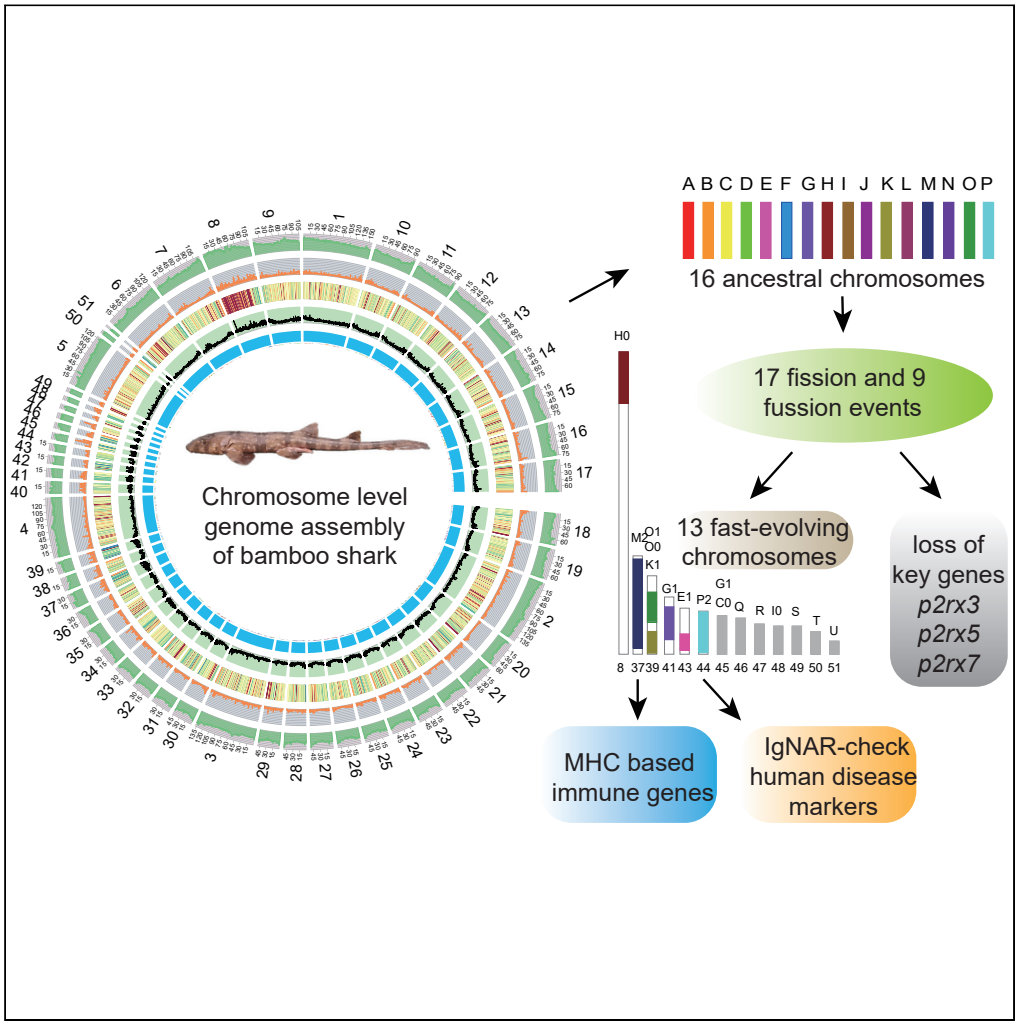
iScience

CellPress
OPEN ACCESS

Article

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




Yaolei Zhang, Haoyang Gao, Hanbo Li, ..., Guangyi Fan, Naibo Yang, Xin Liu

fanguangyi@genomics.cn (G.F.)
yangnaibo@genomics.cn (N.Y.)
liuxin@genomics.cn (X.L.)

## HIGHLIGHTS

Inferred ancestral chromosome karyotypes of cartilaginous fish

Chromosome rearrangements resulted in fast-evolving chromosomes and immune genes

Chromosome rearrangements led to deletion of bone formation-related genes

Proved that single-domain antibodies in shark have great potential application

Zhang et al., iScience *23*, 101754
November 20, 2020 © 2020
https://doi.org/10.1016/j.isci.2020.101754

iScience

CellPress
OPEN ACCESS

Article

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

Yaolei Zhang,[1,4,8,10] Haoyang Gao,[1,4,10] Hanbo Li,[1,4,10] Jiao Guo,[1,4,10] 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 Wei,[5] 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 Shi,[5] Shiping Liu,[2,4] Jamal S.M. Sabir,[9] Mumdooh J. Sabir,[9] Muhummadh Khan,[9] Nahid H. Hajrah,[9] Simon Ming-Yuen Lee,[3] Xun Xu,[2,4] Huanming Yang,[2,7] Jian Wang,[2,7] Guangyi Fan,[1,3,4,*] Naibo Yang,[2,6,*] and Xin Liu[1,2,4,11,*]

## SUMMARY

**Chondrichthyan (cartilaginous fish) occupies a key phylogenetic position and is important for investigating evolutionary processes of vertebrates. However, limited whole genomes impede our in-depth knowledge of important issues such as chromosome evolution and immunity. Here, we report the chromosome-level genome of white-spotted bamboo shark. Combing it with other shark genomes, we reconstructed 16 ancestral chromosomes of bamboo shark and illustrate a dynamic chromosome rearrangement process. We found that genes on 13 fast-evolving chromosomes can be enriched in immune-related pathways. And two chromosomes contain important genes that can be used to develop single-chain antibodies, which were shown to have high affinity to human disease markers by using enzyme-linked immunosorbent assay. We also found three bone formation-related genes were lost due to chromosome rearrangements. Our study highlights the importance of chromosome rearrangements, providing resources for understanding of cartilaginous fish diversification and potential application of single-chain antibodies.**

## INTRODUCTION

The white-spotted bamboo shark, *Chiloscyllium plagiosum*, (hereinafter referred to as bamboo shark) belongs to the class of Chondrichthyes, which is one of the oldest extant jawed vertebrate groups (McKenna, 1988). Cartilaginous fishes including Elasmobranchii and Holocephali shared a common ancestor with other vertebrates about 460–520 Ma, and then evolved independently to distinct lineages (Inoue et al., 2010). The phylogenetic evolution of cartilaginous fishes has been disputed for a long time (Cao et al., 1998; Janvier, 1996; Rasmussen and Arnason, 1999; Zardoya et al., 1998), especially the evolutionary relationships with bony fishes. Furthermore, most cartilaginous fishes have various chromosome karyotypes (2n = 66–104) (Rocco et al., 2003; Schwartz and Maddock, 1986), revealing interesting chromosome evolution processes. As known, immunoglobulins or lymphocyte receptors-based adaptive immunity is restricted to vertebrates (Litman et al., 2010). And as one of the extant early branching jawed vertebrates, cartilaginous fishes developed special immunity (for example, sharks comprise heavy-chain immunoglobulins, which are different from canonical antibodies consisting of both heavy and light chains, Könning et al., 2017), which makes them "immunologist's delight."

The bamboo shark is a nocturnal reef-dwelling species and widely distributed in the Indo-West Pacific from India to Indonesia, southern China, and Japan (Kyne and Burgess, 2006). Its biological features, including docile nature, small body size (24–37 inches in length), convenient reproductivity, and longevity make it ideal for research. Its special immunoglobulins are superior in biological and medical applications and have been proposed for developing antibody drugs (Wesolowski et al., 2009; Zielonka et al., 2014). Despite its biological and application importance, previous researchers mostly focused on addressing its biology of

[1]BGI-Qingdao, BGI-Shenzhen, Qingdao 266555, China

[2]BGI-Shenzhen, Shenzhen 518083, China

[3]State Key Laboratory of Quality Research in Chinese Medicine and Institute of Chinese Medical Sciences, Macao, China

[4]China National GeneBank, BGI-Shenzhen, Shenzhen 518120, China

[5]City University of Hongkong, Kowloon, Hongkong SAR

[6]Complete Genomics, Inc., San Jose, CA 95134, USA

[7]James D. Watson Institute of Genome Sciences, Hangzhou 310058, China

[8]Department of Biotechnology and Biomedicine, Technical University of Denmark, 2800 Lyngby, Denmark

[9]Department of Biological Sciences, King Abdulaziz University (KAU), Jeddah 21589, Saudi Arabia

[10]These authors contributed equally

[11]Lead Contact

*Correspondence: fanguangyi@genomics.cn (G.F.), yangnaibo@genomics.cn (N.Y.), liuxin@genomics.cn (X.L.)

https://doi.org/10.1016/j.isci.2020.101754

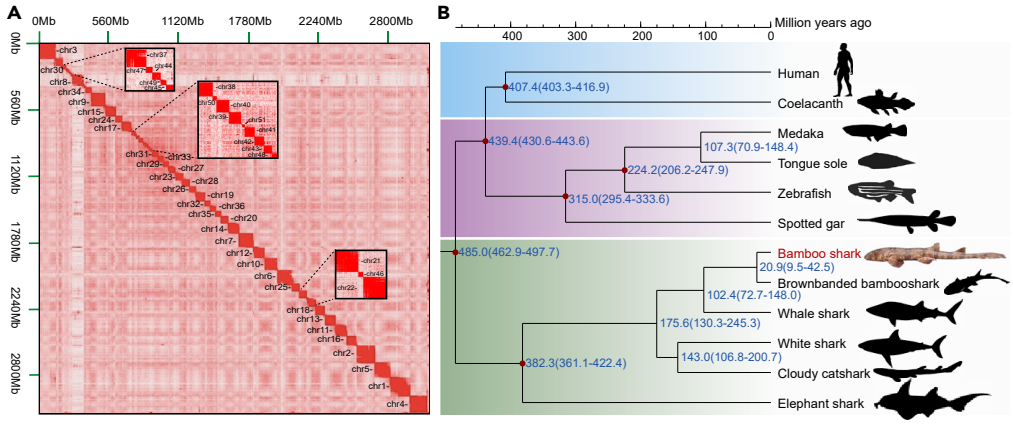



**Figure 1. Genome Assembly and Phylogeny of Bamboo Shark**

(A) Heatmap of chromatin interaction relationships at 125 kb resolution of 51 chromosomes.

(B) Phylogenetic tree and divergence time estimation of cartilaginous fish and representative bony fishes. Red dots mean reference times from TimeTree database (http://www.timetree.org/).

hematology, reproduction, muscle activity, liver regeneration, and anatomy (Alexander et al., 2016; Maia and Wilga, 2013; Straube et al., 2016). Limited whole-genome sequencing of cartilaginous fishes including elephant shark (2014) (Venkatesh et al., 2014), whale shark (2017) (Read et al., 2017), brownbanded bamboo shark (2018) (Hara et al., 2018), cloudy catshark (2018) (Hara et al., 2018), and white shark (2019) (Marra et al., 2019) prevents our further understanding of genetic mechanisms for these species. Furthermore, these five genomes were all assembled to scaffold level, which also limits our investigation of chromosome rearrangements.

To better understand the evolution and special immunity of cartilaginous fishes, we sequenced and assembled a chromosome-level genome of a female bamboo shark, identifying dynamic chromosome rearrangement events and related evolutionary consequences. We carefully analyzed evolution of immune-related genes, which will provide new resources to understand their biology and applications in immunology. We also found that chromosome rearrangements delete three important bone formation-related genes, which may interpret chondrified endoskeleton of cartilaginous fishes.

## RESULTS

### Genome Assembly and Annotation

We assembled a 3.85-Gb genome assembly with 51 chromosomes supported by chromatin interaction relationships with Hi-C sequencing data (Figure 1A and Tables S1 and S2) and karyotype analysis (Ma et al., 2008), and we annotated 19,595 protein coding genes (Table S3) and 63.53% of repeat content (Table S4) in this genome assembly. Comparison of repeat content among cartilaginous fishes and bony fishes shows that cartilaginous fish genomes contain higher proportion of repeated sequences (Table S5). The GC content and repeat and gene density distributed in 51 chromosomes are shown in Figures S1 and S2. And ~95.8% of the Benchmarking Universal Single-Copy Orthologs (BUSCOs) (Simao et al., 2015) were identified to be complete in this genome (Table S6). Syntenic relationships revealed unambiguous alignments of 41 bamboo shark chromosomes to 29 chromosomes of chicken (the tetrapod species with most stable karyotypes, Ellegren, 2010, Figure S3), whereas the alignments between bamboo shark and zebrafish chromosomes (Figure S4) are disordered. Extensive inter-chromosomal rearrangements have been described previously in zebrafish genome (Kasahara et al., 2007). Therefore, these results suggest the bamboo shark genome also kept relatively conserved chromosome karyotypes without many inter-chromosomal rearrangements.

