# Runs of homozygosity in killer whale genomes provide a global record of demographic histories

Andrew D. Foote[1,2,3*#], Rebecca Hooper[4#], Alana Alexander[5], Robin W. Baird[6], Charles Scott Baker[7,8], Lisa Ballance[7,9], Jay Barlow[9], Andrew Brownlow[10], Tim Collins[11], Rochelle Constantine[8], Luciano Dalla Rosa[12], Nicholas J. Davison[10], John W. Durban[7,9], Ruth Esteban[13], Laurent Excoffier[3], Sarah L. Fordyce Martin[1], Karin A. Forney[14,15], Tim Gerrodette[9], M. Thomas P. Gilbert[1,16], Christophe Guinet[17], M. Bradley Hanson[18], Songhai Li[19], Michael D. Martin[1], Kelly M. Robertson[9], Filipa I. P. Samarra[20], Renaud de Stephanis[13], Sara B. Tavares[21,22], Paul Tixier[17,23], John A. Totterdell[24], Paul Wade[25], Jochen B.W. Wolf[26], Guangyi Fan[27,28,29], Yaolei Zhang[27,30], and Phillip A. Morin[9*]

[1]Department of Natural History, NTNU University Museum, Norwegian University of Science and Technology (NTNU), Erling Skakkes gate 47A, 7012 Trondheim, Norway
[2]Molecular Ecology and Fisheries Genetics Laboratory, School of Biological Sciences, Bangor University, Bangor, Gwynedd, LL57 2UW, UK
[3]CMPG, Institute of Ecology and Evolution, University of Bern, Baltzerstrasse 6, 3012 Bern, Switzerland
[4]University of Exeter, Penryn Campus, Penryn, Cornwall, TR10 9FE, UK
[5]Department of Anatomy, School of Biomedical Sciences, University of Otago, Dunedin, New Zealand
[6]Cascadia Research, Olympia, Washington, USA
[7]Marine Mammal Institute, Oregon State University, Newport, Oregon, USA
[8]School of Biological Sciences, University of Auckland, Auckland, New Zealand
[9]Marine Mammal and Turtle Division, Southwest Fisheries Science Center, National Marine Fisheries Service, National Oceanographic and Atmospheric Administration, 8901 La Jolla Shores Drive, La Jolla, California 92037, USA
[10]Scottish Marine Animal Stranding Scheme, Institute of Biodiversity, Animal Health & Comparative Medicine, College of Medical, Veterinary & Life Sciences, University of Glasgow, Glasgow, G12 8QQ, UK
[11]Ocean Giants Program, Wildlife Conservation Society, New York City, New York
[12]Laboratório de Ecologia e Conservação da Megafauna Marinha, Instituto de Oceanografia, Universidade Federal do Rio Grande, Rio Grande, Brazil
[13]CIRCE, Conservation, Information and Research on Cetaceans, Algeciras, Spain
[14]Marine Mammal and Turtle Division, Southwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Moss Landing, CA 95039 USA
[15]Moss Landing Marine Laboratories, San Jose State University, Moss Landing, CA 95039 USA
[16]Section for Evolutionary Genomics, The GLOBE Institute, University of Copenhagen, Øster Farimagsgade 5A, DK-1353 Copenhagen, Denmark
[17]Centre d'Etudes Biologiques de Chizé (CEBC), UMR 7372 La Rochelle Université – CNRS, 79360 Villiers-en-Bois, France
[18]National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Northwest Fisheries Science Center, Seattle, Washington, USA
[19]Marine Mammal and Marine Bioacoustics Laboratory, Institute of Deep-Sea Science and Engineering, Chinese Academy of Science, Sanya, China
[20]University of Iceland's Institute of Research Centres, Ægisgata 2, 900 Vestmannaeyjar, Iceland

[21]Scottish Oceans Institute, East Sands, University of St. Andrews, St. Andrews, UK
[22]Cetacean Research Program, Pacific Biological Station, Fisheries and Oceans Canada, Nanaimo, Canada
[23]MARBEC Université de Montpellier-CNRS-IFREMER-IRD, Avenue Jean Monnet CS 30171, 34203 Sète, France
[24]Cetacean Research Centre (CETREC WA), PO Box 1029 Esperance, 6450 Western Australia
[25]National Marine Mammal Laboratory, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Alaska Fisheries Science Center, Seattle, Washington, USA
[26]Section of Evolutionary Biology, Department of Biology II, Ludwig Maximilian University of Munich, Großhaderner Strasse 2, Planegg-Martinsried 82152, Germany
[27]BGI-Qingdao, BGI-Shenzhen, Qingdao 266555, China
[28]BGI-Shenzhen, Shenzhen, 518083, China
[29]China National GeneBank, BGI-Shenzhen, Shenzhen, 518120, China
[30]Translational Immunology group, Department of Biotechnology and Biomedicine, Technical University of Denmark, Lyngby, Denmark

*email: Andrew.Foote@ntnu.no, Phillip.Morin@noaa.gov
#These authors contributed equally to this manuscript

**ORCID IDs**
Andrew D. Foote: 0000-0001-7384-1634
Paul Tixier: 0000-0002-7325-3573
John A. Totterdell: 0000-0003-4857-3134
Nicholas Davison: 0000-0001-7286-1836
Rochelle Constantine: 0000-0003-3260-539X
Michael D. Martin: 0000-0002-2010-5139
Kelly M. Robertson: 0000-0002-0090-8604
Sara B. Tavares: 0000-0001-7216-6913
Phillip A. Morin: 0000-0002-3279-1519

**Abstract**

Runs of homozygosity (ROH) occur when offspring inherit haplotypes that are identical by descent from each parent. Length distributions of ROH are informative about population history; specifically the probability of inbreeding mediated by mating system and/or population demography. Here, we investigate whether variation in killer whale (*Orcinus orca*) demographic history is reflected in genome-wide heterozygosity and ROH length distributions, using a global dataset of 26 genomes representative of geographic and ecotypic variation in this species, and two F1 admixed individuals with Pacific-Atlantic parentage. We first reconstruct demographic history for each population as changes in effective population size through time using the pairwise sequential Markovian coalescent (PSMC) method. We find a subset of Pacific and Atlantic populations show evidence of decline in effective population size and increasing coalescence rates towards the end of the Pleistocene, an earlier shared decline in effective population size in the Antarctic ecotypes, and more stable demography in a larger geographically widespread set of genomes. These demographic histories were reflected in present day estimates of individual heterozygosity within each genome (estimated as theta). However, runs of homozygosity within each genome provided a more nuanced inference of the impact of ancestral effective population size versus recent inbreeding, due to the length of ROH being an indicator of the time to most recent common ancestor of identical by descent haplotypes. Genomes sampled at high latitudes, inferred to have undergone ancestral declines in effective population size, were autozygous at hundreds of short ROH (<1Mb), reflecting high background relatedness due to coalescence of haplotypes deep within the pedigree. We discuss how these ROH may have less of an influence upon fitness (than longer ROH) due to deleterious homozygotes within these ROH having been exposed to hundreds of generations of selection. In contrast, longer and therefore younger ROH (>1.5 Mb) were found in low latitude populations and populations of known conservation concern. The outlier in this dataset, a Scottish killer whale, had over one-third of the autosomes (37.8%) in ROH >1.5 Mb in length, with a distinct distribution of ROH lengths, indicating generations of inbreeding. The fate of this Scottish killer whale population, in which only two adult males have been sighted in the past five years, and no calves have been born into over the last two decades, may be inextricably linked to its demographic history and consequential inbreeding depression.

## 1 | INTRODUCTION

Species ranges change in response to environmental oscillation (Lorenzen et al., 2011; Foote et al., 2013), and rapid shifts can currently be observed during ongoing global warming (Parmesan & Yohe 2003; Chen et al., 2011). Understanding how range shifts influence the genetic diversity of natural populations at the range edge is an increasingly important conservation consideration (Excoffier, Foll, and Petit, 2009; see North, McGaughran and Jiggins this issue). The changes in genetic diversity experienced by populations during the climatic change and warming following the Last Glacial Maximum (LGM) offer a window into potential future changes due to ongoing global warming (Lorenzen et al., 2011; Foote et al., 2013; Louis et al. 2020; Nykanen et al. 2019).

Whilst summary statistics of genetic diversity can provide a useful metric of the impact of demography, it can be difficult to attribute these metrics to specific past demographic changes or events (although site frequency spectra contain much information; Gutenkunst et al. 2009; Excoffier et al. 2013; Sousa & Hey, 2013). The increasing use of whole genome sequencing (WGS) in molecular ecology, as showcased in this special issue of *Molecular Ecology*, is providing unprecedented resolution of population demographic and evolutionary history in an increasingly broad range of non-model organisms (e.g. Sarabia et al. this issue). This is due to the increased accuracy in the inference of timing of demographic events being drawn from the contiguity, distribution and linkage between loci afforded by WGS datasets (Veeramah & Hammer 2014; Leitwein et al. 2020). The coalescence times of identical by descent haplotypes inherited from both parents, which form runs of homozygosity (ROH), can provide such insights into past demography and more recent inbreeding events (Kardos et al. 2016; Brüniche-Olsen et al. 2018; Ceballos et al. 2018b). ROH can be broken up by subsequent generations of outbreeding, thus the distribution of lengths of ROH across the genome reflects local recombination rate, demographic history (contemporary and ancestral effective population size and thus the rate of coalescence through time), migration rate, and mating behaviour (Kardos et al. 2016; Ceballos et al. 2018b).

In this study, we showcase how patterns of coalescence in individual representative genomes sampled from multiple populations, can be used to reconstruct global variation in demographic histories among killer whale (*Orcinus orca*) populations. Killer whales are comparable to humans in their global distribution, having colonised all the major oceans (Forney & Wade 2006). Killer

whale occurrence is correlated with ocean productivity; highest densities are therefore at high latitudes, decreasing by 1-2 orders of magnitude from the Arctic and Antarctic to the tropics (Figure S1; Forney & Wade 2006; Morin et al. 2015). The northern and southern extremes of this range were covered by ice sheets during the Last Glacial Maximum (Figure S1; Morin et al. 2015) and must therefore have been colonised through range expansion from lower latitudes in the last 16 Kyr (Hoelzel et al. 2007; Morin et al. 2015; Foote et al. 2019). Post-glacial range expansions in different ocean basins offer the possibility to explore the genetic outcomes of independent, parallel demographic histories. Killer whales thus represent a useful system for studying the relationship between demographic history and genetic diversity.

