PhD Thesis

Mikkel Skovrind

Phylogeography, demographic history and hybridization in belugas *Delphinapterus leucas*

Supervisor: Eline D. Lorenzen

Submitted on: 14 June 2019
Phylogeography, demographic history and hybridization in belugas *Delphinapterus leucas*

PhD Thesis
Mikkel Skovrind
Name of department: Natural History Museum of Denmark

Candidate: Mikkel Skovrind

Title: Phylogeography, demographic history and hybridization in belugas *Delphinapterus leucas*

Supervisor: Eline D. Lorenzen

Submitted on: 14 June 2019

The research presented in this thesis was supported by the Carlsberg Foundation Distinguished Associate Professor Fellowship, grant no CF16-0202, and the Natural History Museum of Denmark.

Front cover: Mads Peter Heide-Jørgensen
## Table of contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acknowledgements</td>
<td>6</td>
</tr>
<tr>
<td>Preface</td>
<td>8</td>
</tr>
<tr>
<td>Introduction</td>
<td>9</td>
</tr>
<tr>
<td>Description of belugas</td>
<td>9</td>
</tr>
<tr>
<td>Beluga genetics</td>
<td>12</td>
</tr>
<tr>
<td>English summary</td>
<td>13</td>
</tr>
<tr>
<td>Danish summary (dansk resumé)</td>
<td>15</td>
</tr>
<tr>
<td>Cited literature</td>
<td>16</td>
</tr>
<tr>
<td>Chapter 1 <em>Patterns of mtDNA Variation in Relation to Currently Recognized Stocks of Beluga Whales</em> Delphinapterus leucas</td>
<td>18</td>
</tr>
<tr>
<td>Chapter 2 <em>Mitochondrial genome divergence between beluga whales in Baffin Bay and the Sea of Okhotsk</em></td>
<td>46</td>
</tr>
<tr>
<td>Chapter 3 <em>Habitat fragmentation and secondary contact shape</em> phylodispersal and demographic history of belugas</td>
<td>50</td>
</tr>
<tr>
<td>Chapter 4 <em>Hybridization between two high Arctic cetaceans confirmed by genomic analysis</em></td>
<td>96</td>
</tr>
</tbody>
</table>
Acknowledgements

In the spring of 2007 I had just turned 27 years old and I started to realise that the manual job I was holding at the time was making me miserable. There was no possibility of advancement and what made it even more frustrating was the fact that the other jobs for which I was eligible seemed no better. This was when I knew I had to get some more education. I never attended high school, so I knew it was going to be a long, difficult journey. In hindsight, twelve years later, this thesis is the accumulation of that journey. This journey would never have been possible without the support of all the people around me, colleagues, friends and especially family.

I would like to express my deep gratitude to Eline D. Lorenzen for believing in me, sharing her expertise and experience while guiding me through the academic jungle. I also wish to thank Professor Mads Peter Heide-Jørgensen for sharing his deep knowledge of Arctic whales and taking me to Hjørnedal in Northeast Greenland and letting me take part in the narwhal tagging. Thanks to the rest of Team Hjørnedal 2017, Terry, Susanna, Mikkel, Inuta and HC. Thanks to Rikke, Rasmus, Nynne, Outi and everyone else at Greenland Institute of Natural Resources for your valuable input. I also thank all my co-authors and collaborators from across the world. Special thanks goes to Ilya Meschersky, Lianne Postma and Steve Furgusson for sharing their knowledge on belugas and making me feel welcome in the field of beluga research.

This thesis would not have been completed without the help of my highly skilled colleagues at the University of Copenhagen. A work culture which promotes sharing of data, information and methodologies has led to a sense of community and made my years at UCPH not just inspiring but also enjoyable. I thank every single one of you. Special thanks goes out to James Haile, Marie Louis, Mick Westbury and Jose “Sama” Castruita who all stood by me with help, encouragement and support when things got tough. When working in the lab or sitting in front of a computer to do analyses it is easy to forget the large amount of work required to collect hundreds of whale samples. Without such collection effort, throughout the Arctic, my project would never have been possible. I therefore thank everyone who has collected samples, shared samples or in any other way contributed to the global archive of beluga samples, which my project and this thesis has benefited from.

Thanks to all my friends. Fishing buddies, biologists, hip-hop heads, locals and internationals. I would like to thank the members Darwin’s Armé, Emil Flindt, Lasse Højrup and Nick Juul
Brandtberg. We all started studying biology in 2009 and supported each other through countless classes, exciting exercises and exhausting exams. This meant the world to me and I deeply appreciate that we are still friends. Thanks to my mother, my father, my brother and my sister for always supporting me regardless of the path I chose. When embarking on a journey like the one I have been on, it is paramount to have a strong foundation - and you were surely that. Thanks to Henriette, Per and Kristoffer for welcoming me in into their family and accepting me for who I am. My deepest gratitude goes to my wife, mother of my sons, Karoline who always stands by me, encourages me when I am frustrated and cheers me on when I succeed. Thank you.
Preface

This thesis is the result of a three year PhD project conducted in the Section for Evolutionary Genomics at the Natural History Museum of Denmark, University of Copenhagen. The project was supervised by Eline D. Lorenzen and was conducted from June 15th 2016 to June 14th 2019.

The thesis consists of an introduction and four chapters. Two chapters are published (Chapter 2 and Chapter 4), one is in second round of review (Chapter 2) and the final chapter is a draft manuscript (Chapter 3). The contribution of the candidate to each of the four chapters is listed below.

**Chapter 1**: The candidate participated in all aspects of the publication, including planning, data collection, analyses, interpretation and writing.

**Chapter 2**: The candidate participated in all aspects of the manuscript, including keeping correspondence with data providers, deciding which data to include, choose and perform the analyses, interpret the results and write the manuscript. He also played a major role in reworking the manuscript after a first round of reviews.

**Chapter 3**: The candidate participated in all aspects of the manuscript related to the DNA analyses, including DNA extraction, library build, data analysis, interpretation of results. He also played a role in the parameter selection, GIS plotting and summary statistics of the species distribution model as well as interpretation of the combined results and manuscript writing.

**Chapter 4**: The candidate participated in the planning of the publication, including the selection of individuals for the genomic and stable isotope reference panels. He was heavily involved in the development of the genomic analytic approach and the interpretation of the genomic result. He also played a major role in adjusting the analyses and reworking the manuscript through three rounds of review before publication.

Besides the work included in this thesis the candidate generated a **genome-wide dataset** from the samples included in Chapter 3. This dataset is currently being analysed, promising great new insights to be had. This work is still in an initial, exploratory phase and it is still too early to disclose any results, but a large and time consuming effort has been put in to this dataset during the completion of this thesis.
Introduction

The research underlying this thesis started as a quest to investigate the phylogeography of belugas. The plan was not just a few belugas, but belugas from the whole distribution - and not just a little data, but with as much data as possible. The main focus of my research did revolve around phylogeography, but along the way it also covered other subjects such as hybridization between belugas and narwhals, stable isotope analyses and squeezing every last drop of information from extreme low coverage sequencing data.

Description of belugas

Belugas or white whales (Delphinapterus leucas, Dallas 1776) are toothed whales with an adult size of 3.5-5.5 meters, easily recognized by their melon shaped head, white skin (in adults) and lack of dorsal fin (O’Corry-Crowe 2018). Females can have a single calf every three years from the age of 8. The mating takes place in early spring and gestation lasts ~14 months before calves are born in early summer. At birth their skin is dark grey, but as they grow older their skin becomes white. Belugas form social groups primarily including females and young individuals with males forming their own groups, mainly interacting with females during the mating season (Colbeck et al. 2013). Multiple pods can at times join to form huge beluga aggregations of hundreds or even thousands of individuals. Belugas are known for their vocalization earning them the nickname “canaries of the sea”. The vocalization is believed to reinforce social bonds and results suggest they have Individual signature calls (Vergara & Mikus 2019). Belugas feed on a wide variety of fishes, cephalopods and invertebrates (Quakenbush et al. 2015) and can dive to depths of 900 m (Citta et al. 2013).

Beluga or white whale, Delphinapterus leucas. Photo by Mikkel Skovrind
Belugas and their sister species narwhals (Monodon monoceros) are the only extant members of the *Monodontidae* family. The fossil record only includes few *Monodontids* (Barnes 1984; Vélez-Juarbe & Pyenson 2012; Ichishima et al. 2018; Pesci 2018), but they were found as far apart as Japan, Italy, Belgium and California suggesting a successful family with which evolved in subtropical or temperate climate, and only later adapted to the Arctic waters. Five million years ago when belugas became a separate species (Steeman et al. 2009) the climate was warmer than it is now, with an ice free Arctic (Ballantyne et al. 2013). However, ~2.5 million years ago the Quaternary ice age began, starting continuous cycles of Arctic ice expansions and retractions, which is ongoing to this day (Rahmstorf 2002). Models, estimating the maximum extent of the sea ice during the last glacial expansion suggests that most of the localities which today habor beluga populations, were covered by permanent sea ice (Kaschner et al. 2008). That belugas return to the same summering grounds every year and follow maternally inherited migration routes throughout their lives (O’Corry-Crowe 2018), could be interpreted as a potential threat to populations stuck in unfavorabel habitats. Nevertheless, on a larger time scale, belugas as a species must have had the behavioral plasticity allowing them to adapt to changes in sea ice distribution.

![Belugas among the pack ice.](image)

*Belugas among the pack ice. Photo by Mads Peter Heide-Jørgensen*
Belugas are endemic to the Arctic and Subarctic regions (O’Corry-Crowe 2018). They are found in both the Pacific and Atlantic oceans with a circumpolar distribution, omitting the Greenland Sea (NAMMCO 2018). During summers, belugas congregate in shallower water, usually in bays or estuaries where they can feed on anadromous fishes (Quakenbush et al. 2015), nurture the young of the year and molt an outer layer of dead skin. As winter approaches and the summering grounds get covered by sea ice belugas move to deeper ice free water. For some populations this encompasses long migrations (Citta et al. 2017) while others perform shorter seasonal movements within the same bay or region. Throughout the beluga range site fidelity to summering grounds seems to be a common trait, which has been shown to be stable over several decades (O’Corry-Crowe et al. 2018). Based on these summering grounds the scientific community recently recognized 21 separate stocks, identified by aerial survey, telemetry, genetics and behaviour (NAMMCO 2018).

As a species, belugas has a least concern classification by IUCN (Lowry, L., Reeves, R. & Laidre, K 2017). However, individual stocks as those in St Lawrence Estuary and Cook inlet are endangered; while the Ungava Bay and Southwest Greenland stocks are nearly and fully extirpated,
respectively (NAMMCO 2018). Beluga stocks, which has fallen to low abundance has, despite
decades of conservation efforts, shown no sign of recovery. The species wide abundance is
estimated around 180,000-200,000 with populations having 50 to 55,000 individuals (NAMMCO
2018). The populations that undergo long migrations such as the Beaufort Sea, Eastern High Atlantic
Baffin Bay, Eastern Chukchi Sea and Western Hudson Bay in general have higher abundance than the
more resident populations.

Beluga genetics

Since the nineteen eighties population genetics has been used to differentiate between
summering grounds, analysing primarily fragments of the mitochondrial control region or
microsatellites. The former, which solely reflects female genealogy, has at some point, been applied
in all stocks (Brennin et al. 1997; O’Corry-Crowe et al. 2002; Meschersky et al. 2008; Meschersky et
al. 2018; O’Corry-Crowe et al. 2010). Results revealed that almost all the recognized stocks are
differentiated by haplotype frequencies, but also shows that few stocks has unique mitochondrial
haplotypes found in no other stock. Microsatellites, which are repetitive regions found in the nuclear
genome, was first developed for belugas in the second half of the nineteen nineties and has been
used to distinguish between summer and winter grounds (Gladden et al. 1999), genetic
mark-recapture (Citta et al. 2018) and population movements (O’Corry-Crowe et al. 2018). Nuclear
DNA studies indicates that in some regions several summer stocks interbreed in their shared winter
habitat (Gladden et al. 1999; Meschersky et al. 2013), while populations stay separate in other
regions (O’Corry-Crowe et al. 2018). Despite several centuries of population-level genetic analyses
no range-wide analyses of the diversity and differentiation has been published. Application of
different mtDNA primers has led to different sections of the control region being used for population
genetic studies in different parts of the belugas distribution area. In combination with the relatively
short mtDNA sequences being used this has made direct comparisons of diversity and differentiation
levels across the distribution area difficult. Microsatellites require lab specific standards and data is
not easily combined between labs. Because of the difficulties in combining data from separate
studies and that no global studies of beluga population structure has been performed, our
understanding of the evolutionary and demographic histories of belugas in a global perspective is
limited.
English Summary

The aim of the thesis is to further our understanding of the population structure and
demographic history of beluga whales in a range-wide context, using genomic data. The thesis
includes an introduction, and four chapters written as draft manuscripts; Chapter 1 is in revision
with Marine Fisheries Review, Chapter 2 has been published in Mitochondrial DNA Part B, Chapter 3
is a draft manuscript intended for Molecular Ecology and Chapter 4 is accepted for publication in
Scientific Reports and will be published June 22th 2019.

Chapter 1 Circumpolar mtDNA population structure and variation in belugas: a review. In this
manuscript we review the genetic studies of belugas published to date, and for the first time present
a range-wide analysis of levels of mtDNA diversity and differentiation in beluga populations. Our
analysis of 302 bp of overlapping mtDNA control region sequence includes 2,933 individuals
spanning all 21 recognized beluga stocks, and represents 71 haplotypes defined by 16 variable sites.
Results put the diversity and differentiation in a global perspective and highlights the limitations of
short mtDNA fragments in beluga phylogeography.

Chapter 2 Mitochondrial genome divergence between beluga whales in Baffin Bay and the
Sea of Okhotsk. This publication has the first comparison of two complete mitochondrial sequences
from separate ends of the beluga range and their estimated divergence time. Despite the caveats of
small sample size and simplified methodology, we estimate that the divergence of beluga clades is in
the order of hundreds of thousands of years. The publication is available from:

Chapter 3 Habitat fragmentation and secondary contact shape phylogeography and
demographic history of belugas. In this manuscript, we analyse 2 nuclear genomes and 202
mitogenomes from 23 localities covering the entire beluga distribution, applying population
genetics, bayesian phylogenetic and species distribution modeling analyses. Our results show that
the earliest split within belugas happened around 750 kya with four lineages established before the
start of the last glaciation, and that current population structure is largely a result of secondary
contact of these old lineages. Further, our results show that the effective population size of belugas
has increased six-fold since LGM and that the suitable habitat range has increased since LGM, with
the winter area increasing 3-4 fold.

Chapter 4 Hybridization between two high Arctic cetaceans confirmed by genomic analysis.
We present genome-wide DNA sequence data from a monodontid skull which from its intermediate
morphology was hypothesized to be a beluga/narwhal hybrid. Using DNA retrieved from the skull
and a reference genomic panel of each parental species, we show that the specimen is indeed a
first-generation hybrid. Our analyses reveal that it was a male, with a narwhal mother. We also
present stable isotope analysis of the beluga/narwhal hybrid. The data suggest a unique diet of the hybrid relative to either parental species. These results further our understanding of the interaction between belugas and narwhals, and underscore the importance of natural history collections in monitoring changes in biodiversity.
Danish summary (Dansk resumé)


Kapitel 3 *Habitat fragmentation and secondary contact shape phylogeography and demographic history of belugas* I dette manuskript analyserer vi 2 nukleære genomer og 202 mitogenomer fra 23 lokaliteter dækkende hele hvidhvalens udbredelsesområde. Vi benytter populationsgenetik, fylogenetiske træer og modeller for habitat størrelse. Vores resultater viser at den tidligste deling i hvidhvalens stamtræ skete for omkring 750 tusind år siden, at fire grene fandtes før starten på den sidste istid og at disse gamle grene i høj grad er blandet i moderne hvidhvalbestande. Resultaterne viser også at størrelsen på egnede hvidhvalhabitater er tredoblet siden sidste istid og at hvidhvalerne har været i stand til at udnytte dem.

Kapitel 4 *Hybridization between two high Arctic cetaceans confirmed by genomic analysis.* I dette manuskript præsenterer vi genetisk data fra et hvalkranie, der tidligere, ud fra dets morfologi, er beskrevet som en hybrid mellem en hvidhval og en narhval. Ved hjælp af genetiske analyser af hybridkraniet og et referencepanel af hvidhvaler og narhvaler, viser vi, at der er tale om en førstegenerationshybrid mellem de to arter. Vores resultater viser også at hybriden var hankøn og at moren var en narhval. Vi analyserer også stabile isotoper fra hybridkraniet og sammenligner det med
stabile isotoper fra hvidhvaler og narhvaler. Disse resultater viser, at hybriden havde en kost der var forskellig fra begge forældrearter kost. Resultaterne udbygger vores forståelse af interaktionen mellem de to arter og understreger vigtigheden af naturhistoriske samlinger i monitoreringen af forandringer i biodiversitet.

Cited literature


Chapter 1

A first range-wide look

This manuscript was initiated in the early days of the project. I was reading through the genetic beluga literature and finding it rather difficult to keep it all in my head at once. So I started making tables of genetic markers, sample sizes and reused samples along with lists of Arctic locations with their English and indigenous names and eventually I started downloading mtDNA sequences. This is when the thought of making a range-wide review was born. When attending the Global Review of Monodontids (GROM) meeting in March 2017, I realised that I was not the only one looking for range-wide comparable information. I thought it would be quick and easy to compile the data, analyse it and publish the results, but I could not have been more wrong. Large amounts of tissue samples had arrived and I spent most of my time in the lab, so this project had to wait. In the meantime we had received sequences from Ilya Meschersky and Lianne Postma enabling us to include almost 3,000 samples. Along the way the project was handed to a bachelor student who continued my tables, lists and data collection before I found time to finally write the manuscript. It was decided that we would analyse the data in the context of the 21 stocks defined by the GROM meeting and try to publish it in a special GROM issue of Marine Fisheries Reviews. The manuscript was first submitted July 27th 2018, and again June 14th 2019.
Chapter 1

Patterns of mtDNA Variation in Relation to Currently Recognized Stocks of Beluga Whales *Delphinapterus leucas*

Mikkel Skovrind, Jose Alfredo Samaniego Castruita, Thomas Buur Madsen, Lianne Postma and Eline D. Lorenzen

Mikkel Skovrind, Jose Alfredo Samaniego Castruita, Thomas Buur Madsen and Eline D. Lorenzen are with the Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, 1350 København K, Denmark. Lianne Postma is with Fisheries and Oceans Canada, Central and Arctic Region, 501 University Crescent, Winnipeg, Manitoba, Canada R3T 2N6. Corresponding authors are Mikkel Skovrind (mikkel.skovrind@snm.ku.dk) and Eline Lorenzen (elinelorenzen@snm.ku.dk)

**ABSTRACT** - Belugas are one of only three whale species endemic to Arctic and sub-Arctic seas. They are found in both the Atlantic and Pacific Arctic and sub-Arctic, and are managed at a regional or national level within each of the five Arctic range countries, along with international fora used to share information and management strategies. Genetic data - primarily mtDNA control region sequences and microsatellites - have played an important role in defining appropriate management units. Here, we review the genetic studies published to date, and for the first time present a range-wide analysis of levels of mtDNA diversity and differentiation in beluga stocks. Our meta analysis is based on 302 bp of mtDNA control region sequence and includes 2,933 individuals spanning all 21 recognized beluga stocks. We find that all stocks are significantly differentiated from each other, except in five cases. The belugas in the St. Lawrence Estuary are the most distinct - the stock has the lowest range-wide level of genetic diversity and, with only two haplotypes present not found elsewhere, it is also the most well differentiated. Belugas in the Barents, Kara, and Laptev Seas stock have the highest level of diversity, supporting that this geographically far-ranging stock may
harbor several distinct sub-units. Our study highlights the challenges of compiling and comparing data from various publications with different study designs.

Introduction

Belugas (also known as white whales, *Delphinapterus leucas*) have a discontinuous circumpolar distribution and are one of only two toothed whale species found in the Arctic year round (O’Corry-Crowe, 2018). Belugas can migrate over vast distances - thousands of kilometres in some regions - between summering grounds, where they aggregate in shallow bays, and deeper wintering areas where they stay close to the sea ice fringe (Heide-Jørgensen et al., 2010; Citta et al., 2017). The size of beluga stocks differs markedly, from tens of thousands of individuals in Western Hudson Bay and Eastern Beaufort Sea to only hundreds in Cook Inlet and St. Lawrence Estuary (NAMMCO, 2018).

The Global Review of Monodontids (GROM) status review recently recognized 21 extant beluga stocks and one extirpated stock, in Southwest Greenland, largely defined as separate summer aggregation sites across the species range (NAMMCO, 2018). The GROM meeting took place in March 2017, and was the first of its kind in almost two decades. There, beluga researchers and stakeholders convened and combined data from telemetry, aerial surveys, traditional knowledge, and genetics to estimate the number of distinct beluga stocks. Based on the available information, the GROM panel reevaluated each previously recognized stock, in some cases merging old or recognizing new stocks. The previous GROM meeting in 1999 only included a review of Atlantic stocks, and in contrast to the 2017 meeting also included a review of wintering and mixed aggregations (NAMMCO, 1999). The international whaling commission (IWC) has also published reviews of beluga stocks in 1993 and 2000, recognizing 16 and 29 stocks, respectively (IWC, 1993,
2000), while a more recent review recognized 19 stocks (Laidre et al., 2015). This clearly illustrates that although belugas have been studied for decades, delimiting stocks is not straightforward.

Technological advances in the field of genetics during the 1990’s was readily adopted in beluga research to address questions regarding stock subdivision and connectivity. Several methodologies have been applied, including DNA fingerprinting (Patenaude et al., 1994), variation in the major histocompatibility complex locus DQβ (Murray, 1997) and restriction fragment length polymorphism (RFLP) (Brennin et al., 1997). The majority of studies published over the past two decades have analyzed regions of the maternally inherited mtDNA control region, and microsatellites (Fig. 1, Table 1).

Here, we review the population-level genetic data (mtDNA and microsatellites) published to date on the species. To test the genetic validity of the 21 beluga stocks presented in the GROM report, we compiled a pan-Arctic data set of published mtDNA sequences. Because studies have not always sequenced the same region of the mitochondrial genome, we trimmed the available, aligned sequences to allow the most comprehensive dataset. Our analysis included 302 bp of overlapping control region sequence from 2,933 beluga individuals spanning all 21 GROM stocks. Because microsatellite data from different studies cannot be combined or compared when no standard reference sample has been included, we were unable to carry out a similar range-wide analysis of available microsatellite data. Instead, we present an overview of the published microsatellite studies.
Chapter 1

Methods

Literature Review

We compiled the published population-genetic literature on belugas based on mtDNA and microsatellites. We searched Google Scholar using the search terms: “Beluga whale”, “Delphinapterus leucas”, “population genetics”, “population structure”, “mtDNA”, and “microsatellites”. Furthermore, we looked through the reference lists of the publications retrieved via Google Scholar for any publications that had not been retrieved in our original search. We also went through the recent GROM status review to ensure that we included all relevant papers used to define the 21 beluga stocks (NAMMCO, 2018). For each publication, we compiled information on genetic markers used, sample sites and any sample overlap among studies.

Our Use of the Term Stock

Several terms have been used in the literature to delimit distinct beluga units, including population, subpopulation, management unit and stock. Although they have at times been used interchangeably, their meanings differ. Population and subpopulation indicates different levels of reproductively isolated groups. Management unit indicates that each unit is being managed separately, and stock is a term implying that a natural resource is harvested as a single unit. None of these accurately describe distinct beluga units, however, we adopt the use of stocks as it is defined in the GROM report, which is to identify distinct beluga summering grounds (NAMMCO, 2018).
Chapter 1

mtDNA Data Compilation

A main objective of this study was to compile a range-wide dataset of beluga mtDNA sequences, to test the genetic validity of the GROM stocks, and to further our understanding of pan-Arctic patterns of structuring and variation in the species. When generating our dataset, we wanted to (1) maximise the length of overlapping mtDNA sequence, while also maximising the number of individuals analyzed, (2) ensure our dataset included individuals from all 21 stocks, and (3) avoid duplicate sequences from individuals sequenced in multiple studies. This was achieved using data from eight publications (Fig. 1), including five peer-reviewed publications (Meschersky et al., 2008, 2013, 2018; O’Corry-Crowe et al., 2010, 2018), two conference papers (Meschersky et al., 2012; Yazykova et al., 2012) and a Ph.D. thesis (Postma, 2017). Information on haplotype frequencies within stocks and DNA sequence accession numbers from Meschersky et al. (2008, 2012, 2013, 2018) and Yazykova et al. (2012) were kindly provided by the authors. Two publications (O’Corry-Crowe et al., 2010, 2018) presented tables of haplotype frequencies and haplotype GenBank accession numbers, from which we reconstructed the original datasets. Haplotype frequencies from Postma (2017) were presented in the thesis, and the corresponding sequences were provided by the author. Sequences were assigned to the 21 GROM stocks using sample site information from the original publications. In cases where the same individuals were analyzed in multiple studies, only the sequence from the most recent publication was used. Sequences collected outside recognized GROM stocks, e.g., from migrating individuals or individuals with insufficient information, were omitted from further analysis.
mtDNA Data Analysis

The mtDNA sequences were aligned using ClustaW (Larking et al., 2007) with default settings and trimmed in the 5´ and 3´ end to only include genetic regions covered by all individuals. This resulted in a final dataset of 2,933 individuals with 302 bp of overlapping control region sequence. Haplotypes were named GROM_XX, with XX representing a two-digit sequential number starting at 01 (Supplementary Table 1). Due to the reduced sequence length, some previously unique haplotypes collapsed into the same GROM_XX haplotype (Supplementary Table 2). To visualize the relationships among mtDNA haplotypes, we used the sequences alignment to construct a median spanning haplotype network (Bandelt et al., 1999) using Popart 1.7 (avail. from: http://popart.otago.ac.nz).

We used two parameters to estimate levels of diversity within stocks (H and \( \pi \)), both calculated using Arlequin (Excoffier & Lischer, 2010). These estimators have been widely used in beluga genetics, and report levels of diversity in different ways. Haplotype diversity (H) delimits the proportion of unique haplotypes within each stock, whereas nucleotide diversity (\( \pi \)) uses the average number of nucleotide differences between all possible pairs of sequences within each stock. Thus, stocks with a high proportion of unique haplotypes will have a high level of H, without consideration of how different the haplotypes are, while stocks with many variable sites will have a high level of \( \pi \), ignoring the number of unique haplotypes. Levels of differentiation among all pairs of stocks was estimated by the fixation index \( F_{ST} \) using Arlequin 3.5 (Excoffier & Lischer, 2010). Levels of significance were corrected for multiple testing by dividing 0.05 by the number of tests (210), yielding a corrected \( p \) value of 0.00024.
Chapter 1

To assess the broad-scale, macro-geographic patterns of mtDNA diversity across the species range, and to assist readers less familiar with the Arctic geography, we divided the 21 stocks into four geographic regions ad hoc (Fig. 1):

(1) Sea of Okhotsk, including Sakhalin-Amur, Ulbansky Bay, Tugursky Bay, Udskaya Bay and Shelikov Bay stocks; these stocks all exclusively inhabit the Sea of Okhotsk during winter.

(2) Pacific, including Anadyr Estuary, Cook Inlet, Bristol Bay, Eastern Bering Sea, Eastern Chukchi Sea and Eastern Beaufort Sea stocks, which all inhabit the Bering Sea and adjacent waters during winter.

(3) Western Atlantic, including Eastern High Arctic, Western Hudson Bay, James Bay, Eastern Hudson Bay, Ungava Bay, Cumberland Sound and St. Lawrence Estuary stocks, which spend winters in Baffin Bay, Labrador Sea and Hudson Strait, or remain resident in adjacent bays.

(4) Eastern Atlantic, including Svalbard, Barents, Kara, and Laptev Seas, and White Sea stocks. These three stocks are assumed to inhabit separate geographic regions during winter, but are all located in waters connected to the north-eastern Atlantic Ocean.

Results

Literature Review

We retrieved 27 beluga publications with population-level genetic data (Table 1). In total, > 4,500 beluga individuals have been sequenced for part of their mitochondrial genome. Twenty-one publications are based on the analysis of mtDNA sequences, while 13 of the 27 publications include microsatellite data. Ten publications analyze both mtDNA sequences and microsatellite data (Table 1).

Our literature review revealed a regional publication bias, with some geographic regions more intensely studied than others. Only three of the 27 publications identified in this review include
samples from the Eastern Atlantic region (Fig. 1), by far the least-studied region. The Sea of Okhotsk is included in seven publications. The Pacific and Western Atlantic regions are the most studied regions, and have been included in 16 and 14 publications, respectively. The Eastern Atlantic region is covered by fewer genetic studies, although the Svalbard and White Sea stocks have been investigated using other non-genetic methodologies including telemetry and aerial surveys (Lydersen et al., 2001; Glazov et al., 2010). The Barents, Kara, and Laptev Seas stock has only been analyzed genetically in Meschersky et al. (2018) (NAMMCO, 2018).

Microsatellites

Microsatellites have played an important role in providing valuable information regarding the subdivision of beluga stocks (Brown Gladden et al., 1999; de March & Postma, 2003; O’Corry-Crowe et al., 2010, 2015; Turgeon et al., 2012; Meschersky et al., 2013). Microsatellites are long repeats of short motifs (<10 bp sequences) that can be found throughout the nuclear genome. The variation in microsatellites is captured by the length of the repetitive sequences (allele size), reflecting the varying number of times the short motifs are repeated (Vieira et al., 2016).

Microsatellites evolve at an elevated rate compared to other nuclear markers making them suitable for investigations of recent evolutionary changes (Ellegren, 2004). They have been used in local and regional studies to investigate kinship in pods (Colbeck et al., 2013), genetic mark-recapture (Citta et al., 2018), stock assignment (O’Corry-Crowe et al., 2018) and impacts of changing sea ice conditions (O’Corry-Crowe et al., 2016). However, in a pan-Arctic context the methodology is limited by the fact that allele sizes are influenced by lab-specific practices, and the data generated by different labs are therefore not directly comparable, unless they are calibrated by one or more reference samples. These difficulties have been overcome in work on other species by the inclusion of reference samples by all labs that contribute to the compiled data (Ellis et al., 2011). However, this requires coordinated efforts among research groups prior to the onset of the work, and is perhaps a more
realistic approach for the investigation of commercially important species. Nevertheless, to provide an overview of the microsatellite literature, we compiled an exhaustive list of the studies published to date, and present details of the specific microsatellite markers used in each study in Table 2.

mtDNA Data Analysis

Aligning and trimming the mtDNA sequences resulted in a 302 bp overlapping region (Fig. 1 insert), representing 2,933 individuals and all 21 GROM stocks (Supplementary Table 1). Sample sizes of individual stocks ranged from 22 to 579. Our dataset represented 71 haplotypes defined by 16 variable sites. The median-spanning network revealed two major haplogroups differentiated by four variable sites (Fig. 2). Haplogroup A harbors samples from all four geographic regions, while Haplogroup B only includes samples from the Western Atlantic region. Within Haplogroup A, all haplotypes are differentiated from neighboring haplotypes by only one variable site. This is also the case for all haplotypes within haplogroup B, except for haplotypes found in the St. Lawrence Estuary, which are separated from the other haplogroup B haplotypes by two variable sites (Fig. 2).

