



Extensive Interspecific Gene Flow Shaped Complex Evolutionary History and Underestimated Species Diversity in Rapidly Radiated Dolphins

Weijian Guo¹ · Di Sun¹ · Yang Cao¹ · Linlin Xiao¹ · Xin Huang¹ · Wenhua Ren¹ · Shixia Xu¹ · Guang Yang¹

Accepted: 19 October 2021 / Published online: 17 November 2021
© The Author(s) 2021

Abstract

Recently diverged taxa are often characterized by high rates of hybridization, which can complicate phylogenetic reconstruction. For this reason, the phylogenetic relationships and evolutionary history of dolphins are still not very well resolved; the question of whether the genera *Tursiops* and *Stenella* are monophyletic is especially controversial. Here, we performed re-sequencing of six dolphin genomes and combined them with eight previously published dolphin SRA datasets and six whole-genome datasets to investigate the phylogenetic relationships of dolphins and test the monophyly hypothesis of *Tursiops* and *Stenella*. Phylogenetic reconstruction with the maximum likelihood and Bayesian methods of concatenated loci, as well as with coalescence analyses of sliding window trees, produced a concordant and well-supported tree. Our studies support the non-monophyletic status of *Tursiops* and *Stenella* because the species referred these genera do not form exclusive monophyletic clades. This suggests that the current taxonomy of both genera might not reflect their evolutionary history and may underestimate their diversity. A four-taxon *D*-statistic (ABBA-BABA) test, five-taxon D_{FOIL} test, and tree-based PhyloNet analyses all showed extensive gene flow across dolphin species, which could explain the instability in resolving phylogenetic relationship of oceanic dolphins with different and limited markers. This study could be a good case to demonstrate how genomic data can reveal complex speciation and phylogeny in rapidly radiating animal groups.

Keywords Phylogenomics · *Tursiops* · *Stenella* · Non-monophyly · Gene flow

Introduction

Cetaceans, including whales, dolphins, and porpoises, are a group of secondarily adapted marine mammals with a history of transitioning from terrestrial (land) to fully aquatic habitats, representing one of the most fascinating evolutionary events of the planet (Thewissen et al. 2009; Uhen 2010). Approximately 56–53 million years ago, cetacean ancestors transformed their habitat from a completely terrestrial to a fully aquatic environment and subsequently radiated around the world with a series of outstanding morphological changes (e.g., degeneration of

limbs, lengthening of the skull, thickening of the blubber, loss of hair, etc.) (Reidenberg 2007; Thewissen et al. 2007). After they transitioned to the sea, cetaceans further diverged into two highly distinct groups: Mysticeti (baleen whales, characterized by baleen plates and two blowholes) and Odontoceti (toothed whales, dolphins, and porpoises, which have a single blowhole) (Gatesy et al. 2013).

Establishing correct phylogenetic relationships is an elementary step in unraveling the evolutionary processes responsible for cetacean diversity. The secondarily aquatic nature of whales has inspired numerous phylogenetic studies and sparked debates about their origin from terrestrial mammals (Gatesy and O'Leary 2001; Thewissen et al. 2001; Árnason et al. 2004; O'Leary and Gatesy 2008; Uhen 2010). Several problems concerning the systematics of cetaceans have been well resolved, such as the monophyly of cetaceans, Mysticeti, and Odontoceti (Gatesy et al. 1999; Nikaido et al. 1999; Geisler and Sanders 2003), the sister relationship between cetaceans and Hippopotamidae (Lum et al. 2000; McGowen et al. 2009; Thewissen et al.

Weijian Guo and Di Sun Both authors contributed equally.

✉ Guang Yang
gyang@njnu.edu.cn

¹ Jiangsu Key Laboratory for Biodiversity and Biotechnology, College of Life Sciences, Nanjing Normal University, Nanjing 210046, China

2009; Zhou et al. 2011), and the non-monophyletic status of river dolphins (Geisler et al. 2011; Hassanin et al. 2012; McGowen et al. 2020). Despite widespread popular and scientific interest in some major cetacean lineages, many aspects of their phylogeny and evolutionary history remain unresolved. For example, the phylogenetic relationships within Balaenopteroidea and Ziphiidae have been studied using extensive DNA sequences, but several studies have produced contradictory results (Nikaido et al. 2006; Dalebout et al. 2008; McGowen et al. 2009; Steeman et al. 2009; Árnason et al. 2018). Beyond rorquals and beaked whales, the relationships within Delphinidae remains problematic.

Delphinidae is the largest cetacean family (~37 species), and resolving phylogenetic relationships among its members has been disheartening despite numerous recent attempts because of the rapidity of speciation, which facilitated incomplete lineage sorting (ILS) and hybridization between emerging species (Kingston et al. 2009; Amaral et al. 2012; Perrin et al. 2013). A particular conundrum surrounds the phylogenetic relationships among bottlenose-like dolphins (e.g., subfamily Delphininae), which also have experienced recent, rapid radiations (Steeman et al. 2009; Slater et al. 2010). Although some morphological (Shirakihara et al. 2003; Guidarelli et al. 2014, 2018; Jedensjö et al. 2017) and molecular (McGowen et al. 2009, 2020; Chen et al. 2011; Moura et al. 2013, 2020) analyses have been conducted, the evolutionary relationships among Delphininae generally have remained unclear. In particular, the monophyly of the genera *Tursiops* and *Stenella* is still under debate.

As early as 20 years ago, most taxonomic studies of dolphin evolutionary relationships relied on morphological characters, particularly those of the skull. For example, the studies of skull, teeth, and vertebrae showed that species of *Tursiops* and *Stenella* can be distinguished from those of other genera (Archer and Perrin 1999; Shirakihara et al. 2003). Ever since the comprehensive molecular analysis of Delphinidae by Leduc et al. (1999), which included 33 species and was based on full mitochondrial cytochrome *b* sequences, an increasing number of studies using genetic data including multiple nuclear loci (Kingston et al. 2009), short interspersed elements (SINE) (Chen et al. 2011), and even whole mitochondrial genomes (Hassanin et al. 2012; Moura et al. 2013) have not supported the monophyly of *Tursiops* or *Stenella* (i.e., the species of these two groups do not cluster with their congeners). These results have partly contrasted with other studies integrating nuclear and mitochondrial data (McGowen et al. 2009; McGowen 2011), multiple protein-coding genes (McGowen et al. 2020), and RADseq data (Moura et al. 2020), which have recovered *Tursiops* as a monophyletic group (Fig. 1).

Two evolutionary processes, hybridization and ILS, may trigger the uncertain phylogenetic relationships observed

in some groups of mammals, such as wild pigs, bears, and baboons (Frantz et al. 2013; Kumar et al. 2017; Rogers et al. 2019). Due to the invariable number of chromosomes ($2n = 44$) and the common karyotic arrangement in most cetaceans, a recent review found that hybridization is a common phenomenon among oceanic cetaceans (Crossman et al. 2016). Incongruence of gene trees across the genome results from hybridization and ILS that can be detected by analyzing whole-genome sequences (Degnan and Rosenberg 2009). In this context, it is becoming increasingly important to use genome-scale data sets to resolve phylogenetic relationships among cetaceans.

The genomic era allows us to sample massive genomic data, which undoubtedly helps us to address the phylogenetic issues that cannot be well resolved using limited genetic markers (Delsuc et al. 2005). Moreover, genomic data could also provide a special opportunity to detect the footprints of hybridization and allow exhaustive analyses of how gene flow affects phylogenetic relationships (Payseur and Rieseberg 2016). Recently, Árnason et al. (2018) analyzed the evolutionary history of baleen whales using genomic data and found some gene flow across rorquals. Here, we present new genomic data from six oceanic dolphin species. Using some analytical methods specially developed to handle complex genomic data, the phylogenetic relationships of oceanic dolphins were investigated, providing new insights into whether the genera *Tursiops* and *Stenella* are monophyletic or non-monophyletic as well as signals of recent and ancestral introgression among dolphins.

Materials and Methods

Taxon Sampling

Genome sequences were generated from six dolphin species (nine individuals) belonging to the family Delphinidae (Online Resource 1). We also downloaded previously published sequence data from the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) for *Tursiops truncatus* (common bottlenose dolphin), *Sagmatias obliquidens* (Pacific white-sided dolphin), and published genomes, including *Orcinus orca* (killer whale), *Neophocoena asiaorientalis* (Yangtze finless porpoise), *Monodon monoceros* (narwhal), *Balaenoptera acutorostrata* (common minke whale), *Megaptera novaeangliae* (humpback whale) and *Hippopotamus amphibius* (hippopotamus) (Online Resource 1).

DNA Isolation, Sequencing, and Pseudo-genome Generation

The verified specimens of six species (nine individuals) sequenced in the present study were preserved at Nanjing

