

Article

Whole Genome Sequencing of Chinese White Dolphin (*Sousa chinensis*) for High-Throughput Screening of Antihypertensive Peptides

Kuntong Jia ^{1,†}, Chao Bian ^{2,†}, Yunhai Yi ^{2,†} , Yanping Li ², Peng Jia ¹, Duan Gui ¹, Xiyang Zhang ¹, Wenzhi Lin ¹ , Xian Sun ¹, Yunyun Lv ², Jia Li ², Xinxin You ² , Qiong Shi ^{2,*} , Meisheng Yi ^{1,*} and Yuping Wu ^{1,*}

¹ Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Guangdong Provincial Key Laboratory of Marine Resources and Coastal Engineering, Zhuhai Key Laboratory of Marine Bioresources and Environment, School of Marine Sciences, Sun Yat-Sen University, Zhuhai 519082, China

² Shenzhen Key Lab of Marine Genomics, Guangdong Provincial Key Lab of Molecular Breeding in Marine Economic Animals, BGI Academy of Marine Sciences, BGI Marine, BGI, Shenzhen 518083, China

* Correspondence: shiqiong@genomics.cn (Q.S.); yimsh@mail.sysu.edu.cn (M.Y.); exwyp@mail.sysu.edu.cn (Y.W.); Tel.: +86-185-6627-9826 (Q.S.); +86-138-2566-3933 (M.Y.); +86-139-0253-1137 (Y.W.)

† These authors contributed equally to this work.

Received: 3 June 2019; Accepted: 26 August 2019; Published: 28 August 2019



Abstract: Chinese white dolphin (*Sousa chinensis*), also known as the Indo-Pacific humpback dolphin, has been classified as “Vulnerable” on the IUCN Red List of Threatened Species. It is a special cetacean species that lives in tropical and subtropical nearshore waters, with significant differences from other cetaceans. Here, we sequenced and assembled a draft genome of the Chinese white dolphin with a total length of 2.3 Gb and annotation of 18,387 protein-coding genes. Genes from certain expanded families are potentially involved in DNA replication and repairing, suggesting that they may be related to adaptation of this marine mammal to nearshore environments. We also discovered that its historical population had undergone a remarkable bottleneck incident before the Mindel glaciation. In addition, a comparative genomic survey on antihypertensive peptides (AHTPs) among five representative mammals with various residential habitats (such as remarkable differences in exogenous ion concentrations and sea depth) revealed that these small bioactive peptides were highly conserved among these examined mammals, and they had the most abundant hits in collagen subunit proteins, especially for two putative AHTP peptides Gly-Leu-Pro (GLP) and Leu-Gly-Pro (LGP). Our genome assembly will be a valuable resource for further genetic researches on adaptive ecology and conservation biology of cetaceans, and for in-depth investigations into bioactive peptides in aquatic and terrestrial mammals for development of peptide-based drugs to treat various human cardiovascular diseases.

Keywords: Chinese white dolphin (*Sousa chinensis*); whole genome sequencing; genome assembly; antihypertensive peptide

1. Introduction

Chinese white dolphin (*Sousa chinensis*), also known as the Indo-Pacific humpback dolphin, is a special cetacean species under the genus *Sousa* with a wide distribution from eastern India throughout Southeast Asia to central China. In many regions of China, it commonly inhabits estuaries and inshore waters, but it has more recently become an endangered species. It is believed that a total of six putative populations exist in China, including Xiamen ($n = 86$), the Pearl River Estuary (PRE) ($n = 2,637$),

the Leizhou ($n = 1,485$), the Beibu Gulf (>89), the west coast of Taiwan ($n = 99$) and the west coast of Hainan [1–3]. Unfortunately, the total number of this marine mammal has declined gradually in the past decades. Our recent demographic studies in coastal waters of the Pearl River Delta region of China indicated a declining trend of $\sim 2.5\%$ per annum, suggesting that $\sim 74\%$ of its current population would probably disappear within the lifespan of three generations (~ 60 years) [4]. Thus, this cetacean species has been recently classified as “Vulnerable” on the IUCN Red List of Threatened Species [2,4].

Two major events, climate change and tectonic shifts, have made great contributions to the genetic diversity and structure of marine organisms [5,6]. It is known that numerous changes in the morphological, physiological and behavioral systems of cetaceans, such as *S. chinensis*, have been acquired to adapt to this drastic habitat transition from terrestrial to aquatic environments during the long-term evolution of cetaceans [7–13]. Thus, getting insight into the biological characteristics of cetaceans will benefit the understanding of the evolution of cetaceans and mechanisms for their adaptation to aquatic environments. The white dolphin provides a good example of speciation due to its special biological characteristics. For example, previous studies have shown that this species is primarily distributed in tropical and subtropical waters, and it resides in waters with depths of less than 20 m. It is also famous for the pink skin in adulthood. Similar to other cetaceans in adaptation to the marine environments, its body structure and physiology had exhibited dramatic changes, including streamlined bodies, absence of hind legs and outer ear pinnae, as well as presence of a fluke and blowhole [14,15]. However, the genetic backgrounds that underlies these biological properties are still limited.

Several previous studies showed that genomic analyses of marine mammals provided insights into molecular adaptation to living conditions [16,17]. For example, a comparative genomic analysis of walrus, bottlenose dolphin, killer whale, and manatee determined many convergent amino acid substitutions in these genomes and a portion of these substitutions were in several genes associated with a marine phenotype [16]. In the genome of the minke whale, a number of genes related to stress responses and anaerobic metabolism were expanded, while many genes related to body hair and sensory receptors were contracted [17].

However, the genetic mechanisms of *S. chinensis* underlying adaptation to the aquatic lifestyle is poorly understood. There are only few publications on its genetics and genomics. In the past decade, genetics studies have mainly focused on a single gene or a few genes, such as the major histocompatibility complex class II [18], CYP1A1 and HSP70 [19], as well as the mitochondrial DNA [20]. Recently, we provided the first transcriptome exploration of *S. chinensis* and identified a large number of genes related to adaptive evolution and cetacean-specific traits [9]; the first genome assembly of this marine mammal was also reported previously [21]. One more interesting question is how to regulate blood pressure for cetaceans in different sodium levels and sea depths. Angiotensin-converting enzyme 2 (ACE2), an important enzyme converting vasoconstrictor angiotensin II (Ang II) into inactive Ang (1–7), has been reported to experience changes in adaptation to freshwater [17]. Antihypertensive peptides (AHTPs) have a similar function as ACE2, while they mostly inhibit the production of angiotensin thus leading to the lowering of blood pressure [22]. Many AHTPs, mainly consisting of 2–10 amino acids, are usually digested from products of natural organisms [23]. Endogenous AHTPs may also result from hydrolyzation and degradation of in vivo proteins by certain enzymes, therefore acting as regulators of the renin angiotensin system by binding to angiotensin converting enzymes or related receptors [24]. However, the knowledge about changes of AHTPs in various mammals or cetaceans is still scarce, and there are few studies available for identification of potential protein types that can be hydrolyzed into large amount of AHTPs.

In this study, we report a de novo assembly and primary analyses of the Chinese white dolphin genome based on a cell-line sample that was previously derived from the skin of a dolphin captured by us [25]. Transcriptome sequencing was performed on extracted leucocytes from collected blood samples for assistance to genome annotation. Moreover, a comparative study on AHTPs was performed at a whole proteome level among five representative mammals with different living habitats, including

the terrestrial cow, marine minke whale, epipelagic white dolphin and bottlenose dolphin, as well as the autopotamic Yangtze River dolphin. Here, we aim at providing a valuable genomic resource for genome-wide studies on cetaceans and an in-depth exploration of bioactive peptides for potential development of AHTP-based marine drugs.

2. Results

2.1. Summary of Genome Assembly and Annotation

We generated a total of 318.4 gigabases (Gb) of raw reads (Table S1) by sequencing of seven libraries (see more details in Section 4.2.) in an Illumina HiSeq 2500 platform (Illumina, San Diego, CA, USA). After removal of low-quality reads, we obtained 245.9 Gb of clean data (Table S1), and assembled a 2.3 Gb genome of the Chinese white dolphin, which is close to the estimated genome size (~2.6 Gb) from a *k*-mer analysis (Figure 1). Our genome assembly is composed of 1789 scaffolds (>2000 bp), with a scaffold N50 of 19.2 Mb and a contig N50 of 84.3 Kb (Table 1). Although a primary genome assembly of *S. chinensis* was available, reported by Ming et al. [21], its scaffold N50 (163 kb) and contig N50 (12.9 kb) were shorter. Therefore, in this study, we improved the genome work with a high-quality assembly.

