

Research



Cite this article: Westbury MV, Thompson KF, Louis M, Cabrera AA, Skovrind M, Castruita JAS, Constantine R, Stevens JR, Lorenzen ED. 2021 Ocean-wide genomic variation in Gray's beaked whales, *Mesoplodon grayi*. *R. Soc. Open Sci.* **8**: 201788.

<https://doi.org/10.1098/rsos.201788>

Received: 8 October 2020

Accepted: 18 February 2021

Subject Category:

Ecology, conservation, and global change biology

Subject Areas:

genomics/evolution/ecology

Keywords:

Ziphiidae, mitogenomes, nuclear genomes, population structure, evolution, demographic history

Author for correspondence:

K. F. Thompson

e-mail: k.f.thompson@exeter.ac.uk

[†]These authors contributed equally to this study.

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.5344994>.

Ocean-wide genomic variation in Gray's beaked whales, *Mesoplodon grayi*

M. V. Westbury^{1,†}, K. F. Thompson^{2,†}, M. Louis¹,
A. A. Cabrera¹, M. Skovrind¹, J. A. S. Castruita¹,
R. Constantine³, J. R. Stevens² and E. D. Lorenzen¹

¹GLOBE Institute, University of Copenhagen, Denmark

²Biosciences, University of Exeter, Devon, UK

³School of Biological Sciences and Institute of Marine Science, University of Auckland, New Zealand

MVW, 0000-0003-0478-3930; KFT, 0000-0003-4277-3549;
ML, 0000-0002-4611-5503; AAC, 0000-0001-5385-1114;
RC, 0000-0003-3260-539X; JRS, 0000-0002-1317-6721;
EDL, 0000-0002-6353-2819

The deep oceans of the Southern Hemisphere are home to several elusive and poorly studied marine megafauna. In the absence of robust observational data for these species, genetic data can aid inferences on population connectivity, demography and ecology. A previous investigation of genetic diversity and population structure in Gray's beaked whale (*Mesoplodon grayi*) from Western Australia and New Zealand found high levels of mtDNA diversity, no geographic structure and stable demographic history. To further investigate phylogeographic and demographic patterns across their range, we generated complete mitochondrial and partial nuclear genomes of 16 of the individuals previously analysed and included additional samples from South Africa ($n = 2$) and South Australia ($n = 4$), greatly expanding the spatial range of genomic data for the species. Gray's beaked whales are highly elusive and rarely observed, and our data represents a unique and geographically broad dataset. We find relatively high levels of diversity in the mitochondrial genome, despite an absence of population structure at the mitochondrial and nuclear level. Demographic analyses suggest these whales existed at stable levels over at least the past 1.1 million years, with an approximately twofold increase in female effective population size approximately 250 thousand years ago, coinciding with a period of increased Southern Ocean productivity, sea surface temperature and a potential expansion of suitable habitat. Our results suggest that Gray's beaked whales are likely to be resilient to near-future ecosystem changes, facilitating their

conservation. Our study demonstrates the utility of low-effort shotgun sequencing in providing ecological information on highly elusive species.

1. Introduction

The ocean is undergoing rapid ecological change as a result of climate change, ocean acidification and deoxygenation, poor fisheries management, oil and gas extraction, litter and shipping [1]. In an attempt to protect biodiversity, the United Nations is negotiating a treaty to ensure the conservation and sustainable use of biodiversity in the areas beyond jurisdiction—the ‘high seas’—or the deep ocean beyond the continental shelf edges (<https://www.un.org/bbnj/>). However, despite the desire to protect biodiversity in the high seas, the task is a challenge when the ecology of elusive species is still a mystery.

Currently, the deep sea is one of the last frontiers of exploration on Earth and there are a great many species for which we know very little, including large cetaceans [2,3]. Despite their general public appeal and influence on human culture throughout the years [4], oceanic cetaceans, including beaked whales, dolphins and baleen whales, are generally poorly understood. For the beaked whales, basic data on migrations, foraging ecology and social structures are lacking for most species, leading to difficulties in assessing how they will respond as their key habitats change, which leads to the question of how vulnerable they may be to broad marine ecosystem changes.

The taxonomy of oceanic beaked whales has been substantially revised during the past decade, mainly through the use of genetic tools, and there is much to learn about this family of deep-water cetaceans. Recent examples of newly described or taxonomically resurrected beaked whale species include: spade-toothed whale (*Mesoplodon traversii*) in 2012, Derinayagala’s whale (*Mesoplodon hotaula*) in 2014 and a new species of Baird’s beaked whale (*Berardius minimus*) in 2019 [5–7].

There are 23 currently recognized species of beaked whales within five genera [8]. All species are deep divers that feed on squid, small fish and, for some, crustaceans [9,10]. These whales have the most extreme diving behaviour of all marine mammals, with some species diving to nearly 3000 m deep, and remaining submerged for up to three hours (Cuvier’s beaked whales (*Ziphius cavirostris*) [11]). Beaked whales inhabit areas off the continental shelf edge out to the deep waters. This habitat preference, alongside their elusive surface behaviour, means that only a few species and populations have been well-studied. In general, populations around oceanic islands are better understood [12–16]. Beaked whales that have an oceanic distribution far offshore in the Southern Hemisphere are very rarely observed alive, let alone researched, limiting insights into their ecology [12].

Gray’s beaked whales (*Mesoplodon grayi*) are medium-sized (approx. 5 m) whales, with a Southern Hemisphere, circumpolar distribution south of 33° latitude (figure 1) [20]. There are no estimates of the abundance of Gray’s beaked whales based on census data. The use of genetic data has provided ecological insights into this species, which could not be achieved with more conventional methods due to their elusive nature. Previous work using partial mtDNA control region sequences (530 bp) and 12 microsatellites from 94 Gray’s stranded on the coast of Western Australia and New Zealand uncovered relatively high levels of genetic diversity, and no population differentiation between localities [20]. The study estimated a female effective population size of 460 000 whales, albeit with very large confidence intervals (95% CI 20 000–2 250 000). Moreover, investigations into the genetic kinships within different stranded groups in Western Australia and New Zealand found no indications of related adults stranding together, suggesting both sexes disperse from their natal group to potentially form groups with unrelated individuals [21]. Such a flexible social system combined with the rich oceanic habitat of the Southern Hemisphere could facilitate gene flow and potentially support a surprisingly large number of Gray’s beaked whales.

Here, we present a new reference panel of complete mitochondrial genomes and partial nuclear genomes to test the hypothesis of no clear phylogeographic structuring in the elusive Gray’s beaked whale. Our study builds on previous work through the inclusion of genomic sequencing and an increased geographic range of samples by including an additional two samples from South Africa and four samples from South Australia. Despite the increase in genetic data, and the inclusion of individuals from outside the previous study range, we find a lack of population structure to hold for the wider distribution of the species. Moreover, we provide new insights into the maternal demographic history and genetic diversity of Gray’s beaked whales and find the species has existed at relatively high population densities for at least the past one million years.