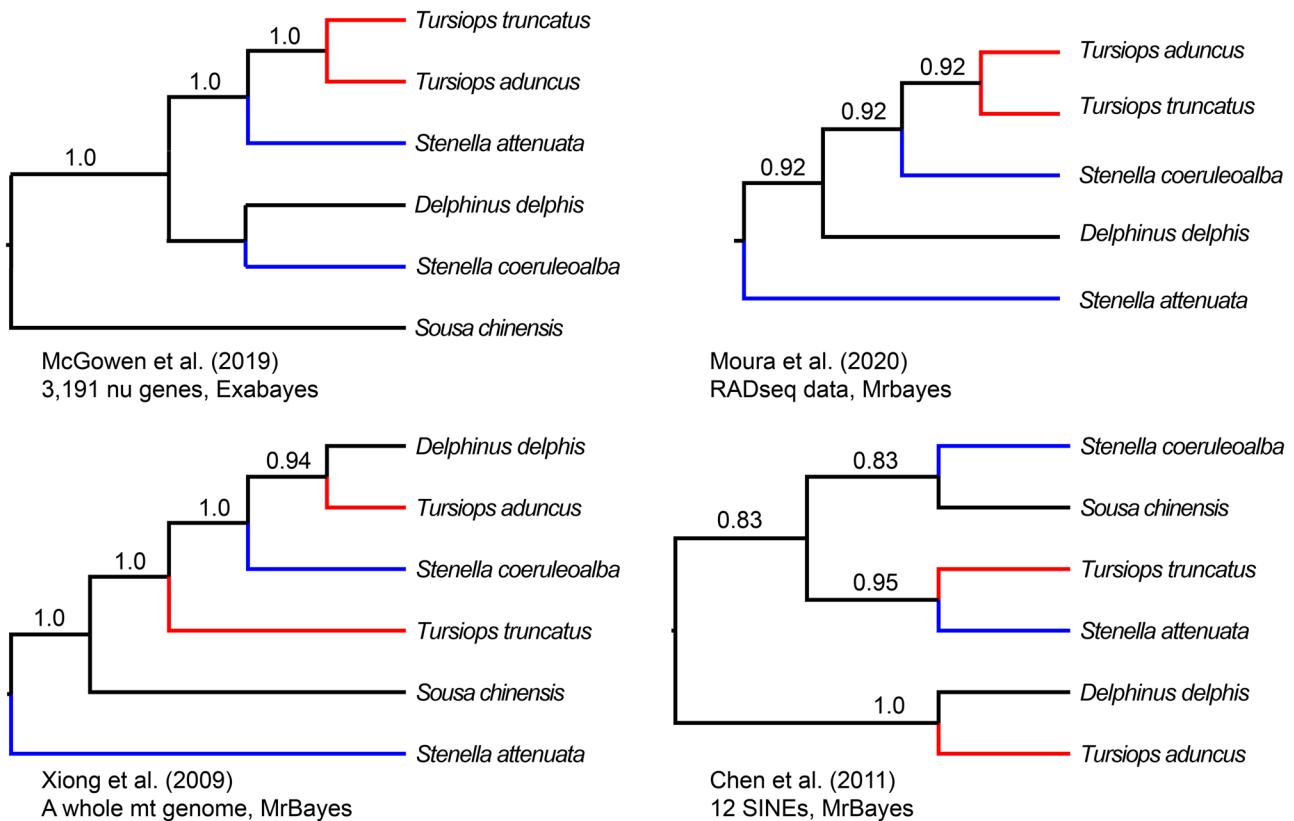


Fig. 1 Recent hypotheses of the inter-relationships of the subfamily Delphininae. The original phylogenies were pruned to include only species used in the current study. Decimal numbers represent poste-

rior probabilities. *Sousa chinensis* was not included in the study of Moura et al. (2020). Abbreviations: nu (nuclear), mt (mitochondrial), SINE (short interspersed elements)

Normal University. All samples were collected from individuals that were incidentally killed by unknown reasons in the wild (and thus no ethical approval was necessary). All tissue samples were frozen at -20°C before this study. DNA was isolated using a standard phenol–chloroform method (Sambrook et al. 1989). Sequencing libraries were prepared with insert sizes of 550 bp and sequenced using Illumina Nextseq 500 technology. The whole genomes were sequenced at $8\times$ to $25.3\times$ depth coverage (Online Resource 1). Quality control was performed using FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc/), and the retrieved raw paired reads were trimmed using Trimmomatic version 0.33 (Bolger et al. 2014) with parameters “SLIDINGWINDOW:4:15” to cut the read once the average quality within the window fell below a threshold and “LEADING:3” as well as “TRAILING:3”, which removed low quality bases from the beginning and the end, respectively. As the common bottlenose dolphin is phylogenetically positioned in the genus *Tursiops*, and a possible mapping bias against its genome is likely to affect phylogenetic and gene flow analyses, the killer whale genome (Foote et al. 2015) with soft mask was used as a reference in order to avoid mapping biases that can affect phylogenetic

analyses. Scaffolds in the killer whale genome shorter than 100 kb were excluded. Paired-end reads were mapped to the killer whale genome with BWA version 0.7.12-r1039 (Li and Durbin 2009) using the “mem” and “-M” option (The “mem” option stands for Maximal Exact Match and the “-M” option means mark shorter split hits as secondary), and duplicates were marked with Picard (<https://github.com/broadinstitute/picard>). From the mapped reads, single-nucleotide variants (SNVs) and short insertion or deletions (Indels) were called by SAMtools (Li et al. 2009) with the default settings. Consensus sequences were created from VCF files using BCFtools (<https://samtools.github.io/bcftools/>), choosing the allele with more reads (≥ 5) in case of heterozygous sites. Indels were removed, and ambiguously called sites were masked as “N”. At this point, the pseudo-genome of each individual was created.

Pseudo-genome Alignment

The generation of multiple sequence alignments is a crucial step for phylogenetic analyses. Our genomic sequence alignment (GSA) of all pseudo-genomes spanned 21 individuals from 12 cetacean species and an outgroup species

(*H. amphibius*) as well as the genome of killer whale. First, the pairwise alignments were conducted by LAST version 963 package (Kiełbasa et al. 2011), with the killer whale genome as the reference genome. Firstly, the killer whale genome “database” was built using “lastdb” with the parameter “-cR11”, which excluded lowercase letters from initial matches but kept the lowercase letters in the sequences. Secondly, each pseudo-genome was aligned to the reference genome using the “lastal” command with the parameter “-E0.05, -m100”. Subsequently, the “maf-swap” command was used to change the order of the sequences in the MAF-format alignments to obtain the best pairwise aligned blocks. Lastly, the pairwise alignments were merged into multiple genome alignments by MULTIZ version 11.2 (Blanchette et al. 2004) using the killer whale genome as a reference. Only 1:1 aligning to the killer whale sequence was allowed, choosing the hits with the highest identity, to reduce the probability of finding paralogs. After the multiple-genome alignments were created, the repetitive regions (lowercase in the killer whale genome) were removed using in-house Python scripts.

Phylogenetic Tree Construction

In total, 2,562,378 syntenic blocks were concatenated using in-house Python scripts, and a FASTA-formatted alignment file containing ~1,014 Mb sequences was then generated in our prior analysis. First, non-overlapping genome fragments of 10 kb were extracted using custom Python scripts, creating 101,408 windows from the GSA. Second, for each window, 200 rapid bootstrap analyses were estimated by using RaxML version 8.2.12 (Stamatakis 2014) with the default parameters under the GTR + GAMMA model. In order to ensure that the segments were appropriate for phylogenomic analysis, all the window trees were checked using a custom script for abnormal clusters in which individuals from the same species showed non-monophyletic status, indicating a certain window provided wrong information. After the abnormal trees were filtered, 14,355 windows remained. The median number of parsimony informative sites across 23 individuals for these 10 kb windows was 366, which was relatively informative for further phylogeny reconstruction, and then a frequency distribution graph (Online Resource 2) was plotted by ggplot2 (Wickham 2016). After individual window-based trees were reconstructed, the average bootstrap support values of these trees were summarized using a custom Python script. Again, these 10 k windows were concatenated using in-house Python scripts to form a new matrix (hereafter referred to as NewGSA). Phylogenetic relationships were inferred using both maximum likelihood (ML) and Bayesian methods as implemented in RaxML and ExaBayes version 1.4 (Aberer et al. 2014) based on NewGSA with an unpartitioned scheme due to thousands of

windows in NewGSA. Moreover, in our pilot analyses (with 100 bootstrap replicates), we found that partitioning schemes had almost no effect on the final phylogenetic results. For ML analyses, 500 bootstrap replicates were estimated by RaxML under the GTR + GAMMA model. For the Bayesian analyses, default parameters and the GTR + GAMMA model of evolution was used to run two independent unpartitioned analyses. Each run was conducted for 10^6 generations, and Metropolis-Coupling (numCoupledChains = 2) was used to speed up convergence. The initial 10% of runs were discarded as burn-in. Results were checked for convergence and effective sampling sizes of all parameters using the tools included with the ExaBayes package. Supporting values are expressed as posterior probabilities.

Standard concatenation approaches did not model discordance among gene trees beyond the differences in sequence evolution history (Gatesy and Baker 2005). Recent studies have suggested that concatenation methods could yield results that differ from those obtained by coalescent-based methods, especially in the presence of ILS (Degnan and Rosenberg 2006; Liu et al. 2015; Roch and Steel 2015). This possible limitation can theoretically be overcome with the gene-tree-based coalescent approach using software such as ASTRAL version 5.6.2 (Zhang et al. 2018). Here, the remained windows were used as the input to ASTRAL. Briefly, for each window, 200 rapid bootstrap analyses were estimated using RaxML with the default parameters under the GTR + GAMMA model. Furthermore, three tree subsets were generated, selecting windows with average bootstrapping values over 70% (8,941 windows), 75% (4,222 windows), and 80% (1,085 windows). ASTRAL was applied to reconstruct the species tree based on all window-based trees and the three tree subsets using the default parameters. For the species tree based on all window-based trees, 200 multi-locus bootstrapping replicates were implemented using ASTRAL-III (option: -r 200) (Seo 2008). In addition, another coalescent-based approach was employed to estimate a species tree. SVDquartets (Chifman and Kubatko 2014, 2015), as implemented in PAUP* version 4.0a (Swofford 2003) was performed on the same 200 bootstrap replicates that we generated window-based tree based on ML analysis.

Divergence Time Calibration

Divergence dating analyses were conducted using the software MCMCTree version 4.9 h, part of the PAML package (Yang 2007), with the approximate likelihood method (dos Reis et al. 2012). Phylogenetic analyses of different genomic segments that were shaped by different evolutionary histories will yield trees with unique clade clusters and branch lengths, which can produce a range of time estimates in divergence dating analyses (Fontaine et al. 2015; Springer

et al. 2019). To overcome difficulties arising from computational efficiency and complicated evolutionary histories, we only used 10 k window sequences that had the same topology as the main topology. Although these divergence times are recent-biased, the amount of error is expected to be relatively small considering the deep time scale in this analysis (Frantz et al. 2013). Meanwhile, we only used species-level datasets (retaining the highest individual coverage of particular species) to ensure a high level of computational efficiency. To do this, a species-level dataset was extracted by SeqKit version 0.12.0 (Shen et al. 2016) from the NewGSA. For our main analyses, correlated rate (clock = 3) was used, and the parameters of the prior birth–death process were fixed at birth rate $\lambda = 1$, death rate $\mu = 1$, and sampling fraction $\rho = 0.1$, which led to a uniform distribution of node ages (Yang and Rannala 2006). The posterior distribution of parameters was estimated using Markov chain Monte Carlo (MCMC) sampling. MCMC runs were conducted twice for 10^8 iterations with a sampling frequency of 1000, with the first 10^4 steps of each run discarded as burn-in. The results were examined in Tracer version 1.7 (Rambaut et al. 2018) to evaluate whether parameters, node ages, and likelihood values had converged. We checked that the estimated sample size (ESS) for each parameter was above 200 (Nascimento et al. 2017).

