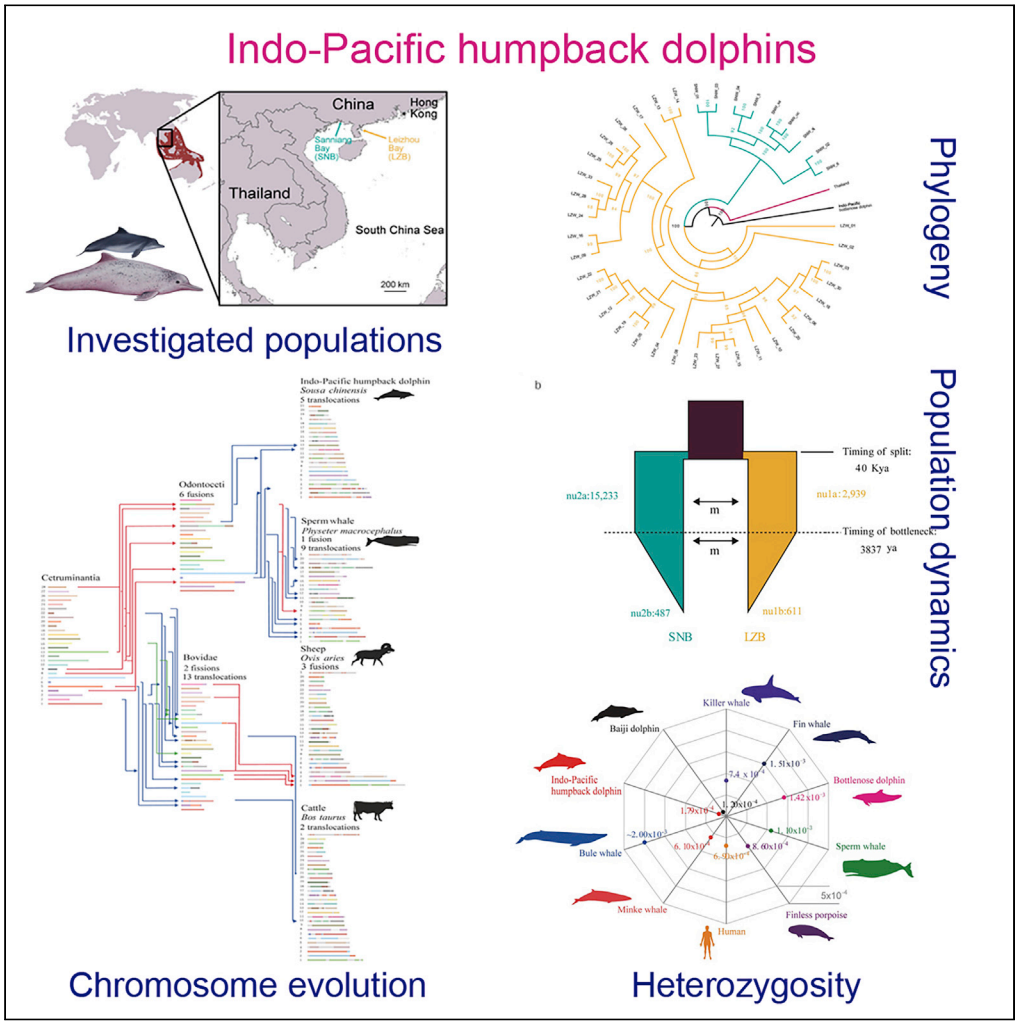
iScience

CellPress
OPEN ACCESS

Article

An Indo-Pacific Humpback Dolphin Genome Reveals Insights into Chromosome Evolution and the Demography of a Vulnerable Species



Peijun Zhang, Yong Zhao, Chang Li, ..., Guangyi Fan, Xin Liu, Songhai Li

fanguangyi@genomics.cn (G.F.)
liuxin@genomics.cn (X.L.)
lish@idsse.ac.cn (S.L.)

HIGHLIGHTS

Deducing chromosome evolution from ancestral Cetruminantia and ancestral Odontoceti

Reconstructing the demographic history of *Sousa chinensis*

Implying high risk of extinction and strong conservation requirement for *S. chinensis*

Zhang et al., iScience *23*, 101640
October 23, 2020
https://doi.org/10.1016/j.isci.2020.101640

iScience

CellPress
OPEN ACCESS

Article

An Indo-Pacific Humpback Dolphin Genome Reveals Insights into Chromosome Evolution and the Demography of a Vulnerable Species

Peijun Zhang,[1,4,23] Yong Zhao,[2,23] Chang Li,[2,8,23] Mingli Lin,[1] Lijun Dong,[1] Rui Zhang,[2] Mingzhong Liu,[1] Kuan Li,[1] He Zhang,[2,20] Xiaochuan Liu,[2] Yaolei Zhang,[2,21] Yuan Yuan,[1,11] Huan Liu,[3] Inge Seim,[9,10] Shuai Sun,[2] Xiao Du,[2] Yue Chang,[2] Feida Li,[3] Shanshan Liu,[2] Simon Ming-Yuen Lee,[5] Kun Wang,[11] Ding Wang,[12] Xianyan Wang,[13] Michael R. McGowen,[14] Thomas A. Jefferson,[15] Morten Tange Olsen,[16] Josefin Stiller,[4] Guojie Zhang,[4,6,19] Xun Xu,[2,3] Huanming Yang,[3,6] Guangyi Fan,[2,5,*] Xin Liu,[2,3,7,18,*] and Songhai Li[1,17,22,24,*]

SUMMARY

The Indo-Pacific humpback dolphin (*Sousa chinensis*) is a small inshore species of odontocete cetacean listed as Vulnerable on the IUCN Red List. Here, we report on the evolution of *S. chinensis* chromosomes from its cetruminant ancestor and elucidate the evolutionary history and population genetics of two neighboring *S. chinensis* populations. We found that breakpoints in ancestral chromosomes leading to *S. chinensis* could have affected the function of genes related to kidney filtration, body development, and immunity. Resequencing of individuals from two neighboring populations in the northwestern South China Sea, Leizhou Bay and Sanniang Bay, revealed genetic differentiation, low diversity, and small contemporary effective population sizes. Demographic analyses showed a marked decrease in the population size of the two investigated populations over the last ~4,000 years, possibly related to climatic oscillations. This study implies a high risk of extinction and strong conservation requirement for the Indo-Pacific humpback dolphin.

INTRODUCTION

A fascinating example of evolution is the return of terrestrial mammals to an aquatic environment. This occurred at least three separate times independently and is manifested by functional adaptations in 129 extant species in three major marine mammal lineages: Cetacea (whales, dolphins, and porpoises), Pinnipedia (walruses, sea lions, and seals), and Sirenia (manatees and dugongs) (Committee on Taxonomy, 2020). Recent genome-scale analyses of marine mammals have provided unprecedented insights into their evolution, adaptations, and demographic histories, highlighting the urgent need for conservation and management of many species (Árnason et al., 2018; Autenrieth et al., 2018; Brüniche-Olsen et al., 2018; Foote et al., 2015; Jones et al., 2017; Keane et al., 2015; Moskalev et al., 2017; Warren et al., 2017; Yim et al., 2014; Zhou et al., 2018a; Zhou et al., 2015; Zhou et al., 2018b; Zhou et al., 2013). Despite these advances, the genome of most marine mammals has yet to be sequenced, and cetacean genome assemblies published to date are often highly fragmented (e.g., see Zhou et al., 2018a), leaving research questions related to genomic structural evolution unanswered. A chromosome-level genome is fundamental for many downstream studies, including the analysis on whole-genome duplications (WGDs) and chromosome rearrangement events (Eichler and Sankoff, 2003). WGDs have been reported to strongly impact genome evolution and species formation in vertebrates. For example, all ray-finned fishes (*Actinopterygia*), accounting for 99% of teleosts, are shaped by a third WGD events (Van de Peer et al., 2009). Chromosome rearrangement usually accompanies the gain/loss of entire genes or regulatory regions, and is believed to play important roles in the evolution of lineage-specific traits and even speciation (Kirkpatrick, 2010). Thus, chromosome-level reference genomes are key for evolutionary analyses and conservation efforts of cetaceans (Moura et al., 2014; Viricel et al., 2014).

The Indo-Pacific humpback dolphin (*Sousa chinensis*; known as the Chinese white dolphin in China) is a small, toothed whale (Odontoceti) that occupies shallow tropical to temperate coastal habitats, from the

[1]Marine Mammal and Marine Bioacoustics Laboratory, Institute of Deep-sea Science and Engineering, Chinese Academy of Sciences, Sanya, Hainan 572000, China

[2]BGI-Qingdao, BGI-Shenzhen, Qingdao, Shandong 266555, China

[3]BGI-Shenzhen, Shenzhen, Guangdong 518083, China

[4]Section for Ecology and Evolution, Department of Biology, University of Copenhagen, Copenhagen 2100, Denmark

[5]State Key Laboratory of Quality Research in Chinese Medicine, Institute of Chinese Medical Sciences, University of Macau, Macao 999078, China

[6]China National GeneBank, BGI-Shenzhen, Shenzhen, Guangdong 518120, China

[7]BGI-Fuyang, BGI-Shenzhen, Fuyang, Anhui 236009, China

[8]BGI Education Center, University of Chinese Academy of Sciences, Shenzhen, Guangdong 518083, China

[9]Integrative Biology Laboratory, College of Life Sciences, Nanjing Normal University, Nanjing, Jiangsu 210023, China

[10]Comparative and Endocrine Biology Laboratory, Translational Research Institute-Institute of Health and Biomedical Innovation, School of Biomedical Sciences, Queensland University of Technology, Brisbane 4102, Australia

[11]Center for Ecological and Environmental Sciences,

Continued

Bay of Bengal to central China and throughout Southeast Asia (Jefferson and Rosenbaum, 2014; Jefferson and Smith, 2016). In recent decades, *S. chinensis* populations have been reported to face threats including by-catch in nets and overfishing (Dans et al., 2003), water pollution (Liu et al., 2018a), heavy marine traffic (Ng and Leung, 2003), and coastal development (Jefferson et al., 2009). Consequently, *S. chinensis* populations are declining (Huang et al., 2012), and the species is considered Vulnerable to extinction by the International Union for Conservation of Nature (Jefferson and Smith, 2016; Li, 2020). *S. chinensis* has also been listed as a Grade 1 National Key Protected Animal since 1988 in China, a distinction shared with the Yangtze River dolphin (baiji; *Lipotes vexillifer*), which was declared functionally extinct in 2007 (Turvey et al., 2007). The management and conservation of this Vulnerable species could be improved by a better understanding of its evolutionary history, inshore adaption, and population dynamics. Here, we traced both genomic reshuffling events and obtained population-level genetic data to illustrate how *S. chinensis* evolved in shallow tropical to temperate coastal habitats and predict its likely future population dynamics.

RESULTS

We generated ~317 Gb (129×) of *S. chinensis* genome data, using 10X Genomics technology (Pleasanton, CA, USA) on the BGISEQ-500 sequencing platform (Table S1) and assembled a 2.46-Gb draft genome sequence with a contig N50 of 114 kb and a scaffold N50 of 27.7 Mb, respectively (Table S2). To assign the draft genome assembly onto chromosomes, we produced ~73-Gb Hi-C data (Figures S1–S4) and anchored ~90.7% of genome scaffolds into 22 chromosomes (Figure S5 and Table S3). Overall, contig and scaffold metrics of this genome is higher than recently reported *S. chinensis* assemblies that were not resolved at the chromosome level (Jia et al., 2019; Ming et al., 2019) (Table S4). We annotated 20,767 protein-coding genes (Figure S6). Assessment with BUSCO (mammalia_odb9 gene set) revealed a high proportion of complete gene models in the genome assembly (~91.9%) and a reliable protein-coding gene set (~97.3%) (Table S5). We identified 17,286 gene families in *S. chinensis* by clustering genes with eight other mammals (human, sheep, cattle, finless porpoise, horse, minke whale, bottlenose dolphin, and sperm whale). Fossil-calibrated phylogenetic analysis, based on 1,915 single-copy gene families, revealed that *S. chinensis* and the bottlenose dolphin diverged ~4.8 million years ago (Figure S7).

As *S. chinensis* represents the first chromosome-level assembly of a 22-chromosome cetacean, we reconstructed the ancestral chromosomes of toothed whales as well as the ancestor of cetaceans and ruminants (Cetruminantia). The evolution of chromosomes is an important driver of speciation and diversification (White, 1969). It is well established that genomic structural rearrangements (often manifested as chromosome copy number variation stemming from fusions, fissions, and translocations) offer valuable insights into the diversification across macroevolutionary scales (Kim et al., 2017; O'Connor et al., 2018a, 2018b; Zhang et al., 2014). These rearrangements can result in the generation of novel genes or transcripts, as well as allowing the co-regulation of previously distantly located genes (Mertens et al., 2015). Cetacea and even-toed ungulates (Artiodactyla) evolved from a common terrestrial ancestor (Cetartiodactyla) over 55 million years ago (Berta et al., 2015); however, across extant species, there are substantial differences in the range of chromosome numbers between Cetacea (n = 21–22) and terrestrial Artiodactyla (n = 3–37) (ÁRnason, 1974; Graphodatsky et al., 2011; Kim et al., 2017; Rubes et al., 2012; Wurster and Benirschke, 1970). Previous studies made efforts to characterize the ancestral karyotype of Cetartiodactyla by relying on molecular cytogenetics (Rubes et al., 2012) and cross-species chromosome painting (Balmus et al., 2007; Kulemzina et al., 2009). A recent study, employing cross-species BAC mapping to the cattle X chromosome, revealed synteny blocks and rearrangements within Cetartiodactyla and a preliminary reconstruction of the ancestral X chromosome (Proskuryakova et al., 2017). However, the evolutionary history of ancient and recent chromosome rearrangements leading to extant cetaceans remains largely unexplored by using chromosome-level genome assemblies (Kim et al., 2017). Therefore, we performed synteny analysis using the genome of *S. chinensis* (n = 22), sperm whale (*Physeter macrocephalus*, n = 21), and cattle (*Bos taurus*, n = 30) (Figure 1A and Tables S6 and S7). This revealed evidence of extensive chromosomal rearrangements among them (Figures S8–S11).

Taking advantage of multiple high-quality chromosome-level assemblies for Cetartiodactyla (represented by sheep, cattle, sperm whale, and *S. chinensis*), as well as their sister group Perissodactyla (represented by the horse), we reconstructed the ancestral karyotypes of Cetruminantia, as well as the ancestors of Bovidae and Odontoceti. We identified 342 shared homologous synteny blocks (HSBs) (Murphy et al., 2005), covering 90.5%, 85.2%, 78.1%, 78.9%, and 83.8% of *S. chinensis*, sperm whale, sheep, cattle, and horse genomes, respectively (Figure 1B). Based on these HSBs, we reconstructed the ancestral chromosomes for

Northwestern Polytechnical University, Xi'an, Shaanxi 710072, China

[12]Key Laboratory of Aquatic Biodiversity and Conservation of Chinese Academy of Sciences, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, Hubei 430072, China

[13]Third Institute of Oceanography, Ministry of Natural Resources, Xiamen, Fujian 361005, China

[14]Department of Vertebrate Zoology, Smithsonian National Museum of Natural History, Washington DC 20560, USA

[15]Clymene Enterprises, Yerba Valley Way, Lakeside, CA 92040, USA

[16]Evolutionary Genomics Section, Globe Institute, University of Copenhagen, Øster Farimagsgade 5, Copenhagen 1353, Denmark

[17]Function Laboratory for Marine Fisheries Science and Food Production Processes, Qingdao National Laboratory for Marine Science and Technology, Qingdao, Shandong 266237, China

[18]State Key Laboratory of Agricultural Genomics, BGI-Shenzhen, Shenzhen, Guandong 518083, China

[19]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan 650223, China

[20]Department of Biology, Hong Kong Baptist University, Hong Kong, China

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

[22]Tropical Marine Science Institute, National University of Singapore, Singapore 119227, Singapore

[23]These authors contributed equally

[24]Lead Contact

*Correspondence: fanguangyi@genomics.cn (G.F.), liuxin@genomics.cn (X.L.), lish@idsse.ac.cn (S.L.)

https://doi.org/10.1016/j.isci.2020.101640

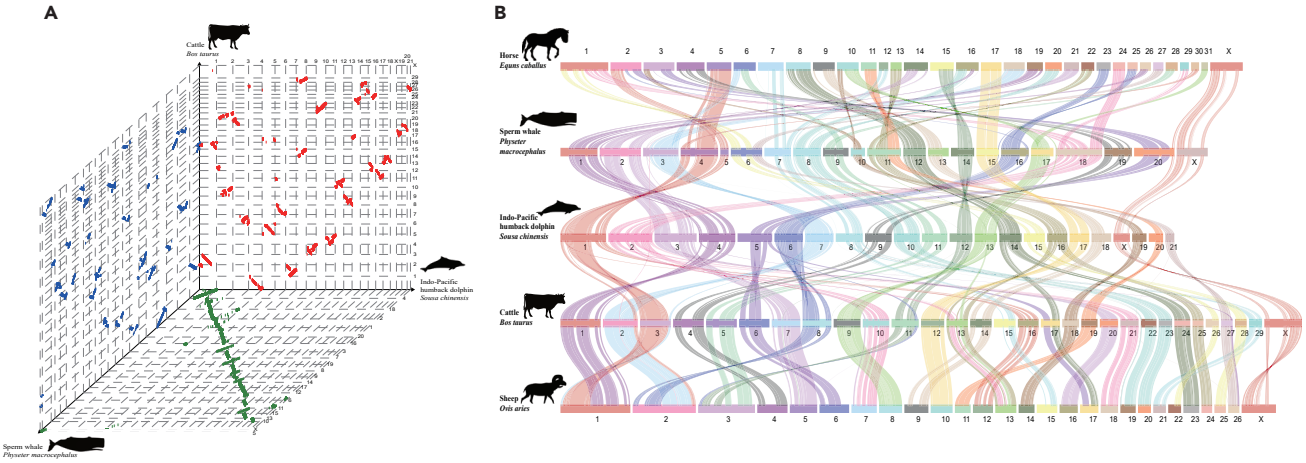


Figure 1. Synteny between the Indo-Pacific Humpback Dolphin and Other Mammal Species

(A) Synteny of genome scaffolds of the Indo-Pacific humpback dolphin and cattle (*Sousa chinensis* and *Bos taurus*; red lines), Indo-Pacific humpback dolphin and sperm whale (*S. chinensis* and *Physeter macrocephalus*; green lines), and sperm whale and cattle (*P. macrocephalus* and *B. taurus*; blue lines).

(B) 342 shared homologous synteny blocks (HSBs) in the Indo-Pacific humpback dolphin (*S. chinensis*), cattle (*B. taurus*), sperm whale (*P. macrocephalus*), sheep (*Ovis aries*), and horse (*Equus caballus*).

Cetruminantia (28), Bovidae (30), and Odontoceti (22) (Tables S8–S10). This analysis confirmed that the number of ancestral chromosomes of Odontoceti is the same as that of most extant cetacean families (22 chromosome pairs) and different from that of superfamily Physeteroidea (ÁRnason, 1974) and family Ziphiidae (Kurihara et al., 2017) (21 chromosome pairs). Thus, we propose a model of chromosome evolution in cetaceans in which a chromosome fusion event likely occurred in the physeterids (i.e., sperm whales).