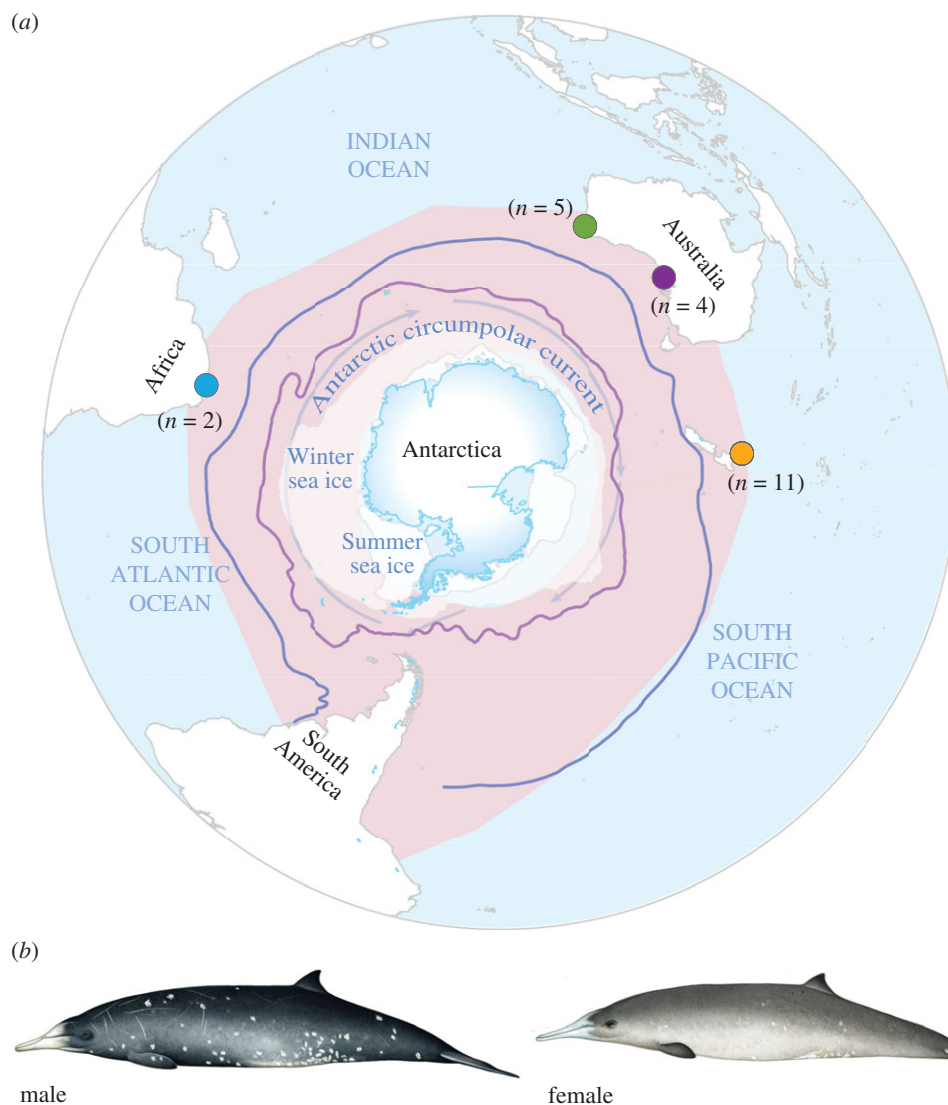


Figure 1. Distribution, sampling locations and morphology of Gray's beaked whale. (a) Pink shaded area represents species distribution based on the IUCN Red List of Threatened Species [17]. Coloured circles denote sample locations: South Africa (blue), Western Australia (green), South Australia (purple), New Zealand (orange), and numbers denote sample sizes. Winter and summer sea ice coverage are indicated by white shading and correspond to average sea ice coverage during the months of March and September 2016 [18]. Subtropical convergence (blue line) and Antarctic polar front (purple line) are indicated [19]. Map scale is 1 : 95 000 000. (b) Male and female Gray's beaked whales; illustration by Mark Camm.

2. Material and methods

2.1. Samples

We analysed a total of 22 samples from South Africa ($n=2$), Western Australia ($n=5$), South Australia ($n=4$) and New Zealand ($n=11$) (figure 1). The samples from Western Australia and New Zealand were included in previous genetic analyses [20]. Samples were collected from beach cast whales during 1999–2013 and provided by the Port Elizabeth Museum (South Africa), South Australian Museum and New Zealand Cetacean Tissue Archive (electronic supplementary material, table S1).

We mapped the distribution of the sampling locations in ArcGIS® (v. 10.3, Esri® Inc.), using a South Pole Lambert azimuthal equal area projection and a World Geodetic System 1984 with map datum at a scale of 1 : 95 000 000, and included the most recent projection of Gray's beaked whale range (figure 1) [20]. The map includes environmental data within Gray's beaked whale habitat; to identify contemporary sea ice extents, ice sheets and glacial projections, we used data from the National Snow and Ice Data Center [18] from Natural Earth 2017. Data from Orsi & Harris [19] were used to identify the location of the subtropical convergence and Antarctic polar front.

2.2. Laboratory analyses

Using the 22 samples available from an approximately 13 000 km range of Gray's beaked whale distribution, we set out to sequence both complete mitochondrial genomes and partial nuclear genomes from shotgun sequencing data. We extracted DNA using a DNeasy blood and tissue kit (Qiagen) following the manufacturer's protocol and sonicated to approximately 450 bp using a Covaris. From the sonicated DNA, we built Illumina sequencing libraries using the BEST protocol [22], using an Illumina adapter mix concentration of 20 μ M and 15 cycles during the indexing PCR step. Indexed libraries were cleaned using a SPRI bead DNA purification method. All samples were multiplexed in approximately equimolar amounts and sequenced on an Illumina Hiseq 4000 at the GeoGenetics Sequencing Core, University of Copenhagen, using 80 bp single-end (SE) and 150 bp paired-end (PE) reads.

2.3. Mitochondrial DNA analyses

2.3.1. Mitochondrial genome assembly

To construct mitochondrial genomes, we initially sequenced the libraries using 80 cycle SE sequencing. We trimmed adapters from the raw reads and removed reads shorter than 30 bp using skewer v. 0.2.2 [23]. We mapped the resultant reads to a Gray's beaked whale mitochondrial genome (Genbank ID: nC_023830.1) using BWA v. 0.7.15 [24] with the mem algorithm, and default parameters. We parsed the output and removed duplicates and reads of mapping quality less than 30 using SAMtools v. 1.6 [25]. We built consensus sequences using a majority rules approach (-doFasta 2) in ANGSD v. 0.921 [26] only considering bases with a base quality score greater than 25 (-minq 25), reads with a mapping quality score greater than 25 (-minmapq 25), and sites with at least 5 x coverage (-minIndepth 5). The resultant consensus sequences were aligned using MAFFT v. 7.392 and default parameters [27].

2.3.2. Diversity statistics

We calculated haplotype diversity (h) and nucleotide diversity (π) in DnaSP v. 6.12.03 [28]. To evaluate the relative level of mitochondrial diversity in Gray's beaked whale, we compared their nucleotide diversity (π) to estimates from 12 other cetaceans obtained using available population-scale mitogenome data from the literature. We used the values previously calculated in Louis *et al.* [29] and added available data from the Blainville's, Cuvier's and Gervais' beaked whales [30] as well as northern bottlenose whales [31], all sampled from across their ranges. Following Louis *et al.* [29], we estimated nucleotide diversity excluding sites with gaps and missing data only in each pairwise comparison in DnaSP.

2.3.3. Haplotype network

We used PopArt v. 1.7 [32] to generate a haplotype neighbour-joining network using a 0.1 reticulation tolerance from the aligned mitochondrial genomes.

2.3.4. Phylogeny

We built a time-calibrated phylogenetic tree in BEAST v. 2.6 [33] to estimate the divergence times among our Gray's beaked whales. For this analysis, we separated the protein-coding genes ($n = 13$), tRNAs ($n = 22$), rRNAs ($n = 2$) and the control region based on the published annotation of the reference Gray's beaked whale mitogenome [34] which we aligned using MAFFT [27]. We manually checked the alignments and, when required, edited them to match the reading frame of the protein-coding genes.

We combined the data into six subsets that were (i) first, (ii) second and (iii) third codon positions of all protein-coding regions, (iv) tRNAs, (v) rRNAs and (vi) control region. We set all regions to have the same topology and clock rate, but specified optimal substitution models for each of the six subsets as identified according to PartitionFinder2 [35] (electronic supplementary material, table S2). We specified a coalescent constant population model and applied a strict clock as we assume little heterogeneity in clock rate within Gray's. We set the clock rate as 0.0038 substitutions per site per million years based on that previously found using an odontocete fossil-calibrated phylogenetic tree [29].

We ran the Markov chain Monte Carlo (MCMC) twice with 50 000 000 steps logged every 5000 steps. We combined the tree files and log files of the two runs using LogCombiner v. 2.5.1 [36] with a burnin of 10%. We checked stationarity (ESS values above 200 for all parameters) and convergence by comparing posterior distributions between the two chains using Tracer v. 1.7 [37]. We used TreeAnnotator v. 2.5.1

[38] to create a maximum clade credibility tree with mean node heights and a posterior probability limit of 0.9 based on the combined trees. We visualized the tree using FigTree v. 1.4.3 (<http://tree.bio.ed.ac.uk/software/figtree/>).

2.3.5. Demographic history

We reconstructed the demographic history of Gray's beaked whales using a coalescent Bayesian skyline plot analysis (BSP) in BEAST v. 2.6 [33] using the fully partitioned mitogenome sequences and all 22 sequences from the dated phylogeny. The substitution models for each of the six partitions were defined as in the previous step. We ran two MCMC chains of 50 000 000 states with sampling every 5000 states, with a burn-in of 10%. We examined ESS values in Tracer v. 1.7.1, and compared the posterior distributions of the two chains to ensure convergence of the runs [37].

