

Supplemental Information

The White-Spotted Bamboo Shark Genome Reveals Chromosome Rearrangements and Fast-Evolving Immune Genes of Cartilaginous Fish

Yaolei Zhang, Haoyang Gao, Hanbo Li, Jiao Guo, Bingjie Ouyang, Meiniang Wang, Qiwu Xu, Jiahao Wang, Meiqi Lv, Xinyu Guo, Qun Liu, Likun Wei, Han Ren, Yang Xi, Yang Guo, Bingzhao Ren, Shanshan Pan, Chuxin Liu, Xiaoyan Ding, Haitao Xiang, Yingjia Yu, Yue Song, Lingfeng Meng, Shanshan Liu, Jun Wang, Yuan Jiang, Jiahai Shi, Shiping Liu, Jamal S.M. Sabir, Mumdooh J. Sabir, Muhummadh Khan, Nahid H. Hajrah, Simon Ming-Yuen Lee, Xun Xu, Huanming Yang, Jian Wang, Guangyi Fan, Naibo Yang, and Xin Liu

Supplemental information for:

The white-spotted bamboo shark genome reveals chromosome rearrangements and fast-evolving immune genes of cartilaginous fish

Yaolei Zhang^{1,4,8,†}, Haoyang Gao^{1,4,†}, Hanbo Li^{1,4,†}, Jiao Guo^{1,4,†}, Bingjie Ouyang^{1,4}, Meiniang Wang^{2,4}, Qiwu Xu^{1,4}, Jiahao Wang^{1,4}, Meiqi Lv^{1,4}, Xinyu Guo^{1,4}, Qun Liu^{1,4}, Likun Wei⁵, Han Ren^{2,4}, Yang Xi^{2,4}, Yang Guo^{1,4}, Bingzhao Ren^{1,4}, Shanshan Pan^{1,4}, Chuxin Liu^{2,4}, Xiaoyan Ding^{1,4}, Haitao Xiang^{2,4}, Yingjia Yu^{1,4}, Yue Song^{1,4}, Lingfeng Meng^{1,4}, Shanshan Liu^{1,4}, Jun Wang^{1,4}, Yuan Jiang^{2,6}, Jiahai Shi⁵, Shiping Liu^{2,4}, Jamal S. M. Sabir⁹, Mumdooh J. Sabir⁹, Muhummadh Khan⁹, Nahid H. Hajrah⁹, Simon Ming-Yuen Lee³, Xun Xu^{2,4}, Huanming Yang^{2,7}, Jian Wang^{2,7}, Guangyi Fan^{1,3,4,*}, Naibo Yang^{2,6,*} & Xin Liu^{1,2,4,10,*}

This file includes: Transparent Methods, Figures S1–S14, Tables S1-S9 and S12-S16.

Transparent Methods	4
DNA, RNA extraction	4
Library construction	4
Sequencing for all libraries	6
Genome assembly and annotation	6
Construction of phylogenetic tree	8
Evolution of chromosomes	9
MHC genes and P2X gene family analysis	9
Identification of IgNAR	10
Bamboo shark vNAR phage display library construction	10
Bamboo shark vNAR phage display library screening	11
Phage ELISA	12
Supplementary Figures	13
Figure S1. GC content of the assembled bamboo shark, elephant shark, and whale shark genomes, Related to Figure 1.	13
Figure S2. Circos plot of the genomic landscape, Related to Figure 1.	14
Figure S3. Circos plot of comparisons between bamboo shark and chicken, Related to Figure 1.....	15
Figure S4. Synteny relationship of bamboo shark and zebrafish, Related to Figure 1.	16
Figure S5. Whole genome duplication in elephant shark, bamboo shark and zebra fish genomes as revealed through 4DTv analyses, Related to Figure 1.	17
Figure S6. HOX gene clusters identified in five species, Related to Figure 1.	18
Figure S7. Distribution of shared paralogous genes of bamboo shark and elephant shark, Related to Figure 2.....	19
Figure S8. Distributions of conserved regions of six cartilaginous fishes (bamboo shark, elephant shark, whale shark, brownbanded bamboo shark, cloudy catshark and white shark) on 51 chromosomes of the bamboo shark, Related to Figure 2.....	20
Figure S9. Distributions of conserved regions between bamboo shark and medaka on 51 chromosomes of the bamboo shark, Related to Figure 2.	21
Figure S10. Distributions of conserved regions between bamboo shark and spotted gar on 51 chromosomes of the bamboo shark, Related to Figure 2.	22
Figure S11. Hypothetical model of MHC genes evolution, Related to Figure 3.....	23
Figure S12. Gene tree of <i>TRIM69</i> , Related to Figure 3.....	24

Figure S13. Re-assembled IgNAR region supported by paired-end reads, Related to Figure 3.....	25
Figure S14. Sequence alignment of the randomly picked clones from the bamboo shark vNAR-phage display library that represent sequences have different CDR3, Related to Figure 3.....	26
Supplementary Tables	27
Table S1. Summary of the sequencing data from WGS libraries and Hi-C libraries, Related to Figure 1.....	27
Table S2. Basic statistics of the assembled bamboo shark genome, Related to Figure 1. .	27
Table S3. Summary of the functional annotation of genes in bamboo shark genome, Related to Figure 1.....	27
Table S4. Summary of the repeat content in bamboo shark genome, Related to Figure 1.	28
Table S5. Summary of the TE content in three cartilaginous fishes and seven bony fishes, Related to Figure 1.....	28
Table S6. Evaluation of the final gene set by using BUSCO, Related to Figure 1.....	29
Table S7. Distribution of paralogous genes in the bamboo shark genome. The yellow highlighted numbers represent gene pairs more than 20, Related to Figure 2.	29
Table S8. Summary of conserved genes among six cartilaginous fishes distributed on each chromosome, Related to Figure 2.	30
Table S9. Enrichment analysis for genes in chromosomes 8, 37, 39, 41, 43, 44, 45, 46, 47,48,49, 50 and 51, Related to Figure 2.....	32
Table S12. List of MHC class II genes in analyzed species, Related to Figure 3.....	34
Table S13. MHC class II fragments detected by using BLAST in the elephant shark genome, Related to Figure 3.....	36
Table S14. Designed primers for amplifying vNARs, Related to Figure 3.....	37
Table S15. Ancestral P2X gene found in amphioxus and ascidiacea genomes, Related to Figure 4.....	37
Table S16. Protein length of P2X genes in bony fishes and sea lamprey, Related to Figure 4.....	38

Transparent Methods

DNA, RNA extraction

Fourteen tissue types including blood, eye, gill, heart, liver, muscle, spleen, stomach, dorsal fin, tail fin, pancreas, leptospira, 2 capsulogenous gland and 2 kidney samples were collected from a female bamboo shark individual. Genomic DNA was extracted from one blood sample using the phenol-chloroform method and its quality and quantity were assessed by pulsed field gel electrophoresis and Qubit Fluorometer. 14 RNAs from these 14 tissue types were extracted by using TRIzol[®] Reagent and were assessed by Agilent 2100 bioanalyzer system.

Library construction

Firstly, for WGS libraries with average insert sizes of 180 bp and 350 bp, a Covaris E220 ultrasonicator (Covaris, Brighton, UK) was used to shear the extracted high-quality DNA and AMPure XP beads (Agencourt, Beverly, USA) were used to select target fragments. Then, fragment end-repairing and A-tailing were performed by T4 DNA polymerase, T4 polynucleotide kinase and rTaq DNA polymerase. Next, PCR amplification of eight cycles was carried out and the single-strand circularization process was performed using T4 DNA Ligase, generating a single-stranded circular DNA library for sequencing.

Secondly, for mate-pair libraries, a Covaris E220 was used to acquire ~2 kb DNA fragments and Hydroshear (GeneMachines, CA, USA) was used to acquire ~5 kb, ~10 kb and ~20 kb DNA fragments. After further selection and purification of DNA, fragments were end-repaired and biotin-labeled. The modified fragments were circularized and re-fragmented using a Covaris E220. Biotin-labeled DNA fragments were captured on M280 streptavidin beads (Invitrogen, CA, USA), end-repaired, A-tailed and adaptor-ligated. Biotin-labeled fragments were PCR-amplified, purified on streptavidin-coated magnetic beads, size-selected by agarose gel

electrophoresis and column purification, single-stranded and re-circularized. The purified PCR products were heat-denatured together with an adapter that was a reverse-complement to a particular strand of the PCR product, and the single-stranded molecule was then ligated using DNA ligase to get a single-stranded circular DNA library.

Thirdly, a blood sample was used for constructing the Hi-C library. The fresh blood cells (5×10^6) were collected by centrifugation and re-suspended in 1 ml of 1x PBS by repetitive pipetting. The cells were cross-linked by adding 37% formaldehyde (SIGMA, USA) to obtain a 1% final concentration, to which was added a 2.5M glycine solution (SIGMA, USA) to a final concentration of 0.2M to quench the reaction. To prepare nuclei, the formaldehyde-fixed powder was resuspended in nuclei isolation buffer (10 mM Tris-HCl pH 8.0 (SIGMA, USA), 10 mM NaCl (BEYOTIME, Shanghai, China), 1× PMSF (SIGMA, St. Louis, USA)) and then incubated in 0.5% SDS for 10 min at 62 °C. SDS was immediately quenched with 10% Triton X-100 (SIGMA, St. Louis, USA) and the nuclei were collected by brief centrifugation. DNA was digested by restriction enzyme (Mbo I) (NEB, Ipswich, USA) and the 5' overhang was repaired using a biotinylated residue (0.4 mM biotin-14-Datp (INVITROGEN, USA)). The resulting blunt-end fragments were ligated in situ (10X NEB T4 DNA ligase buffer (NEB, Ipswich, USA), 10% Triton X-100 (SIGMA, St. Louis, USA), 10 mg/ml BSA (NEB, Ipswich, USA), T4 DNA ligase (NEB, Ipswich, USA)). Finally, the isolated DNA was reverse-crosslinked (adding 10 mg/ml proteinase K (NEB, Ipswich, USA) and 1% SDS (AMBION, Waltham, USA) to the tube followed by incubation at 56°C overnight) and purified (by putting the reverse-crosslinked DNA liquid into three tubes equally, adding 1.5× volumes of AMPure XP (AGENCOURT, Brea, USA) mixture to each tube, vortexing and spinning down briefly, incubating for 10 min. at room temperature, placing on the MPS (INVITROGEN, Waltham, USA) for 5 min. at room temperature, discarding supernatant, washing the beads twice with 1 ml of freshly made 75% ethanol (SINOPHARM, Shanghai, China), air-drying the beads completely and resuspending

the beads in 30 μ l of ddH₂O). The Hi-C library was created by shearing 20 μ g of DNA and capturing the biotin-containing fragments on streptavidin-coated beads using Dynabeads MyOne Streptavidin T1 (INVITROGEN, Waltham, USA). Then DNA fragments were end-repaired and adaptor ligation was performed. After PCR (95°C 3 min.; [98°C 20 sec., 60°C 15 sec., 72°C 15 sec.] (8 cycles); 72°C 10 min.), the standard circularization step required for the BGISEQ-500 platform was carried out and DNBs were prepared as previously described. Fourthly, for RNA library construction, mRNA was extracted from different tissues using TRIzol[®] Reagent, fragmented, and then reverse-transcribed into cDNA by using Hiscript II Reverse Transcriptase (Vazyme Biotech, Nanjing City, P.R. China). Then, all single-stranded circular DNA libraries were constructed by using the same strategy described as above.

Last, the stLFR library was constructed by its standard protocol that was published recently (Wang et al., 2019).

Sequencing for all libraries

All sequencing data were generated using the BGISEQ-500 platform. Libraries with an average insert size of 180 bp and 350 bp were sequenced yielding paired-end reads with 100 bp in length. Mate-pair libraries with average insert sizes of 2k, 5k, 10k and 20k and Hi-C library were sequenced yielding reads with 50 bp in length. RNA libraries were sequenced yielding paired-end reads with 100 bp in length. StLFR reads was also sequenced generating paired-end reads with 100 bp in length.