The assembled genome size of bamboo shark is ~3.85 Gb, larger than elephant shark (~974 Mb) (Venkatesh et al., 2014), whale shark (~2.93 Gb) (Read et al., 2017), and most bony fishes (340 Mb–2.97 Gb). Whole-genome duplication (WGD) is one of reasons that result in larger genome sizes (Grover and Wendel, 2010). Thus we investigate whether there is a third WGD event in bamboo shark apart from the common two rounds of WGD of vertebrates (Grover and Wendel, 2010). First, we found only one peak on the 4-

fold synonymous third-codon transversion rates (4DTv) distribution of bamboo shark genome, which represented the recent common WGD event of all vertebrates (Figure S5). We also checked the 4DTv distribution of elephant shark and zebrafish, finding that elephant shark also has only one peak, which is similar to bamboo shark, but zebrafish has another peak representing the third WGD of teleost fish (Glasauer and Neuhauss, 2014), indicating the reliability of our results. Second, we checked HOX (homeobox) genes, which are highly conserved in vertebrates and always clustered together (Santini et al., 2003) and thus have become reliable markers of WGD events (Kuraku, 2011). We only identified three HOX clusters in bamboo shark, compared with seven clusters in zebrafish, which experienced a third WGD. Similar HOX clusters were also found in cattle, whale shark and elephant shark genomes (Figure S6). These two results suggest that bamboo shark genome did not experience a third WGD event. Therefore, the larger genome size of bamboo shark should result from the burst of repeated sequences.

## Molecular Phylogeny

In consideration of the important evolutionary position of cartilaginous fishes, we inferred phylogenetic relationship of bamboo shark with other five cartilaginous fishes, five representative bony fishes (four ray-finned fishes and one lobefin fish), and humans. Based on coding sequences of 823 single-copy orthologous genes identified using TreeFam (Li et al., 2006), we constructed phylogenetic trees using both maximum-likelihood and Bayesian methods and generated identical results (Figure 1B). The tree topology for cartilaginous and bony fishes is consistent with previous researches (Hara et al., 2018; Marra et al., 2019). And we estimated divergence times of cartilaginous and bony fishes, Elasmobranchii and Holocephali; white spotted bamboo shark; and brownbanded bamboo shark to be about 485.0, 382.3, and 20.9 Ma, respectively.

## Reconstruct Ancestral Chromosome

This chromosome-level genome makes it possible to study chromosome evolution of cartilaginous fishes. Thus we reconstructed ancestral chromosome karyotypes of cartilaginous fishes by identifying paralogous and orthologous genes between the bamboo shark and elephant shark genomes (Venkatesh et al., 2014) following a previously described method (Salse et al., 2009) (Table S7). Finally, we constructed 16 putative ancestral chromosomes and illustrated an evolutionary scenario during which eight fission and five fusion events occurred (Figure 2A, colored arrows), possibly for all cartilaginous fishes. As for the bamboo shark, nine fission and four fusion events (black and dotted arrows) occurred, resulting in six candidate daughter chromosomes (Chr8, Chr29, Chr38 and Chr39, Chr45, Chr48) (Figures 2A and S7 and Table S7). All these rearrangements ultimately gave rise to 51 chromosomes of the bamboo shark genome.

## Fast-Evolving Chromosomes and Immune Genes

Plenty of chromosome rearrangements play a role in fast-evolving gene families and in fostering large-scale changes in gene order (Eichler and Sankoff, 2003). To identify potential causes and consequences of dynamic chromosome rearrangements in cartilaginous fishes, we further analyzed distribution of conserved protein-coding genes of cartilaginous fishes along bamboo shark chromosomes. We identified 2,323 orthologous genes (~12.90% of total genes) shared among bamboo shark, elephant shark, whale shark, brown-banded bamboo shark, cloudy catshark, and white shark (Figure S8). After exclusion of genes shared among these six cartilaginous fishes and representative bony fishes (medaka, Kasahara et al., 2007, Figure S9, and spotted gar Braasch et al. 2015, Figure S10), we finally identified 1,359 genes conserved only in cartilaginous fishes (Figure 2B). Interestingly, we found those genes to be unevenly distributed along bamboo shark chromosomes with conserved genes on chromosomes 8, 37, 39, 41, 43, 44, 45, 46, 47, 48, 49, 50, and 51—notably fewer than (average: 2.6 genes) those of other chromosomes (average: 34.9 genes, (Mann-Whitney U test, p value < 0.001) (Figures 2B and Table S8). We then evaluated the evolutionary rate by calculating $K_S$ (synonymous substitutions per synonymous site) values of orthologous genes on these 13 chromosomes (mean $K_S$ value: 2.79), which was significantly higher than that of other chromosomes (mean $K_S$ value: 1.54, Mann-Whitney U test, p value < 0.001, Figure 2C). In addition, we found heterozygous SNPs in the genome of this individual we sequenced to be notably more frequent on these 13 chromosomes (except Chr43) than other chromosomes (Mann-Whitney U test, p value < 0.001, Figure 2B). All these findings suggest that these 13 chromosomes are fast-evolving. Enrichment analysis (according to the Kyoto Encyclopedia of Genes and Genomes [KEGG]-assigned gene functions and pathways) showed that genes on these 13 fast-evolving chromosomes are significantly enriched in immune-related pathways with 171 immune-related genes (p value < 0.01, Tables S9 and S10). These include allograft rejection, antigen processing and presentation, as well as intestinal immune network for IgA production.

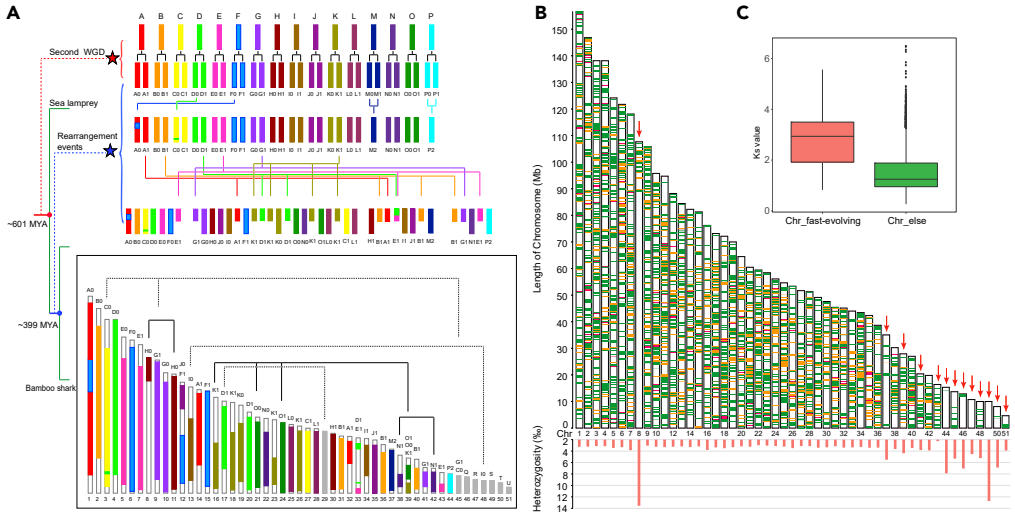



**Figure 2. Chromosome Evolution of Bamboo Shark**

(A) Construction of the ancestral chromosome model of elephant shark and bamboo shark. The lines on the left represent the phylogenetic tree. The dotted lines on the left represent the second WGD event for vertebrates and rearrangement events of ancestral chromosomes of sharks. Letters from A to P represent constructed ancestral chromosomes. Letters Q, R, S, T, and U represent small chromosomes that may be the result of more recent fissions from the larger chromosomes. The colored arrows represent rearrangements of chromosomes. The black arrows represent rearrangements specific to the bamboo shark. The dotted arrows represent potential rearrangements that are supported by gene pairs. The red star represents the second whole-genome duplication occurred before the divergence of sea lamprey and gnathostome (Smith et al., 2013) (earlier than ~601 Ma). The blue star represents rearrangement events of shark ancestral chromosomes. The times are cited from the TimeTree database.

(B) (Top): Distribution of conserved genes of elephant shark, bamboo shark, whale shark, brownbanded bamboo shark, cloudy catshark, white shark and medaka. Magenta, conserved regions shared by four species (R1). Orange, common regions in six shark genomes excluding R1. Green, regions shared between the bamboo shark and medaka excluding R1. The red arrows point out 13 fast-evolving chromosomes. (Bottom) Heterozygosity of 51 chromosomes in bamboo shark genome.

(C) Comparison of *Ks* values of single-copy orthologous genes between 13 fast-evolving chromosomes and other chromosomes.

## Analysis of MHC-Related Genes

Among these 13 fast-evolving chromosomes, we found that Chr37 and Chr44 likely underwent a special self-fusion event after a possible chromosome or large segment duplication event (Figure 2A). We also found that major histocompatibility complex (MHC) genes (11 class I and 3 class II genes) are notably enriched on Chr37 (11 genes, Figure 3A), except for those on unanchored scaffolds. MHC genes were not found in the amphioxus genome, whereas one fragment of a possible MHC class II gene was found in sea lamprey (Gene ID: PMZ_0007681-RA; KEGG function: K06752 MHC, class II) (Smith et al., 2018). Upon further investigation of MHC gene numbers in other species, we found both MHC class I and class II genes in cartilaginous fishes and bony fishes except for the elephant shark genome, which lacked MHC class II genes according to our analysis (Tables S11, S12, and S13 and Figure S11). These results suggest that the innate immune system accompanied with adaptive system based on variable lymphocyte receptors (Pancer et al., 2004; Saha et al., 2010) played a major role in defending against infections in amphioxus and sea lamprey, whereas cartilaginous and bony fishes evolved with acquiring a complete MHC-based adaptive immune system. The differences in these immune systems may have arisen from the fast-evolving chromosomes. Moreover, we suggest that MHC class II genes were likely acquired before MHC class I genes based on our identification of an MHC class II-like fragment in the starlet sea anemone genome (NCBI Accession: XP_001628845.1, identified by aligning PMZ_0007681-RA using BLAST, Altschul et al., 1990) and sea lamprey genome (Smith et al., 2018), potentially resolving a long debate about MHC evolution (Flajnik et al., 1991; Kaufman, 1988, 2011, 2018; Kaufman et al., 1984; Rock et al., 2016; Zhang et al., 2014). In addition, we found that tripartite motif-containing protein 69 (*TRIM69*) gene family was expanded significantly in cartilaginous fishes (average 18 copies) compared with ray-finned fishes (less than 3 copies) (Figure S12). Also, in bamboo shark genome, 13 copies of *TRIM69* were also located on

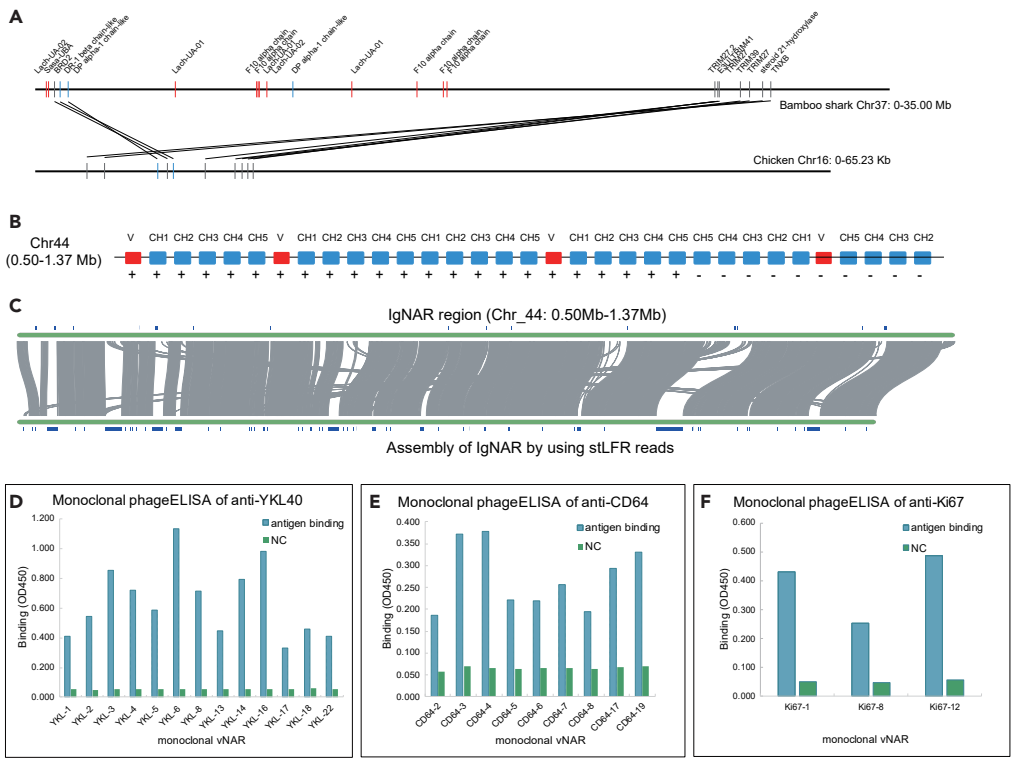



**Figure 3. Specific Immune-Related Genes of Bamboo Shark**

(A) Distribution of MHC genes on chromosome 37. The red and blue rectangles represent MHC class I and class II genes, respectively. The gray rectangles represent non-MHC genes. Here, we only show syntenic genes and shark MHC genes compared with chicken.

