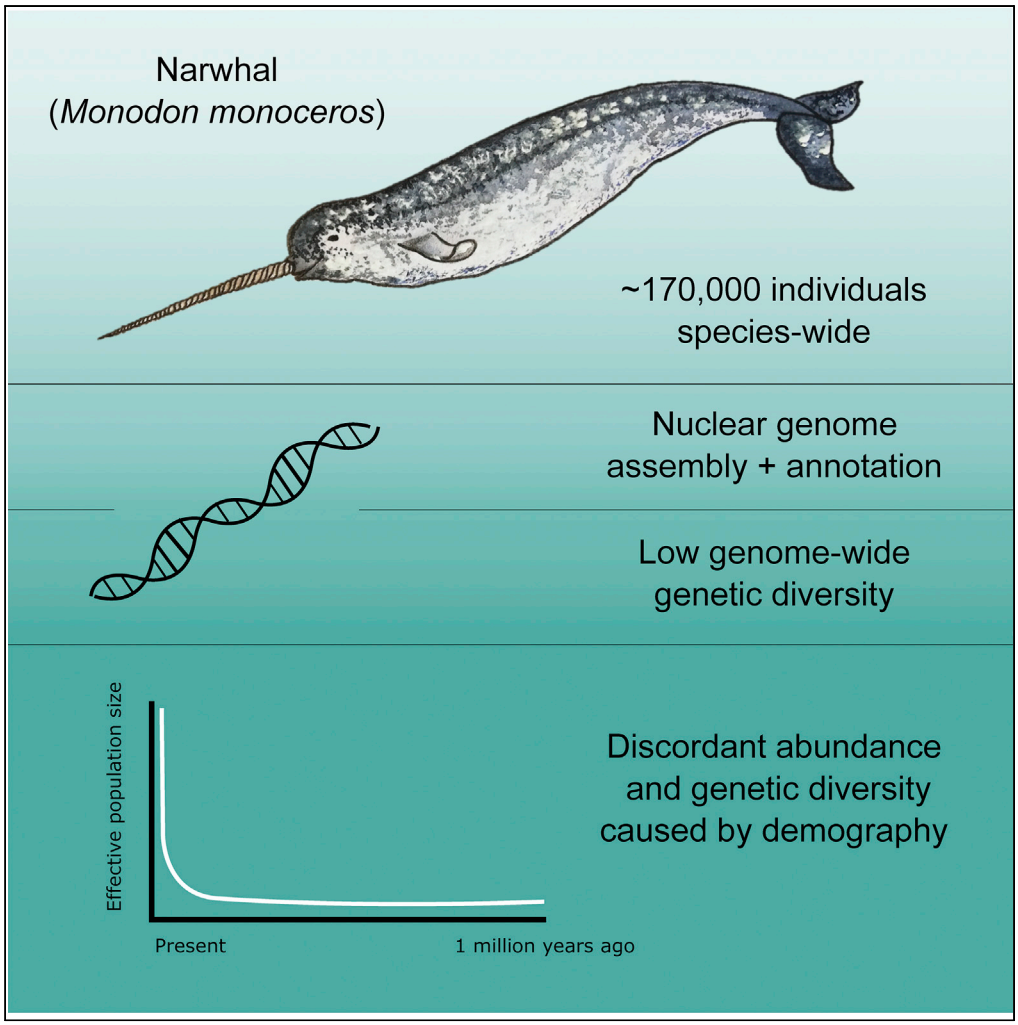
iScience

CellPress

Article

Narwhal Genome Reveals Long-Term Low Genetic Diversity despite Current Large Abundance Size



Michael V. Westbury, Bent Petersen, Eva Garde, Mads Peter Heide-Jørgensen, Eline D. Lorenzen

m.westbury@snm.ku.dk (M.V.W.)
elinelorenzen@snm.ku.dk (E.D.L.)

HIGHLIGHTS

Assembled and annotated narwhal nuclear genome

Low genetic diversity despite large global abundance

Lack of signs of inbreeding

Demography is the driving force behind the low diversity

Westbury et al., iScience *15*, 592–599
May 31, 2019
https://doi.org/10.1016/j.isci.2019.03.023

iScience

CellPress

Article

Narwhal Genome Reveals Long-Term Low Genetic Diversity despite Current Large Abundance Size

Michael V. Westbury,[1,*] Bent Petersen,[1,2] Eva Garde,[3] Mads Peter Heide-Jørgensen,[1,3] and Eline D. Lorenzen[1,4,*]

SUMMARY

The narwhal (*Monodon monoceros*) is a highly specialized endemic Arctic cetacean, restricted to the Arctic seas bordering the North Atlantic. Low levels of genetic diversity have been observed across several narwhal populations using mitochondrial DNA and microsatellites. Despite this, the global abundance of narwhals was recently estimated at ~170,000 individuals. However, the species is still considered vulnerable to changing climates due to its high specialization and restricted Arctic distribution. We assembled and annotated a genome from a narwhal from West Greenland. We find relatively low diversity at the genomic scale and show that this did not arise by recent inbreeding, but rather has been stable over an extended evolutionary timescale. We also find that the current large global abundance most likely reflects a recent rapid expansion from a much smaller founding population.

INTRODUCTION

Adaptive potential in the form of genetic diversity is assumed to be essential (Reed and Frankham, 2003; Ellegren and Galtier, 2016), especially during periods of rapid climatic change. The Arctic and its biodiversity are highly vulnerable to these shifts (Masson-Delmotte et al., 2018; Kovacs et al., 2011). The narwhal (*Monodon monoceros*) is the most specialized endemic Arctic cetacean, well known for its elongated caniniform "tusk" and deep diving capabilities (Heide-Jørgensen, 2009; Heide-Jørgensen & Dietz, 1995; Best, 1981). Narwhals are distributed in the Arctic seas bordering the North Atlantic (Figure 1A) and make annual migrations between coastal summering grounds and offshore ice-covered wintering grounds. Low levels of genetic diversity have been observed across several narwhal populations using mitochondrial DNA (mtDNA) and microsatellites (Palsbøll et al., 1997; Petersen et al., 2011), despite the global abundance of narwhals being relatively high. A recent global abundance estimate of ~170,000 individuals (NAMMCO, 2018) resulted in a change in the category of narwhals in the International Union for Conservation of Nature red list of threatened species, where the species has been downgraded from "near threatened" to "least concern" in 2017 (Lowry et al., 2017). This classification, however, did not take genetic diversity and the future genetic adaptability of the narwhal into account. Therefore despite the downgrade in conservation status, the narwhal should still be considered one of the most vulnerable Arctic marine mammals to ongoing rapid climate changes, especially due to its high specialization and restricted Arctic-Atlantic distribution (Laidre et al., 2008).

The finding of low genetic diversity despite large global abundance appears to contradict the neutral theory of molecular evolution (Kimura, 1968). The neutral theory states that neutral processes (e.g., genetic drift) are the main determinants of genetic diversity. If this is indeed the case, then the larger the population, the lower the chance that genetic drift will randomly remove genetic diversity from the gene pool. This therefore would suggest that genetic diversity is directly proportional to effective population size (Ne). Although this appears not the case in the narwhal, low genetic diversity despite large abundance could be explained by other factors, including (1) the number of breeding individuals, and thereby Ne, being very small despite the species' large abundance; (2) high levels of inbreeding; or (3) a recent population expansion from a small founding population. The first option seems unlikely as this infers that most adults from the ~170,000 individuals do not reproduce. To investigate options 2 and 3, we assembled and annotated a nuclear genome from a narwhal from West Greenland. We find that the low diversity did not arise by recent inbreeding, but rather has been stable over an extended evolutionary timescale. Our analyses show that the current large global abundance most likely arose due to a recent rapid population expansion

[1]Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, 1350 Copenhagen K, Denmark

[2]Centre of Excellence for Omics-Driven Computational Biodiscovery (COMBio), Faculty of Applied Sciences, AIMST University, Kedah, Malaysia

[3]Greenland Institute of Natural Resources, Strandgade 91,2, 1401 Copenhagen K, Denmark

[4]Lead Contact

*Correspondence: m.westbury@snm.ku.dk (M.V.W.), elinelorenzen@snm.ku.dk (E.D.L.)