Genome assembly and annotation

Firstly, we filtered raw sequencing data by discarding low-quality reads (defined as >10% bases with quality values less than 10 and >5% unidentified (N) bases), adaptor-contaminated reads and PCR duplicate reads. We trimmed a few bases at the start and end of reads according to the FastQC (v0.11.2) (Andrews and FastQC, 2015) results. Secondly, we randomly selected about 40X clean reads to carry out k-mer analysis to estimate the

genome size. Thirdly, we used Platanus (v1.2.4) (Kajitani et al., 2014) to perform the initial assembly with WGS clean data with parameters “assemble -k 29 -u 0.2, scaffold -l 3 -u 0.2 -v 32 -s 32 and gap_close -s 34 -k 32 -d 5000”. We filled gaps using KGF (v1.19) and GapCloser (v1.10) (Luo et al., 2012) with default parameters. Fourthly, to obtain a chromosome-level genome, HIC-Pro (v2.8.0) (Servant et al., 2015) was used for quality control of Hi-C sequencing data with parameters [BOWTIE2_GLOBAL_OPTIONS = --very-sensitive -L 30 --score-min L,-0.6,-0.2 --end-to-end -reorder;BOWTIE2_LOCAL_OPTIONS = --very-sensitive -L 20 --score-min L,-0.6,-0.2 --end-to-end -reorder; IGATION_SITE = GATC; MIN_FRAG_SIZE = 100; MAX_FRAG_SIZE = 100000; MIN_INSERT_SIZE = 50; MAX_INSERT_SIZE = 1500]. Finally, the software packages Juicer (Durand et al., 2016) and 3d-dna (v170123) (Dudchenko et al., 2017) were employed to generate contact matrices of chromatin and constructed chromosomes with parameter [-m haploid -s 4 -c 5] based on the karyotype information supplied previously (MA et al., 2008).

After obtaining the final chromosome-level genome, we proceeded with genome annotation including repeat contents, gene models and gene function. For repeat section, both homolog-based and *de novo* prediction strategies were carried out. In detail, RepeatMasker (v 4.0.5) (Tarailo - Graovac and Chen, 2009) and RepeatProteinMasker (v 4.0.5) (Tarailo - Graovac and Chen, 2009) were used to detect interspersed repeats and low complexity sequences against the Repbase database (Jurka et al., 2005) at the nuclear and protein levels, respectively. Then RepeatMasker was further used to detect species-specific repeat elements using a custom database generated by RepeatModeler (v1.0.8) and LTR-FINDER (v1.0.6) (Xu and Wang, 2007). In addition, Tandem Repeat Finder (v4.0.7) (Benson, 1999) was dispatched to predict tandem repeats. The final repeat content result was integrated using in-house scripts. Before gene model construction, we masked the repeat sequences because of their negative effect on gene model prediction. We downloaded protein sets of 13 species

including *Homo sapiens*, *Mus musculus*, *Gallus gallus*, *Xenopus tropicalis*, *Ornithorhynchus anatinus*, *Danio rerio*, *Oryzias latipes*, *Strongylocentrotus purpuratus*, *Ciona intestinalis*, *Rhincodon typus* and *Callorhinchus milii* from RefSeq (release 82), *Petromyzon marinus* from Ensembl (release 84) and *Branchiostoma floridae* from JGI Genome Portal (<http://genome.jgi-psf.org/Brafl1/Brafl1.home.html>) and aligned them to the masked genome with BLAT(Kent, 2002) to identify positive match regions. GeneWise (v2.2.0) (Birney et al., 2004) was then used to do accurate alignments for target regions and to predict homolog-based gene models.

Transcriptome reads from 14 tissues including blood, eye, gill, heart, liver, muscle, spleen, stomach, dorsal fin, tail fin, pancreas, leptospira, 2 capsulogenous gland and 2 kidney samples were mapped to the genome with HISAT2 (Pertea et al., 2016) and StringTie (Pertea et al., 2016) was used to assemble gene transcripts.

TransDecoder (Haas et al., 2013) was then used to predict the candidate complete ORFs. Further, for *de novo* gene prediction, we employed AUGUSTUS (v3.1) (Stanke et al., 2006) to scan the whole genome with a custom training set generated by using 2,000 high quality genes. Subsequently, we combined the homology-based and *de novo*-predicted gene sets using GLEAN (Elsik et al., 2007) and integrated the GLEAN and transcriptome results with in-house scripts to generate a representative and non-redundant gene set. The final gene set was assigned with a potential function by aligning proteins to databases including KEGG, Swissprot, TrEMBL and InterPro.

Construction of phylogenetic tree

To infer phylogenetic tree of bamboo shark, we selected other five published shark genome including elephant shark (2014), whale shark (2017), brownbanded bamboo shark (2018), cloudy catshark (2018) and white shark (2019), five representative bony fishes including four ray-finned fishes and one lobe-finned fish as well as human to identify single copy orthologous genes. we used TreeFam (Li et al., 2006), defining 33,306 gene families

including 823 single copy orthologous genes. Based on coding sequences of these 823 genes, we constructed trees using both Maximum-likelihood and Bayesian methods with GTR+gamma model, which generated identical trees. Next we estimated divergence time of important branches using PAML mcmctree programme (Yang, 2007; Yang and Rannala, 2006) with the approximate likelihood calculation method.

Evolution of chromosomes

For ancestral chromosomes construction, we identified paralogous genes and orthologous genes by using criteria defined by *Salse et al* (Salse et al., 2009) with both Cumulative Identity Percentage (CIP) and Cumulative Alignment Length Percentage (CALP) value of 0.5 and selected genes pairs defined as A match B best and B match A best. Then MCSCAN(Tang et al., 2008) was used to generate synteny blocks with default parameters. We first noted 54 shared duplications (with 414 paralogous gene pairs) on all the chromosomes. After further integration of these duplications and gene pairs, we found 16 duplicated chromosome pairs and 5 single chromosomes. For identification of conserved genes among bamboo shark, elephant shark, whale shark, brownbanded bamboo shark, cloudy catshark and white shark, spot gar, and medaka, the same criteria with both CIP and CALP value of 0.3 were used. The best hit of multiple matches was selected. Ks values of single copy genes were calculated by KaKs Calculator with default parameter. Heterozygosity of each chromosome was calculated by calling heterozygous SNPs generated by BWA (Li and Durbin, 2009) and SAMtools package (Li, 2011).

MHC genes and P2X gene family analysis

We downloaded coding sequences (CDS) and proteins of *Callorhinchus milii* (NCBI Accession:

GCF_000165045.1), *Rhincodon typus* (NCBI Accession: GCF_001642345.1), *Fugu rubripes* (NCBI Accession: GCF_000180615.1) and *Larimichthys crocea* (NCBI Accession: GCF_000972845.1) from NCBI database, and

Danio rerio, *Latimeria chalumnae*, *Oryzias latipes* and *Gasterosteus aculeatus* from Ensembl database (release 84), and sea lamprey from (<https://genomes.stowers.org/organism/Petromyzon/marinus>) and *Branchiostoma floridae* from JGI Genome Portal (<http://genome.jgi-psf.org/Brafl1/Brafl1.home.html>). Next we carried out initial filtering by discarding sequences less than 30 amino acids and kept the longest transcript if one gene contains multiple transcripts. Subsequently, gene function annotation of these gene sets was performed by using the same method to bamboo shark by using BLAST(Altschul et al., 1990). Then we summarized the MHC genes and P2X gene families by integrating the above information. The P2X-like genes were searched against NCBI database.

Identification of IgNAR

We downloaded IgNAR (immunoglobulin new antigen receptor from cartilaginous) nucleotide sequences from NCBI database and then aligned these sequences to the bamboo shark genome by using BLAST (v2.2.26)(Altschul et al., 1990) with parameters “-F F -m 8 -e 1e-05” to get high scoring pairs (HSPs). Then we clustered the HSPs by integrating the overlaps between HSPs to verify the IgNAR loci. To further verify the IgNAR loci, we re-assembled this region by using single tube long fragment reads (stLFR). Firstly, we mapped the stLFR reads to our initial IgNAR region using BWA (Li and Durbin, 2009) and selected these reads. Then we extracted all reads that were marked with the same barcodes as the mapped reads. Finally, we used Supernova (v2.1.1)(Mohr et al., 2017) to assemble these reads and got the re-assembled IgNAR region with default parameters.

Bamboo shark vNAR phage display library construction

Five bamboo shark individuals (3 male and 2 female, with body length ranging from ~31.50-39.37 inch) were bled and dissected out for spleen, which were submerged in RNAlater for subsequent storage at -80°C. Total

RNA was isolated using the TRIzol reagent (Thermo Fisher Scientific) according to the manufacturer's instruction. We reverse-transcribed RNA into cDNA in a total of 20µl volume with the SuperScript III First-Strand Synthesis System (Thermo Fisher Scientific) according to the manufacturer's instruction. Based on the IgNAR sequences of bamboo shark, we designed and synthesized four pairs of primers for four types of the vNAR domain to amplify the vNAR sequence from the cDNA product by PCR. All reactions were pooled on 1.5% TAE agarose gel and purified by QIAEX II Gel Extraction kit 500 (Qiagen). Purified PCR products and pMECS vector were digested by restriction enzyme Pst I and Not I (NEB). The digested PCR products and vector were purified by QIAEX II Gel Extraction kit 500 (Qiagen). Then the PCR products were inserted into the pMECS vector at a molar ratio of 3:1 using T4 ligase (NEB). Ligation products were purified and transformed into electrocompetent TG1 cells (Lucigen Corporation). The TG1 cells were stored at -80°C by addition of glycerol to a final concentration of 15% (v/v).

Bamboo shark vNAR phage display library screening

Nonadjacent wells of a Maxisorp 96-well were coated with 100µl of 5µg/ml human recombinant antigen Ki67, CD64 and YKL40 (Sino Biological). The plate was sealed by microplate sealing tape and was incubated overnight at 4°C. The following morning, the antigen-coated wells were washed three times with 200µl of PBST, then 200µl blocking solution of 0.5% BSA (Genview) was added in the plate which was incubated for 2h on a vibrating platform at RT. After the blocking solution removed from the antigen-coated wells, the plate was washed five times with 200µl of PBST. For each antigen, 100µl rescued phage library were dispensed into both a selection and a control well. The plate was incubated for 2h on a vibrating platform, then washed five times with 200µl PBST. 100µl triethylamine (1.4% v/v) was added into each selection well and control well for 15min at RT to elute phage bound on the wells. The phage solution was neutralized by 100µl of 1M Tris-HCl (pH 7.4).

The rescued phage were titrated to estimate the number by sequentially mixing 10 μ l of the phage into 90 μ l of LB to 10⁶ fold in a low-binding 96-well plate. 10² and 10³ fold infected TG1 were spread onto the other solid selective medium for single colonies.

Phage ELISA

96 individual clones per target in total were picked at 24 candidates per round and inoculated into wells of a 96-deep-well plate with 500 μ l 2TY containing 100 μ g/ml ampicillin and 2% glucose for 3h at 37 $^{\circ}$ C and 180rpm. Sufficient volume of M13KO7 helper phage (1 μ l) were added into each well at 37 $^{\circ}$ C with shaking overnight. The plate was centrifugated at 3200rpm for 10min, and the supernatant was transferred to a clean 96-deep-well plate for ELISA. The antigens were used to coat a 96 well plate at 2 μ g/ml in PBS buffer, 100 μ l/well, at 4 $^{\circ}$ C overnight. The antigen-coated wells were washed three times with 200 μ l of PBST, then 200 μ l blocking solution of 0.5% BSA was added in the plate which was incubated for 2h on a vibrating platform at RT. The each well supernatant of the deep-well plate were added into antigen-coated well and control well, respectively. The ELISA plate was incubated for 2h at RT, and then washed five times with 200 μ l of PBST. Binding was detected by HRP conjugated mouse anti-M13 antibody(Sino Biological) at 1:5000 concentration and read at 450 nm with a microplate reader Mutiskan Go (Thermo Fisher Scientific).

