







## ORIGINAL ARTICLE OPEN ACCESS

# Admixture Increases Genetic Diversity and Adaptive Potential in Australasian Killer Whales

Isabella M. Reeves<sup>1,2</sup>  | John A. Totterdell<sup>2</sup>  | Jonathan Sandoval-Castillo<sup>3</sup>  | Emma L. Betty<sup>4</sup>  | Karen A. Stockin<sup>4</sup>  | Ramari Oliphant Stewart<sup>5</sup> | Muriel Johnstone<sup>6</sup> | Andrew D. Foote<sup>7</sup> 

<sup>1</sup>College of Science and Engineering, Flinders University, Bedford Park, South Australia, Australia | <sup>2</sup>Cetacean Research Centre (CETREC WA), Esperance, Western Australia, Australia | <sup>3</sup>Molecular Ecology Laboratory, College of Science and Engineering, Flinders University, Bedford Park, South Australia, Australia | <sup>4</sup>Cetacean Ecology Research Group, College of Sciences, Massey University, Auckland, New Zealand | <sup>5</sup>Te Kauika Tangaroa Charitable Trust, New Zealand | <sup>6</sup>Ōraka-Aparima Rūnaka, New Zealand | <sup>7</sup>Centre for Ecological and Evolutionary Synthesis, Department of Biosciences, University of Oslo, Oslo, Norway

**Correspondence:** Isabella M. Reeves ([isabella.reeves@flinders.edu.au](mailto:isabella.reeves@flinders.edu.au)) | Andrew D. Foote ([andrew.foote@ibv.uio.no](mailto:andrew.foote@ibv.uio.no))

**Received:** 26 November 2024 | **Revised:** 17 January 2025 | **Accepted:** 27 January 2025

**Handling Editor:** Martin Kardos

**Funding:** This work was supported by the Royal Society of South Australia and American Natural History Museum. We are grateful to the Playford Trust Foundation and the Australian Government for their scholarship and stipend support for IMR. ADF was funded by European Research Council (ERC) Consolidator Grant ERC-COG-101045346 'EXPLOAD'.

**Keywords:** cetacean | effective population size | heterosis | inbreeding | introgression

## ABSTRACT

Admixture is the exchange of genetic variation between differentiated demes, resulting in ancestry within a population coalescing in multiple ancestral source populations. Low-latitude killer whales (*Orcinus orca*) populations typically have higher genetic diversity than those in more densely populated, high productivity and high-latitude regions. This has been hypothesized to be due to episodic admixture between populations with distinct genetic backgrounds. We test this hypothesis by estimating variation in local ancestry of whole genome sequences from three genetically differentiated, low-latitude killer whale populations and comparing them to global genetic variation. We find 'Antarctic-like' ancestry tracts in the genomes of southwestern Australia (SWA) population including recent (within the last 2–4 generations) admixture. Admixed individuals had, on average, shorter and fewer runs of homozygosity than unadmixed individuals and increased effective population size ( $N_e$ ). Thus, connectivity between demes results in the maintenance of  $N_e$  of relatively small demes at a level comparable to the sum of  $N_e$  across demes. A subset of the admixed regions was inferred to be evolving under selection in the SWA population, suggesting that this admixed variation may be contributing to the population's adaptive potential. This study provides important and rare empirical evidence that small populations can maintain genetic diversity due to sporadic admixture between different genetic backgrounds and that admixed ancestry can promote the long-term stability of  $N_e$ .

## 1 | Introduction

Admixture is the exchange of genetic variation between differentiated demes and results in populations comprising ancestry from multiple sources (Korunes and Goldberg 2021). Reported examples of admixture are increasingly common in nature,

and consequential changes to the genetic composition and associated demographic parameters are key to understanding the evolutionary history of admixed populations (Galaverni et al. 2017; Mathieson and Scally 2020). However, understanding evolutionary processes in admixed populations and untangling the source of their constituent ancestry components

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *Molecular Ecology* published by John Wiley & Sons Ltd.

remains challenging (Tan and Atkinson 2023). Thus, despite admixture being potentially common and its evolutionary impact rapid, admixed populations remain relatively understudied (Tan and Atkinson 2023).

Admixture can shape genetic and phenotypic variation within populations (Korunes and Goldberg 2021). Potential benefits of admixed ancestry result from increased population genetic variance and can include increased fitness due to the combined ancestry of different source populations (i.e., heterosis) (Facon et al. 2005), and thereby increased adaptive potential (e.g., Verhoeven et al. 2011). For example, post-admixture selection provides the opportunity for rapid adaptation (as reviewed in Hamid et al. 2023). In addition, admixture can mask the recessive genetic load underlying inbreeding depression (Charlesworth and Willis 2009). However, admixture can also introduce deleterious genetic variation, or mask beneficial recessive alleles within newly-formed heterozygous genotypes leading to reduced fitness and outbreeding depression (Lynch 1991). The impact of admixture is dependent on the prior divergence, diversity, stochastic factors and ecological niche of the two admixing populations (Fontsero et al. 2019). Therefore, understanding the underlying demographic history and evolutionary outcomes of admixture is crucial within the field of molecular ecology. The first step in this process is to understand the composition of admixed ancestry within populations of interest.

Killer whales are a cosmopolitan species with a range that spans from the polar pack ice to the equator (Forney and Wade 2006). Perhaps counter-intuitively, populations in highly productive, high-latitude waters that support the greatest densities of killer whales (Forney and Wade 2006) are typically home to populations with the lowest genetic diversity and effective population size ( $N_e$ ) (Foote et al. 2021) – albeit with localized exceptions due to anthropogenic impacts on census size (e.g., Kardos et al. 2023). This global pattern is consistent with range expansion theory (Excoffier et al. 2009) and the hypothesis that these populations have colonised high-latitude waters after the Last Glacial Maximum (Foote et al. 2021, 2019; Hoelzel et al. 2007; Morin et al. 2015). In contrast, populations with the highest genetic diversity are found at lower latitudes (Foote et al. 2021), where killer whales are typically at low density (Forney and Wade 2006; Morin et al. 2015). This has been hypothesized to be due to rare episodic admixture between divergent populations, resulting in the recombination of distinct genetic backgrounds (Foote et al. 2019). Episodic gene flow occurring amongst demes could theoretically retain  $N_e$  within each individual deme at the level of a panmictic global population (Charlesworth 2009). Even under realistic scenarios of unequal gene flow amongst natural populations, admixture has the potential to increase  $N_e$  of individual demes (e.g., Saremi et al. 2019; Kyriazis et al. 2024).

To date, population genomics studies of killer whale populations have primarily focused on high-latitude populations (e.g., Foote et al. 2016; Garroway et al. 2024; Kardos et al. 2023; Moura et al. 2015). To better understand the processes driving the patterns in genetic diversity and  $N_e$  described above, we investigate if variation in local ancestry is consistent with admixture in genomes from three genetically differentiated low-latitude killer whale populations (Reeves et al. 2022). Our findings shed new

light on how admixed ancestry influences genetic diversity and  $N_e$  in the genomes of these populations. Furthermore, by investigating how admixed ancestry varies across the genome, we gain new insight into the evolutionary processes.

## 2 | Materials and Methods

### 2.1 | Sample Collection, DNA Extraction and Whole Genome Sequencing

Thirty tissue biopsies (skin and blubber) were collected under permit from 20 free-ranging and 10 stranded animals within Australasia between 2013 and 2021. Biopsies from free-ranging animals were taken using a Barnett crossbow with 150 lb. draw weight, 50 cm long darts (designed by F. Larsen, CETA DART, Copenhagen, Denmark) and biopsy tips (length: 40 mm), which provided samples of approximately 25 mm × 4 mm size. Tissue was preserved in 95% ethanol and stored at  $-80^{\circ}\text{C}$ . The samples were previously assigned to three populations: northwestern Australia (NWA), southwestern Australia (SWA) and New Zealand (NZ) based on variation at RADseq markers (Reeves et al. 2022).

Genomic DNA was extracted using a salting-out protocol from tissue samples (Sunnucks and Hales 1996). DNA quality, integrity and quantity were assessed using a NanoDrop 1000 spectrophotometer (Thermo Scientific), gel electrophoresis (2% agarose gels) and a fluorometer (Qubit 2.0, Life Technologies), respectively. Library build and sequencing were completed in three batches. An additional two rounds of extractions and sequencing were performed to increase the coverage of genomic data presented in Reeves et al. (2023). Firstly, DNA extracts were built into genomic libraries by Novogene (Singapore) using NEBNext Ultra II DNA library prep kit and sequenced on an Illumina NovaSeq 6000 S2 platform (150 bp PE) for a total of 547 Gb. We then built new extracts from the same samples into genomic libraries using a UDI primer set and sequenced the samples at QB3 Genomics (University of California, Berkeley) across two lanes of an Illumina NovaSeq 6000 S4 platform (150 bp PE).

### 2.2 | Mapping and Filtering

All bioinformatics and relevant analysis were undertaken using Flinders University's server Deepthought (Flinders University 2021). Sequences were trimmed using Adapter Removal v2.3.2 with a minimum length of 70 bp (Schubert et al. 2016) to remove trim residual adapter sequences and low-quality stretches. Sequence data were then mapped to a high-quality chromosomal reference genome assembly from a North Pacific 'Biggs' or 'transient' (mammal-eating ecotype) killer whale (CNCBdb accession: CNA0050865; Kardos et al. 2023) using BWA-MEM v 0.7.17 (Li and Durbin 2009). Duplicate and ambiguous reads were collapsed using the markdup function in SAMtools v1.14 (Li et al. 2009). Repeat regions were identified using Repeatmasker v4.1 (Smit et al. 2004), as per Foote et al. (2021) and were masked using BEDtools v2.3.0 (Quinlan 2014) to reduce potential bias to downstream population genetics inference. Separate files were generated for both autosomes and sex chromosomes for downstream analysis.

## 2.3 | Published Reference Datasets

To understand the genetic relationship of the Australasian killer whales sampled and sequenced specifically for this study to other killer whale populations, we incorporated two additional datasets of published genome sequences. A dataset of 24 high coverage (> 10×) genomes, in which a single representative individual had been sequenced from populations across the species' range (Foote et al. 2021, 2023). Secondly, a dataset of low coverage (~2×) genomes, comprised of approximately 10 individuals per population from the following ecotypes: North Pacific 'residents', North Pacific 'Bigg's', Antarctic types B1, B2 and C (Foote et al. 2016). These two comparative datasets are hereafter referred to as the high-coverage global dataset and low-coverage ecotype dataset, respectively. Accession numbers are given in Table S2.