Haplotypes GROM_01, GROM_02 and GROM_23 are found at the highest frequencies in our trimmed dataset and are the results of 43, 11 and 7 collapsed haplotypes, respectively (Supplementary Table 1). GROM_01 is found in 809 individuals from 19 stocks, GROM_02 is found in 365 individuals from ten stocks and GROM_23 is found in 461 individuals from 11 stocks. Twenty-four haplotypes are found in less than five individuals.

Levels of haplotype diversity (H) across the 21 beluga stocks range from 0.19 in St. Lawrence Estuary and Bristol Bay, to 0.87 in Barents, Kara, and Laptev Seas (Fig. 3a). Nucleotide diversity (\(\pi\)) ranges from 0.001 in Ulbansky Bay to 0.014 in Eastern Hudson Bay (Fig. 3b). Five stocks have \(\pi < 0.002\) (White Sea, Bristol Bay, Ulbansky Bay, Shelikov Bay and Anadyr Estuary), and two stocks have
The fixation index $F_{ST}$ shows large variation and ranges from 0.00 to 0.93 (Table 3). All but five $F_{ST}$ values are significant. Pairwise comparisons between Tugursky Bay and Udskaya Bay, Ulbansky Bay and Bristol Bay, Ungava Bay and Western Hudson Bay, Ulbansky Bay and Western Hudson Bay and Svalbard and Anadyr are not significant after Bonferroni correction for multiple testing ($\alpha = 0.00024$).

**Discussion**

The mitochondrial genome is a non-recombining, independently evolving genome, which is passed on from mother to offspring (Hutchison et al., 1974). The development of easy-to-use, highly conserved primers that could amplify targeted mtDNA fragments, coupled with high levels of intraspecific diversity, made mtDNA control region the marker of choice in population genetics and phylogeography (Kocher et al., 1989; Ballard & Whitlock, 2004). The analysis of mtDNA has been applied frequently in beluga research, and has revealed high levels of diversity and differentiation across the belugas’ range, establishing the marker as a valuable tool in beluga conservation and management (Brown Gladden et al., 1997; O’Corry-Crowe et al., 1997, 2002; Palsbøll et al., 2002; Meschersky et al., 2008, 2013, 2018). However, mtDNA analysis has inherent limitations (Balloux, 2010). The lack of recombination and the small size of the mitochondrial genome (16,386 bp (Kim et al., 2017)) makes it susceptible to stochastic events, and intraspecific patterns of mtDNA variability may not accurately reflect the evolutionary history of a species (Toews & Breisford, 2012). Further, the maternal inheritance means that only the female lineage is reflected in mtDNA. Nevertheless, as long as these limitations are recognized, mtDNA remains a valuable resource for investigating phylogeographic patterns across a species range.
Pan-Arctic Insights

To carry out a pan-Arctic meta analysis that allowed the direct comparison of stock diversity and differentiation across the beluga range, the complexity of the original data was necessarily reduced as our analysis included 302 bp of sequence alignment from 2,933 individuals and all 21 GROM stocks.

We found 71 haplotypes defined by 16 variable sites. Our haplotype network revealed two major haplogroups, A and B. Haplogroup A included individuals from 20 of the 21 GROM stocks, excluding St Lawrence Estuary, which were exclusively found in Haplogroup B (Fig. 2). Our results lack the genetic structure in Haplogroup A found in previous analyses of both 609 bp sequences (n=2,501) and complete mitochondrial genomes (n=106) from Canadian belugas (Postma, 2017). The longer sequences revealed two well-differentiated haplogroups in our Haplogroup A, highlighting the limited resolution of our 302 bp mtDNA dataset, and hence caution should be exercised when interpreting results.

Our diversity estimates show large variation among beluga stocks both in haplotype diversity (H), which counts the proportion of unique haplotypes in each stock, and nucleotide diversity (\(\pi\)), which describes how different the sequences are. The lowest H was found in Bristol Bay (0.19), St. Lawrence Estuary (0.19), Ulbansky Bay (0.22) and White Sea (0.28) (Fig. 3a). This is supported by corresponding low levels of \(\pi\) in these stocks (Bristol Bay and Ulbansky Bay (0.001), White Sea (0.002) (Fig. 3b)). However, we find relatively high \(\pi\) in St. Lawrence Estuary (0.004), which reflects that this stock has only two unique, but more diverse haplotypes (Fig. 2). The low levels of mtDNA diversity in the afore-mentioned beluga stocks could have arisen from several evolutionary processes, including (1) low levels of diversity of founding individuals, (2) insufficient time since establishment of the stock to accumulate new mutations, (3) low levels of females migrating into the
stock, (4) genetic drift - a stochastic mechanism removing rare haplotypes - which affects small populations to a higher degree than larger populations. The St. Lawrence Estuary stock is less than half the size of the three other low diversity stocks, indicating that genetic drift is to a larger degree affecting this stock (NAMMCO, 2018).

The Barents, Kara, and Laptev Seas stock has the highest level of $H$ (0.87) and $\pi$ (0.14) (Fig. 3). This may be due to the stock harboring multiple distinct, but as-yet undefined stocks, as has been suggested (NAMMCO, 2018). Substructuring within the stock would result in inflated estimates of diversity. This is likely, as the Barents, Kara, and Laptev Seas stock has an extensive geographic range (Fig. 1), yet has to date only been included in one genetic study (Meschersky et al., 2018).

All but five pairs of beluga stocks are significantly differentiated ($p < 0.00024$) indicating that this dataset of only 302 bp is in most cases able to separate the 21 stocks recognized by GROM (Table 3). The lack of significant differentiation between Western Hudson Bay and Ungava Bay is a result of shared haplotypes (Fig. 2). Although Western Hudson Bay has 15 haplotypes and Ungava Bay has 10 haplotypes, 201 out of 228 individuals (88%) found in the two stocks belong to seven shared haplotypes (Supplementary Table 1). Haplotype sharing may be explained by their close geographic proximity (Fig. 1), or reflect that individuals from Western Hudson Bay were sampled as they migrate through the Hudson Strait just outside Ungava Bay (Lewis et al., 2009; NAMMCO, 2018).

Tugursky Bay and Udskaya Bay, which are adjacent stocks in the Sea of Okhotsk, are not differentiated in our $F_{st}$ analysis; 95% of individuals from the two stocks belong to one of four shared haplotypes (Fig. 2, Supplementary Table 1). This could reflect that a single genetic unit is distributed across the two bays, which is further supported by similar $H$ and $\pi$ levels (Fig. 3). A previous study using longer mtDNA fragments and 19 microsatellite loci was also unable to differentiate between
Chapter 1

them (Yazykova et al., 2012). However, GROM evaluations are based on a composite of data, and Tugursky Bay and Udskaya Bay belugas use separate summering grounds and show behavioral differences towards motorized boats (NAMMCO, 2018).

Our analysis was unable to differentiate between several stocks that are otherwise separated by large geographic distances, likely reflecting the low resolution in our data. The pattern could also be explained by shared recent demographic histories, or the colonization of one area from the other, but we consider this less likely. For example, Ulbansky Bay and Bristol Bay are separated by more than 3000 km (Fig. 1). Both stocks harbor haplotype GROM_01 at ~90% frequency (Supplementary Table 1), and additional haplotypes are closely related to GROM_01, resulting in low levels of both H and $\pi$ (Fig. 3). During the trimming of the sequence alignment to 302 bp, a lot of informative sites were lost; GROM_01 includes 43 collapsed haplotypes from the original studies (Supplementary Table 2). This could indicate that longer sequences might enable us to distinguish Ulbansky Bay and Bristol Bay, although when we analysed the available 410 bp mtDNA sequence shared by the two stocks, they remain indistinguishable.

Our $F_{ST}$ analysis supports that St. Lawrence Estuary is the most divergent, underscoring the uniqueness of this stock (Gladden et al., 1999; de March & Postma, 2003; Postma, 2017). This is in agreement with the haplotype network in which the two St. Lawrence haplotypes are not shared with any other stock, and are differentiated from their nearest neighbor by two variable sites (Fig. 2).

The analyses of 302 bp pan-Arctic sequences from 2,933 individuals reveal that, even with large sample sizes and range-wide sampling, short control region mtDNA fragments are a very limited genetic marker. In the light of this, short mtDNA markers may still have a role to play in future beluga genetics, as an initial exploratory tool to help form hypotheses for further testing. Full
mitochondrial genomes offer more information and higher resolution (Postma, 2017; Skovrind et al., 2017). Mitochondrial genome analyses have been used in a number of other whale species to make a wide variety of inferences, including historic female population size, detailed phylogeographic structuring and the existence of subspecies (Morin et al., 2010, 2018; Van Cise et al., 2019), illustrating the potential of using this genetic marker in beluga research. Although complete mitochondrial genomes will certainly add valuable insights, providing a clearer picture of the history and structuring of maternal lineages, it will not be able to elucidate the demography and history of both sexes, including levels of male migration and patterns of gene flow among stocks, which require insights from the nuclear genome. Hence, mtDNA should not be used as a stand-alone tool with which to differentiate stocks, but should be used, with caution, as a component in a multidisciplinary effort including other methodologies, e.g. telemetrics, aerial surveys and behavioral studies, where available, as has been applied in the GROM stock evaluation (NAMMCO, 2018). For example, behavior, social bonds, migration routes, site fidelity and possibly other traits might delineate stocks, even in the absence of genetic differentiation, as is the case for Tugursky Bay and Udskaya Bay in the Sea of Okhotsk.

**Perspectives**

The redistribution of belugas in the face of Arctic climate change is of high conservation concern, and will require cooperation and coordination across the Arctic range of the species (Laidre et al., 2015). Our review highlights that gaining genetic insights across the species range is a challenge with available data. This severely limits range-wide inferences, which is vital for understanding widely distributed species, such as the beluga. Facing this challenge, a natural next step in informing the management and conservation of belugas with genetics is the transition to genome-wide sequencing methodologies, which analyse tens of thousands to millions of genetic markers across the nuclear genome. This will enable high-resolution studies of levels of
differentiation and diversity, which will aid our understanding of current demographic processes in belugas. Future range-wide genetic beluga studies could (1) identify evolutionary lineages and their divergence times to better understand the evolutionary history of the species; (2) estimate gene flow between populations to determine the current connectivity between populations and enable tracking of any future changes in gene flow patterns; and (3) identify genes under selection in belugas to inform managers whether local populations have unique adaptations to their specific habitats. Such work will benefit from the recent release of a high-quality beluga reference nuclear genome (Jones et al., 2017), which offers a valuable resource for future genomic studies of the species (Allendorf, 2017). However, with new methods come new challenges. The transition from mtDNA and microsatellites to genome-wide sequencing will require specialized staff, large computational facilities and it is, despite falling prices over the last decades, still rather expensive compared to methodologies currently used in beluga genetics.

Acknowledgements

The research was funded by the Carlsberg Foundation Distinguished Associate Professor Fellowship, grant no CF16-0202. We would like to thank the participants of the GROM meeting that took place in Hillerød, Denmark, in March 2017, for helpful discussions on beluga stock identification. In addition, we would like to thank Ilya Meschersky and colleagues for providing accession numbers and haplotype frequencies for the Russian stocks. We would also like to thank the beluga research community, all authors of the original publications included in our analyses and indigenous hunters and communities who shared their local knowledge and helped sample many of the belugas included in this publication.

Literature cited

Chapter 1


Bandura, A. (1977) Social learning theory.


through Satellite Telemetry and Nunavik Traditional Ecological Knowledge. ARCTIC. 62: .


Murray W.B. 1997. Major histocompatibility complex class II sequence variation in cetaceans: DQ beta and DR beta variation in beluga (Delphinapterus leucas) and DQ beta variation in North .... .

NAMMCO 1999. Scientific committee working group on the population status of beluga and narwhal in the north atlantic. .


Postma L. 2017. Genetic diversity, population structure and phylogeography among belugas (Delphinapterus leucas) in Canadian waters: broad to fine-scale approaches to inform conservation and management strategies by Lianne D. Postma. The University of Manitoba,


Yazykova M.G., Meschersky I.G., Shpak O.V., Glazov D.M., Litovka D.I., Borisova E.A. & Rozhnov V.V.
Chapter 1

Chapter 1

Figures

Figure 1. Location of beluga stocks recognized by the Global Review of Monodontids (NAMMCO, 2018). 1) Sakhalin Bay, 2) Ulbansky Bay, 3) Tugursky Bay, 4) Udskaya Bay, 5) Shelikov Bay, 6) Anadyr Estuary, 7) Cook Inlet, 8) Bristol Bay, 9) Eastern Bering Sea, 10) Eastern Chukchi Sea, 11) Eastern Beaufort Sea, 12) Eastern High Arctic, 13) Western Hudson Bay, 14) James Bay, 15) Eastern Hudson Bay, 16) Ungava Bay, 17) Cumberland Sound, 18) St. Lawrence Estuary, 19) Svalbard, 20) Barents, Kara, and Laptev Seas, 21) White Sea. Insert map: Eastern Sea of Okhotsk. White insert shows the publications from where we retrieved mtDNA sequences for our circumpolar dataset, the genetic regions included in the original studies, and the position of the 302 bp sequence used in our analysis. Colors indicate the geographic range of the eight publications that contributed mtDNA sequences to our study. The geographic extent of the four ad-hoc regions discussed in the main text are indicated by black lines.
Chapter 1

Figure 2. Median-spanning network of 71 haplotypes across 2,933 mtDNA sequences from the 21 beluga stocks. Each haplotype is represented by a circle colored according to the stocks where the specific haplotype was found. Black dots indicate haplotypes not sampled in the data. Size of circles indicates relative haplotype frequency.
Figure 3. Levels of genetic diversity in the 21 beluga stocks. (a) Haplotype diversity ($H$), (b) nucleotide diversity ($\pi$).
### Tables

Table 1. Publications with population-level genetic data from belugas, including numbers of individuals analyzed and type of genetic marker used.

<table>
<thead>
<tr>
<th>Publication</th>
<th>Genetic marker (n)</th>
<th>Stock in present study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patenaude et al 1994</td>
<td>DNA fingerprints</td>
<td></td>
</tr>
<tr>
<td>Murray et al 1995</td>
<td>Major Histocompatibility Complex (233)</td>
<td></td>
</tr>
<tr>
<td>Buchanan et al 1996</td>
<td>Microsatellites (100)</td>
<td></td>
</tr>
<tr>
<td>O’Corry-Crowe et al 1997</td>
<td>mtDNA (324)</td>
<td></td>
</tr>
<tr>
<td>Brennin et al 1997</td>
<td>RFLP (95)</td>
<td></td>
</tr>
<tr>
<td>Brown Gladden et al 1997</td>
<td>mtDNA (628)</td>
<td></td>
</tr>
<tr>
<td>Brown Gladden et al 1999</td>
<td>Microsatellites (640)</td>
<td></td>
</tr>
<tr>
<td>O’Corry-Crowe et al 2002</td>
<td>mtDNA (427)</td>
<td></td>
</tr>
<tr>
<td>Palsbøll et al 2002</td>
<td>mtDNA (195)</td>
<td></td>
</tr>
<tr>
<td>de March et al 2002</td>
<td>mtDNA (481)</td>
<td></td>
</tr>
<tr>
<td>de March et al 2003</td>
<td>mtDNA (714)</td>
<td></td>
</tr>
<tr>
<td>Meschersky et al 2008</td>
<td>mtDNA (28)</td>
<td></td>
</tr>
<tr>
<td>O’Corry-Crowe et al 2010</td>
<td>mtDNA (122)</td>
<td>White Sea</td>
</tr>
<tr>
<td>Meschersky et al 2013</td>
<td>mtDNA (211)</td>
<td>Anadyr and Shelikhov</td>
</tr>
<tr>
<td>Colbeck et al 2012</td>
<td>mtDNA (256)</td>
<td></td>
</tr>
<tr>
<td>Meschersky et al 2013</td>
<td>mtDNA (211)</td>
<td>Anadyr and Shelikhov</td>
</tr>
<tr>
<td>Colbeck et al 2012</td>
<td>Microsatellites (1524)</td>
<td></td>
</tr>
<tr>
<td>Meschersky et al 2013</td>
<td>Microsatellites (175)</td>
<td></td>
</tr>
<tr>
<td>Meschersky et al 2015</td>
<td>mtDNA (10)</td>
<td></td>
</tr>
<tr>
<td>Meschersky et al 2016</td>
<td>mtDNA (978)</td>
<td></td>
</tr>
<tr>
<td>Skovrind et al 2017</td>
<td>mtDNA (2)</td>
<td></td>
</tr>
<tr>
<td>Glazov et al 2019 (In this issue)</td>
<td>mtDNA</td>
<td>Barents, Kara, and Laptev Seas</td>
</tr>
<tr>
<td>Meschersky et al 2018</td>
<td>mtDNA (58)</td>
<td>White Sea</td>
</tr>
<tr>
<td>Citta et al 2018</td>
<td>Microsatellites (516)</td>
<td></td>
</tr>
<tr>
<td>Shpak et al 2019 (In this issue)</td>
<td>mtDNA (488)</td>
<td>Sakhalin-Amur, Ulbansky Bay, Tugursky Bay and Udskaya Bay</td>
</tr>
</tbody>
</table>
Table 2. Publications with population-level microsatellite data from belugas, including number of individuals and loci analyzed.

<table>
<thead>
<tr>
<th>Publication</th>
<th>n</th>
<th>Number of loci</th>
<th>Markers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buchanan et al 1996</td>
<td>100</td>
<td>15</td>
<td>DirFCB1, DirFCB2, DirFCB3, DirFCB4, DirFCB5, DirFCB6, DirFCB7, DirFCB8, DirFCB10, DirFCB11, DirFCB12, DirFCB13, DirFCB14, DirFCB16, DirFCB17</td>
</tr>
<tr>
<td>Brown Gladden et al 1999</td>
<td>640</td>
<td>5</td>
<td>DirFCB4b, DirFCB5b, 464/465-1, EV37Mna, EV94Mnb</td>
</tr>
<tr>
<td>de March et al 2002</td>
<td>458</td>
<td>15</td>
<td>DirFCB1, DirFCB2, DirFCB3, DirFCB4, DirFCB5, DirFCB8, DirFCB10, DirFCB11, DirFCB13, DirFCB14, DirFCB16, DirFCB17, Gme464/465, MnoEV37Mn, MnoEV94Mn</td>
</tr>
<tr>
<td>de March et al 2003</td>
<td>555</td>
<td>15</td>
<td>DirFCB1, DirFCB2, DirFCB3, DirFCB4, DirFCB5, DirFCB8, DirFCB10, DirFCB11, DirFCB13, DirFCB14, DirFCB16, DirFCB17, Gme464/465, MnoEV37Mn, MnoEV94Mn</td>
</tr>
<tr>
<td>O’Corry-Crowe et al 2010</td>
<td>83</td>
<td>8</td>
<td>CS415, CS417, CS468, EV37, EV94, DirFCB3, DirFCB5, DirFCB17</td>
</tr>
<tr>
<td>Turgeon et al 2012</td>
<td>1605</td>
<td>13</td>
<td>EV37, EV94, FCB1, FCB2, FCB3, FCB4, FCB5, FCB8, FCB10, FCB11, FCB13, FCB14, FCB17</td>
</tr>
<tr>
<td>Yazykova et al 2012</td>
<td>209</td>
<td>19</td>
<td>DirFCB1, DirFCB2, DirFCB3, DirFCB4, DirFCB5, DirFCB6, DirFCB8, DirFCB10, DirFCB11, DirFCB13, DirFCB14, DirFCB16, DirFCB17, EV37Mn, EV94Mn, 415/416, 417/418, 464/465, 468/469</td>
</tr>
<tr>
<td>Colbeck et al 2012</td>
<td>1524</td>
<td>13</td>
<td>EV37, EV94, FCB1, FCB2, FCB3, FCB4, FCB5, FCB8, FCB10, FCB11, FCB13, FCB14, FCB17</td>
</tr>
<tr>
<td>Borisova et al 2012</td>
<td>161</td>
<td>8</td>
<td>DirFCB4, DirFCB5, DirFCB17, EV94Mn, 415/416, 417/418, 464/465, 468/469</td>
</tr>
<tr>
<td>Meschersky et al 2013</td>
<td>175</td>
<td>9</td>
<td>DirFCB4, DirFCB5, DirFCB17, EV37Mn, EV94Mn, 415/416, 417/418, 464/465, 468/469</td>
</tr>
<tr>
<td>O’Corry-Crowe et al 2015</td>
<td>10</td>
<td>8</td>
<td>CS415, CS417, CS468, EV37, EV94, DirFCB3, DirFCB5, DirFCB17</td>
</tr>
<tr>
<td>O’Corry-Crowe et al 2016</td>
<td>816</td>
<td>8</td>
<td>CS415, CS417, CS468, EV37, EV94, DirFCB3, DirFCB5, DirFCB17</td>
</tr>
<tr>
<td>O’Corry-Crowe et al 2018</td>
<td>1116</td>
<td>8</td>
<td>CS415, CS417, CS468, EV37, EV94, DirFCB3, DirFCB5, DirFCB17</td>
</tr>
<tr>
<td>Citta et al 2018</td>
<td>516</td>
<td>7</td>
<td>CS415, CS417, EV37, EV94, DirFCB3, DirFCB5, DirFCB17</td>
</tr>
</tbody>
</table>
Table 3. Levels of differentiation $F_{ST}$ between the 21 beluga stocks recognized by GROM (NAMMCO, 2018). Values in bold are insignificant with $p > 0.00024$.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sakhalin-Amur</td>
<td>0.14</td>
<td>0.15</td>
<td>0.16</td>
<td>0.46</td>
<td>0.55</td>
<td>0.46</td>
<td>0.18</td>
<td>0.17</td>
<td>0.18</td>
<td>0.48</td>
<td>0.29</td>
<td>0.14</td>
<td>0.61</td>
<td>0.55</td>
<td>0.13</td>
<td>0.10</td>
<td>0.86</td>
<td>0.54</td>
<td>0.27</td>
<td>0.61</td>
</tr>
<tr>
<td>Ulbanskiy Bay</td>
<td>0.16</td>
<td>0.17</td>
<td>0.70</td>
<td>0.68</td>
<td>0.51</td>
<td>0.01</td>
<td>0.05</td>
<td>0.22</td>
<td>0.49</td>
<td>0.26</td>
<td>0.03</td>
<td>0.41</td>
<td>0.52</td>
<td>0.03</td>
<td>0.07</td>
<td>0.92</td>
<td>0.72</td>
<td>0.36</td>
<td>0.43</td>
<td></td>
</tr>
<tr>
<td>Tugurskiy Bay</td>
<td>0.00</td>
<td>0.33</td>
<td>0.53</td>
<td>0.39</td>
<td>0.25</td>
<td>0.11</td>
<td>0.18</td>
<td>0.40</td>
<td>0.23</td>
<td>0.10</td>
<td>0.27</td>
<td>0.45</td>
<td>0.13</td>
<td>0.10</td>
<td>0.86</td>
<td>0.49</td>
<td>0.13</td>
<td>0.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Udskaya Bay</td>
<td>0.00</td>
<td>0.25</td>
<td>0.25</td>
<td>0.15</td>
<td>0.19</td>
<td>0.41</td>
<td>0.26</td>
<td>0.14</td>
<td>0.31</td>
<td>0.48</td>
<td>0.16</td>
<td>0.33</td>
<td>0.40</td>
<td>0.47</td>
<td>0.17</td>
<td>0.26</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sletner Bay</td>
<td>0.1</td>
<td>0.62</td>
<td>0.43</td>
<td>0.74</td>
<td>0.47</td>
<td>0.32</td>
<td>0.42</td>
<td>0.45</td>
<td>0.44</td>
<td>0.44</td>
<td>0.28</td>
<td>0.47</td>
<td>0.56</td>
<td>0.35</td>
<td>0.47</td>
<td>0.29</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anadyr Estuary</td>
<td>0.53</td>
<td>0.72</td>
<td>0.52</td>
<td>0.38</td>
<td>0.13</td>
<td>0.75</td>
<td>0.50</td>
<td>0.11</td>
<td>0.56</td>
<td>0.60</td>
<td>0.47</td>
<td>0.86</td>
<td>0.04</td>
<td>0.23</td>
<td>0.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cook Inlet</td>
<td>0.2</td>
<td>0.27</td>
<td>0.17</td>
<td>0.86</td>
<td>0.46</td>
<td>0.44</td>
<td>0.04</td>
<td>0.31</td>
<td>0.72</td>
<td>0.56</td>
<td>0.13</td>
<td>0.33</td>
<td>0.44</td>
<td>0.37</td>
<td>0.16</td>
<td>0.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bristol Bay</td>
<td>0.18</td>
<td>0.42</td>
<td>0.20</td>
<td>0.06</td>
<td>0.39</td>
<td>0.56</td>
<td>0.06</td>
<td>0.07</td>
<td>0.88</td>
<td>0.54</td>
<td>0.19</td>
<td>0.61</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eastern Bering Sea</td>
<td>0.33</td>
<td>0.27</td>
<td>0.21</td>
<td>0.34</td>
<td>0.56</td>
<td>0.21</td>
<td>0.33</td>
<td>0.85</td>
<td>0.38</td>
<td>0.21</td>
<td>0.45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eastern Chukchi Sea</td>
<td>0.18</td>
<td>0.43</td>
<td>0.10</td>
<td>0.59</td>
<td>0.46</td>
<td>0.40</td>
<td>0.81</td>
<td>0.04</td>
<td>0.16</td>
<td>0.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eastern Beaufort Sea</td>
<td>0.21</td>
<td>0.21</td>
<td>0.56</td>
<td>0.22</td>
<td>0.21</td>
<td>0.56</td>
<td>0.28</td>
<td>0.07</td>
<td>0.33</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eastern High Arctic</td>
<td>0.36</td>
<td>0.53</td>
<td>0.01</td>
<td>0.05</td>
<td>0.85</td>
<td>0.50</td>
<td>0.18</td>
<td>0.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Western Hudson Bay</td>
<td>0.35</td>
<td>0.34</td>
<td>0.36</td>
<td>0.73</td>
<td>0.66</td>
<td>0.08</td>
<td>0.12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>James Bay</td>
<td>0.49</td>
<td>0.53</td>
<td>0.43</td>
<td>0.47</td>
<td>0.41</td>
<td>0.53</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Western Hudson Bay</td>
<td>0.04</td>
<td>0.48</td>
<td>0.58</td>
<td>0.23</td>
<td>0.70</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ungava Bay</td>
<td>0.06</td>
<td>0.46</td>
<td>0.18</td>
<td>0.53</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cumberland Sound</td>
<td>0.85</td>
<td>0.84</td>
<td>0.80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S.L. Estuary</td>
<td>0.85</td>
<td>0.84</td>
<td>0.80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Swabard</td>
<td>0.19</td>
<td>0.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Barents, Kara</td>
<td>0.37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>and Laptev Seas</td>
<td>0.37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White Sea</td>
<td>0.37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Supplementary Items

Supplementary Table 1. Haplotypes found in our meta dataset and their frequencies in each of the 21 stocks recognized by GROM (NAMMCO, 2018). Provided as .xlsx file.

Supplementary Table 2. GROM_XX haplotype numbers and the corresponding NCBI accession numbers of sequences included in our 302 bp dataset. Provided as .xlsx file.

Supplementary Items are available from the Electronic Research Data Archive (ERDA):
https://sid.erda.dk/wsgi-bin/ls.py?share_id=fUhpSrLuLT
Making the best of it

When I first started working on this project there were no beluga reference genome available. Not even the mitogenome had been published. So our initial plans was to sequence, assemble and publish a genome. When the first lanes of data were sequenced, we agreed to publish the mitogenome first, while we were waiting for more sequencing and the assembly. This would give my supervisor and I an opportunity to work together on a publication and get familiar with collaborating. We agreed to publish it in a phylogenetic tree with the ~40 other published odontocete mitogenomes. A large number of mitogenomes had recently been published and a new phylogenetic mitogenome of odontocetes were due. I collected all the sequences and was working on the alignments, partitions and model selection when I came into the office one day and saw that we had been scooped. Another group had published the beluga mitogenome. I must admit I was a bit gutted, but I still had high hopes for the nuclear genome (which was also scooped, but that is another story). After realising that the published mitogenome came from the Sea of Okhotsk in the complete opposite end of the beluga range than the one we had sequenced, which came from West Greenland, we decided to tell a story of divergence time. We had very limited data (n=2) and the methodology was rather crude, but we managed to get a rough estimate of when the two individual maternal lineages split from each other.

*Adults and juveniles.* Credit: Carsten Egevang and Rikke Guldberg Hansen.
Mitochondrial genome divergence between beluga whales in Baffin Bay and the Sea of Okhotsk

Mikkel Skovrind*, Jose Alfredo Samaniego Castruita, Mads Peter Heide-Jørgensenb, Love Dalén and Eline Lorenzenab

aNatural History Museum of Denmark, University of Copenhagen, Copenhagen K, Denmark; bGreenland Institute of Natural Resources, Nuuk, Greenland; cDepartment of Bioinformatics and Genetics, Swedish Museum of Natural History, Stockholm, Sweden

ABSTRACT
The beluga whale is one of three endemic Arctic whales. The species is philopatric, and its migration patterns are passed from mother to calf. Management of the species is informed by the levels of genetic structuring among summer aggregation sites based on mitochondrial D-Loop data. To assess the levels of differentiation across the entire mitochondrial genome within belugas, we present a comparison between the first two complete mitochondrial genomes from opposite sides of their distribution range: Baffin Bay and the Russian Far East. Our analyses reveal that additional phylogenetic insights can be gained from expanding the genetic region analyzed. Further, we estimate the divergence time between the two mitochondrial genomes to be 0.469 MYA.

Here, we present the complete mitochondrial genome of a beluga whale (white whale, Delphinapterus leucas) from Baffin Bay. The beluga whale is a toothed whale belonging to the family Monodontidae, which also includes the narwhal Monodon monoceros. It has a discontinuous circumpolar distribution and is endemic to the Arctic region. Individuals can be up to 6.7 m long and are easily recognized by their white skin (Stewart & Stewart 1989). Mitochondrial DNA sequence information is of particular importance in the conservation of the beluga whale; the species is philopatric and does not recolonize summer aggregation sites from which it has been extirpated (Brown Gladden et al. 1997). Previous studies have reported higher levels of genetic differentiation among summer aggregation sites based on mitochondrial D-Loop data than on up to 15 nuclear microsatellite markers (De March & Postma 2003). In light of this, other mitochondrial regions could be used to augment the 400–600 base pairs currently used for management purposes (O’Corry-Crowe et al. 2002; Turgeon et al. 2012; Meschersky et al. 2013).

A beluga whale reference mitochondrial genome from the Sea of Okhotsk, Russia, was recently published (Kim et al. 2017). The mitochondrial genome presented here is from Baffin Bay, and therefore represents the opposite end of the distribution range of the species. The analysis of this specimen may identify mitochondrial genomic regions of interest for further studies. The individual was sampled during the Inuit subsistence hunt in Qeqertarsuaq (69.237835, −53.526519) in western Greenland in April 2008 by staff from Greenland Institute of Natural Resources. The tissue sample is stored at the Natural History Museum of Denmark (ID number CGG_1_017647). The sequence is available from GenBank under accession number KY888944.

