Received: 2 June 2023 | Revised: 20 December 2023 | Accepted: 4 January 2024

DOI: 10.1111/mec.17282

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Whole genome sequencing reveals stepping-stone dispersal buffered against founder effects in a range expanding seabird

Rachael W. Herman[1] | Gemma Clucas[2] | Jane Younger[3] | John Bates[4] | Bryce Robinson[2] | Sushma Reddy[5] | Julia Stepanuk[1] | Katie O'Brien[6] | Krishna Veeramah[1] | Heather J. Lynch[1,7]

[1]Department of Ecology and Evolution, Stony Brook University, Stony Brook, New York, USA

[2]Cornell Lab of Ornithology, Cornell University, Ithaca, New York, USA

[3]Institute for Marine and Antarctic Studies, University of Tasmania, Hobart, Tasmania, Australia

[4]Negaunee Integrative Research Center, The Field Museum of Natural History, Chicago, Illinois, USA

[5]Bell Museum of Natural History and Department of Fisheries, Wildlife and Conservation Biology, University of Minnesota, St. Paul, Minnesota, USA

[6]Milner Centre for Evolution, University of Bath, Bath, UK

[7]Institute for Advanced Computational Sciences, Stony Brook University, Stony Brook, New York, USA

Correspondence
Rachael W. Herman, Department of Ecology and Evolution, Stony Brook University, Nicholls Road, Stony Brook NY, USA.
Email: rachael.herman@stonybrook.edu

Funding information
NASA Earth Sciences Division, Grant/Award Number: 80NSSC20K1605; The Institute for Advanced Computational Science (Stony Brook University); National Geographic Society; Field Museum's Robert A. Pritzker Center for Meteoritics and Polar Studies

Handling Editor: Paul A. Hohenlohe

Abstract

Many species are shifting their ranges in response to climate-driven environmental changes, particularly in high-latitude regions. However, the patterns of dispersal and colonization during range shifting events are not always clear. Understanding how populations are connected through space and time can reveal how species navigate a changing environment. Here, we present a fine-scale population genomics study of gentoo penguins (*Pygoscelis papua*), a presumed site-faithful colonial nesting species that has increased in population size and expanded its range south along the Western Antarctic Peninsula. Using whole genome sequencing, we analysed 129 gentoo penguin individuals across 12 colonies located at or near the southern range edge. Through a detailed examination of fine-scale population structure, admixture, and population divergence, we inferred that gentoo penguins historically dispersed rapidly in a stepping-stone pattern from the South Shetland Islands leading to the colonization of Anvers Island, and then the adjacent mainland Western Antarctica Peninsula. Recent southward expansion along the Western Antarctic Peninsula also followed a stepping-stone dispersal pattern coupled with limited post-divergence gene flow from colonies on Anvers Island. Genetic diversity appeared to be maintained across colonies during the historical dispersal process, and range-edge populations are still growing. This suggests large numbers of migrants may provide a buffer against founder effects at the beginning of colonization events to maintain genetic diversity similar to that of the source populations before migration ceases post-divergence. These results coupled with a continued increase in effective population size since approximately 500–800 years ago distinguish gentoo penguins as a robust species that is highly adaptable and resilient to changing climate.

KEYWORDS
colonization, founder effects, range expansion, seabirds, stepping-stone dispersal, whole genome sequencing

Krishna Veeramah and Heather J. Lynch should be considered joint senior author.

1 | INTRODUCTION

Species-level range shifts are ongoing phenomena that often occur during periods of changing climate. While many shifts in species distributions are attributed to the last deglaciation period approximately 10,000 years ago (Van der Putten, 2012), ongoing rapid climate perturbations are accelerating such events (Germain & Lutz, 2020; Gervais et al., 2021; Sekercioglu et al., 2008). As sea temperatures increase, marine species are shifting their range boundaries poleward as current ranges become unsuitable, and higher latitude waters provide new refugia of suitable habitat (Perry et al., 2005; Pinsky et al., 2019; Poloczanska et al., 2016). In the Southern Ocean, however, we are also observing poleward range expansions in species that are not necessarily shifting their ranges all together, but rather capitalizing on new suitable habitat. For example, king penguins (*Aptenodytes patagonicus*) have colonized areas on the South Shetland Islands, south of their previous range edge, while maintaining historical ranges and exhibiting increases in breeding population size (Foley et al., 2018; Péron et al., 2012; Petry et al., 2013). Meanwhile, king crabs (*Paralomis birsteini*) have colonized the Antarctic continental shelf, a region historically too cold for this species, posing substantial threats to the benthic ecosystem (Aronson et al., 2015; Smith et al., 2012). However, the dispersal patterns involved in these colonization events, particularly in vertebrates, are often not known.

Gentoo penguins (*Pygoscelis papua*), another historically sub-Antarctic distributed species, have increased in population size and are expanding their range south along the Western Antarctic Peninsula (WAP) (Herman et al., 2020; Lynch et al., 2012), where declines in sea ice and increases in precipitation have been linked to a suite of oceanographic and ecological changes over the last 40 years (Lin et al., 2021; Meredith et al., 2017; Turner et al., 2005, 2013). Gentoo penguin populations have grown most quickly in areas with declining sea ice during the pre-breeding season (Lynch et al., 2012), suggesting sea ice concentrations may be one environmental feature driving range expansion in this species.

Eight new colonies of gentoo penguins have been established on the WAP in the last 30 years concurrent with local sea ice decline (Herman et al., 2020). Continued immigration of individuals has been shown to sustain population growth at least during the early colonization phase of such colonies (Herman & Lynch, 2022). This movement of individuals was estimated to account for approximately 7.6% of the entire gentoo penguin population on the WAP, a scale of ongoing range expansion that is to our knowledge unprecedented in this region. This represented a rare opportunity to study the dispersal patterns of range expansion facilitated by changing climate in real time.

Despite the scale of dispersal involved in this recent southward expansion, the source populations of individuals and dispersal pathways are currently unknown. Identifying source populations can help to determine the causes and mechanisms of range shifts, as well as habitat features that facilitate or restrict dispersal. However, real-time observations of dispersal and colonization events can be difficult, particularly in highly mobile, pelagic species such as penguins. While advances in tracking technology have improved our ability to trace the movement of individuals of highly mobile species (Hindell et al., 2020), these methods are not easily applied when source populations are unknown.

Population genomics has been effective at investigating gene flow among populations of seabirds (Clucas et al., 2018; Cristofari et al., 2018; Herman et al., 2022; Kersten et al., 2021), but very few genomics studies address dispersal and colonization patterns (Friesen et al., 2007). Colonial seabirds provide a natural model system to examine the relative roles that genetic divergence and gene flow play in population structure. Seabirds breed along coastlines, island clusters, and island chains, where migration models such as the n-island (random migration among populations independent of distance), continent-island (migration from a source to n-islands), and the stepping-stone models (migration restricted to adjacent habitat patches) can be tested (Kimura & Weiss, 1964; Slatkin, 1993). Analogous models have been proposed to describe seabird colonization, which could take place via geographically adjacent stepping-stones or be unaffected by geographic distance and disperse to form multiple new populations from a single source (Munilla et al., 2016). The patchy colony distribution and recent range expansion of gentoo penguins on the WAP facilitates the application of robust genomic methods to examine historical patterns of colonization and current patterns of migration and identify any barriers to gene flow.

Here, we present a fine-scale population genomics study of 129 gentoo penguin individuals across 12 colonies along the WAP using whole genome sequencing, where we explicitly contrast the fit of our data to (a) a stepping-stone model versus (b) a continental-island model of dispersal. To our knowledge, this is the first study to use whole genome-wide approaches to investigate fine-scale population structure, demographic history, and dispersal in a top marine predator in the Southern Hemisphere. Our detailed examination of a species undergoing rapid and continuing range expansion reveals a pattern of historical and current dispersal that reflects a modified stepping-stone model with some post-divergence gene flow and buffering against founder effects.

2 | METHODS

2.1 | Sample collection

All sampling protocols and procedures employed were ethically reviewed and approved by Stony Brook University's Institutional Animal Care and Use Committee (IACUC). During the 2019/2020 and 2021/2022 breeding seasons, we collected blood samples from 129 breeding gentoo penguins across 12 colonies along the Western Antarctic Peninsula. Nine of these locations represent potential source colonies for individuals dispersing to recently colonized breeding locations: Damoy Point, Danco Island, Neko Harbour, Jougla Point, Port Charcot, Joubin Islands, Gerlache Island, Hannah

Point, and Petermann Island (Figure 1a, circles). Fifteen samples were collected at each of these sites, except for Damoy Point and Hannah Point where we collected ten samples from each, and Petermann Island where we collected three samples. We also collected samples from three newly established colonies: three samples from Moot Point, ten samples from Tuxen Rocks, and five samples from Noble Rocks (Figure 1a, squares). In addition, we collected a blood sample from a non-breeding individual at Armstrong Reef, a location approximately 75 km to the south of their southernmost range limit (Figure 1a). We determined breeding status by observing incubation or chick brooding behaviour to confirm targeted adults were a member of a specific colony. 1–2 mL of blood was then collected for each individual and stored in ethanol or DNA/RNA shield at −20°C.