Five fossil calibrations were used in our molecular dating analyses: 1) the divergence between Cetacea and Hippopotamidae was calibrated at 66–52.4 Ma based on the oldest reported crown artiodactyl, *Himalayacetus subathuensis* (Bajpai and Gingerich 1998; Benton et al. 2015); 2) the divergence between baleen and toothed whales was calibrated using the earliest record of a cetacean from 36.4 ± 1.5 Ma (Lambert et al. 2017); 3) the divergence between Phocoenidae (porpoises) and Monodontidae (narwhals) was calibrated at 19.50–7.50 Ma using the oldest Phocoenidae + Monodontidae record of *Salumiphocoena stocktoni* (Barnes 1985) and the oldest delphinidan record of *Kentriodon pernix* (Kellogg 1927); 4) the age of the root of Delphinidae (exclusive of *Lagenorhynchus albirostris*) was calibrated to 19.50–8.50 Ma based on the oldest stem *Orcinus* record of *Eodelphinus kabatensis* (Murakami et al. 2014) and the oldest delphinidan record of *Kentriodon pernix* (Kellogg 1927); 5) the age of the root of Delphininae (exclusive of *Sotalia guianensis*) was calibrated at 8.50–3.98 Ma using the oldest delphinine record of *Etruri-delphis giulii* (Bianucci 2013) and the oldest stem *Orcinus* record of *Eodelphinus kabatensis* (Murakami et al. 2014).

Gene Flow Analyses

Seemingly conflicting or unstable phylogenetic reconstructions can be caused by gene flow among species (Li et al. 2016). Here, we tried two different types of gene flow

analyses. First, we employed ABBA-BABA based analyses (Green et al. 2010; Durand et al. 2011), which are relatively simple, direct, and robust. Then, we applied the tree-based method, which tries to accommodate ILS and gene flow together to infer evolution history.

The ABBA-BABA tests measured the excess of shared polymorphisms (SNPs) of two closely related species concerning a third species (Green et al. 2010) and discriminated between gene flow and ILS. To examine possible gene flow among different species of Delphinidae, the program ANGSD (Korneliussen et al. 2014) was used for admixture analysis among all the individuals in Delphinidae using killer whale as the outgroup. All possible four-taxon topologies of the Delphinidae species including *Tursiops aduncus*, *T. truncatus*, *Stenella coeruleoalba*, *Stenella attenuata*, *Delphinus delphis*, *Sousa chinensis*, *Grampus griseus*, and *Sa. obliquidens* were involved in the gene flow analysis using *D*-statistics. In total, 369 combinations were used in this analysis. ANGSD allows the calculation of a *Z*-value based on jackknifing with 5 Mb blocks (Efron 1981) for the null hypothesis that the *D*-statistic is 0, which means that the existence of introgression can be rejected if the *Z*-value is below the significance level (an absolute value of the *Z* above 3 is often used as a critical value).

In addition, we analyzed the data using D_{FOIL} -statistics (Pease and Hahn 2015) to detect the direction of introgression. This test required an asymmetric five-taxon tree with a specific topology ($((P_1, P_2), (P_3, P_4)), \text{Outgroup}$), and (P_3, P_4) divergence occurring before the divergence of P_1 and P_2 . Therefore, not all combinations of oceanic dolphin taxa could be analyzed (though all individuals of the genera *Tursiops* and *Stenella* could be), and in total, 48 combinations were calculated. Again, the killer whale was used as the outgroup. For each combination, we used the 100 k bp non-overlap windows, which were extracted from NewGSA using a custom Python script, and SeqKit was used to extract combinations from these 100 k bp windows. Other parameters of D_{FOIL} -statistics were left at default.

To investigate whether hybridization windows contribute to specific biological functions, after each window was identified, the corresponding sequence of the killer whale was retrieved and compared against all coding DNA sequences of killer whale using blastn, requiring an E value $\leq 10^{-40}$, identity $\geq 70\%$, and coverage $\geq 60\%$ or better. Then, all genes from the hybridization windows were enriched to assigned GO terms (Gene Ontology Consortium 2004) for Biological Process (BP). Finally, we kept all genes that met a significance threshold of $P < 0.05$ after false discovery rate (FDR) correction (Benjamini and Hochberg 1995).

In the PhyloNet (Wen et al. 2018) analysis, we used the trees based on 10 k non-overlap windows. The highest individual coverage of particular species was retained from the input window trees because the phylogenetic positions of

other individuals were unambiguous. Two methods were applied to infer phylogenetic networks, including maximum parsimony (MP) and maximum pseudo-likelihood (MPL). The MP and MPL models were run with 50 iterations, and the maximum number of reticulations was set to either one or two, yielding five optimal networks. All other parameters were left at default settings. Analyzing networks with more than two reticulations was too complex and not interpretable from the PhyloNet results.

Results

Phylogenetic Relationships

A ML tree and a Bayesian tree based on the concatenated sequence alignments of ~143 Mbp revealed the following relationship with the highest bootstrap supports or posterior probabilities for all nodes (Fig. 2a): the respective monophyletic status of Cetacea and its two suborders (Mysticeti and Odontoceti), the differentiation between Delphinidae and other two families (Phocoenidae and Monodontidae) within superfamily Delphinoidea, the

placement of *O. orca* as the earliest branching lineage in Delphinidae and the subsequent successive divergence of *Sa. obliquidens*, *G. griseus*, *So. chinensis* and *D. delphis*. In particular, both *Tursiops* and *Stenella* were found to be non-monophyletic, with *T. aduncus* most closely related to *St. attenuata* and *T. truncatus* most closely related to *St. coeruleoalba*.

In addition to the concatenation method above, which did not consider ILS, phylogenetic relationships were also inferred using the coalescent-based method. The ASTRAL tree based on a sliding window approach was congruent with the concatenated analyses (Fig. 2b). All nodes received the highest support values. Unsurprisingly, ASTRAL trees based on the three tree subsets produced identical topologies as the ASTRAL tree based on all window-based trees (Online Resource 3). SVDquartets analysis also supported the non-monophyly of *Tursiops* and *Stenella*, although this result conflicted with the inferred topology from the RaxML, ExaBayes and ASTRAL analyses (Online Resource 4). Moreover, we summed up the topologies that supported non-monophyly or monophyly of *Tursiops* and *Stenella*. We noticed more trees supported the non-monophyletic status of *Tursiops* and *Stenella* than supported the monophyletic

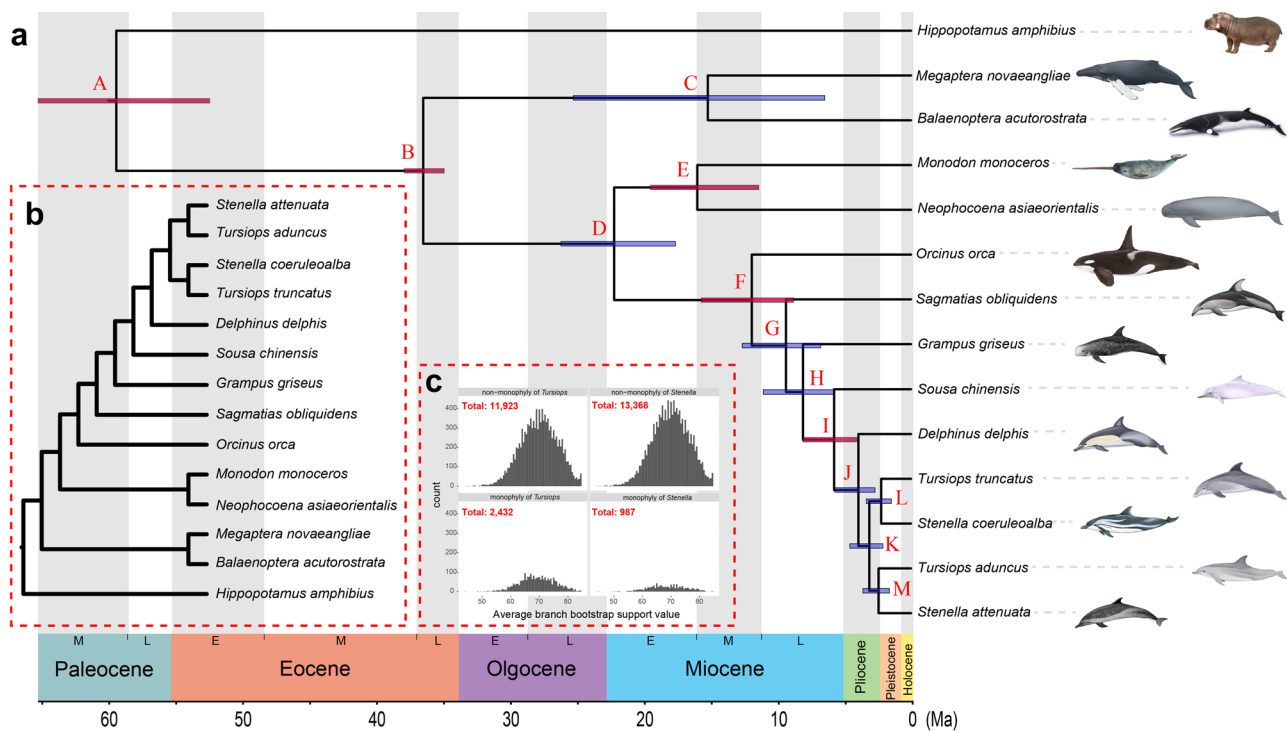


Fig. 2 Phylogeny of cetaceans. **a** The concatenation-based phylogenetic tree from NewGSA data set and 5 fossil calibrations; **b** Coalescent-based tree inferred from 10 k bp window-based trees by ASTRAL; **c** Distribution of average bootstrap support value for individual window. All nodes in **a** have the highest support or posterior probabilities. In **a**, clade letters are identical to those in Table 1. Red boxes in **a** at

each node represent calibration points and blue boxes indicate divergence dates estimated without prior calibration constraints for that node. In **a**, the bounds of the boxes correspond to the 95% highest posterior density (HPD) of each node. All nodes in **b** have the highest support values

status of both genera, which further corroborated the non-monophyly of *Tursiops* and *Stenella* (Fig. 2c).

