PNAS Classification: Biological Sciences, Evolution

Receding ice drove parallel expansions in Southern Ocean penguins.

Theresa L. Cole[1,2,a], Ludovic Dutoit[1,b], Nicolas Dussex[3,4,b], Tom Hart[5,b], Alana Alexander[4,b], Jane L. Younger[6], Gemma V. Clucas[7,8], María José Frugone[9,10], Yves Cherel[11], Richard Cuthbert[12,13], Ursula Ellenberg[14,15], Steven R. Fiddaman[16], Johanna Hiscock[17], David Houston[18], Pierre Jouventin[19], Thomas Mattern[1], Gary Miller[20,21], Colin Miskelly[22], Paul Nolan[23], Michael J. Polito[24], Petra Quillfeldt[25], Peter G. Ryan[26], Adrian Smith[16], Alan J. D. Tennyson[22], David Thompson[27], Barbara Wienecke[28], Juliana A. Vianna[10], Jonathan M. Waters[1]

[1] Department of Zoology, University of Otago, PO Box 56, Dunedin 9054, New Zealand.

[2] Manaaki Whenua Landcare Research, PO Box 69040, Lincoln, Canterbury 7640, New Zealand.

[3] Department of Bioinformatics and Genetics, Swedish Museum of Natural History, Box 50007, Stockholm, Sweden.

[4] Department of Anatomy, University of Otago, PO Box 56, Dunedin 9054, New Zealand.

[5] Department of Zoology, University of Oxford, 11a Mansfield Road, South Parks Road, OX1 3SZ, UK.

[6] Milner Centre for Evolution, University of Bath, Claverton Down, Bath, BA2 7AY, UK.

[7] Atkinson Center for a Sustainable Future, Cornell University, Ithaca, NY 14850, USA.

[8] Cornell Lab of Ornithology, Cornell University, Ithaca, NY 14850, USA.

[9] Laboratorio de Ecología Molecular, Departamento de Ciencias Ecológicas II, Facultad de ciencias, Universidad de Chile. Las Encinas #3770, Ñuñoa, Santiago, Chile.

[10] Pontificia Universidad Católica de Chile, Facultad de Agronomía e Ingeniería Forestal, Departamento de Ecosistemas y Medio Ambiente, Vicuña Mackena 4860, Macul, Santiago, Chile.

[11] Centre d'Etudes Biologiques de Chizé (CEBC), UMR 7372 du CNRS-La Rochelle Université 79360 Villiers-en-Bois, France.

[12] Royal Society for the Protection of Birds, The Lodge, Sandy, Bedfordshire, SG19 2DL, UK.

[13] World Land Trust, Blyth House, Halesworth, Suffolk, IP19 8AB, UK.

[14] Department of Ecology, Environment and Evolution, La Trobe University, Melbourne, Australia.

[15] Global Penguin Society, University of Washington, Seattle, USA.

[16] Department of Zoology, University of Oxford, Peter Medawar Building for Pathogen Research, South Parks Road, OX1 3SY, UK.

[17] Department of Conservation, Murihiku District Office, Invercargill, New Zealand.

[18] Biodiversity, Department of Conservation, Auckland, New Zealand.

[19] Centre d'Ecologie Fonctionnelle et Evolutive, UMR 5175, Campus CNRS, 1919 Route de Mende, 34293 Montpellier Cedex 5, France.

[20] Division of Pathology and Laboratory Medicine, University of Western Australia, Crawley, Western Australia, 6009, Australia.

[21] Institute for Marine and Antarctic Studies, University of Tasmania, Hobart, Tasmania, 7001, Australia.

[22] Museum of New Zealand Te Papa Tongarewa, PO Box 467, Wellington 6140, New Zealand.

[23] Department of Biology, The Citadel, 171 Moultrie St, Charleston, SC, 29409, USA.

[24] Department of Oceanography and Coastal Sciences, Louisiana State University, 1239 Energy, Coast and Environment Building, Baton Rouge, LA 70803, USA.

[25] Justus Liebig Universität Giessen, Heinrich-Buff-Ring 26, 35392, Giessen, Germany.

[26] FitzPatrick Institute of African Ornithology, University of Cape Town, Rondebosch 7701, South Africa.

[27] National Institute of Water and Atmospheric Research Ltd., Private Bag 14901, Kilbirnie, Wellington 6241. New Zealand.

[28] Department of the Environment and Energy, Australian Antarctic Division, 203 Channel Highway, Kingston, TAS 7050, Australia.

ORCID numbers: 0000-0002-0197-286X (T.L. Cole), 0000-0002-8302-0366 (M-J. Frugone), 0000-0001-9469-9489 (Y. Cherel), 0000-0002-6456-7757 (A. Alexander), 0000-0002-1514-7916 (J.M. Waters).

[a] Corresponding author

[b] LD, ND, TH and AA contributed equally to this work

Abstract

Climate shifts are key drivers of ecosystem change. Despite the critical importance of Antarctica and the Southern Ocean for global climate, the extent of climate-driven ecological change in this region remains controversial. In particular, the biological effects of changing sea-ice conditions are poorly understood. We hypothesise that rapid postglacial reductions in sea-ice drove biological shifts across multiple widespread Southern Ocean species. We test for demographic shifts driven by climate events over recent millennia by analysing population genomic datasets spanning three penguin genera (*Eudyptes*, *Pygoscelis* and *Aptenodytes*). Demographic analyses for multiple species (macaroni/royal, eastern rockhopper, Adélie, gentoo, king and emperor) currently inhabiting southern coastlines affected by heavy sea-ice conditions during the Last Glacial Maximum (LGM) yielded genetic signatures of near-simultaneous population expansions associated with post-glacial warming. Populations of the ice-adapted emperor penguin are inferred to have expanded

slightly earlier than those of species requiring ice-free terrain. These concerted high-latitude expansion events contrast with relatively stable/declining demographic histories inferred for four penguin species (northern rockhopper, western rockhopper, Fiordland crested and Snares crested) that apparently persisted throughout the LGM in ice-free habitats. Limited genetic structure detected in all ice-affected species across the vast Southern Ocean may reflect both rapid post-glacial colonisation of sub-Antarctic and Antarctic shores, in addition to recent genetic exchange among populations. Together, these analyses highlight dramatic, ecosystem-wide responses to past Southern Ocean climate change, and suggest potential for further shifts as warming continues.

Keywords

Sphenisciformes, Climate Change, Last Glacial Maximum, Refugia, Genomics.

Significance statement

We analyse population genomic datasets across three penguin genera to test for demographic shifts driven by historical climate events. Numerous species inhabiting coastlines affected by heavy sea-ice during the Last Glacial Maximum show genomic signatures of near-simultaneous population expansions associated with post-glacial warming, contrasting with stable/declining demographic histories inferred for four species occupying consistently ice-free habitats. Shallow population genomic structure detected within species distributed across the vast Southern Ocean likely provides further evidence for recent demographic shifts, and recent genetic exchange among populations. Our results demonstrate dramatic, ecosystem-

wide responses to climate change, and highlight the potential for future biological shifts in the Southern Ocean as global warming continues.

Introduction

Climate change is substantially impacting the abundance and distribution of wildlife, with many species' ranges shifting poleward as a result of climate warming (1). Similar shifts occurred after the Last Glacial Maximum (LGM; 18,000 –25,000 years ago; [2-3]), as temperate refugial populations of many species expanded into high latitudes. While such range shifts may be readily achieved on continents (where terrestrial habitats are essentially continuous [4]), the challenges are more pronounced for isolated or fragmented populations that rely on long-distance dispersal (5-6). For instance, many high-latitude coastal and terrestrial ecosystems of the Southern Hemisphere are isolated by vast ocean gaps (Fig. 1). Southern Ocean circumpolar fronts (including the Subtropical Front and the Antarctic Polar Front) may present additional physical and thermal barriers to southward range expansion of isolated southern coastal populations (10-11).

Understanding past shifts in species distributions is crucial for forecasting responses to contemporary and future climate change. Currently, there is considerable uncertainty surrounding the extent to which high-latitude wildlife populations might have persisted in the Southern Ocean throughout the LGM, versus the extent of post-LGM expansion (6-7, 12). Recent genetic data, however, hint at major ecosystem-wide change following reductions in southern winter sea-ice (7, 13-14). Importantly, past expansions can be reconstructed via genetic analysis of modern populations (2, 15). While several studies of Southern Ocean

species have detected limited population genetic structure, consistent with recent demographic shifts and/or gene flow (9, 13-14, 16-19), a comprehensive genome-wide assessment of Southern Ocean wildlife is lacking. Moreover, as responses to climate change can potentially vary among species (14, 20-21), distinguishing between concerted (multi-species) versus idiosyncratic (single species) shifts may be crucial to forecasting responses to future climate change (22).

Penguins (Sphenisciformes) are iconic marine birds that inhabit all major southern landmasses, with their greatest species diversity in Antarctica and the sub-Antarctic (Fig. 1; SI Appendix, Fig. S1). Although most penguins are natally philopatric (23), some can disperse vast distances traversing major Southern Ocean fronts (24-25), and represent important components of both coastal and marine ecosystems (26). Here we analyse several thousand single nucleotide polymorphisms (SNPs) across 11 Antarctic, sub-Antarctic and temperate penguin species to test for concerted responses to climate change. We detect genomic signatures of population expansion in multiple species currently distributed largely within the LGM sea-ice zone, consistent with concerted re-colonisation of Antarctic and sub-Antarctic coasts during post-LGM warming. In contrast, demographic histories inferred for four temperate penguin species are relatively stable/declining. Our results suggest consistent population dynamics across a species-rich high-latitude assemblage in response to postglacial ice reduction, and demonstrate the potential for rapid change to Southern Ocean ecosystems under future warming.

Results

Demographic reconstructions of effective population sizes (N_e) for 11 penguin species using CubSFS (28), SNAPP (29), Tajima's D (30) and Multi-dice (31) were based on 3,000-13,000 SNPs per species (SI Appendix, Tables S1-S3). Macaroni and royal [*Eudyptes chrysolophus chrysolophus*/*E. c. schlegeli*] penguins were considered a single species based on structure/F_{ST} analyses (see also (19)), whereas Snares-crested [*E. robustus*] and the northern rockhopper [*E. moseleyi*] penguin were excluded from some analyses due to their small sample sizes (Fig. 1, SI Appendix, Tables S4-S5). These analyses revealed comparable postglacial N_e expansions for six southern species (macaroni/royal, eastern rockhopper [*E. filholi*], Adélie [*Pygoscelis adeliae*], gentoo [*P. papua*], king [*Aptenodytes patagonicus*] and emperor [*A. forsteri*] penguins) (Fig. 2, Fig. 3a; SI Appendix, Table S1, Figs. S2-S3), with the emperor penguin expanding slightly earlier. Additionally, two of three demographic analyses supported recent expansion in a seventh species (chinstrap [*Pygoscelis antarctica*]) (Fig. 3a). Notably, these seven species all predominantly occur south of the LGM sea-ice limit (Fig. 1; see [6-8, 23]). By contrast, four species inferred to have relatively stable/declining recent demographic histories (Fig. 3a, Figs. S2-S3) are all predominantly found north of the LGM sea-ice zone (Figs. 1-2): the northern rockhopper (*Eudyptes moseleyi*; Gough and Amsterdam Islands), western rockhopper (*E. chrysocome*; predominantly the Falkland Islands and southern South America), Fiordland-crested (*E. pachyrhynchus*, southern New Zealand) and Snares-crested (*E. robustus*, The Snares and Western Chain) penguins.

The expansion timeframes inferred for most southern lineages (20,000 – 15,000 years ago) correspond to a period of rapid post-LGM warming (27) (Fig. 2a; SI Appendix, Table S1).

These reconstructions suggest populations of the ice-adapted emperor penguin expanded earlier than those of most other southern penguin lineages which require ice-free terrain (see also [16, 32]). The magnitude of inferred postglacial N_e expansions is on average a 2.7-fold increase (ranging from 1.19 – 4.4 fold increase) (Fig. 2a; SI Appendix, Table S1). We detected some variation in the outcomes of different demographic analyses for particular species, perhaps a reflection of varying sensitivity of different model-based approaches and/or biological signal. For example, the CubSFS analysis contrasted with other approaches in suggesting chinstrap penguin populations expanded prior to the LGM, and declined following the LGM. Overall, however, there is broad support for 'stable/declining' demographic trajectories for species inhabiting LGM ice-free regions, versus predominantly 'expanding' trajectories for LGM ice-affected species (Fig. 3a).

We used Multi-dice to test for synchronous versus asynchronous expansions across the seven 'expanded' species identified based on our demographic analyses (Fig. 3a). To this end, we modelled a single expansion event within the last 50,000 years in which up to seven species co-expanded. The synchronous expansion event was inferred to have occurred 20,779 – 24,804 years ago, depending on the summary statistics chosen (Fig. 3b; SI Appendix, Table S3). While only two or three of these southern species were inferred to have expanded simultaneously (SI Appendix, Table S3), minor differences between inferred expansion timings (Fig. 2a) likely hindered the ability for Multi-dice to detect a single expansion event corresponding to all expanding species.

Tests for intraspecific genomic divergence across the ranges of individual species (including previous analyses of *Pygoscelis* and *Aptenodytes* species; see [13, 14, 33-34]) consistently revealed shallow genetic structure within species (Fig 1; SI Appendix, Figs. S4-S7; SI Appendix, Tables S6-S7). In all cases apart from gentoo penguins, we found that panmixia (K=1) was supported, but that using location priors found evidence for additional fine-scale structure, as previously reported [9, 13, 14, 33-34]. Such patterns are consistent with post-LGM demographic and biogeographic expansions (for southern LGM sea-ice species) and recent genetic exchange among populations. Specifically, F_{ST}, PCoA, Structure, DAPC, SNAPP and phylogenetic analyses for *Eudyptes*, *Pygoscelis* and *Aptenodytes* all revealed relatively shallow within-species genomic structure among southern populations (Fig 1; SI Appendix, Figs. S4-S7; SI Appendix, Tables S6-S7; [11, 16, 34-35]). In contrast to the recent genetic exchange inferred within most species, and between macaroni and royal penguins (18-19, 35), these analyses detected little or no admixture among species (Figs. S4-S7; SI Appendix).

Discussion

Our study detected broadly consistent genome-wide signatures of post-LGM expansion across penguin species that currently breed south of the LGM sea-ice zone (Fig. 3a). By contrast, four species currently breeding north of the LGM sea-ice zone exhibited genetic signatures of relatively stable/declining demographies (Fig. 3a). Although estimates of precise LGM breeding ranges for penguins remain elusive (but see [36]), our findings are consistent with the hypothesis of (6) that, during the LGM, many Southern Ocean species retreated to ice-free refugia (e.g. Gough, Amsterdam, Falklands islands, southern South

America, and New Zealand's southern islands (Fig. 2b; see [6-7]). Indeed, several recent studies have suggested that post-LGM reductions in sea-ice were accompanied by rapid re-colonisation of high-latitude shores (7, 11, 14) (Fig. 2c). Recent demographic studies of penguins (Adélie, emperor and king) (16-17, 32) and the southern elephant seal (37), for example, have inferred rapid postglacial recolonisation events. By contrast, recent snow-petrel analyses provide only limited evidence for such postglacial shifts (12). Choice of mutation rate, and possibly time-dependency issues might play some part in these apparently conflicting patterns among taxa. Some contrasting responses among species may also stem from interspecific ecological differences (e.g. variation in feeding ecology, philopatry, habitat preferences). Shifting oceanographic and coastal environmental features associated with postglacial warming may also have impacted local species.

While most LGM coasts are now inundated (see Fig. 2b), some potential LGM refugia may be suggested on the basis of current distributions (e.g. eastern rockhopper penguin likely expanded south from the Auckland, Campbell and Antipodes islands; [Fig. 2c]). Previous studies have concluded that the Southern Ocean's circumpolar fronts can represent important barriers to dispersal for many marine species (10, 11), including penguins (9, 38). However, several penguin species can clearly traverse such boundaries (24-25), and this exceptional dispersal ability may help to explain their apparently rapid biogeographic shifts in response to changing climate (see also [37]).