## 2.4 | Checking for Batch Effects

To screen our data for the impact of batch effects (see Lou and Therkildsen 2022) on the different sequencing runs we estimated covariance as specified in Korneliussen et al. (2014), comparing both amongst samples and amongst sequencing runs for the same sample. Covariance was estimated using the direct sampling method implemented in ANGSD v0.935 (Korneliussen et al. 2014), in which a single allele per site is selected at random. Sites were filtered to only those that had a probability of  $p < 0.000001$  of being variable, had a minimum mapping quality of 30 and a minimum base quality of 30. Sequence data for a subset of individuals from different runs were checked for consistency and batch effects (Figure S1). Sequence data from different sequencing batches for the same individual were then merged using SAMtools *merge* (Li et al. 2009) to create a single high coverage (~10×) genome per sample for use in all downstream analyses. Effective coverage (i.e., after filtering) per genome was estimated with SAMtools v1.14 (Li et al. 2009).

## 2.5 | Inferring Shared Ancestry Using *D*-Statistics

Asymmetry between Australasian populations in the sharing of derived alleles with different outgroups from the high coverage global dataset was quantified using the *D*-statistic (Green et al. 2010). This test compares counts of shared derived alleles between an outgroup and two ingroup populations to test for statistically significant asymmetry, which would suggest either gene flow from the outgroup to one ingroup or that the assumed topology is not supported. For example, if NWA and SWA are considered ingroups and population X is the outgroup sample, the test can estimate if the data support the null hypothesis that the tree topology ((NWA, SWA,) X dolphin) is correct, and that there is no excess gene flow between X and either of the ingroups. We followed the Durand et al. (2011) definition:

$$D = \frac{nBABA - nABBA}{(nBABA + nABBA)}$$

Testing the topology ((SWA, NWA,) X dolphin), nABBA is the number of sites where an NWA genome and the outgroup X share a derived allele, whilst an SWA genome shares the ancestral allele with the bottlenose dolphin; and nBABA is the number of

sites that an SWA genome and X share a derived allele, whilst an NWA genome shares the ancestral allele with the dolphin. If the null hypothesis is supported as true, we expect an approximately equal proportion of sites showing the ABBA and BABA pattern of allele sharing, and that  $D \approx 0$ . We ran tests alternating X with a different sample from a global dataset. Statistical significance was assessed using a Z-score based on block jack-knife (size: 5,000,000) estimates of the standard error of the *D*-statistics and assumes that the statistic is normally distributed with a mean of 0. *D*-statistics were estimated in ANGSD with the following parameters and filters; -doCounts 1 -anc ./Tursiops.fa -minMapQ 30 -minQ 30 -uniqueonly 1 -remove\_bads 1 -blocksize 5,000,000. These tests were implemented in ANGSD to reduce the bias of varying sequencing depth across the genome, which sampled a single base at each position of the genome using a global dataset from Foote et al. (2023); Foote et al. (2021); Foote et al. (2019). We included a bottlenose dolphin (*Tursiops truncatus*) as an outgroup (Short Read Archive accession code: SRX200685; (Foote et al. 2015)) which was processed as per the 'Mapping and filtering' section. The choice of reference genome can introduce reference bias to *D*-statistics in the form of an artificial excess of allele sharing to the populations that are genetically closer to the reference genome. We therefore re-estimated the *D*-statistics using the same short-read data mapped to an alternative high-quality chromosomal reference genome assembly of a Norwegian killer whale (Foote et al. 2022) to check if results were influenced by the choice of reference genome.

## 2.6 | Individual Assignment and Admixture Analysis

Using a pruned dataset for linkage disequilibrium (SNP ≥ 100 kb apart) we examined population structure. First, we used principal component analysis (PCA) in RStudio v2022.07.1 + 554 (Studio 2016) using package FactoMineR v2.3 (Lê et al. 2008) to identify clusters based on allele frequency covariance. We then used the maximum likelihood individual-based assignment method implemented in NGSADMIX v2 (Skotte et al. 2013), which uses genotype likelihoods generated using ANGSD, estimating the posterior genotype likelihood 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 < 0.000001 and with a minimum minor allele frequency of 0.05. We removed reads with mapping quality (MAPQ) < 30, sites with low base quality scores ( $q < 20$ ), and reads that did not map uniquely, adjusting *q*-scores around indels, adjusting the minimum quality score to 50 for excessive mismatches and discarding bad reads (flag ≥ 256). We included available Antarctic killer whales ('type Bs and/or 'type C'), and North Pacific 'resident' and 'Bigg's' killer whale genomes to act as a reference panel (i.e., the low coverage ecotype dataset) in downstream analysis (see Table S2). We ran NGSADMIX assuming 1–8 ancestral populations (*k*) and changed the seed numbers between each of the three runs to check for convergence. We calculated  $\Delta k$  for the number of populations using a custom R-script.

NGSAdmix along with other genetic clustering model-based methods (e.g., ADMIXTURE and STRUCTURE) allow individual ancestries to be inferred. However, they have underlying

assumptions that natural populations do not conform to, such as genetic ancestry being only comprised of populations provided; consequently, results may be misleading (Lawson et al. 2018). We therefore evaluate the fit of the admixture models based on the correlation between the true genotype and the residual difference of the genetic clustering model using evalAdmix v0.95 (Garcia-Erill and Albrechtsen 2020). A good fit of the model results in correlation residuals close to 0. Positively correlated residuals may be indicative of similar demographic histories that are not accurately modelled (e.g., due to the inclusion of close relatives, ghost populations, incorrect K values or other demographic scenarios), whilst negative values can result from a pair of individuals that have different demographic histories but have been modelled to share the same ancestry (Garcia-Erill and Albrechtsen 2020). Here, we used the NGSadmix outputs based on genotype likelihoods to estimate the fit of the inferred ancestry on our population dataset.

## 2.7 | Inferring Shared Ancestry Using PCA

To investigate this potential admixture signal we used PCANGSD v1.10 (Meisner and Albrechtsen 2018), which can use a specified number of eigenvectors to help untangle influences of variation within the dataset. To assess the relationship between the Australasian genomes and the low-coverage ecotype dataset, we generated a PCA inferred from genotype likelihoods in PCANGSD. Additionally, we used the PCANGSD -admix function to look at ancestry proportions of three genetically distinct low-coverage ecotype dataset (resident, Biggs and Antarctic type C) within the three Australasian killer whales. Ancestry proportions are based on the first eigenvector (PC1) using the -e function based on the unpruned dataset. Genetic labelling is often imprecise and therefore misleading (Coop 2023), for instance, failing to accurately account for continuous genetic variation within a species. Therefore, we use the label '-like ancestry' (e.g., 'Antarctic-like' ancestry) to infer ancestry source populations that are genetically similar to, but not necessarily genetically homogenous with, these ecotypes.

## 2.8 | F-Statistic and Treemix Tests for Ancestral Admixture

F-statistics and TreeMix v1.12 were estimated from a VCF based on filtered BAMs to generate a VCF containing called SNPs using BCFtools v1.9 based on the low coverage dataset and Australasian genomes (Danecek et al. 2016). SNPs were kept if they were called in at least 80% of individuals, had an allele balance between 20% and 80% within the heterozygote genotype, and exhibited a minor allele frequency of at least 3%.

The  $f_3$ -statistic uses variation in allele frequencies of population pairs within a tree to assess genetic drift (Patterson et al. 2012; Peter 2016; Reich et al. 2009). The  $f_3$ -statistic can detect admixture, even if gene flow occurred several hundred generations ago (Patterson et al. 2012). These tests are estimated based on  $f_3(A; B, C)$ , where a statistically significant negative  $f_3$ -statistic indicates admixture in population A.  $F_3$ -statistics were estimated as per Patterson et al. (2012) with standard errors using jackknife blocks of 1000 SNPs using TreeMix. Foote et al. (2021) estimated

a mean inverse SNP density (Kb per SNP) of 0.642 (range: 0.204–4.906) from the high-coverage dataset of killer whales. Thus, physical block size is expected to fall within the range of 0.2–5.0 Mbp. We tested three  $f_3$ -statistics forms for SWA, NWA and type C, rotating the position in the form of  $f_3(A; B, C)$ .

When admixture has been detected, the admixed proportions can be estimated using  $f_4$  ratio statistics. The tests based on the form  $f_4(A, B; C, D)$  are expected to be zero when there is no overlap in the paths of allele frequency changes between A and B and between C and D in the tree edges.  $F_4$  indicates admixture when the value statistically significantly deviates from 0, negative or positive estimates are dependent on the position of the source population on the tree. This statistic is insensitive to post-admixture drift, making it useful for detecting admixture events, even if they occurred hundreds of generations ago (Patterson et al. 2012).

Patterson et al. (2012) defined the  $f_4$  ratio as:

$$f_4 \frac{(C, O; X, B)}{(C, O; A, B)}$$

where A and C are a sister group, B is sister to (A, C), X is a mixture of A and B and O is the outgroup. This ratio estimates the ancestry from A, denoted as  $\alpha$ , and the ancestry from B, as  $1-\alpha$  identify source populations of admixture with the three Australasian populations we tested  $f_4$ , as per Patterson et al. (2012) with standard errors using jack-knife blocks of 1000 SNPs. We tested three  $f_4$ -statistics in TreeMix where the Australasian populations and Antarctic type C were rotated in the form of  $f_4(A, B; C, D)$ .

Genetic relationships amongst the reference and Australasian populations were reconstructed using population allele frequency data to generate a bifurcating maximum likelihood tree to illustrate genetic drift using TreeMix v1.12 with Gaussian approximation (Pickrell and Pritchard 2012). The tree's branches represent population relationships based on predominant alleles. Migration edges are introduced between populations that deviate from the tree model, indicating inferred allele exchange. The inferred directionality of gene flow along migration edges is based on asymmetries in a covariance matrix of allele frequencies relative to an ancestral population, as indicated by the tree. We used TreeMix to generate a bifurcating maximum likelihood tree using SNP blocks of 1000 SNPs based on a VCF file for migration events 0–2 to understand potential directional admixture events.

## 2.9 | Local Ancestry Inferred From Haplotypes

We used ELAI v1.01 based on a two-layer hidden Markov Model to detect the structure of haplotypes based on linkage disequilibrium to infer the local ancestry of admixed individuals based on unphased data (Guan 2014) from the VCF file. The program performs independent phasing before calling local ancestry (Guan 2014). Thus, we leveraged this to understand the SNP dosage from the source populations within the admixed populations, using a three-way admixture model from three source populations with -C (upper clusters) 3 -c

(lower clusters) 15 based on  $5 \times C$  as per ELAI documentation and mixed generations (mg) of 10 and 100 for comparison. Source populations ideally need to be highly differentiated from one another, and therefore we choose population reference datasets that represent resident-like, Biggs-like and Antarctic-like ancestry (type C). We used a custom R script to visualise the dosage along the genome with 50 kb sliding windows and 10 kb stepping size.