We extracted DNA from skin tissue using the Kingfisher Duo extraction robot and Cell and Tissue DNA Kit from ThermoFisher Scientific using the manufacturer’s protocol (ThermoFisher Scientific, Waltham, MA). Paired-end sequencing was performed on 180 base pair inserts using the Illumina HiSeq X platform (San Diego, CA). We assembled the mitogenome using a combination of MIRA 4.0.2 (Chevreux et al. 2017) and MITObim v.1.8 (Hahn et al. 2013), using the narwhal reference genome as a reference (Genbank accession: NC005279). We performed the annotation using the MITOS web service (Bernt et al. 2013) using default parameters. We performed a phylogenetic analysis using the 13 protein-coding regions across six closely-related toothed whale species and the beluga reference mitochondrial genome (Kim et al. 2017). The phylogenetic tree was constructed from the best tree among 100 independent runs in RAxML v.8.2 (Stamatakis 2014) using a GTR-GAMMA substitution model (Figure 1). Topology was verified by 100 bootstraps. The number of variable sites in each genomic region was calculated using FaBox 1.4.1 (Villesen 2007). Divergence time (T) between the two mitochondrial genomes was calculated as $T = K/(2i)$, using a mutation rate (i) estimated for killer whales of $2.60 \times 10^{-3}$ substitutions per site per million years (1.50–3.83 $\times 10^{-3}$) (Morin et al. 2008) and a pairwise differentiation (K) of 2.44 $\times 10^{-3}$ calculated as the number of variable sites per site.

Our assembly yielded a 16,386 base-pair circular genome with gene regions in accordance with the findings of Kim et al. 2017.
et al. (2017). We found a total of 40 variable sites between the two beluga mitochondrial genomes, with 4, 2, 6 and 28 variable sites in the tRNA, rRNA, D-loop and protein-coding regions, respectively. The rRNAs had one variable site each and the protein-coding regions had between 1 and 5 variable sites (ND1:1, ND2:5, COX1:3, COX2:2, ATP8:1, ATP6:1, COX3:2, ND3:1 ND4L:1, ND4:2, ND5:4, ND6:3, CYTB:2), indicating that additional phylogenetic information is available in the mitochondrial genome relative to the D-Loop. We estimated a divergence time between the two mitochondrial lineages of 0.469 (0.319–0.814) MYA, which is the same order of magnitude as the divergence time estimation of killer whale lineages (Morin et al. 2008).

Acknowledgements

The authors acknowledge support from Science for Life Laboratory, the Knut and Alice Wallenberg Foundation, the National Genomics Infrastructure funded by the Swedish Research Council, and Uppsala Multidisciplinary Center for Advanced Computational Science for assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure.

Disclosure statement

The authors report no conflicts of interest and are alone responsible for the content and writing of the paper.

Funding

The study was funded by the Carlsberg Foundation Distinguished Associate Professor Fellowship grant no. CF16-0202.

ORCID

Mikkelskovrind et

References

De March BGE, Postma LD. 2003. Molecular genetic stock discrimination of belugas (Delphinapterus leucas) hunted in eastern Hudson Bay, northern Quebec, Hudson Strait, and Sianikulaq (Belcher Islands), Canada, and comparisons to adjacent populations. Arctic. 56:111–124.
A work in progress

This manuscript represents a large collaborative effort, where beluga researchers from across the range agreed to provide samples and join forces to create a range-wide dataset. The samples were all sent to the Natural History Museum of Denmark, where they were made ready for sequencing. When we received the sequencing data we quickly realised that the mitogenomes would most likely drown in a publication also including the nuclear data. We felt that this would be a shame as mitogenomes, despite limitations, can identify signs of old divergence events, which has been lost in the nuclear genome due to subsequent admixture. It was therefore decided that we would try to publish the mitogenomes separately, accompanied by a few nuclear genomes and habitat size estimations. Thus far, we have performed multiple analyses, including phylogenetic trees, demographic reconstructions and predicted habitat models. However, this manuscript is still very much a work in progress.

*Belugas along the ice fringe.* Credit: Carsten Egevang and Rikke Guldborg Hansen
Chapter 3

Habitat fragmentation and secondary contact shape phylogeography and demographic history of belugas

Authors in no particular order

Mikkel Skovrind 1 *, Marie Louis 1, Michael V. Westbury 1, José Alfredo Samaniego Castruita 1, Shyam Gopalakrishnan 1, Steen W. Knudsen 1, James S. Haile 1, Mads Peter Heide-Jørgensen 1,2, Dmitriy M. Glazov 3, Ilya G. Meschersky 3, Viatcheslav V Rozhnov 3, Olga V. Shpak 3,4, Karina K. Tarasyan 3, Dennis I. Litovka 5, Anton D. Chernetsky 6, Vera V. Krasnova 6, V. M. Bel’kovich † 6, Steve Ferguson 7, Lianne Postma 7, Kit Kovacs 8, Christian Lydersen 8, Kristin Kaschner 9, Cristina Garilao 10, Love Dalén 11, Eline D. Lorenzen 1

Affiliations

1: Natural History Museum of Denmark, University of Copenhagen
2 Greenland Institute of Natural Resources
3: A.N. Severtsov Institute of Ecology and Evolution RAS
4: Marine Mammal Council
5: Office of Governor and Government of the Chukotka Autonomous Okrug
6: P.P. Shirshov Institute of Oceanology RAS
7: Fisheries and Oceans Canada
8: Norwegian Polar Institute
9: Department of Biometry and Environmental System Analysis, Albert-Ludwigs-University of Freiburg
10: GEOMAR Helmholtz-Zentrum für Ozeanforschung Kiel
11: Department of Molecular Systematics, Swedish Museum of Natural History

*: Corresponding author
Abstract

Arctic species are predicted to be negatively impacted by rising temperatures and sea ice loss. However, the response of the whales endemic to the Arctic is less certain; they are predicted to face changes in prey availability, increased competition from temperate species and increased predation from killer whales, but they will also gain access to new habitat as the sea ice disappears. To understand the resilience of Arctic whales to future climatic perturbations, we investigate how beluga whales - one of only three whale species found in the Arctic year-round - responded to past climatic fluctuations. We analyze two nuclear genomes and 202 mitochondrial genomes, sampled across the circumpolar distribution range of the species, to investigate their phylogeography and elucidate the role of past climate change in shaping the evolutionary history of the species. We combine demographic reconstructions based on genomic data with estimates of past and present suitable habitat size using species distribution models to investigate range and habitat shifts over time. We find high levels of mitogenome variability within belugas, with up to 162 variable sites differentiating individual sequences. Our results show four well-differentiated mitochondrial lineages that were established before the onset of the last glacial expansion ~115 kya. We argue that lineages diverged due to the isolation of populations during colder climatic periods, when permanent sea ice and the Bering land bridge separated the Pacific and Atlantic oceans, and the Arctic ocean was covered with permanent sea ice. Current population structure is a consequence of secondary contact between these divergent lineages. Our demographic reconstruction reveal that female effective population size has increased six-fold since the Last Glacial Maximum, which we argue is driven by an increase in suitable habitat.

Introduction

Sea ice loss and Arctic marine mammals

Eleven species of marine mammals are endemic to the Arctic, including three whale species (beluga *Delphinapterus leucas*, narwhal *Monodon monoceros* and bowhead whale *Balaena mysticetus*). Polar bears *Ursus maritimus* and pinnipeds *Pinnipedia* (walruses and seals) are closely associated with sea ice, which they rely on for vital life stages, including feeding, hauling, mating and rearing of their young (Gilg et al. 2012). As a consequence, these species are predicted to be directly negatively affected by future sea ice loss, limiting both their distribution ranges and population sizes (Kovacs et al. 2011; Laidre et al. 2015). All three species of Arctic whales exhibit interspecific variation
in their association with sea ice, but in general all three species spend the winter close to the ice fringe or within the pack ice, where the sea ice provides protection from killer whale *Orcinus orca* predation and offers access to prey (Ferguson et al. 2010; Citta et al. 2017; Heide-Jørgensen 2018). During the warmer summer months, belugas most commonly inhabit shallow ice free estuaries, bays and inlets (Shpak et al. 2010; Hauser et al. 2014; G. M. O’Corry-Crowe 2018), whereas narwhals can be found in sea ice free conditions, often in fjords, moving between glacier fronts and deeper foraging areas (Laidre et al. 2004; Heide-Jørgensen 2018). In contrast, bowhead whales are able to feed and thrive in open water (Moore et al. 2010), although studies show they prefer to stay well within the pack ice (Ferguson et al. 2010). All three species have seasonal movements linked to changes in sea ice distribution and predictable ecological resources such as prey availability (O’Corry-Crowe 2018; Heide-Jørgensen 2018; Rugh and Shelden 2009). However, their association with sea ice has been suggested to be primarily indirect through the way it shapes the environment they inhabit (Kovacs et al. 2011; Gilg et al. 2012).

**Climate and distribution of species in the Arctic**

Changes in the geographic distribution of species over time is largely governed by shifts in suitable habitat (Krebs 1994). In the Arctic, the most powerful governing force is the distribution of permanent ice, which affects both terrestrial and marine faunas (*Hewitt 2000*). The beginning of the Quaternary 2.5 My was associated with the expansion of the Arctic ice sheets, and the period has since been characterized by major fluctuations in the extent of glaciers and sea ice (Jansen, Fronval, and Rack 2000; Polyak et al. 2010). These continuous expansions and retractions of ice sheets and sea ice has made the Arctic an ever-changing environment (Cronin and Cronin 2015), posing a challenge for all its inhabitants. Arctic whales rely on open water to breathe. Expanding permanent sea ice during colder climatic periods would have excluded them from large parts of the Arctic, prompting latitudinal habitat shifts. Seven glacials have occurred during the over the past 500 ky (Lisiecki and Raymo 2005). During these colder periods, the habitat of wide-ranging Arctic marine species was further fragmented by the Bering land bridge, which connected Siberia and Alaska (Gladenkov et al. 2002). The land bridge emerged when water levels dropped to 50 m below the current level (Miller et al. 2005). This happened during glacials, when glaciers and sea ice bound large amounts of water (Polyak et al. 2010, 2013).

Although Arctic climates have greatly fluctuated in the past, current climate change is happening at an accelerated pace, with permanent summer sea ice loss predicted by 2050 (J. Stroeve and Notz 2018). This will profoundly alter ecosystems across the planet, with the Arctic being
impacted most (J. C. Stroeve et al. 2012; Wang and Overland 2012; Post et al. 2013). Arctic whales are predicted to be primarily affected by sea ice loss, as it drives redistribution of prey, increased competition with temporal species entering Arctic waters, predation from killer whales expanding their distribution north, and exposure to new diseases (Moore and Huntington 2008; Kovacs et al. 2011; Laidre et al. 2015). However, Arctic whales will also gain access to new habitats uncovered by retreating sea ice, making predictions of resilience to ongoing climate change challenging.

Belugas

Belugas (or white whales, *Delphinapterus leucas*) are mid-sized toothed whales easily recognizable by their white skin and melon shaped forehead (Sergeant and Brodie 1969). They are found in both the Pacific and Atlantic Arctic with a disjunct circumpolar distribution (G. M. O’Corry-Crowe 2018). Belugas are important to indigenous communities across their range, where they are harvested through subsistence hunt and play a vital role in their culture (Dahl 1989; Kawagley 2006; Frost and Suydam 2010). Currently, 21 stocks associated with separate summering grounds are recognized (NAMMCO 2018), although belugas are found and also harvested by indigenous communities outside their summering grounds. Belugas have strong matriarchal site fidelity to these summering grounds, which has been shown to be stable over several decades (Caron and Smith 1990; G. O’Corry-Crowe et al. 2016). Some populations undertake long seasonal migrations to their wintering grounds in kin-structured pods, following predictable migration routes (Colbeck et al. 2013; G. O’Corry-Crowe et al. 2018). Populations that do not migrate have shorter seasonal movements, following the open water within their region (Hobbs et al. 2010; Lefebvre et al. 2012).

For decades, genetic data have been used to differentiate between beluga stocks, and to aid conservation efforts of the species and informing sustainable harvest levels (O’Corry-Crowe et al. 1997; De March and Postma 2003; Turgeon et al. 2012; Colbeck et al. 2013). Analyses of short mtDNA fragments have been used to differentiate summering grounds by differences in haplotype frequencies, but private haplotypes have rarely been identified (Turgeon et al. 2012; Meschersky et al. 2013; O’Corry-Crowe et al. 1997). In the Pacific region, the application of microsatellites has made population assignment possible (G. O’Corry-Crowe et al. 2018), but in the western Atlantic and Sea of Okhotsk populations this has not been possible (Brown Gladden et al. 1999; Yazykova et al. 2012). Genetic analyses have been applied in studies of many aspects of beluga population ecology, but exclusively at a local or regional level (Citta et al. 2018; G. O’Corry-Crowe et al. 2016), and the only
circumpolar attempt to understand beluga population structuring and diversity across the Arctic relied on the analysis of a short 302 bp fragment of mtDNA control region sequence (Chapter 1).

**Aims**

Here, we investigate the evolutionary and demographic history of belugas across its Arctic and sub-Arctic distribution range, and assess the impact of past climatic oscillations on the species. We analyse levels of contemporary genetic variation across 202 mitogenomes and two nuclear genomes to understand phylogeography of the species and reconstruct the demography trajectory of belugas over the past 250 thousand years. To further investigate the impact of past climate oscillations, we use estimates of environmental preferences to estimate the extent of suitable habitat during glacial and interglacial periods.

**Methods**

**Genetics**

**Samples**

Our study included tissue (primarily skin/blubber) samples from 203 belugas covering 23 localities across the species range (Supplementary Table 1). One sample (CGG_1_017647) was used to generate the nuclear genome, and the remaining 202 samples were used to generate mitochondrial genomes. Sample CGG_1_017647 was collected from Qeqertarsuaq, West Greenland. Belugas from this location are believed to be part of the Eastern High Arctic-Baffin Bay population (NAMMCO 2018). Belugas are more geographically separated during the summer months than during other seasons, and we therefore aimed to include samples taken during this season (June-August) for the mitogenome analysis. However, information on collection date, or even collection month, was not always available. For each locality, we aimed to include samples taken from different days and/or years to minimize the chance of sampling closely related individuals, and as far as possible we tried to secure equal representation of males and females. All samples were selected from national sample archives and were originally collected during indigenous subsistence hunts, as part of scientific monitoring programmes, or they were opportunistically collected from beached carcasses. Samples were shipped to Denmark under Cites exemption number DK03, except for Russian samples, which were shipped under CITES permit number IM1006-361/16. Detailed sample information can be found in Supplementary Table 1.
Data generation and bioinformatics

DNA was extracted from tissue using the KingFisher Duo Prime Purification System (Thermo Fisher Scientific Inc., Waltham, USA) and the KingFisher Cell and Tissue DNA purification kit, following the manufacturer’s protocol, with the following minor adjustments. The Protein Kinase K volume was increased to 50 µL, and incubation time was extended to 48 hours. DNA concentrations were estimated using the broad range kit on the Qubit 2.0 fluorometer (Thermo Fisher Scientific Inc., Waltham, USA).

We generated complete mitochondrial genomes from 202 beluga individuals. DNA extracts were fragmented to 300-1000 bp using the Bioruptor Sonicator. Fragment sizes were verified on an Agilent 4200 TapeStation (Agilent Inc.). Libraries were constructed following (Carøe, Gopalakrishnan, and Vinner 2018), using 150-700 ng of input DNA and indexed with unique 6 bp sequences. Libraries were sequenced on the illumina HiSeqX platform. Mapping of the mitochondrial sequences was performed within the Paleomix pipeline 1.2.12 (Schubert et al. 2014). Reads were trimmed with AdapterRemoval 2.2.0 (Schubert, Lindgreen, and Orlando 2016) applying default settings, except minimum read length which was set to 25bp. Read quality was inspected using FastQC (available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and aligned with BWA (Li 2013) using the Backtrack algorithm, while disabling the starting seed length. Reads that mapped to multiple positions or had mapping quality scores lower than 30 were removed with SAMtools (Li et al. 2009). Sequence duplicates were removed using the MarkDuplicates function in Picard (available from: http://broadinstitute.github.io/picard) before realigning around indels with GATK (McKenna et al. 2010). Consensus sequences were extracted for regions covered by more than five reads and heterozygous indels were manually corrected.

To gain nuclear genome-wide insights we generated a high coverage nuclear genome from a single individual from the Disko bay area. DNA was extracted and built into a PCR free Truseq Illumina sequencing library with 180 bp inserts at the National Genomics Infrastructure (NGI) in Stockholm. The library was then sequenced on an Illumina HiSeqX using 150 bp paired-end sequencing technology at the NGI. We additionally downloaded the raw Illumina reads from the European Nucleotide Archive (ENA) from a previously published high coverage beluga individual from Churchill (SRR5197961).
Chapter 3

Adapter sequences were trimmed from the raw Illumina reads with skewer (Jiang et al. 2014), and trimmed reads were mapped to the published beluga nuclear reference genome (Genbank accession: GCA_002288925.2) using BWA v0.7.15 (Li 2013) and the mem algorithm with default parameters. We parsed the output and removed duplicates with Samtools v1.6 (Li et al. 2009). From this we constructed a diploid consensus sequence using Samtools and bcftools (Narasimhan et al. 2016) specifying a minimum base quality score of 20 and minimum coverage of 10.

Diversity and differentiation among populations

To estimate levels of mitogenome diversity within each sampling locality and across populations, we calculated nucleotide diversity (\(\pi\)). Nucleotide diversity (\(\pi\)) delimits the average number of nucleotide differences per site between individuals, and values range between 0 and 1. The differentiation index \(F_{ST}\) was applied to estimate levels of differentiation among the 23 localities. Both parameters were estimated using Arlequin 3.5 (Excoffier and Lischer 2010).

Haplotype network

To identify unique haplotypes and visualize the number of substitutions separating them, we constructed a median spanning network (Bandelt, Forster, and Rohl 1999) using PopART v1.7 (Leigh and Bryant 2015). We performed the analysis on an alignment of the 202 mitogenomes using MEGA X (Kumar et al. 2018).

Bayesian phylogenetic analyses

To estimate the divergence time between beluga clades, we used a two-phase approach. First, we built a fossil calibrated phylogenetic tree including 18 taxa. Fourteen species of toothed whale and four beluga sequences (DL0732, DL1666, DL1700 and DL0926), representing the most divergent clades identified in the haplotype network. The taxa were chosen to represent the diversity within the odontocetes: Delphininidae (n=4), Phocoenidae (n=3), Monodontidae (n=2), Ziphiidae (n=4) and Physeteroidae (Physeteridae + Kogiidae) (n=2). The four river dolphin families Iniidae, Pontoporiniidae, Lipotidae and Platanistidae were not included, due to their paraphyletic topology (Nikaido et al. 2001) and elevated clock rates (Dornburg et al. 2012), which could decrease posterior probabilities and increase credibility intervals of the tree respectively. Second, we constructed a phylogeny of the 179 unique beluga mitogenomes using the oldest divergence date between beluga
clades from our phase one tree as a prior. A similar two-phase approach has been used to estimate intraspecific divergence times in other odontocetes (Morin et al. 2010, 2018).

Phase one: odontocete phylogeny

Published mitogenomes of 14 toothed whales (Supplementary Table 2) were downloaded from NCBI, and gene regions were extracted using information from bed files. All genomic regions including protein coding regions (n=13), control region (n=1), rRNAs (n=2) and tRNAs (n=22) were individually aligned using Mafft 7.3 (Katoh and Standley 2013). For the protein coding regions, alignments were manually corrected, so the length and any gaps matched the reading frame, before the three positions of each gene were split into separate partitions which resulted in a total of 64 subsets of data. Using Dambe (Xia and Xie 2001), all subsets were tested for substitution saturation. We used Partition Finder 2.1.1 (Lanfear et al. 2017) to identify the appropriate partitioning scheme, and substitution models for the subsets. Partitions and substitution model are listed in Supplementary Table 3.

Four fossil calibrations were applied to the phase one phylogenetic tree. (i) Ferecetotherium kelloggi (Mchedlidze 1970) was used to set the minimum age of crown odontoceti to 23 MYA. Similar calibration points has previously been used for both Dornburg et al. (2012) and Galatius et al. (2018). The calibration was applied as a log normal distribution with an offset of 23 MYA and a mean age of 35 MYA (HPD 90% = 27.7-47.1). This mean was set to match consensus age of odontoceti estimated from molecular analyses (McGowen, Spaulding, and Gatesy 2009; Steeman et al. 2009; Xiong et al. 2009; Geisler et al. 2011). (ii) Kentriodon pernix (Kellogg 1927) was used to set the minimum age of Ziphiidae + delphinida to 18 MYA as recommended by Lambert et al. (2017). This calibration point was applied as a log normal distribution with an offset of 18 MYA and a mean of 23.4 MYA (HPD 90% = 18.6-35.0 MYA). (iii) Globicetus hiberus (Bianucci et al. 2013) and Archaeoziphius microglenoideus (Lambert and Louwye 2006) both support a minimum age of crown Ziphiidae of approximately 13.2 MYA. This age constraint on crown Ziphiidae was recommended by both Geisler et al. (2011) and Lambert et al. (2017) and is here applied as a log normal distribution with an offset of 13.2 and a mean of 15.6 MYA (HPD 90% = 13.5-20.7 MYA). Similar node calibrations for crown ziphiiidae has previously been used by Galatius et al. (2018) (iv) Salumiphocaena stocktoni (Wilson 1973) was used to set the minimum age of Monodontoidae (Monodontidae + Phocoenidae) to 7.5 MYA. This calibration point was applied as a log normal distribution with an offset of 7.5 and a mean of 9.9 (HPD 90% = 7.8-15.0 MYA). This node calibration was recommended by both Geisler et al. (2011) and Lambert et al. (2017) and used by both Steeman et al. (2009) and Dornburg et al.
Chapter 3

(2012). Even though several extinct species of monodontidae have been described from the fossil record (Barnes 1984; Vélez-Juarbe and Pyenson 2012; Ichishima et al. 2018; Pesci et al. 2018), no fossil has been described as crown monodontidae, thus we were unable to apply any calibration point to the monodontidae node.

We performed the phylogenetic analysis in Beast2 (Bouckaert et al. 2014) using tRNAs, rRNAs, and first and second codon position of the protein coding regions. Third codon positions and the control region were excluded due to substitution saturation and poor alignment, respectively. A calibrated Yule model (Heled and Drummond 2012) with an estimated birth rate was applied. We used linked trees, linked clocks and unlinked site models with estimated substitution rates. The gamma category count was set to 4 and heterogeneity of clock rates among odontocetes (Dornburg et al. 2012) was accommodated by applying a relaxed log normal clock with an estimated rate. The mean clock prior was set to 0.004 as an initial run with a strict clock gave this as the mean clock rate. The birth rate prior was set as a uniform prior ranging from 0 to 1000. Following Heath (2015) the transversion rate prior on A<-->G mutations were given a gamma distribution with a beta value of 2 and alpha values of 0.5 giving a mean of 1 (C<-->T mutations were fixed to 1). The priors on the transition rates were applied as similar distributions, but the alpha values were set to 0.25 giving mean values of 0.5. The Markov chain Monte Carlo (MCMC) was run twice, each time with 50,000,000 steps storing every 5,000 steps. The tree files and log files of the two runs were combined using LogCombiner v2.5.1 (A. Rambaut and Drummond 2014) excluding the first 10% of each run. Effective sample sizes (ESS) were inspected using Tracer 1.7 (Rambaut et al. 2018) and runs with ESS values above 200 for all parameters where considered converged. A maximum clade credibility tree with mean node heights and a posterior probability limit of 0.9 was constructed using TreeAnnotator v2.5.1 (A. Rambaut and Drummond 2013) and plotted using FigTree 1.4.3 available from (http://tree.bio.ed.ac.uk/software/figtree/).

Phase two: beluga phylogeography

The phase two phylogenetic analysis were performed on 179 unique haplotypes identified using FaBox haplotype converter (Illesen 2007). Duplicate mitogenomes were excluded following Morin (2010, 2018). Thirteen protein coding genes, two rRNAs, 22 tRNAs and the control region were individually extracted using published coordinates (Skovrind et al. 2017), aligned with Mafft 7.3 (Katoh and Standley 2013) and the length and any gaps in the protein coding regions were manually corrected to match the reading frame. To ensure that each data subset had sufficient variation, we combined the data into six subsets, (i) first codon position of all protein coding regions, (ii) second
codon position of all protein coding regions, (iii) third codon position of all protein coding regions, (iv) all tRNAs, (v) the two rRNAs and (vi) the control region. Partition Finder 2.1.1 (Lanfear et al. 2017) was used to identify the best partitioning scheme and substitution models for the six subsets which were subsequently used as input for Beast2 v.2.51 (Bouckaert et al. 2014). Partitions and substitution model are listed in Supplementary Table 4.

The phylogeographic analysis was performed applying a Coalescent Constant Population model suited for single species analyses. A strict clock was used, assuming very little rate heterogeneity within belugas. We used linked trees, linked clocks and unlinked site models with estimated substitution rates. The prior on the clock rate was uniform with lower and upper limits of 0 and 1 respectively. A prior was added to the root as a normal distribution with the mean and 95% credibility intervals matching the oldest divergence within belugas identified in the phase one phylogenetic analysis. The Markov chain Monte Carlo (MCMC) was run twice, each time with 50,000,000 steps storing every 5,000 steps. The tree files and log files of the two runs were combined using LogCombiner v2.5.1 (A. Rambaut and Drummond 2014) excluding the first 10% of each run. Effective sample sizes (ESS) were inspected using Tracer 1.7 (Andrew Rambaut et al. 2018) and runs with ESS values above 200 for all parameters were considered converged. A maximum clade credibility tree with mean node heights and a posterior probability limit of 0.95 was constructed using TreeAnnotator v2.5.1 (A. Rambaut and Drummond 2013) and plotted using FigTree 1.4.3 available from (http://tree.bio.ed.ac.uk/software/figtree/).

Demographic reconstruction

Nuclear genomes

In addition to the nuclear genome from Qeqertarsuaq, West Greenland, we retrieved a previously published nuclear genome from an individual originating from Churchill western Hudson Bay, Canada. To estimate the demographic history of the nuclear genomes, we ran demographic analyses on the autosomal scaffolds of the two beluga genomes using a Pairwise Sequentially Markovian Coalescent model (PSMC) (Li and Durbin 2011). To determine which scaffolds were most likely autosomal in origin, we found putative sex chromosome scaffolds for each of the species under investigation and removed them from future analyses. We found putative sex chromosome scaffolds in the beluga assembly by aligning the assembled genome to the Cow X (Genbank accession: CM0008168.2) and Human Y (Genbank accession: NC_000024.10) chromosomes. Alignments were performed using satsuma synteny (Grabherr et al. 2010) utilising default parameters. We removed
scaffolds found to align to sex chromosomes in the previous step and scaffolds shorter than 100kb. We ran PSMC specifying atomic intervals previously shown to be suitable for human datasets (4+25*2+4+6) and performed 100 bootstrap replicates to investigate support for the resultant demography. To plot the resultant PSMC output, we used the previously published generational mutation rate of 1.65e-08 (Westbury et al. 2019) and a generation time of 32 years (Garde et al. 2015).

Mitogenomes: Bayesian skyline plot

The demographic history of belugas was reconstructed using a coalescent Bayesian skyline analysis in BEAST v2.5.1 (Bouckaert et al. 2014). The analysis included the first, second and third codon position of the protein coding regions, tRNAs, rRNAs and the control region, using the partitioning scheme and substitution models defined in the phase-two phylogenetic analysis. A strict clock with an estimated rate was applied and the root age was calibrated using the mean and 95% highest posterior density (HPD) estimates from the phase two phylogenetic analysis applied as a log-normal distribution. The Markov chain Monte Carlo (MCMC) was run with 50,000,000 steps storing every 5,000 steps. Effective sample sizes (ESS) were inspected using Tracer 1.7 (Andrew Rambaut et al. 2018) and runs with ESS values above 200 for all parameters were considered converged. Median effective maternal population size and 95% HPD interval was plotted applying a generation time of 32 years (Garde et al. 2015).

The two individuals included in the nuclear analyses were not included in the mitogenome analyses. However, we still wanted to identify their positions in the phylogenetic tree. To do this, we aligned the 202 mitogenomes and the mitogenomes of the two samples from the nuclear analyses in Mega X (Kumar et al. 2018) and constructed a pairwise distance matrix. We applied a Tajima-Nie model (Tajima and Nei 1984) with Gamma distributed rates and used the pairwise deletion option for sites with missing data, which we ran for 1,000 bootstraps.

Predicted suitable habitat

In order to identify changes in the size and mean latitude of current and LGM suitable beluga habitat, we applied the AquaMaps method (K. Kaschner et al. 2008). In short, this method uses the annual mean of environmental parameters in the current distribution to identify envelopes describing a suitable habitat in half degree cells. The method then identifies cells matching these envelopes either in the same or different environmental datasets. Once identified, suitable habitat cells are assigned a probability based on how well they match the envelopes, which if above 0.6,
represent habitats of high suitability or core habitat (Kristin Kaschner et al. 2011). Beluga envelopes were extracted for three parameters (depth, ice cover and surface temperature) from the HCAF dataset (K. Kaschner et al. 2008) including cells within the winter distributions described by NAMMCO (2018). The current and LGM suitable beluga habitats were plotted with an Arctic projection using QGIS 3.2 (available from https://qgis.org/en/site/forusers/download.html). Suitable beluga habitat size was estimated by summarizing the geographic size of cells with a probability higher than 0.6 for both current and LGM. To identify any latitudinal differences in suitable habitat between current and LGM, the mean latitude and standard deviation of all cells with a probability higher than 0.6 was calculated. Many caveats are associated with this method. It does not deal with important ecological parameters such as prey availability, competition and predation, but exclusively depend on physical parameters. The method also relies on current habitat, which is hard to precisely determine especially for any migratory arctic whale. Furthermore, it is based on annual mean values for the envelopes, as this is all that is available from LGM, and the model does not account for correlation between parameters. Nevertheless, if we keep these caveats in mind the method offers a way to estimate general trends in habitat size.

Results

Genetics

Mitogenomes

Mitogenomes from 202 individuals were successfully sequenced and mapped to the mitochondrial reference genome (Kim et al. 2017) with coverage between 69x and 14,335x. Mean sample coverage was 879x, with a female: male ratio of 1.04:1 (99 females, 95 males, and 8 samples with unknown sex. Sample sites varied from five to ten individuals, with a mean of 8.8 (Supplementary Table 5).

Nuclear genomes

We downloaded 466,374,135 from the European Nucleotide Archive (ENA), of which 531,535,975 unique reads successfully mapped to the beluga reference genome. This gave us a total of 79,218,897,228bp worth of data and an average genome wide coverage of 34.4734. For the newly sequenced high coverage individual used in the current study, we produced 741,163,239 read pairs,
of which 761,761,252 unique reads successfully mapped to the beluga reference genome. This gave us a total of 111,993,275,909bp worth of data and an average genome wide coverage of 48.8406.

