

## RESEARCH ARTICLE OPEN ACCESS

# Low Heterozygosity and Historical Bottleneck Effect Depicted From the Genome Assembly of the Indus River Dolphin (*Platanista minor*)

Aamir Ibrahim<sup>1</sup> | Simin Chai<sup>2</sup> | Cuijuan Zhong<sup>1</sup> | Kang Jieqiong<sup>1</sup> | Ahsaan Ali<sup>1</sup> | Sajjad Hussain<sup>1,3</sup> | Hassan Ali<sup>3</sup> | Tanveer Hussain<sup>4</sup> | Umer Waqas<sup>5</sup> | Guang Yang<sup>1,2</sup> 

<sup>1</sup>Jiangsu Key Laboratory for the Biodiversity Conservation and Sustainable Utilization in the Middle and Lower Reaches of the Yangtze River Basin, College of Life Sciences, Nanjing Normal University, Nanjing, China | <sup>2</sup>Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Guangzhou, Guangdong, China | <sup>3</sup>Wildlife and Parks Department, Lahore, Punjab, Pakistan | <sup>4</sup>Department of Biological Sciences, Virtual University of Pakistan, Islamabad, Pakistan | <sup>5</sup>Virtual University of Pakistan, Lahore, Pakistan

**Correspondence:** Guang Yang ([gyang@njnu.edu.cn](mailto:gyang@njnu.edu.cn))

**Received:** 28 August 2024 | **Revised:** 29 April 2025 | **Accepted:** 8 May 2025

**Funding:** This work was supported by Priority Academic Program Development of Jiangsu Higher Education Institutions. PI Project of Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), GML2021GD0805. Key Project of the National Natural Science Foundation of China, 32030011. National Key Programme of Research and Development, Ministry of Science and Technology, 2022YFF1301600.

**Keywords:** demography | endangered | freshwater adaptation | homozygosity | threats

## ABSTRACT

The Indus River dolphin (*Platanista minor*) is a highly endangered freshwater dolphin endemic to the Indus River system of the Indian subcontinent. We reported a *de novo* assembly and characterization of the draft genome of the Indus River dolphin by using Illumina short-read sequencing technology. Based on this, for the first time, we conducted the comparative genomics study and identified a selection of genes and gene families that have undergone significant positive selection and expansion or contraction, indicating potential molecular mechanisms associated with freshwater adaptation, such as specialized skin features and their derivatives (e.g., hair loss) and immune adaptations. Additionally, this study estimated that the Indus River dolphin diverged nearly 31.2 million years ago from the most recent common ancestor of Delphinidae and Lipotidae, placing it in a more basal position to other freshwater dolphins (e.g., the baiji *Lipotes vexillifer*). It was suggested that the combined effects of the natural historical bottleneck effect around 40–20 kiloyears ago and anthropogenic activities were the driving factors of inbreeding for this species with very low heterozygosity (0.0218%).

## 1 | Introduction

The riverine *Platanista* lineage is an exclusively extant member of the extinct Platanistidae (de Muizon 1988, Viglino et al. 2021), which might have escaped extinction due to an earlier transition from a marine to a freshwater environment (Hamilton et al. 2001). The phylogeny of the extant *Platanista*

spp. has been a subject of longstanding debate. Now, it is accepted that *Platanista* is an early-diverging lineage, occupying the most basal phylogenetic position among river dolphins; and the observed convergent features among riverine taxa are plesiomorphic characters (Árnason and Gullberg 1996; Cassens et al. 2000; Yan et al. 2005; Werth 2006; McGowen et al. 2009; Steeman et al. 2009; McCurry et al. 2017).

Aamir Ibrahim and Simin Chai contributed equally to this work.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *Ecology and Evolution* published by British Ecological Society and John Wiley & Sons Ltd.

It has been suggested that *Platanista* evolved during the late Oligocene, with high diversity recorded in the early Miocene and markedly decreased by the middle-late Miocene (de Muizon 1988; Fordyce and de Muizon 2001). During the late Neogene epoch, *Platanista* underwent a transition to the freshwater environment of the Ganges River system before subsequently migrating to the Indus River system (Hamilton et al. 2001). The presence of distinct geographical distribution and reproductive isolation has resulted in the recognition of two separate species: the Ganges River dolphin (*Platanista gangetica*) endemic to the Ganges River system, and the Indus River dolphin (*P. minor*) (Figure 1) endemic to the Indus River system; both are endangered species (Braulik et al. 2015; Braulik et al. 2021). Specifically, the drastic negative anthropogenic impacts such as hunting and development of the Indus Basin irrigation system (after 1885) drove the Indus River dolphin to the brink of extinction (Anderson 1878; Braulik et al. 2015). It is gratifying to note that, since the 1970s, conservation efforts have enabled its resurgence, with current estimates exceeding 2000 individuals throughout their distribution range (WWF-Pakistan 2017).

Although former studies focused on the mitochondrial genome of one subpopulation (Sindh Province, Pakistan) and reported high homozygosity (Braulik et al. 2015), investigations into other subpopulations are lacking. Fragmented, smaller, and isolated populations are prone to a higher risk of extinction due to increased inbreeding, resulting in reduced genetic diversity and diminished adaptive potential to sustained viability (Lande 1993; Reed and Frankham 2003; Willi et al. 2006; Melbourne and Hastings 2008). Importantly, evolutionary and demographic histories are essential for the development of an efficient conservation strategy (Nigenda-Morales et al. 2023), although they are still undisclosed for the Indus River dolphin.

Over the last four decades, genetics has played a vital role in promoting the conservation of endangered species. The genetics data has assessed the genetic variation at individual and population levels; the genetics has also provided deep insights in diverse areas of conservation biology including species identification, hybridization, kinship, evolutionary history, effective population size ( $N_e$ ),



**FIGURE 1** | The Indus river dolphin in natural habitat, Indus River, Pakistan (Picture Credit: Dr. Aamir Ibrahim).

population substructure, population connectivity, and inbreeding (Hedrick and Miller 1992; von Der Heyden et al. 2014; Haig et al. 2016). To address these issues, whole genome sequencing serves as a valuable tool for uncovering the unknown biological characteristics of species. A range of cetaceans have been examined through genomics research at both the individual and population levels. For example, through whole genome sequencing and comparative genomics analyses of the Baiji (*Lipotes vexillifer*), the flagship species of the Yangtze River in China, insights into molecular adaptation, decreased heterozygosity, and a historical bottleneck have been gained (Zhou et al. 2013). Other examples include the sperm whale (*Physeter macrocephalus*) (Warren et al. 2017), the bottlenose dolphin (*Tursiops truncatus*) (Foote et al. 2015), the Yangtze finless porpoise (*Neophocaena asiaeorientalis*) (Zhou et al. 2018), and the pygmy right whale (*Caperea marginata*) (Wolf et al. 2023), and so forth. However, whole genome sequencing and annotation are rare for river dolphins. Especially, only one available draft genome is accessible for the Indus River dolphin from an unknown subpopulation ([https://www.ncbi.nlm.nih.gov/datasets/genome/GCA\\_004363435.1/](https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_004363435.1/)).