To compare the results of a coalescent constant population model with the BSP model and determine the best model fit for our data, we ran the analyses again with identical parameters and derived marginal likelihoods for both models. To evaluate the models, we used the Nested Sampling (NS) package from BEAST2 [39] with default settings except for the particleCount parameter which was increased to 50. The marginal likelihoods produced by NS and their standard deviations were used to determine which model was the better fit for our data. If the difference between the marginal likelihoods of two models were larger than $2\sqrt{(SD1*SD1 + SD2*SD2)}$, where $SD1$ and $SD2$ are the standard deviations of the two models being compared, the model with the larger marginal likelihood was favoured. Finally, we ran an extended BSP in BEAST with identical parameters to previous runs to check whether a constant population model was indicated by the sum(indicators.alltrees) parameter that estimated the number of population change steps. To derive N_{ef} we scaled the population size estimates with a generation time of 23.4 years, based on the average for available beaked whale species provided in Taylor *et al.* [40].

2.4. Nuclear DNA analyses

2.4.1. Scaffolding Sowerby's beaked whale assembly

The only available beaked whale assemblies (*Mesoplodon bidens* (GCA_004027085.1—Sowerby's beaked whale), and *Ziphius cavirostris* (GCA_004364475.1—Cuvier's beaked whale)) are very fragmented (scaffold N50 of 33.5 kb and 3.6 kb respectively), which could impact certain downstream analyses and hinder our ability to locate sex chromosomes for sexing of the individuals. Therefore, we first scaffolded the existing Sowerby's beaked whale assembly using the orca (*Orcinus orca*) assembly (Genbank accession: GCA_000331955.2), with a modified version of the methodology set out by cross-species scaffolding [41]. Although the orca is highly divergent from the Gray's beaked whale (approx. 30 Ma [42]), we selected the orca as it is the most closely related species with a high-quality genome assembly available.

We downloaded the raw reads from Sowerby's beaked whale from the European Nucleotide Archive (SRR7704819) and mapped these to the orca genome using the same protocol as mentioned above. We created a consensus sequence using a consensus base call approach with ANGSD, and specifying a minimum read depth of 8 (-minInddepth 8), minimum mapping quality of 25 (-minmapq 25), minimum base quality of 25 (-minq 25) and only considering reads that mapped uniquely to one location (-uniqueonly 1). We constructed in-silico mate-paired libraries with approximately 1, 2, 5, 8, 10, 15 and 20 kb insert sizes from this consensus sequence using seq-scripts (<https://github.com/thackl/seq-scripts>), specifying a read length of 150 bp and read depth of 30x.

Using these *in silico* mate-paired libraries, we scaffolded the existing Sowerby's beaked whale assembly using SSPACE v. 2 [43] and default parameters. We closed assembly gaps using the adapter-trimmed Sowerby's beaked whale raw reads and abyss-sealer [44], specifying a kmer size of 51. We removed all scaffolds shorter than 1 kb from the gap-closed assembly. We assessed the final assembly quality in the form of contiguity using QUAST [45], and in the form of gene content using BUSCO v. 3 [46] and both the eukaryotic and mammalian BUSCO databases.

2.4.2. Mapping to the orca and Sowerby's beaked whale assemblies

To investigate whether our population genomic analyses may be impacted by choice of reference, we mapped our reads (both SE and PE) to the available orca assembly as well as the Sowerby's beaked whale assembly independently. We mapped the 80 bp SE and 150 bp PE reads independently,

following the same procedure as mentioned above for the mitochondrial genomes, and merged the resultant bam files using SAMtools merge.

2.4.3. Population genomics: principal component analysis, admixture proportions, phylogeny, fixation index and D-statistics

We performed principal component analysis (PCA) and admixture proportion analysis on the data mapped to each assembly (both the orca and Sowerby's beaked whale) independently using ANGSD v. 0.921, as ANGSD uses genotype likelihoods, taking uncertainty in genotype calling of low coverage into account, instead of directly calling genotypes. We first performed a PCA and admixture proportion analysis in PCAngsd [47], using a beagle file constructed in ANGSD. To construct the beagle file, we called genotype likelihoods using the GATK algorithm (-GL 2), specified the output as a beagle file (-doGlf 2) and applied the following filters: only include reads with a mapping quality greater than 25 (-minmapQ 25), only include bases with base quality greater than 25 (-minQ 25), only include reads that map to one location uniquely (-uniqueonly 1), a minimum minor allele frequency of 0.05 or greater (-minmaf 0.05), only call a SNP if the p -value is greater than 1×10^{-6} (-SNP_pval 1×10^{-6}), infer major and minor alleles from genotype likelihoods (-doMajorMinor 1), only include sites if at least 50% of individuals are covered (-MinInd 11), remove sex scaffolds and scaffolds shorter than 100 kb (-rf), and call allele frequencies based on a fixed major and an unknown minor allele (-doMaf 2). In PCAngsd, we specified default parameters apart from the addition of the admixture proportions (-admix). The resultant covariance matrix was converted into a PCA and visualized using R v. 3.4.3 [48].

We further evaluated whether the inclusion of sites in linkage disequilibrium (LD) may have an impact on downstream results. To do this, we also ran these analyses using the data mapped to the Sowerby's beaked whale assembly and a subset of 70 140 sites determined not to be in LD using ngsLD [49] specifying the parameters -max_kb_dist of 20 and -min_weight 0.5.

To further investigate geographic structure in our data, we built a neighbour-joining tree using the data mapped to the Sowerby's beaked whale. We built a distance matrix in ANGSD using a consensus base call approach (-doIBS 2), and the same parameters as the beagle file construction for the PCA. We converted the distance matrix into a nexus tree file using PHYLIP neighbour [50].

As an additional method to investigate for the presence (or absence) of population structure based on geography, we computed the fixation index (F_{ST}) between individuals grouped by location. For this analysis, we only used the data mapped to the Sowerby's beaked whale assembly. We computed a consensus haploid call using ANGSD (-dohaplocall 2) and specified the same filtering parameters as the PCA and admixture proportions above. We converted the output haplo file into a geno file by removing the major allele column to be used as input for a python script available online (https://github.com/simonhmartin/genomics_general/blob/master/popgenWindows.py). For this, we grouped individuals based on geographical origin (i.e. South Africa, West Australia, South Australia and New Zealand), and specified a window size of 100 kb, and a minimum number of sites per window as 500 bp. Significance from 0 was evaluated using a one-sample Wilcoxon signed-rank test in R.

As a final test for population structure based on geographical origin, we performed a D-statistics (ABBA/BABA) test for topology (e.g. [51,52]) on the data mapped to the Sowerby's beaked whale assembly. Although D-statistics are most commonly used to find evidence of gene flow, represented by a high D-score, a high D-score can also be caused by more recent common ancestry brought about by an incorrect predefined topology. Taking the latter into account, we placed individuals into four predefined groups based on sampling location (i.e. South Africa, Western Australia, South Australia and New Zealand) and compared the average D values produced from correct topologies, that is, branches H1 and H2 containing individuals from the same locality, and incorrect topologies, that is, branches H2 and H3 have individuals from the same locality, while H1 is from a different locality. Furthermore, to investigate for evidence of subpopulation structure within localities, we also performed this test with individuals from a single locality placed on the H1, H2 and H3 branches.

We computed D-statistics in ANGSD using a random base call approach (-doAbbaBaba 1) and the following parameters: a minimum mapping and quality score of 25 (-minmapQ 25, -minQ 25), only consider reads that map to one location uniquely (-uniqueonly 1), calculate genotype likelihoods using GATK algorithm (-GL 2), calculate major and minor alleles using genotype likelihoods (-doMajorMinor 1), skip sites that are triallelic (-skiptrialelic 1), calculate D using a block size of 1Mb (-blocksize 1 000 000) and specify the Sowerby's beaked whale as the outgroup. We performed a block jackknife approach to test for the significance of the results (Z score) using the Rscript (jackKnife.R)

available as part of the ANGSD toolsuite. ANGSD computes every possible combination of individuals, which we filtered based on the sampling localities of the individuals. Scripts on how to filter the ANGSD output to test for population structure can be found at <https://github.com/Mvwestbury/Dstats-topology-test>. The significance of these results, i.e. whether $|Z|$ was greater than 3, was determined for each comparison using a one-sample Wilcoxon signed-rank test in R.

2.4.4. Relatedness and sex determination

To assess levels of relatedness between individuals originating from the same geographic location, we used NGSrelate v. 2 [53]. As input for this, we calculated genotype likelihoods of our dataset using ANGSD following similar parameters as for the PCA analysis. However, there were some minor adjustments to produce the required input file for NGSrelate, namely the inclusion of `-doGlf 3` which prints the output as a binary likelihood file as opposed to a beagle file. The number of sites used between pairwise comparisons ranged from 302 286 to 651 438.

