UNIVERSITY OF SOUTHAMPTON

FACULTY OF NATURAL AND ENVIRONMENTAL SCIENCES

Ocean and Earth Sciences

Gene flow and glacial history: investigating the processes shaping the population structure of penguins in the Southern Ocean

by

Gemma Clucas

Thesis for the degree of Doctor of Philosophy

November 2016

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF NATURAL AND ENVIRONMENTAL SCIENCES

Ocean and Earth Sciences

Thesis for the degree of Doctor of Philosophy

GENE FLOW AND GLACIAL HISTORY: INVESTIGATING THE PROCESSES SHAPING THE POPULATION STRUCTURE OF PENGUINS IN THE SOUTHERN OCEAN

By Gemma Victoria Clucas

Penguins in Antarctica and the sub-Antarctic have survived multiple glacial to inter-glacial transitions. However, anthropogenic climate change may exceed this natural variability in the rate and extent of warming. To accurately monitor the effects of climate change on natural populations, predict their local or global extinction risk, and design effective conservation management plans, we must first be able to define the borders of breeding populations. Regular dispersal of individuals from their natal colony to breed at another colony may mean that multiple colonies constitute a breeding population. By investigating patterns of intraspecific genetic variation, we can estimate the dispersal of individuals between colonies and identify any barriers to gene flow. Barriers to gene flow may be permanent or transient. Permanent barriers to gene flow may ultimately lead to speciation if populations diverge sufficiently, and so by investigating intraspecific population differentiation, we can potentially gain a window into the speciation process. Using both mitochondrial and nuclear genomic markers (single nucleotide polymorphisms, SNPs), and by comparing closely related species with different life histories across a range of habitats, I identify the historical and contemporary barriers to gene flow in emperor, king, gentoo, chinstrap and Adélie penguins. I find that historical glaciation during the last glacial period was significant in driving population divergence among small, refugial populations of Antarctic emperor penguins, but there was no evidence for this in their congener the sub-Antarctic king penguin. This is likely a result of the greater effects of glaciation in the Antarctic compared to the sub-Antarctic. I also show that high levels of contemporary dispersal among emperor penguin colonies has created at least four metapopulations, which span thousands of kilometres of coastline, and that dispersal between these metapopulations has largely eroded the historical population differentiation. King, chinstrap, and Adélie penguins also showed high levels of dispersal across distances of thousands of kilometres. This is in stark contrast to the gentoo penguin, in which colonies separated by less than 50 km were genetically differentiated, and large stretches of open-ocean appear to form permanent barriers to gene flow, promoting allopatric speciation in gentoo penguins. I find that the at-sea range of these species appears to determine these dispersal patterns. The pelagic foraging emperor, king, chinstrap and Adélie penguins, which range from hundreds to thousands of kilometres away from their colonies during the non-breeding season, show high levels of dispersal and gene flow between colonies. However the coastal foraging gentoo penguin, which is resident at or near colonies year-round, shows very little dispersal and gene flow between colonies. Other barriers to gene flow identified were the Polar Front, which restricts gene flow among king and gentoo penguin colonies, and natal philopatry, which also appears to play a role in restricting gene flow among gentoo penguin colonies. The implications of these patterns of dispersal are discussed with respect to monitoring and management, and taxonomic revision is advised to recognise the incipient allopatric speciation identified between populations of gentoo penguins.

Table of Contents

List of Tables

List of Figures

DECLARATION OF AUTHORSHIP

I, *Gemma Clucas*, declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

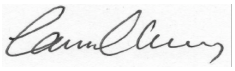
GENE FLOW AND GLACIAL HISTORY: INVESTIGATING THE PROCESSES SHAPING THE POPULATION STRUCTURE OF PENGUINS IN THE SOUTHERN OCEAN.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been published as:

Younger JL, Clucas GV, Kooyman G, Wienecke B, Rogers AD, Trathan PN, Hart T, Miller KJ. Too much of a good thing: sea ice extent may have forced emperor penguins into refugia during the last glacial maximum. Global Change Biology. 2015; 21(6):2215-26.

Clucas GV, Younger JL, Kao D, Rogers AD, Handley J, Miller GD, Jouventin P, Nolan P, Gharbi K, Miller KJ et al. Dispersal in the sub-Antarctic: king penguins show remarkably little population genetic differentiation across their range. BMC Evolutionary Biology. 2016; 16(1):211.