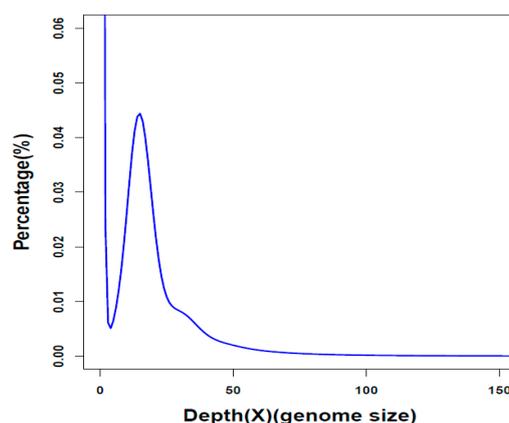


Figure 1. A *K*-mer analysis of the Chinese white dolphin genome. In our present study, the *k*-mer depth is 15, and the estimated genome size is ~2.6 Gb.

Table 1. Summary of the assembly and annotation of the Chinese white dolphin genome.

Genome assembly	Parameter
Contig N50 (kb)	84.3
Scaffold N50 (Mb)	19.2
Assembled genome (Gb)	2.3
Genome coverage (×)	318.4
Longest scaffold (bp)	71,519,079
Genome annotation	Parameter
Number of protein-coding genes	18,387
Transposable elements content (%)	42.3

We further utilized the routine BUSCO (Benchmarking Universal Single-Copy Orthologs) method [26] to check the completeness of our genome assembly, determining that 95% were complete and partial eukaryote BUSCO orthologues. We also identified that repeat sequences account for about 42.3% of the assembled genome (Table S2), and annotated a final complete gene set of 18,387 genes with an average of 44.2 kb in length (Table 1). Approximately 93.9% of the predicted genes have at

least one related function assignment from several public databases, including TrEMBL, SwissProt, KEGG and InterProScan (Table S3).

2.2. Genome Analyses

2.2.1. Phylogenetic Analysis and Divergence Times

A phylogenetic tree was constructed to categorize eleven examined mammals (see more details in Section 4.4) into three major groups (left panel in Figure 2), which is consistent with the traditional taxonomic classification of Cetacea, Artiodactyla and Euarchontoglires. Among these groups, Cetacea appears to have a closer relationship with Artiodactyla than Euarchontoglires, which is strongly supported by the robust node confidence level. Our phylogenetic analysis of 5728 rigorously screened gene orthologs concurred with a recent online phylogenomic report [21] and previous research based on mitochondrial genomes [27].

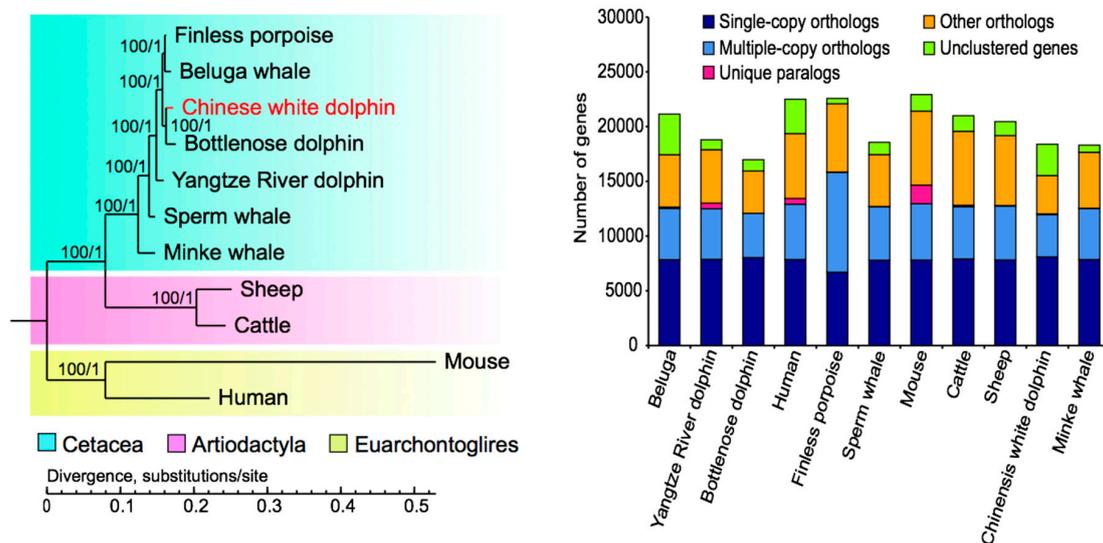


Figure 2. Phylogenetic relationships of Chinese white dolphin and other ten examined mammals (left) in component genes (right). This evolutionary topology indicates a clear division of three major groups of Cetacea, Artiodactyla and Euarchontoglires.

Our divergence time analysis suggested that the cetaceans diverged from Artiodactyla about 53.4 million years ago (Mya; see Figure S1), which is consistent with previous reports [21,25]. Another previous study [28] also considered that the group of cetaceans had diverged from their terrestrial ancestors about 53~56 Mya when they reinvaded aquatic environments. We predicted that the Chinese white dolphin split with the bottlenose dolphin about 7.6 Mya (Figure S1), and since then it began to colonize estuaries and coastal areas. Our present phylogenomic data provide more evidences for the patterns of divergence and evolution in the group of cetaceans.

2.2.2. Expansion and Contraction of Gene Families

We determined the expansion and contraction of gene families (398 and 2505, respectively) in the Chinese white dolphin (see more details in Section 4.6). In comparison with other examined mammals, the largest number of the contracting gene ortholog cluster was inferred from the Chinese white dolphin (Figure 3a), suggesting a possible loss of many gene families during the evolution of this marine mammal. On the other hand, the 398 gene families expanded in the Chinese white dolphin are mainly involved in cell growth and death, transport and catabolism and lipid metabolism (see related annotation in Figure 4). Interestingly, certain genes involved in the nervous and sensory systems were expanded, which is consistent with the fact that the Chinese white dolphin is sensitive to

anthropogenic pressures such as vessel traffic, coastal harbor construction, underwater blasting or dredging, extensive mariculture and fishing activities, as well as agricultural or industrial pollutants in the nearshore environments [21,29,30]. We also identified that certain gene families involved in DNA replication and repairing were also expanded, which may benefit this marine mammal to cope with a high UV exposure condition in the shallow waters.

