







# Genome-Wide Demographic Analyses of Balaenid Whales Revealed Complex History of Gene Flow Associated with Past Climate Oscillation

Bai-Wei Lo <sup>1,\*</sup>, Francisca Martinez Real <sup>1,2</sup>, Andreas Magg <sup>1,3</sup>, John Pierce Wise Sr. <sup>4</sup>, Stefan Mundlos <sup>1,5</sup>, Paolo Franchini <sup>6,\*</sup>

<sup>1</sup>Research Group of Development and Disease, Max Planck Institute for Molecular Genetics, Berlin, Germany

<sup>2</sup>Andalusian Center for Developmental Biology, CABD (UPO-CSIC-JA), Seville, Spain

<sup>3</sup>BCRT, Berlin Institute of Health (BIH), Charité Universitätsmedizin, Berlin, Germany

<sup>4</sup>Wise Laboratory of Environmental and Genetic Toxicology, Department of Pharmacology and Toxicology, University of Louisville, Louisville, KY, USA

<sup>5</sup>Institute of Medical and Human Genetics, Charité Universitätsmedizin, Berlin, Germany

<sup>6</sup>Department of Ecological and Biological Sciences, University of Tuscia, Viale dell'Università s.n.c, Viterbo, Italy

\*Corresponding authors: E-mails: lo@molgen.mpg.de, paolo.franchini@unitus.it

Accepted: April 29, 2025

## Abstract

The balaenid whale, comprising three species of right whales and the bowhead whale, represents an ancient and highly endangered lineage of marine mammals. To unravel the evolutionary history of balaenid whales with respect to gene flow, a comprehensive analysis based on whole-genome data was conducted for all species within this group. Employing population genomic methodologies, we revealed that extant right whales form an unresolved branching pattern, identified evidence of historical transequatorial migration, and provided estimates of the age of the group. Furthermore, we investigated the impact of glacial cycles on the connectivity of bowhead whale populations. By employing multiple complementary approaches to detect gene flow, we identified and characterized gene flow events from bowhead whales to North Atlantic right whales, offering detailed insights into the process. Lastly, we assessed the phenotypic consequences of interspecies gene flow. Our study sheds light on the intricate evolutionary history of modern balaenid whales, which have been profoundly shaped by ancient climate events.

**Key words:** gene flow, introgression, Balaenidae, right whale, bowhead whale, cetaceans.

## Significance

While interspecific gene flow is known to be widespread among animals, including cetaceans, the driver, and impact of these genetic exchanges, especially within the ancient and endangered balaenid whales (bowhead whale and the right whales), remain poorly understood. This study reveals significant waves of gene flow among major lineages in the balaenid whales. These findings enhance our understanding of the evolutionary dynamics and historical gene flow in the balaenid whales, highlighting the profound effects of ancient climate events on their genetic diversity.

## Introduction

Interspecific introgression, the incorporation of genetic material from one species into the gene pool of another through repeated backcrossing (Rieseberg and Carney 1998), has long been considered infrequent. However, it is now recognized as widespread among diverse taxonomic groups (Edelman and Mallet 2021; Dagilis et al. 2022). The increased availability of genomic data from nonmodel organisms, coupled with recent advancements in analytical methodologies, has not only facilitated the identification of interspecific introgression across different taxa but also significantly enhanced our ability to characterize the introgressed genetic material. This, in turn, allows for a more comprehensive assessment of the evolutionary consequences associated with this process (Hibbins and Hahn 2022). Nevertheless, the evolutionary significance of interspecific introgression and its impact on local adaptation and speciation remain active research areas (Edelman and Mallet 2021; Fontaine 2022).

In cetaceans, interspecific introgression has been documented in both baleen and toothed whales, with several studies highlighting ancient introgression events during the early stages of diversification (Árnason et al. 2018; Moura et al. 2020; Guo et al. 2021; Westbury et al. 2023). These introgression events are often the results of extensive gene flow, defined here as movement of genetic material between populations of the same or different species. This process in cetaceans was likely facilitated by the absence of significant extrinsic barriers in the marine environment. However, evaluating the driving forces and evolutionary implications of these ancient events remains challenging. Most studies have focused on historical introgression without fully addressing how contemporary processes, such as ongoing gene flow or adaptation to climate change, contribute to the evolutionary history of cetaceans.

The family Balaenidae presents a unique opportunity to further investigate interspecific introgression while addressing existing challenges. Balaenidae is the sibling group to Plicogulae, which includes rorquals, the gray whale, and the pygmy right whale. Together, these groups comprise modern baleen whales. There are only four extant balaenid species: the bowhead whale (*Balaena mysticetus*) and three species of right whales—the North Pacific right whale (*Eubalaena japonica*), the southern right whale (*Eubalaena australis*) and the North Atlantic right whale (*Eubalaena glacialis*), which is currently one of the most endangered large whale species (IUCN, 2023). Balaenid whales inhabit temperate (right whales) to Arctic (bowhead whale) waters in both hemispheres. The recent divergence of modern balaenid whale species provides a unique opportunity to explore the gene flow that occurred around the time of speciation. Furthermore, the cold-adapted nature

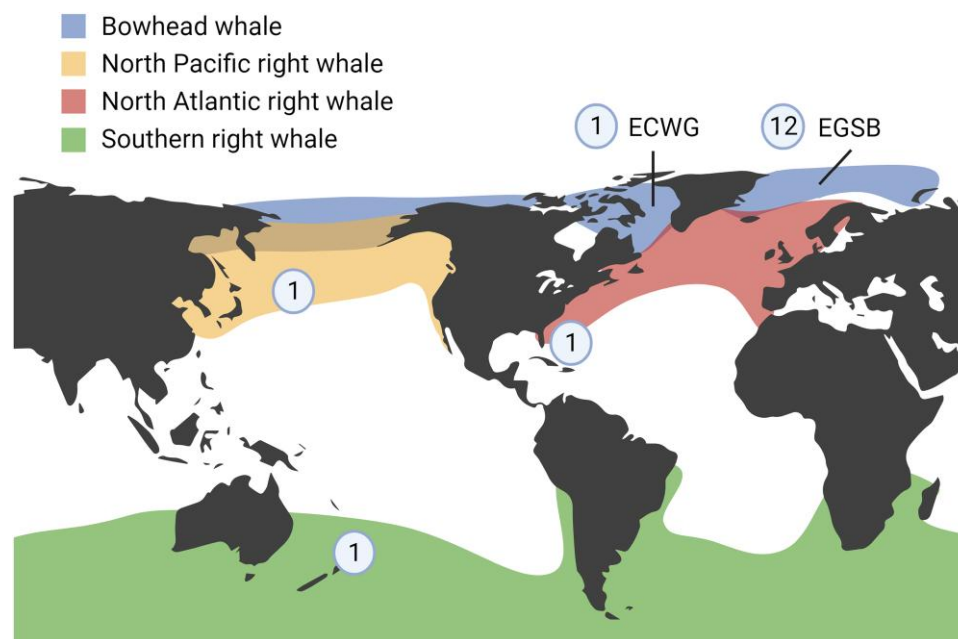
of modern balaenid whales has likely made them more sensitive to global climate events than other cetaceans, enabling a more thorough assessment of the evolutionary significance of introgression.

In this study, we present the first genome-wide demographic analysis of all four balaenid whale species. By employing complementary genomic approaches, we reconstructed a detailed picture of their divergence and gene flow dynamics. We shed light on the phylogenetic relationships among modern right whales, estimated the age of each taxonomic group, described different scenarios of intra and interspecific gene flow, and inferred the phenotypic consequences of interspecific gene flow. Our findings provide novel insights into the evolutionary history of this ancient surviving whale lineage and highlight the role of ancient climate events in shaping patterns of introgression. This research enhances our understanding of the complex dynamics of gene flow and divergence in this group of marine mammals.

## Results

### Phylogenomic Analyses Reveal Conflicting Phylogenetic Signals in Modern-Day Right Whales

To understand the evolutionary relationships of balaenid whales, we collected cetacean genomic data from public databases (Fig. 1; supplementary table S1, Supplementary Material online). We performed phylogenomic analyses on the four focal balaenid species while using three Plicogulae species as outgroups: minke whale (*Balaenoptera acutorostrata*), blue whale (*Balaenoptera musculus*), and pygmy right whale (*Caperea marginata*). We estimated species trees independently using three multispecies coalescent model-based programs. In brief, multispecies coalescent models assume there exists a true dichotomous species tree and infer this species tree by modeling the process of incomplete lineage sorting. First, we used ASTRAL, a summary statistic-based model that infers a species tree from a collection of subtrees (Zhang et al. 2018). We cut the genomes into 28,586 nonoverlapping fragments that were 50 kbp in length and used RAXML to estimate the topology of each subtree. We forwarded the topologies of the subtrees to ASTRAL to estimate the species tree. The result of ASTRAL showed a sibling relationship between North Pacific right whale and southern right whale, with all branches having 100% posterior probability support (Fig. 2). Second, we applied the multispecies coalescent-based models (MCMCs) implemented in SVDQuartets and SNAPP, both of which infer species trees from genome-wide single nucleotide variants (SNVs). Interestingly, the results of SNAPP agreed with those of ASTRAL, whereas SVDQuartets inferred a completely different relationship among the right whales. The tree inferred



**Fig. 1.** Balaenid whale samples used in the study. Geographical distribution of the four balaenid whale species: the bowhead whale (*B. mysticetus*), the North Pacific right whale (*E. japonica*), the southern right whale (*E. australis*), and the North Atlantic right whale (*E. glacialis*). Numbers indicate the sample size for each species, including individuals from the ECWG and EGSB bowhead whale populations. The distribution data were adapted from Tsai and Chang (2019) and George and Thewissen (2020). Created with BioRender.com.