## 2.10 | Tests for Recent Admixture and Timing

Approaches that model the global proportion of an individual's genome based on hybrid indexes or admixture proportions cannot disentangle recent (i.e., within the last five generations) and ancestral admixture (Lawson et al. 2018). The 'Admixture Pedigrees of Hybrids' (apoh) approach (Garcia-Erill et al. 2023) assesses inter-ancestry heterozygosity to detect recently admixed individuals and estimate possible scenarios of recent admixture. This approach uses two complementary models. The paired ancestries model analyses unordered paired ancestries without distinguishing between the possible heterozygous permutations (as reviewed in Garcia-Erill et al. 2023). Whilst the parental admixture model assesses the ordered paired ancestry and considers the possible permutations, apoh automates these models and considers the relative fit of the estimated proportions to detect recent admixture (as reviewed in Garcia-Erill et al. 2023). Therefore, we estimated recent admixture with the possible source populations. Note, as we were trying to understand recent admixture events, we included all population-level whole genome data, including the three Antarctic contemporary populations in this analysis, therefore using  $K=7$  NGSadmixon input.

## 2.11 | Calling Runs of Homozygosity for Individual Genomes

Runs of homozygosity (ROH) were identified using a window-based approach in plink v1.07 (Purcell et al. 2007) for the Australasian populations on a VCF file. We used the high-coverage global dataset when calling the ROHs to act as a reference dataset. Using a sliding window size of 300 kb, a minimum of 50 SNPs with a density of one SNP per 50 kb was required to identify a ROH in the dataset. Up to three heterozygote sites per 300 kb window within called ROHs were allowed to account for genotyping errors, and a 1 kb distance between two SNPs was necessary for them to be considered in separate ROHs.

## 2.12 | Tests for Selection for Admixed Loci

In an attempt to assess if any regions that were admixed were under selection, we used PCANGSD v1.10-selection function that extended the model of FastPCA that uses genotype likelihoods of the Australasian and low coverage ecotype dataset (Galinsky et al. 2016) to perform a genome-wide selection scan along PC1 of the unpruned data (Meisner et al. 2021). We fit the data to a qqplot in Rstudio (RCoreTeam 2022), using chi-square and Lambda to explain variance that statistically significantly

differentiates from genetic drift along the axis. Additionally, we examined any peaks driven by the variance in PC1 using package qqman (Turner 2014) to build a Manhattan plot with the  $-\log_{10}$  transformed  $p$ -values obtained and plotted against chromosome and position. The data were not pruned to remove the effects of linkage disequilibrium (Lotterhos 2019). Therefore, it is important to consider the variance driven by each PCA is inflated by the inclusion of non-independent SNPs.

## 3 | Results

### 3.1 | Genome Datasets and Batch Effects

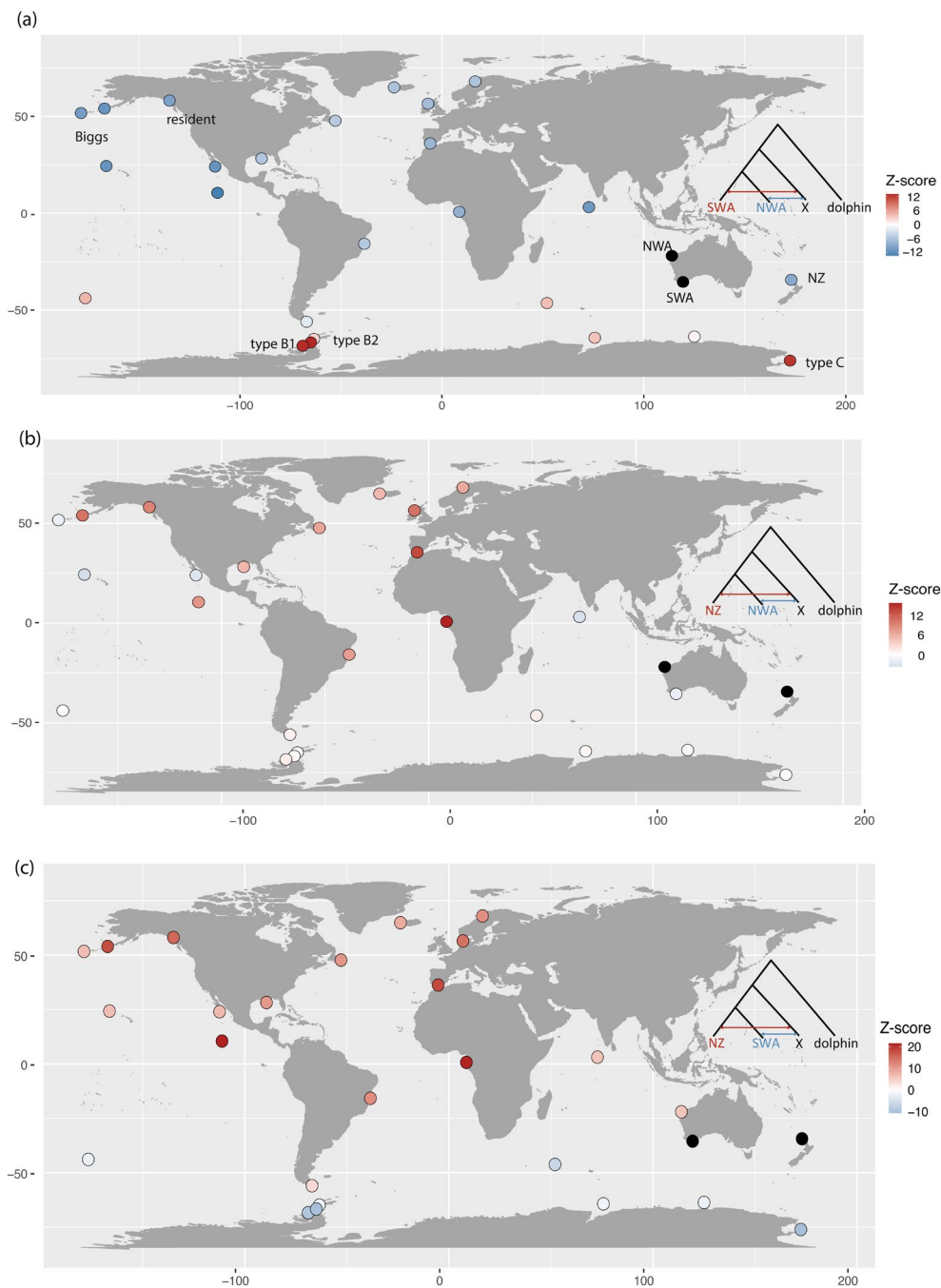
To understand how different library build methods, sequencing platforms and sequencing service providers may bias results between and within runs, we compared replicates between runs estimated with 783,990 SNPs based on genotype likelihoods. Figure S1 illustrates the covariance of replicates within and between runs, as well as replicate samples, sequenced on different platforms, clustered closely together in the adjacent tree. Thus, whilst genotyping may exhibit some small variability between runs, batch effects are unlikely to influence the biological significance of any results here.

### 3.2 | $D$ -Statistics Indicate Admixture Signal for Southwestern Australian Killer Whales

We found that 21 of 25 tests of form  $D(\text{SWA}, \text{NWA}; X, \text{dolphin})$  were statistically significant based on the  $Z$ -score of  $-3 < Z < 3$  and that in 16 of these tests,  $X$  (representing a geographically widespread set of populations, Figure 1a, S2a) shared a statistically significant excess of derived alleles with NWA. The highest  $D$ -statistics being when  $X$  was one of the distinctive Antarctic morphotypes (types B1, B2 and C) (Figure 1a, S2a).

Additionally, we found 11 of 25 tests of the form  $D(\text{NZ}, \text{NWA}, X, \text{dolphin})$  were statistically significant due to an excess of shared derived alleles between the NZ genome and genomes from the North Atlantic, North Pacific resident, offshore, Clipperton Island populations (Figures 1b and S2b). The NWA genome shared an excess of derived alleles with genomes from the Maldives, Hawaii and Mexico (Figures 1b and S2b). These  $D$ -statistic results mirror the genetic affinities of each Australasian genome in the PCA in Figure 1c of Foote et al. (2019). Re-estimation of the  $D$ -statistics after re-mapping to an alternative reference assembly generated from a Norwegian killer whale (Foote et al. 2022) provided comparable results (Figure S3, Tables S4–S6), indicating no major impact of reference bias on our findings.

Estimation of  $D(\text{NZ}, \text{SWA}, X, \text{dolphin})$  revealed that 21 of 25 of the tests statistically significantly deviated from the expected symmetry of derived allele sharing under the tested topology. The same populations that shared an excess of derived alleles with the NWA genome relative to the SWA genome also shared an excess of derived alleles with the NZ genome (Figures 1c and S2b). Likewise, the SWA genome also shared an excess of derived alleles with the Antarctic types in this test. Thus, we see



**FIGURE 1** | A comparison of  $D$ -statistic form **(a)**  $D(\text{SWA}, \text{NWA}, \text{X}, \text{dolphin})$ , **(b)**  $D(\text{NZ}, \text{NWA}, \text{X}, \text{dolphin})$  and **(c)**  $D(\text{NZ}, \text{SWA}, \text{X}, \text{dolphin})$  for Australasian killer whales compared to a global reference dataset. Statistical significance is indicated by a Z-score  $> 3$  or  $< -3$ . ‘SWA’ is southwestern Australia, ‘NWA’ is northwestern Australia and ‘NZ’ is New Zealand. Dots are representative of the sample locality.

that the NWA and NZ genomes share a genetic affinity with populations to the North, whilst the SWA genome shares a genetic affinity with populations to the south.

To test whether X should be an ingroup with an Australasian population we estimated  $D$ -statistics of the form (pop1, X; pop2, dolphin), where pop1 and pop2 are Australasian populations (Figure S4). Consistently, the highest significance was for tests including an Antarctic type as an ingroup (X). Australasian killer whale genomes share a statistically significant excess of shared derived alleles relative to sharing between Australasian and Antarctic genomes. Additionally, like

the previous  $D$ -statistic tests, genomes from the NZ and NWA populations share a statistically significant excess of derived alleles with various northern hemisphere populations, whilst the SWA population share an excess of derived alleles with a single Southern Ocean population (See Table S6 and Figure S4 for population-specific results).