We detected 14 reshuffling events (including six fusions) in the lineage leading from the ancestral Cetruminantia to the Odontoceti ancestor, 32 in the *S. chinensis* lineage, and 42 in the sperm whale lineage. We identified 42 chromosome reshuffling events (including two fissions) in the common ancestor of Bovidae as well as five and eight chromosome reshuffling events in the cattle and sheep lineages, respectively (Tables S11, S12, S13, S14, S15, and S16). Thus, we confirmed a completely opposite karyotype evolution pattern between the lineage leading to the Odontoceti ancestor (reduced chromosome number) and the Bovidae ancestor (increased chromosome number).

Chromosome rearrangements can serve a functional purpose (Mérot et al., 2020). We identified seven fusions, 34 inversions, and 40 breakpoints during the evolution in the ancestral odontocete lineage (Figure 2). Such rearrangements in marine mammals could affect the function (e.g., regulation) of genes, thereby mediating adaptations to an aquatic environment. To explore this phenomenon, we compared ancestral Cetruminantia to ancestral Odontoceti and ancestral Odontoceti to *S. chinensis*. When comparing Odontoceti and Cetruminantia, we found ten genes located in 4-kb flanking regions of the 40 breakpoints (Table S17). For example, adenylate cyclase type 1 (*ADCY1*; *Sochi02981*), which plays a role in kidney filtration (Xiao et al., 2011), is associated with distinct kidney structures of freshwater and marine finless porpoises (Zhou et al., 2018a). Coding variants of the sodium chloride symporter gene *SLC12A3* (*Sochi08789*) are associated with reduced blood pressure in humans (Nandakumar et al., 2018). Glypican-3 precursor (*GPC3*; *Sochi10755*) plays a role in vertebrate limb patterning and skeletal development by controlling the cellular response to bone morphogenic protein 4 (BMP4) and may be related to the lack of hind limbs in cetaceans (Paine-Saunders et al., 2000; Saad et al., 2017).

Compared with the odontocete ancestor, we detected 105 inversions in *S. chinensis* (Figure 2). Six genes were located in 4-kb flanking regions of 138 breakpoints resulting from these HSBs (Table S18). One of these genes, *SERPINB8* (*Sochi06838*), is involved in maintaining the mechanical stability of skin (Pigors et al., 2016). Loss of *SERPINB8* in humans results in peeling skin syndrome, where the outermost layer of the epidermis peels off upon exposure to water or skin occlusion (Pigors et al., 2016). Some *S. chinensis* individuals in Chinese waters have pink skin, likely attributed to the dilation of subcutaneous blood vessels to get rid of excess body heat (Di Meglio et al., 2011). We speculate that *SERPINB8* might mediate this

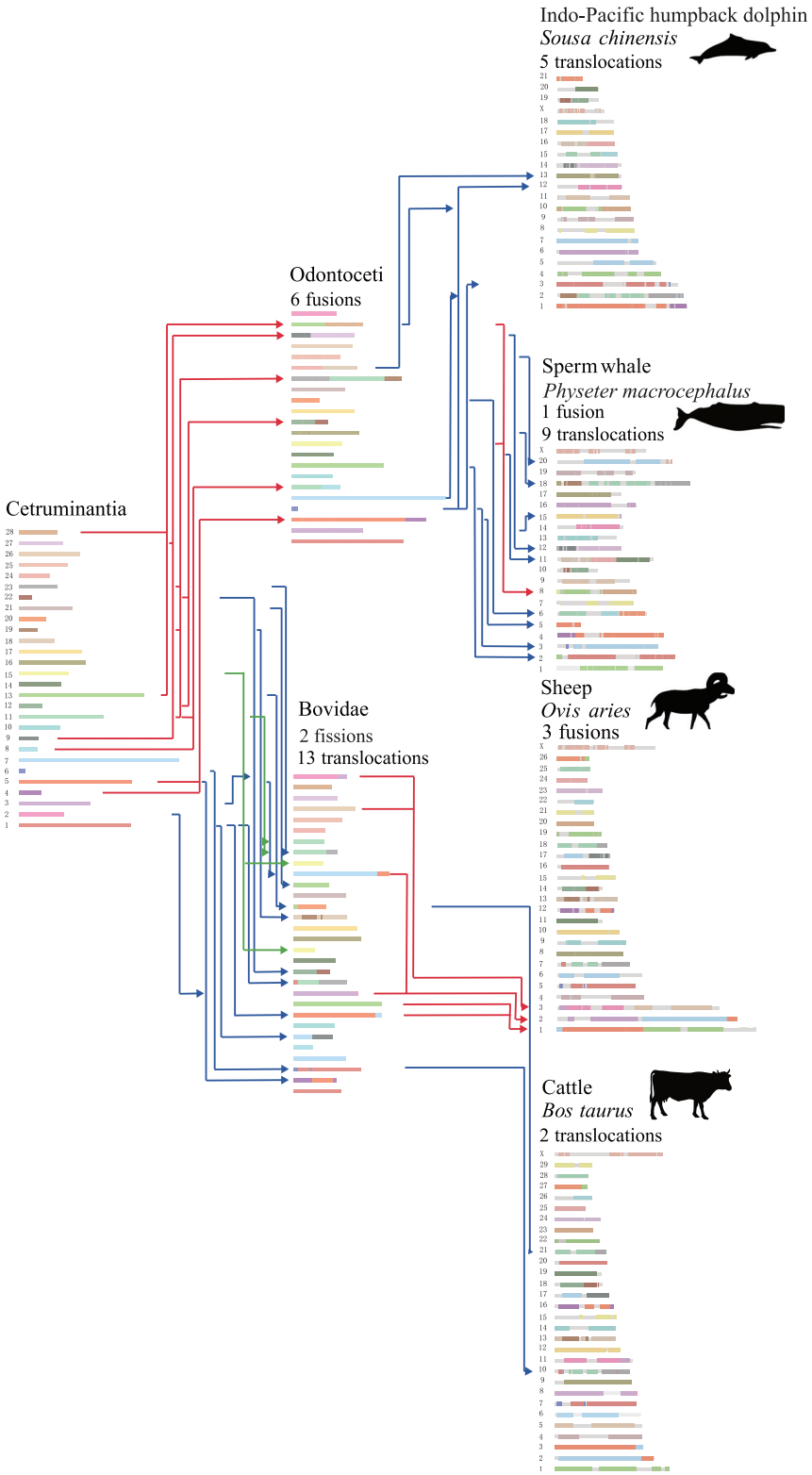


Figure 2. Chromosome Evolution of Cetruminantia

Chromosomes from the same ancestral chromosome are in the same color. Blue lines denote translocations, red lines fusions, and green lines fissions.

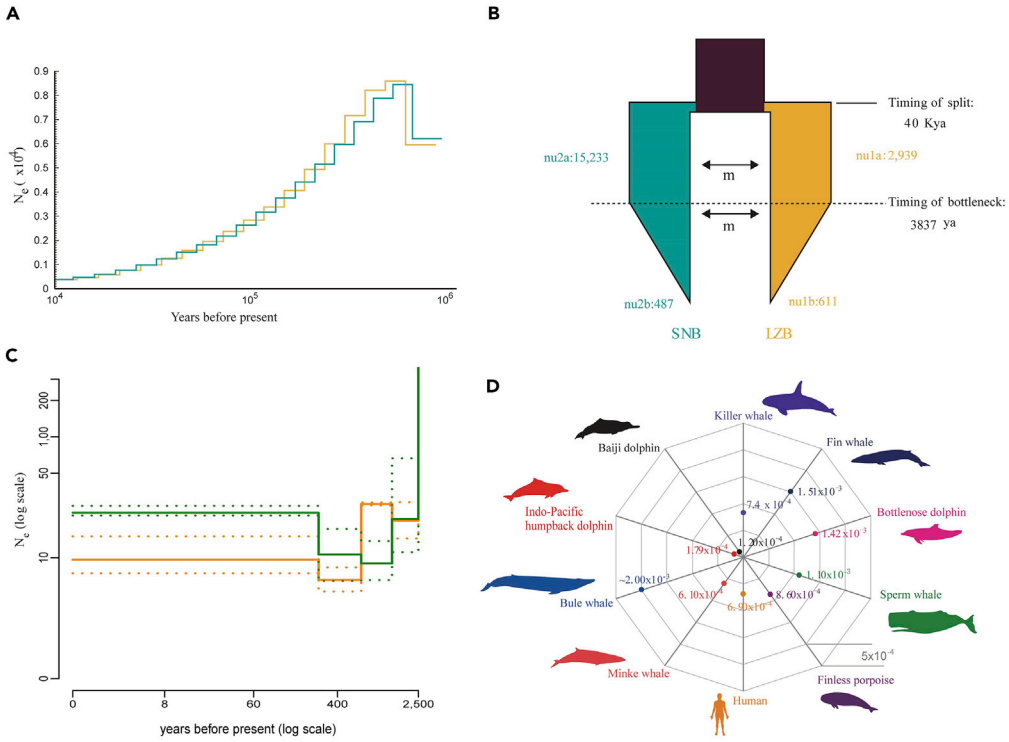


Figure 3. The Heterozygosity and Demographic History of Indo-Pacific Humpback Dolphins in the South China Sea

(A) Demographic histories of Indo-Pacific humpback dolphin populations using the PSMC model. The orange line denotes the LZB population; the green line denotes the SNB population. Settings: g (generation time) = 25 years; μ (neutral mutation rate per generation) = 5.81×10^{-9}.

(B) DaDi analyses of Indo-Pacific humpback dolphin population models between LZB and SNB. We used the "sym_mig_size" model, under which two populations split with symmetric migration and then size change with symmetric migration. nu1a: size of population LZB after split. nu2a: size of population SNB after split. T1: time in the past of split. nu1b: size of LZB after time interval. nu2b: size of population SNB after time interval. T2: time of population size change. m: migration rate between populations.

(C) Effective population sizes (*Ne*) of LZB and SNB in recent time (100 generations before present; approximately 2,500 years). The orange line denotes the LZB population; the green line denotes the SNB population, with 90% confidence intervals for each population as dotted lines. (D) Comparison of heterozygosity between Leizhou Bay (LZB) and Sanniang Bay (SNB) populations and other mammals.

adaptation, rendering *S. chinensis* more susceptible to physical damage, stressors, and pathogens in their present-day environments. *S. chinensis* is vulnerable to environmental pollutants and pathogens (Parsons, 1998; Parsons et al., 2001), threats likely to affect its immune system. Another gene in the breakpoint region is *CDA* (*Sochi00142*), a gene encoding a modulator of innate immunity (Furusho et al., 2018; Liu et al., 2018b).

Ecologically, *S. chinensis* is distributed in several large estuarine areas within Chinese waters: the west coast waters of Taiwan, Xiamen waters, the Pearl River estuary, the southwest waters of Hainan Island, Leizhou Bay (hereafter LZB), and Sanniang Bay (hereafter SNB) (Chen et al., 2009, 2018a; Hemami et al., 2018; Huang et al., 2012; Jefferson and Smith, 2016; Li et al., 2016; Wang et al., 2012; Xu et al., 2015). To gain insights into the demographic history and population genetic status of *S. chinensis*, we sequenced 39 individuals (at an average sequencing depth of ~30×) from two neighboring regions in Chinese waters (LZB and SNB) (Table S19). We reconstructed the demographic history of *S. chinensis* in various ways. First, the pairwise sequentially Markovian coalescent (PSMC) model analysis (N_e estimates reliable at 20 ka to 3 Ma, Li and Durbin, 2011) revealed the onset of a severe *S. chinensis* population decline during the Mid-Pleistocene Transition 1 mya (Figure 3A). Next, we estimated that the populations split at ~40 ka (38.1k ~ 42.2k, 95% confidence interval [CI]), a time with relatively large *S. chinensis* effective population sizes (LZB 2,939; SNB 15,233) (Figure 3B). This timing coincides with an alternating warm and

wet period during the last glacial age, making the LZB and SNB available for colonization and gradual emergence of the two separate populations on either side of the Leizhou Peninsula (Zhong et al., 2010). After the population split, symmetric gene flow between the two populations was detected, suggesting there was still a relatively large genetic pool or mating opportunities between the two populations. This was followed by a marked bottleneck 3,837 (3,645–4,039, 95% CI) years ago, with effective population sizes (N_e) decreasing rapidly to 611 and 487 in LZB and SNB, respectively (Figure 3B, Tables S20 and S21). The bottleneck may be associated with the rapid temperature and sea level decline ~4,000 ago (Li, 2018; Li et al., 2017).

A decreased population size was supported by an approximate Bayesian computation method, PopSizeABC (Boitard et al., 2016), employed to better characterize the contemporary *S. chinensis* effective population size (N_e) dynamics over the last 100 generations. We found that both the LZB and the SNB populations declined over the last 2,500 years, with a contemporary N_e of 12 (11–15, 90% CI) in LZB and 8 in SNB (6–11, 90% CI), respectively (Figure 3C). The extraordinary low population estimate of these populations highlights the urgent need for efficient conservation and management efforts.

In addition, the nucleotide diversity (π) of *S. chinensis* (π: 0.00015 in LZB and 0.00016 in SNB; Figure S12) is much lower than the killer whale estimation (π = 0.0029) (Foote et al., 2016). We also found a relatively low mitochondrial DNA polymorphism (LZB: 0.0034 and SNB: 0.0038), with only 52 SNPs. The average heterozygosity of *S. chinensis* is 1.79×10^{-4}, similar to the recently extinct baiji (1.21×10^{-4}), but notably lower than the common bottlenose dolphins (14.2×10^{-4}) and Yangtze finless porpoises (8.60×10^{-4}) (Figure 3D and Table S22) (Yim et al., 2014; Yuan et al., 2018; Zhou et al., 2013, 2018a). As inbreeding is considered a potential cause of low diversity, we calculated the inbreeding coefficient (F_h) within the two *S. chinensis* populations (Wright, 1922), but found no evidence of inbreeding (Tables S23 and S24).

Early barriers to gene flow (introgression) such as geographic isolation are common drivers of speciation (Poelstra et al., 2014; Via, 2009). SNB and LZB clearly split into two populations isolated by the Leizhou Peninsula (Figures 4A–4D and S13). Furthermore, admixture analysis indicates that LZB subgroups exist (Figure 4B). This observation may reflect the pigmentation differences (Chen et al., 2018b) of individuals in this population (Figure S14); however, further studies are warranted. We found no evidence of gene flow between the LZB and SNB populations from f_3 statistics (Reich et al., 2009) (Table S25). We next assessed the genetic differentiation between populations, and the overall fixation index (F_{ST}) was estimated to be ~0.138, indicating moderate differentiation. To reliably investigate positive selection in the LZB and SNB populations separately under the background of moderate differentiation, we integrated five different methods, including F_{ST}, Tajima's *D*, XP-CLR, XP-EHH, and the μ-statistic (Alachiotis and Pavlos, 2018; Chen et al., 2010; Danecek et al., 2011; Sabeti et al., 2007) (Figures S15 and S16). Finally, we identified seven putative regions with eight genes (*CAMKK2*, *P2RX4*, *P2RX7*, *IFT81*, *ATP2A*, *OLA1*, and *KANSL2*) in the LZB population (Table S26). Among those genes, *CAMKK2* mediates pleiotropic responses to physiologic and pathophysiological processes (Marcelo et al., 2016; Racioppi and Means, 2012), *P2RX4* and *P2RX7* play critical roles in immunity and inflammation (Burnstock, 2016; Raouf et al., 2007), *ATP2A* is involved in the role of maintaining cytoplasmic calcium levels (Pegoraro et al., 2011), and *OLA1* is involved in oxidative stress (Zhang et al., 2009). Correspondingly, in the SNB population, we identified four regions with five genes (two copies of *TBC1D10A*, *CNTRL*, *C5*, and *THBS1*) (Table S26). Among these genes, *C5* and *THBS1* are involved in the regulation of the innate immune response during infection (Noris and Remuzzi, 2013; Zhao et al., 2015). These results suggest that, although geographical isolation led to the differentiation between the two populations, natural selection also played a role in the adaptation to their respective habitats.

Our analyses suggest that the LZB and SNB populations experienced a severe genetic bottleneck and that there has been no migration after the populations split. Although genetic drift caused by a bottleneck is likely to be the dominant evolutionary mechanism driving population divergence between LZB and SNB populations, natural selection also played a role. Our results also suggest that the LZB and SNB populations are differentiating, which may be reflected in morphological differentiation. Indeed, evidence of recent morphological differentiation has been identified between other *S. chinensis* populations (Wang et al., 2015).

DISCUSSION

The high-quality genome assembly of the Indo-Pacific humpback dolphin reported in this study enables a unique view of cetacean chromosome evolution. Although there have been numerous studies on cetacean genome evolution, including gene family expansions/contractions, genes under selection, and their possible roles in adaptation, our work is the first to consider evolution at the chromosome level using

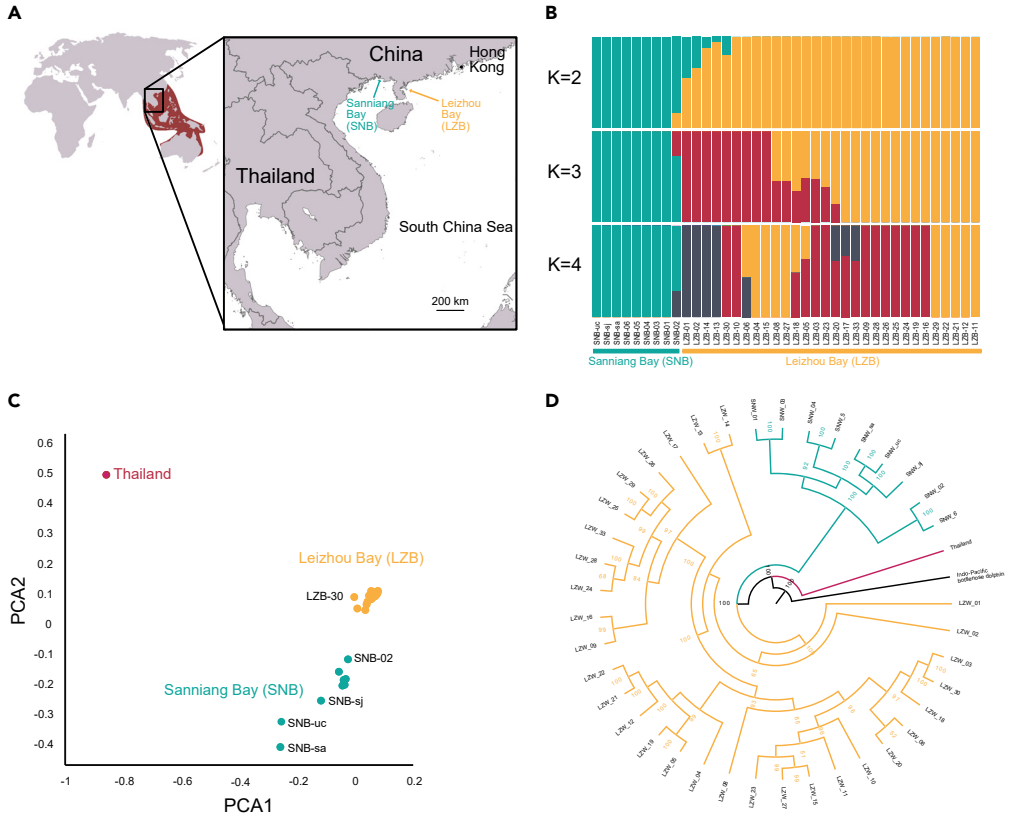


Figure 4. Population Characteristics of Indo-Pacific Humpback Dolphins

(A) Locations of Leizhou Bay (LZB) and Sanniang Bay (SNB) populations in the northwestern South China Sea.