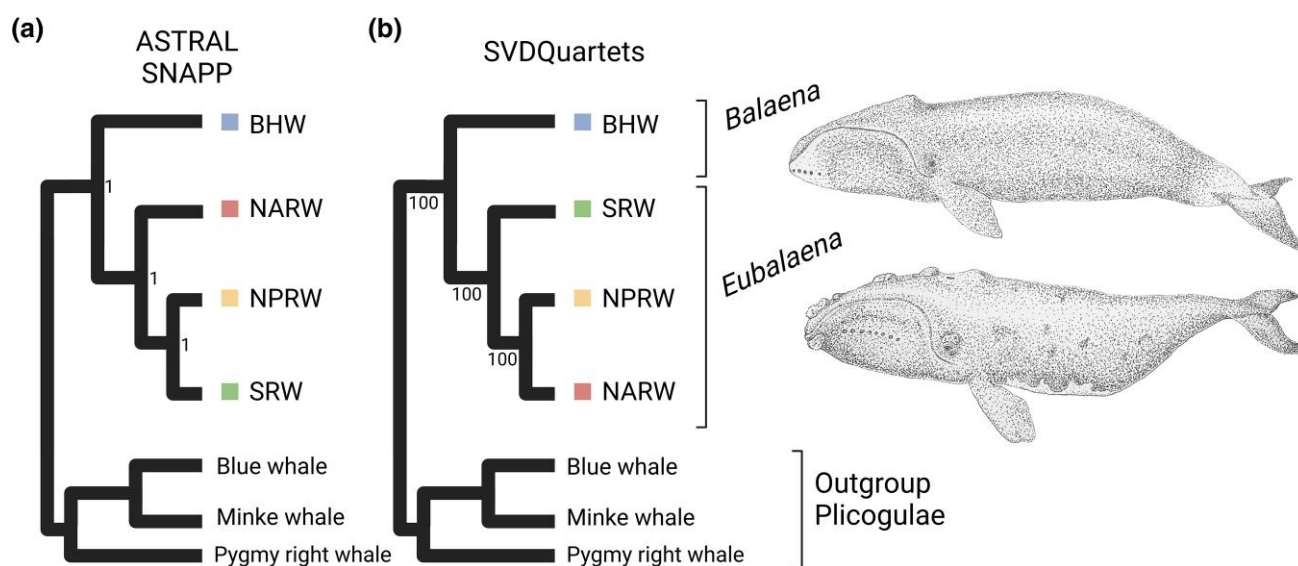
by SVDQuartets revealed a monophyletic northern group comprising the two northern right whale species, with each node supported by a 100% bootstrap value (Fig. 2). The inconsistency in reconstructing species trees, even when using whole-genome data, underscores the challenge of accurately depicting the evolutionary history of balaenid whales through traditional dichotomous phylogeny.

#### Phylogenetic Network and *D*-statistics Show Interspecies Gene Flow among Balaenid Whales and between Major Baleen Whale Lineages

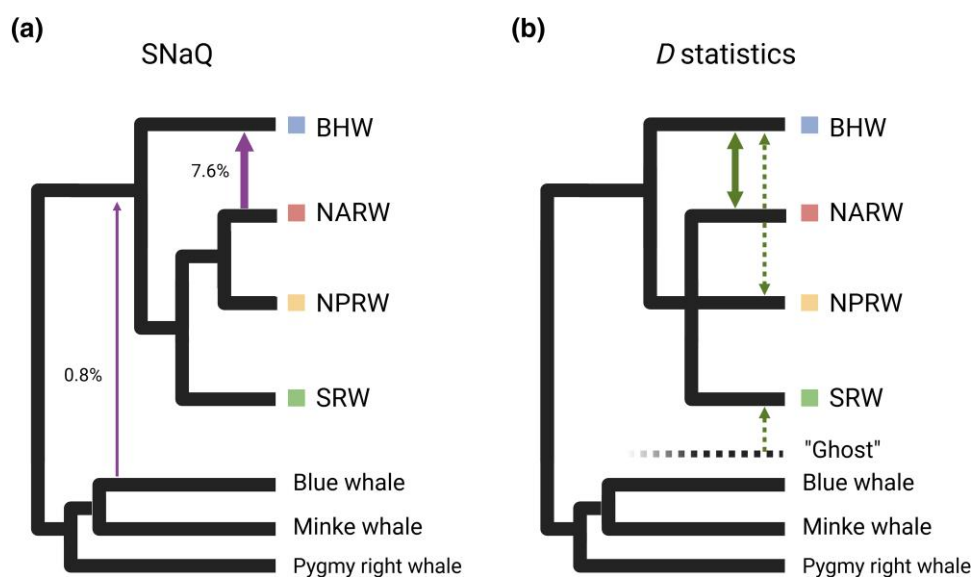
One potential explanation for the discordant topologies obtained with different phylogenomic methods used to resolve the relationship of balaenid whales is the presence of interspecific gene flow. To take gene flow into account, we estimated phylogenetic networks from genome-wide SNVs, again using the three Plicogulæ species as outgroups. Phylogenetic networks have the advantage of simultaneously estimating species tree and gene flow events. To this end, we used the maximum pseudolikelihood algorithm implemented in SNaQ. Because SNaQ can only infer level one networks in order to avoid overfitting, the maximum number of gene flow events that can be predicted in our seven species setting is two. Like SVDQuartets, SNaQ also inferred a sibling relationship of two northern right whales, with 100% bootstrap support for this clade. The outputs of SNaQ are unrooted networks and have no

branch lengths, thus allowing gene flow estimation between Balaenidae and Plicogulæ. The best scored phylogenetic network contains two waves of gene flow: (i) from the North Atlantic right whale to the bowhead whale (7.6% introgression) and (ii) from the blue whale to the common ancestors of balaenid whales (0.8% introgression) (Fig. 3a). The inferred network is able to better explain gene tree discordance than the dichotomous tree (supplementary fig. S1, Supplementary Material online). The result from the phylogenetic network indicates the presence of interspecific gene flow.

To systematically survey for signals of gene flow, we applied genome-wide *D*-statistics in selected trios of balaenid species with *Dsuite*. *D*-statistics is a nonparametric method for detecting gene flow (Patterson et al. 2012). It takes three taxa and an outgroup and tests for excessive allele sharing between the nonsister taxa. Excessive allele sharing between taxa can indicate gene flow. Since the *D*-statistics require a known dichotomous relationship of the tested taxa, we therefore only applied it on the bowhead whale plus a pair of right whales, using the three Plicogulæ species as outgroups. We found significant results in all three tested trios (Table 1). Excessive allele sharing was found between the bowhead whale and North Atlantic right whale and between the bowhead whale and North Pacific right whale. The level of sharing was most prominent between the bowhead whale and North Atlantic right whale (Table 1). This high level of allele sharing between the



**Fig. 2.** Phylogenetic relationships of balaenid whales inferred from whole genomic data. a) As inferred from 50 kbp genomic windows, ASTRAL and SNAPP results show that the North Atlantic right whale is the sister group to the clade that includes the North Pacific and the southern right whales within the *Eubalaena* genus. b) SVDQuartets results, inferred from genome-wide SNVs, suggest a different relationship, where the southern right whale is the sister group to the other two *Eubalaena* species, which form a monophyletic northern group. All branches were supported by 100% posterior probabilities (ASTRAL and SNAPP) or 100% bootstrap values (SVDQuartets). The discrepancy of the results highlights the conflicting phylogenetic signals in the evolutionary history of right whales. Balaenid whales illustrated by B.-V.L. Created with [BioRender.com](https://BioRender.com).



**Fig. 3.** Interspecific introgression inferred from SNaQ and *D*-statistics. a) Phylogenetic network analysis carried out with SNaQ revealed two waves of gene flow (arrows), with introgression levels of 7.6% from the North Atlantic right whale to the bowhead whale and 0.8% from the blue whale to the common ancestors of balaenid whales. b) *D*-statistics again discovered introgression between the bowhead whale and North Atlantic right whale (solid arrow), with potential gene flow events (dashed arrows) explaining excessive allele sharing between the bowhead whale and North Pacific right whale. These events could involve direct gene flow or gene flow from a "ghost" lineage, whose position ranges from being the sister group to the balaenid clade to being within the outgroups. Both analyses support interspecific gene flow among balaenid species. Created with [BioRender.com](https://BioRender.com).

bowhead whale and North Atlantic right whale strongly supports gene flow between the two (Fig. 3b). In contrast, the lesser signal of excessive allele sharing between the bowhead whale and North Pacific right whale can be

explained by direct gene flow between bowhead whale and North Pacific right whale, or nonmutually exclusively by gene flow from an unsampled lineage to southern right whale (Fig. 3b). To explore whether the sex chromosome is

**Table 1** Results of *D*-statistics

P1	P2	P3	D	P value	Z score
NPRW	NARW	BHW	0.113	<0.001	49.3
SRW	NARW	BHW	0.275	<0.001	89.3
SRW	NPRW	BHW	0.173	<0.001	75.1
EGSB BHW	ECWG BHW	NARW	0.297	<0.001	64.8
PRW	BW	SRW	0.03	<0.001	22.9

Note that the tree provided to *D*-statistics is (((P1, P2), P3), Outgroup). NPRW: North Pacific right whale; NARW: North Atlantic right whale; BHW: bowhead whale; SRW: southern right whale; BW: blue whale; PRW: pygmy right whale.

also affected by introgression, we estimated the *D*-statistics on the X chromosome. Surprisingly, the X chromosome harbors more introgression in the bowhead whale—North Atlantic right whale introgression than autosomes (*D* value 0.154 vs. 0.114).