## 2 | MATERIALS AND METHODS

### 2.1 | Data set

Our primary genomic dataset consisted of 26 genomes from samples selected to best represent the known global geographic and genetic diversity of this species (Figure 1; Foote et al. 2019). In addition to previously published, low-coverage whole genome sequence data, new paired-end libraries were built from 20 of the samples using the BEST library build method (Carøe et al. 2018) from genomic DNA that was sheared to an average size of 350-550 bp using a Covaris ME220. Libraries were subsequently dual indexed and amplified for 12 cycles, then purified using ampure beads (Agilent Technologies, CA, USA). The DNA concentration of the libraries was measured using an Agilent TapeStation (Agilent Technologies, CA, USA); these were then equimolarly pooled and sequenced across eight lanes of an Illumina HiSeq4000 platform. An additional two genomes from captive killer whales with mixed parentage of Icelandic mothers and a North Pacific *transient* father were generated as above, with the following differences in protocols. Genomic DNA was sheared to an average size of ~500 bp using a Diagenode Bioruptor Pico sonication device. The sheared DNA extracts were converted to blunt-end Illumina sequencing libraries as above. Libraries were subsequently dual indexed and amplified for 12 cycles using a KAPA HiFi HotStart PCR kit (Kapa Biosystems, Wilmington, Ma. USA) in 50 µl reactions following the manufacturer's guidelines and sequenced on one lane of an Illumina HiSeq4000 platform.

Reads from each individual were processed with AdapterRemoval2 (Schubert, Lindgreen, & Orlando, 2016) to trim residual adapter sequence contamination and to remove adapter dimer sequences as well as low-quality (Q<30) stretches at the read ends. Filtered reads >30 bp were then mapped using BWA-MEM algorithm to a high-quality, chromosomal *O. orca* reference assembly (Zhang et al. in review), requiring a mapping quality greater than 30. Clonal reads were collapsed using the markdup function of the SAMtools (Li et al. 2009). Repetitive elements were identified using RepeatMasker (Smit et al. 1996) and the Cetartiodactyl repeat library from Repbase (Kohany et al. 2006), and masked using BEDtools (Quinlan & Hall 2010). Sites were further filtered to include only autosomes.


**Figure 1: Map of sampling locations of the genomes included in this study.**
Note that sample names used in this study do not necessarily denote discrete populations or geographic/political entities: e.g. Norway and Iceland samples appear to belong to the same metapopulation; Chatham Islands is a part of New Zealand but different names are used to distinguish between the two samples labelled New Zealand and Chatham Islands, which have distinct ancestry (Foote et al. 2019).

## 2.2 | Estimating theta (θ) from individual genomes

To understand the impact of ancestral demography on contemporary killer whale populations, we estimated variation in heterozygosity among our global dataset of individuals. We estimated

heterozygosity of autosomal regions from individual diploid genomes corrected for coverage. Under the infinite sites model (Kimura 1969), individual heterozygosity is a good, unbiased estimator of the population mutation rate, theta (θ; Watterson 1975). Theta was estimated directly from the filtered bam files using the maximum likelihood estimator (Lynch, 2008) implemented in mlRho v.2.9. (Haubold et al. 2010).

### 2.3 | Estimating changes in $N_e$ through time from individual genomes

Changes in effective population size through time were inferred using the pairwise sequential Markovian coalescent (PSMC) (Li & Durbin 2011), a method suitable for studying ancient demographic history from a single unphased genome per population. The PSMC model estimates the Time to Most Recent Common Ancestor (TMRCA) of segmental blocks of the genome and uses information from the rates of the coalescent events to infer $N_e$ at a given time, thereby providing a direct estimate of the past demographic changes of a population (Li & Durbin 2011). We selected the highest coverage genomes, building a consensus sequence of each bam file in fastq format sequentially using: first, SAMtools mpileup command with the –C50 option to reduce the effect of reads with excessive mismatches; second, bcftools view –c to call variants; lastly, vcfutils.pl vcf2fq to convert the vcf file of called variants to fastq format with further filtering to remove sites with less than a third or more than double the average depth of coverage and Phred quality scores less than 30. The PSMC inference was then carried out using the recommended input parameters for human autosomal data, i.e. 25 iterations, with maximum TMRCA ($T$max) = 15, number of atomic time intervals ($n$) = 64 (following the pattern (1*4 + 25*2 + 1*4 + 1*6), and initial theta ratio (r) =5. Plots were scaled to real time as per Foote et al. (2016) assuming a generation time of 25.7 years and a neutral autosomal mutation rate of $2.34 \times 10^{-8}$ substitutions per nucleotide per generation. For depth of coverage less than 20×, negative false detection of heterozygote sites can produce the same effect as a lower mutation rate would have on the plot, causing shifts both along the $x$- and $y$-axes (for example see supplementary figure 10 of Foote et al. 2016). Based on the overlap in our plots, and the uncertainty on which false negative detection rate to apply, we did not apply the uniform correction factor option in PSMC to account for small variations in coverage between our genomes. However, to better understand the impact of coverage on PSMC plot trajectory, we extended the tests run in Foote et al. (2016) by estimating demographic parameters using PSMC on a proportionally (0.1–0.9 in 0.1 increments) randomly

down-sampled genome.

## 2.4 | Calling runs of homozygosity

Runs of homozygous genotypes were identified using the window-based approach implemented in PLINK v1.07 (Purcell et al. 2007). Genotypes were called concurrent to estimating genotype likelihoods for all samples simultaneously, both using ANGSD (Korneliussen et al. 2014) estimating the posterior genotype probability based on the allele frequency as a prior, inferring the major and minor allele from genotype likelihoods, and only outputting SNPs at sites that were identified as polymorphic with likelihood ratio test p-values of <0.000001 and with a minimum minor allele frequency of 0.05. This resulted in two output files, one of called genotypes in T-ped format to be used as input for PLINK, the second of genotype likelihoods in BEAGLE format (Browning & Browning 2007), e.g., likelihoods of the three possible genotypes: AA, Aa and aa, for each inferred biallelic SNP in each individual and normalised so that the likelihood for all three possible genotypes summed to one. The called genotypes in T-ped format and normalised genotype likelihoods in Beagle format were compared to assess confidence in the PLINK input file. Additional filtering steps used during the calling of SNPs were the removal of reads of poor mapping quality (MAPQ < 30), removing sites with low base quality scores (q < 20), discarding reads that did not map uniquely, adjusting q-scores around indels, adjusting minimum quality score to 50 for excessive mismatches, and discarding bad reads (flag >=256). Inferred SNPs were lightly pruned based on linkage disequilibrium (LD) $r^2$ >0.9 using PLINK, which has been found to improve the accuracy of detecting autozygous ROHs (Howrigan et al. 2011). We estimated ROH from pruned and unpruned data and found minimal qualitative difference with our data, and therefore proceeded with unpruned data. Sliding window size was set to 300 kb, with a minimum of 50 SNPs at a minimum density of 1 SNP per 50 kb required to call a ROH. To account for genotyping errors, we allowed up to 3 heterozygote sites per 300 kb window within called ROHs, as per Ceballos et al. (2018a). A length of 1,000 kb between two SNPs was required in order for them to be considered in two different ROHs. Our command line was as follows: ./plink --tfile chr1 --homozyg-snp 50 --homozyg-kb 300 --homozyg-density 50 --homozyg-gap 1000 --homozyg-window-snp 50 --homozyg-window-het 3 --homozyg-window-missing 10 --homozyg-window-threshold 0.05 --out chr1

## 3 | RESULTS

### 3.1 | Genome sequences

Previously published short-read data sequenced to approximately 5× effective coverage (depth of sequencing post QC filtering) of the autosomal region of the killer whale genome (Foote et al., 2019) were combined with published data from a Norwegian (Foote et al. 2015) and Pacific resident killer whale (Moura et al. 2014) and an additional 8 lanes of a HiSeq4000 sequencing to increase samples to 10-20× effective coverage (*i.e.* post-QC filtering). Additionally, two F1 mixed ancestry captive killer whales with Icelandic and Pacific transient parentage were also sequenced across one lane of a HiSeq4000 to a mean of 5× coverage. In total, the dataset comprised 23 genomes sequenced to between 10-20× mean depth of coverage, and five samples sequenced to approximately 5× mean depth of coverage. All depth of coverage estimates are based on bam files for which duplicate reads, low quality mapping regions and repeats have been masked and all filtering steps completed. Distribution of effective depth of coverage for each sample is shown in Figure S2, and distribution of base quality scores for all filtered data is shown in Figure S3. All raw sequencing data used in this study are publicly archived under the Orcanomics BioProject (NCBI accession: PRJNA531206), with the exception of the reads of the Norwegian and Alaskan resident killer whales, which are archived under NCBI BioProjects PRJNA167475 and PRJNA236163 respectively.

### 3.2 | Impact of variation in depth of coverage

Due to the variation in coverage among samples, we down-sampled from a high-coverage genome assembly (Foote et al. 2015) to investigate the impact upon our estimates of θ, inference of $N_e$ through time using PSMC, and our calls of ROH (Figure S4). We found that estimates of θ by mlRho showed a moderate non-linear decline with decreasing depth of coverage (Figure S4c). Our inferred PSMC trajectories of a down-sampled Norwegian killer whale genome were qualitatively similar down to half of the reads being subsampled (corresponding to approximately a mean depth of coverage of 10×), but at lower rates of subsampling, plots showed stronger deviation along both the *x*- and *y*-axes (Figure S4d). Our estimates of the number and sum of ROH > 0.3 Mb in length were relatively consistent down to 0.3× the total number of reads being down-sampled (Figure S4e). Therefore, we report theta estimates and present PSMC plots only for samples with effective depth of coverage >10×, but report ROH estimates for all samples.

### 3.3 | Estimates of theta (θ) from individual genomes

Counterintuitive to the conventional view of genetic diversity, the lowest estimates of θ were found in high latitude populations, *i.e*. in the regions with the greatest habitat suitability and highest density of present-day killer whales, *e.g*. Alaskan resident ecotype, Antarctic types B1, B2 and C, Norway and Scotland (Figure S1; Table S1). However, some of these regions were not suitable habitat for killer whales during the Last Glacial Maximum (Figure S1).

### 3.4 | Pairwise sequential Markovian coalescent inference of ancient (>10KYA) demographic histories

We found that the PSMC plots broadly followed one of three trajectories of changes in effective population size through time (Figure 2). Each of these trajectories was shared among a subset of several individuals. The segregation of genomes based upon PSMC plot trajectory approximately corresponds to the segregation along the major axis of structure detected in PCA and individual assignment (NGSadmix) analyses (Foote et al. 2019). The genomes for which PSMC inferred a decline in effective population size during the last glacial period were intuitively the same as those with markedly lower θ (= $4N_e\mu$) estimated by mlRho (Table S1).

Qualitatively similar demographic histories previously inferred from the Norwegian and North Pacific resident ecotype genomes using PSMC (Moura et al. 2014; Foote et al. 2016) were also shared with a geographically widespread set of Pacific and Atlantic samples (Figure 2A). Foote et al. (2019) had previously found the Pacific resident killer whale shared an excess of shared derived alleles and more recent common ancestor (based on cross coalescence between phased X-chromosome haplotypes) with these same samples relative to degree of shared ancestry with the sympatric Pacific *transient* ecotype. This is yet more evidence of secondary contact between North Pacific lineages, rather than divergence from a common ancestor in sympatry (Foote and Morin, 2016). The point at which PSMC plots become identical can be an indication of the point in time at which alleles in each genome coalesce within a shared ancestral population (Mather et al. 2020). However, as the rate of cross coalescence is not tested in PSMC, this could also reflect similar but independent processes in each individual's demographic histories, and therefore similar rates of coalescence in each individual's genome.