Divergence Time Calibration

To provide a tree-calibrated tree for the gene flow analyses, D_{FOIL} , the ML tree, and corresponding sequence data, which can build the same topology as the ML tree, were used as input for the program MCMCTree in the software package PAML4 to estimate divergence time. All estimated divergence times for nodes with labels from A to M are presented in Fig. 2 and Table 1. We obtained a mean age for Whippomorpha (\bar{x} = 59.45 Ma, 95% highest posterior density (HPD): 66.00–52.46 Ma), which diversified in the middle Paleocene. The split between Mysticeti and Odontoceti was estimated to occur in the late Eocene (\bar{x} = 36.49 Ma, 95% HPD: 37.91–34.92 Ma), which is consistent with the record of the first documented fossil crown cetacean *Mystacodon selenensis* (~36.4 Ma). The age of Delphinoidea was much more recent (\bar{x} = 22.21 Ma, 95% HPD: 26.19–17.62 Ma), less than three million years older than the oldest known delphinidan record of *Kentriodon pernix*. The divergence of Delphinidae started in the early to middle Miocene (\bar{x} = 11.91 Ma, 95% HPD: 15.70–8.77 Ma), whereas Delphininae may have diverged in the Pliocene.

Extensive Gene Flow among Different Dolphin Species

We investigated interspecies gene flow signals using some gene flow analyses. The D -statistics analyses, also called ABBA-BABA tests, found evidence of gene flow among

Table 1 Divergence times of lineages analyzed in this study, estimated from five fossil calibrations using MCMCTREE. Clade letters refer to those shown in Fig. 2. The asterisk indicates that this clade was constrained in the analysis. HPD=highest posterior density. Units are in million years, Ma

Clade	Age	Lower 95% HPD	Upper 95% HPD
A*	59.45	52.46	66.00
B*	36.49	34.92	37.91
C	15.20	6.45	25.26
D	22.21	17.62	26.19
E*	15.99	11.38	19.50
F*	11.91	8.77	15.70
G	9.35	6.76	12.63
H	8.08	5.80	11.05
I*	5.75	4.05	8.10
J	3.94	2.69	5.70
K	3.12	2.10	4.58
L	2.23	1.47	3.35
M	2.43	1.62	3.61

species within the family Delphinidae (Fig. 3, Online Resource 5). Our ABBA-BABA tests showed that the ABBA-BABA comparisons (Z -value > 3) are significantly different from zero, suggesting that gene flow may be an important factor in generating the reticulations among these lineages. These analyses supported the view that D is significant in many of the groups (95%, 352 out of 369 groups; Online Resource 5), which revealed a complex network of ancestors among several dolphin lineages (Fig. 3). For example, trios 37, (((*St. attenuata*, *T. aduncus*), *St. coeruleoalba*), *O. orca*), exhibited significantly positive D values, suggesting that hybridization between *T. aduncus* and *St. coeruleoalba* might lead to sharing some genetic variation. In other words, *T. aduncus* was found to be closer to *St. coeruleoalba* than *St. attenuata* was to *St. coeruleoalba*. In contrast, trios 34, (((*T. aduncus*, *T. truncatus*), *D. delphis*), *O. orca*) exhibited negative D values, showing more derived allele sharing between *T. aduncus* and *D. delphis* than between *T. truncatus* and *D. delphis* (Fig. 3). The fact that the D -statistics analyses found evidence of gene flow between most dolphin sister species may be related to their complex evolutionary history (Fig. 3).

To detect the direction of introgression between different genus pairs, we implemented the D_{FOIL} analyses. In all 48 tests, gene flow can be detected in approximately 9–12% of individual permutations (Fig. 4), showing that gene flow among the four dolphins existed and was strong. Meanwhile, in all 48 tests, among the detected gene flow events, the number of ancestral gene flow scenarios ($P_{12} < = > P_3$ and $P_{12} < = > P_4$) without distinguished direction were higher than the number of directional gene flow scenarios (recent gene flow scenarios, e.g., $P_1 = > P_3$). *Tursiops* and *Stenella* exhibited widespread signals of ancient gene flow, likely due to their broad historical ranges throughout the ocean, overlapping with several congeneric species.

To test whether the introgressed windows in Delphinidae were related to fitness or adaptation, the 314 introgressed windows identified by D_{FOIL} analyses were compared against all killer whale coding-protein sequences and were found to contain notable matches to 6,378 genes. Of these 6,378 genes, 167 genes (e.g., *PBX1*, *IL-17F*) were associated with “skeletal system development” (GO: 0001501, $P=0.00466 < 0.01$), 129 genes (e.g., *BARHL2*, *RBP4*) were associated with “regulation of developmental growth” (GO: 0048638, $P=2.07e^{-05} < 0.01$), and 106 genes (e.g., *PTPRS*, *PAK1*) were associated with the GO pathway “developmental growth involved in morphogenesis” (GO: 0060560, $P=1.74e^{-07} < 0.01$) (Table 2, Online Resource 6). In addition, among 6,378 genes, 134 genes were related to “visual system development” (GO: 0150063, $P=0.000970 < 0.01$) and 130 genes were related to “response to drug” (GO: 0042493, $P=0.00259 < 0.01$) (Online Resource 6). These results might imply that differential introgression of genes

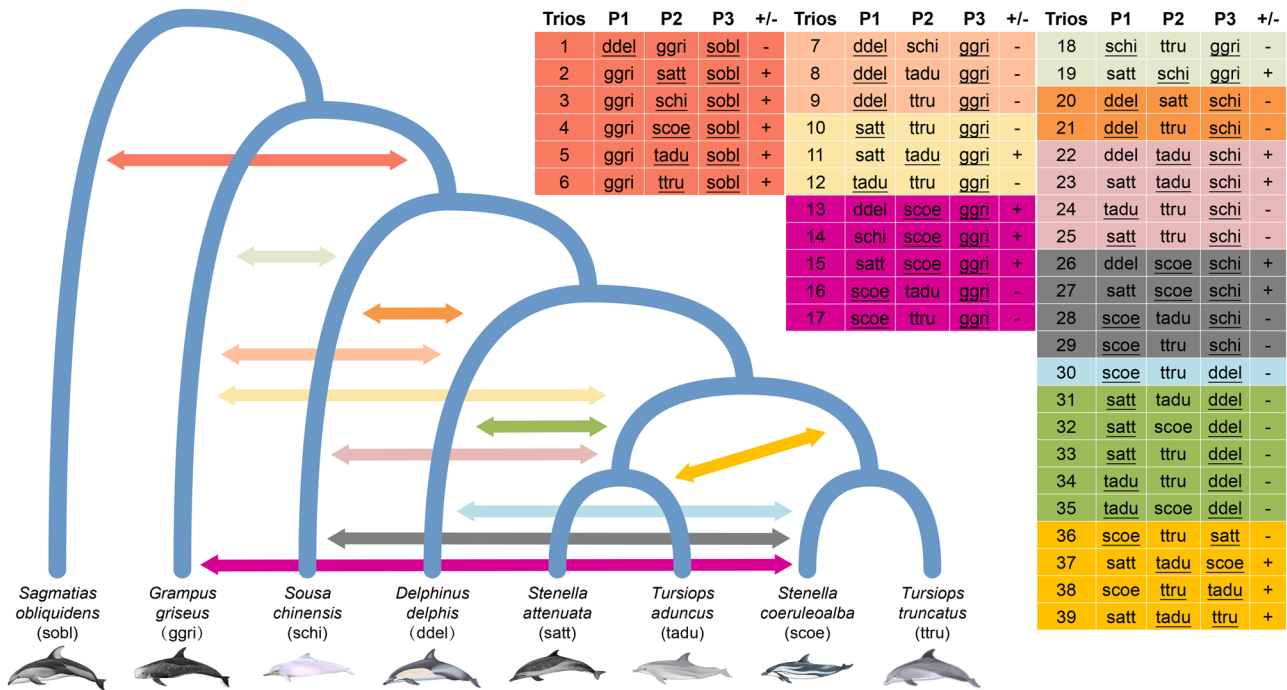


Fig. 3 Gene flow signals for the family Delphinidae inferred by the *D*-statistic. Colored arrows indicate the phylogenetic positions of significant *D* statistics (shown in the table) for different species trios,

using the killer whale as outgroup. Species codes in underlined types indicate pairs for which significant evidence of admixture was detected ($Z > 3$)

gave rise to or contributed to similar morphological features and fitness in dolphins, as suggested for Neanderthal and Denisovan alleles in modern humans (Racimo et al. 2015). For example, *PBX1* plays a crucial role in osteogenesis and matrix mineralization (Cheung et al. 2009). Although our analysis did not explore the specific role of these introgression genes, it did provide pertinent evolutionary hypotheses for future investigation.

In addition to the character-based analysis, gene flow may preferably be studied by tree-based phylogenetic networks

using both MP and MPL models. Although these network topologies differed from the concatenation and coalescent-based method, both pseudolikelihood and parsimony networks with a maximum of two allowed reticulation supported the non-monophyletic status of *Tursiops* and *Stenella*. In both parsimony networks, *St. attenuata* was connected by a hybrid edge to the lineage of *T. truncatus* (Fig. 5a-b). The inferred inheritance probabilities for the hybrid edge differed slightly between the networks with a maximum of one and two allowed reticulations. When we allowed for a

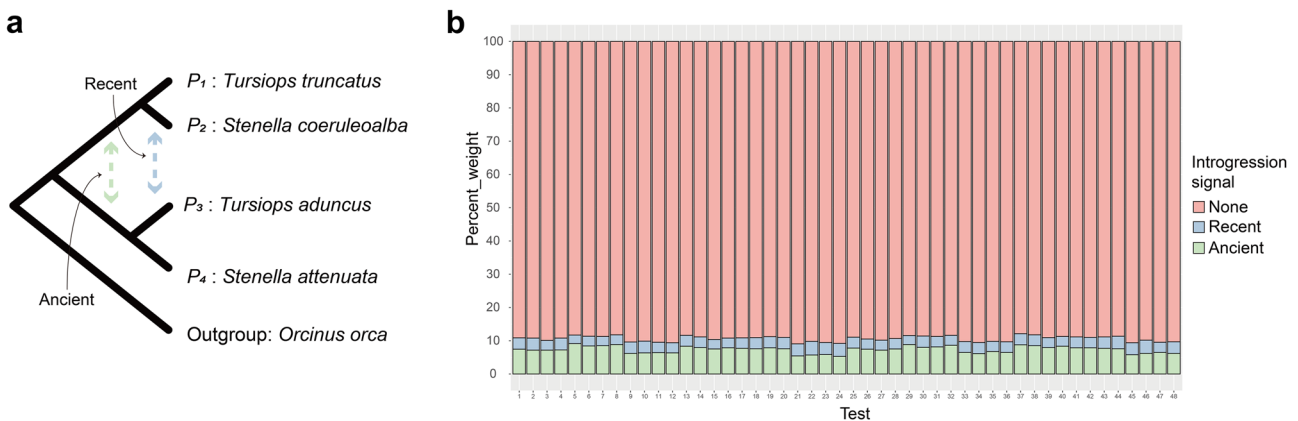


Fig. 4 Gene flow among species in genus *Tursiops* and *Stenella* detected on D_{FOIL} test. **a** Guide tree used to generate D_{FOIL} test comparisons; **b** The results of 48 tests, which were based on 100 k windows

Table 2 Representative enrichment GO categories describing biological processes for introgressed genes

GO Term	Description	Count	p-Value (adjusted)
GO: 0001501	skeletal system development	167	0.00466
GO: 0048638	regulation of developmental growth	129	2.07e ⁻⁰⁵
GO: 0060560	developmental growth involved in morphogenesis	106	1.74e ⁻⁰⁷

maximum of two reticulations in the pseudolikelihood networks, the connection between *T. aduncus* and the ancestor of *T. truncatus* and *St. coeruleoalba* was observed (Fig. 5d). Interestingly, *T. aduncus* had a reticulation edge to the stem lineage of *T. truncatus* and *St. coeruleoalba*, resembling the result of the *D*-statistics analyses, which meant that some loci in *T. truncatus* or *St. coeruleoalba* shared a most recent common ancestor with loci from *T. aduncus*.