To provide further understanding on the blue whale—balaenid whale introgression inferred from phylogenetic network, we applied *D*-statistics on southern right whale, pygmy right whale, and blue whale, using sperm whale (*Physeter macrocephalus*) as an outgroup. We detected significant excessive allele sharing between southern right whale and blue whale (Table 1), which supports the result of phylogenetic network. In summary, our genome-wide surveys from both phylogenetic network and the *D*-statistic painted a similar picture of interspecific gene flow among balaenid whales. Both analyses support (i) North Atlantic right whale—bowhead whale and (ii) blue whale—balaenid whale introgressions. *D*-statistics also suggests the potential occurrence of gene flow between North Pacific right whale and bowhead whale and/or southern right whale and an unsampled lineage (ghost).

### Haplotype-Based Method MSMC-IM Confirms the Increased Connectivity of Atlantic Bowhead Whale Populations during Glacial Periods

We discovered signals of introgression among balaenid whales, which contradicts basic assumptions of traditional phylogenetic analyses. Still, the evolutionary history of modern right whales remains unresolved. Further details regarding the introgression events also need to be elaborated. These details include timing and direction of gene flow and whether the gene flow came from secondary contact or isolation with migration. In order to address these questions, we turned to MSMC-IM, a haplotype-based population genomic approach for demographic analysis. MSMC-IM infers a time-dependent gene flow model between pairs of taxa. MSMC-IM can also detect introgression signals from unsampled divergent lineages, also known as archaic introgression. To demonstrate the power of this haplotype-based approach on inferring migration history, we first used MSMC-IM to reconstruct the intraspecific migration within bowhead whales in the North Atlantic Ocean. The East Greenland-Svalbard-Barents Sea (EGSB)

and East Canada-West Greenland (ECWG) populations are currently predominantly separated by Greenland (Fig. 1), but were predicted to come into contact when they were forced south during the Last Glacial Maximum (Foote et al. 2013). Indeed, we detected periodic changes in migration rate between the two populations using MSMC-IM (Fig. 4a). Previous estimations for the generation time of bowhead whales varied quite differently (35 vs. 52 years) (Rooney et al. 2001; Taylor et al. 2007), and we therefore use the glacial cycles and MSMC-IM result to evaluate which of these two estimates is more accurate. The rises in migration rate aligned almost perfectly with Pleistocene glacial periods when the generation time is set to 35 (Fig. 4a; [supplementary fig. S2, Supplementary Material online](#)). The divergence time estimated for the two populations is 139 Kya (Fig. 4h; [supplementary fig. S3, Supplementary Material online](#)). Our finding provides direct molecular evidence on the influence of glacial cycles on migration in bowhead whales, while also bringing insights into the life history of this species.

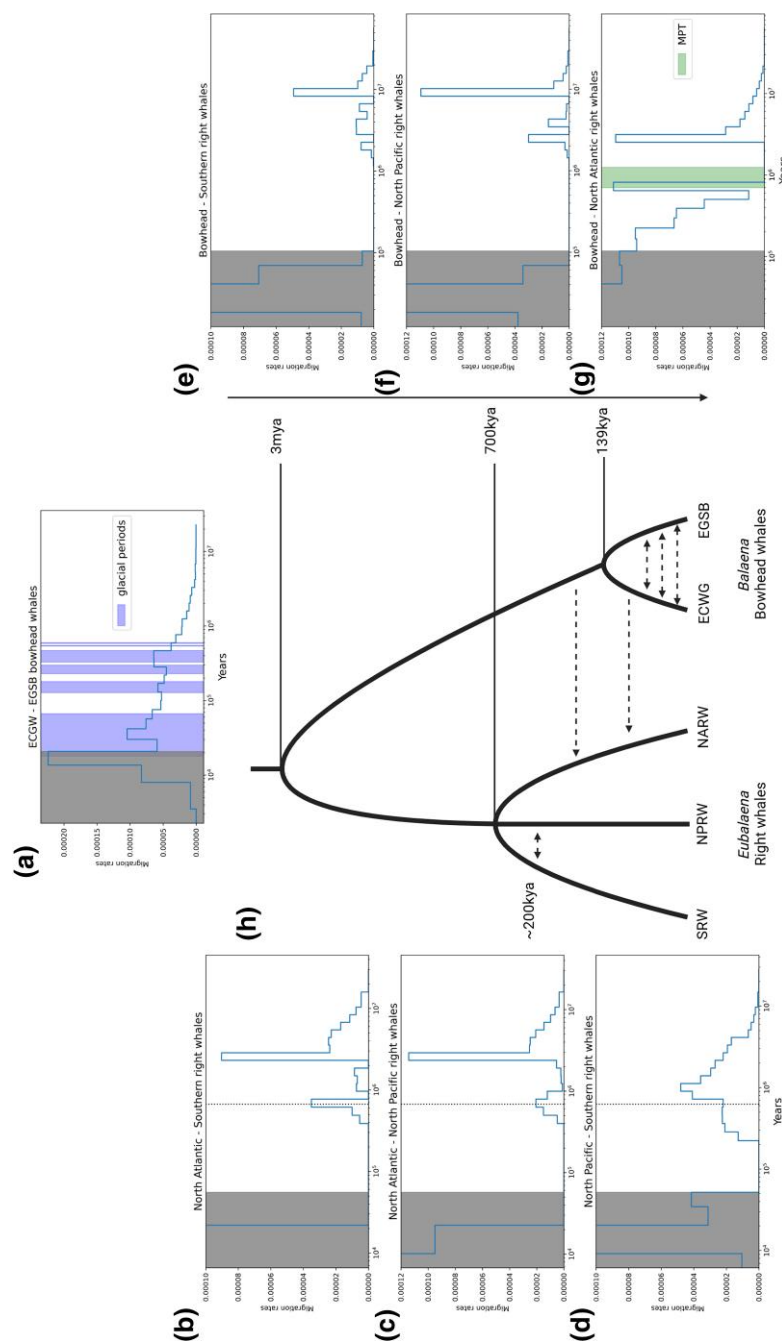
### MSMC-IM Reveals Overlapping Divergence Times among Extant Right Whales and Evidence of Transequatorial Migration through the Pacific Ocean

To understand the detailed divergence history of extant right whales, we applied MSMC-IM on pairs of right whale species. To estimate the divergence time of each species pair, we calculated the 50% cumulative migration rate, a way of representing divergence time between populations. Interestingly, we found that the 50% cumulative migration rates in each species pair are nearly identical (around 0.7 million years ago) with overlapping 25% to 75% cumulative migration intervals (Fig. 4b to d and h; [supplementary fig. S4, Supplementary Material online](#)). Despite the close divergence time of all species pairs, we found continuous gene flow between the North Pacific right whale and southern right whale that lasted until approximately 200,000 years ago, when gene flow in other species pairs already long ceased (Fig. 4d). The difference in gene flow pattern indicates long-lasting isolation with migration between southern right whales and North Pacific right whales, providing solid evidence that late transequatorial migration happened in the Pacific Ocean. Our results reveal a complex divergence pattern of right whales that cannot be represented by a simple dichotomous phylogeny and confirms that the Pacific Ocean was a passage for transequatorial migration.

### MSMC-IM and *D*-statistics Reveal Evidence of Gene flow From the ECWG Bowhead Whale to North Atlantic Right Whale

We detected signals of archaic introgression in MSMC-IM analyses of right whale species pairs (Fig. 4b and c).





**Fig. 4.** Intra- and interspecific migration profiles estimated from MSMC-IM. a) Increases in migration rate between two populations of bowhead whales (ECGW and EGSB) align with Pleistocene glacial periods (highlighted bars), consistent with predicted southward movement during the Last Glacial Maximum. b and c) Signals of archaic introgression are evident in species pairs involving the North Atlantic right whale, with peaks occurring approximately 2 to 3 million years ago. These introgression peaks suggest contributions from an unsampled divergent lineage predating the main divergence. Dashed lines represent the 50% cumulative migration rate, indicating the point of major divergence between species pairs. d) Continuous gene flow between the North Pacific right whale and southern right whale persisted until approximately 200,000 years ago, revealing evidence of late transequatorial migration through the Pacific Ocean. e and f) Minimal postdivergence gene flow is detected between bowhead whales and either the North Pacific or southern right whales. g) Introgression from the ECGW bowhead whale population to the North Atlantic right whale began during the mid-Pleistocene transition (highlighted area) and persisted until the recent past. Gray-shaded areas indicate manually masked regions with potential phasing artifacts. h) Phylogenetic representation showing divergence time inferred from pairwise 50% cumulative migration rate. Arrows represent introgressions and their directions inferred with pairwise MSMC-IM. Created with [BioRender.com](https://BioRender.com).