(B) Distribution of identified IgNAR loci on chromosome 44.

(C) Syntenic relationship of assemblies of IgNAR region by using shotgun WGS and stLFR data. The gray lines represent consistent region. The blue rectangles represent gaps.

(D, E, and F) Binding efficiency of cloned vNARs to human disease markers (D) YKL40, (E) CD64, and (F) Ki67 by using monoclonal phage ELISA. NC represents negative control. The vertical axis represents the signal of binding measured using optical density (OD450).

Chr37. *TRIM69* can function as an antiviral defense (Rihn et al., 2019; Wang et al., 2018), playing important roles in innate immune system and class I MHC-mediated antigen processing and presentation.

## Identification and Amplification of IgNAR

In contrast to MHC genes found on Chr37, we identified that the immunoglobulin new antigen receptor (IgNAR) (Feige et al., 2014) gene loci (four complete IgNAR structure, V-CH1-CH2-CH3-CH4-CH5, and two incomplete IgNAR) was located on Chr44 (Figure 3B). To obtain and verify the complete sequence assembly of IgNAR region, we sequenced ~124-fold new single-tube long fragment reads (stLFR) (Wang et al., 2019) to re-assemble it. The good syntenic relationship (Figure 3C) and the distribution of paired-end reads (Figure S13) reveal high-quality assembly of IgNAR region. Because of the application potential of single-domain antibodies (sdAbs) in biotechnical and therapeutic use, we tried to check the diagnostic potential of IgNAR in bamboo shark. We first designed primers (Table S14) based on IgNAR sequences of bamboo shark and specifically amplified the variable domain of New Antigen Receptors (vNARs) from peripheral blood leukocytes and spleen tissue of five bamboo shark individuals. Amplified vNARs were inserted into phagemid vector pMECS and then into *E. coli* TG1 competent cells to produce the vNAR-phage display library. Sanger sequencing of randomly selected ~100 clones shows low repetition and high diversity of vNARs, especially the complementarity-determining region 3 (CDR3) (Figure S14). We then chose several human disease biomarkers: YKL40 (Kastrup, 2012; Rathcke and Vestergaard, 2009) for cardiovascular disease, CD64 for infectious disease (Hoffmann, 2009), and Ki67 for lung cancer (Li et al., 2015) as targets of four rounds screened monoclonal vNARs. Monoclonal phage enzyme-linked immunosorbent assay

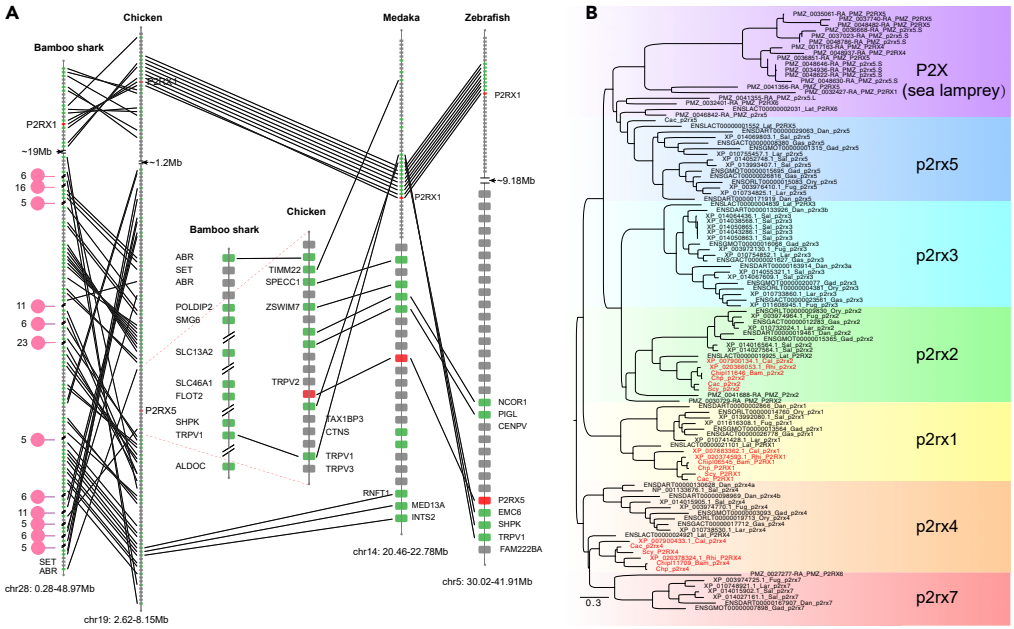



**Figure 4. The Loss of *p2rx3*, *p2rx5*, and *p2rx7* Genes in Cartilaginous Fishes**

(A) The syntenic relationship of the *p2rx5* region among bamboo shark, chicken, medaka, and zebrafish. The red rectangles represent P2X genes. The lines represent syntenic genes. The numbers besides the pink dots represent gene number in the gap regions. From the figure, we can see the gene loss of *p2rx5* in bamboo shark genome after at least four chromosome rearrangements.

(B) Phylogenetic tree of the P2X gene family. Dan: zebrafish, GAD: Atlantic cod, Gas: three-spined sticklebacks, Ory: medada, Lat: coelacanth, Sal: Atlantic salmon, Fug: torafugu, Lar: large yellow croaker, Rhi: whale shark, Cal: elephant shark, Bam: bamboo shark, Chp: brownbanded bamboo shark, Cac: white shark, Scy: cloudy catshark, PMZ: sea lamprey. The shark's P2X genes are marked red. Although many p2rx5 genes exist in PMZ, all these genes were less than 246 aa in length (10 of 14 less than 200 aa). 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).

(ELISA) shows high affinity to targets of those positive clones. In detail, we obtained 13 unique YKL40-binding clones with signal at least ~6-fold (highest: ~20-fold) than negative control (NC) (Figures 3D), 9 unique CD64-binding clones with signal at least ~3-fold than NC (Figure 3E), and 3 unique Ki67-binding clones with signal at least ~5-fold than NC (Figure 3F). Although further study should be carried out to verify the affinity of those positive clones and their diagnostic function by using real samples, we believe that this work is of significance for using genome data to develop sdAbs.

## P2X Gene Family

Chromosome rearrangements would remove genes that may first become pseudogenes because selective pressure acting on them was relaxed when new phenotypic traits arose or they may have very little effect on its adaptations. Syntenic comparison among chicken, zebrafish, medaka, and bamboo shark showed at least four possible genome rearrangement events occurred in bamboo shark that may result in the deletion of gene, *p2rx5* (Purinergic Receptor P2X, Ligand-Gated Ion Channel, 5) after this gene became redundant or non-functioning due to selective pressures acting on it (Figure 4A). The loss of this gene was also supported by checking RNA sequence data of 14 tissues including blood, eye, gill, heart, liver, muscle, spleen, stomach, dorsal fin, tail fin, pancreas, leptospira, two capsulogenous gland, and two kidney samples. And this gene has been previously reported to be involved in bone development and homeostasis (Nicolaidou et al., 2012; Sitcheran et al., 2003; Solle et al., 2001; Sun et al., 2013; Syberg et al., 2012; Takahashi et al., 1988; Thaler et al., 2014). Furthermore, analysis of the whole gene family of purinergic receptor P2X in sea lamprey, six sharks, and representative bony fishes showed that *p2rx3*, *p2rx5*, and *p2rx7* genes were lost in six cartilaginous fishes, whereas at least six paralogs (*p2rx1*, *p2rx2*, *p2rx3*, *p2rx4*, *p2rx5*, *p2rx7*) with multiple copies were found in bony fishes (Figure 4B and Tables S15 and S16). P2X receptors contain ligand-gated ion channels and activate receptor triggers signaling pathways associated with $Ca^{2+}$ influx

(Burnstock, 2012; Jing et al., 2014; Rodrigues-Ribeiro et al., 2015). Moreover, *p2rx3*, *p2rx5*, and *p2rx7* receptors have been shown to participate in differentiation and proliferation of osteoblast (Nakamura et al., 2000; Nicolaidou et al., 2012; Rodrigues-Ribeiro et al., 2015), bone formation, and resorption (Grol et al., 2009; Kim et al., 2018; Syberg et al., 2012). However, *p2rx1* receptor negatively regulates bone mineralization (Lenertz et al., 2015). *P2rx2* receptor, which mainly functions in sensory neurons, neuromuscular junction formation, and hearing (Yan et al., 2013), has nothing to do with bone formation. *P2rx4* also functions in response to ATP binding, and there are few researches that show its role in bone formation to date. Thus, it is reasonable to infer that loss of those genes, together with loss of *spp1* gene identified previously (Venkatesh et al., 2014), may further explain the establishment of chondrification of the endoskeleton in cartilaginous fishes.

## DISCUSSION

Because of the therapeutic potential of single-domain antibodies, sharks have drawn scientists' interest for many years. With ideal biological features of bamboo shark, we selected this species and were able to obtain high-quality samples for further research. Combining paired-end, mate-paired, stLFR, and Hi-C sequencing strategies, we successfully assembled a chromosome-level reference genome of bamboo shark. In the present study, we mainly focused chromosome evolution and t fast-evolving chromosomes and immune gene families of cartilaginous fishes. Also, we investigated that burst of repeat that caused larger genome size of bamboo shark and inferred phylogenetic topologies between sequenced cartilaginous fishes and bony fishes, which is important for exploring evolutionary process of vertebrates.

Using this genome, we inferred ancestral chromosomes of cartilaginous fishes and found dynamical rearrangements. Based on chromosome evolutionary processes and comparative genomic analysis, we were able to identify fast-evolving chromosomes and immune-related genes. Moreover, chromosome fusions and fissions would also cause DNA damages, deletion of genes, and formation of new genes that may be functionally important and closely associated to species-specific features. Based on this, we found gene loss events associated with phenotypic diversity, for example, chondrification of the endoskeleton. Thus, our study highlights the importance of chromosome rearrangements in the diversification of cartilaginous fishes. With effective methods described in this study, more chromosome-level genomes can be obtained in the future to further elucidate the early evolution of jawed vertebrates as well as extant jawed vertebrate lineages.

Shark-specific immunocompetence always attracts researchers. Investigating immune genes could help to understand evolutionary processes of immune system of cartilaginous fishes compared with jawless species. MHC-I-like and MHC-II-like genes found in cartilaginous fishes revealed the possible time of acquirement of MHC-based adaptive immunity. Bamboo shark chromosomes (Chr37 and Chr44) enriched with immune-related gene may play vital roles in its powerful immunity, and more chromosome-level genomes of cartilaginous fishes should be accomplished to further confirm this conclusion. Shark single-domain antibodies have shown prospects in therapeutic use and our ELISA experiments also proved their potential use in human diseases. Thus, assembly of IgNAR sequences will accelerate development of antibodies for future medicine. In summary, our results provide valuable resources and will be significant for future research about vertebrate evolution and pharmaceutical development.

### Limitations of the Study

In this study, we assembled bamboo shark genome, analyzed chromosome rearrangements, and performed ELISA experiments. However, this genome is the only chromosome-level cartilaginous fish genome and more high-quality genomes should be assembled to further verify our chromosome evolution analysis. Besides, more functional experiments should be performed to further validate candidate functional genes.

### Resource Availability

#### *Lead Contact*

Further information and requests for materials should be directed to and will be fulfilled by the lead contact, Xin Liu (liuxin@genomics.cn).

#### *Materials Availability*

This study did not generate new unique reagents.

### *Data and Code Availability*

The accession numbers for the genome sequencing data, RNA sequencing data, and genome assembly reported in this paper NCBI: PRJNA478295.This Whole Genome Shotgun project has been deposited to National Center for Biotechnology Information (NCBI) under the accession: QPFF00000000 referring project: PRJNA478295. Raw RNA sequencing reads have also been uploaded to the SRA database under accession: SRP154403. The assembled genome can also be obtained from CNSA (CNGB Nucleotide Sequence Archive) by assembly ID: CNA0000025.

## METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

## SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2020.101754.

## ACKNOWLEDGMENTS

We would like thank Dr. Lynn Fink for revising these manuscripts. This work was supported by National Key R&D Program of China (2018YFD0900301), Shenzhen-Hongkong Collaboration Fund JCYJ20170412152916724 (20170331), State Key Laboratory of Agricultural Genomics (No. 2011DQ782025), and Strategic Priority Research Program of the Chinese Academy of Sciences (XDA19060403).