(B) Population structure of 39 Indo-Pacific humpback dolphins, varying the number of admixture components (*K*) from 2 to 4. The sample location for each individual is indicated with a population-specific prefix (LZB or SNB).

(C) Principal-component analysis (PCA) of LZB and SNB Indo-Pacific humpback dolphins. In particular, the first eigenvector separated the two populations. (D) Phylogenetic tree of Indo-Pacific humpback dolphins from the LZB and SNB populations.

genomic data. We have established chromosome evolution from the common ancestor of Cetruminantia (28 chromosomes) to the ancestral Odontoceti (22 chromosomes), to *S. chinensis* (22 chromosomes). The ever-increasing number of chromosome-level genomes promises an increasingly more precise picture of the genomes of extant species and their ancestors. Finally, we reveal that *S. chinensis* population sizes in southern China are steadily decreasing, coupled with evidence of divergence of populations in close geographic proximity. The estimated worldwide population size of *S. chinensis* is less than 13,000 individuals (Huang et al., 2012; Jefferson et al., 2017), nearly half of which are in Chinese waters. This study emphasizes the urgency for effective conservation of the Indo-Pacific humpback dolphin, a Vulnerable species.

Limitations of the Study

We reported a high-quality assembly of the Indo-Pacific humpback dolphin. We deduced the chromosome reshuffle events from its cetruminant ancestor and identified 16 genes locating in breakpoint regions during the chromosome evolution, potentially related to an aquatic environment adaptation. More evidence is needed to firmly establish a functional role for these genes, however. Furthermore, samples from wider geographical locations are needed to fully capture the population history of the Indo-Pacific humpback dolphin.

Resource Availability

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Songhai Li (lish@idsse.ac.cn).

Materials Availability

This study did not generate new unique reagents.

Data and Code Availability

The accession number for the Indo-Pacific humpback dolphin chromosomal genome assembly reported in this paper is CNGBdb: CNP0000397.

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.2020.101640.

ACKNOWLEDGMENTS

This work was financially supported by the Ocean Park Conservation Foundation, Hong Kong (Nos. MM03-1415, MM02-1516); the National Natural Science Foundation of China (No. 41422604); the Major Science and Technology Project of Hainan Province (Nos. ZDKJ2016009 and ZDKJ2019011); Chinese White Dolphin Conservation Action Project of and Rural Affairs of the People's Republic of China (No. Y760091HT1); the biodiversity investigation, observation and assessment program (2019–2023) of Ministry of Ecology and Environment, the People's Republic of China; the Youth Innovation Promotion Association of Chinese Academy of Sciences; and the Strategic Priority Research Program of the Chinese Academy of Sciences (No. XDA19060403).

AUTHOR CONTRIBUTIONS

S.L. conceived and designed this study. P.Z., Y. Zhao, and C.L. performed sample preparation and sequencing. P.Z., M. Lin, K.L., M. Liu, and L.D. collected the samples for assembly and re-sequencing. R.Z. and X.L. performed genome assembly. Y. Zhang and Y. Zhao performed genome annotation. Y. Zhao, G.F., and I.S. performed analysis of evolution. C.L., Y. Zhao, R.Z., and H.Z. performed analysis of population. Y. Zhao, P. Z., G.F., X.L., and S.L. wrote the manuscript. All other authors reviewed and revised the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: April 22, 2020
Revised: August 25, 2020
Accepted: September 30, 2020
Published: October 23, 2020

REFERENCES

Alachiotis, N.,A., and Pavlos, P. (2018). RAiSD detects positive selection based on multiple signatures of a selective sweep and SNP vectors. Commun. Biol. *1*, 79.

ÁRnason, Ú. (1974). Comparative chromosome studies in Cetacea. Hereditas *77*, 1–36.

Árnason, Ú., Lammers, F., Kumar, V., Nilsson, M.A., and Janke, A. (2018). Whole-genome sequencing of the blue whale and other rorquals finds signatures for introgressive gene flow. Sci. Adv. *4*, eaap9873.

Autenrieth, M., Hartmann, S., Lah, L., Roos, A., Dennis, A.B., and Tiedemann, R. (2018). High-quality whole-genome sequence of an abundant Holarctic odontocete, the harbour porpoise (Phocoena phocoena). Mol. Ecol. Resour. *18*, 1469–1481.

Balmus, G., Trifonov, V.A., Biltueva, L.S., O'Brien, P.C.M., Alkalaeva, E.S., Fu, B., Skidmore, J.A., Allen, T., Graphodatsky, A.S., Yang, F., et al. (2007). Cross-species chromosome painting among camel, cattle, pig and human: further insights into the putative Cetartiodactyla ancestral karyotype. Chromosome Res. *15*, 499–514.

Berta, A., Sumich, J.L., and Kovacs, K.M. (2015). Marine Mammals: Evolutionary Biology (Academic Press).

Boitard, S., Rodríguez, W., Jay, F., Mona, S., and Austerlitz, F. (2016). Inferring population size history from large samples of Genome-Wide molecular data - an approximate bayesian computation approach. PLoS Genet. *12*, e1005877.

Brüniche-Olsen, A., Westerman, R., Kazmierczyk, Z., Vertyankin, V.V., Godard-Codding, C., Bickham, J.W., and DeWoody, J.A. (2018). The inference of gray whale (*Eschrichtius robustus*) historical population attributes from whole-genome sequences. BMC Evol. Biol. *18*, 87.

Burnstock, G. (2016). P2X ion channel receptors and inflammation. Purinerg. Signal. *12*, 59–67.

Chen, B., Gao, H., Jefferson, T.A., Yi, L., and Yang, G. (2018a). Survival rate and population size of Indo-Pacific humpback dolphins (*Sousa chinensis*) in Xiamen Bay, China. Mar. Mammal Sci. *34*, 1018–1033.

Chen, B., Jefferson, T.A., Wang, L., Gao, H., Zhang, H., Zhou, Y., Xu, X., and Yang, G. (2018b). Geographic variation in pigmentation patterns of Indo-Pacific humpback dolphins (*Sousa chinensis*) in Chinese waters. J. Mammal. *99*, 915–922.

Chen, B., Zheng, D., Yang, G., Xu, X., and Zhou, K. (2009). Distribution and conservation of the Indo-Pacific humpback dolphin in China. Integr. Zool. *4*, 240–247.

Chen, H., Patterson, N., and Reich, D. (2010). Population differentiation as a test for selective sweeps. Genome Res. *20*, 393–402.

Committee on Taxonomy (2020). List of Marine Mammal Species and Subspecies (Society for Marine Mammalogy). www.marinemammalscience.org.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. Bioinformatics *27*, 2156–2158.

Dans, S.L., Koen Alonso, M., Pedraza, S.N., and Crespo, E.A. (2003). Incidental catch of dolphins in trawling fisheries off patagonia, Argentina: can populations persist? Ecol. Appl. *13*, 754–762.

Di Meglio, P., Perera, G.K., and Nestle, F.O. (2011). The multitasking organ: recent insights into skin immune function. Immunity *35*, 857–869.

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

Foote, A.D., Liu, Y., Thomas, G.W.C., Vinař, T., Alföldi, J., Deng, J., Dugan, S., van Elk, C.E., Hunter, M.E., Joshi, V., et al. (2015). Convergent evolution of the genomes of marine mammals. Nat. Genet. *47*, 272–275.

Foote, A.D., Vijay, N., Ávila-Arcos, M.C., Baird, R.W., Durban, J.W., Fumagalli, M., Gibbs, R.A., Hanson, M.B., Korneliussen, T.S., Martin, M.D., et al. (2016). Genome-culture coevolution promotes rapid divergence of killer whale ecotypes. Nat. Commun. *7*, 11693.

Furusho, K., Shibata, T., Sato, R., Fukui, R., Motoi, Y., Zhang, Y., Saitoh, S., Ichinohe, T., Moriyama, M., and Nakamura, S. (2018). Cytidine deaminase enables Toll-like receptor 8 activation by cytidine or its analogs. Int. Immunol. *31*, 167–173.

Graphodatsky, A.S., Trifonov, V.A., and Stanyon, R. (2011). The genome diversity and karyotype evolution of mammals. Mol. Cytogenet. *4*, 22.

Hemami, M.R., Ahmadi, M., Sadegh-Saba, M., and Moosavi, S.M.H. (2018). Population estimate and distribution pattern of Indian Ocean humpback dolphin (*Sousa plumbea*) in an industrialised bay, northwestern Persian Gulf. Ecol. Indic. *89*, 631–638.

Huang, S.L., Karczmarski, L., Chen, J., Zhou, R., Lin, W., Zhang, H., Li, H., and Wu, Y. (2012). Demography and population trends of the largest population of Indo-Pacific humpback dolphins. Biol. Conserv. *147*, 234–242.

Jefferson, T.A., Hung, S.K., and Würsig, B. (2009). Protecting small cetaceans from coastal development: impact assessment and mitigation experience in Hong Kong. Mar. Policy *33*, 305–311.

Jefferson, T.A., and Rosenbaum, H.C. (2014). Taxonomic revision of the humpback dolphins (*Sousa* spp.), and description of a new species from Australia. Mar. Mammal Sci. *30*, 1494–1541.

Jefferson, T.A., and Smith, B.D. (2016). Chapter one - Re-assessment of the conservation status of the Indo-Pacific humpback dolphin (*Sousa chinensis*) using the IUCN red list criteria. In Advances in Marine Biology, T.A. Jefferson and B.E. Curry, eds. (Academic Press), pp. 1–26.

Jefferson, T.A., Smith, B.D., Braulik, G.T., and Perrin, W. (2017). Sousa Chinensis (Errata Version Published in 2018) (The IUCN Red List of Threatened Species). https://doi.org/10.2305/IUCN.UK.2017-3.

Jia, K., Bian, C., Yi, Y., Li, Y., Jia, P., Gui, D., Zhang, X., Lin, W., Sun, X., Lv, Y., et al. (2019). Whole genome sequencing of Chinese White Dolphin (*Sousa chinensis*) for high-throughput screening of antihypertensive peptides. Mar. Drugs *17*, 504.

Jones, S.J.M., Taylor, G.A., Chan, S., Warren, R.L., Hammond, S.A., Bilobram, S., Mordecai, G., Suttle, C.A., Miller, K.M., Schulze, A., et al. (2017). The genome of the Beluga whale (*Delphinapterus leucas*). Genes *8*, 378.

Keane, M., Semeiks, J., Webb, A.E., Li, Y.I., Quesada, V., Craig, T., Madsen, L.B., van Dam, S., Brawand, D., Marques, P.I., et al. (2015). Insights into the evolution of longevity from the bowhead whale genome. Cell Rep *10*, 112–122.

Kim, J., Farré, M., Auvil, L., Capitanu, B., Larkin, D.M., Ma, J., and Lewin, H.A. (2017). Reconstruction and evolutionary history of eutherian chromosomes. Proc. Natl. Acad. Sci. U. S. A *114*, E5379–E5388.

Kirkpatrick, M. (2010). How and why chromosome inversions evolve. PLoS Biol. *8*, e1000501.

Kulemzina, A.I., Trifonov, V.A., Perelman, P.L., Rubtsova, N.V., Volobuev, V., Ferguson-Smith, M.A., Stanyon, R., Yang, F., and Graphodatsky, A.S. (2009). Cross-species chromosome painting in Cetartiodactyla: reconstructing the karyotype evolution in key phylogenetic lineages. Chromosome Res. *17*, 419–436.

Kurihara, N., Tajima, Y., Yamada, T.K., Matsuda, A., and Matsuishi, T. (2017). Description of the karyotypes of Stejneger's beaked whale (*Mesoplodon stejnegeri*) and Hubbs' beaked whale (*M. carlhubbsi*). Genet. Mol. Bio. *40*, 803–807.

Li, H., and Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. Nature *475*, 493–496.

Li, M. (2018). Paleoclimate and Paleoenvironment Evolutions in the Northwestern South China Sea over the Past 36 Kyr BP and the Forcing Mechanisms (Guangzhou Institute of Geochemistry, Chinese Academy of Sciences).

Li, S. (2020). Humpback dolphins at risk of extinction. Science *367*, 1313–1314.

Li, S., Lin, M., Xiao, X., Xing, L., Zhang, P., Gozlan, R.E., Huang, S.L., and Ding, W. (2016). First record of the Indo-Pacific humpback dolphins (*Sousa chinensis*) southwest of Hainan Island, China. Mar. Biodivers. Rec. *9*, 3.

Li, Z., Pospelova, V., Liu, L., Zhou, R., and Song, B. (2017). High-resolution palynological record of Holocene climatic and oceanographic changes in the northern South China Sea. Palaeogeogr. Palaeocl. *483*, 94–124.

Liu, H.X., Huang, J., Sun, X.G., Li, J., Hu, Y.W., Yu, L.Y., Liti, G.N., Tian, D.C., Hurst, L.D., and Yang, S.H. (2018a). Tetrad analysis in plants and fungi finds large differences in gene conversion rates but no GC bias. Nat. Ecol. Evol. *2*, 164–173.

Liu, M.C., Liao, W.Y., Buckley, K.M., Yang, S.Y., Rast, J.P., and Fugmann, S.D. (2018b). AID/APOBEC-like cytidine deaminases are ancient innate immune mediators in invertebrates. Nat. Commun. *9*, 1948.

Mérot, C., Oomen, R.A., Tigano, A., and Wellenreuther, M. (2020). A roadmap for understanding the evolutionary significance of structural genomic variation. Trends Eco. Evol. *35*, 561–572.

Marcelo, K.L., Means, A.R., and York, B. (2016). The Ca2+/Calmodulin/CaMKK2 Axis: nature's metabolic CaMshaft. Trends Endocrin. Met. *27*, 706–718.

Mertens, F., Johansson, B., Fioretos, T., and Mitelman, F. (2015). The emerging complexity of gene fusions in cancer. Nat. Rev. Cancer *15*, 371–381.

Ming, Y., Jian, J., Yu, F., Yu, X., Wang, J., and Liu, W. (2019). Molecular footprints of inshore aquatic adaptation in Indo-Pacific humpback dolphin (*Sousa chinensis*). Genomics *111*, 1034–1042.

Moskalev, A.A., Kudryavtseva, A.V., Graphodatsky, A.S., Beklemisheva, V.R., Serdyukova, N.A., Krutovsky, K.V., Sharov, V.V., Kulakovskiy, I.V., Lando, A.S., Kasianov, A.S., et al. (2017). De novo assembling and primary analysis of genome and transcriptome of gray whale *Eschrichtius robustus*. BMC Evol. Biol. *17*, 258.

Moura, A.E., Kenny, J.G., Chaudhuri, R., Hughes, M.A., Welch, J.A., Reisinger, R.R., de Bruyn, P.N., Dahlheim, M.E., Hall, N., and Hoelzel, A.R. (2014). Population genomics of the killer whale indicates ecotype evolution in sympatry involving both selection and drift. Mol. Ecol. *23*, 5179–5192.

Murphy, W.J., Larkin, D.M., der Wind, A.E.-v., Bourque, G., Tesler, G., Auvil, L., Beever, J.E., Chowdhary, B.P., Galibert, F., Gatzke, L., et al. (2005). Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. Science *309*, 613–617.

Nandakumar, P., Morrison, A.C., Grove, M.L., Boerwinkle, E., and Chakravarti, A. (2018). Contributions of rare coding variants in hypotension syndrome genes to population blood pressure variation. Medicine *97*, e11865.

Ng, S.L., and Leung, S. (2003). Behavioral response of Indo-Pacific humpback dolphin (*Sousa chinensis*) to vessel traffic. Mar. Environ. Res. *56*, 555–567.

Noris, M., and Remuzzi, G. (2013). Overview of complement activation and regulation. Semin. Nephrol. *33*, 479–492.

O'Connor, R.E., Farré, M., Joseph, S., Damas, J., Kiazim, L., Jennings, R., Bennett, S., Slack, E.A., Allanson, E., and Larkin, D.M. (2018a). Chromosome-level assembly reveals extensive rearrangement in saker falcon and budgerigar, but not ostrich, genomes. Genome Biol. *19*, 171.

O'Connor, R.E., Romanov, M.N., Kiazim, L.G., Barrett, P.M., Farré, M., Damas, J., Ferguson-Smith, M., Valenzuela, N., Larkin, D.M., and Griffin, D.K. (2018b). Reconstruction of the diapsid ancestral genome permits chromosome evolution tracing in avian and non-avian dinosaurs. Nat. Commun. *9*, 1883.

Paine-Saunders, S., Viviano, B.L., Zupicich, J., Skarnes, W.C., and Saunders, S. (2000). Glypican-3 controls cellular responses to Bmp4 in limb patterning and skeletal development. Dev. Biol. *225*, 179–187.

Parsons, E.C.M. (1998). Trace metal pollution in Hong Kong: implications for the health of Hong Kong's Indo-Pacific hump-backed dolphins (*Sousa chinensis*). Sci. Total Environ. *14*, 175–184.

Parsons, E.C.M., Overstreet, R.M., and Jefferson, T.A. (2001). Parasites from Indo-Pacific hump-backed dolphins (*Sousa chinensis*) and finless porpoises (*Neophocaena phocaenoides*) stranded in Hong Kong. Vet. Rec. *148*, 776–780.

Pegoraro, C., Pollet, N., and Monsoro-Burq, A.H. (2011). Tissue-specific expression of sarcoplasmic/endoplasmic reticulum calcium ATPases (ATP2A/SERCA) 1, *2*, 3 during Xenopus laevis development. Gene Expr. Patterns *11*, 122–128.

Pigors, M., Sarig, O., Heinz, L., Plagnol, V., Fischer, J., Mohamad, J., Malchin, N., Rajpopat, S., Kharfi, M., Lestringant, G.G., et al. (2016). Loss-of-function mutations in SERPINB8 linked to exfoliative ichthyosis with impaired mechanical stability of intercellular adhesions. Am. J. Hum. Genet. *99*, 430–436.

Poelstra, J.W., Vijay, N., Bossu, C.M., Lantz, H., Ryll, B., Müller, I., Baglione, V., Unneberg, P., Wikelski, M., Grabherr, M.G., et al. (2014). The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. Science *344*, 1410–1414.

Proskuryakova, A., Kulemzina, A., Perelman, P., Makunin, A., Larkin, D., Farré, M., Kukekova, A., Lynn Johnson, J., Lemskaya, N., Beklemisheva, V., et al. (2017). X chromosome evolution in cetartiodactyla. Genes *8*, 216.

Racioppi, L., and Means, A.R. (2012). Calcium/calmodulin-dependent protein kinase kinase 2: roles in signaling and pathophysiology. J. Biol. Chem. *287*, 31658–31665.

Raouf, R., Chabot-Dore, A.-J., Ase, A.R., Blais, D., and Seguela, P. (2007). Differential regulation of microglial P2X4 and P2X7 ATP receptors following LPS-induced activation. Neuropharmacology *53*, 496–504.

Reich, D., Thangaraj, K., Patterson, N., Price, A.L., and Singh, L. (2009). Reconstructing Indian population history. Nature *461*, 489–494.

Rubes, J., Musilova, P., Kopecna, O., Kubickova, S., Cernohorska, H., and Kulemsina, A. (2012). Comparative molecular cytogenetics in Cetartiodactyla. Cytogenet. Genome Res. *137*, 194–207.

Saad, K., Theis, S., Otto, A., Luke, G., and Patel, K. (2017). Detailed expression profile of the six Glypicans and their modifying enzyme, Notum during chick limb and feather development. Gene *610*, 71–79.