Archaic introgression peaks, peaks that predate the main divergence peak, are only present in species pairs that include the North Atlantic right whale, showing that this species is the one that carries genomic material from a divergent species. The archaic lineage that contributed to the introgression diverged from right whales at least 2 to 3 million years ago, based on the position of the peak. We then applied MSMC-IM on the bowhead whale and each of the three right whales. We confirmed that the archaic introgression signal within the North Atlantic right whale indeed came from secondary contact with the bowhead whale (Fig. 4g). This secondary contact started during the mid-Pleistocene transition (Clark et al. 2006) and persisted to the recent past until it mingled with phasing error artifacts. This gene flow is most likely unidirectional since we do not detect any archaic introgression signal in bowhead whales (Fig. 4a). We also used cumulative migration rate to estimate the divergence time between the two genera. Consistent with the position of archaic introgression peak in right whales, the divergence time between right whales and bowhead whale is around 3 million years ago (Fig. 4h; supplementary fig. S5, Supplementary Material online).

To test whether the population structure of bowhead whales affects the introgression with the North Atlantic right whale, we applied the *D*-statistic on the North Atlantic right whale, the ECWG, and the EGSB population of bowhead whales. Our results reveal that the ECWG population, which has a geographical distribution that overlaps with the North Atlantic right whales in the North West Atlantic Ocean, from which the North Atlantic right whale individual was sampled, has strong and significant signals of excessive allele sharing (Table 1). Therefore, we confirmed that the ECWG bowhead whale is the major source of introgression. MSMC-IM does not provide additional insight regarding the excessive allele sharing between bowhead whale and North Pacific right whale. Unlike in the case of bowhead—North Atlantic right whale introgression, we found minimal postdivergence gene flow between the bowhead whale and North Pacific right whale (Fig. 4e). We also did not detect signals from ghost introgression in southern right whales (Fig. 4d). Our analyses provide evidence of unidirectional gene flow from the ECWG bowhead whale to North Atlantic right whale, aligning with the mid-Pleistocene transition. This new information adds more details to the previously predicted introgression.

#### Sliding Window Statistics and Enrichment Analyses Show that Interspecific Introgression May Have Phenotypic Effects

To identify putative introgression regions in the North Atlantic right whale, we calculated genome-wide sliding

**Table 2** GO enrichment analysis of introgression regions in North Atlantic right whale

GO term ID	GO term name	Adjusted <i>P</i> value
0004984	Olfactory receptor activity	$1.9 \times 10^{-6}$
0061778	Intracellular chloride channel activity	$1.6 \times 10^{-4}$
0005229	Intracellular calcium activated chloride channel activity	$1.5 \times 10^{-3}$

window *fdm* statistics. To evaluate the biological and phenotypic consequences of introgression, we extracted genes within regions with high levels of introgression, and performed Gene Ontology (GO) enrichment analysis. Regions with high levels of introgression are defined as sliding windows with a top 5% *fdm* value. The regions with high levels of introgression from the bowhead whale to the North Atlantic right whale are enriched for the GO terms “olfactory receptor activity” and “intracellular chloride channel activity” (Table 2). Our results show the potential phenotypic effect of interspecific introgression in balaenid whale species.

## Discussion

We reconstructed a detailed history of divergence and gene flow within the baleen whale family Balaenidae encompassing all four extant species. Most phylogenetic studies supported either one of the two competing hypotheses regarding the relationship of modern right whales (*Eubalaena* spp.). The first supports a sibling relationship between North Pacific and southern right whale (Rosenbaum et al. 2000; Gaines et al. 2005; Kaliszewska et al. 2005; Churchill et al. 2011; Buono et al. 2017), while the second suggests a sibling relationship between two northern hemisphere species (McGowen et al. 2009; Steeman et al. 2009; McGowen et al. 2020; Wolf et al. 2023). Some studies, however, have reported alternative findings (Bisconti et al. 2017), including instances of reticulated topology (Wolf et al. 2023). In our study, despite employing whole genomic data, we initially encountered challenges in determining the best dichotomous phylogenetic relationship among the three right whales. We successfully addressed this challenge, however, by treating the right whales as populations and leveraging haplotype information. We demonstrated that the differing topologies among these species may result from lasting gene flow between North Pacific and southern right whales. Therefore, we can confirm that the Pacific Ocean served as a passage for late transequatorial migration, which agrees with both fossil evidence from the Taiwan Strait (Tsai and Chang 2019) and divergence estimation from right whales in the Atlantic Ocean (Crossman et al. 2023). The differences in divergence pattern and introgression with archaic species are likely the factors that have hindered past studies from

reaching a conclusion on the two competing topologies, as both conditions violate the basic assumption of most phylogenetic models. Indeed, different algorithms capture different aspects of the data. For instance, the multispecies coalescent models implemented in ASTRAL and SNAPP cluster North Pacific and southern right whales together, possibly because of later gene flow resulting from isolation with migration. In contrast, SVDQuartets, known to be less affected by gene flow from isolation with migration (Long and Kubatko 2018), reported a monophyletic northern group.

We estimated the age of extant *Eubalaena* (0.7 M) and Balaenidae (3 M) from relative cross-coalescence rates between pairs of haplotypes from two populations using the model implemented in MSMC-IM. For this analysis, we used bowhead whale (*B. mysticetus*) generation time recalibrated on late Pleistocene glacial cycles. We first show that despite having a remarkably long lifespan, the generation time of bowhead whales (35 years) is likely similar to that of modern right whales (28 to 35 years) (Taylor et al. 2007). The accuracy of this molecular dating method relies on the assumption that the generation time remains constant throughout evolution. However, fossil evidence supports that gigantism evolved independently in the two extant balaenid genera (Bisconti et al. 2021). As generation time is positively correlated with body size, the common ancestors of extant balaenid whales most likely have shorter generation times. Hence, while the age estimation of *Eubalaena* is not affected since giant body size is a synapomorphy of the extant species, the age of Balaenidae is almost certainly an overestimation. Previous molecular clock-based studies varied in estimating ages of extant *Eubalaena* (0.77 to 4.4 M) and Balaenidae (4.38 to 17.1 M) (Sasaki et al. 2005; McGowen et al. 2009; Slater et al. 2010; Hassanin et al. 2012; Slater et al. 2017; Árnason et al. 2018; McGowen et al. 2020; Wolf et al. 2023). Our results support those estimations that are at the lower end of the spectrum.

The distribution of both extant and extinct species of *Balaena* and *Eubalaena* suggests that the two modern genera likely originated in the Northern Hemisphere (Rooney et al. 2001; Bisconti 2003; Kimura 2009; Churchill et al. 2011). Therefore, the early ancestors of right whales would have needed to cross the equator to reach the Southern Ocean. It has been proposed that climate oscillation during the Pleistocene may have contributed to the antitropical distribution of modern right whales (Davies 1963). Even though our estimated divergence time of *Eubalaena* falls within the Pleistocene, we noticed a deep population structure from the MSMC-IM result. This supports the notion that isolation with migration between southern and North Pacific right whales likely started before the Pleistocene epoch. Consequently, the direct influence of climate oscillations or other geological events on the

distribution of *Eubalaena* remains unclear. While there is no evidence of transequatorial movement in modern balaenid whales, rare occurrences of both fossil and stranded pygmy right whales north of the equator have been documented (Tsai et al. 2017; Tsai and Mead 2018). Pygmy right whale is typically distributed in the temperate waters of the Southern Hemisphere. The occurrence of a specialized temperate cetacean on the other side of the equator offers a glimpse into the possible evolutionary history of right whales. It is likely that similar rare movements, facilitated by periods of cold climate, established the transequatorial colonies and provided subsequent gene flow. Over time, genetic differentiation driven by geographical isolation gradually reduced the amount of effective gene flow, ultimately resulting in the three populations/species we see today (van Weelden et al. 2021).

