PNAS

Supporting Information for
Three amphioxus reference genomes reveal gene and chromosome evolution of chordates

Zhen Huang[1,2,*], Luohao Xu[3,4,5,*,†], Cheng Cai[6,*], Yitao Zhou[1,7,*], Jing Liu[5], Zaoxu Xu[3,4], Zexian Zhu[6], Wen Kang[6], Wab Cen[7], Surui Pei[8], Duo Chen[1,7,9], Chenggang Shi[10], Xiaotong Wu[10], Yongji Huang[11], Chaohua Xu[1], Yanan Yan[1], Ying Yang[1], Ting Xue[1,7,9], Wenjing He[1], Xuefeng Hu[1], Yanding Zhang[1], Youqiang Chen[1,7,9], Changwei Bi[12], Chunpeng He[12], Lingzhan Xue[13], Shijun Xiao[14], Zhicao Yue[15], Yu Jiang[8], Jr-Kai Yu[16,17], Erich D. Jarvis[18,19], Guang Li[10], Gang Lin[1,7,9,†], Qiujin Zhang[1,7,9,†], Qi Zhou[6,5,20,21,†]

1. Fujian Key Laboratory of Special Marine Bio-resources Sustainable Utilization & Fujian Key Laboratory of Developmental and Neurobiology, College of Life Sciences, Fujian Normal University, Fuzhou 350117, Fujian
2. Fujian-Macao Science and Technology Cooperation Base of Traditional Chinese Medicine-Oriented Chronic Disease Prevention and Treatment, Innovation and Transformation Center, Fujian University of Traditional Chinese Medicine, Fuzhou 350108, Fujian
3.Key Laboratory of Freshwater Fish Reproduction and Development (Ministry of Education), Key Laboratory of Aquatic Science of Chongqing, School of Life Sciences, Southwest University, Chongqing 400715, Chongqing
4. Integrative Science Center of Germplasm Creation in Western China (CHONGQING) Science City, Chongqing
5. Department of Neuroscience and Developmental Biology, University of Vienna, Vienna 1090, Vienna
6. 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
7. 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
8. Annoroad Gene Technology Co., Ltd, Beijing 100180, Beijing
9. Center of Engineering Technology Research for Microalgae Germplasm Improvement of Fujian, Southern Institute of Oceanography, Fujian Normal University, Fuzhou 350117, Fujian

10. State Key Laboratory of Cellular Stress Biology, School of Life Sciences, Xiamen University, Xiamen, Fujian 361102, Fujian
11. Institute of Oceanography, Minjiang University, Fuzhou 350108, Fujian
12. State Key Laboratory of Bioelectronics, School of Biological Science and Medical Engineering, Southeast University, Nanjing 210096, Jiangsu
13. Aquaculture and Genetic breeding laboratory, Freshwater Fisheries Research Institute of Fujian, Fuzhou 350002, Fujian
14. College of Plant Protection, Jilin Agricultural University, Changchun 130118, Jilin
15. 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, Shenzhen 518060, Guangdong
16. Institute of Cellular and Organismic Biology, Academia Sinica, Taipei 11529, Taiwan
17. Marine Research Station, Institute of Cellular and Organismic Biology, Academia Sinica, Yilan 26242, Taiwan
18. Laboratory of Neurogenetics of Language, The Rockefeller University, New York 10065, New York
19. Howard Hughes Medical Institute, Chevy Chase, Maryland 20815, Maryland
20. Center for Reproductive Medicine, The 2nd Affiliated Hospital, School of Medicine, Hangzhou 310052, Zhejiang
21. Evolutionary & Organismal Biology Research Center, School of Medicine, Zhejiang University, Hangzhou 310058, Zhejiang

Luohao Xu: luohaox@gmail.com
Gang Lin: lgffz@fjnu.edu.cn
Qiujing Zhang: qjzhang@fjnu.edu.cn
Qi Zhou: zhouqi1982@zju.edu.cn

This PDF file includes:

Figures S1 to S45
Tables S1 to S9
Legends for Datasets S1
SI References

Other supporting materials for this manuscript include the following:

Datasets S1

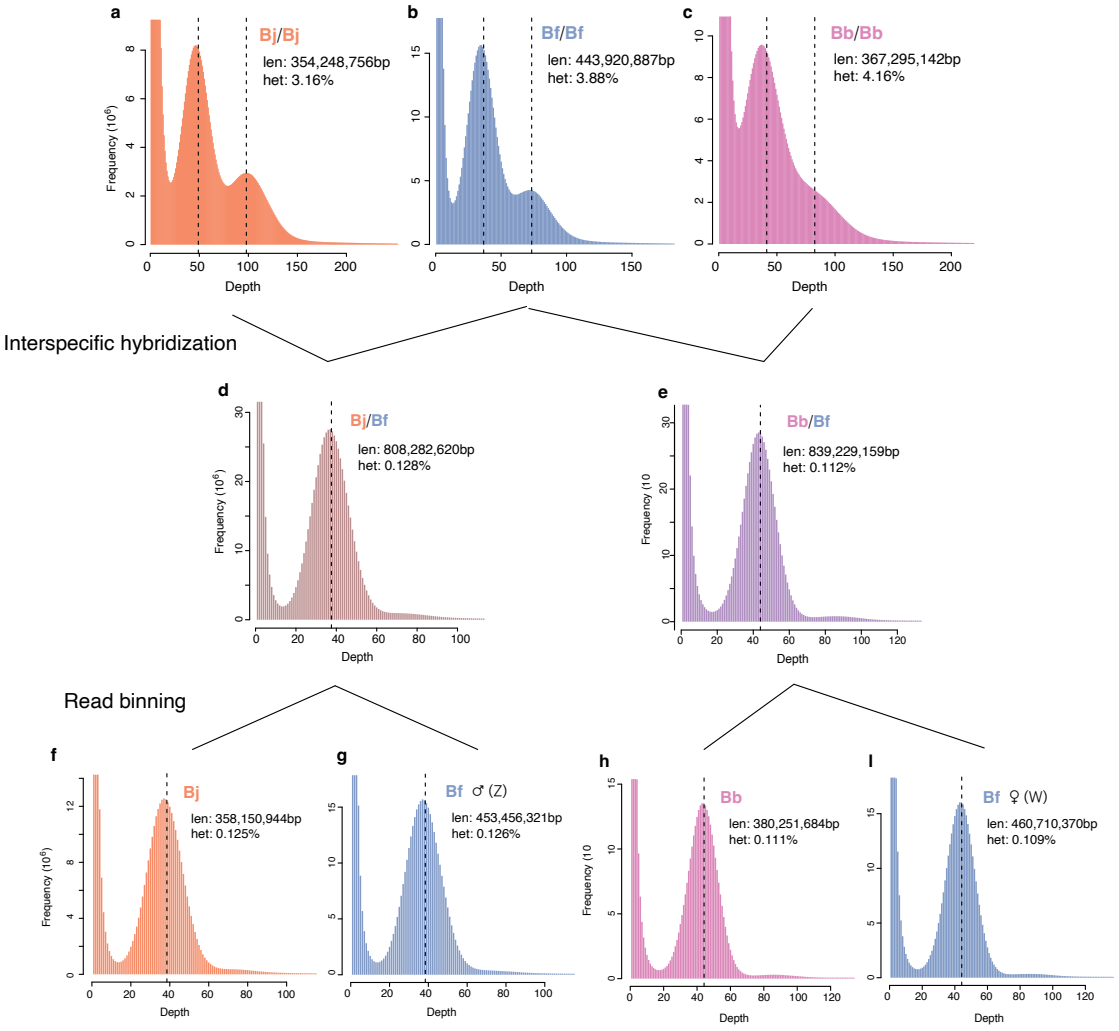


Fig. S1. Estimation of genome size and heterozygosity levels for pure species and the hybrids.

K-mer size of 21 was used. The distribution of k-mer counts were taken as input for genomeScope to estimate the genome size and heterozygosity of the parental species (a-c) and their hybrids (d-e). The parental species have one extra peak at the half-coverage level, suggesting the high levels of heterozygosity, while the hybrids have one single peak, suggesting the interspecific hybrid genomes are effectively haploid. After partitioning the species-specific reads from the sequencing data of the hybrids (f-i), the species-specific reads show a single haploid peak, the estimated genome size represents the haploid genome size of the pure species.

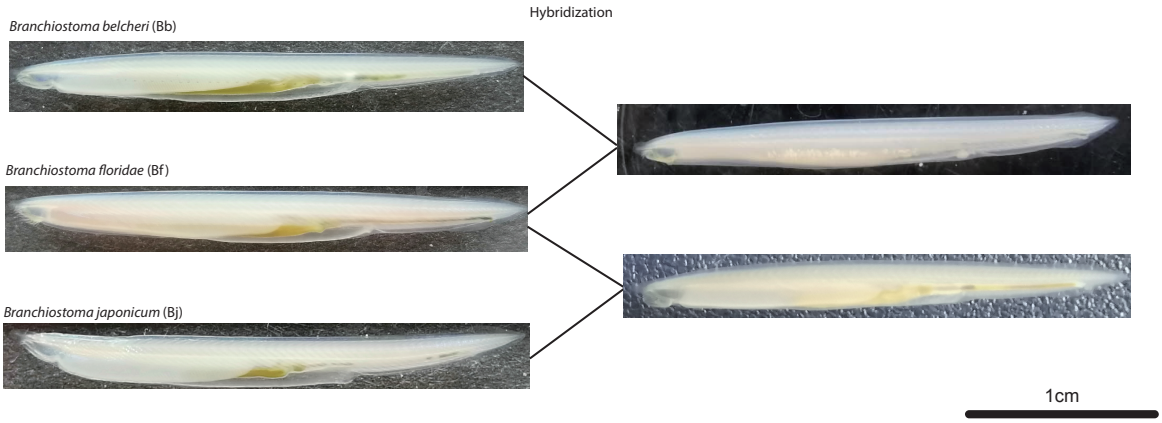


Fig. S2. The pure species of three amphioxus species and their hybrids.
Photo credit: Zhen Huang, Fujian Normal University

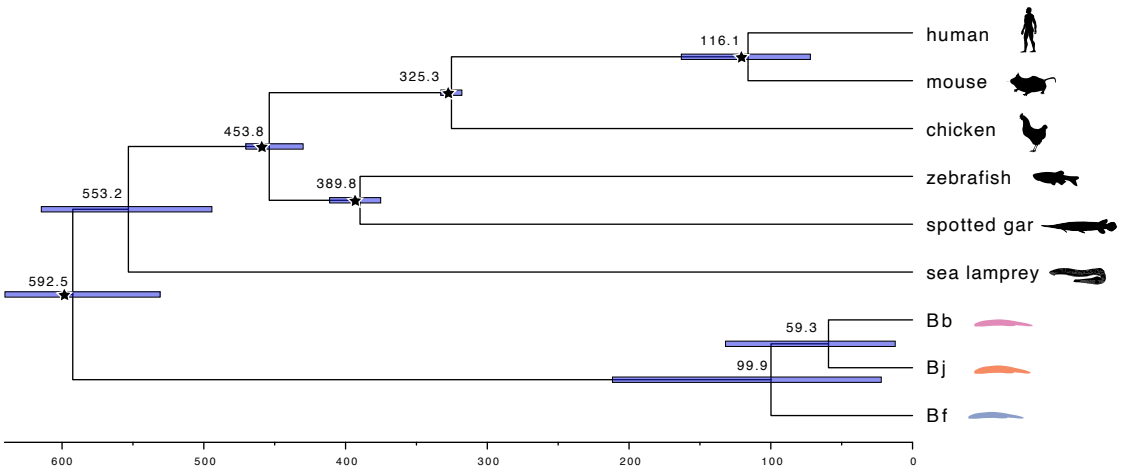


Fig. S3. Molecular dating of amphioxus species.

The 4-fold degenerate (4D) sites were extracted from the whole genome alignments of four vertebrates and three amphioxus species, according to the annotations of human gene models. The bars show the 95% confidence interval of divergence time. The black asterisks represent the nodes whose divergence time have been calibrated with fossil records.

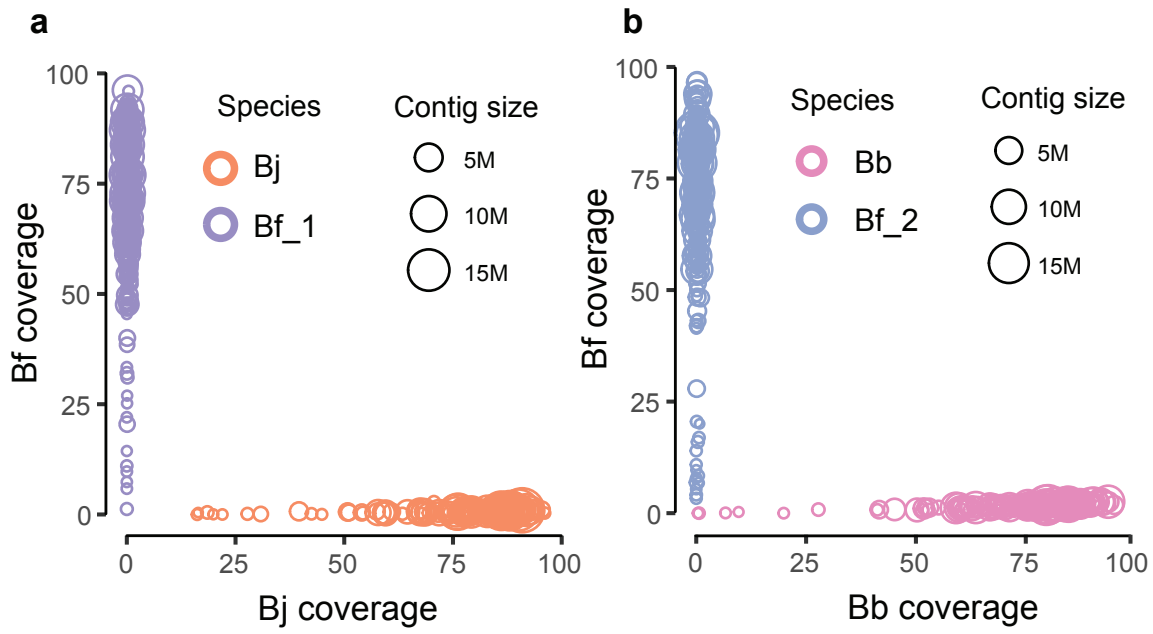


Fig. S4. Identifying species-specific contigs from genome assembly

a-b) 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.

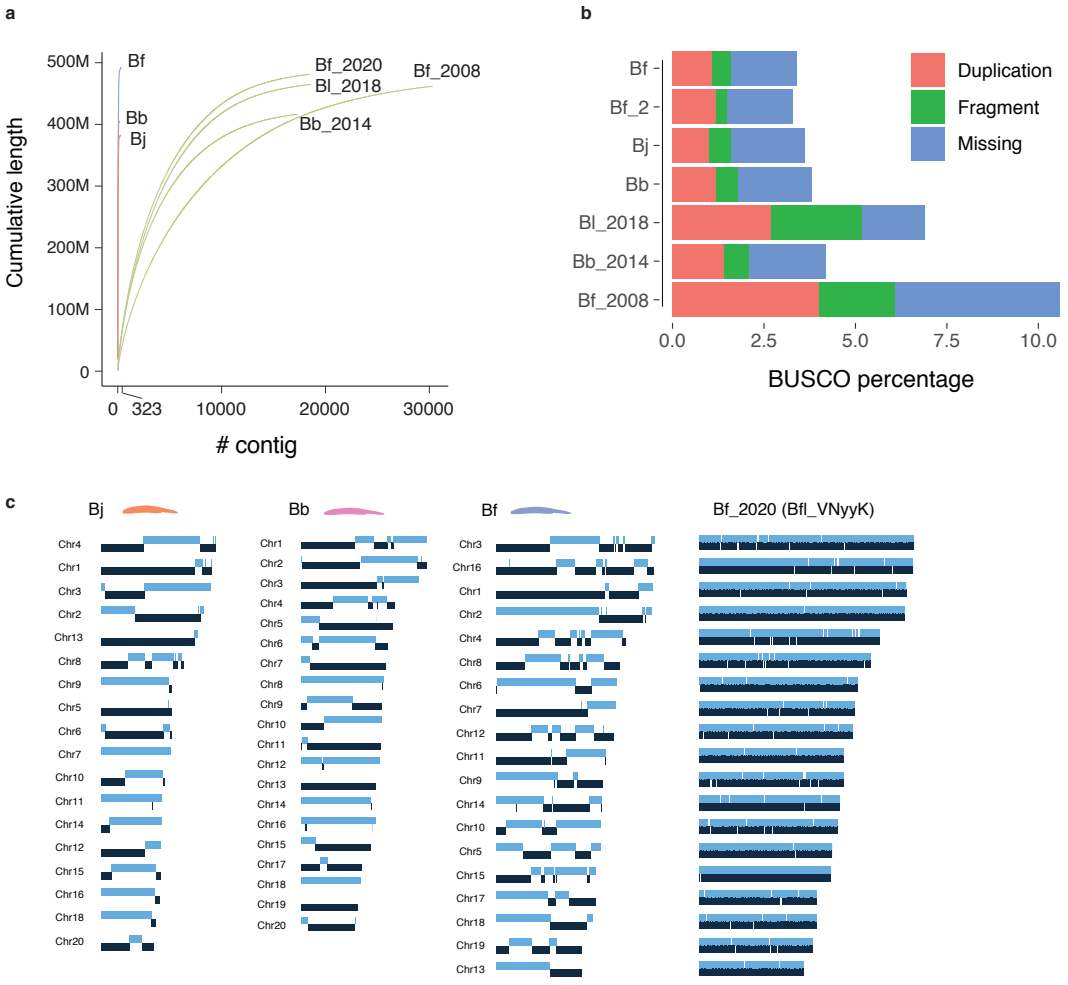


Fig. S5. Improvement of contig length by more than 200 folds.

a) The long-read assemblies produced in this study (Bf, Bb and Bj) have a much longer contig size than the previous short-read assemblies (Bf_2020 GCA_000003815.2, Bf_2008 GCA_000003815.1, Bl_2018 GCA_900088365.1, Bb_2014 GCA_001625305.1). The long-read assemblies have a small contig number (maximum 323), while the short-read assemblies contain at least 16,000 contigs.

b) Percentage of BUSCO genes of the duplication, fragment and missing categories. Our new genome assemblies tend to have a lower percentage for those genes.

c) The distribution of contigs on the chromosomes. The dark and light blue blocks represent contigs. Bf_2020 (Bfl_VNyyK) refers to the Bf genome published by(1)

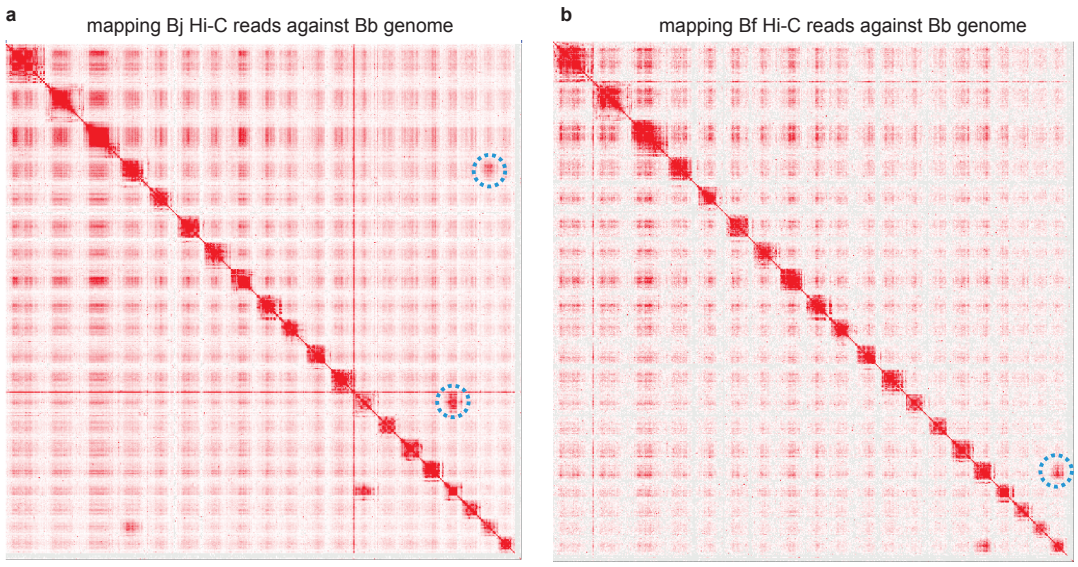


Fig. S6 Chromosome fusions verified by the Hi-C data.

We mapped the Hi-C data of Bj (a) and Bb (b) Hi-C sequencing data against the Bb reference genome. a) In Bj, strong inter-chromosome interactions between chr4 and chr19, and between chr13 and chr17, were highlighted by blue dashed circles. b) In Bf, chr16 and chr20 were found to have strong inter-chromosome interaction (blue dashed circles), consistent with the comparison between chromosome assemblies of Bb and Bf.

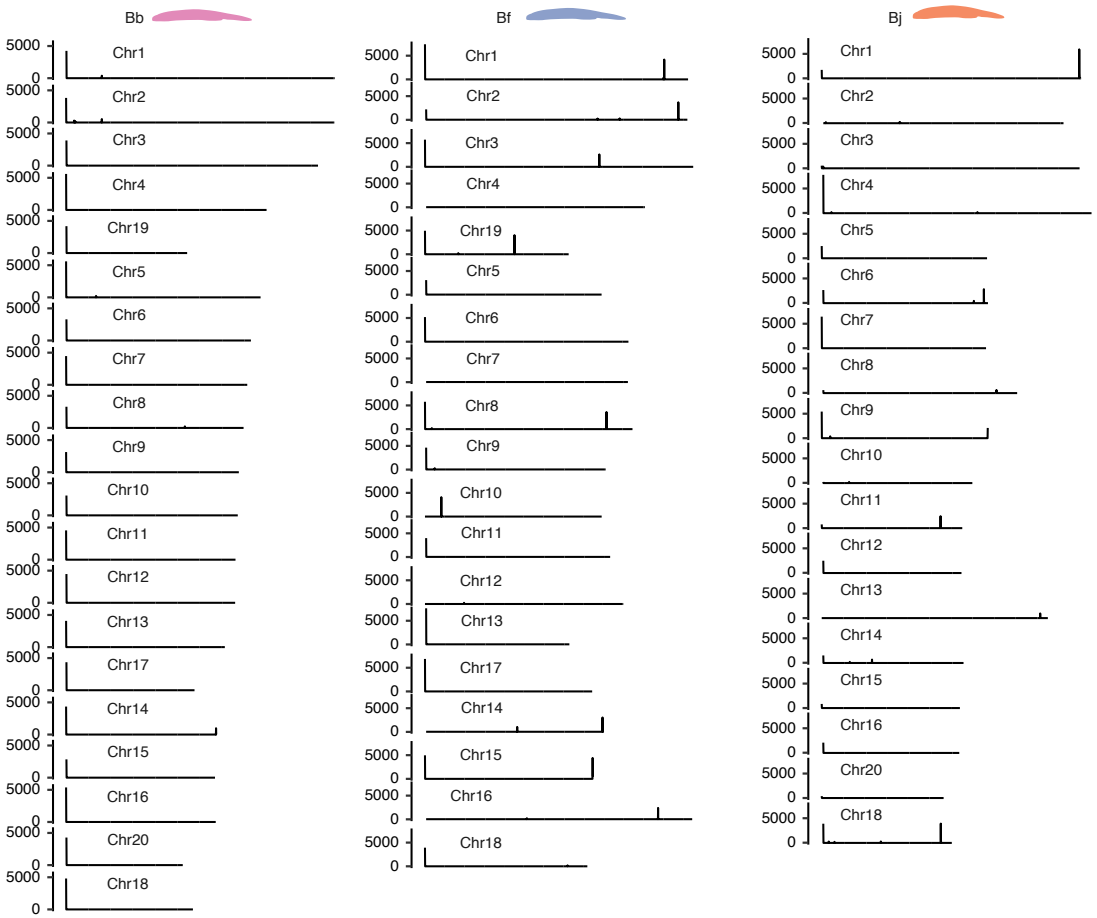


Fig. S7. The distribution of telomere repeats along the chromosomes.

The positions of the telomere repeats (GGGTTA)n were annotated with RepeatMasker. The lengths of the telomere repeats were summed up in every 5 kb window along the chromosomes were plotted. On most chromosomes the telomere repeats are enriched at one chromosome end.

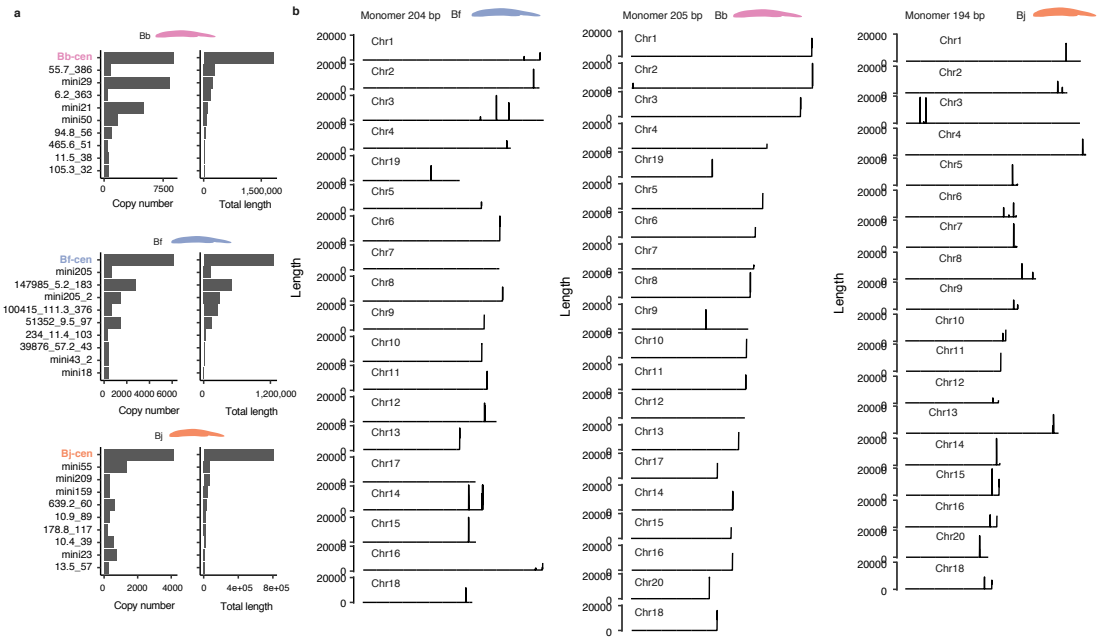


Fig. S8. The identification and distribution of centromeric repeats.

a) The copy number and total lengths of top 10 most abundant satellite repeats. The centromeric satellite repeats (top 1, colored) have the largest copy numbers and longest total lengths.

b) The length of centromeric repeats were summed up in every 20 kb window and plotted along the chromosomes. In most cases, the centromere repeats appear as a single locus on the chromosomes.