In this study, we used the low-coverage genomes to identify the sex of new and previously analysed samples. Our method was as follows: we identified scaffolds in the orca and Sowerby's beak whale assemblies putatively belonging to sex chromosomes using SatsumaSynteny v. 2.01 [54], default parameters, and aligning each assembly independently to the cow X (Genbank accession: CM008168.2) and human Y (Genbank accession: NC_000024.10) chromosomes. Following this, we used the sex chromosome to autosome read depth ratio to determine the sex of the individuals in the current study. This was done by comparing the mean read depth across scaffolds putatively originating from the X chromosome, to that of scaffolds putatively originating from the autosomes. Moreover, to remove biases that may occur due to homology between some regions of the X and Y chromosomes, we removed any scaffolds that aligned to both the X and Y chromosomes from this comparison. This resulted in X-linked scaffolds totalling a length of approximately 120 Mb. We calculated the average read depth of these scaffolds using the depth function in SAMtools. We further selected scaffolds totalling approximately 120 Mb that did not align to the X or Y chromosomes and calculated the average read depth along these scaffolds for comparison.

3. Results

3.1. Mitochondrial DNA

3.1.1. Mitogenomes: coverage, diversity and population structure

We constructed mitochondrial genomes from all 22 individuals to coverages of between 27.0 \times and 112.7 \times , with a mean coverage of 58.3 \times (electronic supplementary material, table S5). Gray's beaked whales exhibit relatively high mitochondrial nucleotide diversity ($\pi=0.00742$) in comparison with the other surveyed odontocete species ($\pi=0.00094$ – 0.00941) (figure 2). The haplotype network and phylogenetic tree show no apparent pattern of geographic structure across the study area (figure 3*a,b*).

3.1.2. Demographic history

We analysed the demographic structure of the mitochondrial genomes with two different models (BSP and constant population). The marginal likelihoods of the models show that the BSP model is the best fit for our data (marginal likelihood (s.d.), BSP = -26590.39 (1.91) and constant = -26611.62 (2.04)). In addition, the extended BSP analyses indicated that we can confidently reject a constant population since the $\text{summ}(\text{indicator.alltrees})$ 95% HPD excluded 0. The BSP analysis suggests that Gray's beaked whales have existed as a relatively stable population over at least the past 1.1 million years (figure 4*a*) with an approximate doubling of female effective population size (N_{ef}) around approximately 250 thousand years ago (kya). The current estimated median N_{ef} is 270 500 whales (95% CI 116 544–1 244 228). Our findings are broadly consistent with those reported by Thompson *et al.* [20,34]. Our estimate of N_{ef} falls within the range reported in Thompson *et al.* [20,34] that used a mutation rate estimated for Cuvier's beaked whale (N_{ef} 433 339 whales, 95% CI 16 667–2 213 248). Our study indicates a BSP trajectory that is similar to previous work, but the current genomic analyses have been able to estimate N_{ef} and the timing of population expansions with greater confidence and more resolution.

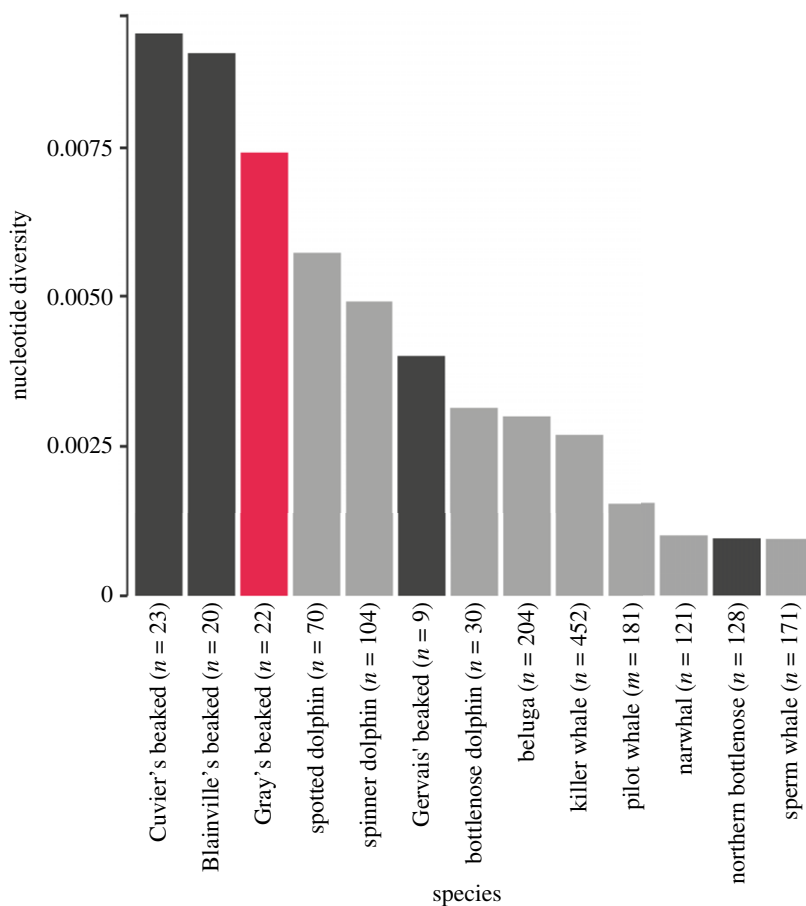


Figure 2. Comparison of nucleotide diversity (π) derived from population-level studies of 13 toothed whale mitochondrial genomes. Gray's beaked whales are shown in red, and other beaked whales are shown in dark grey. Sample sizes for each species are given in the legend.

3.2. Nuclear DNA

3.2.1. Newly scaffolded Sowerby's beaked whale assembly

We scaffolded the publicly available Sowerby's beaked whale contig-level genome assembly (N50 33.5 kb) using *in silico* mate-paired libraries of between 1 and 20 kb in length and SSPACE v. 2 [43], resulting in an assembly of 2 356 367 458 bp in total, with a scaffold N50 of 2.6 Mb (electronic supplementary material, table S3) and a mammalian database BUSCO score of 93% complete BUSCOs (electronic supplementary material, table S4). From this assembly, we successfully identified 145 scaffolds possibly associated with the X chromosome to be used for downstream sex determination analyses.

3.2.2. Nuclear mapping results

We mapped our data independently to two nuclear reference genomes. There was an average of approximately 13% less bp mapping to the orca assembly relative to the Sowerby's beaked whale assembly (tables S5 and S6). The higher mapping count when mapping to the more closely related Sowerby's beaked whale genome is not unexpected, as the orca diverged from the Gray's beaked whale approximately 29.98 Ma (CI 29.0–30.94), while Sowerby's beaked whale diverged only approximately 7.42 Ma (CI 6.67–8.18) [42].

3.2.3. Population genomics: principal component analysis, admixture proportions, phylogeny, fixation index and D-statistics

To investigate the biparental population structure within the Gray's beaked whales, we applied several different population genomic analyses. PCA and admixture proportion analyses separated four New Zealand individuals regardless of the mapping reference and whether a subset of loci determined not

