Biological Journal of the Linnean Society, 2022, **XX**, 1–12. With 4 figures.

Fine-scale population genetic structure and barriers to gene flow in a widespread seabird (*Ardenna pacifica*)

RACHAEL W. HERMAN[1], BENJAMIN M. WINGER[2,3], DONNA L. DITTMANN[4] and MICHAEL G. HARVEY[2,3,*,†]

[1]*Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY, USA*
[2]*Museum of Zoology, University of Michigan, Ann Arbor, MI, USA*
[3]*Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA*
[4]*Museum of Natural Science, Louisiana State University, Baton Rouge, LA, USA*

Received 3 March 2022; revised 27 June 2022; accepted for publication 29 June 2022

Pelagic seabirds are highly mobile, reducing opportunities for population isolation that might promote differentiation and speciation. At the same time, many seabirds are philopatric, and their tendency to return to their natal islands to breed might reduce gene flow sufficiently to permit local adaptation and differentiation. To test the net impact of these competing processes, estimates of differentiation and gene flow based on comprehensive geographical sampling are required. We leveraged diverse source material to achieve comprehensive geographical sampling in a widespread seabird, the Wedge-tailed Shearwater (*Ardenna pacifica*). Using data from sequence capture and high-throughput sequencing of 2402 loci containing 20 780 single nucleotide polymorphisms, we tested for population differentiation and gene flow among breeding areas. We found little evidence of deep divergences within *A. pacifica* but were able to resolve fine-scale differentiation across island groups. This differentiation was sufficient to assign individuals sampled away from breeding areas to their likely source populations. Estimated effective migration surfaces revealed reduced migration between the Indian Ocean and Pacific Ocean, presumably owing to land barriers, and across the equatorial Pacific Ocean, perhaps associated with differences in breeding schedule. Our results reveal that, despite their mobility, pelagic seabirds can exhibit fine-scale population differentiation and reduced gene flow among ocean basins.

ADDITIONAL KEYWORDS: dispersal – marine biology – phylogeography – population genetics – seabirds – sequence capture

INTRODUCTION

Seabirds, like other marine taxa, tend to have broad, even cosmopolitan ranges. Their proclivity for nesting on isolated islands, however, produces highly patchy breeding distributions. Seabirds are extraordinarily mobile and can range widely while foraging and during the non-breeding season (Pinaud & Weimerskirch, 2007; Thaxter *et al.*, 2012), increasing the potential for dispersal and gene flow among breeding areas. At the same time, most seabird species are highly philopatric (Sagar *et al.*, 1998; Coulson, 2002, 2016), and their tendency to breed at their natal nesting grounds might reduce gene flow and present opportunities for population isolation and differentiation. These competing impacts of seabird life history can lead to unpredictable spatial patterns of genetic diversity that differ even among seabird species with otherwise similar biologies (Burg & Croxall, 2001; Friesen *et al.*, 2007; Milot *et al.*, 2008; Danckwerts *et al.*, 2021). Information on the degree and extent of gene flow and population differentiation is crucial, however, because these processes mediate resilience to local declines (Matthiopoulos *et al.*, 2005), adaptation to environmental differences (Dearborn *et al.*, 2003), and the propensity to form new species (Friesen, 2015). Additionally, an understanding of genetic differentiation supports taxonomic decisions and the delimitation of conservation units (Friesen *et al.*, 2007), in addition to the assignment of seabird

*Corresponding author. E-mail: mgharvey@utep.edu
†Current address: Department of Biological Sciences, The University of Texas at El Paso, El Paso, TX, USA.

bycatch, vagrant individuals or victims of mortality events to their source populations (Edwards *et al.*, 2001; Gómez-Díaz & González-Solís, 2007; Baetscher *et al.*, 2022).

Studies of genetic differentiation and gene flow require data from across a species distribution. However, many seabird studies lack comprehensive sampling (Friesen *et al.*, 2007). Preserved genetic samples in natural history collections are lacking for many species, which are represented primarily by specimens collected during exploratory voyages dating to the early 1900s or earlier. Fresh samples are difficult to obtain because the isolation and sheer number of remote islands on which many seabirds breed preclude field expeditions to obtain new material from a sufficient number of breeding areas. New genomic methods have the potential to unlock genetic information stored in diverse types of samples. Samples that might be useful include not only tissue and blood preserved using modern methods, but also feathers, old museum specimens, bones and subfossil remains. Sequence capture, in particular, allows for the recovery of many parts of the genome from degraded samples, because its strategy of enriching DNA by hybridizing short RNA probes can recover even small fragments (Bi *et al.*, 2013; McCormack *et al.*, 2016). The markers captured with sequence capture, even sequence capture targeting conserved genomic loci, are informative for analyses of differentiation and gene flow at shallow timescales, such as within and among populations (Smith *et al.*, 2014). Sequence capture, combined with material from diverse and degraded samples, might permit more comprehensive sampling in genetic studies of seabirds and other organisms for which fresh samples are difficult to obtain.

Here, we present a population genomic study of a widespread seabird using a sequence capture approach and a combination of samples from modern tissues, toe pads from museum specimens and dried whole blood. Our study species, the wedge-tailed shearwater (*Ardenna pacifica* Gmelin, 1789) is a medium-sized, highly pelagic Procellariiform bird with an extensive breeding distribution on islands across the subtropical and tropical Pacific and Indian oceans (Carboneras, 1992). Wedge-tailed shearwaters range widely during the non-breeding season (Catry *et al.*, 2009; Howell, 2012; Ravache *et al.*, 2020; Fig. 1). They exhibit variation in size and in the frequencies of two or more plumage morphs (white-bellied and dark-bellied, with a possible intermediate morph) that corresponds loosely to geography (Murphy, 1951). Two subspecies are generally recognized: the widespread *Ardenna pacifica chlororhyncha* is replaced by the nominate subspecies on Norfolk Island, Kermadec Island, Fiji and Tonga (Murphy, 1951; Carboneras, 1992; Dickinson & Remsen, 2013). Although phylogenetic data are available for the wedge-tailed shearwater (Penhallurick & Wink, 2004; Obiol *et al.*, 2022), there is no prior study of genetic diversity within the species.

We use wide geographical sampling to investigate the population genetics of wedge-tailed shearwaters. Initially, we evaluate the information content recovered across sample types using sequence capture. We then test for population genetic structure across