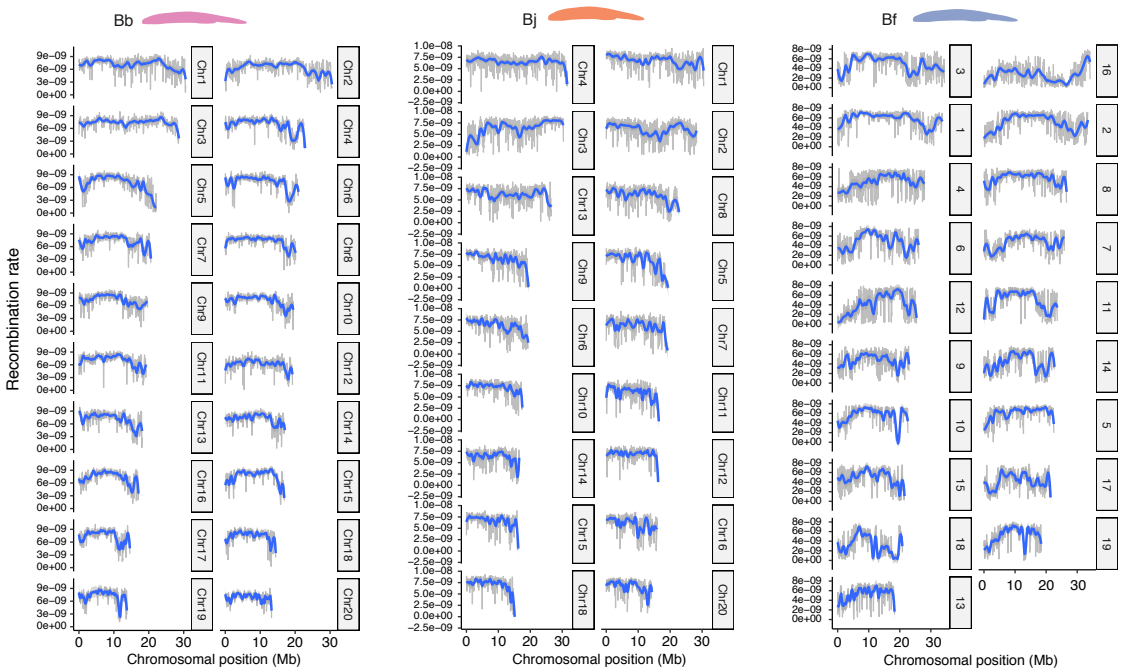


Fig. S9. The low levels of recombination rate at centromeric and pericentromeric regions.

The recombination rates were estimated with the ReLERNN pipeline, using the genome-wide population data. The grey lines show the original estimates while the blue lines show the smoothed means. In most chromosomes the recombination rates tend to be lower in the regions towards the 3' ends where the centromeres are located.

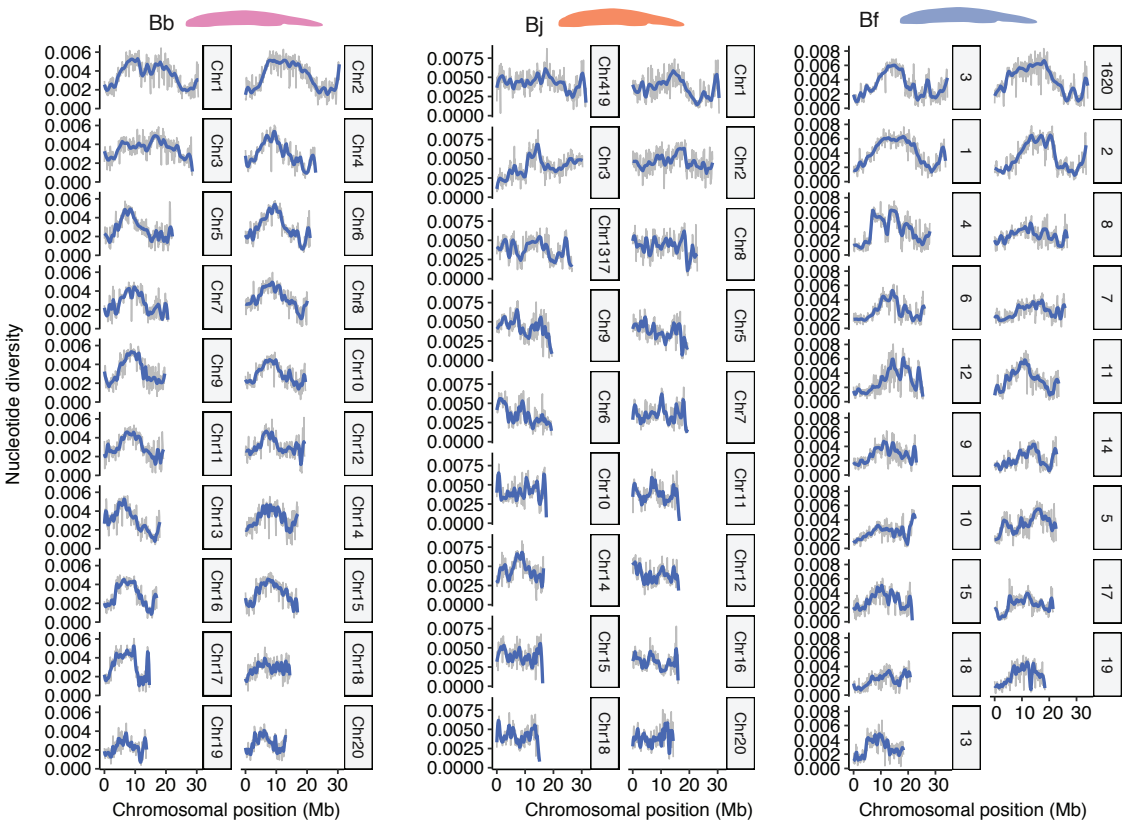


Fig. S10. The low levels of nucleotide diversity at centromeric and pericentromeric regions.

The nucleotide diversity (π) was estimated using the same population data as in Supplementary Fig. S7. We used VCFtools to estimate the nucleotide diversity in 100 kb windows. The π values tend to drop near the regions close to the 3' ends of the chromosomes.

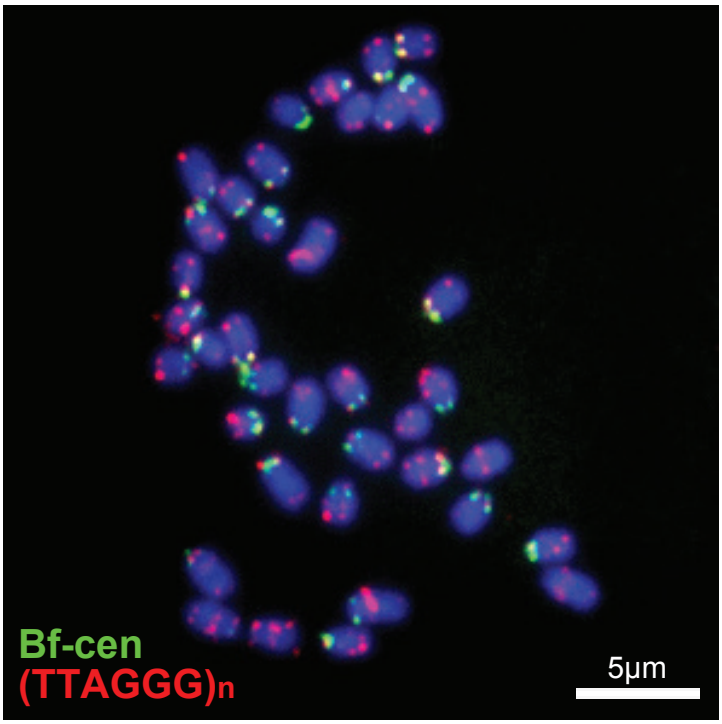


Fig. S11. Fluorescent in situ hybridization (FISH) of centromere and telomere probes to Bf chromosomes.

The 204-bp Bf centromeric monomer sequence (Bf-cen) is shown in Supplementary Fig. S10. The probe (green) is hybridized mostly to the chromosomal ends. The probe of the telomeric motif (TTAGGG)n (red) is hybridized to the other ends of chromosomes, sometimes colocalizing with the centromeric probes. On some chromosomes interstitial telomeric repeats that appear in the middle of the chromosomes are also detected.

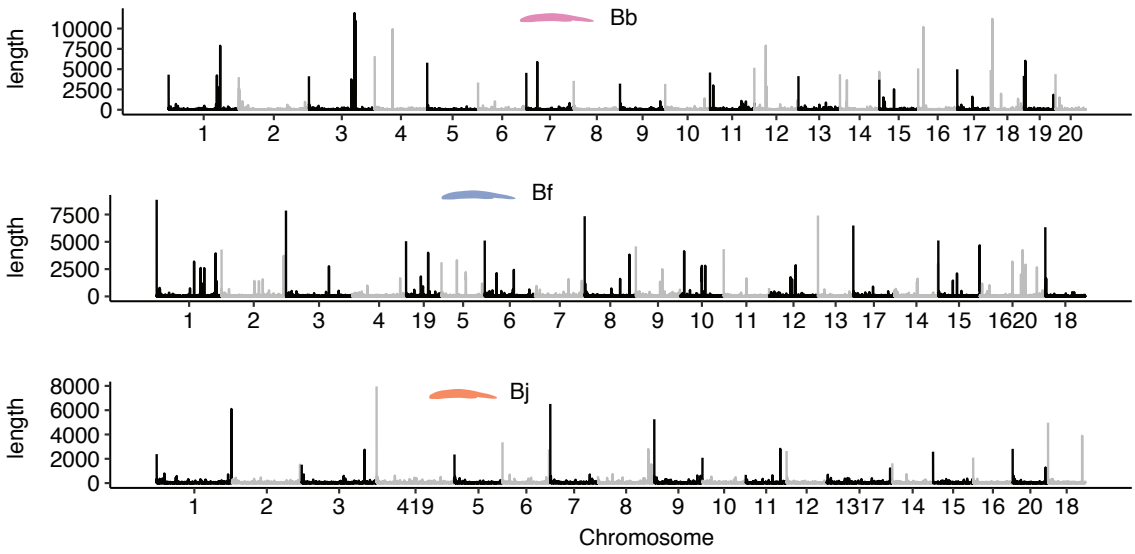


Fig. S12. The distribution of G4 sequences along the chromosomes.

The G4 (G-quadruplex) sites were predicted and the total lengths were summed up in every 20 kb window along the chromosomes. The peaks of G4 are often present at the ends of chromosomes, but sometimes are also seen in the middle of chromosomes.

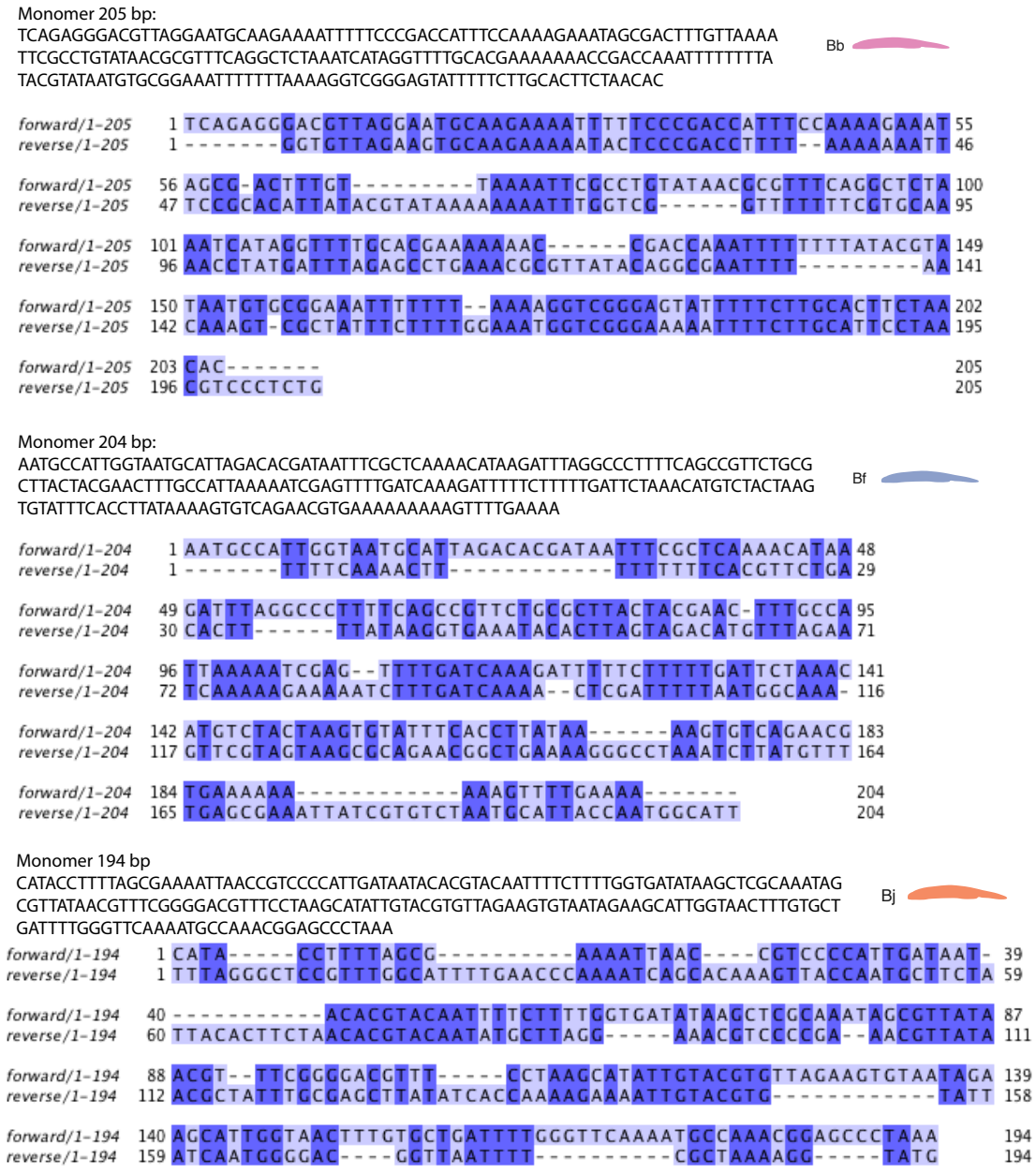


Fig. S13. The centromeric repeat monomer.

The sequences of the centromeric repeat monomer are shown at the top of each panel. The monomers are not homologous among amphioxus species. For each monomer, the reversed complementary sequence (labelled as reverse) was generated and aligned with the original (labelled as forward) monomer. The dark blue highlights the identical nucleotides between the forward and reverse monomer.

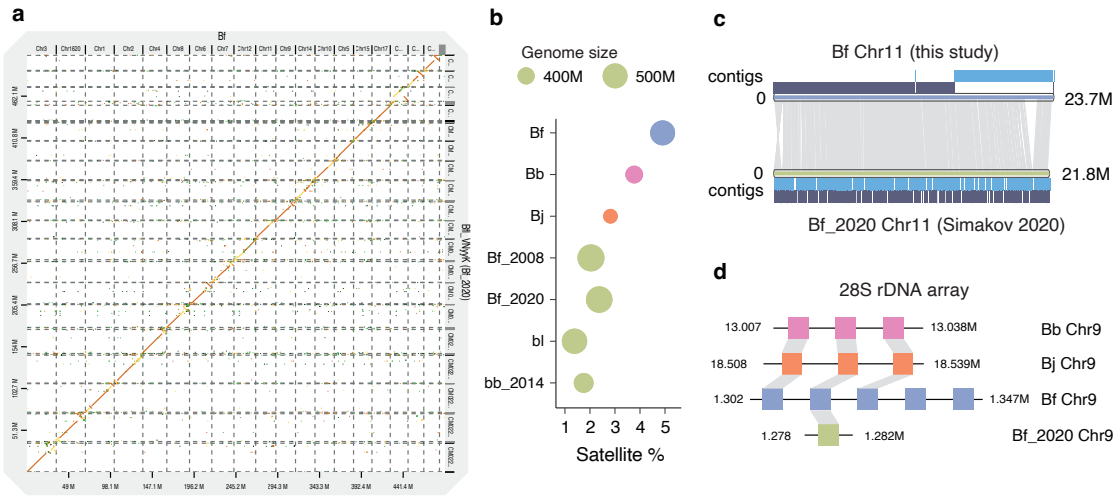


Fig. S14. Complex genomic regions of amphioxus species have been assembled by long-reads.

a) Chromosomal synteny between our new Bf assembly and Bfl_VNyyK (Bf_2020).
b) In the long-read assemblies, the satellite sequences comprise a larger portion of the genome because more satellite sequences were assembled, compared to the previously published genomes. The circle size represents the genome size. The larger genomes have a higher portion of satellite repeats. d) The genomic synteny of Chr11 between the long-read (Bf) and the published short-read (Bf_2020) assemblies. The grey bands represent syntenic blocks. The short-read assembly missed one large sequence fragment at the ~20Mb position. The dark and light blue blocks represent contigs. The Chr11 of long-read assembly is comprised of only 6 contigs, while that of the short- read assembly is comprised of thousands of contigs.
a) The gene synteny of a 28S rRNA gene array. There are 3, 3 and 5 tandem copies of 28S rRNA genes assembled in Bb, Bj and Bf long-read assemblies respectively, but only one copy was assembled in the short-read assembly of Bf.

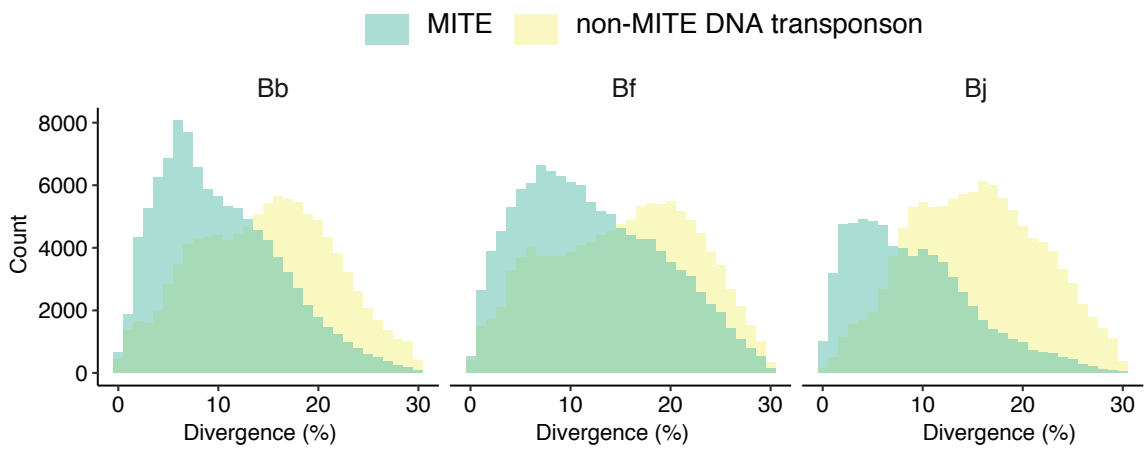


Fig. S15. The divergence of MITE in the amphioxus genomes.

MITE stands for Miniature Inverted-repeat Transposable Element which is a type of DNA transposons. The divergence of repeats (X-axis) were estimated by RepeatMasker. In all amphioxus species, MITEs have a lower divergence level from the consensus sequences, compared with other types of DNA transposon, suggesting more recent activities of MITEs.

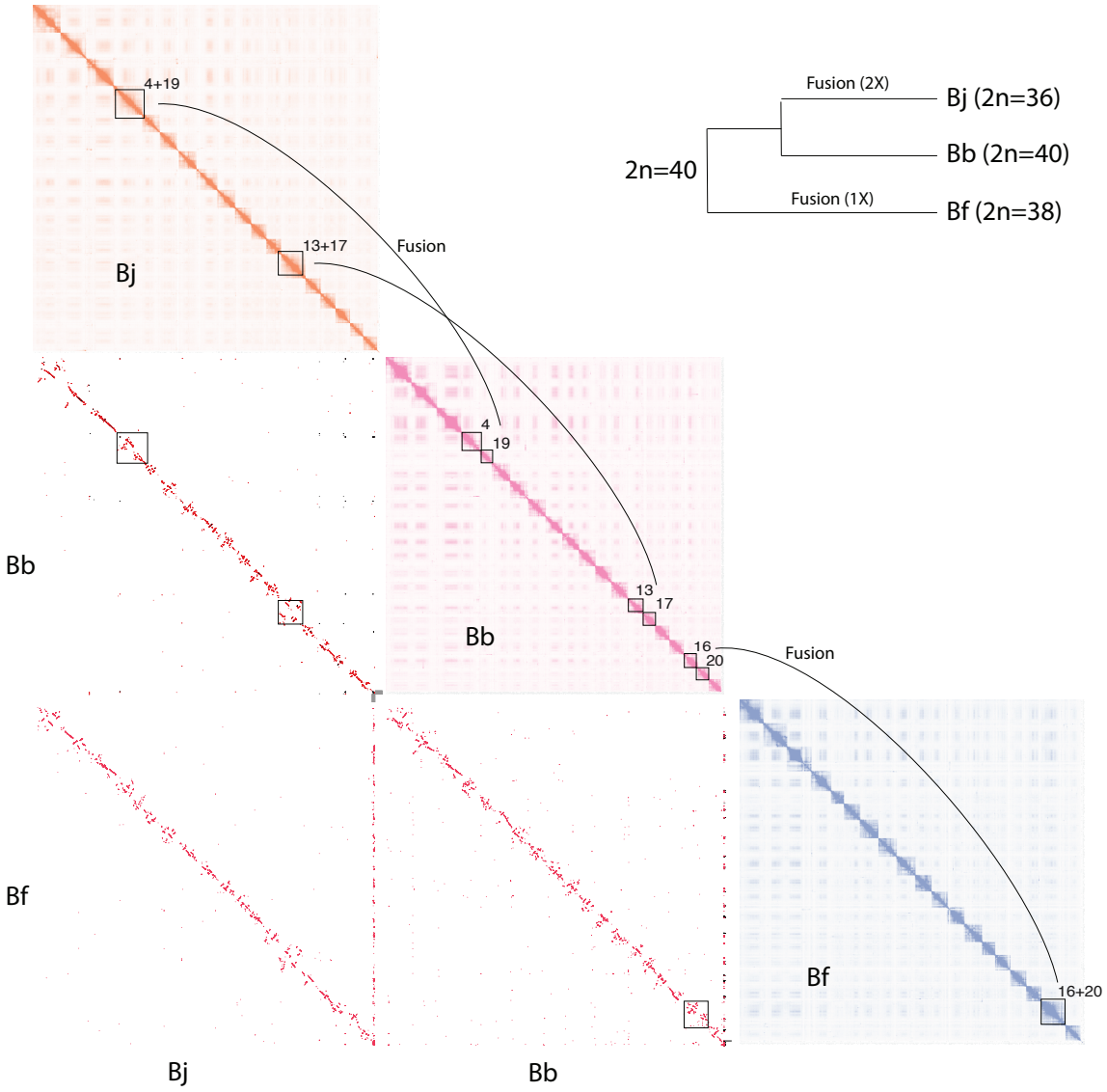


Fig. S16. Chromosomal fusions between amphioxus species.

The Hi-C contact maps were visualized by JuiceBox. The genome synteny between each pair of species were shown underneath the Hi-C contact maps. The synteny dot-plots were visualised using the same methods described in Supplementary Fig. S14. The squares highlight the chromosomes that involve chromosome fusions. The numbers near the highlighted chromosomes in the Hi-C maps represent the chromosome numbers.

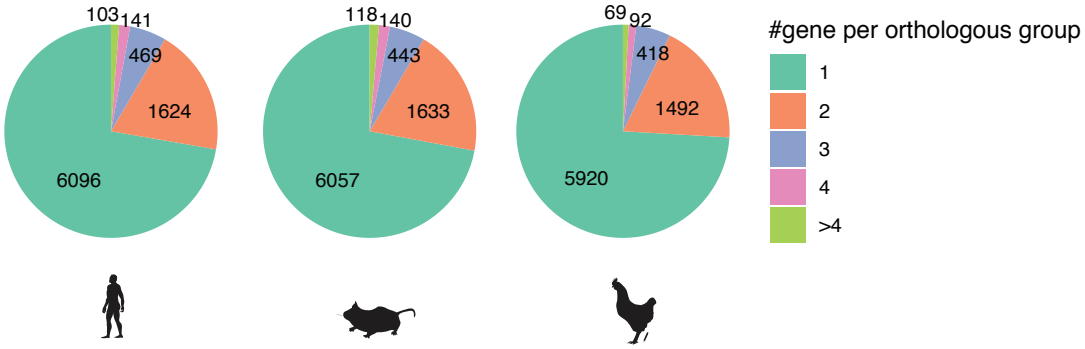


Fig. S17. Number of genes in each orthologous group.

The statistics was directly retrieved from the OrthoFinder result. In most orthologous gene groups the genes are single copy. When the genes are multip-copy, the gene copy number mainly ranges from two to four.