https://doi.org/10.1016/j.isci.2019.03.023

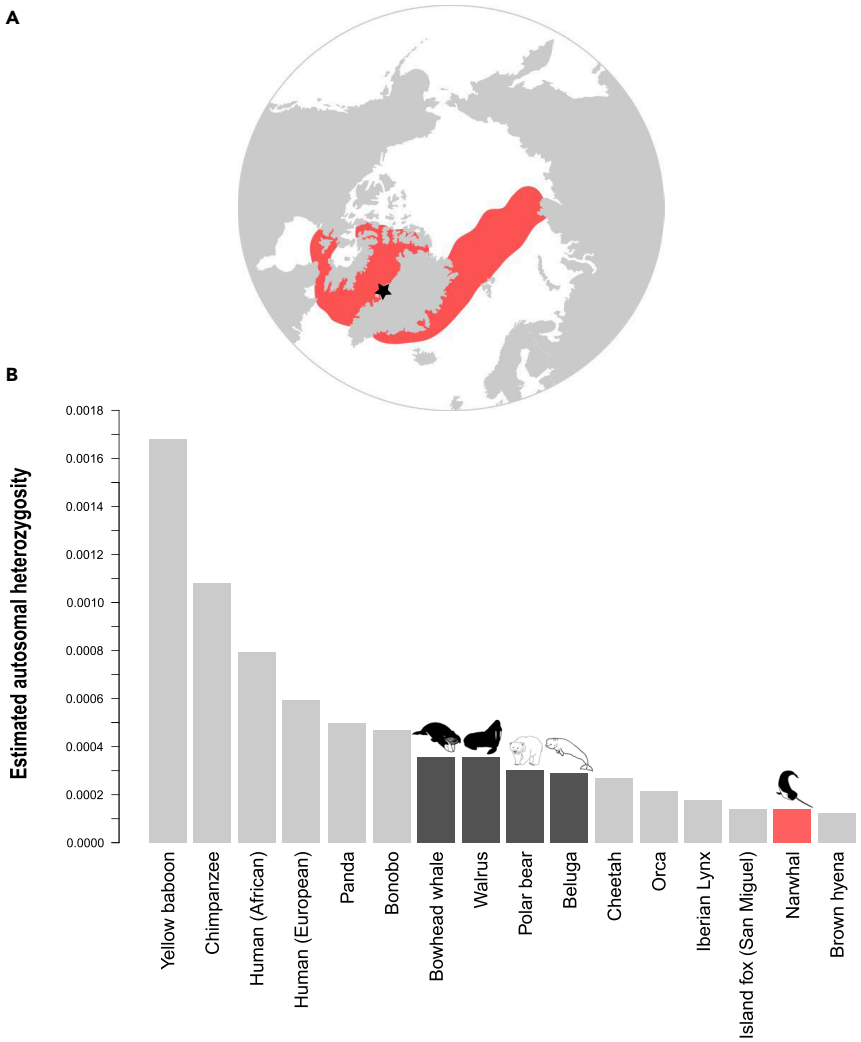


Figure 1. Sampling Location of the Narwhal and Comparative Genome-wide Diversity of 16 Mammalian Species

(A) Map of the narwhal distribution range based on International Union for Conservation of Nature distribution data (Lowry et al., 2017).

(B) Average genome-wide autosomal heterozygosity values of a range of available mammalian genomes. y axis represents the average proportion of sites within the autosomes that are heterozygous. Light gray bar, non-Arctic endemics; dark gray bar, Arctic endemics; red bar, narwhal. Black star in (A) indicates sample location.

around the onset of the Last Glacial Maximum and that genetic diversity may not have had time to increase accordingly.

RESULTS AND DISCUSSION

Narwhal Genome Assembly

We assembled 2,350,959,615 base pairs (bp) of the narwhal nuclear genome (2,156,711,577 bp excluding missing data [N's]) with a scaffold N50 of 1,483,363 bp and contig N50 of 10,481 bp, in a total of 21,007 scaffolds. We used ~100× coverage of multiple short insert and ~32× of mate paired Illumina libraries and *in silico* mate paired libraries (Grau et al., 2018) constructed using the beluga reference genome (Jones et al., 2017) (Genbank: GCA_002288925.2) (Table S1). Investigations into the completeness of the assembly using Benchmarking Universal Single-Copy Orthologs (BUSCO) analyses and the mammalian BUSCO gene set showed a high level of complete BUSCO scores (93%) (Table S2), indicating a fairly complete and high-quality genome. Repeat profiling results showed that our assembled genome consists of 37.9% repetitive elements (Table S3). We identified a total of 21,785 protein-coding genes through genome annotations

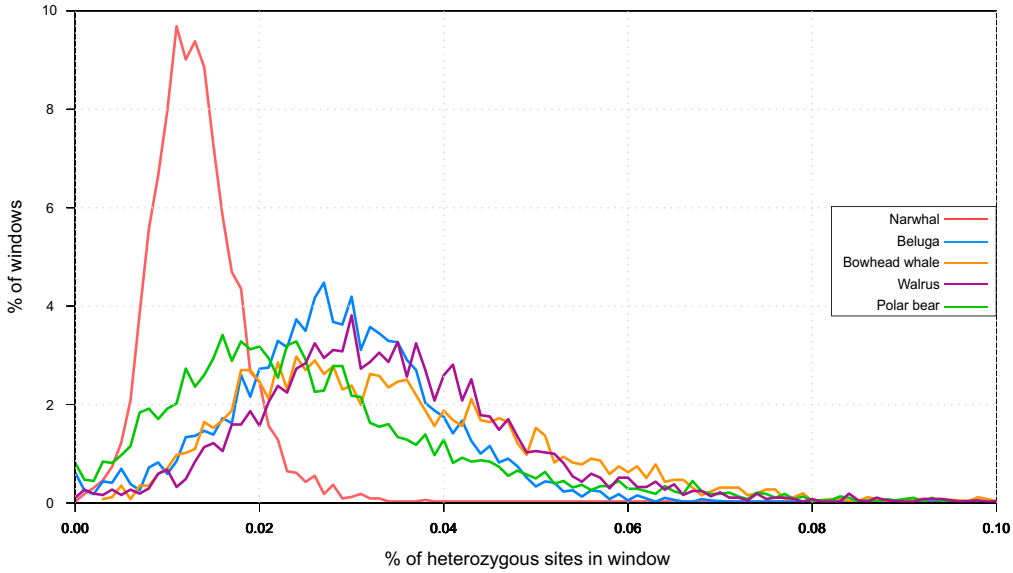


Figure 2. The 500-kb Sliding Window Heterozygosity across the Genomes of Five Arctic Endemic Marine Mammal Species

x axis represents the percentage of the window that is heterozygous, and y axis represents the percentage of the total windows.

with MAKER2 (Holt and Yandell, 2011). These protein-coding genes spanned a total of ~612.8 Mb of the assembled narwhal genome, with exons accounting for ~31.5 Mb.

Genome-wide Genetic Diversity

To investigate genome-wide levels of genetic diversity in the narwhal, we estimated autosomal heterozygosity across the genome using site allele frequency likelihoods with the software, analyzing next generation sequencing data (ANGSD) (Korneliussen et al., 2014). The estimated heterozygosity value of 0.000138 was very low, when compared with available heterozygosity estimates from 11 other mammalian genomes (Westbury et al., 2018) and three additional endemic Arctic marine mammals: beluga (*Delphinapterus leucas*), bowhead whale (*Balaena mysticetus*), and walrus (*Odobenus rosmarus*) (Figure 1B). This result is unexpected, owing to the large census size of narwhals (Lowry et al., 2017; NAMMCO, 2018). The genome-wide low heterozygosity level is, however, in agreement with previous findings based on mtDNA and microsatellites (Palsbøll et al., 1997; Petersen et al., 2011). Low genetic diversity despite wide distribution ranges and large population sizes has been reported in other large mammals (e.g., the brown hyena, Westbury et al., 2018; orca, Hoelzel et al., 2002), and may become a more common finding as more genomic datasets become available.

To investigate whether adaptive potential (i.e., genetic diversity) may be present in certain regions of the narwhal genome, we estimated heterozygosity levels across the genome using 500-kb non-overlapping windows. A relatively high number of regions of high heterozygosity among a general genome-wide pattern of low diversity could suggest that these regions are tied to the adaptive potential of the species and therefore need to retain these higher levels of diversity for the species to persist. We find that heterozygosity follows a normal distribution with very little variation (SD = 0.0000643), indicating an even distribution of diversity across the genome (Figure 2). To uncover whether the evenly distributed genome-wide low diversity is unique to the narwhal, or whether it is somehow associated with its Arctic existence, we compared this distribution to that of four other endemic Arctic marine mammal species (beluga, bowhead whale, walrus, and polar bear [*Ursus maritimus*]) (Figure 2). We find relatively less variability in the distribution of heterozygosity across the narwhal genome (Table S4). This pattern is unique to the narwhal, as the other species have more similar variations in heterozygosity across their genomes. This finding is consistent regardless of whether our narwhal or the beluga genome (Jones et al., 2017) was used as the mapping reference (Figure S1).