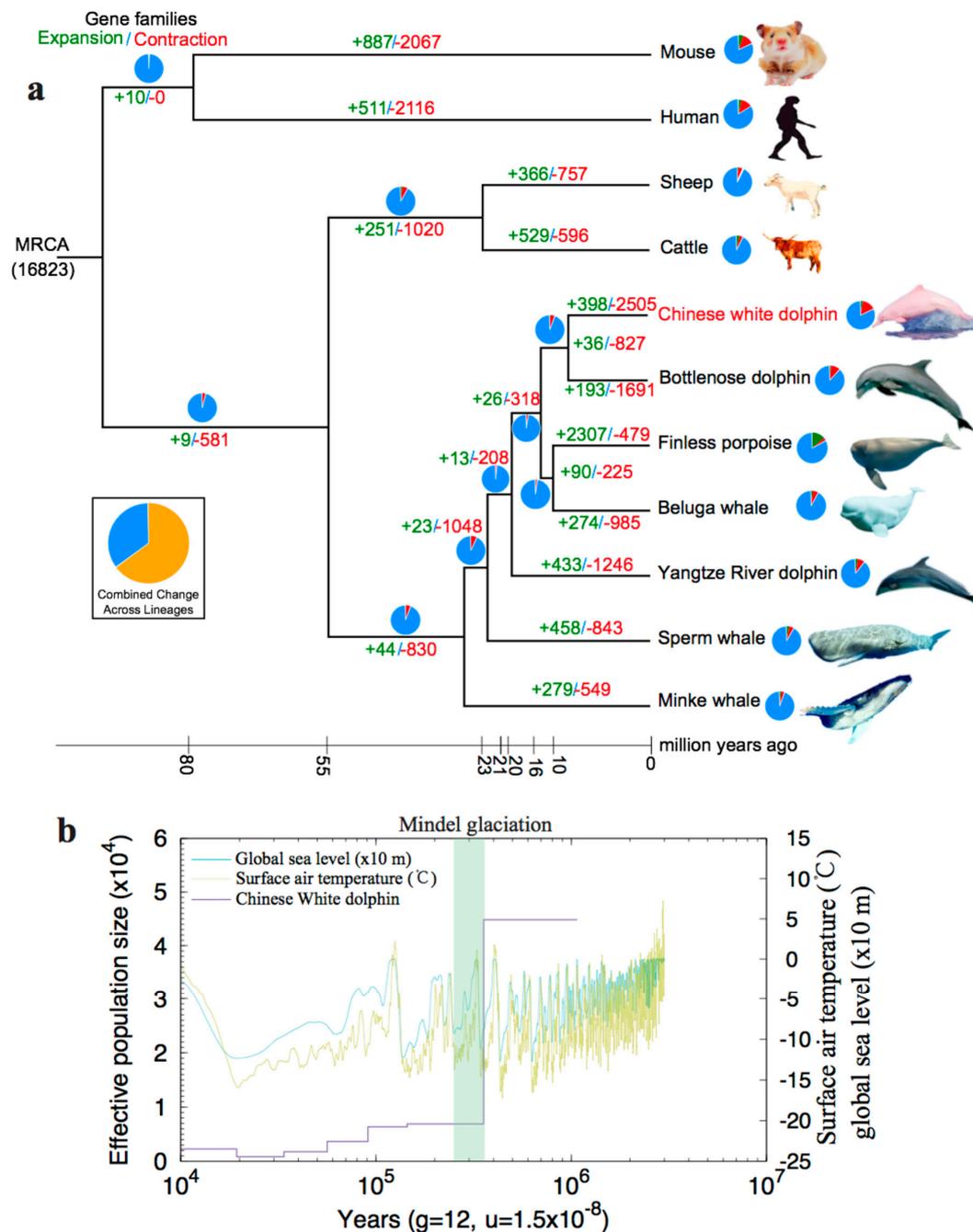


Figure 3. Evolution of the Chinese white dolphin. (a) A phylogenetic tree of the eleven examined mammals. Numbers associated with each branch stand for gene families that have expanded (green) or contracted (red) since splitting from the common ancestor. (b) Predicted population history of the Chinese white dolphin (see more explanations in Section 2.2.3). The purple line represents the population changes. The green and light-yellow lines denote the reported fluctuations of global sea level and surface air temperature. The green box represents the Mindel glaciation period.

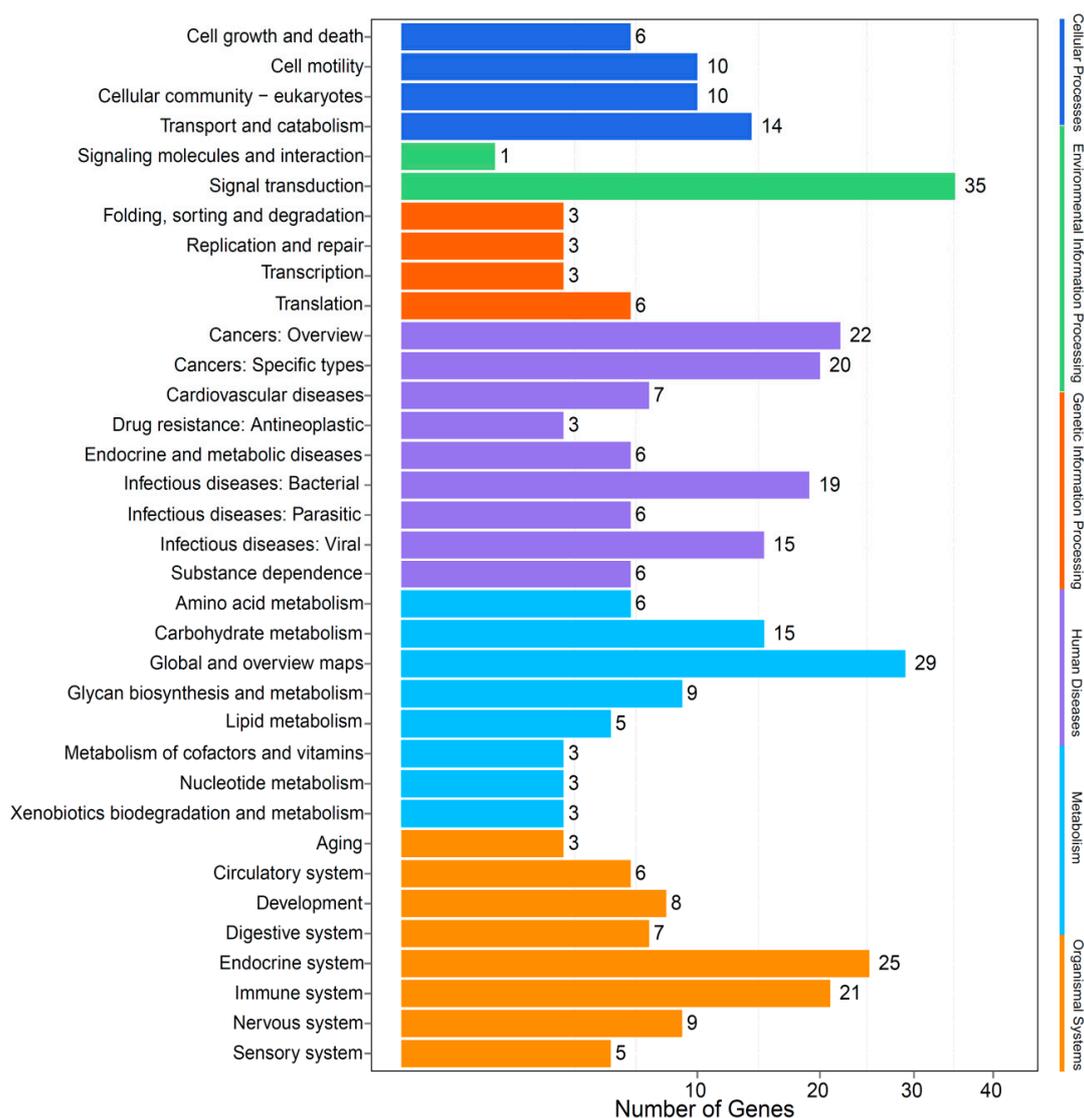


Figure 4. KEGG annotation of the expanded gene families in the Chinese white dolphin.

2.2.3. Population History

We utilized the PSMC approach [31] to reconstruct the population history of the Chinese white dolphin (see more details in Section 4.7). Interestingly, we observed a remarkable bottleneck incident in its historical population that appeared about 0.35 Mya (Figure 3b). Similarly, this phenomenon of a population bottleneck was previously reported in sperm whales and finless porpoises [32,33]. However, the corresponding period of population bottleneck for the Chinese white dolphin was much later (about one Mya) than those of the two reported marine mammals.

After combining the reported data of atmospheric surface air temperature and global relative sea level (from the National Climatic Data Center at <http://www.ncdc.noaa.gov/>), we observed that the population of Chinese white dolphin had sharply declined after a remarkable reduction in temperature and sea level before the Mindel glaciation (middle in Figure 3b). Subsequently, the population size gradually decreased from 0.15 to 0.20 Mya, and this change pattern was similar to the curves of temperature and sea level during this critical period. Therefore, we reach a primary conclusion that the global relative sea levels, possibly due to changes in atmospheric surface air temperature, could have had substantial impacts on the fluctuation of the Chinese white dolphin population.

2.3. Identification of AHTPs

In order to screen potential AHTPs in the deduced proteome of Chinese white dolphin, we established a local database (Table S4) for those AHTPs that have been verified in previous studies. In fact, most of them are tripeptides and usually less than 10 amino acids (Figure 5).

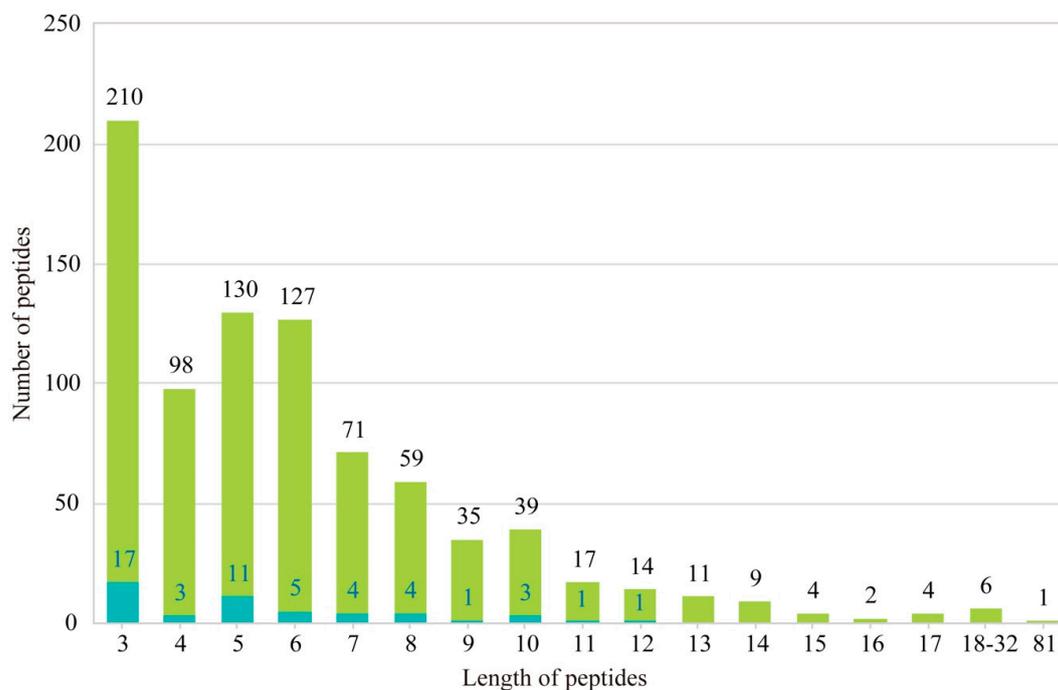


Figure 5. Statistic of antihypertensive peptides (AHTPs) in our local database and those employed for this study. Blue boxes within some groups denote the number of AHTP peptides explored in this study, which potentially have a higher activity for antihypertension based on previous reports.