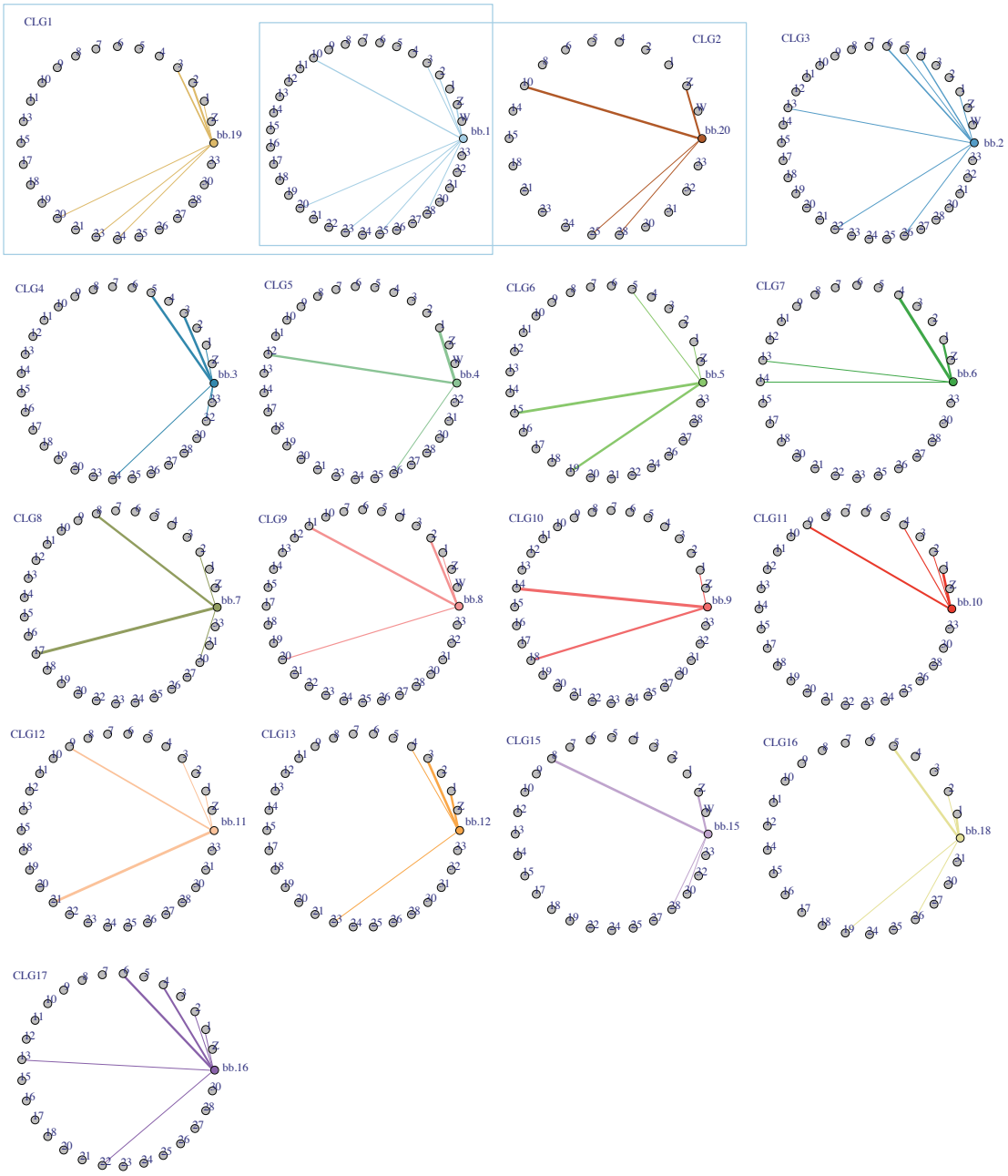


Fig. S18. Homology between chicken and Bb chromosomes.

Each line connecting a Bb and a chicken chromosome represents the proportion of Bb genes (Bb%) of certain Bb chromosome that is homologous to the genes of the chicken chromosome. The thicknesses of the lines indicate the Bb% values. Only the lines with Bb% values larger than 4% are shown. One Bb chromosome usually connects with three or four chicken chromosomes. Each small panel represents one Bb chromosome that represents one CLG (except for CLG1, CLG2 and CLG14). CLG14 is shown in Fig. 2b. A part of Bb chr1 (bb.1) shares the chicken homology with chr19 while the other part shares with chr20 (The blue squares represent CLG1 and CLG2 respectively, both including a part of chr1).

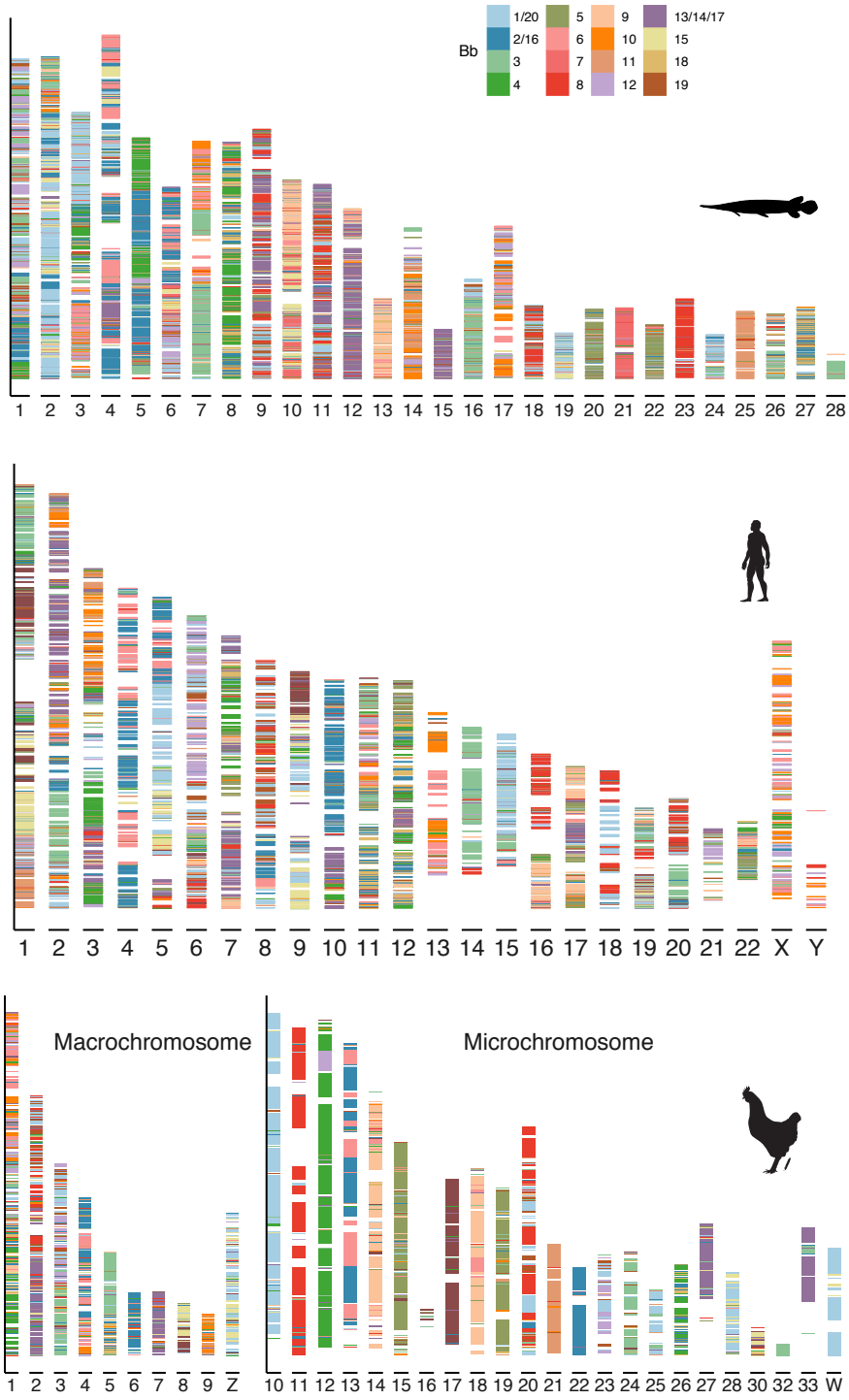


Fig. S19. The composition of human and spotted gar chromosomes by CLG.

The colored bands show the Bb-chicken synteny blocks. A synteny block was built when two syntenic genes are not apart from each other for longer than 5 Mb. The Bb chromosomes were translated into CLGs according to the Bb-CLG relationship (Fig. 2c).

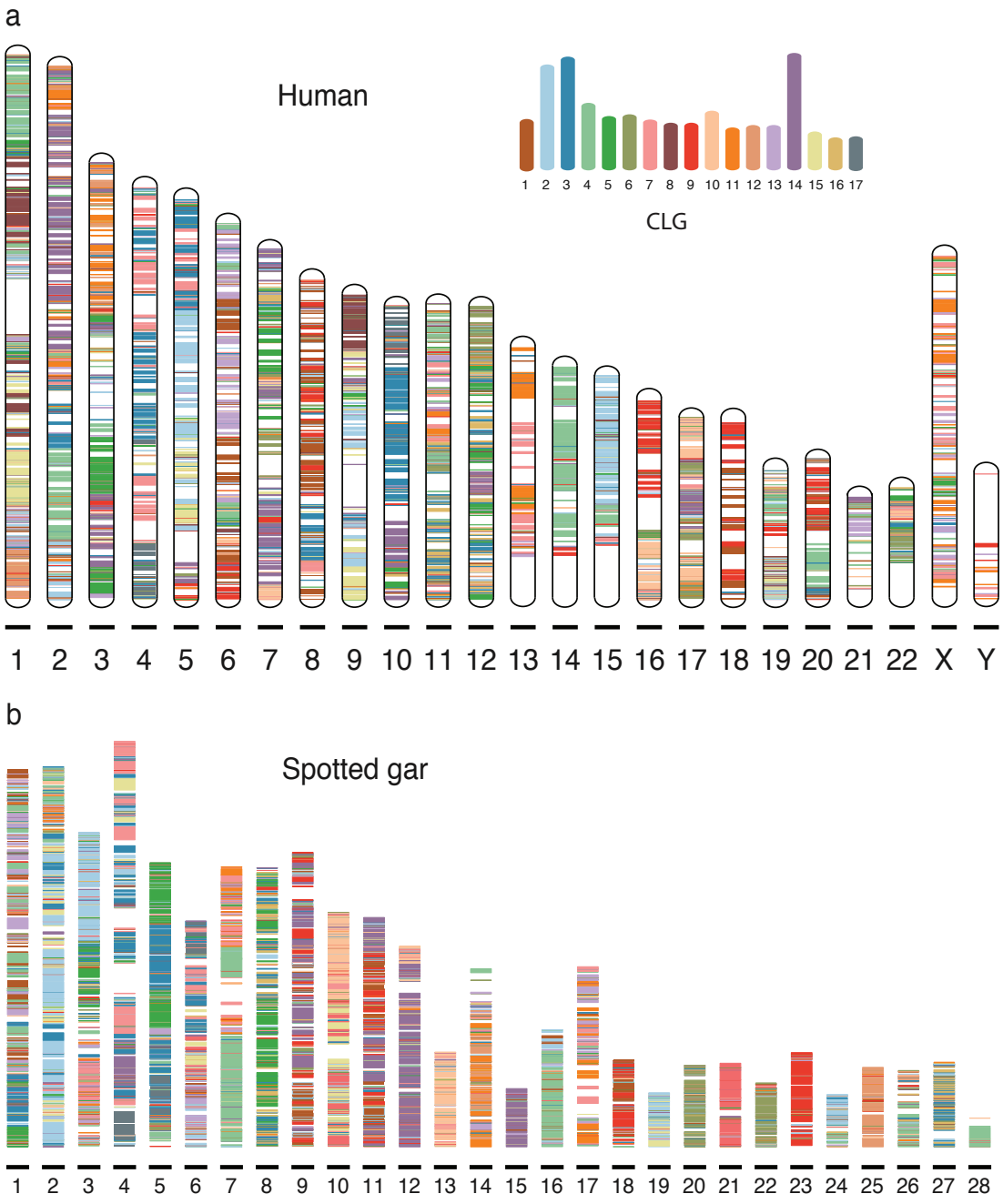


Fig. S20. The composition of human and spotted gar chromosomes by CLGs.

Similar to Fig. 2d, we show the homologous sequences of CLGs across the chromosomes of human and spotted gar.

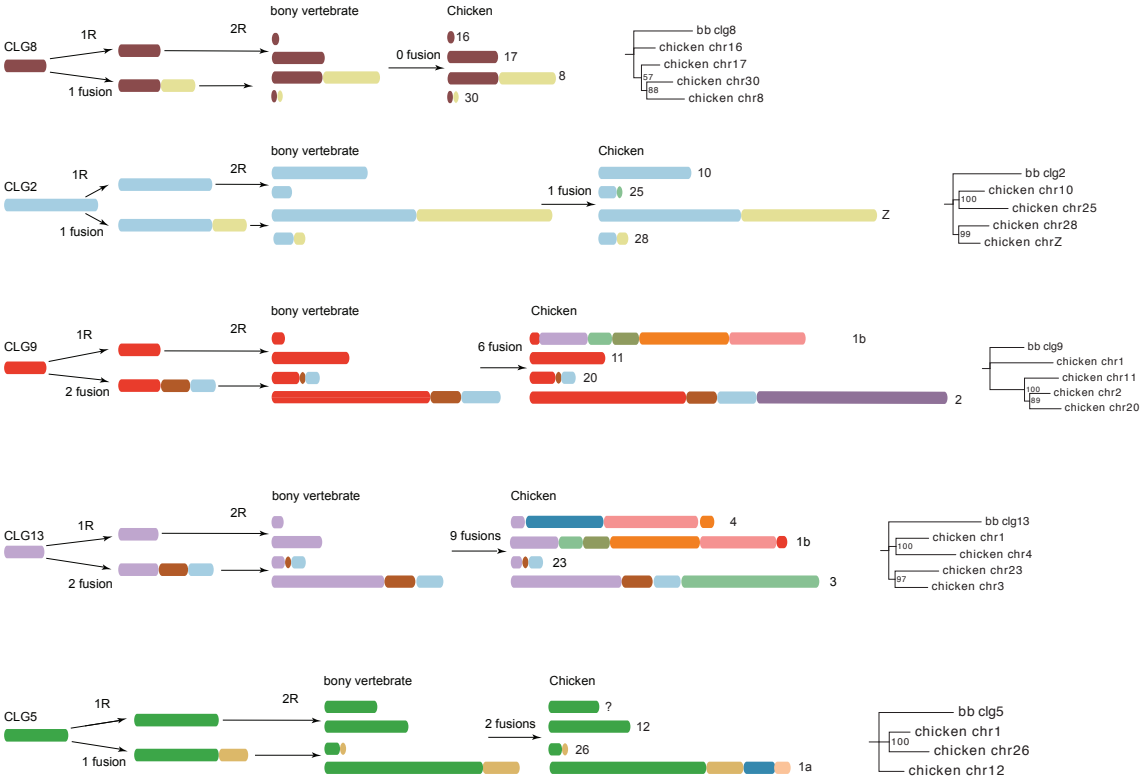


Fig. S21. Post-1R chromosomal fusions.

The schematic plots show the evolutionary history of five CLGs, from the chordate ancestors to extant chicken chromosomes through 1R and 2R. A chromosome block in one color represents a CLG. For all five CLGs, one duplicated chromosome fused with another duplicate chromosome after 1R. Chicken chromosome numbers are labelled at the right end of the plot. In the right panels, the phylogeny was built using the ohnolog genes of chicken, using Bb as an outgroup. The bootstrapping values are shown at the nodes. The topology of the phylogeny supports the evolutionary relationships of ohno-chromosomes inferred based on post-1R fusions.

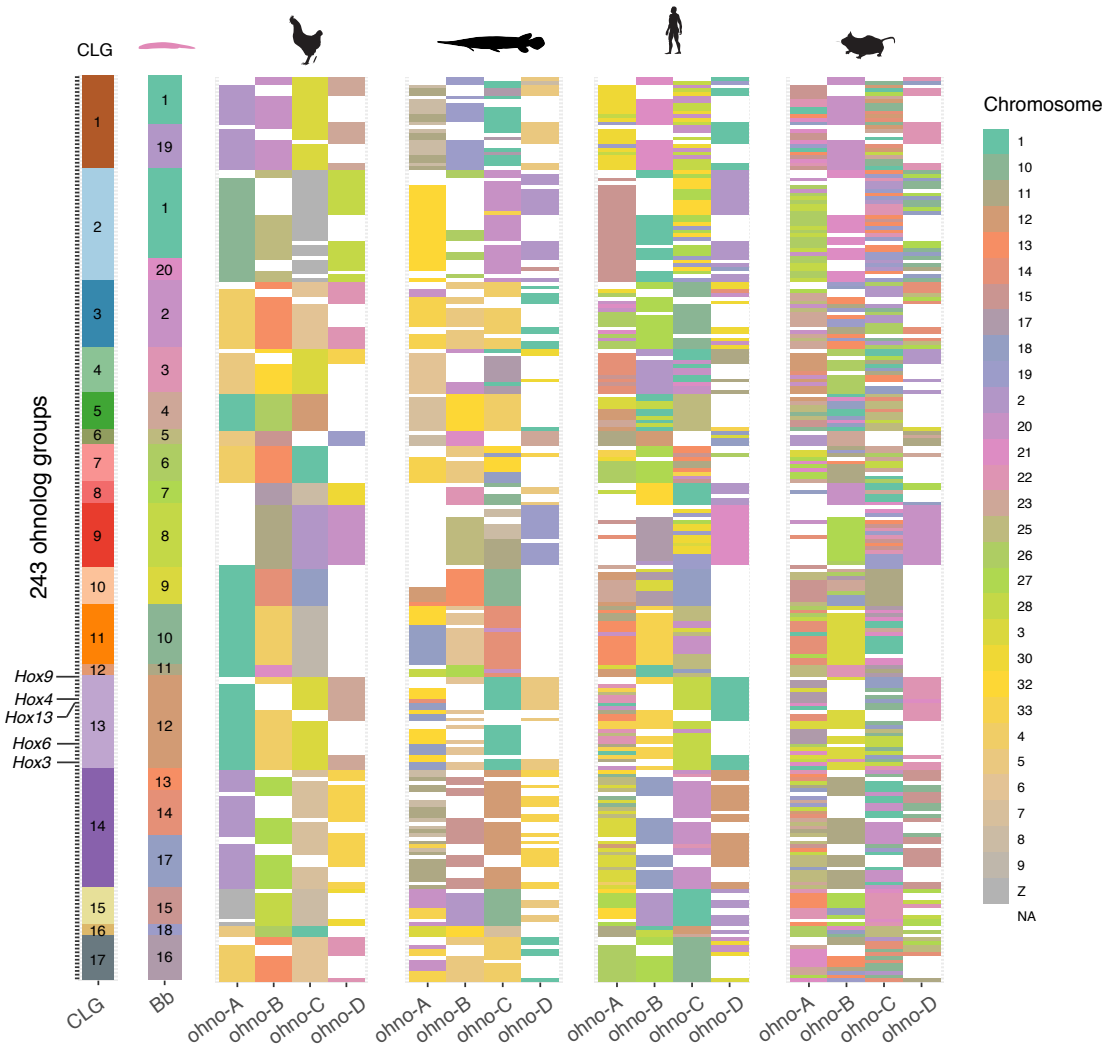


Fig. S22. Ohnolog groups used for dating the time of 1R and 2R.

There are 243 ohnolog groups identified that contain at least three vertebrate ohnologs. Each row represents one ohnolog group which contains one Bb gene, at least three chicken genes, 1-4 genes of spotted gar, human and mouse. The vertebrate genes of the same ohno-chromosome lineage (A/B/C/D) are homologous to each other. The IDs of CLGs and Bb chromosomes are labelled since they are clustered as large blocks, and the legend for vertebrate chromosomes was shown on the right. Five Hox genes are highlighted as examples.

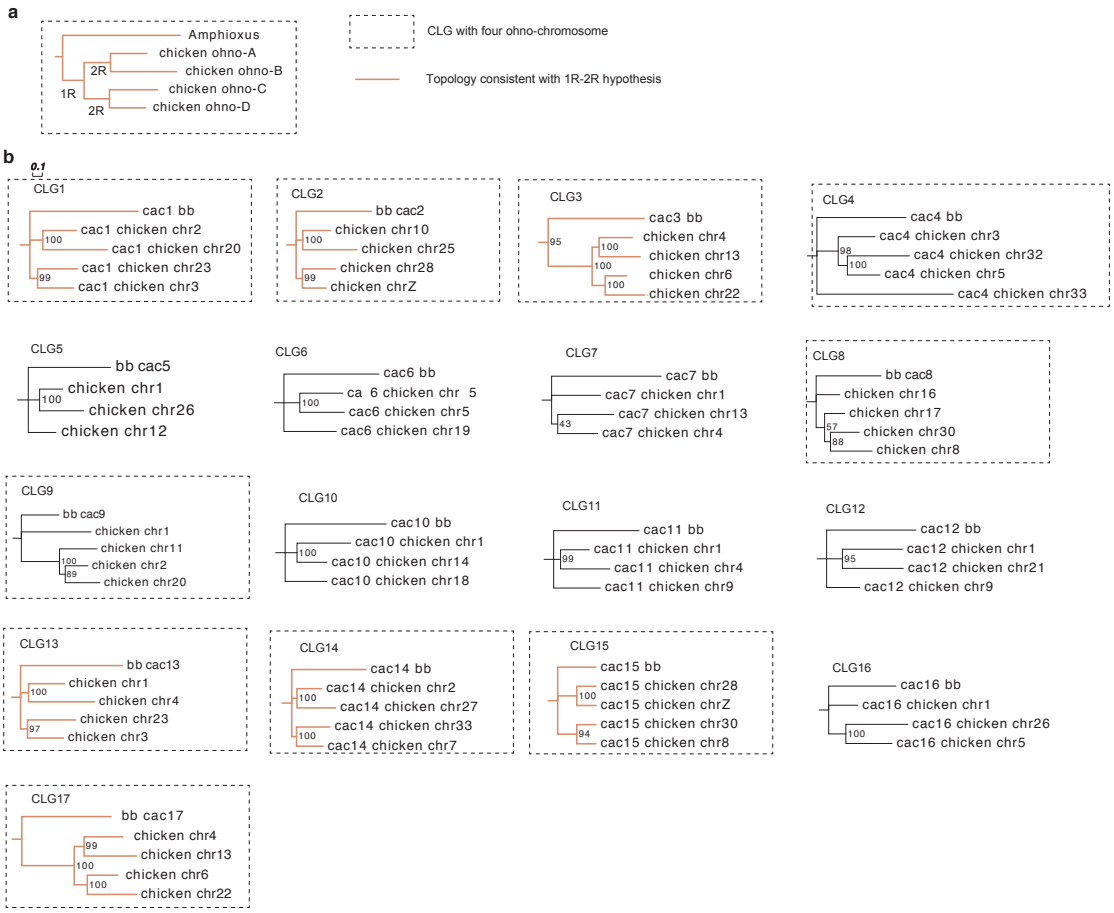


Fig. S23. Phylogeny of ohno-chromosomes built with ohnolog genes.

The phylogenetic trees of chicken ohno-chromosomes were built using the ohnologs gene residing on the same ohno-chromosomes. The list of ohnologs is listed after each ohno-chromosome branch. The phylogeny of the other five CLGs that experienced post-1R fusions in vertebrates is shown in Supplementary Fig. S20. The bootstrapping values are shown at the nodes.

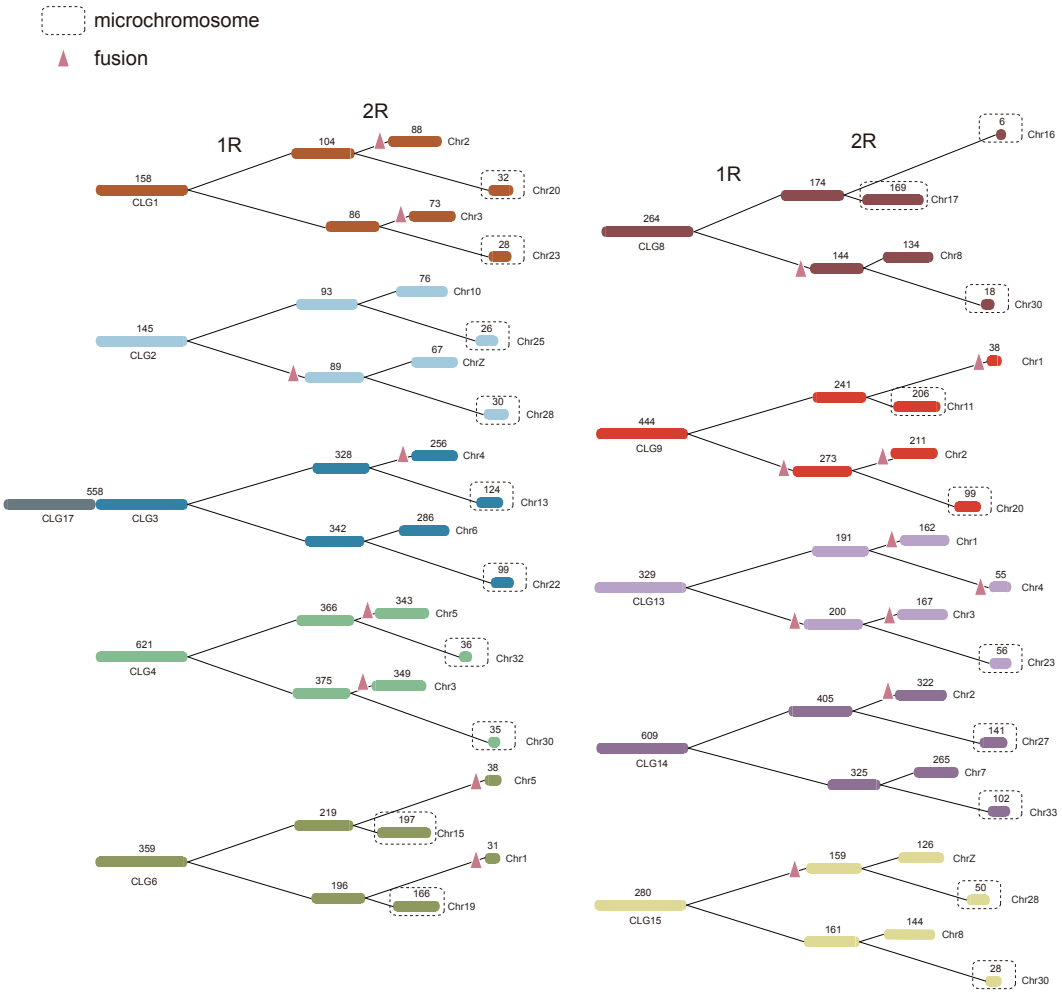


Fig. S24. Reconstruction of gene loss events following 1R and 2R.

The numbers above the CLGs are the original orthologous genes that have homologs in at least one descendant ohno-chromosomes. The numbers above the chicken and post-1R chromosomes show the extant and reconstructed gene numbers respectively. The branch length indicates the numbers of gene loss, i.e. a longer branch means more genes have become lost during or after WGDs in the branch. The pink triangles indicate fusion events along the branches. The microchromosomes are highlighted by dashed squares. In the branches leading to microchromosomes, no fusions have occurred, and the branches are longer (more gene loss).