2.2 | DNA extraction, sequencing, and bioinformatic processing

DNA extractions were performed on collected blood using Qiagen DNEasy blood and tissue kits. Whole genome resequencing with 150 bp paired-end was conducted at Beijing Genomics Institute (BGI) using their DNBseq technology with a target average coverage of ~20×. We aligned raw sequence reads to an indexed gentoo penguin reference genome (Pan et al., 2019) using BWA (Li & Durbin, 2009). We then merged read groups using samtools (Li et al., 2009) and marked duplicates using Picard (2.20.4). We called variants using GATK HaplotypeCaller and hard filtered for high-quality variants following GATK Best Practices recommendations (McKenna et al., 2010; Van der Auwera & O'Connor, 2020). For SNPs, we used the following GATK VariantFiltration expression filters: QUAL < 50.0, QD < 2.0, FS > 60.0, SOR > 3.0, and MQ < 40.0. For INDELs, we used the following GATK VariantFiltration expression filters: QD < 2.0, QUAL < 50.0, and FS > 60.0. We then ran a base quality recalibration with GATK BQSR using the hard-filtered high-quality variants. We then did our final variant calling step using GATK HaplotypeCaller.

We filtered out likely sex chromosomes by testing for bimodality of the depth of coverage within each contig using Hartigan's Dip Test (Hartigan & Hartigan, 1985). Contigs that had a p-value of less than .05 were considered to be located on a sex chromosome and removed. We then took the average depth of coverage across all contigs and removed contigs that were either greater than or less than 5× coverage from the mean genome-wide coverage to filter out repeat elements such as transposable elements. This resulted in a final set containing 5,752,817 SNPs representing ~78% of the full genome across 398 contigs. Individual 10 sampled from Gerlache Island showed an unusually high level of homozygosity, consistent with high levels of inbreeding, and unlike any other individual in the

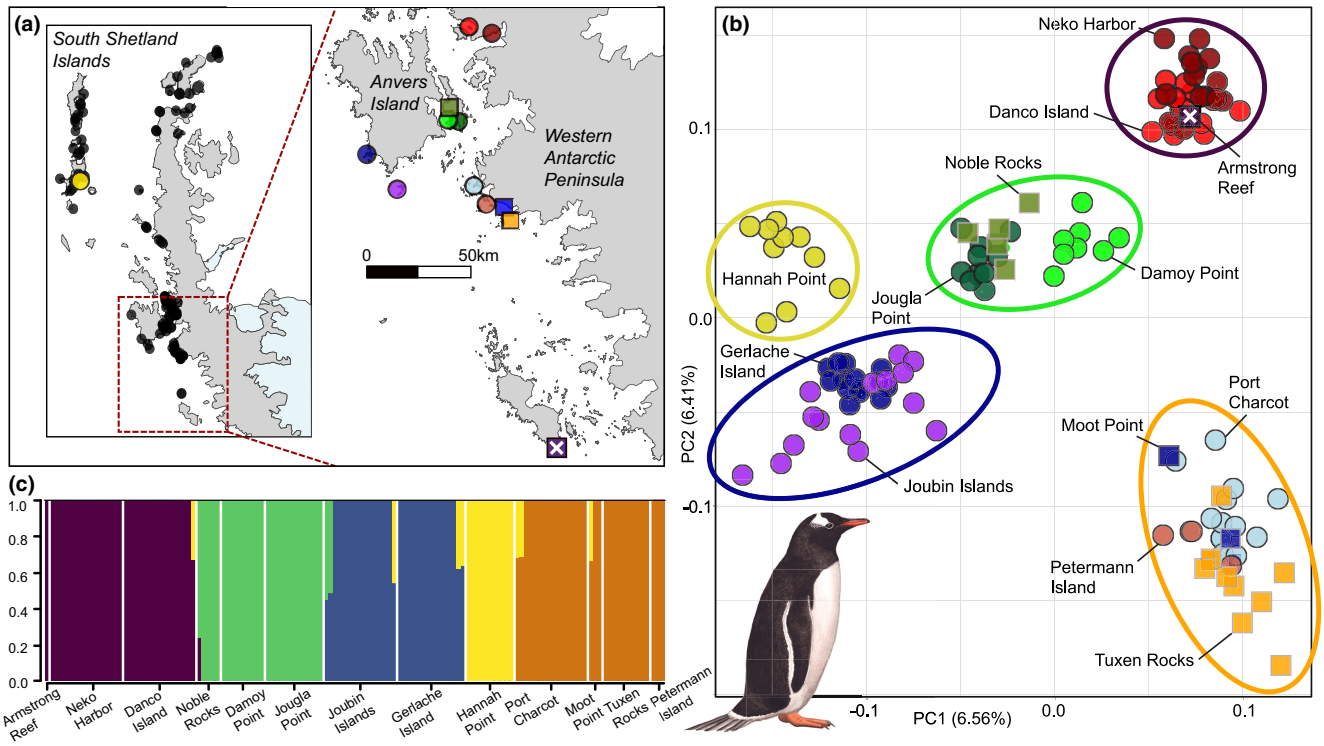

FIGURE 1 (a) Map of colony locations on the Western Antarctic Peninsula. Each coloured circle represents a colony established prior to 20 years ago, and each coloured square represents a recently established colony within the past 20 years. The dark square with a white X indicates Armstrong Reef. Black dots on inset represent all known gentoo penguin colonies. Individual colours are used consistently throughout the figures to represent each colony. (b) Principal component analysis results from smartPCA of 129 individuals projected onto PC axes 1 and 2. (c) ADMIXTURE results with $k = 5$. Bars indicate individual samples and vertical white lines separate individual sample columns by colony. Colours show ancestry clusters with very little admixture. Gentoo penguin artwork by Bryce Robinson.

dataset, including in Gerlache Island. Therefore, we excluded this individual from all downstream population genetic analysis.

2.3 | SNP-based population genetic analysis

For one set of analyses, we randomly thinned the SNP dataset to ~100,000 sites using PLINK (Purcell et al., 2007) to control for linkage disequilibrium. To explore general population structure among individuals across colonies, we conducted a principal component analysis (PCA) using smartPCA (Patterson et al., 2006). We also used ADMIXTURE (Alexander et al., 2009) to perform model-based clustering ancestry analyses assuming $K=2$–12. We ran Treemix on individuals grouped by colony to infer patterns of past population divergence (Pickrell & Pritchard, 2012) and included an Adélie penguin (*P. adeliae*) genome as the outgroup (Vianna et al., 2020) (Figure 2). Treemix returns a maximum likelihood tree based on the allele frequency covariances for the given populations. Migration edges (admixture events) can be included between populations that fit the tree poorly, which are indicated by outlying residuals from the observed allele frequency covariances. Since the maximum likelihood tree for our data with migration edges did not result in any significant outlying residuals, migration was not included in the final model. Tajima's *D* for each population was calculated from allele frequency spectrum generated using ANGSD. We assessed significance by performing 100 simulations of 10 Mbs of sequence (in chunks of 10 kb) under equilibrium using ms (Hudson, 2002) based on the observed Watterson's theta and calculating Tajima's *D* for each simulation. A *p*-value for the observed data was then derived by its percentile in the simulated Tajima's *D* distribution. Finally, we implemented estimated effective migration surfaces (EEMS), an analysis that visualizes regions of a species' distribution where population divergence deviates from uniform isolation by distance (Petkova et al., 2016; Figure 3a). EEMS uses MCMC to estimate migration and diversity parameters by sampling from their posterior distribution given observed genetic dissimilarities. We omitted samples from Hannah Point in order to highlight fine-scale migration rates around Anvers Island and the mainland WAP colonies. We calculated a distance matrix from our thinned SNP dataset using the program bed2diff_v1 from EEMS. We ran three independent chains from different starting seeds for 1×10^7 MCMC iterations (with a burn-in of 1×10^7 iterations and a thinning rate of 5000 iterations) for each of three different grid sizes (250, 450, and 650 demes). We ran multiple short runs to choose proposal values for migration and diversity parameters that had acceptance rates between 20% and 40%. We plotted the log posteriors using the rEEMSplots package in r to confirm the runs converged and then combined runs from the three grid sizes to construct the migration surface visualizations (Petkova et al., 2016).

2.4 | Haplotype-based population genetic analysis

The *P. papua* genome has a low N50, making haplotype-based population genetic analysis difficult. Therefore, we aligned the *P. papua* reference genome (for the sex-chromosome filtered contigs only) to a *Megadyptes antipodes* reference genome (Pan et al., 2019) comprising larger, more complete scaffolds using Ragtag (Alonge et al., 2022) and applied the scaffold merging option. This resulted in 723 *P. papua* contigs placed unambiguously across 77 *Megadyptes*

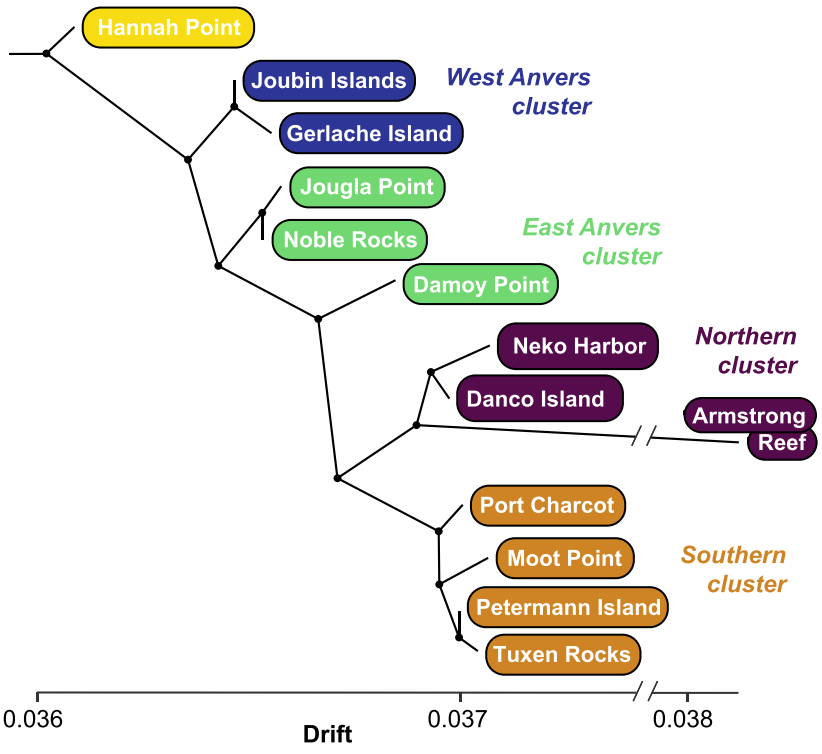

FIGURE 2 Treemix analysis that measures historical relationships among populations. Lengths of branches equal level of genetic drift.