Discussion

Non-monophyly of *Tursiops* and *Stenella*

The phylogenetic relationships of the family Delphinidae have been notoriously difficult to disentangle due to hybridization and ILS arising from recent explosive speciation (Wiens et al. 2006; Whitfield and Lockhart 2007; McGowen et al. 2020). Although a number of morphological and genetic studies have been conducted in an attempt to resolve their evolutionary relationships (e.g., Perrin et al. 1987; Archer and Perrin 1999; Chen et al. 2011; Moura et al. 2013, 2020; Jedensjö et al. 2017), many aspects of Delphinidae phylogeny remain unresolved, especially the monophyly of *Tursiops* and *Stenella*. Traditionally, morphology-based taxonomy has considered *Tursiops* and *Stenella* to be monophyletic groups, based on features of the skull and superficial morphological observations, respectively (Archer and Perrin 1999; Shirakihara et al. 2003). However, Leduc et al. (1999) published a phylogeny for the 33 then-recognized Delphinidae species based on cytochrome *b*, which revealed that *T. truncatus* formed a sister group with a polytomy formed by *St. coeruleoalba* + *Stenella clymene*, *Delphinus* spp., *Stenella frontalis*, and *T. aduncus*. Their results challenged the monophyly of *Tursiops* and *Stenella* for the first time. Since then, some researchers also have suggested that *Tursiops* is a non-monophyletic group (e.g., Kingston et al. 2009; Zurano et al. 2019). For example, Moura et al. (2013) showed *T. aduncus* clustering with *Delphinus capensis* based on the whole mitochondrial genome, while McGowen (2011) found a sister relationship between *T. aduncus* and *T. truncatus* + *St. coeruleoalba*, based on 20 nuclear genes. However, in another study based on 3,191 protein-coding genes, McGowen et al. (2020) suggested that *T. aduncus* and *T. truncatus* form a monophyletic group that

constitutes a sister group within Delphinidae. For *Stenella*, Leduc et al. (1999) suggested *Stenella* may be an artificial assemblage, as members of this genus are both morphologically and genetically very dissimilar. The genus *Stenella* has been rendered non-monophyletic in nearly all analyses since then (e.g., Nishida et al. 2007; Chen et al. 2011; Dornburg et al. 2012). However, due to the considerable phylogenetic uncertainty among the species within *Stenella*, a formal redescription has not been attempted (McGowen et al. 2020). For example, Steeman et al. (2009) showed *St. coeruleoalba* and *St. frontalis* to be more closely related to members of *Delphinus* than to their congeners, based on six mitochondrial and nine nuclear genes; moreover, Zurano et al. (2019) showed that *St. coeruleoalba* clustered with *St. clymene*, and that *St. frontalis* was sister to the clade including *D. capensis* + *D. delphis* and *T. aduncus*, using a whole mitochondrial genome. Differences in molecular markers, taxon sampling, and phylogenetic inference methods can interfere with resolving phylogenetic relationships and result in contrasting conclusions. For this reason, we tried to conduct phylogenetic analyses with large-scale genomic sequence data in this study, avoiding the impact of different markers, and provide a novel opportunity to address the problem of *Tursiops* and *Stenella* monophyly.

In the present study, phylogenetic reconstruction was performed based on concatenation and coalescent-based approaches to infer the relationships within Delphinidae. Consistency between these results based on different tree-building methods should be viewed as better support for a hypothesis than bootstrap values from a single analysis (Suh 2016). Trees derived from the concatenation and coalescent-based method (ASTRAL) recovered a *T. truncatus* + *St. coeruleoalba* clade and a *T. aduncus* + *St. attenuata* clade, both of which were supported by the highest support value (Fig. 2a-b). This non-monophyletic result coincided with previous results based on whole mitogenomes (Moura et al. 2013; Zurano et al. 2019). Furthermore, the window tree statistics showed approximately five times as many window trees supporting the non-monophyly of *Tursiops* as window trees that recovered *Tursiops* as a monophyletic lineage; similarly, window trees overwhelmingly provided support for the non-monophyly of *Stenella* (Fig. 2c). These results implied that the genomic fragments supporting *Tursiops* and *Stenella* as non-monophyletic groups are widely distributed throughout the whole genome. In addition, two phylogenetic

network topologies supported the non-monophyletic status of *Tursiops* and *Stenella* (Fig. 5b, d). Taking all these evidences together, based on different phylogenetic analyses and the largest genomic data so far, our analyses provided novel strong support for the non-monophyly of *Tursiops* and *Stenella* and constitute the best current non-monophyletic hypothesis for the genera *Tursiops* and *Stenella*.

Early molecular studies based on a few nuclear DNA or the whole mitochondrial genomes showed that, with increased taxon sampling, the genera *Tursiops* and *Stenella* are recovered as non-monophyletic groups although internal nodes often have <70% bootstrap support (e.g., Kingston

et al. 2009; Hassanin et al. 2012). In agreement with previous studies (Kingston et al. 2009; Hassanin et al. 2012), we resolved a non-monophyletic *Tursiops* and *Stenella*; in particular, we recovered strong support values for all internal nodes. Nevertheless, some phylogenomic studies have supported *T. aduncus* and *T. truncatus* as a monophyletic group within Delphininae (e.g., McGowen et al. 2020; Moura et al. 2020). Based on the Bayesian inference method, Moura et al. (2020) used RADseq data including nearly all delphinine species and recovered *Tursiops* as a monophyletic clade with 0.92 posterior probability. However, it should be noted that the Bayesian inference method consistently overestimates

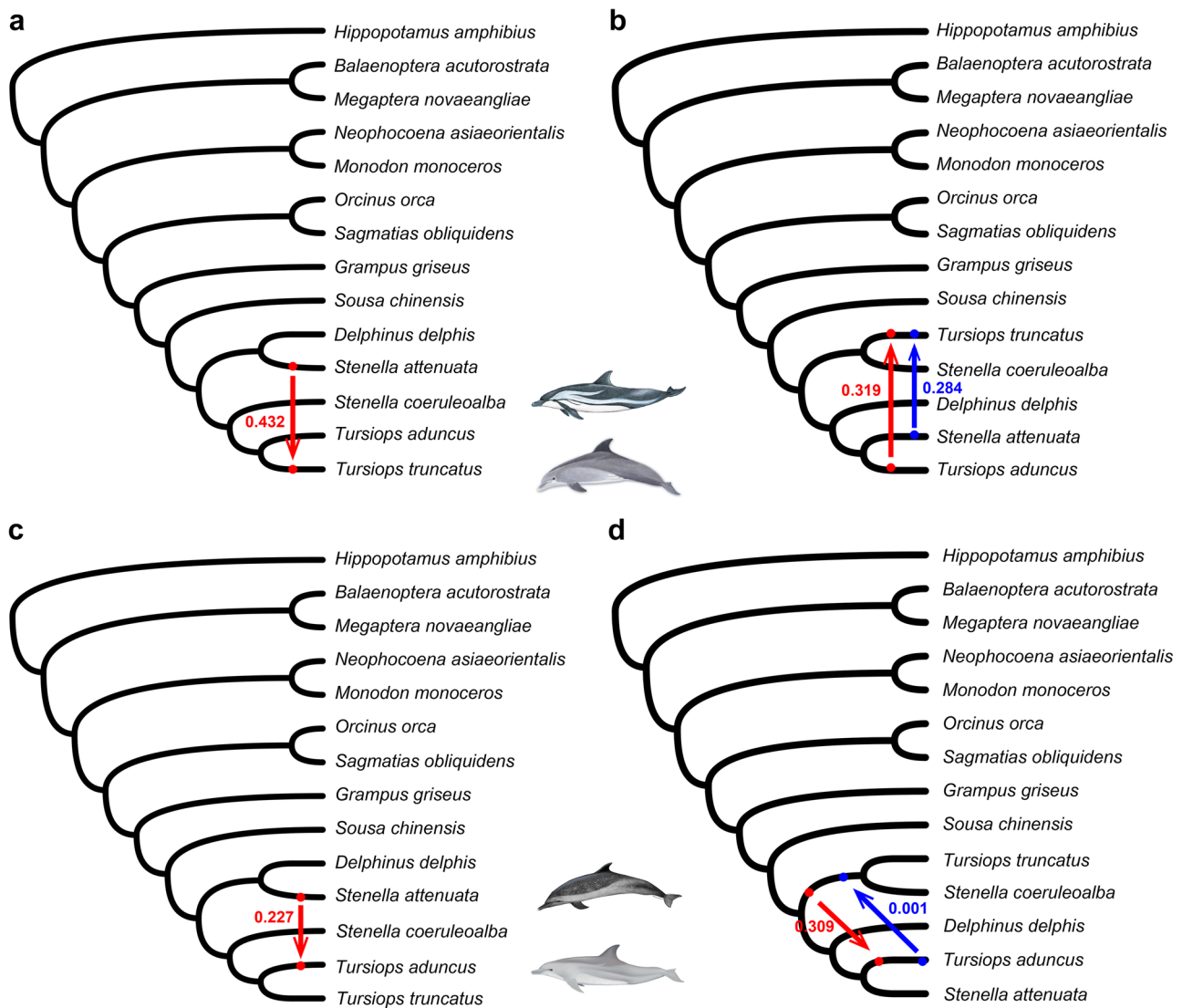


Fig. 5 Gene flows among species inferred by PhyloNet. **a** The optimal network inferred from PhyloNet under MP model with one maximum reticulation; **b** The optimal network inferred from PhyloNet under MP model with two maximum reticulations; **c** The optimal network inferred from PhyloNet under MPL model with one maximum reticulation; **d** The optimal network inferred from