Hybridization in cetaceans has been extensively documented (Miralles et al. 2016; Skovrind et al. 2019; Jossey et al. 2024), yet the extent to which hybrids contribute to gene flow remains debated. Ancient introgressions, which occur at the bases of adaptive radiations, are prevalent across cetacean groups (Árnason et al. 2018; Foote et al. 2019; Moura et al. 2020; Westbury et al. 2023), while more recent interspecific introgressions have also been identified (Westbury et al. 2019; Guo et al. 2021; Jossey et al. 2024). Due to the recent divergence of balaenid whales, we were able to detect not only ancient introgression between balaenid and balaenopterid whales, but also recent introgressions within Balaenidae. We used a combination of likelihood-based, nonparametric, and haplotype-based approaches to detect gene flow among balaenid whales. We characterized both gene flow between sibling taxa (ECWG and EGSB bowhead whale, North Pacific and southern right whale), between more distantly related lineages (North Atlantic right whale and bowhead whale), and ancient gene flow between baleen whale families. Historical shifting of distribution in Atlantic bowhead whales due to glacial cycles is an intuitive hypothesis, but can be hard to verify (Foote et al. 2013). MSMC-IM analysis of the ECWG and EGSB populations of bowhead whales was able to provide direct evidence to the hypothesis, showing how informative and powerful a haplotype-based method is. In terms of interspecific introgression, MSMC-IM was able to identify archaic introgression peaks within the North Atlantic right whale, while also providing details on the timing of the event. On the other hand, no archaic introgression peak was found within the bowhead whales, showing that the interspecific introgression between the two species came from the bowhead whale to the North Atlantic right whale. This unidirectional gene flow mirrors a recent report of fin whale to blue whale introgression (Jossey et al. 2024). The biological implication of such phenomenon in cetaceans warrants further research. One shortcoming of MSMC-IM, like all sequential



Markovian chain-based methods, is the limited ability to explore the recent past due to phasing error (Wang et al. 2020). In a nonmodel species where the reference panel is unavailable, the only way to overcome the phasing problem is to use accurate long read sequencing data such as PacBio HiFi reads. MSMC-IM also failed to detect introgression signals between North Pacific right whale and bowhead whale, while such signals are strong in *D*-statistics. A potential explanation is that the excessive allele sharing did not originate directly from the introgression between North Pacific right whale and bowhead whale. Instead, it might have resulted from the introgression between southern right whale and a more distantly related unsampled lineage. The deep divergence between the involved lineages, and the potentially ancient origin of the introgression, likely hindered MSMC-IM from detecting the “ghost” introgression in southern right whale.

A limitation of this study is the sample size. Except for the bowhead whale, all conclusions were based on a single diploid individual representing the entire species. However, instead of identifying signatures of recent hybridization, we detected effective gene flow that has been broken down by recombination and has persisted in the population for numerous generations. We argue, therefore, that two haplotypes are sufficient to reliably infer introgression at the population level. Nonetheless, we cannot determine whether the level of introgression varies across different populations within a species, as observed in bowhead whales. Population structure associated with loyalty to breeding waters is known to exist in right whales (Carroll et al. 2020). Additionally, variation of introgression linked to deep population structure was reported in other cetaceans, including orcas, blue whales, and fin whales (Foote et al. 2019; Furni et al. 2024; Jossey et al. 2024). Therefore, to gain a more comprehensive understanding of introgression at a species level, it is essential to include broader sampling from different geographical regions, while testing for hidden population structure for sympatric samples.

Drawing definitive conclusions about adaptive introgression based on only two haplotypes remains challenging. However, regions with high levels of introgression may harbor alleles maintained by natural selection. Adaptive introgression, the process by which gene flow introduces beneficial alleles that enhance fitness in the recipient population, is a plausible explanation in this context. For instance, gene flow from bowhead whales (an Arctic species) to North Atlantic right whales (a temperate species) occurred during the mid-Pleistocene transition, a period of global cooling. It is plausible to speculate that this introgression provided adaptive advantages, potentially facilitating ecological or physiological adjustments to changing environmental conditions. Adaptive introgression has been increasingly documented across a wide range of taxa. The

most iconic examples come from our own species, where ancient hybridization events with archaic hominins, such as Neanderthals and Denisovans, have left traces of functional genetic variation (Huerta-Sánchez et al. 2014; Dannemann et al. 2016). Beyond humans, adaptive introgression has been recently reported in several plant and animal lineages, where the genomic basis and molecular mechanisms underlying the associated phenotypes have been largely elucidated (Richards and Martin 2017; Jones et al. 2018; Eberlein et al. 2019).

In our study, GO enrichment analysis suggests that introgression might have influenced olfaction-related genes. Balaenid whales rely on olfaction to sense their prey (Kishida et al. 2015; George and Thewissen 2020), and it is possible that introgressed alleles affect prey identification or prey specificity. However, direct functional validation of these candidate genes is necessary to establish a causal link between introgression and phenotypic adaptation. Testing this hypothesis will require integrating genomic data with ecological observations. Additionally, further exploration of why the X chromosome exhibits an even higher level of introgression than autosomes—a rare phenomenon in mammals (Fraisé and Sachdeva 2021)—will provide insights into the selection mechanisms and gene flow in these species. Future studies with larger sample sizes, incorporating population-level variation, and combining genomic, ecological, and behavioral data will be critical to clarifying the role of selection in shaping introgression between ecologically distinct species.

In conclusion, our work contributes to enhancing our understanding of the evolution of this ancient and endangered whale family. The findings of this research also have broader implications for related fields in biology, for example, research on aging. As the longest-living mammal, the bowhead whale has been a focus of aging-related studies. Our data underscore the importance of considering interspecific introgression in comparative genomic studies involving bowhead whale and right whales. Furthermore, we discovered that certain gene flow events can be associated with ancient global climate events such as mid-Pleistocene transition and Pleistocene glacial cycles, highlighting how climate changes can profoundly impact cold-adapted species.

## Methods

### Data Collection

We downloaded the reference genomes of the southern right whale, bowhead whale (Keane et al. 2015), blue whale (Bukhman et al. 2024), and pygmy right whale (Wolf et al. 2023). We downloaded whole-genome sequencing short reads from six baleen whale species, including all four extant balaenid and two balaenopterid species,

the latter served as outgroups. The balaenid species are North Atlantic right whale, North Pacific right whale, southern right whale, and a bowhead whale from the ECWG population. In addition, we downloaded 12 resequenced bowhead whale samples from the EGSB population (Cerca et al. 2022). The two balaenopterid species were blue whale and minke whale. Detailed information of the samples can be seen in [supplementary table S1, Supplementary Material](#) online.

### SNV Calling

Because raw sequencing reads of the pygmy right whale were not available during the time of analysis, we used wgsim v0.3.1-r13 (<https://github.com/lh3/wgsim>) to simulate 400 million (M) paired short reads from its genome. We aligned short reads of the seven baleen whales to the blue whale genome with bwa mem v0.7.17-r1188 (Li and Durbin 2009) and sorted them with samtools v1.18 (Li et al. 2009). We choose a nonbalaenid species as reference to minimize ascertainment bias. We sampled the mapping files in bam format with samtools so that all balaenid samples matched the sequencing depth of the southern right whale, the species with the lowest sequencing depth (19X). Bcftools v1.18 (Narasimhan et al. 2016) was used to call single nucleotide variations (SNVs) from the aligned sequence files in bam files using default settings. The SNVs were filtered to exclude sites with quality score lower than 20 or a read depth lower than five.

### Phylogenomic Analysis

To estimate the phylogeny of balaenid whales, we used three MCMCs implemented in the programs ASTRAL III v5.7.8 (Zhang et al. 2018), SVDQuartets (Chifman and Kubatko 2014), and SNAPP v1.6.1 (Bryant et al. 2012). The analysis was based on one individual per species and including the ECWG bowhead whale. For the haplotype-based phylogenetic analysis, we first generated consensus autosome genomes for all seven species from the bam files using ANGSD v0.941 (Korneliussen et al. 2014). Second, we used bedtools v2.30.0 (Quinlan 2014) to generate 50 kbp nonoverlapping genomic fragments for each consensus genome. We then apply pargenes v1.2.0 (Morel et al. 2019) on the genomic fragments, which relies on RAXML-NG v1.2.2 (Stamatakis 2014) to estimate individual subtrees from each fragment. Finally, we used ASTRAL III to estimate an overall species tree, where posterior probability was utilized to evaluate the uncertainty in the ASTRAL results. For both SNV-based phylogenetic analyses, we selected biallelic autosomal SNVs and filtered them with vcftools v0.1.17 (Danecek et al. 2011) so that each site was at least 100 bp apart from each other. For SVDQuartets, the Ruby script *convert\_vcf\_to\_nexus.rb* was used to create the input nexus file for the tool

implemented in PAUP v4.0a (Swofford and Sullivan 2009). The minke whale, blue whale, and pygmy right whale were set as outgroups and the standard bootstrapping method with 100 replicates was used to run the analysis. For SNAPP, the Ruby script *snapp\_prep.rb* was used to prepare the input file with five million MCMC generations. We ran SNAPP implemented in BEAST v2.7.1 (Bouckaert et al. 2019) and used posterior probability to evaluate the uncertainty of the analysis.

### Inferring Introgression with Phylogenetic Network

We used autosomal biallelic SNVs from the same seven baleen whale samples used in the phylogenomic analyses for the phylogenetic network analysis. To minimize the linkage in our SNV data set and for computational efficiency, we sampled SNVs with vcftools so that all remaining SNVs were at least 10 kbp apart. We converted the file format with *vcf2phylip* v2.8 (Ortiz 2019) and used the R package SNP2CF v1.6 (Olave and Meyer 2020) to generate an input CF file for the SNaQ (Solís-Lemus and Ané 2016). SNaQ was run in the Julia v1.8.3 environment using the package PhyloNetworks (Solís-Lemus et al. 2017) to estimate the pseudolikelihoods of networks. Since SNaQ can only infer level one network, we set the number of introgression events from zero to two. The starting topology for SNaQ was the one estimated by SVDQuartets. In cases when the best network reported by SNaQ contains gene flow from a younger lineage to an older lineage, which is temporally impossible to occur, we manually examined all the evaluated networks to find the second-best network. The resulting networks were visualized with the Julia package PhyloPlots. We compared the gene tree discordance that can be explained by dichotomous tree (ILS alone) versus the inferred network with fittedQuartetCF implemented in PhyloNetworks and visualized the result with ggplot2 in the R environment called within Julia.