In this study, we reported the newly assembled draft genome and annotation of the Indus River dolphin, which was sequenced at approximately 100× sequencing coverage. Comparative genomic analyses yielded insights into molecular adaptive characteristics in a riverine habitat and shed light on its enigmatic demographic population history. This study will contribute to the genetic conservation efforts for this highly endangered riverine dolphin.

## 2 | Material and Methods

### 2.1 | Sampling and Sequencing

The source of the current genomic study was conducted based on the single adult female Indus River dolphin. This individual was collected on (January-14-2022) from the subpopulation between the Taunsa–Guddu barrage (28.921437°N, 70.541958°E) in the Punjab River section (Pakistan). As this species has endangered status and rarely has availability of fresh samples for genetic study, we preserved it at  $-20^{\circ}\text{C}$  until performing DNA extraction and genome sequencing on June 15, 2023. In our collection, only one adult female sample adequately passed the quality to avail only illumina short-read sequencing technology. Deep skeletal muscle tissues from the abdomen region were selected to process for genomic study, and DNA extraction was performed using the phenol/chloroform DNA extraction method. Paired-end sequencing libraries, with an insert size of  $\sim 350$  bp, were constructed and sequenced by the independent sequencing platform zero PCR DNBSEQ from BGI (Quail et al. 2012).

### 2.2 | Genome Assembly and Annotation

To assemble and annotate the newly sequenced Indus River dolphin genome, SOAPnuke (v1.6.5) was used to filter data and get clean reads, including (1) removing reads with 25% low-quality bases (quality scores  $\leq 7$ ), (2) removing reads with N bases more than 1%, (3) discarding reads with adapter contamination and/or polymerase chain reaction duplicates (Chen et al. 2018). We used the Jellyfish (v2.1.4) (Luo et al. 2012)

software based on a *K*-mer distribution and GenomeScope2 (v2.0) (Vurtture et al. 2017) to estimate the genome size with high-quality reads >Q20. The Platanus (v1.2.4) (Kajitani et al. 2014) was used for *de novo* genome assembly and scaffold construction using the default parameters. To evaluate the quality of the present genome assembly, we assessed its integrity using Benchmarking Universal Single-Copy Orthologs (BUSCO v5.2.2) with a database of mammals (mammalia\_odb10) (Simão et al. 2015).

We annotated the repeat sequences in the present Indus River dolphin genome assembly using both *de novo* predictions and homology-based searching in the known repeat database. RepeatMasker (v4.1.1) and RepeatProteinMasker (v4.1.1) RepBase23.06 (<https://www.girinst.org/repbase/>) were applied to identify repetitive elements at the DNA and protein levels, respectively. RepeatModeler (v2.0.3) was performed for *de novo* prediction, and Tandem Repeats Finder (v4.09.1) (Benson 1999) was applied to annotate tandem repeats.

Protein-coding genes were annotated using homology-based and *de novo* prediction. We annotated protein-coding genes using MAKER (v3.01.03) (Cantarel et al. 2008) based on three independent approaches, including homology-based, ab initio, and RNA sequencing-assisted prediction. The GeMoMa (v1.9) (Keilwagen et al. 2016) software was used to align homologous peptides from cattle *Bos taurus* (GCF\_002263795.3), human *Homo sapiens* (GCF\_000001405.40), baiji (GCF\_000442215.2), mouse *Mus musculus* (GCF\_000001635.27), sperm whale (GCF\_002837175.3), and bottlenose dolphin *Tursiops truncatus* (GCF\_011762595.1) to the assembly and infer the gene structure information. For RNA seq-based gene prediction, mRNA-seq reads from the closely related bottlenose dolphin (DRR513290, DRR513294, DRR513301, and DRR513302) were aligned to the assembly and then assembled to predict open reading frames using Stringtie (v2.2.1) (Pertea et al. 2015). Augustus (Stanke et al. 2008), GeneMark-ETP (Brůna et al. 2020), and SNAP (Korf 2004) were then used for ab initio gene prediction on the training set from GeMoMa.

Gene function information was assigned by comparing them with public databases including Swiss-Prot, Nr, KEGG, and Gene Ontology. To assess the non-coding rRNA, tRNA, snRNA, snoRNA, and microRNA, we used INFERNAL (v1.1.4) to search the Rfam database (Nawrocki et al. 2009).

Through the aforementioned methods of homology-based gene structure annotation and gene function annotation, we proceed to annotate the protein-coding genes of the publicly available draft genome assembly GCA\_004363435.1.

### 2.3 | Phylogeny, Demographic History and Heterozygosity Level

The phylogenetic tree was reconstructed based on the coding sequences of 8806 single-copy orthologous genes identified by Orthofinder (v2.5.3) (Emms and Kelly 2015) with default parameters from the Indus River dolphin, the sperm whale, the blue whale *Balaenoptera musculus*, the Baiji, the bottlenose dolphin, the cattle, the mouse, with the human as an outgroup. We used

MAFFT (v7.520) (Katoh and Standley 2013) for multiple sequence alignment and used RAxML (v8.2.12) (Stamatakis 2014) to reconstruct the phylogenetic tree. The species divergence time was estimated using MCMCTREE in PAML (v4.9) (Yang 2007). Three calibration points were retrieved from Timetree (<http://www.timetree.org/>), including human ~ the mouse (81.3–91.0 MYA), the blue whale ~ the bottlenose dolphin (32.3–35.2 MYA), the sperm whale ~ the bottlenose dolphin (32.1–34.3 MYA). MCMCTREE with parameters “clock=3, model=0, alpha=0, ncatG=4, BDparas=1 1 0, finetune=0.05 0.1 0.12 0.1 0.3, burnin=20000, nsample=100000” was used to infer divergence time.