While CubSFS suggested the chinstrap penguin may have declined following the LGM, Tajima's D and SNAPP supported population expansion for this species, comparable to

results for other southern species (Fig. 3a). This anomaly may perhaps reflect issues with the mutation rate and/or generation time used, or may indicate an idiosyncratic ecological response for this southern species (e.g. variation in feeding ecology, philopatry, habitat preferences, sensitivity to oceanographic fronts). Based on evidence from combined demographic analyses (Fig. 3a), the suggestion that chinstrap penguins have declined since the LGM should be treated with some caution.

A consistent finding of our study is the lack of *major* genome-wide differentiation across the ranges of most penguin species, including several species showing circumpolar near-homogeneity (16-17) (i.e. $K = 1$; $F_{ST} < 0.02$; Fig. 1 and Table S6). These relatively shallow F_{ST} values contrast with more substantial structure, and evidence for multiple Southern Ocean refugia, in white chinned petrels ($K = 3$; $F_{ST} > 0.10$ (39)). While biallelic markers such as the SNPs analysed here are theoretically capable of yielding F_{ST} as high as 1 (i.e. fixed differences at all loci), we note that the upper range of this parameter can be limited by allele frequency distribution (40), and thus these values should be treated with some caution. While use of location priors at higher values of K reveals additional, fine-scale population differentiation (Fig. 1 and Fig. S4), see also (9, 33-34), such structure can potentially evolve rapidly (e.g. 41). Interestingly, the relatively shallow differentiation observed within and among some colonies (e.g. emperor (9, 34)) may also provide additional evidence of recent or ongoing gene flow and admixture, sometimes over vast distances (Fig. 1). Subtle population differentiation detectable with location priors might reflect the influence of contemporary oceanographic fronts and/or changes in local sea-ice conditions, as previously suggested by

(9, 13, 17-18), and may have considerable relevance over ecological timeframes (e.g. conservation management; studies of migration).

Understanding how biota responded to past climate change is essential for predicting species distributions and population sizes under future climate projections, and for developing appropriate conservation management strategies (13, 42). As global temperatures continue to increase, mid-latitude biota will continue to shift towards the poles (11) or alternatively may face extinction (6, 11). Many penguin populations are currently declining, or are predicted to decline as warming continues (43-45). Some of the northernmost colonies of Adélie and emperor penguins have already disappeared (43, 46), and in the case of emperor penguins, these changes have been linked directly to reductions in sea-ice (47). By contrast, populations of gentoo penguin are apparently expanding their ranges southward as the climate warms (48). Our study broadly demonstrates the demographic sensitivity of Southern Ocean wildlife to the effects of past climate change (49), highlighting the potential for future shifts under anthropogenic climate change.

Materials and Methods

DArT-Seq™ library preparation and filtering: DNA was extracted from 428 *Eudyptes* penguin samples spanning six species (Fig. 1; SI Appendix, Fig. S1; SI Appendix, Table S4; macaroni/royal penguins were combined; see [18-19]) using a modified Qiagen DNeasy Blood and Tissue kit. Library preparation and SNP discovery was performed on the 282 highest quality DNA extracts using Diversity Arrays Technology Pty Ltd (DArT-seq™) in Canberra, Australia (50). Each sample was processed following (51), and was sequenced

across three lanes on an Illumina Hiseq 2500. Sequences were processed using in-house proprietary DArT analytical pipelines. We used DartR v1.1.6 (52) in R v.3.5.1 (R Core Team, 2018) to filter the DArT-seq™ data for ten separate *Eudyptes* datasets (based on previous systematic discussions [18, 35], SI Appendix, Table S5). For these *Eudyptes* datasets, we filtered on reproducibility (t=1), and filtered out monomorphic loci, loci with call rates <0.95%, all individuals with call rates <0.90%, all loci with trimmed sequence tags, and all loci that departed from Hardy Weinberg Equilibrium in any colony (P = 0.05 following Bonferroni correction). We also obtained filtered RAD-seq datasets from an additional five penguin species generated and examined by (9, 33-34), comprising Adélie (*Pygoscelis adeliae*; *n*=87), gentoo (*P. papua*; *n*=36), chinstrap (*P. antarctica*; *n*=44), king (*Aptenodytes patagonicus*; *n*=64) and emperor (*A. forsteri*; *n*=110) penguins (SI Appendix, Table S8). See SI Appendix for details.

Phylogenomic analysis and population structure: To clarify the evolutionary relationships among our *Eudyptes* samples newly sequenced in this study, we created a maximum likelihood phylogeny using RAxML-HPC v.8.2.1 (53) (SI Appendix, Fig. S5). We undertook similar population structure analyses for *Eudyptes* as previously implemented for *Pygoscelis* and *Aptenodytes* in (9), as follows: we calculated population summary statistics, including the number of private alleles, observed and expected heterozygosity, the inbreeding coefficient, and global and pairwise F_{ST} (Fig. 1, Tables S5-S7). Genetic clusters were visualised using three methods: principal coordinates analyses (PCoA) using adegenet (54) (SI Appendix, Fig. S4); the Evanno method (55) in Structure v.2.3.4 (56), to estimate the most likely K (Fig. 1; SI Appendix, Fig. S4); and discriminant analysis of principal components (DAPC) using

adegenet (SI Appendix, Fig. S4). We used the SNAPP tree set analyser in BEAST v.2.4.7 (30, 57) to investigate gene flow between closely related *Eudyptes* species (SI Appendix, Fig. S6), based on our results and systematic discussions of (18, 35). While SNAPP analyses have been previously generated for the emperor, king and gentoo datasets (9, 33-34), we also undertook SNAPP analyses for the chinstrap and Adélie penguin datasets obtained from (9) (SI Appendix, Fig. S7). See SI Appendix for details.

Testing for demographic expansions: We reconstructed population histories for 11 *Eudyptes*, *Pygoscelis* and *Aptenodytes* species over the last 1,000,000 years, by estimating the time and magnitude of demographic changes using four different approaches (northern rockhopper and Snares crested penguin were excluded from some analyses due to low sample size). Specifically, we reconstructed the demographic histories using CubSFS (Fig. 2A; SI Appendix, Figs. S2-S3; SI Appendix, Table S1); obtained Tajima's D (SI Appendix, Table S2); identified the change in theta values as inferred by our previous SNAPP analyses (SI Appendix, Figs. S8-S9); and tested for synchronous expansion using Multi-dice (Fig. 3b; SI Appendix, Table S3). As the *Eudyptes* and *Pygoscelis*/*Aptenodytes* datasets were obtained using different pipelines (DArT-seq™ versus RAD-seq), we applied further stringent filtering to ensure consistency between the datasets (SI Appendix). While previous studies have reported shallow population genetic structure within most *Pygoscelis* and *Aptenodytes* species (9, 13, 16-17, 33-34) (Fig. 1), given the relatively shallow F_{ST} values involved, and reported $K = 1$ ([9, 33-34]), we consider this fine-scale structure likely to have evolved recently, and broadly consistent with a scenario of high gene flow, suitable for combined demographic analysis (with the exception of the gentoo penguin (9)). To account for the

deeper genetic structure observed in gentoo populations (e.g. four distinct lineages [9]; Fig. 1), we limited most subsequent analyses of this species to one lineage (Fig. 1; see [9]). For each *Pygoscelis, Eudyptes* and *Aptenodytes* vcf file, we projected the folded allele frequency spectrum down to increase the number of segregating sites using EasySFS. We then adjusted the number of monomorphic sites in our allele frequency spectrum to reflect the total number of monomorphic loci within each species following down projection. For all analyses, we assumed a generation time of 8 – 14 years (from [42]). For all models, we used a mutation rate of 2.6×10^{-7} per locus per generation (14, 16).

Acknowledgements

We thank C. Bazjak, A-S. Coquel, N. Dehnhard, M. Fawcett, H. Irvine, K. Morrison and M. Nicolaus for sample collection, T. King, C. Mitchell and K. Trought for laboratory assistance, V. Chhatre, A. Georges, B. Gruber, F. Pinamartins, B. Roberts, A. Savary, J. Wood and A. Xue for bioinformatics assistance and L. Beheregaray, C. Fraser, A. Kilian, M. Knapp, A. Ruzzante, A. Santure, P. Scofield, P. Sunnucks, J. Wilmshurst and J. Wood for discussions. The research was approved by the Otago University Animal Ethics Committee (AEC) 61/2016, Oxford University Local AEC, University of Western Australia AEC, Woods Hole Oceanographic Institution Animal Care and Use Committee, IPEV ethics committee, Otago University Ngāi Tahu Research Consultation Committee and The Zoological Society of London. Work was carried out under NZ Department of Conservation Permits (OT-25557-DOA, IACUC-18958.00, 32202-FAU, 35682-FAU, 37312-LND, 50437-DOA, 50436-FAU, 50464-DOA, and 38882-RES), NZ Ministry of Primary Industries (2015056535, 2016060908, 2017064905), a Permit to Possess Threatened Fauna for Scientific Purposes No.

TFA 15086, DPIWE Permit to Take Wildlife for Scientific Purposes No. FA05246, Tristan Da Cunha Conservation Department, South African Department of Environmental Affairs, Government of South Georgia and the South Sandwich Islands Restricted Activity Permits (GSGSSI RAP 2018/018, GSGSSI), US NSF Department of Polar Programs ACA permits (ACA 2016-011, ACA 2016-012), Falkland Islands Government (R05/2009, R014/2006) and UK Antarctic Permits. We thank N. Fowke and B. McKinlay for NZ permits. The research was supported by Manaaki Whenua Landcare Research, the University of Otago, Museum of NZ Te Papa Tongarewa, US National Science Foundation (OPP-012-8913), Quark Expeditions, Cheesemans Ecology Safaris, Golden Fleece Expeditions, New Island Conservation Trust, Deutsche Forschungsgemeinschaft, The Citadel Foundation, The Dalio Foundation, donations to Penguin Watch, the Institut Polaire Français Paul Emile Victor (IPEV, Programme N°109, P. Jouventin, H. Weimerskirch, and N°134, C.A. Bost), The Royal Society of NZ Hutton Fund, The Ornithological Society of NZ and an Alumni of Otago in America Award. TLC was supported by an Otago University Postgraduate Award.

Author contributions

TLC, TH, JMW and ND conceived and designed the study. TLC, LD, ND, AA, JLY and GVC analysed the data. TLC and SRF undertook laboratory work. TLC, TH, ND, JLY, GVC, YC, RC, UE, SRF, DH, PJ, TM, GM, CM, PN, MJP, PQ, PGR, AS, AJDT, DT and BW collected samples. All authors contributed to the manuscript.

Data availability

Raw fastq reads are available from the Short Read Archive (DOIXXX). Additional DArT-Seq files, Structure and SNAPP input and output files, and the original and amended SFS are also available on FigShare (DOI:10.6084/m9.figshare.c.4475300).

References

1. Chen I-C, Hill JK, Ohlemuller R, Roy DB, Thomas CD. Rapid range shifts of species associated with high levels of climate warming. *Science.* **333**, 1024–1026 (2011).

2. Hewitt G. The genetic legacy of the Quaternary ice ages. *Nature.* **405**, 907–913 (2000).

3. Davis MB, Shaw RG. Range shifts and adaptive responses to Quaternary climate change. *Science.* **292**, 673–679 (2001).

4. Parmesan C, Yohe GA. globally coherent fingerprint of climate change impacts across natural systems. *Nature.* **421**, 37–42 (2003).

5. Trakhenbrot A, Nathan R, Perry G, Richardson DM. The importance of long-distance dispersal in biodiversity conservation. *Divers. Distrib.* **11**, 173 (2005).

6. Fraser CI, Nikula R, Ruzzante DE, Waters JM. Poleward bound: biological impacts of Southern Hemisphere glaciation. *Trends Ecol. Evol.* **27**, 462–471 (2011).

7. Fraser CI, Nikula R, Spencer HG, Waters JM. Kelp genes reveal effects of subantarctic sea ice during the Last Glacial Maximum. *P. Nat. Acad. Sci. USA.* **106**, 3249–3253 (2009).

8. Gersonde R, Crosta X, Abelmann A, Armand L. Sea-surface temperature and sea ice distribution of the Southern Ocean at the EPILOG Last Glacial Maximum – a circum-Antarctic view based on siliceous microfossil records. *Quat. Sci. Rev.* **24**, 869–896 (2005).

9. Clucas GV, et al. Comparative population genomics reveals key barriers to dispersal in Southern Ocean penguins. *Mol. Ecol*. **27**, 4680–4697 (2018).

10. Munro KJ, Burg TM. A review of historical and contemporary processes affecting population genetic structure of Southern Ocean seabirds. *Emu.* **117**, 4–18 (2017).

11. Fraser CI, et al. Antarctica's ecological isolation will be broken by storm-driven dispersal and warming. *Nat. Clim. Change.* **8**, 604–708 (2018).

12. Carrea C, et al. High vagility facilitates population persistence and expansion prior to the Last Glacial Maximum in an Antarctic top predator: the snow petrel (*Pagodroma nivea*). *J. Biogeog*. **46**, 442–453 (2019).

13. Clucas GV, et al. A reversal of fortunes: climate change 'winners' and 'losers' in Antarctic Peninsula penguins. *Sci. Rep.* **4**, 5024 (2014).

14. Trucchi E, et al. King penguin demography since the last glaciation inferred from genome-wide data. *Proc. R. Soc. Lond. B.* **281**, 20140528 (2014).

15. Waters JM, Fraser CI, Hewitt GM. Founder takes all: density-dependent processes structure biodiversity. *Trends Ecol. Evol.* **28**, 78–85 (2013).

16. Cristofari R, et al. Full circumpolar migration ensures evolutionary unity in the emperor penguin. *Nat. Commun.* **7**, 11842 (2016).

17. Cristofari R, et al. Climate-driven range shifts of the king penguin in a fragmented ecosystem. *Nat. Clim. Change.* **8**, 245–251 (2018).

18. Frugone M-J, et al. Contrasting phylogeographic pattern among *Eudyptes* penguins around the Southern Ocean. *Sci. Rep.* **8**, 17481 (2018).

19. Frugone N-J, et al. More than the eye can see: Genomic insights into the drivers of genetic differentiation in Royal/Macaroni penguins across the Southern Ocean. *Mol. Phylogenet. Evol*. **139**, 106563 (2019).

20. Maggs CA, et al. Evaluating signatures of glacial refugia for North Atlantic benthic marine taxa. *Ecology.* **89**, S108–S122 (2008).

21. Stewart JR, Lister AM, Barnes I, Dalén L. Refugia revisited: individualistic responses of species in space and time. *Proc. R. Soc. Lond. B: Biol. Sci.* **277**, 661–671 (2010).

22. Walther G-R, et al. Ecological responses to recent climate change. *Nature.* **416**, 389 (2002).

23. Borboroglu PG, Boersma PD (eds). Biology and Conservation of the World's penguins, University of Washington Press, Seattle U.S.A. (2013).

24. Thiébot JB, Cherel Y, Trathan PN, Bost CA. Coexistence of oceanic predators on wintering areas explained by population-scale foraging segregation in space or time. *Ecology.* **93**, 122–130 (2012).

25. Thiébot JB, Cherel Y, Trathan PN, Bost CA. Inter-population segregation in the wintering areas of macaroni penguins. *Mar. Ecol. Prog. Ser.* **421**, 279–290 (2011).

26. Woehler EJ, et al. A statistical assessment of the status and trends of Antarctic and subantarctic seabirds. Scientific Committee on Antarctic Research, Cambridge, UK. (2011).

27. Jouzel J, et al. Orbital and millennial Antarctic climate variability over the past 800,000 years. *Science.* **317**, 793–796 (2007).

28. Waltoft BL, Hobolth A. Non-parametric estimation of population size changes from the site frequency spectrum. *Stat. Appl. Genet. Mo. Biol.* **17**, 3 (2018).

29. Bouckaert R, Heled J. DensiTree2: seeing trees through the forest. *BioRxiv*. 012401. (2014).

30. Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data. *PLOS Genet.* **5**, e1000695 (2009).

31. Xue AT, Hickerson MJ. Multi-dice: r package for comparative population genomic inference under hierarchical co-demographic models of independent single-population size changes. *Mol. Ecol. Res.* **17**, e212–e224 (2017).

32. Li C, et al. Two Antarctic penguin genomes reveal insights into their evolutionary history and molecular changes related to the Antarctic environment. *GigaScience.* **3**, 27 (2014).

33. Clucas GV, et al. Dispersal in the sub-Antarctic: king penguins show remarkably little population genetic differentiation across their range. *BMC Evol. Biol.* **16**, 211 (2016).