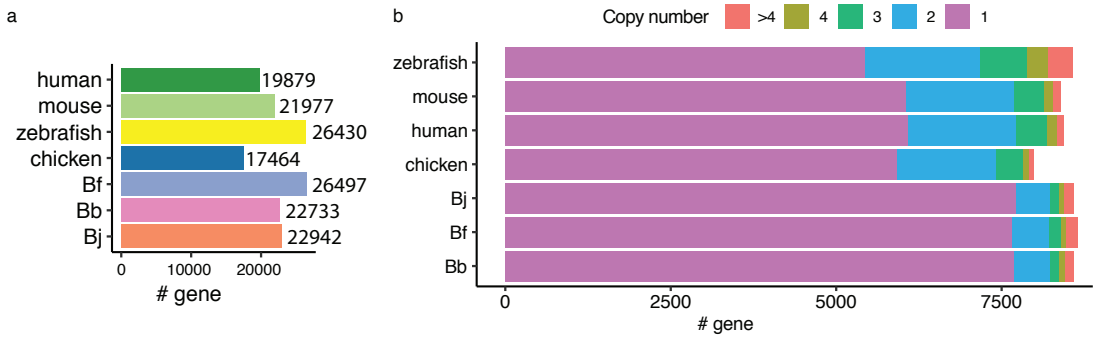
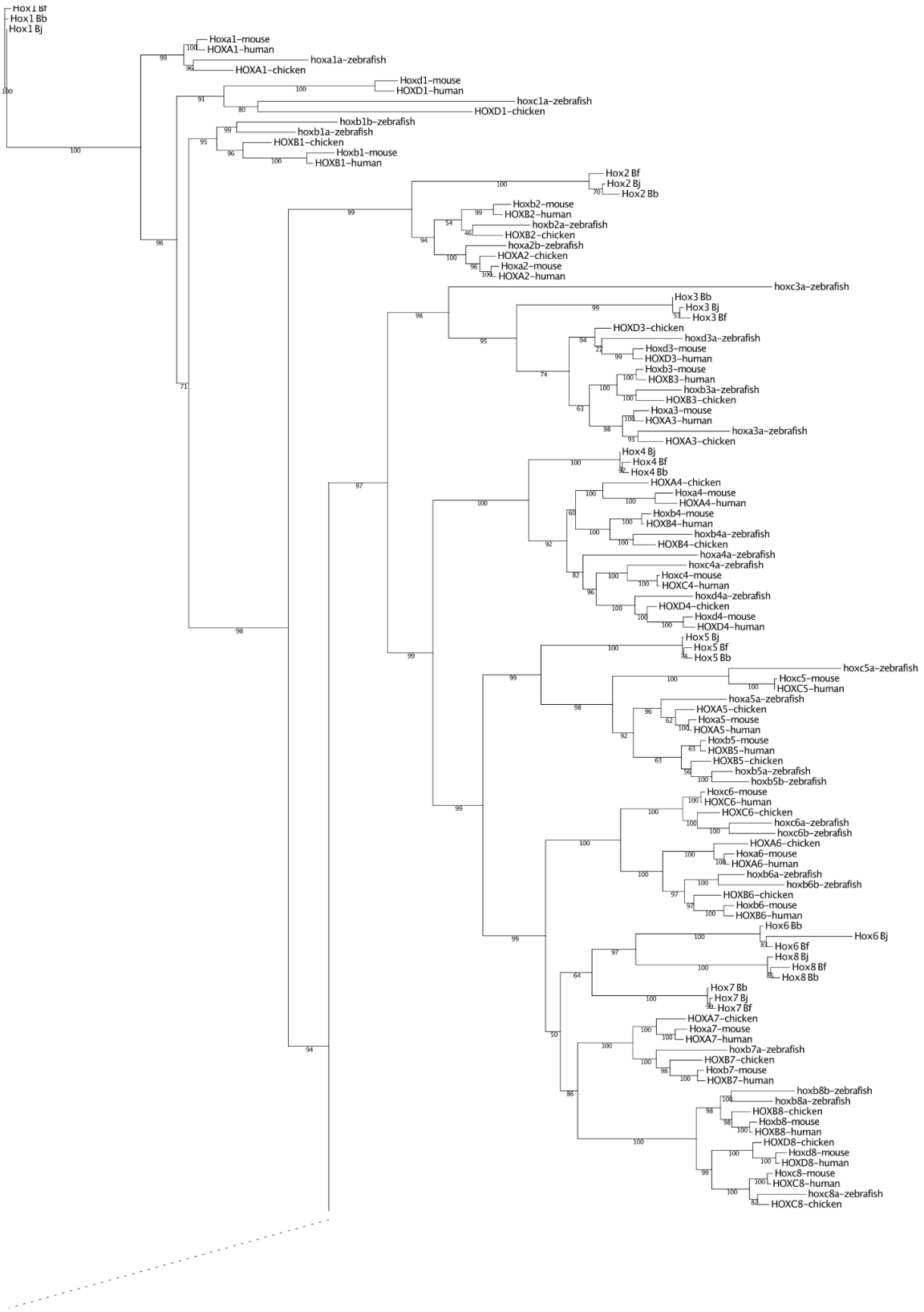


Fig. S25. Gene number of chordate genomes.

a) The total number of protein-coding genes of four vertebrate species and three amphioxus species.b) Statistics of gene counts in orthogroups. We included the orthologous gene groups in which at least one amphioxus gene and one vertebrate gene are present (i.e. the orthogroups are shared by amphioxus and vertebrates). Amphioxus has a larger proportion of single-copy genes, but also has a considerable amount of duplicated genes given the lack of WGDs.


Hox1 Bf
Hox1 Bb
Hox1 Bj
Hoxa1–mouse
HOXA1–human
hoxa1a–zebrafish
HOXA1–chicken
Hoxd1–mouse
HOXD1–human
hoxc1a–zebrafish
HOXD1–chicken
hoxb1b–zebrafish
hoxb1a–zebrafish
HOXB1–chicken
Hoxb1–mouse
HOXB1–human
Hox2 Bf
Hox2 Bj
Hox2 Bb
Hoxb2–mouse
HOXB2–human
hoxb2a–zebrafish
HOXB2–chicken
hoxa2b–zebrafish
HOXA2–chicken
Hoxa2–mouse
HOXA2–human
hoxc3a–zebrafish
Hox3 Bb
Hox3 Bj
Hox3 Bf
HOXD3–chicken
hoxd3a–zebrafish
Hoxd3–mouse
HOXD3–human
Hoxb3–mouse
HOXB3–human
hoxb3a–zebrafish
HOXB3–chicken
Hoxa3–mouse
HOXA3–human
hoxa3a–zebrafish
HOXA3–chicken
Hox4 Bj
Hox4 Bf
Hox4 Bb
HOXA4–chicken
Hoxa4–mouse
HOXA4–human
Hoxb4–mouse
HOXB4–human
hoxb4a–zebrafish
HOXB4–chicken
hoxa4a–zebrafish
hoxc4a–zebrafish
Hoxc4–mouse
HOXC4–human
hoxd4a–zebrafish
HOXD4–chicken
Hoxd4–mouse
HOXD4–human
Hox5 Bj
Hox5 Bf
Hox5 Bb
hoxc5a–zebrafish
Hoxc5–mouse
HOXC5–human
hoxa5a–zebrafish
HOXA5–chicken
Hoxa5–mouse
HOXA5–human
Hoxb5–mouse
HOXB5–human
HOXB5–chicken
hoxb5a–zebrafish
hoxb5b–zebrafish
Hoxc6–mouse
HOXC6–human
HOXC6–chicken
hoxc6a–zebrafish
hoxc6b–zebrafish
HOXA6–chicken
Hoxa6–mouse
HOXA6–human
hoxb6a–zebrafish
hoxb6b–zebrafish
HOXB6–chicken
Hoxb6–mouse
HOXB6–human
Hox6 Bb
Hox6 Bj
Hox6 Bf
Hox8 Bj
Hox8 Bf
Hox8 Bb
Hox7 Bb
Hox7 Bj
Hox7 Bf
HOXA7–chicken
Hoxa7–mouse
HOXA7–human
hoxb7a–zebrafish
HOXB7–chicken
Hoxb7–mouse
HOXB7–human
hoxb8b–zebrafish
hoxb8a–zebrafish
HOXB8–chicken
Hoxb8–mouse
HOXB8–human
HOXD8–chicken
Hoxd8–mouse
HOXD8–human
Hoxc8–mouse
HOXC8–human
hoxc8a–zebrafish
HOXC8–chicken

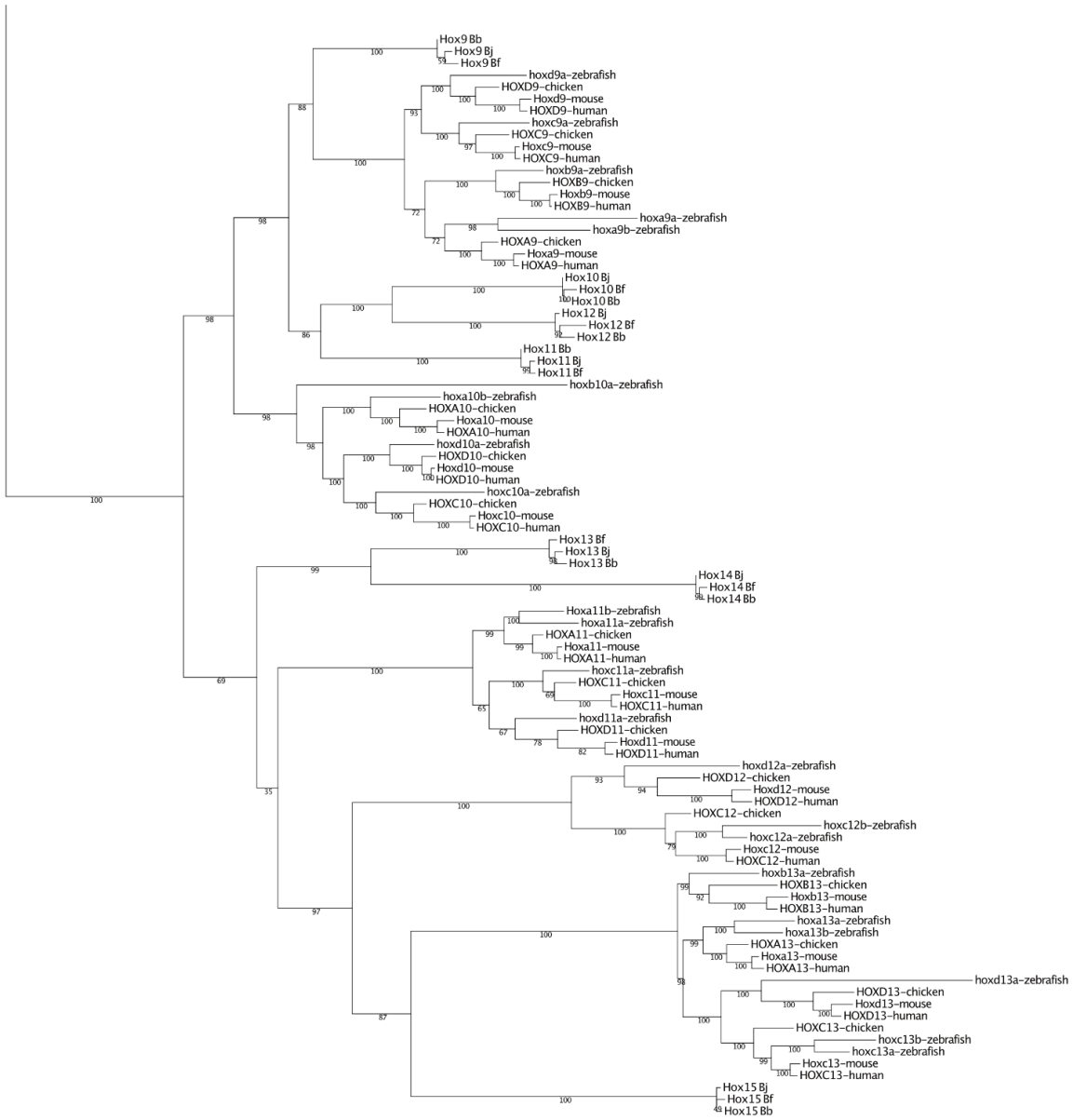


Fig. S26. Phylogenetic tree of Hox genes.

The homologous Hox gene group of amphioxus and vertebrates shown in rectangular mode. Numbers at the internal nodes indicate the bootstrapping value under 1000 replicates.

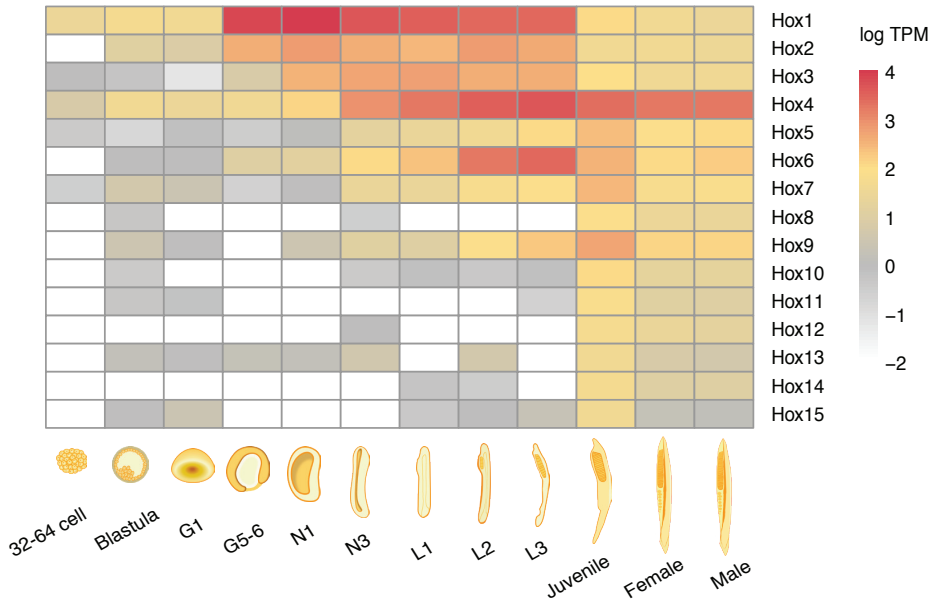


Fig. S27. The temporal expression of Hox genes.

The expression profile of Hox genes of Bf across 11 developmental stages. In earlier stages only anterior Hox genes are expressed, and the posterior Hox genes are generally expressed in later stages.

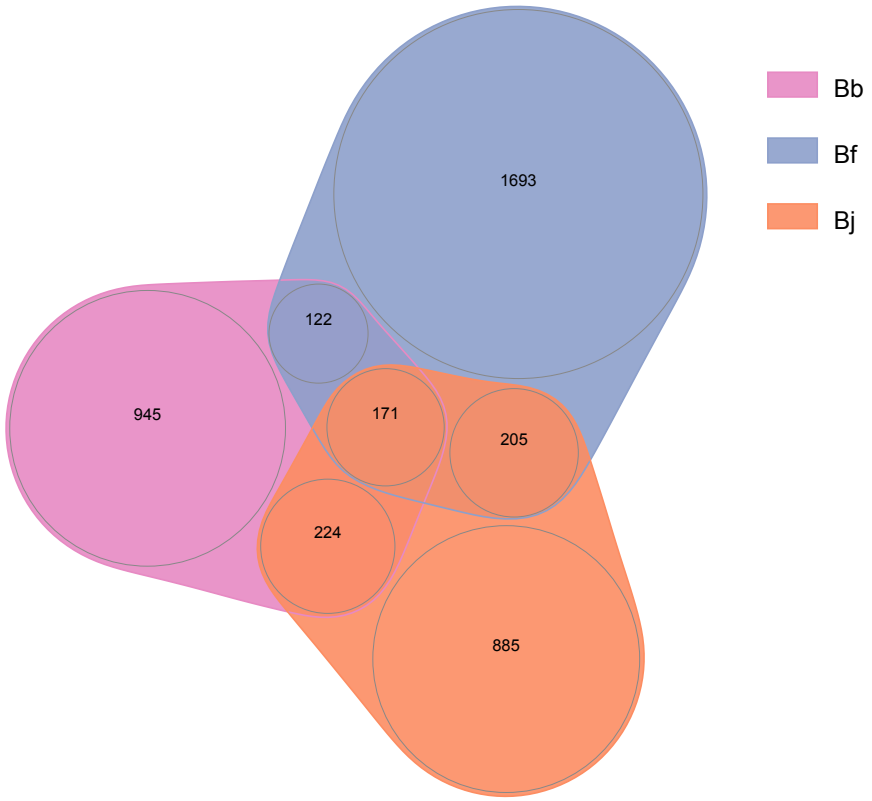


Fig. S28. The Venn diagram of genes involved in segmental duplications of three amphioxus species.

For those genes whose coding regions are overlapped for 60% or more of the segmental duplication regions in three different amphioxus species, we identified the orthologous genes that are shared between or specific to the three species. We found that most of the genes are unique in one of the amphioxus species, suggesting the segmental duplications in amphioxus tend to be of more recent origin.

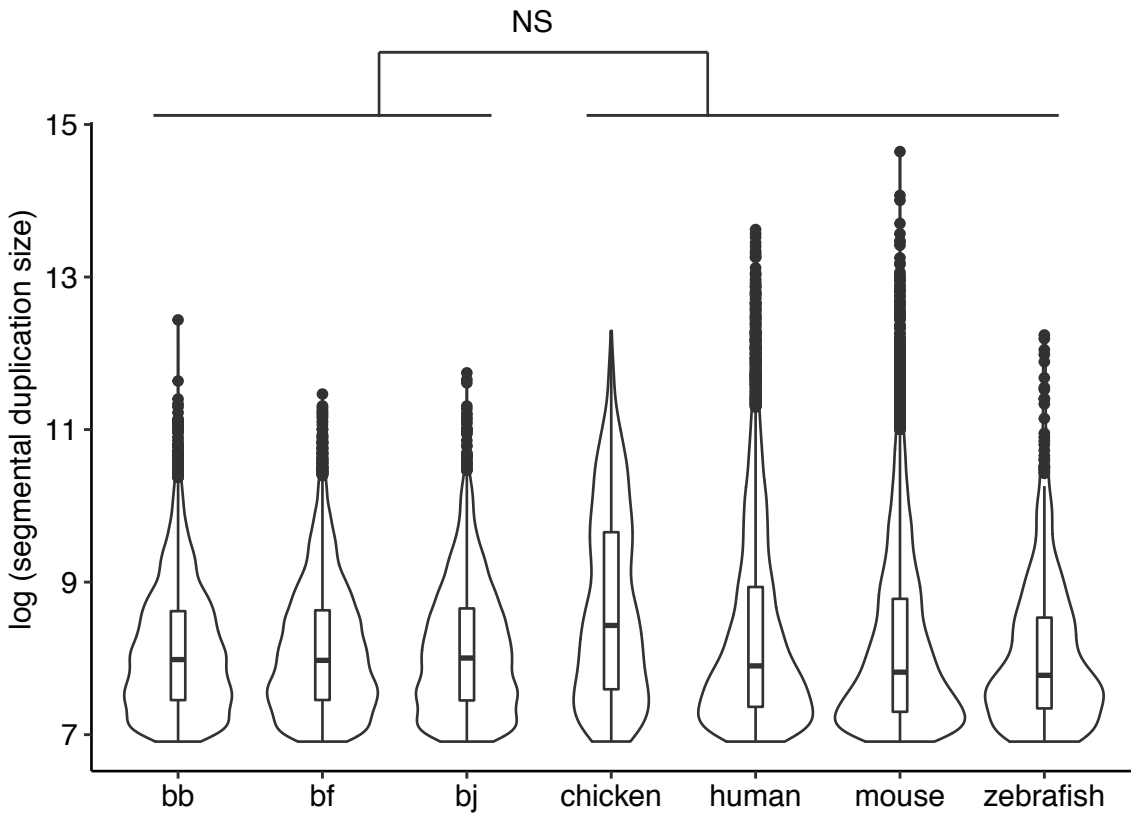


Fig. S29. The size of segmental duplications.

The lengths of segmental duplications of the amphioxus genomes were compared with those from the vertebrate genomes. By definition all segmental duplications studied are larger than 1 kb. The lengths of segmental duplications of amphioxus are not significantly different from those of vertebrates (Wilcoxon test, $p > 0.05$).

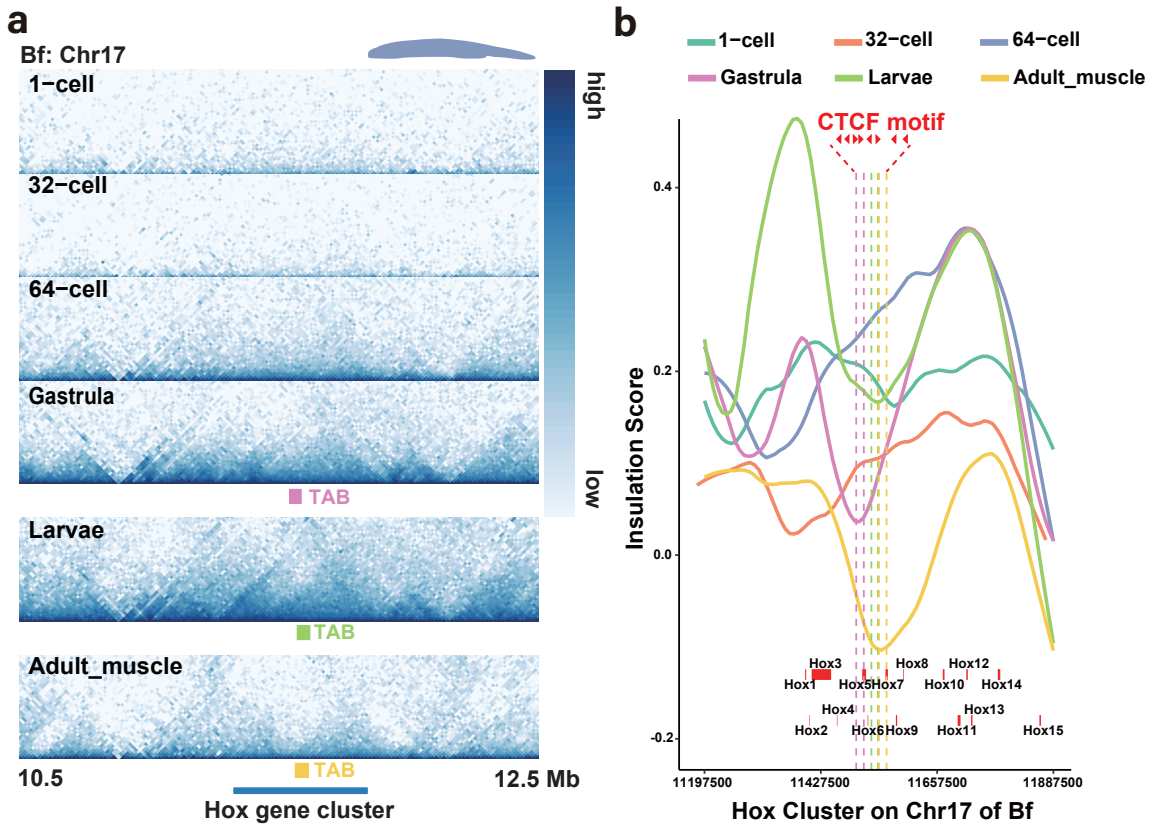


Fig. S30. TAD and insulation score at 15kb resolution within the amphioxus Hox gene cluster.

a) The interaction heat map for high-order chromatin structures of chromosome 17(10500000-12500000) showing the TAD borders (TABs) within *Hox* cluster is shown for 1-cell, 32-cell, 64-cell, gastrula, larvae and adult (muscle tissues) stages of Bf at 15-kb resolution. The dots under heat maps of Gastrula, Larvae and Adult_muscle indicate the TABs of these three developmental stages.

b) The horizontal solid lines indicate the distribution of insulation scores along the *Hox* gene cluster, the smaller the insulation score, the stronger the TABs. The vertical dotted lines represent the TABs of gastrula (pink), larvae (green) and adult (yellow) stages of amphioxus. The length and location of the *Hox* genes are marked with thick red lines. The TAB at the *Hox* gene cluster is shifted between different developmental stages, but between the regions of *Hox5* and *Hox7*.

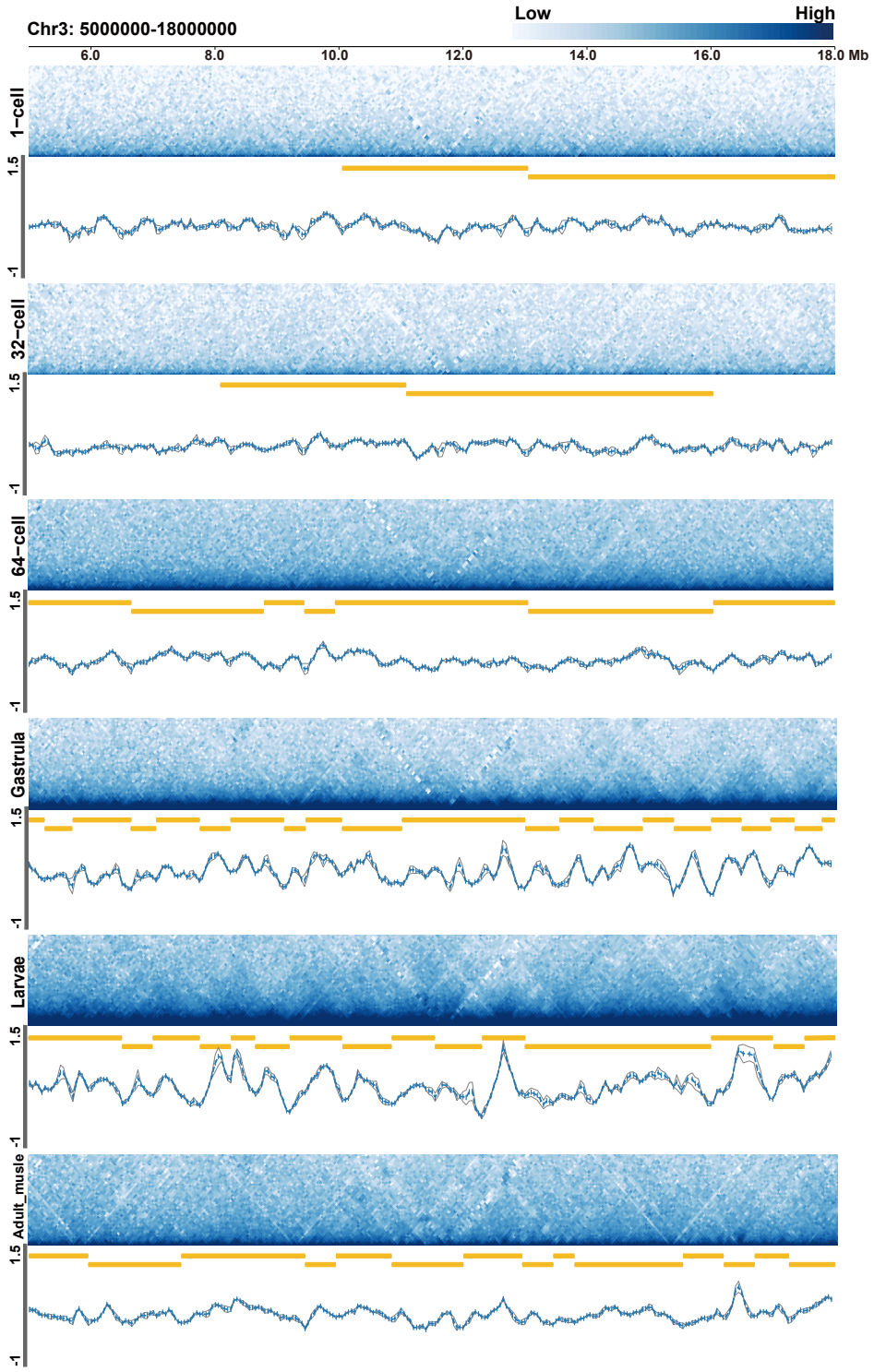


Fig. S31. Three-dimensional chromatin structures during embryonic development of amphioxus.