Supplementary Figures

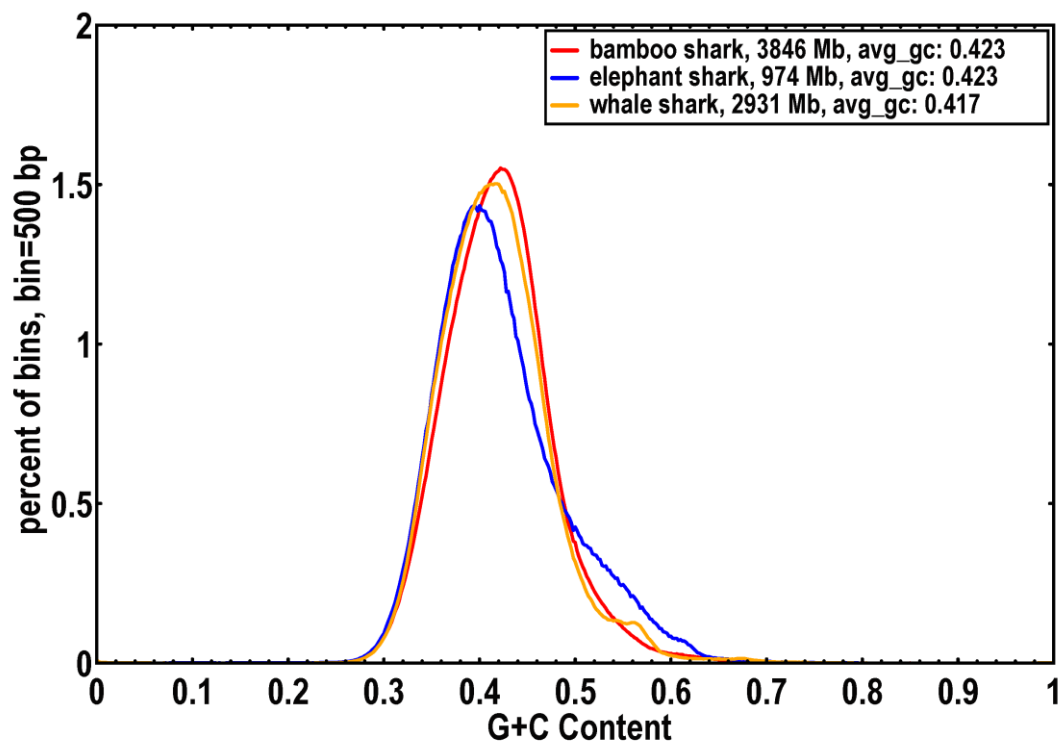


Figure S1. GC content of the assembled bamboo shark, elephant shark, and whale shark genomes, Related to Figure 1.

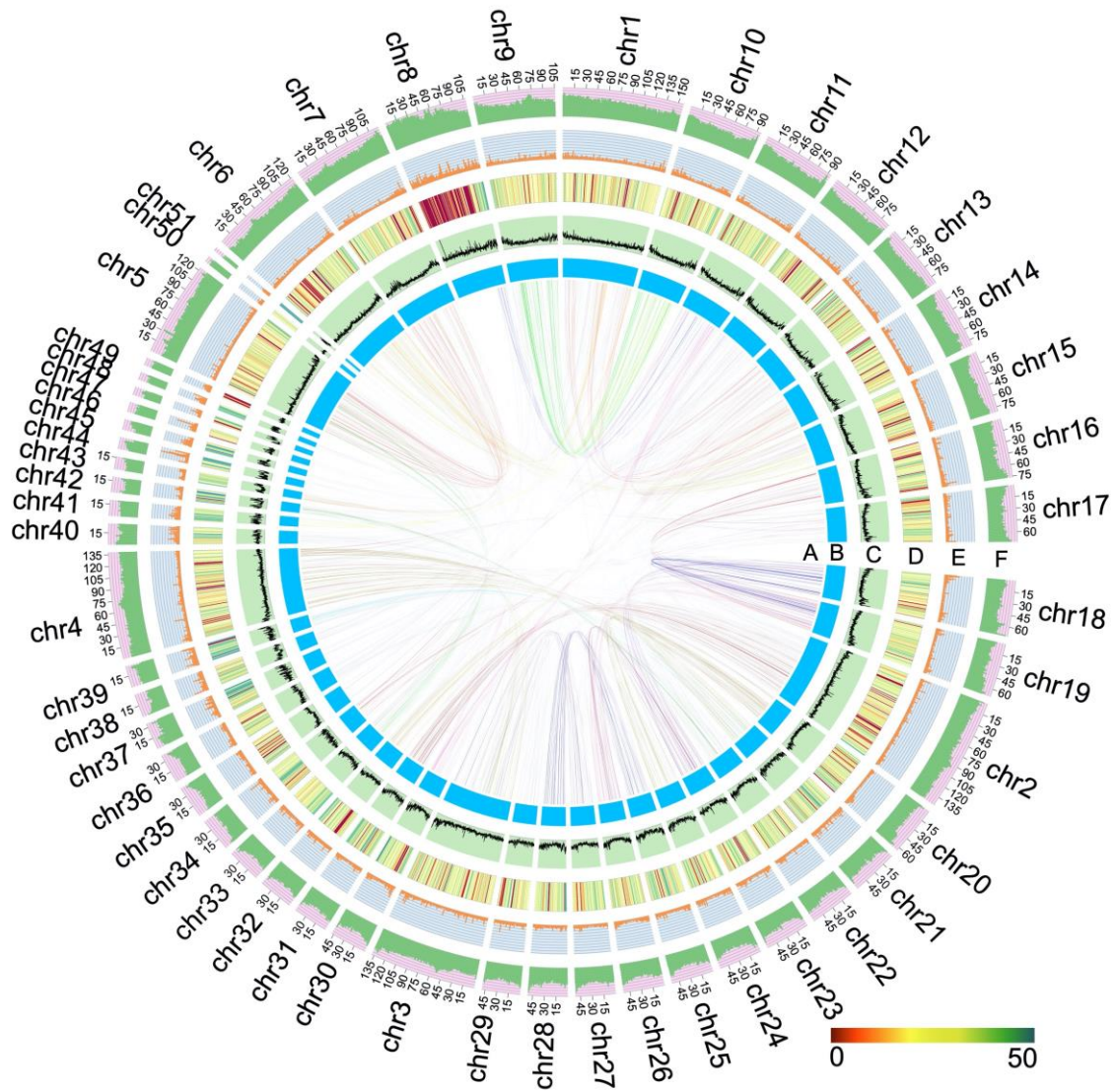


Figure S2. Circos plot of the genomic landscape, Related to Figure 1.

A, Paralogous gene pairs of 51 chromosomes. **B**, 51 chromosomes of the assembled genome. **C**, GC content of the assembled genome at 100kb windows. **D**, gene density per Mb on each chromosome. **E**, histogram of DNA transposon ratios. **F**, histogram of retrotransposon ratios.

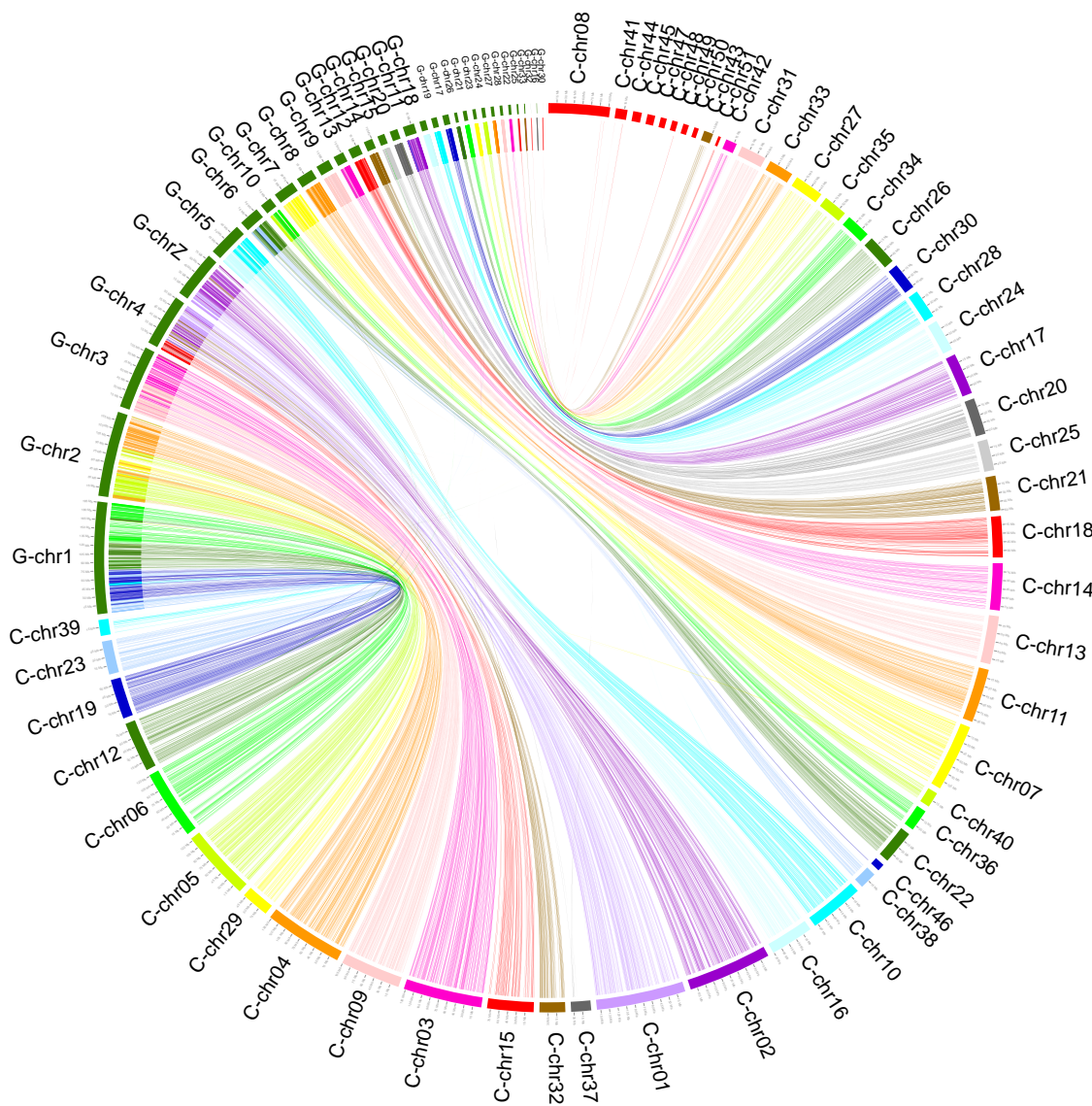


Figure S3. Circos plot of comparisons between bamboo shark and chicken, Related to Figure 1.

G-chr and C-chr represent chicken and bamboo shark chromosomes, respectively. The lines represent gene pairs identified with both CIP and CALP (defined by *Salse et al (Salse et al., 2009)*) value of 0.3.

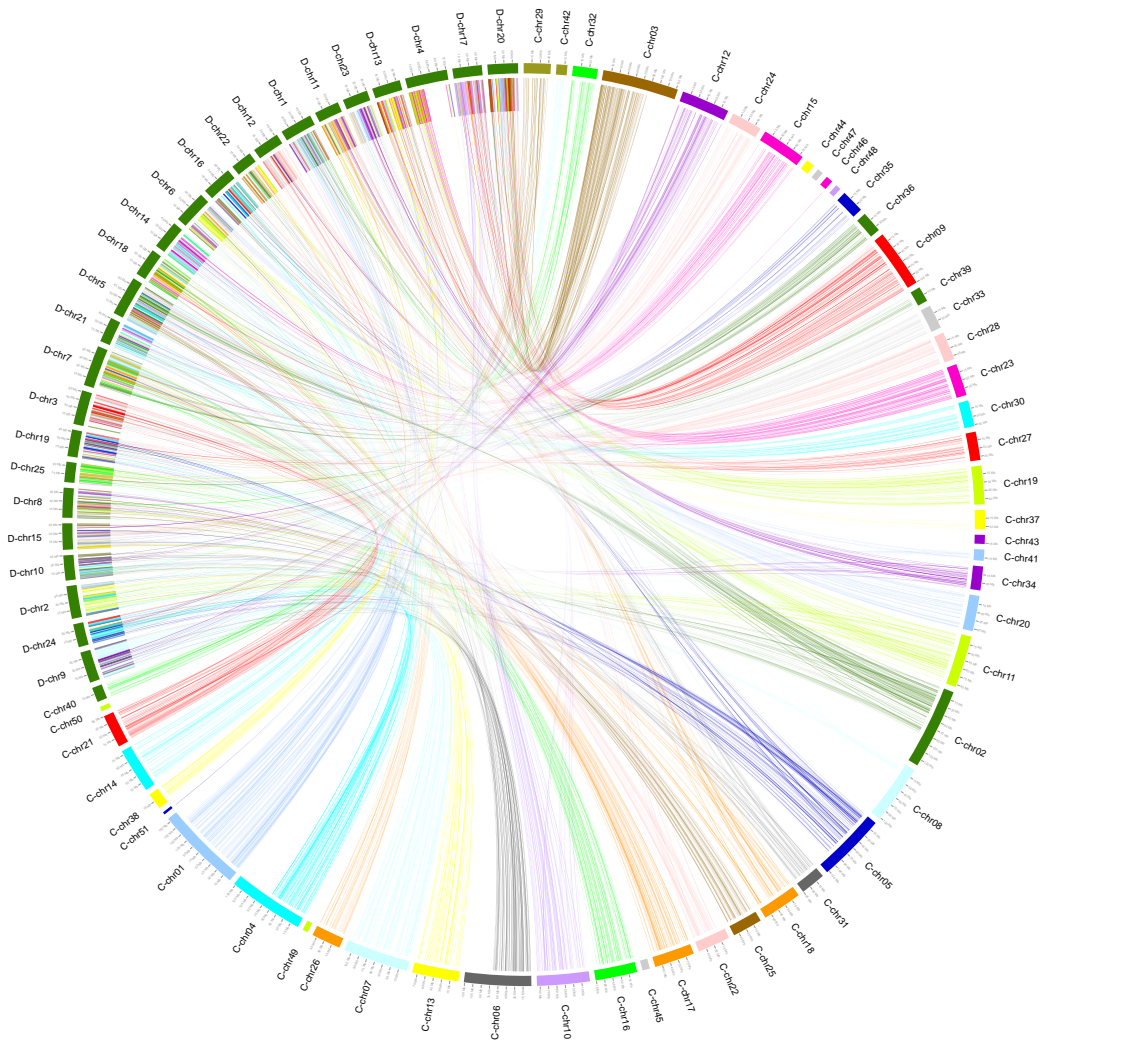


Figure S4. Synteny relationship of bamboo shark and zebrafish, Related to Figure 1.