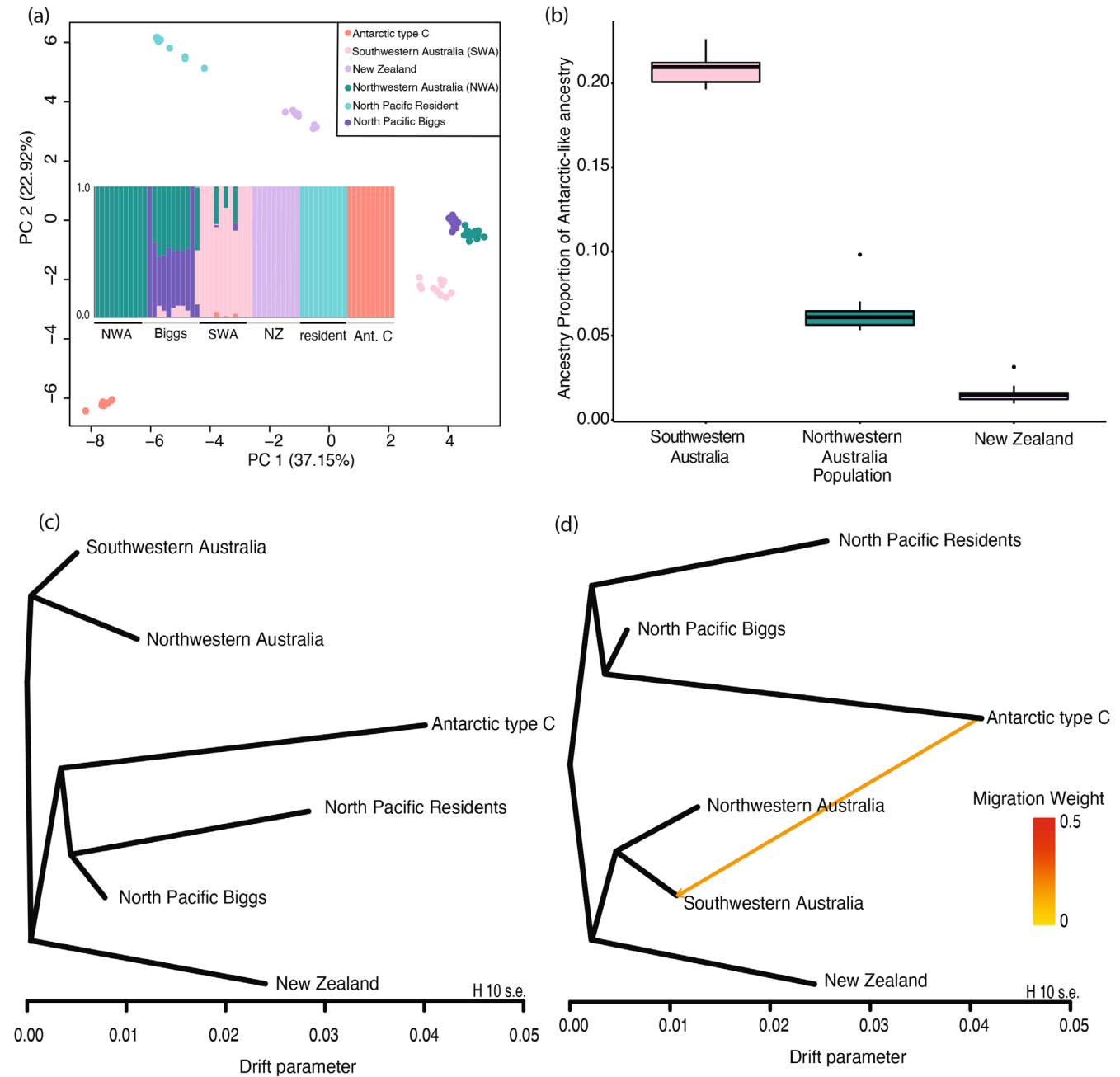
In summary, based on the above results from all  $D$ -statistic tests, there is statistically greater Antarctic-like ancestry in the SWA population relative to the NZ or NWA populations. However, the three Australasian populations share more ancestry with each other than any does with the Antarctic populations.

### 3.3 | Population Assignment

Population assignment was estimated from genotype likelihoods from 18,355 unlinked SNPs. Congruent with Foote et al. (2019), results from Figure 2a (see Figure S5 for additional type B genomes) demonstrate that the greatest differentiated clusters are the Antarctic killer whales along PC1 (37.15% explained variance), with the residents driving the variance along PC2 (22.92% explained variance). This pattern is also consistent when assessing the additional type B1 and B2 genomes, however, variation is

increased along PC1 (56.15%) and decreased along PC2 (14.75%). We used PCANGSD to highlight that SWA shares a higher proportion of Antarctic-like ancestry (one-way ANOVA,  $p < 0.05$ ) relative to NWA and NZ (Figure 2b) based on the variation explained on PC1.

NGSAdmix reflected the same hierarchical structure as seen in Foote et al. (2019) with  $K = 2$  and  $K = 3$  being the best-supported number of genetic clusters (Figures S6 and S7). EvalAdmix highlighted residual differences within populations inferred



**FIGURE 2** | (a) Principal components analysis and NGSAdmix plot ( $K=6$ ) of Australasian killer whale genomes and reference population-level data available from Foote et al. (2016) based upon genotype likelihoods from 18,355 pruned SNPs. Variation along PC1 was used to quantify the (b) ancestry proportion of Antarctic-like ancestry within Australasian populations using genotype likelihoods SNPs. (c) TreeMix maximum likelihood graph from whole-genome sequencing data (1,093,713 unpruned VCF-called SNPs). Horizontal branch lengths are proportional to the amount of genetic drift that has occurred along that branch. The scale bar shows 10 times the average standard error of the entries in the sample covariance matrix. (d) TreeMix maximum likelihood graph with directional migration edge inferred using TreeMix and coloured by migration weight.

by NGSadmix, suggesting that the data violate at least some of the assumptions of the underlying model (Figure S8). These assumptions encompass Hardy–Weinberg equilibrium, random mating, linkage equilibrium, the comprehensive inclusion of all potential source populations, unrelated individuals and the independence of alleles at loci (Ainley and Blight 2009; Garcia-Erill and Albrechtsen 2020). EvalAdmix correlates residuals between true and predicted genotypes to identify violations of the admixture type models. Here, some samples from the Australasian and resident populations have positive residual differences within their population, suggesting high relatedness which was not unexpected based on their small population sizes. Positive residuals may also reflect unrepresented ghost populations. However, three residents (#ID 8–10), from the southern resident cluster have negative residual differences with the rest of the population, likely due to shared population-specific demographic histories, which include inbreeding (see Kardos et al. 2023). Similarly, a single Antarctic type C individual (#ID 3) appears to have negative residuals with the individuals in the cluster and appears to have the lowest coverage genome from the population. The remainder of the samples demonstrate close to zero if not zero residual differences, suggesting that samples generally are a good fit to the NGSadmix model.

### 3.4 | Evidence for Ancient Admixture Between Distinct Genetic Backgrounds

The statistical test of the form  $f_3(\text{SWA}; \text{typeC}, \text{NWA})$  was negative suggesting admixture, but was not statistically significant ( $Z$ -score of  $-2.06$ ; Table S9).  $F_4$ -statistics highlighted the complexity of relationships, for example, with  $f_4(\text{SWA}, \text{typeC}; \text{NWA}; \text{NZ})$  showing significance ( $Z$ -score of  $-17.75$ ; Table S10). However, these tests were less suitable for hypothesis testing in this case than the  $D$ -statistics.

TreeMix graphs reflect that resident, Biggs and Australasian populations cluster together consistently across all tested trees (Figures 1d,e and S9). However, the position of New Zealand and Antarctic populations in the topology changes depending on the number of migration edges. The migration edges support the  $D$ -statistics, with inferred gene flow of Antarctic-like ancestry into SWA and resident-like ancestry into the NZ population. However, it is important to note the residual fit of the observed data versus the predicted squared allele frequency difference demonstrates that none of the topologies perfectly fit the data (Figure S10). TreeMix models are complex, and outputs are for a single best-fit model when there may be several alternative evolutionary histories that fit the data almost as well but which are not outputted by the program. However, the TreeMix results are broadly consistent with the suite of alternative and less parameterized analyses run here. Specifically, the  $D$ -statistics, TreeMix and NGSadmix provide evidence that there is a clear and consistent signal of Antarctic-like ancestry in the SWA genomes arising from admixture.

To examine the extent of ancient admixture ( $n$  generations 100 and 10) we used ELAI to infer local ancestry and its dosage from the reference population into recipient populations.

‘Biggs-like’ ancestry was the major ancestry component within the SWA population, suggesting a recent common ancestor between the SWA and North Pacific Biggs killer populations (Figures 3a and S11). The highest minor ancestry component within the SWA genomes was represented by the Antarctic-like ancestral dosage. The distribution of Antarctic-like ancestry found at high frequencies in the SWA population is in small blocks along the genome, which suggests it is not recent in origin (Figure 3b).

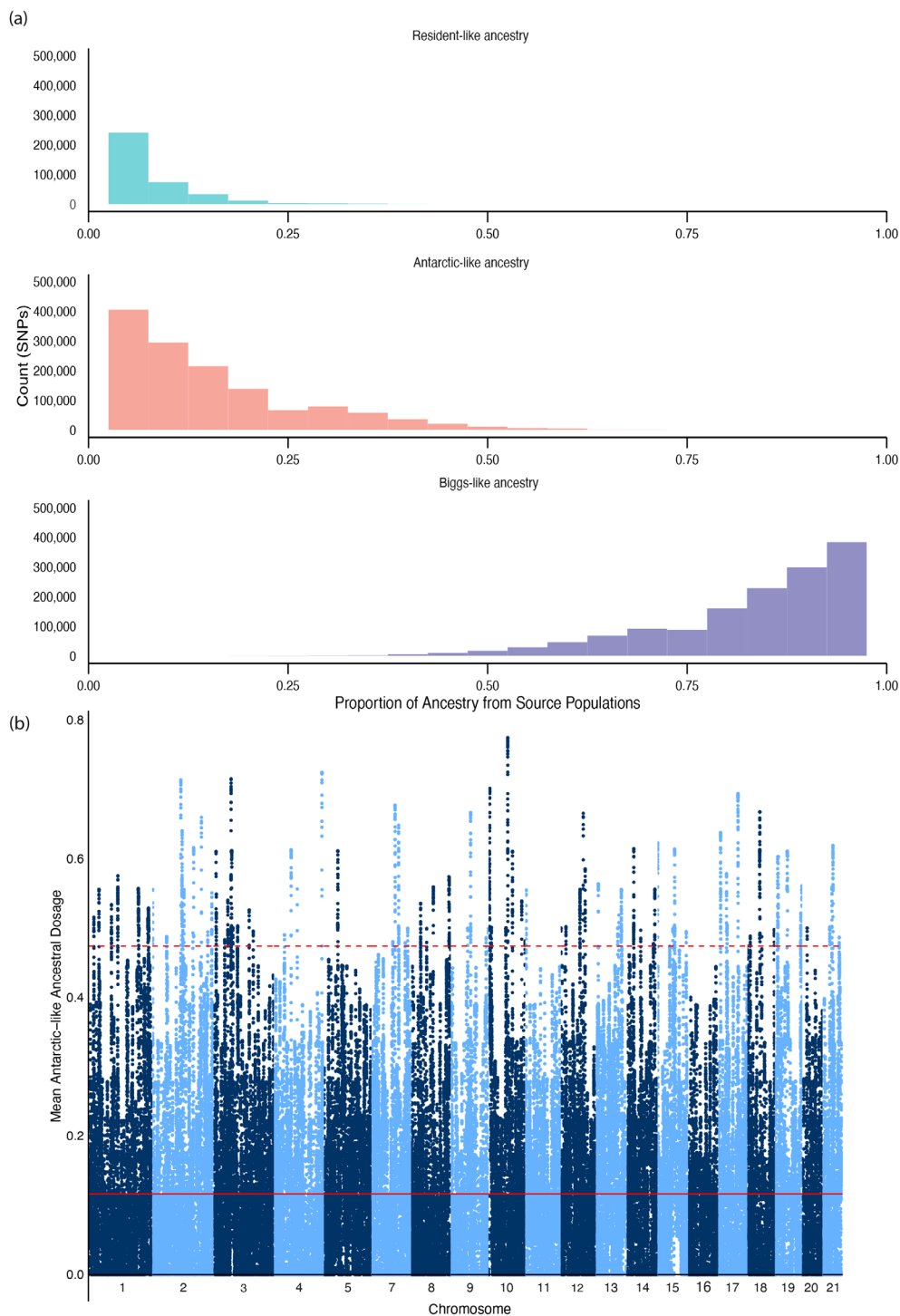
### 3.5 | Evidence for Ongoing or Recent Admixture