34. Younger J, et al. The challenges of detecting subtle population structure and its importance for the conservation of Emperor penguins. *Mol. Ecol*. **26**, 3883–3897 (2017).

35. Cole TL, et al. Mitogenomes uncover extinct penguin taxa and reveal island formation as a key driver of speciation.*Mol. Biol. Evol*. **36**, 784–797 (2019).

36. Emslie SD, McKenzie A & Patterson WP. The rise and fall of an ancient Adélie penguin 'supercolony' at Cape Adare, Antarctica. *Roy. Soc. Open Sci*. **54**, 172032 (2018).

37. de Bruyn M, et al. Rapid response of a marine mammal species to Holocene climate and habitat change. *Plos. Gen.* **5**, e1000554 (2009).

38. de Dinechin M, Ottval R, Quillfeldt P, Jouventin P. Speciation chronology of rockhopper penguins inferred from molecular, geological and palaeoceanographic data. *J. Biogeog*. **36**, 693 – 702 (2009).

39. Rexer-Huber, et al. Genomics detects population structure within and between ocean basins in a circumpolar seabird: the white-chinned petrel. *Mol. Ecol.* (In Press).

40. Alcala N & Rosenberg NA. Mathematical constraints on F_{ST}: biallelic markers in arbitrarily many populations. *Genetics.* **206**, 1581-1600 (2017).

41. Boessenkool S, Star B, Waters JM, Seddon PJ. Multilocus assignment analyses reveal multiple units and rare migration events in the recently expanded yellow-eyed penguin (*Megadyptes antipodes*). *Mol. Ecol.* **18**, 2390–2400 (2009).

42. Forcada J, Trathan PN. Penguin responses to climate change in the Southern Ocean. *Glob. Change Biol.* **15**, 1618–1630 (2009).

43. Forcada J, Trathan PN, Reid K, Murphy EJ, Croxall JP. Contrasting population changes in sympatric penguin species in association with climate warming. *Glob. Change Biol.* **12**, 411–423 (2006).

44. Robert-Coudert, et al. Happy feet in a hostile world? The future of penguins depends on proactive management of current and expected threats. *Front. Mar. Sci.* **6**, 248 (2019).

45. Boersma PD, et al. Applying science to pressing conservation needs for penguins. *Cons. Biol.* (In Press).

46. Barbraud C, Weimerskirch H. Emperor penguins and climate change. *Nature.* **411**, 183–186 (2001).

47. Fretwell PT & Trathan PN. Emperors on thin ice: three years of breeding failure at Halley Bay. *Antarct. Sci*. (In Press).

48. Lynch HJ, Naveen R, Trathan PN, Fagan WF. Spatially integrated assessment reveals widespread changes in penguin populations on the Antarctic Peninsula. *Ecology.* **93**, 1367–1377 (2012).

49. Younger JL, van den Hoff J, Wienecke B, Hindell M, Miller KJ. Contrasting responses to climate regime change by sympatric, ice-dependent predators. *BMC Evol. Biol*. 16, 61 (2016).

50. Kilian A, et al. Diversity Arrays Technology: a generic genome profiling technology on open platforms. *Meth. Mol. Biol.* **888**, 67–89 (2012).

51. Sansaloni C, et al. Diversity Arrays Technology (DArT) and next generation sequencing combined: genome-wide, high-throughput, highly informative genotyping for molecular breeding of *Eucalyptus. BMC Proceed.* **5**, 54 (2011).

52. Gruber B, Unmack PJ, Berry OF, Georges A. DARTR: An R package to facilitate analysis of SNP data generated from reduced representation genome sequencing. *Mol. Ecol. Res.* **18**, 691–699 (2018).

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

54. Jombart T, Collins C. Analysing genome-wide SNP data using adegenet 2.0.0. (2015).

55. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* **14**, 2611–2620 (2005).

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

57. Bouckaert R, et al. BEAST 2: A software platform for Bayesian evolutionary analysis. *PLOS Comput. Biol.* **10**, e1003537 (2014).

Figure Legends

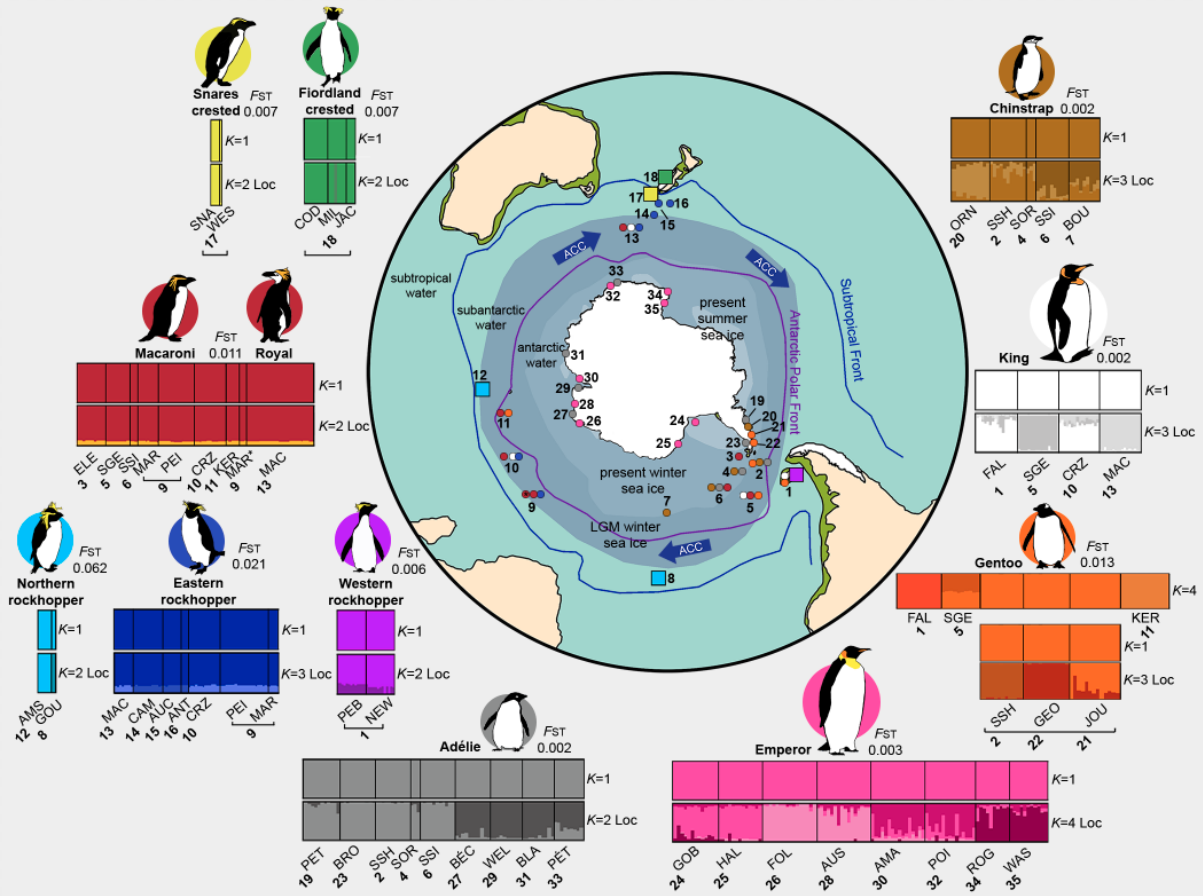

Fig. 1. Sampling locations and genetic Structure plots for 11 penguin species (royal/macaroni are considered one species). The map (adapted from [6]) shows the Antarctic Circumpolar Current (ACC), the Subtropical Front (blue line), the Antarctic Polar Front (purple line), present summer (light blue shading) and winter sea-ice (mid blue shading), Last Glacial Maximum (LGM) winter sea-ice (dark blue shading) (see [6-7, 8]), LGM land extent (green) and glaciation during the LGM (white). Four species (indicated by squares) breed north of the LGM sea-ice limit, whereas seven species (indicated by circles) breed in southern regions affected by LGM sea-ice. The top Structure plot for each species (top two for gentoo) represents the most likely number of genetic clusters as determined via the Evanno method. The bottom Structure plot for each species shows a higher value of *K* to illustrate recently-evolved fine-scale genetic structure that can only be detected using location priors (Loc), as demonstrated by (9). Structure plots for Adélie, emperor, gentoo, king and chinstrap penguins are adapted from (9). With the exception of the gentoo penguin, all analyses demonstrated a

most likely K of 1, with relatively shallow F_{ST} values (global F_{ST} is shown beside each species) (see [9]). Numerical codes for sampling locations (details in SI Appendix, Fig. S1) are indicated on the map and underneath structure plots. Sampling localities: Falkland Islands (FAL, PEB, NEW); South Shetland Islands (SSH); Elephant Island (ELE); South Orkney Islands (SOR); South Georgia (SGE); South Sandwich Islands (SSI); Bouvet (BOU); Gough Island and Tristan da Cunha (GOU); Marion Island (MAR), Prince Edward Islands (PEI); Crozet (CRZ); Kerguelen (KER); Amsterdam Island (AMS); Macquarie Island (MAC); Campbell Island (CAM); Auckland Islands (AUC); Antipodes Islands (ANT); The Snares (SNA), Western Chain (WES); Codfish Island (COD), Milford Sound (MIL), Jackson Head (JAC); Peterman Island (PET); Orne Harbour (ORN); Jougla Point (JOU); George's Point (GEO); Brown Bluff (BRO); Gould Bay (GOB); Halley Bay (HAL); Fold Island (FOL); Béchervaise Island (BÉC); Auster (AUS); Welch Island (WEL); Amanda Bay (AMA); Blakeney Point (BLA); Point Géologie (POI); Pétrels island (PÉT); Cape Roget (ROG); Cape Washington (WAS). The asterix on Marion Island indicates the "white-faced" phenotype of macaroni/royal penguin. Coloured symbols (squares/circles) are consistent with Figs. 2 - 3.

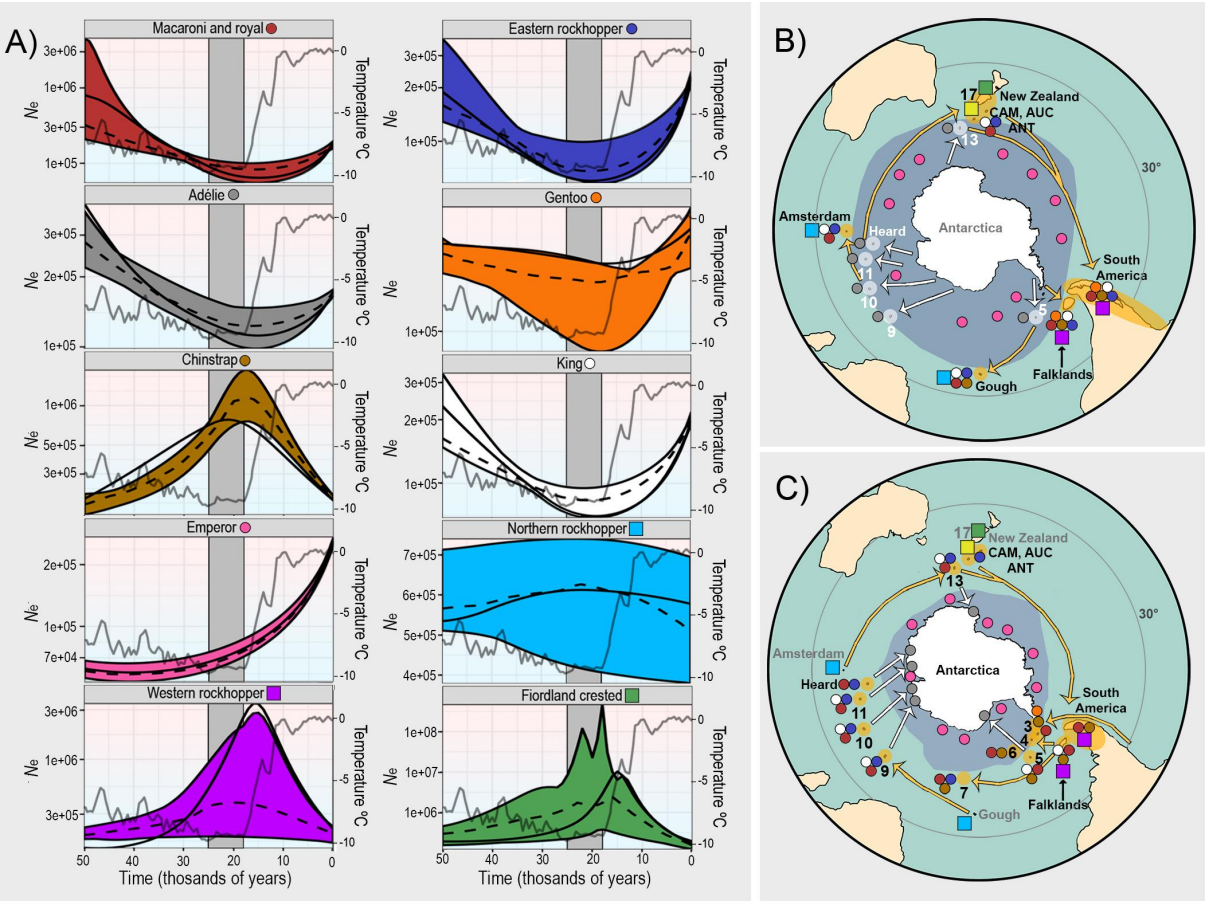

Fig. 2. Population expansions/contractions of penguin species in relation to the LGM. Species breeding south of the LGM sea-ice limit are represented by circles in A, B and C, and species breeding north of the LGM sea-ice limit are represented by squares in A, B and C. A) CubSFS demographic reconstructions for 10 penguin species (Snares crested penguin is excluded due to low sample size). 95% confidence intervals are given by solid colour intervals. Median for bootstrap replicates is given by the dotted line, and the solid line gives the demographic reconstruction for the amended SFS. A 50 thousand year record of Antarctic temperature change (grey line in each plot) as estimated from the EPICA Dome C Ice Core [27] is shown in each plot. The grey bar in each plot shows the LGM. B) shows the winter sea-ice and sea-level during the LGM, with putative refugia shown (orange ellipses for sub-Antarctic penguins; grey points outlined in opaque white for all Antarctic penguins except the emperor penguin). Arrows indicate likely glacial retractions of southern species in response

to LGM sea-ice (white arrows indicate retractions of Antarctic penguins to the fringes of the summer sea-ice during the LGM [except the emperor penguin]; orange arrows indicate retraction of sub-Antarctic penguins to refugial islands north of LGM sea-ice). The emperor penguin presumably bred on the fringes of the summer sea-ice during the LGM (indicated by pink points). Site names in black indicate possible refugia regions for sub-Antarctic penguins, while site names in white indicate possible refugia regions for Antarctic penguins. C) shows the present sea level and winter sea-ice extent, with possible post-LGM routes of recolonization back to Antarctic and southern island habitats (white arrows for penguins breeding in Antarctica [except the emperor penguin]; yellow arrows for penguins breeding on southern islands). Regions where penguins likely persisted are shown with orange ellipses. The emperor penguin breeds on the fringes of the summer sea-ice, which is marked with pink points. Site names in black indicate where each penguin species currently breeds, while sites names marked in grey indicate locations where penguins may have bred during the LGM (as shown in B]). Note, these LGM breeding ranges in both B and C are uncertain. The maps have been adapted from (6). As the Snares crested penguin was included in other demographic analyses (see Fig. 3), the species is shown in both B and C. Coloured symbols (squares/circles) are consistent with Figs. 1 and 3.