D-chr and C-chr represent zebrafish and bamboo shark chromosomes, respectively. The lines represent gene pairs identified with both CIP and CALP (defined by *Salse et al (Salse et al., 2009)*) value of 0.3.

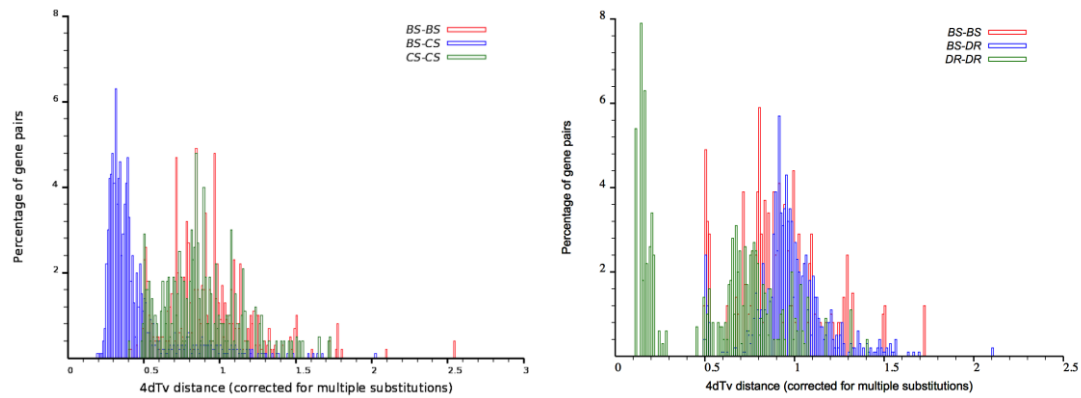


Figure S5. Whole genome duplication in elephant shark, bamboo shark and zebra fish genomes as revealed through 4DTV analyses, Related to Figure 1.

We took zebra fish (DR), elephant shark (CS) and bamboo shark (BS) as examples, detected gene pair blocks by using Mscan and calculated 4DTV values of gene pairs. From the figure we can clearly see that zebra fish has one extra WGD also supported by seven HOX gene clusters (the third WGD event). But in the elephant shark and bamboo shark genomes, we only detected one peak (the second WGD event) which indicates that the first WGD event cannot be detected possibly due to too old WGD event (too high divergence of the paralogous gene pairs) and only can be detected by some conserved genes (for example, HOX genes).

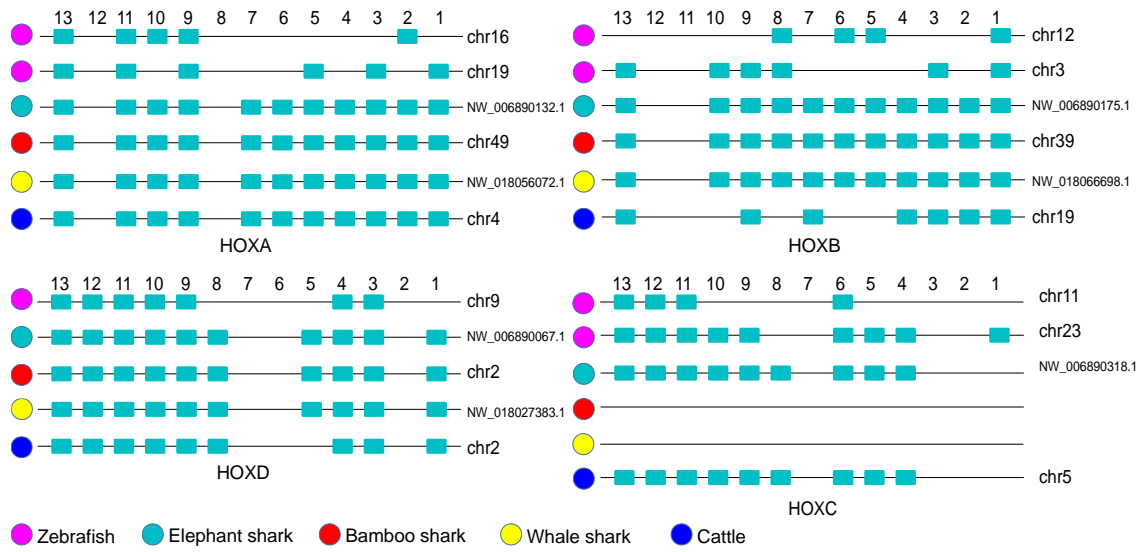


Figure S6. HOX gene clusters identified in five species, Related to Figure 1.

We identified four HOX gene clusters in elephant shark, three in bamboo shark and whale shark (HOXC lost) which suggest sharks have only two rounds of WGD events.

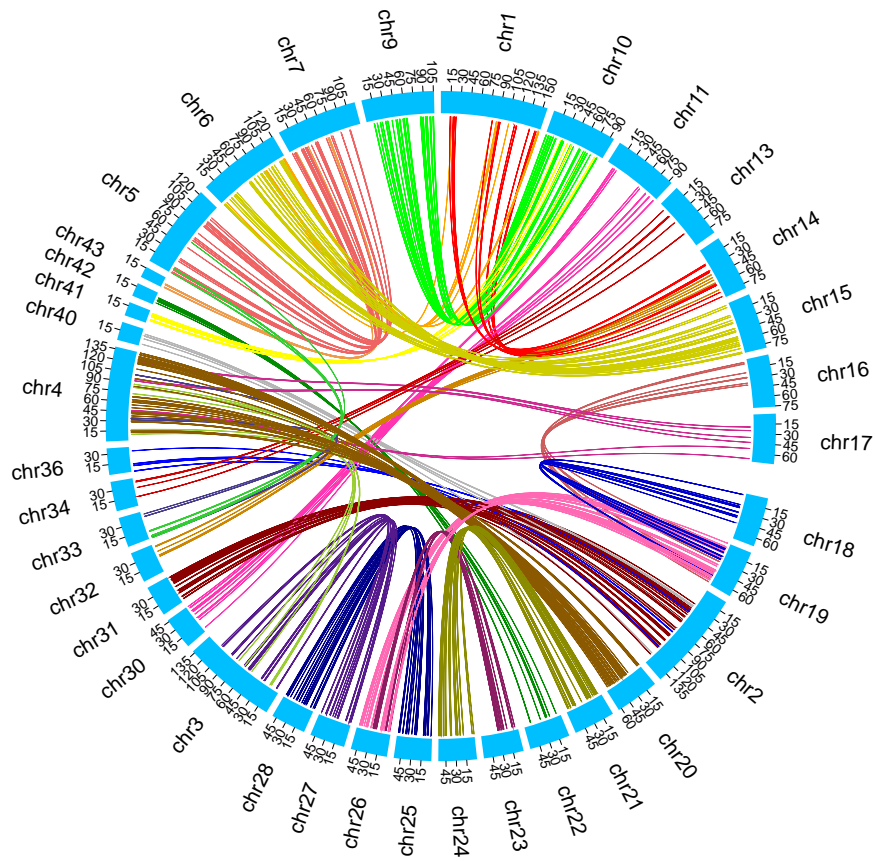


Figure S7. Distribution of shared paralogous genes of bamboo shark and elephant shark, Related to Figure 2.

The lines represent gene pairs obtained by using blastp with both CIP and CALP (defined by *Salse et al*(*Salse et al., 2009*)) value of 0.5.

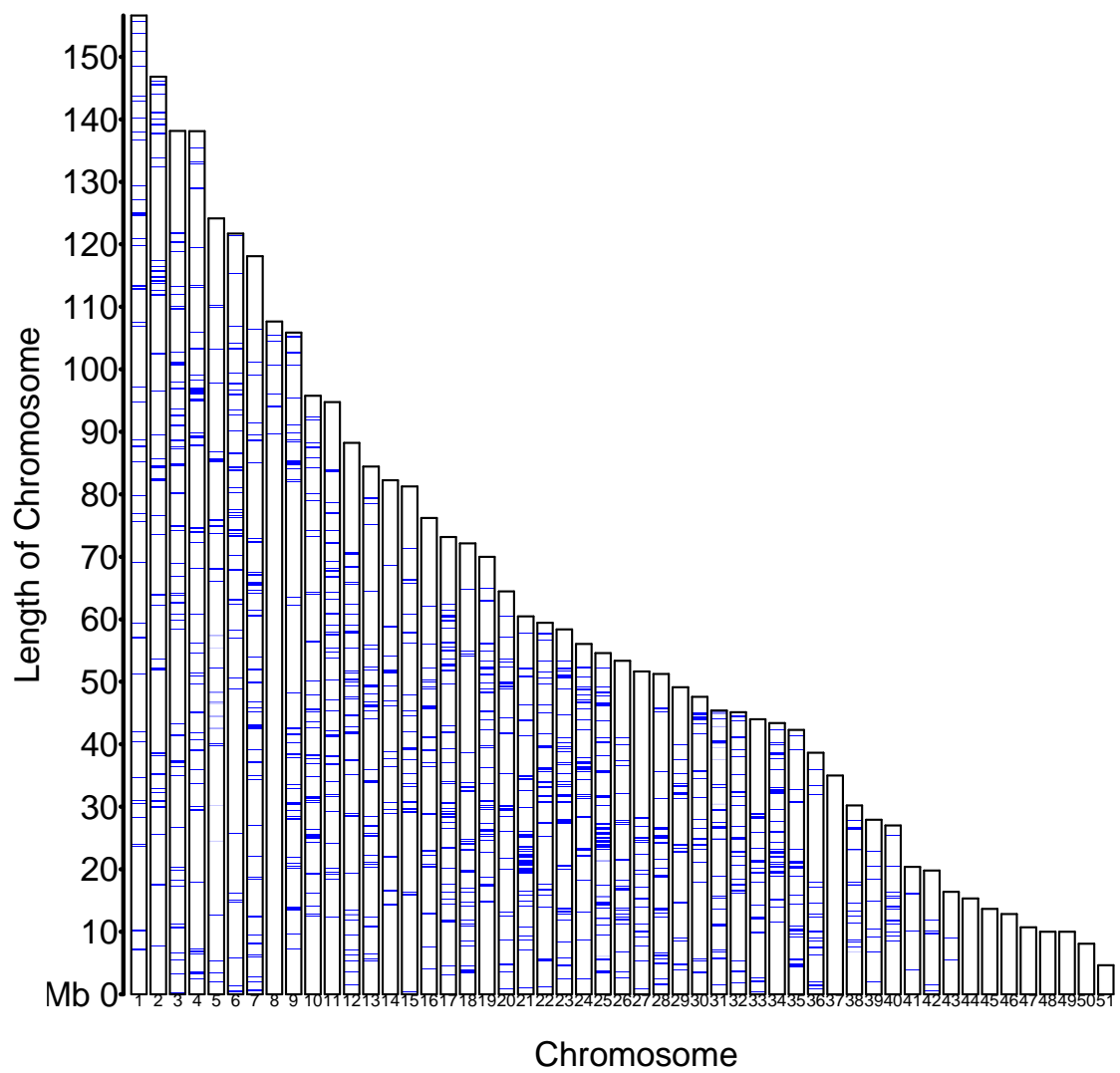


Figure S8. Distributions of conserved regions of six cartilaginous fishes (bamboo shark, elephant shark, whale shark, brownbanded bamboo shark, cloudy catshark and white shark) on 51 chromosomes of the bamboo shark, Related to Figure 2.