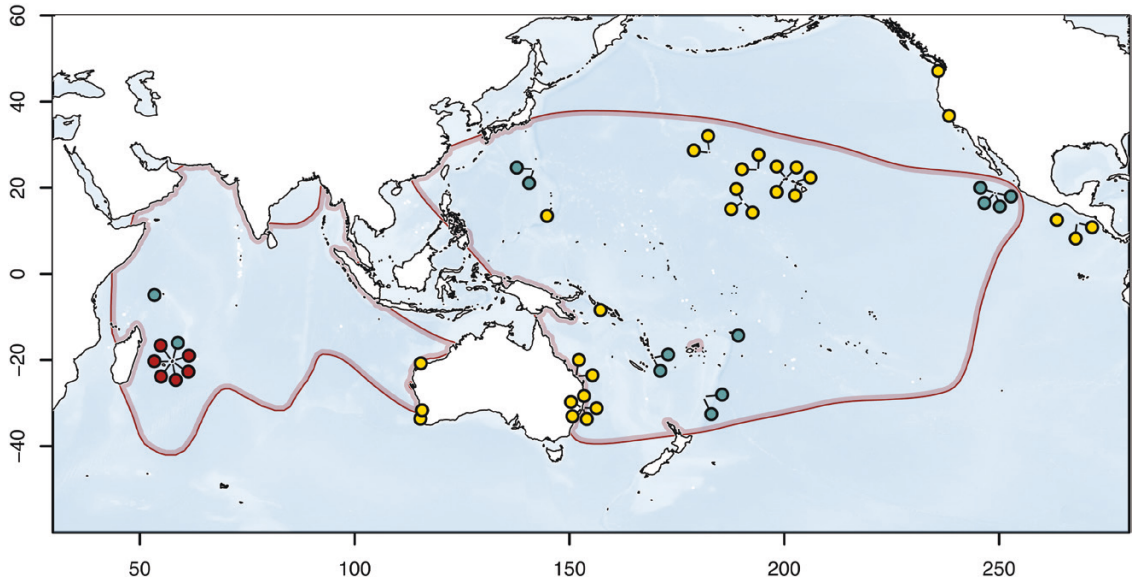

Figure 1. A map showing the distribution of wedge-tailed shearwaters (red outline) and sampling localities used in this study, distinguished by source material (yellow, preserved tissue; red, whole blood; blue, toe pad). For sites with multiple samples, markers are separated from one another, with lines pointing to the locality.

the distribution of wedge-tailed shearwaters, assess patterns of gene flow among breeding areas and evaluate evidence for the role of geographical barriers in impeding gene flow and facilitating differentiation. Finally, we conduct a preliminary examination of genetic diversity of each breeding area and associations between allelic variation across sampled loci and the plumage morphs of individuals.

MATERIAL AND METHODS

Sampling and laboratory methods

We sampled wedge-tailed shearwaters spanning the entire breeding distribution (Fig. 1; Supporting Information, Table S1). We collected new tissue samples and prepared voucher specimens ($N = 7$) from individuals salvaged in the north-western Hawaiian islands and Johnston Atoll, obtained existing frozen tissue samples ($N = 22$) that are associated with voucher specimens from museum collections, collected toe pad material from round skin specimens ($N = 13$) for populations lacking vouchered tissue material, and obtained unvouchered blood samples ($N = 6$) from a conservation project run in Mauritius by the Durrell Wildlife Conservation Trust and Mauritius Wildlife Foundation. These samples were all from breeding localities except for two collected off the coast of Guatemala; one off the coast of Mexico; one off the coast of Washington, USA; and one from a grounded individual in California, USA. We also sampled two vouchered tissues from Buller's shearwater (*Ardenna bulleri*), the sister species of the wedge-tailed shearwater (Penhallurick & Wink, 2004), to be used as an outgroup.

We extracted whole genomic DNA from all samples using a DNeasy blood and tissue kit and its standard protocols (Qiagen, Valencia, CA, USA). Blood samples were stored in ethanol and were air-dried before extraction. We conducted toe pad extractions separately from those for tissue and blood in a dedicated ancient DNA laboratory at the University of Michigan Biodiversity Lab. Toe pad extractions were completed on days without prior access to laboratories housing fresh samples and using clothing, equipment and reagents that were not exposed to fresh material. We sliced toe pad samples finely and adjusted the standard DNeasy protocol by extending the digestion step to 48 h, adding 40 µL of proteinase K, warming the elution buffer to 56 °C, allowing the elution buffer sit in filter tubes for 30 min, and eluting twice with 75 µL of elution buffer. We quantified DNA using a Qubit fluorometer (Thermo-Fisher, Waltham, MA, USA). We then sent samples to Rapid Genomics (Gainesville, FL, USA) for sequence capture and Illumina sequencing (Illumina, San Diego, CA, USA) following the protocol described by Faircloth *et al.* (2012) and Smith *et al.* (2014). Our capture array included 4715 RNA probes targeting 2321 ultraconserved elements and 96 exons (Harvey *et al.*, 2017). Samples were multiplexed at 100 samples per lane and sequenced using a paired-end HiSeq 2500 run.

Sequence read assembly and genotyping

Reads were demultiplexed by Rapid Genomics using custom scripts with strict barcode matching. We then cleaned reads using ILLUMIPROCESSOR (Faircloth, 2013). We used the seqcap_pop pipeline (Harvey *et al.*, 2017) for dataset assembly, substituting sqcl (Singhal *et al.*, 2017) scripts for the steps in which contigs are mapped to RNA probe sequences and a contig from one sample at each locus is selected for inclusion in the pseudo-reference genome. This pipeline depends on software including VELVET (Zerbino & Birney, 2008) and VELVETOPTIMISER (Gladman & Seemann, 2008) for contig assembly in each sample, BLAT (Kent 2002) for matching contigs to probes to generate the pseudo-reference genome, BWA (Li & Durbin, 2009) for mapping reads to the pseudo-reference genome, SAMTOOLS (Li *et al.*, 2009) and PICARD (Broad Institute, Cambridge, MA, USA) for processing read pile-ups, GATK (McKenna *et al.*, 2010) for calling and phasing alleles and filtering variants, and MAFFT (Katoh *et al.*, 2005) with custom seqcap_pop scripts to obtain sequence alignments. Using the GATK VARIANTFILTRATION tool, we removed variants with overall quality < 99.9% accuracy, with quality normalized by read depth (QD) < 5.0, and with ≥ 10% of the reads exhibiting a mapping quality score of zero.

Estimation of population genetic structure

Initially, we examined genetic differentiation among geographical localities using principal components analysis (PCA). We ran PCAs of all single nucleotide polymorphisms (SNPs), both with and without the outgroup individuals included. We then used discriminant analysis of principal components (DAPC; Jombart *et al.*, 2010) to cluster ingroup individuals into discrete populations. The DAPC uses discriminant analysis to partition samples in order to maximize the ratio of between-group to within-group genetic difference. Both PCA and DAPC were run using the *adegenet* package (Jombart, 2008) in R (R Core Team, 2017), following the authors' recommendations. We used STRUCTURE (Pritchard *et al.*, 2000) both to cluster individuals into populations and to estimate admixture among populations. We used the linkage model and specified the distance between linked sites and provided phase information. We ran ten individual runs with 200 000 sampling iterations after

50 000 burn-in iterations at each value of number of populations (K) from one to ten. We assessed convergence by examining likelihood and parameter values within and across runs, and we determined the best K value using the Evanno *et al.* (2005) method in STRUCTURE HARVESTER (Earl & vonHoldt, 2012). We combined assignment probabilities across runs for summary and visualization using CLUMPP (Jakobsson & Rosenberg, 2007).

We also assessed differentiation by calculating the fixation index (F_{ST}) between localities using the formulas of Weir & Cockerham (1984) in VCFTOOLS (Danecek *et al.*, 2011) and the formula of Reich *et al.* (2009), which is robust to small sample sizes (Willing *et al.*, 2012). Finally, we used RAxML v.8.2.11 (Stamatakis, 2014) to estimate a phylogenetic tree from concatenated nuclear SNPs. We used the GTRGAMMA substitution model and used 100 rapid bootstrap replicates to evaluate support. For nuclear SNPs, we assigned the two alleles from each individual randomly to one of two haplotypes for that sample and built trees both for all SNPs and for only those SNPs with no missing data.