## AUTHOR CONTRIBUTIONS

X.L., N.Y., G.F., and X.X. designed and managed this project. M.W., C.L., H.X., L.W., H.R., Y.X., Q.X., and S.P. were responsible for collecting samples, library construction, sequencing, and co-drafting the manuscript. Y.Z., H.G., and J.G. worked on genome assembly, annotation, chromosome evolution, gene family analysis, transcriptome, and co-drafting the manuscript. J.W., M.L., X.G., Q.L., and Y.S. performed data processing, whole-genome duplication, Hox gene clusters, and repeat analysis. H.L., B.O., Y.G., B.R., X.D., and Y.Y. performed ELSA experiment. S.L., J.W., Y.J., J.S., S.L., L.M., J.S.M.S., M.J.S., M.K., N.H.H., H.Y., J.W., and S.M.-Y.L. helped to revise the manuscript. All authors took part in the interpretation of data.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 16, 2020
Revised: September 17, 2020
Accepted: October 28, 2020
Published: November 20, 2020

## REFERENCES

Alexander, A.B., Parkinson, L.A., Grant, K.R., Carlson, E., and Campbell, T.W. (2016). The hemic response of white-spotted bamboo sharks (Chiloscyllium plagiosum) with inflammatory disease. Z. Biol. *35*, 251–259.

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.

Braasch, I., Gehrke, A.R., Smith, J.J., Kawasaki, K., Manousaki, T., Pasquier, J., Amores, A., Desvignes, T., Batzel, P., and Catchen, J. (2015). The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. Nat. Genet. *47*, 427.

Burnstock, G. (2012). Purinergic signalling: its unpopular beginning, its acceptance and its exciting future. Bioessays *34*, 218–225.

Cao, Y., Janke, A., Waddell, P.J., Westerman, M., Takenaka, O., Murata, S., Okada, N., Pääbo, S., and Hasegawa, M. (1998). Conflict among individual mitochondrial proteins in resolving the phylogeny of eutherian orders. J. Mol. Evol. *47*, 307–322.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797.

Eichler, E.E., and Sankoff, D. (2003). Structural dynamics of eukaryotic chromosome evolution. science *301*, 793–797.

Ellegren, H. (2010). Evolutionary stasis: the stable chromosomes of birds. Trends Ecol. Evol. *25*, 283–291.

Feige, M.J., Gräwert, M.A., Marcinowski, M., Hennig, J., Behnke, J., Ausländer, D., Herold, E.M., Peschek, J., Castro, C.D., and Flajnik, M. (2014). The structural analysis of shark IgNAR antibodies reveals evolutionary principles of immunoglobulins. Proc. Natl. Acad. Sci. U S A *111*, 8155–8160.

Flajnik, M.F., Canel, C., Kramer, J., and Kasahara, M. (1991). Which came first, MHC class I or class II? Immunogenetics *33*, 295–300.

Glasauer, S.M., and Neuhauss, S.C. (2014). Whole-genome duplication in teleost fishes and its evolutionary consequences. Mol. Genet. genomics *289*, 1045–1060.

Grol, M.W., Panupinthu, N., Korcok, J., Sims, S.M., and Dixon, S.J. (2009). Expression, signaling, and function of P2X7 receptors in bone. Purinergic Signal. *5*, 205.

Grover, C.E., and Wendel, J.F. (2010). Recent insights into mechanisms of genome size change in plants. J. Bot. *2010*, 1–8.

Hara, Y., Yamaguchi, K., Onimaru, K., Kadota, M., Koyanagi, M., Keeley, S.D., Tatsumi, K., Tanaka, K., Motone, F., and Kageyama, Y. (2018). Shark genomes provide insights into elasmobranch evolution and the origin of vertebrates. Nat. Ecol. Evol. *2*, 1761.

Hoffmann, J.J. (2009). Neutrophil CD64: a diagnostic marker for infection and sepsis. Clin. Chem. Lab. Med. *47*, 903–916.

Inoue, J., Donoghue, P.C., and Yang, Z. (2010). The impact of the representation of fossil calibrations on Bayesian estimation of species divergence times. Syst. Biol. *59*, 74–89.

Janvier, P.. (1996). Early Vertebrates (Oxford University Press).

Jing, D., Baik, A.D., Lu, X.L., Zhou, B., Lai, X., Wang, L., Luo, E., and Guo, X.E. (2014). In situ intracellular calcium oscillations in osteocytes in intact mouse long bones under dynamic mechanical loading. FASEB J. *28*, 1582–1592.

Kasahara, M., Naruse, K., Sasaki, S., Nakatani, Y., Qu, W., Ahsan, B., Yamada, T., Nagayasu, Y., Doi, K., and Kasai, Y. (2007). The medaka draft genome and insights into vertebrate genome evolution. Nature *447*, 714.

Kastrup, J. (2012). Can YKL-40 be a new inflammatory biomarker in cardiovascular disease? Immunobiology *217*, 483–491.

Kaufman, J. (1988). Vertebrates and the evolution of the major histocompatibility complex (MHC) class I and class II molecules. Verh. Dtsch. Zool. Ges. *81*, 131–144.

Kaufman, J. (2011). The evolutionary origins of the adaptive immune system of jawed vertebrates. In The Immune Response to Infection, S.H.E. Kaufmann, B.T. Rouse, and D.L. Sacks, eds. (American Society of Microbiology), pp. 41–55.

Kaufman, J. (2018). Unfinished business: evolution of the MHC and the adaptive immune system of jawed vertebrates. Annu. Rev. Immunol. *36*, 383–409.

Kaufman, J.F., Auffray, C., Korman, A.J., Shackelford, D.A., and Strominger, J. (1984). The class II molecules of the human and murine major histocompatibility complex. Cell *36*, 1–13.

Kim, H., Kajikawa, T., Walsh, M.C., Takegahara, N., Jeong, Y.H., Hajishengallis, G., and Choi, Y. (2018). The purinergic receptor P2X5 contributes to bone loss in experimental periodontitis. BMB Rep. *51*, 468.

Kuraku, S. (2011). Hox gene clusters of early vertebrates: do they serve as reliable markers for genome evolution? Genomics, proteomics bioinformatics *9*, 97–103.

Kyne, P.M., and Burgess, G.H. (2006). Chiloscyllium plagiosum. . The IUCN Red List of Threatened Species, *2006*, eT60222A12325334.

Könning, D., Zielonka, S., Grzeschik, J., Empting, M., Valldorf, B., Krah, S., Schröter, C., Sellmann, C., Hock, B., and Kolmar, H. (2017). Camelid and shark single domain antibodies: structural features and therapeutic potential. Curr. Opin. Struct. Biol. *45*, 10–16.

Lenertz, L.Y., Baughman, C.J., Waldschmidt, N.V., Thaler, R., and van Wijnen, A.J. (2015). Control of bone development by P2X and P2Y receptors expressed in mesenchymal and hematopoietic cells. Gene *570*, 1–7.

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–D580.

Li, L.T., Jiang, G., Chen, Q., and Zheng, J.N. (2015). Ki67 is a promising molecular target in the diagnosis of cancer. Mol. Med. Rep. *11*, 1566–1572.

Litman, G.W., Rast, J.P., and Fugmann, S.D. (2010). The origins of vertebrate adaptive immunity. Nat. Rev. Immunol. *10*, 543–553.

Ma, Q., Wang, S.-f., Wang, J., and Su, Y.-q. (2008). Analysis of the karyotype of Chiloscyllium plagiosum [J]. J. Xiamen Univ. (Nat. Sci.) *6*.

Maia, A., and Wilga, C.D. (2013). Anatomy and muscle activity of the dorsal fins in bamboo sharks and spiny dogfish during turning maneuvers. J. Morphol. *274*, 1288–1298.

Marra, N.J., Stanhope, M.J., Jue, N.K., Wang, M., Sun, Q., Pavinski Bitar, P., Richards, V.P., Komissarov, A., Rayko, M., Kliver, S., et al. (2019). White shark genome reveals ancient elasmobranch adaptations associated with wound healing and the maintenance of genome stability. Proc. Natl. Acad. Sci. U S A *116*, 4446–4455.

McKenna, M.C. (1988). The vertebrates updated: vertebrate paleontology and evolution. Science *239*, 512–513.

Nakamura, E.I., Uezono, Y., Narusawa, K.I., Shibuya, I., Oishi, Y., Tanaka, M., Yanagihara, N., Nakamura, T., and Izumi, F. (2000). ATP activates DNA synthesis by acting on P2X receptors in human osteoblast-like MG-63 cells. Am. J. Physiol. Cell Physiol. *279*, C510–C519.

Nicolaidou, V., Wong, M.M., Redpath, A.N., Ersek, A., Baban, D.F., Williams, L.M., Cope, A.P., and Horwood, N.J. (2012). Monocytes induce STAT3 activation in human mesenchymal stem cells to promote osteoblast formation. PLoS One *7*, e39871.

Pancer, Z., Amemiya, C.T., Ehrhardt, G.R., Ceitlin, J., Gartland, G.L., and Cooper, M.D.J.N. (2004). Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey *430*, 174.

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

Rasmussen, A.-S., and Arnason, U. (1999). Molecular studies suggest that cartilaginous fishes have a terminal position in the piscine tree. Proc. Natl. Acad. Sci. U S A *96*, 2177–2182.

Rathcke, C.N., and Vestergaard, H. (2009). YKL-40-an emerging biomarker in cardiovascular disease and diabetes. Cardiovasc. diabetol. *8*, 61.

Read, T.D., Petit, R.A., Joseph, S.J., Alam, M.T., Weil, M.R., Ahmad, M., Bhimani, R., Vuong, J.S., Haase, C.P., and Webb, D.H. (2017). Draft sequencing and assembly of the genome of the world's largest fish, the whale shark: rhincodon typus Smith 1828. BMC genomics *18*, 532.

Rihn, S.J., Aziz, M.A., Stewart, D.G., Hughes, J., Turnbull, M.L., Varela, M., Sugrue, E., Herd, C.S., Stanifer, M., and Sinkins, S.P. (2019). TRIM69 inhibits vesicular stomatitis Indiana virus. J. Virol. *93*, e00951–00919.

Rocco, L., Costagliola, D., Liguori, I., and Stingo, V. (2003). Cytogenetic and molecular studies in the nurse shark, Ginglymostoma cirratum (Galeomorphii, Heterodontiformes). Annales de Genetique *46*, 98.

Rock, K.L., Reits, E., and Neefjes, J. (2016). Present yourself! by MHC class I and MHC class II molecules. Trends Immunol. *37*, 724–737.

Rodrigues-Ribeiro, R., Alvarenga, E.C., Calio, M.L., Paredes-Gamero, E.J., and Ferreira, A.T. (2015). Dual role of P2 receptors during osteoblast differentiation. Cell Biochem. Biophys. *71*, 1225–1233.

Saha, N.R., Smith, J., and Amemiya, C.T. (2010). Evolution of adaptive immune recognition in jawless vertebrates. In Seminars in Immunology (Elsevier), pp. 25–33.

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. Brief. Bioinformatics *10*, 619–630.

Santini, S., Boore, J.L., and Meyer, A. (2003). Evolutionary conservation of regulatory elements in vertebrate Hox gene clusters. Genome Res. *13*, 1111–1122.

Schwartz, F.J., and Maddock, M.B. (1986). Comparisons of karyotypes and cellular DNA contents within and between major lines of elasmobranchs (Indo-Pacific fish biology Ichthyological Society of Japan), pp. 148–157.

Simao, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics *31*, 3210–3212.

Sitcheran, R., Cogswell, P.C., and Baldwin, A.S. (2003). NF-κB mediates inhibition of mesenchymal cell differentiation through a posttranscriptional gene silencing mechanism. Genes Dev. *17*, 2368–2373.

Smith, J.J., Kuraku, S., Holt, C., Sauka-Spengler, T., Jiang, N., Campbell, M.S., Yandell, M.D., Manousaki, T., Meyer, A., and Bloom, O.E. (2013). Sequencing of the sea lamprey (Petromyzon marinus) genome provides insights into vertebrate evolution. Nat. Genet. *45*, 415.

Smith, J.J., Timoshevskaya, N., Ye, C., Holt, C., Keinath, M.C., Parker, H.J., Cook, M.E., Hess, J.E., Narum, S.R., and Lamanna, F. (2018). The sea lamprey germline genome provides insights into programmed genome rearrangement and vertebrate evolution. Nat. Genet. *50*, 270.

Solle, M., Labasi, J., Perregaux, D.G., Stam, E., Petrushova, N., Koller, B.H., Griffiths, R.J., and

Gabel, C.A. (2001). Altered cytokine production in mice lacking P2X7Receptors. J. Biol. Chem. *276*, 125–132.

Straube, N., Lampert, K.P., Geiger, M.F., Weiss, J.D., and Kirchhauser, J.X. (2016). First record of second-generation facultative parthenogenesis in a vertebrate species, the whitespotted bambooshark Chiloscyllium plagiosum. J. fish Biol. *88*, 668–675.