We picked out the top 50 active AHTPs (Table S4) to map the deduced proteome of Chinese white dolphin and the downloaded protein datasets (Table S5), and identified 35 AHTPs in the five representative mammals (Table S9), including the terrestrial cow (34), marine minke whale (34), epipelagic Chinese white dolphin (31) and bottlenose dolphin (31), as well as the autopotamic Yangtze River dolphin (31). All the mapping results were listed in Table S6, and the hit numbers of AHTPs in each protein of the five examined mammals were sorted in Table S7.

As shown in Table S9, the composition of AHTPs seems to determine their presence in the whole proteome datasets. In general, Leu-Gly-Pro (LGP) had a much higher frequency than Leu-Lys-Pro (LKP) and Leu-Arg-Pro (LRP) in the five examined mammals, while the occurrence of LRP was closer to LKP than to Leu-Arg-Trp (LRW). Among the 35 characterized AHTPs, Gly-Leu-Pro (GLP), LGP, Val-Ser-Val (VSV), LRP and LKP were the major components in the five mammal protein datasets (Table S9, Figure 6). There were fewer mapping results for longer peptides, including four peptides that only existed in the minke whale and/or terrestrial cow. Interestingly, for the most mapped AHTPs, the white dolphin had the least hit numbers (Table S9, Figure 6a), which is consistent with the least annotated gene/protein number among the five mammal genomes.

Minke whale and cow possessed the most abundant AHTP hits (60,820 and 61,028, respectively) and mapped proteins (25,079 and 25,012, respectively), while the Chinese white dolphin had the least with 27,260 hits in 18,387 mapped proteins (Table 2). However, their mapping rates were at a narrow range from about 0.62 to 0.67. Average AHTP numbers of all mapped proteins in the five mammals were also between 2.3 and 2.5. Our mapping results revealed that titin and collagen type IV alpha protein were at the top with the most abundant AHTP hits in these mammals (Table S7), which is consistent with our previous report in 18 fishes [34]. It therefore seems that the occurrence of AHTPs

on their mapped protein sequences may be highly conserved among vertebrates, which may suggest conservation of AHTP-dependent antihypertensive mechanisms in various animals.

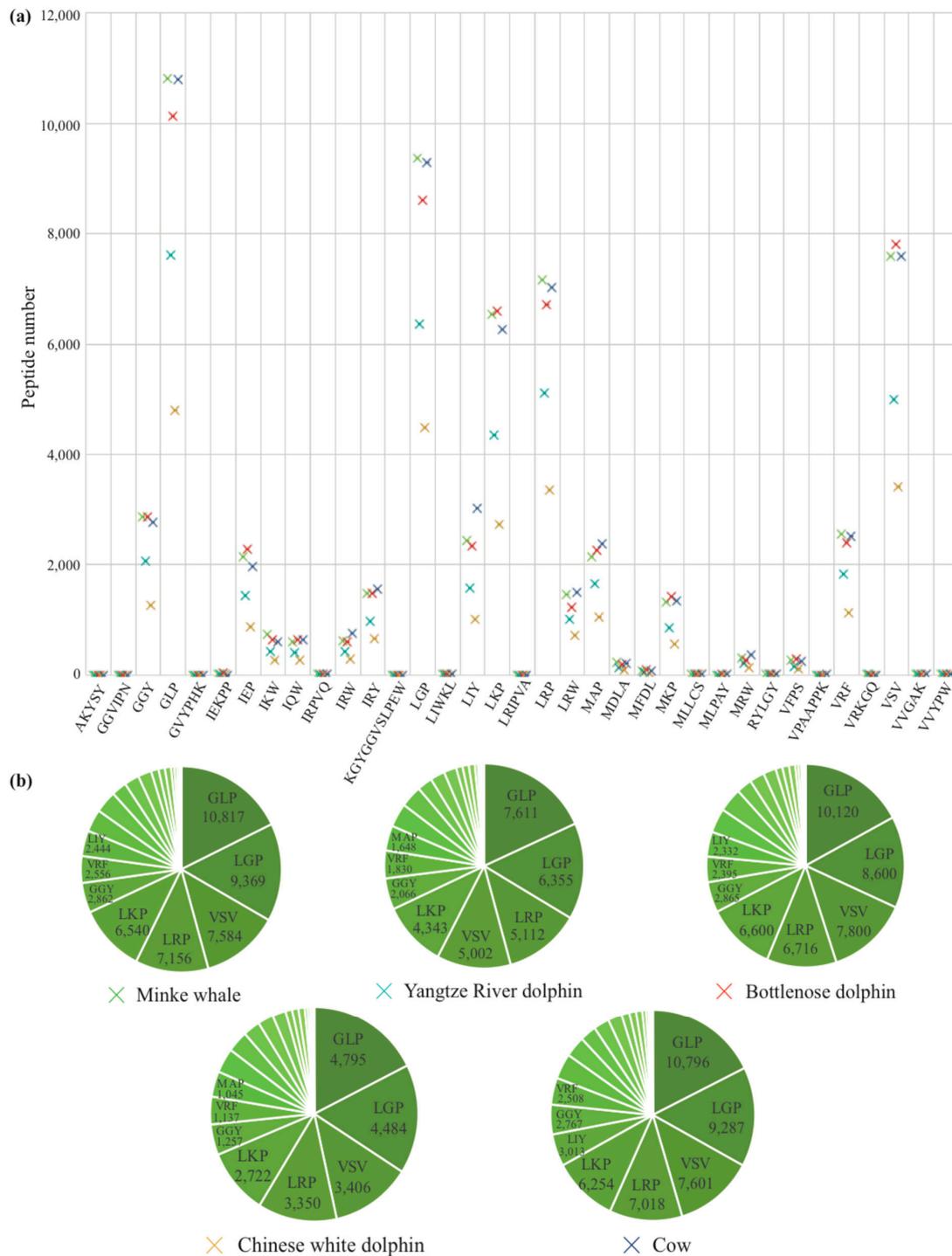


Figure 6. Comparison of each AHTP hit number in the five examined mammals (a) and summary of the top eight abundant AHTPs in each mammal species (b).

Generally speaking, the minke whale mapped more collagen subunit proteins than the other four mammals, although the Chinese white dolphin had 48 mapped collagen subunits (Table 2). The longest peptide, KGYGGVSLPEW, isolated from whole whey proteins, was only identified in the lactalbumin

alpha protein of cow in this study. Our data consolidate the reliability of our genomic approach in high-throughput discovery of bioactive peptides.

Table 2. Summary of the AHTP mapping results in the five representative mammals.

Parameter	Minke Whale	Yangtze River Dolphin	Bottlenose Dolphin	Chinese White Dolphin	Cow
Total hits	60,820	41,733	58,992	27,260	61,028
Mapped protein	25,079	17,633	25,593	11,323	25,012
Total protein	37,625	26,901	38,849	18,387	37,525
Annotated protein number	3206	3105	2768	1692	13,435
Mapping rate	0.6666	0.6555	0.6588	0.6158	0.6665
Average AHTPs number in mapped protein	2.4251	2.3668	2.3050	2.4075	2.4399
Collagen subunit number in mapped protein	92	66	69	48	75

Functional annotation of AHTP-containing proteins revealed a similar distribution pattern in the five examined mammals (Figure 7, Table S8). However, the annotated protein number of cow in each functional category far more exceeded that of the others. It seems that their functions focused on metabolic process and regulation of biological process; that is, they were mainly components of cells, membranes and organelle parts, and possessed binding and catalytic activity in a molecular function term. These results proved that the categories of AHTP-containing proteins in the five mammals were much similar, which conforms to their similar mapping rates at the whole proteome level. Although these mammals are from different habitats, our comparisons of AHTPs and AHTP-containing proteins support the high conservation between aquatic and terrestrial animals.