PhyloNet under MPL model with two maximum reticulations. The reticulation edges are colored in red and blue, respectively. The inheritance probabilities are shown on the reticulation edges. Illustrations represented (from top to bottom): *Stenella coeruleoalba*, *Tursiops truncatus*, *Stenella attenuata* and *Tursiops aduncus*

support and therefore should not be interpreted as probabilities that clades are correctly resolved (Simmons et al. 2004). In particular, compared with Moura et al. (2020), phylogenetic trees inferred from concatenation (ML and Bayesian) and coalescent-based approaches were concordant with each other in the present study, which supported non-monophyly of *Tursiops*. Based on the largest genetic assessment of the *Tursiops* species to date, we showed that the current taxonomic framework of *Tursiops* greatly underestimates its diversity. Previously, Charlton-Robb et al. (2011) found *Tursiops* to be non-monophyletic and named *Tursiops australis* sp. nov. with the common name of ‘Burrnun Dolphin’, based on morphological and genetic differentiation; Gray et al. (2018) also discovered a new ‘*aduncus*’ type lineage in the Arabian Sea. Considering that traditional species groupings of dolphins were based largely on the skull or superficial morphological observations (Perrin et al. 2013), it is not surprising that the taxonomy of *Tursiops* and *Stenella* has been surrounded by uncertainty and indicates gene flow in their evolutionary history (see the next section). In addition, we are aware that poor taxon sampling can result in an ambiguous result (Philippe et al. 2011); therefore, the current non-monophyletic status of the genera *Tursiops* and *Stenella* still needs further validation when more genome data for *Tursiops* and *Stenella* are available.

Widespread Gene Flow among Oceanic Dolphins

Hybridization might produce phylogenetic discordance in genomes, which has been especially observed in species undergoing rapid speciation such as big cats, butterflies, and cichlids (Figueiró et al. 2017; Malinsky et al. 2018; Edelman et al. 2019). Oceanic dolphins are iconic mammals with a complex evolutionary history due to a recent and rapid diversification process (Steeman et al. 2009; Zhou et al. 2011). A large number of phylogenetic discordances extensively reported in dolphins are commonly explained by rapid radiation associated with ILS or hybridization among lineages (e.g., Amaral et al. 2012; Perrin et al. 2013). However, only very few studies have detected ILS or hybridization in dolphins with molecular data (Kingston et al. 2009; Moura et al. 2020). For example, Kingston et al. (2009) applied the STRUCTURE program (Pritchard et al. 2000; Falush et al. 2003) to 418 polymorphic markers and indicated some hybridization events between the two spotted dolphin species. However, the potential hybridization events outside these two species may be hidden because not all of their studied species were tested. Instead, it is necessary to take more combinations among dolphins into consideration to accurately explore hybridization events and their effects on dolphin evolution.

In this study, we applied different statistical methods, i.e., D -statistic, D_{FOIL} and PhyloNet analyses, to illuminate the

existence of gene flow in dolphins. The D -statistic found remarkable gene flow among all members of Delphinidae (Fig. 3). Although only one *So. chinensis* sample was included, the Z -values > 3 of most combinations including *So. chinensis* (74/75, Online Resource 5) were massive. Besides, only one individual of *St. attenuata* was sampled. Of course, small sample size may cause biased results due to SNP (Hoggart et al. 2008) and thus, the current gene flow signals needed to be further confirmed in the future. Furthermore, considering the wide overlapping geographical distribution among different dolphin species (Jefferson and LeDuc 2018), as well as several studies showing successful mating between different dolphin species (Crossman et al. 2016; Gridley et al. 2018), it has been regarded that extensive gene flow could characterize all oceanic dolphins rather than just closely related species, and it may still be ongoing. Furthermore, ancestral gene flow is expected to leave signatures in genomes of their descendants, causing some phylogenetic conflict (Kumar et al. 2017). The instance of ancestral gene flow between *T. aduncus* and the ancestor of *T. truncatus* and *St. coeruleoalba* (Figs. 3 and 5d), might be the factor leading previous studies to infer an incorrect relationship between *T. aduncus* and *T. truncatus* in previous cases (e.g., McGowen et al. 2009; Amaral et al. 2012). Although it was inferred in the present study that gene flow might be responsible for part of the non-concordance within the dolphin evolutionary tree because the D values calculated by D -statistic and D_{FOIL} were significantly different from zero (Figs. 3 and 4), phylogenetic discordance might also be attributed to ILS (Whitfield and Lockhart 2007). Therefore, future studies are required to further identify the role of these two processes. Based on our data, our study strongly suggested pervasive gene flow in oceanic dolphins, and this may help explain why previous phylogenetic studies with limited markers (e.g., Xiong et al. 2009; McGowen 2011; Amaral et al. 2012) were unable to reliably resolve the phylogenetic relationships of this highly rapidly radiated group of animals.

Conclusion

In this study, we produced a high-resolution phylogenetic reconstruction for dolphins with a special focus on addressing the monophyly hypothesis of *Tursiops* and *Stenella*. Based on the genomic sequences, we were able to obtain a well-resolved tree that revealed the power of multiple complete genomes from closely related species to comprehensively infer their evolutionary history. Based on the concordant phylogenetic relationships in the concatenation and coalescent-based analysis, the non-monophyletic status of both *Tursiops* and *Stenella* was strongly suggested, which implied that the current

taxonomy of both genera might not reflect their evolutionary history and might underestimate their diversity. Extensive inter-specific gene flow among species detected with different methods could explain interspecific convergence among species from different genera and also instability in resolving phylogenetic relationships of oceanic dolphins with different and limited markers.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10914-021-09581-6>.

Acknowledgements We are particularly grateful to Xinrong Xu and Bingyao Chen for their contributions to sample collection. Many thanks are also given to Simin Chai and Zhenpeng Yu for support and discussions.

Author Contributions Weijian Guo and Di Sun performed the data analyses and wrote the manuscript. Wenhua Ren, Shixia Xu and Guang Yang designed the study and improved the manuscript. Yang Cao, Lin-Lin Xiao and Xin Huang downloaded and participated in the analysis of the genomic data. All authors read and approved the final manuscript.

Funding This work was supported by the Key Project of the National Natural Science Foundation of China (NSFC) (grant nos. 31630071, 32030011 to G.Y.), the National Key Program of Research and Development, Ministry of Science and Technology of China (grant no. 2016YFC0503200 to G.Y. and S.X.), the National Natural Science Foundation of China (NSFC) (grant nos. 31772448 and 32070409 to S.X., 31872219 to W.R. and 31900321 to D.S.), the Qingnian project of Jiangsu Province to S.X., and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

Data Accessibility Raw sequences have been deposited in GenBank under accession number PRJNA715930. All the 10 k windows, NewGSA, the trees for RaxML, the results of D_{FOIL} analyses and the custom scripts were uploaded to the Figshare digital repository <https://figshare.com/s/34bd4df789e0743fb15d>

Code Availability Sofeware: FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), Trimmomatic (<http://www.usadellab.org/cms/?page=trimmomatic>), SAMtools (<http://www.htslib.org/download/>), BCFtools (<https://samtools.github.io/bcftools/>), LAST (<https://gitlab.com/mcfrith/last>), MULTIZ (<https://github.com/multiz/multiz>), RaxML (<https://github.com/stamatak/standard-RAxML>), ExaBayes (<https://cme.h-its.org/exelixis/web/software/exabayes/>), ASTRAL (<https://github.com/smirarab/ASTRAL>), PAML (<http://abacus.gene.ucl.ac.uk/software/paml.html>), SeqKit (<https://github.com/shenwei356/seqkit>), ANGSD (<https://github.com/ANGSD/angsd/releases>), D_{FOIL} (<https://github.com/jbpease/dfoil>) and PhyloNet (<https://bioinfocsc.rice.edu/phylo-net>). Code: Stored in the Figshare digital repository <https://figshare.com/s/34bd4df789e0743fb15d>

Declarations

Ethics Approval Not applicable.

Consent to Participate Not applicable.

Conflicts of Interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aberer AJ, Kobert K, Stamatakis A (2014) ExaBayes: massively parallel Bayesian tree inference for the whole-genome era. *Mol Biol Evol* 31:2553–2556
- Amaral AR, Jackson JA, Möller LM, Beheregaray LB, Manuela Coelho M (2012) Species tree of a recent radiation: the subfamily Delphininae (Cetacea, Mammalia). *Mol Phylogenet Evol* 64:243–253
- Archer FI, Perrin WF (1999) *Stenella coeruleoalba*. *Mamm Species* 603:1–9
- Árnason Ú, Gullberg A, Janke A (2004) Mitogenomic analyses provide new insights into cetacean origin and evolution. *Gene* 333:27–34
- Árnason Ú, Lammers F, Kumar V, Nilsson MA, Janke A (2018) Whole-genome sequencing of the blue whale and other rorquals finds signatures for introgressive gene flow. *Sci Adv* 4:eaa9873
- Bajpai S, Gingerich PD (1998) A new Eocene archaeocete (Mammalia, Cetacea) from India and the time of origin of whales. *Proc Natl Acad Sci USA* 95:15464–15468
- Barnes LG (1985) Evolution, taxonomy and antitropical distributions of the porpoises (Phocoenidae, Mammalia). *Mar Mammal Sci* 1:149–165
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol* 57:289–300
- Benton MJ, Donoghue PC, Asher RJ, Friedman M, Near TJ, Vinther J (2015) Constraints on the timescale of animal evolutionary history. *Palaeontol Electron* 18:1–106
- Bianucci G (2013) *Septidelphis morii*, n. gen. et sp., from the Pliocene of Italy: new evidence of the explosive radiation of true dolphins (Odontoceti, Delphinidae). *J Vertebr Paleontol* 33:722–740
- Blanchette M, Kent WJ, Riemer C, Elnitski L, Smit AF, Roskin KM, Baertsch R, Rosenbloom K, Clawson H, Green ED, Haussler D, Miller W (2004) Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res* 14:708–715
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120
- Charlton-Robb K, Gershwin L-a, Thompson R, Austin J, Owen K, McKechnie S (2011) A new dolphin species, the Burrunan dolphin *Tursiops australis* sp. nov., endemic to southern Australian coastal waters. *PLoS ONE* 6:e24047
- Chen Z, Xu S, Zhou K, Yang G (2011) Whale phylogeny and rapid radiation events revealed using novel retroposed elements and their flanking sequences. *BMC Evol Biol* 11:314
- Cheung CL, Chan BY, Chan V, Ikegawa S, Kou I, Ngai H, Smith D, DK Luk K, Huang Q-Y, Mori S, Sham P-C, Kung AWC (2009) Pre-B-cell leukemia homeobox 1 (*PBX1*) shows functional and