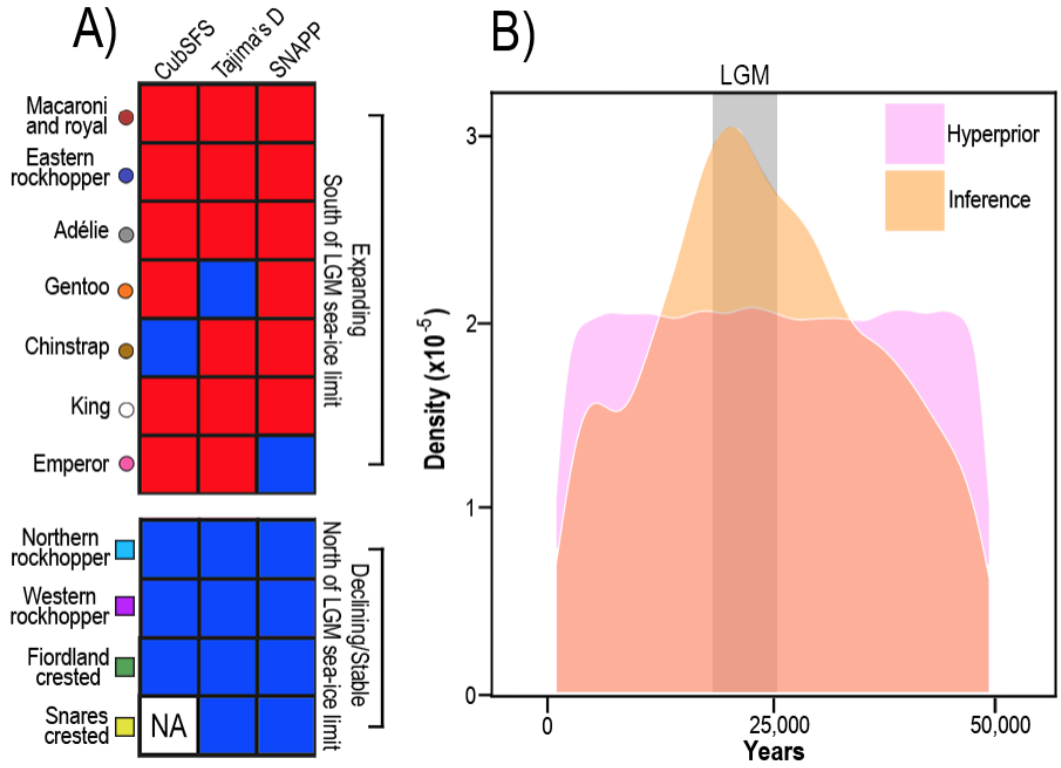

Fig. 3. Summary of demographic results for 11 penguin species. A) shows the combined results of CubSFS, Tajima's D and SNAPP theta values. Species are broadly classified as 'expanding' (red: macaroni/royal, eastern rockhopper, Adélie, gentoo, chinstrap, king and emperor penguin [all south of the LGM sea-ice, represented by circles]) or 'declining/stable' (blue: northern rockhopper, western rockhopper, Fiordland crested, Snares crested penguin [all north of the LGM sea-ice, represented by squares]) on the basis of a majority of these analytical outputs. 'NA' indicates when a species was excluded from an analysis due to limited sample size. All analyses specifically address post-LGM demographic change, with the exception of Tajima's D which may also be influenced by earlier demographic events. B) Multi-dice results, suggesting a LGM expansion, with the mean of the co-expansion time parameter inferred at 24,065 years (mode: 20,778; median: 24,065). Coloured symbols (squares/circles) are consistent with Figs. 1 - 2.

Supplementary Information for

Receding ice drove parallel expansions in Southern Ocean penguins.

Theresa L. Cole, Ludovic Dutoit, Nic Dussex, Tom Hart, Alana Alexander, Jane L. Younger, Gemma V. Clucas, María José Frugone, Yves Cherel, Richard Cuthbert, Ursula Ellenberg, Steven R. Fiddaman, Johanna Hiscock, David Houston, Pierre Jouventin, Thomas Mattern, Gary Miller, Colin Miskelly, Paul Nolan, Michael J. Polito, Petra Quillfeldt,Peter G. Ryan, Adrian Smith, Alan JD Tennyson, David Thompson, Barbara Wienecke, Juliana A. Vianna, Jonathan M. Waters

Corresponding author: Theresa L. Cole
Email: tesscole1990@gmail.com

This PDF file includes:

Supplementary Information Text

Methodology

Sampling and DNA Extraction for six *Eudyptes* penguin species

Blood, tissue and feathers were obtained from 428 *Eudyptes* penguins encompassing entire breeding distributions across six species: macaroni (*E. chrysolophus chrysolophus*)/royal (*E. c. schlegeli*) (considered one species), eastern rockhopper (*E. filholi*), northern rockhopper (*E. moseleyi*), western rockhopper (*E. chrysocome*), Fiordland crested (*E. pachyrhynchus*) and Snares crested (*E. robustus*); [Fig. 1; Fig. S1; Table S4). DNA was extracted from each sample using the Qiagen DNeasy Blood and Tissue kit at Manaaki Whenua Landcare Research

(New Zealand) or at the University of Oxford (United Kingdom). Several modifications were made to the user protocol, depending on the tissue type (blood suspended in Queen's lysis buffer, blood suspended in ethanol, freeze-dried blood cells, tissue or feathers). Specifically, we 1) increased the amount of Proteinase K to 40 µL and added 5 µL RNAse to all samples; 2) we did not include any phosphate buffered saline to blood samples that had been suspended in Queen's lysis buffer; 3) for freeze-dried blood samples, we increased the lysis incubation to 45 minutes, and increased the purification incubation to 15 minutes; 4) for feathers, we increased the digestion to 24 hours at 56°C, and then raised the temperature to 60°C for the final 15 minutes of the digestion; 5) after adding AL buffer, we incubated all samples at 70°C for 45 minutes, and precipitated DNA in cold 100% ethanol, and 6) we eluted the feather samples in 100 µL AE buffer twice, first at 70°C for 15 minutes, and then we recycled the buffer to the spin column membrane, incubating for an additional 5 minutes. We did not modify the user protocol for tissue samples.

The quality and quantity of each DNA sample was assessed using a qubit fluorometer, and on a 1% agarose gel in 0.5X TBE. We diluted samples to a final concentration between 50-100 ng/µL, and for those samples that were too dilute, we combined multiple DNA extractions if they had been undertaken from the same individual, and concentrated the combined DNA using a speedvac to reduce the volume to a minimum of 10 µL. We tested for buffer-activated nucleases in our DNA extractions by combining 1 µL of each DNA sample with 10X restriction enzyme buffer and 8 µL H20, incubating each sample at 37°C for two hours. Following

incubation, the quality of each sample was visualised on a 0.8% agarose gel in 0.5X TBE buffer. 282 of the highest quality samples, which encompassed all *Eudyptes* species (except the erect-crested penguin [*E. sclateri*]) were retained for library preparation.

DArT-Seq™ library preparation, SNP discovery and filtering for *Eudyptes*penguins

Library preparation for SNP discovery was performed using Diversity Arrays Technology Pty Ltd (DArT-seq™) in Canberra, Australia. DArT-seq™ represents a combination of DArT complexity reduction methods and Next Generation Sequencing platforms (1-4). DArT-seq™ is a relatively recent reduced genomic representation method, comparable to Restriction-site Associated DNA Sequencing (RAD-seq [5]) and Genotyping by Sequencing (GBS [6]). DArT-seq™ SNP selection is optimised for each organism by selecting the most appropriate complexity reduction method, including the size of the representation and the fraction of the genome selected for assays. For our *Eudyptes* samples, the *PstI-HpaII* restriction enzyme combination was chosen. Each DNA sample was processed following (1), however, replacing a single *PstI*-compatible adaptor with two additional adaptors that corresponded to two different Restriction Enzyme overhangs (7). The *PstI*-compatible adaptor was designed to include the Illumina flow-cell attachment sequence, the sequencing primer sequence and the barcode region (similar to [6]), while the reverse adaptor contained a flow-cell attachment region and the *HpaII*-compatible overhanging sequence. Initial denaturation during Polymerase Chain

Reaction (PCR) was undertaken at 94°C for 1 minute, followed by 30 cycles of 94°C for 20 seconds, 58°C for 30 seconds, 72°C for 45 seconds, with a final extension at 72°C for 7 minutes. Following PCR, equimolar amounts of amplification products from each sample were bulked and applied to c-Bot (Illumina) bridge PCR, which was run for 77 cycles, and was followed by single-read sequencing across three lanes, on an Illumina Hiseq 2500.

Sequences generated from each lane were initially processed using an in-house DArT analytical pipeline, as follows. Poor quality sequences were removed from the fastq files, while applying stringent criteria to the barcode region, in comparison to the rest of the genome. This ensured that the assignments of the sequences to our specific DNA samples were reliable. Approximately 1,290,000 (+/- 7%) sequences per library, and 1,550,000 sequences per sample were used in the marker calling. The bioinformatic procedure for 53 samples was replicated, to ensure repeatability. Specific filtering included; 1) filtering of the barcode region, with a minimum Phred mass score of 30, and a minimum pass percentage of 75; 2) filtering of the whole read quality, with a minimum Phred score of 10, and a minimum pass percentage of 50, while identical sequences were collapsed into the fastqcall file. The fastqcall file was then used in a secondary in-house pipeline for DArT PL's proprietary SNP and SilicoDArT calling algorithms, which included the presence/absence of restriction fragments in representation. All unique sequences from the set of Fastqcol files were clustered by sequence similarity, at a distance threshold of 3 bases. The sequence clusters were then parsed into SNP and silicoDArT markers, while utilising a range of

metadata parameters derived from the quantity and distribution of each sequence across all samples, combined with previous experience of Mendelian behaviours of DArT-seq™ markers. Given a high level of technical replication was included in the DArT-seq™ genotyping process, reproducibility scores were able to be calculated for each candidate marker. The candidate markers output by ds14 were further filtered on the basis of the reproducibility values (>0.95), average count for each sequence (based on the sequencing depth), and the call rate (the proportion of samples for which the marker was scored). Only 21 samples failed the sequencing pipeline, so our resulting dataset consisted of 134,907 SNPs across 261 *Eudyptes* penguins (Table S4).

We then used DartR v1.1.6(8) in R v.3.5.1 (R Core Team, 2018) to further filter the ten *Eudyptes* datasets separately, based on previous systematic discussions ([9-17], Table S5); 1) all samples; 2) macaroni and royal; 3) northern rockhopper, western rockhopper and eastern rockhopper; 4) eastern rockhopper and western rockhopper; 5) eastern rockhopper; 6) western rockhopper; 7) northern rockhopper; 8) Fiordland crested and Snares crested; 9) Fiordland crested and 10) Snares crested. Specifically, we filtered 1) reproducibility (*gl.filter.repavg*; t=1 [a measure applicable only to DArT-seq™]); 2) removed monomorphic loci that had resulted from the removal of populations or individuals when creating each initial dataset (*gl.filter.monomorphs*); 3) all loci that had call rates of less than ninety-five percent (*gl.filter.callrate*; method "*loc*"); 4) discarded all individuals with a call rate of less than ninety percent (*gl.filter.callrate*; method "*ind*"); 5) all loci with trimmed sequence tags

that could be paralogues (*gl.filter.hamming*), and 6) all loci that exhibited departures from Hardy Weinberg Equilibrium in any colony (*gl.filter.hwe*;*P* = 0.05 following Bonferroni correction). DartR, PGDSpider (18) and easySFS (https://github.com/isaacovercast/easySFS) were used to convert our SNPs to other formats. Computing analyses were undertaken on the New Zealand eScience Infrastructure and the CIPRES scientific gateway (19).

Phylogenomic analysis and summary statistics for *Eudyptes*penguins:

To clarify evolutionary relationships among our samples, we created a maximum likelihood phylogeny using RAxML-HPC v.8.2.1 (20-21). We removed missing sites, and heterozygous states were resolved by randomly assigning one or the other SNP variant to the individual using *gl2fasta*in DartR. We used Phrynomics v.2.0 (https://github.com/bbanbury/phrynomics) to remove all invariable sites. We created 20 independent likelihood tree inferences, drawing bootstrap support from 1000 replicates on the best scoring topology. Searches were conducted under the GTRGAMMA nucleotide substitution model, with the Lewis correction applied for ascertainment bias. The best scoring tree was visualised using FigTree v.1.4.2 (http://tree.bio.ed.ac.uk/) (Fig. S5).

Population summary statistics were calculated from the datasets in Table S5. We calculated the number of private alleles using poppr v.2.8.1 (22-23). We used Genodive v2.0b27 (24-25) to calculate the observed (H_O) and expected heterozygosity (H_S), the inbreeding coefficient (G_{IS}), and AMOVA-based global F_{ST}

(26) using 999 permutations (Table S7). We used StAMPP v.1.5.1 (27) to measure pairwise F_{ST} population differentiation. 95% confidence intervals and *P*-values were generated over 10,000 bootstraps. We corrected the critical *P*-values using the Holm-Bonferroni sequential criterion (28) (Table S6).

Genetic structure analysis for *Eudyptes*, *Pygoscelis* and *Aptenodytes* penguins.

In addition to our novel *Eudyptes* dataset generated here, we also obtained filtered RAD-seq datasets from an additional five penguin species generated and examined by (29-31), comprising chinstrap (*Pygoscelis antarctica*; *n*=44),Adélie (*P. adeliae*; *n*=87), gentoo (*P. papua*; *n*=69), emperor (*Aptenodytes forsteri*; *n*=110) and king (*A. patagonicus*;*n*=64) penguins (Table S8). For consistency, we followed similar methods to those described in (31) to explore genetic structure in our *Eudyptes* dataset. Specifically, genetic clusters within each dataset were assessed and visualised using three methods (Figs. S4-S6). We implemented principal coordinates analyses (PCoA) using adegenet (32). We also used Structure v.2.3.4 (33) via Structure Threader (https://github.com/StuntsPT/Structure_threader[34]). For our Structure analysis, we implemented the admixture model, with allele frequencies correlated across populations (i.e. F model [35]). Each analysis was run 20 times for 150,000 MCMC steps, with the first 50,000 iterations discarded as burnin. With the exception of the dataset that encompassed all samples, each analysis was run twice, with and without location priors. We used the Evanno method (36) in Structure to estimate the most likely number of genetic clusters (*K*), accounting for all putative

populations plus two, using Structure Harvester core version vA.2 (37). Some studies have recommended exploring alternative *K* values for species exhibiting limited or shallow genetic structure (31, 38-39). We therefore present different values of *K* for our results (see below). We undertook a discriminant analysis of principal components (DAPC) using adegenet, with and without priors (Fig. S4). The most appropriate number of principal components was retained based on cross-validation with 1000 replicates. Finally, we used the SNAPP tree set analyser in BEAST v.2.4.7 (40-41), to investigate gene flow between closely related *Eudyptes* species (Fig. S6). In addition to the SNAPP analysis for *Eudyptes*, we also used SNAPP to test for gene flow within the Adélie and chinstrap penguin datasets obtained from (31) (SNAPP analyses have previously been undertaken by [29-31] for gentoo, king and emperor penguin datasets). Mutation rates were based on the data in the alignment. We ran the analysis up to 100,000,000 MCMC states, logging parameters every 10,000 trees with a pre-burnin of 5000 states. We assessed stationarity of the given chain using Tracer v.1.6.0 (http://beast.bio.ed.ac.uk/) until ESS values were >200. Each analysis was run three times for *Eudyptes*, and two times for Adélie and chinstrap penguins, assigning different individuals per colony to each of the replicates. DensiTree v.2.4.7(41) was used to visualize the trees (Figs. S6-S7).

Demographic analyses

We used four demographic approaches to test for population expansions/contractions following the LGM in all 11 penguin species (six *Eudyptes* species generated here [macaroni/royal considered one species], two *Aptenodytes*

species generated in [29-30] and three *Pygoscelis* species generated in [31]). While previous studies have reported shallow genetic structure within *Pygoscelis* and *Aptenodytes* species (29-31, 42-47) (Fig. 1), we interpret this subtle differentiation, based on relatively shallow F_{ST} and estimated K=1 (as suggested by [29-31]), to represent a scenario of high gene flow, and with the exception of gentoo penguin (31), we consider each species (chinstrap, Adélie, emperor and king) as panmictic. With the exception of our SNAPP analysis, we only analysed gentoo penguin individuals derived from the South Shetland Islands, George's Point and Jougla Point (Fig. 1; see [31]), as they show apparent panmixia between locations. We did not analyse the other lineages identified by (31) as they encompassed <12 individuals. We re-filtered the gentoo penguin dataset using vcftools (48) with a minor allele frequency of 0.01, so that it was consistent with (31).