For the demographic history, we combined both the already available genome and newly generated genome of the Indus River dolphin; the data was inferred with the PSMC model (Li and Durbin 2009). We set  $N=25$ ,  $t=15$ ,  $r=5$ , and  $p=“4+25*2+4+6”$  for PSMC analysis with 100 bootstrap replicates. The value of  $N$ , set at 25, signifies that the maximum number of iterations is 25; with the assumption of 15 years per generation, the notation  $t=15$  represents the conversion of a generation into years;  $r$  is the ratio of initial theta to rho, set at 5 and near the default setting;  $p$  signifies the pattern of parameters and was set as default. GATK was used for the single nucleotide polymorphisms (SNPs) calling and filtering (Poplin et al. 2017). The heterozygosity level was calculated as the ratio of the number of heterozygous SNPs divided by the number of totally filtered SNPs.

### 2.4 | Comparative Genomics and Evolutionary Analyses

We employed PAML (v4.9) (Yang 2007) to detect positively selected genes (PSGs) using the branch-site model. This model aims to detect positive selection that affects only a few sites on the prespecified foreground branch(es), the Indus River dolphin in the present study. For the background branch, there are two classes of sites, the conserved sites with  $0 < \omega_0 < 1$  and the neutral sites with  $\omega_1 = 1$ . For the foreground branch, some of those sites become under positive selection with  $\omega_2 > 1$ . Positive selection or the presence of sites with  $\omega_2 > 1$  is tested by comparing this model with a null model in which  $\omega_2 = 1$  is fixed, using a 50:50 mixture of 0 and  $\chi^2$  as the null distribution. The BEB (Bayes empirical Bayes) method was used to identify codon sites potentially under positive selection on the foreground branches. While genes with the false discovery rate (FDR) of  $\chi^2$  less than 0.05 were identified as positive selection genes.

Gene family evolution was determined using CAFÉ (v4.2.1) (Mendes et al. 2020). The program uses a birth and death process to model gene gain and loss over a phylogeny. For each family in the data file, CAFÉ computes a probability (*p*-value) of observing the data given the average rate of gain and loss of genes. Expansions and contractions of gene families with *p*-values less than 0.05 were significant.

Enrichment of KEGG (<https://www.genome.jp/kegg/>) and GO (<https://geneontology.org/docs/go-enrichment-analysis/>) terms for the gene families which were significantly expanded or contracted in the Indus River dolphin was performed using

clusterProfiler v4.0 (Wu et al. 2021). KEGG and GO categories with p-adjust less than 0.05 were significant.

### 3 | Results

After filtering low-quality data and duplicate reads, we finally obtained 265.31 Gb from the Illumina platform for subsequent analyses (Table S1). An estimated genome size of 3.35 Gb was deduced from the detection of a *K*-mer (*K*=21) peak at 40× (Figure S1). We obtained the final draft genome assembly of a total length of 2.74 Gb and the GC-content of 41.9%, similar to the previously reported assembly of the Indus River dolphin (PlaMin\_v1\_BIUU, GCA\_0044363435.1). The present genome assembly exhibits

**TABLE 1** | Summary statistics of the current draft genome assembly and assemblies' comparison of the Indus River dolphin.

Draft assembly	The present draft assembly	PlaMin_v1_BIUU
Genome size	2.74 Gb	2.67 Gb
Assembly level	Scaffold	Scaffold
Number of scaffolds	3,275,841	1,098,790
Scaffold N50	40.6 kb	23.9 kb
Number of contigs	3,511,952	1,110,492
Contig N50	10.7 kb	20.9 kb
Genome coverage	40×	28×
GC percent	41.9%	41.5%

**TABLE 2** | BUSCO evaluation of the present Indus River dolphin genome assembly.

BUSCO statistics	Genes	Percentage
Completeness	6440	69.8%
Single-copied	6367	69.0%
Duplicated	73	0.8%
Fragmented	414	4.5%
Missing	2372	25.7%

**TABLE 3** | Comparison of BUSCO evaluations (mammalia\_odb) in cetaceans.

Species	Data source	BUSCO Evaluation	Composed of repetitive sequence	Protein-coding genes
Indus River dolphin	This study	C: 69.8%	35.64%	25,591
Baiji	GCF_003031525	C: 94.6%	43.20%	22,168
Sperm whale	GCA_000472045	C: 94.7%	40.40%	18,686
Bowhead whale	<a href="http://bowhead-whale.org">bowhead-whale.org</a>	C: 90.0%	41.00%	22,677
Indo-Pacific humpback dolphin	GCA_007760645	C: 94.3%	37.41%	24,640

contig and scaffold N50 values of 10.7kb and 40.6kb, respectively, which are comparable in quality to the assembly of the Indus River dolphin published previously (Table 1 and Table S2). The BUSCO analysis showed that 69.8% (single-copied gene: 69%, duplicated gene: 0.8%) of the mammalia\_odb10 gene set was identified as complete, 4.52% of genes were fragmented, and 25.7% of genes were missing in the assembled genome (Table 2 and Table S3). Although not the best BUSCO assessment for cetaceans or river dolphins, this is the first published genome annotation and genome completeness evaluation of the Indus River dolphin (Table 3).