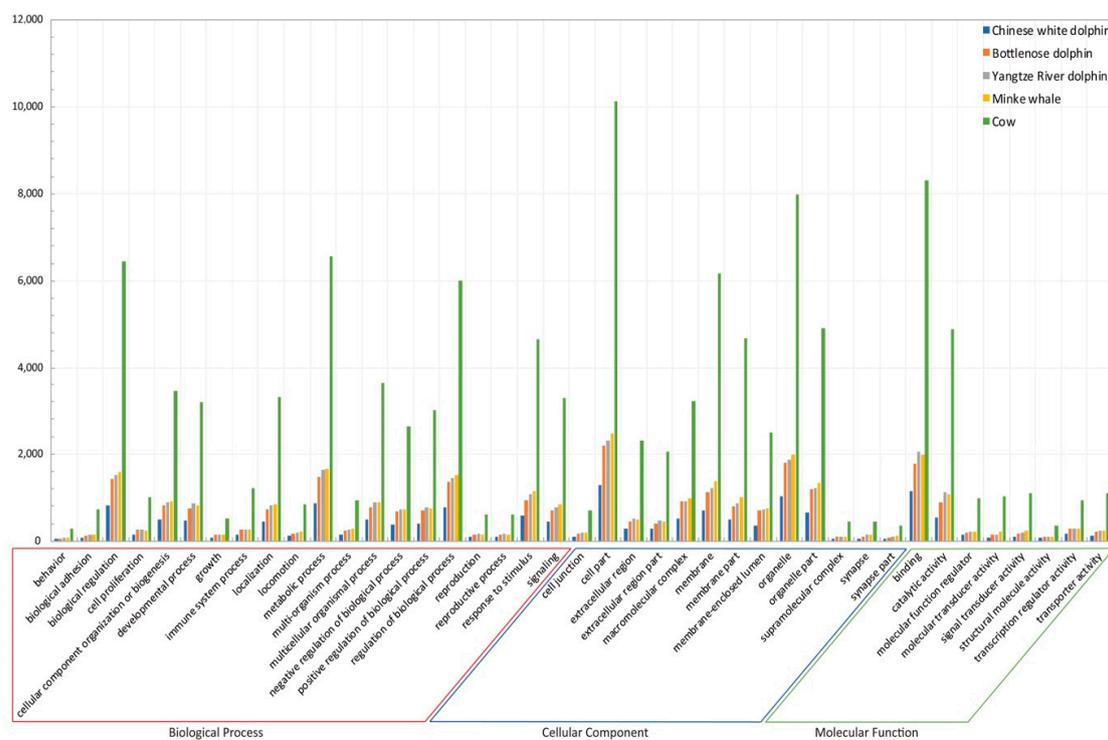


Figure 7. Functional annotation of AHTP-mapped proteins in the five examined mammals.

3. Discussion

Previous researches have indicated that expanded gene families to some extent may reflect specific physiological adaptation and special traits in the studied lineages [35]. The cetaceans that live in

offshore environments may often face many physiological challenges, such as sensory disruption. They have to evolve and reserve some special characters to adapt to various conditions. An expansion of gene families related to nervous and sensory systems was found in the Chinese white dolphin ([21] and this study), suggesting that this marine mammal has developed a very sensitive sensory system for anthropogenic or prey pressures. We also identified some expanded gene families involved in DNA replication and repairing, indicating that the Chinese white dolphin may have a great capacity to replicate and repair DNA due to the potential serious damages from a high level of UV stimuli in tropical regions and nearshore environments. These data are consistent with the previous genome report of the Chinese white dolphin [21].

Our genome assembly of the Chinese white dolphin also provides a valuable genetic resource for AHTP research in aquatic mammals for the first time. The comparisons in this study among the five representative mammals, including the terrestrial cow, marine minke whale, freshwater Yangtze River dolphin, as well as the epipelagic bottlenose dolphin and Chinese white dolphin, showed that AHTPs in these mammals were highly conserved. In our previous study of AHTPs in whole proteome datasets of 18 fish species [34], we also observed a conserved mapping pattern.

The living conditions of these examined mammals have no obvious association with the amount of AHTPs. The mapping rate and average AHTP hit number of each mapped protein were similar in this study. Collagen subunit proteins among the longest proteins possess the top abundant AHTPs here (Table S7), especially for the type IV alpha 5 (col4a5) and type VIII alpha 1 (col8a1) subunits from both our studies of fish and mammals. Moreover, the most abundant AHTP categories involved GLP, LGP and VSV in the five examined mammals. We previously found that Atlantic salmon had the most abundant AHTPs in comparison with other fishes [34]. In this study, cow and minke whale had an equal high level of AHTP hit numbers. These AHTP-containing proteins were reported to participate in multiple biological and metabolic functions. Several mechanisms contribute to the maintenance of blood pressure, at steady state and during diving [36]. Thus, the richness of collagen subunit proteins in the minke whale (Table 2), the longest diving cetacean among the analyzed species, may suggest their importance towards adaptation to diving-induced hypoxia [37]. These proteins are also a potential resource for development of AHTP-based marine drugs.

4. Materials and Methods

4.1. Sample Collection and Preparation

A cell line derived from the skin of a Chinese white dolphin was prepared as we reported previously [25]. The skin sample was collected by using a noninvasive method from a male adult individual, which was rescued for rehabilitation from an animal live-stranding event in a shallow river near Foshan City, Guangdong Province, China. Under the permission of the Pearl River Estuary Chinese White Dolphin National Nature Reserve (No. 2017A030308005), we obtained assistance from local veterinarians to collect samples. The sampling site on the dolphin's back was sterilized by surgical cottons with 70% alcohol, and the tissue fragments were sheared off aseptically by scraping with a blade. The wound (approximately 0.2 cm²) was treated immediately with haemostatic and anti-inflammatory ointments. The skin tissue was immersed into Dulbecco's modified Eagle's medium (DMEM; ThermoFisher Scientific, Waltham, MA, USA) including penicillin (100 U/ml), streptomycin (100 µg/ml) and amphotericin B (5 µg/ml), and then immediately transported on ice within 2 h to our laboratory for a subsequent cell-line preparation [25].

Genomic DNA was extracted from cells using a Tissue DNA Kit (Omega, Norcross, GA, USA) according to the manufacturer's protocol. All experiments were performed in accordance with the Regulations of the Animal Ethics Committee and were approved by the Institutional Review Board on Bioethics and Biosafety of Sun Yat-Sen University, China.

4.2. Genome Sequencing and Assembling

We employed the traditional whole-genome shotgun sequencing strategy and constructed seven paired-end libraries with diverse insert sizes, including three short-insert libraries (270, 500 and 800 bp) and four long-insert libraries (2, 5, 10 and 20 kb), for genome sequencing of the extracted genomic DNA on an Illumina HiSeq 2500 platform. About 318.4 Gb of raw reads were generated. After removal of low-quality and redundant reads, we obtained 245.9 Gb of clean data for further de novo assembly (Table S1).

We employed SOAP-denovo2 [38] (with -k 65) to build contigs and primary scaffolds by utilizing reads from the short-insert libraries (250, 500 and 800 bp). Subsequently, reads from the long-insert libraries (2, 5, 10 and 20 kb) were mapped onto contigs to shape corresponding scaffolds. Gapcloser in the package of SOAP-denovo2 was employed to fill the gaps within those achieved scaffolds. Our genome assembly of the Chinese white dolphin was deposited in NCBI with the accession number of RWJT00000000.

4.3. Genome Annotation

We first identified repeat sequences in our genome assembly using the Tandem Repeats Finder [39], LTR_FINDER [40], RepeatProteinMask and RepeatMasker (version 3.2.9, Institute for System Biology, Seattle, CA, USA) [41]. The Tandem Repeat Finder was employed to search the tandem repeats from the genome assembly with the following parameters: Match = 2, Mismatch = 7, Delta = 7, PM = 80, PI = 10, Minscore = 50 and MaxPeriod = 2000. A de novo repeat library was built by the LTR_FINDER software (version 1.0.6, parameter: -w 2; University of Fudan, Shanghai, China). Subsequently, the RepeatMasker software was utilized to align our genome sequences onto the Repbase TE (version 3.2.9; Genetic Information Research Institute, Mountain View, CA, USA) [42] to search for known repeat sequences, which were also mapped onto the de novo repeat libraries to identify novel types of repeat sequences.