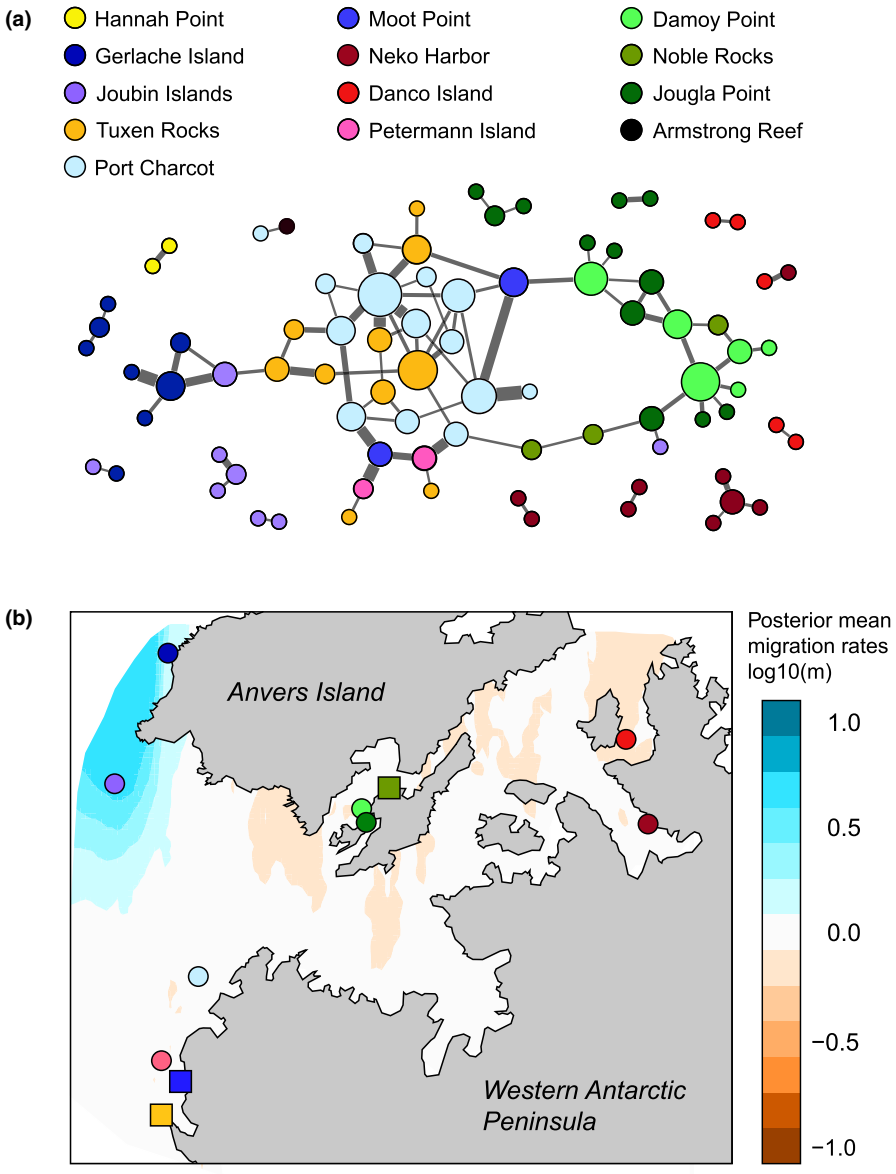

FIGURE 3 (a) EEMS results visualization of gene flow. Circle colours indicate respective colony locations consistent with Figure 1. (b) Gephi network analysis using RefinedIBD shared IBD segments. Nodes represent individuals and edges represent shared IBD segments. Individual node sizes increase with respect to the number of other nodes with which it shares IBD segments. Thickness of edges increases with the number of shared IBD segments between two individual nodes.

scaffolds. We then calculated pairwise linkage disequilibrium (estimated as the r^2 value using a Pearson's correlation coefficient from diploid genotypes (Purcell et al., 2007)) for each pair of SNPs spanning 10 kb on either side of those *P. papua* contigs proposed to be adjacent (including all SNPs with minor allele frequency greater than 10% and with a Hardy–Weinberg *p*-value greater than 1×10^{-8}). The average r^2 across all pairs was weighted by the inverse of their physical distance assuming no gap between contigs and compared to 1000 randomly chosen pairs of contigs to generate an empirical *p*-value for the target pair. Those pairs with $p<.05$ (i.e., greater linkage disequilibrium expected than by chance) were placed on the same scaffold; otherwise, the proposed Megadyptes scaffold was broken into two. We only considered final scaffolds >5 Mb, resulting in a new build (Ppap.V1toMegaLD_5MB) containing 72 scaffolds spanning ~766 Mb (~58% of the complete genome). Seven scaffolds in this new build were between 20 and 40 Mb and 26 were greater than 10 Mb (compared to 0 and 12 scaffolds in the original build, respectively).

We created a chain file and converted genotyped coordinates from *P. papua* to Ppap.V1toMegaLD_5MB using the LiftOverVCF function in GATK and phased the resulting VCF using Shapeit4 (Delaneau et al., 2019). We then used LDHelmet to construct a *linkage-disequilbrium*-based recombination map (Chan et al., 2012) using the combined samples of Neko Harbour and Danco Island. A grid of ρ values was set to −r 0.0 0.1 10.0 1.0 and 100.0 and set to −t 0.0007. We calculated 11 Padé coefficients using pade (Chan et al., 2012) with a population scale mutation rate of 0.0076. We then estimated a recombination map by running rjmcmc (Chan et al., 2012) for 1×10^6 MCMC iterations with burn-in rate of 10,000, a window size of 50, and block penalty of 50.0. We determined the final ρ-scaled recombination map from the 50% percentile of the sampling distribution, with interpolation used to determine ρ between SNPs. Finally, we converted the recombination map to centiMorgans assuming $=4N_eu$ and $\rho=4\ N_eu$, where $=0.0007$, $u=1.5\times10^{-8}$ (Bergeron et al., 2023) and thus $N_e=11{,}667$. We controlled for a minority of regions with spuriously high estimated ρ by capping

the maximum value by the 99th percentile of the genome-wide distribution.

We inferred tracts of identity-by-descent (IBD) using Refined IBD and the recombination map generated above (Browning & Browning, 2013). We accepted any IBD segment with a LOD score>3 and constructed a haplotype network of tracts using ForceAtlas2 edge-node algorithm in Gephi (Jacomy et al., 2014, Figure 3b). We used PLINK to determine runs of homozygosity (ROH) (--homozyg --homozyg-window-snp 1000 --homozyg-window-missing 0 --homozyg-density 1 --homozyg-gap 1000 --homozyg-window-threshold 0.05 --homozyg-window-het 1) and summed the total number of ROH per population using all segments (Figure 4a), and then only segments greater than 5 centimorgans (Figure S6). We excluded Moot Point, Petermann Island, and Armstrong Reef due to low sample sizes. We estimated genome-wide geneaologies using Relate (Speidel et al., 2019, Figure 4c). We estimated historical population size and pairwise separation histories between populations assuming a per-generation mutation rate of 1.5×10^{-8} based on Adélie penguins (Bergeron et al., 2023) and assumed a generation time of eight years. Finally, we used FineSTRUCTURE to perform chromosome painting on dense haplotype data and perform population clustering analyses (Lawson et al., 2012) (Figure S3).