Diversity and differentiation

Mitogenome nucleotide diversity (π) across all 202 samples was 0.003. Five localities (Shelikhov Bay, Bristol Bay, James Bay, Churchill and the White Sea) had π values below 0.0001 and three sample sites (Repulse Bay, Nastapoka River and Qaanaaq) had π values above 0.0030 (Fig. 1). Levels of differentiation (FST) among localities ranged from 0 to 0.98; 45 of the 253 pairwise comparisons were nonsignificant (Supplementary Table 6). Within the Sea of Okhotsk, the FST analyses could not differentiate between Sakhalinsky Bay and Ulbansky Bay, although both were highly differentiated from the Shelikhov Bay belugas (FST=0.88-0.79, respectively). In the Pacific region, Hendrickson Island could not be differentiated from Anadyr Estuary and Cook Inlet. All other comparisons in the Pacific region were significant, with the highest levels of diversity found between Bristol Bay and Anadyr Estuary (FST=0.72). Within the Western Atlantic region only two sample sites (James Bay and St. Lawrence Estuary) were significantly differentiated from all other sample sites. All other sample sites within this region were indistinguishable from one to eight other sample sites. In the Eastern Atlantic region the comparison of Svalbard and Yenisei Gulf showed that they were not significantly different, while they were both different from the White Sea with FST values of 0.42 and 0.46, respectively.

Haplotype network

The haplotype network included 281 informative sites with haplotypes in the network representing between one and eight samples (Fig. 2). The network had 132 haplotypes revealed that maternal beluga lineages have a lot of structure, with four major clades, separated by up to 131 variable sites. The most distant clade included all nine belugas from St. Lawrence Bay, seven out of nine belugas from Nastapoka river and one to three belugas from Repulse Bay, Qaanaaq and Arviat. Belugas from ten different sample sites (Sakhalinsky Bay, Ulbansky Bay, Bristol Bay, Churchill, James Bay, Qeqertarsuaq, St. Lawrence Bay, Svalbard, White Sea and Yenisey Gulf) were only found within a single clade, while belugas from 13 sample sites (Anadyr Estuary, Arviat, Cook Inlet, Cunningham Inlet, Grise Fjord, Hendrickson Island, Iqaluit, Nastapoka River, Nelson River, Pangnirtung, Qaanaaq

---

1 The different number of unique haplotypes identified by FaBox (179, used in the phylogenetic analyses) and PopArt (132, used in the haplotype network) is a result of the different ways they handle missing data. FaBox treats sites with missing data (N) as a fifth base, and will identify sequences as unique even if they only differ from other sequences in positions with N. PopArt will exclude sites that have N in any individual before identifying unique sequences. Before this manuscript is submitted for publication a uniform treatment of missing data will be applied to both analyses.
Chapter 3

and Repulse Bay) was found in two or more clades. As an example of the uneven haplotype
distribution among sample sites, belugas found in Shelikhov Bay were separated by a maximum of
four mutations and all found within a single clade, whereas haplotypes found in Qaanaaq were
separated by up to 150 mutations and found in three clades.

Divergence times

The phylogeny of 15 odontocete taxa had posterior values of 1 for all nodes between the
root and the four beluga sequences (Supplementary Fig. 1). Our analysis revealed that belugas
diverged from their sister species narwhals 4.8 Ma (95% HPD=3.1-6.6 Mya) and estimated the
earliest divergence within belugas to have happened 801 kya (95% HPD=490-1,160 kya). The phase
two dated phylogeny of 179 unique beluga haplotypes resulted in 22 clades with posterior values
above 0.9 (Fig. 2). The oldest divergence event was 745 kya (95% HPD =445-1,030 kya), with the three
subsequent divergence events between 203 kya (95% HPD=110-292 kya) and 125 kya (95%
HPD=65-185 kya). All other divergence events have mean divergence times later than 36 kya.

The pairwise distance matrix revealed that the two individuals were closest related to individuals in
different clades. The individual from Qeqertarsuaq (CGG_1_017647) was closest related to
CGG_1_010755 from Grise Fjord with a pairwise distance of 0.00006, and belonged to Clade D. The
Individual from Churchill was closest related to seven samples (CGG_1_023435, CGG_1_023054,
CGG_1_023037, CGG_1_023043, CGG_1_023045, CGG_1_023046, CGG_1_020812) all found in
Clade C, with pairwise distances below 0.00001.

Demographic reconstruction

The median effective population size (Ne) based on the mitogenomes, was near-constant at
around 6,000 individuals between 250-25 kya, after which it increased six-fold, to ~38,000 individuals
1,000 years ago (Fig. 4a). The PSMC analysis of the two nuclear genomes follow identical trajectories;
both genomes show Ne of ~18,000 individuals until 150 kya, after followed by a steady decline in Ne,
ending at 4,200 individuals ~25 kya (Fig. 4a). From 25 kya to present, Ne increased again 3-4 fold.

Predicted suitable habitat

The present day geographic areas identified as suitable beluga habitat (p>0.6) by the
aquamaps model did, in general, match the actual winter beluga distributions (Fig. 5). Exceptions
were that the model identified the waters off the Greenlandic east coast as suitable habitat even
though belugas are only occasionally observed there and a small area in the Gulf of Bothnia, which is
also not beluga habitat. When comparing the suitable habitat distribution between present day and LGM there is a clear pattern, with the LGM suitable habitat being smaller and further south than the present day suitable habitat (Fig. 5). The suitable current winter habitat was estimated at 3,877,000 square kilometers with a mean latitude of 63.3 (SD=7.6) degrees north, while the suitable LGM winter habitat was estimated at 1,381,000 square kilometers with a mean latitude of 57.8 (SD=6.4) degrees north.

**Discussion**

We analyzed 202 mitochondrial genomes from belugas samples across their circumpolar range, and find a high degree of diversity and differentiation within the species. The 132 haplotypes found across samples were structured into four, well-differentiated clades (Fig. 3). The number of mutations separating mitogenome clades ranged from 24 to 131, whereas levels of differentiation within clades only ranged up to 27 mutations. The four clades varied markedly in size from 14 to 91 individuals. Our phylogenetic analyses revealed that the four clades were already established before the beginning of the last colder climatic period ~115 kya (Fig. 2). Results also reveal that the two oldest clades have non overlapping distributions on opposite sides of the beluga range, while the two youngest clades are found across the distribution (Fig. 2, Fig. 3). The demographic reconstructions from the 202 mitogenomes and the two nuclear genomes support a decline in effective population size during colder climatic periods and an increase during warmer climatic periods (Fig. 4a). The habitat model indicates that the predicted suitable beluga habitat size followed a similar trajectory, suggesting that increases in effective population size during warmer climatic periods was achieved through range expansions (Fig. 5).

**Genetics**

The phylogenetic analysis identifies four clades and distinguishes Clade A as a basal lineage (Fig 2). Twentythree individuals belong to Clade A, which are separated by 1-21 mutations, including all individuals from St. Lawrence Estuary in Quebec, Canada. This locality has a resident and highly divergent population with less than a thousand individuals, which is endangered and protected under the Canadian Species At Risk Act (Béland 1996; Findlay et al. 2009). Clade A comprises samples from populations that overwinter in Baffin Bay, Hudson Strait, Davis Strait or Labrador Sea (G. M. O’Corry-Crowe 2018; NAMMCO 2018). These include seven out of nine samples from Nastapoka River, a location in Eastern Hudson Bay, which harbors migratory belugas that has previously been
Chapter 3

shown to share shorter mtDNA haplotypes with the St. Lawrence population (De March and Postma 2003; Colbeck et al. 2013). The divergence between Clade A and Clades B-D is dated to ~750 (745 HPD =1,026-451) kya. A divergence time of ~1025-450 kya coincides with a period of 500 ky (925-425 kya), where the Bering land bridge separated the Bering sea and the Arctic ocean for more than 80% of the time (Miller et al. 2005). In addition, the mean divergence time between Clade A and Clades B-D of ~750 kya matches the beginning of a 150 ky period with continuous separation between the two sea basins, which is the longest such separation in the last million years (Fig. 2). In other marine mammal species such as killer whales and walruses, similar intraspecific divergence estimates has been found between subspecies or ecotypes (Morin et al. 2010; Andersen et al. 2017).

Clade B includes 14 individuals separated from each other by 1-6 mutations (Fig. 3), and separated from Clade A by 145 mutations. Clade B includes samples from three localities, all within the Sea of Okhotsk, including all ten samples from Shelikhov Bay (Fig. 3). Hence, the two most basal beluga lineages, Clades A and B are found on separate sides of the Arctic with non overlapping distributions. Clade A is found exclusively in the western Atlantic region and Clade B is exclusively found on the Pacific region, suggesting they originated in these regions. The presence of isolated beluga populations in the Atlantic and Pacific during periods of large sea ice extent is supported by our species distribution model analysis, which shows the presence of suitable habitat in both regions during the LGM (Fig 5). This could indicate that the fragmentation of beluga habitat, be it from sea ice or the Bering land bridge, was driving early divergence in the species. This would explain why belugas have large amounts of intraspecific diversity compared to their sister species narwhals, which are exclusively found in the Atlantic part of the Arctic (Garde 2011; Heide-Jørgensen 2018).

Clades C and D diverged from Clade B ~200 (202 HPD=110-292) kya and subsequently diverged from each other ~175 (173 HPD=98-259) kya. The uncertainties surrounding these estimates are large and cover both warm and cold climatic periods (Fig. 4b), but the mean ages of ~200 and ~175 kya coincides with the beginning of global cooling period which lasted from ~215-140 kya (Snyder 2016), with sea ice expansion and separation of the Bering sea and the Arctic ocean (Miller et al. 2005; Polyak et al. 2010) (Fig. 4b, Fig. 4c). Clades C and D include 74 and 91 individuals, respectively. Thirtyone mutations separates Clades C and D from Clade B, and Clade C and D are separated by 27 mutations (Fig. 3). The geographic distribution of Clades C and D reveal secondary contact with both clade A and Clade B. (Fig. 2). In four localities (Nastapoka River, Repulse Bay, Arviat and Qaanaaq), we observe secondary contact between Clade A and one or both of Clades C and D (Fig. 2, Fig. 3). The ~750 ky time-span and more than 135 nucleotide differences separating Clades C
Chapter 3

and D from Clade A are reflected in the genetic diversity ($\pi$), which is substantially higher in these four localities ($\pi=0.003-0.005$) than in the remaining 19 localities ($\pi=0.000-0.002$). Both Clade C and D are found throughout the beluga distribution with no pattern of geographic structuring (Fig. 2, Fig. 3). Clades C and D both have wide distributions across the beluga range and secondary contact between them is evident by the haplotype composition in ten localities (Ulbanski Bay, Cook Inlet, Pangnirtung, Iqaluit, Hendrickson Island, Svalbard, Yenisei Gulf, Nelson River, Cunningham inlet and Grise Fjord) (Fig. 3).

The large geographic distribution of Clades C and D is contrasting the limited distributions of Clades A and B and indicates that Clades C and D were more successful at colonizing new habitat as it became available. During LGM ~20 kya the majority of localities currently inhabited by belugas were inaccessible due to the sea ice extent and the lower sea levels. These include the localities in Western Hudson Bay, Eastern High Arctic, and the Northeast Atlantic, which are currently predominantly inhabited by Clades C and D. Despite this larger pattern of secondary contact throughout the beluga range, clades and subclades are found in uneven proportions (Fig. 2). In five localities only a single subclade is found (Shelikhov Bay, Churchill, James Bay, Bristol Bay and the White Sea), which is reflected in the low genetic diversity in these localities ($\pi=0.00003-0.00012$) (Fig. 1). This uneven distribution of clades and subclades could be explained by a first colonizer effect maintained by maternal site fidelity (Caron and Smith 1990). The first arriving lineage would be able to fill out the ecological niche through a population expansion - which is supported in our demographic reconstruction (Fig. 4a). When other maternal lineages subsequently arrived, the habitat would already be occupied, limiting their potential population expansion and thereby keeping their frequencies low (Waters, Fraser, and Hewitt 2013). The lower frequencies of newly arrived lineages, would make them harder for us to detect, either because they have been removed by stochastic events or simply not among the 5-10 samples per location in our study.

The shape of the two nuclear demographic reconstructions are identical despite the two individuals originating from different localities and belonging to separate mitogenome clades. The individual from Qeqertarsuaq on the west coast of Greenland belongs to Clade C, while the individual from Churchill in Western Hudson Bay, Canada belongs to Clade D (Fig. 2). This is easiest explained by post-LGM colonization of the newly available habitat by matriarchal lineages with strong site fidelity combined with substantial male dispersal or even panmixia in this region. Present day belugas exhibit such matriarchal site fidelity, which remain stable over decades (Caron and Smith 1990; G. O’Corry-Crowe et al. 2016). Migrational movements are conducted in pods that include related individuals, with female and young belugas forming the core of the social structure (Colbeck et al.
2013). Thus, the site fidelity is inherited from mother to young, and are hypothesised to be maintained in matriarchal lineages over time, mimicking the inheritance mode of mitogenomes (G. M. O’Corry-Crowe 2018). This is consistent with the male biased dispersal described in belugas (O’Corry-Crowe et al. 1997; Colbeck et al. 2013) and a previous microsatellites analysis which was not able to differentiate samples from Churchill and Qeqertarsuaq (Brown Gladden et al. 1999). However, the two individuals used in the nuclear demographic reconstructions are both from the western Atlantic region and patterns of interbreeding between lineages might be different in other regions. In the Pacific, belugas exhibit the same site fidelity and male biased dispersal as in the western Atlantic region (G. O’Corry-Crowe et al. 2018), but microsatellite and telemetry studies suggest that they stay separate during winter and that gene flow between them is limited (Citta et al. 2017; O’Corry-Crowe et al. 2002).

The demographic reconstruction of the 202 mitogenomes shows a six-fold increase in the effective female population size since LGM (Fig. 4a), indicating that the colonization was accompanied by a species-wide population expansion. However, Clades C and D include 82% of all samples making it likely that their geographic expansion was driving the demographic trajectory of the species. This is supported by our haplotype network, where we see star shaped patterns with no or small centers in Clades C and D, suggesting that they went through a recent population expansion, a pattern which is less pronounced in Clades A and B (Fig. 3). The nuclear demographic reconstructions (PSMCs) suggest a gradual 3-4 fold decline in effective population size from ~115-20 kya, coinciding with the latest expansion of sea ice which reached its maximum during LGM ~20 kya (Fig 4a). After LGM, when the sea ice retracted, the PSMCs suggest a rapid 3-4 fold increase in population size before 10 kya (Fig. 4a), at which the PSMC method stops being informative (Li and Durbin 2011). This could reflect actual fluctuations in effective population size, decreased connectivity with other populations or a combination of the two. Combined, the nuclear and mitogenomic demographic reconstructions support population expansions during periods with warmer climate and population decline and/or fragmentation during colder climatic periods.

**Predicted suitable habitat**

To further investigate the response in belugas to climatic oscillations, we estimated the size and mean latitude of potential suitable beluga habitat during the latest cold (LGM) and warm periods (present). There are many caveats with the model that we applied. It only deals with a limited number of physical parameters (depth, sea surface temperature, sea ice cover) and it does not consider any ecological parameters such as prey, predation and competition. However, it was able to
identify the general extent of the known current beluga winter habitat with a few exceptions (Fig. 5), the main exceptions being the waters of South-East Greenland and a small area in the northern part of the Baltic Sea, giving us overall confidence in the model. Keeping the caveats in mind, the suitable habitat model did offer information on general patterns of habitat change through time. Our results reveal that the potential suitable winter habitat was located further south during LGM (mean latitude=57.8 degrees north) than it is today (mean latitude=63.3 degrees north), indicating a substantial habitat shift between colder and warmer climatic periods. Habitat shifts, as the ones belugas has gone through, have been hypothesised to reduce haplotype diversity (Pauls et al. 2013). If the habitat becomes unfavorable and colonization of the newly available habitat begins, only a smaller fraction of the population is able to track the habitat, hence the mtDNA diversity is reduced (Hampe and Petit 2005; Bálint et al. 2011). However, our results show that despite the repeated habitat shifts, belugas have retained a high level of mitochondrial diversity through time, indicating that multiple matriarchal lineages were able to trace their habitat as it shifted. Belugas have been suggested to be flexible during climate fluctuations due to their generalist status (Laidre et al. 2008; Kovacs et al. 2011). Some beluga populations have already made changes to their habitat use, utilizing new habitats as they have become ice free and adjusting migration timing in response to changes in sea ice distribution (Hauser et al. 2017; Heide-Jørgensen et al. 2010; Hamilton et al. 2019), but links to fitness or abundance is yet to be identified.

The predicted habitat model also indicated that the potential suitable winter habitat was substantially smaller during LGM ($1.4 \times 10^6$ Km$^2$) than it is today ($3.9 \times 10^6$ Km$^2$). Thus, combined the effective population size estimates from the PSMC and BSP analyses and the species distribution model suggests that colder climatic periods could have been disadvantageous for belugas whereas the warmer post-LGM climate was beneficial. The exact mechanism driving this is cannot be disclosed by our data, but belugas in many locations rely on shallow protected water in bays and estuaries when they nurture their young (Goetz et al. 2007; Smith et al. 2017). During colder climatic periods many such areas were covered by sea ice and inaccessible to belugas, which could have limited their breeding success and reduced their effective population size. Similar responses to post-LGM retraction of the Arctic sea ice has previously been reported in both narwhals and bowhead whales (Foote et al. 2013; Westbury et al. 2019), indicating a common positive response to the latest retraction of sea ice in the three endemic Arctic whales.
Chapter 3

Conclusion

Arctic whales evolved under an ever-changing climate, repeatedly excluding them from their habitat by expanding sea ice (Harington 2008). Our investigation of range-wide genetic variability in belugas provide new perspectives to how they responded to these fluctuations. Our results show that belugas possibly did not benefit from colder periods. The size of their predicted suitable habitat and their effective population size decreased during these colder climatic periods (Fig. 5, Fig. 4). However, our results also suggests that many separate lineages were able to colonize new larger habitats, through population expansions during warmer climatic periods (Fig. 2, Fig 3). The Arctic has already experienced extensive sea ice loss (J. Stroeve and Notz 2018), and the first signs of belugas altering their behavior are being documented (Heide-Jørgensen et al. 2010; G. O’Corry-Crowe et al. 2016; Hauser et al. 2017; Hamilton et al. 2019), but it is unknown if these changes are affecting their fitness. Belugas, as a species, has experienced many past climate oscillations, but never at the current pace. Predictions suggests that there will not be any permanent sea ice left within the lifetime of belugas born today (Wang and Overland 2012). Our results suggest that belugas are adaptable, but whether they will be able to accommodate environmental changes at this elevated pace is uncertain. In addition to the rapid warming and accompanying ecological changes to their environment, belugas will also face new challenges from anthropogenic activity (McQuinn et al. 2011; Castellote et al. 2014, 2019), making predictions of long-term outcomes difficult.

Acknowledgements

This research was supported by the Carlsberg Foundation Distinguished Associate Professor Fellowship, grant no CF16-0202.
Figures

Figure 1. Sample site, sample sizes and nucleotide diversity ($\pi$) of each locality. Colours correspond to sampling localities shown in Fig 2a.
Figure 2. Dated phylogeny of beluga mitogenomes. (a) Map of sample sites and (b) dated phylogeny of 202 mitochondrial genomes sampled across the Arctic distribution range of the species. Sample localities are indicated with coloured squares. In (b), nodes with posterior values below 0.9 have been collapsed and blue bars indicate 90% HPD of node ages. The age of the node separating Clade A from Clades B, C and D is 745 (HPD=450-1026) kya.
Chapter 3

Figure 3. Haplotype network of beluga mitogenomes. (a) Map of the 23 sample localities, (b) median-spanning network of 132 beluga mitogenome haplotypes found across 202 individuals. Each haplotype is represented by a circle and circle size indicates relative frequency of haplotypes. Circles are colored according to the sample sites where it is found. Black dots indicate haplotypes not present in the data. Hatches indicate <6 nucleotide differences between haplotypes, numbers indicate >5 differences between haplotypes. Size of circles indicates relative haplotype frequency.
Figure 4: Demographic reconstruction, global temperature and ocean sea levels change over the last 250 thousand years. (a) Demographic reconstruction based on two nuclear genomes (pink: genome from Qeqertarsuaq, orange: genome from Churchill) and 202 mitochondrial genomes (blue). (b) Global average surface temperature as published by Snyder (2016). (c) Ocean sea level change adapted from Miller et al. (2005). When water levels fell 50 m below current sea levels, as indicated by the red line, the Bering land bridge would have been exposed.
Figure 5: Predicted suitable winter habitat for belugas during the LGM (red) and at present (green). Darker colors indicate higher probabilities. LGM sea ice extent and LGM land exposed by lowered sea levels are adapted from the HCAF dataset (K. Kaschner et al. 2008), used in the modeling of the predicted suitable habitat.
Cited literature


Colbeck, Gabriel J., Pierre Duchesne, Lianne D. Postma, Véronique Lesage, Mike O. Hammill, and Julie Turgeon. 2013. “Groups of Related Belugas (Delphinapterus Leucas) Travel Together during Their
Chapter 3

Seasonal Migrations in and around Hudson Bay.” *Proceedings of the Royal Society B* 280 (1752): 20122552.


Hamilton, Charmain D., Jade Vacquié-Garcia, Kit M. Kovacs, Rolf A. Ims, Jack Kohler, and Christian


Chapter 3


McQuinn, Ian H., Véronique Lesage, Dominic Carrier, Geneviève Larrivée, Yves Samson, Sylvain
Chapter 3


Climate Change on Genetic Diversity within Populations and Species.” *Molecular Ecology* 22 (4): 925–46.


Chapter 3


Supplementary figure 1. Phylogenetic tree of 18 odontocete taxa including four distant related belugas. Numbers indicate the mean node ages. Blue bars and numbers in parentheses indicate the 90% HPD of node ages.
Chapter 3

Supplementary figure 2. Phylogenetic tree of 179 unique beluga haplotypes: Identical sequences are added next to the names. Branch colors indicate posterior values.
### Supplementary tables

**Supplementary Table 1. Sample information.** CGG ID: Center for Geogenetics identification number. Sample collector: Institution or individual who collected the tissue sample, (1: SWFSC, Southwest Fisheries Science Center, Alaska, 2: Fisheries and Oceans Canada, 3: D. I Litovka, 4: A.N. Severtsov Institute of Ecology and Evolution Research, 5: LLC Utrish Dolphinarium, 6: Shirshov Institute of Oceanology Research, 7: Greenland Institute of Natural Resources, 8: Norwegian Polar Institute.) Collector ID: Sample identification number given by the collector. Sex: Putative sex as informed by the collector. Locality: Site where the sample was collected. Sample day, Sample month, Sample year and Sample season indicates when the sample was collected, ? indicates unavailable information.

<table>
<thead>
<tr>
<th>CGG ID</th>
<th>ID</th>
<th>Sample Collector</th>
<th>Collector ID</th>
<th>Sex</th>
<th>Locality</th>
<th>Sample day</th>
<th>Sample month</th>
<th>Sample year</th>
<th>Sample season</th>
</tr>
</thead>
<tbody>
<tr>
<td>CGG_1_020217</td>
<td>D0217</td>
<td>1</td>
<td>61075</td>
<td>F</td>
<td>Bristol Bay</td>
<td>?</td>
<td>9</td>
<td>2006</td>
<td>Autumn</td>
</tr>
<tr>
<td>CGG_1_020218</td>
<td>D0218</td>
<td>1</td>
<td>61076</td>
<td>M</td>
<td>Bristol Bay</td>
<td>?</td>
<td>9</td>
<td>2006</td>
<td>Autumn</td>
</tr>
<tr>
<td>CGG_1_020219</td>
<td>D0219</td>
<td>1</td>
<td>61079</td>
<td>M</td>
<td>Bristol Bay</td>
<td>?</td>
<td>9</td>
<td>2006</td>
<td>Autumn</td>
</tr>
<tr>
<td>CGG_1_020220</td>
<td>D0220</td>
<td>1</td>
<td>43391</td>
<td>F</td>
<td>Bristol Bay</td>
<td>22</td>
<td>5</td>
<td>2004</td>
<td>Spring</td>
</tr>
<tr>
<td>CGG_1_020221</td>
<td>D0221</td>
<td>1</td>
<td>43392</td>
<td>M</td>
<td>Bristol Bay</td>
<td>21</td>
<td>5</td>
<td>2004</td>
<td>Spring</td>
</tr>
<tr>
<td>CGG_1_020222</td>
<td>D0222</td>
<td>1</td>
<td>43382</td>
<td>M</td>
<td>Bristol Bay</td>
<td>21</td>
<td>5</td>
<td>2004</td>
<td>Spring</td>
</tr>
<tr>
<td>CGG_1_020223</td>
<td>D0223</td>
<td>1</td>
<td>33593</td>
<td>M</td>
<td>Bristol Bay</td>
<td>9</td>
<td>5</td>
<td>2003</td>
<td>Spring</td>
</tr>
<tr>
<td>CGG_1_020224</td>
<td>D0224</td>
<td>1</td>
<td>47964</td>
<td>F</td>
<td>Bristol Bay</td>
<td>20</td>
<td>5</td>
<td>2005</td>
<td>Spring</td>
</tr>
<tr>
<td>CGG_1_020225</td>
<td>D0225</td>
<td>1</td>
<td>55825</td>
<td>F</td>
<td>Bristol Bay</td>
<td>20</td>
<td>5</td>
<td>2006</td>
<td>Spring</td>
</tr>
<tr>
<td>CGG_1_020226</td>
<td>D0226</td>
<td>1</td>
<td>17296</td>
<td>M</td>
<td>Cook Inlet</td>
<td>13</td>
<td>9</td>
<td>2000</td>
<td>Autumn</td>
</tr>
<tr>
<td>CGG_1_020227</td>
<td>D0227</td>
<td>1</td>
<td>28925</td>
<td>M</td>
<td>Cook Inlet</td>
<td>22</td>
<td>7</td>
<td>2002</td>
<td>Summer</td>
</tr>
<tr>
<td>CGG_1_020228</td>
<td>D0228</td>
<td>1</td>
<td>16604</td>
<td>M</td>
<td>Cook Inlet</td>
<td>19</td>
<td>6</td>
<td>2000</td>
<td>Summer</td>
</tr>
<tr>
<td>CGG_1_020229</td>
<td>D0229</td>
<td>1</td>
<td>10918</td>
<td>F</td>
<td>Cook Inlet</td>
<td>1</td>
<td>10</td>
<td>1998</td>
<td>Autumn</td>
</tr>
<tr>
<td>CGG_1_020230</td>
<td>D0230</td>
<td>1</td>
<td>10921</td>
<td>M</td>
<td>Cook Inlet</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>CGG_1_020231</td>
<td>D0231</td>
<td>1</td>
<td>10044</td>
<td>F</td>
<td>Cook Inlet</td>
<td>15</td>
<td>6</td>
<td>1998</td>
<td>Summer</td>
</tr>
<tr>
<td>CGG_1_020232</td>
<td>D0232</td>
<td>1</td>
<td>34357</td>
<td>F</td>
<td>Cook Inlet</td>
<td>4</td>
<td>8</td>
<td>2003</td>
<td>Summer</td>
</tr>
<tr>
<td>CGG_1_020233</td>
<td>D0233</td>
<td>1</td>
<td>49105</td>
<td>F</td>
<td>Bristol Bay</td>
<td>2</td>
<td>9</td>
<td>2005</td>
<td>Autumn</td>
</tr>
<tr>
<td>CGG_1_020234</td>
<td>D0234</td>
<td>2</td>
<td>ARAR-03-1064</td>
<td>M</td>
<td>Arviat</td>
<td>27</td>
<td>8</td>
<td>2003</td>
<td>Summer</td>
</tr>
<tr>
<td>CGG_1_020235</td>
<td>D0235</td>
<td>2</td>
<td>ARAR-03-1065</td>
<td>F</td>
<td>Arviat</td>
<td>6</td>
<td>8</td>
<td>2003</td>
<td>Summer</td>
</tr>
<tr>
<td>CGG_1_020236</td>
<td>D0236</td>
<td>2</td>
<td>ARAR-03-1067</td>
<td>F</td>
<td>Arviat</td>
<td>4</td>
<td>8</td>
<td>2003</td>
<td>Summer</td>
</tr>
<tr>
<td>CGG_1_020237</td>
<td>D0237</td>
<td>2</td>
<td>ARAR-03-1068</td>
<td>M</td>
<td>Arviat</td>
<td>14</td>
<td>8</td>
<td>2003</td>
<td>Summer</td>
</tr>
<tr>
<td>CGG_1_020238</td>
<td>D0238</td>
<td>2</td>
<td>ARAR-03-1071</td>
<td>M</td>
<td>Arviat</td>
<td>6</td>
<td>8</td>
<td>2003</td>
<td>Summer</td>
</tr>
<tr>
<td>CGG_1_020239</td>
<td>D0239</td>
<td>2</td>
<td>ARAR-03-1074</td>
<td>F</td>
<td>Arviat</td>
<td>17</td>
<td>8</td>
<td>2003</td>
<td>Summer</td>
</tr>
<tr>
<td>CGG_1_020240</td>
<td>D0240</td>
<td>2</td>
<td>ARAR-03-1077</td>
<td>F</td>
<td>Arviat</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>CGG_1_020241</td>
<td>D0241</td>
<td>2</td>
<td>DL-CI-96-2</td>
<td>F</td>
<td>Cunningham Inlet</td>
<td>18</td>
<td>7</td>
<td>1996</td>
<td>Summer</td>
</tr>
<tr>
<td>CGG_1_020242</td>
<td>D0242</td>
<td>2</td>
<td>DL-CI-96-3</td>
<td>F</td>
<td>Cunningham Inlet</td>
<td>19</td>
<td>7</td>
<td>1996</td>
<td>Summer</td>
</tr>
<tr>
<td>CGG_1_020243</td>
<td>D0243</td>
<td>2</td>
<td>DL-CI-96-4</td>
<td>F</td>
<td>Cunningham Inlet</td>
<td>22</td>
<td>7</td>
<td>1996</td>
<td>Summer</td>
</tr>
<tr>
<td>CGG_1_020244</td>
<td>D0244</td>
<td>2</td>
<td>DL99-CI-17</td>
<td>M</td>
<td>Cunningham Inlet</td>
<td>20</td>
<td>7</td>
<td>1999</td>
<td>Summer</td>
</tr>
<tr>
<td>CGG_1_020245</td>
<td>D0245</td>
<td>2</td>
<td>DL99-CI-13</td>
<td>F</td>
<td>Cunningham Inlet</td>
<td>20</td>
<td>7</td>
<td>1999</td>
<td>Summer</td>
</tr>
<tr>
<td>CGG_1_020246</td>
<td>D0246</td>
<td>2</td>
<td>DL99-CI-28</td>
<td>M</td>
<td>Cunningham Inlet</td>
<td>20</td>
<td>7</td>
<td>1999</td>
<td>Summer</td>
</tr>
<tr>
<td>CGG_1_020247</td>
<td>D0247</td>
<td>2</td>
<td>ARGF-00-1034</td>
<td>F</td>
<td>Grise Fjord</td>
<td>13</td>
<td>9</td>
<td>2000</td>
<td>Autumn</td>
</tr>
</tbody>
</table>
Chapter 3