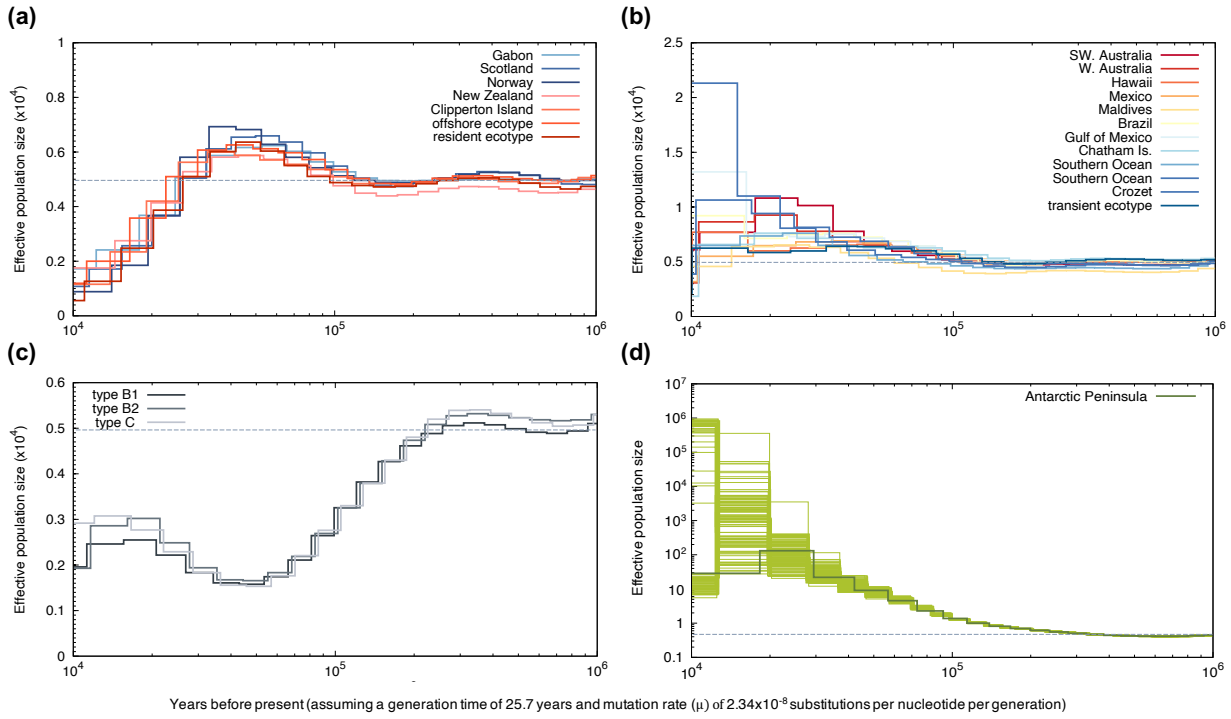



**Figure 2.** Pairwise sequentially Markovian coalescent (PSMC) plots of changes in effective population size through time inferred from coalescence rates across individual genomes. (a) PSMC plots for the genomes of Atlantic (blue) and Pacific (red) killer whales, which show a similar trajectory. (b) PSMC plots for the genomes of a geographically widespread subset of killer whales, which show a similar stable trajectory until approximately 30 KYA. (c) PSMC plots for one genome each of an Antarctic type B1, Antarctic type B2 and Antarctic type C killer whale, which show a similar trajectory until the Last Glacial Maximum (LGM). (d) PSMC plot inferred from the genome of an Antarctic type A killer whale sampled off the Antarctic Peninsula. The thick line represents the median. Due to the unusually high estimate of effective population size in plot (d), we chose to visualise uncertainty around the median, and thin lines correspond to 100 rounds of bootstrapping. Note that the *y*-axis has a logarithmic scale in plot (d). Each plot has a horizontal dashed line drawn at $0.5 \times 10^4$ to facilitate visual comparison between the different plots.

The second pattern observed in the PSMC plots (Figure 2b) encompassed a wide geographically distributed set of samples, which share much of the trajectory in $N_e$ through time with the plots in Figure 2A, but remain stable or increase at ~40,000 years BP. The plots then separate and show distinct trajectories roughly 10,000 years either side of the LGM. Note that the time calibration in all of these PSMC plots is dependent upon the accuracy of the mutation rate and generation time estimates, so should not be interpreted too literally, but rather as a broad indication of the temporal scale of events.

The third pattern of inferred changes in effective population size through time was shared among the Antarctic morphotypes: types B1, B2 and C (Figure 2c). We infer a longer-term low effective population size in the genealogical ancestors of these three genomes compared to the apparent baseline effective population size ($N_e \approx 0.5 \times 10^4$) inferred in most other genomes included in this study. A shared demographic history which includes extensive genetic drift has previously been inferred from the high covariance in allele frequencies among these three Antarctic types, and their strong genetic differentiation from all other killer whale populations (Foote et al. 2016, 2019). The timing of inferred changes in effective population size (Figure 2c) correspond closely to timing of cross coalescence due to admixture and divergence visualised in Figure 3c & d of Foote et al. (2019).

The genome from a 'type A' killer whale sampled off the tip of Antarctic Peninsula had a unique demographic history inferred by PSMC. Inferred effective population size increased exponentially by two orders of magnitude approximately 30,000 years ago, before declining again 20,000 years ago. The sudden inferred increase in the plot resembles PSMC plots of unphased pseudo-diploid genomes generated by combining two randomly sampled haploid genomes from two different individuals, in which the upsweep can be coarsely indicative of the timing of cessation of cross coalescence after populations diverge (Cahill et al. 2016). Similar PSMC patterns are reported by Sarabia et al. in this special issue, for an Egyptian African golden wolf (*Canis lupaster*), with apparent introgressed ancestry (Sarabia et al. 2021). QC checks performed when combining the sequencing data generated in 2016 and 2021 for each individual included randomly selecting an allele from each site of the genome for filtered bam files corresponding to the 2016 and 2021 data, and then estimating covariance between the 2016 and 2021 sequencing data for the same individual (to test for batch effects). We found high covariance between alleles randomly drawn from the 2016 and 2021 sequence data for all other individuals. However, the covariance of randomly drawn alleles from the 2016 and 2021 batches of data for the Antarctic Peninsula sample was similar to random draws from different individuals. This could be consistent with either contamination of one of the sequence runs for this sample, or unusually high genetic diversity in this individual. We further checked for signs of contamination of sequencing data for this individual by comparing the haploid mitochondrial genome sequence, both between 2016 and 2021 data, and with earlier

published targeted sequencing the of the mitogenome (Morin et al. 2010, 2015). The mitogenome sequence was haploid and consistent across sequencing datasets, suggesting this high genetic diversity in the nuclear genome was not an artefact of contamination during sequencing. We therefore interpret the PSMC plot of the Antarctic Peninsula killer whale as evidence that this individual has admixed ancestry comprised of two distinct genetic backgrounds, that have remained mostly isolated from one another for approximately 1,000-2,000 generations.

### 3.4 | Calling Runs of Homozygosity

In total, 3,960,271 SNPs were called from the repeat-masked, QC filtered autosomes. The mean inverse SNP density (Kb per SNP) was 0.642 (range:0.204 to 4.906; Figure S5), *i.e.* on average there was a physical distance of 642 bp between SNPs. The mean number of SNPs per ROH was 1,310 (range:69 to 55,309; Figure S5). As we only consider ROH of length ≥0.3Mb, a called ROH would, based on the average SNP densities, require calling homozygous genotypes at approximately >467 consecutive SNPs. We found that increasing the number of missing windows (--homozyg-window-missing) increased the number of ROH called in the lower coverage (<10×) genomes, which plateaued at 30 missing sites per 300,000 bp window (*i.e.* allowing windows to be called as ROH when there is ≤0.01% missing data). There was no difference in the ROH of higher coverage (>10×) genomes above 10 missing sites per window.

### 3.4 | ROH density

ROH density along chromosome 1 highlights the overlap in ROH among the genomes (Figure 3). ROH in up to 17 genomes overlapped at some 100-kb windows along chromosome 1, with a mean ROH density of 8.3 overlapping ROH per 100-kb window (when considering only windows with at least one ROH). In particular, ROH overlapped in those genomes that showed a similar decline in $N_e$ during the last glacial period as represented in Figure 2a. However, this does not necessarily signify shared IBD tracts due to recent shared ancestry among genomes. When we estimated pairwise genetic distance based upon IBS (identical by state) at sites within ROH in the Norwegian killer whale genome across all genomes, only the closely related Icelandic killer whale clustered by IBS with the Norwegian killer whale, and the two likely share identical ROH tracts through descent from a common ancestor (Figure S6). Variation in ROH density therefore likely reflects variation in recombination rate along the genome (Kardos et al. 2017).

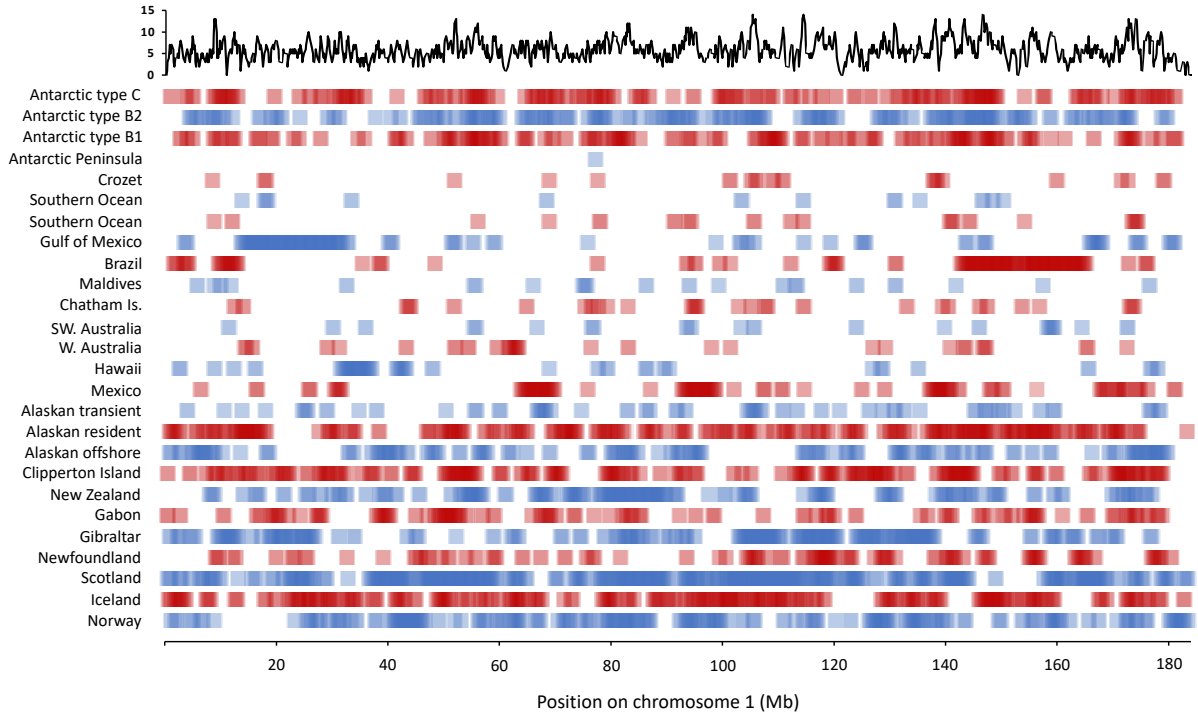



**Figure 3.** Physical mapping of ROH density along chromosome 1 of all ROH > 0.3Mb in 26 globally sampled killer whale genomes. The top panel shows counts of overlapping ROH in non-overlapping 100-kb windows for all 26 genomes. The bottom panel presents ROH density per individual genome. Each 100-kb window containing a ROH is shaded red and blue in alternating rows. Genomes are ordered so that those in different panels in Figure 2 are grouped together, e.g. the first three rows are Antarctic types for which a long-term low effective population size was inferred; row 4 is the admixed Antarctic Peninsula genome, with just a single ROH in chr1, etc.