2.5 | Model fitting using *fastsimcoal2*

To test which dispersal patterns could best explain the observed genetic structure, we used an approach implemented in *fastsimcoal2* (Excoffier et al., 2009) that estimates a composite-likelihood based on coalescent simulations in order to fit a demographic model and associated parameters to observed site frequency spectrum (SFS). We compared three dispersal models: a stepping-stone (SS) model of colonization, continent-islands models with independent migration rates (Ci-IM) among populations, and a continent-islands model with a single migration rate (Ci-SM) among populations (Figure 5). For this analysis, we used Hannah Point as the primary source population for both models. We chose the Gerlache Island population over the Joubin Islands population to represent the West Anvers cluster due to the former having fewer ROH (Figure 4a). We randomly selected Jougla Point and Neko Harbour to represent the East Anvers cluster and Northern cluster, respectively. We omitted the Southern cluster from analysis in order to reduce the complexity of the demographic model, as our assumption was that the Northern and Southern clusters would have the same divergence time from East Anvers cluster under our proposed SS model. We generated six folded pairwise joint site frequency spectra (2D-SFS) among Hannah Point, Gerlache Island, Jougla Point, and Neko Harbour using ANGSD (Korneliussen et al., 2014). Based on the significantly positive Tajima's *D* values for all populations and RELATE Ne trajectories, we included a historical population decline event between the ancestral population and Hannah Point in both models. All Ne parameters were set to be constant over time and were drawn from a log-uniform distribution with range [10, 10^6]. Ne for extant lineages were set as constant over time since their divergence from a most recent common ancestor population. We simulated three historical divergence events for the stepping-stone model in which each divergence time parameter was drawn from log-uniform distribution with range [10, 10^5]. We bounded the lower range of each divergence event by the preceding divergence event to simulate a stepping-stone pattern. For both Ci-IM and Ci-SM models, we simulated one historical divergence event where Gerlache Island, Jougla Point, and Neko Island diverged simultaneously from Hannah Point. We allowed for independent migration rates between adjacent populations in the SS model, while allowing for independent pairwise migration rates among all populations in the Ci-IM, and only a single migration rate among all populations in the Ci-SM model (Figure 5). Parameter estimates for each migration rate were drawn from a uniform distribution with range [0, 10^{-2}]. When fitting our observed data, we performed 100,000 simulations per set of parameters with 40 conditional maximization cycles, and, as suggested by the manual of *fastsimcoal2*, each dispersal model was run independently 100 times with different seeds to ensure they were converging in the global maximum likelihood rather than a local maximum. We then selected the best simulation runs based on the highest maximum estimated likelihoods for each model and conducted Akaike information criterion (AIC) model comparisons. Summary statistics estimation and plotting of 1D and 2D SFS were performed using tools in *dadi* (Gutenkunst et al., 2009).

3 | RESULTS

3.1 | Genomic population structure

We performed whole genome resequencing for 129 individuals from 12 colonies at ~18× mean coverage. A PCA revealed five discrete clusters in genotypic (PC1 and PC2) space that corresponded closely with geographic location (Figure 1b). Gentoo penguins at Hannah Point formed their own cluster. The Northern cluster on the WAP consisted of Danco Island and Neko Harbour. Interestingly the non-breeding individual sampled at Armstrong Reef more than 75 km away fell within this cluster. The cluster on the eastern side of Anvers Island consisted of Jougla Point, Noble Rocks, and Damoy Point, with the latter being clearly distinguishable in the PC space. Gerlache and Joubin Islands clustered on the western side of Anvers Island, though the latter showed greater spread. The Southern cluster consisted of Port Charcot and Petermann Island, two larger and older established colonies, and Moot Point and Tuxen Rocks, two colonies recently established within the last 15 years (Herman et al., 2020). While Port Charcot and Tuxen Rocks overlapped in PC space, the latter were generally more shifted along PC2, corresponding with their more southern latitude.

Model-based clustering analysis via ADMIXTURE with $k=5$ confirmed the strong geographic clustering identified by the PCA, with little to no evidence of significant admixture (Figure 1c; Figure S1). Beyond $k=6$, no meaningful structure was observed. Results also indicated that the most recently established colonies

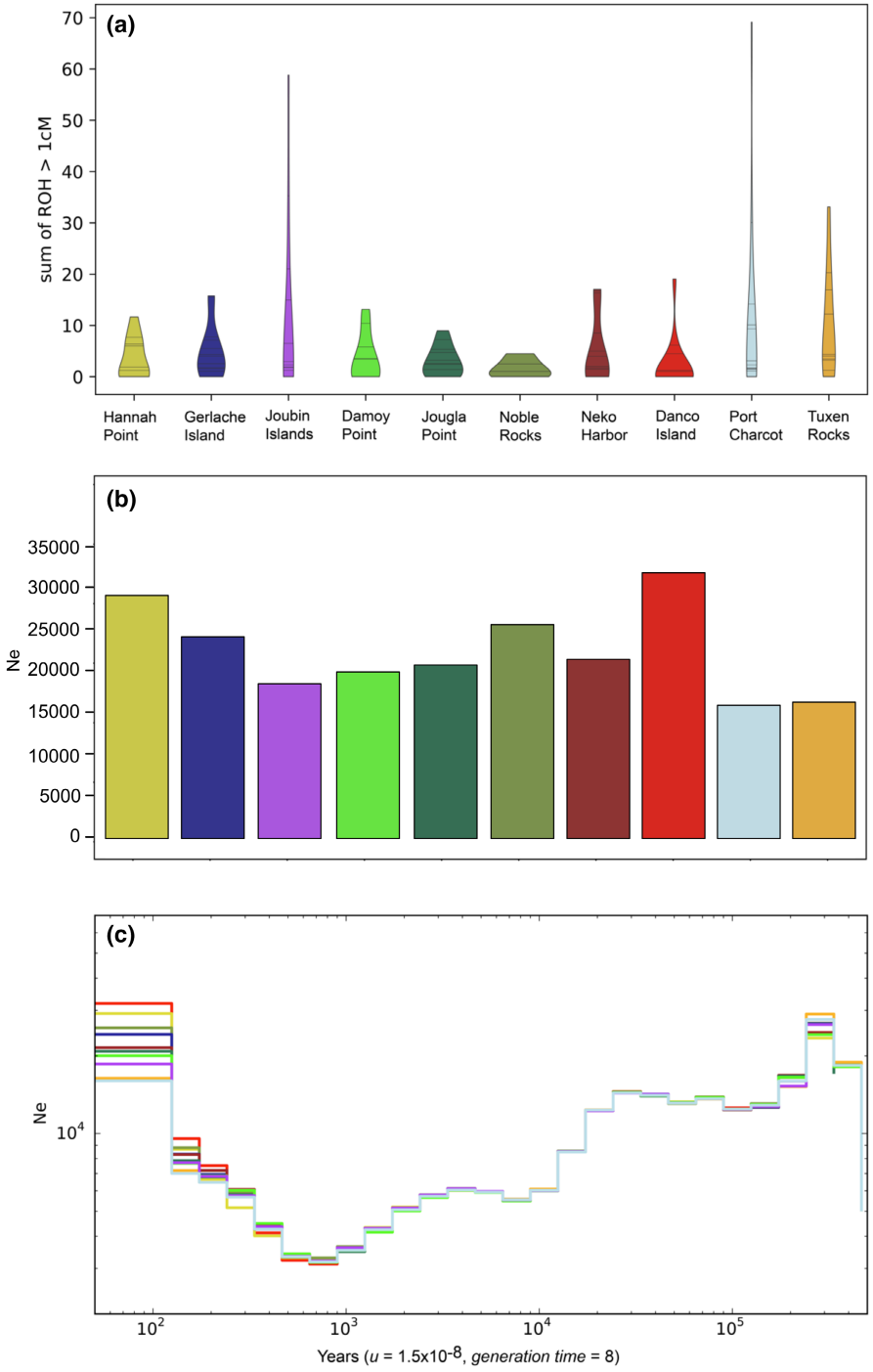

FIGURE 4 (a) Violin plots showing sums of all runs of homozygosity for each population. (b) Most recent estimated effective population sizes (N_e) for each population. (c) Skyline plots of effective population size (N_e) through time for each population in the last 500,000 years.

sampled at the range edge, Moot Point and Tuxen Rocks, were likely sourced from the closest adjacent colonies sampled, Port Charcot and Petermann Island. This indicated that gentoo penguin southern range expansion is likely sourced by the nearest range-edge colonies. However, the non-breeding individual sampled at Armstrong Reef again clustered with the northern colonies, supporting the PCA results.

Reconstruction of population divergences based on patterns of genetic drift using Treemix (Figure 2) supported the five discrete clusters found in the PCA and ADMIXTURE results. Introducing migration edges did not improve the model fit based on remaining residuals with no migration and were therefore omitted (Figure S2). The topology of the resulting dendrogram revealed an early split between Hannah Point in the South Shetland Islands and all colonies sampled from the Western Antarctic Peninsula (Figure 2). The East Anvers cluster split from the West Anvers cluster, and the Southern cluster and Northern cluster formed a clade that split from an ancestral population likely closely related to the Damoy Point colony (Figure 2). These results suggest gentoo penguins colonized Anvers Island first and moved from west to east in a stepping-stone fashion before colonizing this region of the mainland WAP. An unrooted dendrogram of individuals constructed using FineSTRUCTURE

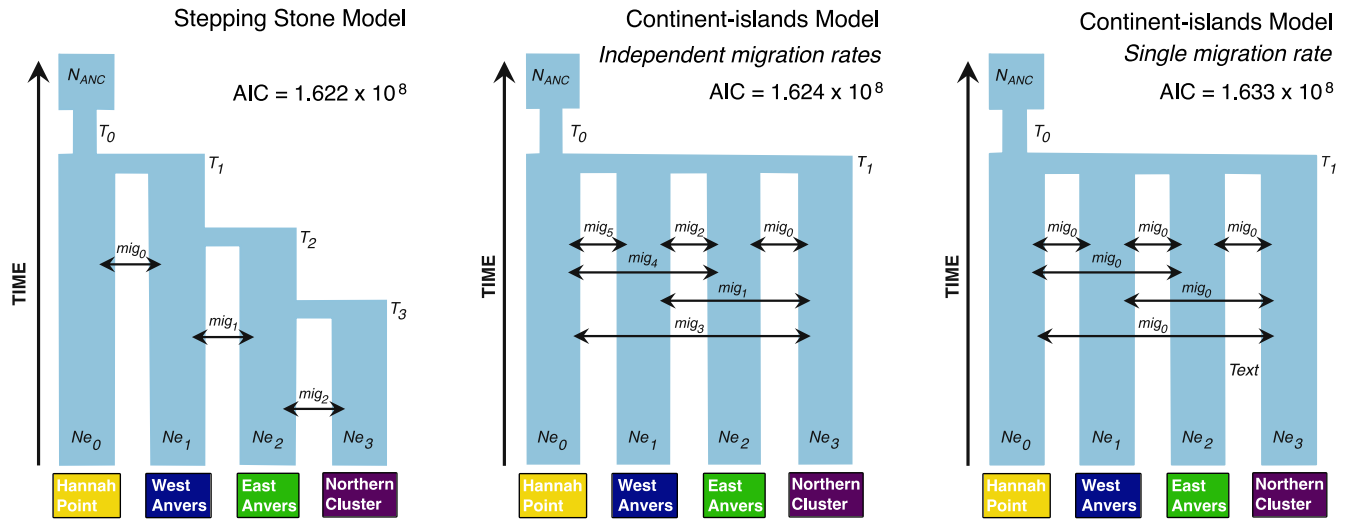

FIGURE 5 Schematics of the three dispersal models. Parameters are labelled accordingly and respective AIC values from best simulation runs are provided.