### Inferring Introgression with *D*-statistics and Sliding Window Statistics

We used Dsuite v0.5r47 (Malinsky et al. 2021) Dtrios to estimate genome-wide *D*-statistics. For introgression among balaenid whales, we used autosomal biallelic SNVs from selected trios of the balaenid species/populations and used the minke whale, blue whale, and pygmy right whale as outgroups. For introgression among major baleen whale lineages, we generated another set of SNVs using the sperm whale genome as a reference. We applied *D*-statistics on the trios of baleen whales: [blue whale, pygmy right whale, southern right whale], using sperm whale as an outgroup. To identify putative introgression regions within the North Atlantic right whale, we estimated genome-wide sliding window *f*<sub>dm</sub> statistics with Dinvestigate in the selected set of trios: [North Pacific right

whale, North Atlantic right whale, bowhead whale]. Each window is 50 SNVs in size with a step size of 5 SNVs. To infer introgression in the sex chromosome, we extracted SNVs located on the X chromosome of the blue whale. We estimated the level of introgression with SNaQ using the same methods described above.

### Multiple Sequentially Markovian Coalescent and MSMC-IM

To minimize ascertainment bias, we used the bowhead whale genome, sister group to the three right whales among the extant Balaenidae lineages, as the reference for MSMC and MSMC-IM analyses. To identify scaffolds that belong to autosomes and to improve the contiguity of the assembly, we used a chromosome-level assembly of the southern right whale for reference-guided scaffolding of the published bowhead whale reference genome. We aligned the bowhead whale genome to the southern right whale genome using minimap2.1 (Li 2018) and scaffolded the bowhead whale genome into pseudochromosomes with ragtag v2.1.0 (Alonge et al. 2022). We then performed MSMC analysis on each of the four balaenid species (Schiffels and Wang 2020). We separately analyzed the two populations of bowhead whales. For the EGSB bowhead whales, we selected two deeply sequenced individuals as representatives (accession numbers: SRR15669485 and SRR15669487) and aligned the short reads to the improved bowhead whale genome. We generated SNVs-only vcf and bed files of masked regions for each autosome based on sequencing coverage with bcftools and bamCaller.py scripts contained in the MSMC package. We generated inputs for MSMC for each autosome from vcf and masked region bed files using generate\_multihetsep.py. We then ran MSMC2 on all autosomes of each species. For MSMC-IM, we first phased the SNVs containing all balaenid whale samples (Browning and Browning 2013). We then used the phased SNVs to estimate cross-coalescence rates for all possible species and population pairs with MSMC2. The MSMC results of the analyzed population/species pairs mentioned above and their cross-coalescence rate were combined with combineCrossCoal.py. We utilized the combined data as input to MSMC-IM (Wang et al. 2020) to estimate migration rate over time in all balaenid species and population pairs. Trio-based estimation of mutation rate was only available in bowhead whale, we therefore used  $1.25 \times 10^{-8}$ , the mean mutation rate of studied baleen whales (Suárez-Menéndez et al. 2023), for MSMC-IM. Because phasing error disproportionally affects estimation of both MSMC and MSMC-IM in more recent times, leading to notable artifacts, after examining the outputs of MSMC-IM, we manually masked the four most recent time segments in all results. We plotted the results of MSMC and MSMC-IM with a custom python script.

We used the following generation times for each species: (i) North Atlantic right whale: 35.7, (ii) North Pacific right whale: 28, (iii) southern right whale: 29, and (iv) bowhead whale: 35 and 52. Average generation time for a pair of species was calculated using the following equation:

$$G_{12} = 2G_1G_2/(G_1 + G_2) \quad (1)$$

where  $G_1G_1$  and  $G_2G_2$  are generation times from each species. We transformed the time unit from generation to year by multiplying generation by average generation time. We used the same methods to perform MSMC-IM analysis between ECWG and EGSB bowhead whales. We calibrated the bowhead whale generation time by manually inspecting the overlapping of migration peaks and glacial periods. We estimated 50% cumulative migration rates in all pairs of right whales, the two bowhead whale populations, and between SRW and ECWG bowhead whales, as an estimation for divergence date. The reason we choose to use SRW to estimate divergence time between BHW is that SRW has no evidence of introgression. The estimation was based on the output of MSCM-IM and calculated using interpolation with a custom python script. When calculating cumulative migration rate, we excluded migration rate estimated in the masked region and the archaic introgression peak region.

### Enrichment Analyses of Genes within Introgression Regions

We identified the sliding windows with top 5% *f<sub>dm</sub>* values in the trios [North Pacific right whale, North Atlantic right whale, Bowhead whale]. We then extracted the genes that fall within these regions. We used g:Profiler (Reimand et al. 2007) to perform GO and enrichment analyses on the extracted genes based on the Ensembl annotation of blue whale. We corrected for multitesting using the default g:SCS correction and set the significance cutoff to 0.05.

### Supplementary Material

Supplementary material is available at *Genome Biology and Evolution* online.

### Acknowledgments

This work was supported by the National Institute of Health grant R35ES032876 (J.P.W.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Environmental Health Sciences or National Institute of Health. We would like to acknowledge Cheng-Hsiu Tsai for his input of ideas from the perspective of a paleontologist, Chen-Ruei Lee and Lo-Yu Chang for commenting on

the manuscript, Joshua Kim for proofreading the manuscript, and Fiona Puntieri for the advice on the graphic designs.

## Author Contributions

B.-W.L. designed the study. A.M. collected the data. B.-W.L. conducted the analyses and wrote the draft manuscript. F.M.-R. provided guidance during the process and edited the manuscript. J.P.W. and S.M. secured the funding. B.-W.L. and P.F. finalized the manuscript.

## Data Availability

All sequencing data used in this study came from open databases. Scripts used for the analyses can be found in <https://github.com/BaiweiLo/Balaenidae-demography>.