- possible genetic association with bone mineral density variation. *Hum Mol Genet* 18:679–687
- Chifman J, Kubatko L (2014) Quartet inference from SNP data under the coalescent model. *Bioinformatics* 30:3317–3324
- Chifman J, Kubatko L. (2015) Identifiability of the unrooted species tree topology under the coalescent model with time-reversible substitution processes, site-specific rate variation, and invariable sites. *J Theor Biol* 374:35–47
- Crossman CA, Taylor EB, Barrett-Lennard LG (2016) Hybridization in the Cetacea: widespread occurrence and associated morphological, behavioral, and ecological factors. *Ecol Evol* 6:1293–1303
- Dalebout ML, Steel D, Baker CS (2008) Phylogeny of the beaked whale genus *Mesoplodon* (Ziphiidae: Cetacea) revealed by nuclear introns: implications for the evolution of male tusks. *Syst Biol* 57:857–875
- Degnan JH, Rosenberg NA (2009) Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol Evol* 24:332–340
- Degnan JH, Rosenberg NA (2006) Discordance of species trees with their most likely gene trees. *PLoS genet* 2:e68
- Delsuc F, Brinkmann H, Philippe H (2005) Phylogenomics and the reconstruction of the tree of life. *Nat Rev Genet* 6:361–375
- Dornburg A, Brandley M, McGowen M, Near T (2012) Relaxed clocks and inferences of heterogeneous patterns of nucleotide substitution and divergence time estimates across whales and dolphins (Mammalia: Cetacea). *Mol Biol Evol* 29:721–736
- dos Reis M, Inoue J, Hasegawa M, Asher RJ, Donoghue PCJ, Yang Z (2012) Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proc R Soc B* 279:3491–3500
- Durand EY, Patterson N, Reich D, Slatkin M (2011) Testing for ancient admixture between closely related populations. *Mol Biol Evol* 28:2239–2252
- Edelman NB, Frandsen PB, Miyagi M, Clavijo B, Davey J, Dikow RB, García-Accinelli G, Van Belleghem SM, Patterson N, Neafsey DE, Challis R, Kumar S, Moreira GRP, Salazar C, Chouteau M, Counterman BA, Papa R, Blaxter M, Reed RD, Dasmahapatra KK, Kronforst M, Joron M, Jiggins CD, McMillan WO, Di Palma F, Blumberg AJ, Wakeley J, Jaffe D, Mallet J (2019) Genomic architecture and introgression shape a butterfly radiation. *Science* 366:594–599
- Efron B (1981) Nonparametric estimates of standard error: the jack-knife, the bootstrap and other methods. *Biometrika* 68:589–599
- Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164:1567–1587
- Figueiró HV, Li G, Trindade FJ, Assis J, Pais F, Fernandes G, Santos SHD, Hughes GM, Komissarov A, Antunes A, Trinca CS, Rodrigues MR, Linderoth T, Bi K, Silveira L, Azevedo FCC, Kantek D, Ramalho E, Brassaloti RA, Villela PMS, Nunes ALV, Teixeira RHF, Morato RG, Loska D, Saragüeta P, Gabaldón T, Teeling EC, O'Brien SJ, Nielsen R, Coutinho LL, Oliveira G, Murphy WJ, Eizirik E (2017) Genome-wide signatures of complex introgression and adaptive evolution in the big cats. *Sci Adv* 3:e1700299
- Fontaine MC, Pease JB, Steele A, Waterhouse RM, Neafsey DE, Sharakhov IV, Jiang X, Hall AB, Catteruccia F, Kakani E, Mitchell SN, Wu Y-C, Smith HA, Love RR, Lawniczak MK, Slotman MA, Emrich SJ, Hahn MW, Besansky NJ (2015) Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science* 347:1258524
- Foote AD, Liu Y, Thomas GW, Vinar T, Alfoldi J, Deng J, Dugan S, van Elk CE, Hunter ME, Joshi V, Khan Z, Kovar C, Lee SL, Lindblad-Toh K, Mancia A, Nielsen R, Qin X, Qu J, Raney BJ, Vijay N, Wolf JB, Hahn MW, Muzny DM, Worley KC, Gilbert MT, Gibbs RA (2015) Convergent evolution of the genomes of marine mammals. *Nat Genet* 47:272–275
- Frantz LAF, Schraiber JG, Madsen O, Megens H-J, Bosse M, Paudel Y, Semiadi G, Meijaard E, Li N, Crooijmans RPMA, Archibald AL, Slatkin M, Schook LB, Larson G, Groenen MAM (2013) Genome sequencing reveals fine scale diversification and reticulation history during speciation in *Sus*. *Genome Biol* 14:1–12
- Gatesy J, Baker RH (2005) Hidden likelihood support in genomic data: can forty-five wrongs make a right? *Syst Biol* 54:483–492
- Gatesy J, Milinkovitch M, Waddell V, Stanhope M (1999) Stability of cladistic relationships between Cetacea and higher-level artiodactyl taxa. *Syst Biol* 48:6–20
- Gatesy J, O'Leary MA (2001) Deciphering whale origins with molecules and fossils. *Trends Ecol Evol* 16:562–570
- Gatesy J, Geisler JH, Chang J, Buell C, Berta A, Meredith RW, Springer MS, McGowen MR (2013) A phylogenetic blueprint for a modern whale. *Mol Phylogenet Evol* 66:479–506
- Geisler JH, McGowen MR, Yang G, Gatesy J (2011) A supermatrix analysis of genomic, morphological, and paleontological data from crown Cetacea. *BMC Evol Biol* 11:1–33
- Geisler JH, Sanders AE (2003) Morphological evidence for the phylogeny of Cetacea. *J Mamm Evol* 10:23–129
- Gene Ontology Consortium (2004) The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* 32:D258–D261
- Gray HWI, Nishida S, Welch AJ, Moura AE, Tanabe S, Kiani MS, Culloch R, Möller L, Natoli A, Ponnampalam LS, Minton G, Gore M, Collins T, Willson A, Baldwin R, Hoelzel AR (2018) Cryptic lineage differentiation among Indo-Pacific bottlenose dolphins (*Tursiops aduncus*) in the northwest Indian Ocean. *Mol Phylogenet Evol* 122:1–14
- Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH-Y, Hansen NF, Durand EY, Malaspina A-S, Jensen JD, Marques-Bonet T, Alkan C, Prüfer K, Meyer M, Burbano HA, Good JM, Schultz R, Aximu-Petri A, Butthof A, Höber B, Höffner B, Siegemund M, Weihmann A, Nusbaum C, Lander ES, Russ C, Novod N, Affourtit J, Egholm M, Verna C, Rudan P, Brajkovic D, Kucan Ž, Gušić I, Doronichev VB, Golovanova LV, Lalueva-Fox C, de la Rasilla M, Fortea J, Rosas A, Schmitz RW, Johnson PLF, Eichler EE, Falush D, Birney E, Mullikin JC, Slatkin M, Nielsen R, Kelso J, Lachmann M, Reich D, Pääbo S (2010) A draft sequence of the Neandertal genome. *Science* 328:710–722
- Gridley T, Elwen SH, Harris G, Moore DM, Hoelzel AR, Lampen F (2018) Hybridization in bottlenose dolphins—A case study of *Tursiops aduncus* × *T. truncatus* hybrids and successful back-cross hybridization events. *PLoS ONE* 13:e0201722
- Guidarelli G, Nicolosi P, Fusco G, de Francesco MC, Loy A (2014) Morphological variation and modularity in the mandible of three Mediterranean dolphin species. *Ital J Zool* 81:354–367
- Guidarelli G, Colangelo P, de Francesco MC, Nicolosi P, Meloro C, Loy A (2018) Phenotypic changes across a geographic gradient: the case of three sympatric dolphin species. *Evol Biol* 45:113–125
- Hassanin A, Delsuc F, Ropiquet A, Hammer C, Jansen van Vuuren B, Matthee C, Ruiz-Garcia M, Catzeflis F, Areskou V, Nguyen TT, Couloux A (2012) Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a comprehensive analysis of mitochondrial genomes. *C R Acad Sci* 335:32–50
- Hoggart CJ, Clark TG, De Iorio M, Whittaker JC, Balding DJ (2008) Genome-wide significance for dense SNP and resequencing data. *Genet Epidemiol* 32:179–185
- Jedensjö M, Kemper CM, Krützen M (2017) Cranial morphology and taxonomic resolution of some dolphin taxa (Delphinidae) in Australian waters, with a focus on the genus *Tursiops*. *Mar Mammal Sci* 33:187–205