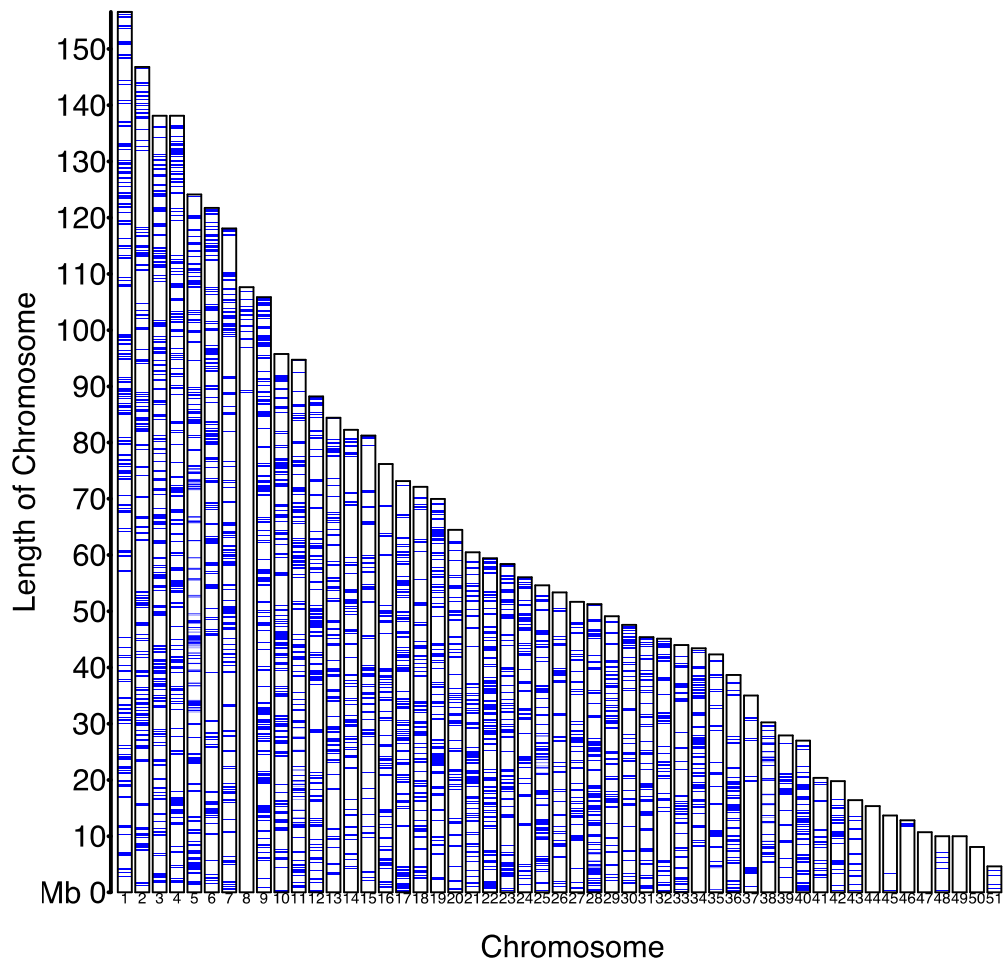


Figure S9. Distributions of conserved regions between bamboo shark and medaka on 51 chromosomes of the bamboo shark, Related to Figure 2.

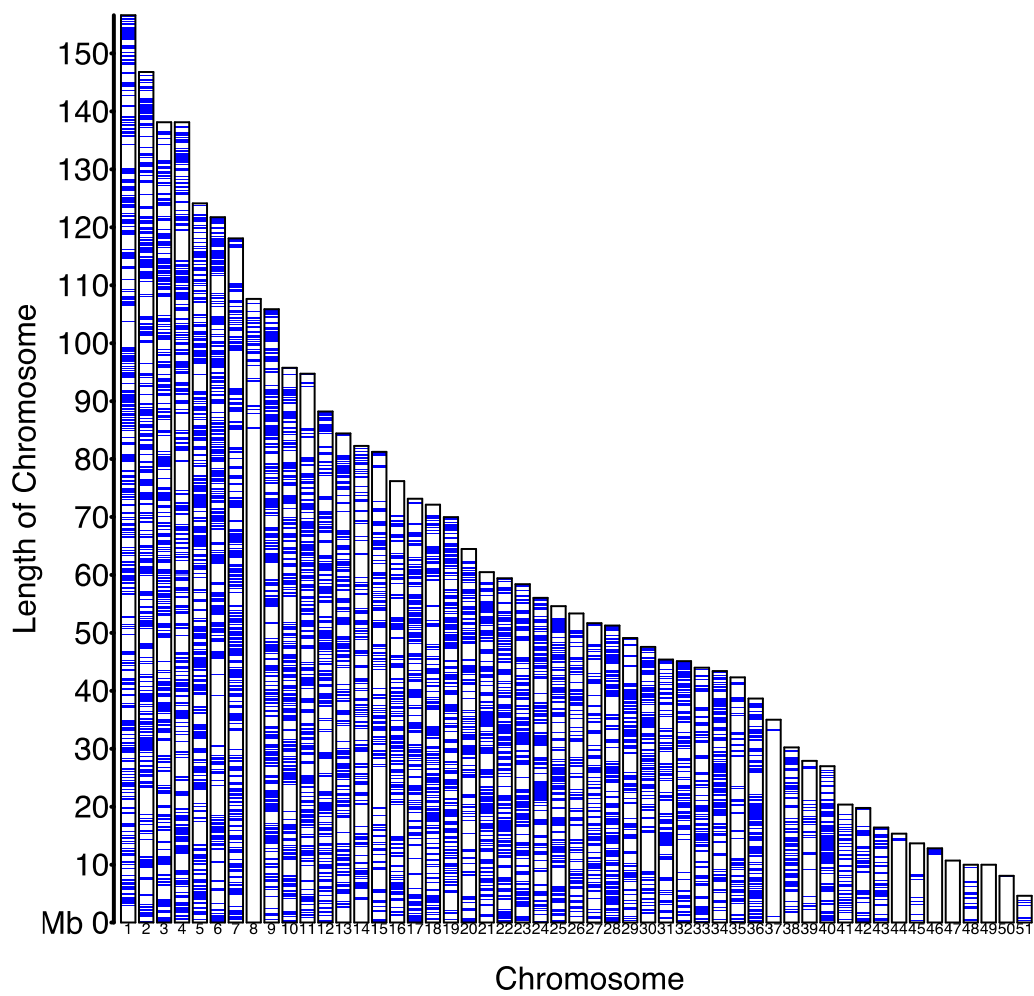


Figure S10. Distributions of conserved regions between bamboo shark and spotted gar on 51 chromosomes of the bamboo shark, Related to Figure 2.

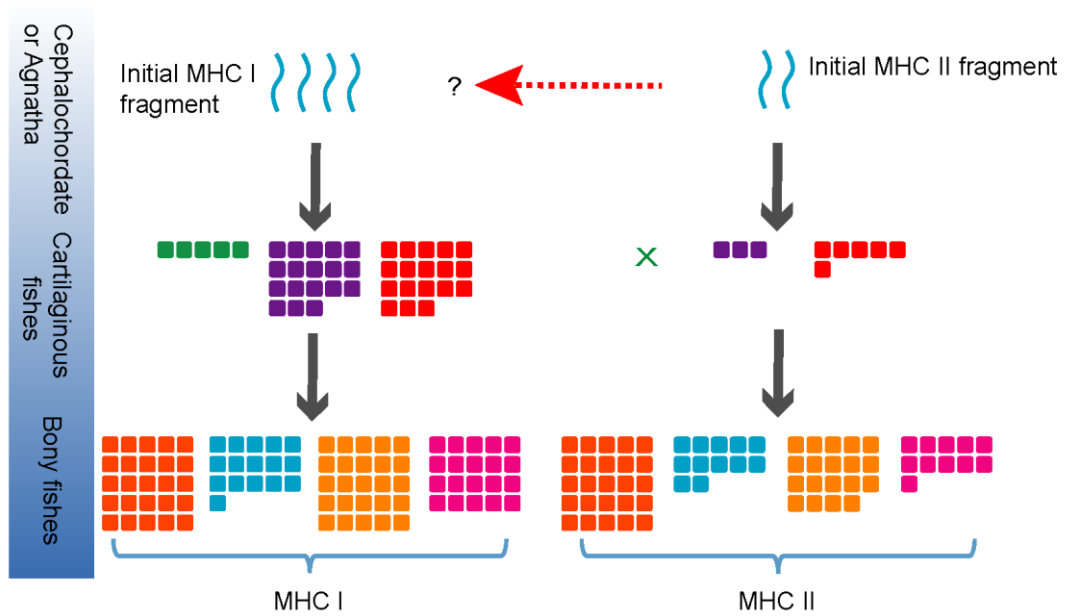


Figure S11. Hypothetical model of MHC genes evolution, Related to Figure 3.

The square numbers represent gene numbers. Green, elephant shark. Purple, whale shark. Red, bamboo shark. Orange, zebrafish. Blue, medaka. Yellow, coelacanth, magenta, three-spined sticklebacks. The green X mark represents there are no MHC class II genes found in the elephant shark genome. The red arrow and question mark represent the hypothesis that MHC class I molecules may be derived from MHC class II while more researches should be done in the future.

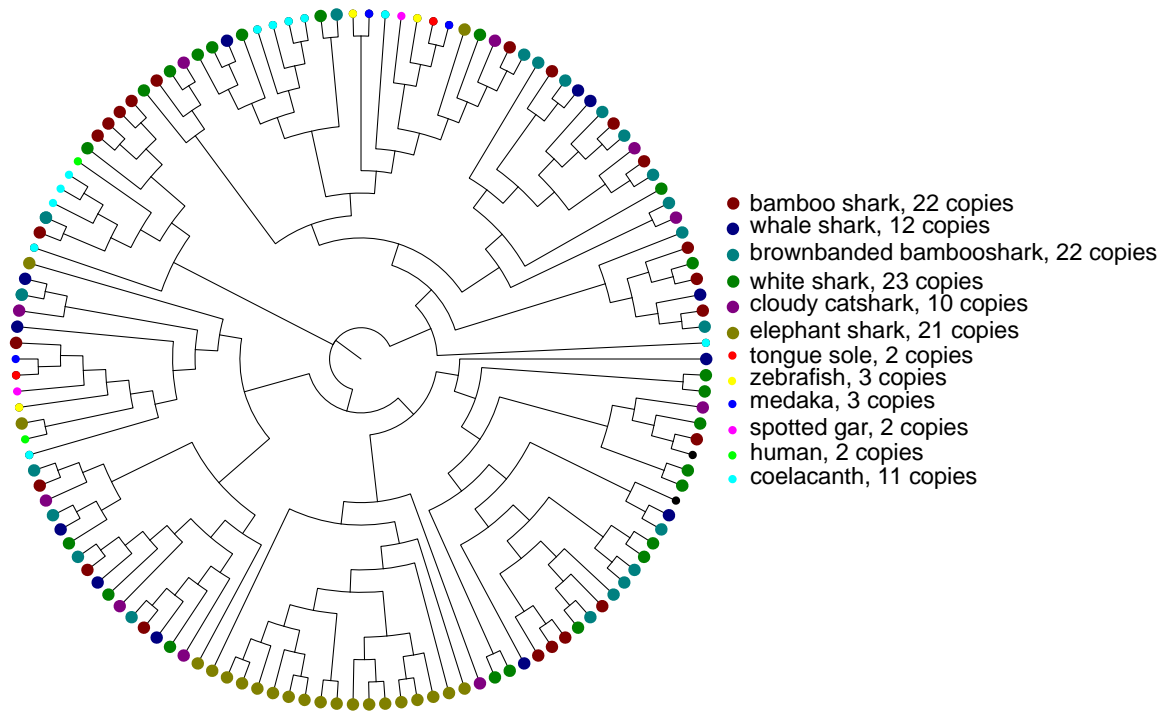


Figure S12. Gene tree of *TRIM69*, Related to Figure 3.