Sun, D., Junger, W.G., Yuan, C., Zhang, W., Bao, Y., Qin, D., Wang, C., Tan, L., Qi, B., and Zhu, D. (2013). Shockwaves induce osteogenic differentiation of human mesenchymal stem cells through ATP release and activation of P2X7 receptors. Stem Cells *31*, 1170–1180.

Syberg, S., Petersen, S., Beck Jensen, J.-E., Gartland, A., Teilmann, J., Chessell, I., Steinberg, T.H., Schwarz, P., and Jørgensen, N.R. (2012). Genetic background strongly influences the bone phenotype of P2X7 receptor knockout mice. J. Osteoporos. *2012*, 391097.

Takahashi, N., Akatsu, T., Udagawa, N., Sasaki, T., Yamaguchi, A., Moseley, J.M., Martin, T.J., and Suda, T. (1988). Osteoblastic cells are involved in osteoclast formation. Endocrinology *123*, 2600–2602.

Thaler, R., Sturmlechner, I., Spitzer, S., Riester, S.M., Rumpler, M., Zwerina, J., Klaushofer, K., Van Wijnen, A.J., and Varga, F. (2014). Acute-phase protein serum amyloid A3 is a novel paracrine coupling factor that controls bone homeostasis. FASEB J. *29*, 1344–1359.

Venkatesh, B., Lee, A.P., Ravi, V., Maurya, A.K., Lian, M.M., Swann, J.B., Ohta, Y., Flajnik, M.F., Sutoh, Y., and Kasahara, M. (2014). Elephant shark genome provides unique insights into gnathostome evolution. Nature *505*, 174.

Wang, K., Zou, C., Wang, X., Huang, C., Feng, T., Pan, W., Wu, Q., Wang, P., and Dai, J. (2018). Interferon-stimulated TRIM69 interrupts dengue virus replication by ubiquitinating viral nonstructural protein 3. PLoS Pathog. *14*, e1007287.

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 Res. *29*, 798–808.

Wesolowski, J., Alzogaray, V., Reyelt, J., Unger, M., Juarez, K., Urrutia, M., Cauerhff, A., Danquah, W., Rissiek, B., and Scheuplein, F. (2009). Single domain antibodies: promising experimental and therapeutic tools in infection and immunity. Med. Microbiol. Immunol. *198*, 157–174.

Yan, D., Zhu, Y., Walsh, T., Xie, D., Yuan, H., Sirmaci, A., Fujikawa, T., Wong, A.C.Y., Loh, T.L., and Du, L. (2013). Mutation of the ATP-gated P2X2 receptor leads to progressive hearing loss and increased susceptibility to noise. Proc. Natl. Acad. Sci. U S A *110*, 2228–2233.

Zardoya, R., Cao, Y., Hasegawa, M., and Meyer, A. (1998). Searching for the closest living relative (s) of tetrapods through evolutionary analyses of mitochondrial and nuclear data. Mol. Biol. Evol. *15*, 506–517.

Zhang, P., Julia, I., Leu, J., Murphy, M.E., George, D.L., and Marmorstein, R. (2014). Crystal structure of the stress-inducible human heat shock protein 70 substrate-binding domain in complex with peptide substrate. PLoS One *9*, e103518.

Zielonka, S., Weber, N., Becker, S., Doerner, A., Christmann, A., Christmann, C., Uth, C., Fritz, J., Schäfer, E., and Steinmann, B. (2014). Shark attack: high affinity binding proteins derived from shark vNAR domains by stepwise in vitro affinity maturation. J. Biotechnol. *191*, 236–245.

**iScience, Volume *23***

## 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 Wei[5], 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 Shi[5], Shiping Liu[2,4], Jamal S. M. Sabir[9], Mumdooh J. Sabir[9], Muhummadh Khan[9], Nahid H. Hajrah[9], Simon Ming-Yuen Lee[3], 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

### 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 x 106) 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℃ 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 μl of ddH2O). The Hi-C library was created by shearing 20 ug 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℃ 3 min.; [98℃ 20 sec., 60℃ 15 sec., 72℃ 15 sec.] (8 cycles); 72℃ 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 lobefin 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 databsse (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 (immunoglobin 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℃. 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 $_{V}$NAR domain to amplify the $_{V}$NAR sequence from the cDNA product by PCR. All reactions were pooled on 1.5% TAE agaose 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℃ 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℃. 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 10µl of the phage into 90µl of LB to $10^6$ fold in a low-binding 96-well plate. $10^2$ and $10^3$ 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 µl 2TY containing 100 µg/ml ampicillin and 2% glucose for 3h at 37℃ and 180rpm. Sufficient volume of M13KO7 helper phage (1µl) were added into each well at 37℃ 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 2µg/ml in PBS buffer, 100µl/well, at 4℃ overnight. The antigen-coated wells were washed three times with 200µl of PBST, then 200µl 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 200µl 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