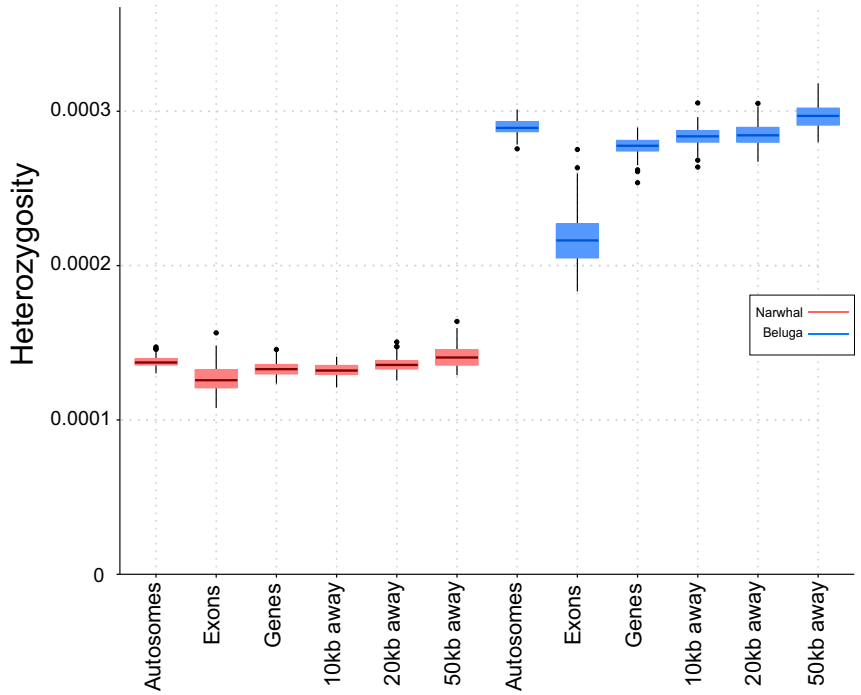


Figure 3. Box Plot Graphs Showing the Proportions of Heterozygosity across Different Genomic Regions in the Narwhal and Beluga

Data were based on sampling and resampling 10% of the windows in the stated region 100 times randomly.

Inbreeding has been shown to be a major driver in the loss of genetic diversity following population bottlenecks (Palkopoulou et al., 2015; Abascal et al., 2016), and is a well-known factor in the reduction of survivability and reproductive success in naturally outbreeding species (Frankham, 2005). We investigated the 500-kb sliding windows for any indications of inbreeding. We find that only 0.03% of the 500-kb windows lack heterozygosity, indicating that inbreeding is most likely not the cause of low heterozygosity in the narwhal. To further validate this, we manually investigated for runs of homozygosity across the five largest scaffolds and found no contiguous stretches of homozygosity (Figure S2). Thus, it appears that, even though the genetic diversity of the narwhal is low, it may have arisen in a manner that allowed the species to slowly adapt over an extended time period, rather than rapidly through inbreeding. This could have implications for the long-term survivability of the species. Inbreeding can rapidly purge alleles from a gene pool that has adaptive potential. However, the slower decrease in diversity in the narwhal could have given the species time to adapt to the decreasing diversity by selectively removing alleles that had little to no influence on the adaptive potential of the species. Our findings appear to reflect the current large abundance of the narwhal, rather than the low levels of genomic diversity. Moreover, similar findings of no inbreeding despite low diversity have previously been reported in other species (Westbury et al., 2018), indicating that this may be an important method of adaptation in multiple species.

To investigate whether the narwhal exhibited higher diversity in regions of putatively higher selective pressure, as opposed to putatively neutral regions, we compared heterozygosity in protein-coding regions (exons) versus gene regions (introns plus exons) and non-coding flanking regions 10, 20, and 50 kb up- and downstream of protein-coding genes (Figure 3, Table S5). To elucidate whether this pattern is unique to the narwhal lineage, we ran the same suite of analyses on the beluga, its sister species (Figure 3, Table S6). The significance of the differences detected in these regions was tested using an unpaired two-sample t test. Significant differences in average heterozygosity between the exons and all other regions tested were found in both the narwhal (Table S7) and the beluga (Table S8). This finding is as expected, as coding regions are more conserved than non-coding regions, due to higher selective pressures. However, even though the p scores suggested significant differences in coding regions compared with non-coding regions in both the narwhal and the beluga, we note that the t scores uncovered in the narwhal are much lower than those found while comparing the same genomic regions in the beluga. This may imply

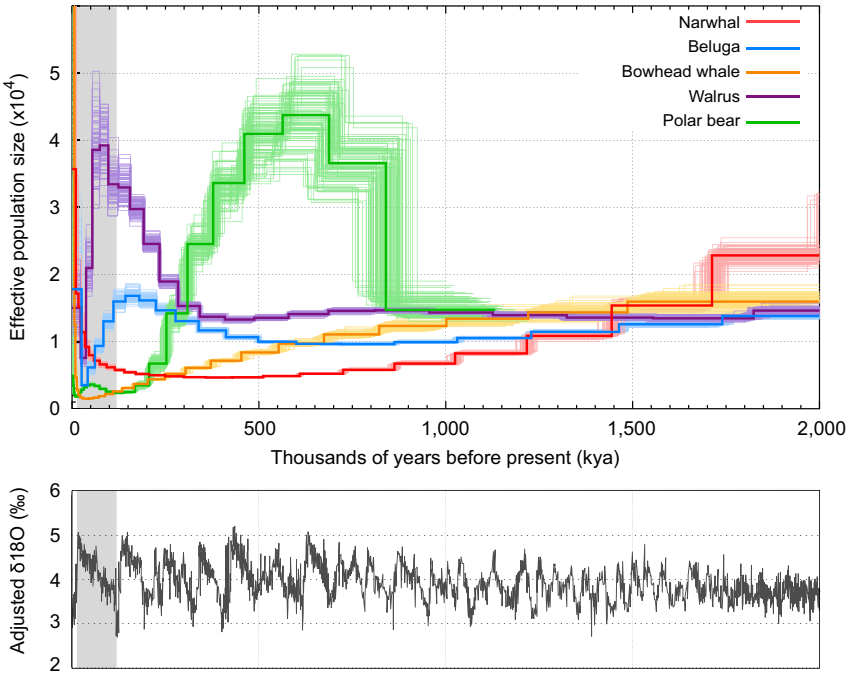


Figure 4. Demographic History of Five Arctic Marine Mammal Species

Pairwise Sequentially Markovian Coalescent model of the narwhal and four other endemic Arctic marine mammal species and $\delta^{18}O$ levels through time based on data found in Zachos et al. (2001). x axis represents thousands of years before present (kya), upper y axis represents effective population size (×10,000), lower y axis represents $\delta^{18}O$ (‰), and faded lines represent bootstrap values. Shaded gray area represents the last glacial period.

a level of retention of diversity in coding regions relative to the non-coding regions. Such retained diversity could help to maintain some adaptive potential in the narwhal, enabling the species to adapt to changes in its environment, despite low genome-wide diversity levels. The overall low levels of heterozygosity and lack of variation in diversity levels between coding and non-coding regions across the narwhal genome, relative to the beluga, may suggest that heterozygosity levels are at a diversity stasis across the genome. Hence, in the narwhal, any further decrease in genomic diversity could lead to survivability problems.

Demographic History

A recent bottleneck could offer an explanation for the current low genetic diversity in the narwhal. We therefore estimated the demographic history of the narwhal genome using the pairwise sequentially Markovian coalescent (PSMC) model (Li and Durbin, 2011) (Figure 4). We find no drastic recent bottlenecks, but rather find an ancient, gradual but constant decline in Ne beginning at ~2.0 million years ago (Ma). The trajectory culminated in a persistently low Ne of ~5,000 individuals starting ~600 thousand years ago (kya), followed by an increase in Ne ~100 kya and a subsequent rapid expansion ~30–40 kya. The initial expansion coincides with the onset of the last glacial period ~115 kya (Rasmussen et al., 2014), suggesting an environmental driver, possibly linked to an increase in Arctic sea ice. Similar patterns of expansion ~30–40 kya have been documented in Arctic terrestrial mammals, e.g., mammoth, horse, and reindeer (Lorenzen et al., 2011), suggesting a significant ecological shift across the Arctic region at this time. The rapid recent increase in narwhal Ne is in agreement with previous hypotheses based on population genetic analysis of mtDNA control region sequences. Inferences suggested a rapid recent expansion from a small founding population at the end of the last glaciation (~25 kya) (Palsbøll et al., 1997). The pattern of a long steady decline followed by constant low population size could explain the low genetic diversity in current populations. However, differentiating true changes in population size from changes in population structure and gene flow is difficult based on a single genome (Mazet et al., 2015). Therefore caution should be exercised when interpreting these demographic results from a single individual. Further work involving multiple individuals from several populations will provide essential additional information about the findings of the current study.

To investigate whether the demographic trajectory is unique to the narwhal, or is associated with an Arctic lifestyle, we investigated the demographic history of the published genomes of beluga, bowhead whale, walrus, and polar bear. We find each species had its own unique demographic trajectory (Figure 4). This result is not unexpected, as each species has its own particular ecology, which would influence its demographic history. However, this result does suggest that simply living in the Arctic environment is in itself not the main driver behind the low genetic diversity seen in the narwhal.

Comparative History of the Narwhal and Beluga

The narwhal and its closest living relative, the beluga, are the only toothed whales found in the Arctic year round. The molecularly dated divergence time between the two species based on nuclear and mitochondrial markers is estimated at ~5.5 Ma (Steeman et al., 2009). However, despite this deep divergence time, the narwhal and beluga share some similarities. Both species are well adapted to life in the high Arctic, have long lifespans (narwhal up to ~100 years, Garde et al., 2007; beluga ~75 years, Stewart et al., 2006), and have similar body sizes (Garde et al., 2007), reproductive strategies, and generation times (Heide-Jørgensen, 2009; O'corry-Crowe, 2009; Garde et al., 2015). Hybridization between the two species has been reported (Heide-Jørgensen and Reeves, 1993), but the reproductive success of the hybrid offspring remains unknown.