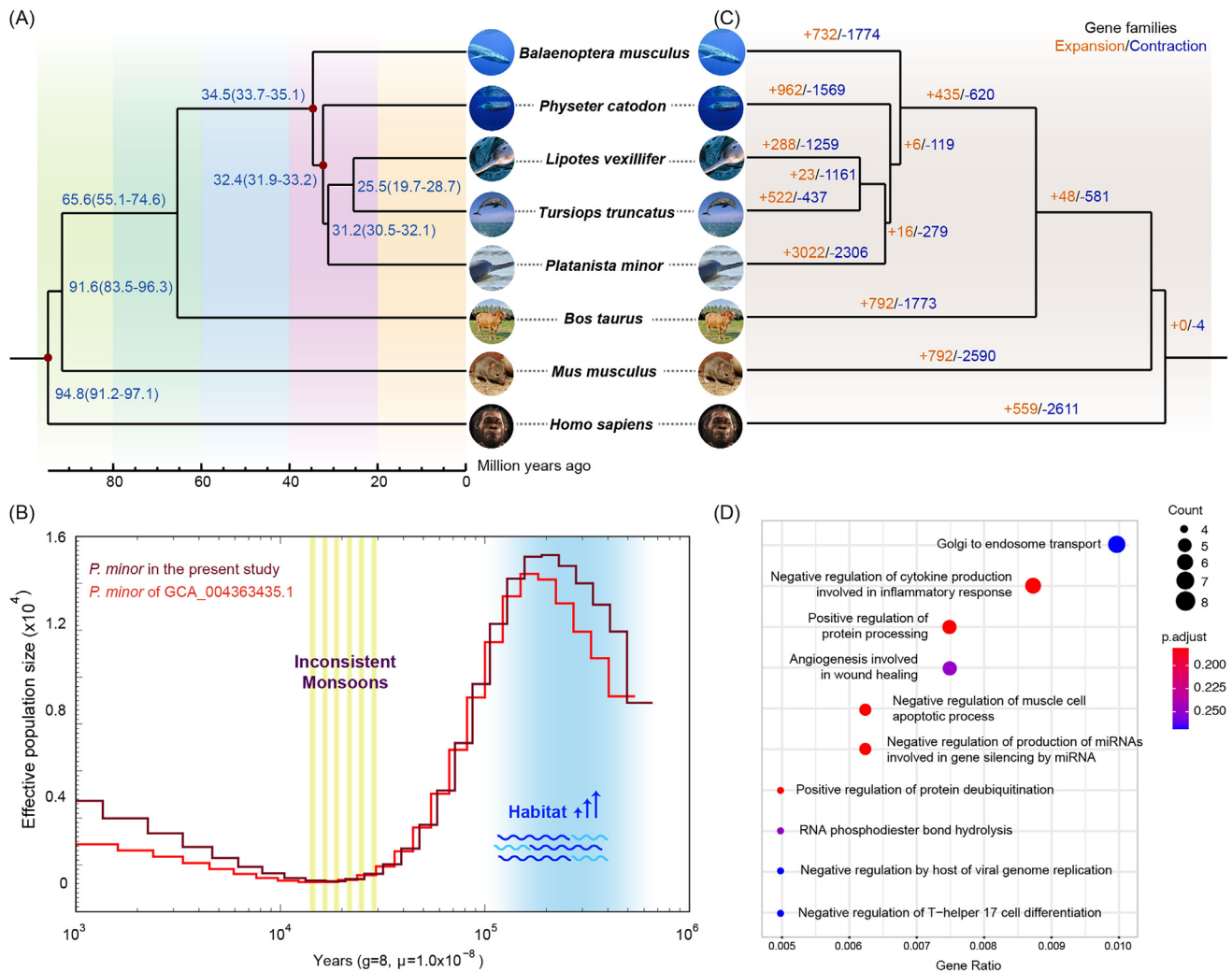
A total of 35.64% of the Indus River dolphin genome is composed of repetitive sequences, with the highest value recorded for long interspersed nuclear elements (LINEs, 19.75%), followed by short interspersed nuclear elements (SINEs, 6.45%), long terminal repeats (LTRs, 4.50%), DNA transposons (2.37%), and the lowest value for simple repeats (0.89%) (Figure S2 and Table S4). In the non-coding RNA annotation, a total of 474 miRNAs, 1063 tRNAs, 180 rRNAs, and 1672 snRNAs were identified (Table S5). Our predictions indicated the presence of 25,591 protein-coding genes, with an average gene length of 9640.82 bp. The average lengths of exons and introns were determined to be 174.86 and 1750.31 bp, respectively (Table S6).

Further, the phylogeny inferred from concatenated alignments of the single-copy orthologues showed that the Indus River dolphin is in the most basal position among the riverine dolphins examined in our dataset (Figure 2A and Figure S3). Additionally, we estimated that the Indus River dolphin and the sperm whale diverged around 32.4 million years ago (MYA), while the Indus River dolphin and the common ancestor of Lipotidae and Delphinidae diverged ~31.2 MYA (Figure 2A).

GATK estimated a total of 977,685 SNPs, and after filtering, we obtained 597,874 SNP across the current genome assembly (2,746,718,295 bp). As a result, a low heterozygosity level of 0.0218% was detected. Utilizing genomic data from individuals from the present study and publicly available assemblies, the PSMC method was employed to infer changes in the effective population size ( $N_e$ ) of the Indus River dolphin. The demographic inference showed a population peak at around 0.3–0.2 MYA after a persistent increase (Figure 2B).

Based on the present assembly and genome annotation, we identified 933 expanded gene families (4238 genes) and 53 contracted gene families (46 genes) specifically in the Indus River





**FIGURE 2** | The divergence, historical population demography, and genome evolution of the Indus River dolphin. (A) A phylogram showing the phylogeny of the Indus River dolphin and other representative mammals, using the human as an outgroup. (B) Pairwise sequential Markovian coalescent result displays the historical demography. (C) Gene family evolution. Gains (in orange) and losses (in blue) of gene families are denoted along nodes and branches. (D) Gene Ontology enrichment of the positively selected genes.

dolphin (Figure 2C; Table S7; Figure S4). Further functional enrichment analyses showed the expanded gene families were significantly enriched in GO and KEGG terms related to immunity, nervous system, cell growth and death, translation, oxidative stress, and glycolipid metabolism. Genes within the contracted gene families were significantly enriched in keratinization, adaptive immune response, and olfactory system (Tables S8, S9, and S10; Figures S5 and S6). Moreover, we identified 878 positively selected genes ( $p < 0.05$ ) in the Indus River dolphin (Table S11). These positively selected genes were functionally enriched in wound healing, muscle cell apoptotic process, inflammatory response, and immunity (Figure 2D; Tables S12 and S13).

## 4 | Discussion

High-quality genomics data from unstudied species can provide answers to various biological questions in conservation biology, including population demographic history,  $N_e$  estimation,

genealogical tracing of mutations, genetics, and local adaptations (Nagasaki et al. 2015). In the present study, we showed a newly sequenced and annotated draft genome of the single adult female Indus River dolphin. Our assembly metrics are comparable to the previously reported one, which was generated using next-generation sequencing technology similarly (Table 1). However, due to limitations in sequencing platform and technology, the continuity and completeness of our genome are not as high as the genome of the baiji, a fellow freshwater dolphin, nor do they exceed the quality of other cetaceans (Table 2). This could be addressed in the future by integrating more advanced sequencing strategies. Nevertheless, based on the current genomic information, we can to some extent comprehend the evolutionary history, genomic features, and adaptation mechanisms of the Indus River dolphin.