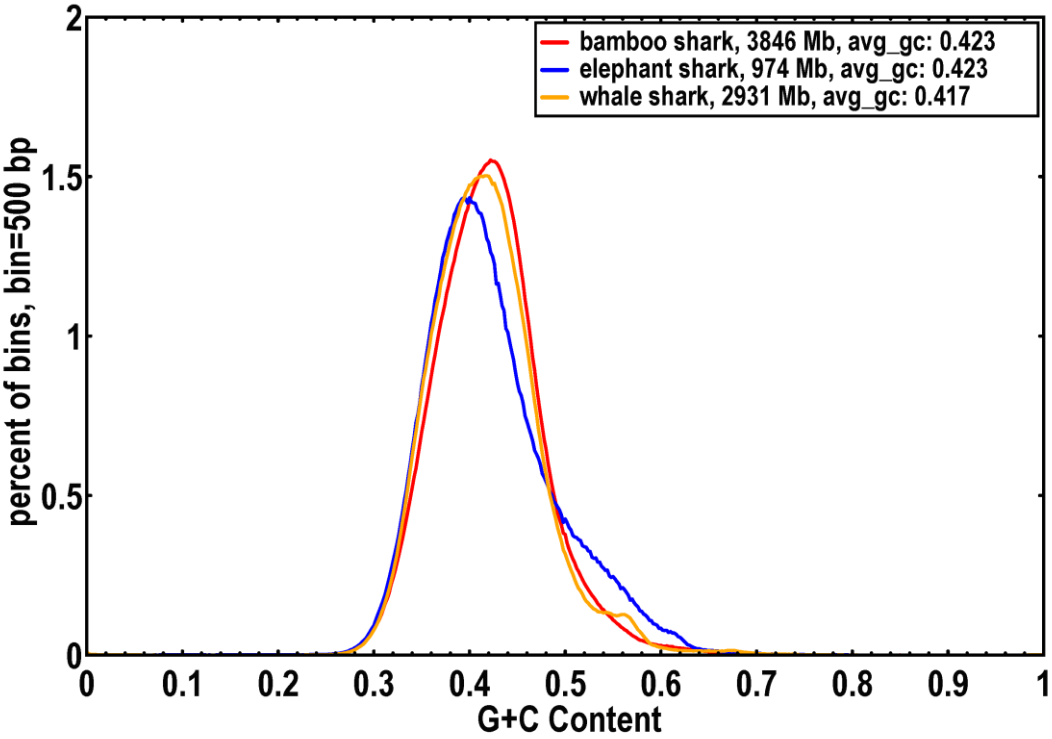



**Figure S1.** GC content of the assembled bamboo shark, elephant shark, and whale shark genomes, Related to Figure 1.

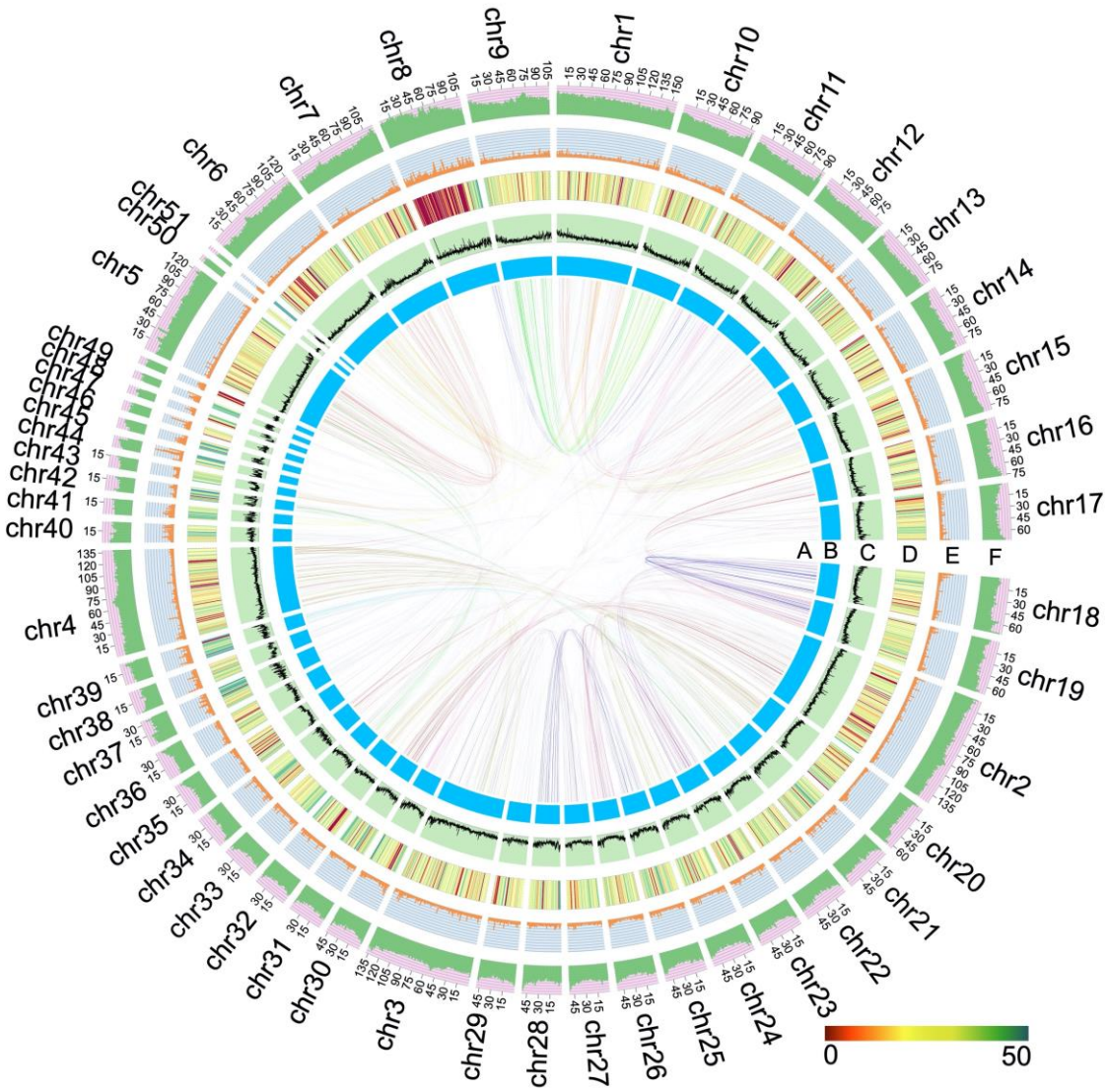



**Figure S2.** Circos plot of the genomic landscape, Related to Figure 1.

**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.

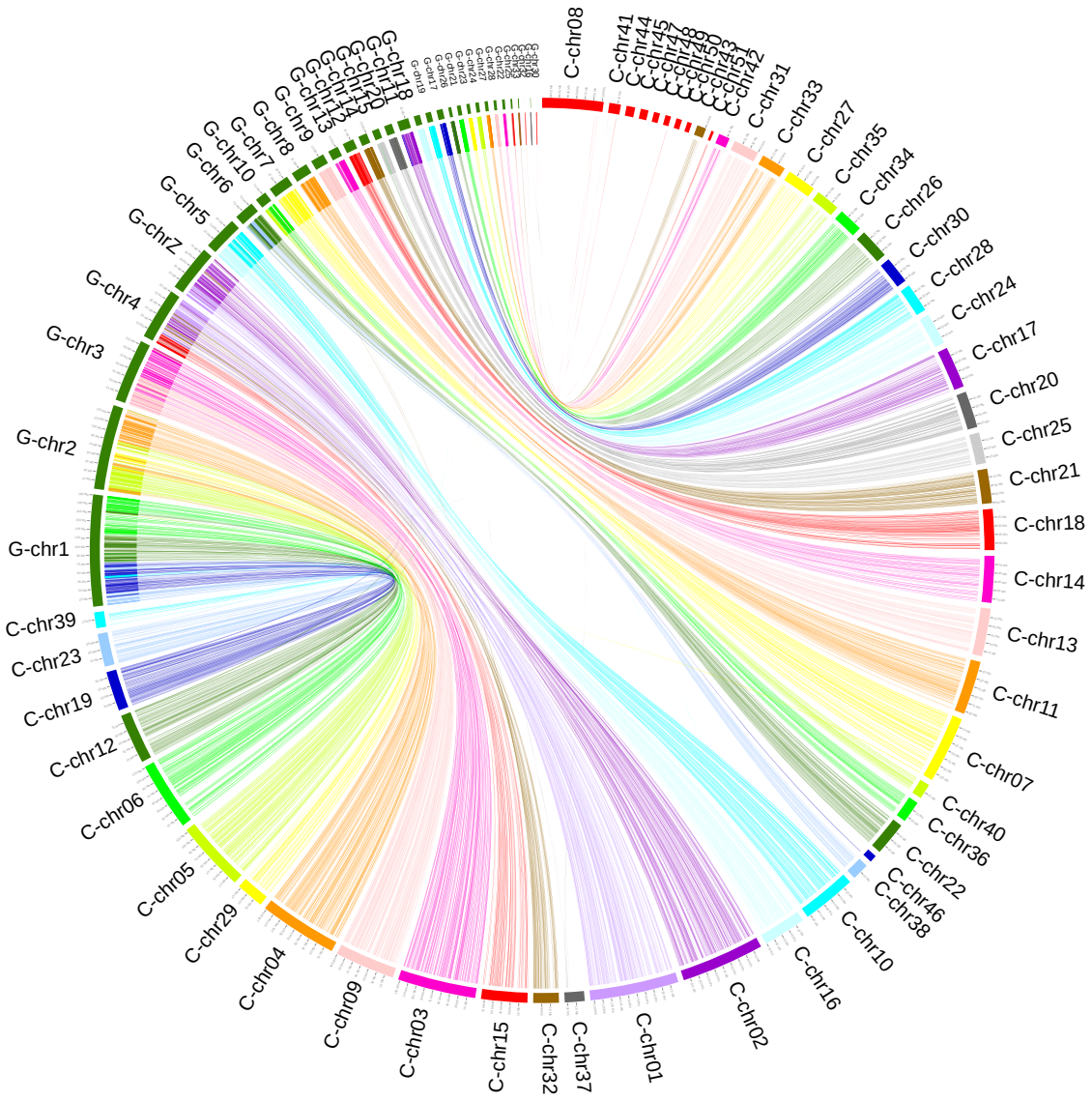



**Figure S3.** Circos plot of comparisons between bamboo shark and chicken, Related to Figure 1.

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.

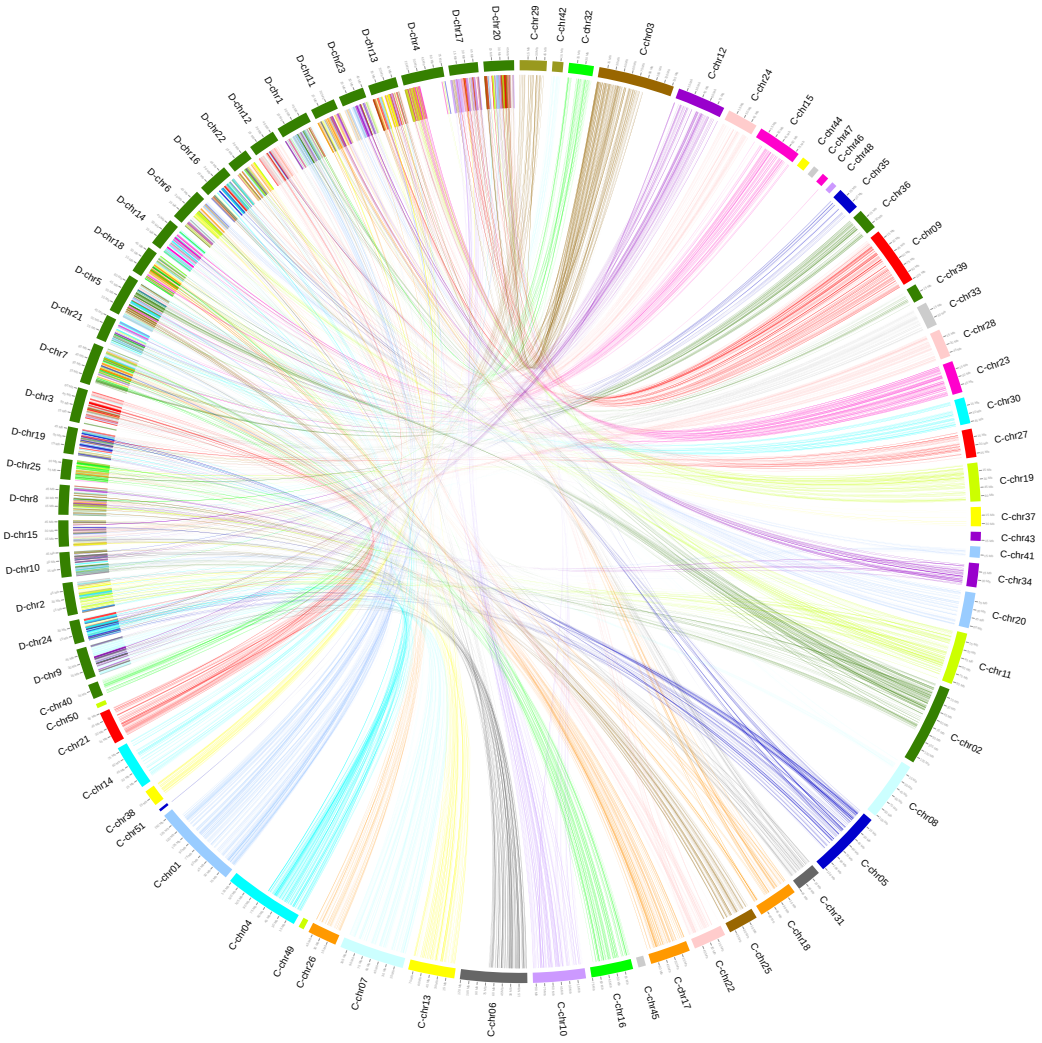



**Figure S4.** Synteny relationship of bamboo shark and zebrafish, Related to Figure 1.

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.

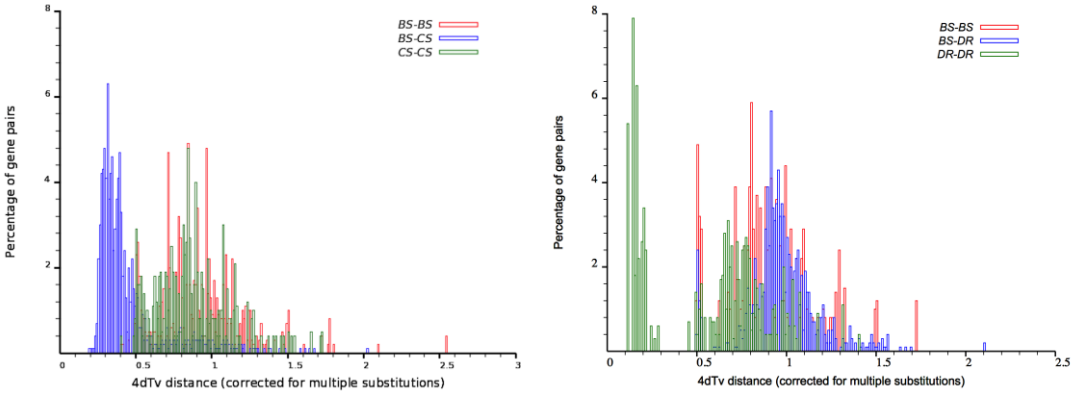



**Figure S5.** Whole genome duplication in elephant shark, bamboo shark and zebra fish genomes as revealed through 4DTv analyses, Related to Figure 1.

We took zebra fish (DR), elephant shark (CS) and bamboo shark (BS) as examples, detected gene pair blocks by using Mcscan 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).

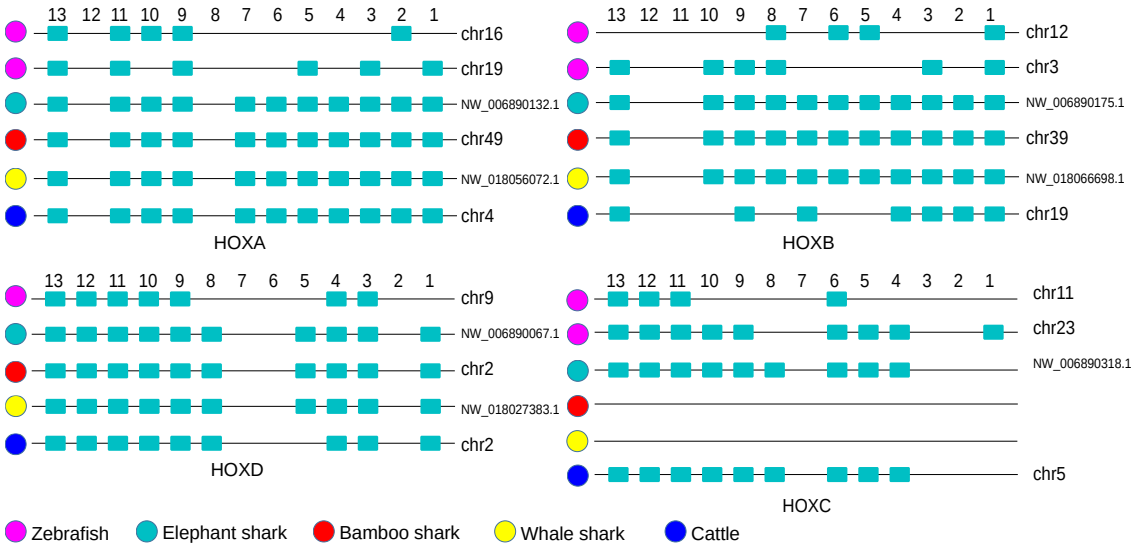



**Figure S6.** HOX gene clusters identified in five species, Related to Figure 1.

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.

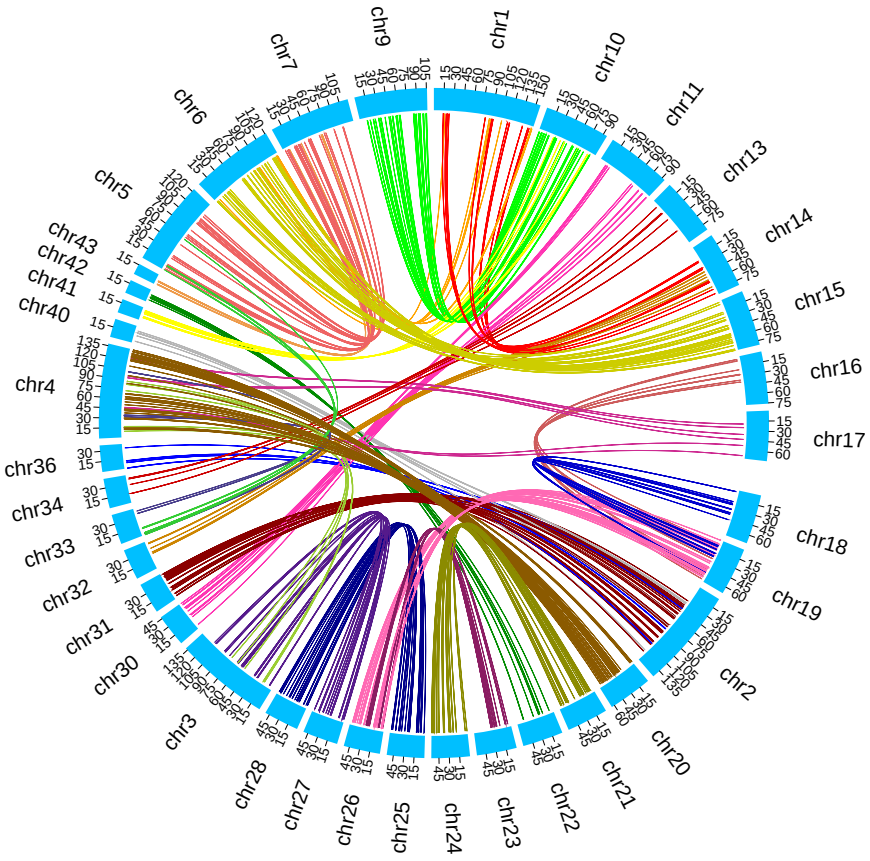



**Figure S7.** Distribution of shared paralogous genes of bamboo shark and elephant shark, Related to Figure 2. The lines represent gene pairs obtained by using blastp with both CIP and CALP (defined by *Salse et al(Salse et al., 2009)*) value of 0.5.