Signed:  .....

Date: 3rd November 2016 .....

Acknowledgements

Firstly I'd like to thank all my collaborators for all the samples they have provided and the feedback they have given on these manuscripts. My particular thanks go to Karen Miller, Damian Kao, Alex Rogers and Mike Polito for their insightful comments throughout. I'd also like to thank Duncan Purdie for going beyond his role as my panel chair and helping me through difficult times at Southampton.

I couldn't have achieved this amount of work without Jane Younger, who has co-authored these manuscripts with me. After thousands of emails, hundreds of hours talking about the data, trips back and forth between the UK and Australia, and just a few disagreements, we have climbed the steep bioinformatics learning curve together, and had a lot of fun along the way.

To all of the Oxford Ocean Research and Conservation group (the ORCs), thank you for all of the wonderful meals and drinks we enjoyed all over Oxford, the advice on my work, and, most importantly, your friendship. Thank you to Alex for welcoming me into the group, and particular thanks to Michelle, Hila and Jesse for always looking out for me.

Although at times I felt like I was doing a PhD in rowing, this journey wouldn't have been half so much fun without the times I spent on the river with Linacre Boat Club, and particularly Gemma, Hannah, Beth and Marie. Some of my happiest days during this PhD were spent on the water with the Linacre rowers, and after all the early mornings, gruelling training, and well-deserved meals together, I have come away with some of my closest friends. Oh, and a blade!

I owe so much of this PhD to the support of my family. Always at the end of the phone, willing to drop everything to come to Oxford or have me over to stay in times of crisis (or when I just needed to escape), you have all been there for me and I'm incredibly lucky to have such wonderful people around me. Words can't express my gratitude to you.

Finally, I'd like to thank my supervisor Tom Hart. Tom took a chance on me when he initially took me on as a research assistant freshly out of my undergrad, and then allowed me to stick around for a Master's and this PhD. During that time he's taken me to Antarctica six times and sunken huge sums of money into this project. I owe him a great deal for all the samples he was able to gather and the amazing opportunities I've had during this PhD, from conferences to kite-surfing. Thank you for trusting me to get on with things and being there when I needed help.

Contributions and Publications

All data chapters in this thesis have been published, submitted or are almost ready for submission. I share the first-author position on all of them. Below is an estimate of my percentage contribution to each chapter and the contributions of others, broken down by the different aspects of the work. The published versions of the papers are included as appendices.

Chapter 2 has been published in *Global Change Biology*:

Younger JL, Clucas GV*, Kooyman G, Wienecke B, Rogers AD, Trathan PN, Hart T, Miller KJ. Too much of a good thing: sea ice extent may have forced emperor penguins into refugia during the last glacial maximum. Glob Change Biol. 2015; 21(6):2215-26.*

Conceived the study: GC - 15%; lab work: GC - 50%; data analysis: GC - 50% [manual checking of sequencing data, estimation of population structure and summary statistics, haplotype networks], JY - 50% [estimation of demographic history and phylogeography]; prepared the manuscript: GC - 60% [abstract, introduction, relevant sections of methods and results, figures], JY - 40% [relevant sections of methods and results, and discussion]. The published version is included as Appendix A, and was included as a chapter in JY's thesis.

Chapter 3 has been published in *BMC Evolutionary Ecology*:

Clucas GV, Younger JL*, Kao D, Rogers AD, Handley J, Miller GD, Jouventin P, Nolan P, Gharbi K, Miller KJ, Hart T. Dispersal in the sub-Antarctic: king penguins show remarkably little population genetic differentiation across their range. BMC Evol Biol. 2016; 16(1):211.*

Conceived the study: GC [50%]; lab work: GC [100%]; analysed the data: GC - 75% [sequencing quality control, manual checking of mtDNA haplotypes, alignment of reads to the genome, SNP calling and filtering, estimation of population structure and summary statistics], JY - 25% [outlier detection and phylogeography]; prepared the manuscript: GC - 50% [relevant sections of methods and results, discussion, figures], JY - 50% [abstract, introduction, and relevant sections of methods and results, figures]. The published version is included as Appendix B.