## Literature Cited

- Alonge M, Lebeigle L, Kirsche M, Jenike K, Ou S, Aganezov S, Wang X, Lippman ZB, Schatz MC, Soyk S. Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biol.* 2022;23(1):1–19. <https://doi.org/10.1186/s13059-022-02823-7>.
- Árnason Ú, Lammers F, Kumar V, Nilsson MA, Janke A. Whole-genome sequencing of the blue whale and other rorquals finds signatures for introgressive gene flow. *Sci Adv.* 2018;4(4):eaap9873. <https://doi.org/10.1126/sciadv.aap9873>.
- Bisconti M. Evolutionary history of Balaenidae. *Cranium.* 2003;20:9–50.
- Bisconti M, Lambert O, Bosselaers M. Revision of “Balaena” belgica reveals a new right whale species, the possible ancestry of the northern right whale, *Eubalaena glacialis*, and the ages of divergence for the living right whale species. *PeerJ.* 2017;5:e3464. <https://doi.org/10.7717/peerj.3464>.
- Bisconti M, Pellegrino L, Carnevale G. Evolution of gigantism in right and bowhead whales (Cetacea: Mysticeti: Balaenidae). *Biol. J. Linn. Soc.* 2021;134(2):498–524. <https://doi.org/10.1093/biolinnean/blab086>.
- Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, Heled J, Jones G, Kühnert D, De Maio N, et al. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Comput Biol.* 2019;15(4):e1006650. <https://doi.org/10.1371/journal.pcbi.1006650>.
- Browning BL, Browning SR. Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics.* 2013;194(2):459–471. <https://doi.org/10.1534/genetics.113.150029>.
- Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, Roychoudhury A. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Mol Biol Evol.* 2012;29(8):1917–1932. <https://doi.org/10.1093/molbev/mss086>.
- Bukhman YV, Morin PA, Meyer S, Chu L-F, Jacobsen JK, J Antosiewicz-Bourget, Mamott D, Argus C, Bolin J, Berres ME, et al. A high-quality blue whale genome, segmental duplications, and historical demography. *Mol Biol Evol.* 2024;41(3):msae036 <https://doi.org/10.1093/molbev/msae036>.
- Buono MR, Fernández MS, Cozzuol MA, Cuitiño JI, Fitzgerald EMG. The early Miocene balaenid *Morenocetus parvus* from Patagonia (Argentina) and the evolution of right whales. *PeerJ.* 2017;5:e4148. <https://doi.org/10.7717/peerj.4148>.
- Carroll EL, Ott PH, McMillan LF, Vernazzani BG, Neveceralova P, Vermeulen E, Gaggiotti OE, Andriolo A, Scott Baker C, Bamford C, et al. Genetic diversity and connectivity of southern right whales (*Eubalaena australis*) found in the Brazil and Chile–Peru wintering grounds and the south Georgia (Islas Georgias del Sur) feeding ground. *J Hered.* 2020;111(3):263–276. <https://doi.org/10.1093/jhered/esaa010>.
- Cerca J, Westbury MV, Heide-Jørgensen MP, Kovacs KM, Lorenzen ED, Lydersen C, Shpak OV, Wiig Ø, Bachmann L. High genomic diversity in the endangered East Greenland Svalbard Barents Sea stock of bowhead whales (*Balaena mysticetus*). *Sci Reports.* 2022;12:1–11. <https://doi.org/10.1038/s41598-022-09868-5>.
- Chifman J, Kubatko L. Quartet inference from SNP data under the coalescent model. *Bioinformatics.* 2014;30(23):3317–3324. <https://doi.org/10.1093/bioinformatics/btu530>.
- Churchill M, Berta A, Deméré T. The systematics of right whales (Mysticeti: Balaenidae). *Mar. Mammal Sci.* 2011;28(3):497–521. <https://doi.org/10.1111/j.1748-7692.2011.00504.x>.
- Clark PU, Archer D, Pollard D, Blum JD, Rial JA, Brovkin V, Mix AC, Pisas NG, Roy M. The middle Pleistocene transition: characteristics, mechanisms, and implications for long-term changes in atmospheric pCO<sub>2</sub>. *Quat Sci Rev.* 2006;25(23–24):3150–3184. <https://doi.org/10.1016/j.quascirev.2006.07.008>.
- Crossman CA, Fontaine MC, Frasier TR. A comparison of genomic diversity and demographic history of the North Atlantic and Southwest Atlantic southern right whales. *Mol Ecol.* 2023;00:1–17. <https://doi.org/10.1111/mec.17099>.
- Dagilis AJ, Peede D, Coughlan JM, Jofre GI, D’Agostino ERR, Mavengere H, Tate AD, Matute DR. A need for standardized reporting of introgression: insights from studies across eukaryotes. *Evol Lett.* 2022;6(5):344. <https://doi.org/10.1002/evl3.294>.
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. The variant call format and VCFtools. *Bioinformatics.* 2011;27(15):2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>.
- Dannemann M, Andrés AM, Kelso J. Introgression of Neandertal- and Denisovan-like haplotypes contributes to adaptive variation in human toll-like receptors. *Am J Hum Genet.* 2016;98(1):22–33. <https://doi.org/10.1016/j.ajhg.2015.11.015>.
- Davies JL. The antitropical factor in cetacean speciation. *Evolution (N Y).* 1963;17:107–116. <https://doi.org/10.1111/j.1558-5646.1963.tb03258.x>.
- Eberlein C, Hénault M, Fijarczyk A, Charron G, Bouvier M, Kohn LM, Anderson JB, Landry CR. Hybridization is a recurrent evolutionary stimulus in wild yeast speciation. *Nat Commun.* 2019;10(1):1–14. <https://doi.org/10.1038/s41467-018-07882-8>.
- Edelman NB, Mallet J. Prevalence and adaptive impact of introgression. *Annu Rev Genet.* 2021;55(1):265–283. <https://doi.org/10.1146/annurev-genet-021821-020805>.
- Fontaine M. Reticulated evolution marks the rapid diversification of the Delphinoidae. *Peer Commun Evol Biol.* 2022;100139. <https://doi.org/10.24072/pci.evolbiol.100139>.
- Footo AD, Kaschner K, Schultze SE, Garilao C, Ho SYW, Post K, Higham TFG, Stokowska C, Van Der Es H, Embling CB, et al. Ancient DNA reveals that bowhead whale lineages survived Late Pleistocene climate change and habitat shifts. *Nat Commun.* 2013;4(1):1–7. <https://doi.org/10.1038/ncomms2714>.
- Footo AD, Martin MD, Louis M, Pacheco G, Robertson KM, Sinding MHS, Amaral AR, Baird RW, Baker CS, Ballance L, et al. Killer whale genomes reveal a complex history of recurrent admixture and vicariance. *Mol Ecol.* 2019;28(14):3427–3444. <https://doi.org/10.1111/mec.15099>.



- Fraïsse C, Sachdeva H. The rates of introgression and barriers to genetic exchange between hybridizing species: sex chromosomes vs autosomes. *Genetics*. 2021;217(2):iyaa025. <https://doi.org/10.1093/genetics/iyaa025>.
- Furni F, Secchi ER, Speller C, Dendanto D, Ramp C, Larsen F, Mizroch S, Robbins J, Sears R, Urbán J, et al. Phylogenomics and pervasive genome-wide phylogenetic discordance among fin whales (*Balaenoptera physalus*). *Syst Biol*. 2024;73(6):873–885. <https://doi.org/10.1093/sysbio/syae049>.
- Gaines CA, Hare MP, Beck SE, Rosenbaum HC. Nuclear markers confirm taxonomic status and relationships among highly endangered and closely related right whale species. *Proc R Soc B Biol Sci*. 2005;272(1562):533–542. <https://doi.org/10.1098/rspb.2004.2895>.
- George JC, Thewissen JGM. The bowhead whale: *balaena mysticetus*: biology and human interactions. Academic Press; 2020.
- Guo W, Sun D, Cao Y, Xiao L, Huang X, Ren W, Xu S, Yang G. 2021. Extensive interspecific gene flow shaped complex evolutionary history and underestimated species diversity in rapidly radiated dolphins. *J Mamm Evol*. 2021;29(2):353–367. <https://doi.org/10.1007/s10914-021-09581-6>.
- Hassanin A, Delsuc F, Ropiquet A, Hammer C, Jansen Van Vuuren B, Matthee C, Ruiz-Garcia M, Catzeffis F, Areskoug V, Nguyen TT, et al. Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a comprehensive analysis of mitochondrial genomes. *C R Biol*. 2012;335(1):32–50. <https://doi.org/10.1016/j.crv.2011.11.002>.
- Hibbins MS, Hahn MW. Phylogenomic approaches to detecting and characterizing introgression. *Genetics*. 2022;220(2):iyab173. <https://doi.org/10.1093/genetics/iyab173>.
- Huerta-Sánchez E, Jin X, Asan, Bianba Z, Peter BM, Vinckenbosch N, Liang Y, Yi X, He M, Somel M, et al. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nordisk Alkohol Nark*. 2014;512(7513):194–197. <https://doi.org/10.1038/nature13408>.
- IUCN. 2023. The IUCN red list of threatened species. Version 2023-1. <https://www.iucnredlist.org>. [Accessed 2023 December 1]
- Jones MR, Scott Mills L, Alves PC, Callahan CM, Alves JM, Lafferty DJR, Jiggins FM, Jensen JD, Melo-Ferreira J, Good JM. Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. *Science*. 2018;360(6395):1355–1358. <https://doi.org/10.1126/science.aar5273>.
- Jossey S, Hadrath O, Loureiro L, Weir JT, Lim BK, Miller J, Scherer SW, Goksøyr A, Lille-Langøy R, Kovacs KM, et al. Population structure and history of North Atlantic Blue whales (*Balaenoptera musculus musculus*) inferred from whole genome sequence analysis. *Conserv Genet*. 2024;25(2):357–371. <https://doi.org/10.1007/s10592-023-01584-5>.
- Kaliszewska ZA, Seger J, Rowntree VJ, Barco SG, Benegas R, Best PB, Brown MW, Brownell RL, Carribero A, Harcourt R, et al. Population histories of right whales (Cetacea: Eubalaena) inferred from mitochondrial sequence diversities and divergences of their whale lice (Amphipoda: Cyamus). *Mol Ecol*. 2005;14(11):3439–3456. <https://doi.org/10.1111/j.1365-294X.2005.02664.x>.
- Keane M, Semeiks J, Webb AE, Li YI, Quesada V, Craig T, Madsen LB, van Dam S, Brawand D, Marques PI, et al. Insights into the evolution of longevity from the bowhead whale genome. *Cell Rep*. 2015;10(1):112. <https://doi.org/10.1016/j.celrep.2014.12.008>.
- Kimura T. Review of the fossil balaenids from Japan with a re-description of *Eubalaena shinshuensis* (Mammalia, Cetacea, Mysticeti). *Quad del Mus di Stor Nat di Livorno*. 2009;22:3–21.
- Kishida T, Thewissen J, Hayakawa T, Imai H, Agata K. Aquatic adaptation and the evolution of smell and taste in whales. *Zool Lett*. 2015;1(1):1–10. <https://doi.org/10.1186/s40851-014-0002-z>.
- Korneliussen TS, Albrechtsen A, Nielsen R. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics*. 2014;15(1):356. <https://doi.org/10.1186/s12859-014-0356-4>.
- Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018;34(18):3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Subgroup 1000 Genome Project Data Processing. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
- Long C, Kubatko L. The effect of gene flow on coalescent-based species-tree inference. *Syst Biol*. 2018;67(5):770–785. <https://doi.org/10.1093/sysbio/syy020>.
- Malinsky M, Matschiner M, Svardal H. Dsuite—fast D-statistics and related admixture evidence from VCF files. *Mol Ecol Resour*. 2021;21(2):584–595. <https://doi.org/10.1111/1755-0998.13265>.
- McGowen MR, Spaulding M, Gatesy J. Divergence date estimation and a comprehensive molecular tree of extant cetaceans. *Mol Phylogenet Evol*. 2009;53(3):891–906. <https://doi.org/10.1016/j.ympev.2009.08.018>.
- McGowen MR, Tsagkogeorga G, Álvarez-Carretero S, Dos Reis M, Struebig M, Deaville R, Jepson PD, Jarman S, Polanowski A, Morin PA, et al. Phylogenomic resolution of the cetacean tree of life using target sequence capture. *Syst Biol*. 2020;69(3):479–501. <https://doi.org/10.1093/sysbio/syz068>.
- Miralles L, Oremus M, Silva MA, Planes S, Garcia-Vazquez E. Interspecific hybridization in pilot whales and asymmetric genetic introgression in northern globicephala melas under the scenario of global warming. *PLoS One*. 2016;11(8):e0160080. <https://doi.org/10.1371/journal.pone.0160080>.
- Morel B, Kozlov AM, Stamatakis A. ParGenes: a tool for massively parallel model selection and phylogenetic tree inference on thousands of genes. *Bioinformatics*. 2019;35(10):1771–1773. <https://doi.org/10.1093/bioinformatics/bty839>.
- Moura AE, Shreves K, Pilot M, Andrews KR, Moore DM, Kishida T, Möller L, Natoli A, Gaspari S, McGowen M, et al. Phylogenomics of the genus *Tursiops* and closely related Delphininae reveals extensive reticulation among lineages and provides inference about eco-evolutionary drivers. *Mol Phylogenet Evol*. 2020;146:106756. <https://doi.org/10.1016/j.ympev.2020.106756>.
- Narasimhan V, Danecek P, Scally A, Xue Y, Tyler-Smith C, Durbin R. BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics*. 2016;32(11):1749–1751. <https://doi.org/10.1093/bioinformatics/btw044>.
- Olave M, Meyer A. Implementing large genomic single nucleotide polymorphism data sets in phylogenetic network reconstructions: a case study of particularly rapid radiations of cichlid fish. *Syst Biol*. 2020;69(5):848–862. <https://doi.org/10.1093/sysbio/syaa005>.
- Ortiz EM. 2019. vcf2phylip v2.0: convert a VCF matrix into several matrix formats for phylogenetic analysis. Zenodo
- Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D. Ancient admixture in human history. *Genetics*. 2012;192(3):1065–1093. <https://doi.org/10.1534/genetics.112.145037>.
- Quinlan AR. BEDTools: the Swiss-Army tool for genome feature analysis. *Curr. Protoc. Bioinforma*. 2014;47(1):11.12.1–11.12.34. <https://doi.org/10.1002/0471250953.bi1112s47>.
- Reimand J, Kull M, Peterson H, Hansen J, Vilo J. g:Profiler—a web-based toolset for functional profiling of gene lists from large-scale