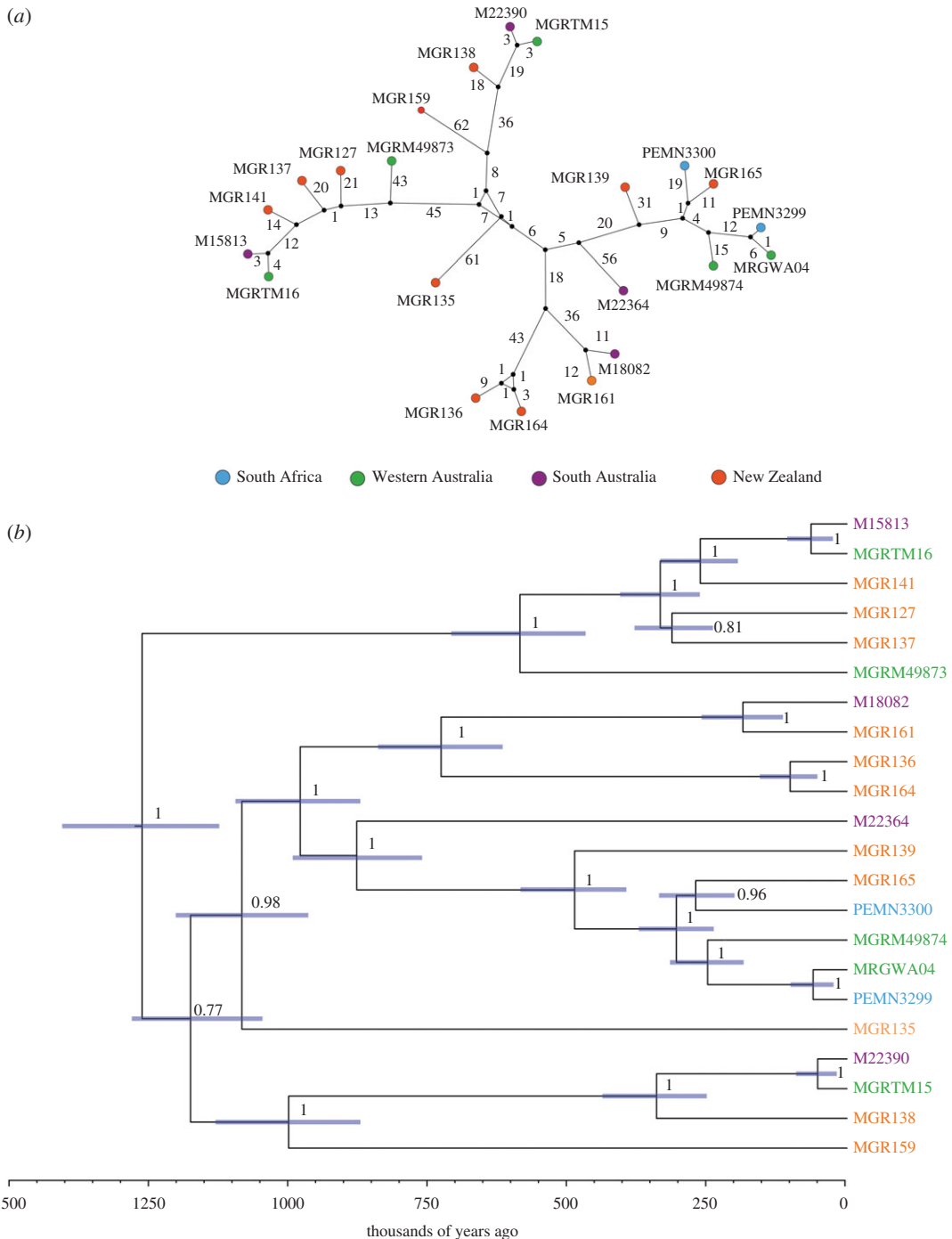


Figure 3. The absence of geographic population structure in 22 Gray's beaked whale mitochondrial genomes. (a) Haplotype network of mitochondrial genomes. Nodes represent haplotypes, and colours correspond to sampling localities; black nodes represent theoretical, unsampled haplotypes; numbers adjacent to lines indicate the number of mutations between haplotypes. Distances between haplotypes are not to scale. (b) Dated mitochondrial phylogeny with node age error bars shown as blue indicating 95% highest posterior density for each divergence, and numbers show posterior probability values.

to be in linkage disequilibrium was considered (electronic supplementary material, figures S1 and S2). Moreover, to investigate for population structure without any *a priori* restrictions on number of populations, we also computed a genome-wide distance matrix and neighbour-joining tree using all individuals mapped to Sowerby's beaked whale. The resultant tree showed no obvious phylogeographic population structure (figure 5a). In addition, the fixation index (F_{ST}) between individuals grouped by geographical origins did not show any significant differences from 0 (figure 5b).

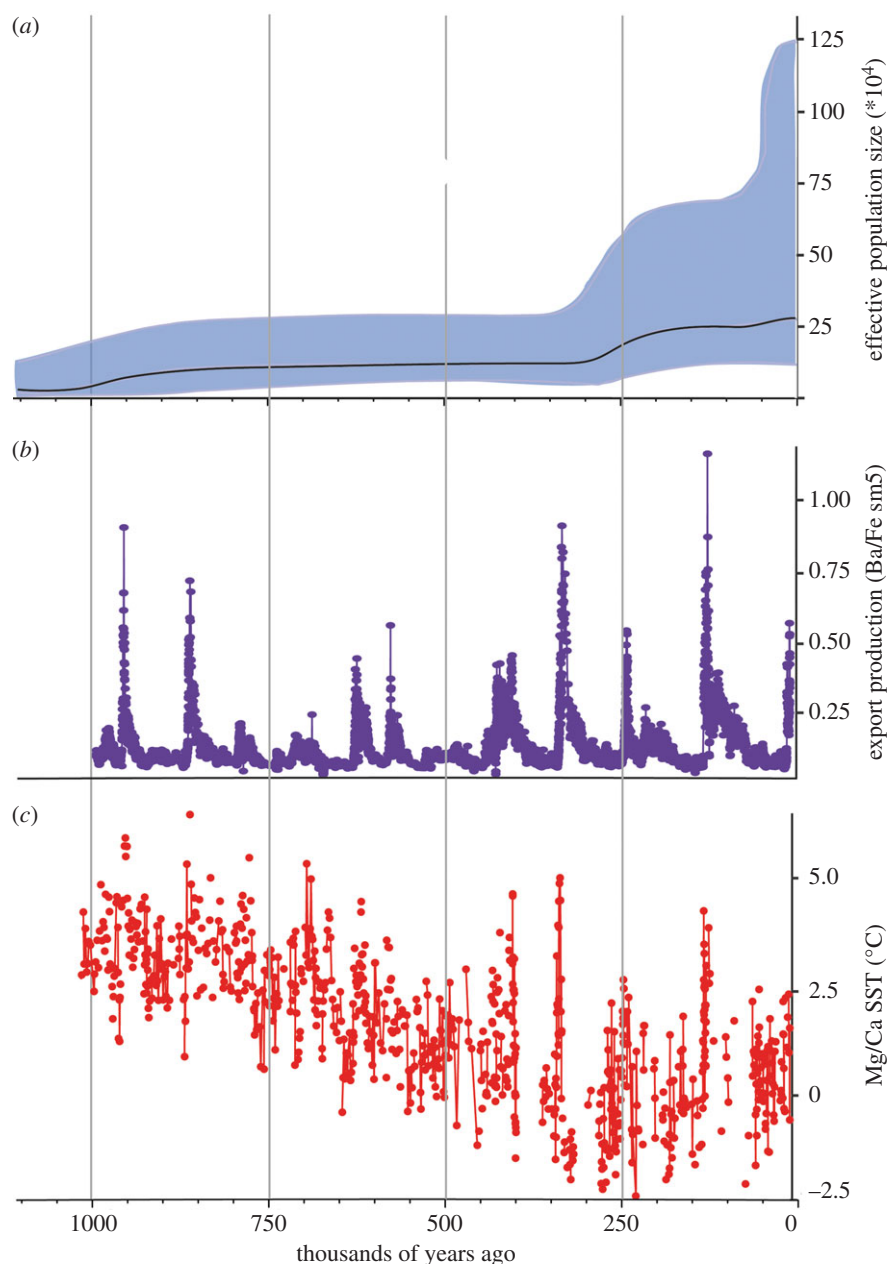


Figure 4. Matrilineal demographic history of 22 Gray's beaked whales based on mitochondrial genomes and environmental context. (a) Bayesian skyline plot showing median female effective population size and 95% confidence intervals through time; (b) export production as described by marine barite (Ba/Fe) as a qualitative proxy across time [55] and (c) Mg/Ca in *Noctiluca pachyderma* as a quantitative estimate of past sea surface temperature (SST) [56]. Current SST in the Southern Ocean sampling locations is -2.5°C .

To further investigate the presence of population structure, we used the D-statistics analysis to test for population structure, as this has previously been shown to be more sensitive to small genetic differences among populations [51] (electronic supplementary material, figures S4–S7). This analysis did not uncover any significant differences from 0, regardless of whether we tested 'correct' or 'incorrect' topologies. The only significant difference was retrieved when comparing the 'incorrect' topology of samples from Western Australia and South Africa (p -value=0.024) (electronic supplementary material, figure S4). This result means that when computing D-statistics on the predefined topology [[[Western Australia, South Africa], South Africa], outgroup], the two South African samples appeared more closely related to each other than they were to any of the individuals from Western Australia. However, we had only two samples from South Africa, and this was not seen when comparing South Africa to any other sampling localities. Therefore, it may be a product of not having enough individuals for this specific comparison.