The interaction heat map for high-order chromatin structures of chromosome 3(5M-18M), as an example, is shown for 1-cell, 32-cell, 64-cell, gastrula, larvae and adult(muscle tissues) stages of Bf at 50-kb resolution. The lines indicate the insulation score of each development stages.

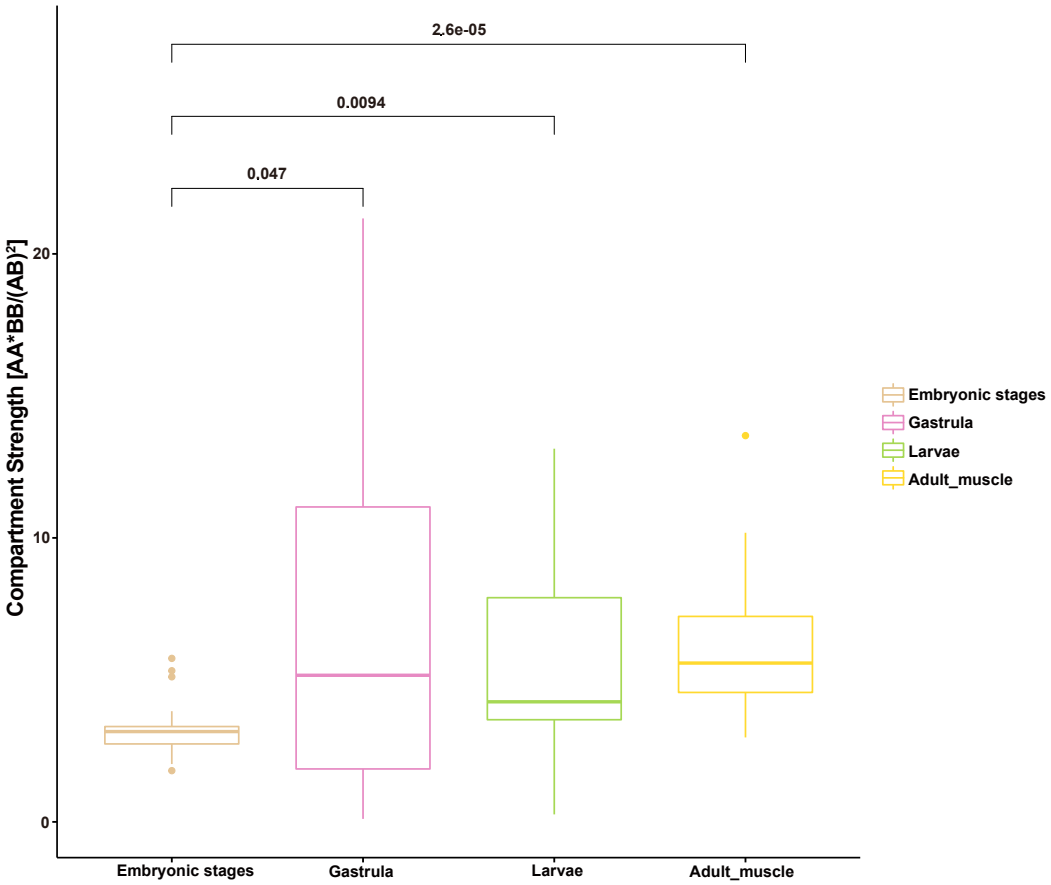


Fig. S32. A/B compartment strength between different developmental stages

Compartment strengths were calculated as $AA*BB/AB^2$ for each chromosome. The value in embryonic stages is the average value of compartments strength identified in 1-cell, 32-cell and 64-cell.

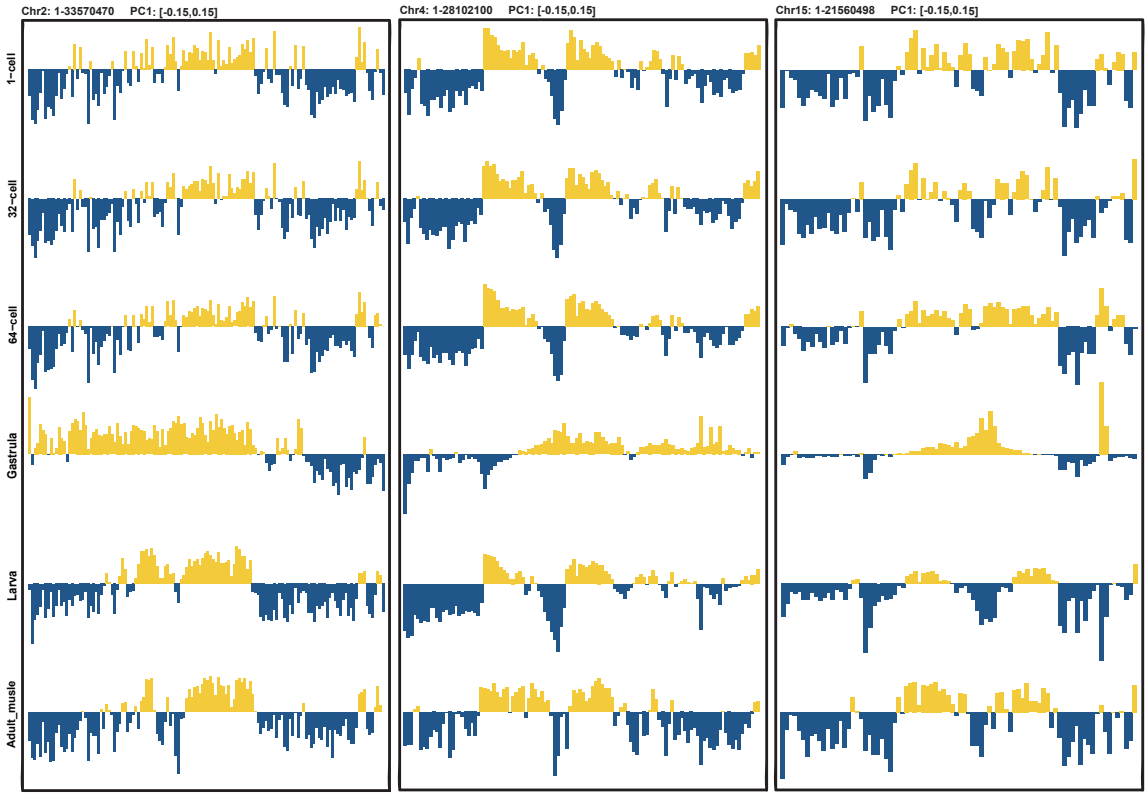


Fig. S33. Eigenvector 1 value tracks for amphioxus chromosomes at 250kb resolution at six different developmental stages.

Eigenvector 1 values (PC1, 250kb resolution) across the amphioxus chromosome 2, chromosome 4 and chromosome 15, representing A (yellow) and B (blue) compartments. On some chromosomes, e.g., chr2, compartment strength is significantly reduced at the gastrula stage.

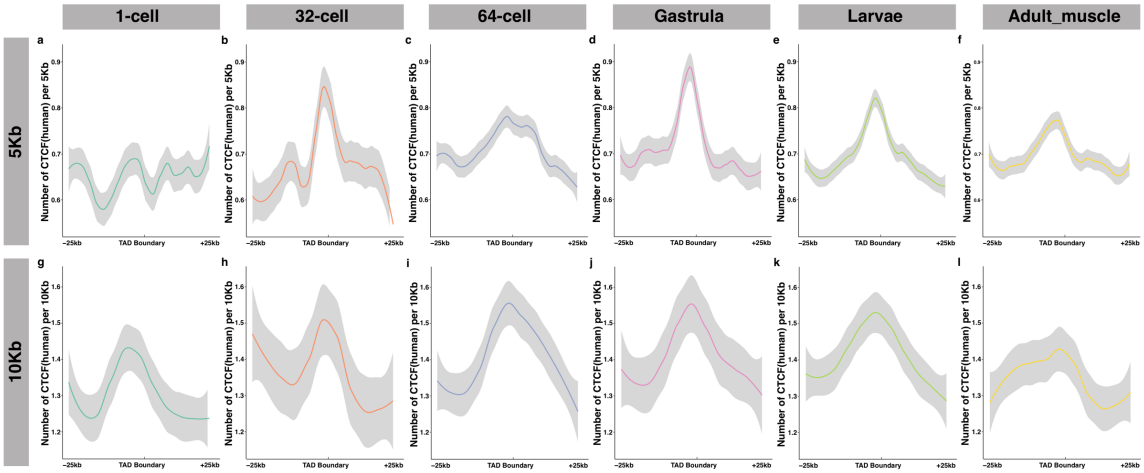


Fig. S34. Enrichment of putative CTCF-binding sites at TAD boundaries in the amphioxus genome.

a-f) Putative CTCF binding site density was calculated in 5kb windows overlapping each other 50% along the up- and down-stream (25kb) regions of amphioxus TAD boundaries. For each window, the mean value of the CTCF density was used to generate the plots. Panel a-f represent 1-cell, 32cell, 64-cell, gastrula, larvae and adult muscle respectively.

g-l) The same as panel a-f but in 10kb resolution.

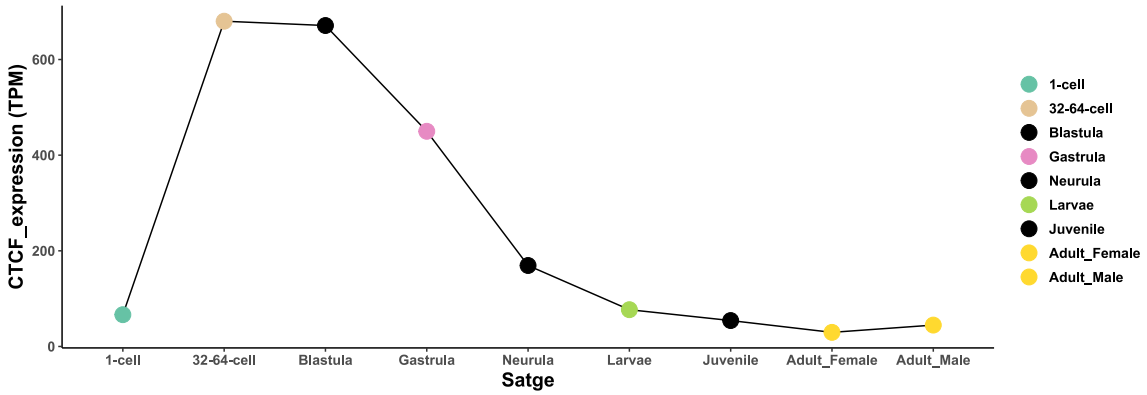


Fig. S35. Expression of CTCF in Bf during developmental stages.

The expression of the CTCF gene is increased at zygotic genome activation (ZGA) around the 64-cell stage.

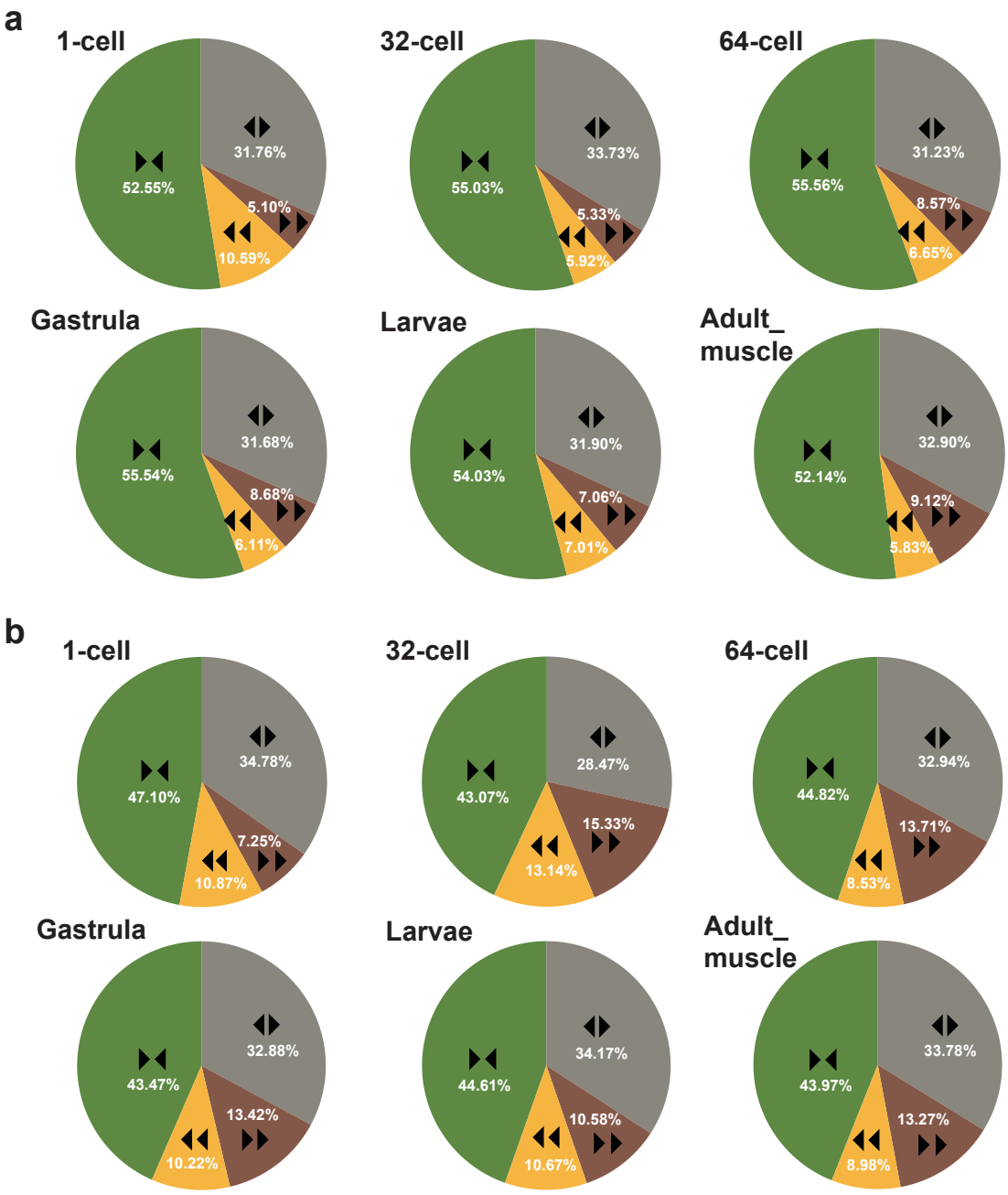


Fig. S36. The orientation of CTCF-binding sites pairs in amphioxus TABs.

a) Pie chart shows four types of orientation (arrowheads and different colors, green color for the convergent CTCF site pairs) for the paired CTCF sites in the TABs. The TABs are identified at the 5kb resolution and extended 5kb up- and down-stream, respectively. The convergent CTCF-binding site pairs are the major type(>52%).

b) The same as a, but in 10kb resolution. The ratio of convergent CTCF-binding site pairs reduced to about 45%.

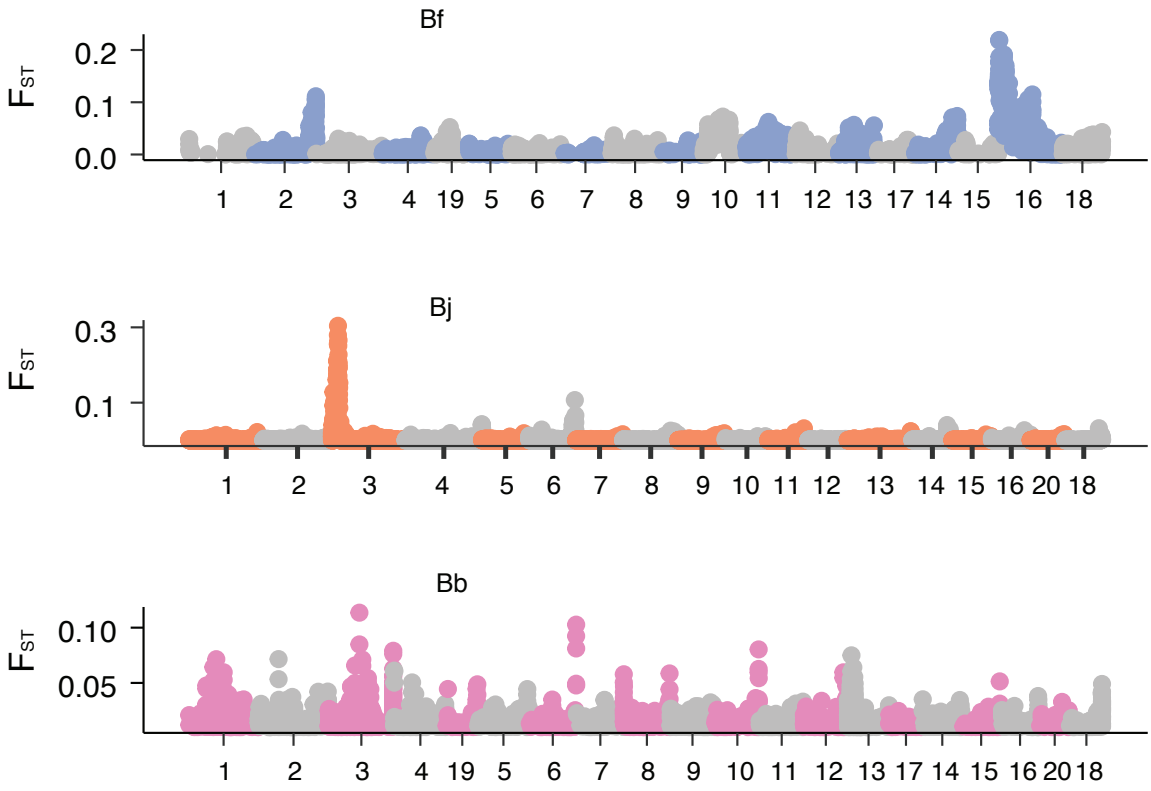


Fig. S37. The F_{ST} statistics between male and female populations.

Only biallelic SNPs were considered for F_{ST} calculation. The F_{ST} values were estimated in 10 kb windows for Bj and Bf, but 5 kb for Bb because of the smaller size of the non-recombining region (~1.5 kb) in Bb. The highest peaks of F_{ST} appear at Chr16, Chr3 and Chr3 of Bf, Bj, and Bb respectively. The number of individuals re-sequenced can be found in the Supplementary Table S5.

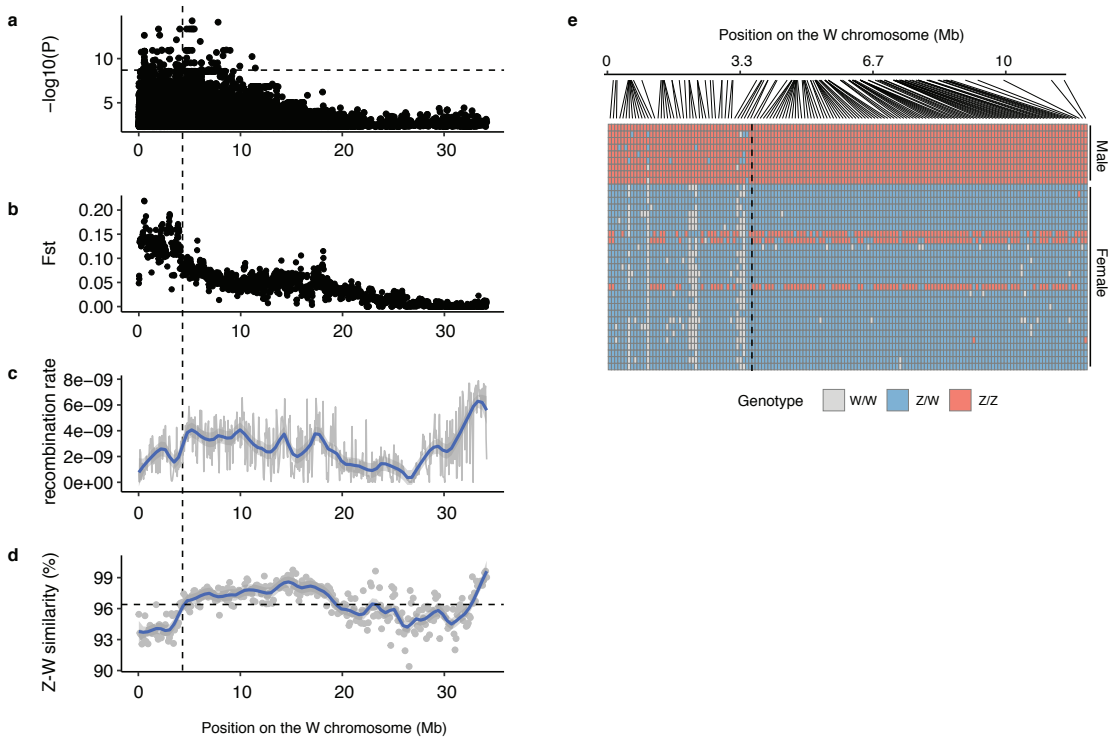


Fig. S38. The sex-linked region in Bf.

a) A zoom-in view of Fig. 5a on the Chr16 (ChrW). b) A zoom-in view of Fig. S37 on the Chr16 chromosome. c) This panel was retrieve from Fig. S7 for Bf Chr16. d) We used Lastz to align the ChrZ and ChrW sequences, and calculated the sequence similarity in 100 kb windows (grey dot). The blue line shows the smoothed mean values. The horizontal dashed line shows the mean sequence similarity between the Bf and Bf_2 genomes. The vertical dashed line shows the boundary of the first evolutionary stratum. e) The position and genotype of the female-associated SNPs. Each column in the lower panel represents a SNP site and each raw represents an individual. The vertical dashed line shows the boundary of the first evolutionary stratum.

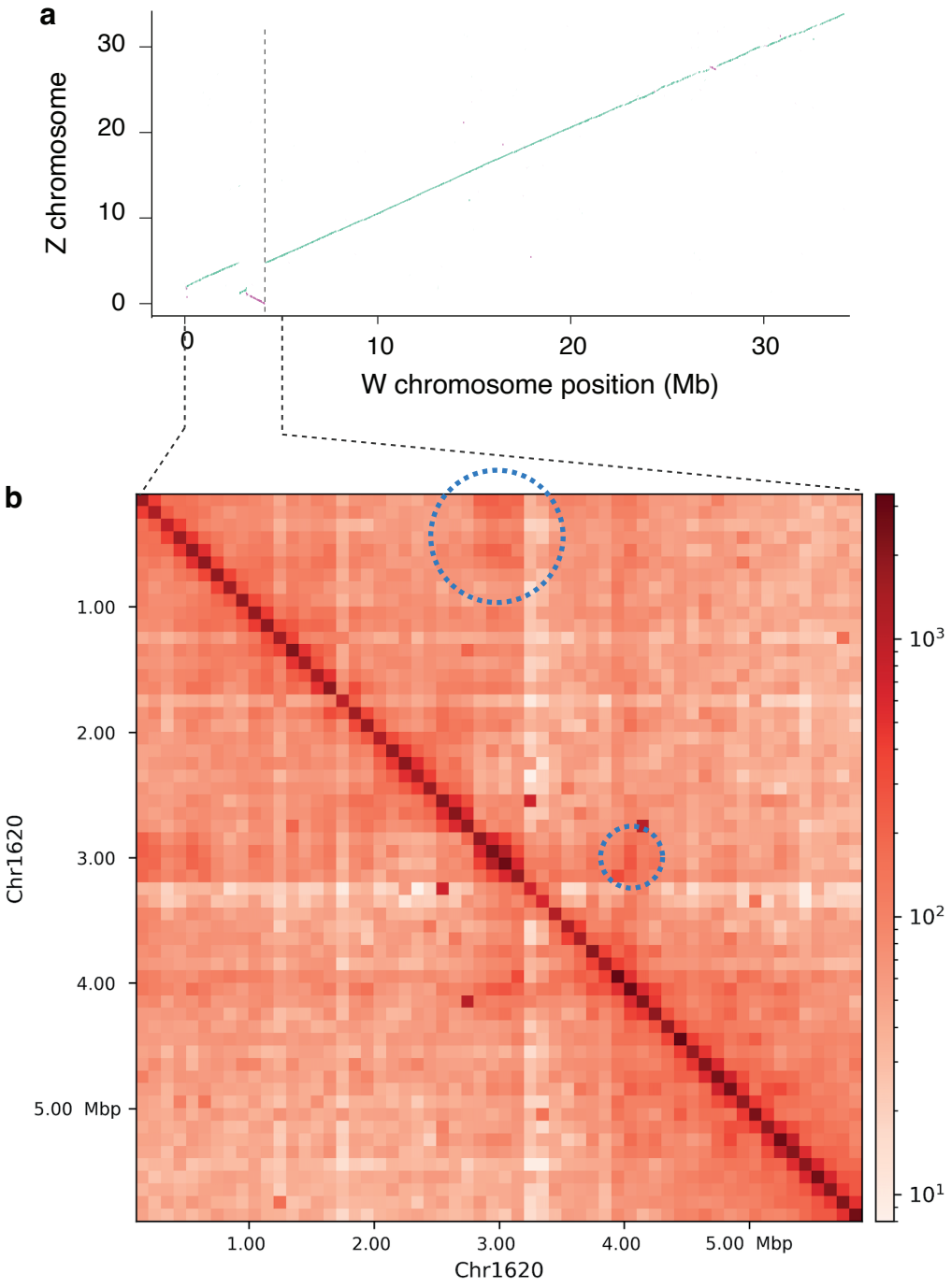


Fig. S39. the inversion between the Z and W on the oldest stratum.

a) a dot-plot figure for the Z-W chromosome alignment, similar to the figure 5e. b) The Hi-C interaction heatmap for the first 6 Mb of the W chromosome. The blue dashes circles point to the signals of inversions at 3.2 Mb and 3.9 Mb.