<table>
<thead>
<tr>
<th>Code</th>
<th>Date</th>
<th>Location</th>
<th>Species</th>
<th>Age</th>
<th>Sex</th>
<th>Year</th>
<th>Season</th>
</tr>
</thead>
<tbody>
<tr>
<td>CGG_1_020749</td>
<td>DL0749</td>
<td>2</td>
<td>ARGF-00-1043</td>
<td>F</td>
<td>Grise Fjord</td>
<td>13</td>
<td>9</td>
</tr>
<tr>
<td>CGG_1_020751</td>
<td>DL0751</td>
<td>2</td>
<td>ARGF-00-1059</td>
<td>F</td>
<td>Grise Fjord</td>
<td>13</td>
<td>9</td>
</tr>
<tr>
<td>CGG_1_020755</td>
<td>DL0755</td>
<td>2</td>
<td>ARGF-01-1077</td>
<td>M</td>
<td>Grise Fjord</td>
<td>19</td>
<td>9</td>
</tr>
<tr>
<td>CGG_1_020758</td>
<td>DL0758</td>
<td>2</td>
<td>ARGF-00-1044</td>
<td>F</td>
<td>Grise Fjord</td>
<td>13</td>
<td>9</td>
</tr>
<tr>
<td>CGG_1_020759</td>
<td>DL0759</td>
<td>2</td>
<td>GF93-507_GF33</td>
<td>F</td>
<td>Grise Fjord</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>CGG_1_020763</td>
<td>DL0763</td>
<td>2</td>
<td>GF93-517_GF37</td>
<td>M</td>
<td>Grise Fjord</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>CGG_1_020764</td>
<td>DL0764</td>
<td>2</td>
<td>B95-429-GF</td>
<td>M</td>
<td>Grise Fjord</td>
<td>?</td>
<td>8</td>
</tr>
<tr>
<td>CGG_1_020766</td>
<td>DL0766</td>
<td>2</td>
<td>ARGF-00-1033</td>
<td>M</td>
<td>Grise Fjord</td>
<td>13</td>
<td>8</td>
</tr>
<tr>
<td>CGG_1_020770</td>
<td>DL0770</td>
<td>2</td>
<td>HI-05-27</td>
<td>M</td>
<td>Hendrickson Island</td>
<td>?</td>
<td>2005</td>
</tr>
<tr>
<td>CGG_1_020771</td>
<td>DL0771</td>
<td>2</td>
<td>HI-04-01</td>
<td>F</td>
<td>Hendrickson Island</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>CGG_1_020772</td>
<td>DL0772</td>
<td>2</td>
<td>HI-04-1B</td>
<td>M</td>
<td>Hendrickson Island</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>CGG_1_020773</td>
<td>DL0773</td>
<td>2</td>
<td>HI-04-02</td>
<td>M</td>
<td>Hendrickson Island</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>CGG_1_020774</td>
<td>DL0774</td>
<td>2</td>
<td>HI-04-05</td>
<td>F</td>
<td>Hendrickson Island</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>CGG_1_020775</td>
<td>DL0775</td>
<td>2</td>
<td>HI-04-06</td>
<td>F</td>
<td>Hendrickson Island</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>CGG_1_020776</td>
<td>DL0776</td>
<td>2</td>
<td>HI-04-07</td>
<td>M</td>
<td>Hendrickson Island</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>CGG_1_020782</td>
<td>DL0782</td>
<td>2</td>
<td>MDH97-11</td>
<td>F</td>
<td>Hendrickson Island</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>CGG_1_020783</td>
<td>DL0783</td>
<td>2</td>
<td>MDH97-12</td>
<td>F</td>
<td>Hendrickson Island</td>
<td>25</td>
<td>7</td>
</tr>
<tr>
<td>CGG_1_020785</td>
<td>DL0785</td>
<td>2</td>
<td>MDH97-16</td>
<td>M</td>
<td>Hendrickson Island</td>
<td>26</td>
<td>7</td>
</tr>
<tr>
<td>CGG_1_020797</td>
<td>DL0797</td>
<td>2</td>
<td>IQ99-16-DL</td>
<td>M</td>
<td>Iqaluit</td>
<td>?</td>
<td>9</td>
</tr>
<tr>
<td>CGG_1_020798</td>
<td>DL0798</td>
<td>2</td>
<td>IQ99-19-DL</td>
<td>F</td>
<td>Iqaluit</td>
<td>?</td>
<td>9</td>
</tr>
<tr>
<td>CGG_1_020805</td>
<td>DL0805</td>
<td>2</td>
<td>B96-214-IQ</td>
<td>F</td>
<td>Iqaluit</td>
<td>16</td>
<td>7</td>
</tr>
<tr>
<td>CGG_1_020808</td>
<td>DL0808</td>
<td>2</td>
<td>ARIQ-04-1010</td>
<td>M</td>
<td>Iqaluit</td>
<td>?</td>
<td>2004</td>
</tr>
<tr>
<td>CGG_1_020810</td>
<td>DL0810</td>
<td>2</td>
<td>ARIQ-04-1093</td>
<td>M</td>
<td>Iqaluit</td>
<td>?</td>
<td>2004</td>
</tr>
<tr>
<td>CGG_1_020812</td>
<td>DL0812</td>
<td>2</td>
<td>ARIQ-04-1106</td>
<td>M</td>
<td>Iqaluit</td>
<td>?</td>
<td>2004</td>
</tr>
<tr>
<td>CGG_1_020815</td>
<td>DL0815</td>
<td>2</td>
<td>ARIQ-2005-03</td>
<td>M</td>
<td>Iqaluit</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>CGG_1_020816</td>
<td>DL0816</td>
<td>2</td>
<td>ARIQ-2005-04</td>
<td>F</td>
<td>Iqaluit</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>CGG_1_020819</td>
<td>DL0819</td>
<td>2</td>
<td>DL2008-JB-05</td>
<td>M</td>
<td>James Bay</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>CGG_1_020820</td>
<td>DL0820</td>
<td>2</td>
<td>DL2008-JB-06</td>
<td>F</td>
<td>James Bay</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>CGG_1_020821</td>
<td>DL0821</td>
<td>2</td>
<td>DL2008-JB-07</td>
<td>F</td>
<td>James Bay</td>
<td>16</td>
<td>8</td>
</tr>
<tr>
<td>CGG_1_020824</td>
<td>DL0824</td>
<td>2</td>
<td>DL2009-JB-04</td>
<td>F</td>
<td>James Bay</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>CGG_1_020827</td>
<td>DL0827</td>
<td>2</td>
<td>JBDL-07-02</td>
<td>M</td>
<td>James Bay</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>CGG_1_020828</td>
<td>DL0828</td>
<td>2</td>
<td>JBDL-07-04</td>
<td>F</td>
<td>James Bay</td>
<td>14</td>
<td>8</td>
</tr>
<tr>
<td>CGG_1_020830</td>
<td>DL0830</td>
<td>2</td>
<td>JBDL-07-06</td>
<td>M</td>
<td>James Bay</td>
<td>25</td>
<td>8</td>
</tr>
<tr>
<td>CGG_1_020831</td>
<td>DL0831</td>
<td>2</td>
<td>DL2008-JB-04</td>
<td>M</td>
<td>James Bay</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>CGG_1_020834</td>
<td>DL0834</td>
<td>2</td>
<td>DL2008-03</td>
<td>M</td>
<td>James Bay</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>CGG_1_020835</td>
<td>DL0835</td>
<td>2</td>
<td>DL 2007-JB-01</td>
<td>F</td>
<td>James Bay</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>CGG_1_020837</td>
<td>DL0837</td>
<td>2</td>
<td>DLN85-06</td>
<td>M</td>
<td>Nastapoka River</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>CGG_1_020838</td>
<td>DL0838</td>
<td>2</td>
<td>DLN85-08</td>
<td>M</td>
<td>Nastapoka River</td>
<td>13</td>
<td>7</td>
</tr>
<tr>
<td>CGG_1_020841</td>
<td>DL0841</td>
<td>2</td>
<td>DLN85-16</td>
<td>M</td>
<td>Nastapoka River</td>
<td>28</td>
<td>7</td>
</tr>
<tr>
<td>CGG_1_020842</td>
<td>DL0842</td>
<td>2</td>
<td>DLN85-21</td>
<td>F</td>
<td>Nastapoka River</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>CGG_1_020843</td>
<td>DL0843</td>
<td>2</td>
<td>DLN85-23</td>
<td>M</td>
<td>Nastapoka River</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>Code</td>
<td>Date</td>
<td>Species</td>
<td>Lat</td>
<td>Long</td>
<td>Month</td>
<td>Year</td>
<td>Location</td>
</tr>
<tr>
<td>-------------</td>
<td>----------</td>
<td>---------</td>
<td>----------</td>
<td>----------</td>
<td>-------</td>
<td>------</td>
<td>---------------------</td>
</tr>
<tr>
<td>CGG_1_020844</td>
<td>DL0844</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Nastapoka River</td>
</tr>
<tr>
<td>CGG_1_020845</td>
<td>DL0845</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Nastapoka River</td>
</tr>
<tr>
<td>CGG_1_020850</td>
<td>DL0850</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Nastapoka River</td>
</tr>
<tr>
<td>CGG_1_020852</td>
<td>DL0852</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Nastapoka River</td>
</tr>
<tr>
<td>CGG_1_020867</td>
<td>DL0867</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Pangnirtung</td>
</tr>
<tr>
<td>CGG_1_020868</td>
<td>DL0868</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Pangnirtung</td>
</tr>
<tr>
<td>CGG_1_020869</td>
<td>DL0869</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Pangnirtung</td>
</tr>
<tr>
<td>CGG_1_020871</td>
<td>DL0871</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Pangnirtung</td>
</tr>
<tr>
<td>CGG_1_020872</td>
<td>DL0872</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Pangnirtung</td>
</tr>
<tr>
<td>CGG_1_020873</td>
<td>DL0873</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Pangnirtung</td>
</tr>
<tr>
<td>CGG_1_020874</td>
<td>DL0874</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Pangnirtung</td>
</tr>
<tr>
<td>CGG_1_020876</td>
<td>DL0876</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Pangnirtung</td>
</tr>
<tr>
<td>CGG_1_020877</td>
<td>DL0877</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Pangnirtung</td>
</tr>
<tr>
<td>CGG_1_020897</td>
<td>DL0897</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Repulse Bay</td>
</tr>
<tr>
<td>CGG_1_020898</td>
<td>DL0898</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Repulse Bay</td>
</tr>
<tr>
<td>CGG_1_020899</td>
<td>DL0899</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Repulse Bay</td>
</tr>
<tr>
<td>CGG_1_020900</td>
<td>DL0900</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Repulse Bay</td>
</tr>
<tr>
<td>CGG_1_020902</td>
<td>DL0902</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Repulse Bay</td>
</tr>
<tr>
<td>CGG_1_020903</td>
<td>DL0903</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Repulse Bay</td>
</tr>
<tr>
<td>CGG_1_020904</td>
<td>DL0904</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Repulse Bay</td>
</tr>
<tr>
<td>CGG_1_020905</td>
<td>DL0905</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Repulse Bay</td>
</tr>
<tr>
<td>CGG_1_020919</td>
<td>DL0919</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>St. Lawrence River</td>
</tr>
<tr>
<td>CGG_1_020925</td>
<td>DL0925</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>St. Lawrence River</td>
</tr>
<tr>
<td>CGG_1_020926</td>
<td>DL0926</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>St. Lawrence River</td>
</tr>
<tr>
<td>CGG_1_020928</td>
<td>DL0928</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>St. Lawrence River</td>
</tr>
<tr>
<td>CGG_1_020931</td>
<td>DL0931</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>St. Lawrence River</td>
</tr>
<tr>
<td>CGG_1_020933</td>
<td>DL0933</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>St. Lawrence River</td>
</tr>
<tr>
<td>CGG_1_020934</td>
<td>DL0934</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>St. Lawrence River</td>
</tr>
<tr>
<td>CGG_1_020935</td>
<td>DL0935</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>St. Lawrence River</td>
</tr>
<tr>
<td>CGG_1_020936</td>
<td>DL0936</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>St. Lawrence River</td>
</tr>
<tr>
<td>CGG_1_021656</td>
<td>DL1656</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Anadyr estuary</td>
</tr>
<tr>
<td>CGG_1_021657</td>
<td>DL1657</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Anadyr estuary</td>
</tr>
<tr>
<td>CGG_1_021658</td>
<td>DL1658</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Anadyr estuary</td>
</tr>
<tr>
<td>CGG_1_021659</td>
<td>DL1659</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Anadyr estuary</td>
</tr>
<tr>
<td>CGG_1_021660</td>
<td>DL1660</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Anadyr estuary</td>
</tr>
<tr>
<td>CGG_1_021661</td>
<td>DL1661</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Anadyr estuary</td>
</tr>
<tr>
<td>CGG_1_021662</td>
<td>DL1662</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Anadyr estuary</td>
</tr>
<tr>
<td>CGG_1_021663</td>
<td>DL1663</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Anadyr estuary</td>
</tr>
<tr>
<td>CGG_1_021664</td>
<td>DL1664</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Anadyr estuary</td>
</tr>
<tr>
<td>CGG_1_021665</td>
<td>DL1665</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Anadyr estuary</td>
</tr>
<tr>
<td>CGG_1_021666</td>
<td>DL1666</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Anadyr estuary</td>
</tr>
<tr>
<td>CGG_1_021667</td>
<td>DL1667</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Anadyr estuary</td>
</tr>
<tr>
<td>CGG_1_021668</td>
<td>DL1668</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Anadyr estuary</td>
</tr>
<tr>
<td>CGG_1_021669</td>
<td>DL1669</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Anadyr estuary</td>
</tr>
<tr>
<td>Code</td>
<td>Location</td>
<td>Date</td>
<td>Type</td>
<td>Season</td>
<td>Notes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>----------------</td>
<td>------</td>
<td>--------</td>
<td>--------</td>
<td>--------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGG_1_021670</td>
<td>DL1670</td>
<td>4</td>
<td>2476</td>
<td>M</td>
<td>Shelikhov</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGG_1_021671</td>
<td>DL1671</td>
<td>4</td>
<td>3339</td>
<td>F</td>
<td>Shelikhov</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGG_1_021672</td>
<td>DL1672</td>
<td>4</td>
<td>3341</td>
<td>F</td>
<td>Shelikhov</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGG_1_021673</td>
<td>DL1673</td>
<td>4</td>
<td>3342</td>
<td>F</td>
<td>Shelikhov</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGG_1_021674</td>
<td>DL1674</td>
<td>4</td>
<td>3334</td>
<td>M</td>
<td>Shelikhov</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGG_1_021675</td>
<td>DL1675</td>
<td>4</td>
<td>3335</td>
<td>M</td>
<td>Shelikhov</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGG_1_021676</td>
<td>DL1676</td>
<td>4</td>
<td>2562</td>
<td>F</td>
<td>Ulbanskiy bay</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGG_1_021677</td>
<td>DL1677</td>
<td>4</td>
<td>2564</td>
<td>F</td>
<td>Ulbanskiy bay</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGG_1_021678</td>
<td>DL1678</td>
<td>4</td>
<td>2568</td>
<td>F</td>
<td>Ulbanskiy bay</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGG_1_021679</td>
<td>DL1679</td>
<td>4</td>
<td>2584</td>
<td>M</td>
<td>Ulbanskiy bay</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGG_1_021680</td>
<td>DL1680</td>
<td>4</td>
<td>2585</td>
<td>M</td>
<td>Ulbanskiy bay</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGG_1_021681</td>
<td>DL1681</td>
<td>4</td>
<td>3217</td>
<td>F</td>
<td>Ulbanskiy bay</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGG_1_021682</td>
<td>DL1682</td>
<td>4</td>
<td>3225</td>
<td>F</td>
<td>Ulbanskiy bay</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGG_1_021683</td>
<td>DL1683</td>
<td>4</td>
<td>3215</td>
<td>M</td>
<td>Ulbanskiy bay</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGG_1_021684</td>
<td>DL1684</td>
<td>4</td>
<td>3224</td>
<td>M</td>
<td>Ulbanskiy bay</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGG_1_021686</td>
<td>DL1686</td>
<td>4</td>
<td>1338</td>
<td>F</td>
<td>Sakhalinsky bay</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGG_1_021687</td>
<td>DL1687</td>
<td>4</td>
<td>1346</td>
<td>F</td>
<td>Sakhalinsky bay</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGG_1_021688</td>
<td>DL1688</td>
<td>4</td>
<td>1350</td>
<td>F</td>
<td>Sakhalinsky bay</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGG_1_021689</td>
<td>DL1689</td>
<td>4</td>
<td>1343</td>
<td>M</td>
<td>Sakhalinsky bay</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGG_1_021690</td>
<td>DL1690</td>
<td>4</td>
<td>1348</td>
<td>M</td>
<td>Sakhalinsky bay</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGG_1_021691</td>
<td>DL1691</td>
<td>4</td>
<td>3713</td>
<td>F</td>
<td>Sakhalinsky bay</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGG_1_021692</td>
<td>DL1692</td>
<td>4</td>
<td>3723</td>
<td>F</td>
<td>Sakhalinsky bay</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGG_1_021694</td>
<td>DL1694</td>
<td>4</td>
<td>3696</td>
<td>M</td>
<td>Sakhalinsky bay</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGG_1_021695</td>
<td>DL1695</td>
<td>4</td>
<td>3700</td>
<td>M</td>
<td>Sakhalinsky bay</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGG_1_021696</td>
<td>DL1696</td>
<td>5</td>
<td>2749</td>
<td>M</td>
<td>White Sea</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGG_1_021697</td>
<td>DL1697</td>
<td>5</td>
<td>2750</td>
<td>M</td>
<td>White Sea</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGG_1_021700</td>
<td>DL1700</td>
<td>6</td>
<td>3627</td>
<td>F</td>
<td>White Sea</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGG_1_021701</td>
<td>DL1701</td>
<td>6</td>
<td>3629</td>
<td>F</td>
<td>White Sea</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGG_1_021702</td>
<td>DL1702</td>
<td>6</td>
<td>3630</td>
<td>F</td>
<td>White Sea</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGG_1_021703</td>
<td>DL1703</td>
<td>6</td>
<td>3749</td>
<td>F</td>
<td>White Sea</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGG_1_021704</td>
<td>DL1704</td>
<td>6</td>
<td>3751</td>
<td>F</td>
<td>White Sea</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGG_1_021705</td>
<td>DL1705</td>
<td>6</td>
<td>3752</td>
<td>F</td>
<td>White Sea</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGG_1_023037</td>
<td>DL3037</td>
<td>2</td>
<td>CH89-C1</td>
<td>M</td>
<td>Churchill</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGG_1_023038</td>
<td>DL3038</td>
<td>2</td>
<td>CH89-C2</td>
<td>F</td>
<td>Churchill</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGG_1_023039</td>
<td>DL3039</td>
<td>2</td>
<td>DLC90-09</td>
<td>M</td>
<td>Churchill</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGG_1_023041</td>
<td>DL3041</td>
<td>2</td>
<td>PTT-7277</td>
<td>M</td>
<td>Churchill</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGG_1_023042</td>
<td>DL3042</td>
<td>2</td>
<td>PTT-7278</td>
<td>M</td>
<td>Churchill</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGG_1_023043</td>
<td>DL3043</td>
<td>2</td>
<td>ARCHL-15-128157</td>
<td>M</td>
<td>Churchill</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGG_1_023044</td>
<td>DL3044</td>
<td>2</td>
<td>ARCHL-15-28158</td>
<td>F</td>
<td>Churchill</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGG_1_023045</td>
<td>DL3045</td>
<td>2</td>
<td>ARCHL-15-128160</td>
<td>F</td>
<td>Churchill</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGG_1_023046</td>
<td>DL3046</td>
<td>2</td>
<td>ARCHL-15-128161</td>
<td>M</td>
<td>Churchill</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGG_1_023047</td>
<td>DL3047</td>
<td>2</td>
<td>ARCHL-15-128162</td>
<td>F</td>
<td>Churchill</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGG_1_023048</td>
<td>DL3048</td>
<td>2</td>
<td>ARNR-03-01</td>
<td>M</td>
<td>Nelson River</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Chapter 3
<table>
<thead>
<tr>
<th>Dataset</th>
<th>ID</th>
<th>Location</th>
<th>Image</th>
<th>Date</th>
<th>Season</th>
</tr>
</thead>
<tbody>
<tr>
<td>CGG_1_023049</td>
<td>DL3049</td>
<td>Nelson River</td>
<td>F</td>
<td>2003</td>
<td>Summer</td>
</tr>
<tr>
<td>CGG_1_023050</td>
<td>DL3050</td>
<td>Nelson River</td>
<td>F</td>
<td>2003</td>
<td>Summer</td>
</tr>
<tr>
<td>CGG_1_023051</td>
<td>DL3051</td>
<td>Nelson River</td>
<td>M</td>
<td>2003</td>
<td>Summer</td>
</tr>
<tr>
<td>CGG_1_023052</td>
<td>DL3052</td>
<td>Nelson River</td>
<td>M</td>
<td>2004</td>
<td>Summer</td>
</tr>
<tr>
<td>CGG_1_023053</td>
<td>DL3053</td>
<td>Nelson River</td>
<td>M</td>
<td>2004</td>
<td>Summer</td>
</tr>
<tr>
<td>CGG_1_023054</td>
<td>DL3054</td>
<td>Nelson River</td>
<td>F</td>
<td>2004</td>
<td>Summer</td>
</tr>
<tr>
<td>CGG_1_023055</td>
<td>DL3055</td>
<td>Nelson River</td>
<td>M</td>
<td>2004</td>
<td>Summer</td>
</tr>
<tr>
<td>CGG_1_023056</td>
<td>DL3056</td>
<td>Nelson River</td>
<td>M</td>
<td>2005</td>
<td>?</td>
</tr>
<tr>
<td>CGG_1_023057</td>
<td>DL3057</td>
<td>Nelson River</td>
<td>F</td>
<td>2005</td>
<td>?</td>
</tr>
<tr>
<td>CGG_1_023431</td>
<td>DL3431</td>
<td>Yenisei Gulf</td>
<td>M</td>
<td>2015</td>
<td>?</td>
</tr>
<tr>
<td>CGG_1_023432</td>
<td>DL3432</td>
<td>Yenisei Gulf</td>
<td>F</td>
<td>2012</td>
<td>?</td>
</tr>
<tr>
<td>CGG_1_023433</td>
<td>DL3433</td>
<td>Yenisei Gulf</td>
<td>F</td>
<td>2012</td>
<td>?</td>
</tr>
<tr>
<td>CGG_1_023434</td>
<td>DL3434</td>
<td>Yenisei Gulf</td>
<td>F</td>
<td>2012</td>
<td>?</td>
</tr>
<tr>
<td>CGG_1_023435</td>
<td>DL3435</td>
<td>Yenisei Gulf</td>
<td>F</td>
<td>2012</td>
<td>?</td>
</tr>
<tr>
<td>CGG_1_017637</td>
<td>DL7637</td>
<td>Qeqertarsuaq</td>
<td>M</td>
<td>2008</td>
<td>Spring</td>
</tr>
<tr>
<td>CGG_1_017638</td>
<td>DL7638</td>
<td>Qeqertarsuaq</td>
<td>F</td>
<td>2008</td>
<td>Spring</td>
</tr>
<tr>
<td>CGG_1_017640</td>
<td>DL7640</td>
<td>Qeqertarsuaq</td>
<td>M</td>
<td>2008</td>
<td>Spring</td>
</tr>
<tr>
<td>CGG_1_017641</td>
<td>DL7641</td>
<td>Qeqertarsuaq</td>
<td>F</td>
<td>2008</td>
<td>Spring</td>
</tr>
<tr>
<td>CGG_1_017642</td>
<td>DL7642</td>
<td>Qeqertarsuaq</td>
<td>F</td>
<td>2008</td>
<td>Spring</td>
</tr>
<tr>
<td>CGG_1_017643</td>
<td>DL7643</td>
<td>Qeqertarsuaq</td>
<td>M</td>
<td>2008</td>
<td>Spring</td>
</tr>
<tr>
<td>CGG_1_017644</td>
<td>DL7644</td>
<td>Qeqertarsuaq</td>
<td>F</td>
<td>2008</td>
<td>Spring</td>
</tr>
<tr>
<td>CGG_1_017645</td>
<td>DL7645</td>
<td>Qeqertarsuaq</td>
<td>F</td>
<td>2008</td>
<td>Spring</td>
</tr>
<tr>
<td>CGG_1_017646</td>
<td>DL7646</td>
<td>Qeqertarsuaq</td>
<td>M</td>
<td>2008</td>
<td>Spring</td>
</tr>
<tr>
<td>CGG_1_017647</td>
<td>DL7647</td>
<td>Qeqertarsuaq</td>
<td>F</td>
<td>2008</td>
<td>Spring</td>
</tr>
<tr>
<td>CGG_1_017648</td>
<td>DL7648</td>
<td>Qeqertarsuaq</td>
<td>F</td>
<td>2008</td>
<td>Spring</td>
</tr>
<tr>
<td>CGG_1_017649</td>
<td>DL7649</td>
<td>Qeqertarsuaq</td>
<td>F</td>
<td>2008</td>
<td>Spring</td>
</tr>
<tr>
<td>CGG_1_017650</td>
<td>DL7650</td>
<td>Qeqertarsuaq</td>
<td>M</td>
<td>2008</td>
<td>Spring</td>
</tr>
<tr>
<td>CGG_1_017651</td>
<td>DL7651</td>
<td>Qeqertarsuaq</td>
<td>M</td>
<td>2008</td>
<td>Spring</td>
</tr>
<tr>
<td>CGG_1_017652</td>
<td>DL7652</td>
<td>Qeqertarsuaq</td>
<td>M</td>
<td>2008</td>
<td>Spring</td>
</tr>
<tr>
<td>CGG_1_017653</td>
<td>DL7653</td>
<td>Qeqertarsuaq</td>
<td>M</td>
<td>2008</td>
<td>Spring</td>
</tr>
<tr>
<td>CGG_1_017654</td>
<td>DL7654</td>
<td>Qeqertarsuaq</td>
<td>M</td>
<td>2008</td>
<td>Spring</td>
</tr>
<tr>
<td>CGG_1_017655</td>
<td>DL7655</td>
<td>Qeqertarsuaq</td>
<td>M</td>
<td>2008</td>
<td>Spring</td>
</tr>
<tr>
<td>CGG_1_017656</td>
<td>DL7656</td>
<td>Qeqertarsuaq</td>
<td>M</td>
<td>2008</td>
<td>Spring</td>
</tr>
<tr>
<td>CGG_1_017657</td>
<td>DL7657</td>
<td>Qeqertarsuaq</td>
<td>M</td>
<td>2008</td>
<td>Spring</td>
</tr>
<tr>
<td>CGG_1_017658</td>
<td>DL7658</td>
<td>Qeqertarsuaq</td>
<td>M</td>
<td>2008</td>
<td>Spring</td>
</tr>
<tr>
<td>CGG_1_017659</td>
<td>DL7659</td>
<td>Qeqertarsuaq</td>
<td>M</td>
<td>2008</td>
<td>Spring</td>
</tr>
<tr>
<td>CGG_1_017660</td>
<td>DL7660</td>
<td>Qeqertarsuaq</td>
<td>M</td>
<td>2008</td>
<td>Spring</td>
</tr>
<tr>
<td>CGG_1_017661</td>
<td>DL7661</td>
<td>Qeqertarsuaq</td>
<td>M</td>
<td>2008</td>
<td>Spring</td>
</tr>
<tr>
<td>CGG_1_017662</td>
<td>DL7662</td>
<td>Qeqertarsuaq</td>
<td>M</td>
<td>2008</td>
<td>Spring</td>
</tr>
<tr>
<td>CGG_1_017663</td>
<td>DL7663</td>
<td>Qeqertarsuaq</td>
<td>M</td>
<td>2008</td>
<td>Spring</td>
</tr>
<tr>
<td>CGG_1_017664</td>
<td>DL7664</td>
<td>Qeqertarsuaq</td>
<td>M</td>
<td>2008</td>
<td>Spring</td>
</tr>
<tr>
<td>CGG_1_017665</td>
<td>DL7665</td>
<td>Qeqertarsuaq</td>
<td>M</td>
<td>2008</td>
<td>Spring</td>
</tr>
<tr>
<td>CGG_1_017666</td>
<td>DL7666</td>
<td>Qeqertarsuaq</td>
<td>M</td>
<td>2008</td>
<td>Spring</td>
</tr>
<tr>
<td>CGG_1_017667</td>
<td>DL7667</td>
<td>Qeqertarsuaq</td>
<td>M</td>
<td>2008</td>
<td>Spring</td>
</tr>
<tr>
<td>CGG_1_017668</td>
<td>DL7668</td>
<td>Qeqertarsuaq</td>
<td>M</td>
<td>2008</td>
<td>Spring</td>
</tr>
<tr>
<td>CGG_1_017669</td>
<td>DL7669</td>
<td>Qeqertarsuaq</td>
<td>M</td>
<td>2008</td>
<td>Spring</td>
</tr>
<tr>
<td>CGG_1_017670</td>
<td>DL7670</td>
<td>Qeqertarsuaq</td>
<td>M</td>
<td>2008</td>
<td>Spring</td>
</tr>
<tr>
<td>CGG_1_017671</td>
<td>DL7671</td>
<td>Qeqertarsuaq</td>
<td>M</td>
<td>2008</td>
<td>Spring</td>
</tr>
<tr>
<td>CGG_1_017672</td>
<td>DL7672</td>
<td>Qeqertarsuaq</td>
<td>M</td>
<td>2008</td>
<td>Spring</td>
</tr>
<tr>
<td>CGG_1_017673</td>
<td>DL7673</td>
<td>Qeqertarsuaq</td>
<td>M</td>
<td>2008</td>
<td>Spring</td>
</tr>
<tr>
<td>CGG_1_017674</td>
<td>DL7674</td>
<td>Qeqertarsuaq</td>
<td>M</td>
<td>2008</td>
<td>Spring</td>
</tr>
<tr>
<td>CGG_1_017675</td>
<td>DL7675</td>
<td>Qeqertarsuaq</td>
<td>M</td>
<td>2008</td>
<td>Spring</td>
</tr>
<tr>
<td>CGG_1_017676</td>
<td>DL7676</td>
<td>Qeqertarsuaq</td>
<td>M</td>
<td>2008</td>
<td>Spring</td>
</tr>
<tr>
<td>CGG_1_017677</td>
<td>DL7677</td>
<td>Qeqertarsuaq</td>
<td>M</td>
<td>2008</td>
<td>Spring</td>
</tr>
<tr>
<td>CGG_1_019652</td>
<td>DL9652</td>
<td>Svalbard</td>
<td>M</td>
<td>1998</td>
<td>Summer</td>
</tr>
<tr>
<td>CGG_1_019679</td>
<td>DL9679</td>
<td>Svalbard</td>
<td>M</td>
<td>2001</td>
<td>Autumn</td>
</tr>
<tr>
<td>CGG_1_019680</td>
<td>DL9680</td>
<td>Svalbard</td>
<td>F</td>
<td>2001</td>
<td>Autumn</td>
</tr>
<tr>
<td>CGG_1_019681</td>
<td>DL9681</td>
<td>Svalbard</td>
<td>M</td>
<td>2001</td>
<td>Autumn</td>
</tr>
<tr>
<td>CGG_1_019686</td>
<td>DL9686</td>
<td>Svalbard</td>
<td>M</td>
<td>2001</td>
<td>Autumn</td>
</tr>
<tr>
<td>CGG_1_019689</td>
<td>DL9689</td>
<td>Svalbard</td>
<td>M</td>
<td>2013</td>
<td>Summer</td>
</tr>
<tr>
<td>CGG_1_019691</td>
<td>DL9691</td>
<td>Svalbard</td>
<td>F</td>
<td>2014</td>
<td>Summer</td>
</tr>
</tbody>
</table>
Supplementary Table 2. Taxa included in phase one phylogenetic analyses.