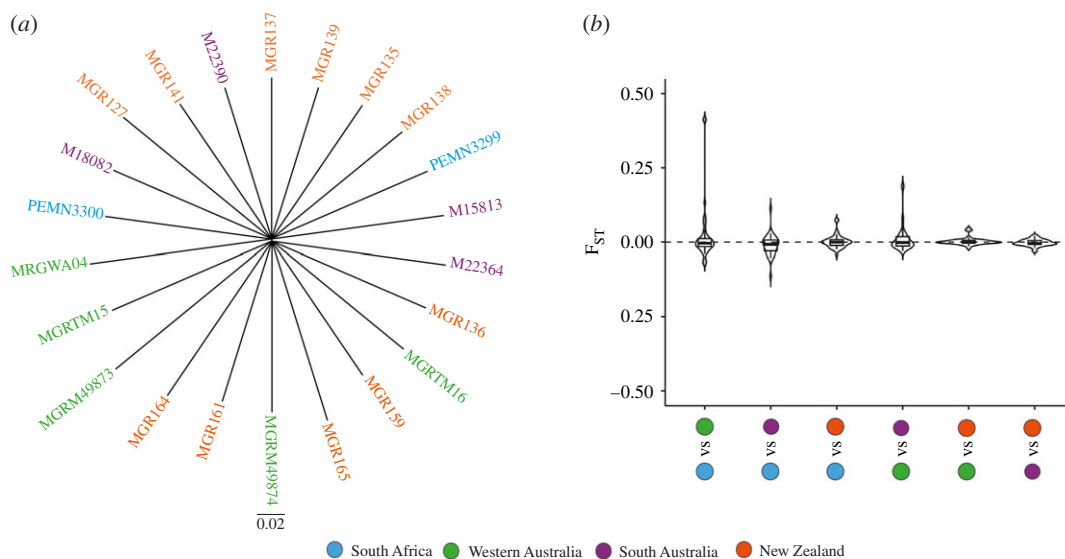


Figure 5. The absence of geographic population structure in nuclear genomes. (a) Neighbour-joining phylogenetic tree constructed from 3 477 967 segregating sites across autosomes; branch lengths and scale axis show the proportion of site differences between individuals; (b) fixation index (F_{ST}) between sample localities based on 1 581 396 segregating sites. Variance across the genome was calculated using a window size of 100 kb and a minimum number of sites per window of 500. Significance from 0 was evaluated using a one-sample Wilcoxon signed-rank test in R.

3.2.4. Relatedness and sex determination

The highest level of relatedness between two individuals in our dataset was 0.013. This value was recovered from the West Australian individuals MGRTM16 and MGRTM15. Although this is the highest level in our dataset, it is still very low and suggests these individuals are only related to the sixth or seventh degree. A relatedness of one indicates monozygotic twins while 0.5 indicates parent-offspring or siblings.

To determine the sex of the Gray's beaked whales in this study, we used the ratio of the mean coverage on the scaffolds putatively originating from the X chromosome compared to autosomal scaffolds. Despite some differences in the recovered ratios based on mapping reference, the assigned sexes do not differ (electronic supplementary material, tables S5 and S6). The results are concordant with previous methods based on the amplification of the SRY gene for a subsample of the samples (electronic supplementary material, tables S5 and S6) [20]. Concordance between methods shows that low-effort shotgun sequencing is an effective method to assess the sex of Gray's beaked whale individuals. In total, we found 12 males and 10 females.

4. Discussion

By performing shallow shotgun sequencing of 22 Gray's beaked whale individuals sampled across approximately 13 000 km of the species' range, we were able to successfully recover complete mitochondrial genomes and partial nuclear genomes from all individuals. This increases the amount of genetic data available for this species, increasing the power of evolutionary inference we can make even with few available samples [57]. Furthermore, by including individuals from new sampling localities in South Africa and South Australia, we are able to broaden inferences to the wider Southern Hemisphere, rather than Australasia specifically.

We recovered no discernible geographic population structure in either the mitochondrial or nuclear genome data, suggesting a lack of both maternal and paternal population structure across the study area in this potentially highly mobile marine predator (figures 3 and 5, electronic supplementary material, figures S1–S7). We acknowledge that our data were derived from stranded animals, and as such we assigned each whale a broad geographic location. However, we do not expect that dead whales would drift across regions, based on the fact that we have so many genome-wide markers, if there was subtle structure, we would expect some degree of genetic clustering between sampling localities.

The absence of clear maternal and biparental population structure across the Southern Hemisphere, as well as a lack of closely related individuals being found at the same site, suggests that both sexes disperse from their natal group, perhaps forming affiliations with unrelated individuals over vast distances. Such a dispersal method has previously been suggested based on genetic kinship data from Gray's beaked whale individuals found in stranded groups [21]. High rates of dispersal, a flexible social system in combination with the rich oceanic habitat of the Southern Hemisphere, could facilitate gene flow and potentially support a surprisingly large number of Gray's beaked whales. The seemingly range-wide gene flow suggested by our data could be facilitated by broad-scale oceanographic features, such as the subtropical convergence that supports large populations of squid [58], and which runs through the centre of Gray's beaked whale distribution (figure 1). High levels of gene flow have also been noted in several Antarctic and Southern Ocean species: highly vagile snow petrels (*Pagodroma nivea*) [59] and central-place foragers such as penguins (several species) [60]. By contrast, other Southern Hemisphere whales, such as the wide-ranging blue whale (*Balaenoptera musculus*), do exhibit moderate genetic structure across its range, even though large-scale movements have been reported (up to 6650 km) [61].

Comparisons of mitochondrial genome diversity between Gray's beaked whale and several other cetacean species revealed relatively high levels of maternal genetic diversity within Gray's beaked whale, consistent with previous results using both partial mitochondrial and microsatellite markers (figure 2) [20]. While the same study also indicated a lack of population structure [20], inferences on wider patterns of population structure were not possible as sampling was restricted to two localities only. Here, we further explore the wider population structure in Gray's beaked whales by including whales from South Africa and South Australia. We find that, despite a distance of up to approximately 13 000 km between sampling locations, Gray's beaked whales recovered in South Africa could not be differentiated from the Australasian individuals at either the mitochondrial or nuclear genomic level (figures 3 and 5, electronic supplementary material, figures S1–S7).

To explore the genetic diversity of Gray's beaked whales through time, we performed demographic history analysis on the mitochondrial genomes. We uncover relatively stable and high levels of female effective population size over at least the past approximately 1.1 million years, but an approximate doubling of N_{ef} approximately 250 kya [20] (figure 4a).

The increase in N_{ef} coincides with an interglacial period during a proposed increase in Southern Ocean productivity [55]. This increased productivity is shown as export production, inferred by the presence of marine barite in sediments (Ba/Fe ratio) used as a qualitative proxy [55,62] (figure 4b). During this time, sea surface temperature (SST) [56] also showed a marked increase (figure 4c). Moreover, the peak in sea surface temperature 250 kya, shows the ocean being warmer or at least as warm as it is today. A plausible explanation for the increase in N_{ef} could be that productivity increased in polar waters at the same time as a shift in temperature, which may have allowed an expansion in suitable habitat at this time. Increases in primary production are likely to have provided more food for whale populations, resulting in increased reproduction with subsequent expansion and divergence of existing lineages.

Aside from the doubling in N_{ef} approximately 250 kya, the long-term stability of N_{ef} of the Gray's beaked whale could reflect that its circumpolar habitat has been relatively stable over the long term. Oceanographic features, such as the Antarctic circumpolar current, subtropical convergence and Antarctic polar front (shown in figure 1), have all existed since at least 32–30 Ma, and the land masses of the Southern Ocean have not significantly altered their position for at least 28 Ma [63,64]. However, during periods of recent climate change, for example, the Pleistocene/Holocene transition of approximately 11.7 kya, we do not see any significant changes in maternal N_{ef} (figure 4a).

Our results indicate that Gray's beaked whales have existed as a diverse and stable population across their distribution over the long term, accruing relatively high genetic diversity in comparison with other toothed whales. It is likely that a flexible social system, rich Southern Hemisphere feeding grounds and permanent oceanographic features facilitate gene flow in this species. Given that these whales have never been the focus of whaling—whalers named them 'scamperdown whales' due to their elusive behaviour—Gray's beaked whales appear to have been able to maintain ancestral levels of high genetic diversity and abundance.