We then annotated the *S. chinensis* genome assembly using three routine approaches, including homology-based, transcriptome-based and ab initio annotations. We selected several representative animal species to perform the homology annotation, including the Baiji dolphin (*Lipotes vexillifer*), sperm whale (*Physeter catodon*), bottlenose dolphin (*Tursiops truncatus*), cattle (*Bos Taurus*), human (*Homo sapiens*), mouse (*Mus musculus*), sheep (*Ovis aries*), pig (*Sus scrofa*) and zebrafish (*Danio rerio*). Related protein sequences were aligned onto our genome sequences utilizing TblastN [43] with an E-value < $1.0e^{-5}$. Genewise 2.2.0 [44] was subsequently employed to predict possible gene structures based on the TblastN results.

Total RNA extracted from leucocytes was sequenced on an Illumina HiSeq 4000 platform. These transcriptome reads were aligned onto our genome assembly using HISAT [45]. We utilized Cufflinks (version 2.2.1; University of Maryland, College Park, MD, USA) [46] to identify the preliminary genes. Meanwhile, Augustus [47] and Genscan [48] were employed for ab initio annotation by using the repeat-masked genome sequences. Finally, we employed GLEAN [49] to integrate all predicted genes from the three annotation procedures.

All protein sequences of the GLEAN results were mapped onto the public TrEMBL, SwissProt [50] and KEGG [51] databases using BLASTP with an E-value $\leq 1.0e^{-5}$. We also applied the InterProScan [52] to predict potential functions of these protein sequences with Pfam [53], PRINTS [54], PANTHER [55], ProDom [56] and SMART [57].

4.4. Phylogenetic Relationships of the Chinese White Dolphin

To understand the phylogenetic relationships of Chinese white dolphin, we constructed a phylogenetic tree using Chinese white dolphin and other ten mammals, including six Cetacea species, i.e., Beluga whale (*Delphinapterus leucas*), Yangtze finless porpoise (*Neophocaena asiaeorientalis*), Baiji dolphin, bottlenose dolphin, sperm whale and minke whale (*Balaenoptera acutorostrata*)), as well as

two Artiodactyla species (cattle and sheep) and two Euarchontoglires species (human and mouse). Whole-genome gene sets for the other ten mammals were available online, and thus we downloaded them from NCBI. These datasets and the gene sets predicted from the Chinese white dolphin genome were aligned with each other by BLAST (version 2.2.6; Genome Research Center, Cold Spring Harbor, NY, USA) [43] to determine homologous genes. In this way, we obtained 5728 single-copy gene families with 63,008 genes in total. These single-copy genes from each species were concatenated together to constitute a super-length gene that yielded 3,657,951 aligned sites. Finally, the four-fold degenerated sites were extracted from these aligned sites to construct the phylogenetic tree with the maximum likelihood (ML) method using PhyML (version 3.0; Université de Montpellier, Montpellier, France) [58].

4.5. Molecular Dating

Bayesian molecular dating was adopted to estimate the neutral evolutionary rate and species divergence time with MCMCTREE from PAML (version 4.4b; University College London, London, UK) [59]. Five nodes (C1~C5) were considered as time-calibrated points with normal distributions and soft constraint bands, allowing a small probability (0.025) of violation. Based on previous research [35], we calibrated the phylogenetic tree with two time points, 61.7~100.5 Mya for Human–Mouse and 71.2–113 Mya for Laurasiatheria and Euarchontoglires, which were retrieved from the TimeTree database [60]. The divergence time between Cetacea and Artiodactyla using the Ypresian fossil Pakicetus (Eocene: 55.8–48.6 Mya) [61,62] was applied as the C3 calibrate point; the C4 calibration point was 10~30 Mya for sheep and cattle based on a previous report [27]; and the divergence between Phocoenidae (such as Finless porpoise) and Monodontidae (such as Beluga whale), calibrated on the basis of the earliest record of a Phocoenid (10.0~11.2 Mya) [63], was chosen as the C5 calibration point. The MCMC (Markov chain Monte Carlo) chain length was set to 200,000 generations and sampled every 200. The first 20% of samples were discarded as burns in. An independent rate model (clock = 2) following a lognormal distribution was employed for the MCMC search.

4.6. Identification of Gene Family Changes

To determine the evolutionary dynamics of gene families, especially for expansion and contraction of gene ortholog clusters, we employed the program CAFÉ [64] to identify gene family changes among the Chinese white dolphin, Beluga whale, finless porpoise, Baiji dolphin, bottlenose dolphin, sperm whale, minke whale, cattle, sheep, human and mouse.

4.7. Prediction of Historical Population

We realigned the reads from short-insert libraries (500 and 800 bp) onto our genome assembly using the package SOAP [65]. Subsequently, we called heterozygous SNPs (single nucleotide polymorphisms) using the package SOAPSnp [66] with an optimized threshold (coverage depth ≥ 4 and ≤ 150 , genotype quality ≥ 20 , copy number ≤ 2 and distance of adjacent SNPs ≥ 5). We then employed the Hidden Markov model (HMM) approach to implement Pairwise Sequentially Markovian Coalescence (PSMC) on the basis of SNP distribution [31].

We used these heterozygous SNPs to reconstruct a demographic history. The generation time ($g = 12$ years) and neutral mutation rate per generation ($\mu = 1.5 \times 10^{-8}$) were based on a previous report [35]. We obtained atmospheric surface air temperature ($^{\circ}\text{C}$) and global relative sea level (10 m) data of the past one million years from National Climatic Data Center (NCDC), and then combined them together with the demographic data to generate a single plot. However, it is difficult for the PSMC simulation to detect population changes within 10,000 years ago. We hence didn't predict historical population during this recent period.

4.8. Identification of AHTPs

The top 50 previously reported AHTPs with high antihypertensive activities (Table S2) were selected to map our target protein datasets. These protein datasets were downloaded from NCBI

and Ensemble for four mammals with different living habitats (Table S3), including cow (on land), minke whale (in deep sea), Yangtze River dolphin (in freshwater), and Bottlenose dolphin (in shallow sea). In-house scripts were compiled to identify AHTP-mapped proteins by in situ mapping, and the localization of each target peptide was marked for further statistics and analysis.

5. Conclusions

In summary, we report a high-quality genome with a relatively complete gene set for the Chinese white dolphin, an endangered nearshore marine mammal in China. Expansion of certain gene families, especially the increase in immune and sensory genes, could partly shed light on the molecular mechanisms for adaptation to the nearshore environments. We also observed a serious bottleneck in the demographic population history of Chinese white dolphin about 350,000 years ago. The identification of AHTPs broadens our knowledge about the potential of mammal proteins for development of antihypertensive peptides. Our genome assembly will provide a genetic resource for further researches on adaptive ecology, conservation biology of cetaceans and development of marine peptide-based drugs for treatment of various human cardiovascular diseases.

Supplementary Materials: The following materials are available online at <http://www.mdpi.com/1660-3397/17/9/504/s1>. Figure S1: Molecular dating of the Chinese white dolphin and other ten examined mammals. Table S1: Statistics of raw reads and clean data for the whole genome sequencing. Table S2: Summary of repeat sequences in the assembled genome. Table S3: Statistics of functional annotations. Table S4: A local database of AHTPs. Table S5: Summary of the downloaded protein datasets for other four mammals. Table S6: Mapped AHTPs in the five mammalian proteome datasets. Table S7: Hit numbers of AHTPs in the five examined mammals. Table S8: GO annotation of AHTP-mapped proteins in the five examined mammals. Table S9: Numbers of each mapped AHTP in the five examined mammals.

Author Contributions: Y.W., M.Y. and Q.S. conceived the project; P.J. isolated the genomic DNA; C.B., K.J., Y.Y., Y.L., Y.L., J.L. and X.Y. performed data analysis; P.J., D.G., W.L., X.Z. and X.S. participated in data analysis and figure preparation; C.B., K.J. and Y.Y. prepared the manuscript; Y.W., Q.S. and M.Y. revised the manuscript; and all the authors have read and approved the manuscript.

Funding: This work was supported by Natural Science Foundation of China (No. 41576128), Natural Science Foundation of Guangdong Province in China (No. 2017A030308005), Zhuhai Scholar Professor Program (2015), and Shenzhen Special Program for Upgrading Key Links to Strategies for the Emerging and Future Industries (No. 20170428173357698).

Conflicts of Interest: The authors declare no conflict of interest.