<table>
<thead>
<tr>
<th>Species scientific name</th>
<th>Species common name</th>
<th>Accession number (NCBI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Berardius bairdii</td>
<td>Baird’s beaked whale</td>
<td>NC005274</td>
</tr>
<tr>
<td>Cephalorhynchus heavisidii</td>
<td>Heaviside’s Dolphin</td>
<td>NC020696</td>
</tr>
<tr>
<td>Delphinus capensis</td>
<td>Long-beaked Common Dolphin</td>
<td>NC012061</td>
</tr>
<tr>
<td>Globicephala melas</td>
<td>Long-finned Pilot Whale</td>
<td>HM060334</td>
</tr>
<tr>
<td>Hyperoodon ampullatus</td>
<td>Northern Bottlenose Whale</td>
<td>NC005273</td>
</tr>
<tr>
<td>Kogia breviceps</td>
<td>Pygmy Sperm Whale</td>
<td>NC005272</td>
</tr>
<tr>
<td>Mesoplodon densirostris</td>
<td>Blainville’s Beaked Whale</td>
<td>KF032860</td>
</tr>
<tr>
<td>Mesoplodon grayi</td>
<td>Gray’s Beaked Whale</td>
<td>KF981442</td>
</tr>
<tr>
<td>Monodon monoceros</td>
<td>Narwhal</td>
<td>NC005279</td>
</tr>
<tr>
<td>Neophocaena asiapontialis</td>
<td>Narrow-ridged Finless Porpoise</td>
<td>KP170488</td>
</tr>
<tr>
<td>Neophocaena phocaenoides</td>
<td>Indo-Pacific Finless Porpoise</td>
<td>NC021461</td>
</tr>
<tr>
<td>Orcinus orca</td>
<td>Killer Whale</td>
<td>GU187155</td>
</tr>
<tr>
<td>Phocoena phocoena</td>
<td>Harbour Porpoise</td>
<td>NC005280</td>
</tr>
<tr>
<td>Physeter catodon</td>
<td>Sperm Whale</td>
<td>NC005230</td>
</tr>
<tr>
<td>Delphinapterus leucas</td>
<td>Beluga Whale</td>
<td>CGG_1_017166</td>
</tr>
<tr>
<td>Delphinapterus leucas</td>
<td>Beluga Whale</td>
<td>CGG_1_017194</td>
</tr>
<tr>
<td>Delphinapterus leucas</td>
<td>Beluga Whale</td>
<td>CGG_1_017254</td>
</tr>
<tr>
<td>Delphinapterus leucas</td>
<td>Beluga Whale</td>
<td>CGG_1_017288</td>
</tr>
</tbody>
</table>
Supplementary Table 3. Partitioning scheme and substitution models for analysis of 15 species of toothed whales identified by Partition Finder 2.1.1 (Lanfear et al. 2017). Partitions 14, 17, 20, 22, 23, 26, 27, 29 and 30 only included third codon position of protein coding regions or the control region and were excluded due to substitution saturation or poor alignment. +I: Estimate the proportion of invariable sites. +X: Estimate the base frequencies using maximum likelihoods. +G: Estimate the shape of the gamma distribution describing the evolutionary rate of sites.

<table>
<thead>
<tr>
<th>Partition</th>
<th>Model</th>
<th>Length</th>
<th>Subsets</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>TRN+G+X</td>
<td>207</td>
<td>tRNA_Cys, tRNA_Glu, tRNA_Ala</td>
</tr>
<tr>
<td>2</td>
<td>GTR+G+X</td>
<td>284</td>
<td>tRNA_Arg, tRNA_Lys, tRNA_His, tRNA_Phe</td>
</tr>
<tr>
<td>3</td>
<td>TRN+iG+X</td>
<td>213</td>
<td>tRNA_Asn, tRNA_Ser_1, tRNA_Leu_2</td>
</tr>
<tr>
<td>4</td>
<td>TRN+G+X</td>
<td>137</td>
<td>tRNA_Asp, tRNA_Ile</td>
</tr>
<tr>
<td>5</td>
<td>HKY+G+X</td>
<td>143</td>
<td>tRNA_Gln, tRNA_Tyr</td>
</tr>
<tr>
<td>6</td>
<td>GTR+iG+X</td>
<td>1673</td>
<td>tRNA_Gly, rRNA_16S</td>
</tr>
<tr>
<td>7</td>
<td>GTR+iG+X</td>
<td>775</td>
<td>tRNA_Leu_1, CDS_CYTB_1, CDS_ND1_1</td>
</tr>
<tr>
<td>8</td>
<td>TRNEF+iG</td>
<td>588</td>
<td>CDS_COX1_1, tRNA_Met</td>
</tr>
<tr>
<td>9</td>
<td>HKY+iG+X</td>
<td>245</td>
<td>CDS_ND6_2, tRNA_Pro</td>
</tr>
<tr>
<td>10</td>
<td>GTR+G+X</td>
<td>385</td>
<td>tRNA_Val, tRNA_Thr, CDS_ND3_1, tRNA_Ser_2, tRNA_Trp</td>
</tr>
<tr>
<td>11</td>
<td>TRN+iG+X</td>
<td>982</td>
<td>rRNA_12S</td>
</tr>
<tr>
<td>12</td>
<td>GTR+G+X</td>
<td>835</td>
<td>CDS_ND5_1, CDS_ATP6_1</td>
</tr>
<tr>
<td>13</td>
<td>GTR+iG+X</td>
<td>1049</td>
<td>CDS_ND5_2, CDS_ND3_2, CDS_ND4L_2, CDS_ATP6_2</td>
</tr>
<tr>
<td>14</td>
<td>TRN+iG+X</td>
<td>671</td>
<td>CDS_ATP8_3, CDS_ND3_3, CDS_ATP6_3, CDS_COX3_3</td>
</tr>
<tr>
<td>15</td>
<td>GTR+G+X</td>
<td>482</td>
<td>CDS_ND2_1, CDS_ATP8_2, CDS_ATP8_1</td>
</tr>
<tr>
<td>16</td>
<td>HKY+iG+X</td>
<td>518</td>
<td>CDS_COX1_2</td>
</tr>
<tr>
<td>17</td>
<td>GTR+iG+X</td>
<td>518</td>
<td>CDS_COX1_3</td>
</tr>
<tr>
<td>18</td>
<td>TRN+iG+X</td>
<td>490</td>
<td>CDS_COX3_1, CDS_COX2_1</td>
</tr>
<tr>
<td>19</td>
<td>TRN+iG+X</td>
<td>228</td>
<td>CDS_COX2_2</td>
</tr>
<tr>
<td>20</td>
<td>TRN+iG+X</td>
<td>228</td>
<td>CDS_COX2_3</td>
</tr>
<tr>
<td>21</td>
<td>GTR+iG+X</td>
<td>1420</td>
<td>CDS_COX3_2, CDS_ND4_2, CDS_CYTB_2, CDS_ND1_2</td>
</tr>
<tr>
<td>22</td>
<td>TRN+iG+X</td>
<td>380</td>
<td>CDS_CYTB_3</td>
</tr>
<tr>
<td>23</td>
<td>GTR+iG+X</td>
<td>667</td>
<td>CDS_ND1_3, CDS_ND2_3</td>
</tr>
<tr>
<td>24</td>
<td>TRN+G+X</td>
<td>348</td>
<td>CDS_ND2_2</td>
</tr>
<tr>
<td>25</td>
<td>TRN+G+X</td>
<td>558</td>
<td>CDS_ND4_1, CDS_ND4L_1</td>
</tr>
<tr>
<td>26</td>
<td>TRN+iG+X</td>
<td>459</td>
<td>CDS_ND4_3</td>
</tr>
<tr>
<td>27</td>
<td>TRN+iG+X</td>
<td>707</td>
<td>CDS_ND4L_3, CDS_NDS_3</td>
</tr>
<tr>
<td>28</td>
<td>TRN+G+X</td>
<td>176</td>
<td>CDS_ND6_1</td>
</tr>
<tr>
<td>29</td>
<td>HKY+G+X</td>
<td>176</td>
<td>CDS_ND6_3</td>
</tr>
<tr>
<td>30</td>
<td>HKY+G+X</td>
<td>972</td>
<td>Ctrl-Reg_D-loop</td>
</tr>
</tbody>
</table>
Supplementary Table 4. Partitioning scheme and substitution models for analysis of 202 beluga mitochondrial genomes identified by Partition Finder 2.1.1 (Lanfear et al. 2017). +I: Estimate the proportion of invariable sites. +X: Estimate the base frequencies using maximum likelihoods. +G: Estimate the shape of the gamma distribution describing the evolutionary rate of sites.

<table>
<thead>
<tr>
<th>Partition</th>
<th>Model</th>
<th>Length</th>
<th>Subset</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>TRN+X</td>
<td>1516</td>
<td>tRNA_ALL</td>
</tr>
<tr>
<td>2</td>
<td>TRN+I+G+X</td>
<td>2553</td>
<td>rRNA_ALL</td>
</tr>
<tr>
<td>3</td>
<td>HKY+I+X</td>
<td>3802</td>
<td>CDS_ALL_1st</td>
</tr>
<tr>
<td>4</td>
<td>HKY+X</td>
<td>3802</td>
<td>CDS_ALL_2st</td>
</tr>
<tr>
<td>5</td>
<td>TRN+G+X</td>
<td>3802</td>
<td>CDS_ALL_3st</td>
</tr>
<tr>
<td>6</td>
<td>TRN+I+G+X</td>
<td>917</td>
<td>Ctrl-Reg_D-loop</td>
</tr>
</tbody>
</table>
Supplementary Table 5. Geographic location and sample size of localities included in the study. Locality: Geographic location where samples were collected. Lat: Latitude in decimal degrees, Lon: Longitude in decimal degrees. N: sample size.

<table>
<thead>
<tr>
<th>Locality</th>
<th>Lat</th>
<th>Lon</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anadyr estuary</td>
<td>64.551</td>
<td>177.731</td>
<td>10</td>
</tr>
<tr>
<td>Arviat</td>
<td>61.115</td>
<td>-94.033</td>
<td>7</td>
</tr>
<tr>
<td>Bristol Bay</td>
<td>58.373</td>
<td>-158.116</td>
<td>10</td>
</tr>
<tr>
<td>Churchill</td>
<td>58.824</td>
<td>-94.246</td>
<td>10</td>
</tr>
<tr>
<td>Cook Inlet</td>
<td>61.333</td>
<td>-150.75</td>
<td>7</td>
</tr>
<tr>
<td>Cunningham Inlet</td>
<td>74.109</td>
<td>-93.801</td>
<td>9</td>
</tr>
<tr>
<td>Grise Fjord</td>
<td>76.413</td>
<td>-83.101</td>
<td>9</td>
</tr>
<tr>
<td>Hendrickson Island</td>
<td>69.466</td>
<td>-133.636</td>
<td>10</td>
</tr>
<tr>
<td>Iqaluit</td>
<td>63.669</td>
<td>-68.601</td>
<td>8</td>
</tr>
<tr>
<td>James Bay</td>
<td>52.336</td>
<td>-80.074</td>
<td>10</td>
</tr>
<tr>
<td>Nastapoka River</td>
<td>56.911</td>
<td>-76.546</td>
<td>9</td>
</tr>
<tr>
<td>Nelson River</td>
<td>57.127</td>
<td>-92.353</td>
<td>10</td>
</tr>
<tr>
<td>Pangnirtung</td>
<td>66.058</td>
<td>-66.084</td>
<td>9</td>
</tr>
<tr>
<td>Qeqertarsuaq</td>
<td>69.232</td>
<td>-53.618</td>
<td>7</td>
</tr>
<tr>
<td>Qaanaaq</td>
<td>76.932</td>
<td>-71.939</td>
<td>10</td>
</tr>
<tr>
<td>Repulse Bay</td>
<td>66.408</td>
<td>-85.828</td>
<td>8</td>
</tr>
<tr>
<td>St. Lawrence River</td>
<td>48.084</td>
<td>-69.535</td>
<td>9</td>
</tr>
<tr>
<td>Svalbard</td>
<td>78.392</td>
<td>19.32</td>
<td>9</td>
</tr>
<tr>
<td>Shelikhov</td>
<td>57.44</td>
<td>155.893</td>
<td>10</td>
</tr>
<tr>
<td>Sakhalinsky bay</td>
<td>52.99</td>
<td>140.97</td>
<td>9</td>
</tr>
<tr>
<td>Ulbanskiy bay</td>
<td>53.862</td>
<td>137.972</td>
<td>9</td>
</tr>
<tr>
<td>White Sea</td>
<td>66.25</td>
<td>36.954</td>
<td>8</td>
</tr>
<tr>
<td>Yenisei Gulf</td>
<td>72.58</td>
<td>79.813</td>
<td>5</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>202</td>
</tr>
</tbody>
</table>
## Chapter 3

Supplementary Table 6: Fixation index $F_{ST}$ Levels of differentiation $F_{ST}$ between the 23 localities included in the study. Green, yellow and red cells represent high, intermediate and low levels of differentiation. Values in bold are insignificant with $p > 0.05$.

| Sea of Okhotsk | Seattle | Sakhalinsky bay | Ulbanovsky bay | Pacific | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 |
| Sea of Okhotsk | Shelikhov | 1 | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
|                | Sakhalinsky bay | 2 | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
|                | Ulbanovsky bay | 3 | 0.69 | 0.86 | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
| Pacific        | Anadyr estuary | 4 | 0.68 | 0.64 | 0.54 | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
|                | Cook Inlet | 5 | 0.67 | 0.24 | 0.22 | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
|                | Tortugas | 6 | 0.71 | 0.52 | 0.43 | 0.10 | 0.08 | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
|                | Bristol Bay | 7 | 0.66 | 0.53 | 0.37 | 0.72 | 0.47 | 0.62 | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
| Western Atlantic | Ogeg sharks | 8 | 0.68 | 0.25 | 0.66 | 0.14 | 0.25 | 0.11 | 0.86 | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
|                | Green.Fjord | 9 | 0.67 | 0.28 | 0.17 | 0.03 | 0.02 | 0.45 | 0.18 | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
|                | Nelson River | 10 | 0.83 | 0.35 | 0.23 | 0.58 | 0.30 | 0.45 | 0.33 | 0.69 | 0.24 | + | + | + | + | + | + | + | + | + | + | + | + |
|                | James Bay | 11 | 0.67 | 0.89 | 0.62 | 0.39 | 0.53 | 0.36 | 0.58 | 0.47 | 0.43 | 0.82 | + | + | + | + | + | + | + | + | + | + | + | + |
|                | St. Lawrence River | 12 | 0.68 | 0.06 | 0.95 | 0.02 | 0.91 | 0.92 | 0.58 | 0.95 | 0.91 | 0.95 | 0.98 | + | + | + | + | + | + | + | + | + | + | + | + |
| Eastern Atlantic | Cunningham Bank | 13 | 0.70 | 0.29 | 0.19 | 0.35 | 0.09 | 0.19 | 0.32 | 0.41 | 0.06 | 0.05 | 0.62 | 0.92 | + | + | + | + | + | + | + | + | + | + | + |
|                | Rhode Island | 14 | 0.49 | 0.30 | 0.26 | 0.37 | 0.22 | 0.32 | 0.33 | 0.28 | 0.25 | 0.27 | 0.51 | 0.59 | 0.23 | + | + | + | + | + | + | + | + | + | + | + |
|                | Kiel | 15 | 0.62 | 0.32 | 0.02 | 0.54 | 0.25 | 0.40 | 0.42 | 0.65 | 0.20 | 0.37 | 0.08 | 0.25 | 0.05 | + | + | + | + | + | + | + | + | + | + | + |
|                | Churchill | 16 | 0.56 | 0.50 | 0.25 | 0.75 | 0.46 | 0.60 | 0.42 | 0.85 | 0.41 | 0.09 | 0.37 | 0.08 | 0.32 | 0.31 | 0.09 | + | + | + | + | + | + | + | + | + | + | + |
|                | Aviary | 17 | 0.61 | 0.15 | 0.10 | 0.41 | 0.15 | 0.30 | 0.15 | 0.45 | 0.15 | 0.03 | 0.62 | 0.04 | 0.05 | 0.00 | 0.60 | 0.06 | + | + | + | + | + | + | + | + | + | + | + |
|                | Nantucket | 18 | 0.72 | 0.68 | 0.66 | 0.64 | 0.59 | 0.64 | 0.72 | 0.64 | 0.61 | 0.68 | 0.72 | 0.63 | 0.63 | 0.65 | 0.65 | 0.71 | 0.47 | + | + | + | + | + | + | + | + | + | + | + |
|                | George Washington | 19 | 0.45 | 0.37 | 0.03 | 0.10 | 0.16 | 0.17 | 0.42 | 0.15 | 0.16 | 0.35 | 0.29 | 0.03 | 0.23 | 0.04 | 0.31 | 0.42 | 0.15 | 0.25 | + | + | + | + | + | + | + | + | + | + | + |
|                | Channel Islands | 20 | 0.77 | 0.15 | 0.10 | 0.41 | 0.06 | 0.25 | 0.27 | 0.43 | 0.50 | 0.04 | 0.68 | 0.92 | 0.02 | 0.23 | 0.61 | 0.17 | 0.24 | 0.63 | 0.26 | + | + | + | + | + | + | + | + | + | + | + |
| Eastern Atlantic | Yersey.Gulf | 21 | 0.61 | 0.63 | 0.62 | 0.67 | 0.09 | 0.03 | 0.76 | 0.63 | 0.01 | 0.54 | 0.45 | 0.04 | 0.22 | 0.26 | 0.48 | 0.75 | 0.29 | 0.59 | 0.31 | + | + | + | + | + | + | + | + | + | + | + |
|                | White Sea | 22 | 0.56 | 0.88 | 0.31 | 0.41 | 0.52 | 0.39 | 0.37 | 0.62 | 0.43 | 0.82 | 0.92 | 0.06 | 0.60 | 0.48 | 0.81 | 0.90 | 0.70 | 0.26 | 0.67 | 0.46 | + | + | + | + | + | + | + | + | + | + | + |
|                | Svalbard | 23 | 0.62 | 0.70 | 0.61 | 0.10 | 0.23 | 0.04 | 0.89 | 0.63 | 0.12 | 0.63 | 0.38 | 0.95 | 0.36 | 0.39 | 0.59 | 0.79 | 0.40 | 0.66 | 0.17 | 0.44 | 0.05 | 0.42 | + | + | + | + | + | + | + | + | + | + | + |
To be or not to be - a hybrid

When I was first employed at the Natural History Museum of Denmark I was truly fascinated by the vast collections. It spoke to my childhood curiosity, the same curiosity that made me want to study biology. Rows upon rows of natural wonders, insects, fish, mammal skulls and even a basement full of whale skeletons. I had not been at the Museum for long, before a very special specimen was brought to my attention. A putative whale hybrid skull. There was a paper from the early nineties describing its morphology and indicating that it was a beluga-narwhal hybrid. I quickly realised that the advances in methodologies would not just enable us to verify the hybrid status, but also get additional information about the skull. My initial thought was that this is just a single sample and that it would be a rather quick, but I could not have been more wrong. There were problems with the DNA concentrations and the subsequent coverage being very low and we therefore had to be innovative in the way we analysed the data. As a consequence the manuscript went through four rounds of review and five reviewers before it was finally accepted.

*Single beluga in narwhal territory.* Beluga observed in Dove Bugt, northeast Greenland. Photo by Rikke Guldborg Hansen.
Hybridization between two high Arctic cetaceans confirmed by genomic analysis

Authors
Mikkel Skovrind¹, Jose Alfredo Samaniego Castruita¹, James Haile¹, Eve C. Treadaway¹, Shyam Gopalakrishnan¹, Michael V. Westbury¹, Mads Peter Heide Jørgensen¹,², Paul Szpak³, Eline D. Lorenzen¹

¹ Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, 1350 København K, Denmark
² Greenland Institute of Natural Resources, Strandgade 91,2, DK-1401 Copenhagen K, Denmark
³ Department of Anthropology, Trent University, 1600 West Bank Drive, Peterborough, Ontario, Canada K9L0G2

Corresponding authors: Mikkel Skovrind (mikkel.skovrind@snm.ku.dk), Eline D. Lorenzen (elinelorenzen@snm.ku.dk)

Author contribution statement
EDL conceived the study. JH and ECT generated the genomic data, which was analysed by MS, MVW, SG and JCS, PS generated and analyzed the stable isotope data. MPHJ collected the skull in 1990 and contributed tissue samples. MS, PS and EDL wrote the manuscript, with input from the other co-authors.

Abstract
In 1990 a skull from a morphologically unusual Monodontid was found in West Greenland and collected for the Natural History Museum of Denmark, University of Copenhagen. From its intermediate morphology, the skull was hypothesized to be a beluga/narwhal hybrid. If confirmed, the specimen would, to our knowledge, represent the sole evidence of hybridization between the only two toothed whale species endemic to the Arctic. Here we present genome-wide DNA sequence data from the specimen and investigate its origin using a genomic reference panel of eight belugas and eight narwhals. Our analyses reveal that the specimen is a male, first-generation hybrid between a female narwhal and a male beluga. We use stable carbon and nitrogen isotope analysis to investigate the dietary niche of the hybrid and find a higher δ¹³C value than in both belugas and narwhals, suggesting a foraging strategy unlike either parental species. These results further our understanding of the interaction between belugas and narwhals, and
Chapter 4

underscore the importance of natural history collections in monitoring changes in biodiversity. In addition, our study exemplifies how recent major advances in population genomic analyses using genotype likelihoods can provide key biological and ecological insights from low-coverage data (down to 0.05x).

Introduction

Of the 89 extant cetacean species, only three are found in Arctic waters year-round. Belugas, or white whales (*Delphinapterus leucas*), and narwhals (*Monodon monoceros*), are medium-sized toothed whales and the sole representatives of the Monodontidae family. The Monodontids constitute a characteristic component of Arctic ecosystems along with the bowhead whale (*Balaena mysticetus*), a baleen whale.

Belugas have a disjunct distribution, with populations in both Pacific and Atlantic sectors of the Arctic, with a hiatus in the Greenland Sea (Fig. 1a) 1. Narwhals have a more limited distribution in the Atlantic sector (Fig. 1b) 2. Based on a combination of mitochondrial and nuclear genes, the two species are estimated to have diverged ~5 MYA 3, and a recent study of their nuclear genomes showed that subsequent gene flow between belugas and narwhals ceased between 1.25 and 1.65 MYA 4.

Belugas and narwhals are similar-sized whales (3.5-5.5 meters), and both exhibit inherited migratory patterns following the annual break up and formation of sea ice 5,6. For both species, mating takes place in spring during the sea ice break up, as the whales are heading towards their summering grounds. Due to the inaccessibility of the whales in the pack ice during this period, their mating is not well understood.

However, both species have similar breeding and nursing behavior, and females may calve every two to four years 7. Nonetheless, despite their similarities, belugas and narwhals differ in several key aspects. Adult belugas have white skin 1 whereas narwhals have spotted skin with brown, black, grey and white flecks (Fig. 1) 2. Belugas have up to 40 teeth, whereas narwhals have no teeth in the lower jaw and males have an elongated front tooth that protrudes through their upper lip and can grow up to 2.5 meters in length (Fig. 2) 8. The two species differ in prey selection and diving capabilities; belugas mainly feed on fish down to 500 meters 9, whereas narwhals have the capacity to feed on fish and squids at depths > 800 meters 10.

In 1990, an unusual whale skull was collected from a hunter on an island in Disko Bay, West Greenland (Fig. 2). Using 12 morphological measurements as well as teeth counts, the skull was quantitatively compared to a reference panel of beluga and narwhal skulls 8. Although results showed that the skull was larger than those of belugas and narwhals, the measurements revealed that the skull’s relative dimensions were intermediate, in particular the dentition, which showed longitudinal grooves and horizontal orientation of some teeth, resembling the tusk and the vestigial tooth of narwhals 11 (Supplementary Fig. S1). A hunter caught the whale during a subsistence hunt, and in the near vicinity were two similar-looking whales (both
of which were also harvested, but no samples were preserved). All reportedly had an evenly grey coloration and pectoral flippers shaped like those of belugas, and a tail shaped like that of a narwhal. The skull was assumed to belong to a fully grown individual due to the fusion of its premaxilla-maxilla and its large size, and was therefore initially described as either a beluga/narwhal hybrid or an anomalous beluga.

The skull is stored at the Natural History Museum of Denmark, University of Copenhagen (specimen ID MCE1356, Fig. 2b). It is to our knowledge the only evidence of potential hybridization between belugas and narwhals. To investigate whether MCE1356 indeed represents a hybrid, we analyzed genome-wide DNA sequence data retrieved from the specimen against a genomic reference panel of eight belugas and eight narwhals sampled from the same area in West Greenland. In addition, we used stable carbon and nitrogen isotope analysis of bone collagen extracted from the specimen and a reference panel of 18 belugas and 18 narwhals also collected from West Greenland, to investigate the dietary niche of the individual relative to either putative parental species.

**Methods**

**Sampling**

We analyzed (i) genome-wide DNA data extracted from teeth of the skull (MCE1356, Fig. 2b, Supplementary Fig. S1) and from tissue samples of eight belugas and eight narwhals sampled in Disko Bay, West Greenland, and (ii) stable carbon and nitrogen isotopic compositions of bone collagen from MCE1356 and a reference panel of 18 belugas and 18 narwhal skulls from West Greenland (Fig. 1c). Tissue samples (skin) were stored in 96% ethanol. The samples were collected by scientists from the Greenland Institute of Natural Resources under the general permit for biological sampling of the Inuit from the Greenland Government and exported to Denmark under CITES permit IM 0401-897/04, IM 0721-199/08, IM 0330-819/09 and 116.2006. Sample information is detailed in Supplementary Table S2.

**DNA extraction, library preparation and sequencing**

**Tissue samples**

DNA from tissue samples was extracted using the Qiagen Blood and Tissue Kit following the manufacturer's protocol. The DNA was fragmented using the Covaris M220 Focused-ultrasonicator to create ~350-550 base pair (bp) fragment lengths. Libraries were built from the fragmented DNA extracts using Illumina NeoPrep following the NeoPrep Library Prep System Guide applying default settings. PCR amplification, quantification, and normalization were all carried out by the NeoPrep Library Prep System. The libraries
were screened for size distribution on an Agilent 2100 Bioanalyzer and pooled in equimolar ratios before sequencing on an Illumina HiSeq 2500 with 80bp SE technology.

**Tooth/bone samples**

Unlike the tissue samples, old bone and teeth have relatively low DNA concentrations, which necessitates different extraction and library build protocols. Approximately 0.5 g of bone powder from five teeth and one bone shard from specimen MCE1356 was drilled using a hand-held Dremel. DNA was extracted from the tooth/bone powder in a dedicated ancient DNA clean laboratory at the Natural History Museum of Denmark, University of Copenhagen, using the extraction buffer described in \(^{12}\) with the addition of a 30 minute predigest stage \(^{13}\). Instead of using Zymo-Spin V columns (Zymo Research), the extraction buffer was concentrated using Amicon Ultra 30 kDa Centrifugal Filter Units and further concentrated and cleaned using Qiagen Minelute tubes. The purified extracts were then built into Illumina libraries following the protocol described by \(^{14}\). We used QPCR to check that the library build was successful, to select which libraries to sequence, and to calculate the appropriate number of PCR cycles required to sufficiently amplify each library without causing overamplification. In total, four libraries were amplified with unique 6bp indexes, and screened for endogenous content on the Illumina MiSeq platform using 250bp SE sequencing. The best libraries were then re-sequenced on the Illumina HiSeq 2500 platform using 80bp SE technology.

**Bioinformatic analysis**

All mapping and DNA damage analyses were performed within the Paleomix pipeline 1.2.12 \(^{15}\). Reads were trimmed with AdapterRemoval 2.2.0 \(^{16}\) using default settings except minimum read length which was set to 25bp. Reads were inspected using FastQC and aligned with BWA \(^{17}\) applying the Backtrack algorithm and disabling the starting seed length. If reads mapped to multiple positions or had mapping quality scores (MAPQ score from BWA) less than 30, they were removed using SAMtools \(^{18}\). Sequence duplicates were removed using MarkDuplicates from Picard (available from: [http://broadinstitute.github.io/picard](http://broadinstitute.github.io/picard)) and the final alignment was realigned around indels using GATK \(^{19}\). Deamination of cytosine to uracil in specimen MCE1356 was assessed using the output from mapDamage v2.0.6 \(^{20}\). MapDamage results did not show any clear signal of deamination in the sequences (Supplementary Fig. S3).

**Mitochondrial analysis**

To determine the maternal lineage of MCE1356, DNA sequencing reads were mapped to the beluga (GenBank accession: KY444734) and narwhal (GenBank accession: NC_005279) mitochondrial reference genomes and mean coverage was compared. Reads from the eight beluga and eight narwhal samples, that comprised the reference panel, were mapped to their respective mitochondrial reference genomes. We constructed mitochondrial consensus sequences of MCE1356 and the 16 reference panel samples with regions covered by more than five reads using BEDtools \(^{21}\), SAMtools \(^{18}\) and GATK \(^{19}\). We created two
sequence alignments using ClustalW\textsuperscript{22}, applying default settings, which included the 16 reference panel samples and either the version of MCE1356 mapped to the beluga mitochondrial reference genome or the the version of MCE1356 mapped to the narwhal mitochondrial reference genome. We used the two alignments to construct median-joining haplotype networks\textsuperscript{23} using Popart 1.7\textsuperscript{24} (available from: http://popart.otago.ac.nz), excluding any sites with indels or missing data. Subsequently, both post-mapping coverage and the two haplotype networks were used to determine the species of MCE1356’s maternal lineage.

**Nuclear DNA analyses**

We mapped the DNA sequencing reads from all samples to the killer whale (Orcinus orca) reference genome (GCA\_000331955.2). A high-quality beluga whale genome was recently published\textsuperscript{25}, but mapping to one of the two potential parental species could create biases in the analyses. Hence we mapped the reads to the killer whale genome, as it is well assembled and killer whales are still relatively closely related to belugas and narwhals (12 MYA)\textsuperscript{26}, yet distant enough to minimize the risk of introgression that would complicate our analyses.