As our *Eudyptes* DArT-seq™ dataset was generated using slightly different filtering to the *Pygoscelis* and *Aptenodytes* RAD-seq dataset, we applied further filtering to ensure our datasets (DArT-Seq versus RAD-Seq) were consistent with one another. Therefore, we used vcftools to apply further filtering to the five DArT-Seq *Eudyptes* datasets (macaroni/royal, western rockhopper, eastern rockhopper, northern rockhopper and Fiordland crested), the RAD-Seq *Aptenodytes* datasets (king and emperor) and the RAD-Seq *Pygoscelis* datasets (Adélie, chinstrap and gentoo). Specifically, we filtered the *Eudyptes* datasets with a minor allele frequency of 0.01 (following the criteria described in [31]), reflecting the filtering of the *Aptenodytes* and *Pygoscelis* datasets. For the *Aptenodytes* and *Pygoscelis* datasets, we filtered loci

with call rates <95%, and individuals with call rates <90%, reflecting the filtering criteria for the *Eudyptes* samples. The SFS were generated using easySFS, maximising the amount of variation in the downprojected samples.

Demographic analyses using CubSFS

We used CubSFS v1.0 (49) to test for demographic expansions among all *Eudyptes*, *Pygoscelis* and *Aptenodytes* species (except for Snares crested penguin due to limited sample sizes). We read the folded, MAF filtered (0.01) SFS into R v3.5.1, using tidyverse v1.2.1 functions (https://github.com/tidyverse/tidyverse.org). After obtaining the sample size and total number of sites from the SFS, the monomorphic site category was removed, and CubSFS was used to estimate the demographic history for each species, using 29 knots, and a t_m of 0.25 coalescent units. Because of variance in how the monomorphic site category was recorded per dataset due to the differing RAD-seq methodologies employed in this study, we used the expected SFS generated and a negative exponential model to estimate the 'true' number of monomorphic sites. We also amended the observed SFS site categories likely affected by the MAF by extracting the expected SNP counts for these sites. Following this, we used CubSFS to re-estimate demographic history from the amended SFS, again using 29 knots and a t_m of 0.25 coalescent units. We then generated 10 bootstrap replicates per species, and used CubSFS to estimate the demography for each of these replicates with the same parameters as used for the amended SFS. Finally, we obtained demographic reconstructions in units of effective population size (N_e) against year by using the estimated number of monomorphic

sites, a per-generation mutation rate of 2.6×10^{-7} [44, 50], and species-specific generation times (8 years: macaroni/royal, eastern rockhopper, northern rockhopper, western rockhopper, Fiordland crested, Adélie, gentoo, and chinstrap; 9 years: king; 14 years: emperor [51]). These demographic histories were plotted using R, and the tidyverse, gridextra (https://cran.r-project.org/web/packages/gridExtra/index.html), and scales (https://cran.r-project.org/web/packages/scales/index.html) libraries (Fig. 2a; Figs. S2-S3). Refer to https://github.com/laninsky/penguinsfor the code that was used for running and plotting the CubSFS analyses.

Demographic analyses using Tajima's D

Tajima's D was computed for all species using δaδi (53) from the downprojected allelic frequency spectrum, ignoring allele frequencies below a MAF of 0.01. Confidence intervals were obtained by re-sampling 1000 times each site frequency spectrum dataset using in-house scripts.

Demographic analyses using SNAPP

We investigated the change in N_e in all of the species as represented by the change in theta between the terminal lineages and nearest internal branches derived from the previous SNAPP analyses. For *Eudyptes* spp., Adélie and chinstrap penguins, we used results from the SNAPP analyses described above (three replicates each for the *Eudyptes* spp., two replicates for the Adélie and chinstrap). For the remaining species, we used previously published SNAPP analyses (two replicates) obtained from (31) (gentoo), (29) (king) and (30) (emperor). While our previous demographic

analyses (CubSFS, Tajima's D [above] and Multi-dice [below]) explored population expansions across the combined sample locations for each species, we explored the change in theta inferred from SNAPP separately for each sampling locality e.g. for gentoo penguin, the four lineages represented in Fig. 1 (see also [31]). To do this, we constructed maximum clade credibility (MCC) trees for each species using common ancestor heights through TreeAnnotator v2.5.0. This MCC tree was used as input into the SNAPP tree set analyser with 10% burnin to calculate the theta values for each terminal tip (sampling location) and the nearest internal ancestral branches. Using the MCC trees, we also estimated the length of time (in years) that the change of theta had occurred across for each sampling location by extracting the age of the nearest ancestral node (in number of substitutions), dividing this by the per-generation mutation rate (see above), and then multiplying by the estimated generation time (in years) for each species (see above). Refer to https://github.com/laninsky/penguinsfor the code that was used for running the SNAPP delta theta analyses.

Demographic analyses using Multi-dice

Finally we used the R package Multi-dice (54) to test for a population expansion among the expanding penguin species following the LGM. We tested scenarios of one co-expansion within the last 50,000 years that could be synchronous or not across seven 'expanding' species (macaroni/royal, eastern rockhopper, Adélie, gentoo, chinstrap, king and emperor, based on the combined results of CubSFS, Tajima's D and SNAPP). We did not analyse northern rockhopper, Fiordland crested,

western rockhopper or Snares crested penguin as we consider those species to have had a relatively stable/declining population size since the LGM (based on the combined results of CubSFS, Tajima's D and SNAPP). We projected all SNPs datasets to 70 haploid samples using easySFS, maximising the variation in the chinstrap penguin dataset, which had the smallest number of sampled individuals and equalising the SFS across species as required by Multi-dice. We performed 100,000 simulations under a simple scenario comparable to (54). We allowed for one synchronous expansion event where one to seven co-species expand while the others expand idiosyncratically. Expansions were constrained to have a post-expansion N_e within 10 – 100 times ancestral N_e, and for species-specific expansions to occur within 3000 years of each other to be considered as co-expanding. Both the proportions of co-expanding species and the timing of the synchronous event were recorded along with the aggregate frequency spectrum (54) for inference. We then sampled the best 5% (5000 simulations) using hierarchical approximate bayesian computation as implemented in ABC (55) and the aggregate site-frequency spectrum of each simulation as a single pseudo-observed dataset.

Results

Genetic diversity, structure and admixture across all *Eudyptes* penguins

We tested for population genetic structure and possible admixture (as inferred by [56-59]) between all our sampled *Eudyptes* colonies using F_{ST} (Table S6), RAxML (Fig. S5), Structure (with and without location priors) (Fig. 1; Fig. S4), DAPC (with and without location priors) (Fig. S4) and PCoA (Fig. S4).

As demonstrated by several recent studies (e.g. [9, 14-16]), macaroni and royal penguins consistently clustered together, with little evidence to indicate population genetic structure can be used to distinguish between the two subspecies (Fig. 1; Figs. S4-S6; Table S6), despite clear phenotypic differences (60-61). The closely related Fiordland crested and Snares crested penguin (12) remained difficult to distinguish under our global Structure analyses (including K = 5-7), initial global DAPC analyses, and to a lesser extent with our RAxML phylogeny (bootstrap support: 92). However, hierarchical downstream analyses of just the Fiordland crested and Snares crested penguin dataset supports recognition of two distinct species (Fig. S4). In addition, our results consistently support findings of several previous studies (9, 17), which recognise three distinct rockhopper penguin species (the northern, western and eastern rockhopper) (Figs. S4-S6).

With the exception of the macaroni/royal penguin (which are probably in the earliest stages of speciation (17) and which we consider a single species; see [9, 16-17]), our Structure analysis (Fig. S4) revealed limited evidence for admixture between the six *Eudyptes* species. Very shallow levels of admixture (<0.5%) (based on Structure) were detected from the Snares crested penguin to the Fiordland crested, the northern rockhopper, western rockhopper and eastern rockhopper penguin; from the Fiordland crested penguin to the northern rockhopper and eastern rockhopper penguin; from the northern rockhopper penguin to the western rockhopper and macaroni/royal penguin; from the eastern rockhopper penguin to the Snares crested

penguin; from the western rockhopper penguin to the northern rockhopper penguin, and from the macaroni/royal penguin to the eastern rockhopper, western rockhopper, Fiordland crested and the Snares crested penguin. Slightly higher levels of admixture (0.6-0.9%) were observed from the eastern rockhopper penguin to the northern rockhopper penguin (mean 0.6%), from the western rockhopper penguin to the eastern rockhopper penguin (mean 0.7%), and from the northern rockhopper penguin to the Snares crested penguin (mean 0.9%), while the highest levels of admixture were detected between the most closely related species, such as the eastern rockhopper penguin to the western rockhopper penguin (mean 5.1%), from the northern rockhopper penguin to the eastern rockhopper penguin (mean 1.9%) and from the Fiordland crested penguin to the Snares crested penguin (mean 68%, however see Fig. S4). To explore the patterns among closely related *Eudyptes* species and within each species, we undertook further analyses on smaller datasets (Table S5).

Macaroni and royal penguins

The phylogenetic placement of macaroni and royal penguins together clustering into a single panmictic population (9, 16-17) was consistently supported by pairwise F_{ST}, PCoA, Structure, DAPC, and SNAPP analyses, supporting recent conclusions of (9, 16-17) (Fig. 1; Figs. S4-S6). Our Structure analyses using the Evanno method suggested the most likely number of populations (*K*) is 2. However, the corresponding *K*=2 Structure plot revealed no molecular distinction between macaroni and royal penguins (Fig. 1, Fig. S4), even though the subspecies exhibit

clear phenotypic differences (macaroni penguins are smaller than royal penguins and have a black face, while royal penguins have a white face). Very subtle population genetic structure between the two species was only evident when we examined alternative Structure plots of $K>3$ (with location priors), when we undertook DAPC analyses also using location priors, or when we examined PCoA plots (Fig. S4). Our PCoA and DAPC analysis (when location priors were used) (Fig. S4) revealed the possible presence of three very shallow genetic clusters of macaroni and royal penguins, which fall into three distinct geographical regions. These correspond to 1) Elephant, South Georgia and South Sandwich Islands (macaroni penguins); 2) Marion, Prince Edward, Crozet and Kerguelen (macaroni penguins), and 3) Macquarie Island and Marion Island "white-faced penguins" (royal penguins). These shallow clusters were also generally supported by our SNAPP analyses (Fig. S6). While our Structure analyses suggest a scenario of high gene flow among these regions, our DAPC analysis (with location priors) indicate that there may be limited movement of macaroni penguins between Elephant, South Georgia and the South Sandwich Islands, with only few individuals potentially representing migrants (Fig. S4). Moreover, our results support a recent study (16) that also demonstrated that the white-faced penguins breeding on Marion Island consistently cluster with the Macquarie Island royal penguin individuals (see Fig. S4; Table S6), suggesting recent dispersal of royal penguins across the Southern Ocean (see Fig. 2c). Nevertheless, as our analyses consistently support a lack of strong genetic structuring between macaroni and royal penguins, we consider these subspecies as a single species for our demographic modelling approaches.

Northern, western and eastern rockhopper penguins

Species-level phylogenomic separation of the three inferred rockhopper penguin species (the northern, western and eastern rockhopper) was consistent among F_{ST}, PCoA, Structure, DAPC, and SNAPP analyses, when we examined all rockhopper penguin species in a single dataset. Our Structure analysis that encompassed all rockhopper penguin species suggested a most likely *K* of 3, supporting the species-level recognition of the northern, western and eastern rockhopper penguins (as suggested by [9, 11, 17, 62]). However, when we examined the Structure plot for a lower value of *K* for this dataset (*K*=2; Fig. S4), the northern rockhopper penguin clustered with the eastern rockhopper penguin. This is surprising, as the northern rockhopper penguin is sister to the combined lineage of the eastern and western rockhopper (17), and also because the eastern rockhopper penguin is occasionally considered a subspecies, or an incipient species of the western rockhopper penguin. This result, of the northern rockhopper penguin being clustered with the eastern rockhopper penguin may reflect the difficulty for Structure to choose two distinct clusters, when three are present (as suggested by all analyses). Further Structure results of *K*=3 (Fig. S4) revealed possible shallow levels of admixture within the three rockhopper penguins. Very shallow levels of admixture (<0.1%) were observed from the western rockhopper penguin to the northern rockhopper penguin and from the eastern rockhopper penguin to the northern rockhopper penguin. Further shallow levels of admixture were observed from the eastern rockhopper penguin to the western rockhopper penguin (mean 1.9%), from the western rockhopper penguin to

the eastern rockhopper penguin (mean 0.6%) and from the northern rockhopper penguin to the eastern rockhopper penguin (mean 1.4%). When we examined the Structure plots of*K*=3 (and *K*=4) (Fig. S4), it appears that only eastern rockhopper penguin individuals from the New Zealand Campbell, Auckland and Antipodes Islands revealed admixture with the western rockhopper penguin, while all other eastern rockhopper penguin samples appear to have experienced admixture with the northern rockhopper penguin. This pattern may reflect possible historical admixture (and possibly historical migration) of the eastern rockhopper (Campbell, Auckland and Antipodes Islands) populations with the western rockhopper penguin (Falkland Islands) populations, given those four populations were not impacted by LGM-winter sea-ice, and that the Falkland Islands is the closest island group to Campbell, Auckland and Antipodes Islands (in the direction of the Antarctic Circumpolar Current). This geographical pattern may also explain why the remaining eastern rockhopper penguin populations (Macquarie, Crozet, Prince Edward and Marion Islands) show admixture with the northern rockhopper penguin (Amsterdam and Gough Islands), as Amsterdam and Gough Islands are geographically closer to Amsterdam and Gough Islands, than they are to the Falkland Islands.

Eastern rockhopper penguin

Our Structure analysis suggested that *K*=1 for the eastern rockhopper penguin. However, when we explored the Structure plots for *K*=3 (with and without location priors), we revealed subtle population structure corresponding to 1) Campbell, Auckland and Antipodes Islands and 2) Macquarie, Crozet, Prince Edward and

Marion Islands (Fig. S4). This pattern was also observed in our PCoA and SNAPP analyses, and to a lesser extent within our DAPC analysis (Figs. S4, S6). Our DAPC analysis (with location priors) revealed only limited evidence of movement between populations of the eastern rockhopper penguin, although it appears that a single individual from Antipodes Island could have migrated to Campbell Island, while there also appears to be some migration between Crozet, Prince Edward and Marion Islands.

Western rockhopper penguin

Samples for the western rockhopper penguin were obtained only from the Falkland Islands (New Island and Pebble Beach). While our Structure analyses suggest the most likely K=2 for the western rockhopper penguin, when we examined K=4 (with location priors), subtle population structure between the two sites was observed. This result is also reflected in our DAPC analysis (with location priors). In addition, our PCoA and Structure analyses revealed two Pebble Beach individuals as consistent outliers to all other western rockhopper penguin samples (Fig. S4), while our DAPC analysis revealed a different single individual from Pebble Beach as a possible migrant. This observation may suggest the presence of further genetic structure within other, unsampled western rockhopper penguin colonies, within the Falkland Islands, or on different island groups.

Northern rockhopper penguin

PCoA, Structure and DAPC analyses also revealed the possibility of population clustering between populations of the northern rockhopper penguin belonging to Amsterdam and Gough Islands (Fig. S4). This pattern may reflect the patterns of other penguin species that inhabited islands north of the LGM sea-ice (such as the western rockhopper penguin). It is likewise conceivable that the northern rockhopper penguin also did not experience rapid population expansions and post-glacial recolonisation following the LGM. These potentially genetically structured populations may also reflect the historical and contemporary population isolation of the two island groups. This possible pattern could not be analysed further due to the low sample size (*n*=6 for Amsterdam Island and *n*=2 for Gough Island).

Fiordland crested and Snares crested penguins

Our initial global Structure and RAxML analyses detected possible admixture between the Fiordland crested penguin and the Snares crested penguin. However, all additional analyses conducted on just the two species (F_{ST}, PCoA, Structure, DAPC and SNAPP) (Fig. S4; Table S6) supported species-level distinction for both the Fiordland crested and the Snares crested penguin. This result was also suggested by (15), who used mitochondrial cytochrome oxidase subunit 1 and control region to analyse the same Fiordland crested and Snares crested penguin individuals under a phylogenetic and haplotype framework.

Fiordland crested penguin

Within the Fiordland crested penguin, we found very little evidence to suggest there may be contemporary population genetic structure (Fig. 1; Fig. S4) among colonies from Codfish Island, Milford Sound and Jackson Head, supporting previous findings (15) of a panmictic population, which also analysed the same individuals. When we examined the Structure plots for *K*=2 and *K*=3 (with and without location priors), we revealed a single individual from Milford Sound as a consistent outlier to all other samples, and when we examined the Structure plot for *K*=3 (without location priors), a second individual from Codfish Island was revealed as an outlier (Fig. S4). These individuals are possibly migrants from unsampled populations, such as Stewart Island, or may be relicts from recently extinct historical populations that may have once existed in the northern New Zealand South Island (15).