In conclusion, through generating complete mitochondrial genomes, a new low-coverage reference panel of nuclear genomes, and a newly scaffolded Sowerby's beaked whale reference genome, we contribute new and complementary insights into the ecology of a relatively unknown whale. We show that Gray's beaked whales have relatively high levels of maternal genetic diversity, which they have retained over the long term, putatively due to the ability of the species to move across oceans. Such

characteristics may facilitate the future survival of this species in a habitat that is undergoing rapid change. Our study shows the ecological information gained from genetic analyses of beached specimens, even with relatively low sequencing effort, and suggests this is a valuable approach for studying highly elusive species where samples are available.

Data accessibility. Genbank accession codes for mitogenomes are MW645443–MW645463 and raw reads for nuclear genomes can be found in Genbank BioProject ID PRJNA702760. Additional Data and relevant code for this research work are stored in GitHub: <https://github.com/Mvwestbury/Dstats-topology-test> and have been archived within the Zenodo repository: <https://doi.org/10.5281/zenodo.4320997>

Authors' contributions. K.F.T. conceived the project; M.V.W. and K.F.T. performed the laboratory work; M.V.W., K.F.T., M.L., A.A.C., M.S. and J.A.S.C. performed bioinformatic analyses. R.C. provided samples. K.F.T., E.D.L. and J.R.S. provided funding. E.D.L. provided supervision. M.V.W., K.F.T. and E.D.L. wrote the manuscript with input from all other authors.

Competing interests. We declare we have no competing interests.

Funding. The work was funded by the Lerner Grey Memorial Fund for Marine Research of the American Natural History Museum grant and a University of Exeter European Network Fund (#ENF15) grant to K.F.T. The work was also supported by the Villum Fonden Young Investigator Programme, grant no. 13151, the Carlsberg Foundation Distinguished Associate Professor Fellowship, grant no CF16-0202 and the Independent Research Fund Denmark | Natural Sciences, Forskningsprojekt 1, grant no. 8021-00218B to E.D.L. A.A.C. was funded by the Rubicon-NWO grant (project 019.183EN.005).

Acknowledgements. We thank the following institutions and staff who provided samples for this study: New Zealand Cetacean Tissue Archive—University of Auckland (UoA), South Australia Museum (SAM), West Australia Museum and Port Elizabeth Museum (PEM). Specifically, we thank C. Kemper, S. South, S. Donnellan and D. Stemmer from SAM and G. Hofmeyr from PEM. West Australian samples were collected by N. Gales, R. O'Shea and J. Bannister. New Zealand samples were collected by the Department of Conservation staff and local Māori. Thanks also to S. Jaccard for help with interpretation of Southern Ocean environmental data and C. S. Baker and D. Steel for management of the UoA samples.

References

- Pörtner HO *et al.* 2019 *IPCC special report on the ocean and cryosphere in a changing climate*. Geneva, Switzerland: IPCC Intergovernmental Panel on Climate Change.
- Ramirez-Llodra E *et al.* 2010 Deep, diverse and definitely different: unique attributes of the world's largest ecosystem. *Biogeosciences* **7**, 2851–2899. (doi:10.5194/bg-7-2851-2010)
- Costello MJ, Chaudhary C. 2017 Marine biodiversity, biogeography, deep-sea gradients, and conservation. *Curr. Biol.* **27**, 2051. (doi:10.1016/j.cub.2017.06.015)
- Parsons ECM, Monaghan-Brown D. 2017 From Hunting to watching: human interactions with cetaceans. In *Marine mammal welfare: human induced change in the marine environment and its impacts on marine mammal welfare* (ed. A Butterworth), pp. 67–89. Cham, Switzerland: Springer International Publishing.
- Thompson K, Baker CS, van Helden A, Patel S, Millar C, Constantine R. 2012 The world's rarest whale. *Curr. Biol.* **22**, R905. (doi:10.1016/j.cub.2012.08.055)
- Dalebout ML *et al.* 2014 Resurrection of *Mesoplodon hotaula* Deraniyagala 1963: a new species of beaked whale in the tropical Indo-Pacific. *Mar. Mamm. Sci.* **30**, 1081–1108. (doi:10.1111/mms.12113)
- Yamada TK, Kitamura S, Abe S, Tajima Y, Matsuda A, Mead JG, Matsuishi TF. 2019 Description of a new species of beaked whale (*Berardius*) found in the North Pacific. *Sci. Rep.* **9**, 12723. (doi:10.1038/s41598-019-46703-w)
- The society for Marine Mammalogy. In press. List of Marine Mammal Species and Subspecies. See <https://marinemammalscience.org/species-information/list-marine-mammal-species-subspecies/> (accessed on 9 August 2020).
- Pitman R. 2009 Mesoplodont whales: (*Mesoplodon* spp.). In *Encyclopedia of marine mammals*, 2nd edn (eds WF Perrin, B Würsig, JGM Thewissen), pp. 721–726. London, UK: Academic Press.
- Würsig B, Perrin WF, Thewissen JGM. 2009 *Encyclopedia of marine mammals*. New York, NY: Academic Press.
- Schorr GS, Falcone EA, Moretti DJ, Andrews RD. 2014 First long-term behavioral records from Cuvier's beaked whales (*Ziphius cavirostris*) reveal record-breaking dives. *PLoS ONE* **9**, e92633. (doi:10.1371/journal.pone.0092633)
- Hooker SK *et al.* 2019 Future directions in research on beaked whales. *Front. Mar. Sci.* **5**, 514. (doi:10.3389/fmars.2018.00514)
- Baird RW. 2019 Behavior and ecology of not-social odontocetes: Cuvier's and Blainville's beaked whales. In *Ethology and behavioral ecology of odontocetes*, pp. 305–329. Berlin, Germany: Springer.
- Baird RW, Webster DL, McSweeney DJ, Ligon AD, Schorr GS, Barlow J. 2006 Diving behaviour of Cuvier's (*Ziphius cavirostris*) and Blainville's (*Mesoplodon densirostris*) beaked whales in Hawaii. *Can. J. Zool.* **84**, 1120–1128. (doi:10.1139/z06-095)
- Dinis A, Marques R, Dias L, Sousa D, Gomes C, Abreu N, Alves F. 2017 Site fidelity of Blainville's beaked whale (*Mesoplodon densirostris*) off Madeira Island (Northeast Atlantic). *Aquat. Mamm.* **43**, 387. (doi:10.1578/AM.43.4.2017.387)
- Reyes SC. 2018 *Abundance estimate, survival and site fidelity patterns of Blainville's (Mesoplodon densirostris) and Cuvier's (Ziphius cavirostris) beaked whales off El Hierro (Canary Islands)*. Unpublished PhD thesis, University of St Andrews, UK. See <https://research-repository.st-andrews.ac.uk/handle/10023/14189>.
- Pitman RL, Taylor BL. 2020 *Mesoplodon grayi*. The IUCN Red List of Threatened Species 2020: e.T13247A50366236. See <https://dx.doi.org/10.2305/IUCN.UK.2020-3.RLTS.T13247A50366236.en> (accessed 14 March 2021).
- Fetterer F, Knowles K, Meier W, Savoie M, Windnagel AK. 2017 *Sea ice index, version 3*. Boulder, CO: National Snow and Ice Data Center (NSIDC).
- Orsi AH, Harris U. 2019 Fronts of the Antarctic Circumpolar Current – GIS data, Ver. 1, Australian Antarctic Data Centre. See https://data.aad.gov.au/metadata/records/antarctic_circumpolar_current_fronts
- Thompson KF, Patel S, Baker CS, Constantine R, Millar CD. 2016 Bucking the trend: genetic analysis reveals high diversity, large population size and low differentiation in a deep ocean cetacean. *Heredity* **116**, 277–285. (doi:10.1038/hdy.2015.99)
- Patel S, Thompson KF, Santure AW, Constantine R, Millar CD. 2017 Genetic kinship analyses reveal that Gray's beaked whales strand in