The genomic characteristics of the Indus River dolphin are, on the whole, similar to those of other cetaceans, showing a lower proportion of repetitive sequences and a higher proportion of protein-coding genes compared to baiji, and a closer

resemblance to the Indo-Pacific humpback dolphin (Table 3) (Zhou et al. 2013; Keane et al. 2015; Warren et al. 2017; Ming et al. 2019).

The phylogenetic placement of *Platanista* has been the subject of ongoing debate and controversy. The distinctive morphological and skeletal characteristics of this lineage establish *Platanista* as an early-diverging lineage, positioning it at the most basal level among river dolphins (Árnason and Gullberg 1996; Cassens et al. 2000; McHenry et al. 2006; Werth 2006; McGowen et al. 2009; Steeman et al. 2009; Xiong et al. 2009; McCurry et al. 2017). Importantly, this study supported the placement that river dolphins are not a monophyletic group and that *Platanista* has a distant relationship with another representative riverine dolphin lineage, that is, Lipotidae. This observation highlighted the independent origination and adaptation of river dolphin lineages as well as the convergent evolution. Our phylogeny and divergence times overall concurred with tree topologies and estimates in previous studies based on the fossil record and a small set of molecular markers from the mitochondrial and nuclear genes (Árnason and Gullberg 1996; Hamilton et al. 2001; Nikaido et al. 2001; McGowen et al. 2009; Xiong et al. 2009). The fossil records of extinct Platanistidae species from the late Oligocene supported the divergence time of the Indus River dolphin estimated in the present study (Braulik et al. 2015; Braulik et al. 2021).

It has been suggested that the *Platanista* lineage entered the freshwater of the Indian subcontinent during the late Neogene trend of sea-level regression (Hamilton et al. 2001). Combined with the hypothesis that the tributaries started to shift from the Ganges nearly 5 MYA and completely connected to the Indus River around 0.3 MYA (Clift and Blusztajn 2005), this detected flourishing of the Indus River dolphin might be due to habitat expansion. There was an apparent reduction in  $N_e$  of the Indus River dolphin 40–20 kiloyears ago (KYA) (Figure 1B). This period is highly consistent with the upper Paleolithic, marked by the inconsistent monsoons and colder climate in the Indian context (Singhvi and Kale 2010; Khara et al. 2021), thus it might impact the habitat of the Indus River dolphin. This situation parallels that of the baiji, which also experienced a significant bottleneck approximately 10 KYA as a result of severe adverse environmental influences (Zhou et al. 2013). To advance understanding, it is necessary to obtain more high-quality genomic data and employ various methods to estimate  $N_e$  of the Indus River dolphin across different time frames in the future.

It has been shown that populations with loss of genetic diversity and inbreeding can lead to sharp declines in adaptability, reproductive capacity, and disease resistance, thus heightening the risk of extinction (Lacy 1987). This observed genetic diversity of the Indus River dolphin is similar to the baiji (0.0121%) (Zhou et al. 2013), and is less than that of the pygmy right whale (*Caperea marginanta*) (0.11%) (Wolf et al. 2023), while it exceeds that of the minke whale (*Balaenoptera acutorostrata*) (0.00069%) (Yim et al. 2014), the sperm whale (0.0011%) (Warren et al. 2017), the fin whale (*Balaenoptera physalus*) (0.00142%), and the bottlenose dolphin (0.0014%). The heterozygosity of the Indus River dolphin is similar to that of the baiji, an extinct river dolphin over past decades, indicating its lower evolutionary potential.

This highlights the need for increased attention to the survival status of the Indus River dolphin populations.

Further, the endangered populations facing the risk of anthropogenic impacts, such as habitat fragmentation, are usually threatened by population size reduction and inbreeding depression, leading to the lowered genetic diversity (Grueber et al. 2008). After 1885, the development of Indus Basin irrigation fragmented the overall habitat and population; additionally, hunting pressure declined the overall population to less than 200 individuals in the 1970s (Anderson 1878; Braulik et al. 2015). However, the subpopulation sizes increased rapidly over time. For example, during four decades from 1979 to 2017, the isolated subpopulation of the Taunsa-Guddu barrage has expanded from 36 (Pilleri and Bhatt 1980) to more than 1000 individuals (WWF-Pakistan 2017). The bottleneck effect of only a few generations is enough to lead to a loss of heterozygosity (Wolf et al. 2023). As the expected generation time of the Indus River dolphin is 8 years, about five generations have passed since 1979. Thus, the cumulative effect of historic and recent bottleneck events has a strong probability of the loss of heterozygosity due to inbreeding.

Cetaceans, including river dolphins, are a group of highly aquatic mammals that have evolved numerous remarkable adaptations (Würsig 1989; Springer et al. 2021). Thereinto, hair, a highly keratinized tissue in mammals, plays a crucial role in cetaceans by reducing friction and enhancing locomotion through significant evolutionary adaptations that lead to hair loss (Chen et al. 2013; Sun et al. 2017). We identified that the contracted gene families in the Indus River dolphin are significantly related to keratinization (GO:0031424), cornification (GO:0070268), and hair follicle morphogenesis (GO:0031069), supporting the evolution of the degenerated skin derivatives of the Indus River dolphin. Moreover, the degenerated olfactory system is another significant adaptation for fully aquatic life (De Vreese et al. 2023). The contracted gene families associated with olfactory receptor activity (GO:0004984) aligned with olfactory degeneration in the Indus River dolphin, as observed in another river dolphin, the baiji (Zhou et al. 2013).

Besides, river dolphins including *Platanista* have successfully adapted to and colonized the freshwater environment such as the Indus River system. Notably, in comparison to the marine ecosystem, the freshwater environment demonstrates a higher diversity and distinction of bacteria and other microorganisms (Auguet et al. 2010; Barberán and Casamayor 2010). In this study, we identified specific genes and gene families associated with immunity that have undergone positive selection and significant turnover, respectively (Figure 2C and Tables S7–S10). These findings indicated a specific immune response tailored to the freshwater environment of the Indus River dolphin.