Snares crested penguin

In addition, analyses of Structure, PCoA, DAPC and RAxML found little support to suggest that there is population genetic structure within the Snares crested penguin populations breeding on The Snares and the Western Chain (Figs. S4-S6). However, our sample size is small (*n*=4), and only a single sample was genotyped from the Western Chain (Fig. S1). Our results, therefore, support recent findings of a single species, and panmictic population of the Snares crested penguin (15).

Demographic results

We tested for demographic expansions following the LGM for all *Eudyptes*,*Pygoscelis* and *Aptenodytes* penguins using CubSFS, Tajima's D, SNAPP

and Multi-dice. Although there are a few inconsistencies between the different demographic approaches, the general pattern of these analyses suggests that the ice-adapted emperor penguin began expanding during or just prior to the LGM, that seven penguin species experienced population expansions in the immediate aftermath of the LGM (macaroni/royal, eastern rockhopper, Adélie, gentoo, chinstrap and king) and four species did not experience post-LGM population expansions (northern rockhopper, western rockhopper, Fiordland crested and Snares crested). In general, this pattern suggests that penguin species inhabiting regions south of the LGM sea-ice extent rapidly expanded following reductions in LGM sea-ice, while penguin species inhabiting regions north of the LGM sea-ice extent did not experience population expansions (Fig. 3a).

CubSFS

The demographic histories inferred by CubSFS for 10 penguin species (the Snares crested penguin was excluded due to small sample sizes) revealed concerted demographic expansions in macaroni/royal, eastern rockhopper, Adélie, gentoo, king and emperor (Fig. 2a; Fig. S2-S3; Table 1). All of these recently 'expanded' species are predominantly found south of the LGM sea-ice limit (Fig. 1). Crucially, five of these near-simultaneous demographic expansions began approximately 17.2 thousand years ago (ranging from 20 – 15 thousand years ago) (Fig. 2a; Fig. S2-S3; Table 1), at a time of rapid warming in the immediate aftermath of the LGM (63). By contrast, analysis of emperor penguin shows population expansion earlier than other penguin lineages, suggesting that this ice-adapted species was able to exploit

southern regions earlier; see also (44, 64). The magnitude of these population expansions following the LGM is on average a 2.77-fold increase (ranging from 1.18 – 4.40-fold increase; [Table 1]). In addition, analyses of four species (northern rockhopper, western rockhopper, Fiordland crested, Snares crested) found north of the LGM sea-ice zone (Fig. 1), and analysis of chinstrap penguin found south of the LGM sea-ice zone revealed decreasing or relatively stable population sizes (Fig. 2; Fig. S2-S3; Table 1).

Tajima's D

While Tajima's D can infer population expansion events, it cannot test for the timing of those expansions. Tajima's D for all penguin species were generally consistent with previous demographic results. Past population expansions, inferred by a negative Tajima's D were detected in macaroni/royal, eastern rockhopper, Adélie, chinstrap, king and emperor penguins, and possibly Fiordland crested penguin, while stable/declining population sizes were detected in the gentoo, northern rockhopper, western rockhopper, Snares crested (Table S2). Although Tajima's D was negative for the Fiordland crested penguin, which could suggest a population expansion, this result was not statistically significant (P = 0.135) (Table S2).

SNAPP

While the theta value analysis using our SNAPP results did infer some demographic changes (Fig. S8), the internal node ages both within and between penguin species vary widely (Fig. S9). For those species with older inferred internal node ages (e.g. >

600 thousand years between some macaroni/royal populations) the change in theta represents all the fluctuations in N_e that may have occurred over the last 600 thousand years. In addition, several species that experienced a post-LGM demographic expansion as inferred by CubSFS, Tajima's D (above) and Multi-dice (see below); e.g. gentoo penguin, showed reductions in theta for some sampling localities. However, the internal node ages that at least some of these inferred reductions in theta are based on are far younger than the LGM (see also [31]), increasing the likelihood that contemporary factors impacting N_e are influencing these estimates downward (Fig. S9). For these reasons, it is difficult to use these SNAPP analyses to directly assess the response of $N_e$to the LGM, and therefore we focus our interpretation on the CubSFS, Tajima's D (see above) and Multi-dice analysis (see below).

Multi-Dice

Based on the combined results of CubSFS, Tajima's D and SNAPP (see above; Fig 3a), we investigated how synchronous the expansion event was among seven species, as well as the timing of the inferred expansion. For our seven "expanding" species (macaroni/royal, eastern rockhopper, Adélie, gentoo, chinstrap, king and emperor), Multi-dice inferred a co-expansion event during the last 21 - 25 Kya while other species expanded idiosyncratically (Table S3). Interestingly, our multi-dice results did not reveal a largely synchronous expansion event, as only 29% - 39% of species were inferred to co-expand. There are several reasons that may limit our ability to detect a potentially synchronous expansion event. Mainly, small

inaccuracies in the generation times used for each species might lead to large time differences over several thousand of years. Although the 'expanding' species might show a general pattern of co-expansion, if the time scale of this co-expansion varied by more than 3,000 years (either due to generation time inaccuracy, or due to real biological signal e.g. the emperor penguin and chinstrap penguin expanding earlier), then multi-dice would fail to detect a single co-expansion event.

Supporting Figures

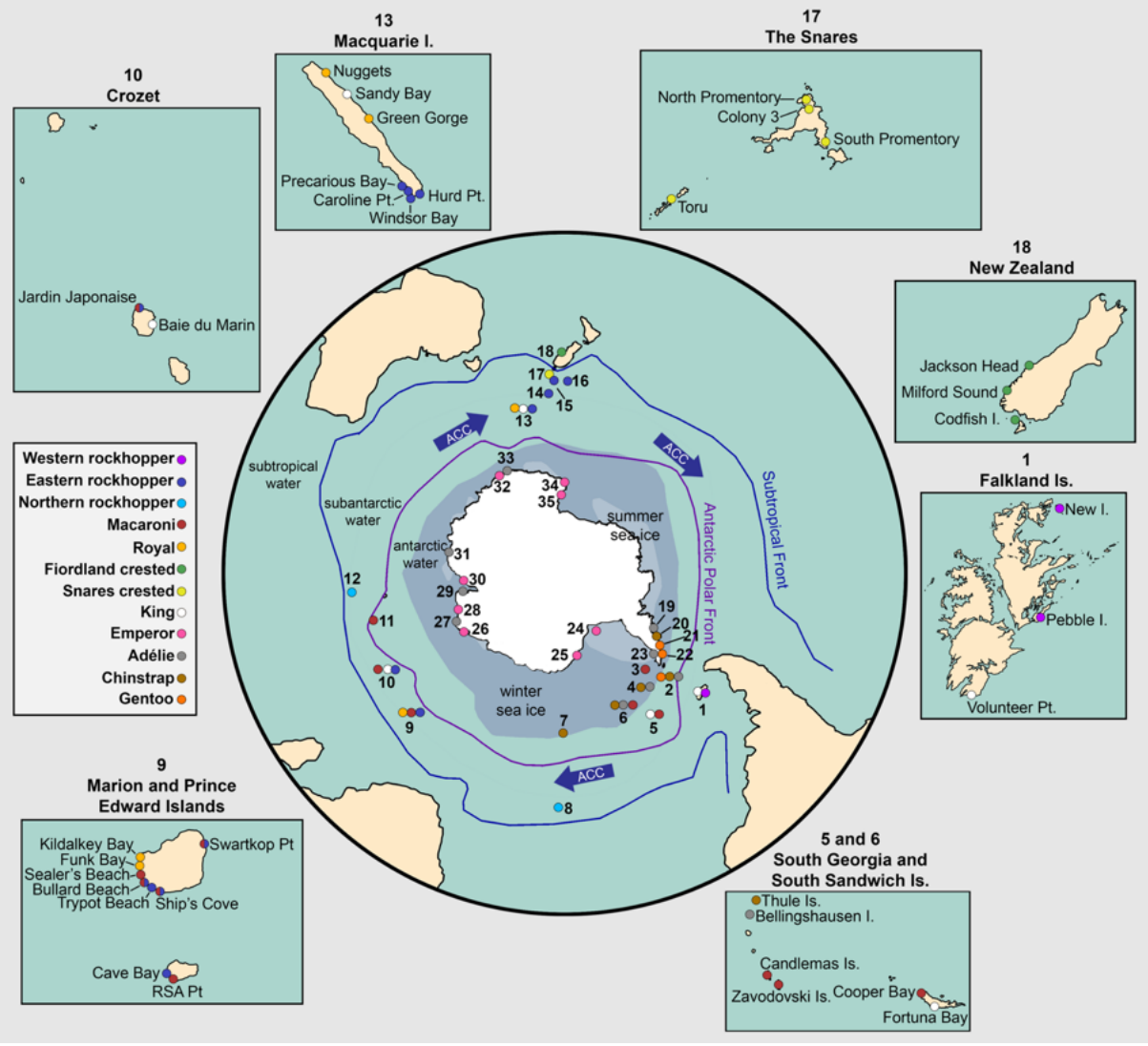

Fig. S1. Map of sampling locations of twelve penguin species representing *Eudyptes*, *Aptenodytes*and *Pygoscelis*penguin genera. Island groups that contained

multiple sample collections from different sites are shown. Note Adélie and chinstrapSouth Orkney Islands samples all came from Signy Island, and Adélie, gentoo and chinstrap South Shetland Islands samples all came from Admiralty Bay (see Table S12). The main map has been modified from (65) and shows current winter and summer sea-ice, the Subtropical Front, the Antarctic Polar Front and the Antarctic Circumpolar Current (ACC).

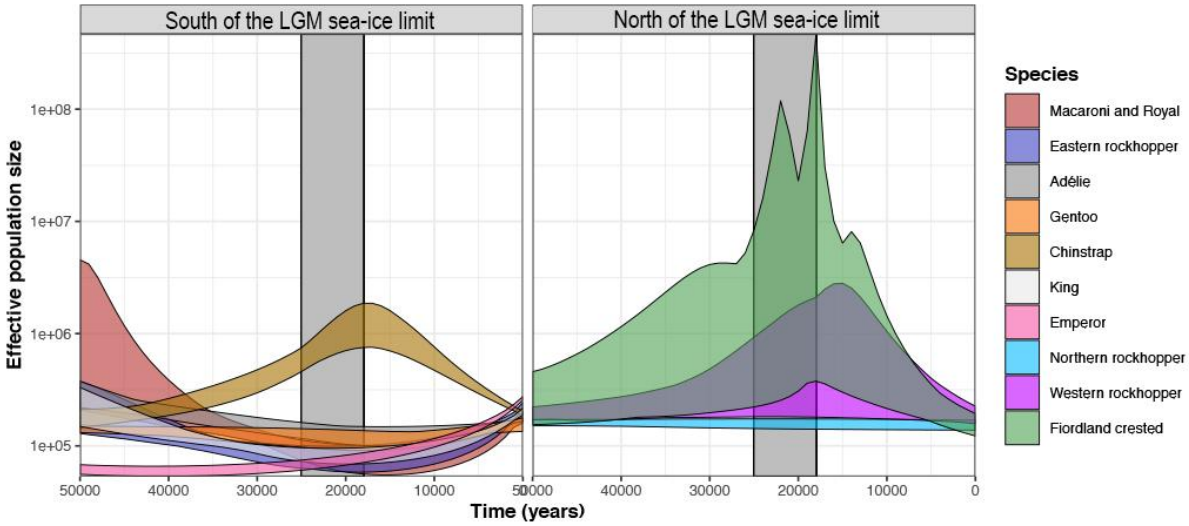

Fig. S2. Overlaid 95% CIs for all penguin species for CubSFS demographic reconstructions, grouped by breeding locations (south or north of the LGM sea-ice limit). Note the extremely wide CI for the Fiordland crested penguin, and two broad categories of demographic reconstruction: recent expansion (macaroni/royal, eastern rockhopper, Adélie, gentoo, king and emperor) or constant/decline since the LGM (chinstrap, northern rockhopper, western rockhopper and Fiordland crested). The solid bar indicates the LGM.

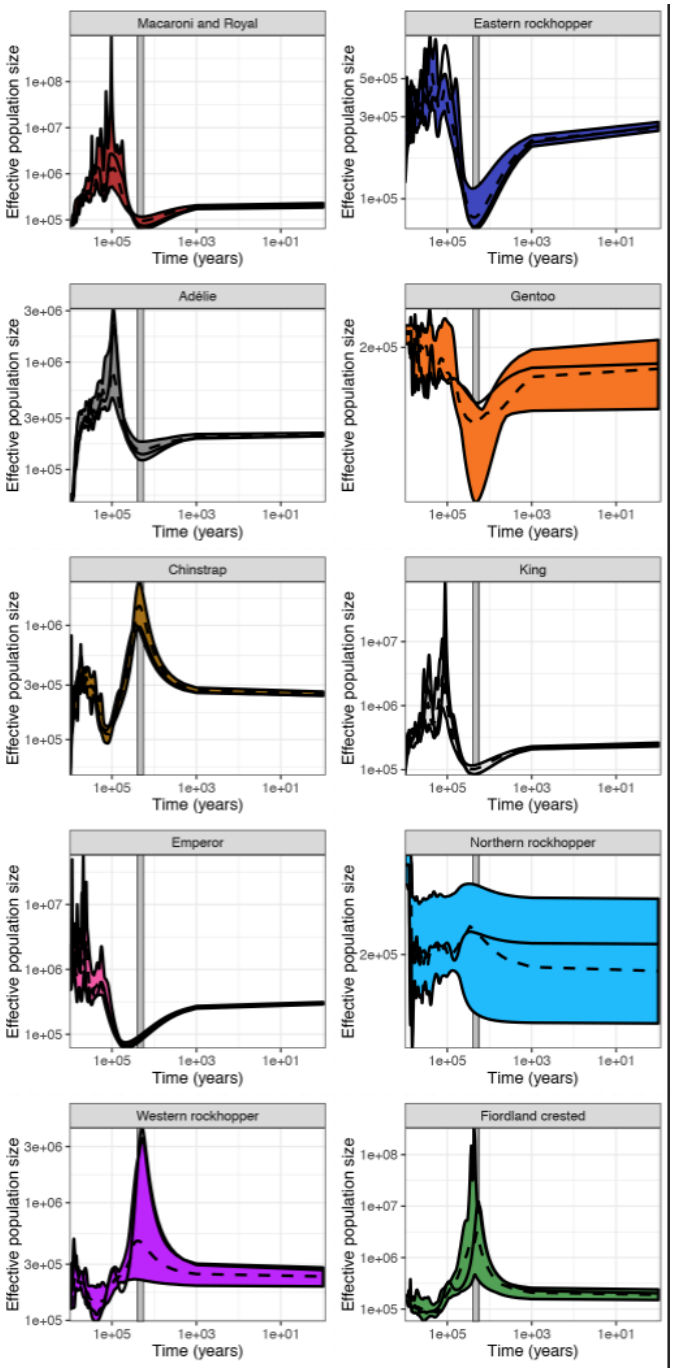

Fig. S3. CubSFS demographic reconstructions for 10 penguin species spanning 1 million years. 95% confidence intervals are given by solid colour intervals. Median for bootstrap replicates is given by the dotted line, and the solid line gives the demographic reconstruction for the amended SFS. Note that macaroni and royal penguins are combined in the same analysis. Snares crested penguin was not analysed due to low sample size. The grey bar indicates the LGM (18 - 25 kya).