Chapter 4 is in review with *Nature Communications* at the time of submission:

Younger JL, Clucas GV*, Kao D, Rogers AD, Gharbi K, Hart T, Miller KJ. Recognising metapopulation structure in emperor penguins is critical for their future conservation. Nat Comms.*

Conceived the study: GC - 25%; analysed the data: GC - 75% [sequencing quality control, alignment of reads to the genome, SNP calling and filtering, estimation of population structure and summary statistics], JY - 25% [outlier detection and phylogeography]; prepared the manuscript: GC - 50% [abstract, relevant sections of methods and results, sections of the discussion, figures], JY - 50% [introduction, relevant sections of methods and results, sections of the discussion]. A chapter using this sequencing data was submitted as part of JY's thesis, but all of the data analysis and the writing has been altered for this version.

Chapter 5 is written in the style of *Proceedings of the National Academy of Sciences of the United States of America* but has not yet been submitted.

Clucas GV, Younger JL*, Kao D, Rogers AD, Handley J, Miller GD, Polito MJ, Bost CA, Kooyman G, Wienecke B, Emmerson L, Southwell C, Phillips R, Dunn MJ, Gharbi K, Miller KJ, Hart T. Comparative population genomics in Antarctica: ecological and evolutionary factors driving patterns of intraspecific genetic variation in Antarctic and sub-Antarctic penguins.*

Conceived the study: GC - 30%; lab work: GC – 100%; analysed the data: GC 75% [sequencing quality control, alignment of reads to the genome, SNP calling and filtering, estimation of population structure and summary statistics], JY - 25% [outlier detection and phylogeography]; prepared the manuscript: GC - 60% [abstract, introduction, relevant sections of the methods and results, figures], JY – 40% [relevant sections of the methods and results, discussion].

I have also included, as appendices, the publications that are relevant to this body of work that I wrote or contributed to during the period of candidature, but do not form part of the PhD thesis.

Appendix C: *Clucas GV, Dunn MJ, Dyke G, Emslie SD, Levy H, Naveen R, Polito MJ, Pybus OG, Rogers AD, Hart T. A reversal of fortunes: climate change 'winners' and 'losers' in Antarctic Peninsula penguins. Scientific Reports. 2014; 4:5024.* This work was undertaken for my MRes at the University of Southampton but I reanalysed the data and published it during my PhD.

Appendix D: *Levy H*, Clucas GV*, Rogers AD, Leaché AD, Ciborowski KL, Polito MJ, Lynch HJ, Dunn MJ, Hart T. Population structure and phylogeography of the Gentoo Penguin (Pygoscelis papua)*

across the Scotia Arc. Ecology & Evolution. 2016; 6(6):1834. This work was undertaken during Hila Levy's MSc at the University of Oxford but we reanalysed and published it during my PhD.

Appendix E: *Freer JJ, Mable BK, Clucas G, Rogers AD, Polito MJ, Dunn M, Naveen R, Levy H, Hart T. Limited genetic differentiation among chinstrap penguin (Pygoscelis antarctica) colonies in the Scotia Arc and Western Antarctic Peninsula. Polar Biol. 2015; 38(9):1493-502.* This was Jennifer Freer's Master's dissertation project, and I helped with lab work and writing up the manuscript.

*** Indicates that I hold joint-first authorship.

Chapter 1: Introduction

The aim of this thesis was to investigate the roles of external and intrinsic factors in restricting gene flow across a range of Antarctic and sub-Antarctic penguin species, and to determine the role of glacial history in creating patterns of intraspecific genetic variation.

1.1 Intraspecific genetic variation

Understanding the micro-evolutionary processes that contribute to the creation of patterns of intraspecific genetic variation is one of the central goals of intraspecific phylogeographic research (Avise *et al.* 1987). By investigating the evolutionary history of populations within a species, through space and through time, we can identify the processes that create patterns of genetic diversity. The genetic divergence of populations from one another is the first step towards speciation (Wright 1940). Hence, studying the micro-evolutionary forces that create or diminish intraspecific genetic differentiation can provide us with a window into the speciation process. The study of genetic variation is an old topic (Dobzhansky 1937; Fisher 1930), yet new technologies now allow a whole-genome understanding of genetic divergence (Seehausen *et al.* 2014), and the role that gene flow may play in the process (Nosil 2008).