## 5 | Conclusion

The endangered status of the Indus River dolphin (*P. minor*) is attributed to the reduction in its distributional range, loss of habitat, and escalating severe anthropogenic pressures along the Indus River in Pakistan. The newly reported draft genome

of the endangered species led to the identification of a peak in population size around 0.3–0.2 MYA, as indicated by the PSMC model. This finding implies a possible migration from the Ganges to the Indus River system at that time. As a consequence of climate change, the population experienced a gradual decline and encountered a bottleneck effect approximately 40–20 KYA. Additionally, our comparative genomic findings emphasize that the Indus River dolphin, residing in freshwater habitats, has evolved a highly adapted immune system. In conclusion, this study provides new insights into the molecular mechanisms involved in the freshwater adaptation, divergence, phylogeny, genetic diversity, and demographic history of the Indus River dolphin. These findings will contribute to conservation efforts for this species in the Indian subcontinent. In the future, advanced genome sequencing, comprehensive comparative, and population genomics studies are essential to elucidate the ecology, evolution, and adaptation of this species to freshwater. These studies will establish a robust scientific foundation for the conservation of this endangered species.

### Author Contributions

**Aamir Ibrahim:** conceptualization (equal), data curation (equal), writing – original draft (equal), writing – review and editing (equal). **Simin Chai:** conceptualization (equal), data curation (equal), visualization (lead), writing – original draft (equal), writing – review and editing (equal). **Cuijuan Zhong:** resources (equal), writing – review and editing (equal). **Kang Jieqiong:** investigation (equal), project administration (equal). **Ahsaan Ali:** investigation (equal), resources (equal). **Sajjad Hussain:** investigation (equal), resources (equal). **Hassan Ali:** investigation (equal), resources (equal). **Tanveer Hussain:** investigation (equal), resources (equal). **Umer Waqas:** investigation (equal), resources (equal). **Guang Yang:** conceptualization (lead), supervision (lead), writing – review and editing (equal).

### Acknowledgments

This work was supported by the Key Project of the National Natural Science Foundation of China (grant no. 32030011); the National Key Programme of Research and Development, Ministry of Science and Technology (grant no. 2022YFF1301600), the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD), and PI Project of Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou) (GML2021GD0805). The authors are thankful to Shahid Iqbal, Prof. Dr. Abdul Qadir, Saira Aamir, Komal Mushtaq, Abdul Qadeer, and the Wildlife and Parks Department, Punjab, Pakistan (Dr. Muhammad Rizwan Khan, Dr. Waqas Noor, and Dr. Warda Gill) for helping us throughout the processing of the genome sequencing.

### Conflicts of Interest

The authors declare no conflicts of interest.

### Data Availability Statement

The data is uploaded and available at the National Genomics Data Center (<https://www.cncb.ac.cn/>) with an accession number PRJCA027994.

### References

Anderson, J. 1878. Anatomical and Zoological Researches: Comprising an Account of the Zoological Results of the Two Expeditions to Western Yunnan in 1868 and 1875; and a Monograph of the Two Cetacean Genera, Platanista.

Árnason, Ú., and A. Gullberg. 1996. “Cytochrome b Nucleotide Sequences and the Identification of Five Primary Lineages of Extant Cetaceans.” *Molecular Biology and Evolution* 13, no. 2: 407–417.

Auguet, J.-C., A. Barberan, and E. O. Casamayor. 2010. “Global Ecological Patterns in Uncultured Archaea.” *ISME Journal* 4, no. 2: 182–190.

Barberán, A., and E. O. Casamayor. 2010. “Global Phylogenetic Community Structure and  $\beta$ -Diversity Patterns in Surface Bacterioplankton Metacommunities.” *Aquatic Microbial Ecology* 59, no. 1: 1–10.

Benson, G. 1999. “Tandem Repeats Finder: A Program to Analyze DNA Sequences.” *Nucleic Acids Research* 27, no. 2: 573–580.

Braulik, G., R. Barnett, V. Odon, A. Islas-Villanueva, A. Hoelzel, and J. Graves. 2015. “One Species or Two? Vicariance, Lineage Divergence and Low mtDNA Diversity in Geographically Isolated Populations of South Asian River Dolphin.” *Journal of Mammalian Evolution* 22: 111–120.

Braulik, G. T., F. I. Archer, U. Khan, et al. 2021. “Taxonomic Revision of the South Asian River Dolphins (Platanista): Indus and Ganges River Dolphins Are Separate Species.” *Marine Mammal Science* 37, no. 3: 1022–1059.

Brůna, T., A. Lomsadze, and M. Borodovsky. 2020. “GeneMark-EP+: Eukaryotic Gene Prediction With Self-Training in the Space of Genes and Proteins.” *NAR Genomics and Bioinformatics* 2, no. 2: lqaa026.

Cantarel, B. L., I. Korf, S. M. Robb, et al. 2008. “MAKER: An Easy-To-Use Annotation Pipeline Designed for Emerging Model Organism Genomes.” *Genome Research* 18, no. 1: 188–196.

Cassens, I., S. Vicario, V. G. Waddell, et al. 2000. “Independent Adaptation to Riverine Habitats Allowed Survival of Ancient Cetacean Lineages.” *Proceedings of the National Academy of Sciences of the United States of America* 97, no. 21: 11343–11347.

Chen, Y., Y. Chen, C. Shi, et al. 2018. “SOAPnuke: A MapReduce Acceleration-Supported Software for Integrated Quality Control and Preprocessing of High-Throughput Sequencing Data.” *GigaScience* 7, no. 1: gix120.

Chen, Z., Z. Wang, S. Xu, K. Zhou, and G. Yang. 2013. “Characterization of Hairless (Hr) and FGF5 Genes Provides Insights Into the Molecular Basis of Hair Loss in Cetaceans.” *BMC Evolutionary Biology* 13: 1–11.

Clift, P. D., and J. Blusztajn. 2005. “Reorganization of the Western Himalayan River System After Five Million Years Ago.” *Nature* 438, no. 7070: 1001–1003.

de Muizon, C. 1988. “Les Relations Phylogénétiques des Delphinida (Cetacea, Mammalia).” *Annales de Paleontologie* 74, no. 4: 159–227.