Estimation of migration and genetic diversity

We examined spatial variation in migration among populations (demes) and diversity within populations using estimation of effective migration surfaces (EEMS; Petkova *et al.*, 2016), which represents genetic differentiation as a function of migration rates among geo-referenced genetic samples and enables visual analytics of potential barriers to gene flow in geographical space. For EEMS, we removed the outgroup samples and the vagrant individuals to focus solely on breeding populations (N = 43 samples). We used PLINK v.1.9 (Chang *et al.*, 2015) to convert VCF format to binary BED files. Given that missing data, especially non-randomly distributed missing data, can be problematic in EEMS, we reduced the SNP set to only those 2495 SNPs with high-quality genotypes in all 43 individuals. We calculated a distance matrix from these SNPs using the program bed2diff_v1 from EEMS. We then ran EEMS using either 200 or 400 demes and a Markov chain Monte Carlo with 10 million burn-in iterations followed by 50 million sampling iterations. We conducted a series of runs, each time examining traces within and across replicate runs to evaluate convergence and using proposal acceptance rates to adjust the proposal variance of each parameter for subsequent runs. The settings for the final run, which was based on a model with 400 demes, are presented in the Supporting Information (Table S2). Migration rates were presented on a $\log_{10}$ scale relative to the overall migration rate across the habitat, such that a rate of one was equivalent to a tenfold higher migration rate relative to the average. We supplemented the EEMS estimates of genetic diversity, which are based only on divergence between samples within a population and on SNPs without missing data, with simple summaries of nucleotide diversity (calculated in VCFTOOLS) and observed heterozygosity (calculated in ROHAN; Renaud *et al.*, 2019) at each locality.

Morph-associated SNPs

We scanned for any sequence capture loci associated with the wedge-tailed shearwater plumage polymorphism using a subset of 31 individuals for which we were able to obtain morph information (17 dark morphs and 14 light morphs). We used VCFTOOLS to convert the VCF containing both autosomal and sex-linked data for all individuals to formats readable by PLINK. We then used PLINK with the command '--assoc' to run a simple association test comparing allele frequencies at all SNPs between the two morphs. Given that morph frequencies vary geographically, we also ran a stratified association analysis (command '--mh') controlling for population structure based on the set of 12 breeding populations identified in Figure 3.

RESULTS

We recovered an average of 3.78 (SD = 1.51) million reads per sample after cleaning (Supporting Information, Table S3). Variant calling identified 20 780 high-quality SNPs across 2373 loci. An average of 16 772 (SD = 5398) SNPs had high-quality genotypes in each sample, resulting in a matrix with 19.3% of missing data. Read depth averaged 24.4 (SD = 11.6) at genotyped SNPs (Supporting Information, Table S3). Read depths and the number of genotyped SNPs were similar across blood and tissue samples, but both read depth ($t = -6.97$, $P < 0.01$) and the number of genotyped SNPs ($t = -15.59$, $P < 0.01$) were lower in toe pads from museum skins (Fig. 2).

Population genetic structure

Visualization of genetic variation across samples using PCA revealed a surprising pattern, in which samples were well separated in multidimensional genotype space, but this separation did not correspond to geography. Further investigation revealed that separation on the first principal component axis (PC1) was associated with sex (Supporting Information, Fig. S1). Based on the knowledge that many bird species have non-degenerate W chromosomes, we hypothesized that the Z-linked locus assemblies

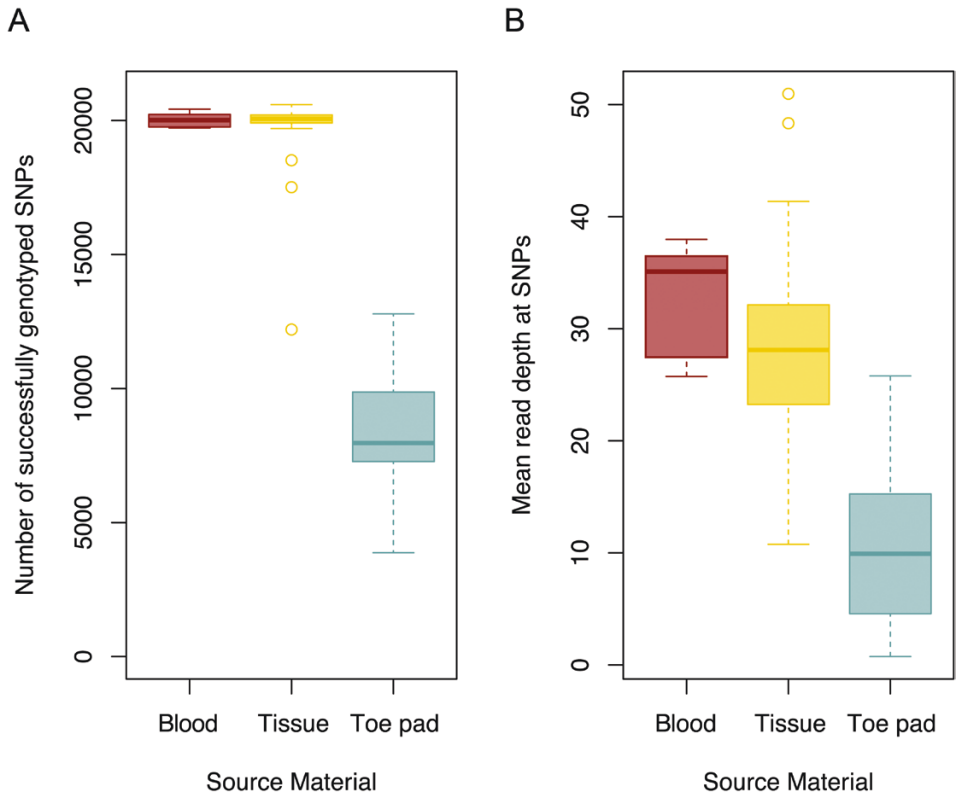

Figure 2. Comparisons of data recovery from samples with different source materials, showing both the number of single nucleotide polymorphisms (SNPs; A) and the mean read depth at those SNPs (B).

in female individuals might contain high levels of contamination from W-linked paralogues producing spurious alternative alleles. To confirm this, we mapped all loci to the zebra finch bTaeGut2.pat.W (Warren *et al.*, 2010) and chicken Gallus_gallus-5.0 (Warren *et al.*, 2017) genomes and identified loci that mapped to either autosomal or sex-linked regions in both cases. We found that females, which are heterogametic and should be homozygous at sex-linked sites, averaged 4.30 times greater heterozygous calls at sex-linked SNPs than at autosomal SNPs, compared with 0.75 times the heterozygosity at sex-linked vs. autosomal SNPs in males. This suggests that paralogous alleles are included in female assemblies at sex-linked loci, presumably owing to incorporation of W chromosome reads. In order to avoid spurious downstream effects, we therefore removed sex-linked loci for all analyses except for the morph association genome scan. This resulted in 19 127 autosomal SNPs across 2206 loci.