To verify if any admixture was recent (i.e., within the last five generations), we focused on three candidate samples (BC30, BC34 and BC47 from southwestern Australia) that through population assignment tests appeared to have ancestry arising from recent admixture (see Figures 2a and S12). These candidate samples' parent admixture demonstrated that all parents were estimated to have consistent admixture patterns, with clear remnants of ‘Antarctic-like’ ancestry (Figure 4). When modelling recent admixture within the last five generations for the candidates we used the three best-fit pedigrees and contrasted them with three not recently admixed samples (BC14, BC28 and BC51) (Table S11, Figure S13). Whilst comparing the three selected candidates that were suggested to not be recently admixed, we find that the inconsistency index and distance to independent pedigrees are marginally higher (Table S8), supporting the lack of recent admixture within these three samples (BC14, BC28, BC51; Figure S13). Figure S14 demonstrates the potential admixture of a single NWA animal; however, it does not support admixture with any of our low-coverage ecotype population data. The admixture index infers that outbreeding occurred between 2.95 and 4.19 generations ago ( $\leq 100$  years based on a generation time of 25.7 years, Taylor et al. (2007)). Therefore, it appears that at least some SWA animals carry ancestry inherited from a great grandparent or great great grandparent with Antarctic-like ancestry.

### 3.6 | Admixture Affects Genetic Diversity

ROHs were estimated to understand how recent admixture may be influencing genetic diversity and  $N_e$  of populations. Admixture can break up haplotypes that form ROH, thereby reducing ROH in the genome (Ceballos et al. 2018) and increasing heterozygosity (Charlesworth 2009). Statistically significant differences in ROH lengths were observed amongst Australasian populations (mean ROH length in kb; NWA:  $565 \pm 25.64\text{SD}$ , NZ:  $536 \pm 12.43\text{SD}$ , SWA:  $489 \pm 20.46\text{SD}$ ; one-way ANOVA,  $p < 0.05$ , see Figure 5a) based on 1000,517 SNPs. We directly compare the three samples that appeared recently admixed (Figure 2a, BC30, BC34 and BC47), with three seemingly unadmixed individuals (BC14, BC28 and BC51) from the population. The mean ROH for recently admixed individuals was 425 kb, compared with 524 kb for other SWA individuals (one-way ANOVA,  $p < 0.05$ ; Figure 5a).

The sum and number of ROH (SROH and NROH, respectively) within individual genomes were highly correlated ( $R = 0.99$ ,  $p = < 2.2 \times 10^{-16}$ ; Figure 5b). SROH and NROH reflect population

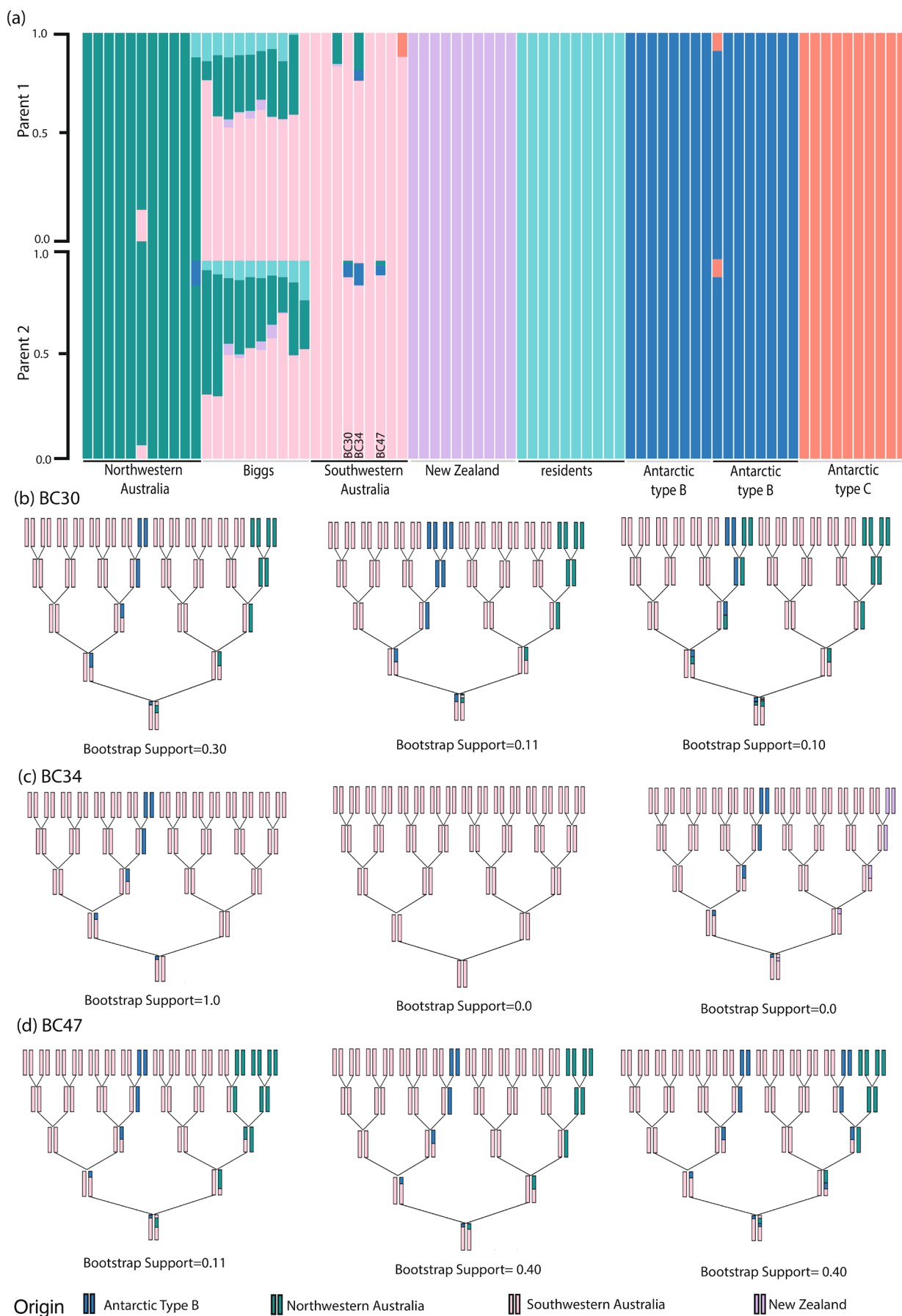


**FIGURE 3** | The distribution of local ancestry in the southwestern Australian population by SNP ( $n = 1000,517$ ) by (a) each of the three reference panel populations for SNP dosage genome-wide and (b) mean Antarctic-like ancestry across the genome per chromosome for 100 mixing generations. The solid red line is the mean ancestry proportion, and the dotted lines are  $\pm 3$  standard deviations based on 50 kb sliding windows and a step size of 10 kb.

demography (Ceballos et al. 2018). Larger populations tend to have lower SROH and NROH, which are further reduced by admixture, particularly between divergent lineages with different haplotypes (Ceballos et al. 2018). In contrast, populations experiencing bottlenecks show shorter and fewer ROHs, as seen in NZ (Foote et al. 2021). As expected, the more admixed the individual, the less the number and sum of ROHs, as evident in BC30 and NI1.

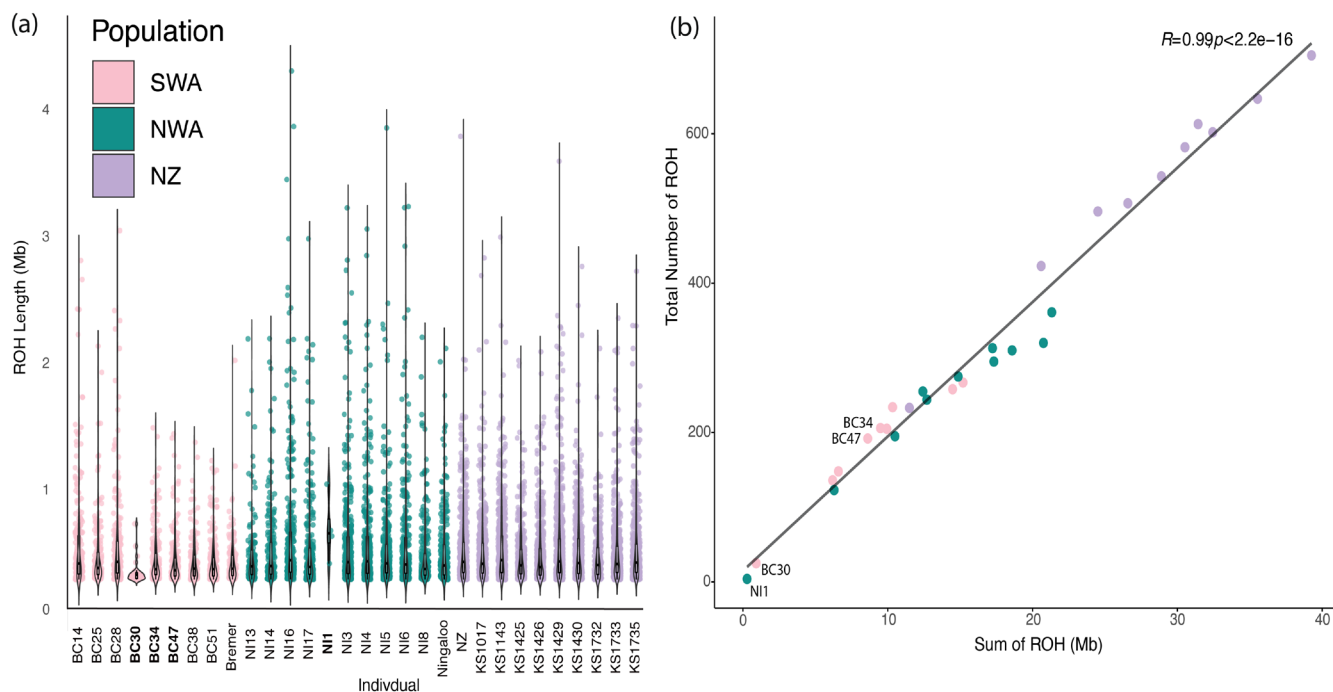
### 3.7 | Potential Signal of Selection of Admixed Regions

