap
3
8
12
Gpatch8
Smug1
4
9
13
15
Nfe2
Mtx2
5
14
Copz
Hox gene cluster
Bb: Chr17
Inversion
Adult_muscle
10.5
8.5 Mb
Meox
10
6
1
Lnp
Gpatch8
Cbx
11
7
2
Evxb
12
8
3
Skap
Nfe2
15
13
4
Smug1
14
9
5
Mtx2
Copz
Hox gene cluster
b
400
200
0
1−cell
32−cell
64−cell
Gastrula
Larvae
Adult_mus
2000
1000
0
Total TABs
Uniq TABs
Shared TABs


a
Bf
−log10 (P)
14
7
1 2 3 4 19 5 6 7 8 9 10 11 12 13 17 14 15 16 18
b
Bj
−log10 (P)
20
10
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 20 18
c
Bb
−log10 (P)
10
7
4
1 2 3 4 19 5 6 7 8 9 10 11 12 13 17 14 15 16 20 18
d
Fst
0.20
0.15
0.10
0.05
0.00
e
Z chromosome
30
20
10
0
0 10 20 30
W chromosome position (Mb)
f
Fst
0.3
0.2
0.1
0
0 5 10 15 Mb
Z chromosome position (Mb)
i
Ovary
Testis
TPM
Bb
Bj
Bf
20 15 10 5 0
20 15 10 5 0
80 60 40 20 0
immature mature
immature mature
immature mature
j
g
Wnt4
β-catenin
Fst
Rspo1
Foxl2
Cyp19a1
Ovary
Bipotential gonad
Dmrt1
Sf1
Sox9
Amh
Testis
h
log1p TPM
7 6 5 4 3 2 1 0
Tissue
Species
Wnt4
Sox9
Fst
Foxl2
Cyp19a1
Sf1
β−catenin
immature mature immature mature immature mature immature mature immature mature immature mature 5dah 30dah 180dah 5dah 30dah 180dah e6 e19 e45 e6 e19 e45 12.5dpc 13.5dpc 16.5dpc 6dpp 12.5dpc 13.5dpc 16.5dpc 6dpp
Tissue
muscle
testis
ovary
Species
bb
bj
bf
chicken
mouse
Nile_tilapia