To gain further insight into the demographic history of the narwhal and its evolutionary relationship to the beluga, we estimated the timing of when viable admixture between the two species may have ceased. We used an F1 hybrid PSMC (hPSMC) model. Results from the hPSMC model and simulation analysis suggest that gene flow between narwhal and beluga, at least in the populations that these genomes come from, ceased between 1.25 and 1.65 Ma (Figure S3). This result suggests that viable gene flow between the two species continued for a long time after divergence (~4 Ma), but the time frame also suggests that any contemporary hybrids are likely unable to reproduce, or have such high selective pressures against them that they fail to add anything to the parental species' gene pools. The narwhal in this study was sampled in West Greenland, and the beluga individual was sampled in Hudson Bay, Canada, where the species ranges do not overlap. Investigating narwhal and beluga individuals from populations where their ranges do overlap (e.g., West Greenland), may uncover more recent dates.

As more genomes become available, the notion that genetic diversity is directly associated with a species' long-term viability is being called into question. The demographic history of a species has been shown to be an important factor in long-term survival, despite seemingly detrimental low diversity (Westbury et al., 2018; Xue et al., 2015; Robinson et al., 2016). As narwhals are found only in the Atlantic sector of the Arctic, and not in the Pacific (Figure 1A), unlike the two other endemic Arctic cetaceans, the beluga and the bowhead whale, the limited geographic range may increase the species' vulnerability (Laidre et al., 2008). However, previous studies investigating wild species diversity using large genomic datasets found no detectable influence of geographic range on genomic diversity estimates (Romiguier et al., 2014), therefore the limited range most likely does not offer a suitable explanation for the low diversity. The same study also reported significant correlations between body mass, longevity and reproductive strategy, and genetic diversity (Romiguier et al., 2014). However, narwhals and belugas share many similarities in these traits (Heide-Jørgensen, 2009; O'corry-Crowe, 2009; Garde et al., 2015; Garde et al., 2007; Stewart et al., 2006), albeit the narwhal still has a much lower estimated genetic diversity. Hence these do not appear to be the main forces driving low diversity in the narwhal and indicate that somatic growth and life history parameters of the two species alone cannot explain the observed pattern of genetic diversity.

Although life history does not offer a definite explanation as to the low genetic diversity in narwhal, demographic history could offer a viable explanation. Previous studies have suggested that long-term, slow declines in population size (Westbury et al., 2018) or long-term low population size (Xue et al., 2015; Robinson et al., 2016) can allow a species to persist despite low diversity, by reducing the strain of deleterious recessive alleles. The narwhal has exhibited both of these demographic characteristics. Furthermore, our findings of an increase in Ne starting ~100 kya offers an explanation for the current abundance, despite the low genetic diversity in the species today. The low contemporary genetic diversity may in part reflect the longevity and long generation time in narwhals (Garde et al., 2015), which could slow the increase of genomic diversity.

The sequencing and assembly of a narwhal nuclear genome provides insights into genome-wide genetic diversity patterns and the demographic history of the species. In a comparison with 14 other mammalian species, including four endemic Arctic marine mammals, we find the narwhal to have low genome-wide diversity, and a unique demographic history not shared with any other endemic Arctic marine mammals. Our analyses reveal that demographic history has been an important factor influencing patterns of genetic diversity in the narwhal and offers an explanation for the low diversity in this individual despite the large global abundance of the species. We do not find evidence that the low diversity was caused by rapid bottlenecks or inbreeding, both of which are common mechanisms used to explain this pattern. Rather, we propose that low diversity has been present in narwhals for an extended period of time. The species may have adapted to cope with this at a genome-wide level over time and has potentially reached a stasis as to how low its genetic diversity can go before influencing the long-term survival of the species. Our study sets the groundwork for future studies into the evolutionary history of narwhals, which will hopefully uncover more details as to the causes of low diversity and how it influences the long-term population survival of the species.

Limitations of the Study

All inferences of the narwhal at a species level were based on a single individual. Although our inferences were based on the entire nuclear genome, which consists of many independently evolving loci, which may have been present in multiple populations in the past, there could be differences between contemporary narwhal populations that could not be uncovered from our individual. To fully uncover whether the conclusions drawn in this article are representative of the current species as a whole, further studies using genomic data from multiple narwhal individuals and populations will need to be investigated.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

DATA AND SOFTWARE AVAILABILITY

The narwhal assembly and short reads are available under the GenBank BioProject ID PRJNA508363.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2019.03.023.

ACKNOWLEDGMENTS

DNA sequencing of the narwhal genome was supported by Dronning Margrethes og Prins Henriks Fond, grant no. 2015/2-II-13. The work was further supported by Villum Fonden Young Investigator Programme, grant no. 13151; and Carlsberg Foundation Distinguished Associate Professor Fellowship, grant no. CF16-0202. We would like to thank Binia De Cahsan for drawing the animal images in Figure 1B and the narwhal illustration in the graphical abstract. The collection of samples from the Inuit hunt of narwhals was funded and conducted by the Greenland Institute of Natural Resources. Finally, we would like to thank Rheon Slade for his assistance in the design and layout of the graphical abstract.

AUTHOR CONTRIBUTIONS

The project was conceptualized by E.D.L. Genome assembly, annotation, and analyses were performed by M.V.W. and B.P. Sample acquisition and information on the said sample was provided by E.G. and M.P.H.J. The manuscript was written by M.V.W. with significant input from E.D.L. All co-authors read and agreed on the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 6, 2018
Revised: February 4, 2019
Accepted: March 22, 2019
Published: May 1, 2019

REFERENCES

Abascal, F., Corvelo, A., Cruz, F., Villanueva-Cañas, J.L., Vlasova, A., Marcet-Houben, M., Martínez-Cruz, B., Cheng, J.Y., Prieto, P., Quesada, V., et al. (2016). Extreme genomic erosion after recurrent demographic bottlenecks in the highly endangered Iberian lynx. Genome Biol. *17*, 251.

Best, R.C. (1981). The tusk of the narwhal (*Monodon monoceros* L.): interpretation of its function (Mammalia: Cetacea). Can. J. Zool. *59*, 2386–2393.

Ellegren, H., and Galtier, N. (2016). Determinants of genetic diversity. Nat. Rev. Genet. *17*, 422–433.

Frankham, R. (2005). Genetics and extinction. Biol. Conserv. *126*, 131–140.

Garde, E., Heide-Jørgensen, M.P., Hansen, S.H., Nachman, G., and Forchhamme, M.C. (2007). Age-specific growth and remarkable longevity in narwhals (*Monodon monoceros*) from west Greenland as estimated by aspartic acid racemization. J. Mammal. *88*, 49–58.

Garde, E., Hansen, S.H., Ditlevsen, S., Tvermosegaard, K.B., Hansen, J., Harding, K.C., and Heide-Jørgensen, M.P. (2015). Life history parameters of narwhals (*Monodon monoceros*) from Greenland. J. Mammal. *96*, 866–879.

Grau, J.H., Hackl, T., Koepfli, K.P., and Hofreiter, M. (2018). Improving draft genome contiguity with reference-derived in silico mate-pair libraries. GigaScience *7*, giy029.

Heide-Jørgensen, M.P. (2009). N - narwhal: *Monodon monoceros*. In Encyclopedia of Marine Mammals, Second Edition, W.F. Perrin, B. Würsig, and J.G.M. Thewissen, eds. (Academic Press), pp. 754–758.

Heide-Jørgensen, M.P., and Dietz, R. (1995). Some characteristics of narwhal, *Monodon monoceros*, diving behaviour in Baffin Bay. Can. J. Zool. *73*, 2120–2132.

Heide-Jørgensen, M.P., and Reeves, R.R. (1993). Description of an anomalous monodontid skull from west Greenland: a possible hybrid? Marine Mammal Science *9*, 258–268.

Hoelzel, A.R., Natoli, A., Dahlheim, M.E., Olavarria, C., Baird, R.W., and Black, N.A. (2002). Low worldwide genetic diversity in the killer whale (*Orcinus orca*): implications for demographic history. Proc. Biol. Sci. *269*, 1467–1473.

Holt, C., and Yandell, M. (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC bioinformatics *12*, 491.

Jones, S.J.M., Taylor, G.A., Chan, S., Warren, R.L., Hammond, S.A., Bilobram, S., Mordecai, G., Suttle, C.A., Miller, K.M., Schulze, A., et al. (2017). The genome of the beluga whale (*Delphinapterus leucas*). Genes (Basel) *8*, 378.

Kimura, M. (1968). Evolutionary rate at the molecular level. Nature *217*, 624–626.

Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: analysis of next generation sequencing data. BMC Bioinformatics *15*, 356.