Sabeti, P.C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E.H., McCarroll, S.A., Gaudet, R., et al. (2007). Genome-wide detection and characterization of positive selection in human populations. Nature *449*, 913–918.

Turvey, S.T., Pitman, R.L., Taylor, B.L., Barlow, J., Akamatsu, T., Barrett, L.A., Zhao, X., Reeves, R.R., Stewart, B.S., and Wang, K. (2007). First human-caused extinction of a cetacean species? Biol. Lett. *3*, 537–540.

Van de Peer, Y., Maere, S., and Meyer, A. (2009). The evolutionary significance of ancient genome duplications. Nat. Rev. Genet. *10*, 725–732.

Via, S. (2009). Natural selection in action during speciation. Proc. Natl. Acad. Sci. U S A *106*, 9939–9946.

Viricel, A., Pante, E., Dabin, W., and Simon-Bouhet, B. (2014). Applicability of RAD - tag genotyping for interfamilial comparisons: empirical data from two cetaceans. Mol. Eco. Resour. *14*, 597–605.

Wang, J.Y., Chu, S., Fruet, P.F., Daura-Jorge, F.G., and Secchi, E.R. (2012). Mark-recapture analysis of the critically endangered eastern Taiwan strait population of Indo-Pacific humpback dolphins (*Sousa chinensis*): implications for conservation. B. Mar. Sci. *88*, 885–902.

Wang, J.Y., Yang, S.C., and Hung, S.K. (2015). Diagnosability and description of a new subspecies of Indo-Pacific humpback dolphin, *Sousa chinensis* (Osbeck, 1765), from the Taiwan Strait. Zool. Stud. *54*, 36.

Warren, W.C., Kuderna, L., Alexander, A., Catchen, J., Pérez-Silva, J.G., López-Otín, C., Quesada, V., Minx, P., Tomlinson, C., Montague, M.J., et al. (2017). The novel evolution of the sperm whale genome. Genome Biol. Evol. *9*, 3260–3264.

White, M. (1969). Chromosomal rearrangements and speciation in animals. Annu. Rev. Genet. *3*, 75–98.

Wright, S. (1922). Coefficients of inbreeding and relationship. Am. Nat. *56*, 330–338.

Wurster, D.H., and Benirschke, K. (1970). Indian Momtjac, Muntiacus muntiak: a deer with a low diploid chromosome number. Science *168*, 1364–1366.

Xiao, Z., He, L., Takemoto, M., Jalanko, H., Chan, G.C., Storm, D.R., Betsholtz, C., Tryggvason, K., and Patrakka, J. (2011). Glomerular podocytes express type 1 adenylate cyclase: inactivation results in susceptibility to proteinuria. Nephron Exp. Nephrol. *118*, e39–e48.

Xu, X., Song, J., Zhang, Z., Li, P., Yang, G., and Zhou, K. (2015). The world's second largest population of humpback dolphins in the waters of Zhanjiang deserves the highest conservation priority. Sci. Rep. *5*, 8147.

Yim, H.S., Cho, Y.S., Guang, X., Kang, S.G., Jeong, J.Y., Cha, S.S., Oh, H.M., Lee, J.H., Yang, E.C., Kwon, K.K., et al. (2014). Minke whale genome and aquatic adaptation in cetaceans. Nat. Genet. *46*, 88–92.

Yuan, Y., Zhang, P., Wang, K., Liu, M., Li, J., Zheng, J., Wang, D., Xu, W., Lin, M., and Dong, L. (2018). Genome sequence of the freshwater Yangtze finless porpoise. Genes *9*, 213.

Zhang, G., Li, C., Li, Q., Li, B., Larkin, D.M., Lee, C., Storz, J.F., Antunes, A., Greenwold, M.J., and Meredith, R.W. (2014). Comparative genomics reveals insights into avian genome evolution and adaptation. Science *346*, 1311–1320.

Zhang, J., Rubio, V., Lieberman, M.W., and Shi, Z.Z. (2009). OLA1, an Obg-like ATPase, suppresses antioxidant response via nontranscriptional mechanisms. Proc. Natl. Acad. Sci. U S A *106*, 15356–15361.

Zhao, Y., Olonisakin, T.F., Xiong, Z., Hulver, M., Sayeed, S., Yu, M.T., Gregory, A.D., Kochman, E.J., Chen, B.B., Mallampalli, R.K., et al. (2015). Thrombospondin-1 restrains neutrophil granule serine protease function and regulates the innate immune response during Klebsiella pneumoniae infection. Mucosal Immunol. *8*, 896–905.

Zhong, W., Xue, J., Ouyang, J., Zheng, Y., Ma, Q., and Yu, X. (2010). Last glacial climate variations on the tropical Leizhou Peninsula, South China. J. Paleolimnol. *44*, 777–788.

Zhou, X., Guang, X., Sun, D., Xu, S., Li, M., Seim, I., Jie, W., Yang, L., Zhu, Q., Xu, J., et al. (2018a). Population genomics of finless porpoises reveal an incipient cetacean species adapted to freshwater. Nat. Commun. *9*, 1276.

Zhou, X., Seim, I., and Gladyshev, V.N. (2015). Convergent evolution of marine mammals is associated with distinct substitutions in common genes. Sci. Rep. *5*, 16550.

Zhou, X., Sun, D., Guang, X., Ma, S., Fang, X., Mariotti, M., Nielsen, R., Gladyshev, V.N., and Yang, G. (2018b). Molecular footprints of aquatic adaptation including bone mass changes in cetaceans. Genome Biol. Evol. *10*, 967–975.

Zhou, X., Sun, F., Xu, S., Fan, G., Zhu, K., Liu, X., Chen, Y., Shi, C., Yang, Y., Huang, Z., et al. (2013). Baiji genomes reveal low genetic variability and new insights into secondary aquatic adaptations. Nat. Commun. *4*, 2708.

iScience, Volume *23*

Supplemental Information

An Indo-Pacific Humpback Dolphin Genome Reveals Insights into Chromosome Evolution and the Demography of a Vulnerable Species

Peijun Zhang, Yong Zhao, Chang Li, Mingli Lin, Lijun Dong, Rui Zhang, Mingzhong Liu, Kuan Li, He Zhang, Xiaochuan Liu, Yaolei Zhang, Yuan Yuan, Huan Liu, Inge Seim, Shuai Sun, Xiao Du, Yue Chang, Feida Li, Shanshan Liu, Simon Ming-Yuen Lee, Kun Wang, Ding Wang, Xianyan Wang, Michael R. McGowen, Thomas A. Jefferson, Morten Tange Olsen, Josefin Stiller, Guojie Zhang, Xun Xu, Huanming Yang, Guangyi Fan, Xin Liu, and Songhai Li

Supplementary Information

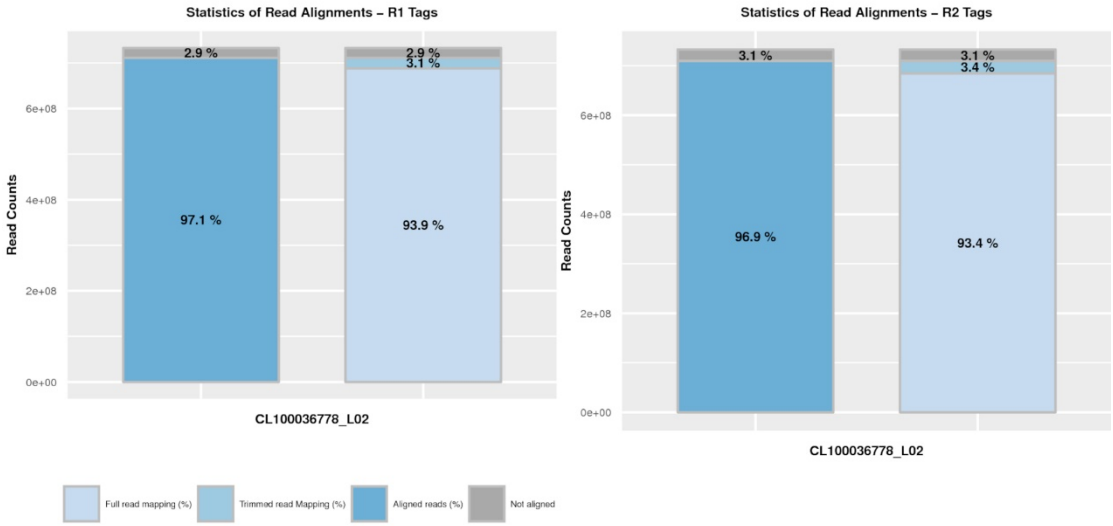


Figure S1. Related to Figure 1. Statistics of Hi-C reads alignments of *S. chinensis* genome. Read1 and read2 mapping rate are 97.1% and 96.9%, respectively.

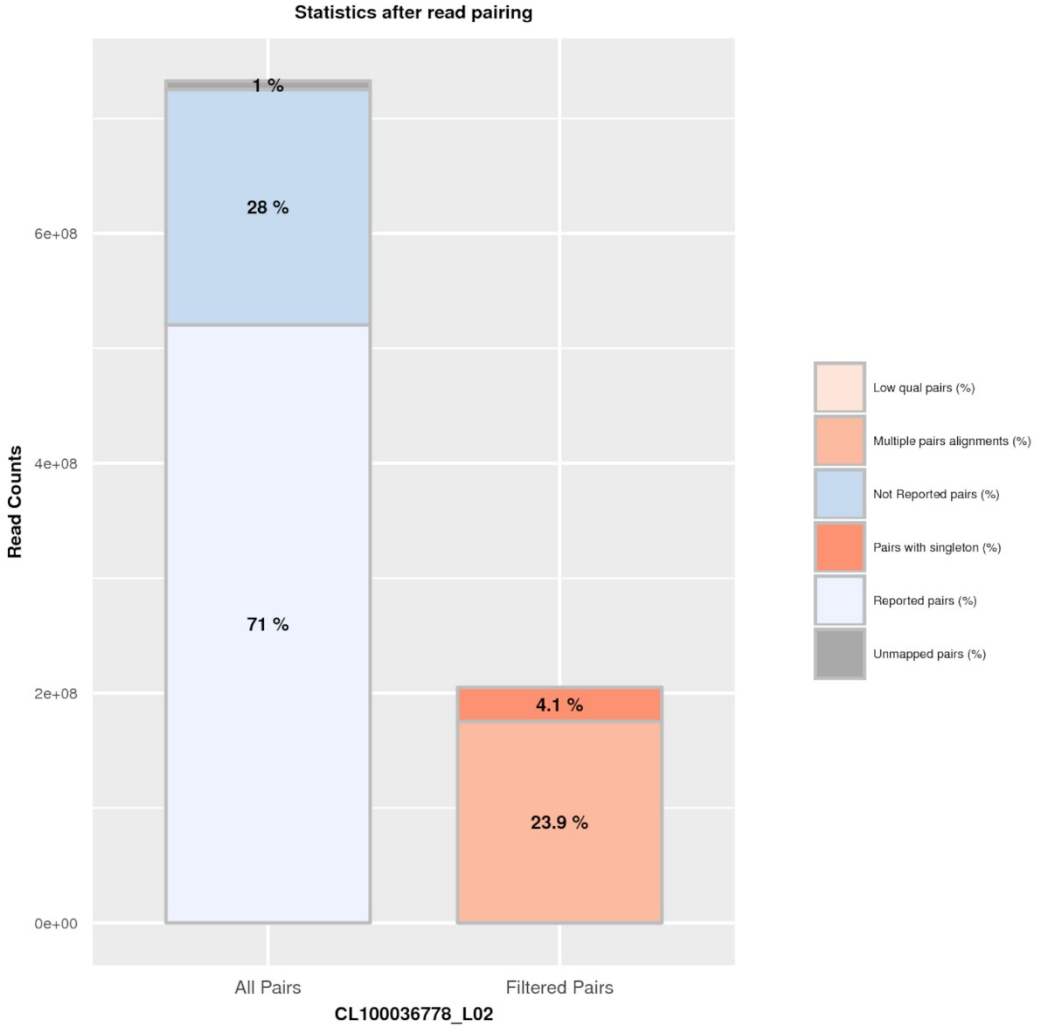


Figure S2. Related to Figure 1. Statistics of read1 and read2 merged result of Hi-C alignment. The ratio of reported unique pairs, total ratio of the multiple and singleton pairs, ratio of unmapped pairs is 71.02%, 27.98%, 1.00%.

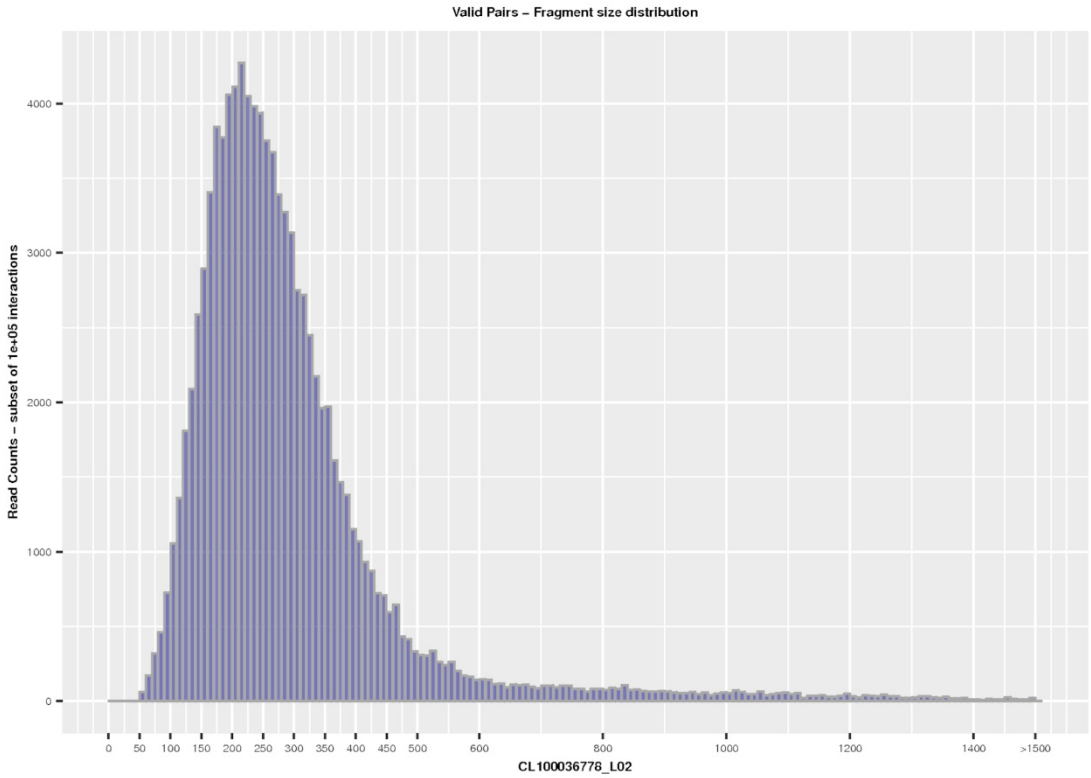


Figure S3. Related to Figure 1. The distribution of Hi-C reads aligned to the restriction fragments. Most of Hi-C reads insert size are distributed between 50-600 bp.

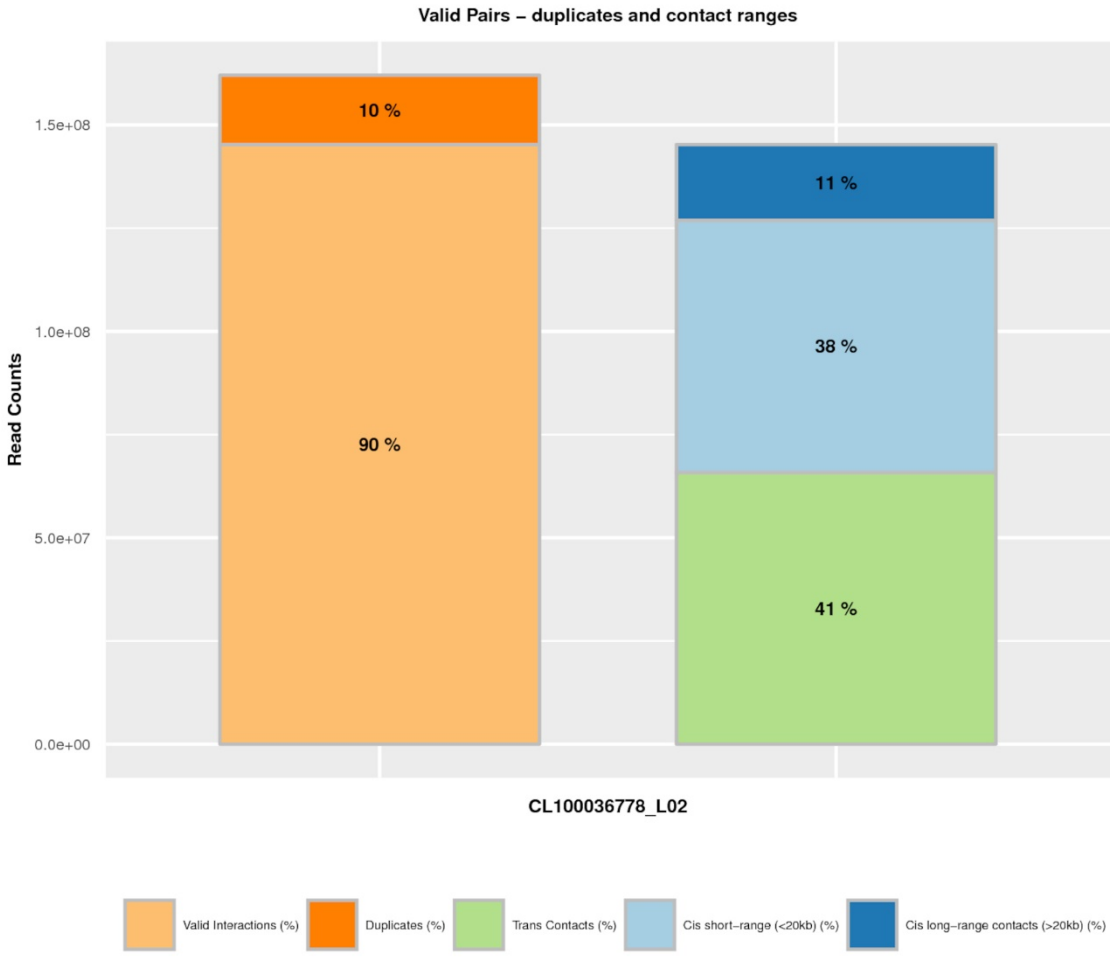


Figure S4. Related to Figure 1. The ratio of duplicated pairs of Hi-C reads. There are only 10% duplicates of in all pairs reads, 90% are valid interactions. 41% are trans contacts, 38% are cis short-range contacts and 11% are cis long-range contacts.