For all further filtering we used ANGSD v0.923\textsuperscript{27}, a software package that uses genotype likelihoods instead of called genotypes, which is particularly useful when analysing low-coverage NGS data. We used the SAMtools\textsuperscript{18} method implemented in ANGSD to estimate genotype likelihoods, and inferred major and minor alleles directly from the genotype likelihoods using a maximum likelihood approach as described in\textsuperscript{28}.

To visualise the mapped data, we plotted the read depth distribution from all individuals, excluding sites with more than two alleles and sites with a Phred score below 25. Visualising the data from all individuals combined enabled us to estimate the mean read depth of 4.14x and identify sites with elevated read depth. Such sites were more likely to come from paralog and repetitive regions of the genome. The dataset was visually inspected and further filtered, excluding all sites with read depth greater than nine (6.9\% of sites). In ANGSD, SNPs were called based on their allele frequencies. In ANGSD SNPs were called based on their allele frequencies. The minor allele frequency (MAF) was estimated from the genotype likelihoods and a likelihood ratio test was used to determine if the MAF was different from zero. If the \( p \) value from the likelihood ratio test was \(<1e-4\) the site was considered polymorphic and retained in the dataset. Applying this SNP filter meant that no sites with less than four reads were retained, as sites covered by fewer reads could not obtain \( p \) values this low. These filters were applied to a data set including all 17 samples, and a dataset without MCE1356.
To determine whether MCE1356 was a beluga/narwhal hybrid, we further filtered the dataset of 17 individuals, excluding sites with no reads in MCE1356. At this point the mean read depth of MCE1356 was only 1.15x, so in order to ensure that we were not analyzing multicopy loci, we excluded sites covered by more than one read in MCE1356.

Our aim was to compare the alleles found in MCE1356 to the alleles in the reference panel, so we estimated the probability of assigning the allele found in MCE1356 to the wrong parental species given different reference panel minimum unique read depths and allele frequencies. Unique read depth were defined as the number of reads covering a specific site, where all reads came from a unique individual. This probability $P$ was calculated as in Equation 1:

$$P = f_{(PS)}^{urd_{(PS)}} \times (1 - f_{(PS)})$$

Where $f_{(ps)}$ is the allele frequency in the parental species, and $urd_{(ps)}$ is the species specific unique read depth in the reference panel.

The parental species allele frequency giving the highest probability $f_{(ps-max)}$ could be described as in Equation 2:

$$f_{(ps-max)} = \left( \frac{urd_{(PS)}}{urd_{(PS)}+1} \right)$$

By inserting Equation 2 into equation 1, the maximum probability $P_{(max)}$ of assigning the allele found in MCE1356 to the wrong parental species was calculated as in Equation 3:

$$P_{(max)} = \left( \frac{urd_{(PS)}}{urd_{(PS)}+1} \right) \times (1 - \left( \frac{urd_{(PS)}}{urd_{(PS)}+1} \right))$$

Results revealed that with a unique read depth of two, three and four in each parental species, the maximum probability of assigning the allele found in MCE1356 to the wrong parental species was 0.148, 0.105 and 0.082, respectively. These maximum values should not be interpreted as error rates, as they would only be obtained if all sites were variable within the parental species, and all sites had a MAF of exactly $(1-(urd/urd+1))$. Furthermore, assuming that the MAF distribution in belugas and narwhals was similar, erroneous assignment of alleles found in MCE1356 would affect both species equally, and therefore have a minimal influence on inferences of hybridization. A benefit of using the unique read depth in equations 2 and 3 was that it combined the number of individuals and number of reads in the estimation of the probability of assigning the allele found in MCE1356 to the wrong parental species. subsequently, the
read depth distribution, MAF (with fixed major and minor allele), and number of individuals with data in each site were calculated separately for each parental species.

Three datasets were constructed, which besides MCE1356 included, parental species panels with minimum unique read depths of two, three and four. The number of sites retained in the three datasets were 61,105, 11,764 and 360, respectively. As a compromise between maximizing the number of sites and minimizing the wrong assignment of alleles found in MCE1356, we chose to use the dataset with one read in MCE1356 and minimum unique read depths of three in each parental species. That dataset included 11,764 sites, which were used in subsequent analyses.

Summary statistics were performed on the dataset with 17 individuals, including number of sites that were (i) fixed for different alleles in the beluga and narwhal species panels; (ii) polymorphic in belugas, but not in narwhals; (iii) polymorphic in narwhals, but not in belugas; (iv) polymorphic in both belugas and narwhals. The sites that are estimated to be fixed between the two parental species panels will be enriched for markers that are highly differentiated, i.e. have large allele frequency differences, between the two parental species. These markers, although not necessarily fixed differences between the two parental species, are still highly informative for ancestry in MCE1356.

We used the genotype likelihoods from the dataset without MCE1356 to verify that the belugas and narwhals in the reference panel were not themselves recently admixed individuals. We estimated their individual admixture coefficients while specifying two populations (K=2) using NgsAdmix. One hundred runs were performed and mean and standard deviation of the admixture coefficients were used for subsequent interpretation. To confirm that the filters had not revealed previously hidden admixed genetic profiles in the reference panel, this analysis was performed both before and after the unique read depth filters were applied.

We analysed the admixture proportions of MCE1356 using fastNGSadmix, applying 1,000 bootstraps. fastNGSadmix uses allele frequencies of reference populations/species and the genotype likelihoods of a single individual to estimate its admixture proportions. The software has proven robust with NGS data with coverage as low as 0.00015X, and was therefore ideal for our study.

We further estimated the hybrid status of MCE1356 by investigating sites fixed for different alleles in the beluga panel and the narwhal panel (9,178 sites), and comparing the observed proportion of reads matching the beluga-specific allele and the narwhal-specific allele in MCE1356 to the expected proportions under different hybridization scenarios. To determine how well seven different hybridization scenarios
matched the observed data, we computed a Pearson’s Chi-square goodness of fit statistic. The test statistic is computed as in Equation 4:

$$T = \sum \left( \frac{(O_i - E_i)^2}{E_i} \right)$$

where $O_i$ and $E_i$ are the observed and expected counts of alleles derived from parental species $i$, respectively. Under the null hypothesis where the chosen scenario corresponds well to the observed data, the test statistic $T$ follows a central $\chi^2$ distribution. Thus, the scenario that corresponds best with the observed data would lead to the lowest test statistic.

To further investigate the seven different hybridization scenarios, we computed the likelihood of the observed alleles at sites that were fixed for different alleles in the parental populations. Specifically, we computed the likelihood of the observed alleles in the hybrid MCE1356, under a binomial model for the inheritance of the alleles from the parental species, further assuming independence between markers. We would like to note, however, that the violation of the independence assumption still leads to unbiased estimates. Assuming that the hybrid MCE1356 is composed of a proportion $b$ of beluga ancestry and $(1-b)$ of narwhal ancestry, we can write the likelihood of $b$, given the number of reads $n_{ib}$ that match the beluga allele at site $i$ and $n_{in}$ that match narwhal ancestry, as in Equation 5.

$$L(b|n_{ib}, n_{in}) = \prod_{i=1}^{k} L(b|n_{ib}, n_{in}) = \prod_{i=1}^{k} P(n_{ib}, n_{in} | b) = \prod_{i=1}^{k} b^{n_{ib}}(1-b)^{n_{in}}$$

We computed the log-likelihood of $b$ across its whole range, from 0 to 1, and compared the likelihoods across the seven different hybridization scenarios.

**Sex determination**

To determine the sex of MCE1356 and the individuals in the beluga and narwhal reference panels, we investigated X chromosome to autosomal coverage ratios. This was done by comparing the coverage across scaffolds putatively originating from the X chromosome, to that of scaffolds putatively originating from the autosomes. We determined which scaffolds originate from the X chromosome by aligning the killer whale genome to the Cow (Bos taurus) X chromosome (CM008168.2) with SatsumaSynteny \(^{31}\). Moreover, to remove biases that may occur due to homology between some regions of the X and Y chromosomes, we further aligned the resultant putative X chromosome scaffolds to the human Y chromosome (NC_000024.10) and removed these scaffolds from further analysis. We then calculated coverage across the remaining X scaffolds and the four largest scaffolds not aligning to the X chromosome (i.e. autosomal scaffolds), using the samtools depth function in Samtools \(^{18}\). To compensate for any variation in coverage across the genome, we randomly selected 10,000,000 sites, calculated the average coverage for these sites,
and repeated this step 100 times. For each individual, 5% and 95% confidence intervals, as well as first and third quartiles, were calculated and used for subsequent interpretation.

**Stable carbon and nitrogen isotope analysis**

Powdered bone samples (~ 100 mg) from MCE1356, 18 beluga and 18 narwhal skulls were treated with 10 ml 2:1 chloroform/methanol (v/v) under sonication for 1 h to remove lipids. After removing the solvent, the samples were dried under normal atmospheric pressure for 24 h. Samples were then demineralized in 10 ml of 0.5 M HCl for 4 h while being agitated by an orbital shaker. After demineralization, samples were rinsed to neutrality with Type I water, and then heated at 75°C for 36 h in 10⁻³ M HCl to solubilize the collagen. The water soluble collagen was then freeze dried. The collagen extract was then treated with 10 ml 10:5:4 chloroform/methanol/water (v/v/v) under sonication for 1 h to remove any residual lipids. After centrifugation, the chloroform/methanol layer was removed and the methanol was evaporated out of the water layer at 60°C for 24 h. The samples were again freeze dried and weighed into tin capsules for elemental and isotopic analysis. Carbon and nitrogen stable isotopic and elemental compositions were determined using a Nu Horizon (Nu Instruments, UK) continuous flow isotope ratio mass spectrometer at Trent University, Canada. Stable carbon and nitrogen isotopic compositions were calibrated relative to the VPDB and AIR scales using USGS40 and USGS41a. Standard uncertainty was determined to be ±0.17 ‰ for \( \delta^{13}C \) and ±0.22 ‰ for \( \delta^{15}N \); additional analytical details are provided in Supplementary Document S4.

The statistical significance of differences between beluga and narwhal isotopic compositions were assessed using unpaired t-tests.

**Data availability**

Data is available from Electronic Research Data Archive:


**Results**

**Mitochondrial analysis**

Mapping the reads of MCE1356 to the narwhal mitochondrial reference genome yielded a ~3-fold higher coverage (19.5x) than when reads were mapped to the beluga reference genome (6.6x), indicating a narwhal maternal lineage (Supplementary Table S5). We retrieved an average coverage of the mitochondrial genomes of the beluga and narwhal reference panel of 69x and 84x, respectively (Supplementary Table S5). MCE1356 grouped with the narwhal clade both when it was mapped to the beluga (Supplementary Fig. S6) and narwhal reference genome (Fig. 3), confirming a narwhal maternal lineage.
Nuclear analysis
The beluga and narwhal tissue samples were sequenced to an average coverage of 0.24x and 0.18x, respectively, and the sequencing of the four MCE1356 libraries yielded a combined coverage of 0.05x (Supplementary Table S5). The read depth distribution of the combined dataset, excluding reads below 25 bp, duplicate reads, reads with Phred score below 30, mapping quality (MAPQ from BWA) below 25, reads mapping to multiple locations in the reference genome, but including sites covered only by a single read and non-variable sites, had a mean read depth of 4.14x (Supplementary Fig. S7). After removing sites that were not variable with a \( p \) value below \( 1 \times 10^{-4} \), and sites with a read depth greater than nine, the dataset including 17 samples contained 2,700,875 polymorphic sites and the dataset without MCE1356 included 2,671,704 sites.

Excluding sites with no reads in MCE1356 reduced the number of sites to 107,997, with 105,588 sites covered by one read in MCE1356, 2,325 sites covered by two reads and 84 sites covered by three or more reads. Filtering to only include nuclear sites with (i) one read in MCE1356, (ii) minimum beluga unique read depth of three, (iii) minimum narwhal unique read depth of three, resulted in a final dataset including 11,764 SNPs, of which 9,178 sites (78.0%) were fixed for alternate alleles in belugas and narwhals; 1,553 sites (13.2%) were polymorphic in belugas, but not in narwhals; 724 sites (6.2%) were polymorphic in narwhals, but not in belugas; 309 sites (2.6%) were polymorphic in both belugas and narwhals. The mean read depth in the final dataset was 8.15x and the mean number of belugas and narwhals per SNP in the reference panels was 3.34 (SD=0.56) and 3.26 (SD=0.49), respectively.

In the NGSAdmix analyses performed on 2,671,704 sites, all belugas and narwhals had mean admixture coefficients > 0.999 (SD < \( 10^{-6} \)) indicating that none of the individuals in the reference panel were recently admixed. This result was retained when applying the filters based on read depth in the reference panel, reducing the number of sites to 11,764.

The FastNGSadmix analysis that estimated the admixture proportions in MCE1356 from its genotype likelihoods and allele frequencies of belugas and narwhals across 11,764 sites, estimated that MCE1356 is 54% beluga and 46% narwhal (fig. 4), and indicating that it is a first generation hybrid.

Among the 9,178 sites that were called fixed between the beluga and narwhal species panels and covered by a single read in MCE1356, 4,679 sites (51%) carried the beluga-specific allele and 4,499 sites (49%) had the narwhal-specific allele (Fig. 5a). When comparing this observed read distribution with the expected read distribution under seven different hybridization scenarios (Fig. 5b), the \( T \) statistics and the associated \( p \) values revealed that only the first-generation-hybrid scenario could not be rejected (\( T \) value = 3.58, \( p \) value = 0.06) (Fig. 5a, 5c). The log-likelihood of the proportion of beluga/narwhal ancestry, given the observed
read counts, under the seven hybridization scenarios further supported that MCE1356 was a first
generation hybrid (Fig 5d).

**Sex determination**

We used the X chromosome to autosome coverage ratios to determine the sex of MCE1356. A 1:2 ratio
(0.5) indicates a single copy of the X chromosome and a male individual, and a 1:1 ratio (1) indicates two
copies of the X chromosome and a female individual. We find a ratio of ~0.58 in MCE1356, suggesting the
specimen was likely male. To investigate this ratio further, we performed the same analysis on the 16
individuals of the reference panel (Fig. 6). We found no individuals to have a perfect 1:2 (0.5) X
chromosome to autosome coverage ratio. This discrepancy could arise due to a multitude of factors,
including random chance, sequencing bias, and difficulties stemming from the genome assembly and
correct X chromosome scaffold determination. However, it was obvious that there were clear differences
between a number of individuals, and these differences could be separated into two clusters (sexes) around
~0.5 and ~1, with very little variation within each individual. These results gave us confidence in our
method. In the reference panel, we found eight males and eight females. Furthermore, despite the much
lower average coverage of MCE1356, the X chromosome to autosome coverage ratio of ~0.58, with little
variation between different subsamples, still clearly clustered with the putative males from the reference
panel.

**Stable isotope analysis**

The carbon and nitrogen isotopic compositions of beluga, narwhal, and MCE1356 are summarized in Fig. 7
and presented in full in the Supplementary Information, along with elemental compositions and collagen
yields (Supplementary Document S4). Belugas and narwhals were characterized by significantly different
$\delta^{13}C$ ($p < 0.001$) and $\delta^{15}N$ ($p = 0.02$) values. MCE1356 had a much higher $\delta^{13}C$ value than any of the
belugas or narwhals analyzed (> 3 $\sigma$ higher than the beluga mean; > 4 $\sigma$ higher than the narwhal mean),
but a comparable $\delta^{15}N$ value.

**Discussion**

Our analyses of genome-wide DNA retrieved from MCE1356 and a reference panel of belugas and narwhals
showed that the abnormal skull is a F1 hybrid of the two species (Fig. 4, Fig. 5). Mating strategies of belugas
and narwhals are not well understood, reflecting the logistical challenges associated with studying the
behavior of Arctic marine endemics; mating takes place in spring, when the whales are particularly difficult
to access due to sea ice break up. However, the existence of a beluga/narwhal hybrid suggests that the two
species can mate and produce viable offspring.
Our analysis of the mitochondrial genome shows that the mother of MCE1356 was a narwhal (Fig. 3). Male narwhals have one, and in rare cases two, spiraled tusks representing a protruding canine tooth. In contrast, belugas have up to 40 homodont teeth (Fig. 2). The narwhal tusk has been hypothesized to be a secondary sexual characteristic of the species \[33\], which could suggest that male belugas would be less successful in securing cross-species matings than male narwhals. In addition, recent work has suggested that belugas and narwhals have different mating systems, with sperm competition being more important in belugas than in narwhals \[34\]. Our finding of a narwhal mother and beluga father of MCE1356 suggests that even with the absence of a tusk and different mating systems, successful mating can still occur between a male beluga and a female narwhal.

The elongated mating period (late winter to late spring) and migratory nature of both species makes it difficult to evaluate the overlap in their respective distributions during the mating season. However, Disko Bay in West Greenland, where the hybrid skull was collected (Fig. 1c), is one of only a few areas globally where belugas and narwhals are known to occur in large and predictable numbers during the mating season \[7\]. Hybridization could also occur in mixed species aggregations, as belugas are occasionally found in narwhal pods \[7\], and narwhals have also been continuously observed in beluga pods. The absence of publications describing introgression between belugas and narwhals is not necessarily a consequence of it not occurring, but genetic studies performed on narwhals have so far been limited to mitochondrial sequences; this study represents the first genomic data from the species.

The admixture proportions inferred from 11,764 variable sites indicated that MCE1356 was a first-generation beluga/narwhal hybrid. This was supported by our analyses of 9,178 sites fixed for alternative alleles in the two parental species, where the observed distribution of reads matched the expected distribution in a first-generation hybrid (Fig. 5a, 5c). The coverage of the X chromosome was roughly half of the autosome, revealing that MCE1356 was a male (Fig 6). Considering the large size of the MCE1356 skull \[8\], it seems credible, as both beluga and narwhal males are larger than females \[1,2\].

Our finding of hybridisation between belugas and narwhals is unexpected, as a recent genomic analysis showed that gene flow between the two species ceased 1.25-1.65 MYA \[4\]. However, hybridization among cetacean species is relatively common \[35\]; there are at least 16 described cases of hybridization between wild or captive cetacean species \[36\]. Evidence has been based on morphological traits in the offspring and genetic analyses \[37–39\]. A recent genomic study of eight rorqual species (baleen whales) showed pronounced hybridization throughout the evolution and speciation of the group \[40\].

The differences between the carbon and nitrogen isotopic compositions in belugas and narwhals, while relatively small, suggest distinct foraging behaviors, consistent with what has been reported for these two
species in the North Water Polynya. The unusually high $\delta^{13}C$ value for MCE1356 suggests a unique diet for this individual. Because bone collagen remodels at a slow rate, reflecting the diet over the last several years of an animal’s life, this unique diet reflects sustained differences in prey consumption or habitat use, and cannot be ascribed to unusual short-term variation. High $\delta^{13}C$ values are observed in marine animals that forage to a greater extent on benthic relative to pelagic prey. Accounting for the differences in fractionation with trophic level, bearded seals ($Erignathus barbatus$) and walrus ($Odobenus rosmarus$) tend to have the highest $\delta^{13}C$ values of any Arctic marine mammal, consistent with their strong reliance on benthic prey.

The high $\delta^{13}C$ value of the hybrid MCE1356 therefore suggests a greater use of benthic relative to pelagic prey compared to either belugas or narwhals foraging in the same area, the magnitude of which may be comparable to the difference in benthic prey consumption between bearded ($Erignathus barbatus$) and ringed ($Pusa hispida$) seals, which is substantial. The unique foraging behavior of the beluga/narwhal hybrid may have been driven by its peculiar dental morphology (Fig. 2b, Supplementary Fig. S1). Although the precise prey species that formed part of its diet whilst alive are difficult to discern, the $\delta^{15}N$ value suggests that the mean trophic level of prey consumed would have been comparable to belugas and narwhals, for which fish and squid are important prey items.

Conclusions

Recent major advances in analytical methods that use genotype likelihoods, which take uncertainty in genotype calling into account rather than using directly called genotypes, are revolutionizing the quality of insights gained from low-coverage genomic data. Even though we had an average coverage of less than 30% of the genome per individual in the beluga and narwhal reference panel, and only 5% of the hybrid genome, we were still able to retain 11,764 informative sites. From these sites, we confirm hybridization between belugas and narwhals based on DNA retrieved from an anomalous skull. We analyzed 9,178 sites fixed for alternative alleles in each parental species and ascertain that specimen MCE1356 is an F1 hybrid, used mitochondrial genomes to show the mother was a narwhal, and investigated the coverage of scaffolds mapping to the X chromosome versus those of the autosomes to infer the hybrid was male. In addition, the stable isotopic signature of carbon and nitrogen indicates a unique dietary niche of the hybrid unlike that of either parental species, supported by its unique dentition.
Acknowledgements
This research was supported by the Carlsberg Foundation Distinguished Associate Professor Fellowship, grant no CF16-0202, the Villum Fonden Young Investigator Programme, grant no 13151, and the Canada Research Chairs program.

Competing interests
The authors declare no competing financial or non-financial interests
References


11. Heide-Jørgensen, M. P. & Reeves, R. R. Description of an anomalous Monodontid skull from West


Figures

Figure 1
Distribution ranges of (a) belugas and (b) narwhals, and (c) insert map of sampling localities of the beluga (blue) and narwhal (green) used in the study. Dots represent tissue samples used in the genomic analyses, triangles represent the skulls used in the stable isotope analysis. The collection site of specimen MCE1356 is indicated with a red circle. Sample sizes at each locality are indicated. Hatched grey area in Fig 2c represents Disko Bay and adjacent waters. Whale illustrations in (a) and (b) by Larry Foster.
Figure 2
Skull morphology of (a) beluga, (b) MCE1356, and (c) narwhal. Photos: Mikkel Høegh Post.
Figure 3
Median-spanning haplotype network of complete mitochondrial genomes of eight belugas (blue), eight narwhals (green) and MCE1356 (red). Black dots indicate intermediate haplotypes not found in the data. The size of the circles indicate the relative number of specimens sharing a haplotype. Numbers indicate number of variable sites between haplotypes.
Figure 4
Admixture proportions of MCE1356 from fastNGSadmix. Error bars are standard deviations estimated from 1000 bootstraps.
Figure 5
Distribution of reads in MCE1356 at 9,178 polymorphic sites fixed for altering alleles in reference panel belugas and narwhals. a: Observed number of MCE1356 reads matching the beluga allele and narwhal allele, and expected number of reads matching the beluga allele and narwhal allele under seven hybridization scenarios (I-VII). \(T\) values from Pearson’s Chi-square goodness of fit statistic, where MCE1356 is the observed and the seven different hybridization scenarios are used to compute the expected read counts. The only hybridization scenario that could not be rejected is presented in bold. b: Schematic illustration of the seven hybridization scenarios. c: Observed proportion of reads matching the beluga allele and the narwhal allele in MCE1356 and the expected proportions matching the beluga allele and the narwhal allele under seven different hybridization scenarios. d: Log-likelihood of \(b\), the proportion of beluga ancestry in MCE1356, across its whole range (see Equation 5 in main text) and the log-likelihoods of the seven different hybridization scenarios. The seven hybrid scenarios are represented by colored squares with colors matching 5a, MCE1356 is represented by an open red square.
Figure 6
X chromosome to autosome coverage ratio in MCE1356 and the eight belugas and eight narwhals in the reference panel. Values close to 0.5 indicate the individual is a male. Values ~1 indicate a female.
Figure 7
Stable carbon and nitrogen isotopic compositions of belugas (blue), narwhals (green) and MCE1356 (red).
Hybridization between two high Arctic cetaceans confirmed by genomic analysis

Authors
Mikkel Skovrind, Jose Alfredo Samaniego Castruita, James Haile, Eve C. Treadaway, Shyam Gopalakrishnan, Michael V. Westbury, Mads Peter Heide Jørgensen, Paul Szpak, Eline D. Lorenzen

Supplementary Figure S1

Teeth from MCE1356 from which DNA was retrieved. The skull and teeth are housed at the Natural History Museum of Denmark, University of Copenhagen
### Supplementary Table S2

Sample information. All samples are housed at the Natural History Museum of Denmark, University of Copenhagen.

<table>
<thead>
<tr>
<th>Specimen ID</th>
<th>Lab ID</th>
<th>Species</th>
<th>Sample site</th>
<th>LAT</th>
<th>LON</th>
<th>Year</th>
<th>DNA study</th>
<th>Isotope study</th>
<th>Sample</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCE1356</td>
<td>CGG_1_021706</td>
<td>Delphinapterus leucas x Monodon monoceros</td>
<td>Kitisissuarsuit</td>
<td>68.901</td>
<td>-53.022</td>
<td>1987-1988</td>
<td>X</td>
<td>X</td>
<td>Tooth/bone</td>
<td></td>
</tr>
<tr>
<td>GINR3800</td>
<td>CGG_1_017580</td>
<td>Delphinapterus leucas</td>
<td>Ilulissat</td>
<td>69.210</td>
<td>-51.191</td>
<td>1999</td>
<td>X</td>
<td>Tissue</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GINR3801</td>
<td>CGG_1_017581</td>
<td>Delphinapterus leucas</td>
<td>Ilulissat</td>
<td>69.210</td>
<td>-51.191</td>
<td>1999</td>
<td>X</td>
<td>Tissue</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GINR3802</td>
<td>CGG_1_017582</td>
<td>Delphinapterus leucas</td>
<td>Ilulissat</td>
<td>69.210</td>
<td>-51.191</td>
<td>1999</td>
<td>X</td>
<td>Tissue</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GINR3803</td>
<td>CGG_1_017583</td>
<td>Delphinapterus leucas</td>
<td>Ilulissat</td>
<td>69.210</td>
<td>-51.191</td>
<td>1999</td>
<td>X</td>
<td>Tissue</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GINR3804</td>
<td>CGG_1_017584</td>
<td>Delphinapterus leucas</td>
<td>Ilulissat</td>
<td>69.210</td>
<td>-51.191</td>
<td>1999</td>
<td>X</td>
<td>Tissue</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GINR3805</td>
<td>CGG_1_017585</td>
<td>Delphinapterus leucas</td>
<td>Ilulissat</td>
<td>69.210</td>
<td>-51.191</td>
<td>1999</td>
<td>X</td>
<td>Tissue</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GINR3806</td>
<td>CGG_1_017586</td>
<td>Delphinapterus leucas</td>
<td>Ilulissat</td>
<td>69.210</td>
<td>-51.191</td>
<td>1999</td>
<td>X</td>
<td>Tissue</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GINR3807</td>
<td>CGG_1_017587</td>
<td>Delphinapterus leucas</td>
<td>Ilulissat</td>
<td>69.210</td>
<td>-51.191</td>
<td>1999</td>
<td>X</td>
<td>Tissue</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GINR3808</td>
<td>CGG_1_017557</td>
<td>Monodon monoceros</td>
<td>Qeqertarsuaq</td>
<td>69.232</td>
<td>-53.603</td>
<td>1999</td>
<td>X</td>
<td>Tissue</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GINR3809</td>
<td>CGG_1_017558</td>
<td>Monodon monoceros</td>
<td>Qeqertarsuaq</td>
<td>69.232</td>
<td>-53.603</td>
<td>1999</td>
<td>X</td>
<td>Tissue</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GINR3810</td>
<td>CGG_1_017559</td>
<td>Monodon monoceros</td>
<td>Qeqertarsuaq</td>
<td>69.232</td>
<td>-53.603</td>
<td>1999</td>
<td>X</td>
<td>Tissue</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GINR3244</td>
<td>CGG_1_017560</td>
<td>Monodon monoceros</td>
<td>Qeqertarsuaq</td>
<td>69.232</td>
<td>-53.603</td>
<td>1999</td>
<td>X</td>
<td>Tissue</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GINR3245</td>
<td>CGG_1_017561</td>
<td>Monodon monoceros</td>
<td>Qeqertarsuaq</td>
<td>69.232</td>
<td>-53.603</td>
<td>1999</td>
<td>X</td>
<td>Tissue</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GINR3246</td>
<td>CGG_1_017562</td>
<td>Monodon monoceros</td>
<td>Qeqertarsuaq</td>
<td>69.232</td>
<td>-53.603</td>
<td>1999</td>
<td>X</td>
<td>Tissue</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GINR3247</td>
<td>CGG_1_017563</td>
<td>Monodon monoceros</td>
<td>Qeqertarsuaq</td>
<td>69.232</td>
<td>-53.603</td>
<td>1999</td>
<td>X</td>
<td>Tissue</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GINR3248</td>
<td>CGG_1_017564</td>
<td>Monodon monoceros</td>
<td>Qeqertarsuaq</td>
<td>69.232</td>
<td>-53.603</td>
<td>1999</td>
<td>X</td>
<td>Tissue</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCE1312</td>
<td>Delphinapterus leucas</td>
<td>Kullorsuaq</td>
<td>74.583</td>
<td>-57.434</td>
<td>1990</td>
<td>X</td>
<td>Bone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCE1313</td>
<td>Delphinapterus leucas</td>
<td>Kullorsuaq</td>
<td>74.583</td>
<td>-57.434</td>
<td>1990</td>
<td>X</td>
<td>Bone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCE1314</td>
<td>Delphinapterus leucas</td>
<td>Kullorsuaq</td>
<td>74.583</td>
<td>-57.434</td>
<td>1991</td>
<td>X</td>
<td>Bone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCE1316</td>
<td>Delphinapterus leucas</td>
<td>Nuussuaq</td>
<td>76.173</td>
<td>-51.846</td>
<td>1994</td>
<td>X</td>
<td>Bone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCE1318</td>
<td>Delphinapterus leucas</td>
<td>Nuussuaq</td>
<td>76.173</td>
<td>-51.846</td>
<td>1994</td>
<td>X</td>
<td>Bone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCE1319</td>
<td>Delphinapterus leucas</td>
<td>Nuussuaq</td>
<td>76.173</td>
<td>-51.846</td>
<td>1994</td>
<td>X</td>
<td>Bone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCE1320</td>
<td>Delphinapterus leucas</td>
<td>Nuussuaq</td>
<td>76.173</td>
<td>-51.846</td>
<td>1994</td>
<td>X</td>
<td>Bone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCE1321</td>
<td>Delphinapterus leucas</td>
<td>Nuussuaq</td>
<td>76.173</td>
<td>-51.846</td>
<td>1994</td>
<td>X</td>
<td>Bone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCE1322</td>
<td>Delphinapterus leucas</td>
<td>Nuussuaq</td>
<td>76.173</td>
<td>-51.846</td>
<td>1994</td>
<td>X</td>
<td>Bone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCE1323</td>
<td>Delphinapterus leucas</td>
<td>Nuussuaq</td>
<td>76.173</td>
<td>-51.846</td>
<td>1994</td>
<td>X</td>
<td>Bone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCE1324</td>
<td>Delphinapterus leucas</td>
<td>Nuussuaq</td>
<td>76.173</td>
<td>-51.846</td>
<td>1994</td>
<td>X</td>
<td>Bone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCE1325</td>
<td>Delphinapterus leucas</td>
<td>Nuussuaq</td>
<td>76.173</td>
<td>-51.846</td>
<td>1994</td>
<td>X</td>
<td>Bone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCE1326</td>
<td>Delphinapterus leucas</td>
<td>Nuussuaq</td>
<td>76.173</td>
<td>-51.846</td>
<td>1994</td>
<td>X</td>
<td>Bone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCE1329</td>
<td>Delphinapterus leucas</td>
<td>Nuussuaq</td>
<td>76.173</td>
<td>-51.846</td>
<td>1994</td>
<td>X</td>
<td>Bone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCE1330</td>
<td>Delphinapterus leucas</td>
<td>Nuussuaq</td>
<td>76.173</td>
<td>-51.846</td>
<td>1994</td>
<td>X</td>
<td>Bone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCE1331</td>
<td>Delphinapterus leucas</td>
<td>Nuussuaq</td>
<td>76.173</td>
<td>-51.846</td>
<td>1994</td>
<td>X</td>
<td>Bone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCE1332</td>
<td>Delphinapterus leucas</td>
<td>Nuussuaq</td>
<td>76.173</td>
<td>-51.846</td>
<td>1994</td>
<td>X</td>
<td>Bone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCE1333</td>
<td>Delphinapterus leucas</td>
<td>Nuussuaq</td>
<td>76.173</td>
<td>-51.846</td>
<td>1994</td>
<td>X</td>
<td>Bone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCE1358</td>
<td>Monodon monoceros</td>
<td>Uummannaq</td>
<td>70.888</td>
<td>-52.722</td>
<td>1993</td>
<td>X</td>
<td>Bone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCE1359</td>
<td>Monodon monoceros</td>
<td>Uummannaq</td>
<td>70.888</td>
<td>-52.722</td>
<td>1993</td>
<td>X</td>
<td>Bone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCE1360</td>
<td>Monodon monoceros</td>
<td>Uummannaq</td>
<td>70.888</td>
<td>-52.722</td>
<td>1993</td>
<td>X</td>
<td>Bone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCE1361</td>
<td>Monodon monoceros</td>
<td>Uummannaq</td>
<td>70.888</td>
<td>-52.722</td>
<td>1993</td>
<td>X</td>
<td>Bone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCE1362</td>
<td>Monodon monoceros</td>
<td>Uummannaq</td>
<td>70.888</td>
<td>-52.722</td>
<td>1993</td>
<td>X</td>
<td>Bone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCE1363</td>
<td>Monodon monoceros</td>
<td>Uummannaq</td>
<td>70.888</td>
<td>-52.722</td>
<td>1993</td>
<td>X</td>
<td>Bone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCE1364</td>
<td>Monodon monoceros</td>
<td>Balgoni</td>
<td>76.173</td>
<td>-62.446</td>
<td>1993</td>
<td>X</td>
<td>Bone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCE1365</td>
<td>Monodon monoceros</td>
<td>Balgoni</td>
<td>76.173</td>
<td>-62.446</td>
<td>1993</td>
<td>X</td>
<td>Bone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCE1366</td>
<td>Monodon monoceros</td>
<td>Balgoni</td>
<td>76.173</td>
<td>-62.446</td>
<td>1993</td>
<td>X</td>
<td>Bone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCE1380</td>
<td>Monodon monoceros</td>
<td>Qaanuaq</td>
<td>77.400</td>
<td>-69.166</td>
<td>2003</td>
<td>X</td>
<td>Bone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCE1382</td>
<td>Monodon monoceros</td>
<td>Qaanuaq</td>
<td>77.400</td>
<td>-69.166</td>
<td>2003</td>
<td>X</td>
<td>Bone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCE1383</td>
<td>Monodon monoceros</td>
<td>Qaanuaq</td>
<td>77.400</td>
<td>-69.166</td>
<td>2003</td>
<td>X</td>
<td>Bone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCE1384</td>
<td>Monodon monoceros</td>
<td>Qaanuaq</td>
<td>77.400</td>
<td>-69.166</td>
<td>2003</td>
<td>X</td>
<td>Bone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCE1386</td>
<td>Monodon monoceros</td>
<td>Qaanuaq</td>
<td>77.400</td>
<td>-69.166</td>
<td>2003</td>
<td>X</td>
<td>Bone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCE1388</td>
<td>Monodon monoceros</td>
<td>Qaanuaq</td>
<td>77.400</td>
<td>-69.166</td>
<td>2003</td>
<td>X</td>
<td>Bone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCE1389</td>
<td>Monodon monoceros</td>
<td>Qaanuaq</td>
<td>77.400</td>
<td>-69.166</td>
<td>2003</td>
<td>X</td>
<td>Bone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCE1374</td>
<td>Monodon monoceros</td>
<td>Qaanuaq</td>
<td>77.400</td>
<td>-69.166</td>
<td>2003</td>
<td>X</td>
<td>Bone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCE1377</td>
<td>Monodon monoceros</td>
<td>Qaanuaq</td>
<td>77.400</td>
<td>-69.166</td>
<td>2003</td>
<td>X</td>
<td>Bone</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Supplementary Figure S3