Principal components analysis of autosomal data revealed low separation of the sexes in multidimensional genotype space, but clear fine-scale geographical separation (Fig. 3). Separation between Buller's and wedge-tailed shearwaters was high in comparison to separation between samples within species (Supporting Information, Fig. S2). However, the PCA of only wedge-tailed shearwaters clearly separated populations from many of the breeding regions sampled. This was true for the analysis of the full SNP dataset and for an analysis reducing the impact of linkage by using one randomly selected SNP per locus (Supporting Information, Fig. S3). Breeding regions in which samples were overlapping or in close proximity in multidimensional space, for example Hawaii and Guam or eastern Australia and the Solomon Islands, were generally in close proximity geographically. Principal component 1 appeared to separate individuals largely by latitude, with the North Pacific Ocean samples at one end and the South Pacific Ocean and Indian Ocean samples at the other. Principal component 2 (PC2), in turn, separated South Pacific Ocean samples from Indian Ocean samples. Samples from the tropical Pacific were clustered in the middle of both PC1 and PC2. Notably, migrant or vagrant shearwaters collected at sea or wrecked onshore along the Pacific coast of North America largely clustered with samples from Hawaii and Guam, their presumed source population. However, one individual collected on a beach on 9 September 2005 in Gray's Harbor County, WA, USA (UWBM 63735) clustered with eastern Australian and Solomon Islands samples and, presumably, originated

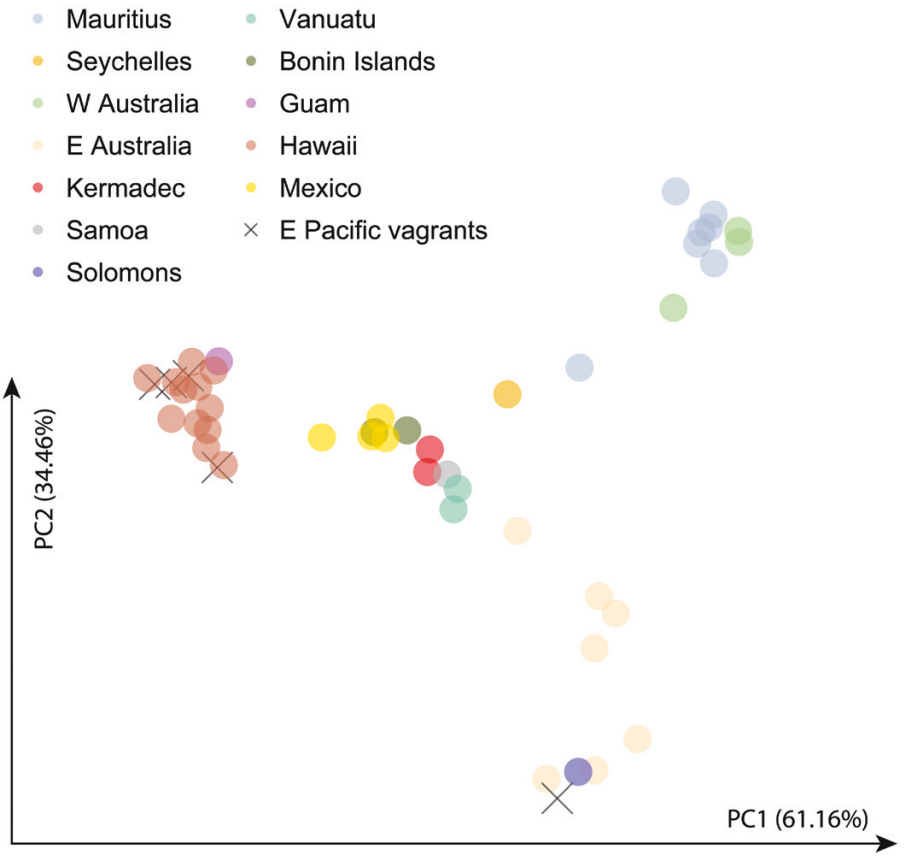

Figure 3. The results of a principal components analysis of genomic SNPs in wedge-tailed shearwaters, with samples coloured by geographical origin. Vagrant individuals collected away from breeding localities are marked with an 'X'.

from a south-western Pacific breeding area. The F_{ST} results largely corroborated those of the PCA but appeared, even using the Reich *et al.* (2009) estimator, to vary according to the relative sizes of the populations in a given comparison (Supporting Information, Tables S4–S6).

Despite the fine-scale geographical differentiation evident in the PCA, clustering of samples into larger, discrete populations was less clear. Bayesian information criteria across different numbers of clusters (K) in DAPC provided support for two or three populations (Supporting Information, Fig. S4). With $K = 2$, these clusters separated the North Pacific samples from the South Pacific plus Indian Ocean individuals. However, $K = 3$ combined the North Pacific samples with much of the tropical Pacific, separating only eastern Australia and the Solomons Islands into a second population, with Indian Ocean samples in a third. Although the best value of K in STRUCTURE was three, no K value between two and ten resulted in strong differences in assignment probabilities across samples (Supporting Information, Fig. S5). At $K = 2$, a step in probabilities between a cluster containing individuals from the North Pacific and another containing individuals from the South Pacific plus Indian Ocean was evident, but the difference in assignment probabilities to the two clusters was < 40% across individuals. Higher values of K did not reveal additional geographical clustering. RAxML trees of concatenated SNPs separated the North Pacific, South Pacific and Indian Oceans into separate clades, but the branch lengths between them were relatively short, and bootstrap support values were low (Supporting Information, Fig. S6).

Migration among populations

The EEMS revealed two major regions of reduced rates of migration (Fig. 4). Migration rates between the Indian and Pacific Ocean were at least tenfold lower (1 $\log_{10}$ unit) than the average. This area of reduced migration coincides with the continental land barriers created by Asia and Australia. Low migration was even inferred between western and eastern Australian populations, which are geographically proximate and separated by a short distance by sea. Migration rates across the equator in the Pacific Ocean were also lower than the average across the distribution by ≥ 30 times

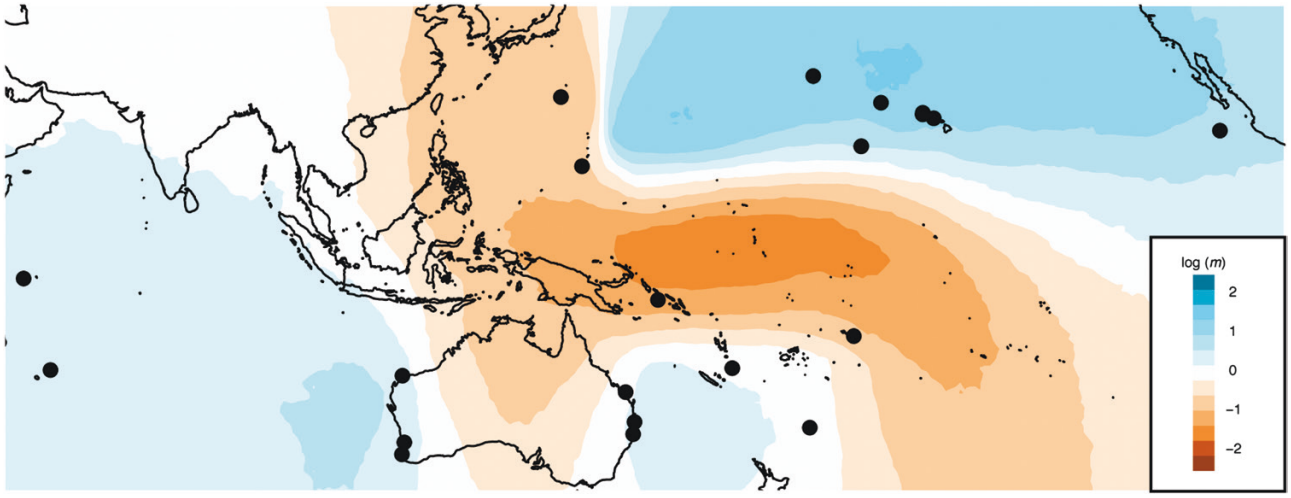

Figure 4. A smooth contour map of the posterior mean migration rate from estimation of effective migration surfaces (EEMS).