- unrelated groups. *J. Hered.* **108**, 456–461. (doi:10.1093/jhered/esx021)
22. Carøe C, Gopalakrishnan S, Vinner L, Mak SST, Sinding MHS, Samaniego JA, Wales N, Sicheritz-Pontén T, Gilbert MTP. 2018 Single-tube library preparation for degraded DNA. *Methods Ecol. Evol.* **9**, 410–419. (doi:10.1111/2041-210X.12871)
 23. Jiang H, Lei R, Ding S-W, Zhu S. 2014 Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinf.* **15**, 182. (doi:10.1186/1471-2105-15-182)
 24. Li H, Durbin R. 2009 Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760. (doi:10.1093/bioinformatics/btp324)
 25. Li H *et al.* 2009 The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079. (doi:10.1093/bioinformatics/btp352)
 26. Korneliusen TS, Albrechtsen A, Nielsen R. 2014 ANGSD: analysis of next generation sequencing data. *BMC Bioinf.* **15**, 356. (doi:10.1186/s12859-014-0356-4)
 27. Katoh K, Standley DM. 2013 MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780. (doi:10.1093/molbev/mst010)
 28. Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R. 2003 DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**, 2496–2497. (doi:10.1093/bioinformatics/btg359)
 29. Louis M *et al.* 2020 Influence of past climate change on phylogeography and demographic history of narwhals, *Monodon monoceros*. *Proc. R. Soc. B* **287**, 20192964. (doi:10.1098/rspb.2019.2964)
 30. Morin PA, Duchene S, Lee N, Durban J, Claridge D. 2013 Preliminary analysis of mitochondrial genome phylogeography of Blainville's, Cuvier's and Gervais' beaked whales. *Report to the Scientific Committee of the International Whaling Commission (No. SC64/SM14)*.
 31. Feyrer LJ, Bentzen P, Whitehead H, Paterson IG, Einfield A. 2019 Evolutionary impacts differ between two exploited populations of northern bottlenose whale (*Hyperoodon ampullatus*). *Ecol. Evol.* **9**, 13 567–13 584. (doi:10.1002/ece3.5813)
 32. Leigh JW, Bryant D. 2015 popart: full-feature software for haplotype network construction. *Methods Ecol. Evol.* **6**, 1110–1116. (doi:10.1111/2041-210X.12410)
 33. Bouckaert R *et al.* 2019 BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **15**, e1006650. (doi:10.1371/journal.pcbi.1006650)
 34. Thompson KF, Patel S, Williams L, Tsai P, Constantine R, Baker CS, Millar CD. 2016 High coverage of the complete mitochondrial genome of the rare Gray's beaked whale (*Mesoplodon grayi*) using Illumina next generation sequencing. *Mitochondrial DNA A DNA Mapp Seq. Anal.* **27**, 128–129. (doi:10.3109/19401736.2013.878908)
 35. Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B. 2017 PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol.* **34**, 772–773.
 36. Rambaut A, Drummond AJ. 2014 LogCombiner v. 2.1.3. Institute of Evolutionary Biology, University of Edinburgh, UK.
 37. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018 Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* **67**, 901–904. (doi:10.1093/sysbio/syy032)
 38. Rambaut A, Drummond AJ. 2013 TreeAnnotator v. 1.7.0. Available as part of the BEAST package. See <http://beast.bio.ed.ac.uk>
 39. Russel PM, Brewer BJ, Klaere S, Bouckaert RR. 2019 Model selection and parameter inference in phylogenetics using nested sampling. *Syst. Biol.* **68**, 219–233. (doi:10.1093/sysbio/syy050)
 40. Taylor BL, Chivers SJ, Larese J, Perrin WF. 2007 *Generation length and percent mature estimates for IUCN assessments of cetaceans*. Administrative Report LJ-07-01. La Jolla, CA: National Marine Fisheries Service, Southwest Fisheries Science Center.
 41. Grau JH, Hackl T, Koepfli K-P, Hofreiter M. 2018 Improving draft genome contiguity with reference-derived in silico mate-pair libraries. *Gigascience* **7**, gij029.
 42. McGowen MR *et al.* 2020 Phylogenomic resolution of the cetacean tree of life using target sequence capture. *Syst. Biol.* **69**, 479–501. (doi:10.1093/sysbio/syz068)
 43. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011 Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**, 578–579. (doi:10.1093/bioinformatics/btq683)
 44. Paulino D, Warren RL, Vandervalk BP, Raymond A, Jackman SD, Birol I. 2015 Sealer: a scalable gap-closing application for finishing draft genomes. *BMC Bioinf.* **16**, 230. (doi:10.1186/s12859-015-0663-4)
 45. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013 QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075. (doi:10.1093/bioinformatics/btt086)
 46. Seppy M, Manni M, Zdobnov EM. 2019 BUSCO: assessing genome assembly and annotation completeness. *Methods Mol. Biol.* **1962**, 227–245. (doi:10.1007/978-1-4939-9173-0_14)
 47. Meisner J, Albrechtsen A. 2018 Inferring population structure and admixture proportions in low-depth NGS Data. *Genetics* **210**, 719–731. (doi:10.1534/genetics.118.301336)
 48. R Core Team. 2013 R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. See <https://www.R-project.org/>.
 49. Fox EA, Wright AE, Fumagalli M, Vieira FG. 2019 ngsLD: evaluating linkage disequilibrium using genotype likelihoods. *Bioinformatics* **35**, 3855–3856. (doi:10.1093/bioinformatics/btz200)
 50. Felsenstein J. 2005 PHYLIP (Phylogeny Inference Package) version 3.6. Department of Genome Sciences, University of Washington, Seattle, WA. See <http://evolution.genetics.washington.edu/phylib.html>.
 51. Westbury MV *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. (doi:10.1093/molbev/msy037)
 52. Barlow A *et al.* 2018 Partial genomic survival of cave bears in living brown bears. *Nat. Ecol. Evol.* **2**, 1563–1570. (doi:10.1038/s41559-018-0654-8)
 53. Hanghøj K, Moltke I, Andersen PA, Manica A, Korneliusen TS. 2019 Fast and accurate relatedness estimation from high-throughput sequencing data in the presence of inbreeding. *Gigascience* **8**. (doi:10.1093/gigascience/gjz034)
 54. Grabherr MG, Russell P, Meyer M, Mauceli E, Alfoldi J, Di Palma F, Lindblad-Toh K. 2010 Genome-wide synteny through highly sensitive sequence alignment: Satsuma. *Bioinformatics* **26**, 1145–1151. (doi:10.1093/bioinformatics/btq102)
 55. Jaccard SL, Hayes CT, Martínez-García A, Hodell DA, Anderson RF, Sigman DM, Haug GH. 2013 Two modes of change in Southern Ocean productivity over the past million years. *Science* **339**, 1419–1423. (doi:10.1126/science.1227545)
 56. Hasenfratz AP *et al.* 2019 The residence time of Southern Ocean surface waters and the 100 000-year ice age cycle. *Science* **363**, 1080–1084. (doi:10.1126/science.aat7067)
 57. Willing E-M, Dreyer C, van Oosterhout C. 2012 Estimates of genetic differentiation measured by F_{ST} do not necessarily require large sample sizes when using many SNP markers. *PLoS ONE* **7**, e42649. (doi:10.1371/journal.pone.0042649)
 58. Butler ECV, Butt JA, Lindstrom EJ, Teldesley PC, Pickmere S, Vincent WF. 1992 Oceanography of the subtropical convergence zone around southern New Zealand. *N. Z. J. Mar. Freshw. Res.* **26**, 131–154. (doi:10.1080/00288330.1992.9516509)
 59. Carrea C, Burridge CP, Wienecke B, Emmerson LM, White D, Miller KJ. 2019 High vagility facilitates population persistence and expansion prior to the Last Glacial Maximum in an Antarctic top predator: the snow petrel (*Pagodroma nivea*). *J. Biog.* **46**, 442–453. (doi:10.1111/jbi.13513)
 60. Cole TL *et al.* 2019 Receding ice drove parallel expansions in Southern Ocean penguins. *Proc. Natl Acad. Sci. USA* **116**, 26 690–26 696. (doi:10.1073/pnas.1904048116)
 61. Sremba AL, Hancock-Hanser B, Branch TA, LeDuc RL, Baker CS. 2012 Circumpolar diversity and geographic differentiation of mtDNA in the critically endangered Antarctic blue whale (*Balaenoptera musculus intermedia*). *PLoS ONE* **7**, e32579. (doi:10.1371/journal.pone.0032579)
 62. Paytan A, Griffith EM. 2007 Marine barite: recorder of variations in ocean export productivity. *Deep Sea Res. Part 2 Top. Stud. Oceanogr.* **54**, 687–705. (doi:10.1016/j.dsr2.2007.01.007)
 63. Buening N, Carlson SJ, Spero HJ, Lee DE. 1998 Evidence for the Early Oligocene formation of a proto-Subtropical Convergence from oxygen isotope records of New Zealand Paleogene brachiopods. *Palaeoogeogr. Palaeoecol. Palaeoecol.* **138**, 43–68. (doi:10.1016/S0031-0182(97)00113-2)
 64. Halanych KM, Mahon AR. 2018 Challenging dogma concerning biogeographic patterns of Antarctica and the Southern Ocean. *Annu. Rev. Ecol. Evol. Syst.* **49**, 355–378. (doi:10.1146/annurev-ecolsys-121415-032139)