This tree was constructed with protein sequences of these genes by using software of MUSCLE (v3.8.31) (Edgar, 2004) and FastTree (v2.1.10) (Price et al., 2010).

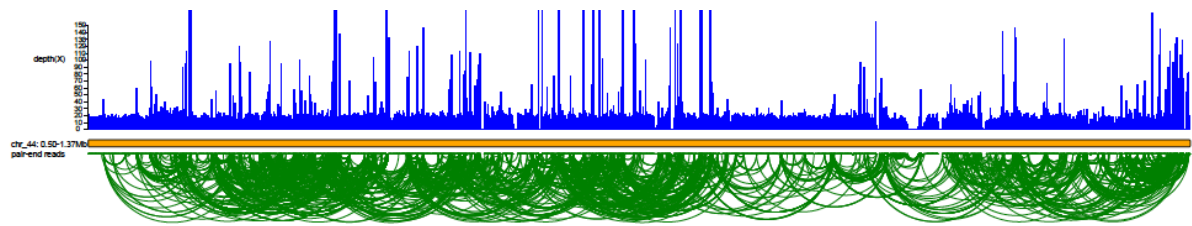


Figure S13. Re-assembled IgNAR region supported by paired-end reads, Related to Figure 3.

All of the non-N bases were supported by reads with average depth of ~33 times. The blue histogram represents sequencing depth of this region. The green curves represent relationship of pair-end reads.

```

1-1-C11 QTPRTITKEAGESLTIINVLKDALFALSSTIYWYLTKLKDATKRDRIISIGGRYSETVNGKSKSFSFLRLRDLRVEDSGAYHCEA-----ETA---AP-----R-----YHE-GRGTF
1-1-F11 QTPKTIIRGATESLTIINVLKDPYSYAFSTIYWYLTKLKDATNKESMSISIGGRYSETVNGKSKSFSFLRITDLRVEDSGTYHCEA-----C---DGYVTRNTWS--E-GSGTI
1-1-B12 QTPRTIITKEIGESLTIINVLKDASYALSSTIYWYLTKLKDATKWDRIISIGGRYSETVNGKSKSFSFLRLRDLRVEDSGAYHCEA-Y--AG-TAAMT-VC-----C---DGYVTRNTWS--E-GSGTI
1-1-G10 QTPRTIITKETGESLTIINVLKDASYALSSTIYWYLTKLKDATKWDRIISIGGRYSETVNGKSKSFSFLRLRDLRVEDSGAYHCEA-Y--GG-----C---DRGG-Y-----YHE-GRGTF
1-1-H12 QTPRTIITKETGESLTIINVLKDASYALSSTIYWYLTKLKDATKWDRIISIGGRYSETVNGKSKSFSFLRLRDLRVEDSGAYHCEA-YV-----C---TAA-IARAN-----YHE-GRGTF
2-0-4 QTPRTIITKETGESLTIINVLKDASYALSSTIYWYLTKLKDATKWDRIISIGGRYSETVNGKSKSFSFLRLRDLRVEDSGAYHCEA-Y--GG-----C---DGGG-Y-----YHE-GRGTF
2-0-6 QTPRTIITKEAGESLTIINVLKDPYSALSSTIYWYLTKLKDATKGDRIISIGGRYSETVSRGSKSFSFLRLRDLRVEDSGAYHCEA-YV-ALYS-----C---ED-----R-YHE-GRGTF
2-0-33 QTPRTIITKETGESLTIINVLKDASYALSSTIYWYLTKLKDATKWDRIISIGGRYSETVNGKSKSFSFLRLRDLRVEDSGAYHCEA-YV-GD-S-----C---EDW-----YHE-GRGTF
1-2-2 QKPRSAITSGESLTIINVLRDVNVGLYSTIDWYWKLGSTNEESISTSERVSVNKGSLKSFSLRISDLRVEDSGTYHCEA-----DDMT--GHS-----LFYVQKSGSTV
2-0-24 QTPRTIITKETGESLTIINVLKGAQVALSSTIYWYLTKLKDATKWDRIISIGGRYSETVNGKSKSFSFLRLRDLRVEDIGVYHCEAND-DDR-----S-----YHE-GRGTF
2-0-5 QTPRTIITKEAGESLTIINVLKGSSTALDSTIYWFYFKKGATKKEASLSTIGGRYSETVNGKSKSFSFLRISDLRVEDSGTYHCEA-YS-SQLGWVSWN-----D---E-GGGTI
1-1-G12 QTPRTIITKEAGESLTIINVLKGSSTALDSTIYWFYFKKGATKKEASLSTIGGRYSETVNGKSKSFSFLRISDLRVEDSGTYHCEA-Y-----TPL---WNP-----D---YHE-GRGTF
2-0-2 QTPRTIITKEAGESLTIINVLKGSSTALDSTIYWFYFKKGATKKEASLSTIGGRYSDTKNTASKSFSFLRISDLRVEDSGTYHCEA-YARD--SWDDL---RGR-----N-GI---E-GGGTI
2-0-8 QTPRTIITKETGESLTIINVLKDSSTALDNTIYWFYFKKGATKKEASLSTIGGRYAETVKNKASKSFSFLRIRDLRVEDSGTYHCEA-YSRVAFD-----LTRR-----SVE-GGSI
2-0-10 QTPRTIITKEAGESLTIINVLKDSSTALDSTIYWFYFKKGATKKEASLSTIGGRYAETVKNKASKSFSFLRISDLRVEDSGTYHCEA--S-GRHSWVR---YTP--GMSY--I---E-GGGTI
2-0-11 QTPRTIITKEAGESLTIINVLKDSSTALDSTHWYFKEDATKKEASLSTIGGRYAETVKNKASKSFSFLRISDLRVEDSGTYHCEA-YSWDVVRHWN---R-D-----I---E-GGGTI
2-0-17 QTPRTIITKEAGESLTIINVLKGSSTALDSTIYWFYFKKGATKKEASLSTIGGRYSDTKNTASKSFSFLRISDLRVEDSGTYHCEA-Y----TAG---TGN--SWS--S-YSE-AGGTI
2-0-26 QTPRTIITKEAGESLTIINVLKGSSTALDSTIYWFYFKKGATKKEASLSTIGGRYSDTKNTASKSFSFLRISDLRVEDSGTYHCEALY-----C---SDLG-----I---E-GGGTI
2-0-29 QTPRTIITKEAGESLTIINVLKRSNIALRSTIYWFYFKKGATKKEASLSTIGGRYSDTKNTASKSFSFLRISDLGVEDSGTYHCEA-Y-RWYDG-----Y---WRE-GGGTI
1-1-A11 QTPRTIITKEAGESLTIINVLKGSSTALDNTIYWFYFKKGATKKEASLSTIGGRYAETVKNKASKSFSFLRISDLRVEDSGTYHCEA-----RAAGTARGYSP-----Y---YHE-GRGTF
1-1-F10 QTPRTIITKEAGESLTIINVLKGSSTALDNTIYWFYFKKGATKKEASLSTIGGRYSDTKNTASKSFSFLRISDLRVEDSGTYHCEA-Y-R-RGWDERAY-----G-IG-----E-GGGTI
1-1-E11 QTPRTIITKEAGESLTIINVLKGSSTALDNTIYWFYFKKGATKKEASLSTIGGRYSDTKNTASKSFSFLRISDLRVEDSGTYHCEA-Y-RAQM-PLR-----C---DTSS--D--Y-E-GGGTI
1-1-G11 QTPRTIITKEAGESLTIINVLKGSSTALDNTIYWFYFKKGATKKEASLSTIGGRYSDTKNTASKSFSFLRISDLRVEDSGTYHCEA-Y-RGVYSWN-----C---E-GS--D-FK-E-GGGTI
1-1-C12 QTPRTIITKEAGESLTIINVLKDSSTALDNTIYWFYFKKGATKKEASLSTIGGRYAETVKNKASKSFSFLRISDLRVEDSGTYHCEAY-----GSA--GR--G-YHE-GRGTF
1-1-E12 QTPRTIITKEAGESLTIINVLKGSSTALDNTIYWFYFKKGATKKEASLSTIGGRYSDTKNTASKSFSFLRISDLRVEDSGTYHCEALYSRDD-----C---TINELGR-YD--YHE-GRGTF
2-0-9 QTPRTIITKEAGESLTIINVLKGSSTALDNTIYWFYFKKGATKKEASLSTIGGRYAETVKNKASKSFSFLRISDLRVEDSGTYHCEA-----AGMT---GVAIAGP-----YHE-GRGTF
2-0-15 QTPRTIITKEAGESLTIINVLKGSSTALDNTIYWFYFKKGATKKEASLSTIGGRYAETVKNKASKSFSFLRIIDL RVEDSGTYHCEA-----LE-----GT-RIAGL-----YHE-GRGTF
2-0-28 QTPRTIITKEAGESLTIINVLKGSSTALDNTIYWFYFKKGATKKEASLSTIGGRYAETVKNKASKSFSFLRISDLRVEDSGTYHCEA-----PHSWDA---GT-RIAGL-----IHYE-GRGTF
2-0-30 QTPRTIITKEAGESLTIINVLKGSSTALDNTIYWFYFKKGATKKEASLSTIGGRYAETVKNKASKSFSFLRISDLRVEDSGTYHCEA-----C---DSYVEFP-----HYE-GRGTF
2-0-31 QTPRTIITKEAGESLTIINVLKDSSTALDNTIYWFYFKKGATKKEASLSTIGGRYAETVKNKASKSFSFLRINLRSEERGTYHCEA-----RLGR--N-----RV-----P--D---YHE-GRGTF

```

Figure S14. Sequence alignment of the randomly picked clones from the bamboo shark vNAR-phage display library that represent sequences have different CDR3, Related to Figure 3.

The areas highlighted by yellow are the CDRs determined by IMGT database. The areas highlighted by red are the cysteines.

Supplementary Tables

Table S1. Summary of the sequencing data from WGS libraries and Hi-C libraries, Related to Figure 1.

Strategy	Insert	Reads	Total
	Size (bp)	Length (bp)	Data (Gb)
WGS	170	100_100	383.98
	350	100_100	387.24
	2,000	50_50	65.36
	5,000	50_50	142.20
	10,000	50_50	71.34
	20,000	50_50	57.56
HIC	\	50_50	71.95
Total	\	\	1179.63

Table S2. Basic statistics of the assembled bamboo shark genome, Related to Figure 1.

	WGS		HIC	
	Contig (bp)	Scaffold (bp)	Contig (bp)	Scaffold (bp)
N90	1,473	2,360	1,415	2,218
N80	9,471	172,678	8,377	20,708,781
N70	18,406	430,414	17,000	44,175,687
N60	27,604	684,074	25,851	49,995,320
N50	37,247	962,547	35,451	57,918,702
N40	48,650	1,251,191	46,889	71,834,548
N30	62,981	1,595,392	60,770	80,155,495
N20	82,569	2,081,779	80,154	93,009,420
N10	115,410	2,827,051	112,378	99,330,644
max_length	7,042,168	438,259	438,259	136,536,876
Total length	3,647,281,518	3,846,300,841	3,647,255,807	3,851,610,035

Table S3. Summary of the functional annotation of genes in bamboo shark genome, Related to Figure 1.

	Number	Percent (%)
Total	19,595	100

InterPro	16,938	86.44
GO	13,147	67.09
KEGG	17,171	87.63
Swiss-Prot	18,507	94.45
TrEMBL	15,622	79.72
Annotated	19,371	98.86
Unannotated	224	1.24

Table S4. Summary of the repeat content in bamboo shark genome, Related to Figure 1.

Type	Repeat Size	% of genome
Trf	193,033,335	5.01
Repeatmasker	945,921,870	24.56
Proteinmask	855,315,381	22.24
De novo	2,266,569,286	58.85
Total	2,447,074,981	63.53

Table S5. Summary of the TE content in three cartilaginous fishes and seven bony fishes, Related to Figure 1.