De Vreese, S., K. Orekhova, M. Morell, T. Gerussi, and J.-M. Graïc. 2023. “Neuroanatomy of the Cetacean Sensory Systems.” *Animals* 14, no. 1: 66.

Emms, D. M., and S. Kelly. 2015. “OrthoFinder: Solving Fundamental Biases in Whole Genome Comparisons Dramatically Improves Orthogroup Inference Accuracy.” *Genome Biology* 16: 1–14.

Foote, A. D., Y. Liu, G. W. Thomas, et al. 2015. “Convergent Evolution of the Genomes of Marine Mammals.” *Nature Genetics* 47, no. 3: 272–275.

Fordyce, R., and C. de Muizon. 2001. “Evolutionary History of Cetaceans: A Review.” *Secondary Adaptation of Tetrapods to Life in Water*: 169–233.

Grueber, C. E., G. P. Wallis, and I. G. Jamieson. 2008. “Heterozygosity–Fitness Correlations and Their Relevance to Studies on Inbreeding Depression in Threatened Species.” *Molecular Ecology* 17, no. 18: 3978–3984.

Haig, S. M., M. P. Miller, R. Bellinger, H. M. Draheim, D. M. Mercer, and T. D. Mullins. 2016. “The Conservation Genetics Juggling Act: Integrating Genetics and Ecology, Science and Policy.” *Evolutionary Applications* 9, no. 1: 181–195.



- Hamilton, H., S. Caballero, A. G. Collins, and R. L. Brownell Jr. 2001. "Evolution of River Dolphins." *Proceedings of the Royal Society of London. Series B: Biological Sciences* 268, no. 1466: 549–556.
- Hedrick, P. W., and P. S. Miller. 1992. "Conservation Genetics: Techniques and Fundamentals." *Ecological Applications* 2, no. 1: 30–46.
- Kajitani, R., K. Toshimoto, H. Noguchi, et al. 2014. "Efficient de Novo Assembly of Highly Heterozygous Genomes From Whole-Genome Shotgun Short Reads." *Genome Research* 24, no. 8: 1384–1395.
- Katoh, K., and D. M. Standley. 2013. "MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability." *Molecular Biology and Evolution* 30, no. 4: 772–780.
- Keane, M., J. Semeiks, A. E. Webb, et al. 2015. "Insights Into the Evolution of Longevity From the Bowhead Whale Genome." *Cell Reports* 10, no. 1: 112–122.
- Keilwagen, J., M. Wenk, J. L. Erickson, M. H. Schattat, J. Grau, and F. Hartung. 2016. "Using Intron Position Conservation for Homology-Based Gene Prediction." *Nucleic Acids Research* 44, no. 9: e89.
- Khara, A., M. Khanyari, A. Ghoshal, et al. 2021. "The Forgotten Mountain Monarch? Understanding Conservation Status of the Vulnerable Ladakh Urial in India." *European Journal of Wildlife Research* 67: 1–10.
- Korf, I. 2004. "Gene Finding in Novel Genomes." *BMC Bioinformatics* 5: 1–9.
- Lacy, R. C. 1987. "Loss of Genetic Diversity From Managed Populations: Interacting Effects of Drift, Mutation, Immigration, Selection, and Population Subdivision." *Conservation Biology* 1, no. 2: 143–158.
- Lande, R. 1993. "Risks of Population Extinction From Demographic and Environmental Stochasticity and Random Catastrophes." *American Naturalist* 142, no. 6: 911–927.
- Li, H., and R. Durbin. 2009. "Fast and Accurate Short Read Alignment With Burrows–Wheeler Transform." *Bioinformatics* 25, no. 14: 1754–1760.
- Luo, R., B. Liu, Y. Xie, et al. 2012. "SOAPdenovo2: An Empirically Improved Memory-Efficient Short-Read de Novo Assembler." *GigaScience* 1, no. 1: 2047–2217X-2041–2018.
- McCurry, M. R., A. R. Evans, E. M. Fitzgerald, J. W. Adams, P. D. Clausen, and C. R. McHenry. 2017. "The Remarkable Convergence of Skull Shape in Crocodylians and Toothed Whales." *Proceedings of the Royal Society B: Biological Sciences* 284, no. 1850: 20162348.
- McGowen, M. R., M. Spaulding, and J. Gatesy. 2009. "Divergence Date Estimation and a Comprehensive Molecular Tree of Extant Cetaceans." *Molecular Phylogenetics and Evolution* 53, no. 3: 891–906.
- McHenry, C. R., P. D. Clausen, W. J. Daniel, M. B. Meers, and A. Pendharkar. 2006. "Biomechanics of the Rostrum in Crocodylians: A Comparative Analysis Using Finite-Element Modeling." *Anatomical Record Part A: Discoveries in Molecular, Cellular, and Evolutionary Biology* 288, no. 8: 827–849.
- Melbourne, B. A., and A. Hastings. 2008. "Extinction Risk Depends Strongly on Factors Contributing to Stochasticity." *Nature* 454, no. 7200: 100–103.
- Mendes, F. K., D. Vanderpool, B. Fulton, and M. W. Hahn. 2020. "CAFE 5 Models Variation in Evolutionary Rates Among Gene Families." *Bioinformatics* 36, no. 22–23: 5516–5518.
- Ming, Y., J. Jian, X. Yu, J. Wang, and W. Liu. 2019. "The Genome Resources for Conservation of Indo-Pacific Humpback Dolphin, *Sousa chinensis*." *Scientific Data* 6, no. 1: 68.
- Nagasaki, M., J. Yasuda, F. Katsuoka, et al. 2015. "Rare Variant Discovery by Deep Whole-Genome Sequencing of 1,070 Japanese Individuals." *Nature Communications* 6, no. 1: 8018.
- Nawrocki, E. P., D. L. Kolbe, and S. R. Eddy. 2009. "Infernal 1.0: Inference of RNA Alignments." *Bioinformatics* 25, no. 10: 1335–1337.
- Nigenda-Morales, S. F., M. Lin, P. G. Nuñez-Valencia, et al. 2023. "The Genomic Footprint of Whaling and Isolation in Fin Whale Populations." *Nature Communications* 14, no. 1: 5465.
- Nikaido, M., F. Matsuno, H. Hamilton, et al. 2001. "Retroposon Analysis of Major Cetacean Lineages: The Monophyly of Toothed Whales and the Paraphyly of River Dolphins." *Proceedings of the National Academy of Sciences of the United States of America* 98, no. 13: 7384–7389.
- Pertea, M., G. M. Pertea, C. M. Antonescu, T.-C. Chang, J. T. Mendell, and S. L. Salzberg. 2015. "StringTie Enables Improved Reconstruction of a Transcriptome From RNA-Seq Reads." *Nature Biotechnology* 33, no. 3: 290–295.
- Pilleri, G., and M. U. Bhatt. 1980. "Status of the Indus Dolphin population (*Platanista indi*, BLYTH, 1859) between Sukkur and Taunsa Barrages." *Investigations on Cetacea* 13: 245–252.
- Poplin, R., V. Ruano-Rubio, M. A. DePristo, et al. 2017. "Scaling Accurate Genetic Variant Discovery to Tens of Thousands of Samples." *BioRxiv*: 201178.
- Quail, M. A., M. Smith, P. Coupland, et al. 2012. "A Tale of Three Next Generation Sequencing Platforms: Comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq Sequencers." *BMC Genomics* 13: 1–13.
- Reed, D. H., and R. Frankham. 2003. "Correlation Between Fitness and Genetic Diversity." *Conservation Biology* 17, no. 1: 230–237.
- Simão, F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, and E. M. Zdobnov. 2015. "BUSCO: Assessing Genome Assembly and Annotation Completeness With Single-Copy Orthologs." *Bioinformatics* 31, no. 19: 3210–3212.
- Singhvi, A., and V. S. Kale. 2010. *Paleoclimate Studies in India: Last Ice Age to the Present*. Indian National Science Acad.
- Springer, M. S., C. F. Guerrero-Juarez, M. Huelsmann, et al. 2021. "Genomic and Anatomical Comparisons of Skin Support Independent Adaptation to Life in Water by Cetaceans and Hippos." *Current Biology* 31, no. 10: 2124–2139.
- Stamatakis, A. 2014. "RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies." *Bioinformatics* 30, no. 9: 1312–1313.
- Stanke, M., M. Diekhans, R. Baertsch, and D. Haussler. 2008. "Using Native and Syntenically Mapped cDNA Alignments to Improve de Novo Gene Finding." *Bioinformatics* 24, no. 5: 637–644.
- Steeman, M. E., M. B. Hebsgaard, R. E. Fordyce, et al. 2009. "Radiation of Extant Cetaceans Driven by Restructuring of the Oceans." *Systematic Biology* 58, no. 6: 573–585.
- Sun, X., Z. Zhang, Y. Sun, J. Li, S. Xu, and G. Yang. 2017. "Comparative Genomics Analyses of Alpha-Keratins Reveal Insights Into Evolutionary Adaptation of Marine Mammals." *Frontiers in Zoology* 14: 1–8.
- Viglino, M., M. Gaetán, M. R. Buono, R. E. Fordyce, and T. Park. 2021. "Hearing From the Ocean and Into the River: The Evolution of the Inner Ear of Platanistoidea (Cetacea: Odontoceti)." *Paleobiology* 47, no. 4: 591–611.
- von Der Heyden, S., M. Beger, R. J. Toonen, et al. 2014. "The Application of Genetics to Marine Management and Conservation: Examples From the Indo-Pacific." *Bulletin of Marine Science* 90, no. 1: 123–158.
- Vurtture, G. W., F. J. Sedlazeck, M. Nattestad, et al. 2017. "GenomeScope: Fast Reference-Free Genome Profiling From Short Reads." *Bioinformatics* 33, no. 14: 2202–2204.
- Warren, W. C., L. Kuderna, A. Alexander, et al. 2017. "The Novel Evolution of the Sperm Whale Genome." *Genome Biology and Evolution* 9, no. 12: 3260–3264.