Fig. S4. Genetic structure among all sampled *Eudyptes* penguins inferred by Structure, DAPC and PCoA. A) to J) (represented in separate blocks of 2-4 plots) are Structure and DAPC plots. Each block of plots represents a single dataset, analysed to explore relatedness and admixture within the *Eudyptes* genus (e.g. A), between closely related species (e.g. B, C, E and H) and within each species (e.g. B, D, F, G, I and J). Within A) to J), *K* indicates the number of genetic clusters inferred by that corresponding Structure plot, Loc indicates that the corresponding Structure plot used Location Priors, and DAPC indicates that the corresponding plot was created with DAPC, also using location priors. A) is all *Eudyptes* penguins together (without location priors) for *K*=5 to *K*=7; B) is macaroni/royal penguins which are occasionally considered separate species, subspecies, or incipient species, for *K*=2 to *K*=3 (the most likely *K* is *K*=1); C) is the northern, western and eastern rockhopper penguin for *K*=2 to *K*=4 (the most likely *K* is*K*=3); D) is the northern rockhopper penguin, for *K*=2 to *K*=3 (the most likely *K* is *K*=1); E) is the western and eastern rockhopper penguin, which are occasionally considered subspecies, for *K*=2 to *K*=4 (the most likely *K* is

K=2); F) is the western rockhopper penguin, for *K*=2 to *K*=4 (the most likely *K* is *K*=1); G) is the eastern rockhopper penguin for *K*=2 to *K*=3 (the most likely *K* is *K*=1); H) is the Fiordland and Snares crested penguin (which are occasionally considered subspecies) for *K*=2 (the most likely *K*is *K*=2); I) is the Fiordland crested penguin, for *K*=2 to *K*=3 (the most likely *K* is *K*=1); and J) is the Snares crested penguin, for *K*=2 (the most likely *K* is *K*=1). Each Structure run was repeated 20 times. Structure plots for *K*=1 are shown in Fig. 1. Species names and the corresponding population is marked under the final DAPC plot; ELE is Elephant Island, SGE is South Georgia, SSI is South Sandwich Islands, MAR is Marion Island, PEI is Prince Edward Islands, CRZ is Crozet, KER is Kerguelen, MAR* is Marion Island “white-faced” penguins, MAC is Macquarie Island, AMS is Amsterdam Island, GOU is Gough Island, NEW is New Island, PEB is Pebble Beach, CAM is Campbell Island, AUC is Auckland Islands, ANT is Antipodes Islands, COD is Codfish Island, MIL is Milford Sound, JAC is Jackson Head, SNA is the Snares and WES is the Western Chain. The coloured bar underneath the population names indicates which population belongs to which species (e.g. for macaroni and royal penguins). K) to N) are Principal Components Analyses (PCoA); K) encompasses all *Eudyptes* samples and corresponds to the same dataset used to generate the Structure and DAPC plots in A); L) encompasses all macaroni/royal samples, and corresponds to the dataset used to generate the Structure and DAPC plots in B); M) encompasses the rockhopper penguin species, and corresponds to the dataset used to generate the Structure and DAPC plots in C); and N) encompasses the Fiordland and Snares crested penguins, and corresponds to the dataset used to generate the Structure and DAPC plots in H). Eigenvectors one and two are plotted. Each sample in the PCoA corresponds to the colours shown in the DAPC plot.

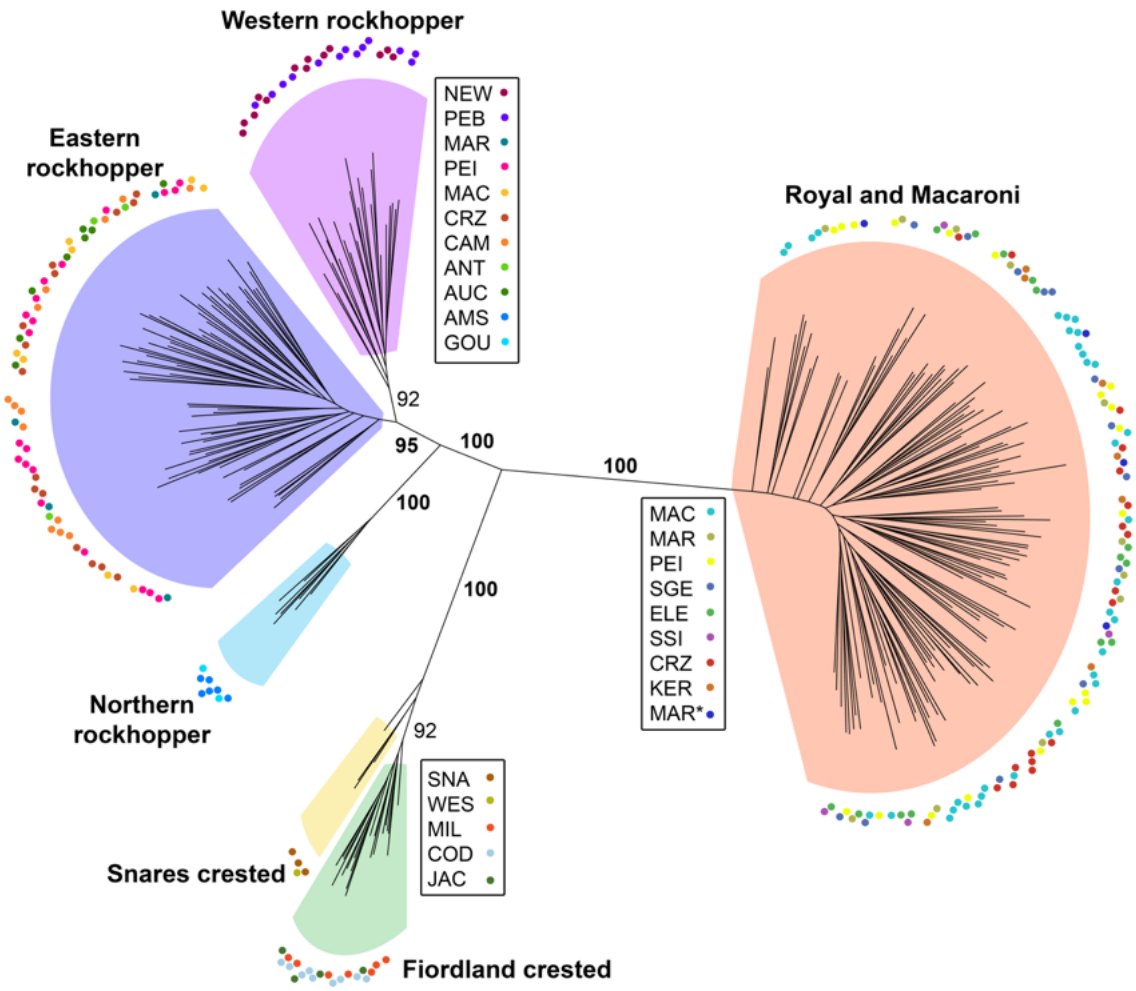

Fig. S5. Unrooted RAxML maximum likelihood phylogenetic tree of *Eudyptes* penguins inferred from 2241 concatenated variable SNPs. Bootstrap values for all major clades are shown. Small circles represent the population that individual was derived from, which is shown in the legends besides each species complex. MAR* indicates Marion Island “white-faced” penguins (see [9]).

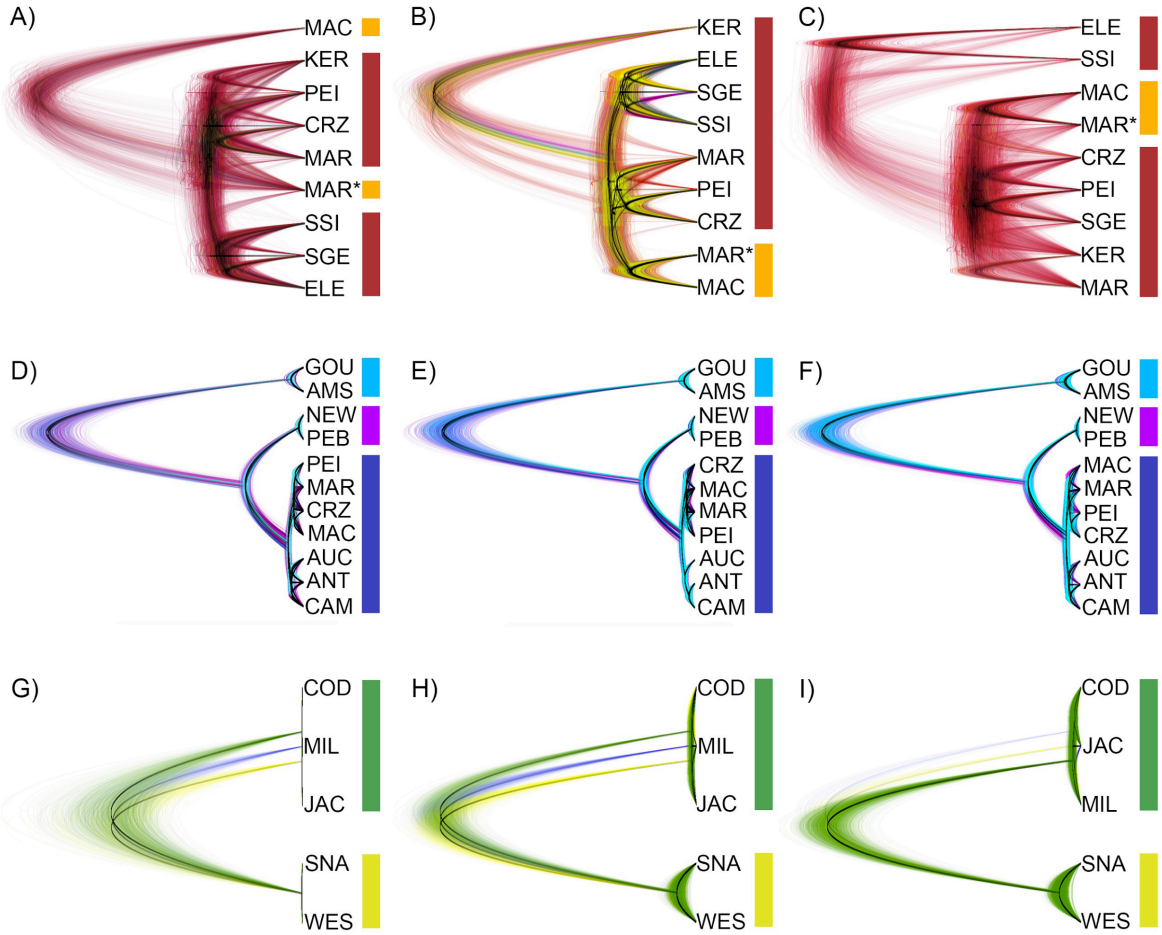

Fig. S6. Alternative SNAPP analyses for closely related *Eudyptes* species. A-C) is macaroni penguin (red) and royal (penguin orange); D-F) is northern rockhopper penguin (light blue), western rockhopper penguin (purple) and eastern rockhopper penguin (dark blue); G-I) is Fiordland crested penguin (green) and Snares crested penguin (yellow). MAR* indicates Marion Island "white-faced" penguins (see [9]).

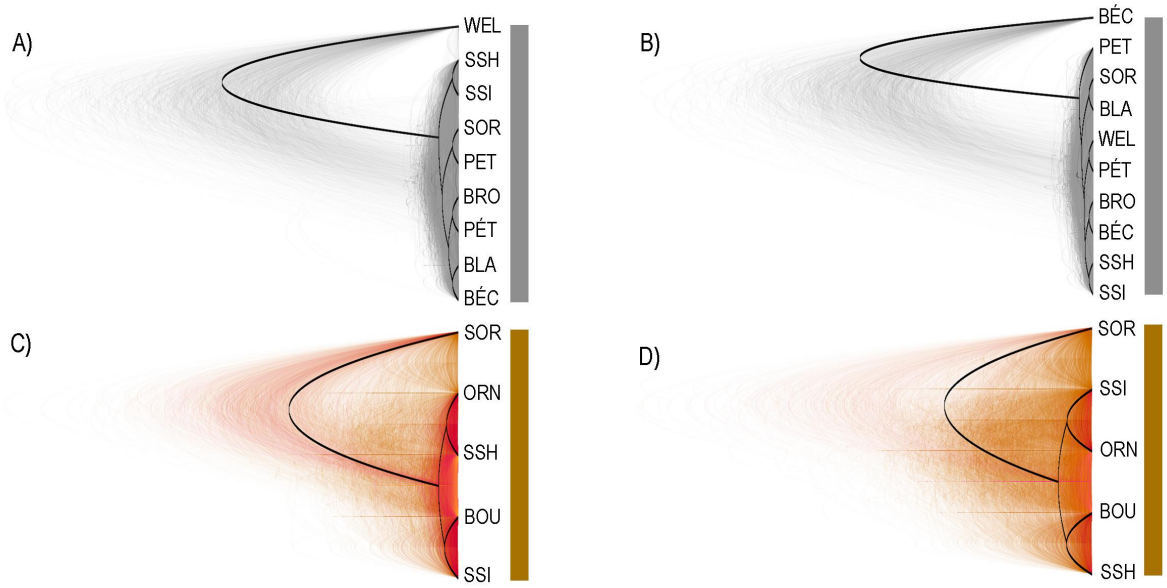

Fig. S7. Alternative SNAPP analyses for Adélie and chinstrap penguins. A-B) is Adélie penguin (grey); C-D) is chinstrap penguin (brown). Note SNAPP analyses for gentoo penguin are published in (31), SNAPP analyses for king penguin are published in (43) and SNAPP analyses for emperor penguin are published in (30).

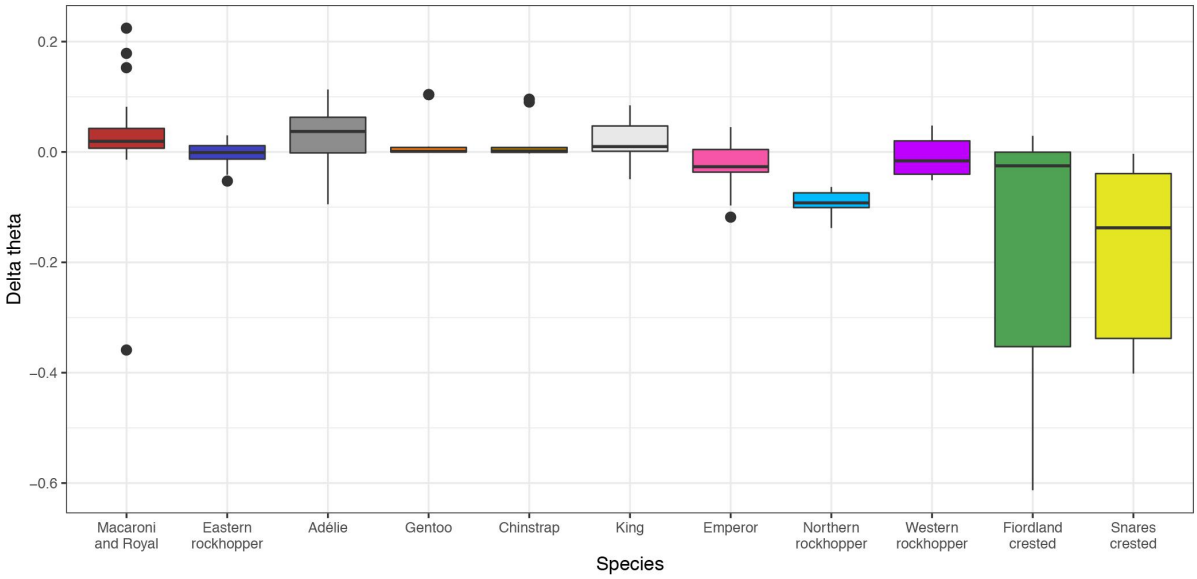

Fig. S8. SNAPP delta theta values for *Eudyptes*, *Pygoscelis* and *Aptenodytes* penguins, based on delta theta of individual sampled locations. Positive theta values indicate expansions, negative, population contractions. Based on this analysis, species that are inferred to have undergone a population expansion include macaroni/royal, eastern rockhopper, Adélie, gentoo, chinstrap and king. Species that are not inferred to have undergone a population expansion following the LGM include northern rockhopper, western rockhopper, Fiordland crested, Snares crested and emperor). Black dots represent theta inferences of locations that are more than 1.5x the size of the inter-quartile range above or below the upper or lower quartile, respectively.