### 3.5 | Sum and number of ROH within individual genomes

The sum and number of ROH (SROH and NROH respectively) within individual genomes were intuitively highly correlated (Spearman correlation $\rho = 0.91$, $p = 1.3 \times 10^{-6}$; Figure 4). The SROH and NROH are expected to reflect population demography (Ceballos et al. 2018b). Large populations are expected to have low SROH and NROH, which will be further reduced by admixture, in particular admixture between highly diverged lineages carrying different haplotypes will reduce the number and sum of ROH in the genome (Ceballos et al. 2018b). In contrast, populations that have experienced population bottlenecks will have higher SROH and NROH (Ceballos et al. 2018b). Qualitatively similar ROH length distributions were observed among individuals from high latitude populations in different ocean basins, which were inferred using PSMC to have undergone through a decline in $N_e$ during the last glacial period (Figure 2a & c;

Figure S7). In this subset of genomes, autozygosity was spread across hundreds of short (less than 1Mb in length) ROH in high latitude populations from the North Atlantic (Iceland and Norway), Antarctic (types B1, B2 and C) and North Pacific (resident ecotype) (Figures 3, 4, 5a, S7).

Deviation from the regression line can be informative (assuming equivalent recombination rates) as to whether demographic history (contemporary and ancestral effective population size), or mating behaviour, or a combination, are factors in determining the length distribution of ROH across the genome (Ceballos et al. 2018b). For example, Antarctic types B1, B2 and C have an excess of smaller ROH, suggesting a high rate of out-breeding relative to populations along the regression line. The genome of a female from the 'west coast community' of killer whales of Scotland was a marked outlier below the regression line, indicating an excess of long ROH relative to populations along the regression line; consistent with consanguineous mating and long-term low effective population size.

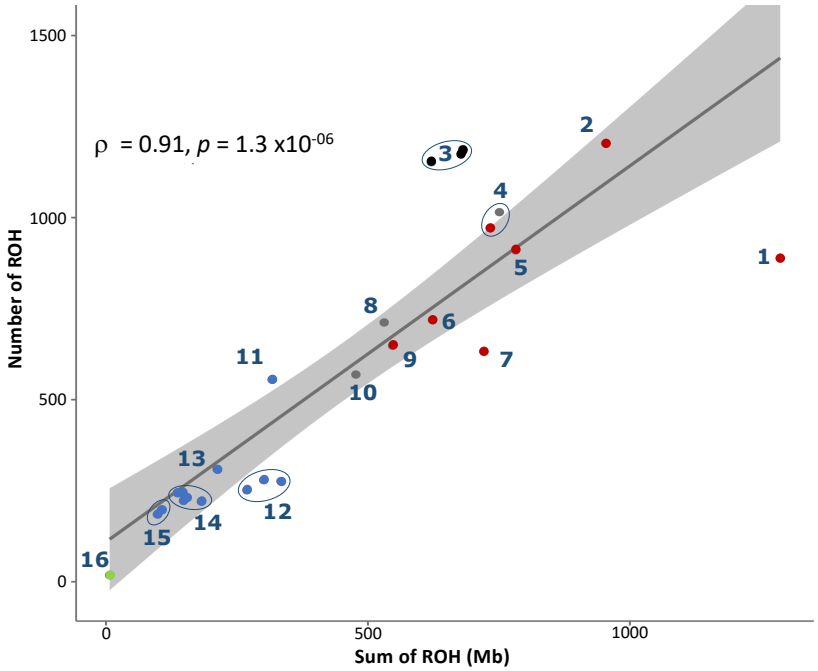
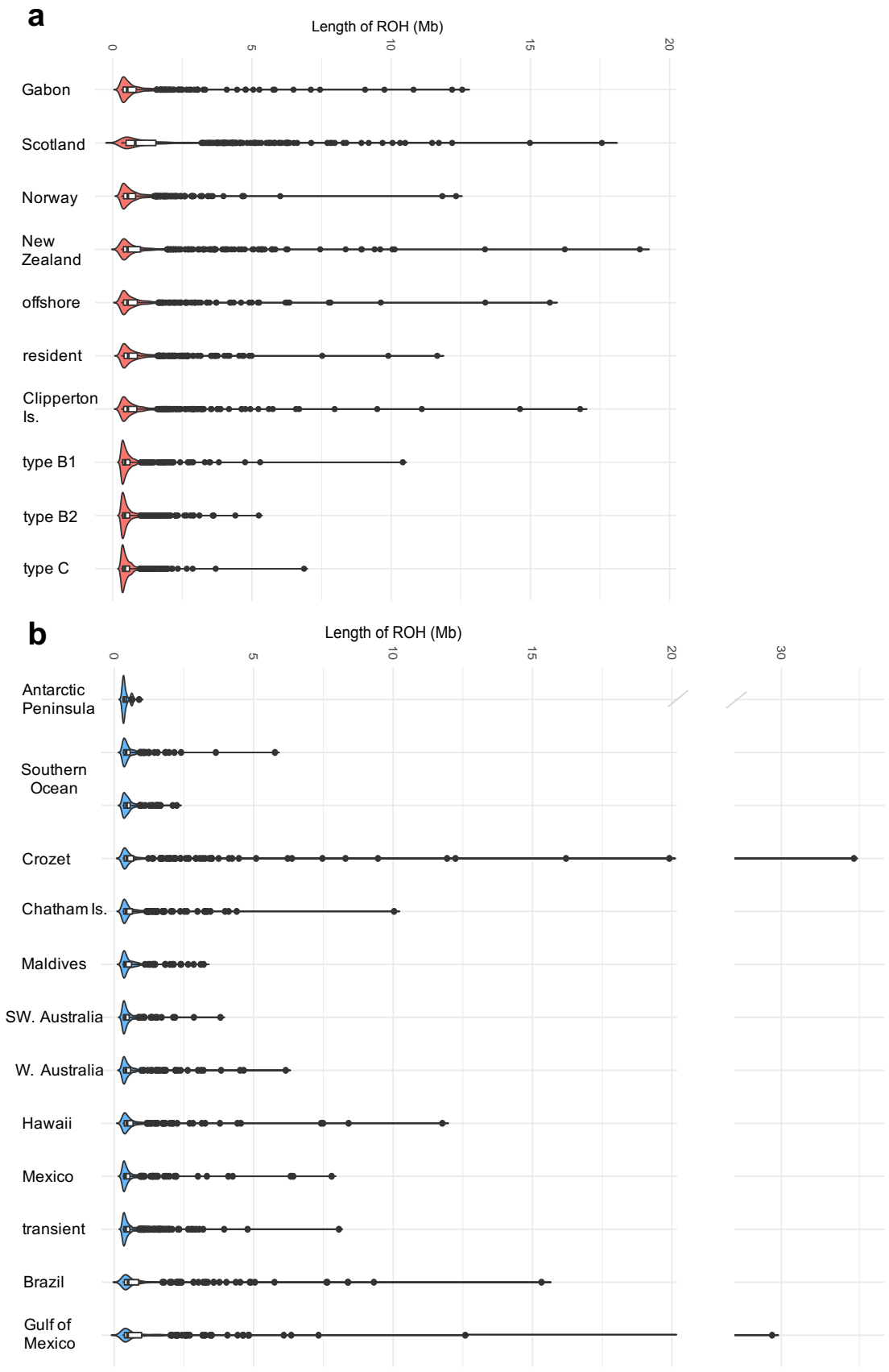



**Figure 4.** Number of ROH compared to the sum of the length of ROH across the autosomes. Data points are coloured by PSMC trajectory; those genomes included in Figure 2a are coloured red, those in Figure 2b are coloured blue, those in Figure 2c are coloured black, and the admixed Antarctic Peninsula type A individual in Figure 2d is coloured green. The lower coverage genomes not included in the PSMC analyses are coloured grey. Numbered data points indicate genomes from the following populations/locations: **1.** Scotland, **2.** Alaskan resident ecotype, **3.** Antarctic types B1, B2 and C, **4.** Norway and Iceland, **5.** Eastern Tropical Pacific (ETP) - Clipperton Island, **6.** Alaskan offshore ecotype, **7.** New Zealand, **8.** Gibraltar, **9.** Gabon, **10.** Newfoundland, **11.** Alaskan transient ecotype, **12.** Crozet, Brazil and Gulf of Mexico, **13.** Chatham Islands, **14.** Hawaii, ETP – Mexico, W. Australia, Southern Ocean and Maldives, **15.** SW. Australia and Southern Ocean, **16.** Antarctic Peninsula.


a
Length of ROH (Mb)
0
5
10
15
20
Gabon
Scotland
Norway
New Zealand
offshore
resident
Clipperton Is.
type B1
type B2
type C
b
Length of ROH (Mb)
0
5
10
15
20
30
Antarctic Peninsula
Southern Ocean
Crozet
Chatham Is.
Maldives
SW. Australia
W. Australia
Hawaii
Mexico
transient
Brazil
Gulf of Mexico

438

**Figure 5.** Kernel density (violin) plots of the length of individual ROH in each genome included in the PSMC analyses. White rectangle shows the interquartile range, and the black bar the median of the data. Samples are grouped by PSMC trajectory, so that panel **a.** includes genomes inferred by PSMC to have an ancestral decline in effective population size deep in their pedigrees (*i.e.* those from Figure 2a & c); and panel **b.** includes samples with a more stable effective population size, and the putatively admixed Antarctic Peninsula sample (*i.e.* those from Figure 2b & d).