When identifying and untangling admixture with southwestern Australian killer whales, we ran simple selection tests to see if any admixed regions may be under selection. We used PCANGSD to elucidate the variance that seems to be separating the Australasian killer whales based on genotype likelihoods estimated at 1,093,713



**FIGURE 4** | Legend on next page.

**FIGURE 4** | Estimates for recent admixture of three candidate admixed southwestern Australia killer whales assuming  $K = 7$  ancestral components. **(a)** Admixture proportions using the parental admixture model in apoh using only the offspring's genotypes for the candidate and **(b)** three most compatible recent admixture pedigrees, including their bootstrap support based on 18,355 SNPs with bootstrapping.



**FIGURE 5** | Genetic diversity estimates of Australasian killer whales. **(a)** Kernel density (violin) plots of the length of individual ROH in each genome based on 1000,517 SNPs. White rectangle shows the interquartile range, and the black bar is the median of the data and **(b)** depicts the number of ROH compared to the sum of the length of ROH across the autosomes 'SWA' is southwestern Australia, 'NWA' is northwestern Australia and 'NZ' New Zealand. The admixed individuals are labelled in bold font, in **(a)** and labelled in **(b)**. Note the 'NZ', 'Ningaloo' and 'Bremer' labelled samples are from Foote et al. (2021).

unlinked SNPs. When looking at unpruned SNPs inferred as evolving under selection along PC1, we see several peaks along the genome (Figure 6), however,  $\lambda$  estimates suggest that this analysis only explains 71% of the observed variance (Figure S15) implying there are factors beyond PC1 driving this. Conversely, when identifying ancestral dosage, there were several regions that could be inferred to be evolving under selection. This is somewhat consistent with findings from PCANGSD (Figure 3 and S10). This suggests potential evidence for the selection of admixed regions; however, these loci remain relatively unexplored here.

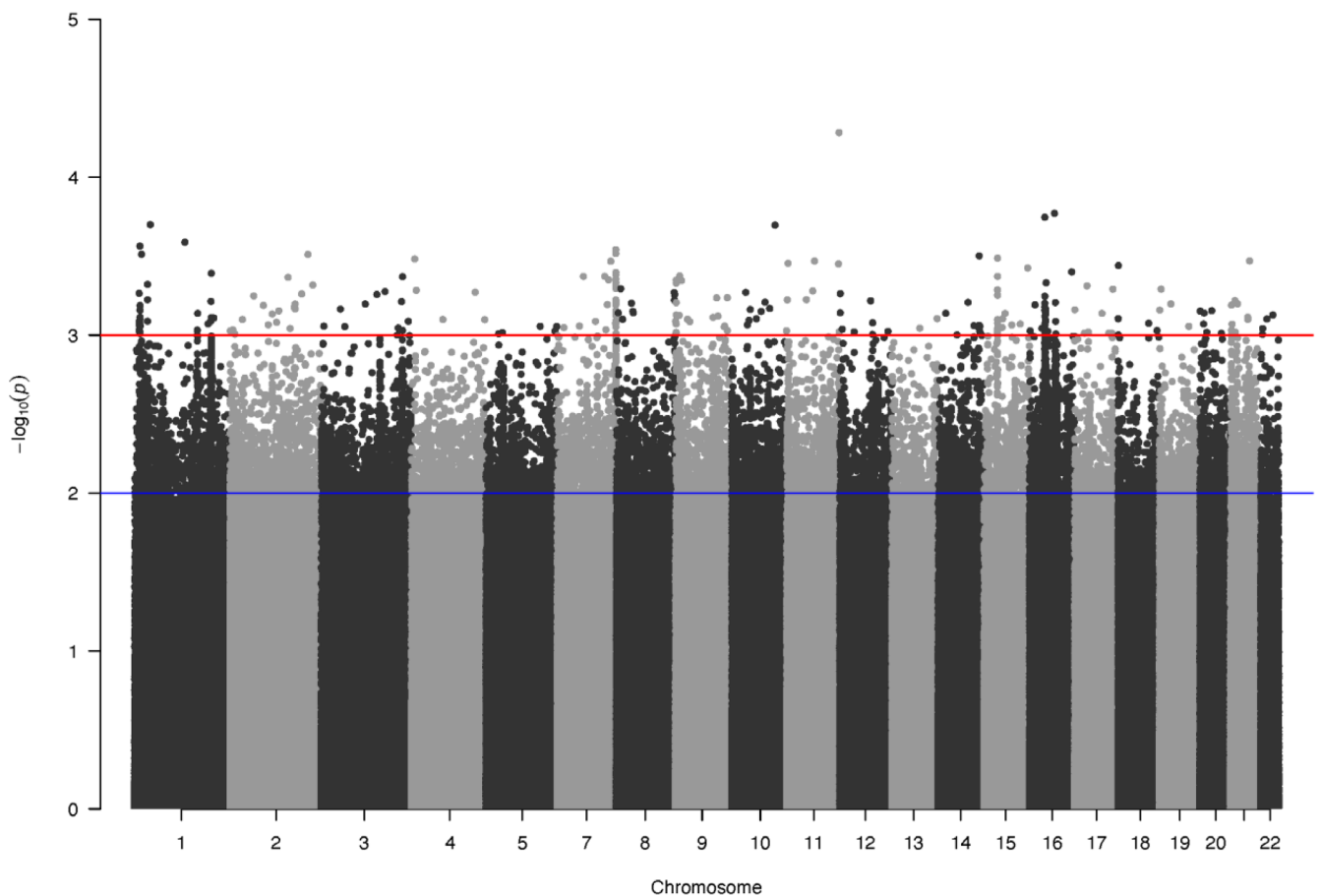
#### 4 | Discussion

This study provides direct empirical evidence that sporadic admixture between different genetic backgrounds increases genetic diversity in low-latitude Australasian killer whale populations. We find that admixture influences present-day genetic variation within and amongst Australasian killer whale populations. An admixture signal is most strongly evident in the SWA population, and Antarctic-like ancestry influences patterns of genetic diversity within the population.

Admixture between two populations with distinct genetic backgrounds can maintain higher  $N_e$  (Charlesworth 2009).

The genetic variation contributed by the donating population may play a role in the maintenance of individual population health. Specifically, admixture between divergent lineages can provide a genetic 'rescue' effect: reducing recessive mutation load (i.e., dominance heterosis) (Fitzpatrick et al. 2020). Longer runs of homozygosity are enriched for homozygous deleterious recessive mutations (Szpiech et al. 2013) and consistent with a genetic rescue effect, we found that recently admixed individuals had fewer long ROH. However, whilst admixture reduces the realised genetic load, it could increase the masked genetic load so that in the event of future inbreeding, introduced deleterious recessive alleles are expressed in homozygous genotypes (Hedrick et al. 2019; Robinson et al. 2019; Szpiech et al. 2019).

Ancestry acts as a source of standing genetic variation and can influence population fitness (Eyre-Walker and Keightley 2007). Introgressed ancestry can increase adaptive potential (Brauer et al. 2023; Fitzpatrick et al. 2020). We found Antarctic-like ancestry at higher than expected frequencies in some genomic regions in the SWA population, consistent with, if not conclusive proof of positive selection on Antarctic ancestry. This highlights the need for further research to understand the evolutionary processes shaping the phenotypic and genetic variation within these populations.



**FIGURE 6** | Selection test for SNPs underpinning genetic variance along PC1 of the PCA in Figure 2. The Manhattan plot is derived from a selection scan using PCAngsd of PC1, showing the log-transformed  $p$ -values per chromosome. The blue line indicated a significance line of 0.01 and the red line was 0.001. The analyses are based on the input of genotype likelihoods at 1,093,713 (non-LD-pruned) SNPs from the combined Australasian and low-coverage ecotype datasets.

Detecting, characterising and untangling admixture between populations remains challenging, especially when the admixing populations are genetically similar (Lawson et al. 2018). Admixture can also become increasingly difficult to identify over time as recombination can erode the length of introgressed tracts (Gopalan et al. 2022). However, recent admixture events can be easier to identify and to characterise their role in shaping the genetic variation of populations by looking for inter-ancestry patterns (Garcia-Erill et al. 2023). This is especially true when admixture is between highly distinct genetic backgrounds (e.g., modern and archaic humans, Racimo et al. (2015)).

Signals of recent admixture are evident in SWA genomes and a single NWA genome but are not detected within the tested NZ genomes. This is consistent with coalescent-based estimates of changes in  $N_e$  (Foote et al. 2021), which combined with other genomic analyses suggest the NZ population to be a small, relatively insular population (Reeves et al. 2022, 2023). In contrast, NWA and SWA populations have relatively stable long-term  $N_e$  with recent increases (Foote et al. 2021), consistent with potential admixture or population expansion. When assessing admixture for the Australian populations, the reference datasets included lineages genetically similar to those that individuals within the SWA population have admixed with (i.e., Antarctic

type C killer whales). Admixture in the genome of NI1 from the NWA population was detected, but the source population was not identified. This is likely due to ancestry being donated from an unsampled lineage not represented within our reference dataset, highlighting the need for increased sampling and sequencing of additional killer whale populations to fully understand the scale of connectivity.

Despite the challenges of identifying ancestry in genomes, we had multiple lines of evidence of sporadic admixture between SWA and Antarctic-like lineages. Antarctic killer whales have highly divergent genetic backgrounds from other killer whales. Their genomes carry an excess of private (relative to non-Antarctic killer whales) alleles, some clustered within 'archaic' tracts (regions of outlier older TMRCA relative to the genome-wide mean), which contribute to the genetic differentiation between them and other killer whale populations (Foote et al. 2021, 2019). The gene flow from Antarctic-like lineages represents ~20% of the ancestry of the SWA population, with varying dosages along the genome suggesting ancestral and ongoing gene flow between these lineages.