(1.5 $\log_{10}$ units). This area of reduced migration does not coincide with a land barrier or expanse of island-free ocean. Both areas of reduced migration inferred with EEMS correspond to areas across which differentiation was observed in the PCA, DAPC and RAxML analyses.

Population sizes and diversity

The diversity rate (q) from EEMS represents the expected within-deme coalescence time, or the amount of diversity within a population attributable to differences between individuals while accounting for migration. The diversity rate was much higher, however, in the Bonin Islands and Kermadec populations than elsewhere (Supporting Information, Fig. S7). The samples from these two localities (each of which was represented by two toe pad samples) were of particularly poor quality based on read depths at SNPs (Supporting Information, Table S3). Both population nucleotide diversity (π) and the proportion of SNPs that were heterozygous showed an opposing pattern (Supporting Information, Fig. S8), with lower levels of diversity in individuals and populations represented by only toe pads. However, calling heterozygotes using ROHan resulted in more similar estimates of heterozygosity across sample types (Supporting Information, Fig. S9). Within a particular sample type, we observed little variation among populations in nucleotide diversity or observed heterozygosity.

Morph-associated SNPs

We identified five SNPs that exhibited a strong association with plumage morph at an α-level of 0.05 (Supporting Information, Fig. S10). These SNPs were not significant at an α-level of 0.01, and these associations were not present after controlling for population structure in a stratified analysis (Supporting Information, Fig. S11). This is perhaps not surprising given the wide spacing of the target conserved loci across the genome, small sample of individuals, and lack of within-population plumage polymorphism.

DISCUSSION

We found marked genetic differentiation among populations of wedge-tailed shearwaters breeding on different islands and archipelagos. Discrimination was sufficient to identify the regions serving as the sources for five non-breeding or vagrant individuals collected along the Pacific coast of North America. The differentiation we identified with sequence capture should therefore also be sufficient to identify the source populations of bycatch from fisheries, a challenge with traditional genetic markers (Edwards *et al.*, 2001; Gómez-Díaz & González-Solís, 2007; Techow *et al.*, 2016; Ellis *et al.*, 2020). Despite this fine-scale genetic structure, we did not find evidence for deep evolutionary divergences within wedge-tailed shearwaters. The separation of populations in the PCA, for example, was much less than between wedge-tailed and Buller's shearwaters or than between closely related avian sister species in some other studies using similar genomic datasets (e.g. Irwin *et al.*, 2018; Linck *et al.*, 2020). However, STRUCTURE and RAxML results did identify at least some division between two or three major groups. The North Pacific was distinct from the South Pacific plus Indian Ocean, and the latter two

areas were also somewhat separated in the RAxML results and based on EEMS migration patterns. Further study of morphological variation might reveal differences warranting subspecies status for these three groups. However, these genetic groups are highly discordant with current subspecies taxonomy, which was based primarily on morphometric characters and is in clear need of revision (Murphy, 1951; Dickinson & Remsen, 2013). Regardless, North Pacific, South Pacific and Indian Ocean populations merit treatment as separate conservation units based on their genetic distinctness and infrequent exchange of migrants.

Analysis of migration detected reduced gene flow corresponding roughly to the genetic subdivisions identified in differentiation analyses, but the explicitly spatial EEMS results provide additional insight into what might drive these discontinuities. The EEMS is agnostic with respect to geography (Petkova, 2017), hence it is of interest when areas of reduced migration overlap with apparent geographical barriers. As predicted, there is limited migration between the Pacific and Indian Oceans, presumably owing to land barriers, because pelagic seabirds generally avoid flying over landmasses (Steeves *et al.*, 2005; Friesen *et al.*, 2007; Lombal *et al.*, 2020). Surprisingly, migration is also very limited across the equator in the Pacific, suggesting that birds from the North Pacific are not exchanging genes with birds from the South Pacific. This might be attributable to a lack of movement between these areas, but the distance between North and South Pacific populations is relatively small, and wedge-tailed shearwaters from the North Pacific are known to migrate to the South Pacific in the non-breeding season (Carboneras, 1992; Howell, 2012). In addition, we identified an instance of a vagrant individual in Washington state that was assigned to the south-west Pacific populations, suggesting the possibility of transequatorial movement in the other direction. Instead, the equatorial barrier to migration might be attributable to historical or ecological factors, including ocean regime, currents and wind dynamics, environmental differences or breeding phenology (Friesen, 2015; Lombal *et al.*, 2020; Torres *et al.*, 2021). For example, there is evidence that a mismatch in the timing of breeding between hemispheres plays a role. Most breeding of wedge-tailed shearwaters in Hawaii is between March and November, whereas breeding in southern Australia is from August to May (Carboneras, 1992; Brooke, 2004). Differences in the timing of breeding have been detected within other Procellariiform species, in some cases associated with genetically differentiated populations breeding in the same areas (allochrony; Friesen *et al.*, 2007; Rayner *et al.*, 2011; Garrett *et al.*, 2020). Discordant annual life cycles might be a more general explanation for reduced gene flow and differentiation among populations in seabirds.

We found different patterns of genetic diversity across breeding populations based on EEMS and summary statistics (nucleotide diversity and observed heterozygosity). These different patterns appear to be driven by biases in the data introduced by the different sampling sources. Toe pad samples and populations represented only by toe pads had lower heterozygosity and nucleotide diversity than those sampled with tissue and blood. This is likely to be because degradation of DNA combined with low read coverage in these poor-quality samples resulted in a failure to recover some alleles, producing a downward bias in both summary statistics. Accurate genotyping, particularly of heterozygotes, can be challenging below about 20 times coverage (Choi *et al.*, 2009; Green *et al.*, 2009), and most of our toe pad samples fell within this range (Fig. 2; Supporting Information, Table S3). Conversely, the populations sampled with particularly degraded material had higher diversity in the EEMS results. The diversity rate in EEMS reflects the expected dissimilarity between individuals within a locality (Petkova, 2017), and therefore could be impacted greatly by contamination in a single individual. The two populations with higher diversity in the EEMS results (Bonin Islands and Kermadec) included the two poorest-quality samples based on sequence recovery (Supporting Information, Table S3), and spurious private alleles are likely to have driven their higher diversity estimates. However, these spurious alleles do not appear to have impacted the migration surface, which is estimated from differences between sites (Petkova, 2017), because the Bonin Islands and Kermadec do not show evidence of reduced migration with other localities (Fig. 4).