Gene flow has often been viewed as a constraint on evolution due to its homogenising effect on populations, undoing the work of selection or genetic drift in altering allele frequencies among populations (Fisher 1930; Slatkin 1987; Wright 1931, 1932). However, it has long been recognised that adaptation can occur in the presence of substantial gene flow (Ehrlich & Raven 1969), provided that the strength of selection outweighs the influence of migrants entering a population (Haldane 1930). Migration and recombination will eventually homogenise the genomes of individuals in different populations, and in the inverse situation, genetic drift will eventually differentiate populations throughout the genome. However when selection acts in the presence of gene flow, narrow islands in the genome may become genetically differentiated whilst the majority of the genome remains relatively homogeneous (Feder *et al.* 2012). For example, Toews *et al.* (2016) found just six regions of the genomes of golden-winged and blue-winged warblers that were genetically differentiated from one another, after a long history of hybridisation and gene flow between the species. The genomic regions that had remained differentiated whilst the rest of the genomes became homogenised were primarily associated with feather development and plumage colouration, and hence were likely under sexual selection for features associated with mate choice in each species. As an increasing number of examples of speciation with gene flow emerge, along with theoretical studies investigating these mechanisms and advances in

methods to detect its occurrence ("isolation with migration" (Hey 2006, 2010)) it seems even more relevant to study intraspecific population differentiation, which is where we are most likely to discover examples of incipient speciation.

Genetic drift, the stochastic process that results in population differentiation throughout the genome, can be overcome by migration into a population by a small number of individuals per generation (Wright 1932). Therefore genetic drift can only occur where there are barriers to gene flow between populations. These barriers can be categorised as either extrinsic barriers, where some external force prevents gene flow, or intrinsic barriers, where gene flow is prevented despite the lack of a physical barrier existing. Extrinsic barriers can be either biotic or abiotic in nature. External abiotic barriers to gene flow in the terrestrial realm are common, ranging from rivers, mountain ranges, and glaciers, to smaller-scale habitat discontinuities such as roads, urban areas, dams, and natural breaks in habitat (Storfer *et al.* 2007). In the marine realm, where dispersal is often thought to be much more widespread (Palumbi 1992), external abiotic barriers to gene flow include oceanic fronts between different water masses (de Dinechin *et al.* 2009; Thornhill *et al.* 2008), depth (Shaw *et al.* 2004), and habitat fragmentation, such as discontinuities in reefs caused by areas of soft sediment (Bernardi 2000). Examples of external biotic barriers to gene flow include predation, resource competition, and interspecific interference, all of which can prevent the movement of individuals from one population to another. Intrinsic barriers to gene flow include behaviours such as natal philopatry (Piertney *et al.* 1998), differential timing of breeding (Friesen *et al.* 2007b), differential mate choice (Seehausen *et al.* 1997) and environmental adaptation (Pespeni & Palumbi 2013).

Barriers to gene flow can be transient, such as the advancement and retreat of glaciers, or the local extinction of a predator or competitor. When these barriers break down, populations that were isolated and subject to genetic drift can come into contact with one another. If the populations were isolated for a sufficient period of time, then on secondary contact, hybridisation between the two populations may not occur due to other intrinsic barriers such as behavioural incompatibilities, pre-zygotic isolation due to mechanical or gametic incompatibilities, or post-zygotic isolation, such as reduced hybrid fitness (Templeton 1981). Isolated populations can be subject to both genetic drift and selection for their local environment, and so, for a full understanding of the mechanisms that have resulted in population divergence, studies should investigate population divergence at both selectively neutral and putatively adaptive genetic loci. This is even more important in marine species, where large population sizes preclude rapid divergence by genetic drift, but significant divergence at loci putatively under selection has been observed (Pespeni & Palumbi 2013; Pogson & Fevolden 2003).

Patterns of genetic diversity must also be considered through the lens of demographic history. Population expansions, declines, range shifts and changes in gene flow will leave their own footprints on the genome (Mazet *et al.* 2016), such as runs of homozygosity (Kirin *et al.* 2010). Therefore the mechanisms underlying population differentiation cannot be understood without consideration of demographic history (Luikart *et al.* 2003). Global climate has fluctuated dramatically between glacial and inter-glacial periods during the Quaternary period (2.58 Ma to the present), and these fluctuations have largely shaped the genetic structure of species and communities from temperate to high latitudes (Hewitt 2000). Throughout the Pleistocene and early Holocene, climatic shifts were one of the major drivers of changes in species' ranges, changes in abundances, species extinctions and in the formation of genetically distinct populations (Hewitt 1996; Parmesan & Yohe 2003; Twitchett 2006; Walther *et al.* 2002). The isolation of small populations within glacial refugia can lead to rapid divergence through genetic drift (Petit *et al.* 2003). Indeed, much of the genetic differentiation among terrestrial populations in North America and Northern Europe can be attributed to range contraction and isolation of small populations in refugia during the last glacial period, followed by range expansions, founder effects and secondary contact during the postglacial and current interglacial periods (Hewitt 1996; Petit *et al.* 2003). These patterns are more apparent in species found at higher latitudes, where the effects of climate changes are more extreme because of the advance and retreat of polar ice-caps (Williams 1998).