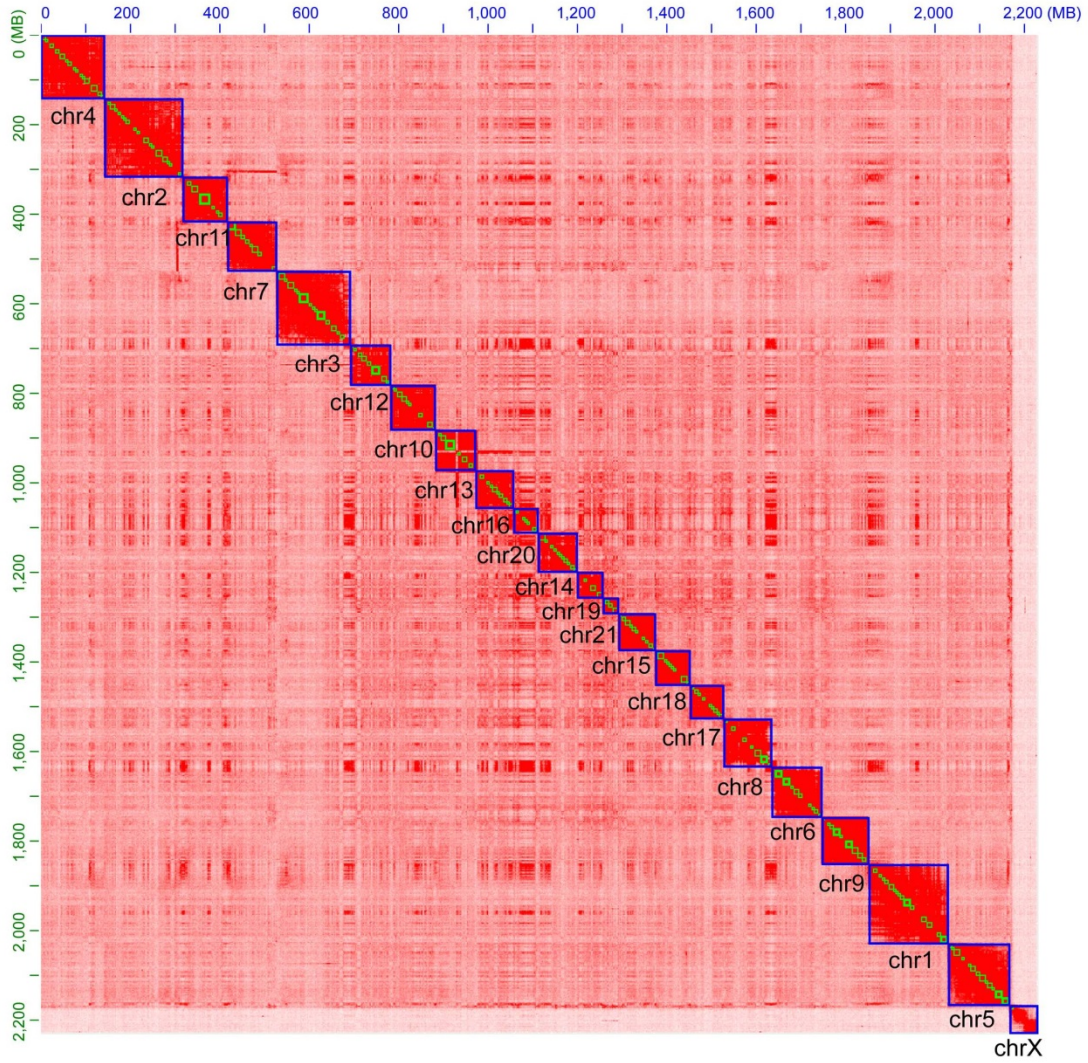


Figure S5. Related to Figure 1. Hi-C contact heatmap of 22 chromosomes of the *S. chinensis*.

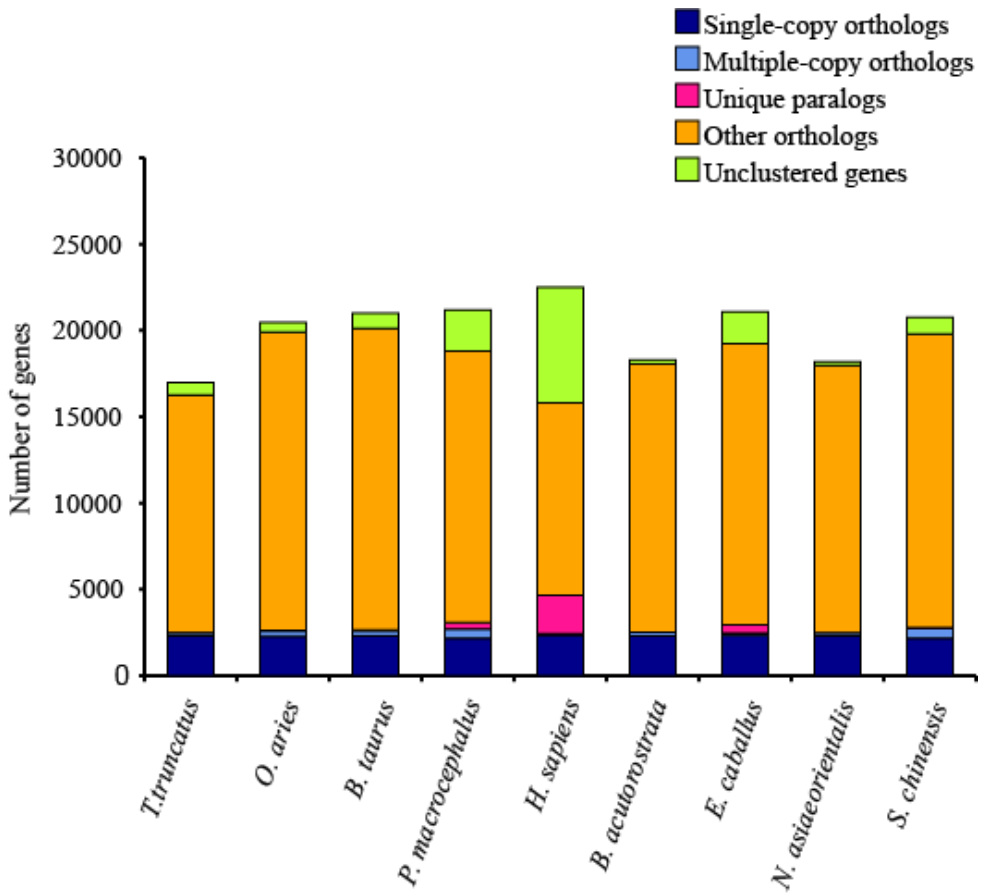


Figure S6. Related to Figure 1. Comparison of the number of orthologous genes in the *S. chinensis* and the other related species.

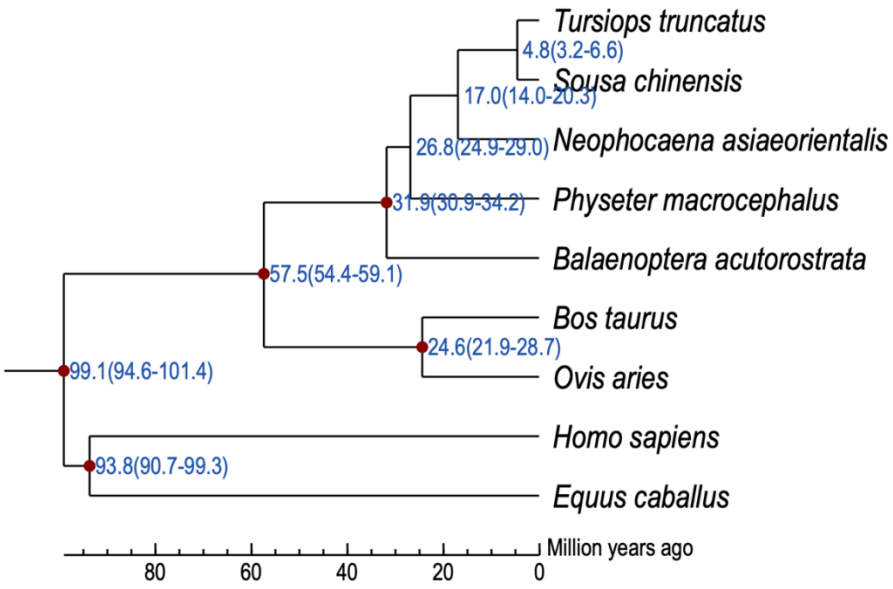


Figure S7. Related to Figure 1. Phylogenetic tree and divergence time of *S. chinensis* and related species.

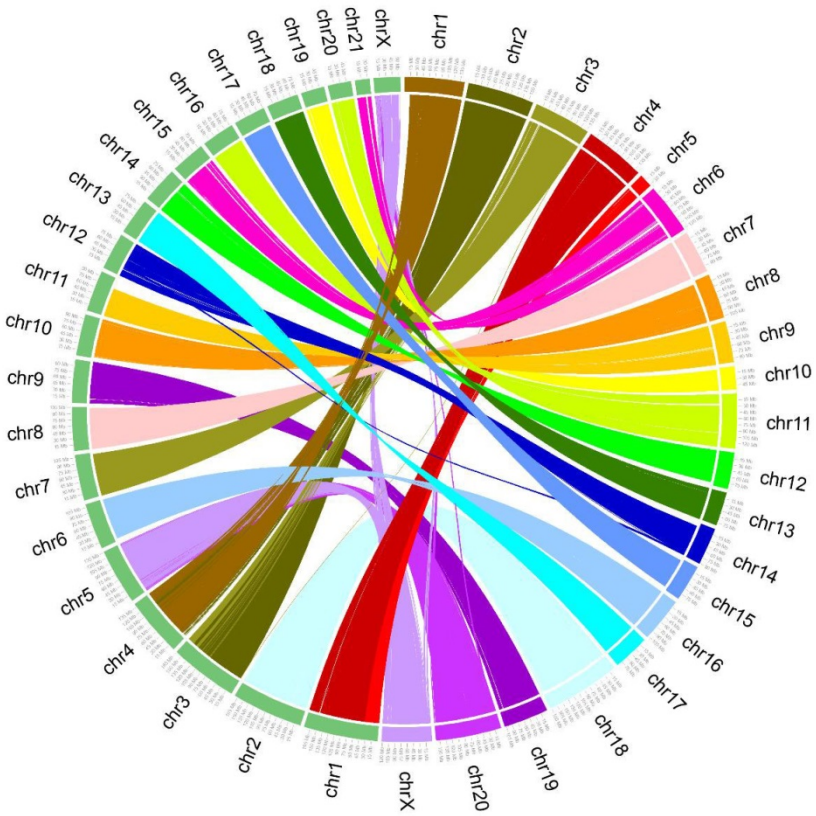


Figure S8. Related to Figure 1. The synteny between *S. chinensis* and sperm whale chromosomes. The left green block presents the 22 *S. chinensis* chromosomes and right various colors block presents 21 sperm whale chromosomes.

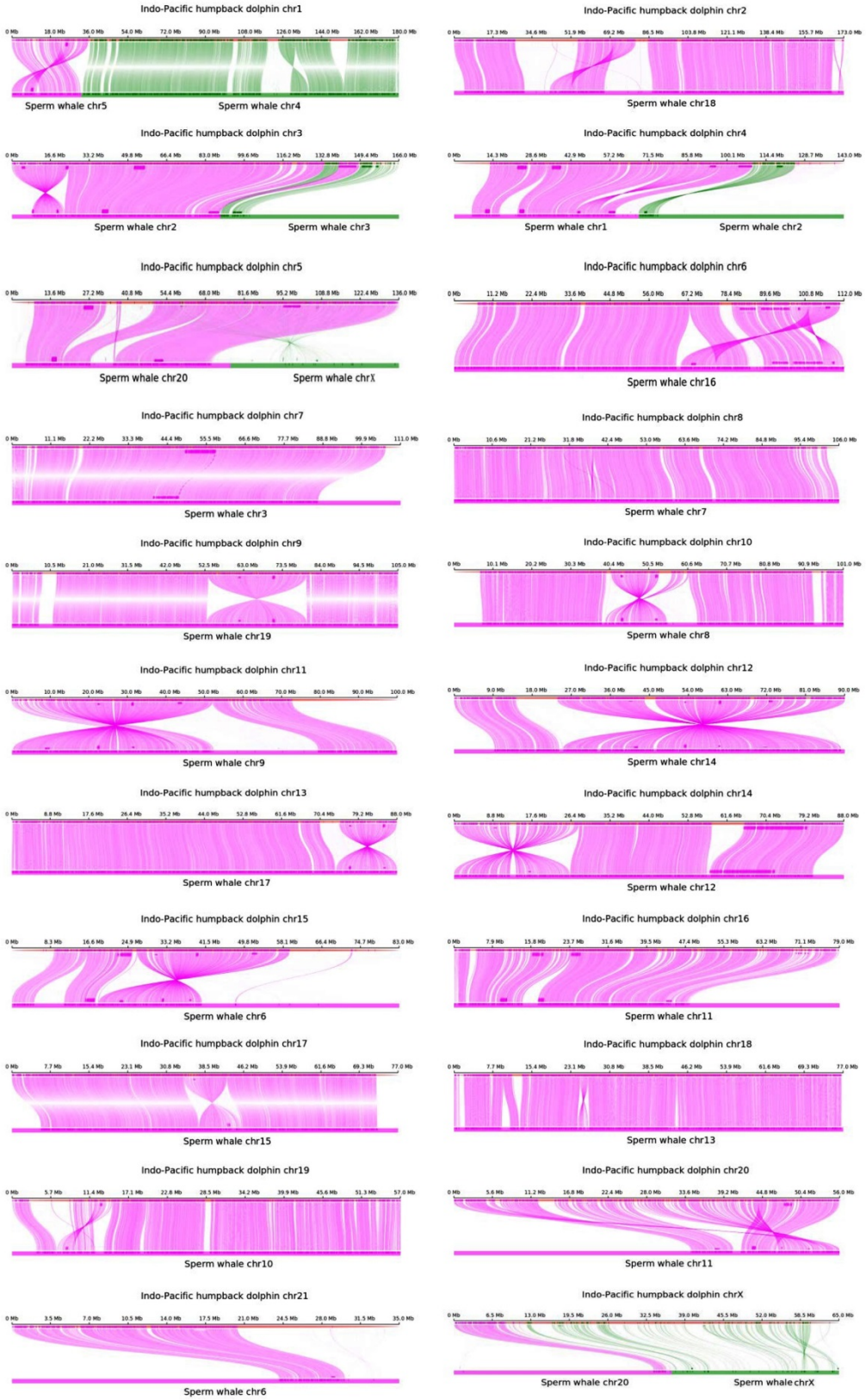


Figure S9. Related to Figure 1. 22 *S. chinensis* chromosomes in synteny with sperm whale chromosomes. The upper red lines represent *S. chinensis* chromosomes, while the lower lines in various colors represent sperm whale chromosomes.

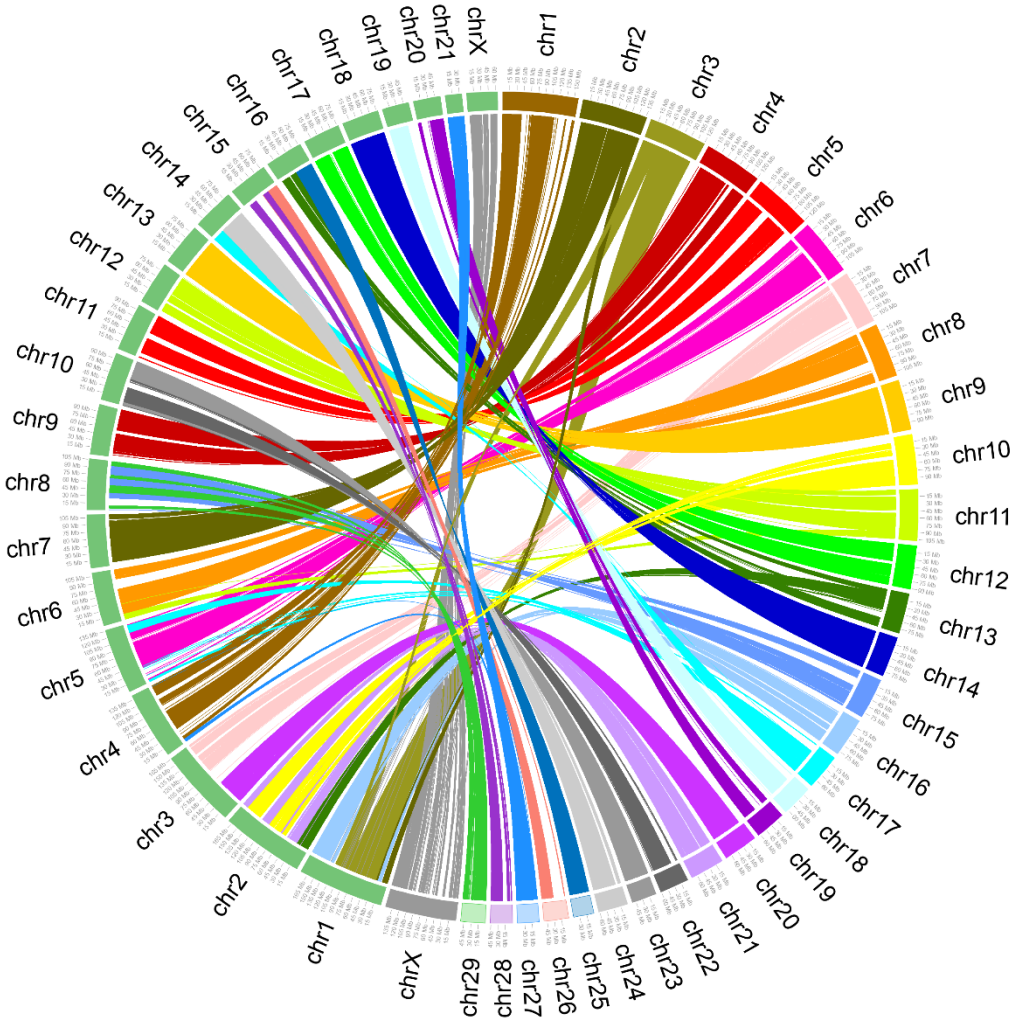


Figure S10. Related to Figure 1. The synteny between *S. chinensis* and cattle chromosomes. The left green blocks represent the 22 *S. chinensis* chromosomes and right various color blocks represent 30 cattle chromosomes.