### 3.5 | Inferring approximate TMRCA of ROH of differing length

Since homozygous tract length ($L$) declines as a function of recombination rate ($r$) and time ($t$), ROH due to background inbreeding during an ancestral bottleneck are expected to be shorter than ROH caused by recent inbreeding (Kardos et al. 2017). The expected length of ROH is $L = 100/2t$ cM, where $t$ is the time back to the parental common ancestor in generations (Thompson 2013). So that for example, 0.5 Mb ROH, which we found are common in some high latitude populations, would correspond to two IBD tracts coalescing in a parental common ancestor ≈88 generations ago (assuming a recombination rate equivalent to that in humans, 1.133cM/Mb; Dumont & Payseur, 2008), which equates to 2,270 years ago (assuming a generation time of 25.7 years; Taylor et al. 2007). However, it should be noted that mendelian sampling and the stochasticity of recombination will generate variation around this expected relationship between ROH length, recombination rate and time, increasingly so for older coalescent times (Thompson 2013; Stoffel et al. 2021). Assuming a range of constant mammalian recombination rates (0.649-1.554 cM/Mb) (Dumont and Payseur, 2008), we estimate that ROH less than 1 Mb in length correspond to haplotypes representing coalescence greater than 1,000 years ago, whilst ROH longer than 2Mb are estimated to result from inbreeding during the past millennium (Figure S8). Furthermore, these estimates of the timing of coalescence of ROH should be reconsidered when recombination maps are developed for the killer whale genome assembly, as recombination cold or hotspots could fall outside the recombination rates used in these estimates.

### 3.6 | The effect of admixture on ROH

Genomes sampled at low latitudes, where killer whales are found at low density, were typically characterised as being less autozygous than those sampled at high latitudes (Figures 3, S7).

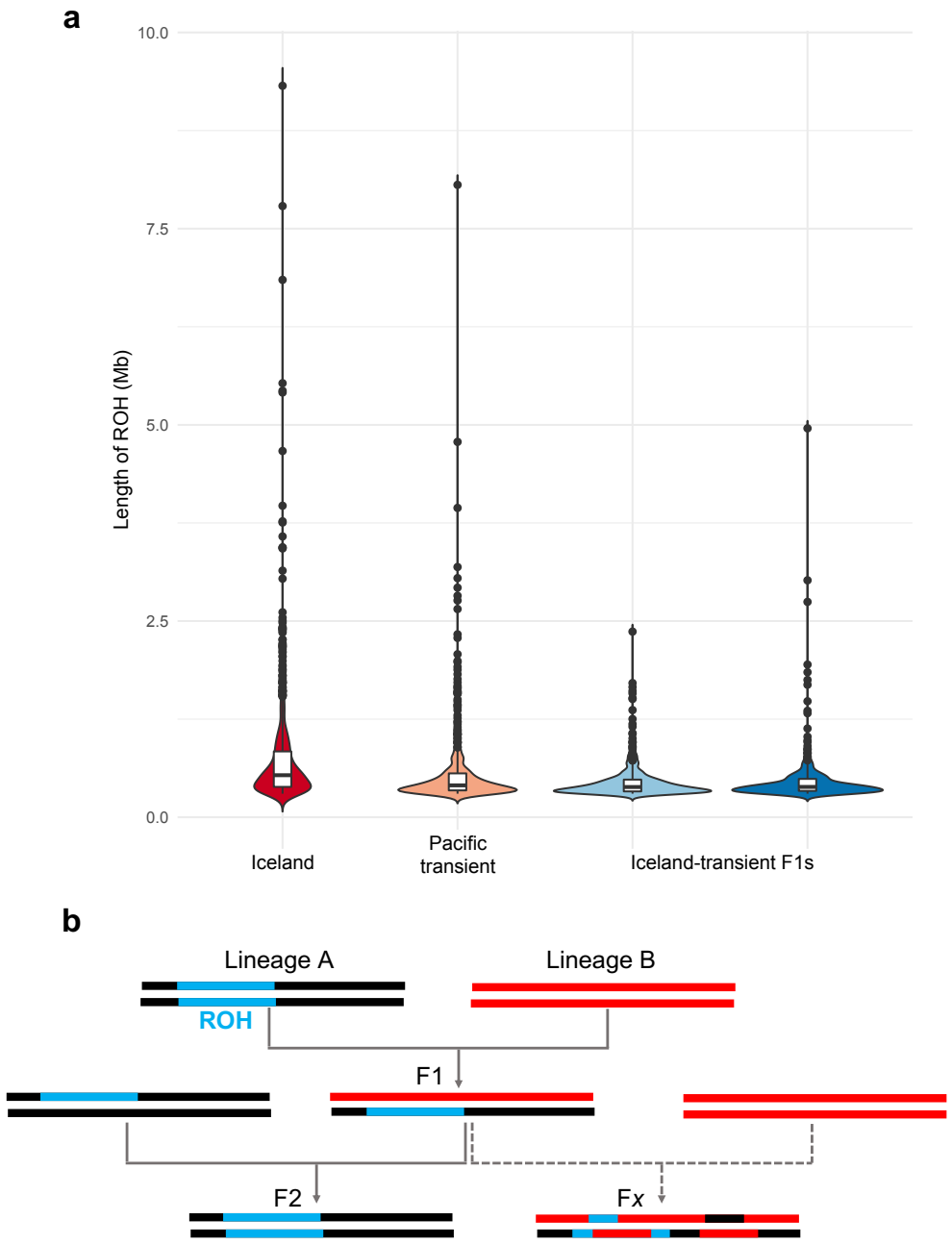



**Figure 6a.** Kernel density (violin) plots of the Atlantic (Icelandic) and Pacific (transient) parental genomes and two F1 Iceland-transient admixed individuals bred in captivity. White rectangle shows the interquartile range, and the black bar the median of the data. **b.** Schematic diagram of a ROH due to two IBD tracts (blue lines) in a diploid genome sampled from lineage A. In the next generation, the tract is in a heterozygous state after recombination with a divergent haplotype due to admixture with lineage B. After a subsequent *x* generations the ROH is either further broken up by recombination if subsequent recombination from generations F1 to F*x* occur within lineage B. If instead the F1 mates within lineage A then the F2 offspring could inherit the blue IBD tract from both parents, re-establishing the same ROH within the F2 generation (see for example Robinson et al. 2019; Hasselgren et al. 2021).

Admixture between lineages with divergent evolutionary histories, and therefore distinct genetic backgrounds, is predicted to break up ROH, through the recombination of different haplotypes (Ceballos et al. 2018b). For example, comparison of patterns of ROH in a global dataset of human genomes found that African Americans had notably shorter mean ROH length than most Europeans or Africans (Pemberton et al. 2012). Our interpretation of the PSMC plot generated from the Antarctic Peninsula type A killer whale genome was that the individual had mixed ancestry due to recent admixture between two divergent lineages that shared a common ancestor approximately 1,000-2,000 generations ago. Accordingly, the genome of this individual had lowest sum and number of ROH in the dataset. We further illustrate this outcome of episodic admixture of different genetic backgrounds by comparing the ROH of F1 offspring from Icelandic and North Pacific *transient* parents, with the ROH of the parental lineages (Figure 6, S7). The two admixed individuals show a reduction in longer (>1 Mb) ROH relative to the parental lineages.

### 3.7 | Individual genomic inbreeding coefficients

Individual genomic inbreeding coefficients were estimated from ROH ($F_{ROH}$) as the percentage of the autosomes that are autozygous and in ROH > 1.5Mb, so as to exclude shorter ROH which result from background relatedness caused by ancestral bottlenecks, and are less likely to reduce individual fitness relative to the population average (*i.e.* cause inbreeding depression). A geographically widespread set of 21 of the 26 samples had estimated inbreeding coefficients of $F_{ROH} > 0.015$ (Table S1). Our finding of widespread inbreeding across the species range is consistent with the close kin structure and high social philopatry observed in most killer whale populations studied to date (Bigg et al. 1990; Barrett-Lennard et al. 2000; Parsons et al. 2009; Ford et al. 2011; Ivkovich et al. 2010; Baird & Whitehead 2000; Pilot et al. 2010; Baird et al. 2006; Beck et al. 2012; Esteban et al. 2016a; Reisinger et al. 2017; Busson et al. 2019).

When considering the correlation between NROH and SROH (Figure 3), a killer whale from a small population studied off the West coast of Scotland was a clear outlier below the regression line, indicating an excess of long ROH which is consistent with consanguineous mating (Ceballos et al. 2018b). We found that 37.8% of the autosomal regions of the Scottish killer whales' genome was comprised of ROH > 1.5 Mb (Table S1). The distinct distribution of ROH lengths in this individual (Figures 3, 4 & S7) suggest long-term inbreeding.

The genome sampled from the Antarctic Peninsula type A killer whale, in addition to the two captive F1 admixed individuals with mixed Atlantic and Pacific killer whale parentage (Table S1), highlight the low $F_{ROH}$ inbreeding coefficients in admixed genomes, where long ROH are broken up by the lack of identical by descent haplotypes shared by the two divergent parental lineages.

### 3.7 | Correlates of runs of homozygosity

Although short ROH (<1Mb) were found in higher abundance in a subset of genomes sampled at high latitude, the relationship between the sum of <1Mb ROH and latitude was not significant correlated (Spearman correlation $\rho = 0.30$, $p = 0.13$). We therefore fitted General Linear Models in base R to test whether latitude or demography (as a categorical variable of either a demographic decline during the last glacial period, or stable demography during that same period) are predictors of the sum of ROH < 1Mb and >2Mb. Whilst latitude did not significant predict the sum of short ROH (<1Mb) per genome ($t = 0.294$, $p = 0.771$), demography was a significant predictor ($t = -3.156$, $p = 0.005$; for a single term model $p = 0.00229$). Neither demography or latitude predicted the sum of long (>2Mb) ROH when compared together in a GLM, but demography was a marginally significant predictor in a single term model ($p = 0.0345$).

## 4 | DISCUSSION

We find that the present-day genetic variation and diversity within and among killer whale genomes sampled across the globe was shaped by ancestral demography, hundreds to thousands of generations ago. Killer whales have radiated globally in a timescale comparable to anatomically modern humans. In this study, we have shown that, like human populations that expanded out of Africa (Li and Durbin 2011; Pemberton et al. 2012; Ceballos et al. 2018b), present day killer whale populations at the edge of the species' range also share the genomic footprints of past demographic history. In a subset of high latitude populations, we found increasing coalescence (inferred as declining $N_e$ by PSMC) through the last glacial period, and many short ROH that we estimate coalesce in a common ancestor deep in the pedigree of these individuals.

Our findings from a global dataset of genomes reinterpret previous inference based upon just two genomes; this earlier study interpreted this sparse data as evidence for a widespread genetic bottleneck across most of the species range during the LGM (Moura et al. 2014). Our results are

instead consistent with the scenario suggested by Hoelzel et al. (2007), in which distinct genetic bottlenecks occurred in the ancestors of present-day high latitude populations, caused by a continuous range expansion or serial founder effect during the colonisation of emerging habitat post-LGM. This interpretation is supported by observations of these same high latitude populations being the most differentiated based on allele frequencies (Morin et al. 2015; Foote et al. 2019).