References

- Xu, X.; Song, J.; Zhang, Z.; Li, P.; Yang, G.; Zhou, K. The world's second largest population of humpback dolphins in the waters of Zhanjiang deserves the highest conservation priority. *Sci. Rep.* **2015**, *5*, 8147. [[CrossRef](#)] [[PubMed](#)]
- Jefferson, T.A.; Smith, B.D. 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*; Jefferson, T.A., Curry, B.E., Eds.; Academic Press: Cambridge, MA, USA, 2016; Volume 73, pp. 1–26. ISBN 0065-2881.
- Minton, G.; Zulkifli Poh, A.N.; Peter, C.; Porter, L.; Krebs, D. Chapter Six Indo-Pacific humpback dolphins in Borneo: A review of current knowledge with emphasis on Sarawak. In *Advances in Marine Biology*; Jefferson, T.A., Curry, B.E., Eds.; Academic Press: Cambridge, MA, USA, 2016; Volume 73, pp. 141–156. ISBN 0065-2881.
- Huang, S.-L.; Karczmarski, L.; Chen, J.; Zhou, R.; Lin, W.; Zhang, H.; Li, H.; Wu, Y. Demography and population trends of the largest population of Indo-Pacific humpback dolphins. *Biol. Conserv.* **2012**, *147*, 234–242. [[CrossRef](#)]
- Steeman, M.E.; Hebsgaard, M.B.; Fordyce, R.E.; Ho, S.Y.; Rabosky, D.L.; Nielsen, R.; Rahbek, C.; Sorensen, M.V.; Willerslev, E. Radiation of extant cetaceans driven by restructuring of the oceans. *Syst. Biol.* **2009**, *58*, 573–585. [[CrossRef](#)] [[PubMed](#)]
- Pelc, R.A.; Warner, R.R.; Gaines, S.D. Geographical patterns of genetic structure in marine species with contrasting life histories. *J. Biogeogr.* **2009**, *36*, 1881–1890. [[CrossRef](#)]

7. Noren, S.R.; Williams, T.M. Body size and skeletal muscle myoglobin of cetaceans: Adaptations for maximizing dive duration. *Comp. Biochem. Physiol. Part A Mol. Integr. Physiol.* **2000**, *126*, 181–191. [[CrossRef](#)]
8. Hoffman, J.M.; Ponnampalam, L.S.; Araújo, C.C.; Wang, J.Y.; Kuit, S.H.; Hung, S.K. Comparison of Indo-Pacific humpback dolphin (*Sousa chinensis*) whistles from two areas of western Peninsular Malaysia. *J. Acoust. Soc. Am.* **2015**, *138*, 2829–2835. [[CrossRef](#)] [[PubMed](#)]
9. Gui, D.; Jia, K.; Xia, J.; Yang, L.; Chen, J.; Wu, Y.; Yi, M. De novo assembly of the Indo-Pacific humpback dolphin leucocyte transcriptome to identify putative genes involved in the aquatic adaptation and immune response. *PLoS ONE* **2013**, *8*, e72417. [[CrossRef](#)]
10. Arnason, U.; Gullberg, A.; Janke, A. Mitogenomic analyses provide new insights into cetacean origin and evolution. *Gene* **2004**, *333*, 27–34. [[CrossRef](#)]
11. LeDuc, R.G.; Perrin, W.F.; Dizon, A.E. Phylogenetic relationships among the delphinid cetaceans based on full cytochrome b sequences. *Mar. Mammal. Sci.* **1999**, *15*, 619–648. [[CrossRef](#)]
12. May-Collado, L.; Agnarsson, I. Cytochrome b and Bayesian inference of whale phylogeny. *Mol. Phylogenet. Evol.* **2006**, *38*, 344–354. [[CrossRef](#)]
13. Montgelard, C.; Catzeflis, F.M.; Douzery, E. Phylogenetic relationships of artiodactyls and cetaceans as deduced from the comparison of cytochrome b and 12S rRNA mitochondrial sequences. *Mol. Phylogenet. Evol.* **1997**, *14*, 550–559. [[CrossRef](#)] [[PubMed](#)]
14. Orbach, D.N.; Hedrick, B.; Würsig, B.; Mesnick, S.L.; Brennan, P.L.R. The evolution of genital shape variation in female cetaceans. *Evolution* **2018**, *72*, 261–273. [[CrossRef](#)] [[PubMed](#)]
15. Park, T.; Evans Alistair, R.; Gallagher Stephen, J.; Fitzgerald Erich, M.G. Low-frequency hearing preceded the evolution of giant body size and filter feeding in baleen whales. *Proc. R. Soc. B Biol. Sci.* **2017**, *284*, 20162528.
16. 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. Convergent evolution of the genomes of marine mammals. *Nat. Genet.* **2015**, *47*, 272–275. [[CrossRef](#)] [[PubMed](#)]
17. 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. Minke whale genome and aquatic adaptation in cetaceans. *Nat. Genet.* **2013**, *46*, 88–92. [[CrossRef](#)] [[PubMed](#)]
18. Gui, D.; Zhou, R.; Lin, W.; Yu, X.; Zhang, X.; Wu, Y. Low major histocompatibility complex class II variation in the endangered Indo-Pacific humpback dolphin (*Sousa chinensis*): Inferences about the role of balancing selection. *J. Hered.* **2016**, *107*, 143–152.
19. Jia, K.; Ding, L.; Zhang, L.; Zhang, M.; Yi, M.; Wu, Y. In vitro assessment of environmental stress of persistent organic pollutants on the Indo-Pacific humpback dolphin. *Toxicol. In Vitro* **2015**, *30*, 529–535. [[CrossRef](#)]
20. Brown, A.M.; Kopps, A.M.; Allen, S.J.; Bejder, L.; Littleford-Colquhoun, B.; Parra, G.J.; Cagnazzi, D.; Thiele, D.; Palmer, C.; Frère, C.H. Population differentiation and hybridisation of Australian snubfin (*Orcaella heinsolmi*) and Indo-Pacific humpback (*Sousa chinensis*) dolphins in North-Western Australia. *PLoS ONE* **2014**, *9*, e101427. [[CrossRef](#)]
21. Ming, Y.; Jian, J.; Yu, F.; Yu, X.; Wang, J.; Liu, W. Molecular footprints of inshore aquatic adaptation in Indo-Pacific humpback dolphin (*Sousa chinensis*). *Genomics* **2018**, in press. [[CrossRef](#)]
22. Majumder, K.; Wu, J. Molecular targets of antihypertensive peptides: Understanding the mechanisms of action based on the pathophysiology of hypertension. *Int. J. Mol. Sci.* **2015**, *16*, 256–283. [[CrossRef](#)]
23. Lee, S.Y.; Hur, S.J. Antihypertensive peptides from animal products, marine organisms, and plants. *Food Chem.* **2017**, *228*, 506–517. [[CrossRef](#)]
24. Hussain, M.; Awan, F.R. Hypertension regulating angiotensin peptides in the pathobiology of cardiovascular disease. *Clin. Exp. Hypertens.* **2018**, *40*, 344–352. [[CrossRef](#)] [[PubMed](#)]
25. Jin, W.; Jia, K.; Yang, L.; Chen, J.; Wu, Y.; Yi, M. Derivation and characterization of cell cultures from the skin of the Indo-Pacific humpback dolphin *Sousa chinensis*. *Vitr. Cell. Dev. Biol. Anim.* **2013**, *49*, 449–457. [[CrossRef](#)]
26. Kriventseva, E.V.; Zdobnov, E.M.; Simão, F.A.; Ioannidis, P.; Waterhouse, R.M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **2015**, *31*, 3210–3212.
27. Xiong, Y.; Brandley, M.C.; Xu, S.; Zhou, K.; Yang, G. Seven new dolphin mitochondrial genomes and a time-calibrated phylogeny of whales. *BMC Evol. Biol.* **2009**, *9*, 20. [[CrossRef](#)] [[PubMed](#)]
28. Thewissen, J.G.M.; Cooper, L.N.; Clementz, M.T.; Bajpai, S.; Tiwari, B.N. Whales originated from aquatic artiodactyls in the Eocene epoch of India. *Nature* **2007**, *450*, 1190–1194. [[CrossRef](#)]