(Figure S3) supported the general topology of the population-based results of Treemix and grouped the individual from Armstrong Reef with individuals sampled from Danco Island, suggesting the vagrant likely originated from this colony rather than Neko Harbour.

IBD segments were generally short and ranged from 1.5 to 3.8 centimorgans, indicating no major recent episodes of gene flow. A total of 89 shared segments were inferred, with 36 segments shared among individuals from different colonies. Individuals from Hannah Point, Danco Island, and Neko Harbour shared no segments with individuals from other colonies (including those belonging to the same PC and ADMIXTURE clusters), suggesting these colonies had experienced extensive genetic drift since colonization events, with little post-divergence gene flow from the other colonies sampled. In contrast, a large network of shared IBD tracts connected individuals from all colonies around Anvers Island and the Southern cluster (Figure 3b). Notably, the network was made up of three primary components: (i) some shared segments among the southern colonies and the East Anvers cluster (including Damoy despite no overlap in the PCA), (ii) some shared segments among the Western Anvers cluster, and (iii) many shared tracts among individuals from all four populations in the Southern cluster, likely reflecting the recent common ancestry of these newer colonies. The latter had some shared tracts with both the East and West Anvers cluster possibly reflecting limited gene flow between Anvers Island and the mainland after divergence had occurred.

Explicitly modelling gene flow spatially using EEMS migration suggested reduced migration rates between Anvers Island and mainland WAP (Figure 3a), while migration rates between Joubin Islands and Gerlache Island were at least tenfold higher than the region average, with no areas of reduced migration rates between the West Anvers cluster and the Southern cluster. However, these results are likely a vast oversimplification of migration surfaces. Because EEMS requires the input of a single closed polygon for spatial analysis, we were not able to include all of the islands and islets present around the Antarctic Peninsula, making it difficult to infer all physical barriers to gene flow.

Estimates of population genetic diversity (Θ_W,Θ_π) and individual genome-wide heterozygosity for the two newest colonies, Moot Point and Tuxen Rocks, were not markedly different from the other more established historical colonies in the Southern cluster, Port Charcot and Peterman Island, showing that despite being only recently founded (<15 years), they maintained much of the same genetic diversity as their progenitors, rather than evidence of showing strong founder effects (Table S5).

While the above estimates provide an overall measure of genetic diversity over the long term, we then performed a series of analyses that examined how diversity may have been structured at different points in gentoo population history. Individuals from Joubin Islands and Port Charcot had the highest number of ROH greater than both 1 and 5 cm, followed by Tuxen Rocks (Figure 4a, Figure S6), suggesting relatively small recent effective population sizes (N_e). Gerlache Island had the next highest number of ROH greater than 5 cm (Figure S6). All other colonies had much fewer ROH greater than 1 cm (Figure 4a), and Jougla Point, Noble Rocks, and Neko Harbour had zero ROH greater than 5 cm (Figure S6). Allele frequency spectra for each population demonstrate a relative deficiency of singletons, and Tajima's *D* estimates were all significantly above zero (Figure S4, Table S5), indicating that all populations went through a population decline at some point in the past. We used Relate to determine genome-wide genealogies and used the coalescent rate to model changes in N_e for each individual population over the last ~500,000 years (Figure 4c) as well as cross-coalescent rates between pairs of populations to model periods of reduced gene flow (Figure S7). All populations followed the same declining N_e trajectory until approximately ~500–1000 years ago, presumably reflecting the positive Tajima's *D* values. Populations appeared to then begin diverging and all populations experienced significant increases in N_e. Danco Island and Hannah Point had the highest most recent

N_e interval estimate, while Port Charcot and Tuxen Rocks followed by Joubin Island had the lowest (Figure 4b), supporting the observation of increased ROH in these latter populations (Figure 4a). Among populations from different clusters, cross-coalescent rates showed clear evidence of reduced population gene flow, while pairs of populations within clusters were largely indistinguishable. Interestingly the timing of onset of reduced gene flow between all five clusters appears quite similar, suggesting the expansion across the WAP occurred quite rapidly.

4 | MODEL COMPARISON

We used the composite-likelihood approach implemented in *fastsimcoal2* to compare the fit of three potential colonization models to our observed data: a stepping-stone model (SS) of colonization, continent-islands models with independent migration rates (Ci-IM) among populations, and a null continent-islands model with a single migration rate (Ci-SM) among populations (Figure 5). We included a historical population decline event between the ancestral population and Hannah Point in all three models based on the significant Tajimas' *D* values and RELATE analysis (Figure 5). To ensure our model was not overparameterized and prevent overfitting, we did not model growth in individual clusters, included only one representative population per cluster, and excluded the Southern cluster under the assumption that it would be similarly diverged from the East Anvers cluster as the Northern cluster under the SS model. Thus, we fit 2D-SFS for Hannah Point, Gerlache Island, Jougla Point, and Neko Harbour.

The AIC estimator for the SS model was found to be the best fit of the data (AIC = 1.622×10^8), followed by the Ci-IM (AIC = 1.624×10^8) and the Ci-SM (AIC = 1.633×10^8). In fact, the likelihood of the SS model was higher than the Ci-IM model despite possessing less parameters (Table S8). Visual inspection of the observed versus best model 1D (Figure S9) and 2D-SFS (Figure S10), and their residuals showed generally a very good fit, the only major discrepancy being a slightly wider model 2D-SFS, which may be due to an excess of intermediate frequency singletons in Hannah Point that could not be fit by any of three models SFS-based summary statistics (ΘW, $\Theta\pi$, Tajima's *D*, and F_{ST}) were also in broad agreement, apart from the model pairwise F_{ST}s involving Hannah Point, which, consistent with the wider model 2D-SFS, were larger than the observed values.

The estimated divergence times for the stepping-stone model (T_1, T_2, and T_3) were all fairly recent (~1–20 kya, assuming 8 years per generation), though these were older than that inferred from RELATE. However, this discrepancy is almost certainly a result of the fact that RELATE allows populations to separate gradually with reduced gene flow rather than enforcing an instant divergence with subsequence long-term gene flow. Similarly, the time of the population bottleneck (T0) was estimated on the order of 1×10^5 years, which would be compatible with the long-term decline over hundreds of thousands of years observed in our RELATE analysis. Migration was found to be higher between the West and East Anvers cluster ($mig_1 = \sim 1 \times 10^{-2}$ n units of proportion of migrants per generation) than between Hannah Point and West Anvers ($mig_2 = \sim 2 \times 10^{-3}$), while the lowest rate between East Anvers and the Northern cluster ($mig_0 = \sim 5 \times 10^{-4}$).

Though it could not fit the data as well as the SS model, it was noteworthy that allowing migration rates to vary among populations significantly improved the model fit in the Ci-IM compared to the Ci-SM, and the estimated rates showed a pattern mimicking a stepping-stone model of migration, with relatively high migration between neighbouring demes ($\sim 1 \times 10^{-2}$, 4×10^{-3}), and like the SS model, and low migration with Hannah Point ($\sim 1 \times 10^{-5}$).

5 | DISCUSSION

We tracked the ongoing climate-driven range expansion of gentoo penguins in Antarctica genetically, revealing a stepping-stone model of dispersal buffered against founder effects. We characterized the population genetic structure and evolutionary history of gentoo penguin colonies along the WAP using a high-coverage whole genome sequencing approach. We identified a region of fine-scale genetic differentiation among five discrete population clusters that correspond closely to geographic location (Figure 1a,b). Demographic modelling using *fastsimcoal2* provided evidence for a stepping-stone pattern of dispersal and colonization, in which Anvers Island was likely colonized by gentoo penguins that originated in the South Shetland Islands (represented by Hannah Point) and then continued to disperse east quite rapidly along the southern coast of Anvers Island before colonizing the mainland of the WAP (Figure 6). Gentoo penguins from the Northern cluster and Southern cluster were likely sourced by individuals from the cluster east of Anvers Island based on the Treemix results. Our results also indicated that the most recently established colonies sampled at the range edge, Moot Point and Tuxen Rocks, were likely sourced from the closest adjacent colonies, Port Charcot and Petermann Island. Thus, gentoo penguin southern range expansion is likely continuing in a stepping-stone pattern of dispersal, with the leading edge of expansion acting as the source of colonizing individuals for continued expansion. Surprisingly, a vagrant individual found at Armstrong Reef likely came from the Danco Island in the Northern cluster (Figure 1b,c, Figure S3), suggesting that gentoo penguins may sometimes disperse beyond the expected stepping-stone pattern, though this individual was not breeding.