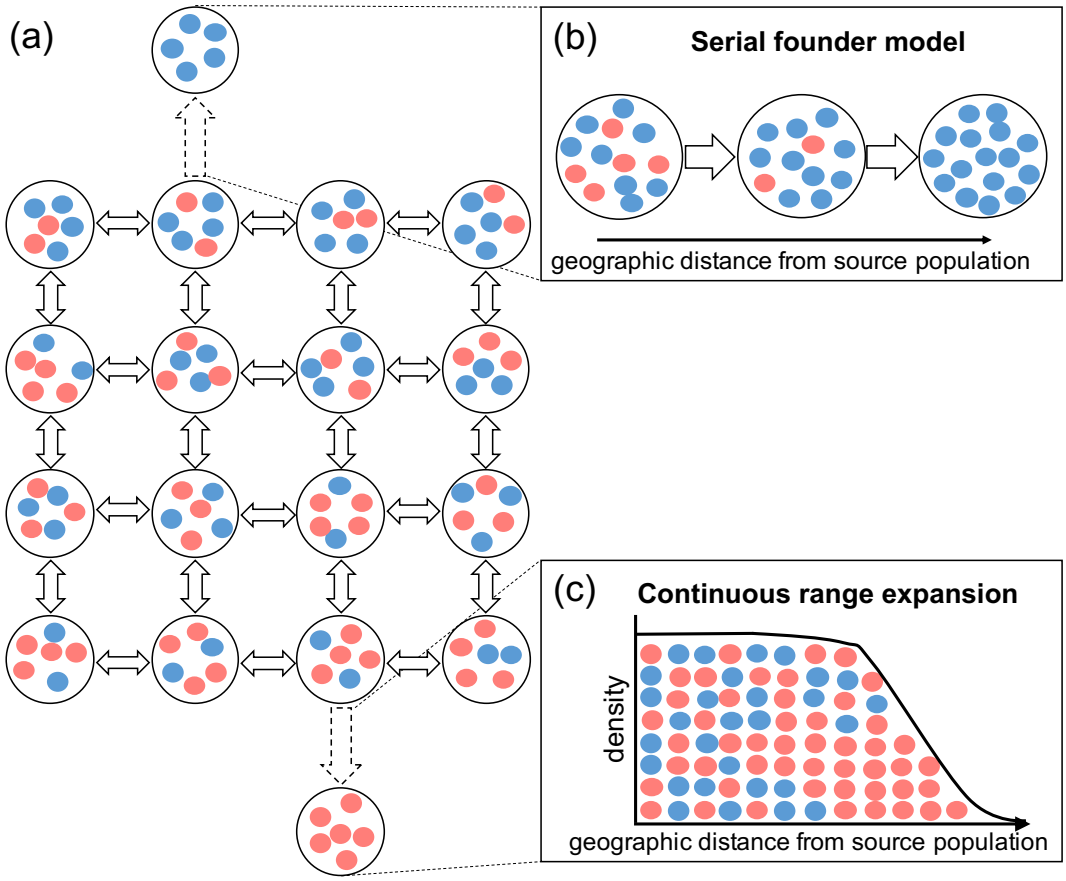



**Figure 6a.** A simple schematic model showing allele frequencies (for two alleles, red and blue) at equilibrium due to low levels of episodic admixture between demes in the core range, whilst alleles drift to extinction or fixation in demes at the range margins due to **b.** serial founder effects or **c.** continuous range expansion (adapted from Peischl et al. 2016). Post expansion variation will coalesce in the founder population.

Whilst the patterns of coalescence within the genomes of a subset of mainly high latitude killer whales were consistent with founder effects, much of our genome dataset showed a more constant long-term effective population size ($N_e$) through time, shared by a geographically widespread set of populations. Notably, we find that all PSMC plots infer an ancestral effective population size

of $N_e \approx 5{,}000$ in deep time (>100-200 KYA). How do we interpret this aspect of the demographic history of killer whales? The *y* axis of the PSMC plot is an estimation of $N_e$ derived from the rate of coalescence between the two chromosomes of a diploid genome. The basis of the PSMC method (Li & Durbin 2011) is provided by the theory of the coalescent process (Hein et al. 2005; Wakeley 2008), in which the rate of the coalescent process in a Wright–Fisher population is also inversely related to the population size (Charlesworth 2009). Coalescence rates are thus dependent upon effective population size, but are also influenced by non-panmictic population structure and selection. Under an *n*-island model of *n* panmictic populations separated by a fixed rate of migration (Wright 1931), and where this rate of migration is low, coalescence events between alleles in different island populations will be rare, which will lead to estimates of large coalescent effective population size (Nei & Takahata 1993; Li & Durbin 2011). Low levels of gene flow among demes that retain their relative sizes can equilibrate the allele frequencies, coalescence times and genetic diversity of neutral loci across demes (Charlesworth 2009). Thus, a single deme can reflect the genetic diversity in a panmictic population with equivalent $N_e$ to the sum of all individual demes in a structured population (Charlesworth 2009).

Whilst killer whale populations may violate some of the underlying assumptions of these expectations from model populations, low levels of migration among killer whale populations in low density regions appears to retain genetic diversity (estimated as $\theta$ by mlRho and $N_e$ by PSMC in our analyses) across ocean basins. Although our results are based upon a low number of individuals, we do find evidence of a rare admixture event in the genome of the Antarctic Peninsula 'type A' individual, in which inferred $N_e$ exponentially increases during the last glacial period and PSMC bootstraps show a wide confidence interval around the median plot. We interpret this result as evidence of rare and recent admixture between two diverged lineages with distinct demographic histories, which are reflected in the distribution of individual bootstraps. We propose these rare admixture events between divergent lineages maintain the 'ancestral' or 'baseline' $N_e$ of approximately 5,000 in our PSMC plots and break up ROH. Admixture between divergent lineages as described above has been proposed to 'rescue' populations and reduce the recessive mutation load (dominance heterosis), in particular upon secondary contact after a period of allopatry and when the two admixing populations' demographic histories do not include shared bottlenecks (Crow 1948; Whitlock et al. 2000; Ingvarsson & Whitlock 2000; Bierne et al. 2002).

The physical position of ROH in the genome were found to overlap in different individuals. ROH have been found to be more abundant in genomic regions with low recombination rate in other study systems, often associated with strong selection (Pemberton et al. 2012; Kardos et al. 2017). Heterogeneity in genetic diversity across the genomic landscape has been shown to be highly correlated across killer whale populations, through comparison of inter-dependent variables: nucleotide diversity ($\pi$), divergence ($D_{xy}$) and differentiation ($F_{ST}$), across multiple pairwise comparisons (see Figure 4 of Foote et al. 2016). Such patterns are typically correlated with the recombination landscape in other study systems (Burri et al. 2015; Samuk et al. 2017); and Foote et al. (2016) concluded that overall the landscape of genome-wide differentiation in killer whales is a result of population-specific genetic drift, regionally elevated by ancestral linked selection (e.g., background selection or repeated selective sweeps shared among populations). As such, $N_e$ estimates and ROH density and length will vary with genomic position reflecting the combined changes in population size, population structure and the strength of selection (including background selection) through time (Mather et al. 2020).

The mutation load from post-bottleneck expansion (expansion load) is expected to be comprised of many weakly deleterious recessive mutations in homozygous genotypes, which may have only a small or negligible effect on fitness (Peischl & Excoffier 2015), making it difficult to disentangle their impact from extrinsic influences on fitness. Additionally, deleterious recessive mutations within homozygous genotypes become exposed to selection and can be purged over time (Szpiech et al. 2013; Hedrick & Garcia-Dorado 2016; Stoffel et al. 2021). Deleterious derived mutations are typically rare (*i.e.* at low frequency) within populations and are therefore predominantly found in heterozygous genotypes in outbred populations, but in homozygous derived genotypes in inbred populations/tracts. The fitness of the deleterious mutation in a heterozygous genotype is determined by both the dominance ($h$) and selection coefficients ($s$), having fitness ($1–hs$); whereas in a homozygous genotype, uninfluenced by dominance, fitness is $1–s$ (Huber et al. 2018). Thus, selection will be more efficacious in purging deleterious recessive alleles from ROH than from heterozygous genotypes. Evidence of such purging has been found by comparing Early Holocene and modern high latitude North Atlantic killer whale genomes (Foote unpublished data). Thus, shorter ROH, for example <1 Mb, have typically been subject to more generations of selection

622 (than longer ROH), which can act upon deleterious recessive alleles in ROH through rescue or
623 purging (Stoffel et al. 2021).
624
625 However, potential fitness impacts of inbreeding are evident in the small (<10 individuals) Scottish
626 West Coast population, which has had zero recruitment over the past two decades (Beck et al.
627 2014). The reduction in genetic diversity and individual fitness caused by inbreeding depression
628 can impact population growth and increase the chance of population extinction (Hedrick &
629 Kalinowski 2000; Keller & Waller 2002). Our finding of >1.5Mb long ROH encompassing a high
630 proportion of the genome combined with the demographic decline in this population suggests the
631 extreme level of long-term inbreeding detected in the Scottish sample exceeds a threshold above
632 which killer whale populations become unviable (see Beck et al. 2014 for consideration of other
633 threats versus inbreeding in this population). Consanguineous mating generates new recessive
634 homozygotes in long ROH, which are in human populations enriched for weakly deleterious alleles
635 (Morten et al. 1956; Szpiech et al. 2013). Yet, even weakly deleterious recessive mutations can
636 contribute to the heritability of complex traits such as genetic diseases and morphology
637 (Lohmueller 2014). Autosomal recessive disorders, for example Chédiak-Higashi syndrome
638 (Taylor and Farrell 1973), have been identified in wild killer whales, but remain under-studied.
639 The genetic impacts of inbreeding depression should therefore be of key concern for populations
640 already impacted by contaminants, depleted prey populations and lethal fishery interactions
641 (Matkin et al. 2008; Beck et al. 2014; Esteban et al. 2016b; Tixier et al. 2017).
642
643 We concur with previous studies (Kardos et al. 2016; Brüniche-Olsen et al. 2018) that ROH
644 estimated from whole genome sequences can provide more nuanced estimates of genetic diversity
645 than other diversity metrics that do not take contiguity and linkage into account. ROH provide a
646 means to differentiate IBD coalescing deep in the pedigree due to ancestral demography, versus
647 coalescing in a more recent ancestor. It is the longer, younger ROH tracts that are likely to harbour
648 more deleterious variation (Szpiech et al. 2013; Stoffel et al. 2021) and which we suggest should
649 be of greater concern when assessing the conservation and management status of endangered
650 populations.
651
652

## REFERENCES


Baird, R.W., McSweeney, D.J., Bane, C., Barlow, J., Salden,D.R., La'Ren, A.K., LeDuc, R.G., and Webster, D.L. (2006). Killer Whales in Hawaiian Waters: Information on Population Identity and Feeding Habits. Pacific Science *60*, 523-530.