In summary, this study provides novel insight into the evolutionary history of Australasian killer whales. Analyses of high-coverage genomes of these low-latitude killer whales provide

evidence that sporadic admixture with distinct genetic lineages is helping to maintain genetic diversity. Thus, admixture could promote genetic rescue within populations of Australasian killer whales. Sporadic gene flow between killer whale populations with distinct genetic backgrounds, as reported here, has been suggested as a broader pattern of behaviour that could explain why populations at low latitudes, low killer whale density and low productivity are able to maintain genetic diversity equal to, or even greater than, in high latitude, high productivity regions with high killer whale density (Foote et al. 2021). Whilst this requires testing (as done here) across more populations to confirm the generality of this pattern amongst killer whale populations, our study does provide support for the concept. More broadly, our study highlights the importance of maintaining connectivity between small populations to their genetic health and persistence.

### Author Contributions

A.D.F. and I.M.R. conceived the study. J.A.T., K.A.S. and E.L.B. provided samples from the contemporary Australasian populations. I.M.R. conducted laboratory work. I.M.R. conducted genomic and bioinformatic analysis with guidance from A.D.F. and J.S.-C., I.M.R. wrote the manuscript with input from all coauthors.

### Acknowledgements

We thank Subject Editor Marty Kardos, Genís Garcia-Erill and an anonymous reviewer for their constructive feedback on an earlier draft of this manuscript. Biopsy samples were collected under approval of the Flinders University Animal Ethics Committee (Approval number: E460/17), Department of Agriculture, Environment (Biological resources permit approval number: AU-COM2016-317/AU-COM 2019-459; Cetacean research: Approval number CP-2013-0003/CP 2018-0017) and western Australian Department of Environment and Conservation (Approval number: DPAW 08-000870-1). New Zealand tissue samples were collected under research permit 39239-MAR issued to Massey University by the Department of Conservation Te Papa Atawhai, and with the support of mana whenua, shared with Flinders University via CITES Scientific transfer in August 2019. We are particularly indebted to Ngati Whatua, Oraka-Aparima Rūnaka, Te Whānau a Kauaetangohia and Te Rūnanga o Te Whānau ā Apanui for entrusting Massey University with post-mortem tissue collection from their stranded taonga. Particular thanks are owed to Glenn Wilcox, Riki Dallas, Iain MacCullum, Rika Mato and Rawiri Waititi, respectively, for their ongoing support of this mahi. Additional thanks are further extended to Ramari Oliphant Stewart (Kauika Tangaroa Charitable Trust), Ros Cole (Department of Conservation) and Stu Hunter (Massey University) for facilitating tissue collection and supporting this kaupapa. Open access publishing facilitated by Flinders University, as part of the Wiley - Flinders University agreement via the Council of Australian University Librarians.

### Conflicts of Interest

The authors declare no conflicts of interest.

### Data Availability Statement

Sequencing data generated from Australian samples for this study has been archived at the National Centre for Biotechnology Information ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) under the Orconomics BioProject (NCBI accession: PRJNA531206). Sequencing data generated from New Zealand samples have been uploaded to the Genomics Aotearoa Repository (<https://www.genomics-aotearoa.org.nz>; accession: <https://doi.org/10.57748/3vba-rg35>).

### References

- Ainley, D. G., and L. K. Blight. 2009. "Ecological Repercussions of Historical Fish Extraction From the Southern Ocean." *Fish and Fisheries* 10, no. 1: 13–38. <https://doi.org/10.1111/j.1467-2979.2008.00293.x>.
- Brauer, C. J., J. Sandoval-Castillo, K. Gates, et al. 2023. "Natural Hybridization Reduces Vulnerability to Climate Change." *Nature Climate Change* 13, no. 3: 282–289. <https://doi.org/10.1038/s41558-022-01585-1>.
- Ceballos, F. C., S. Hazelhurst, and M. Ramsay. 2018. "Assessing Runs of Homozygosity: A Comparison of SNP Array and Whole Genome Sequence Low Coverage Data." *BMC Genomics* 19: 1–12. <https://doi.org/10.1186/s12864-018-4489-0>.
- Charlesworth, B. 2009. "Effective Population Size and Patterns of Molecular Evolution and Variation." *Nature Reviews Genetics* 10, no. 3: 195–205. <https://doi.org/10.1038/nrg2526>.
- Charlesworth, D., and J. H. Willis. 2009. "The Genetics of Inbreeding Depression." *Nature Reviews Genetics* 10, no. 11: 783–796. <https://doi.org/10.1038/nrg2664>.
- Coop, G. 2023. "Genetic Similarity and Genetic Ancestry Groups as Sample Descriptors in Human Genetics." arXiv preprint arXiv:2207.11595.
- Danecek, P., S. Schiffels, and R. Durbin. 2016. "Multiallelic Calling Model in Bcftools (-m)." <https://samtools.github.io/bcftools/call-m.pdf>.
- Durand, E. Y., N. Patterson, D. Reich, and M. Slatkin. 2011. "Testing for Ancient Admixture Between Closely Related Populations." *Molecular Biology and Evolution* 28, no. 8: 2239–2252. <https://doi.org/10.1093/molbev/msr048>.
- Garcia-Erill, G., K. Hanghøj, R. Heller, C. Wiuf, and A. Albrechtsen. 2023. "Estimating Admixture Pedigrees of Recent Hybrids Without a Contiguous Reference Genome." *Molecular Ecology Resources* 23, no. 7: 1604–1619. <https://doi.org/10.1111/1755-0998.13830>.
- Excoffier, L., M. Foll, and R. J. Petit. 2009. "Genetic Consequences of Range Expansions." *Annual Review of Ecology, Evolution, and Systematics* 40: 481–501. <https://doi.org/10.1146/annurev.ecolsys.39.110707.173414>.
- Eyre-Walker, A., and P. D. Keightley. 2007. "The Distribution of Fitness Effects of New Mutations." *Nature Reviews Genetics* 8, no. 8: 610–618. <https://doi.org/10.1038/nrg2146>.
- Facon, B., P. Jarne, J. Pointier, and P. David. 2005. "Hybridization and Invasiveness in the Freshwater Snail *Melanoides Tuberculata*: Hybrid Vigour Is More Important Than Increase in Genetic Variance." *Journal of Evolutionary Biology* 18, no. 3: 524–535. <https://doi.org/10.1111/j.1420-9101.2005.00887.x>.
- Fitzpatrick, S. W., G. S. Bradburd, C. T. Kremer, P. E. Salerno, L. M. Angeloni, and W. C. Funk. 2020. "Genomic and Fitness Consequences of Genetic Rescue in Wild Populations." *Current Biology* 30, no. 3: 517–522. <https://doi.org/10.1016/j.cub.2019.11.062>.
- Flinders University. 2021. "DeepThought (HPC)." <https://doi.org/10.25957/FLINDERS.HPC.DEEPTHOUGHT>.
- Fontseré, C., M. de Manuel, T. Marques-Bonet, and M. Kuhlwilm. 2019. "Admixture in Mammals and How to Understand Its Functional Implications: On the Abundance of Gene Flow in Mammalian Species, Its Impact on the Genome, and Roads Into a Functional Understanding." *BioEssays* 41, no. 12: 1900123. <https://doi.org/10.1002/bies.201900123>.
- Foote, A. D., A. Alexander, L. T. Ballance, et al. 2023. "'Type D' Killer Whale Genomes Reveal Long-Term Small Population Size and Low Genetic Diversity." *Journal of Heredity* 114, no. 2: 94–109. <https://doi.org/10.1093/jhered/esac070>.
- Foote, A., and P. Bunskoek, Wellcome Sanger Institute Tree of Life Programme, & Darwin Tree of Life Consortium, Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective, Tree of Life