29. Jefferson, T.A.; Hung, S.K. A Review of the status of the Indo-Pacific humpback dolphin (*Sousa chinensis*) in Chinese Waters. *Aquat. Mamm.* **2004**, *30*, 149–158. [[CrossRef](#)]
30. Chen, B.; Zheng, D.; Zhai, F.; Xu, X.; Sun, P.; Wang, Q.; Yang, G. Abundance, distribution and conservation of Chinese white dolphins (*Sousa chinensis*) in Xiamen, China. *Mamm. Biol.* **2008**, *73*, 156–164. [[CrossRef](#)]
31. Li, H.; Durbin, R. Inference of human population history from individual whole-genome sequences. *Nature* **2011**, *475*, 493–496. [[CrossRef](#)]
32. 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. The novel evolution of the sperm whale genome. *Genome Biol. Evol.* **2017**, *9*, 3260–3264. [[CrossRef](#)]
33. Zhou, X.; Guang, X.; Sun, D.; Xu, S.; Li, M.; Seim, I.; Jie, W.; Yang, L.; Zhu, Q.; Xu, J.; et al. Population genomics of finless porpoises reveal an incipient cetacean species adapted to freshwater. *Nat. Commun.* **2018**, *9*, 1276. [[CrossRef](#)] [[PubMed](#)]
34. Yi, Y.; Lv, Y.; Zhang, L.; Yang, J.; Shi, Q. High throughput identification of antihypertensive peptides from fish proteome datasets. *Mar. Drugs* **2018**, *16*, 365. [[CrossRef](#)] [[PubMed](#)]
35. Zhou, X.; Sun, F.; Xu, S.; Fan, G.; Zhu, K.; Liu, X.; Chen, Y.; Shi, C.; Yang, Y.; Huang, Z.; et al. Baiji genomes reveal low genetic variability and new insights into secondary aquatic adaptations. *Nat. Commun.* **2013**, *4*, 2708. [[CrossRef](#)] [[PubMed](#)]
36. Panneton, W.M. The mammalian diving response: An enigmatic reflex to preserve life? *Physiology* **2013**, *28*, 284–297. [[CrossRef](#)] [[PubMed](#)]
37. Tian, R.; Wang, Z.; Niu, X.; Zhou, K.; Xu, S.; Yang, G. Evolutionary genetics of hypoxia tolerance in cetaceans during diving. *Genome Biol. Evol.* **2016**, *8*, 827–839. [[CrossRef](#)] [[PubMed](#)]
38. Luo, R.; Liu, B.; Xie, Y.; Li, Z.; Huang, W.; Yuan, J.; He, G.; Chen, Y.; Pan, Q.; Liu, Y.; et al. SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler. *GigaScience* **2012**, *1*, 18. [[CrossRef](#)]
39. Benson, G. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **1999**, *27*, 573–580. [[CrossRef](#)] [[PubMed](#)]
40. Xu, Z.; Wang, H. LTR_FINDER: An efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **2007**, *35*, W265–W268. [[CrossRef](#)]
41. Tarailo-Graovac, M.; Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinform.* **2009**, *25*. [[CrossRef](#)]
42. Jurka, J.; Kapitonov, V.V.; Pavlicek, A.; Klonowski, P.; Kohany, O.; Walichiewicz, J. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **2005**, *110*, 462–467. [[CrossRef](#)]
43. Mount, D.W. Using the Basic Local Alignment Search Tool (BLAST). *Cold Spring Harb. Protoc.* **2007**, *2007*. [[CrossRef](#)] [[PubMed](#)]
44. Birney, E.; Clamp, M.; Durbin, R. GeneWise and Genomewise. *Genome Res.* **2004**, *14*, 988–995. [[CrossRef](#)] [[PubMed](#)]
45. Kim, D.; Langmead, B.; Salzberg, S.L. HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **2015**, *12*, 357–360. [[CrossRef](#)] [[PubMed](#)]
46. Trapnell, C.; Williams, B.A.; Pertea, G.; Mortazavi, A.; Kwan, G.; van Baren, M.J.; Salzberg, S.L.; Wold, B.J.; Pachter, L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **2010**, *28*, 511–515. [[CrossRef](#)] [[PubMed](#)]
47. Stanke, M.; Keller, O.; Gunduz, I.; Hayes, A.; Waack, S.; Morgenstern, B. AUGUSTUS: Ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **2006**, *34*, W435–W439. [[CrossRef](#)] [[PubMed](#)]
48. Cai, Y.; González, J.V.; Liu, Z.; Huang, T. Computational systems biology methods in molecular biology, chemistry biology, molecular biomedicine, and biopharmacy. *BioMed. Res. Int.* **2014**, *2014*, 746814. [[CrossRef](#)]
49. Elsik, C.G.; Mackey, A.J.; Reese, J.T.; Milshina, N.V.; Roos, D.S.; Weinstock, G.M. Creating a honey bee consensus gene set. *Genome Biol.* **2007**, *8*, R13. [[CrossRef](#)]
50. Bairoch, A.; Estreicher, A.; Boeckmann, B.; O'Donovan, C.; Gasteiger, E.; Phan, I.; Michoud, K.; Martin, M.J.; Blatter, M.-C.; Schneider, M.; et al. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* **2003**, *31*, 365–370.
51. Kanehisa, M.; Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30. [[CrossRef](#)]

52. Hunter, S.; Apweiler, R.; Attwood, T.K.; Bairoch, A.; Bateman, A.; Binns, D.; Mitchell, A.; Quinn, A.F.; Laugraud, A.; Wu, C.H.; et al. InterPro: The integrative protein signature database. *Nucleic Acids Res.* **2008**, *37*, D211–D215. [[CrossRef](#)]
53. Finn, R.D.; Bateman, A.; Clements, J.; Coggill, P.; Eberhardt, R.Y.; Eddy, S.R.; Heger, A.; Hetherington, K.; Holm, L.; Mistry, J.; et al. Pfam: The protein families database. *Nucleic Acids Res.* **2013**, *42*, D222–D230. [[CrossRef](#)] [[PubMed](#)]
54. Attwood, T.K. The PRINTS database: A resource for identification of protein families. *Brief. Bioinform.* **2002**, *3*, 252–263. [[CrossRef](#)] [[PubMed](#)]
55. Thomas, P.D.; Kejariwal, A.; Campbell, M.J.; Mi, H.; Diemer, K.; Guo, N.; Ladunga, I.; Ulitsky-Lazareva, B.; Muruganujan, A.; Rabkin, S.; et al. PANTHER: A browsable database of gene products organized by biological function, using curated protein family and subfamily classification. *Nucleic Acids Res.* **2003**, *31*, 334–341. [[CrossRef](#)] [[PubMed](#)]
56. Bru, C.; Courcelle, E.; Carrère, S.; Beausse, Y.; Dalmar, S.; Kahn, D. The ProDom database of protein domain families: More emphasis on 3D. *Nucleic Acids Res.* **2005**, *33*, D212–D215. [[CrossRef](#)]
57. Letunic, I.; Copley, R.R.; Schmidt, S.; Ciccarelli, F.D.; Doerks, T.; Schultz, J.; Ponting, C.P.; Bork, P. SMART 4.0: Towards genomic data integration. *Nucleic Acids Res.* **2004**, *32*, D142–D144. [[PubMed](#)]
58. Guindon, S.; Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **2003**, *52*, 696–704. [[CrossRef](#)] [[PubMed](#)]
59. Yang, Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* **2007**, *24*, 1586–1591. [[CrossRef](#)] [[PubMed](#)]
60. Hedges, S.B.; Dudley, J.; Kumar, S. TimeTree: A public knowledge-base of divergence times among organisms. *Bioinformatics* **2006**, *22*, 2971–2972. [[CrossRef](#)]
61. Gingerich, P.D.; Russell, D.E. Pakicetus inachus, a new archaeocete (Mammalia, Cetacea) from the early-middle Eocene Kuldana Formation of Kohat (Pakistan). *Mus. Paleontol. Univ. Mich.* **1981**, *25*, 235–246.
62. Thewissen, J.G.M.; Williams, E.M.; Roe, L.J.; Hussain, S.T. Skeletons of terrestrial cetaceans and the relationship of whales to artiodactyls. *Nature* **2001**, *413*, 277–281. [[CrossRef](#)] [[PubMed](#)]
63. Barnes, L.G. Evolution, taxonomy and antitropical distributions of the porpoises (phocoenidae, mammalia). *Mar. Mammal. Sci.* **1985**, *1*, 149–165. [[CrossRef](#)]
64. De Bie, T.; Cristianini, N.; Demuth, J.P.; Hahn, M.W. CAFE: A computational tool for the study of gene family evolution. *Bioinformatics* **2006**, *22*, 1269–1271. [[CrossRef](#)] [[PubMed](#)]
65. Li, R.; Yu, C.; Li, Y.; Lam, T.-W.; Yiu, S.-M.; Kristiansen, K.; Wang, J. SOAP2: An improved ultrafast tool for short read alignment. *Bioinformatics* **2009**, *25*, 1966–1967. [[CrossRef](#)] [[PubMed](#)]
66. Li, R.; Li, Y.; Fang, X.; Yang, H.; Wang, J.; Kristiansen, K.; Wang, J. SNP detection for massively parallel whole-genome resequencing. *Genome Res.* **2009**, *19*, 1124–1132. [[CrossRef](#)] [[PubMed](#)]