These issues revealed here with genetic diversity estimated from low-quality samples might be alleviated in the future by optimizing laboratory protocols or implementing more robust genotyping models for low-coverage data in bioinformatics pipelines. Indeed, calling heterozygotes using a method that more explicitly factored in sequence coverage and missing data (ROHAN) resulted in estimates that appeared to less be biased among sample sources. However, strand or allele bias in DNA preservation or capture will be hard to eliminate entirely with downstream approaches. In datasets with mixed sample types, researchers might prefer to focus on analysing patterns of differentiation and gene flow rather than genetic diversity. The similarity in some measures of genetic diversity within a sample type in our data along with the lack of strong patterns in the ROHAN heterozygosity results suggest that wedge-tailed shearwaters exhibit little variation in genetic diversity or effective population size across their breeding distribution. This result is perhaps not surprising given the large population size of this species (Birdlife International, 2020) and the relatively extensive gene flow we observed among localities.

We have demonstrated that sequence capture and Illumina sequencing can be combined with a diverse set of sample sources to complete a comprehensive

phylogeographical and population genetic study of a widespread seabird. Although more degraded samples, toe pads from old specimens, had fewer data recovered than blood and tissue samples, those data were sufficient for examining patterns of genetic differentiation and gene flow (albeit less so for genetic diversity). In the wedge-tailed shearwater, we were able to identify fine-scale differentiation among breeding areas in addition to somewhat deeper divergences and reduced migration between the North Pacific, South Pacific and Indian Oceans. These results are significant because they will enable identification of the sources of future bycatch or vagrants; they will be useful in any revision of conservation units in the species, and they provide a framework for future work on population dynamics, adaptation and speciation. We expect that studies of differentiation and gene flow using sequence capture will prove a powerful tool for research in seabirds and other organisms for which comprehensive sampling is challenging.

ACKNOWLEDGEMENTS

For samples and specimen information, we thank the University of Michigan Museum of Zoology (Janet Hinshaw); the American Museum of Natural History (Paul Sweet, Thomas Trombone, Peter Capainolo and Brian Tilston Smith); the Louisiana State University Museum of Natural Science (Robb Brumfield and Fred Sheldon); the University of Florida Natural History Museum (Andy Kratter); the Smithsonian National Museum of Natural History (Christopher Milensky, Brian Schmidt and Jacob Saucier); the San Diego Natural History Museum (Phil Unitt); San Diego State University (Kevin Burns); the University of Washington Burke Museum (Sharon Birks); the Western Australian Museum (Rebecca Bray and Ron Johnstone); the California Academy of Sciences (Maureen Flannery); the Durrell Wildlife Conservation Trust and Mauritian Wildlife Foundation (Nik Cole, Carl Jones and Vikash Tatayah); and the International Zoo Veterinary Group pathology division. For permissions, we are indebted to the National Parks and Conservation Service, Ministry of Agro-Industry, Government of Mauritius. DNA extractions were conducted at the University of Michigan Biodiversity Laboratory. Bioinformatics and analyses were conducted with high-performance computing provided by University of Michigan Advanced Research Computing—Technology Services (ARC-TS) and the Derryberry Lab and National Institute for Computational Sciences (NICS) at the University of Tennessee. Funding was provided by National Science Foundation Postdoctoral Research Fellowship in Biology to M.G.H. (DBI-1523893) and the University of Michigan Museum of Zoology. We thank the Lynch Lab at Stony Brook University for discussion and feedback on this manuscript. We have no conflicts of interest to declare.

DATA AVAILABILITY

The raw sequence data used in this work (Herman *et al.*, 2022) are available in the NCBI short read archive (BioProject PRJNA849173).

REFERENCES

Baetscher DS, Beck J, Anderson EC, Ruegg K, Ramey AM, Hatch S, Nevins H, Fitzgerald SM, Garza JC. 2022. Genetic assignment of fisheries bycatch reveals disproportionate mortality among Alaska Northern Fulmar breeding colonies. *Evolutionary Applications* **15:** 447–458.

Bi K, Linderoth T, Vanderpool D, Good JM, Nielsen R, Moritz C. 2013. Unlocking the vault: next-generation museum population genomics. *Molecular Ecology* **22:** 6018–6032.

BirdLife International. 2020. *Species factsheet:* Ardenna pacifica. Available at: http://datazone.birdlife.org/species/factsheet/wedge-tailed-shearwater-ardenna-pacifica

Brooke M. 2004. *Albatrosses and petrels across the world*. Oxford: Oxford University Press.

Burg TM, Croxall JP. 2001. Global relationships amongst black-browed and grey-headed albatrosses: analysis of population structure using mitochondrial DNA and microsatellites. *Molecular Ecology* **10:** 2647–2660.

Carboneras C. 1992. Family Procellariidae (petrels and shearwaters). In: del Hoyo J, Elliott A, Sargatal J, eds. *Handbook of the birds of the world*, *Vol. 1*. Barcelona: Lynx Edicions, 216–257.

Catry T, Ramos JA, Le Corre M, Phillips RA. 2009. Movements, at-sea distribution and behaviour of a tropical pelagic seabird: the wedge-tailed shearwater in the western Indian Ocean. *Marine Ecology Progress Series* **391:** 231–242.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4:** s13742-015-0047-8.

Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, Nayir A, Bakkaloğlu A, Özen S, Sanjad S, Nelson-Williams C. 2009. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proceedings of the National Academy of Sciences of the United States of America* **106:** 19096–19101.

Coulson JC. 2002. Colonial breeding in seabirds. In: Schreiber EA, Burger J, eds. *Biology of marine birds*. Boca Raton: CRC Press, 87–113.

Coulson JC. 2016. A review of philopatry in seabirds and comparisons with other waterbird species. *Waterbirds* **39:** 229–240.

Danckwerts DK, Humeau L, Pinet P, McQuaid CD, Le Corre M. 2021. Extreme philopatry and genetic

diversification at unprecedented scales in a seabird. *Scientific Reports* **11:** 6834.

Danecek P, **Auton A**, **Abecasis G**, **Albers CA**, **Banks E**, **DePristo MA**, **Handsaker RE**, **Lunter G**, **Marth GT**, **Sherry ST**, **McVean G. 2011.** The variant call format and VCFtools. *Bioinformatics* **27:** 2156–2158.

Dearborn DC, **Anders AD**, **Schreiber EA**, **Adams RM**, **Mueller UG. 2003.** Inter-island movements and population differentiation in a pelagic seabird. *Molecular Ecology* **12:** 2835–2843.

Dickinson EC, **Remsen JV. 2013.** *The Howard & Moore complete checklist of the birds of the World, Fourth Edition, Volume 1: non-passerines*. Eastbourne: Aves Press.

Earl DA, **von Holdt BM. 2012.** STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4:** 359–361.

Edwards SV, **Silva MC**, **Burg T**, **Friesen V**, **Warheit KI. 2001.** Molecular genetic markers in the analysis of seabird bycatch populations. In: Melvin EF, Parrish JK, eds. *Seabird bycatch: trends, roadblocks and solutions*. Fairbanks: University of Alaska Sea Grant, 115–140.

Ellis CD, **Jenkins TL**, **Svanberg L**, **Eriksson SP**, **Stevens JR. 2020.** Crossing the pond: genetic assignment detects lobster hybridisation. *Scientific Reports* **10:** 7781.

Evanno G, **Regnaut S**, **Goudet J. 2005.** Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14:** 2611–2620.