The Northern and Southern clusters did not share any IBD segments despite being the most genetically similar, whereas the Southern cluster shared IBD segments with the East Anvers cluster. Our stepping-stone dispersal and colonization model could explain these IBD results. If colonization of mainland WAP stemmed from a single historical colonization event originating from the East Anvers Island cluster, individuals may have dispersed and colonized in a stepping-stone pattern north along the mainland WAP, eventually colonizing the Northern cluster and becoming genetically differentiated (Figure 6). The presence of multiple established gentoo penguin colonies along the mainland WAP south of the Northern cluster

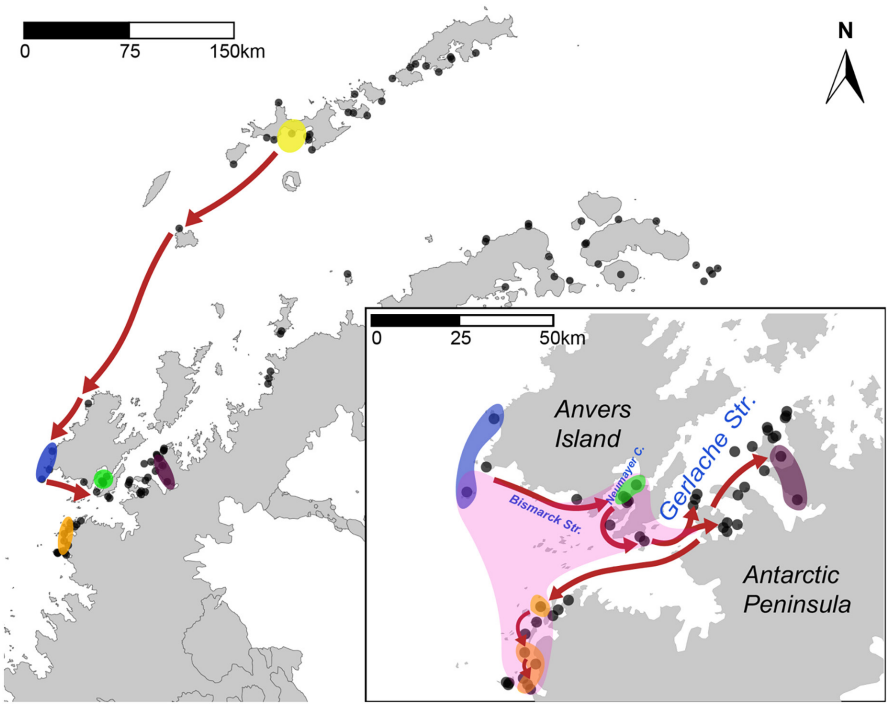

FIGURE 6 Map of inferred dispersal pathways (red arrows) among major gentoo penguin population clusters. Black circles indicate all other gentoo penguin colonies. Pink-shaded region indicates post-divergence gene flow.

supports this concept (Figure 6). Meanwhile, gentoo penguins from the same historical colonization event on the mainland WAP began dispersing south and colonizing the area at the range edge where the Southern cluster is located, resulting in divergence between the north and south (Figure 6). Some limited post-divergence gene flow from the West and East Anvers Island clusters may have also occurred during this southward expansion, reflecting our IBD results (Figure 3b). This is also supported by the EEMS visualization showing patches of reduced gene flow between the Northern and Southern clusters, indicative of the Northern cluster becoming isolated over time, and slight extension of increased gene flow from the Western Anvers cluster towards the Southern cluster with no patches of reduced gene flow between them (Figure 3a).

The earliest known presence of gentoo penguins on the Antarctic Peninsula was at Potter Peninsula on King George Island dating back to the mid-Holocene, or 4450–4550 B.P. (Del Valle et al., 2002), while the earliest known gentoo penguin breeding colony was located on Byers Peninsula on Livingston Island dating back 1150 B.P. (Emslie et al., 2011), at least 2000 years after the island was deglaciated (Ingólfsson et al., 2003). Emslie et al. (2011) suggested that gentoo penguins may have been slow to colonize recently deglaciated regions at that time, which is supported by our findings of a signal of population expansion started roughly 500–800 years ago after a decline in effective population size since approximately 20,000 B.P. While Relate results are highly dependent on input parameters and should therefore be interpreted with caution, the decline in historical population size may have coincided with the last glacial maximum (LGM) during which the ice shelf extended to the continental shelf (Davies et al., 2012), making the region unsuitable for gentoo penguins that had to take refuge on the closest sub-Antarctic islands such as South Georgia and the South Sandwich Islands. In contrast, two studies that used mitochondrial marker HVR-1 (Clucas et al., 2014; Peña M et al., 2014) detected signals of population expansion estimated around 13,000 and 15–20,000 years ago, respectively, suggesting that the populations were expanding soon after the last LGM out of glacial refugia. This order of magnitude difference in results may be due to known issues with mitochondrial markers as unreliable for inferring population history without the inclusion of fossil calibrations (Allio et al., 2017; Galtier et al., 2009) and also may reflect differences in effective population size and gene flow as a result of sex-specific processes (Weir et al., 2016). Whole genome sequencing, however, provides high resolution with greater accuracy for inferring demographic histories, and our results likely provide an update on these historical population sizes and demographic history of gentoo penguins to the literature. Furthermore, the lack of paleofossil evidence prior to 4550 B.P. also supported our inferred timing of historical population expansion (Del Valle et al., 2002; Emslie et al., 2011), though it is possible that earlier fossil evidence has yet to be discovered.

While range expansion could be considered beneficial for a species, there can be genetic consequences such as low genetic diversity at the leading edge. In the case of a species following a stepping-stone dispersal pattern during range expansion, genetic diversity can diminish with distance (Excoffier et al., 2009; White et al., 2013). Surprisingly, the gentoo penguin colonies sampled exhibited very little evidence of founder effects. Instead, the effective population sizes of colonies increased at a similar rate as populations diverged over time (Figure 4c). Multiple estimates of genetic diversity (Table S5) were relatively high and similar across all colonies, including those less than 15 years old where such founder effects would be expected to be most pronounced, suggesting large numbers of individuals migrate to new locations during the initial colonization phase rather than a few individuals. This is true even

of the Southern cluster colonies, which are known to be recently established, and yet had only a moderate increase in ROH and short between-colony IBD segments consistent with only a minor decrease in effective population size. Herman and Lynch (2022) found that continued – and in some cases increasing – immigration is essential for observed population growth in recently colonized gentoo penguin colonies, which is supported by our findings. This may help to buffer nascent colonies from Allee effects, as colonial seabird fledglings are vulnerable to predation and larger cohorts reduce the probability of being depredated (Schippers et al., 2011). In the closely related Adélie penguin, it has been shown that fledgling survival is inversely related to cohort size due to a combination of increased predation risk by aerial predators at the colony, by marine predators at sea, and by reduced cooperative behaviour for foraging efficiency (Emmerson & Southwell, 2022). The ongoing dispersal behaviour of gentoo penguins may therefore be an evolutionary adaptation that has allowed the species to increase in number and expand their range in recent decades, demonstrating robustness to environmental changes and predator presence.

A meta-analysis by Friesen et al. (2007) found that isolation by distance was present in many colonial seabird species, but evidence of long-range colonization in other species suggested that distance is not a universal barrier to dispersal and colonization. This was demonstrated by the gentoo penguin sampled at Armstrong Reef that likely came from Danco Island, approximately 215 km to the north. Seabirds are known to carry out prospecting behaviours to evaluate future potential breeding locations, with this behaviour particularly common in pre- or failed breeders (Campioni et al., 2017; Ponchon et al., 2017). Individual variation in the spatial scale of prospecting and dispersal behaviours is found throughout seabirds (Phillips et al., 2017; Ponchon et al., 2017), and it is possible that such variation in gentoo penguins may be inherent plasticity in response to environmental changes. There is some evidence of selection against immigration in wild seabird populations. For example, individuals who immigrate from their natal breeding colonies to other locations have lower fitness than individuals who remain at their natal breeding grounds (Barbraud & Delord, 2021). However, dispersal may outweigh this disadvantage if the current breeding locations have reached nest-site carrying capacity, forcing individuals to disperse to other locations to breed (Acosta Alamo et al., 2022), or in response to changes in local environmental conditions (Barros et al., 2013; Garnier et al., 2023; Møller et al., 2006). This suggests there is likely an adaptive advantage in individual variation in dispersal behaviours that override distance barriers, and gentoo penguins may sustain this variation in behaviour.

In conclusion, our study showed a fine-scale and complex network of historical and recent dispersal events indicative of the stepping-stone model of dispersal, illustrating how gentoo penguins have spread and colonized along the Western Antarctic Peninsula. However, the detection of post-divergence gene flow and minimal population bottlenecks suggests a less dramatic form of the classic stepping-stone model. This method of dispersal likely serves as a buffer against founder effects, exemplified by the robustness of gentoo penguin populations and their continued southward expansion. To our knowledge, this study is the first to use whole genome resequencing to investigate such fine-scale networks of dispersal in a species undergoing range expansion. Harnessing high-resolution genomic data to explore such systems provides an effective tool for determining dispersal networks in species that are otherwise difficult or impossible to track using alternative methods. Such results provide a foundation for the forecasting of distributions of species undergoing range expansion and shifts in response to climate-driven environmental changes. Our findings of sustained genetic diversity and population growth at the range edge, coupled with evidence of more southerly prospecting, highlight the need for ongoing monitoring of gentoo penguins on the West Antarctic Peninsula as it is clear their range expansion may be ongoing.