MapDamage plot of base frequencies and substitutions in DNA sequence reads from MCE1356. (a) Base frequencies inside (surrounded by grey outline) and outside the reads. (b) Substitutions from the 5’ (left) and the 3’ end (right). C -> T substitutions (red), G -> A substitutions (blue), insertions relative to the killer whale reference (purple), and all other substitutions (grey).
Supplementary Document S4

Supplementary Material

**Stable Isotope Analysis — Calibration, Accuracy, and Precision**

Carbon and nitrogen isotopic and elemental compositions were determined using a Nu Horizon continuous flow isotope ratio mass spectrometer. Sample measurements were calibrated relative to VPDB (δ\textsuperscript{13}C) and AIR (δ\textsuperscript{15}N) using USGS40 and USGS41a (Table S1; Qi et al., 2003; Qi et al., 2016).

**Table S1.** Standard reference materials used for calibration of δ\textsuperscript{13}C relative to VPDB and δ\textsuperscript{15}N relative to AIR.

<table>
<thead>
<tr>
<th>Standard</th>
<th>Material</th>
<th>Accepted δ\textsuperscript{13}C (‰, VPDB)</th>
<th>Accepted δ\textsuperscript{15}N (‰, AIR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>USGS40</td>
<td>Glutamic Acid</td>
<td>−26.39±0.04</td>
<td>−4.52±0.06</td>
</tr>
<tr>
<td>USGS41a</td>
<td>Glutamic Acid</td>
<td>+36.55±0.08</td>
<td>+47.55±0.15</td>
</tr>
</tbody>
</table>

The following standards were used to monitor accuracy and precision (Table S2). The isotopic compositions for the internal keratin standard represent long-term averages.

**Table S2.** Standard reference materials used to monitor internal accuracy and precision.

<table>
<thead>
<tr>
<th>Standard</th>
<th>Material</th>
<th>Mean δ\textsuperscript{13}C (‰, VPDB)</th>
<th>Mean δ\textsuperscript{15}N (‰, AIR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRM−1</td>
<td>Caribou bone collagen</td>
<td>−19.36±0.11</td>
<td>+1.81±0.10</td>
</tr>
<tr>
<td>SRM−2</td>
<td>Walrus bone collagen</td>
<td>−14.76±0.12</td>
<td>+15.59±0.11</td>
</tr>
<tr>
<td>SRM−4</td>
<td>Gluten</td>
<td>−26.76±0.08</td>
<td>+5.25±0.11</td>
</tr>
<tr>
<td>SRM−14</td>
<td>Polar bear bone collagen</td>
<td>−13.66±0.06</td>
<td>+21.63±0.11</td>
</tr>
</tbody>
</table>

Table S3 summarizes the mean and standard deviation of carbon and nitrogen isotopic compositions for all check standards, as well as the standard deviation for all calibration standards – the mean of the calibration standard for an individual run is predetermined to calibrate the data.
**Table S3.** Mean and standard deviation of all check and calibration standards for all analytical sessions containing data presented in this paper. Note that means for calibration standards are not presented as they are pre-determined to be equal to the known value.

<table>
<thead>
<tr>
<th>Run ID</th>
<th>Standard</th>
<th>n</th>
<th>$\delta^{13}$C (‰, VPDB)</th>
<th>$\delta^{15}$N (‰, AIR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>18-02</td>
<td>USGS40</td>
<td>5</td>
<td>± 0.01</td>
<td>± 0.25</td>
</tr>
<tr>
<td>18-03</td>
<td>USGS40</td>
<td>5</td>
<td>± 0.13</td>
<td>± 0.06</td>
</tr>
<tr>
<td>18-06</td>
<td>USGS40</td>
<td>8</td>
<td>± 0.11</td>
<td>± 0.06</td>
</tr>
<tr>
<td>18-02</td>
<td>USGS41a</td>
<td>5</td>
<td>± 0.05</td>
<td>± 0.28</td>
</tr>
<tr>
<td>18-03</td>
<td>USGS41a</td>
<td>5</td>
<td>± 0.14</td>
<td>± 0.06</td>
</tr>
<tr>
<td>18-06</td>
<td>USGS41a</td>
<td>8</td>
<td>± 0.05</td>
<td>± 0.03</td>
</tr>
<tr>
<td>18-06</td>
<td>SRM−1</td>
<td>3</td>
<td>−19.39 ± 0.02</td>
<td>+1.72 ± 0.06</td>
</tr>
<tr>
<td>18-02</td>
<td>SRM−2</td>
<td>4</td>
<td>−14.86 ± 0.07</td>
<td>+15.79 ± 0.12</td>
</tr>
<tr>
<td>18-03</td>
<td>SRM−2</td>
<td>4</td>
<td>−14.84 ± 0.08</td>
<td>+15.62 ± 0.07</td>
</tr>
<tr>
<td>18-06</td>
<td>SRM−2</td>
<td>6</td>
<td>−14.78 ± 0.03</td>
<td>+15.50 ± 0.03</td>
</tr>
<tr>
<td>18-02</td>
<td>SRM−4</td>
<td>6</td>
<td>−26.76 ± 0.02</td>
<td>+5.19 ± 0.11</td>
</tr>
<tr>
<td>18-03</td>
<td>SRM−4</td>
<td>5</td>
<td>−26.72 ± 0.12</td>
<td>+5.28 ± 0.04</td>
</tr>
<tr>
<td>18-02</td>
<td>SRM−14</td>
<td>4</td>
<td>−13.65 ± 0.04</td>
<td>+21.78 ± 0.07</td>
</tr>
<tr>
<td>18-03</td>
<td>SRM−14</td>
<td>4</td>
<td>−13.70 ± 0.09</td>
<td>+21.59 ± 0.11</td>
</tr>
<tr>
<td>18-06</td>
<td>SRM−14</td>
<td>6</td>
<td>−13.71 ± 0.03</td>
<td>+21.41 ± 0.05</td>
</tr>
</tbody>
</table>

Twenty-eight samples were analyzed in duplicate. The pooled standard deviation for the duplicate pairs was ±0.15 ‰ for $\delta^{13}$C and 0.12 ‰ for $\delta^{15}$N.

**Analytical Uncertainty**

Standard uncertainty was calculated using the method present by Szpak et al. (2017b), which largely follows Magnusson et al.’s (2012) approach. Standard uncertainty was determined to be ±0.17 ‰ for $\delta^{13}$C and ±0.22 ‰ for $\delta^{15}$N.

**Results**


The carbon and nitrogen isotopic and elemental compositions of all samples analyzed are presented in Table S4. Most of the samples were analyzed in duplicate and the results of each analysis is presented separately in Table S4. Four samples that were analyzed were excluded from subsequent analysis because they presented unusually high atomic C:N ratios and are identified with strikethrough text in Table S4. The C:N ratios of archaeological specimens are used to assess the integrity of the collagen as contaminated or degraded collagen will tend to have isotopic compositions that are significantly different from the original, endogenous material (Ambrose, 1990; DeNiro, 1985; Szpak et al., 2017a; van Klinken and Hedges, 1995). For modern material, collagen degradation is not an issue but contamination with lipids is a potential concern as the $\delta^{13}C$ values of lipids are significantly lower than those of proteins such as collagen (Post et al., 2007). Even though these samples were treated with solvents that are effective at removing lipids from bone (Guiry et al., 2016), some still had atomic C:N ratios that suggested contamination with additional material that was rich in carbon, lacking in nitrogen, and had a low $\delta^{13}C$ value – most likely lipids. The samples that were excluded had atomic C:N ratios in the range of 3.56 to 3.92, whereas the theoretical value for pure mammalian bone collagen is 3.23 (Szpak, 2011). These samples also displayed the lowest $\delta^{13}C$ values of any members of the taxon to which they belonged. Consequently the isotopic compositions were considered unreliable and were excluded from plots and statistical calculations.
Table S4. Stable isotopic and elemental compositions for all samples analyzed in this study.

<table>
<thead>
<tr>
<th>Lab ID</th>
<th>Sample ID (MCE)</th>
<th>Location</th>
<th>Year</th>
<th>Taxon</th>
<th>δ(^{13})C (‰)</th>
<th>δ(^{15})N (‰)</th>
<th>wt%C</th>
<th>wt%N</th>
<th>Atomic C:N</th>
<th>δ(^{13})C (‰)</th>
<th>δ(^{15})N (‰)</th>
<th>wt%C</th>
<th>wt%N</th>
<th>Atomic C:N</th>
</tr>
</thead>
<tbody>
<tr>
<td>571</td>
<td>1312</td>
<td>Kullorsuaq</td>
<td>1990</td>
<td>Delphinapterus leucas</td>
<td>-15.04</td>
<td>17.46</td>
<td>49.30</td>
<td>16.90</td>
<td>3.40</td>
<td>-15.12</td>
<td>17.25</td>
<td>35.0</td>
<td>12.4</td>
<td>3.29</td>
</tr>
<tr>
<td>572</td>
<td>1313</td>
<td>Kullorsuaq</td>
<td>1990</td>
<td>Delphinapterus leucas</td>
<td>-14.77</td>
<td>17.19</td>
<td>36.40</td>
<td>12.90</td>
<td>3.29</td>
<td>-14.76</td>
<td>17.25</td>
<td>35.0</td>
<td>12.4</td>
<td>3.29</td>
</tr>
<tr>
<td>573</td>
<td>1314</td>
<td>Kullorsuaq</td>
<td>1991</td>
<td>Delphinapterus leucas</td>
<td>-14.47</td>
<td>17.00</td>
<td>34.03</td>
<td>12.58</td>
<td>3.15</td>
<td>-14.56</td>
<td>17.19</td>
<td>34.7</td>
<td>12.7</td>
<td>3.19</td>
</tr>
<tr>
<td>587</td>
<td>1316</td>
<td>Nuusuaaq</td>
<td>1994</td>
<td>Delphinapterus leucas</td>
<td>-14.68</td>
<td>18.44</td>
<td>44.45</td>
<td>15.38</td>
<td>3.37</td>
<td>-14.76</td>
<td>17.25</td>
<td>34.7</td>
<td>12.7</td>
<td>3.19</td>
</tr>
<tr>
<td>588</td>
<td>1318</td>
<td>Nuusuaaq</td>
<td>1994</td>
<td>Delphinapterus leucas</td>
<td>-15.16</td>
<td>15.97</td>
<td>46.09</td>
<td>15.77</td>
<td>3.41</td>
<td>-15.12</td>
<td>15.93</td>
<td>44.8</td>
<td>15.5</td>
<td>3.37</td>
</tr>
<tr>
<td>589</td>
<td>1319</td>
<td>Nuusuaaq</td>
<td>1994</td>
<td>Delphinapterus leucas</td>
<td>-14.72</td>
<td>18.07</td>
<td>44.84</td>
<td>15.70</td>
<td>3.33</td>
<td>-15.09</td>
<td>18.07</td>
<td>46.3</td>
<td>15.8</td>
<td>3.42</td>
</tr>
<tr>
<td>590</td>
<td>1320</td>
<td>Nuusuaaq</td>
<td>1994</td>
<td>Delphinapterus leucas</td>
<td>-14.41</td>
<td>17.87</td>
<td>43.77</td>
<td>15.50</td>
<td>3.29</td>
<td>-14.78</td>
<td>18.20</td>
<td>46.0</td>
<td>15.9</td>
<td>3.37</td>
</tr>
<tr>
<td>591</td>
<td>1321</td>
<td>Nuusuaaq</td>
<td>1994</td>
<td>Delphinapterus leucas</td>
<td>-15.07</td>
<td>18.61</td>
<td>44.97</td>
<td>15.29</td>
<td>3.43</td>
<td>-14.78</td>
<td>18.20</td>
<td>46.0</td>
<td>15.9</td>
<td>3.37</td>
</tr>
<tr>
<td>592</td>
<td>1322</td>
<td>Nuusuaaq</td>
<td>1994</td>
<td>Delphinapterus leucas</td>
<td>-15.24</td>
<td>15.97</td>
<td>44.98</td>
<td>15.03</td>
<td>3.49</td>
<td>-15.21</td>
<td>16.01</td>
<td>45.4</td>
<td>15.0</td>
<td>3.53</td>
</tr>
<tr>
<td>593</td>
<td>1323</td>
<td>Nuusuaaq</td>
<td>1994</td>
<td>Delphinapterus leucas</td>
<td>-14.35</td>
<td>17.12</td>
<td>43.29</td>
<td>15.79</td>
<td>3.20</td>
<td>-14.50</td>
<td>17.39</td>
<td>44.2</td>
<td>15.9</td>
<td>3.24</td>
</tr>
<tr>
<td>594</td>
<td>1324</td>
<td>Nuusuaaq</td>
<td>1994</td>
<td>Delphinapterus leucas</td>
<td>-14.37</td>
<td>17.59</td>
<td>43.43</td>
<td>16.01</td>
<td>3.16</td>
<td>-14.76</td>
<td>17.73</td>
<td>44.7</td>
<td>16.0</td>
<td>3.26</td>
</tr>
<tr>
<td>595</td>
<td>1325</td>
<td>Nuusuaaq</td>
<td>1994</td>
<td>Delphinapterus leucas</td>
<td>-14.38</td>
<td>17.17</td>
<td>44.78</td>
<td>16.16</td>
<td>3.23</td>
<td>-14.52</td>
<td>17.27</td>
<td>45.3</td>
<td>16.1</td>
<td>3.28</td>
</tr>
<tr>
<td>596</td>
<td>1326</td>
<td>Nuusuaaq</td>
<td>1994</td>
<td>Delphinapterus leucas</td>
<td>-14.77</td>
<td>17.34</td>
<td>44.75</td>
<td>15.82</td>
<td>3.30</td>
<td>-15.06</td>
<td>17.68</td>
<td>45.9</td>
<td>16.0</td>
<td>3.35</td>
</tr>
<tr>
<td>599</td>
<td>1329</td>
<td>Nuusuaaq</td>
<td>1994</td>
<td>Delphinapterus leucas</td>
<td>-14.71</td>
<td>17.67</td>
<td>45.08</td>
<td>15.76</td>
<td>3.34</td>
<td>-14.81</td>
<td>16.55</td>
<td>43.9</td>
<td>15.3</td>
<td>3.35</td>
</tr>
<tr>
<td>600</td>
<td>1330</td>
<td>Nuusuaaq</td>
<td>1994</td>
<td>Delphinapterus leucas</td>
<td>-14.71</td>
<td>16.79</td>
<td>44.32</td>
<td>15.56</td>
<td>3.32</td>
<td>-14.69</td>
<td>16.85</td>
<td>44.1</td>
<td>15.9</td>
<td>3.24</td>
</tr>
<tr>
<td>601</td>
<td>1331</td>
<td>Nuusuaaq</td>
<td>1994</td>
<td>Delphinapterus leucas</td>
<td>-14.68</td>
<td>17.72</td>
<td>44.93</td>
<td>16.46</td>
<td>3.18</td>
<td>-14.83</td>
<td>17.81</td>
<td>43.4</td>
<td>15.6</td>
<td>3.24</td>
</tr>
<tr>
<td>602</td>
<td>1332</td>
<td>Nuusuaaq</td>
<td>1994</td>
<td>Delphinapterus leucas</td>
<td>-14.51</td>
<td>16.35</td>
<td>43.65</td>
<td>15.92</td>
<td>3.20</td>
<td>-14.81</td>
<td>16.55</td>
<td>43.9</td>
<td>15.3</td>
<td>3.35</td>
</tr>
<tr>
<td>No.</td>
<td>Location</td>
<td>Year</td>
<td>Species</td>
<td>T1</td>
<td>T2</td>
<td>T3</td>
<td>T4</td>
<td>T5</td>
<td>T6</td>
<td>T7</td>
<td>T8</td>
<td>T9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>------------</td>
<td>------</td>
<td>--------------------------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td></td>
<td></td>
</tr>
<tr>
<td>603</td>
<td>Nuusuaaq</td>
<td>1994</td>
<td>Delphinapterus leucas</td>
<td>-14.53</td>
<td>16.99</td>
<td>44.58</td>
<td>16.18</td>
<td>3.21</td>
<td>-14.61</td>
<td>17.06</td>
<td>43.3</td>
<td>15.5</td>
<td>3.26</td>
<td></td>
</tr>
<tr>
<td>575</td>
<td>Kitsissuarsuit</td>
<td>1986</td>
<td>Delphinapterus leucas X Monodon monoceros</td>
<td>-13.78</td>
<td>17.40</td>
<td>44.84</td>
<td>16.33</td>
<td>3.20</td>
<td>-13.91</td>
<td>17.64</td>
<td>44.8</td>
<td>16.5</td>
<td>3.17</td>
<td></td>
</tr>
<tr>
<td>576</td>
<td>Uummannaq</td>
<td>1993</td>
<td>Monodon monoceros</td>
<td>-14.77</td>
<td>17.17</td>
<td>44.27</td>
<td>15.83</td>
<td>3.26</td>
<td>-14.97</td>
<td>17.31</td>
<td>45.8</td>
<td>16.3</td>
<td>3.28</td>
<td></td>
</tr>
<tr>
<td>577</td>
<td>Uummannaq</td>
<td>1993</td>
<td>Monodon monoceros</td>
<td>-15.83</td>
<td>17.48</td>
<td>45.5</td>
<td>15.3</td>
<td>3.47</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>580</td>
<td>Uummannaq</td>
<td>1993</td>
<td>Monodon monoceros</td>
<td>-15.15</td>
<td>17.72</td>
<td>44.04</td>
<td>15.53</td>
<td>3.31</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>581</td>
<td>Uummannaq</td>
<td>1993</td>
<td>Monodon monoceros</td>
<td>-15.07</td>
<td>16.79</td>
<td>45.45</td>
<td>16.49</td>
<td>3.21</td>
<td>-15.21</td>
<td>17.04</td>
<td>44.7</td>
<td>16.1</td>
<td>3.24</td>
<td></td>
</tr>
<tr>
<td>582</td>
<td>Uummannaq</td>
<td>1993</td>
<td>Monodon monoceros</td>
<td>-15.14</td>
<td>16.68</td>
<td>42.78</td>
<td>15.58</td>
<td>3.20</td>
<td>-15.34</td>
<td>16.70</td>
<td>44.4</td>
<td>15.8</td>
<td>3.28</td>
<td></td>
</tr>
<tr>
<td>583</td>
<td>Uummannaq</td>
<td>1993</td>
<td>Monodon monoceros</td>
<td>-15.40</td>
<td>16.62</td>
<td>44.30</td>
<td>15.74</td>
<td>3.28</td>
<td>-15.72</td>
<td>16.91</td>
<td>45.8</td>
<td>15.8</td>
<td>3.38</td>
<td></td>
</tr>
<tr>
<td>584</td>
<td>Balgoni</td>
<td>1993</td>
<td>Monodon monoceros</td>
<td>-15.81</td>
<td>16.38</td>
<td>42.84</td>
<td>15.04</td>
<td>3.32</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>585</td>
<td>Balgoni</td>
<td>1993</td>
<td>Monodon monoceros</td>
<td>-15.34</td>
<td>17.38</td>
<td>44.79</td>
<td>15.94</td>
<td>3.28</td>
<td>-15.51</td>
<td>17.70</td>
<td>44.7</td>
<td>15.9</td>
<td>3.28</td>
<td></td>
</tr>
<tr>
<td>586</td>
<td>Balgoni</td>
<td>1993</td>
<td>Monodon monoceros</td>
<td>-14.82</td>
<td>17.50</td>
<td>43.65</td>
<td>15.75</td>
<td>3.23</td>
<td>-15.05</td>
<td>17.61</td>
<td>45.2</td>
<td>16.1</td>
<td>3.27</td>
<td></td>
</tr>
<tr>
<td>605</td>
<td>Qaanaaq</td>
<td>2003</td>
<td>Monodon monoceros</td>
<td>-15.42</td>
<td>16.55</td>
<td>42.71</td>
<td>15.04</td>
<td>3.31</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>607</td>
<td>Qaanaaq</td>
<td>2003</td>
<td>Monodon monoceros</td>
<td>-15.14</td>
<td>16.68</td>
<td>44.56</td>
<td>15.54</td>
<td>3.34</td>
<td>-15.13</td>
<td>16.71</td>
<td>44.9</td>
<td>15.7</td>
<td>3.33</td>
<td></td>
</tr>
<tr>
<td>608</td>
<td>Qaanaaq</td>
<td>2003</td>
<td>Monodon monoceros</td>
<td>-15.22</td>
<td>16.05</td>
<td>44.68</td>
<td>15.87</td>
<td>3.28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>609</td>
<td>Qaanaaq</td>
<td>2003</td>
<td>Monodon monoceros</td>
<td>-14.82</td>
<td>16.90</td>
<td>43.93</td>
<td>16.16</td>
<td>3.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>610</td>
<td>Qaanaaq</td>
<td>2003</td>
<td>Monodon monoceros</td>
<td>-15.03</td>
<td>16.79</td>
<td>43.46</td>
<td>15.95</td>
<td>3.18</td>
<td>-15.20</td>
<td>16.84</td>
<td>43.9</td>
<td>15.7</td>
<td>3.26</td>
<td></td>
</tr>
<tr>
<td>611</td>
<td>Qaanaaq</td>
<td>2003</td>
<td>Monodon monoceros</td>
<td>-14.92</td>
<td>16.65</td>
<td>42.02</td>
<td>15.67</td>
<td>3.13</td>
<td>-14.83</td>
<td>16.61</td>
<td>43.2</td>
<td>15.7</td>
<td>3.21</td>
<td></td>
</tr>
<tr>
<td>612</td>
<td>Qaanaaq</td>
<td>2003</td>
<td>Monodon monoceros</td>
<td>-14.80</td>
<td>16.14</td>
<td>44.72</td>
<td>16.48</td>
<td>3.16</td>
<td>-14.92</td>
<td>16.12</td>
<td>43.3</td>
<td>15.8</td>
<td>3.20</td>
<td></td>
</tr>
<tr>
<td>613</td>
<td>Qaanaaq</td>
<td>2003</td>
<td>Monodon monoceros</td>
<td>-15.11</td>
<td>16.24</td>
<td>42.84</td>
<td>15.49</td>
<td>3.22</td>
<td>-15.60</td>
<td>16.41</td>
<td>45.4</td>
<td>15.4</td>
<td>3.44</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Qaanaaq</td>
<td>2003</td>
<td>Monodon monoceros</td>
<td>-15.86</td>
<td>16.73</td>
<td>44.61</td>
<td>16.12</td>
<td>3.23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
References


Supplementary Table S5

Mapping coverage information. The genomic data were mapped to the mitochondrial reference genomes of beluga (KY444734) and narwhal (NC_005279), and the nuclear genome of killer whale (GCA_000331955.2).

<table>
<thead>
<tr>
<th>Lab ID</th>
<th>Species</th>
<th>Delphinapterus leucas mitogenome</th>
<th>Monodon monoceros mitogenome</th>
<th>Orcinus orca nuclear genome</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Total number of reads</td>
<td>Total number of retained reads</td>
<td>Total number of unique reads mapping</td>
</tr>
<tr>
<td>CGG 1 021706</td>
<td>Delphinapterus leucas x Monodon monoceros</td>
<td>8507904</td>
<td>8287568</td>
<td>1687</td>
</tr>
<tr>
<td>CGG 1 017580</td>
<td>Delphinapterus leucas</td>
<td>11533738</td>
<td>9261136</td>
<td>19303</td>
</tr>
<tr>
<td>CGG 1 017581</td>
<td>Delphinapterus leucas</td>
<td>12621088</td>
<td>12941166</td>
<td>15553</td>
</tr>
<tr>
<td>CGG 1 017582</td>
<td>Delphinapterus leucas</td>
<td>12283228</td>
<td>1255413</td>
<td>21591</td>
</tr>
<tr>
<td>CGG 1 017583</td>
<td>Delphinapterus leucas</td>
<td>5086697</td>
<td>5079700</td>
<td>10779</td>
</tr>
<tr>
<td>CGG 1 017584</td>
<td>Delphinapterus leucas</td>
<td>10705418</td>
<td>10690868</td>
<td>9899</td>
</tr>
<tr>
<td>CGG 1 017585</td>
<td>Delphinapterus leucas</td>
<td>13025709</td>
<td>13597923</td>
<td>9696</td>
</tr>
<tr>
<td>CGG 1 017586</td>
<td>Delphinapterus leucas</td>
<td>11537672</td>
<td>11456320</td>
<td>15730</td>
</tr>
<tr>
<td>CGG 1 017587</td>
<td>Delphinapterus leucas</td>
<td>12128988</td>
<td>12100941</td>
<td>14519</td>
</tr>
<tr>
<td>CGG 1 017557</td>
<td>Monodon monoceros</td>
<td>8263484</td>
<td>825347</td>
<td>12603</td>
</tr>
<tr>
<td>CGG 1 017558</td>
<td>Monodon monoceros</td>
<td>6456222</td>
<td>643277</td>
<td>12501</td>
</tr>
<tr>
<td>CGG 1 017559</td>
<td>Monodon monoceros</td>
<td>6141963</td>
<td>608543</td>
<td>26177</td>
</tr>
<tr>
<td>CGG 1 017560</td>
<td>Monodon monoceros</td>
<td>8426163</td>
<td>8388874</td>
<td>12227</td>
</tr>
<tr>
<td>CGG 1 017561</td>
<td>Monodon monoceros</td>
<td>1036194</td>
<td>1032587</td>
<td>28340</td>
</tr>
<tr>
<td>CGG 1 017562</td>
<td>Monodon monoceros</td>
<td>8232802</td>
<td>8144938</td>
<td>20861</td>
</tr>
<tr>
<td>CGG 1 017563</td>
<td>Monodon monoceros</td>
<td>8702405</td>
<td>8684282</td>
<td>10694</td>
</tr>
<tr>
<td>CGG 1 017564</td>
<td>Monodon monoceros</td>
<td>10165750</td>
<td>10148446</td>
<td>10261</td>
</tr>
</tbody>
</table>
Supplementary Figure S6

Median-spanning haplotype network of complete mitochondrial genomes of eight belugas (blue), eight narwhals (green) and MCE1356 mapped to the beluga reference (red). Black dots indicate intermediate haplotypes not found in the data. The size of the circles indicate the relative number of specimens sharing a haplotype. Numbers indicate number of variable sites between haplotypes.
Supplementary Figure S7

Combined read depth distribution of MCE1356, eight belugas and eight narwhals. Mean value is 4.14.