- Jefferson TA, LeDuc R (2018) Delphinids, overview. In: Würsig B, Thewissen JGM, and Kovacs KM (ed) Encyclopedia of Marine Mammals, 3rd edn. Elsevier Amsterdam, Netherlands, pp. 242-246
- Kellogg R (1927) *Kentriodon pernix*, a Miocene porpoise from Maryland. Proc US Natl Mus 69:1-55
- Kiełbasa SM, Wan R, Sato K, Horton P, Frith MC (2011) Adaptive seeds tame genomic sequence comparison. Genome Res 21:487-493
- Kingston SE, Adams LD, Rosel PE (2009) Testing mitochondrial sequences and anonymous nuclear markers for phylogeny reconstruction in a rapidly radiating group: molecular systematics of the Delphininae (Cetacea: Odontoceti: Delphinidae). BMC Evol Biol 9:245
- Korneliusen TS, Albrechtsen A, Nielsen R (2014) ANGSD: analysis of next generation sequencing data. BMC Bioinform 15:356
- Kumar V, Lammers F, Bidon T, Pfenninger M, Kolter L, Nilsson MA, Janke A (2017) The evolutionary history of bears is characterized by gene flow across species. Sci Rep 7:46487
- Lambert O, Martínez-Cáceres M, Bianucci G, Di Celma C, Salas-Gismondi R, Steurbaut E, Urbina M, de Muizon C (2017) Earliest mysticete from the late Eocene of Peru sheds new light on the origin of baleen whales. Curr Biol 27:1535-1541
- Leduc RG, Perrin WF, Dizon AE (1999) Phylogenetic relationships among the delphinid cetaceans based on full cytochrome *b* sequences. Mar Mammal Sci 15:619-648
- Li G, Davis BW, Eizirik E, Murphy WJ (2016) Phylogenomic evidence for ancient hybridization in the genomes of living cats (Felidae). Genome Res 26:1-11
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754-1760
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) the sequence alignment/map format and SAMtools. Bioinformatics 25: 2078-2079
- Liu L, Xi Z, Davis CC (2015) Coalescent methods are robust to the simultaneous effects of long branches and incomplete lineage sorting. Mol Biol Evol 32:791-805
- Lum JK, Nikaïdo M, Shimamura M, Shimodaira H, Shedlock AM, Okada N, Hasegawa M (2000) Consistency of SINE insertion topology and flanking sequence tree: quantifying relationships among cetartiodactyls. Mol Biol Evol 17:1417-24
- Malinsky M, Svardal H, Tyers AM, Miska EA, Genner MJ, Turner GF, Durbin R (2018) Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. Nat Ecol Evol 2:1940-1955
- McGowen MR (2011) Toward the resolution of an explosive radiation—A multilocus phylogeny of oceanic dolphins (Delphinidae). Mol Phylogenet Evol 60:345-357
- McGowen MR, Tsagkogeorga G, Álvarez-Carretero S, dos Reis M, Struebig M, Deaville R, Jepson PD, Jarman S, Polanowski A, Morin PA, Rossiter SJ (2020) Phylogenomic resolution of the cetacean tree of life using target sequence capture. Syst Biol 69:479-501
- McGowen MR, Spaulding M, Gatesy J (2009) Divergence date estimation and a comprehensive molecular tree of extant cetaceans. Mol Phylogenet Evol 53:891-906
- Moura AE, Shreves K, Pilot M, Andrews KR, Moore DM, Kishida T, Möller L, Natoli A, Gaspari S, McGowen M, Chen I, Gray H, Gore M, Culloch RM, Kiani MS, Sarrouf Willson M, Bulushi A, Collins T, Baldwin R, Willson A, Minton G, Ponnampalam L, Rus Hoelzel A (2020) Phylogenomics of the genus *Tursiops* and closely related Delphininae reveals extensive reticulation among lineages and provides inference about eco-evolutionary drivers. Mol Phylogenet Evol 146:106756
- Moura AE, Nielsen SCA, Vilstrup JT, Moreno-Mayar JV, Gilbert MTP, Gray HWI, Natoli A, Möller L, Hoelzel AR (2013) Recent diversification of a marine genus (*Tursiops* spp.) tracks habitat preference and environmental change. Syst Biol 62:865-877
- Murakami M, Shimada C, Hikida Y, Soeda Y, Hirano H (2014) *Eodelphis kabatensis*, a new name for the oldest true dolphin *Stenella kabatensis* Horikawa, 1977 (Cetacea, Odontoceti, Delphinidae), from the upper Miocene of Japan, and the phylogeny and paleobiogeography of Delphinoidea. J Vertebr Paleontol 34:491-511
- Nascimento FF, Reis Md, Yang Z (2017) A biologist's guide to Bayesian phylogenetic analysis. Nat Ecol Evol 1:1446-1454
- Nikaïdo M, Hamilton H, Makino H, Sasaki T, Takahashi K, Goto M, Kanda N, Pastene LA, Okada N (2006) Baleen whale phylogeny and a past extensive radiation event revealed by SINE insertion analysis. Mol Biol Evol 23:866-873
- Nikaïdo M, Rooney AP, Okada N (1999) Phylogenetic relationships among cetartiodactyls based on insertions of short and long interspersed elements: hippopotamuses are the closest extant relatives of whales. Proc Natl Acad Sci USA 96:10261-10266
- Nishida S, Goto M, Pastene LA, Kanda N, Koike H (2007) Phylogenetic relationships among cetaceans revealed by Y-chromosome sequences. Zoolog Sci 24:723-732
- O'Leary MA, Gatesy J (2008) Impact of increased character sampling on the phylogeny of Cetartiodactyla (Mammalia): combined analysis including fossils. Cladistics 24:397-442
- Payseur BA, Rieseberg LH (2016) A genomic perspective on hybridization and speciation. Mol Ecol 25:2337-60
- Pease JB, Hahn MW (2015) Detection and polarization of introgression in a five-taxon phylogeny. Syst Biol 64:651-662
- Perrin WF, Mitchell ED, Mead JG, Caldwell DK, Caldwell MC, van Bree PJH, Dawbin WH (1987) Revision of the spotted dolphins, *Stenella* spp. Mar Mammal Sci 3:99-170
- Perrin WF, Rosel PE, Cipriano F (2013) How to contend with paraphyly in the taxonomy of the delphinine cetaceans? Mar Mammal Sci 29:567-588
- Philippe H, Brinkmann H, Lavrov DV, Littlewood DTJ, Manuel M, Wörheide G, Baurain D (2011) Resolving difficult phylogenetic questions: why more sequences are not enough. PLoS Biol 9:e1000602
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945-959
- Racimo F, Sankararaman S, Nielsen R, Huerta-Sánchez E (2015) Evidence for archaic adaptive introgression in humans. Nat Rev Genet 16:359-371
- Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA (2018) Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst Biol 67:901-904
- Reidenberg JS (2007) Anatomical adaptations of aquatic mammals. Anat Rec (Hoboken) 290:507-513
- Roch S, Steel M (2015) Likelihood-based tree reconstruction on a concatenation of aligned sequence data sets can be statistically inconsistent. Theor Popul Biol 100:56-62
- Rogers J, Raveendran M, Harris RA, Mailund T, Leppälä K, Athanasiadis G, Schierup MH, Cheng J, Munch K, Walker JA, Konkel MK, Jordan V, Stealy CJ, Beckstrom TO, Bergey C, Burrell A, Schrepf D, Noll A, Kothe M, Kopp GH, Liu Y, Murali S, Billis K, Martin FJ, Muffato M, Cox L, Else J, Disotell T, Muzny DM, Phillips-Conroy J, Aken B, Eichler EE, Marques-Bonet T, Kosiol C, Batzer MA, Hahn MW, Tung J, Zinner D, Roos C, Jolly CJ, Gibbs RA, Worley KC (2019) The comparative genomics and complex population history of *Papio* baboons. Sci Adv 5:eaau6947
- Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, New York
- Seo T-K (2008) Calculating bootstrap probabilities of phylogeny using multilocus sequence data. Mol Biol Evol 25:960-971
- Shen W, Le S, Li Y, Hu F (2016) SeqKit: a cross-platform and ultrafast toolkit for fasta/q file manipulation. PLoS ONE 11:e0163962

- Shirakihara M, Yoshida H, Shirakihara K (2003) Indo-Pacific bottlenose dolphins *Tursiops aduncus* in Amakusa, western Kyushu, Japan. *Fish Sci* 69:654–656
- Simmons MP, Pickett KM, Miya M (2004) How meaningful are Bayesian support values? *Mol Biol Evol* 21:188–199
- Slater GJ, Price SA, Santini F, Alfaro ME (2010) Diversity versus disparity and the radiation of modern cetaceans. *Proc Biol Sci B* 277:3097–104
- Springer MS, Foley NM, Brady PL, Gatesy J, Murphy WJ (2019) Evolutionary models for the diversification of placental mammals across the KPg boundary. *Front Genet* 10:1241
- Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313
- Steeaman ME, Hebsgaard MB, Fordyce RE, Ho SY, Rabosky DL, Nielsen R, Rahbek C, Glenner H, Sorensen MV, Willerslev E (2009) Radiation of extant cetaceans driven by restructuring of the oceans. *Syst Biol* 58:573–585
- Suh A (2016) The phylogenomic forest of bird trees contains a hard polytomy at the root of Neoaves. *Zool Scr* 45:50–62
- Swofford DL (2003) PAUP*. Phylogenetic analysis using parsimony (*and Other Methods). Version 4. Sinauer Associates Press, Sunderland
- Thewissen JG, Williams EM, Roe LJ, Hussain ST (2001) Skeletons of terrestrial cetaceans and the relationship of whales to artiodactyls. *Nature* 413:277–281
- Thewissen JGM, Cooper LN, George JC, Bajpai S (2009) From land to water: the origin of whales, dolphins, and porpoises. *Evo Edu Outreach* 2:272–288
- Thewissen JGM, Cooper LN, Clementz MT, Bajpai S, Tiwari BN (2007) Whales originated from aquatic artiodactyls in the Eocene epoch of India. *Nature* 450:1190–1194
- Uhen MD (2010) The origin(s) of whales. *Annu Rev Earth Planet Sci* 38:189–219
- Wen D, Yu Y, Zhu J, Nakhleh L (2018) Inferring phylogenetic networks using PhyloNet. *Syst Biol* 67:735–740
- Whitfield JB, Lockhart PJ (2007) Deciphering ancient rapid radiations. *Trends Ecol Evol* 22:258–265
- Wickham, H (2016) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. Retrieved from <https://ggplot2.tidyverse.org>
- Wiens JJ, Engstrom TN, Chippindale PT (2006) Rapid diversification, incomplete isolation, and the “speciation clock” in North American salamanders (genus *Plethodon*): testing the hybrid swarm hypothesis of rapid radiation. *Evolution* 60:2585–2603
- Xiong Y, Brandley MC, Xu S, Zhou K, Yang G (2009) Seven new dolphin mitochondrial genomes and a time-calibrated phylogeny of whales. *BMC Evol Biol* 9:20
- Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24:1586–1591
- Yang Z, Rannala B (2006) Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Mol Biol Evol* 23:212–226
- Zhang C, Rabiee M, Sayyari E, Mirarab S (2018) ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinform* 19:153
- Zhou X, Xu S, Yang Y, Zhou K, Yang G (2011) Phylogenomic analyses and improved resolution of Cetartiodactyla. *Mol Phylogenet Evol* 61:255–264
- Zurano JP, Magalhães FM, Asato AE, Silva G, Bidau CJ, Mesquita DO, Costa GC (2019) Cetartiodactyla: updating a time-calibrated molecular phylogeny. *Mol Phylogenet Evol* 133:256–262