- Core Informatics collective, and Darwin Tree of Life Consortium. 2022. "The Genome Sequence of the Killer Whale, *Orcinus orca* (Linnaeus, 1758)." *Wellcome Open Research* 7: 250.
- Foote, A. D., R. Hooper, A. Alexander, et al. 2021. "Runs of Homozygosity in Killer Whale Genomes Provide a Global Record of Demographic Histories." *Molecular Ecology* 30: 6162–6177. <https://doi.org/10.1111/mec.16137>.
- Foote, A. D., Y. Liu, G. W. Thomas, et al. 2015. "Convergent Evolution of the Genomes of Marine Mammals." *Nature Genetics* 47, no. 3: 272. <https://doi.org/10.1038/ng.3198>.
- Foote, A. D., M. Martin, M. Louis, et al. 2019. "Killer Whale Genomes Reveal a Complex History of Recurrent Admixture and Vicariance." *Molecular Ecology* 28, no. 14: 3427–3444. <https://doi.org/10.1111/mec.15099>.
- Foote, A. D., N. Vijay, M. C. Ávila-Arcos, et al. 2016. "Genome-Culture Coevolution Promotes Rapid Divergence of Killer Whale Ecotypes." *Nature Communications* 7, no. 1: 11693. <https://doi.org/10.1038/ncomm511693>.
- Forney, K., and P. Wade. 2006. "World-Wide Abundance and Density of Killer Whales." In *Whales, Whaling, and Ocean Ecosystems*. University of California Press.
- Galaverni, M., R. Caniglia, L. Pagani, E. Fabbri, A. Boattini, and E. Randi. 2017. "Disentangling Timing of Admixture, Patterns of Introgression, and Phenotypic Indicators in a Hybridizing Wolf Population." *Molecular Biology and Evolution* 34, no. 9: 2324–2339. <https://doi.org/10.1093/molbev/msx169>.
- Galinsky, K. J., G. Bhatia, P.-R. Loh, et al. 2016. "Fast Principal-Component Analysis Reveals Convergent Evolution of ADH1B in Europe and East Asia." *American Journal of Human Genetics* 98, no. 3: 456–472. <https://doi.org/10.1371/journal.pgen.1007741>.
- Garcia-Erill, G., and A. Albrechtsen. 2020. "Evaluation of Model Fit of Inferred Admixture Proportions." *Molecular Ecology Resources* 20, no. 4: 936–949. <https://doi.org/10.1111/1755-0998.13171>.
- Garroway, C. J., E. de Greef, K. J. Lefort, et al. 2024. "Climate Change Introduces Threatened Killer Whale Populations and Conservation Challenges to the Arctic." *Global Change Biology* 30, no. 6: e17352. <https://doi.org/10.1111/gcb.17352>.
- Gopalan, S., S. P. Smith, K. Korunes, I. Hamid, S. Ramachandran, and A. Goldberg. 2022. "Human Genetic Admixture Through the Lens of Population Genomics." *Philosophical Transactions of the Royal Society B* 377, no. 1852: 20200410. <https://doi.org/10.1098/rstb.2020.0410>.
- Green, R. E., J. Krause, A. W. Briggs, et al. 2010. "A Draft Sequence of the Neandertal Genome." *Science* 328, no. 5979: 710–722. <https://doi.org/10.1126/science.1188021>.
- Guan, Y. 2014. "Detecting Structure of Haplotypes and Local Ancestry." *Genetics* 196, no. 3: 625–642. <https://doi.org/10.1534/genetics.113.160697>.
- Hamid, I., K. L. Korunes, D. R. Schrider, and A. Goldberg. 2023. "Localizing Post-Admixture Adaptive Variants With Object Detection on Ancestry-Painted Chromosomes." *Molecular Biology and Evolution* 40: 4. <https://doi.org/10.1093/molbev/msad074>.
- Hedrick, P., J. Robinson, R. O. Peterson, and J. A. Vucetich. 2019. "Genetics and Extinction and the Example of Isle Royale Wolves." *Animal Conservation* 22, no. 3: 302–309. <https://doi.org/10.1111/acv.12479>.
- Hoelzel, A. R., J. Hey, M. E. Dahlheim, C. Nicholson, V. Burkanov, and N. Black. 2007. "Evolution of Population Structure in a Highly Social Top Predator, the Killer Whale." *Molecular Biology and Evolution* 24, no. 6: 1407–1415. <https://doi.org/10.1093/molbev/msm063>.
- Kardos, M., Y. Zhang, K. M. Parsons, et al. 2023. "Inbreeding Depression Explains Killer Whale Population Dynamics." *Nature Ecology & Evolution* 7, no. 5: 675–686. <https://doi.org/10.1038/s41559-023-01995-0>.
- Korneliussen, T. S., A. Albrechtsen, and R. Nielsen. 2014. "ANGSD: Analysis of Next Generation Sequencing Data." *BMC Bioinformatics* 15, no. 1: 1–13. <https://doi.org/10.1186/s12859-014-0356-4>.
- Korunes, K. L., and A. Goldberg. 2021. "Human Genetic Admixture." *PLoS Genetics* 17, no. 3: e1009374.
- Kyriazis, C. C., L. E. Serieys, J. M. Bishop, et al. 2024. "The Influence of Gene Flow on Population Viability in an Isolated Urban Caracal Population." *Molecular Ecology* 33, no. 9: e17346.
- Lawson, D., L. van Dorp, and D. Falush. 2018. "A Tutorial on How Not to Over-Interpret STRUCTURE and ADMIXTURE Bar Plots." *Nature Communications* 9, no. 1: 3258. <https://doi.org/10.1038/s41467-018-05257-7>.
- Lê, S., J. Josse, and F. Husson. 2008. "FactoMineR: An R Package for Multivariate Analysis." *Journal of Statistical Software* 25, no. 1: 1–18. <https://doi.org/10.18637/jss.v025.i01>.
- Li, H., and R. Durbin. 2009. "Fast and Accurate Short Read Alignment With Burrows–Wheeler Transform." *Bioinformatics* 25, no. 14: 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.
- Li, H., B. Handsaker, A. Wysoker, et al. 2009. "The Sequence Alignment/Map Format and SAMtools." *Bioinformatics* 25, no. 16: 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
- Lotterhos, K. E. 2019. "The Effect of Neutral Recombination Variation on Genome Scans for Selection." *G3: Genes, Genomes, Genetics* 9, no. 6: 1851–1867. <https://doi.org/10.1016/j.cub.2019.11.062>.
- Lou, R. N., and N. O. Therkildsen. 2022. "Batch Effects in Population Genomic Studies With Low-Coverage Whole Genome Sequencing Data: Causes, Detection and Mitigation." *Molecular Ecology Resources* 22, no. 5: 1678–1692. <https://doi.org/10.1111/1755-0998.13559>.
- Lynch, M. 1991. "The Genetic Interpretation of Inbreeding Depression and Outbreeding Depression." *Evolution* 45, no. 3: 622–629. <https://doi.org/10.1111/j.1558-5646.1991.tb04333.x>.
- Mathieson, I., and A. Scally. 2020. "What Is Ancestry?" *PLoS Genetics* 16, no. 3: e1008624. <https://doi.org/10.1371/journal.pgen.1008624>.
- Meisner, J., and A. Albrechtsen. 2018. "Inferring Population Structure and Admixture Proportions in Low-Depth NGS Data." *Genetics* 210, no. 2: 719–731. <https://doi.org/10.1002/bies.201900123>.
- Meisner, J., A. Albrechtsen, and K. Hanghøj. 2021. "Detecting Selection in Low-Coverage High-Throughput Sequencing Data Using Principal Component Analysis." *BMC Bioinformatics* 22, no. 470: 1–13. <https://doi.org/10.1186/s12859-021-04375-2>.
- Morin, P. A., K. M. Parsons, F. I. Archer, et al. 2015. "Geographic and Temporal Dynamics of a Global Radiation and Diversification in the Killer Whale." *Molecular Ecology* 24, no. 15: 3964–3979. <https://doi.org/10.1111/mec.13284>.
- Moura, A., J. Kenny, R. Chaudhuri, et al. 2015. "Phylogenomics of the Killer Whale Indicates Ecotype Divergence in Sympatry." *Heredity* 114, no. 1: 48. <https://doi.org/10.1038/hdy.2014.67>.
- Patterson, N., P. Moorjani, Y. Luo, et al. 2012. "Ancient Admixture in Human History." *Genetics* 192, no. 3: 1065–1093. <https://doi.org/10.1534/genetics.112.145037>.
- Peter, B. M. 2016. "Admixture, Population Structure, and F-Statistics." *Genetics* 202, no. 4: 1485–1501. <https://doi.org/10.1371/journal.pgen.1007741>.
- Pickrell, J., and J. Pritchard. 2012. "Inference of Population Splits and Mixtures From Genome-Wide Allele Frequency Data." *Nature Precedings* 1: 1. <https://doi.org/10.1038/npre.2012.6956.1>.
- Purcell, S., B. Neale, K. Todd-Brown, et al. 2007. "PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses." *American Journal of Human Genetics* 81, no. 3: 559–575. <https://doi.org/10.1086/519795>.

Quinlan, A. R. 2014. "BEDTools: The Swiss-Army Tool for Genome Feature Analysis." *Current Protocols in Bioinformatics* 47, no. 1: 11.12.11–11.12.34. <https://doi.org/10.1002/0471250953.b11112s47>.

Racimo, F., S. Sankararaman, R. Nielsen, and E. Huerta-Sánchez. 2015. "Evidence for Archaic Adaptive Introgression in Humans." *Nature Reviews Genetics* 16, no. 6: 359–371. <https://doi.org/10.1038/nrg3936>.

RCoreTeam. 2022. *R: A Language and Environment for Statistical Computing, Version 4.2.1*. R Foundation for Statistical Computing.

Reeves, I. M., J. A. Totterdell, A. Barcelo, et al. 2022. "Population Genomic Structure of Killer Whales (*Orcinus orca*) in Australian and New Zealand Waters." *Marine Mammal Science* 38, no. 1: 151–174. <https://doi.org/10.1111/mms.12851>.

Reeves, I. M., J. A. Totterdell, E. L. Betty, et al. 2023. "Ancestry Testing of "Old tom," a Killer Whale Central to Mutualistic Interactions With Human Whalers." *Journal of Heredity* 114, no. 6: 598–611. <https://doi.org/10.1093/jhered/esad058>.

Reich, D., K. Thangaraj, N. Patterson, A. L. Price, and L. Singh. 2009. "Reconstructing Indian Population History." *Nature* 461, no. 7263: 489–494. <https://doi.org/10.1038/nature08365>.

Robinson, J. A., J. Räikkönen, L. M. Vucetich, et al. 2019. "Genomic Signatures of Extensive Inbreeding in Isle Royale Wolves, a Population on the Threshold of Extinction." *Science Advances* 5, no. 5: eaau0757. <https://doi.org/10.1126/sciadv.aau0757>.

Saremi, N. F., M. A. Supple, A. Byrne, et al. 2019. "Puma Genomes from North and South America Provide Insights into the Genomic Consequences of Inbreeding." *Nature Communications* 10, no. 1: 4769.

Schubert, M., S. Lindgreen, and L. Orlando. 2016. "AdapterRemoval v2: Rapid Adapter Trimming, Identification, and Read Merging." *BMC Research Notes* 9, no. 1: 88. <https://doi.org/10.1186/s13104-016-1900-2>.

Skotte, L., T. S. Korneliussen, and A. Albrechtsen. 2013. "Estimating Individual Admixture Proportions From Next Generation Sequencing Data." *Genetics* 195, no. 3: 693–702. <https://doi.org/10.1534/genetics.113.154138>.

Smit, A., R. Hubley, and P. Green. 2004. "RepeatMasker Open-3.0." <http://www.repeatmasker.org>.

Studio, R. 2016. Boston, MA, USA.

Sunnucks, P., and D. F. Hales. 1996. "Numerous Transposed Sequences of Mitochondrial Cytochrome Oxidase I-II in Aphids of the Genus *Sitobion* (Hemiptera: Aphididae)." *Molecular Biology and Evolution* 13, no. 3: 510–524.

Szpiech, Z. A., A. C. Mak, M. J. White, et al. 2019. "Ancestry-Dependent Enrichment of Deleterious Homozygotes in Runs of Homozygosity." *American Journal of Human Genetics* 105, no. 4: 747–762. <https://doi.org/10.1016/j.ajhg.2019.08.011>.

Szpiech, Z. A., J. Xu, T. J. Pemberton, et al. 2013. "Long Runs of Homozygosity Are Enriched for Deleterious Variation." *American Journal of Human Genetics* 93, no. 1: 90–102. <https://doi.org/10.1016/j.ajhg.2013.05.003>.

Tan, T., and E. G. Atkinson. 2023. "Strategies for the Genomic Analysis of Admixed Populations." *Annual Review of Biomedical Data Science* 6: 105–127. <https://doi.org/10.1146/annurev-biodatasci-020722-014310>.

Taylor, B. L., S. J. Chivers, J. Larese, and W. F. Perrin. 2007. "Generation Length and Percent Mature Estimates for IUCN Assessments of Cetaceans." IUCN Assessments of Cetaceans.

Turner, S. D. 2014. "qqman: An R Package for Visualizing GWAS Results Using QQ and Manhattan Plots." *BioRxiv*. no. 005165. <https://doi.org/10.1101/005165>.

Verhoeven, K. J., M. Macel, L. M. Wolfe, and A. Biere. 2011. "Population Admixture, Biological Invasions and the Balance Between Local Adaptation and Inbreeding Depression." *Proceedings of the Royal*

*Society B: Biological Sciences* 278, no. 1702: 2–8. <https://doi.org/10.1098/rspb.2010.1272>.

## Supporting Information

Additional supporting information can be found online in the Supporting Information section.