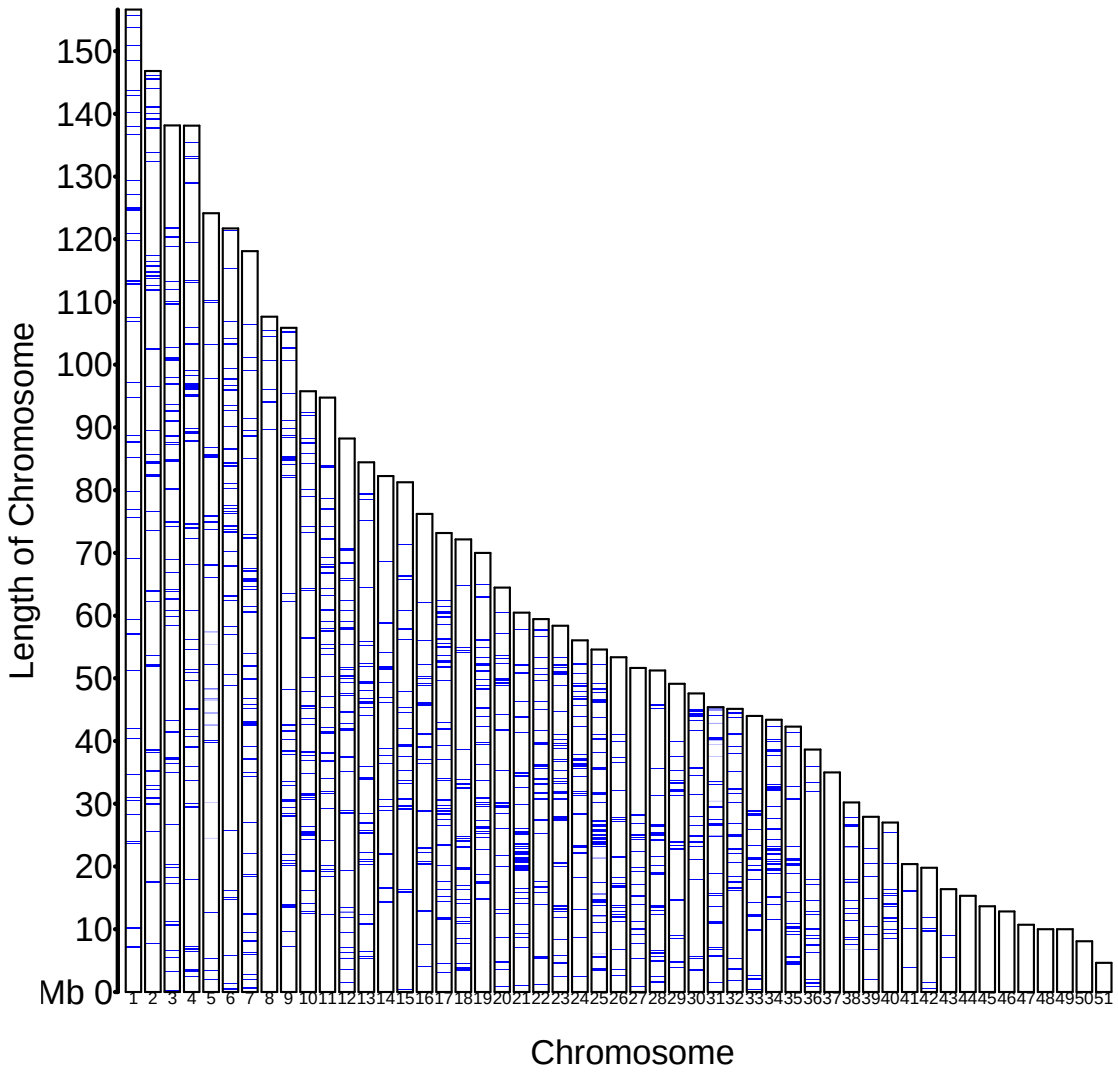



**Figure S8.** Distributions of conserved regions of six cartilaginous fishes (bamboo shark, elephant shark, whale shark, brownbanded bamboo shark, cloudy catshark and white shark) on 51 chromosomes of the bamboo shark, Related to Figure 2.

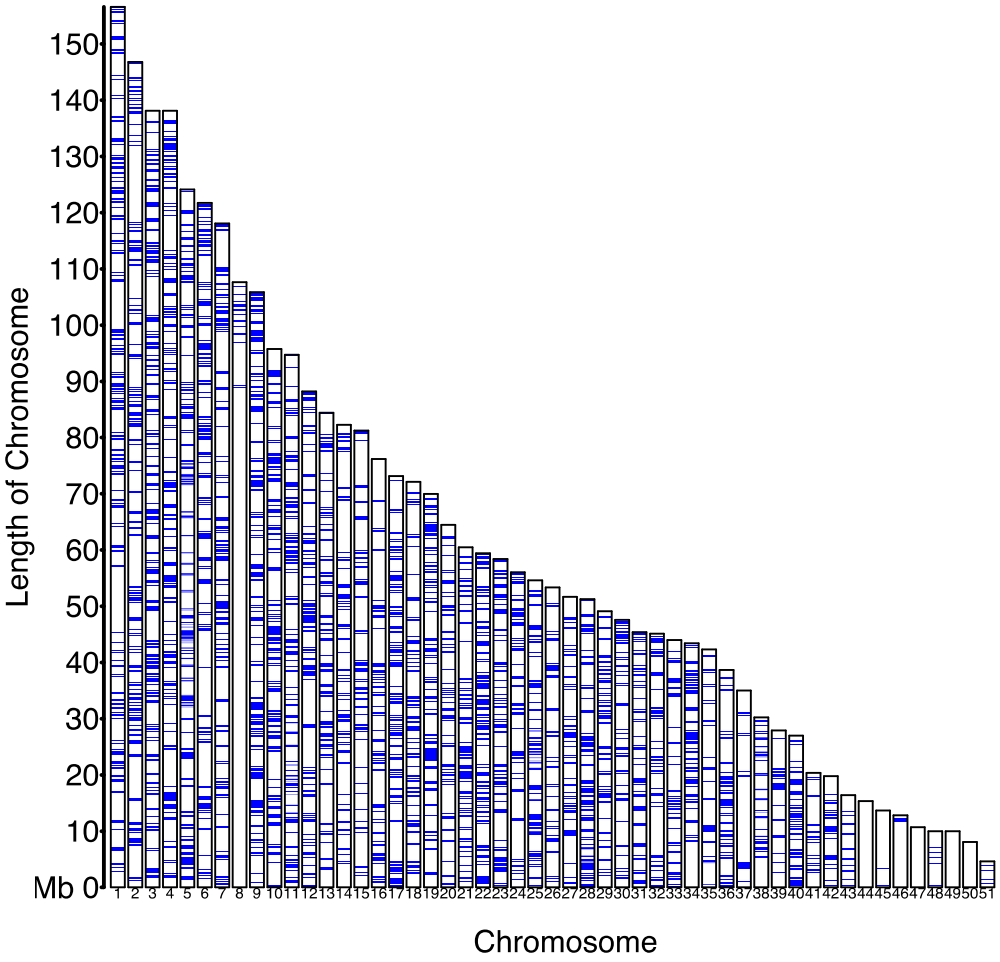



**Figure S9.** Distributions of conserved regions between bamboo shark and medaka on 51 chromosomes of the bamboo shark, Related to Figure 2.

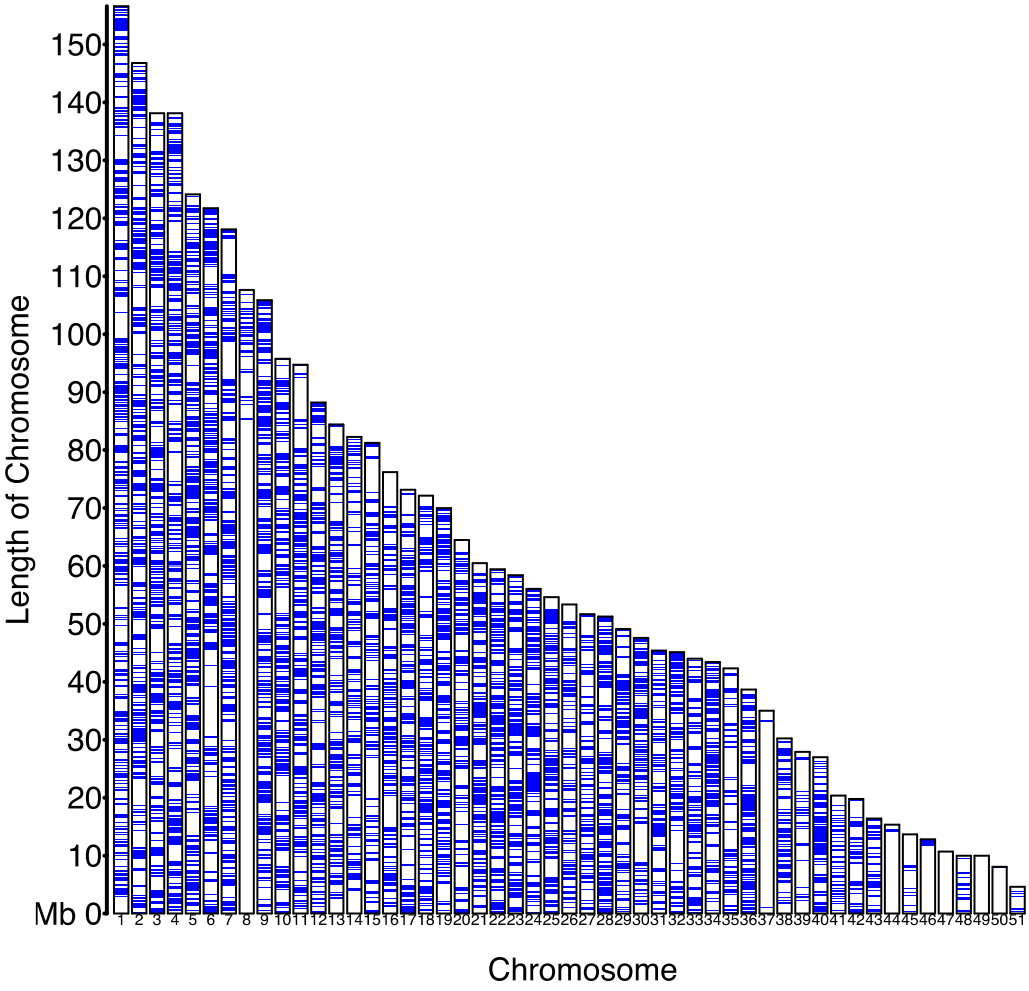



**Figure S10.** Distributions of conserved regions between bamboo shark and spotted gar on 51 chromosomes of the bamboo shark, Related to Figure 2.

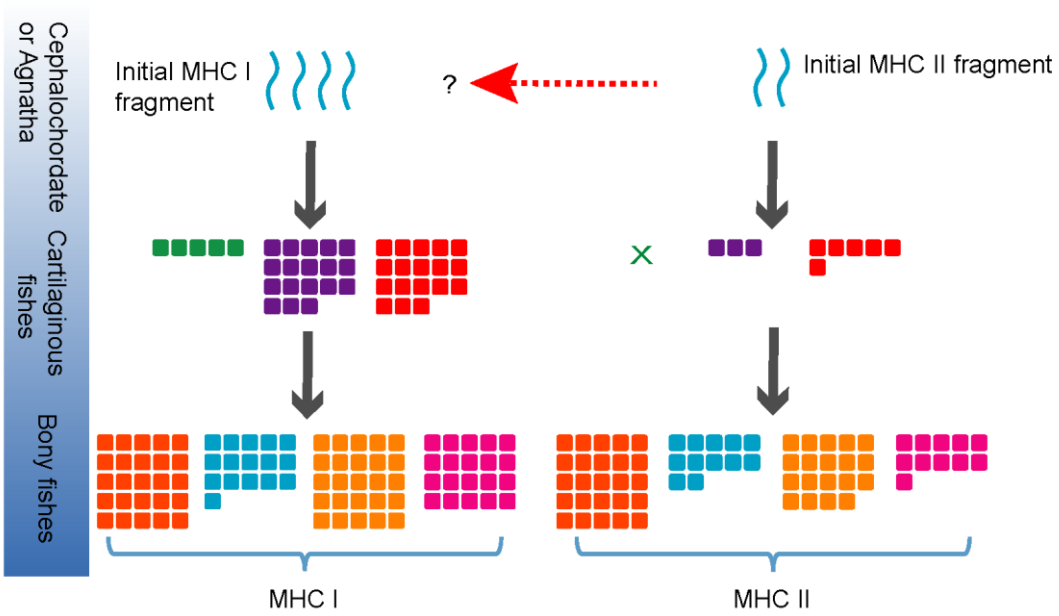



**Figure S11.** Hypothetical model of MHC genes evolution, Related to Figure 3.

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.