	Type	DNA	LINE	SINE	LTR	Total
Bamboo shark	Length (bp)	84,246,842	1,841,705,606	151,630,200	707,421,972	2,240,930,640
	% in genome	2.19	47.88	3.94	18.39	58.26
Elephant shark	Length (bp)	20,045,841	375,567,981	150,517,703	119,161,620	444,674,297
	% in genome	2.06	38.54	15.45	12.23	45.63
Whale shark	Length (bp)	44,888,512	1,482,979,651	102,877,514	120,791,609	1,608,053,993
	% in genome	1.53	50.59	3.51	4.12	54.85
Coelacanth	Length (bp)	682,694,273	654,465,678	345,675,101	219,465,253	1,253,477,861
	% in genome	23.87	22.88	12.08	7.67	43.82
Three-spined sticklebacks	Length (bp)	35,650,096	28,697,415	3,533,002	30,466,795	82,399,819
	% in genome	7.72	6.22	0.77	6.6	17.85
	Length (bp)	50,367,111	26,418,884	3,139,446	17,332,054	85,823,340

Large yellow croaker	% in genome	7.77	4.07	0.48	2.67	13.24
Medaka	Length (bp)	126,716,317	90,435,719	9,001,570	62,654,967	253,215,905
	% in genome	14.58	10.41	1.04	7.21	29.14
Torafugu	Length (bp)	21,275,275	22,316,804	1,056,847	15,374,370	49,743,515
	% in genome	5.43	5.7	0.27	3.93	12.71
Zebrafish	Length (bp)	599,266,189	59,287,717	36,817,925	79,463,691	744,738,398
	% in genome	43.69	4.32	2.68	5.79	54.29
Tongue soles	Length (bp)	36,780,534	17,599,748	1,876,835	11,627,464	59,151,167
	% in genome	7.82	3.74	0.4	2.47	12.58

Table S6. Evaluation of the final gene set by using BUSCO, Related to Figure 1.

Description	Gene number	Percent (%)
Complete BUSCOs (C)	2,477	95.78
Complete and single-copy BUSCOs (S)	2,372	91.72
Complete and duplicated BUSCOs (D)	105	4.06
Fragmented BUSCOs (F)	68	2.62
Missing BUSCOs (M)	41	1.59
Total BUSCO groups searched	2,586	100.00

Table S7. Distribution of paralogous genes in the bamboo shark genome. The yellow highlighted numbers represent gene pairs more than 20, Related to Figure 2.

chr5	575	124,146,585	23	4.00%	1,105,727	0.89%
chr6	585	121,738,290	50	8.55%	2,069,570	1.70%
chr7	622	118,101,210	47	7.56%	2,127,210	1.80%
chr8	243	107,642,081	12	4.94%	480,197	0.45%
chr9	558	105,880,166	39	6.99%	2,021,688	1.91%
chr10	538	95,765,976	42	7.81%	1,588,937	1.66%
chr11	572	94,731,789	61	10.66%	2,023,263	2.14%
chr12	518	88,215,905	39	7.53%	2,187,022	2.48%
chr13	528	84,432,200	50	9.47%	2,150,712	2.55%
chr14	412	82,264,610	29	7.04%	1,163,437	1.41%
chr15	397	81,258,972	39	9.82%	1,354,341	1.67%
chr16	394	76,207,864	30	7.61%	1,873,298	2.46%
chr17	441	73,167,756	38	8.62%	1,830,134	2.50%
chr18	404	72,140,527	31	7.67%	1,036,258	1.44%
chr19	430	69,975,007	35	8.14%	1,285,421	1.84%
chr20	362	64,470,864	34	9.39%	1,182,523	1.83%
chr21	446	60,475,494	47	10.54%	2,309,047	3.82%
chr22	366	59,434,692	23	6.28%	847,973	1.43%
chr23	350	58,390,607	43	12.29%	2,161,640	3.70%
chr24	357	56,061,091	36	10.08%	1,334,444	2.38%
chr25	386	54,600,476	55	14.25%	2,741,762	5.02%
chr26	325	53,348,395	26	8.00%	1,286,581	2.41%
chr27	303	51,668,882	32	10.56%	810,993	1.57%
chr28	348	51,251,456	35	10.06%	1,381,797	2.70%
chr29	249	49,130,325	21	8.43%	979,109	1.99%
chr30	324	47,581,304	26	8.02%	1,099,286	2.31%
chr31	389	45,411,694	34	8.74%	1,264,812	2.79%
chr32	255	45,129,243	20	7.84%	990,196	2.19%
chr33	307	43,998,411	27	8.79%	859,422	1.95%
chr34	311	43,401,492	37	11.90%	1,292,584	2.98%
chr35	217	42,304,929	29	13.36%	1,096,282	2.59%
chr36	239	38,655,811	16	6.69%	689,636	1.78%
chr37	360	35,007,441	1	0.28%	1,574	0.00%

chr38	205	30,225,729	12	5.85%	441,029	1.46%
chr39	340	27,915,349	11	3.24%	259,694	0.93%
chr40	205	26,992,261	12	5.85%	358,743	1.33%
chr41	212	20,367,195	5	2.36%	142,612	0.70%
chr42	225	19,796,787	12	5.33%	221,910	1.12%
chr43	149	16,392,901	5	3.36%	109,590	0.67%
chr44	177	15,333,346	0	0.00%	0	0.00%
chr45	188	13,660,452	0	0.00%	0	0.00%
chr46	57	12,823,346	0	0.00%	0	0.00%
chr47	73	10,712,325	0	0.00%	0	0.00%
chr48	124	9,987,384	0	0.00%	0	0.00%
chr49	13	9,983,820	0	0.00%	0	0.00%
chr50	49	8,079,144	0	0.00%	0	0.00%
chr51	97	4,637,688	0	0.00%	0	0.00%

Table S9. Enrichment analysis for genes in chromosomes 8, 37, 39, 41, 43, 44, 45, 46, 47,48,49, 50 and 51, Related to Figure 2.

Pathway	Level	Gene number	Q-value
Systemic lupus erythematosus	Immune diseases	66	6.08E-23
Staphylococcus aureus infection	Infectious diseases: Bacterial	47	1.40E-14
Asthma	Immune diseases	36	1.84E-14
Intestinal immune network for IgA production	Immune system	40	4.66E-14
Malaria	Infectious diseases: Parasitic	37	1.05E-12
Measles	Infectious diseases: Viral	59	2.02E-10
Mineral absorption	Digestive system	48	3.52E-10
Allograft rejection	Immune diseases	36	1.15E-09

NF-kappa B signaling pathway	Signal transduction	54	1.34E-09
Autoimmune thyroid disease	Immune diseases	37	6.19E-09
Graft-versus-host disease	Immune diseases	28	1.25E-08
Inflammatory bowel disease (IBD)	Immune diseases	32	4.87E-08
T cell receptor signaling pathway	Immune system	45	6.48E-08
Chagas disease (American trypanosomiasis)	Infectious diseases: Parasitic	43	7.64E-08
Type I diabetes mellitus	Endocrine and metabolic diseases	28	2.04E-07
Viral myocarditis	Cardiovascular diseases	40	2.04E-06
Rheumatoid arthritis	Immune diseases	41	2.39E-06
Antigen processing and presentation	Immune system	34	6.26E-06
Cell adhesion molecules (CAMs)	Signaling molecules and interaction	51	5.33E-05
Alcoholism	Substance dependence	39	5.71E-05
HTLV-I infection	Infectious diseases: Viral	63	2.90E-04
Transcriptional misregulation in cancer	Cancers: Overview	56	7.29E-04
Ras signaling pathway	Signal transduction	56	7.22E-03
Calcium signaling pathway	Signal transduction	52	8.39E-03
Leishmaniasis	Infectious diseases: Parasitic	22	8.68E-03
Rap1 signaling pathway	Signal transduction	57	8.90E-03
Viral carcinogenesis	Cancers: Overview	44	2.35E-02
Tuberculosis	Infectious diseases: Bacterial	37	6.91E-02

Table S12. List of MHC class II genes in analyzed species, Related to Figure 3.

Species	Gene ID	Description from KEGG
Medaka	ENSORLT0000000027	HLA class II histocompatibility antigen, DP alpha 1 chain-like;
	ENSORLT00000016063	rano class II histocompatibility antigen, A beta chain-like;
	ENSORLT00000005537	CD74 molecule, major histocompatibility complex, class II invariant chain a
	ENSORLT00000016021	Orla-DDA; RLA class II histocompatibility antigen, DP alpha-1 chain;
	ENSORLT00000000030	Orla-DCB; H-2 class II histocompatibility antigen, E-S beta chain;
	ENSORLT00000023575	Orla-DAA; mamu class II histocompatibility antigen, DR alpha chain;
	ENSORLT00000023543	Orla-DAB; H-2 class II histocompatibility antigen, E-S beta chain;
	ENSORLT00000011500	Orla-DEA; RLA class II histocompatibility antigen, DP alpha-1 chain-like;
	ENSORLT00000016052	Orla-DDB; rano class II histocompatibility antigen, A beta chain;
	ENSORLT00000024164	Orla-DFA; mamu class II histocompatibility antigen, DR alpha chain;
	ENSORLT00000006056	H-2 class II histocompatibility antigen, A-U alpha chain-like;
	ENSORLT00000024129	H-2 class II histocompatibility antigen, E-S beta chain-like;
Coelacanth	ENSLACT00000001762	H-2 class II histocompatibility antigen, E-S beta chain-like;
	ENSLACT00000000411	DLA class II histocompatibility antigen, DR-1 beta chain-like;
	ENSLACT00000000328	SLA class II histocompatibility antigen, DQ haplotype C beta chain-like;
	ENSLACT00000000754	HLA class II histocompatibility antigen, DP alpha 1 chain-like;
	ENSLACT00000000891	rano class II histocompatibility antigen, D-1 beta chain-like;
	ENSLACT00000003230	RLA class II histocompatibility antigen, DP alpha-1 chain-like;
	ENSLACT00000000197	H-2 class II histocompatibility antigen, A-D beta chain-like;
	ENSLACT00000000493	DLA class II histocompatibility antigen, DR-1 beta chain-like;
	ENSLACT00000000162	rano class II histocompatibility antigen, D-1 beta chain-like;
	ENSLACT00000002416	RLA class II histocompatibility antigen, DP alpha-1 chain-like;
	ENSLACT00000002328	DLA class II histocompatibility antigen, DR-1 beta chain-like;
	ENSLACT00000009069	DLA class II histocompatibility antigen, DR-1 beta chain-like;
ENSLACT00000004598	SLA class II histocompatibility antigen, DQ haplotype C beta chain-like;	
ENSLACT00000000968	H-2 class II histocompatibility antigen, A-B alpha chain-like;	
ENSLACT00000000721	H-2 class II histocompatibility antigen, A-U alpha chain-like;	

	ENSLACT00000000561	DLA class II histocompatibility antigen, DR-1 beta chain-like;
	ENSLACT00000008726	DLA class II histocompatibility antigen, DR-1 beta chain-like;
	ENSLACT00000000289	SLA class II histocompatibility antigen, DQ haplotype C beta chain-like;
	ENSLACT00000006607	H-2 class II histocompatibility antigen, E-S beta chain-like;
	ENSGACT00000000421	H-2 class II histocompatibility antigen, A-U alpha chain-like;
	ENSGACT00000025242	Orla-DAA; mamu class II histocompatibility antigen, DR alpha chain;
	ENSGACT00000000431	H-2 class II histocompatibility antigen, E-S beta chain-like;
	ENSGACT00000004860	rano class II histocompatibility antigen, A beta chain-like;
Three-	ENSGACT00000025238	Orla-DAB; H-2 class II histocompatibility antigen, E-S beta chain;
spined	ENSGACT00000000450	Orla-DAB; H-2 class II histocompatibility antigen, E-S beta chain;
sticklebacks	ENSGACT00000000437	HLA class II histocompatibility antigen, DRB1-8 beta chain-like;
	ENSGACT00000023783	H-2 class II histocompatibility antigen, E-S beta chain-like;
	ENSGACT00000004910	H-2 class II histocompatibility antigen, A-U alpha chain-like;
	ENSGACT00000000439	Orla-DAA; mamu class II histocompatibility antigen, DR alpha chain;
	ENSGACT00000000434	H-2 class II histocompatibility antigen, A-U alpha chain-like;
	ENSDART00000006898	novel protein with a Class II histocompatibility antigen, alpha domain and a Immunoglobulin C1-set domain;
	ENSDART00000161194	MHCII, si:busm1-48c11.4, si:dz194e12.12, si:dz48c11.4, zgc:123061; si:busm1-194e12.12;
	ENSDART00000102847	si:dz48c11.3; si:busm1-48c11.3;
	ENSDART00000160609	mhc2dab, major histocompatibility complex class II DAB gene;
	ENSDART00000129901	novel protein with a Class II histocompatibility antigen, alpha domain and a Immunoglobulin C1-set domain;
Zebrafish	ENSDART00000109439	MHCII, si:busm1-160c18.10, si:dz160c18.10, si:dz194e12.8, zgc:101701; si:busm1-194e12.8;
	ENSDART00000099281	mhc2dcb, major histocompatibility complex class II DCB gene;
	ENSDART00000097932	MHCII, si:busm1-48c11.4, si:dz194e12.12, si:dz48c11.4, zgc:123061; si:busm1-194e12.12;
	ENSDART00000159361	MHCII, si:busm1-48c11.4, si:dz194e12.12, si:dz48c11.4, zgc:123061; si:busm1-194e12.12;
	ENSDART00000162877	si:dz48c11.3; si:busm1-48c11.3;