1.2 Intraspecific genetic variation of seabirds

High-latitude seabirds are an excellent system for the study of population differentiation and speciation. The majority of seabirds are long-lived and capable of long-distance dispersal (Schreiber & Burger 2001). Therefore, extrinsic abiotic barriers to gene flow between populations are few, allowing the more subtle and/or locus-specific mechanisms of population differentiation to be studied. Seabirds also tend to have different distributions in their breeding and non-breeding seasons; with populations often dispersing over large areas of open ocean during the non-breeding season, yet annually contracting into discrete terrestrial colonies during the breeding season (Coulson 2001). Furthermore, seabirds inhabiting higher latitudes will have experienced multiple cycles of population expansion and contraction throughout the Quaternary period, and so the effects of demographic changes on population differentiation can also be studied.

In a recent review Friesen (2015) identified the factors most important for restricting gene flow between seabird populations. Gene flow was restricted by physical barriers, such as discontinuities in their oceanic habitat, such as land or ice (e.g. northern fulmar (Kerr & Dove

2013), Trinidade petrel (Brown *et al.* 2010), sooty terns (Avise *et al.* 2000)), but a variety of non-physical barriers also limited gene flow. The main non-physical barriers included differences in the oceanic regimes experienced by populations (e.g. Cory's shearwaters (Gómez-Díaz *et al.* 2009), the distribution of individuals during the non-breeding season (e.g. black-browed albatross (Burg & Croxall 2001)), their foraging distributions during the breeding season (e.g. Hawaiian petrels (Wiley *et al.* 2012)), differences in breeding phenology (e.g. band-rumped storm petrel (Friesen *et al.* 2007b), and philopatry (e.g. shy albatross (Abbott & Double 2003). Many of these barriers are intrinsic (behavioural) in nature, and, without fine-scale, detailed population genetic analyses, could have gone unnoticed. Population genetics therefore has a crucial role to play in illuminating potential mechanisms of population differentiation and speciation. To date, most studies of seabird population differentiation and speciation have so far relied on small numbers of neutrally evolving loci, and so the role of adaptation with gene flow may have been underestimated, as signals of local adaptation will have been missed.

1.3 Penguin evolution and ecology

Penguins are colonial breeding, flightless seabirds, highly adapted for swimming and diving. Their bones are dense rather than pneumatic to aid diving, their wings have been converted to flippers for swimming, their legs are set far back on the body to act as a rudder, and their feathers have been modified to form a dense, insulating cover, allowing them to survive long periods at sea and temperatures below -40 °C. There are currently 18 recognised species, distributed throughout the Southern Hemisphere from Antarctica to the Galapagos Islands (Borboroglu & Boersma 2013); with all species associating with cool ocean currents. The greatest numbers of penguins are found in the Antarctic and sub-Antarctic, where they form 80% of the avian biomass (Williams 1995), but the greatest diversity of species is found between 45 and 60 °S, particularly around the mainland and islands of New Zealand.

Penguins originated more than 60 Ma (Slack *et al.* 2006), during a warm period in Earth's history. The fossil record reveals that they dispersed rapidly from New Zealand to reach Antarctica, South America and Australia by the Middle Eocene, 50 Ma (Clarke *et al.* 2007). The first penguins to arrive in Antarctica were distantly related to extant penguins (Spheniscidae) and arrived at a time before permanent ice sheets were established (Clarke *et al.* 2007; Ksepka *et al.* 2006). All extant penguins (the crown group, Figure 1.1) derived from a lineage that originated in the Middle Miocene around 16 Ma. The most recent common ancestor of the crown group has been estimated at 12.45 Ma (95% HPD 10.44 – 15.08) (Gavryushkina *et al.* 2015), which is also supported by fossil evidence (Stucchi 2002). The recent emergence of crown penguins places

many of the divergences among modern taxa within the last 2 Myr, with 13 of 19 extant species emerging within this period (Gavryushkina *et al.* 2016).