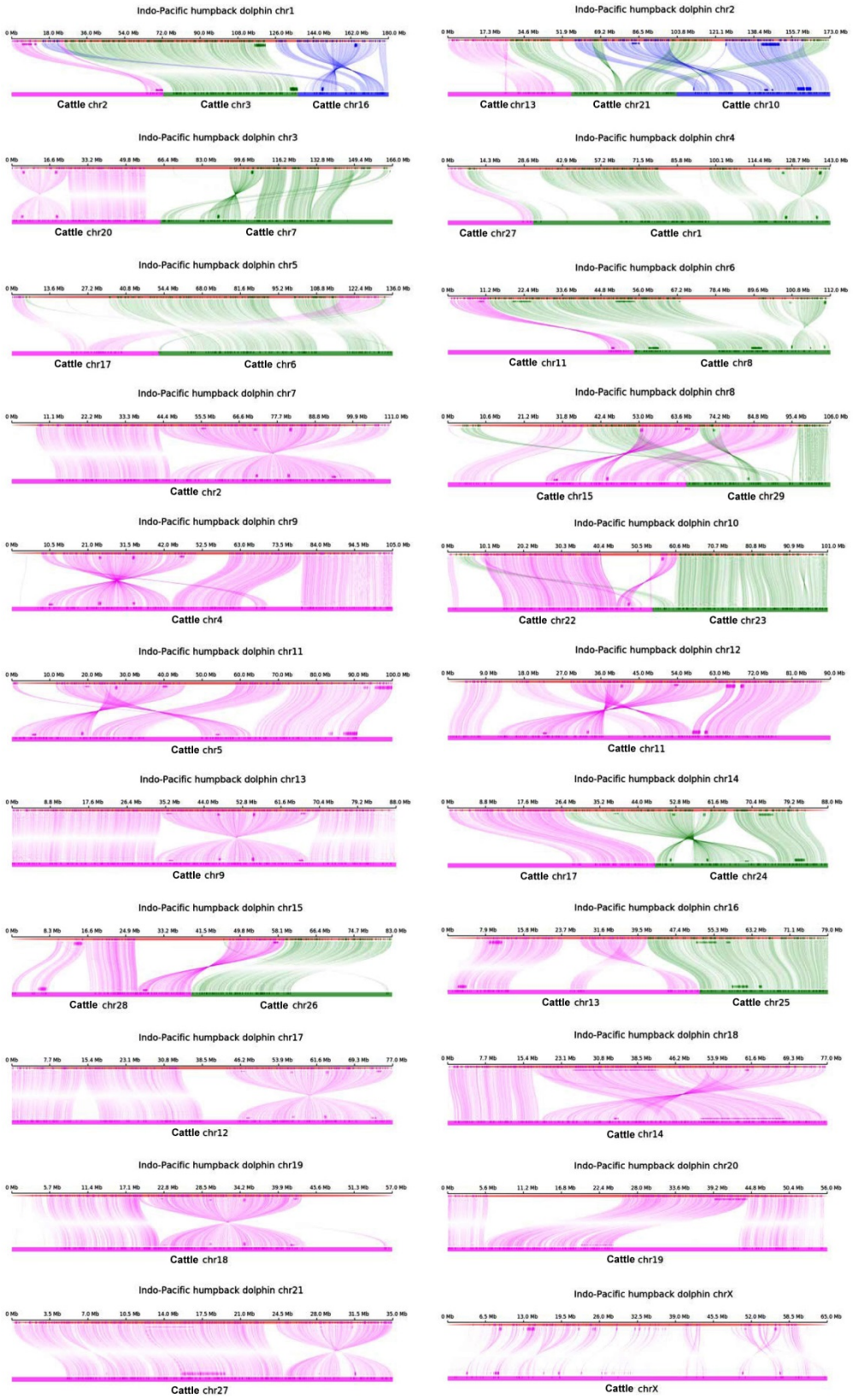


Figure S11. Related to Figure 1. 22 *S. chinensis* chromosomes in synteny with cattle chromosome. The upper red lines *S. chinensis* chromosomes, while the lower lines in various colors represent cattle chromosomes.

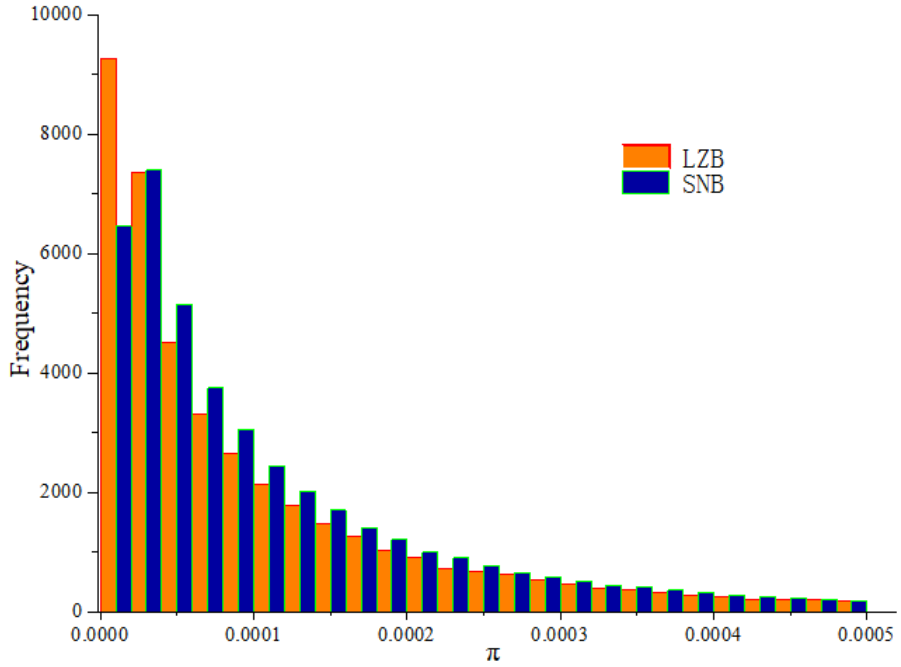


Figure S12. Related to Figure 3. Nucleotide diversity frequency distribution of *S. chinensis* populations from Leizhou Bay and Sanniang Bay. Brown bars represent data from Leizhou Bay while blue bars present data from Sanniang Bay.

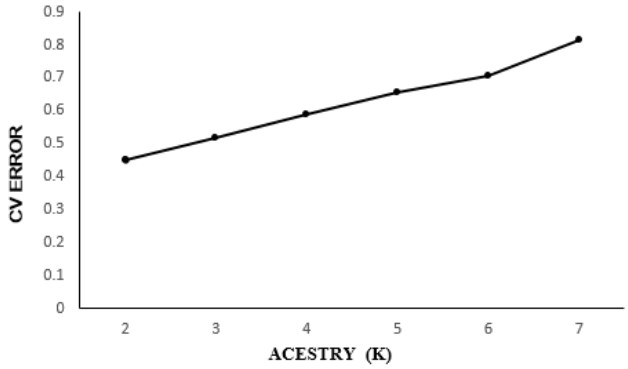


Figure S13. Related to Figure 4. Cross-validation (CV) error for varying values of K in the ADMIXTURE analysis. Minimum of estimated CV error on K=2 suggests the most suitable number of ancestral populations.

Figure S14. Related to Figure 4. Pigmentation pattern difference of Indo-Pacific humpback dolphins in Leizhou Bay and Sanniang Bay. a) Pattern 1: white/pink spots are firstly found on the skin and then enlarged gradually when youth grows old. b) Pattern 2: skin color gradually turns light and finally becomes pink/white when youth grows old.

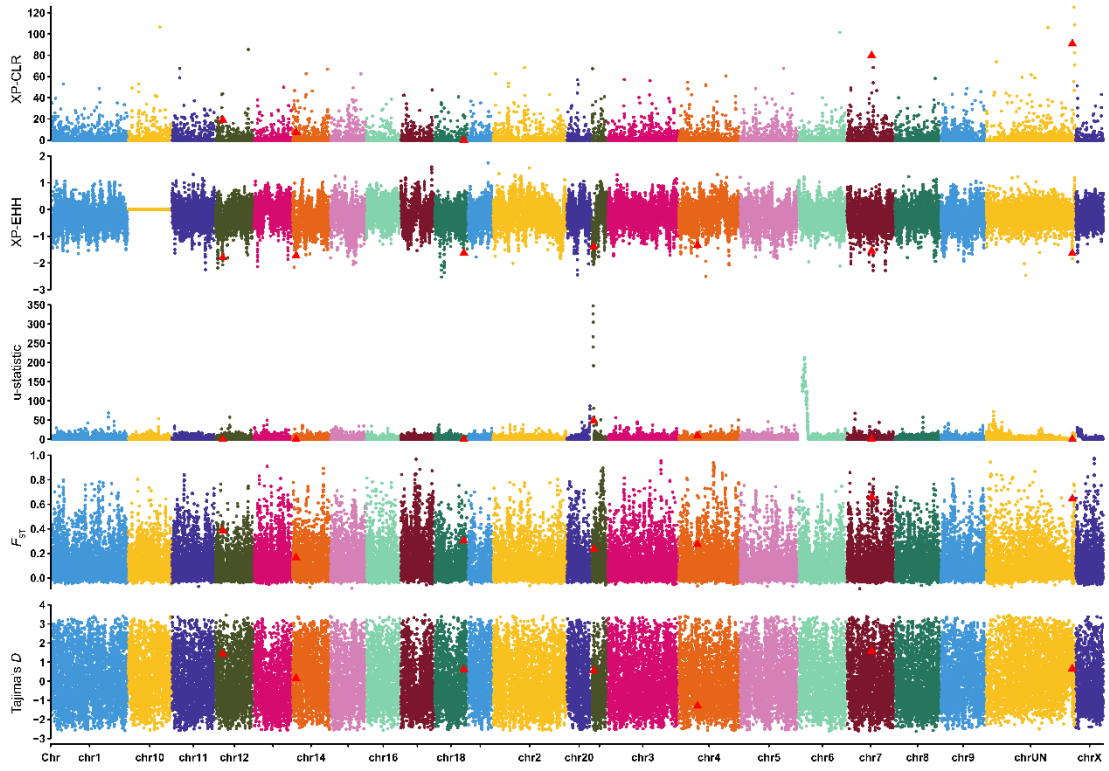


Figure S15. Related to Figure 4. Detection of positive selection of LZB population at genome-wide level. The red triangles indicate the identified selection regions.

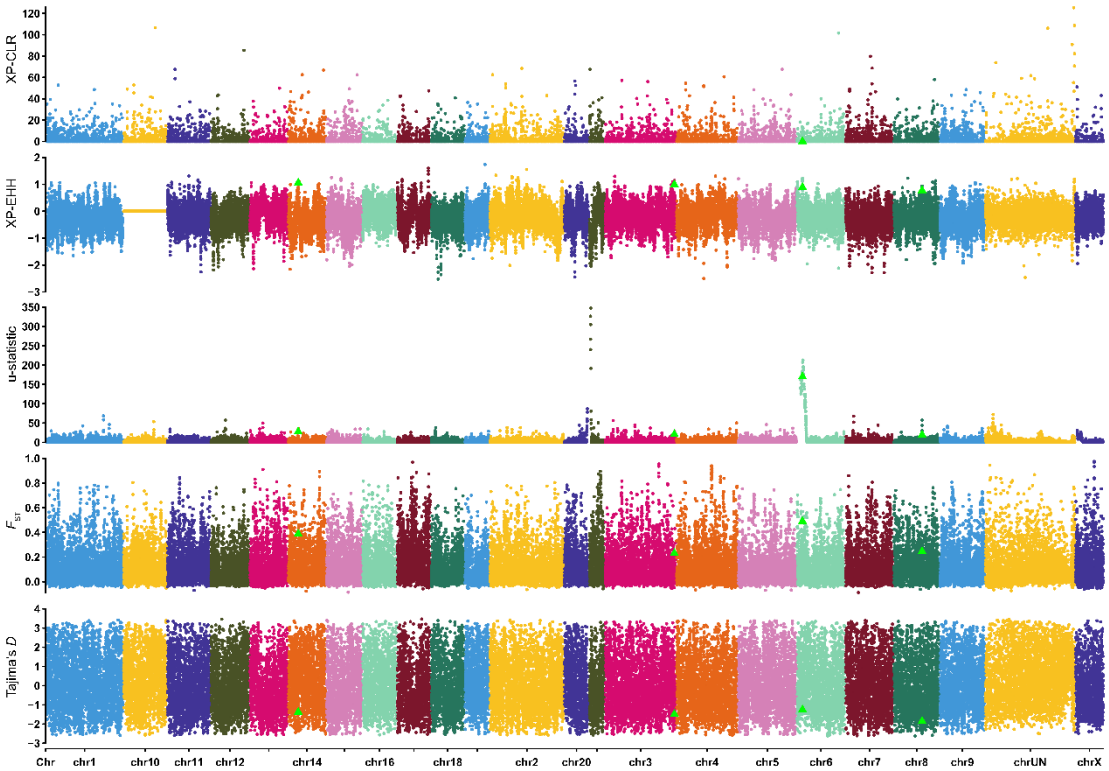


Figure S16. Related to Figure 4. Detection of positive selection of SNB population at genome-wide level. The green triangles indicate the identified selection regions.

Table S1. Related to Figure 1. Summary of the sequencing data obtained for WGS libraries, 10X Genomics Chromium™ library, Hi-C libraries and transcriptome libraries.

Pair-end Libraries	Insert Size	Average Reads Length(bp)	Total Data(Gb)	Sequence Depth(X)
WGS	200-400bp	100	71.592	29
10X Genomics Chromium™	680bp	150	317.499	129
Hi-C	200-400bp	50	73.267	29
Transcriptome	250	100	8.5	-

Table S2. Related to Figure 1. Statistics of assembly after gap closing.

Statistical level	Scaffold	Contig
Total number	76,346	113,866
Total length (bp)	2,458,790,580	2,350,835,813
Gap number (bp)	107,954,767	0
Average length (bp)	32,205.89	20,645.63
N50 Length (bp)	27,695,174	114,025
N90 Length (bp)	619,950	19,971
Maximum length (bp)	106,452,494	1,269,737
Minimum length (bp)	100	48
GC content (%)	41.33	41.33

Table S3. Related to Figure 1. Length of the 22 chromosomes of *S. chinensis* genome.

Chromosome ID	Length (bp)
chr1	179,798,283
chr2	172,887,203
chr3	165,834,538
chr4	143,218,580
chr5	136,344,689
chr6	112,186,916
chr7	111,358,879
chr8	106,383,039
chr9	104,828,680
chr10	101,190,361
chr11	100,036,560
chr12	89,822,583
chr13	88,266,666
chr14	87,587,368
chr15	82,807,222
chr16	79,253,017
chr17	77,431,126
chr18	77,392,672
chr19	56,525,992
chr20	56,438,207
chr21	35,025,918
chrX	64,669,492

Table S4. Related to Figure 1. Comparison of our genome and the previously published *S. chinensis* genomes

Items	Ming et al., 2019	Jia et al., 2019	The present study
Genome size	2.34 Gb	2.6 Gb	2.46 Gb
Scaffold N50	9 Mb	19.2 Mb	27.7 Mb
Contig N50	67 Kb	84.3 Kb	114 Kb
Genome BUSCO	94.3 % (mammalian)	94.5 % (mammalian)	91.9% (mammalian)
Gene BUSCO	95.1% (mammalian)	86.1% (mammalian)	97.3% (mammalian)
Genome coverage	92.31%	88.39%	Reference genome

Table S5. Related to Figure 1. BUSCO evaluation of the *S. chinensis* genome and genes.

	Genome		Gene	
	Number	Percentage (%)	Number	Percentage (%)
Complete BUSCOs	3,772	91.9	3,991	97.3
Complete and single-copy BUSCOs	3,695	90.0	3,890	94.8
Complete and duplicated BUSCOs	77	1.9	101	2.5
Fragmented BUSCOs	154	3.8	65	1.6
Missing BUSCOs	178	4.3	48	1.1
Total BUSCO genes	4,104	100	4,104	100

Table S6. Related to Figure 1. Summary of matchup between *S. chinensis* and sperm whale (*Physeter macrocephalus*) chromosomes.

Chromosome	The optimal blast results in *P. macrocephalus*	Length (bp)	Coverage (%)	The second optimal blast results in *P. macrocephalus*	Length (bp)	Coverage (%)
chr1	chr4	110,177,592	74.1	chr5	22,771,017	15.3
chr10	chr8	76,216,064	96.0	chr14	1,553,268	2.0
chr11	chr9	61,594,447	94.9	chr21	503,872	0.8
chr12	chr14	66,116,983	91.4	chr3	4,299,851	5.9
chr13	chr17	74,772,291	91.1	chr18	4,886,887	6.0
chr14	chr12	65,512,022	97.6	chr21	383,790	0.6
chr15	chr6	45,436,031	96.9	chr21	382,887	0.8
chr16	chr11	57,721,150	98.5	chr21	178,682	0.3
chr17	chr15	65,536,405	94.5	chr10	1,528,016	2.2
chr18	chr13	65,869,194	97.0	chr21	376,454	0.6
chr19	chr10	40,375,616	94.1	chr6	1,234,108	2.9
chr2	chr18	122,968,524	91.2	chr11	2,840,736	2.1
chr20	chr11	36,524,106	96.4	chr12	348,299	0.9
chr21	chr6	17,062,791	92.7	chr21	249,361	1.4
chrX	chrX	12,948,533	55.1	chr20	7,954,072	33.9
chr3	chr2	124,123,628	76.1	chr3	14,995,279	9.2
chr4	chr1	88,408,478	65.7	chr2	15,552,910	11.5
chr5	chr20	104,644,775	74.7	chr21	12,821,859	9.2
chr6	chr16	86,528,512	91.1	chr15	4,141,016	4.4
chr7	chr3	92,287,627	94.8	chrX	1,036,599	1.1
chr8	chr7	85,967,770	93.1	chr8	2,197,731	2.4
chr9	chr19	89,504,942	96.4	chr21	701,122	0.8

Table S7. Related to Figure 1. Summary of collinearity between *S. chinensis* and cattle (*Bos taurus*) chromosomes.

Chromosome	The optimal blast results in *B. taurus*	Length (bp)	Coverage (%)	The second optimal blast results in *B. taurus*	Length (bp)	Coverage (%)
chr1	chr3	65,056,851	57.2	chr16	36,911,969	32.4
chr10	chr23	30,017,618	50.2	chr22	28,320,227	47.4
chr11	chr5	50,405,868	97.0	chrX	220,862	0.4
chr12	chr11	47,904,601	96.3	chr8	519,458	1.0
chr13	chr9	59,648,962	97.8	chrX	131,470	0.2
chr14	chr24	34,479,960	63.8	chr17	18,354,426	34.0
chr15	chr28	19,803,289	52.3	chr26	16,906,071	44.6
chr16	chr25	25,326,201	51.6	chr13	22,621,722	46.1
chr17	chr12	44,737,217	97.4	chrX	161,480	0.4
chr18	chr14	48,305,537	97.3	chrX	198,031	0.4
chr19	chr18	25,879,342	95.7	chrX	137,576	0.5
chr2	chr10	49,020,914	47.0	chr21	35,638,731	34.2
chr20	chr19	24,191,537	97.0	chr7	67,827	0.3
chr21	chr27	22,540,472	94.6	chr21	483,791	2.0
chrX	chrX	14,226,322	87.2	chr7	152,519	0.9
chr3	chr20	38,203,316	46.3	chr7	37,825,273	45.8
chr4	chr1	56,810,981	90.6	chr27	3,831,858	6.1
chr5	chr6	49,710,506	76.6	chr17	13,480,201	20.8
chr6	chr8	50,084,201	83.9	chr11	7,709,543	12.9
chr7	chr2	67,826,831	91.3	chr3	4,602,624	6.2
chr8	chr15	32,350,228	56.2	chr29	23,605,291	41.0
chr9	chr4	58,840,842	96.8	chr22	547,719	0.9

Table S8. Related to Figure 1. Inferred length of ancestral chromosomes of Cetruminantia.

Chromosome	Length (bp)
chr1	154,474,779
chr2	61,905,470
chr3	98,056,067
chr4	29,858,164
chr5	155,821,731
chr6	7,924,970
chr7	220,874,184
chr8	25,091,812
chr9	26,407,148
chr10	56,103,623
chr11	115,978,101
chr12	31,739,070
chr13	172,675,727
chr14	57,310,100
chr15	68,579,386
chr16	92,089,847
chr17	86,440,286
chr18	48,023,814
chr19	24,720,582
chr20	37,358,162
chr21	72,722,789
chr22	17,467,707
chr23	52,362,696
chr24	42,316,857
chr25	66,412,430
chr26	83,725,353
chr27	60,005,105
chr28	51,879,593

Table S9. Related to Figure 1. Inferred length of ancestral chromosomes of Bovidae.

Chromosome	Length (bp)
chr1	65,197,042
chr2	57,517,106
chr3	92,279,129
chr4	70,751,086
chr5	25,091,812
chr6	52,576,165
chr7	56,103,623
chr8	121,108,196
chr9	120,273,145
chr10	87,754,165
chr11	72,761,895
chr12	49,206,777
chr13	57,310,100
chr14	28,354,261
chr15	92,089,847
chr16	86,440,286
chr17	72,744,396
chr18	43,875,462
chr19	71,404,992
chr20	47,203,079
chr21	131,008,674
chr22	40,225,125
chr23	59,255,844
chr24	41,246,636
chr25	42,316,857
chr26	66,412,430
chr27	83,725,353
chr28	60,005,105
chr29	51,879,593
chr30	72,207,372

Table S10. Related to Figure 1. Inferred length of ancestral chromosomes of Odontoceti.