	ENSDART00000159276	MHCII, si:busm1-48c11.4, si:dz194e12.12, si:dz48c11.4, zgc:123061; si:busm1-194e12.12;
	ENSDART00000123658	mhc2dcb, si:busm1-243a08.1, si:busm1-37i06.8, si:dz243a08.1, si:dz37i06.8; major histocompatibility complex class II DCB gene;
	ENSDART00000053205	CD74 molecule, major histocompatibility complex, class II invariant chain a;
	ENSDART00000168831	mhc2dab, major histocompatibility complex class II DAB gene;
	ENSDART00000168076	mhc2dab,major histocompatibility complex class II DAB gene;
	ENSDART00000171760	si:dz48c11.3; si:busm1-48c11.3;
	ENSDART00000026021	CD74 molecule, major histocompatibility complex, class II invariant chain a;
	ENSDART00000158400	MHCII, si:busm1-160c18.10, si:dz160c18.10, si:dz194e12.8, zgc:101701; si:busm1-194e12.8;
	ENSDART00000158784	mhc2dcb, si:busm1-243a08.1, si:busm1-37i06.8, si:dz243a08.1, si:dz37i06.8; major histocompatibility complex class II DCB gene;
	ENSDART00000048448	uncharacterized LOC791723;
	ENSDART00000105025	H-2 class II histocompatibility antigen, E-D beta chain-like;
Whale shark	XP_020390554.1	ORAN-DRA; MHC class II DR alpha;
	XP_020390555.1	DRB, DRB1, Mamu-DRB, Mamu-DRB1, Mane-DRB; HLA class II histocompatibility antigen, DRB1-3 chain;
	XP_020392206.1	DLA class II histocompatibility antigen, DR-1 beta chain-like;
Bamboo shark	Chipl12839	DLA class II histocompatibility antigen, DR-1 beta chain-like;
	Chipl12840	RLA class II histocompatibility antigen, DP alpha-1 chain-like;
	Chipl12907	RLA class II histocompatibility antigen, DP alpha-1 chain-like;
	Chipl01708	class II histocompatibility antigen, B-L beta chain-like;
	Chipl14140	SAHA-DAB; HLA class II histocompatibility antigen, DRB1-15 beta chain;
	Chipl18694	HLA class II histocompatibility antigen, DR alpha chain-like;

Table S13.MHC class II fragments detected by using BLAST in the elephant shark genome, Related to Figure 3.

Target scaffolds	Query sequences	Identity	Target Start	Target End	Query start	Query End	E value
NW_006895284.1	ENSGACT00000025242	93.55	2641	2671	32	2	2.00E-06
NW_006895810.1	ENSGACT00000025242	93.55	3175	3205	32	2	2.00E-06

NW_006896611.1	ENSGACT00000025242	93.55	2160	2190	2	32	2.00E-06
NW_006896785.1	ENSGACT00000025242	93.55	283	313	2	32	2.00E-06
NW_006896886.1	ENSGACT00000025242	93.55	1221	1251	32	2	2.00E-06
NW_006898245.1	Chip112840	86.96	355	400	299	344	4.00E-06
NW_006898335.1	ENSGACT00000025242	93.33	175	204	32	3	4.00E-06
NW_006900613.1	ENSGACT00000025242	93.55	354	384	2	32	9.00E-07
NW_006910024.1	ENSGACT00000025242	93.55	622	652	32	2	6.00E-07

Table S14. Designed primers for amplifying vNARs, Related to Figure 3.

Forward primer	Reverse primer
1 F1:GGGTTGAACACCGACAAACTGCAGGGGT TGAACAAACACCGACA	R1:ATAAGAATGCGGCCGCAATCCATTTG CCCTCTGTTCT
2 F2:AACTGCAGCCAAACACCGAGAACAATAAC G	R2:ATAAGAATGCGGCCGCGAAATCCGT TTGCTCTCTGTTCT
3 F3:AACTGCAGCCAAACACCGAAAACGATAAT AA	R3:ATAAGAATGCGGCCGCGCCAGAAATCC GTTTACTCTCTGTTCT
4 F4:AACTGCAGCCGAGATCAGCAACAAAAC TC	R4:ATAAGAATGCGGCCGCAATCCGTTTG CTCTCTTTTCTTC

Table S15. Ancestral P2X gene found in amphioxus and ascidiacea genomes, Related to Figure 4.

Name/Gene ID	Description
LOC109481458	P2X purinoceptor 7-like [Branchiostoma belcheri (Belcher's lancelet)]
LOC109474372	P2X purinoceptor 7-like [Branchiostoma belcheri (Belcher's lancelet)]
LOC109472980	P2X purinoceptor 7-like [Branchiostoma belcheri (Belcher's lancelet)]
LOC109464342	P2X purinoceptor 4-like [Branchiostoma belcheri (Belcher's lancelet)]
LOC109462847	P2X purinoceptor 7-like [Branchiostoma belcheri (Belcher's lancelet)]
LOC109462011	P2X purinoceptor 7-like [Branchiostoma belcheri (Belcher's lancelet)]
BRAFLDRAFT_84310(P2RX4)	hypothetical protein [Branchiostoma floridae (Florida lancelet)]
BRAFLDRAFT_103732	hypothetical protein [Branchiostoma floridae (Florida lancelet)]
LOC101242976	P2X purinoceptor 7-like [Ciona intestinalis (vase tunicate)]
LOC101242780	P2X purinoceptor 7-like [Ciona intestinalis (vase tunicate)]

Table S16. Protein length of P2X genes in bony fishes and sea lamprey, Related to Figure 4.

	Gene ID	Protein Length(aa)	Gene Name
Coelacanth	ENSLACT00000004839	399	P2RX3
	ENSLACT00000021101	403	P2RX1
	ENSLACT00000019925	388	P2RX2
	ENSLACT00000025971	134	P2RX7
	ENSLACT00000001552	363	P2RX5
	ENSLACT00000002031	419	P2RX6
	ENSLACT00000024921	391	P2RX4
	ENSDART00000167907	597	p2rx7
Zebrafish	ENSDART00000002866	398	p2rx1
	ENSDART00000098969	401	p2rx4b
	ENSDART00000019461	400	p2rx2
	ENSDART00000130628	398	p2rx4a
	ENSDART00000171919	482	p2rx5
	ENSDART00000163914	411	p2rx3a
	ENSDART00000133926	412	p2rx3b
	ENSDART00000029063	391	p2rx5
	ENSORLT00000014760	397	p2rx1
	ENSORLT00000004381	405	p2rx3
Medaka	ENSORLT00000019713	392	p2rx4
	ENSORLT00000015083	458	p2rx5
	ENSORLT00000009830	412	p2rx2
	PMZ_0041355-RA_PMZ	79	p2rx5.L
Sea lamprey	PMZ_0041356-RA_PMZ	205	P2RX5
	PMZ_0048630-RA_PMZ	169	p2rx5.S
	PMZ_0048646-RA_PMZ	160	p2rx5.S
	PMZ_0037740-RA_PMZ	225	P2RX5
	PMZ_0034936-RA_PMZ	160	p2rx5.S
	PMZ_0048622-RA_PMZ	160	p2rx5.S
	PMZ_0046842-RA_PMZ	74	p2rx5
	PMZ_0036668-RA_PMZ	99	p2rx5.S

PMZ_0035061-RA_PMZ	246	P2RX5
PMZ_0048786-RA_PMZ	97	p2rx5.S
PMZ_0037023-RA_PMZ	98	p2rx5.S
PMZ_0036851-RA_PMZ	244	P2RX5
PMZ_0048482-RA_PMZ	154	P2RX5

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J Mol Biol* 215, 403-410.

Andrews, S., and FastQC, A. (2015). A quality control tool for high throughput sequence data. 2010. Google Scholar.

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic acids research* 27, 573.

Birney, E., Clamp, M., and Durbin, R. (2004). GeneWise and genomewise. *Genome research* 14, 988-995.

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

Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S., Huntley, M.H., Lander, E.S., and Aiden, E.L. (2016). Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* 3, 95-98.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* 32, 1792-1797.

Elsik, C.G., Mackey, A.J., Reese, J.T., Milshina, N.V., Roos, D.S., and Weinstock, G.M. (2007). Creating a honey bee consensus gene set. *Genome biology* 8, 1.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., *et al.* (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* 8, 1494-1512.

Jurka, J., Kapitonov, V.V., Pavlicek, A., Klonowski, P., Kohany, O., and Walichiewicz, J. (2005). Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and genome research* 110, 462-467.

Kajitani, R., Toshimoto, K., Noguchi, H., Toyoda, A., Ogura, Y., Okuno, M., Yabana, M., Harada, M., Nagayasu, E., Maruyama, H., *et al.* (2014). Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Res* 24, 1384-1395.

Kent, W.J. (2002). BLAT--the BLAST-like alignment tool. *Genome Res* 12, 656-664.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27, 2987-2993.

Li, H., Coghlan, A., Ruan, J., Coin, L.J., Heriche, J.K., Osmotherly, L., Li, R., Liu, T., Zhang, Z., Bolund, L., *et al.* (2006). TreeFam: a curated database of phylogenetic trees of animal gene families. *Nucleic Acids Res* 34, D572-580.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25, 1754-1760.

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., and Liu, Y. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* 1, 18.

MA, Q., WANG, S.-f., WANG, J., and SU, Y.-q.J.J.o.X.U. (2008). Analysis of the Karyotype of *Chiloscyllium plagiosum* [J]. 6.

Mohr, D.W., Naguib, A., Weisenfeld, N., Kumar, V., Shah, P., Church, D.M., Jaffe, D., and Scott, A.F. (2017). Improved de novo Genome Assembly: Synthetic long read sequencing combined with optical mapping produce a high quality mammalian genome at relatively low cost. *bioRxiv*, 128348.

Pertea, M., Kim, D., Pertea, G.M., Leek, J.T., and Salzberg, S.L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc* 11, 1650-1667.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2—approximately maximum-likelihood trees for large alignments. *PloS one* 5, e9490.

Salse, J., Abrouk, M., Murat, F., Quraishi, U.M., and Feuillet, C. (2009). Improved criteria and comparative genomics tool provide new insights into grass paleogenomics. *Briefings in bioinformatics* 10, 619-630.

Servant, N., Varoquaux, N., Lajoie, B.R., Viara, E., Chen, C.J., Vert, J.P., Heard, E., Dekker, J., and Barillot, E. (2015). HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol* 16, 259.

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. (2006). AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic acids research* 34, W435-W439.

Tang, H., Wang, X., Bowers, J.E., Ming, R., Alam, M., and Paterson, A.H. (2008). Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. *Genome research*, gr. 080978.080108.

Tarailo-Graovac, M., and Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics*, 4.10. 11-14.10. 14.

Wang, O., Chin, R., Cheng, X., Wu, M.K.Y., Mao, Q., Tang, J., Sun, Y., Anderson, E., Lam, H.K., and Chen, D. (2019). Efficient and unique cobarcoding of second-generation sequencing reads from long DNA molecules enabling cost-effective and accurate sequencing, haplotyping, and de novo assembly. *Genome research* 29, 798-808.

Xu, Z., and Wang, H. (2007). LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic acids research* 35, W265-W268.

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24, 1586-1591.

Yang, Z., and Rannala, B. (2006). Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Mol Biol Evol* 23, 212-226.