Kovacs, K.M., Lydersen, C., Overland, J.E., and Moore, S.E. (2011). Impacts of changing sea-ice conditions on Arctic marine mammals. Marine Biodiversity *41*, 181–194.

Laidre, K.L., Stirling, I., Lowry, L.F., Wiig, O., Heide-Jørgensen, M.P., and Ferguson, S.H. (2008). Quantifying the sensitivity of Arctic marine mammals to climate-induced habitat change. Ecol. Appl. *18* (2 Suppl), S97–S125.

Li, H., and Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. Nature *475*, 493–496.

Lorenzen, E.D., Nogués-Bravo, D., Orlando, L., Weinstock, J., Binladen, J., Marske, K.A., Ugan, A., Borregaard, M.K., Gilbert, M.T., Nielsen, R., et al. (2011). Species-specific responses of Late Quaternary megafauna to climate and humans. Nature *479*, 359–364.

Lowry, L., Laidre, K., and Reeves, R. (2017). Monodon monoceros. The IUCN Red List of Threatened Species 2017: e.T13704A50367651, Available at: https://doi.org/10.2305/IUCN.UK.2017-3.RLTS.T13704A50367651.en.

Masson-Delmotte, V., Zhai, P., Pörtner, H.-O., Roberts, D., Skea, J., Shukla, P.R., Pirani, A., Moufouma-Okia, W., Péan, C., Pidcock, R., et al. (2018). Global warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty. Intergovernmental panel Clim. Change (ipcc).

Mazet, O., Rodríguez, W., and Chikhi, L. (2015). Demographic inference using genetic data from a single individual: Separating population size variation from population structure. Theor. Popul. Biol. *104*, 46–58.

NAMMCO. (2018). Report of the NAMMCO Global Review of Monodontids (NAMMCO Scientific Publications), Available at: https://nammco.no/topics/sc-working-group-reports/.

O'corry-Crowe, G.M. (2009). Beluga whale: *Delphinapterus leucas*. In Encyclopedia of Marine Mammals, Second Edition, W.F. Perrin, B. Würsig, and J.G.M. Thewissen, eds. (Academic Press), pp. 108–112.

Palkopoulou, E., Mallick, S., Skoglund, P., Enk, J., Rohland, N., Li, H., Omrak, A., Vartanyan, S., Poinar, H., Götherström, A., et al. (2015). Complete genomes reveal signatures of demographic and genetic declines in the woolly mammoth. Curr. Biol. *25*, 1395–1400.

Palsbøll, P.J., Heide-Jørgensen, M.P., and Dietz, R. (1997). Population structure and seasonal movements of narwhals, *Monodon monoceros*, determined from mtDNA analysis. Heredity *78* (Pt 3), 284–292.

Petersen, S.D., Tenkula, D., and Ferguson, S.H. (2011). Population genetic structure of narwhal (*Monodon monoceros*), Fisheries and Oceans Canada. Science.

Rasmussen, S.O., Bigle, M., Blockley, S.P., Blunier, T., Buchardt, S.L., Clausen, H.B., Cvijanovic, I., Dahl-Jensen, D., and Johnsen, S.J. (2014). A stratigraphic framework for abrupt climatic changes during the Last Glacial period based on three synchronized Greenland ice-core records: refining and extending the INTIMATE event stratigraphy. Quat. Sci. Rev. *106*, 14–28.

Reed, D.H., and Frankham, R. (2003). Correlation between fitness and genetic diversity. Conserv. Biol. *17*, 230–237.

Robinson, J.A., Ortega-Del Vecchyo, D., Fan, Z., Kim, B.Y., vonHoldt, B.M., Marsden, C.D., Lohmueller, K.E., and Wayne, R.K. (2016). Genomic flatlining in the endangered Island fox. Curr. Biol. *26*, 1183–1189.

Romiguier, J., Gayral, P., Ballenghien, M., Bernard, A., Cahais, V., Chenuil, A., Chiari, Y., Dernat, R., Duret, L., Faivre, N., et al. (2014). Comparative population genomics in animals uncovers the determinants of genetic diversity. Nature *515*, 261–263.

Steeman, M.E., Hebsgaard, M.B., Fordyce, R.E., Ho, S.Y., Rabosky, D.L., Nielsen, R., Rahbek, C., Glenner, H., Sørensen, M.V., Willerslev, E., et al. (2009). Radiation of extant cetaceans driven by restructuring of the oceans. Syst. Biol. *58*, 573–585.

Stewart, R.E.A., Campana, S.E., Jones, C.M., and Stewart, B.E. (2006). Bomb radiocarbon dating calibrates beluga (*Delphinapterus leucas*) age estimates. Can. J. Zool. *84*, 1840–1852.

Westbury, M.V., Hartmann, S., Barlow, A., Wiesel, I., Leo, V., Welch, R., Parker, D.M., Sicks, F., Ludwig, A., Dalén, L., et al. (2018). Extended and continuous decline in effective population size results in low genomic diversity in the world's rarest Hyena species, the brown Hyena. Mol. Biol. Evol. *35*, 1225–1237.

Xue, Y., Prado-Martinez, J., Sudmant, P.H., Narasimhan, V., Ayub, Q., Szpak, M., Frandsen, P., Chen, Y., Yngvadottir, B., Cooper, D.N., et al. (2015). Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. Science *348*, 242–245.

Zachos, J., Pagani, M., Sloan, L., Thomas, E., and Billups, K. (2001). Trends, rhythms, and aberrations in global climate 65 Ma to present. Science *292*, 686–693.

ISCI, Volume *15*

Supplemental Information

Narwhal Genome Reveals Long-Term Low Genetic Diversity despite Current Large Abundance Size

Michael V. Westbury, Bent Petersen, Eva Garde, Mads Peter Heide-Jørgensen, and Eline D. Lorenzen

Supplemental information

Supplemental tables:

Supplemental table S1: Illumina and cross-mate libraries used for the assembly, Related to Results and discussion section Narwhal genome assembly.

Library	Approximate insert size	# of reads
Short insert 01	-	384,563,392
Short insert 02	-	379,311,470
Short insert 03	-	381,357,017
Mate-pair	3kb	152,848,424
Mate-pair	5kb	108,544,278
Mate-pair	10kb	100,789,266
Cross-mate	500bp	48,445,804
Cross-mate	1kb	45,509,780
Cross-mate	1.5kb	44,383,867
Cross-mate	2kb	43,559,749
Cross-mate	3kb	42,428,532
Cross-mate	4kb	41,623,118
Cross-mate	5kb	41,082,099
Cross-mate	8kb	40,265,781
Cross-mate	10kb	40,091,817
Cross-mate	15kb	39,771,013
Cross-mate	20kb	39,635,226

Supplemental table S2: BUSCO scores of the assembled narwhal genome when using the BUSCOv3 mammal dataset, Related to Results and discussion section Narwhal genome assembly.

Category	Number of BUSCO	Percentage
Complete	3819	93.00%
Complete and single copy	3772	91.90%
Complete and duplicated	47	1.10%
Fragmented	140	3.40%
Missing	145	3.60%
Total searched	4104	

Supplemental table S3: Narwhal genome repeat profile, Related to Results and discussion section Narwhal genome assembly.

Repeat type	De novo repeats (%)	Model based repeats (%)	Total (%)
Total	34.87	3.03	37.90
SINEs	6.26	0.22	6.48
LINEs	20.43	1.63	22.06
LTR elements	4.87	0.70	5.57
DNA elements	3.17	0.40	3.57
Unclassified	0.05	0.08	0.13
Small RNA	3.05	0.21	3.26
Satellites	0.07	0.00	0.07

Supplemental table S4: Standard deviation of heterozygosity in 500kb windows across the autosomes of the narwhal and other endemic Arctic marine mammals, Related to Figure 2.

Species	Standard deviation
Narwhal	0.0000642673
Beluga	0.000114535
Bowhead	0.000166954
Walrus	0.000191573
Polar bear	0.000272893

Supplemental table S5: Distribution of heterozygosity in different regions across the narwhal genome, Related to Figure 3.

Feature/location	Mean heterozygosity	Standard deviation
Autosomes	0.000138	0.00000333
Exons	0.000127	0.00000913
Exons and introns	0.000133	0.00000480
10kb away	0.000132	0.00000410
20kb away	0.000136	0.00000500
50kb away	0.000141	0.00000721

Supplemental table S6: Distribution of heterozygosity in different regions across the beluga genome, Related to Figure 3.

Feature/location	Mean heterozygosity	Standard deviation
Autosomes	0.000289	5.31e-06
Exons	0.000221	1.60e-05
Exons and introns	0.000277	6.21e-06
10kb away	0.000284	6.34e-06
20kb away	0.000285	7.36e-06
50kb away	0.000294	7.81e-06

Supplementary table S7: Unpaired two sample t-test to test for significant differences in heterozygosity between different genomic regions in the narwhal, Related to Figure 3.