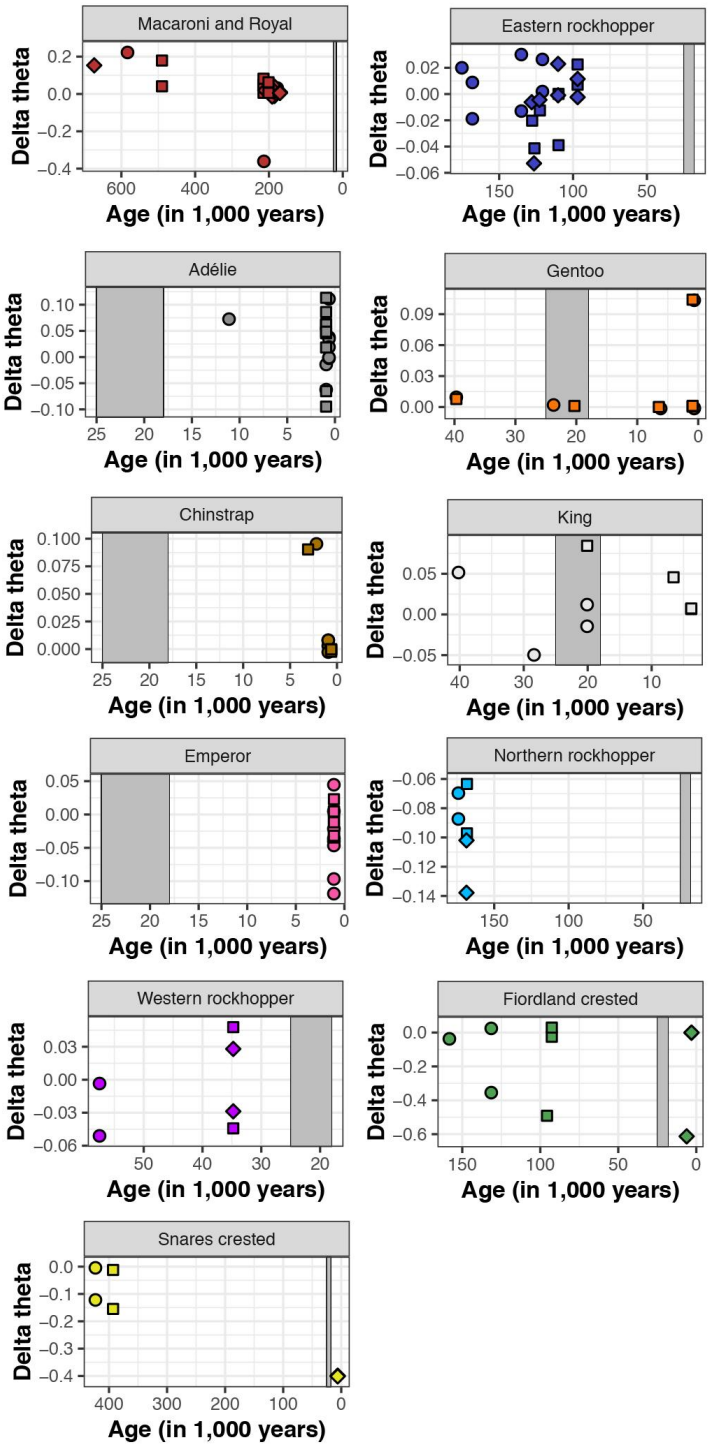

Fig. S9. SNAPP delta theta values for *Eudyptes*, *Pygoscelis* and *Aptenodytes*penguins, in relation to node age. Age is taken from the node ages in number of substitutions from the annotated tree files, divided by the mutation rate and multiplied by the species-specific generation time. Circles are replicate 1 (*Eudyptes*, *Pygoscelis* and *Aptenodytes*), squares are replicate 2 (*Eudyptes*, *Pygoscelis* and *Aptenodytes*) and diamonds are replicate 3 (*Eudyptes* only). The solid grey bar indicates the LGM. Note all emperor, Adélie and chinstrap penguins have post-LGM node ages.

Supplementary Tables

Table S1. Summary of the demographic histories of eleven penguins as inferred by CubSFS. The summary includes whether CubSFS infers an expansion, the time of the expansion, the $N_e$at the start of expansion, the $N_e$at time 1000 years before present and the $N_e$at time 0 years before present, and the fold increase. Results are based on the observed values.

Table S2. Tajima's D for 11 penguins as inferred by δaδi. Negative values suggest a population expansion and positive values suggest a stable or negative N_e. *Note the Fiordland-crested penguin did not have a statistically significant *P*-value, even though it had a negative Tajima's D.

Table S3. Summary of Multi-dice results showing the median, mean and mode for the time of a synchronous expansion event (in years), and the proportion of co-expanding taxa.

Table S4. Sample details for 428 individual penguins across seven *Eudyptes* species. DNA samples in category 1 were too low quality for DArT-seq™ library preparation, DNA samples in category 2 were sent for DArT-seq™ (but failed quality control), and DNA samples in category 3 were sequenced and genotyped by DArT-seq™. DNA samples that were extracted several times are indicated with multiple lab codes.

Table S5. Number of samples, number of variant SNPs and global F_{ST} for each *Eudyptes*, *Aptenodytes* and *Pygoscelis* dataset. *n* individuals 'structure' and *n* SNPs 'structure' refers to the datasets used to generate the Global F_{ST} and corresponding *P*-value (this table), as well as genetic diversity indices, pairwise F_{ST}, PCoA, Structure, DAPC, RAxML and SNAPP analyses (except for Gentoo, see [31]). Global F_{ST} and the corresponding *P*-value for emperor, king, Adélie and chinstrap penguins are derived from (31). Note that some *Eudyptes* individuals were removed in the filtering process, depending on the analysis. Note also that the gentoo penguin dataset has been re-filtered from (31) to only include samples from the South Shetland Islands and Jougla and George's Point on the West Antarctic Peninsula (Fig. 1), except for the Structure and SNAPP analysis (which was undertaken in [31]), which instead included all gentoo penguin individuals spanning 6 localities (*n* individuals = 69, *n* SNPs = 10,108, Global F_{ST} = 0.22, *P*-value = 0.001). *n* individuals 'demographic' and *n* SNPs 'demographic' refers to the further filtered datasets that were used in the demographic analyses of CubSFS, Tajima's D and Multi-dice.

Table S6. Pairwise F_{ST} between all populations of *Eudyptes* penguins calculated across 10,000 bootstraps. F_{ST} is below, corrected *P* value are above.*P* value <0.05 is shown in bold.

Table S7. Genetic Diversity indices calculated for each *Eudyptes* colony. The number of private alleles per each colony was also calculated across all *Eudyptes* species together (global). In all cases, macaroni and royal penguins were analysed together. H_S = expected heterozygosity and H_O = observed heterozygosity. Note one individual each from Jackson Head and The Snares were removed in the global analysis during filtering.

Table S8. Locality, sampling date and sample size information for two *Aptenodytes* and five *Pygoscelis* species used in this study. Genomic information (Single Nucleotide Polymorphisms; SNP) were derived from (31). Note, we did not use gentoo penguin samples derived from Kerguelen (Pointe du Morne), Falkland Islands (Cow Bay) or South Georgia (Bird Island) because they are considered separate taxa to those on the South Shetland Islands (Admiralty Bay) and the West Antarctic Peninsula (George's Point and Jougla Point) (Fig. 1).

References:

1. Kilian A, et al. Diversity Arrays Technology: a generic genome profiling technology on open platforms. *Meth. Mol. Biol.* **888**, 67–89 (2012).

2. Courtois B, et al. Genome-wide association mapping of root traits in a Japonica rice panel. *PLoS One.* **8**, e78037 (2013).

3. Raman H, et al. Genome-wide delineation of natural variation for pod shatter resistance in *Brassica napis*. *PLoS One.* **9**, e101673 (2014).

4. Cruz VM, Kilian A, Dierig DA. Development of DArT marker platforms and genetic diversity assessment of the U.S. collection of the new oilseed crop*lesquerella*and related species. *PLoS One.* **8**, e64062 (2013).

5. Baird NA, et al. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One.* **3**, e3376 (2008).

6. Elshire RJ, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One.* **6**, e19379 (2011).

7. Sansaloni C, et al. Diversity Arrays Technology (DArT) and next generation sequencing combined: genome-wide, high-throughput, highly informative genotyping for molecular breeding of *Eucalyptus. BMC Proceed.* **5**, 54 (2011).

8. Gruber B, Unmack PJ, Berry OF, Georges A. DARTR: An R package to facilitate analysis of SNP data generated from reduced representation genome sequencing. *Mol. Ecol. Res.* **18**, 691–699 (2018).

9. Frugone M-J, et al. Contrasting phylogeographic pattern among *Eudyptes*penguins around the Southern Ocean. *Sci. Rep.* **8**, 17481 (2018).

10. Jouventin P, Cuthbert RJ, Ottvall R. Genetic isolation and divergence in sexual traits: evidence for the northern rockhopper penguin *Eudyptes moseleyi* being a sibling species. *Mol. Ecol.* **15**, 3413–3423 (2006).

11. Banks J, Van Buren A, Cherel Y, Whitfield JB. Genetic evidence for three species of rockhopper penguins, *Eudyptes chrysocome. Polar Biol.* **30**, 61–67 (2006).

12. Christidis L, Boles W. Systematics and taxonomy of Australian birds. CSIRO Publishing. (2008).

13. Jouventin P, Dobson FS. Why penguins communicate: The evolution of visual and vocal signals. Elsevier. (2018).

14. Cole TL, et al. Ancient DNA reveals that the 'extinct' Hunter Island penguin (*Tasidyptes hunteri*) is not a distinct taxon. *Zool. J. Linnean Soc.***182**, 459–464 (2018).

15. Cole TL, et al. Ancient DNA of crested penguins: Testing for temporal genetic shifts in the world's most diverse penguin clade. *Mol. Phylogen. Evol.* **131**, 72–79 (2019).

16. Frugone N-J, et al. More than the eye can see: Genomic insights into the drivers of genetic differentiation in Royal/Macaroni penguins across the Southern Ocean. *Mol. Phylogenet. Evol*. **139**, 106563 (2019).

17. Cole TL, et al. Mitogenomes uncover extinct penguin taxa and reveal island formation as a key driver of speciation. *Mol. Biol. Evol*. **36**, 784–797 (2019).

18. Lischer HEL, Excoffier L. PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics.***28**, 298-299 (2012).

19. Miller MA, Pfeiffer W, Schwartz T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *Gateway Computing Environments Workshop (GCE)*. (2010).

20. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics.* **22**, 2688–2690 (2006).

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

22. Kamvar ZN, Tabima JF, Grünwald NJ. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and or/sexual reproduction. *PeerJ.* **2**, e281 (2014).

23. Kamvar ZN, Brooks JC, Grünwald NJ. Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Front. Gen.* **6**, 208 (2015).

24. Meirmans PG, Van Tienderen PH. GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Mol. Ecol. Notes.* **4**, 792–794 (2004).

25. Nei M. Molecular evolutionary genetics. Columbia University Press. (1987).

26. Excoffier L. AMOVA 1.55 (analysis of molecular variance). *University of Geneva, Geneva*. (1995).

27. Pembleton LW, Cogan NO, Forster JW. StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Mol. Ecol. Res.* **13**, 946–952 (2013).

28. Holm SA. Simple sequentially rejective multiple test procedure. *Scand. J. Stat.* 65–70 (1979).

29. Clucas GV, et al. Dispersal in the sub-Antarctic: king penguins show remarkably little population genetic differentiation across their range. *BMC Evol. Biol.* **16**, 211 (2016).

30. Younger J, et al. The challenges of detecting subtle population structure and its importance for the conservation of Emperor penguins. *Mol. Ecol.* **26**, 3883–3897 (2017).

31. Clucas GV, et al. Comparative population genomics reveals key barriers to dispersal in Southern Ocean penguins. *Mol. Ecol*. **27**, 4680–4697 (2018).

32. Jombart T, Collins C. Analysing genome-wide SNP data using adegenet 2.0.0. (2015).

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

34. Pina-Martins F, Silva DN, Fino J, Paulo OS. Structure_threader: An improved method for automation and parallelization of programs structure, fastStructure and MavericK on multicore CPU systems. *Mol. Ecol. Res.***17**, e268–e274 (2017).

35. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics.* **164**, 1567–1587 (2003).

36. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* **14**, 2611–2620 (2005).

37. Earl DA, von Holdt BM. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Cons. Gen. Res.***4**, 359–361 (2012).

38. Gilbert KJ, et al. Recommendations of utilizing and reporting population genetic analyses: the reproducibility of genetic clustering using the program STRUCTURE. *Mol. Ecol.* **21**, 4925–4930 (2012).

39. Janes JK, et al. The K = 2 conundrum. *Mol. Ecol.* **26**, 3594–3602 (2017).

40. Bouckaert R, et al. BEAST 2: A software platform for Bayesian evolutionary analysis. *PLOS Comput. Biol.* **10**, e1003537 (2014).

41. Bouckaert R, Heled J. DensiTree2: seeing trees through the forest. *BioRxiv*. 012401 (2014).

42. Clucas GV, et al. A reversal of fortunes: climate change 'winners' and 'losers' in Antarctic Peninsula penguins. *Sci. Rep.* **4**, 5024 (2014).

43. Mura-Jornet I, et al. Chinstrap penguin population genetic structure: one or more populations along the Southern Ocean? *BMC Evol. Biol*. **18**, 90 (2018).

44. Cristofari R, et al. Full circumpolar migration ensures evolutionary unity in the emperor penguin. *Nat. Commun.* **7**, 11842 (2016).

45. Cristofari R, et al. Climate-driven range shifts of the king penguin in a fragmented ecosystem. *Nat. Clim. Change.* **8**, 245–251 (2018).

46. Levy H, et al. Population structure and phylogeography of the Gentoo Penguin (*Pygoscelis papua*) across the Scotia Arc. *Ecol. Evol.***6**, 1834–1853 (2016).

47. Vianna JA, et al. Marked phylogeographic structure of Gentoo penguin reveals an ongoing diversification process along the Southern Ocean. *Mol. Phylogenet. Evol.***107**, 486–498 (2017).

48. Danecek P, et al. The variant call format and VCFtools. *Bioinformatics.***27**, 2156–2158 (2011).

49. Waltoft BL, Hobolth A. Non-parametric estimation of population size changes from the site frequency spectrum. *Stat. Appl. Genet. Mo. Biol.* **17**, 3 (2018).

50. Trucchi E, et al. King penguin demography since the last glaciation inferred from genome-wide data. *Proc. R. Soc. Lond. B.* **281**, 20140528 (2014).

51. Forcada J, Trathan PN. Penguin responses to climate change in the Southern Ocean. *Glob. Change Biol.* **15**, 1618–1630 (2009).

52. Lisiecki LE, Raymo ME. A Pliocene-Pleistocene stack of 57 globally distributed benthic δ18O records. *Paleoceanography*. **20**, PA1003 (2005).

53. Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data. *PLOS Genet.* **5**, e1000695 (2009).

54. Xue AT, Hickerson MJ. Multi-dice: r package for comparative population genomic inference under hierarchical co-demographic models of independent single-population size changes. *Mol. Ecol. Res.* **17**, e212–e224 (2017).

55. Csilléry K, François O, Blum MG. abc: an R package for approximate Bayesian computation (ABC). *Methods. Ecol. Evol*. **3**, 475–9 (2012).

56. Napier R. Erect-crested and rockhopper penguins interbreeding in the Falkland Islands. *Brit. Antarct. Surv. B.***16**, 71–72 (1968).

57. Woehler E, Gilbert C. Hybrid rockhopper-macaroni penguins, interbreeding and mixed species pairs at Heard and Marion Islands. *Emu.***90**, 198–201 (1990).

58. White RW, Clausen AP. Rockhopper *Eudyptes chrysocome chrysocome* x macaroni *E. chrysolophus* penguin hybrids apparently breeding in the Falkland Islands. *Mar. Ornith.* **30**, 40–42 (2002).

59. Morrison KW, Sagar PM. First record of interbreeding between a Snares crested (*Eudyptes robustus*) and erect-crested penguin (*E. sclateri*).*Notornis.* **61**, 109–112 (2014).

60. Warham J. The crested penguins. In: The biology of penguins. pp. 189–269. (1975)

61. Shaughnessy PD. Variation in facial colour of the Royal Penguin. *Emu.***75**, 147–152 (1975).

62. de Dinechin M, Ottvall R, Quillfeldt P, Jouventin P. Speciation chronology of rockhopper penguins inferred from molecular, geological and palaeoceanographic data. *J. Biogeog.***36**, 693–702 (2009).

63. Jouzel J, et al. Orbital and millennial Antarctic climate variability over the past 800,000 years. *Science.***317**, 793–796 (2007).

64. Li C, et al. Two Antarctic penguin genomes reveal insights into their evolutionary history and molecular changes related to the Antarctic environment.*GigaScience.* **3**, 27 (2014).

66. Fraser CI, Nikula R, Ruzzante DE, Waters JM. Poleward bound: biological impacts of Southern Hemisphere glaciation. *Trends Ecol. Evol.* **27**, 462–471 (2011).