AUTHOR CONTRIBUTIONS

R.H., H.L., and K.V. conceived the study. R.H., G.C., J.Y., J.B., B.R., S.R., J.S., and K.O. collected genetic samples. R.H. and K.V. analysed the data. R.H., H.L., and K.V wrote the manuscript with input from all coauthors.

ACKNOWLEDGEMENTS

We gratefully acknowledge National Geographic Lindblad Expeditions, Hurtigruten Expeditions, Skip Novak and Pelagic Expeditions, and the crew of the Vinson of Antarctica (Alec Hazel, Kenneth Perigón, and Jose Gritti) for providing logistical support for fieldwork in Antarctica. We thank Kerry Reid for providing wet lab training and guidance, and Irby Lovette and Jacob Cooper for their assistance in the field. We acknowledge Stony Brook University's SeaWulf supercomputing resources (https://it.stonybrook.edu/services/high-performance-computing) made available for conducting the research reported in this paper. We thank Ron Naveen at Oceanites for providing permitting and funding for sample collection during the 2019/2020 field season (Antarctic Conservation Act Permit No. 2019-0001). Sampling during the 2021/2022 season was permitted through the UK Foreign and Commonwealth Office (Permit No. 20-2021-22). This work was funded by NASA FINESST Grant Award No. 80NSSC20K1605, the Stony Brook University Faculty-Staff Dissertation Fellowship, the Institute for Advanced Computational Sciences Endowment for Ecology & Evolution, the National Geographic Society, and the Field Museum's Robert A. Pritzker Center of Meteoritics and Polar Studies.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Raw sequence reads and BAM files for all individuals are available on the sequence read archive (SRA) at NCBI under BioProject PRJNA1047030.

BENEFIT-SHARING STATEMENT

Benefits from this research accrue from the sharing of our data on the public database NCBI Sequence Read Archive (SRA) under BioProject PRJNA1047030.

ORCID

Rachael W. Herman https://orcid.org/0000-0002-5914-6907
Jane Younger https://orcid.org/0000-0001-5974-7350

REFERENCES

Acosta Alamo, M., Manne, L. L., & Veit, R. R. (2022). Does population size drive changes in transatlantic vagrancy for gulls? A study of seven North Atlantic species. *Frontiers in Ecology and Evolution*, *10*, 275.

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664.

Allio, R., Donega, S., Galtier, N., & Nabholz, B. (2017). Large variation in the ratio of mitochondrial to nuclear mutation rate across animals: Implications for genetic diversity and the use of mitochondrial DNA as a molecular marker. *Molecular Biology and Evolution*, *34*(11), 2762–2772.

Alonge, M., Lebeigle, L., Kirsche, M., Jenike, K., Ou, S., Aganezov, S., Wang, X., Lippman, Z. B., Schatz, M. C., & Soyk, S. (2022). Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biology*, *23*(1), 1–19.

Aronson, R. B., Smith, K. E., Vos, S. C., McClintock, J. B., Amsler, M. O., Moksnes, P. O., Ellis, D. S., Kaeli, J., Singh, H., Bailey, J. W., Schiferl, J. C., van Woesik, R., Martin, M. A., Steffel, B. V., Deal, M. E., Lazarus, S. M., Havenhand, J. N., Swalethorp, R., Kjellerup, S., & Thatje, S. (2015). No barrier to emergence of bathyal king crabs on the Antarctic shelf. *Proceedings of the National Academy of Sciences*, *112*(42), 12997–13002.

Barbraud, C., & Delord, K. (2021). Selection against immigrants in wild seabird populations. *Ecology Letters*, *24*(1), 84–93.

Barros, A., Alvarez, D., & Velando, A. (2013). Climate influences fledgling sex ratio and sex-specific dispersal in a seabird. *PLoS One*, *8*(8), e71358.

Bergeron, L. A., Besenbacher, S., Zheng, J., Li, P., Bertelsen, M. F., Quintard, B., Hoffman, J. I., Li, Z., Leger, J. S., Shao, C., Stiller, J., Gilbert, M. T. P., Schierup, M. H., & Zhang, G. (2023). Evolution of the germline mutation rate across vertebrates. *Nature*, *615*(7951), 285–291.

Browning, B. L., & Browning, S. R. (2013). Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics*, *194*(2), 459–471.

Campioni, L., Granadeiro, J. P., & Catry, P. (2017). Albatrosses prospect before choosing a home: Intrinsic and extrinsic sources of variability in visit rates. *Animal Behaviour*, *128*, 85–93.

Chan, A. H., Jenkins, P. A., & Song, Y. S. (2012). Genome-wide fine-scale recombination rate variation in *Drosophila melanogaster*. *PLoS Genetics*, *8*(12), e1003090.

Clucas, G. V., Dunn, M. J., Dyke, G., Emslie, S. D., Levy, H., Naveen, R., Polito, M. J., Pybus, O. G., Rogers, A. D., & Hart, T. (2014). A reversal of fortunes: Climate change 'winners' and 'losers' in Antarctic peninsula penguins. *Scientific Reports*, *4*(1), 1–7.

Clucas, G. V., Younger, J. L., Kao, D., Emmerson, L., Southwell, C., Wienecke, B., Rodgers, A. D., Bost, C. A., Miller, G. D., Polito, M. J., Lelliott, P., Handley, J., Crofts, S., Phillips, R. A., Dunn, M., Miller, K. J., & Hart, T. (2018). Comparative population genomics reveals key barriers to dispersal in Southern Ocean penguins. *Molecular Ecology*, *27*(23), 4680–4697.

Cristofari, R., Liu, X., Bonadonna, F., Cherel, Y., Pistorius, P., Le Maho, Y., Raybaud, V., Stenseth, N. C., Le Bohec, C., & Trucchi, E. (2018). Climate-driven range shifts of the king penguin in a fragmented ecosystem. *Nature Climate Change*, *8*(3), 245–251.

Davies, B. J., Hambrey, M. J., Smellie, J. L., Carrivick, J. L., & Glasser, N. F. (2012). Antarctic peninsula ice sheet evolution during the Cenozoic era. *Quaternary Science Reviews*, *31*, 30–66.

Del Valle, R. A., Montalti, D., & Inbar, M. (2002). Mid-Holocene macrofossil-bearing raised marine beaches at potter peninsula, King George Island, South Shetland Islands. *Antarctic Science*, *14*(3), 263–269.

Delaneau, O., Zagury, J. F., Robinson, M. R., Marchini, J. L., & Dermitzakis, E. T. (2019). Accurate, scalable and integrative haplotype estimation. *Nature Communications*, *10*(1), 5436.

Emmerson, L., & Southwell, C. (2022). Environment-triggered demographic changes cascade and compound to propel a dramatic decline of an Antarctic seabird metapopulation. *Global Change Biology*, *28*(24), 7234–7249.

Emslie, S. D., Baumann, K., & van Tuinen, M. (2011). Late Holocene occupation of gentoo penguins (*Pygoscelis papua*) on byers peninsula, Livingston Island, Antarctica. *Polar Biology*, *34*, 283–290.

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

Foley, C. M., Hart, T., & Lynch, H. J. (2018). King penguin populations increase on South Georgia but explanations remain elusive. *Polar Biology*, *41*, 1111–1122.

Friesen, V. L., Burg, T. M., & McCoy, K. D. (2007). Mechanisms of population differentiation in seabirds. *Molecular Ecology*, *16*(9), 1765–1785.

Galtier, N., Nabholz, B., Glémin, S., & Hurst, G. D. D. (2009). Mitochondrial DNA as a marker of molecular diversity: A reappraisal. *Molecular Ecology*, *18*(22), 4541–4550.

Garnier, J., Clucas, G., Younger, J., Sen, B., Barbraud, C., Larue, M., Fraser, A. D., Labrousse, S., & Jenouvrier, S. (2023). Massive and infrequent informed emigration events in a species threatened by climate change: The emperor penguins.

Germain, S. J., & Lutz, J. A. (2020). Climate extremes may be more important than climate means when predicting species range shifts. *Climatic Change*, *163*(1), 579–598.

Gervais, C. R., Champion, C., & Pecl, G. T. (2021). Species on the move around the Australian coastline: A continental-scale review of climate-driven species redistribution in marine systems. *Global Change Biology*, *27*(14), 3200–3217.

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, *5*(10), e1000695.

Hartigan, J. A., & Hartigan, P. M. (1985). The dip test of unimodality. *The Annals of Statistics*, *13*(1), 70–84.

Herman, R. W., Borowicz, A., Lynch, M., Trathan, P., Hart, T., & Lynch, H. (2020). Update on the global abundance and distribution of breeding gentoo penguins (*Pygoscelis papua*). *Polar Biology*, *43*, 1947–1956.

Herman, R. W., & Lynch, H. J. (2022). Age-structured model reveals prolonged immigration is key for colony establishment in gentoo penguins. *Ornithological Applications*, *124*(3), duac014.

Herman, R. W., Winger, B. M., Dittmann, D. L., & Harvey, M. G. (2022). Fine-scale population genetic structure and barriers to gene flow in a widespread seabird (*Ardenna pacifica*). *Biological Journal of the Linnean Society*, *137*(1), 125–136.