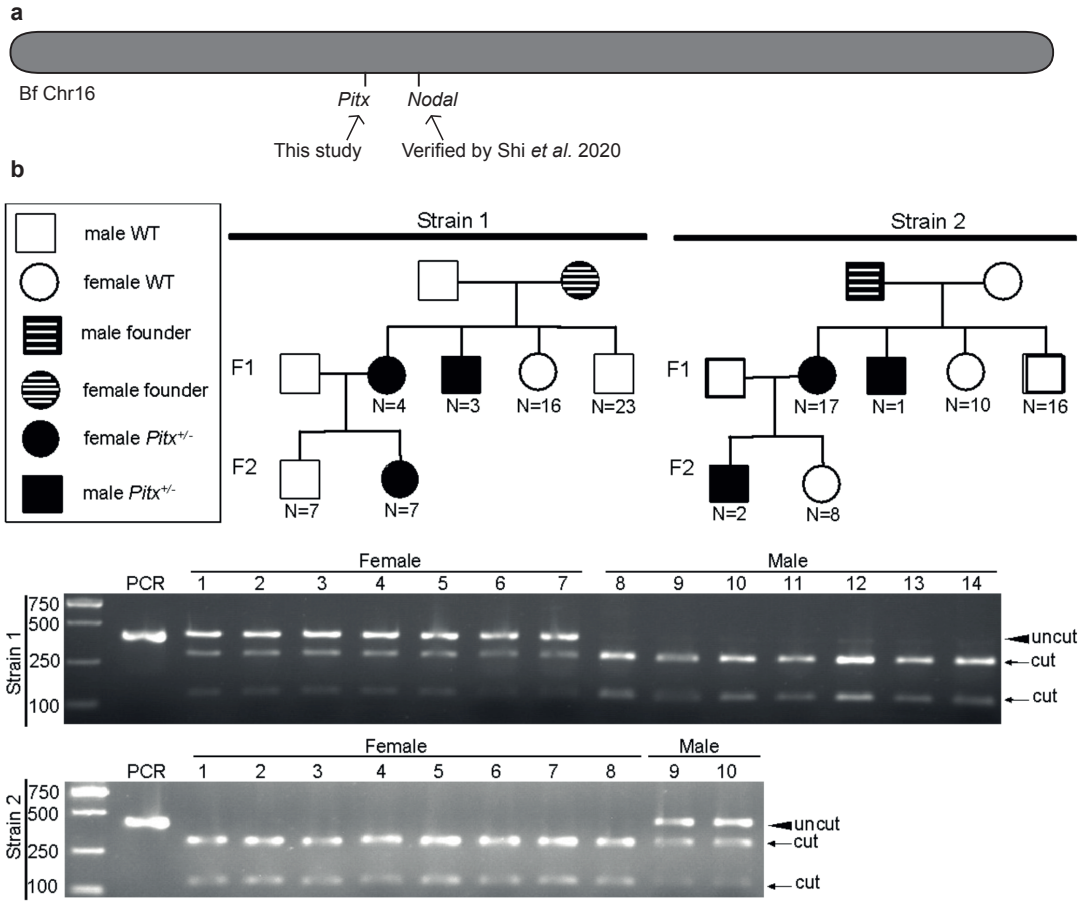


Fig. S40. Experiment design to verify the sex-linkage of *Pitx* in *B. floridae*.

a) Both *Pitx* and *Nodal* are located on the ZW chromosome (Chr16) in Bf.

b) Pedigree and genotyping of *Pitx* mutant strains. The mutation types carried in $Pitx^{+/-}$ heterozygotes of this figure include frame shift and non-frame shift mutations. The animal numbers analyzed are noted. Strain 1: a female founder injected with TALEN pair Pitx-Fw3/Rv3 (primer Fw3: 5'-GCAACCGTTCGACGAC-3' and Rv3 5'-TGTAGGCCGGCGAGTA-3') was crossed with a WT male. One F1 female $Pitx^{+/-}$ heterozygote carrying mutation (-14 bp) was crossed a WT male.

Strain 2: a male founder injected was crossed with a WT female. One F1 female $Pitx^{+/-}$ heterozygote carrying mutation (-7 bp, +18 bp) was crossed with a WT male. Genotyping analysis of F2 progenies from both strain 1 and 2 is present. Amplicons from were digested with *Tat*I enzyme and separated on 2% agarose gel to identify their genotypes.

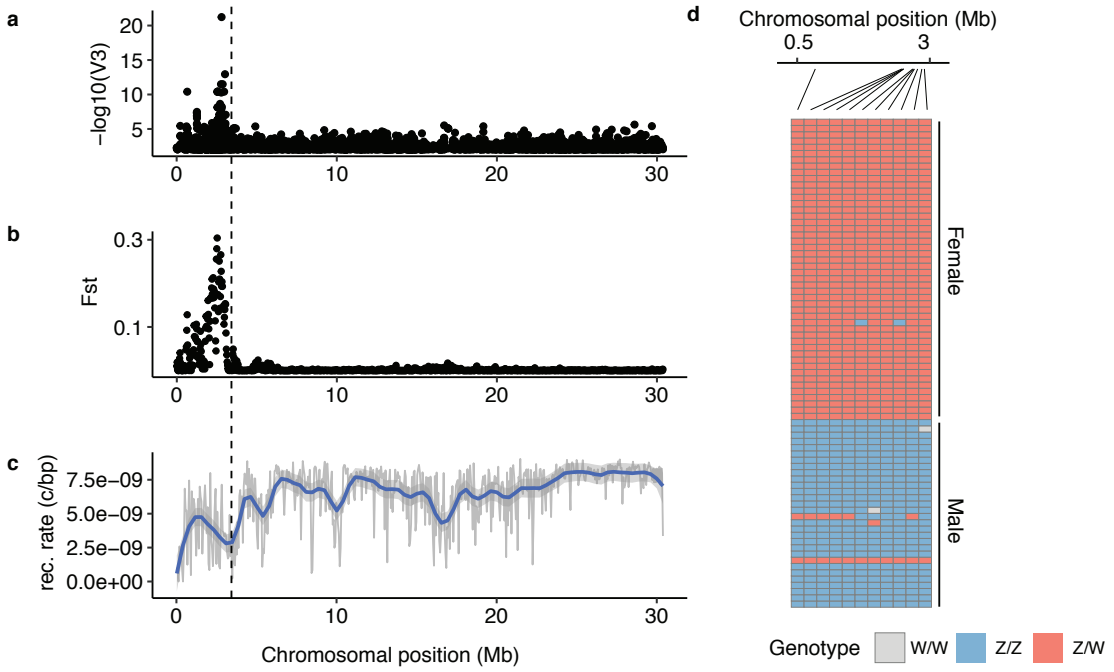


Fig. S41. The candidate sex determining region of Bj.

a) A zoom-in view of Fig. 5b on the Chr3 (ChrZ). b) A zoom-in view of Supplementary Fig. S37 on the Chr3 chromosome of Bj. c) This panel (recombination rate) was retrieved from Supplementary Fig. S7 for Bj Chr3. The vertical dashed line indicates the boundary of the sex-linked region. d) The position and genotype of the female-associated SNPs. Each column in the lower panel represents a SNP site and each row represents one individual.

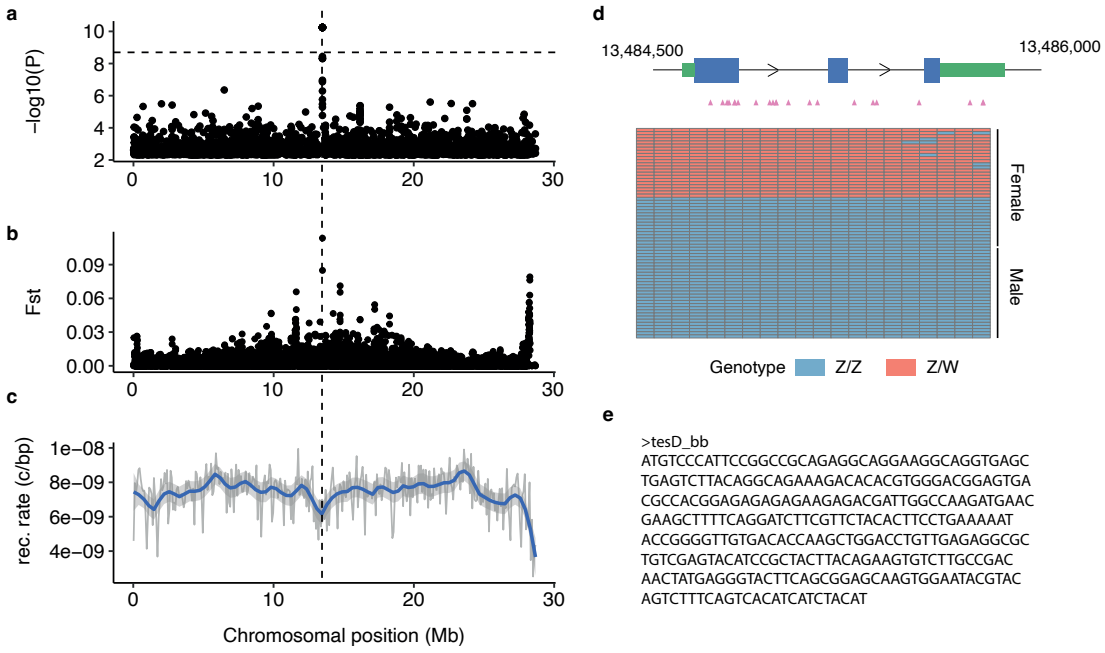


Fig. S42. The candidate sex determining region of Bb.

a) A zoom-in view of Fig. 5c on the Chr3 (ChrZ). b) A zoom-in view of Supplementary Fig. S37 on the Chr3 chromosome of Bb. c) This panel (recombination rate) was retrieved from Supplementary Fig. S7 for Bb Chr3. The vertical dashed line indicates the position of the Bb sex-linked locus. d) The sex-linked variants (pink triangles) are all located in the gene body of the Bb candidate sex-determining gene *tesD*. Each column in the lower panel represents a SNP site and each row represents an individual. The SNP sites are ordered according to their position in the genome. e) the coding sequence of *testD*.

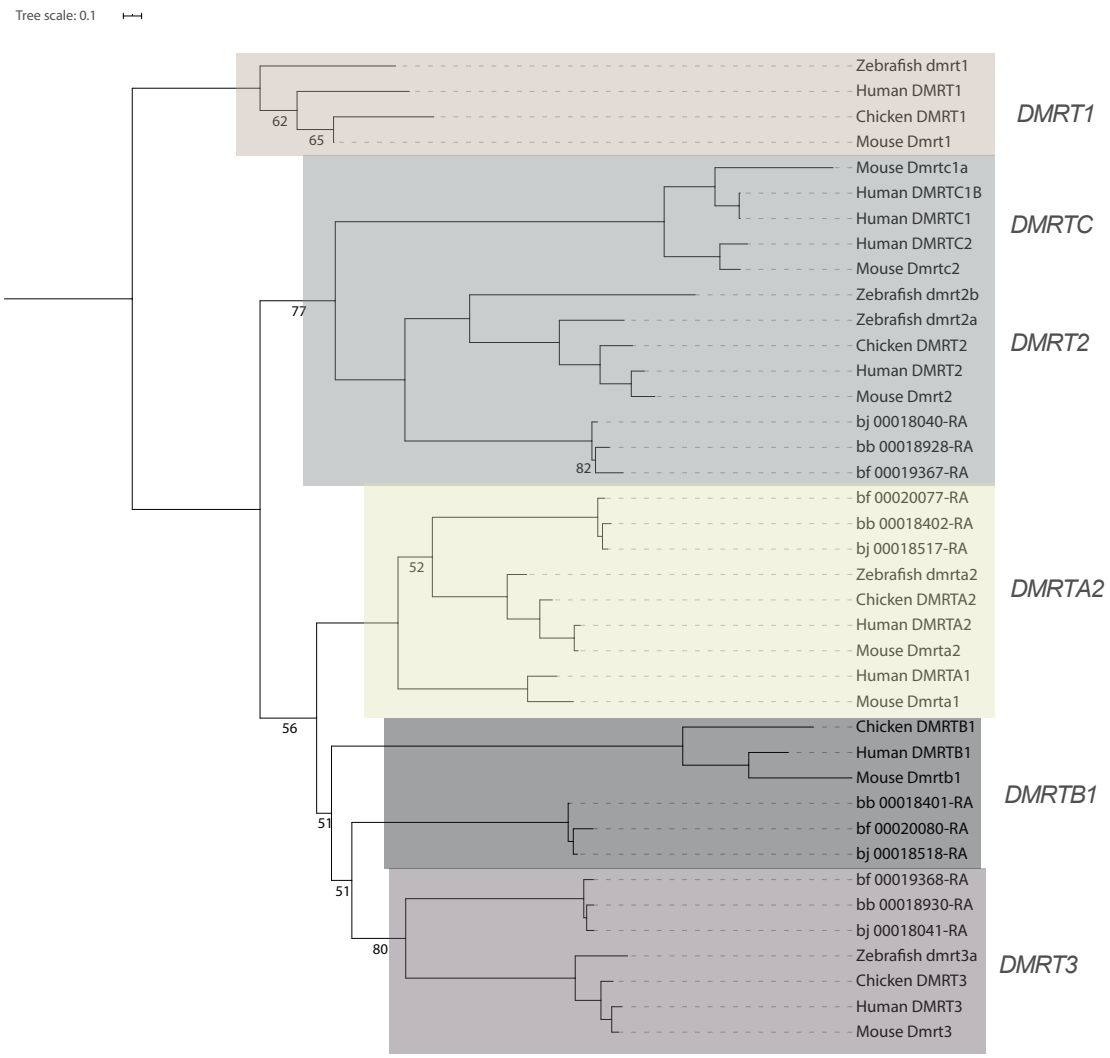


Fig. S43. The phylogeny of DMRT family genes across chordates.

The phylogeny was built from the coding sequences of the DMRT family genes of Bj, Bb, Bf, human, mouse, chicken and zebrafish. The root of DMRT family phylogeny was decided according to (2). The same families were highlighted in the same colored background. Bootstrapping values higher than 90 were not labelled at the nodes. Amphioxus lacks the *DMRT1* gene.

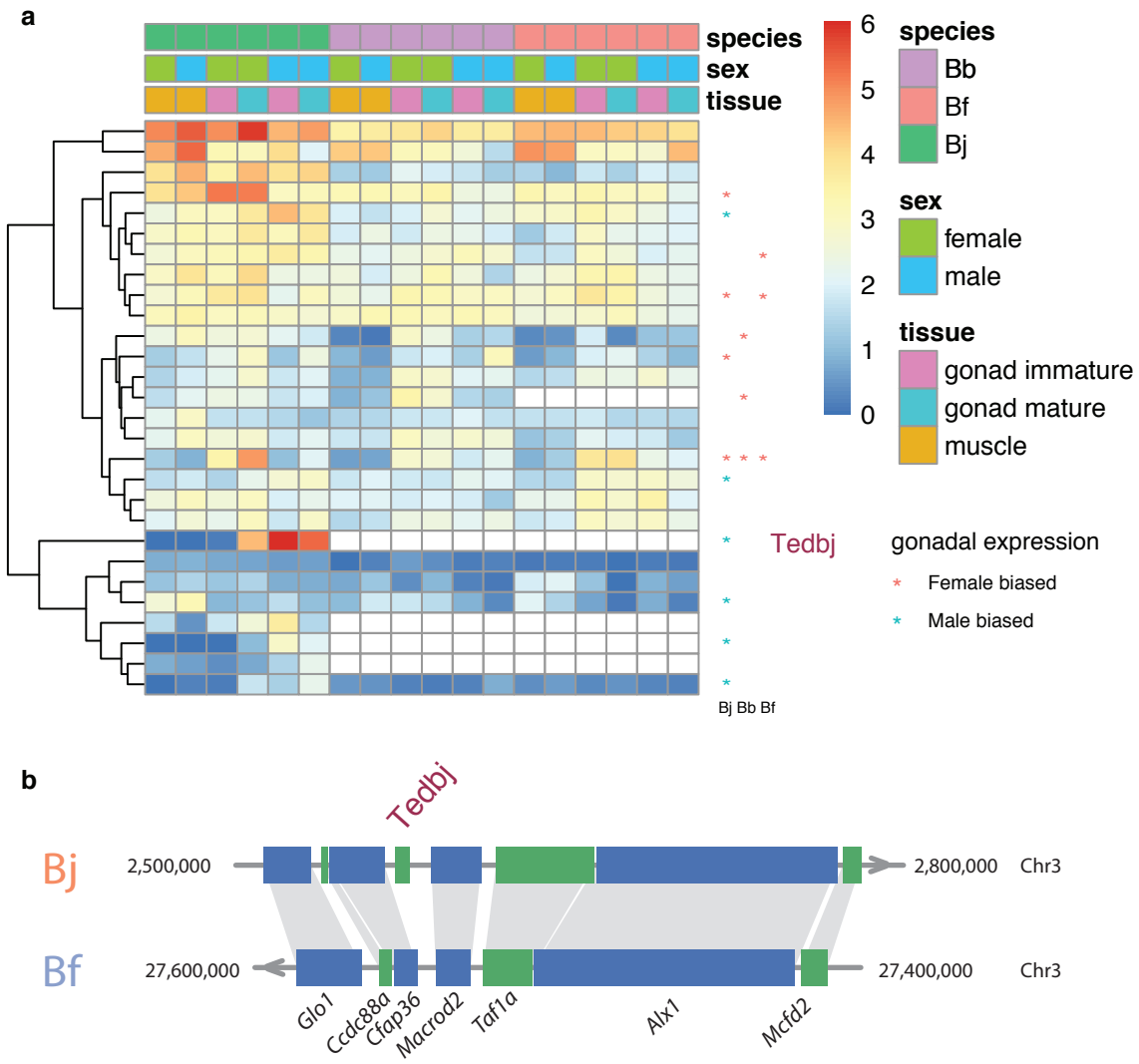


Fig. S44. The expression profile of sex-linked genes in Bj.

a) The heatmap shows the expression levels of the sex-linked genes of Bj stratum I in immature and mature gonads, as well as muscle, in three amphioxuses. The expression levels were log1p transformed. The blank tiles in the Bb and Bj panels represent those genes whose orthologs are absent in certain species. The asterisks at the end of the rows indicate whether the genes show differential expression between the immature gonads of males and females in Bj, Bb and Bf. *Tedbj* was highlighted which has a strong testis-biased expression.

b) Gene synteny around *Tedbj* between Bj and Bf. *Tedbj* is found to be specific to Bj which is also lacking in Bb (not shown).

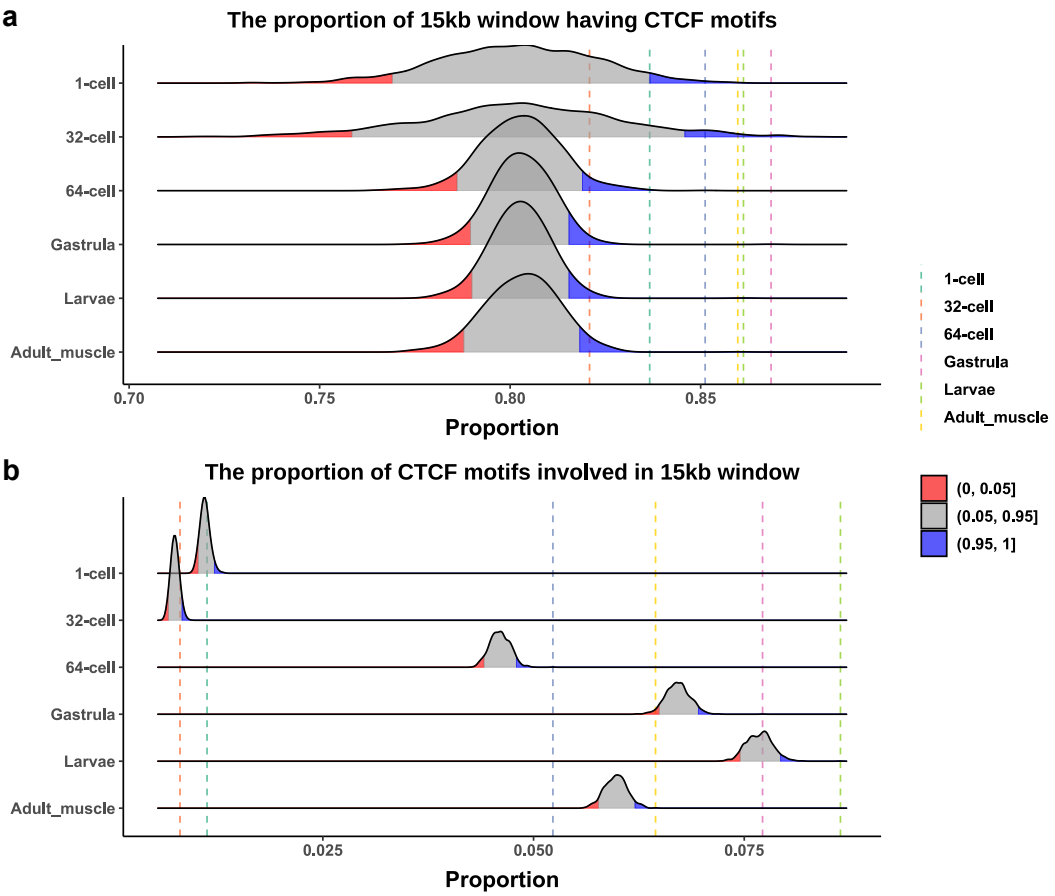


Fig. S45. Significance test of CTCF enrichment at TABs.

a) For each developmental stage, we randomly selected 15kb windows with the same number of TABs across the genome and calculated the proportion of 15kb windows having CTCF motifs. We repeated this step for 1000 times and obtained the density of all numbers of proportions. The red part and blue part showed the lowest 5% and highest 5% proportions, respectively. The proportion of TABs having CTCF motifs of each stage was indicated by vertical dotted lines in different colors.

b) We used the same significance test method as a), but we calculated the proportion of CTCF motifs involved in 15kb windows.

Table S1. Statistics of genome assembly

Species	Read type	Assembly	Genome size (Mb)	NG50 contig (Mb)	NG50 scaffold (Mb)	% assigned to chromosomes	BUSCO complete (%)	BUSCO duplicate (%)	BUSCO fragment (%)	BUSCO missing (%)
Branchiostoma floridae	Long read	Bf-hap	490.3	7.6	25.5	99.5	97.7	1.1	0.5	1.8
Branchiostoma floridae	Long read	Bf-hap-2	490.8	6.4	24.6	98.7	97.9	1.2	0.3	1.8
Branchiostoma belcheri	Long read	Bb-hap	406.5	13.9	19.7	98.8	97.4	1.2	0.6	2
Branchiostoma japonicum	Long read	Bj-hap	382.0	14.2	19.6	99.1	97.4	1	0.6	2
Branchiostoma floridae	Short read	Bfl_VNyyK (Bf_2020)	513.4	0.0504	25.4	94.2	95.6	1	1.3	3.1
Branchiostoma floridae	Short read	Version 2 (Bf_2008)	521.9	0.0279	2.6	-	93.4	4	2.1	4.5
Branchiostoma belcheri	Short read	Haploidv1 8h27 (Bb_2014)	426.1	0.0472	2.3	-	97.2	1.4	0.7	2.1
Branchiostoma lanceolatum	Short read	Bl71nemr (Bl_2018)	495.2	0.0533	1.3	-	95.8	2.7	2.5	1.7

Table S2. Estimations for the divergence time using orthologous genes.

Node name	constraints [min, max] (Mya)	Mean	95% CI (Mya)
chordate	[514, 636.1]	592.5	[530,8 640.3]
Vertebrate	NA	553.1	[494.3, 614.7]
Gnathostomata	[420.7, 468.4]	453.8	[429.9, 470.4]
Amniota	[318, 332.9]	325.3	[317.9, 322.8]
Actinopterygii	[378.19, 422.4]	389.8	[375.2, 411.3]
Euarchontoglires	[65.6, 164.6]	116.1	[72.1, 163.1]
Branchiostoma	NA	99.9	[22.1, 211.7]
Bb/Bj	NA	59.3	[12.2, 132.0]

Table S3. GO enrichment for lineage-specific orthologs.