- experiments. *Nucleic Acids Res.* 2007;35(suppl\_2):W193. <https://doi.org/10.1093/nar/gkm226>.
- Richards EJ, Martin CH. Adaptive introgression from distant Caribbean islands contributed to the diversification of a microendemic adaptive radiation of trophic specialist pupfishes. *PLOS Genet.* 2017;13(8):e1006919. <https://doi.org/10.1371/journal.pgen.1006919>.
- Rieseberg LH, Carney SE. Plant hybridization (Tansley Review No. 102). *New Phytol.* 1998;140(4):599–624. <https://doi.org/10.1046/j.1469-8137.1998.00315.x>.
- Rooney AP, Honeycutt RL, Derr JN. Historical population size change of bowhead whales inferred from DNA sequence polymorphism data. *Evolution* (N Y). 2001;55:1678–1685. <https://doi.org/10.1111/j.0014-3820.2001.tb00687.x>.
- Rosenbaum HC, Brownell RL, Brown MW, Schaeff C, Portway V, White BN, Malik S, Pastene LA, Patenaude NJ, Baker CS, et al. World-wide genetic differentiation of Eubalaena: questioning the number of right whale species. *Mol Ecol.* 2000;9(11):1793–1802. <https://doi.org/10.1046/j.1365-294x.2000.01066.x>.
- Sasaki T, Nikaido M, Hamilton H, Goto M, Kato H, Kanda N, Pastene LA, Cao Y, Fordyce RE, Hasegawa M, et al. Mitochondrial phylogenetics and evolution of mysticete whales. *Syst Biol.* 2005;54(1):77–90. <https://doi.org/10.1080/10635150590905939>.
- Schiffels S, Wang K. MSMC and MSMC2: the multiple sequentially Markovian coalescent. *Methods Mol Biol.* 2020;2090:147–166. [https://doi.org/10.1007/978-1-0716-0199-0\\_7](https://doi.org/10.1007/978-1-0716-0199-0_7).
- Skovrind M, Castruita JAS, Haile J, Treadaway EC, Gopalakrishnan S, Westbury M V, Heide-Jørgensen MP, Szpak P, Lorenzen ED. Hybridization between two high Arctic cetaceans confirmed by genomic analysis. *Sci Rep.* 2019;9(1):7729. <https://doi.org/10.1038/s41598-019-44038-0>.
- Slater GJ, Goldbogen JA, Pyenson ND. Independent evolution of baleen whale gigantism linked to Plio-Pleistocene ocean dynamics. *Proc R Soc B Biol Sci.* 2017;284(1855):20170546. <https://doi.org/10.1098/rspb.2017.0546>.
- Slater GJ, Price SA, Santini F, Alfaro ME. Diversity versus disparity and the radiation of modern cetaceans. *Proc R Soc B Biol Sci.* 2010;277(1697):3097–3104. <https://doi.org/10.1098/rspb.2010.0408>.
- Solís-Lemus C, Ané C. Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. *PLOS Genet.* 2016;12(3):e1005896. <https://doi.org/10.1371/journal.pgen.1005896>.
- Solís-Lemus C, Bastide P, Ané C. PhyloNetworks: a package for phylogenetic networks. *Mol Biol Evol.* 2017;34(12):3292–3298. <https://doi.org/10.1093/molbev/msx235>.
- Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 2014;30(9):1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>.
- Steeman ME, Hebsgaard MB, Fordyce RE, Ho SYW, Rabosky DL, Nielsen R, Rahbek C, Glenner H, Sørensen M V, Willerslev E. Radiation of extant cetaceans driven by restructuring of the oceans. *Syst Biol.* 2009;58(6):573–585. <https://doi.org/10.1093/sysbio/syp060>.
- Suárez-Menéndez M, Bérubé M, Furni F, Rivera-León VE, Heide-Jørgensen MP, Larsen F, Sears R, Ramp C, Eriksson BK, Etienne RS, et al. Wild pedigrees inform mutation rates and historic abundance in baleen whales. *Science.* 2023;381(6661):990–995. <https://doi.org/10.1126/science.adf2160>.
- Swofford D, Sullivan J. Phylogeny inference based on parsimony and other methods using PAUP\*. In: *The Phylogenetic Handbook*; 2009. p. 267–312.
- Taylor BL, Chivers SJ, Larese J, Perrin WF. Generation length and percent mature estimates for IUCN assessments of cetaceans. NOAA, NMFS, Southwest Fisheries Science Center Administrative Report LJ-07-01. 2007:21.
- Tsai CH, Chang CH. A right whale (Mysticeti, Balaenidae) from the Pleistocene of Taiwan. *Zool Lett.* 2019;5(1):1–7. <https://doi.org/10.1186/s40851-019-0153-z>.
- Tsai CH, Collareta A, Fitzgerald EMG, Marx FG, Kohno N, Bosselaers M, Insacco G, Reitano A, Catanzariti R, Oishi M, et al. Northern pygmy right whales highlight Quaternary marine mammal interchange. *Curr Biol.* 2017;27(19):R1058–R1059. <https://doi.org/10.1016/j.cub.2017.08.056>.
- Tsai CH, Mead JG. Crossing the equator: a northern occurrence of the pygmy right whale. *Zool. Lett.* 2018;4(1):1–4. <https://doi.org/10.1186/s40851-018-0117-8>.
- van Weelden C, Towers JR, Bosker T. Impacts of climate change on cetacean distribution, habitat and migration. *Clim. Chang. Ecol.* 2021;1:100009. <https://doi.org/10.1016/j.ecochg.2021.100009>.
- Wang K, Mathieson I, O'Connell J, Schiffels S. Tracking human population structure through time from whole genome sequences. *PLOS Genet.* 2020;16(3):e1008552. <https://doi.org/10.1371/journal.pgen.1008552>.
- Westbury MV, Cabrera AA, Rey-Iglesia A, De Cahsan B, Duchêne DA, Hartmann S, Lorenzen ED. A genomic assessment of the marine-speciation paradox within the toothed whale superfamily Delphinoidea. *Mol Ecol.* 2023;32(17):4829–4843. <https://doi.org/10.1111/mec.17069>.
- Westbury MV, Petersen B, Lorenzen ED. Genomic analyses reveal an absence of contemporary introgressive admixture between fin whales and blue whales, despite known hybrids. *PLoS One.* 2019;14(9):e0222004. <https://doi.org/10.1371/journal.pone.0222004>.
- Wolf M, Zapf K, Gupta DK, Hiller M, Árnason Ú, Janke A. The genome of the pygmy right whale illuminates the evolution of rorquals. *BMC Biol.* 2023;21(1):1–18. <https://doi.org/10.1186/s12915-023-01579-1>.
- Zhang C, Rabiee M, Sayyari E, Mirarab S. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics.* 2018;19(S6):153. <https://doi.org/10.1186/s12859-018-2129-y>.

Associate editor: Eran Elhaik