Regions compared	t-score	p-value
Autosomes vs. exons	11.389	< 2.2e-16
Autosomes vs. genes	7.6497	8.59e-13
Autosomes vs. 10kb away	10.466	< 2.2e-16
Autosomes vs. 20kb away	3.013	0.002924
Autosomes vs. 50kb away	-4.1801	4.37e-05
Exons vs. genes	-6.3992	1.11e-09
Genes vs. 10kb away	1.6733	0.09585
Genes vs. 20kb away	-3.8386	0.0001664

Supplementary table S8: Unpaired two sample t-test to test for significant differences in heterozygosity between different genomic regions in the beluga, Related to Figure 3.

Regions compared	t-score	p-value
Autosomes vs. exons	42.691	2.20e-16
Autosomes vs. genes	14.707	2.20e-16
Autosomes vs. 10kb away	6.9906	4.09e-11
Autosomes vs. 20kb away	5.6237	6.30e-08
Autosomes vs. 50kb away	-7.4233	3.31e-12
Exons vs. genes	-42.691	2.20e-16
Genes vs. 10kb away	-34.921	2.20e-16
Genes vs. 20kb away	-38.454	2.20e-16

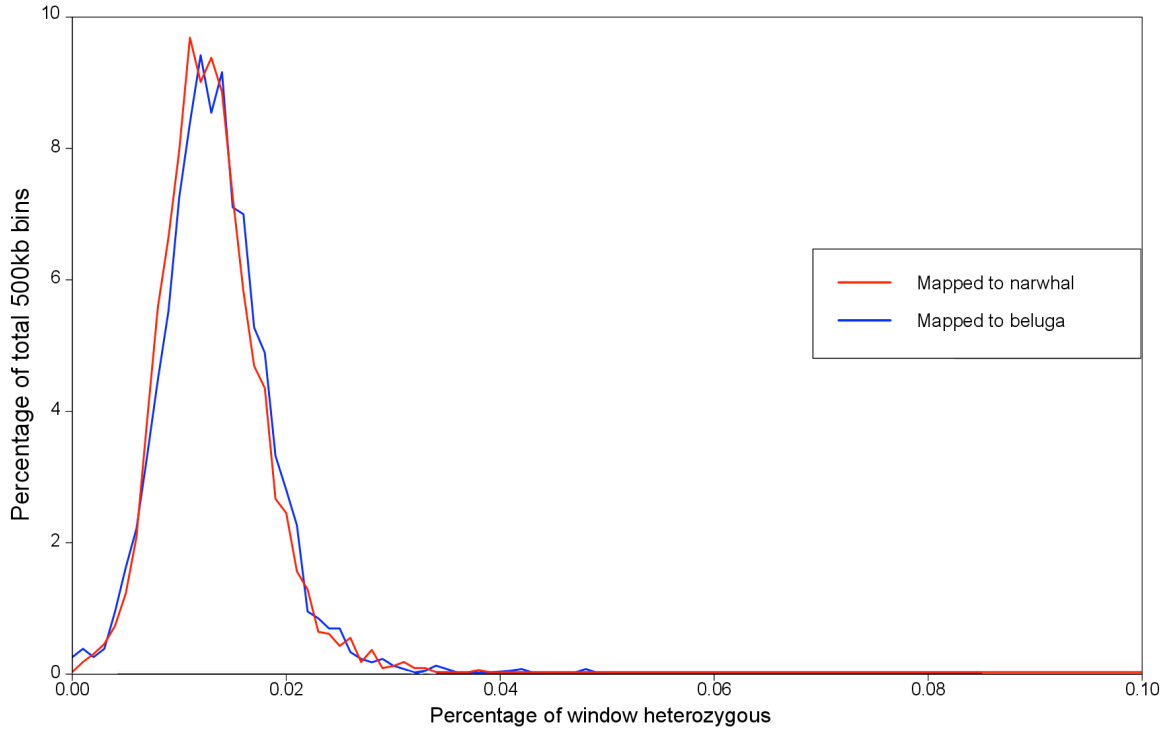
Supplemental table S9: Pairwise distances (PWD) and species divergence times used to calculate mutation rates and the resultant mutation rates per year, Related to Figure 4.

	PWD	Divergence time (Ma)	u/year
Bowhead-right whale	0.0067	4.38	0.0000000007676940639
Beluga-Narwhal	0.0057	5.5	0.0000000005181818182
Seal-Walrus	0.023	18	0.0000000006259166667

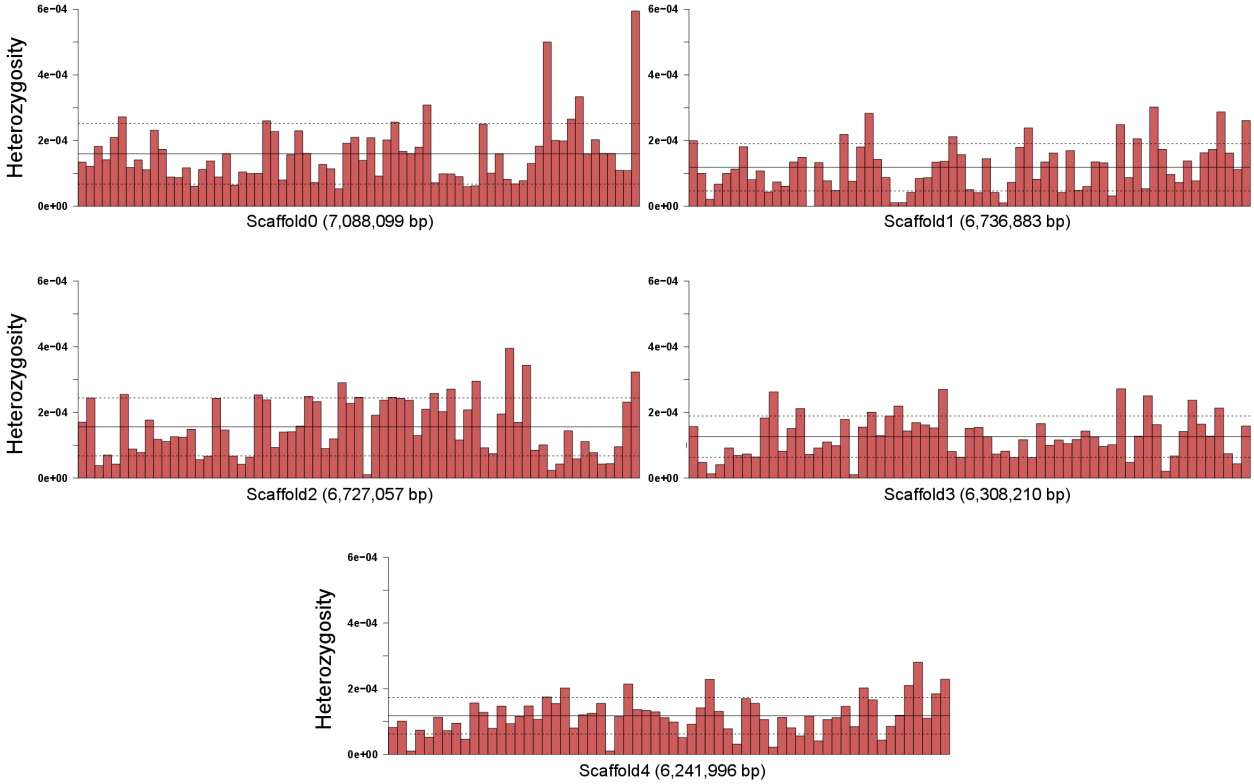
Supplemental table S10: Mutation rates and generation times used for plotting PSMC, Related to Figure 4.

Common name	Genus	Species	Generation time	u/generation
Narwhal	*Monodon*	*monoceros*	30	1.56e-08
Beluga	*Delphinapterus*	*leucas*	32	1.65e-08
Polar bear	*Ursus*	*maritimus*	11.2	1.83e-08
Walrus	*Odobenus*	*rosmarus*	15	9.40e-09
Bowhead whale	*Balaena*	*mysticetus*	35	2.69e-08