Faircloth BC. 2013. *illumiprocessor: a trimmomatic wrapper for parallel adapter and quality trimming*. Available at: http://dx.doi.org/10.6079/J9ILL

Faircloth BC, **McCormack JE**, **Crawford NG**, **Harvey MG**, **Brumfield RT**, **Glenn TC. 2012.** Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology* **61:** 717–726.

Friesen VL. 2015. Speciation in seabirds: why are there so many species…and why aren't there more? *Journal of Ornithology* **156:** 27–39.

Friesen VL, **Burg TM**, **McCoy KD. 2007.** Mechanisms of population differentiation in seabirds. *Molecular Ecology* **16:** 1765–1785.

Garrett LJH, **Myatt JP**, **Sadler JP**, **Dawson DA**, **Hipperson H**, **Colbourne JK**, **Dickey RC**, **Weber SB**, **Reynolds SJ. 2020.** Spatio-temporal processes drive fine-scale genetic structure in an otherwise panmictic seabird population. *Scientific Reports* **10:** 20725.

Gladman S., **Seemann T. 2008.** *VelvetOptimiser, Version 2.2.4*. Available at: https://github.com/tseemann/VelvetOptimiser

Gómez-Díaz E, **González-Solís J. 2007.** Geographic assignment of seabirds to their origin: combining morphologic, genetic, and biogeochemical analyses. *Ecological Applications* **17:** 1484–1498.

Green RE, **Briggs AW**, **Krause J**, **Prüfer K**, **Burbano HA**, **Siebauer M**, **Lachmann M**, **Pääbo S. 2009.** The Neandertal genome and ancient DNA authenticity. *The EMBO Journal* **28:** 2494–2502.

Harvey MG, **Aleixo A**, **Ribas CC**, **Brumfield RT. 2017.** Habitat association predicts genetic diversity and population divergence in Amazonian birds. *The American Naturalist* **190:** 631–648.

Herman RW, **Winger BM**, **Dittmann DL**, **Harvey MG. 2022.** *Data from: Fine-scale population genetic structure and barriers to gene flow in a widespread seabird (*Ardenna pacifica*)*. NCBI SRA (doi pending acceptance).

Howell SN. 2012. *Petrels, albatrosses, and storm-petrels of North America: a photographic guide*. Princeton: Princeton University Press.

Irwin DE, **Milá B**, **Toews DP**, **Brelsford A**, **Kenyon HL**, **Porter AN**, **Grossen C**, **Delmore KE**, **Alcaide M**, **Irwin JH. 2018.** A comparison of genomic islands of differentiation across three young avian species pairs. *Molecular Ecology* **27:** 4839–4855.

Jakobsson M, **Rosenberg NA. 2007.** CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23:** 1801–1806.

Jombart T. 2008. *adegenet*: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24:** 1403–1405.

Jombart T, **Devillard S**, **Balloux F. 2010.** Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* **11:** 94.

Katoh K, **Kuma KI**, **Toh H**, **Miyata T. 2005.** MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research* **33:** 511–518.

Kent WJ. 2002. BLAT—the BLAST-like alignment tool. *Genome Research* **12:** 656–664.

Li H, **Durbin R. 2009.** Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25:** 1754–1760.

Li H, **Handsaker B**, **Wysoker A**, **Fennell T**, **Ruan J**, **Homer N**, **Marth G**, **Abecasis G**, **Durbin R. 2009.** The sequence alignment/map format and SAMtools. *Bioinformatics* **25:** 2078–2079.

Linck E, **Freeman BG**, **Dumbacher JP. 2020.** Speciation and gene flow across an elevational gradient in New Guinea kingfishers. *Journal of Evolutionary Biology* **33:** 1643–1652.

Lombal AJ, **O'dwyer JE**, **Friesen V**, **Woehler EJ**, **Burridge CP. 2020.** Identifying mechanisms of genetic differentiation among populations in vagile species: historical factors dominate genetic differentiation in seabirds. *Biological Reviews* **95:** 625–651.

Matthiopoulos J, **Harwood J**, **Thomas LEN. 2005.** Metapopulation consequences of site fidelity for colonially breeding mammals and birds. *Journal of Animal Ecology* **74:** 716–727.

McCormack JE, **Tsai WL**, **Faircloth BC. 2016.** Sequence capture of ultraconserved elements from bird museum specimens. *Molecular Ecology Resources* **16:** 1189–1203.

McKenna A, **Hanna M**, **Banks E**, **Sivachenko A**, **Cibulskis K**, **Kernytsky A**, **Garimella K**, **Altshuler D**, **Gabriel S**, **Daly M**, **DePristo MA. 2010.** The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20:** 1297–1303.

Milot E, **Weimerskirch H**, **Bernatchez L. 2008.** The seabird paradox: dispersal, genetic structure and population

dynamics in a highly mobile, but philopatric albatross species. *Molecular Ecology* **17:** 1658–1673.

Murphy RC, **Irving S. 1951.** The populations of the wedge-tailed shearwater (*Puffinus pacificus*). *American Museum Novitates* **1512:** 1–21.

Obiol JF, **James HF**, **Chesser RT**, **Bretagnolle V**, **González-Solís J**, **Rozas J**, **Welch AJ**, **Riutort M. 2022.** Palaeoceanographic changes in the late Pliocene promoted rapid diversification in pelagic seabirds. *Journal of Biogeography* **49:** 171–188.

Penhallurick J, **Wink M. 2004.** Analysis of the taxonomy and nomenclature of the Procellariiformes based on complete nucleotide sequences of the mitochondrial cytochrome *b* gene. *Emu* **104:** 125–147.

Petkova D. 2017. *Instruction Manual - EEMS: a method to visualize patterns of non-stationary isolation by distance, Version 0.0.9000*. Available at: https://github.com/dipetkov/eems/blob/master/Documentation/EEMS-doc.pdf

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

Pinaud D, **Weimerskirch H. 2007.** At-sea distribution and scale-dependent foraging behaviour of petrels and albatrosses: a comparative study. *Journal of Animal Ecology* **76:** 9–19.

Pritchard JK, **Stephens M**, **Donnelly P. 2000.** Inference of population structure using multilocus genotype data. *Genetics* **155:** 945–959.

R Core Team. 2017. *R: a language and environment for statistical computing*. Available at: https://www.R-project.org/

Ravache A, **Bourgeois K**, **Thibault M**, **Dromzée S**, **Weimerskirch H**, **de Grissac S**, **Prudor A**, **Lorrain A**, **Menkes C**, **Allain V**, **Bustamante P**, **Letourneur Y**, **Vidal É. 2020.** Flying to the moon: Lunar cycle influences trip duration and nocturnal foraging behavior of the wedge-tailed shearwater *Ardenna pacifica*. *Journal of Experimental Marine Biology and Ecology* **525:** 151322.

Rayner MJ, **Hauber ME**, **Steeves TE**, **Lawrence HA**, **Thompson DR**, **Sagar PM**, **Bury SJ**, **Landers TJ**, **Phillips RA**, **Ranjard L**, **Shaffer SA. 2011.** Contemporary and historical separation of transequatorial migration between genetically distinct seabird populations. *Nature Communications* **2:** 232.

Reich D, **Thangaraj K**, **Patterson N**, **Price AL**, **Singh L. 2009.** Reconstructing Indian population history. *Nature* **461:** 489–494.