Taxa	GO term	GO description	p value	FDR
Vertebrate	GO:0007155	cell adhesion	9.47E-20	1.44E-16
	GO:0007204	positive regulation of cytosolic calcium ion concentration	6.43E-15	3.26E-12
	GO:0007218	neuropeptide signaling pathway	6.13E-14	2.33E-11
	GO:0005179	hormone activity	4.10E-12	1.04E-09
	GO:0004896	cytokine receptor activity	5.94E-12	1.29E-09
	GO:0007267	cell-cell signaling	1.07E-11	2.17E-09
	GO:0007200	phospholipase C-activating G protein-coupled receptor signaling pathway	1.66E-11	3.05E-09
	GO:0019221	cytokine-mediated signaling pathway	5.63E-11	9.51E-09
	GO:0001525	angiogenesis	2.24E-10	3.24E-08
	GO:0001227	DNA-binding transcription repressor activity, RNA polymerase II-specific	3.32E-10	4.39E-08
	GO:0008201	heparin binding	2.83E-09	3.44E-07
	GO:0008284	positive regulation of cell population proliferation	3.92E-09	4.51E-07
	GO:0008083	growth factor activity	4.00E-09	4.51E-07
	GO:0001228	DNA-binding transcription activator activity, RNA polymerase II-specific	4.58E-09	4.97E-07
	GO:0005102	signaling receptor binding	7.06E-09	7.28E-07
	GO:0019955	cytokine binding	7.18E-09	7.28E-07
	GO:0007156	homophilic cell adhesion via plasma membrane adhesion molecules	9.96E-09	9.17E-07
	GO:0005184	neuropeptide hormone activity	1.34E-08	1.14E-06
	GO:0018146	keratan sulfate biosynthetic process	1.34E-08	1.14E-06
	GO:0007601	visual perception	1.68E-08	1.38E-06
	GO:0044212	transcription regulatory region DNA binding	2.27E-08	1.80E-06
	GO:0035025	positive regulation of Rho protein signal transduction	2.37E-08	1.80E-06
	GO:0007268	chemical synaptic transmission	2.37E-08	1.80E-06
	GO:0035589	G protein-coupled purinergic nucleotide receptor signaling pathway	5.76E-08	4.27E-06
	GO:0005243	gap junction channel activity	6.09E-08	4.31E-06
	GO:0038023	signaling receptor activity	7.56E-08	5.07E-06
	GO:0070374	positive regulation of ERK1 and ERK2 cascade	7.66E-08	5.07E-06
	GO:0045028	G protein-coupled purinergic nucleotide receptor activity	8.96E-08	5.80E-06
	GO:0051482	positive regulation of cytosolic calcium ion concentration involved in phospholipase C-activating G protein-coupled signaling pathway	1.06E-07	6.73E-06
	GO:0005581	collagen trimer	1.18E-07	7.34E-06
	GO:0030335	positive regulation of cell migration	1.66E-07	1.01E-05
	GO:0016338	calcium-independent cell-cell adhesion via plasma membrane cell-adhesion molecules	2.94E-07	1.75E-05
	GO:0016339	calcium-dependent cell-cell adhesion via plasma membrane cell adhesion molecules	3.29E-07	1.92E-05
	GO:0005518	collagen binding	3.51E-07	2.01E-05
	GO:0019233	sensory perception of pain	3.70E-07	2.05E-05
	GO:0019838	growth factor binding	3.95E-07	2.15E-05
	GO:0044267	cellular protein metabolic process	4.83E-07	2.58E-05
	GO:0051965	positive regulation of synapse assembly	5.22E-07	2.70E-05
	GO:0030021	extracellular matrix structural constituent conferring compression resistance	5.33E-07	2.70E-05
	GO:0008013	beta-catenin binding	1.08E-06	5.35E-05
	GO:0007189	adenylate cyclase-activating G protein-coupled receptor signaling pathway	1.11E-06	5.35E-05
	GO:0003429	growth plate cartilage chondrocyte morphogenesis	1.33E-06	6.31E-05
	GO:0042730	fibrinolysis	1.38E-06	6.35E-05
	GO:0097755	positive regulation of blood vessel diameter	1.38E-06	6.35E-05
	GO:0051897	positive regulation of protein kinase B signaling	1.47E-06	6.58E-05
	GO:0007162	negative regulation of cell adhesion	1.69E-06	7.33E-05
	GO:0007015	actin filament organization	1.74E-06	7.45E-05
	GO:0007422	peripheral nervous system development	1.98E-06	8.37E-05
	GO:0043406	positive regulation of MAP kinase activity	3.21E-06	1.32E-04
	GO:0002576	platelet degranulation	3.38E-06	0.000137079
	GO:0007565	female pregnancy	3.60E-06	0.000144093
	GO:0016525	negative regulation of angiogenesis	3.94E-06	0.000155641
	GO:0042060	wound healing	4.20E-06	0.000163764
	GO:0007631	feeding behavior	5.89E-06	0.000224042
	GO:0051930	regulation of sensory perception of pain	5.89E-06	0.000224042
	GO:0016327	apicolateral plasma membrane	6.27E-06	0.00023528
	GO:0070102	interleukin-6-mediated signaling pathway	6.51E-06	0.000241426
	GO:0008092	cytoskeletal protein binding	7.37E-06	0.000270133
	GO:0006954	inflammatory response	8.32E-06	0.000297838
	GO:0008285	negative regulation of cell population proliferation	8.49E-06	0.000300061
	GO:0002687	positive regulation of leukocyte migration	1.12E-05	0.000392836

Taxa	GO term	GO description	p value	FDR
	GO:0007409	axonogenesis	1.24E-05	0.000429782
	GO:0043194	axon initial segment	1.39E-05	0.000475368
	GO:0000902	cell morphogenesis	1.45E-05	0.0004897
	GO:0035633	maintenance of blood-brain barrier	1.55E-05	0.00051837
	GO:0016324	apical plasma membrane	1.68E-05	0.000554027
	GO:0043005	neuron projection	2.10E-05	0.000686843
	GO:0098982	GABA-ergic synapse	2.20E-05	0.000712294
	GO:0005104	fibroblast growth factor receptor binding	2.26E-05	0.000723652
	GO:0007229	integrin-mediated signaling pathway	2.37E-05	0.000752044
	GO:0009612	response to mechanical stimulus	2.54E-05	0.000795008
	GO:0031093	platelet alpha granule lumen	3.34E-05	0.001035399
	GO:0050731	positive regulation of peptidyl-tyrosine phosphorylation	3.45E-05	0.001048882
	GO:0034113	heterotypic cell-cell adhesion	3.45E-05	0.001048882
	GO:0007411	axon guidance	3.52E-05	0.0010591
	GO:0002040	sprouting angiogenesis	3.69E-05	0.001090492
	GO:0006953	acute-phase response	3.73E-05	0.001090492
	GO:0071392	cellular response to estradiol stimulus	3.73E-05	0.001090492
	GO:0005540	hyaluronic acid binding	3.98E-05	0.001151542
	GO:0008360	regulation of cell shape	4.08E-05	0.001171766
	GO:0005178	integrin binding	4.29E-05	0.00121105
	GO:0014068	positive regulation of phosphatidylinositol	4.30E-05	0.001417713
	GO:0032496	response to lipopolysaccharide	4.48E-05	1.24E-03
	GO:0007193	adenylate cyclase-inhibiting G protein-coupled receptor signaling pathway	4.49E-05	0.001240922
	GO:0002682	regulation of immune system process	5.12E-05	0.001390149
	GO:0030594	neurotransmitter receptor activity	5.19E-05	0.001395936
	GO:0051378	serotonin binding	6.18E-05	0.001634395
	GO:0045295	gamma-catenin binding	6.83E-05	0.001789526
	GO:0001817	regulation of cytokine production	7.16E-05	0.001862134
	GO:0008528	G protein-coupled peptide receptor activity	7.71E-05	0.001985917
	GO:0050877	nervous system process	7.81E-05	0.001995776
	GO:0006936	muscle contraction	7.97E-05	0.002019753
	GO:0003779	actin binding	8.89E-05	0.002215183
	GO:0019957	C-C chemokine binding	1.09E-04	0.002648187
	GO:0042923	neuropeptide binding	1.09E-04	0.002648187
	GO:0007166	cell surface receptor signaling pathway	1.12E-04	0.00271518
	GO:0005242	inward rectifier potassium channel activity	1.16E-04	0.002787667
	GO:0035987	endodermal cell differentiation	1.18E-04	0.002805401
	GO:0035914	skeletal muscle cell differentiation	1.34E-04	0.003167765
	GO:0005865	striated muscle thin filament	1.41E-04	0.003263933
	GO:0050930	induction of positive chemotaxis	1.41E-04	0.003263933
	GO:0070757	interleukin-35-mediated signaling pathway	1.50E-04	0.003462637
	GO:0090050	positive regulation of cell migration involved in sprouting angiogenesis	1.52E-04	0.003465429
	GO:0019956	chemokine binding	1.71E-04	0.003852374
	GO:0045987	positive regulation of smooth muscle contraction	2.05E-04	0.00452882
	GO:0098641	cadherin binding involved in cell-cell adhesion	2.05E-04	0.00452882
	GO:0001501	skeletal system development	2.06E-04	0.00452882
	GO:0007596	blood coagulation	2.34E-04	0.005109771
	GO:0005044	scavenger receptor activity	2.40E-04	0.005184503
	GO:0051781	positive regulation of cell division	2.40E-04	0.005184503
	GO:0033627	cell adhesion mediated by integrin	2.45E-04	0.005241036
	GO:0018298	protein-chromophore linkage	2.66E-04	0.005579343
	GO:0035994	response to muscle stretch	2.66E-04	0.005579343
	GO:0008307	structural constituent of muscle	2.75E-04	0.005737607
	GO:0010628	positive regulation of gene expression	3.10E-04	0.00642159
	GO:0034707	chloride channel complex	3.16E-04	0.006425898
	GO:0030208	dermatan sulfate biosynthetic process	3.19E-04	0.006425898
	GO:0042340	keratan sulfate catabolic process	3.19E-04	0.006425898
	GO:0050927	positive regulation of positive chemotaxis	3.19E-04	0.006425898
	GO:0060078	regulation of postsynaptic membrane potential	3.45E-04	0.00689942
	GO:0001666	response to hypoxia	3.54E-04	0.007043744
	GO:0016493	C-C chemokine receptor activity	3.81E-04	0.007388006
	GO:0042755	eating behavior	3.81E-04	0.007388006
	GO:0045672	positive regulation of osteoclast differentiation	3.81E-04	0.007388006
	GO:0051602	response to electrical stimulus	3.87E-04	0.007453419
	GO:0005604	basement membrane	3.90E-04	0.007453419

Taxa	GO term	GO description	p value	FDR
	GO:0042531	positive regulation of tyrosine phosphorylation of STAT protein	4.00E-04	0.007607192
	GO:0098978	glutamatergic synapse	4.18E-04	0.007858617
	GO:0045766	positive regulation of angiogenesis	4.21E-04	0.007858617
	GO:0005198	structural molecule activity	4.21E-04	0.007858617
	GO:0000165	MAPK cascade	4.51E-04	0.008342114
	GO:0007417	central nervous system development	4.53E-04	0.008342114
amphioxus	GO:0005230	extracellular ligand-gated ion channel activity	4.34E-26	3.44E-24
	GO:0034220	ion transmembrane transport	7.38E-26	4.39E-24
	GO:0004888	transmembrane signaling receptor activity	8.25E-24	3.93E-22
	GO:0004970	ionotropic glutamate receptor activity	4.74E-18	1.88E-16
	GO:0038023	signaling receptor activity	3.32E-14	1.13E-12
	GO:0008373	sialyltransferase activity	1.77E-13	4.87E-12
	GO:0004983	neuropeptide Y receptor activity	1.92E-11	3.80E-10
	GO:0006486	protein glycosylation	2.77E-10	5.07E-09
	GO:0051260	protein homooligomerization	4.08E-10	6.93E-09
	GO:0042981	regulation of apoptotic process	7.69E-09	1.14E-07
	GO:0004252	serine-type endopeptidase activity	7.97E-08	9.99E-07
	GO:0007160	cell-matrix adhesion	3.71E-06	4.20E-05
	GO:0002020	protease binding	1.81E-05	0.000187172
	GO:0070513	death domain binding	1.81E-05	0.000187172
	GO:0004867	serine-type endopeptidase inhibitor activity	2.16E-05	0.000205676
	GO:0071805	potassium ion transmembrane transport	2.16E-05	0.000205676
	GO:0008237	metallopeptidase activity	3.07E-05	0.000280703
	GO:0008188	neuropeptide receptor activity	4.78E-05	0.000417836
	GO:0006508	proteolysis	4.92E-05	0.000417836
	GO:0007165	signal transduction	6.04E-05	0.000495163
	GO:0009190	cyclic nucleotide biosynthetic process	0.000113832	0.000873549
	GO:0016849	phosphorus-oxygen lyase activity	0.000113832	0.000873549
	GO:0004857	enzyme inhibitor activity	0.000180566	0.001310251
	GO:0001733	galactosylceramide sulfotransferase activity	0.000187261	0.001310251
	GO:0009247	glycolipid biosynthetic process	0.000187261	0.001310251
	GO:0005267	potassium channel activity	0.000275214	0.001870626
	GO:0030414	peptidase inhibitor activity	0.00037529	0.002479984
	GO:0042246	tissue regeneration	0.000390304	0.002509495
	GO:0004222	metalloendopeptidase activity	0.000466924	0.002923126
	GO:0004553	hydrolase activity, hydrolyzing O-glycosyl compounds	0.001027761	0.006269204
	GO:0016702	oxidoreductase activity, acting on single donors with incorporation of molecular oxygen, incorporation of two atoms of oxygen	0.001101749	0.006552506
	GO:0005125	cytokine activity	0.001209654	0.007018787
	GO:0004383	guanylate cyclase activity	0.001403582	0.007765229
	GO:0006182	cGMP biosynthetic process	0.001403582	0.007765229

Table S4. number of amphixous orthologous groups with and without vertebrate ohnologs.

	all gene	amphioxus duplicates	all gene	amphioxus duplicates	P-value of Fisher's exact test
	with human ohnolog		without human ohnolog		
Bb	2198	284	6132	433	1.70E-13
Bf		303		511	1.43E-10
Bj		279		403	5.29E-15
	with chicken ohnolog		without chicken ohnolog		
Bb	1973	294	5949	397	6.38E-22
Bf		330		459	7.32E-23
Bj		286		370	2.28E-23

Table S5. GO enrichment for genes in amphixous segmental duplications.

Species	GO term	GO description	p value	FDR
Bf	GO:0015074	DNA_integration	9.70E-14	5.50E-11
	GO:0006310	DNA_recombination	3.90E-09	1.10E-06
	GO:0004713	protein_tyrosine_kinase_activity	1.30E-17	6.20E-15
	GO:0003676	nucleic_acid_binding	5.80E-15	1.40E-12
	GO:0008270	zinc_ion_binding	5.10E-07	8.40E-05
	GO:0004930	G-protein_coupled_receptor_activity	2.10E-05	0.0025
	GO:0003691	double-stranded_telomeric_DNA_binding	2.50E-05	0.0025
	GO:0003677	DNA_binding	7.60E-05	6.20E-03
Bj	GO:0015074	DNA_integration	5.50E-07	0.00034
	GO:0007186	G-protein_coupled_receptor_protein_signaling_pathway	4.00E-06	0.0012
	GO:0003001	generation_of_a_signal_involved_in_cell-cell_signaling	8.40E-05	0.0095
	GO:0006836	neurotransmitter_transport	8.40E-05	0.0095
	GO:0006310	DNA_recombination	9.10E-05	0.0095
	GO:0007269	neurotransmitter_secretion	5.50E-05	0.0095
	GO:0008773	[protein-PII]_uridylyltransferase_activity	4.10E-08	2.00E-05
	GO:0070569	uridylyltransferase_activity	8.90E-08	2.20E-05
	GO:0004930	G-protein_coupled_receptor_activity	6.50E-07	0.00011
	GO:0016779	nucleotidyltransferase_activity	8.90E-06	0.0011
Bb	GO:0015074	DNA_integration	1.50E-15	8.00E-13
	GO:0060249	anatomical_structure_homeostasis	1.50E-10	1.60E-08
	GO:0000723	telomere_maintenance	1.50E-10	1.60E-08
	GO:0006310	DNA_recombination	1.30E-10	1.60E-08
	GO:0032200	telomere_organization	1.50E-10	1.60E-08
	GO:0006468	protein_amino_acid_phosphorylation	1.30E-08	1.20E-06
	GO:0006259	DNA_metabolic_process	6.80E-08	5.20E-06
	GO:0016310	phosphorylation	4.90E-07	3.30E-05
	GO:0006464	protein_modification_process	5.70E-06	0.00034
	GO:0043412	macromolecule_modification	2.10E-05	0.0012
	GO:0006796	phosphate_metabolic_process	2.60E-05	0.0013
	GO:0006793	phosphorus_metabolic_process	2.80E-05	0.0013
	GO:0043413	macromolecule_glycosylation	6.10E-05	0.0022
	GO:0009101	glycoprotein_biosynthetic_process	6.60E-05	0.0022
	GO:0006486	protein_amino_acid_glycosylation	6.10E-05	0.0022
	GO:0070085	glycosylation	6.10E-05	0.0022
	GO:0009100	glycoprotein_metabolic_process	7.00E-05	0.0022
	GO:0043687	post-translational_protein_modification	0.00013	0.0038
	GO:0006493	protein_amino_acid_O-linked_glycosylation	0.00024	0.0068
	GO:0004713	protein_tyrosine_kinase_activity	3.00E-19	1.20E-16
	GO:0003676	nucleic_acid_binding	4.80E-09	9.30E-07
	GO:0003678	DNA_helicase_activity	6.80E-09	9.30E-07
	GO:0004386	helicase_activity	1.30E-08	1.30E-06
	GO:0004672	protein_kinase_activity	2.50E-08	2.10E-06
	GO:0016772	transferase_activity,_transferring_phosphorus-containing_groups	1.30E-07	9.00E-06
	GO:0016301	kinase_activity	2.20E-07	1.30E-05
	GO:0016773	phosphotransferase_activity,_alcohol_group_as_acceptor	2.60E-07	1.30E-05
	GO:0016740	transferase_activity	4.90E-05	0.0022

Table S6. QC of HiC data

Species	Development Stages	Paired Ends	Sequencing Depth(fold)	Total Read Pairs	Mapped Rates	Valid Pairs	PCR Dup Read Pairs	Cis/Trans Ratio	Estimated Resolution/Kb
Bf	**1-cell**	R1	81	131,530,545	62.13%	51.89%	17.25%	3.61	2.3
		R2			61.63%				
	32-cell	R1	84	136,779,099	63.65%	52.39%	14.95%	2.89	
		R2			63.16%				
	64-cell	R1	94	154,347,440	84.94%	60.72%	16.67%	3.43	2.05
		R2			84.39%				
	Gastrula	R1	61	100,186,189	93.60%	70.17%	14.17%	7.98	2.27
		R2			93.14%				
	Larvae	R1	108	175,939,611	94.79%	74.73%	14.39%	2.44	1.91
		R2			93.79%				
	Adult_muscle	R1	79	128,214,144	92.77%	68.49%	8.03%	2.03	2.2
		R2			92.14%				
BB	**Adult_muscle**	R1	90	121,294,118	94.71%	76.04%	7.72%	2.03	1.98
		R2			93.25%				
Bj	**Adult_muscle**	R1	84	106,681,598	96.45%	69.79%	9.23%	1.9	1.35
		R2			96.14%				

Table S7. Sequencing data produced in this study

Species	Sequencing type	SRA accession	title
BjBf	Pacbio	SRR11013311	Pacbio reads of the B. japonicum x B. floridae hybrid
BbBf	Pacbio	SRR11012663	Pacbio reads of B. floridae x B. belcheri hybrid
B. belcheri	Hi-C	SRR12007920	B. belcheri muscle HiC
B. floridae	Hi-C	SRR12007919	B. floridae muscle HiC
B. japonicum	Hi-C	SRR12007918	B. japonicum muscle HiC
B. floridae	Hi-C	SRR12059955	Bf 1 cell stage Hi-C
B. floridae	Hi-C	SRR12059954	Bf 32 cell stage Hi-C
B. floridae	Hi-C	SRR12059953	Bf 64 cell stage Hi-C
B. floridae	Hi-C	SRR12059952	Bf gastrulation Hi-C
B. floridae	Hi-C	SRR12059951	Bf Larval Hi-C
B. japonicum	ISO-seq	SRR11023658	iso-seq for bj female
B. japonicum	ISO-seq	SRR11023657	iso-seq for bj male
B. belcheri	ISO-seq	SRR11023656	iso-seq for bb male
B. belcheri	ISO-seq	SRR11023655	iso-seq for bb female
B. floridae	ISO-seq	SRR11023654	iso-seq for bf female
B. floridae	ISO-seq	SRR11023653	iso-seq for bf male
B. japonicum	RNA-seq	SRR12011621	Bj gonad-mature_male-1 RNA-seq
B. japonicum	RNA-seq	SRR12011620	Bj gonad-mature_male-2 RNA-seq
B. japonicum	RNA-seq	SRR12011609	Bj gonad-mature_male-3 RNA-seq
B. japonicum	RNA-seq	SRR12011598	Bj gonad-mature_female-2 RNA-seq
B. japonicum	RNA-seq	SRR12011587	Bj gonad-mature_female-1 RNA-seq
B. japonicum	RNA-seq	SRR12011576	Bj gonad-mature_female-3 RNA-seq
B. japonicum	RNA-seq	SRR12011565	Bj gonad-mature_female-4 RNA-seq
B. japonicum	RNA-seq	SRR12011554	Bj gonad-immature_male-2 RNA-seq
B. japonicum	RNA-seq	SRR12011548	Bj gonad-immature_male-1 RNA-seq
B. japonicum	RNA-seq	SRR12011547	Bj gonad-immature_male-3 RNA-seq
B. japonicum	RNA-seq	SRR12011619	Bj gonad-immature_female-4 RNA-seq
B. japonicum	RNA-seq	SRR12011618	Bj gonad-immature_female-5 RNA-seq
B. japonicum	RNA-seq	SRR12011617	Bj gonad-immature_female-1 RNA-seq
B. japonicum	RNA-seq	SRR12011616	Bj gonad-immature_female-2 RNA-seq
B. japonicum	RNA-seq	SRR12011615	Bj gonad-immature_female-3 RNA-seq
B. japonicum	RNA-seq	SRR12011614	Bj whole-body_male-1 RNA-seq
B. japonicum	RNA-seq	SRR12011613	Bj whole-body_male-2 RNA-seq
B. japonicum	RNA-seq	SRR12011612	Bj whole-body_male-3 RNA-seq
B. japonicum	RNA-seq	SRR12011611	Bj whole-body_male-4 RNA-seq
B. japonicum	RNA-seq	SRR12011610	Bj whole-body_female-1 RNA-seq
B. japonicum	RNA-seq	SRR12011608	Bj whole-body_female-2 RNA-seq
B. japonicum	RNA-seq	SRR12011607	Bj whole-body_female-3 RNA-seq
B. japonicum	RNA-seq	SRR12011606	Bj muscle_male-1 RNA-seq
B. japonicum	RNA-seq	SRR12011605	Bj muscle_female-1 RNA-seq
B. japonicum	RNA-seq	SRR12011604	Bj muscle_male-2 RNA-seq
B. japonicum	RNA-seq	SRR12011603	Bj muscle_male-3 RNA-seq
B. japonicum	RNA-seq	SRR12011602	Bj muscle_male-4 RNA-seq
B. japonicum	RNA-seq	SRR12011601	Bj muscle_female-2 RNA-seq
B. japonicum	RNA-seq	SRR12011600	Bj muscle_female-3 RNA-seq
B. japonicum	RNA-seq	SRR12011599	Bj muscle_female-4 RNA-seq
B. belcheri	RNA-seq	SRR12011597	bb_gonad-immature_male-1 RNA-seq
B. belcheri	RNA-seq	SRR12011596	bb_gonad-immature_male-2 RNA-seq
B. belcheri	RNA-seq	SRR12011595	bb_gonad-immature_male-3 RNA-seq
B. belcheri	RNA-seq	SRR12011594	bb_gonad-immature_male-4 RNA-seq
B. belcheri	RNA-seq	SRR12011593	bb_gonad-mature_male-1 RNA-seq
B. belcheri	RNA-seq	SRR12011592	bb_gonad-mature_male-2 RNA-seq
B. belcheri	RNA-seq	SRR12011591	bb_gonad-mature_male-3 RNA-seq
B. belcheri	RNA-seq	SRR12011590	bb_gonad-mature_male-4 RNA-seq
B. belcheri	RNA-seq	SRR12011589	bb_gonad-mature_male-5 RNA-seq
B. belcheri	RNA-seq	SRR12011588	bb_gonad-mature_male-6 RNA-seq
B. belcheri	RNA-seq	SRR12011586	bb_gonad-mature_female-1 RNA-seq
B. belcheri	RNA-seq	SRR12011585	bb_gonad-mature_female-2 RNA-seq
B. belcheri	RNA-seq	SRR12011584	bb_gonad-mature_female-3 RNA-seq
B. belcheri	RNA-seq	SRR12011583	bb_gonad-immature_female-1 RNA-seq
B. belcheri	RNA-seq	SRR12011582	bb_gonad-immature_female-2 RNA-seq
B. belcheri	RNA-seq	SRR12011581	bb_gonad-immature_female-3 RNA-seq
B. belcheri	RNA-seq	SRR12011580	Bb muscle_male-1 RNA-seq
B. belcheri	RNA-seq	SRR12011579	Bb muscle_male-2 RNA-seq
B. belcheri	RNA-seq	SRR12011578	Bb muscle_male-3 RNA-seq
B. belcheri	RNA-seq	SRR12011577	Bb muscle_female-1 RNA-seq
B. belcheri	RNA-seq	SRR12011575	Bb muscle_female-2 RNA-seq

Species	Sequencing type	SRA accession	title
B. belcheri	RNA-seq	SRR12011574	Bb muscle_female-3 RNA-seq
B. floridae	RNA-seq	SRR12011573	Bf gonad-mature_female-1 RNA-seq
B. floridae	RNA-seq	SRR12011572	Bf gonad-immature_female-1 RNA-seq
B. floridae	RNA-seq	SRR12011571	Bf gonad-immature_female-7 RNA-seq
B. floridae	RNA-seq	SRR12011570	Bf gonad-immature_male-1 RNA-seq
B. floridae	RNA-seq	SRR12011569	Bf gonad-immature_male-2 RNA-seq
B. floridae	RNA-seq	SRR12011568	Bf gonad-immature_male-3 RNA-seq
B. floridae	RNA-seq	SRR12011567	Bf gonad-immature_male-4 RNA-seq
B. floridae	RNA-seq	SRR12011566	Bf gonad-immature_female-2 RNA-seq
B. floridae	RNA-seq	SRR12011564	Bf gonad-immature_female-3 RNA-seq
B. floridae	RNA-seq	SRR12011563	Bf gonad-immature_female-4 RNA-seq
B. floridae	RNA-seq	SRR12011562	Bf gonad-immature_female-5 RNA-seq
B. floridae	RNA-seq	SRR12011561	Bf gonad-immature_female-6 RNA-seq
B. floridae	RNA-seq	SRR12011560	Bf gonad-mature_male-1 RNA-seq
B. floridae	RNA-seq	SRR12011559	Bf gonad-mature_male-2 RNA-seq
B. floridae	RNA-seq	SRR12011558	Bf gonad-mature_male-3 RNA-seq
B. floridae	RNA-seq	SRR12011557	Bf muscle_male-1 RNA-seq
B. floridae	RNA-seq	SRR12011556	Bf muscle_female-1 RNA-seq
B. floridae	RNA-seq	SRR12011555	Bf muscle_male-2 RNA-seq
B. floridae	RNA-seq	SRR12011553	Bf muscle_male-3 RNA-seq
B. floridae	RNA-seq	SRR12011552	Bf muscle_male-4 RNA-seq
B. floridae	RNA-seq	SRR12011551	Bf muscle_female-2 RNA-seq
B. floridae	RNA-seq	SRR12011550	Bf muscle_female-3 RNA-seq
B. floridae	RNA-seq	SRR12011549	Bf muscle_female-4 RNA-seq
BbBf	Resequencing	SRR11012662	Illumina reads of B. floridae x B. belcheri hybrid
BjBf	Resequencing	SRR11013310	Illumina reads of the B. japonicum x B. floridae hybrid
B. japonicum	Resequencing	SRR12010284	Bj_male_1_reseq
B. japonicum	Resequencing	SRR12010283	Bj_male_2_reseq
B. japonicum	Resequencing	SRR12010211	Bj_male_3_reseq
B. japonicum	Resequencing	SRR12010200	Bj_male_4_reseq
B. japonicum	Resequencing	SRR12010189	Bj_male_5_reseq
B. japonicum	Resequencing	SRR12010178	Bj_male_6_reseq
B. japonicum	Resequencing	SRR12010167	Bj_male_7_reseq
B. japonicum	Resequencing	SRR12010156	Bj_male_8_reseq
B. japonicum	Resequencing	SRR12010145	Bj_male_9_reseq
B. japonicum	Resequencing	SRR12010134	Bj_male_10_reseq
B. japonicum	Resequencing	SRR12010282	Bj_male_11_reseq
B. japonicum	Resequencing	SRR12010271	Bj_male_12_reseq
B. japonicum	Resequencing	SRR12010260	Bj_male_13_reseq
B. japonicum	Resequencing	SRR12010249	Bj_male_14_reseq
B. japonicum	Resequencing	SRR12010238	Bj_male_15_reseq
B. japonicum	Resequencing	SRR12010227	Bj_male_16_reseq
B. japonicum	Resequencing	SRR12010216	Bj_male_17_reseq
B. japonicum	Resequencing	SRR12010214	Bj_male_18_reseq
B. japonicum	Resequencing	SRR12010213	Bj_male_19_reseq
B. japonicum	Resequencing	SRR12010212	Bj_male_20_reseq
B. japonicum	Resequencing	SRR12010210	Bj_male_21_reseq
B. japonicum	Resequencing	SRR12010209	Bj_male_22_reseq
B. japonicum	Resequencing	SRR12010208	Bj_male_23_reseq
B. japonicum	Resequencing	SRR12010207	Bj_male_24_reseq
B. japonicum	Resequencing	SRR12010206	Bj_male_25_reseq
B. japonicum	Resequencing	SRR12010205	Bj_male_26_reseq
B. japonicum	Resequencing	SRR12010204	Bj_male_27_reseq
B. japonicum	Resequencing	SRR12010203	Bj_male_28_reseq
B. japonicum	Resequencing	SRR12010202	Bj_male_29_reseq
B. japonicum	Resequencing	SRR12010201	Bj_male_30_reseq
B. japonicum	Resequencing	SRR12010199	Bj_male_31_reseq
B. japonicum	Resequencing	SRR12010198	Bj_male_32_reseq
B. japonicum	Resequencing	SRR12010197	Bj_male_33_reseq
B. japonicum	Resequencing	SRR12010196	Bj_male_34_reseq
B. japonicum	Resequencing	SRR12010195	Bj_male_35_reseq
B. japonicum	Resequencing	SRR12010194	Bj_male_36_reseq
B. japonicum	Resequencing	SRR12010193	Bj_male_37_reseq
B. japonicum	Resequencing	SRR12010192	Bj_male_38_reseq
B. japonicum	Resequencing	SRR12010191	Bj_male_39_reseq
B. japonicum	Resequencing	SRR12010190	Bj_male_40_reseq
B. japonicum	Resequencing	SRR12010188	Bj_male_41_reseq
B. japonicum	Resequencing	SRR12010187	Bj_male_42_reseq
B. japonicum	Resequencing	SRR12010186	Bj_male_43_reseq