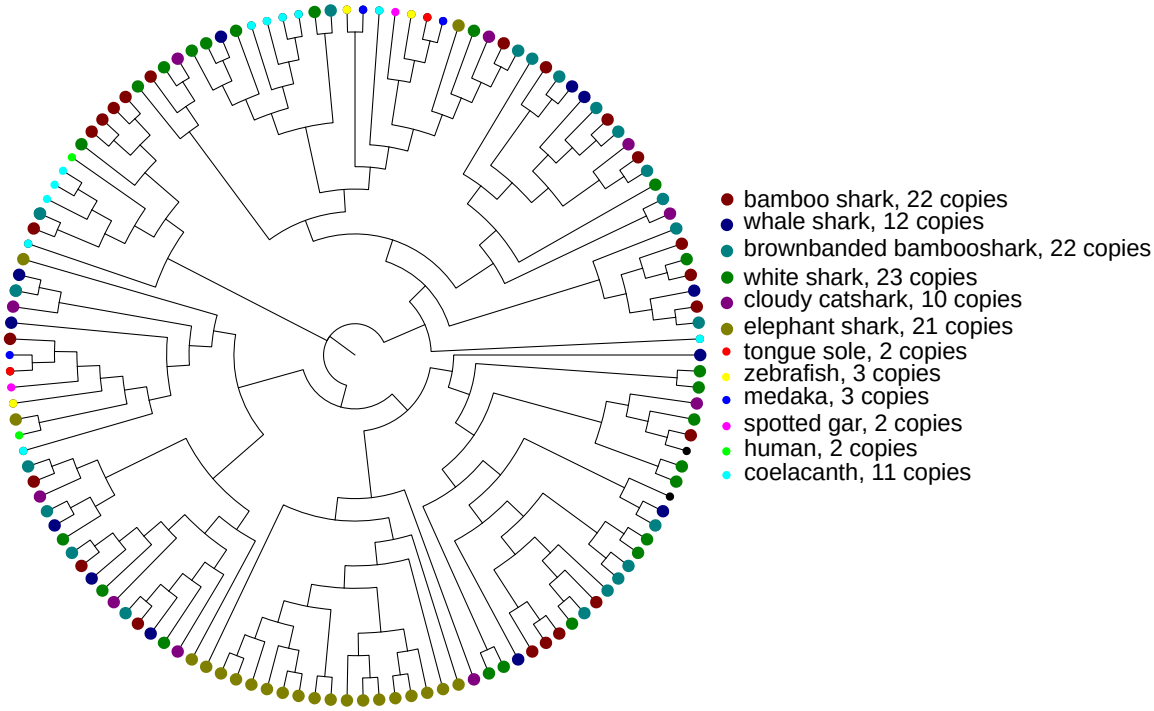



**Figure S12.** Gene tree of *TRIM69*, Related to Figure 3.

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).

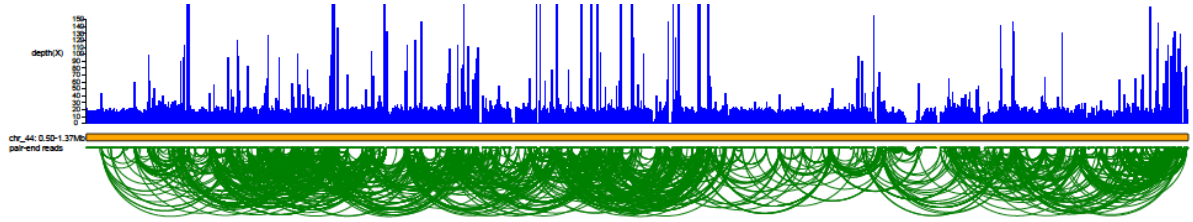



**Figure S13.** Re-assembled IgNAR region supported by paired-end reads, Related to Figure 3.

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.

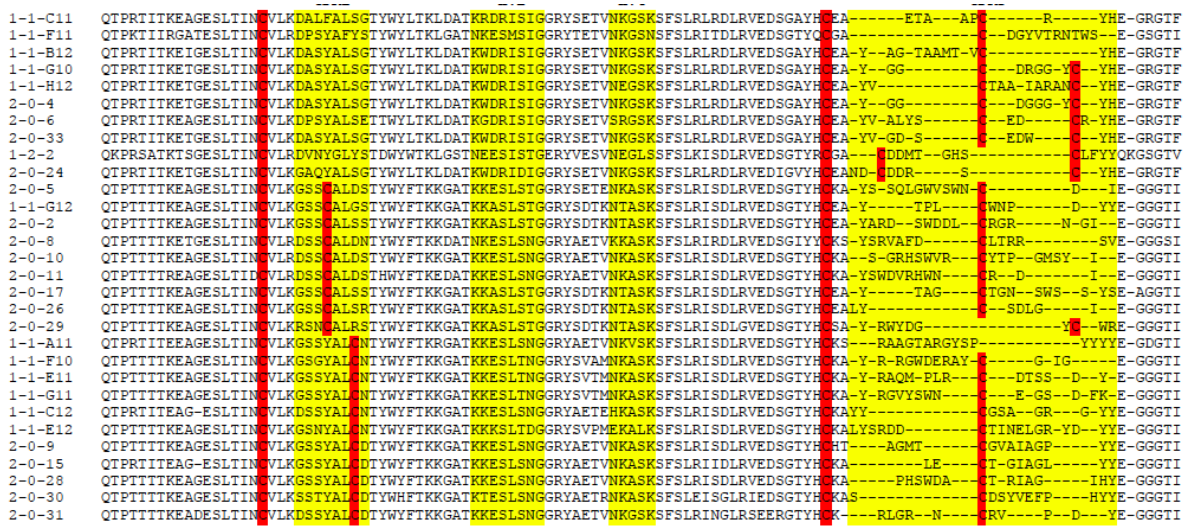



**Figure S14.** Sequence alignment of the randomly picked clones from the bamboo shark vNAR-phage display library that represent sequences have different CDR3, Related to Figure 3.

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 Size (bp) | Reads Length (bp) | Total 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 | 136,536,87 6 |
| 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.

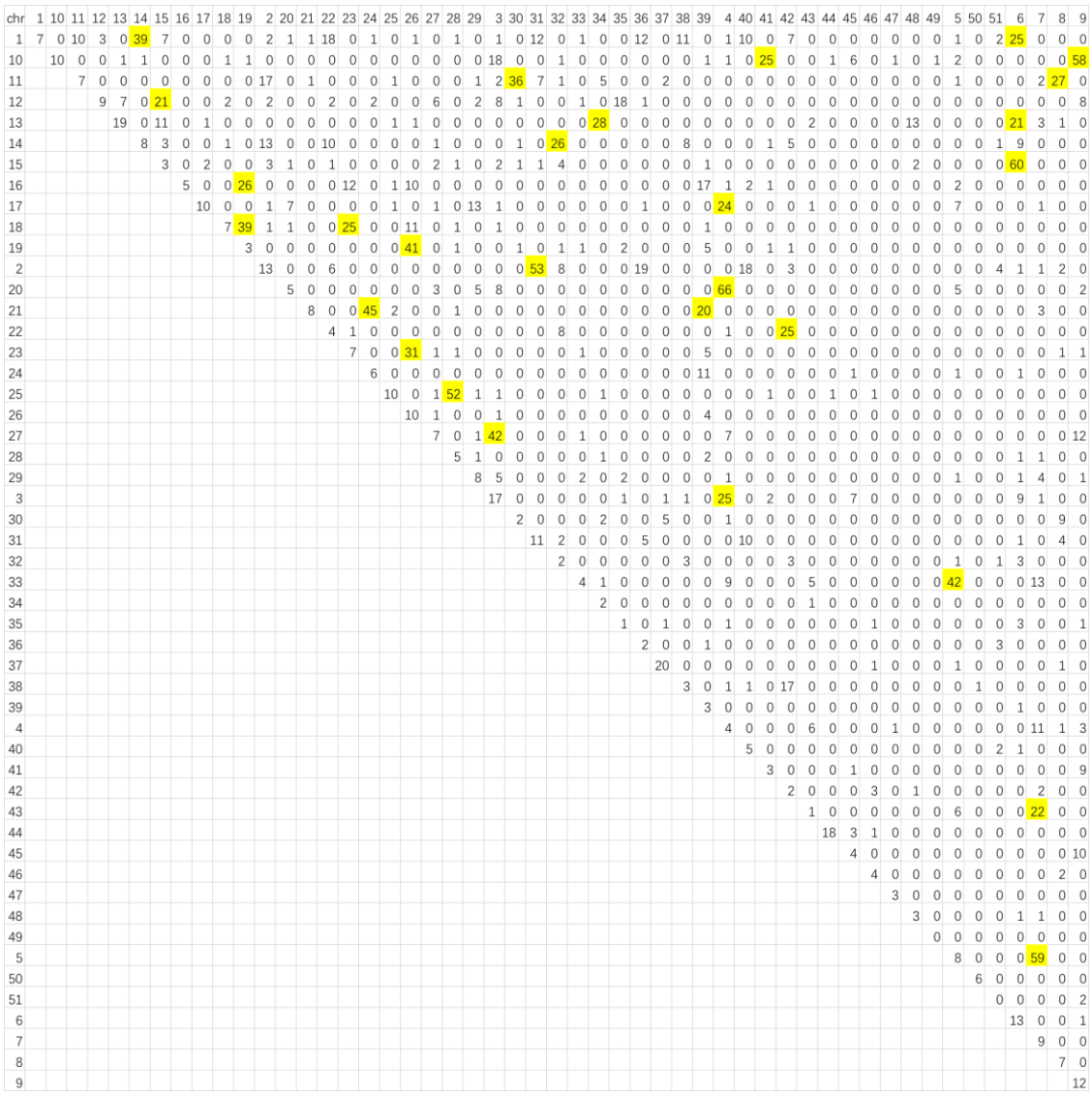


**Table S8.** Summary of conserved genes among six cartilaginous fishes distributed on each chromosome, Related to Figure 2.

| Chr | Total gene number | Chr length (bp) | Conserved gene number | Percentage of conserved gene number | Conserved gene length (bp) | Percentage of conserved gene length |
|---|---|---|---|---|---|---|
| chr1 | 749 | 156,605,781 | 45 | 6.01% | 2,317,785 | 1.48% |
| chr2 | 671 | 146,811,608 | 51 | 7.60% | 2,234,064 | 1.52% |
| chr3 | 714 | 138,146,672 | 46 | 6.44% | 2,109,842 | 1.53% |
| chr4 | 650 | 138,123,871 | 51 | 7.85% | 3,003,819 | 2.17% |

| | | | | | | |
|---|---|---|---|---|---|---|
| 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 | ENSORLT00000000027 | 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; |
| Three-spined sticklebacks | 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; |
| | ENSGACT00000025238 | Orla-DAB; H-2 class II histocompatibility antigen, E-S beta chain; |
| | ENSGACT00000000450 | Orla-DAB; H-2 class II histocompatibility antigen, E-S beta chain; |
| | 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; |
| Zebrafish | 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; |
| | 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 | Chipl12840 | 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:GGGTTGAACCACCGACAAACTGCAGGGGTTGAACAAACACCGACA | R1:ATAAGAATGCGGCCGCAATCCATTTGCCCTCTGTTCT |
| 2 | F2:AACTGCAGCCAAACACCGAGAACAATAACG | R2:ATAAGAATGCGGCCGCGAAATCCGTTTGCTCTCTGTTCT |
| 3 | F3:AACTGCAGCCAAACACCGAAAACGATAATAA | R3:ATAAGAATGCGGCCGCCAGAAATCCGTTTACTCTCTGTTCT |
| 4 | F4:AACTGCAGCCGAGATCAGCAACAAAAACTTC | R4:ATAAGAATGCGGCCGCAATCCGTTTGCTCTCTTTTCTTC |

**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 |
| Zebrafish | ENSDART00000167907 | 597 | p2rx7 |
| | ENSDART00000002866 | 398 | p2rx1 |
| | ENSDART00000098969 | 401 | p2rx4b |
| | ENSDART00000019461 | 400 | p2rx2 |
| | ENSDART00000130628 | 398 | p2rx4a |
| | ENSDART00000171919 | 482 | p2rx5 |
| | ENSDART00000163914 | 411 | p2rx3a |
| | ENSDART00000133926 | 412 | p2rx3b |
| | ENSDART00000029063 | 391 | p2rx5 |
| Medaka | ENSORLT00000014760 | 397 | p2rx1 |
| | ENSORLT00000004381 | 405 | p2rx3 |
| | ENSORLT00000019713 | 392 | p2rx4 |
| | ENSORLT00000015083 | 458 | p2rx5 |
| | ENSORLT00000009830 | 412 | p2rx2 |
| Sea lamprey | PMZ_0041355-RA_PMZ | 79 | p2rx5.L |
| | 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.