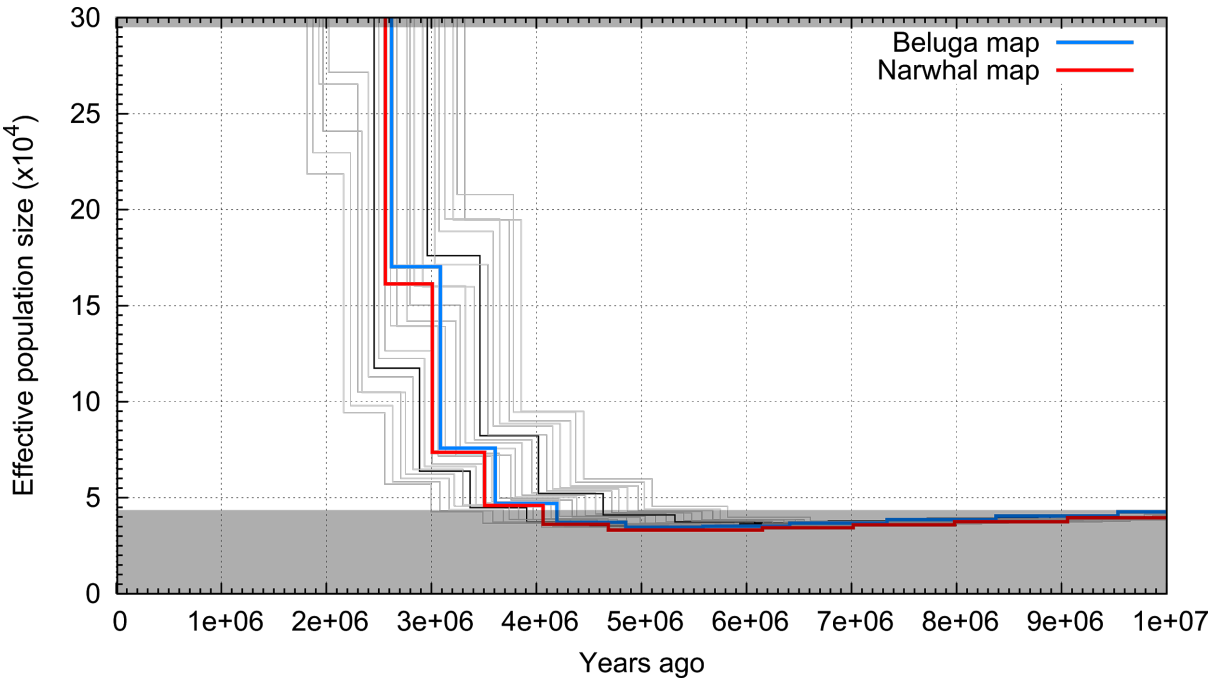
Supplemental figures:



Supplemental figure S1: Sliding window heterozygosity in the narwhal when using different mapping reference genomes, Related to Figure 2. Red shows when mapped to the narwhal. Blue shows when mapped to the beluga.



Supplemental figure S2: Heterozygosity in 100kb non-overlapping sliding windows across the five largest scaffolds from the narwhal assembly, Related to Figure 2. Black line represents the mean heterozygosity value for said scaffold and the dotted line represents one standard deviation above and below the mean heterozygosity value.



Supplemental figure S3: hPSMC plot between the beluga and the narwhal and simulations of various different divergences, Related to Results and discussion section Comparative history of the narwhal and beluga. Greyed out regions represent 1.5x and 10x the pre divergence effective population size, grey lines represent the simulated data, black line represents the simulations closest to the real data without overlapping it, blue line represents the hPSMC result when both the narwhal and beluga were mapped to the beluga reference genome, red line represents the hPSMC result when both the narwhal and the beluga were mapped to the narwhal reference genome.

Transparent methods:

Sample information

The narwhal individual was sampled in Uummannaq/West Greenland in 1993, and originated from the Somerset Island stock. It was collected from the Greenland Institute of Natural Resources under the general permit for biological sampling of the Inuit from the Greenland Government. The sample was exported to Denmark under CITES permit number 15GL1003549. The Somerset Island stock is one of the largest narwhal stocks with current population levels being estimated at ~50,000 individuals (NAMMCO 2018).

Genome assembly

Whole genomic DNA was extracted from a frozen liver tissue from a single narwhal individual using the QIAGEN DNeasy blood and tissue kit following the manufacturer's protocol with slight modifications (2x volume of reagents (except AW1 and AW2)). Extracts were built into three short insert Illumina sequencing libraries and three mate-paired Illumina sequencing libraries (~3kb ~5kb, ~10kb) by the UC Davis genome center (http://genomecenter.ucdavis.edu/). Libraries were sequenced at the UC Davis genome center on an Illumina HiSeq platform. Additionally, we constructed cross-species 100bp mate paired reads of insert sizes between 500bp and 20kb (Table S1) utilising the repeat masked beluga genome (Genbank: GCA_002288925.2) (Jones et al. 2017) and the software Cross-Species Scaffolding (Grau et al. 2018). We removed adapter sequences from the short insert and mate paired libraries using skewer (Jiang et al. 2014) and removed PCR duplicates with prinseq (Schmieder & Edwards 2011). We performed an error correction step using a kmer size of 31 in tadpole from the bbtools toolsuite (Bushnell 2014). We constructed a *de novo* assembly using these error corrected reads, the three mate paired libraries and the cross species mate paired libraries using SOAPdenovo2 (Luo et al. 2012) and specified a kmer size of 51. The short insert libraries were used in both the contig construction and scaffolding steps while the mate paired libraries were only used in the scaffolding step. We removed all contigs shorter than 1000bp from the final assembly. We performed gap closing on the assembly with Sealer (Paulino et al. 2015), utilising various kmer sizes (50, 60, 70, 80, 90, 100) and the error corrected short insert library reads. The assembly continuity was assessed using quast v4.5 (Gurevich et al. 2013) and gene content was assessed using BUSCO v3 (Waterhouse et al. 2017) and the mammalian BUSCO gene set database.

Repeatmasking and annotation

Repeats and low complexity DNA sequences were masked in the genome prior to gene annotation using RepeatMasker version open-4.0.7 (Smit et al. 2013-2015) using the species repeat database 'narwhal' with RepBase database version 20170127. Remaining specific repetitive elements were predicted *de novo* using RepeatModeler version 1.0.11 (Smit & Hubley 2008-2015) on the masked genome. Subsequently, a second round of RepeatMasker was run with the model generated from RepeatModeler as custom library input on the previously masked genome.

Genome annotation was performed using the genome annotation pipeline MAKER2 version 2.31.9 (Holt & Yandell 2011) with ab-initio and homology-based gene predictions. Protein sequences from killer whale (*Orcinus orca*), beluga whale (*Delphinapterus leucas*), cattle (*Bos taurus)*, dog, (*Canis lupus familiaris*), humans (*Homo sapiens*), minke whale (*Balaenoptera acutorostrata*) and the finless porpoise (*Neophocaena asiaeorientalis*) were used for homology-based gene prediction. As no training gene models were available for narwhals, we used CEGMA (Parra et al. 2007; Parra et al. 2009) to train the ab-initio gene predictor SNAP (Korf 2004), rather than using the de-novo gene predictor in Augustus

(Stanke & Waack 2003). MAKER2 was run with "model_org=simple, softmask=1, augustus_species=human" and the "snaphmm" parameter was set to the HMM generated in the manual training of SNAP. As EST evidence we used a published transcriptome skin sample of a beluga whale (Genbank: PRJNA414234).

Heterozygosity estimates

We estimated autosomal heterozygosity from our narwhal genome and four endemic Arctic marine mammals. We downloaded the assembled genomes and raw Illumina reads from the beluga (*Delphinapterus leucas* Genbank: GCA_002288925.2), bowhead whale (*Balaena mysticetus*) (Keane et al. 2015) and walrus (*Odobenus rosmarus*, Genbank: GCF_000321225.1) (Foote et al. 2015). Genome-wide average autosomal heterozygosity for the polar bear (*Ursus maritimus*, Genbank: GCF_000687225.1) (Liu et al. 2014), was taken from Westbury et al, 2018 (Westbury et al. 2018), while the following methods were implemented for the other species. 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 narwhal, beluga, and bowhead whale by aligning the assembled genomes to the Cow X (Genbank: CM008168.2) and Human Y (Genbank: NC_000024.10) chromosomes. We found the putative sex chromosome scaffolds in the polar bear, and walrus by aligning the assembled genomes to the Human Y and the Dog X (Genbank: CM000039.3) chromosomes. Alignments were performed using satsuma synteny (Grabherr et al. 2010) and utilising default parameters.

We trimmed adapter sequences from the downloaded raw reads using skewer, mapped the trimmed reads to each respective reference genome using BWA v0.7.15 (Li & Durbin 2009) and the mem algorithm. We parsed the output and removed duplicates with samtools v1.6 (Li et al. 2009). Furthermore, to ensure comparability with previous heterozygosity estimates and to remove biases in heterozygosity levels that could arise due to different global coverages between the genomes of the individuals being investigated, we subsampled all of the resultant alignments down to 20x using samtools. We estimated the autosomal heterozygosity using sample allele frequencies in ANGSDv0.921 (Korneliussen et al. 2014), taking genotype likelihoods into account and specifying the following filters -minq 25 -minmapq 25 -uniqueOnly 1 -baq 1 -remove_bads 1 as was previously done in Westbury et al 2018 (Westbury et al. 2018). We computed the heterozygosity using ANGSD as it can overcome biases that may arise due to differential coverage across the genome. Instead of relying on direct SNP/genotype calling from the data, ANGSD uses genotype likelihoods data in downstream analyses and allows for the incorporation of statistical uncertainties into the analysis. This feature should reduce the biases caused by differential coverage across the genome.

The resultant values were compared alongside previously reported values from 10 other mammalian species (Westbury et al. 2018). We investigated heterozygosity in 500kb non-overlapping windows across the genomes of the five marine mammal species, using bedtools (Quinlan 2014). When plotting the results, we only considered windows from within the autosomes, scaffolds over 500kb in length, and windows with more than 70% data. Each window was treated individually and the percentage of heterozygous within each window was calculated. To investigate whether the heterozygosity results of the narwhal were a result of the quality of the genome, we mapped the short reads of our narwhal to the published beluga genome and repeated the above steps.