- Werth, A. J. 2006. "Mandibular and Dental Variation and the Evolution of Suction Feeding in Odontoceti." *Journal of Mammalogy* 87, no. 3: 579–588.
- Willi, Y., J. Van Buskirk, and A. A. Hoffmann. 2006. "Limits to the Adaptive Potential of Small Populations." *Annual Review of Ecology, Evolution, and Systematics* 37, no. 1: 433–458.
- Wolf, M., K. Zapf, D. K. Gupta, M. Hiller, Ú. Árnason, and A. Janke. 2023. "The Genome of the Pygmy Right Whale Illuminates the Evolution of Rorquals." *BMC Biology* 21, no. 1: 79.
- Wu, T., E. Hu, S. Xu, et al. 2021. "clusterProfiler 4.0: A Universal Enrichment Tool for Interpreting Omics Data." *Innovation* 2: 3.
- Würsig, B. 1989. "Cetaceans." *Science* 244, no. 4912: 1550–1557.
- WWF-Pakistan. 2017. Signs of Hope for the Endemic and Endangered Bhulan Comprehensive Population Assessment of the Indus River Dolphin (*Platanista gangetica* Minor). Indus River-2017 Preliminary Results Report.
- Xiong, Y., M. C. Brandley, S. Xu, K. Zhou, and G. Yang. 2009. "Seven New Dolphin Mitochondrial Genomes and a Time-Calibrated Phylogeny of Whales." *BMC Evolutionary Biology* 9: 1–13.
- Yan, J., K. Zhou, and G. Yang. 2005. "Molecular Phylogenetics of 'River Dolphins' and the Baiji Mitochondrial Genome." *Molecular Phylogenetics and Evolution* 37, no. 3: 743–750.
- Yang, Z. 2007. "PAML 4: Phylogenetic Analysis by Maximum Likelihood." *Molecular Biology and Evolution* 24, no. 8: 1586–1591.
- Yim, H.-S., Y. S. Cho, X. Guang, et al. 2014. "Minke Whale Genome and Aquatic Adaptation in Cetaceans." *Nature Genetics* 46, no. 1: 88–92.
- Zhou, X., X. Guang, D. Sun, et al. 2018. "Population Genomics of Finless Porpoises Reveal an Incipient Cetacean Species Adapted to Freshwater." *Nature Communications* 9, no. 1: 1276.
- Zhou, X., F. Sun, S. Xu, et al. 2013. "Baiji Genomes Reveal Low Genetic Variability and New Insights Into Secondary Aquatic Adaptations." *Nature Communications* 4, no. 1: 2708.

### Supporting Information

Additional supporting information can be found online in the Supporting Information section.