Chromosome	Length (bp)
chr1	154,474,779
chr2	98,056,067
chr3	185,679,895
chr4	7,924,970
chr5	220,874,184
chr6	66,338,448
chr7	56,103,623
chr8	126,790,445
chr9	57,310,100
chr10	68,579,386
chr11	92,089,847
chr12	49,206,777
chr13	86,440,286
chr14	37,358,162
chr15	72,722,789
chr16	151,814,743
chr17	90,340,671
chr18	66,412,430
chr19	83,725,353
chr20	86,412,253
chr21	97,764,875
chr22	61,905,470

Table S17. Related to Figure 2.The function of genes located in 4 kb regions spanning 40 evolution breakpoints in ancestral Odontoceti compared to ancestral chromosomes of Cetruminantia.

Gene	Symbol	Identity	Descriptions
Sochi10755	*GPC3*	84 %	Glypican-3 precursor; Cell surface proteoglycan that bears heparan sulfate. Plays a role in limb patterning and skeletal development by controlling the cellular response to BMP4. Modulates the effects of growth factors BMP2, BMP7 and FGF7 on renal branching morphogenesis.

			Required for coronary vascular development.
Sochi01840	*TMEM9*	84 %	Transmembrane protein 9 precursor
Sochi03223	*CLDN12*	97 %	Claudin-12; Plays a major role in tight junction-specific obliteration of the intercellular space, through calcium-independent cell-adhesion activity
Sochi02981	*ADCY1*	96 %	Adenylate cyclase type 1; Catalyzes the formation of the signaling molecule cAMP in response to G-protein signaling. Mediates responses to increased cellular Ca(2+)/calmodulin levels. May be involved in regulatory processes in the central nervous system. May play a role in memory and learning. Plays a role in the regulation of the circadian rhythm of daytime contrast sensitivity probably by modulating the rhythmic synthesis of cyclic AMP in the retina.
Sochi04817	*DLGAP1*	49 %	Disks large-associated protein 1, Part of the postsynaptic scaffold in neuronal cells
Sochi08789	*SLC12A3*	95 %	Solute carrier family 12 member 3. Target of thiazide diuretics used in the treatment of high blood pressure.

Sochi10119	*PTGDR*	88 %	Prostaglandin D2 receptor; Receptor for prostaglandin D2 (PGD2). The activity of this receptor is mainly mediated by G(s) proteins that stimulate adenylate cyclase, resulting in an elevation of intracellular cAMP.
Sochi10473	*VRK1*	90 %	Serine/threonine-protein kinase VRK1
Sochi15252	*IL9R*	76 %	The protein encoded by this gene is a cytokine receptor that specifically mediates the biological effects of interleukin 9 (IL9). Genetic studies suggested an association of this gene with the development of asthma.
Sochi06544	*RTN4R*	88 %	Reticulon 4 Receptor, this receptor mediates axonal growth inhibition and may play a role in regulating axonal regeneration and plasticity in the adult central nervous system

Table S18. Related to Figure 2. The function of genes located in 4 kb regions spanning 138 evolution breakpoints in *S. chinensis* compared to ancestral chromosomes of Odontoceti.

Gene	Symbol	Identity	Descriptions
Sochi00142	*CDA*	100%	Forms a homotetramer that catalyzes the irreversible hydrolytic deamination of cytidine and deoxycytidine to uridine and deoxyuridine, respectively.
Sochi06838	*SERPINB8*	87 %	Serpin B8

Sochi10473	*VRK1*	90 %	Serine/threonine-protein kinase VRK1
Sochi17998	*IMPDH2*	94 %	Inosine-5'-monophosphate dehydrogenase 2; plays an important role in the regulation of cell growth. Could also have a single-stranded nucleic acid-binding activity and could play a role in RNA and/or DNA metabolism. It may also have a role in the development of malignancy and the growth progression of some tumors
Sochi18043	*NAT6*	90 %	NAT6 is a putative tumour suppressor, and plays a critical role in the maturation of actins by carrying out the acetylation of their N-terminal acidic residue
Sochi18044	*HYAL2*	41 %	Hyaluronidase-2; Hydrolyzes high molecular weight hyaluronic acid to produce an intermediate-sized product which is further hydrolyzed by sperm hyaluronidase to give small oligosaccharides. Displays very low levels of activity. Associates with and negatively regulates MST1R

Table S20. Related to Figure 3. The best parameter to fit the model 'sym_mig_size' of Dadi.

Parameter	Maximum likelihood	95% CI
θ	4091.59	3966.65-4216.53
nu1a	1.3061	2761.95-3118.13
nu2a	6.7683	10162.28-20306.52
T1	0.356	38058.90-42170.53
nu1b	0.2718	581.16-643.94
nu2b	0.2168	463.56-513.64
T2	0.0341	3645.15-4038.95
m	16.6216	3.51E-3-3.89E-3

Table S22. Related to Figure 3. Heterozygosity of *S. chinensis* and related species.

Species	heterozygosity	μ	Samples	g(years)
Human[10,12]	0.00069	2.35×10^{-8}	--	25
Sperm whale[13]	0.0011	--	4	--
Blue whale[14]	~0.002	1.39×10^{-8}	--	30.8
Fin whale[15]	0.00151	2.77×10^{-8}	--	25.9
Minke whale[15]	0.00061	2.75×10^{-8}	4	30.8
Bottlenose dolphin[15]	0.00142	2.54×10^{-8}	--	21.0
Killer whale	0.00074	1.1×10^{-8}	1	27
Finless porpoise[15]	0.00086	1.1×10^{-8}	48	16.5
Baiji[16]	0.000121	1.5×10^{-8}	3	12
S. chinensis (the present study)	0.000179	5.81×10^{-9}	39	12

Table S23. Related to Figure 3. Inbreeding coefficients ($F_{h)}$ of *S. chinensis* in Sanniang Bay.

Individual	O(HOM)	E(HOM)	N(NM)	F
SNB-1B	818826	8.87E+05	1292005	-0.1683
SNB-2B	781121	8.87E+05	1292005	-0.2614
SNB-3B	811280	8.87E+05	1292005	-0.1869
SNB-4B	811515	8.87E+05	1292005	-0.1863
SNB-5B	830267	8.87E+05	1292005	-0.14
SNB-6A	792275	8.87E+05	1292005	-0.2338
SNB-saA	713894	8.87E+05	1292005	-0.4274
SNB-sjA	780633	8.87E+05	1292005	-0.2626
SNB-ucA	735015	8.87E+05	1292005	-0.3752

Table S24. Related to Figure 3. Inbreeding coefficients ($F_{h)}$ of *S. chinensis* in Leizhou Bay.

Individual	O(HOM)	E(HOM)	N(NM)	F
LZB-01B	692943	7.34E+05	1133826	-0.1032
LZB-02B	697368	7.34E+05	1134028	-0.09245
LZB-03B	637037	7.35E+05	1134321	-0.2438
LZB-04B	672191	7.34E+05	1133995	-0.1553
LZB-05B	666780	7.35E+05	1134601	-0.1698
LZB-06B	634238	7.34E+05	1133462	-0.2495
LZB-08B	647218	7.34E+05	1134079	-0.218
LZB-09B	645475	7.34E+05	1134213	-0.2225
LZB-10B	630991	7.34E+05	1133972	-0.2584
LZB-11B	666563	7.34E+05	1133423	-0.1685
LZB-12B	660018	7.34E+05	1133661	-0.1852
LZB-13B	665687	7.34E+05	1133085	-0.1702
LZB-14B	657890	7.34E+05	1133210	-0.1898
LZB-15B	637694	7.34E+05	1133300	-0.2405
LZB-16B	693313	7.35E+05	1134274	-0.103
LZB-17B	643846	7.35E+05	1134486	-0.227
LZB-18B	623691	7.35E+05	1134643	-0.2776
LZB-19B	669735	7.34E+05	1133754	-0.1612
LZB-20B	645513	7.34E+05	1134172	-0.2224
LZB-21B	650587	7.34E+05	1134182	-0.2097
LZB-22B	655485	7.34E+05	1134000	-0.1972
LZB-23B	641979	7.34E+05	1133908	-0.2308
LZB-24B	630906	7.34E+05	1133922	-0.2584
LZB-25B	666698	7.34E+05	1133603	-0.1684
LZB-26B	678044	7.34E+05	1133030	-0.1391
LZB-27B	627112	7.34E+05	1134202	-0.2684
LZB-28B	684419	7.34E+05	1133826	-0.1245
LZB-29B	632192	7.34E+05	1133258	-0.2542
LZB-30A	640436	7.32E+05	1131198	-0.2304
LZB-33A	626180	7.34E+05	1134100	-0.2706

Table S25. Related to Figure 4. *f*3 statistic of Indo-Pacific populations between Leizhou Bay and Sanniang Bay.

Source1	Source2	Target	*f*3	std.err	Z	SNPs
SNB	REF	LZB	0.107176	0.005478	19.565	2550599
LZB	REF	SNB	0.062828	0.005043	12.459	2550599
LZB	SNB	REF	0.614164	0.017815	34.475	2550599

Table S26. Related to Figure 4. Putative regions of positive selection in the Leizhou Bay and Sanniang Bay populations.

Population	Region	Window_chr	Window_start (bp)	Window_end (bp)	F_{ST}	Tajima's *D*	XP-CLR	XP-EHH	μ-statistic	Gene	Gene function annotated by KEGG ortholog database
LZB	chr12:17400001-17450000	chr12	17,400,001	17,450,000	0.39	-0.50	NA	1.80(*)	34.00 (*)	NA	NA
	chr14:8150001-8450000	chr14	8,150,001	8,200,000	0.17	-0.36	1.87	1.73(*)	32.01 (*)	Sochi06245	K07359|1|0.0|1050|lve:103090099|calcium/calmodulin-dependent protein kinase kinase 2 [EC:2.7.11.17], CAMKK2
										Sochi06246	K05218|1|0.0|743|bacu:103001338|P2X purinoceptor 4, P2RX4
										Sochi06247	NA
		chr14	8,250,001	8,300,000	0.24	-0.45	NA	1.36(*)	29.10 (*)	Sochi06248	K05220|1|0.0|1149|lve:103091614|P2X purinoceptor 7, P2RX7
		chr14	8,350,001	8,400,000	0.35	-1.44	NA	1.63(*)	23.76 (*)	Sochi06249	K19677|1|0.0|1365|lve:103068493|intraflagellar transport protein 81 (IFT81)
		chr14	8,400,001	8,450,000	0.38	-0.89	NA	1.46(*)	21.55 (*)	Sochi06250	K05853|1|0.0|2058|ecb:791242|P-type Ca2+ transporter type 2A [EC:7.2.2.10], ATP2A
	chr18:69200001-69250000	chr18	69,200,001	69,250,000	0.30	-0.47	NA	1.62(*)	31.02 (*)	NA	NA
	chr21:4000001-4050	chr21	4,000,001	4,050,00	0.23	-0.66	NA	1.40	28.30	NA	NA

	000			0				(*)	(*)		
	chr4:44150001-44300000	chr4	44,150,001	44,200,000	0.36	-1.36	NA	1.42 (*)	41.75 (*)	NA	NA
		chr4	44,200,001	44,250,000	0.39	-1.09	NA	1.44 (*)	77.22 (*)	NA	NA
		chr4	44,250,001	44,300,000	0.27	-0.84	NA	1.34 (*)	96.38 (*)	NA	NA
	chr7:58200001-58250000	chr7	58,200,001	58,250,000	0.66	-1.56	29.43	1.55 (*)	0.39	Sochi17249	K19788|1|0.0|837|pps:100990203|obg-like ATPase 1, OLA1
	chrUN:204100001-204150000	chrUN	204,100,001	204,150,000	0.64	-0.09	48.03(*)	1.64 (*)	0.01	Sochi10448	K18401|1|6e-27|126|acs:100558460|KAT8 regulatory NSL complex subunit 2, KANSL2
SNB	chr14:22800001-22850000	chr14	22,800,001	22,850,000	0.39	-1.37	NA	1.05 (*)	27.58 (*)	Sochi06427	K19944|1|6e-158|462|gfr:102042531|TBC1 domain family member 10, TBC1D10A
										Sochi06428	K19944|1|0.0|977|bacu:103016384|TBC1 domain family member 10, TBC1D10A
	chr3:162750001-162800000	chr3	162,750,001	162,800,000	0.23	-1.49	NA	0.99 (*)	20.62 (*)	NA	NA
	chr6:12400001-12450000	chr6	12,400,001	12,450,000	0.49	-1.25	0	0.89 (*)	170.00 (*)	Sochi14780	K16770|1|0.0|4399|bacu:103008830|centriolin, CNTRL
										Sochi14781	K03994|1|0.0|3304|lve:103085597|complement component 5, C5
	chr8:66150001-66200000	chr8	66,150,001	66,200,000	0.25	-1.85	NA	0.77 (*)	19.28 (*)	Sochi14048	K16857|1|1e-29|132|aqu:100632678|thrombospondin 1, THBS1

Note: an asterisk in parentheses indicates a significance level of 0.01

Transparent Methods

1 Samples and Genome sequencing

High-quality DNA was extracted from a blood sample of an adult male Indo-Pacific humpback dolphin (*S. chinensis*) in captivity, which was originally from Thai waters. We constructed a paired-end library (with insert-size of 200-400 bp), a 10X Genomics Chromium library, a Hi-C library, and a transcriptome library. All DNA libraries were sequenced on the BGISEQ-500 sequencing platform. Re-sequencing of 39 *S. chinensis* individuals (30 from LZB and 9 from SNB) was also performed on the BGISEQ-500 sequencing platform (100 bp paired-end reads, ~30× coverage).

1.1 DNA, RNA extraction, and library construction

DNA was extracted from blood of an adult male *S. chinensis* using the cetyl trimethylammonium bromide (CTAB) method. For the short insert size WGS libraries, DNA was sheared into fragments between 50 bp and 800 bp in size using a Covaris E220 ultrasonicator (Covaris, Brighton, UK) according to the manufacturer's instructions. Fragments between 200 to 400 bp were selected using AMPure XP beads (Agencourt, Beverly, USA) and then repaired using T4 DNA polymerase (ENZYMATICS, Beverly, USA) 30 min. at 20 °C to obtain blunt ends which were then 3'-adenlyated to create sticky ends. These DNA fragments were ligated at both ends to T-tailed adapters and amplified for eight cycles using KAPA HiFi HotStart ReadyMix (Kapa Biosystems, Wilmington, USA). The temperature profile was 3 min. at 95 °C followed by 8 cycles of 20 sec. at 98 °C, 15 sec. at 60 °C, 30 sec. at 72 °C, and more 10 min. at 72 °C for further elongation. These amplification products were subjected to a single-strand circularization process using T4 DNA Ligase, (ENZYMATICS, Beverly, USA) 60 min. at 20 °C to generate a single-stranded circular DNA library.

RNA was extracted using TRIzol Reagent (manufacturer). Blood cells (5×10^6 to 1×10^7) were collected by centrifugation and lysed in 1 ml of TRIzol Reagent by repetitive pipetting. After incubation of the homogenized samples for 5 min. at room temperature to permit the complete dissociation of nucleoprotein complexes, 0.2 ml of

chloroform was added per 1 ml of TRIzol Reagent. The samples were mixed intensely and then centrifuged at 12,000 × *g* for 15 min. at 4 °C. Centrifugation separated the biphasic mixtures into the lower red, phenol-chloroform phase and the upper colorless, aqueous phase. The RNA was precipitated from the aqueous phase by mixing with 0.5 ml of isopropanol. The samples were incubated at room temperature for 10 min. and centrifuged at 12,000 × *g* for 10 min. at 4 °C. The supernate was removed and the RNA pellet was washed twice with 75% ethanol. The pellet was air dried and dissolved in diethyl pyrocarbonate (DEPC)- treated water. mRNA was then fragmented and reverse-transcribed into cDNA using Hiscript II Reverse Transcriptase (Vazyme Biotech, Nanjing City, P.R. China) 10 min. at 25 °C, 40 min. at 42 °C, 15 min. at 70 °C and hold at 4 °C. A library of single-stranded circular molecules was created following the same strategy as above for the single-stranded circular DNA libraries. All these single-stranded circular DNA were then subjected to DNA loading and sequencing according to the sequencer manufacturer's instructions (Slater and Birney, 2005).

1.2 10X Genomics Chromium Genome library preparation, sequencing and analysis

High molecular weight (HMW) DNA was isolated from the blood sample using a RecoverEase DNA Isolation Kit (Agilent, PN 720203, San Clara, USA) with some modifications, in which a plastic pestle (Corning Axygen, PES-15-B-SI, Corning, USA) and 1.5ml microcentrifuge tube (VIOX scientific, V1501-C, London, UK) were applied during the first step of tissue homogenization, replacing the glass tissue grinders. The molecular weight of isolated DNA was then assayed by pulsed field electrophoresis.

For preparation of the Chromium library, the isolated HMW DNA was quantitated and 1 ng was denatured according to the manufacturer's recommendations (Chromium Genome Chip Kit v1, PN-120229, 10X Genomics, Pleasanton, USA). The denatured DNA was spiked into the reaction master mix, and mixed with gel beads and emulsification oil to generate droplets within a Chromium Genome chip. Then we finished the rest steps of Chromium library preparation following the

manufacturer's protocols, with one modified PCR primer to introduce a 5' phosphorylation site on one amplified strand. After PCR, the standard circularization step for BGISEQ-500 sequencer was carried out and DNA nanoballs (DNB) were prepared as previously described (Drmanac et al., 2010) followed by sequencing at the length of 150 bp pair-end. The raw fastq file was converted into a fastq format that could then be recognized by 10X Genomics Supernova (v2.0.0)[3].

1.3 Hi-C library preparation and sequencing

The blood sample was fixed in 1% formaldehyde and the reaction was stopped with glycine. Then the restriction enzyme (*Mbo* I) (NEB, Ipswich, USA) was added to digest the DNA, followed by repairing the 5′ overhang (10 mM dCTP, 10 mM dGTP, 10 mM dTTP, (INVITROGEN, Waltham, USA) 5U/μl DNA Polymerase I, Large (Klenow) Fragment (NEB, Ipswich, USA)) using a biotinylated residue (0.4 mM biotin-14-dATP(INVITROGEN, Waltham, USA)), and 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 (Add 10 mg/ml proteinase K (NEB, Ipswich, USA) and 1% SDS (AMBION, Waltham, USA) to the tube and incubated at 56 °C overnight). And then we purified it by splitting the reverse-crosslinked DNA into three tube and added 1.5× volume of AMpure XP (AGENCOURT, Brea, USA) mixture to each tube, vortexed and spun down briefly, incubated for 10 min. at room temperature. We placed them on the MPS (INVITROGEN, Waltham, USA) for 5 min. at room temperature, discarded supernatant, wash the beads twice with 1 ml of freshly made 70% ethanol (SINOPHARM, Shanghai, China), air-dried the beads completely and re-suspended in 30 μl ddH2O). The Hi-C library was created by shearing 20 μg of DNA and capturing the biotin-containing fragments on streptavidin-coated beads using Dynabeads MyOne Streptavidin T1 (INVITROGEN, Waltham, USA). DNA fragment end repair (10X NEB T4 DNA ligase buffer with 10 mM ATP (NEB, Ipswich, USA), 25 mM dNTP mix (ENZYMATICS, Beverly, USA), 10 U/μl NEB T4 PNK (NEB, Ipswich, USA), 3 U/μl NEB T4 DNA polymerase I (NEB, Ipswich, USA), 5 U/μl NEB DNA

polymerase I, Large (Klenow) Fragment (NEB, Ipswich, USA)), adenylation (10X NEBuffer 2 (NEB, Ipswich, USA), 10 mM dATP (Invitrogen, Waltham, USA), 5 U/μl NEB Klenow exo minus (NEB, Ipswich, USA)), and adaptor ligation were performed using 10X T4 PNK Reaction Buffer(NEB, Ipswich, USA), 100mM ATP (FERMENTAS, USA), 600 U/ul T4 DNA Ligase (NEB, Ipswich, USA), 50% PEG8000 (RIGAKU, Tokyo, Japan), 50 uM Ad153 barcode oligo_2B mix (BGI, Shenzhen, China), and followed by PCR (95 °C 3 min.; [98 °C 20 sec., 60 °C 15 sec., 72 °C 15 sec.] (8 cycles); 72 °C 10 min.). After PCR, the standard circularization step required for BGISEQ-500 was carried out and DNB were prepared as previously described (Drmanac et al., 2010).