Hindell, M. A., Reisinger, R. R., Ropert-Coudert, Y., Hückstädt, L. A., Trathan, P. N., Bornemann, H., Charrassin, J. B., Chown, S. L., Costa, D. P., Danis, B., Lea, M. A., Thompson, D., Torres, L. G., Dan de Putte, A. P., Alderman, R., Andrews-Goff, V., Arthur, B., Ballard, G., Bengston, J., … Raymond, B. (2020). Tracking of marine predators to protect Southern Ocean ecosystems. *Nature*, *580*(7801), 87–92.

Hudson, R. R. (2002). Generating samples under a Wright-fisher neutral model. *Bioinformatics*, *18*, 337–338.

Ingólfsson, Ó., Hjort, C., & Humlum, O. (2003). Glacial and climate history of the Antarctic peninsula since the last glacial maximum. *Arctic, Antarctic, and Alpine Research*, *35*, 175–186.

Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS One*, *9*(6), e98679.

Kersten, O., Star, B., Leigh, D. M., Anker-Nilssen, T., Strøm, H., Danielsen, J., Descamps, S., Erikstad, K. E., Fitzsimmons, M. G., Fort, J., Hansen, E. S., Harris, M. P., Irestedt, M., Kleven, O., Mallory, M. L., Jakobsen, K., & Boessenkool, S. (2021). Complex population structure of the Atlantic puffin revealed by whole genome analyses. *Communications Biology*, *4*(1), 922.

Kimura, M., & Weiss, G. H. (1964). The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics*, *49*(4), 561–576.

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics*, *15*(1), 1–13.

Lawson, D. J., Hellenthal, G., Myers, S., & Falush, D. (2012). Inference of population structure using dense haplotype data. *PLoS Genetics*, *8*(1), e1002453.

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

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & 1000 Genome Project Data Processing Subgroup. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079.

Lin, Y., Moreno, C., Marchetti, A., Ducklow, H., Schofield, O., Delage, E., Meredith, M., Li, Z., Eveillard, D., Chaffron, S., & Cassar, N. (2021). Decline in plankton diversity and carbon flux with reduced sea ice extent along the Western Antarctic peninsula. *Nature Communications*, *12*(1), 4948.

Lynch, H. J., Naveen, R., Trathan, P. N., & Fagan, W. F. (2012). Spatially integrated assessment reveals widespread changes in penguin populations on the Antarctic peninsula. *Ecology*, *93*(6), 1367–1377.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303.

Meredith, M. P., Stammerjohn, S. E., Venables, H. J., Ducklow, H. W., Martinson, D. G., Iannuzzi, R. A., Leng, M. J., ven Wessem, J. M., Reijmer, C. H., & Barrand, N. E. (2017). Changing distributions of sea ice melt and meteoric water west of the Antarctic peninsula. *Deep Sea Research Part II: Topical Studies in Oceanography*, *139*, 40–57.

Møller, A. P., Flensted-Jensen, E., & Mardal, W. (2006). Dispersal and climate change: A case study of the Arctic tern Sterna paradisaea. *Global Change Biology*, *12*(10), 2005–2013.

Munilla, I., Genovart, M., Paiva, V. H., & Velando, A. (2016). Colony foundation in an oceanic seabird. *PLoS One*, *11*(2), e0147222.

Pan, H., Cole, T. L., Bi, X., Fang, M., Zhou, C., Yang, Z., Ksepka, D. T., Hart, T., Bouzat, J. L., Argilla, L. S., Bertelsen, M. F., Boersma, P. D., Bost, C. A., Cherel, Y., Dann, P., Fiddaman, S. R., Haward, P., Labuschagne, K., Mattern, T., … Zhang, G. (2019). High-coverage genomes to elucidate the evolution of penguins. *GigaScience*, *8*(9), giz117.

Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genetics*, *2*(12), e190.

Peña M, F., Poulin, E., Dantas, G. P., Gonzalez-Acuna, D., Petry, M. V., & Vianna, J. A. (2014). Have historical climate changes affected gentoo penguin (*Pygoscelis papua*) populations in Antarctica? *PLoS One*, *9*(4), e95375.

Péron, C., Weimerskirch, H., & Bost, C. A. (2012). Projected poleward shift of king penguins' (*Aptenodytes patagonicus*) foraging range at the Crozet Islands, southern Indian Ocean. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1738), 2515–2523.

Perry, A. L., Low, P. J., Ellis, J. R., & Reynolds, J. D. (2005). Climate change and distribution shifts in marine fishes. *Science*, *308*(5730), 1912–1915.

Petkova, D., Novembre, J., & Stephens, M. (2016). Visualizing spatial population structure with estimated effective migration surfaces. *Nature Genetics*, *48*(1), 94–100.

Petry, M. V., Basler, A. B., Valls, F. C. L., & Krüger, L. (2013). New southerly breeding location of king penguins (*Aptenodytes patagonicus*) on Elephant Island (maritime Antarctic). *Polar Biology*, *36*, 603–606.

Phillips, R. A., Lewis, S., González-Solís, J., & Daunt, F. (2017). Causes and consequences of individual variability and specialization in foraging and migration strategies of seabirds. *Marine Ecology Progress Series*, *578*, 117–150.

Picard Tools Broad Institute. http://broadinstitute.github.io/picard/

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, *8*(11), e1002967.

Pinsky, M. L., Eikeset, A. M., McCauley, D. J., Payne, J. L., & Sunday, J. M. (2019). Greater vulnerability to warming of marine versus terrestrial ectotherms. *Nature*, *569*(7754), 108–111.

Poloczanska, E. S., Burrows, M. T., Brown, C. J., García Molinos, J., Halpern, B. S., Hoegh-Guldberg, O., Kappel, C. V., Moore, P. J., Richardson, A. J., Schoeman, D. S., & Sydeman, W. J. (2016). Responses of marine organisms to climate change across oceans. *Frontiers in Marine Science*, *3*, 62.

Ponchon, A., Iliszko, L., Grémillet, D., Tveraa, T., & Boulinier, T. (2017). Intense prospecting movements of failed breeders nesting in an unsuccessful breeding subcolony. *Animal Behaviour*, *124*, 183–191.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., de Bakker, W. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81*(3), 559–575.

Schippers, P., Stienen, E. W., Schotman, A. G., Snep, R. P., & Slim, P. A. (2011). The consequences of being colonial: Allee effects in metapopulations of seabirds. *Ecological Modelling*, *222*(17), 3061–3070.

Sekercioglu, C. H., Schneider, S. H., Fay, J. P., & Loarie, S. R. (2008). Climate change, elevational range shifts, and bird extinctions. *Conservation Biology*, *22*(1), 140–150.

Slatkin, M. (1993). Isolation by distance in equilibrium and non-equilibrium populations. *Evolution*, *47*(1), 264–279.

Smith, C. R., Grange, L. J., Honig, D. L., Naudts, L., Huber, B., Guidi, L., & Domack, E. (2012). A large population of king crabs in palmer deep on the west Antarctic peninsula shelf and potential invasive impacts. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1730), 1017–1026.

Speidel, L., Forest, M., Shi, S., & Myers, S. R. (2019). A method for genome-wide genealogy estimation for thousands of samples. *Nature Genetics*, *51*(9), 1321–1329.

Turner, J., Lachlan-Cope, T., Colwell, S., & Marshall, G. J. (2005). A positive trend in western Antarctic peninsula precipitation over the last 50 years reflecting regional and Antarctic-wide atmospheric circulation changes. *Annals of Glaciology*, *41*, 85–91.

Turner, J., Maksym, T., Phillips, T., Marshall, G. J., & Meredith, M. P. (2013). The impact of changes in sea ice advance on the large winter warming on the western Antarctic peninsula. *International Journal of Climatology*, *33*(4), 852–861.

Van der Auwera, G. A., & O'Connor, B. D. (2020). *Genomics in the cloud: Using docker, GATK, and WDL in Terra*. O'Reilly Media.

Van der Putten, W. H. (2012). Climate change, aboveground-belowground interactions, and species' range shifts. *Annual Review of Ecology, Evolution, and Systematics*, *43*, 365–383.

Vianna, J. A., Fernandes, F. A., Frugone, M. J., Figueiró, H. V., Pertierra, L. R., Noll, D., Bi, K., Wang-Claypool, C. Y., Lowther, A., Parker, P., Le Bohec, C., Bonadonna, F., Wienecke, B., Pistorius, P., Steinfurth, A., Burridge, C. P., Dantas, G. P. M., Poulin, E., Simison, W. B., … Bowie, R. C. (2020). Genome-wide analyses reveal drivers of penguin diversification. *Proceedings of the National Academy of Sciences*, *117*(36), 22303–22310.

Weir, J. T., Haddrath, O., Robertson, H. A., Colbourne, R. M., & Baker, A. J. (2016). Explosive ice age diversification of kiwi. *Proceedings of the National Academy of Sciences*, *113*(38), E5580–E5587.

White, T. A., Perkins, S. E., Heckel, G., & Searle, J. B. (2013). Adaptive evolution during an ongoing range expansion: the invasive bank vole (Myodes glareolus) in Ireland. *Molecular Ecology*, *22*(11), 2971–2985. Portico. https://doi.org/10.1111/mec.12343

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Herman, R. W., Clucas, G., Younger, J., Bates, J., Robinson, B., Reddy, S., Stepanuk, J., O'Brien, K., Veeramah, K., & Lynch, H. J. (2024). Whole genome sequencing reveals stepping-stone dispersal buffered against founder effects in a range expanding seabird. *Molecular Ecology*, *00*, e17282. https://doi.org/10.1111/mec.17282