Finally, we investigated the distribution of heterozygosity across the genome, considering only autosomes and scaffolds longer than 500kb. This was done by independently calculating heterozygosity in five different partitions; exons, genes (exons +

introns), 10kb windows 10kb away, 20kb away, and 50kb away from the nearest protein-coding gene. We calculated variance in these results by randomly sampling 10% of the windows in each partition 100 times and plotting box plots using R. Using these 100 random samplings we additionally performed eight unpaired two sample t-tests per species to investigate the significance of differences between the different partitions. The comparisons included exons vs. autosomes, genes vs. autosomes, 10kb away vs. autosomes, 20kb away vs. autosomes, 50kb away vs. autosomes, exons vs. genes, genes vs, 10kb away, and genes vs. 20kb away. Differences were deemed significant by a p-value < 0.05.

Demographic history

We ran demographic analyses on diploid genomes from single individual species representatives of the narwhal, beluga, bowhead whale, walrus, and polar bear using a Pairwise Sequentially Markovian Coalescent model (PSMC)(Li & Durbin 2011). We called diploid genome sequences using samtools and bcftools (Narasimhan et al. 2016) specifying a minimum quality score of 20 and minimum coverage of 10. 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 estimate the mutation rate per generation for each species, we computed pairwise distances between closely related species, using a consensus base call in ANGSD and applying the filters -minQ 25 -minmapq 25 -uniqueonly 1 -remove_bads 1. Mutation rate per generation was calculated as follows: mutation rate = pairwise distance x generation time / 2 x divergence time. To estimate the narwhal and beluga mutation rates, short reads of both species were mapped to the narwhal genome, and mutation rate was calculated from the pairwise distances, assuming a divergence date of 5.5 Ma (Steeman et al. 2009). We assumed a narwhal generation time of 30 years and a beluga generation time of 32 years (Garde et al. 2015). To estimate the bowhead whale mutation rate, we downloaded short reads from the right whale (Genbank: SRR5665640) (Árnason et al. 2018) and mapped them to the bowhead whale genome. We calculated the mutation rate assuming a divergence date between the right whale and bowhead whale of 4.38 Ma (Árnason et al. 2018). We assumed a bowhead generation time of 35 years (Rooney et al. 2001). To estimate the walrus mutation rate, we mapped the northern fur seal (Genbank: SRR7278673) to the walrus genome and calculated the mutation rate assuming a divergence date between the walrus and the northern fur seal of 18 Ma (Higdon et al. 2007). We assumed a walrus generation time of 15 years (Andersen et al. 2009). For the polar bear, we used the previously published generational mutation rate of 1.825728e-08 and generation time of 11.2 years (Liu et al. 2014). Results and calculations can be seen in Supplemental tables S9 and S10.

Dating the end of gene flow between narwhal and beluga

To calculate when gene flow ceased between the narwhal and beluga, we used hPSMC (Cahill et al. 2016). To overcome any biases that may occur due to differences in reference qualities, we replicated this analysis twice, once with both species mapped to the narwhal genome and once with both species mapped to the beluga. We constructed haploid consensus sequences using ANGSD by considering the base with the highest effective depth, the following quality filters; -minQ 25, -minmapq 25, -uniqueonly 1, -remove_bads 1, -setMinDepthInd 10, and only considering autosomes and scaffolds over 100kb. These haploid consensus sequences were merged together using the hPSMC toolsuite into a pseudo diploid sequence, run through PSMC and plotted using a narwhal/beluga intermediate

mutation rate per generation of $1.6e^{-08}$ and an intermediate generation time of 31 years. From this output we estimated the pre-divergence Ne of the narwhal and beluga to be ~29,000 individuals. We ran simulations using this pre-divergence Ne with various divergence times between 1Ma and 2Ma in 50,000 year intervals using ms (Hudson 2002). Results were plotted and the simulations with an exponential increase in Ne closest to the real data, within 1.5x and 10x of the pre-divergence Ne, were taken as the time interval in which gene flow stopped.

References:

Andersen, L.W. et al., 2009. Genetic signals of historic and recent migration between sub-populations of Atlantic walrus *Odobenus rosmarus rosmarus* west and east of Greenland. *Endangered species research*, 9, pp.197–211.

Árnason, Ú. et al., 2018. Whole-genome sequencing of the blue whale and other rorquals finds signatures for introgressive gene flow. *Science advances*, 4(4), p.eaap9873.

Bushnell, B., 2014. BBTools software package. *URL http://sourceforge. net/projects/bbmap*.

Cahill, J.A. et al., 2016. Inferring species divergence times using pairwise sequential Markovian coalescent modelling and low-coverage genomic data. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 371(1699), p.20150138.

Foote, A.D. et al., 2015. Convergent evolution of the genomes of marine mammals. *Nature genetics*, 47(3), pp.272–275.

Garde, E. et al., 2015. Life history parameters of narwhals (*Monodon monoceros*) from Greenland. *Journal of mammalogy*, 96(4), pp.866–879.

Grabherr, M.G. et al., 2010. Genome-wide synteny through highly sensitive sequence alignment: Satsuma. *Bioinformatics* , 26(9), pp.1145–1151.

Grau, J.H. et al., 2018. Improving draft genome contiguity with reference-derived in silico mate-pair libraries. *GigaScience*, 7(5), p.giy029.

Gurevich, A. et al., 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* , 29(8), pp.1072–1075.

Higdon, J.W. et al., 2007. Phylogeny and divergence of the pinnipeds (Carnivora: Mammalia) assessed using a multigene dataset. *BMC evolutionary biology*, 7, p.216.

Holt, C. & Yandell, M., 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC bioinformatics*, 12, p.491.

Hudson, R.R., 2002. Generating samples under a Wright–Fisher neutral model of genetic variation. *Bioinformatics* . Available at: https://academic.oup.com/bioinformatics/article-abstract/18/2/337/225783.

Jiang, H. et al., 2014. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC bioinformatics*, 15, p.182.

Jones, S.J.M. et al., 2017. The Genome of the Beluga Whale (*Delphinapterus leucas*). *Genes*, 8(12), p.378.

Keane, M. et al., 2015. Insights into the evolution of longevity from the bowhead whale genome. *Cell reports*, 10(1), pp.112–122.

Korf, I., 2004. Gene finding in novel genomes. *BMC bioinformatics*, 5, p.59.

Korneliussen, T.S., Albrechtsen, A. & Nielsen, R., 2014. ANGSD: Analysis of Next Generation Sequencing Data. *BMC bioinformatics*, 15, p.356.

Li, H. et al., 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* , 25(16), pp.2078–2079.

Li, H. & Durbin, R., 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* , 25(14), pp.1754–1760.

Li, H. & Durbin, R., 2011. Inference of human population history from individual whole-genome sequences. *Nature*, 475(7357), pp.493–496.

Liu, S. et al., 2014. Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears. *Cell*, 157(4), pp.785–794.

Luo, R. et al., 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience*, 1(1), p.18.

NAMMCO, 2018. Report of the NAMMCO Global Review of Monodontids, Hillerød, Denmark. Available at: https://nammco.no/topics/sc-working-group-reports/.

Narasimhan, V. et al., 2016. BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics* , 32(11), pp.1749–1751.

Parra, G. et al., 2009. Assessing the gene space in draft genomes. *Nucleic acids research*, 37(1), pp.289–297.

Parra, G., Bradnam, K. & Korf, I., 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* , 23(9), pp.1061–1067.

Paulino, D. et al., 2015. Sealer: a scalable gap-closing application for finishing draft genomes. *BMC bioinformatics*, 16, p.230.

Quinlan, A.R., 2014. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Current protocols in bioinformatics*, 47, pp.11.12.1–34.

Rooney, A.P., Honeycutt, R.L. & Derr, J.N., 2001. Historical population size change of bowhead whales inferred from DNA sequence polymorphism data. *Evolution; international journal of organic evolution*, 55(8), pp.1678–1685.

Schmieder, R. & Edwards, R., 2011. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* , 27(6), pp.863–864.

Smit, A.F.A. & Hubley, R., 2008-2015. RepeatModeler Open-1.0. Available at: http://www.repeatmasker.org.

Smit, A.F.A., Hubley, R. & Green, P., 2013-2015. RepeatMasker Open-4.0. Available at: http://www.repeatmasker.org.

Stanke, M. & Waack, S., 2003. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* , 19, pp.215–225.

Steeman, M.E. et al., 2009. Radiation of extant cetaceans driven by restructuring of the oceans. *Systematic biology*, 58(6), pp.573–585.

Waterhouse, R.M. et al., 2017. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular biology and evolution*, 35(3), pp.543–548.

Westbury, M.V. et al., 2018. Extended and Continuous Decline in Effective Population Size Results in Low Genomic Diversity in the World's Rarest Hyena Species, the Brown Hyena. *Molecular biology and evolution*, 35(5), pp.1225–1237.