Baird, R.W., and Whitehead, H. (2000). Social organization of mammal-eating killer whales: group stability and dispersal patterns. Can. J. Zool. *78*, 2096-2105.

Barrett-Lennard, L.G., (2000). Population structure and mating patterns of killer whales (*Orcinus orca*) as revealed by DNA analysis (Doctoral dissertation, University of British Columbia).

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. J. Stat. Softw. *67*, 1–48.

Beck, S., Foote, A.D., Koetter, S., Harries, O., Mandleberg, L., Stevick, P.T., Whooley, P. and Durban, J.W. (2014). Using opportunistic photo-identifications to detect a population decline of killer whales (*Orcinus orca*) in British and Irish waters. J. Mar. Biol. Assoc. UK *94*, 1327–1333.

Beck, S., Kuningas, S., Esteban, R., and Foote, A.D. (2012). The influence of ecology on sociality in the killer whale (*Orcinus orca*). Behavioral Ecology, 23(2), pp.246-253.

Bierne, N., Lenormand, T., Bonhomme, F., and David, P. (2002). Deleterious mutations in a hybrid zone: can mutational load decrease the barrier to gene flow? Genet. Res. Camb. *80*, 197–204.

Bigg, M.A., Olesiuk, P.F., Ellis, G.M., Ford, J.K.B., and Balcomb, K.C. (1990). Social organization and genealogy of resident killer whales (*Orcinus orca*) in the coastal waters of British Columbia and Washington State. Report of the International Whaling Commission *12*, 383-405.

Browning, S.R. and Browning, B.L. (2007) Rapid and accurate haplotype phasing and missing data inference for whole genome association studies by use of localized haplotype clustering. Am J Hum Genet 81:1084-97.

Brüniche-Olsen, A., Kellner, K.F., Anderson, C.J. and DeWoody, J.A., 2018. Runs of homozygosity have utility in mammalian conservation and evolutionary studies. *Conservation Genetics*, *19*, 1295-1307.

Burri, R., Nater, A., Kawakami, T., Mugal, C. F., Olason, P. I., Smeds, L., ... & Ellegren, H. (2015). Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. Genome Research, 25, 1656-1665.

Busson, M., Authier, M., Barbraud, C., Tixier, P., Reisinger, R.R., Janc, A., and Guinet, C. (2019). Role of sociality in the response of killer whales to an additive mortality event. Proc. Natl. Acad. Sci. *116*, 11812-11817.

Cahill, J. A., Soares, A. E., Green, R. E., & Shapiro, B. (2016). Inferring species divergence times using pairwise sequential Markovian coalescent modelling and low-coverage genomic data. *Phil. Trans. R. Soc. B*: Biol. Sci., 371(1699), 20150138.

Carøe, C., Gopalakrishnan, S., Vinner, L., Mak, S.S., Sinding, M.H., Samaniego, J.A., Wales, N., Sicheritz-Pontén, T., and Gilbert, M.T.P. (2018). Single-tube library preparation for degraded DNA. *Methods in Ecology and Evolution*. 9:410-419.

Ceballos, F.C., Hazelhurst, S. and Ramsay, M., 2018a. Assessing runs of Homozygosity: a comparison of SNP Array and whole genome sequence low coverage data. *BMC Genomics*, *19*, 106.

Ceballos, F.C., Joshi, P.K., Clark, D.W., Ramsay, M. and Wilson, J.F., 2018b. Runs of homozygosity: windows into population history and trait architecture. *Nature Reviews Genetics*, *19*, 220.

Charlesworth, B. (2009). Effective population size and patterns of molecular evolution and variation. Nature Reviews Genetics, 10(3), 195-205.

Chen, I.C., Hill, J.K., Ohlemüller, R., Roy, D.B. and Thomas, C.D., 2011. Rapid range shifts of species associated with high levels of climate warming. *Science*, *333*, 1024-1026.

Crow, J.F. (1948). Alternative hypotheses of hybrid vigour. Genetics *33*, 477–487.

Dumont, B.L., and Payseur, B.A. (2008). Evolution of the genomic rate of recombination in mammals. Evol.: Int. J. Organic Evol. *62*, 276–294.

Esteban, R., Verborgh, P., Gauffier, P., Giménez, J., Foote, A.D., and De Stephanis, R. (2016a). Maternal kinship and fisheries interaction influence killer whale social structure. Behav. Ecol. Sociobiol. *70*, 111-122.

Esteban, R., Verborgh, P., Gauffier, P., Alarcón, D., Salazar-Sierra, J.M., Giménez, J., Foote, A.D., and de Stephanis, R. (2016b). Conservation status of killer whales, Orcinus orca, in the Strait of Gibraltar. Adv. Mar. Biol. *75*, 141–172.

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V.C. and Foll, M., 2013. Robust demographic inference from genomic and SNP data. *PLoS Genet*, *9*, e1003905.

Excoffier, L., Foll, M. and Petit, R.J., 2009. Genetic consequences of range expansions. *Annual Review of Ecology, Evolution, and Systematics*, *40*, 481-501.

Foote, A.D., Kaschner, K., Schultze, S.E., Garilao, C., Ho, S.Y., Post, K., Higham, T.F., Stokowska, C., Van Der Es, H., Embling, C.B., Gregersen, K., et al. 2013. Ancient DNA reveals that bowhead whale lineages survived Late Pleistocene climate change and habitat shifts. *Nature Communications*, *4*, 1-7.

Foote, A.D., Liu, Y., Thomas, G.W., Vinař, T., Alföldi, J., Deng, J., Dugan, S., van Elk, C.E., Hunter, M.E., Joshi, V. and Khan, Z., 2015. Convergent evolution of the genomes of marine mammals. Nature Genetics, 47, 272.

Foote, A.D., Martin, M.D., Louis, M., Pacheco, G., Robertson, K.M., Sinding, M.H.S., Amaral, A.R., Baird, R.W., Baker, C.S., Ballance, L., Barlow, J. et al., 2019. Killer whale genomes reveal a complex history of recurrent admixture and vicariance. *Molecular Ecology*, *28*, 3427-3444.

Foote, A. D., & Morin, P. A. (2016). Genome-wide SNP data suggest complex ancestry of sympatric North Pacific killer whale ecotypes. Heredity, 117(5), 316. https://doi.org/10.1038/hdy.2016.54

Foote, A. D., Vijay, N., Ávila-Arcos, M. C., Baird, R. W., Durban, J. W., Fumagalli, M., … Wolf, J. B. W. (2016). Genome-culture coevolution promotes rapid divergence of killer whale ecotypes. Nature Communications, 7, 11693. https://doi.org/10.1038/ncomms11693

Ford, M.J., Hanson, M.B., Hempelmann, J.A., Ayres, K.L., Emmons, C.K., Schorr, G.S., Baird, R.W., Balcomb, K.C., Wasser, S.K., Parsons, K.M., and Balcomb-Bartok, K., (2011). Inferred paternity and male reproductive success in a killer whale (*Orcinus orca*) population. J Hered *102*, 537-553.

Forney, K.A. and Wade, P.R., 2006. World-wide abundance and density of killer whales. *Whales, Whaling and Ecosystems. University of California Press, Berkeley*, pp.143-160.

Gutenkunst, R.N., Hernandez, R.D., Williamson, S.H. and Bustamante, C.D., 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS genet*, *5*, p.e1000695.

Hasselgren, M., Dussex, N., von Seth, J., Angerbjörn, A., Olsen, R.-A., Dalén, L., and Norén, K. (2021). Genomic and fitness consequences of inbreeding in an endangered carnivore. Mol. Ecol. 30, 2790-2799.

Haubold, B., Pfaffelhuber, P. & Lynch, M. (2010). MlRho–a program for estimating the population mutation and recombination rates from shotgun-sequenced diploid genomes. *Mol. Ecol.* **19**, 277–284.

Hedrick, P. W., and Garcia-Dorado, A. (2016). Understanding inbreeding depression, purging, and genetic rescue. Trends Ecol. Evol. *31*, 940–952.

783 Hedrick, P.W., and Kalinowski, S.T. (2000). Inbreeding depression in conservation biology. Annu.
784 Rev.Ecol. Syst. *31*, 139–162.
785
786 Hein, J., Schierup, M. H. & Wiuf, C. *Gene Genealogies, Variation and Evolution* (Oxford
787 Univ. Press, Oxford, 2005).
788
789 Hoelzel, A.R., Hey, J., Dahlheim, M.E., Nicholson, C., Burkanov, V. and Black, N., 2007. Evolution
790 of population structure in a highly social top predator, the killer whale. Molecular Biology and
791 Evolution, 24, 1407-1415.
792
793 Howrigan, D. P., Simonson, M. A., & Keller, M. C. (2011). Detecting autozygosity through runs of
794 homozygosity: a comparison of three autozygosity detection algorithms. *BMC genomics*, *12*(1),
795 1-15.
796
797 Huber, C.D., Durvasula, A., Hancock, A.M. et al. (2018). Gene expression drives the evolution of
798 dominance. Nat Commun 9, 2750.
799
800 Ingvarsson, P.K., and Whitlock, M.C. (2000). Heterosis increases the effective migration rate.
801 Proc. R. Soc. Lond. B *267*, 1321–1326.
802
803 Ivkovich, T., Filatova, O.A., Burdin, A.M., Sato, H., and Hoyt, E. (2010). The social organization of
804 resident-type killer whales (*Orcinus orca*) in Avacha Gulf, Northwest Pacific, as revealed through
805 association patterns and acoustic similarity. Mammalian Biology *75*, 198-210.
806
807 Kardos, M., Qvarnström, A., & Ellegren, H. (2017). Inferring individual inbreeding and
808 demographic history from segments of identity by descent in *Ficedula* flycatcher genome
809 sequences. *Genetics*, 205(3), 1319-1334.
810
811 Kardos, M., Taylor, H.R., Ellegren, H., Luikart, G., and Allendorf, F.W. (2016). Genomics advances
812 the study of inbreeding depression in the wild. *Evol. Appl*. 9, 1205-1218.
813
814 Kaschner, K., Tittensor, D.P., Ready, J., Gerrodette, T., and Worm, B. (2011). Current and future
815 patterns of global marine mammal biodiversity. PLoS ONE *6*, e19653
816
817 Keller, L.F., and Waller, D.M. (2002). Inbreeding effects in wild populations. Trends Ecol. Evol.
818 *17*, 230–241.
819
820 Kimura, M. (1969). The number of heterozygous nucleotide sites maintained in a finite
821 population due to steady flux of mutations. Genetics *61*, 893–903.
822
823 Kohany, O., Gentles, A.J., Hankus, L. and Jurka, J., 2006. Annotation, submission and screening
824 of repetitive elements in Repbase: RepbaseSubmitter and Censor. BMC Bioinformatics, 7, 474.
825