Species	Sequencing type	SRA accession	title
B. japonicum	Resequencing	SRR12010185	Bj_male_44_reseq
B. japonicum	Resequencing	SRR12010184	Bj_male_45_reseq
B. japonicum	Resequencing	SRR12010183	Bj_male_46_reseq
B. japonicum	Resequencing	SRR12010182	Bj_male_47_reseq
B. japonicum	Resequencing	SRR12010181	Bj_male_48_reseq
B. japonicum	Resequencing	SRR12010180	Bj_female_1_reseq
B. japonicum	Resequencing	SRR12010179	Bj_female_2_reseq
B. japonicum	Resequencing	SRR12010177	Bj_female_3_reseq
B. japonicum	Resequencing	SRR12010176	Bj_female_4_reseq
B. japonicum	Resequencing	SRR12010175	Bj_female_5_reseq
B. japonicum	Resequencing	SRR12010174	Bj_female_6_reseq
B. japonicum	Resequencing	SRR12010173	Bj_female_7_reseq
B. japonicum	Resequencing	SRR12010172	Bj_female_8_reseq
B. japonicum	Resequencing	SRR12010171	Bj_female_9_reseq
B. japonicum	Resequencing	SRR12010170	Bj_female_10_reseq
B. japonicum	Resequencing	SRR12010169	Bj_female_11_reseq
B. japonicum	Resequencing	SRR12010168	Bj_female_12_reseq
B. japonicum	Resequencing	SRR12010166	Bj_female_13_reseq
B. japonicum	Resequencing	SRR12010165	Bj_female_14_reseq
B. japonicum	Resequencing	SRR12010164	Bj_female_15_reseq
B. japonicum	Resequencing	SRR12010163	Bj_female_16_reseq
B. japonicum	Resequencing	SRR12010162	Bj_female_17_reseq
B. japonicum	Resequencing	SRR12010161	Bj_female_18_reseq
B. japonicum	Resequencing	SRR12010160	Bj_female_19_reseq
B. japonicum	Resequencing	SRR12010159	Bj_female_20_reseq
B. japonicum	Resequencing	SRR12010158	Bj_female_21_reseq
B. japonicum	Resequencing	SRR12010157	Bj_female_22_reseq
B. japonicum	Resequencing	SRR12010155	Bj_female_23_reseq
B. japonicum	Resequencing	SRR12010154	Bj_female_24_reseq
B. japonicum	Resequencing	SRR12010153	Bj_female_25_reseq
B. japonicum	Resequencing	SRR12010152	Bj_female_26_reseq
B. japonicum	Resequencing	SRR12010151	Bj_female_27_reseq
B. japonicum	Resequencing	SRR12010150	Bj_female_28_reseq
B. japonicum	Resequencing	SRR12010149	Bj_female_29_reseq
B. japonicum	Resequencing	SRR12010148	Bj_female_30_reseq
B. belcheri	Resequencing	SRR12010147	Bb_male_1_reseq
B. belcheri	Resequencing	SRR12010146	Bb_male_2_reseq
B. belcheri	Resequencing	SRR12010144	Bb_male_3_reseq
B. belcheri	Resequencing	SRR12010143	Bb_male_4_reseq
B. belcheri	Resequencing	SRR12010142	Bb_male_5_reseq
B. belcheri	Resequencing	SRR12010141	Bb_male_6_reseq
B. belcheri	Resequencing	SRR12010140	Bb_male_7_reseq
B. belcheri	Resequencing	SRR12010139	Bb_male_8_reseq
B. belcheri	Resequencing	SRR12010138	Bb_male_9_reseq
B. belcheri	Resequencing	SRR12010137	Bb_male_10_reseq
B. belcheri	Resequencing	SRR12010136	Bb_male_11_reseq
B. belcheri	Resequencing	SRR12010135	Bb_male_12_reseq
B. belcheri	Resequencing	SRR12010133	Bb_male_13_reseq
B. belcheri	Resequencing	SRR12010132	Bb_male_14_reseq
B. belcheri	Resequencing	SRR12010131	Bb_male_15_reseq
B. belcheri	Resequencing	SRR12010130	Bb_male_16_reseq
B. belcheri	Resequencing	SRR12010129	Bb_male_17_reseq
B. belcheri	Resequencing	SRR12010128	Bb_male_18_reseq
B. belcheri	Resequencing	SRR12010127	Bb_male_19_reseq
B. belcheri	Resequencing	SRR12010126	Bb_male_20_reseq
B. belcheri	Resequencing	SRR12010125	Bb_female_1_reseq
B. belcheri	Resequencing	SRR12010124	Bb_female_3_reseq
B. belcheri	Resequencing	SRR12010281	Bb_female_5_reseq
B. belcheri	Resequencing	SRR12010280	Bb_female_7_reseq
B. belcheri	Resequencing	SRR12010279	Bb_female_8_reseq
B. belcheri	Resequencing	SRR12010278	Bb_female_9_reseq
B. belcheri	Resequencing	SRR12010277	Bb_female_10_reseq
B. belcheri	Resequencing	SRR12010276	Bb_female_11_reseq
B. belcheri	Resequencing	SRR12010275	Bb_female_12_reseq
B. belcheri	Resequencing	SRR12010274	Bb_female_13_reseq
B. belcheri	Resequencing	SRR12010273	Bb_female_14_reseq
B. belcheri	Resequencing	SRR12010272	Bb_female_15_reseq
B. belcheri	Resequencing	SRR12010270	Bb_female_16_reseq
B. belcheri	Resequencing	SRR12010269	Bb_female_17_reseq

Species	Sequencing type	SRA accession	title
B. belcheri	Resequencing	SRR12010268	Bb_female_19_reseq
B. belcheri	Resequencing	SRR12010267	Bb_female_20_reseq
B. belcheri	Resequencing	SRR12010266	Bb_female_21_reseq
B. belcheri	Resequencing	SRR12010265	Bb_female_22_reseq
B. belcheri	Resequencing	SRR12010264	Bb_female_23_reseq
B. belcheri	Resequencing	SRR12010263	Bb_female_24_reseq
B. belcheri	Resequencing	SRR12010262	Bb_female_25_reseq
B. belcheri	Resequencing	SRR12010261	Bb_female_26_reseq
B. belcheri	Resequencing	SRR12010259	Bb_female_27_reseq
B. belcheri	Resequencing	SRR12010258	Bb_female_28_reseq
B. belcheri	Resequencing	SRR12010257	Bb_female_29_reseq
B. belcheri	Resequencing	SRR12010256	Bb_female_30_reseq
B. floridae	Resequencing	SRR12010255	Bf_male_1_reseq
B. floridae	Resequencing	SRR12010254	Bf_male_2_reseq
B. floridae	Resequencing	SRR12010253	Bf_male_3_reseq
B. floridae	Resequencing	SRR12010252	Bf_male_4_reseq
B. floridae	Resequencing	SRR12010251	Bf_male_5_reseq
B. floridae	Resequencing	SRR12010250	Bf_male_6_reseq
B. floridae	Resequencing	SRR12010248	Bf_male_7_reseq
B. floridae	Resequencing	SRR12010247	Bf_male_8_reseq
B. floridae	Resequencing	SRR12010246	Bf_male_9_reseq
B. floridae	Resequencing	SRR12010245	Bf_female_1_reseq
B. floridae	Resequencing	SRR12010244	Bf_female_2_reseq
B. floridae	Resequencing	SRR12010243	Bf_female_3_reseq
B. floridae	Resequencing	SRR12010242	Bf_female_4_reseq
B. floridae	Resequencing	SRR12010241	Bf_female_5_reseq
B. floridae	Resequencing	SRR12010240	Bf_female_6_reseq
B. floridae	Resequencing	SRR12010239	Bf_female_7_reseq
B. floridae	Resequencing	SRR12010237	Bf_female_8_reseq
B. floridae	Resequencing	SRR12010236	Bf_female_9_reseq
B. floridae	Resequencing	SRR12010235	Bf_female_10_reseq
B. floridae	Resequencing	SRR12010234	Bf_female_11_reseq
B. floridae	Resequencing	SRR12010233	Bf_female_12_reseq
B. floridae	Resequencing	SRR12010232	Bf_female_13_reseq
B. floridae	Resequencing	SRR12010231	Bf_female_14_reseq
B. floridae	Resequencing	SRR12010230	Bf_female_15_reseq
B. floridae	Resequencing	SRR12010229	Bf_female_16_reseq
B. floridae	Resequencing	SRR12010228	Bf_female_17_reseq
B. floridae	Resequencing	SRR12010226	Bf_female_18_reseq
B. floridae	Resequencing	SRR12010225	Bf_female_19_reseq
B. floridae	Resequencing	SRR12010224	Bf_female_20_reseq
B. floridae	Resequencing	SRR12010223	Bf_female_21_reseq
B. floridae	Resequencing	SRR12010222	Bf_female_22_reseq
B. floridae	Resequencing	SRR12010221	Bf_female_23_reseq
B. floridae	Resequencing	SRR12010220	Bf_female_24_reseq
B. floridae	Resequencing	SRR12010219	Bf_female_25_reseq
B. floridae	Resequencing	SRR12010218	Bf_female_26_reseq
B. floridae	Resequencing	SRR12010217	Bf_female_27_reseq
B. floridae	Resequencing	SRR12010215	Bf_female_28_reseq

Table S8. The list of Bf fully sex-linked genes and their sex-biased expression

Gene ID	Start	End	Gene name	Sex biased (immature gonad)
bf_00005984	25770	28110	Similar to ZNF626	unbiased
bf_00005985	28279	28898	Similar to PLEKHH1	unbiased
bf_00005986	32629	76251	unknown	unbiased
bf_00005987	78779	81223	Similar to ZNF160	female-biased
bf_00005988	86987	97961	unknown	unbiased
bf_00005989	99614	120042	NFKB2	unbiased
bf_00005990	119886	124449	RNLS	female-biased
bf_00005991	128618	142909	SGPL1	male-biased
bf_00005992	144804	248870	PSD2	unbiased
bf_00005993	145943	146783	Similar to FGFR1OP	unbiased
bf_00005994	270552	276389	Similar to COLEC11	unbiased
bf_00005995	281987	286544	unknown	unbiased
bf_00005996	290042	297045	CASP7	female-biased
bf_00005997	302604	304101	unknown	unbiased
bf_00005998	306213	309543	Similar to DDX24	unbiased
bf_00005999	310091	313561	Similar to C2CD6	unbiased
bf_00006000	314097	315644	Similar to MAP4K1	unbiased
bf_00006001	316492	371775	PTPN13	male-biased
bf_00006002	374019	375212	Similar to ISLR2	unbiased
bf_00006003	453232	482224	Similar to FGFR2	female-biased
bf_00006004	492188	558734	PRSS12	unbiased
bf_00006005	561180	585923	Similar to HABP2	unbiased
bf_00006006	582919	592650	Similar to CASP7	unbiased
bf_00006007	596706	598707	unknown	unbiased
bf_00006008	604321	607724	Similar to LAMP1	unbiased
bf_00006009	608992	632949	Similar to PAQR3	unbiased
bf_00006010	638496	645100	unknown	unbiased
bf_00006011	657029	662915	unknown	unbiased
bf_00006012	665902	670319	Similar to HPGDS	unbiased
bf_00006013	671306	673798	Similar to HPGDS	unbiased
bf_00006014	674899	677345	Similar to HPGDS	unbiased
bf_00006015	678606	685111	Similar to HPGDS	female-biased
bf_00006016	701114	704281	Similar to HPGDS	unbiased
bf_00006017	705383	705589	Similar to HPGDS	unbiased
bf_00006018	706263	709741	unknown	unbiased
bf_00006019	716198	718309	Similar to MYO16	unbiased
bf_00006020	731360	736577	Similar to AKT3	unbiased
bf_00006021	738850	743633	C4H4ORF45	male-biased
bf_00006022	743444	759735	GMCL1	male-biased
bf_00006023	760882	763715	unknown	female-biased
bf_00006024	769153	788647	EIF4E	female-biased
bf_00006025	795176	812068	unknown	unbiased
bf_00006026	811708	817190	LOC422643	male-biased
bf_00006027	823106	823396	Similar to LOXL2	unbiased
bf_00006028	832528	833835	unknown	unbiased
bf_00006029	834403	836018	Similar to LRRC8E	unbiased
bf_00006030	837397	842500	unknown	unbiased
bf_00006031	842747	870172	LOC422643	unbiased
bf_00006032	878016	879224	unknown	unbiased
bf_00006033	887726	895652	unknown	unbiased
bf_00006034	898440	917209	NCAPG	unbiased
bf_00006035	919708	969288	unknown	female-biased
bf_00006036	959371	968788	unknown	unbiased
bf_00006037	972405	998401	unknown	male-biased
bf_00006038	999430	1003874	Similar to HPN	unbiased
bf_00006039	1006988	1014562	Similar to ST14	unbiased
bf_00006040	1016906	1117146	CAMK2A	female-biased
bf_00006041	1130921	1141957	Similar to CTRB1	unbiased
bf_00006042	1150421	1161241	Similar to KLKB1	unbiased
bf_00006043	1176830	1231659	unknown	unbiased
bf_00006044	1198940	1201398	unknown	unbiased
bf_00006045	1199512	1200122	Similar to TANGO6	unbiased
bf_00006046	1229244	1233367	Similar to KLC3	unbiased
bf_00006047	1234587	1290879	Similar to TTC24	unbiased
bf_00006048	1297756	1306162	unknown	unbiased
bf_00006049	1307211	1351153	UNC5B	unbiased
bf_00006050	1484998	1520556	unknown	unbiased

Gene ID	Start	End	Gene name	Sex biased (immature gonad)
bf_00006051	1525327	1553340	Similar to STPG2	male-biased
bf_00006052	1555816	1658625	FSTL5	unbiased
bf_00006053	1662184	1690275	ANXA11	female-biased
bf_00006054	1691379	1693675	Similar to OR51B5	unbiased
bf_00006055	1761746	1769217	DYDC1	male-biased
bf_00006056	1768928	1778660	RBM22	unbiased
bf_00006057	1777346	1790060	Similar to USP39	female-biased
bf_00006058	1790142	1793274	unknown	unbiased
bf_00006059	1812551	1866964	PRKG1	unbiased
bf_00006060	1868732	1872340	unknown	unbiased
bf_00006061	1873709	1884340	LOC112530190	unbiased
bf_00006062	1896109	1993532	CDH23	unbiased
bf_00006063	2026449	2037641	Similar to DHRS7	female-biased
bf_00006064	2043266	2056048	DNAJB12	unbiased
bf_00006065	2056257	2064362	LIPA	unbiased
bf_00006066	2065928	2072000	LIPA	unbiased
bf_00006067	2075035	2078487	H2AFZ	unbiased
bf_00006068	2080598	2082653	Similar to IKZF3	unbiased
bf_00006069	2097686	2103895	DDIT4	female-biased
bf_00006070	2108176	2142643	WAPL	male-biased
bf_00006071	2217378	2348138	Similar to COL24A1	unbiased
bf_00006072	2369262	2373427	FGL1	unbiased
bf_00006073	2374896	2383042	unknown	unbiased
bf_00006074	2391500	2397794	IPMK	female-biased
bf_00006075	2397617	2416994	DDX21	female-biased
bf_00006076	2419941	2433562	Similar to IRF1	female-biased
bf_00006077	2450401	2450907	Similar to DDX60	unbiased
bf_00006078	2465461	2472045	STAMBPL1	unbiased
bf_00006079	2473923	2513340	AAK1	unbiased
bf_00006080	2520060	2524911	unknown	unbiased
bf_00006081	2525417	2525752	Similar to ITGAM	unbiased
bf_00006082	2525786	2526583	unknown	unbiased
bf_00006083	2527496	2528833	unknown	unbiased
bf_00006084	2529263	2530066	unknown	unbiased
bf_00006085	2568464	2592853	Similar to HNRNPC	unbiased
bf_00006086	2626989	2693251	NRG2	female-biased
bf_00006087	2651479	2652197	unknown	unbiased
bf_00006088	2700731	2702881	unknown	unbiased
bf_00006089	2716543	2781971	Similar to DAW1	unbiased
bf_00006090	2801908	2820266	unknown	unbiased
bf_00006091	2822370	2832111	HGSNAT	female-biased
bf_00006092	2850844	2867705	unknown	unbiased
bf_00006093	2878484	2898440	Similar to CACNA1A	unbiased
bf_00006094	2899230	3069267	Similar to HMCN2	unbiased
bf_00006095	3077534	3109449	Similar to PLAT	unbiased
bf_00006096	3085917	3090352	Similar to ZNF845	unbiased
bf_00006097	3109033	3110160	unknown	unbiased
bf_00006098	3113650	3147295	Mar-01	unbiased
bf_00006099	3151862	3157985	FUT11	unbiased
bf_00006100	3158069	3161260	MRPS24	unbiased
bf_00006101	3163044	3166793	LOC112530190	unbiased
bf_00006102	3166894	3176486	Similar to ZNF845	unbiased
bf_00006103	3176731	3182072	Similar to ZNF569	unbiased
bf_00006104	3183833	3187935	Similar to ZNF845	unbiased
bf_00006105	3191639	3223167	Similar to FCN1	unbiased
bf_00006106	3226442	3256518	Similar to ANGPT2	unbiased
bf_00006107	3259670	3261469	unknown	unbiased
bf_00006108	3261471	3264177	unknown	unbiased
bf_00006109	3264948	3265283	Similar to PDK2	unbiased
bf_00006110	3265317	3266111	unknown	unbiased
bf_00006111	3266895	3268232	unknown	unbiased
bf_00006112	3268662	3269084	Similar to RBM12	unbiased
bf_00006113	3269087	3269449	unknown	unbiased
bf_00006114	3273627	3276017	Similar to COL6A5	unbiased
bf_00006115	3287007	3320673	Similar to DMBT1	unbiased
bf_00006116	3320866	3328910	Similar to COL6A5	unbiased
bf_00006117	3334831	3335824	Similar to IVL	unbiased
bf_00006118	3336219	3337532	unknown	unbiased
bf_00006119	3339416	3341485	unknown	unbiased

Gene ID	Start	End	Gene name	Sex biased (immature gonad)
bf_00006120	3343369	3345438	unknown	unbiased
bf_00006121	3346668	3349278	Similar to ANGPTL2	unbiased
bf_00006122	3358888	3384500	LOC107053715	unbiased
bf_00006123	3392628	3433662	LOC107053715	unbiased
bf_00006124	3435568	3439522	Similar to ZNF160	unbiased
bf_00006125	3441063	3442454	Similar to ZNF431	female-biased
bf_00006126	3443913	3447349	unknown	unbiased
bf_00006127	3448351	3454148	ACTA2	male-biased
bf_00006128	3457784	3463168	SKP1	male-biased
bf_00006129	3464706	3470562	Similar to CRIP1	unbiased
bf_00006130	3472693	3478227	unknown	unbiased
bf_00006131	3479268	3482047	Similar to SOAT2	unbiased
bf_00006132	3482108	3483148	Similar to ZNF75D	unbiased
bf_00006133	3497962	3511283	unknown	unbiased
bf_00006134	3514020	3537993	SEC24D	female-biased
bf_00006135	3538414	3541257	Similar to SAR1B	male-biased
bf_00006136	3544856	3547356	SAR1B	unbiased
bf_00006137	3553192	3559008	unknown	unbiased
bf_00006138	3561175	3577293	SEC24B	unbiased
bf_00006139	3579351	3583111	unknown	unbiased
bf_00006140	3582172	3594755	UBE2D1	male-biased
bf_00006141	3600256	3611655	unknown	unbiased
bf_00006142	3614264	3616678	ELOVL6	female-biased
bf_00006143	3618449	3626276	AP1AR	unbiased
bf_00006144	3628349	3633464	C4H4ORF47	male-biased
bf_00006145	3633840	3654936	CCAR1	male-biased
bf_00006146	3657722	3662483	TXNDC15	female-biased
bf_00006147	3664132	3664398	Similar to FAT2	female-biased
bf_00006148	3664592	3666031	Similar to ZNF568	female-biased
bf_00006149	3676986	3677345	Similar to CSNK1A1	unbiased
bf_00006150	3696390	3722175	Similar to EDIL3	unbiased
bf_00006151	3734195	3759624	Similar to PKD1L2	unbiased
bf_00006152	3767859	3768423	Similar to CCDC168	unbiased
bf_00006153	3770221	3873067	FRAS1	female-biased
bf_00006154	3877109	3880947	Similar to ZNF347	female-biased
bf_00006155	3882872	3897882	SH2D4B	female-biased
bf_00006156	3901826	3903393	Similar to PAXBP1	male-biased
bf_00006157	3905719	3927140	DNA2	unbiased
bf_00006158	3927393	3932782	DCK	unbiased
bf_00006159	3934841	3940372	Similar to ELOVL5	unbiased
bf_00006160	3940530	3958514	USO1	female-biased
bf_00006161	3961846	3967907	Similar to TRIO	unbiased
bf_00006162	3968440	3973987	Similar to ZNF845	unbiased
bf_00006163	3980656	3985941	ARHGAP24	female-biased
bf_00006164	3991229	4006653	MAPK9	unbiased
bf_00006165	4011266	4017269	Similar to ST8SIA1	unbiased
bf_00006166	4018636	4023863	COX18	female-biased
bf_00006167	4030424	4065856	unknown	unbiased
bf_00006168	4069210	4074849	SLC25A16	female-biased
bf_00006169	4075850	4076158	Similar to GRM5	unbiased
bf_00006170	4077247	4083225	ADH5	female-biased
bf_00006171	4084210	4094090	METAP1	female-biased
bf_00006172	4093730	4099919	LRRC20	unbiased
bf_00006173	4102116	4103302	unknown	unbiased
bf_00006174	4107523	4109459	unknown	unbiased
bf_00006175	4109744	4127205	unknown	unbiased
bf_00006176	4127333	4128688	Similar to E2F7	unbiased
bf_00006177	4133625	4142775	EIF4EBP2	female-biased

Table S9. The list of Bj fully sex-linked genes and their sex-biased expression

Gene ID	Start	End	Gene name	Sex biased (immature gonad)
bj_00002076	2110238	2112225	Similar to ABCA1	unbiased
bj_00002077	2112958	2132529	ZFYVE19	unbiased
bj_00002078	2132655	2153864	TYRO3	unbiased
bj_00002079	2156622	2232922	Similar to KNTC1	female-biased
bj_00002080	2234883	2244565	unknown	male-biased
bj_00002081	2247090	2250784	Similar to ZDHHC24	unbiased
bj_00002082	2251509	2262747	KATNBL1	female-biased
bj_00002083	2274785	2360274	WDR27	unbiased
bj_00002084	2370813	2394250	THBS1	unbiased
bj_00002085	2374705	2375252	Similar to VIP	unbiased
bj_00002086	2401971	2466792	FSIP1	unbiased
bj_00002087	2469411	2473829	FAM228	unbiased
bj_00002088	2501223	2517859	C5H14ORF169	unbiased
bj_00002089	2528747	2539333	Similar to MCFD2	unbiased
bj_00002090	2554478	2568127	Similar to PAX3	unbiased
bj_00002091	2571040	2572994	MCFD2	unbiased
bj_00002092	2573419	2589307	TAF1A	unbiased
bj_00002093	2592460	2596653	unknown	male-biased
bj_00002094	2596724	2597732	Similar to MMP9	male-biased
bj_00002095	2602791	2617210	MACROD2	male-biased
bj_00002096	2621438	2649601	CFAP36	unbiased
bj_00002097	2650447	2719785	CCDC88C	unbiased
bj_00002098	2721245	2726433	GLO1	female-biased
bj_00002099	2738972	2739172	Similar to VPS13B	unbiased
bj_00002100	2747625	2762158	SLC30A6	unbiased
bj_00002101	2768033	2768546	GNG4	unbiased
bj_00002102	2775145	2779688	Similar to ZNF26	unbiased
bj_00002103	2790003	2798638	BCKDHA	unbiased
bj_00002104	2799415	2802926	ADCK1	male-biased
bj_00002105	2803623	2811605	EXOSC5	unbiased
bj_00002106	2815085	2817881	Similar to PCGF1	unbiased
bj_00002107	2840607	2855400	Similar to CALM1	unbiased
bj_00002108	2878706	2889477	Similar to CALM1	male-biased
bj_00002109	2946552	2986265	Similar to CALM1	unbiased
bj_00002110	3017590	3091283	SCFD1	female-biased

Dataset S1. Ohnolog group gene list

SI References

1. O. Simakov, *et al.*, Deeply conserved synteny resolves early events in vertebrate evolution. *Nat. Ecol. Evol.* **4**, 820–830 (2020).

2. S. Mawaribuchi, Y. Ito, M. Ito, Independent evolution for sex determination and differentiation in the DMRT family in animals. *Biol. Open* **8**, bio041962 (2019).