The library was then sequenced on the BGISEQ-500 sequencer. The raw fastq files were converted into the fastq format that could be recognized by HiC-Pro pipeline for subsequent quality control of sequence data.

2 Genome assembly and annotation

2.1 Genome assembly

Supernova (v2.0.0) (Weisenfeld et al., 2017) was used to *de novo* assemble the *S. chinensis* with reads from the 10X Genomics Chromium library. Paired-end sequencing data were applied to fill gaps using KGF (v1.19) and GapCloser (Luo et al., 2012) (v1.12) with default parameters. After filtering Hi-C library data by HiC-Pro (Servant et al., 2015) (v2.9.0), we used 3D-DNA pipeline (Dudchenko et al., 2017) to anchor the gap-filled genome assembly into chromosomes. Mitochondria of *S. chinensis* individuals were assembled by NOVOPlasty (v2.7.1) (Dierckxsens et al., 2017)

2.2 Gene annotation

Repeat elements were annotated based on *de novo* and homolog method. *De novo* repeats were annotated by RepeatModeler (v1.0.8) and long terminal repeats were annotated by LTR-FINDER (Xu and Wang, 2007) (v1.0.6). DNA and protein transposable elements (TEs) were annotated by RepeatMasker (Price et al., 2005) (v4.0.5) and RepeatProteinMasker (v4.0.5) to search Repbase (Bao et al., 2015) database, respectively. Tandem repeats were annotated by Tandem Repeat Finder

(Benson, 1999) (v4.07).

To perform the gene prediction, we generated ~8.5 Gb RNA-seq data from blood of one SNB *S. chinensis* (SNB-06). We used Trinity (Grabherr et al., 2011; Haas et al., 2013) (v2.5.1) to assemble the transcriptome and TGI Clustering Tool (Pertea et al., 2003) (TGICL; v2.1) to generate 16 Mb unique (non-redundant) cDNA sequences. Also, we predicted genes of *S. chinensis* by Augustus (Stanke et al., 2006) (v3.3) for *de novo* annotation, and downloaded proteins (i.e. homologous proteins) of nine reference species from NCBI including common bottlenose dolphin (*Tursiops truncatus*), minke whale (*Balaenoptera acutorostrata*), Beluga (*Delphinapterus leucas*), finless porpoise (*Neophocaena asiaeorientalis*), baiji (*Lipotes vexillifer*), killer whale (*Orcinus orca*), goat (*Capra aegagrus hircus*), cattle (*Bos taurus*) and human (*Homo sapiens*) to perform homologous annotation by GeneWise (v2.4.1) (Birney et al., 2004). cDNA sequences, *de novo* and homologous evidences were combined by GLEAN to get final gene set (Elsik et al., 2007). The final *S. chinensis* gene set was functionally annotated by mapping against Kyoto Encyclopedia of Genes and Genome (Kanehisa and Goto, 2000) (KEGG), Swiss-Prot (Bairoch and Apweiler, 2000), TrEMBL (Bairoch and Apweiler, 2000), COG (Tatusov et al., 1997), and NR (Pruitt et al., 2005) databases using BLASTp (*E*-value threshold was 1×10^{-5}). Gene domains and motifs were identified using InterProScan (Finn et al., 2017; Zdobnov and Apweiler, 2001) against PRINTS (Attwood et al., 2002), ProDom (Bru et al., 2005), Pfam (Finn et al., 2008), SMART (Ponting et al., 1999), PANTHER (Mi et al., 2005), and PROSITE (Hulo et al., 2006). GO (The Gene Ontology et al., 2000) terms were obtained from the InterPro entries in the study. We identified 17, 286 gene families in the Indo-Pacific humpback dolphin genome through BLAST searching against eight other species genomes (*H. sapiens, E. cabllus, T. truncatus, N. asiaeorintalis, P. macrocephlus, B. acutorostrata, B. taurus,* and *O. aries*), with the human genome and *E. cabllus* genome as outgroups. We then identified 1,915 single copy gene families to build species phylogenetic trees using maximum likelihood method. Divergence time were estimated by MCMCTREE (v4.4), and calibration divergence times were used between *O. aries* and *B. taurus* (18.3 ~28.5 Mya), *O.*

aries and *H. sapiens* (953 ~ 113 Mya), and *E. caballus* and *H. sapiens* (95.3 ~ 113 Mya).

3 Genome analyses

3.1 Evolutionary analysis

Gene families was identified in *S. chinensis* genome through BLAST search against six other genomes including bottlenose dolphin (*Tursiops truncatus*), sperm whale (*Physeter macrocephalus*), common minke whale (*Balaenoptera acutorostrata*), Finless porpoise (*Neophocaena asiaeorientalis*), sheep (*Ovis aries*) and cattle (*Bos taurus*), as well as human (*Homo sapiens*) and horse (*Equns caballus*) genome as outgroups. We then identified 1,915 single copy gene families with TreeFam to build species phylogenetic trees (Li et al., 2006). Divergence time were estimated by MCMCTREE (v4.4), and calibration divergence times were used between *O. aries* and *B. taurus* (18.3 ~28.5 Mya), *O. aries* and *H. sapiens* (953 ~ 113 Mya), *E. caballus* and *H. sapiens*(Benton and Donoghue, 2007) (95.3 ~ 113 Mya).

3.2 Synteny analysis

All *S. chinensis* chromosomes were aligned to cattle and sperm whale chromosomes using LASTZ (Harris, 2007a) (version 1.04) with parameters 'T=2 C=2 H=2000 Y=3400 L=6000 K=2200', and output with maf format. After filtering the aligned blocks shorter than 2 kb in length, we plotted the synteny relationships using Circos (version 0.69) (Krzywinski et al., 2009).

3.3. Reconstruction of ancestral chromosome

In order to reconstruct the ancestral chromosome of extant cetaceans and their terrestrial sister taxa, chromosome-level genomes of cattle (*B. taurus*; assembly bosTau8), sheep (*O. aries*; assembly oviAri3), and horse (*E. caballus*; assembly equCab2) were downloaded from UCSC. Whereas a chromosome-level sperm whale (*P. macrocephalus*) genome was obtained from CNGBdb (https://db.cngb.org/search/assembly/CNA0002349) (Fan et al., 2019). The cattle genome was served as the reference genome, onto which all other genomes were aligned using LASTZ (Harris, 2007b) (v1.04) with parameters 'T=2 C=2 H=2000 Y=3400 L=6000 K=2200'. We defined sequence fragments of at least 300 kb length

found in all investigated species as a homologous synteny block (HSB) according to the LASTZ results. In total, 342 homologous synteny blocks (HSBs) were identified at the 300-kb resolution, with the horse genome serving as outgroup. These HSBs were used to construct the ancestral chromosomes for Cetruminantia, Bovidae, and Odontoceti by the Multiple Genomes Rearrangements and Ancestors tool (MGRA2) (Alekseyev and Pevzner, 2009; Avdeyev et al., 2016) with default parameter settings. MGRA2 outputs served as inputs in MGR and Genome Rearrangements In Man and Mouse (MGR/GRIMM) (Bourque and Pevzner, 2002) to find the most parsimonious scenarios of reshuffling events. We defined reshuffling events as chromosome evolutionary changes of reversal, translocation, fusion, and fission events; and each reshuffling event can include more than one HSB movement. Evolutionary breakpoint regions (EBRs) were intervals demarked by two adjacent HSB boundaries on the same chromosome. The analysis results of Cetruminantia chromosome evolutionary changes with simplified changing steps were shown in Fig.2.

To identify positive selection genes, the non-synonymous to synonymous substitutions (dN/dS) ratios for foreground branch (*S. chinensis*) and background branch (cattle, sheep and pig) were estimated by the free ratio model of Codeml in PAML (v4.8) (Yang, 2007). The branch-site likelihood ratio test was performed with the null model. Genes with P value less than 0.05 were treated as candidates of positive selection. The faster evolving genes ($P<0.05$, $\omega>1$) were confirmed with the branch model in PAML with the null model and the alternative model.

4 Population resequencing

4.1 Variant calling and population structure

Genome variant calling was conducted using Illumina DRAGEN pipeline based on the high-quality reads. In order to reduce the errors on variant calling, we used VCFtools (Danecek et al., 2011) (v0.1.14) to filter out low-quality (allele frequency lower than 0.01, quality value lower than 50, and more than 30% missing genotypes) SNPs. Mitochondrial variants were called using the GATK (v3.6). Principal component analysis (PCA) was conducted using PLINK (Chang et al., 2015) (v1.90).

Population structure was inferred by ADMIXTURE (v1.3.0) with 1000 cross-validation replicates.

To identify the potential positive selection region, the fixation index (F_{ST}) between the LZB and SNW populations was calculated using VCFtools (v0.1.14) (Danecek et al., 2011) with a non-overlapping sliding window of 50 kb length with parameter "--fst-window-size 50000 --fst-window-step 50000". Tajima's *D* of LZB and SNB was calculated using VCFtools (v0.1.14) (Danecek et al., 2011) with a non-overlapping sliding window of 50 kb length with the parameter "--TajimaD 50000", separately. XP-CLR score of LZB and SNB was calculated using xpclr (v1.1.2) (Chen et al., 2010; Sabeti et al., 2007) by assuming uniform recombination of 1 cM per 1 Mbp with default parameters, separately. XP-EHH score of LZB and SNB was calculated using xpehh (Sabeti et al., 2007) with default parameters. μ-statistic of LZB and SNB was calculated using RAiSD (v2.5) (Alachiotis and Pavlidis, 2018) with the parameter "-R -f -D -P -A 0.99". Because scores of XP-CLR, XP-EHH, and μ-statistic were obtained from SNP sites, we re-calculated the average scores by a non-overlapping sliding window of 50 kb length using our costumed Perl script. Finally, the windows with a significant level of 1% at least two supports from XP-CLR, XP-EHH or μ-statistic, and meanwhile with F_{ST}>0.15 and negative Tajima's *D*, were identified as positive selection regions, and the genes on those regions were extracted as potential positive selection genes.

SNPs were filtered – removing individuals with more than 10% missing genotypes, a high rate of missing genotype calls (10%), and minor allele frequency (MAF) value above 0.05 – and analysed using PLINK (v1.90) to generate the inbreeding coefficient (F_h) of each population (Keller et al., 2011). To build an individual-level tree, we used the Indo-Pacific bottlenose dolphin (*Tursiops aduncus*) genome as an outgroup and excluded SNP sites with missing data. The remaining 6,995,513 SNPs were used to calculate uncorrected distances with an UPGMA tree search and 100 bootstrap replicates in PAUP (v4a164) (https://paup.phylosolutions.com).

4.2 Demographic history reconstruction and gene flow

The effective population size was estimated by heterozygous sites' changes across the genome in PSMC model. The f3 test was calculated using ADMIXTURE. The μ (neutral mutation rate per generation) was calculated using the distmat software (http://emboss.sourceforge.net/apps/release/6.6/emboss/apps/distmat.html). The divergence time of the two populations and migration rates were firstly inferred by DaDi (v1.7.0) through 32 models to evaluate which model was the best (**Supplementary Table 22**), and then "sym_mig_size" model was confirmed and used to calculate our results. The effective population sizes (*Ne*) of contemporary and their change over the past 100 generations were estimated by PopSizeABC with parameters: nb_rep (number of simulated datasets) is 1000; L (size of each segment, in bp) is 2,000,000; mac (minor allele count threshold for AFS and IBS statistics computation) is 0; nb_seg (number of independent segments in each dataset) is 100; mac_ld (minor allele count threshold for LD statistics computation) is 6; nb_times (number of time window) is 5; T_{max} (the oldest time window will start at $_{Tmax}$) is 100; all others as default.

Supplemental references

Alachiotis, N., and Pavlidis, P. (2018). RAiSD detects positive selection based on multiple signatures of a selective sweep and SNP vectors. Communications Biology *1*.

Alekseyev, M.A., and Pevzner, P.A. (2009). Breakpoint graphs and ancestral genome reconstructions. Genome Research *19*, 943-957.

Attwood, T.K., Blythe, M.J., Flower, D.R., Gaulton, A., Mabey, J.E., Maudling, N., McGregor, L., Mitchell, A.L., Moulton, G., Paine, K., *et al.* (2002). PRINTS and PRINTS-S shed light on protein ancestry. Nucleic Acids Research *30*, 239-241.

Avdeyev, P., Jiang, S., Aganezov, S., Hu, F., and Alekseyev, M.A. (2016). Reconstruction of ancestral genomes in presence of gene gain and loss. bioRxiv.

Bairoch, A., and Apweiler, R. (2000). The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Research *28*, 45-48.

Bao, W., Kojima, K.K., and Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. Mobile DNA *6*, 11.

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

Benton, M.J., and Donoghue, P.C.J. (2007). Paleontological evidence to date the

tree of life. Molecular Biology and Evolution *24*, 26-53.

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

Bourque, G., and Pevzner, P.A. (2002). Genome-Scale Evolution: Reconstructing Gene Orders in the Ancestral Species. Genome Research *12*, 26-36.

Bru, C., Courcelle, E., Carrère, S., Beausse, Y., Dalmar, S., and Kahn, D. (2005). The ProDom database of protein domain families: more emphasis on 3D. Nucleic Acids Research *33*, D212-D215.

Chang, C.C., Chow, C.C., Tellier, L.C.A.M., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience *4*, 7.

Chen, H., Patterson, N., and Reich, D. (2010). Population differentiation as a test for selective sweeps. Genome Research *20*, 393-402.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., *et al.* (2011). The variant call format and VCFtools. Bioinformatics *27*, 2156-2158.

Dierckxsens, N., Mardulyn, P., and Smits, G. (2017). NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Research *45*.

Drmanac, R., Sparks, A.B., Callow, M.J., Halpern, A.L., Burns, N.L., Kermani, B.G., Carnevali, P., Nazarenko, I., Nilsen, G.B., Yeung, G., *et al.* (2010). Human Genome Sequencing Using Unchained Base Reads on Self-Assembling DNA Nanoarrays. Science *327*, 78-81.

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.

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*, R13.

Fan, G., Zhang, Y., Liu, X., Wang, J., Sun, Z., Sun, S., Zhang, H., Chen, J., Lv, M., Han, K., *et al.* (2019). The first chromosome-level genome for a marine mammal as a resource to study ecology and evolution. Mol Ecol Resour.

Finn, R.D., Attwood, T.K., Babbitt, P.C., Bateman, A., Bork, P., Bridge, A.J., Chang, H.-Y., Dosztányi, Z., El-Gebali, S., Fraser, M., *et al.* (2017). InterPro in 2017—beyond protein family and domain annotations. Nucleic Acids Research *45*, D190-D199.

Finn, R.D., Tate, J., Mistry, J., Coggill, P.C., Sammut, S.J., Hotz, H.-R., Ceric, G., Forslund, K., Eddy, S.R., Sonnhammer, E.L.L., *et al.* (2008). The Pfam protein families database. Nucleic Acids Research *36*, D281-D288.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., *et al.* (2011). Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nature biotechnology *29*, 644-652.

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: reference generation and analysis with Trinity. Nature protocols *8*, 10.1038/nprot.2013.1084.

Harris, R.S. (2007a). Improved Pairwise Alignmnet of Genomic DNA. In The Graduate School College of Engineering (State College: The Pennsylvania State University).

Harris, R.S. (2007b). Ph.D. Thesis: improved pairwise alignment of genomic DNA.

Hulo, N., Bairoch, A., Bulliard, V., Cerutti, L., De Castro, E., Langendijk-Genevaux, P.S., Pagni, M., and Sigrist, C.J.A. (2006). The PROSITE database. Nucleic Acids Research *34*, D227-D230.

Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research *28*, 27-30.

Keller, M.C., Visscher, P.M., and Goddard, M.E. (2011). Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. Genetics *189*, 237-249.

Krzywinski, M., Schein, J., Birol, İ., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: An information aesthetic for comparative genomics. Genome Research *19*, 1639-1645.

Li, H., Coghlan, A., Ruan, J., Coin, L.J., Heriche, J.-K., Osmotherly, L., Li, R., Liu, T., Zhang, Z., and Bolund, L. (2006). TreeFam: a curated database of phylogenetic trees of animal gene families. Nucleic acids research *34*, D572-D580.

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

Mi, H., Lazareva-Ulitsky, B., Loo, R., Kejariwal, A., Vandergriff, J., Rabkin, S., Guo, N., Muruganujan, A., Doremieux, O., Campbell, M.J., *et al.* (2005). The PANTHER database of protein families, subfamilies, functions and pathways. Nucleic Acids Research *33*, D284-D288.

Pertea, G., Huang, X., Liang, F., Antonescu, V., Sultana, R., Karamycheva, S., Lee, Y., White, J., Cheung, F., Parvizi, B., *et al.* (2003). TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. Bioinformatics *19*, 651-652.

Ponting, C.P., Schultz, J., Milpetz, F., and Bork, P. (1999). SMART: identification and annotation of domains from signalling and extracellular protein sequences. Nucleic Acids Research *27*, 229-232.

Price, A.L., Jones, N.C., and Pevzner, P.A. (2005). De novo identification of repeat families in large genomes. Bioinformatics *21 Suppl 1*, i351-358.

Pruitt, K.D., Tatusova, T., and Maglott, D.R. (2005). NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Research *33*, D501-D504.

Sabeti, P.C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E.H., McCarroll, S.A., Gaudet, R., *et al.* (2007). Genome-wide detection and characterization of positive selection in human populations. Nature *449*,

913-U912.

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 Biology *16*.

Slater, G.S.C., and Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. In BMC bioinformatics, pp. 31.

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*, W435-439.

Tatusov, R.L., Koonin, E.V., and Lipman, D.J. (1997). A Genomic Perspective on Protein Families. Science *278*, 631.

The Gene Ontology, C., Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., *et al.* (2000). Gene Ontology: tool for the unification of biology. Nature genetics *25*, 25-29.

Weisenfeld, N.I., Kumar, V., Shah, P., Church, D.M., and Jaffe, D.B. (2017). Direct determination of diploid genome sequences. Genome Research *27*, 757-767.

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

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Molecular biology and evolution *24*, 1586-1591.

Zdobnov, E.M., and Apweiler, R. (2001). InterProScan--an integration platform for the signature-recognition methods in InterPro. Bioinformatics *17*, 847-848.