826 Korneliussen, T. S., Albrechtsen, A., Nielsen, R. 2014. ANGSD: Analysis of Next Generation
827 Sequencing Data. *BMC Bioinformatics*. **15**, 356.
828
829 Leitwein, M., Duranton, M., Rougemont, Q., Gagnaire, P.A. and Bernatchez, L., 2020. Using
830 haplotype information for conservation genomics. *Trends in Ecology & Evolution*, *35*, 245-258.
831
832 Li, H., and Durbin, R. (2011). Inference of human population history from individual whole-
833 genome sequences. Nature 475, 493–496.
834
835 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Durbin, R. (2009). The
836 sequence alignment/map format and SAMtools. *Bioinformatics*, **25**(16), 2078– 2079.
837
838 Lohmueller, K.E. (2014). The impact of population demography and selection on the genetic
839 architecture of complex traits. PLoS Genetics 10, e1004379.
840
841 Lorenzen, E.D., Nogués-Bravo, D., Orlando, L., Weinstock, J., Binladen, J., Marske, K.A., Ugan, A.,
842 Borregaard, M.K., Gilbert, M.T.P., Nielsen, R., Ho, S.Y., et al. 2011. Species-specific responses of
843 Late Quaternary megafauna to climate and humans. *Nature*, *479*, 359-364.
844
845 Louis, M., Skovrind, M., Samaniego Castruita, J.A., Garilao, C., Kaschner, K., Gopalakrishnan, S.,
846 Haile, J.S., Lydersen, C., Kovacs, K.M., Garde, E., Heide-Jørgensen, M.P. et al. 2020. Influence of
847 past climate change on phylogeography and demographic history of narwhals, Monodon
848 monoceros. *Proceedings of the Royal Society B*, *287*, p.20192964.
849
850 Lynch M (2008) Estimation of nucleotide diversity, disequilibrium coefficients, and mutation
851 rates from high-coverage genomic-sequencing projects. Molecular Biology and Evolution, 25,
852 2409– 2419.
853
854 Mather, N., Traves, S. M., & Ho, S. Y. (2020). A practical introduction to sequentially Markovian
855 coalescent methods for estimating demographic history from genomic data. *Ecology and*
856 *evolution*, *10*(1), 579-589.
857
858 Matkin, C.O., Saulitis, E.L., Ellis, G.M., Olesiuk, P., and Rice, S.D. (2008). Ongoing population-
859 level impacts on killer whales Orcinus orca following the 'Exxon Valdez' oil spill in Prince William
860 Sound, Alaska. Mar. Ecol. Prog. Series *356*, 269–281.
861
862 Morin, P. A., Archer, F. I., Foote, A. D., Vilstrup, J., Allen, E. E., Wade, P., … Harkins, T. (2010).
863 Complete mitochondrial genome phylogeographic analysis of killer whales (Orcinus orca)
864 indicates multiple species. Genome Research, 20(7), 908– 916
865
866 Morin, P.A., Parsons, K.M., Archer, F.I., Ávila-Arcos, M.C., Barrett-Lennard, L.G., Dalla Rosa, L.,
867 Duchêne, S., Durban, J.W., Ellis, G.M., Ferguson, S.H. and Ford, J.K. et al. 2015. Geographic and
868 temporal dynamics of a global radiation and diversification in the killer whale. *Molecular*
869 *Ecology*, *24*, 3964-3979.

Morton, N.E., Crow, J.F., and Muller, H.J. (1956). An estimate of the mutational damage in man from data on consanguienous marriages. Proc. Natl Acad. Sci. USA *42*, 855–863.

Moura, A.E., Janse van Rensburg, C., Pilot, M., Tehrani, A., Best, P.B., Thornton, M., Plön, S., de Bruyn, P.N., Worley, K.C., Gibbs, R.A., *et al*. (2014). Killer whale nuclear genome and mtDNA reveal widespread population bottleneck during the last glacial maximum. Mol. Biol. Evol. *31*, 1121–1131.

Nei, M., & Takahata, N. (1993). Effective population size, genetic diversity, and coalescence time in subdivided populations. *Journal of Molecular Evolution*, 37, 240-244.

North, H.L., McGaughran, A, and Jiggins, C. (2021) Insights into species from whole-genome resequencing. Mol. Ecol. This special issue.

Nykänen, M., Kaschner, K., Dabin, W., Brownlow, A., Davison, N.J., Deaville, R., Garilao, C., Kesner-Reyes, K., Gilbert, M.T.P., Penrose, R., Islas-Villanueva, V. et al. 2019. Postglacial Colonization of Northern Coastal Habitat by Bottlenose Dolphins: A Marine Leading-Edge Expansion?. *Journal of Heredity*, *110*(6), pp.662-674.

Parmesan, C. and Yohe, G., 2003. A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, *421*, 37-42.

Parsons, K.M., Balcomb Iii, K.C., Ford, J.K.B., and Durban, J.W. (2009). The social dynamics of southern resident killer whales and conservation implications for this endangered population. Anim. Behav. *77*, 963-971.

Peischl, S., Dupanloup, I., Bosshard, L., & Excoffier, L. (2016). Genetic surfing in human populations: from genes to genomes. Current Opinion in Genetics & Development, 41, 53-61.

Peischl, S., and Excoffier, L. (2015). Expansion load: recessive mutations and the role of standing genetic variation. Mol. Ecol. *24*, 2084–2094.

Pemberton, T.J., Absher, D., Feldman, M.W., Myers, R.M., Rosenberg, N.A., and Li, J.Z. (2012). Genomic patterns of homozygosity on worldwide human populations. Am. J. Human Genet. 91, 275–292.

Pilot, M., Dahlheim, M.E., and Hoelzel, A.R. (2010). Social cohesion among kin, gene flow without dispersal and the evolution of population genetic structure in the killer whale (*Orcinus orca*). J Evol Biol *23*, 20-31.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., De Bakker, P.I., Daly, M.J. and Sham, P.C., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American journal of human genetics*, *81*, 559-575.

Quinlan, A.R. & Hall, I.M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics, 26, 841-842.

R Core Team. (2017). R: A language and environment for statistical computing, https://www.r-project.org/

Ready, J., Kaschner, K., South, A.B., Eastwood, P.D., Rees, T., Rius, J., Agbayani, E., Kullander, S., and Froese, R. (2010). Predicting the distributions of marine organisms at the global scale. Ecological Modelling *221*, 467-478.

Reisinger, R.R., Beukes, C., Hoelzel, A.R., and de Bruyn, P.N. (2017). Kinship and association in a highly social apex predator population, killer whales at Marion Island. Behav. Ecol. *28*, 750-759.

Ringbauer, H. Novembre, J., and Steinruecken, M., 2020. Detecting runs of homozygosity from low-coverage ancient DNA. *bioRxiv*. https://doi.org/10.1101/2020.05.31.126912

Robinson, J.A., Räikkönen, J., Vucetich, L.M., Vucetich, J.A., Peterson, R.O., Lohmueller, K.E., and Wayne, R.K. (2019). Genomic signatures of extensive inbreeding in Isle Royale wolves, a population on the threshold of extinction. Science Advances 5, eaau0757.

Samuk, K., Owens, G. L., Delmore, K. E., Miller, S. E., Rennison, D. J., & Schluter, D. (2017). Gene flow and selection interact to promote adaptive divergence in regions of low recombination. Molecular Ecology, 26(17), 4378-4390.

Sarabia, C., voHoldt, B., Larrasoaña, J.C., Uríos V, and Leonard, J.A. (2020). Pleistocene climate fluctuations drove demographic history of African golden wolves (*Canis lupaster*). Mol. Ecol. https://doi.org/10.1111/mec.15784

Schubert, M., Lindgreen, S. and Orlando, L., 2016. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC research notes*, *9*, 1-7.

Smit, A., Hubley, R. & Green, P. 1996. RepeatMasker Open-3.0 www.repeatmasker.org.

Stoffel MA, Johnston SE, Pilkington JG, Pemberton JM. (2021). Mutation load decreases with haplotype age in wild Soay sheep. Evolution Letters, 5, 187-195.

Sousa, V. and Hey, J., 2013. Understanding the origin of species with genome-scale data: modelling gene flow. Nature Reviews Genetics, 14, 404-414.

Szpiech, Z.A., Xu, J., Pemberton, T.J., Peng, W., Zöllner, S., Rosenberg, N.A., and Li, J.Z. (2013). Long runs of homozygosity are enriched for deleterious variation. Am. J. Human Genet. *93*, 90–102.

Taylor, B.L., Chivers, S.J., Larese, J., and Perrin, W.F. (2007). Generation length and percent mature estimates for IUCN assessments of cetaceans. *NOAA, NMFS, Southwest Fisheries Science Center Administrative Report LJ-07–01*, *21*.

Taylor, R.F., and Farrell, R.K. (1973). Light and electron microscopy of peripheral blood neutrophils in a killer whale affected with Chediak-Higashi syndrome. Fed. Proc. *32*, 822a

Thompson, E.A. (2013). Identity by descent: variation in meiosis, across genomes, and in populations. Genetics, 194, 301-326.

Tixier, P., Barbraud, C., Pardo, D., Gasco, N., Duhamel, G., and Guinet, C. (2017). Demographic consequences of fisheries interaction within a killer whale (*Orcinus orca*) population. Mar. Biol. *164*, 170.

Veeramah, K.R. and Hammer, M.F., 2014. The impact of whole-genome sequencing on the reconstruction of human population history. *Nature Reviews Genetics*, *15*, 149-162.

Wakeley, J. *Coalescent Theory. An Introduction* (Ben Roberts, Greenwood Village, Colorado, 2008).

Watterson, G. (1975). On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. *7*, 256–276.

Whitlock, M.C., Ingvarsson, P.K., and Hatfield, T. (2000). Local drift load and the heterosis of interconnected populations. Heredity *84*, 452–457.

Wright, S. (1931). Evolution in Mendelian populations. Genetics 16, 97-159.

**Data Availability Statement**

All genomic data and associated metadata will be deposited in the NCBI sequence read archive.

**Acknowledgements**

The sequencing service was provided by GeneWiz (genewiz.com) and the Genomics Core Facility (GCF), Norwegian University of Science and Technology (NTNU). GCF is funded by the Faculty of Medicine and Health Sciences at NTNU and Central Norway Regional Health Authority. A.D.F. was supported by a short visit grant from the European Science Foundation–Research Networking Programme ConGenOmics and by a Swiss National Science Foundation grant (31003A-143393) to L. Excoffier, and by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 663830. MTPG acknowledges ERC Consolidator Award 681396-Extinction Genomics. Additional genomic sequencing was funded by Mridula Srinivasan, Protected Species Science Branch, NOAA Fisheries Office of Science and Technology. We also thank James McBain and Debbie Duffield for providing DNA. We thank Anna Hewett for feedback on our Biorxiv preprint, Marty Kardos and 2 anonymous reviewers for their helpful feedback on an earlier draft submitted to Current Biology, and Martin Stoffel and two

1002 anonymous reviewers for their constructive comments on our submission to Molecular Ecology
1003 – all of this input greatly improved the manuscript. Marty Kardos also generously provided R-
1004 code for estimating the relationship between ROH length and recombination rate.
1005