Renaud G, **Hanghøj KM**, **Korneliussen TS**, **Willerslev E**, **Orlando L. 2019.** Joint estimates of heterozygosity and runs of homozygosity for modern and ancient samples. *Genetics* **212:** 587–614.

Sagar PM, **Stahl JC**, **Molloy J. 1998.** Sex determination and natal philopatry of Southern Buller's Mollymawks (*Diomedea bulleri bulleri*). *Notornis* **45:** 271–278.

Singhal S, **Grundler M**, **Colli G**, **Rabosky DL. 2017.** Squamate Conserved Loci (SqCL): A unified set of conserved loci for phylogenomics and population genetics of squamate reptiles. *Molecular Ecology Resources* **17:** e12–e24.

Smith BT, **Harvey MG**, **Faircloth BC**, **Glenn TC**, **Brumfield RT. 2014.** Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. *Systematic Biology* **63:** 83–95.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30:** 1312–1313.

Steeves TE, **Anderson DJ**, **Friesen VL. 2005.** A role for nonphysical barriers to gene flow in the diversification of a highly vagile seabird, the masked booby (*Sula dactylatra*). *Molecular Ecology* **14:** 3877–3887.

Techow NM, **O'Ryan C**, **Robertson CJ**, **Ryan PG. 2016.** The origins of white-chinned petrels killed by long-line fisheries off South Africa and New Zealand. *Polar Research* **35:** 21150.

Thaxter CB, **Lascelles B**, **Sugar K**, **Cook AS**, **Roos S**, **Bolton M**, **Langston RH**, **Burton NH. 2012.** Seabird foraging ranges as a preliminary tool for identifying candidate Marine Protected Areas. *Biological Conservation* **156:** 53–61.

Torres L, **Pante E**, **González-Solís J**, **Viricel A**, **Ribout C**, **Zino F**, **MacKin W**, **Precheur C**, **Tourmetz J**, **Calabrese L**, **Militão T**, **Zango L**, **Shirahai H**, **Bretagnolle V. 2021.** Sea surface temperature, rather than land mass or geographic distance, may drive genetic differentiation in a species complex of highly dispersive seabirds. *Ecology and Evolution* **11:** 14960–14976.

Warren WC, **Clayton DF**, **Ellegren H**, **Arnold AP**, **Hillier LW**, **Künstner A**, **Searle S**, **White S**, **Vilella AJ**, **Fairley S**, **Heger A**, **Kong L**, **Ponting CP**, **Jarvis ED**, **Mello CV**, **Minx P**, **Lovell P**, **Velho TAF**, **Ferris M**, **Balakrishnan CN**, **Sinha S**, **Blatti C**, **London SE**, **Li Y**, **Lin YC**, **George J**, **Sweedler J**, **Southey B**, **Gunaratne P**, **Watson M**, **Nam K**, **Backström N**, **Smeds L**, **Nabholz B**, **Itoh Y**, **Whitney O**, **Pfenning AR**, **Howard J**, **Völker M**, **Skinner BM**, **Griffin DK**, **Ye L**, **McLaren WM**, **Flicek P**, **Quesada V**, **Velasco G**, **Lopez-Otin C**, **Puente XS**, **Olender T**, **Lancet D**, **Smit AFA**, **Hubley R**, **Konkel MK**, **Walker JA**, **Batzer MA**, **Gu W**, **Pollock DD**, **Chen L**, **Cheng Z**, **Eichler EE**, **Stapley J**, **Slate J**, **Ekblom R**, **Birkhead T**, **Burke T**, **Burt D**, **Scharff C**, **Adam I**, **Richard H**, **Sultan M**, **Soldatov A**, **Lehrach H**, **Edwards SV**, **Yang SP**, **Li X**, **Graves T**, **Fulton L**, **Nelson J**, **Chinwalla A**, **Hou S**, **Mardis ER**, **Wilson RK. 2010.** The genome of a songbird. *Nature* **464:** 757–762.

Warren WC, **Hillier LW**, **Tomlinson C**, **Minx P**, **Kremitzki M**, **Graves T**, **Markovic C**, **Bouk N**, **Pruitt KD**, **Thibaud-Nissen F**, **Schneider V. 2017.** A new chicken genome assembly provides insight into avian genome structure. *G3: Genes, Genomes, Genetics* **7:** 109–117.

Weir BS, **Cockerham CC. 1984.** Estimating *F*-statistics for the analysis of population structure. *Evolution* **38:** 1358–1370.

Willing EM, **Dreyer C**, **van Oosterhout C. 2012.** Estimates of genetic differentiation measured by FST do not necessarily require large sample sizes when using many SNP markers. *PLoS One* **7:** e42649.

Zerbino DR, **Birney E. 2008.** Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research* **18:** 821–829.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Figure S1. A principal components analysis of all single nucleotide polymorphisms, including sex-linked regions, showing separation of samples by sex on principal component 1 (PC1).
Figure S2. A principal components analysis containing both the ingroup (wedge-tailed shearwater) and outgroup samples, showing much greater divergence between than within species.
Figure S3. A principal components analysis (PCA) based on an unlinked analysis with only one randomly selected single nucleotide polymorphism per autosomal locus, showing similar patterns to the PCA using the full dataset.
Figure S4. Discriminant analysis of principal components (DAPC) results with $K = 2$ (top) and $K = 3$ (bottom), showing the assignment of individuals to different populations (indicated with different shades) using bar plots (left) and how those populations map onto the principal components analysis plots (right), using circles shaded to match the corresponding populations in the bar plots.
Figure S5. STRUCTURE results for $K = 1$ to $K = 10$, showing assignment probabilities of individuals to each of K groups, indicated with different shades of grey.
Figure S6. A RAxML phylogeny of concatenated genomic single nucleotide polymorphisms, showing clustering of individuals by genetic distance.
Figure S7. A smooth contour map of the posterior mean diversity rate from estimation of effective migration surfaces (EEMS).
Figure S8. Plots of the proportion of called single nucleotide polymorphisms with heterozygous genotypes across individuals (A) and per-population nucleotide diversity (π; B), with the sizes of circles indicating relative diversity. Toe pad samples are coloured blue, blood samples red and tissue samples yellow.
Figure S9. Plots of relative observed heterozygosity across individuals based on ROHAN results (A) and the same values averaged for each locality and plotted on a map (B), with the sizes of circles indicating relative diversity. Toe pad samples are coloured blue, blood samples red and tissue samples yellow.
Figure S10. A plot of *P*-values for the association between wedge-tailed shearwater colour morph and allele frequencies at each sampled single nucleotide polymorphism. Points above the blue line are significant at $\alpha = 0.05$ and points above the red line at $\alpha = 0.01$.
Figure S11. A plot of *P*-values for the association between colour morph and allele frequencies at each sampled single nucleotide polymorphism in a stratified analysis based on the 12 breeding populations depicted in Figure 3. Points above the blue line are significant at $\alpha = 0.05$ and points above the red line at $\alpha = 0.01$.
Table S1. Sample information.
Table S2. Settings used in final estimation of effective migration surfaces (EEMS) analysis.
Table S3. Dataset attributes by sample.
Table S4. Weir & Cockerham's (1984) mean F_{ST} between localities calculated using VCFTOOLS.
Table S5. Weir & Cockerham's (1984) weighted F_{ST} between localities calculated using VCFTOOLS.
Table S6. Reich *et al.*'s (2009) F_{ST} between localities.