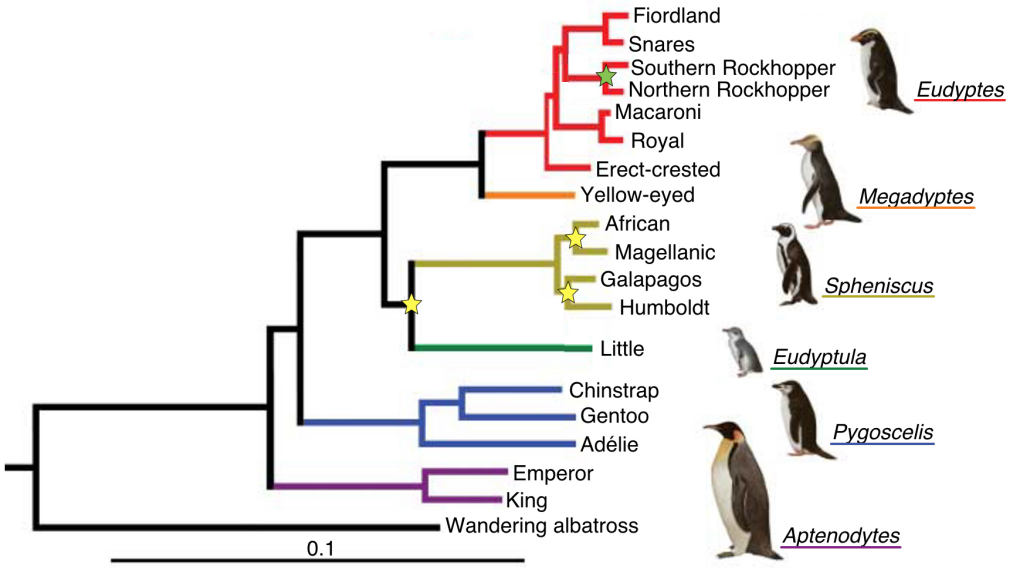


Figure 1.1. Phylogeny of the extant crown group penguins based on one nuclear and four mitochondrial genes, with the wandering albatross as an outgroup, adapted from Baker *et al.* (2006). Scale bar represents the expected number of substitutions per site. Yellow stars mark speciation via dispersal-vicariance events, green star marks speciation following ecological isolation by water mass.

Dispersal-vicariance events (Ronquist 1997) appear to be responsible for much of the diversity within the "coastal" crown penguins, that is, those that tend to remain in continental shelf waters year-round. Major oceanic currents are thought to have facilitated dispersal to new habitats (Ksepka *et al.* 2006), followed by allopatric speciation (Figure 1.1). The circumpolar current and West Wind drift may be responsible for the split between the *Spheniscus* (banded penguin) and *Eudyptula* (little penguin) genera, when the ancestor of *Spheniscus* crossed the south Pacific from Australia/New Zealand to South America (Bertelli & Giannini 2005; Ksepka *et al.* 2006). The north-westward Humboldt current may then have been responsible for the dispersal of *Spheniscus* to the Galapagos, resulting in the divergence between the Pacific-coast Humboldt penguin (*Spheniscus humboldti*) and Galapagos penguin (*Spheniscus mendiculus*) (Bertelli & Giannini 2005; Ksepka *et al.* 2006). Finally the westward Brazil-Benguela current may be responsible for the dispersal-vicariance event that gave rise to the Magellanic penguin (*Spheniscus magellanicus*) from the African penguin (*Spheniscus demersus*) (Ksepka *et al.* 2006).

Dispersal-vicariance events seem unlikely to have been the mechanism for speciation among the more "pelagic" species that annually disperse into the open ocean during the non-breeding season, as they are more likely to maintain gene flow between disparate populations, low levels of which would preclude allopatric speciation (Slatkin 1987). Instead, barriers to gene flow are likely to include those identified by Friesen (2015) such as oceanic regime, phenology, philopatry and distributions during the non-breeding season, in addition to local adaptation. Evidence that oceanic regimes have played a role in the radiation of pelagic penguins comes from a study into the speciation of rockhopper penguins (genus: *Eudyptes*) by de Dinechin *et al.* (2009). The mid-Pleistocene transition, around 0.9 Ma, was accompanied by a major shift southwards of the subtropical convergence. The convergence represents a change in salinity and rise in sea surface temperature of 5 °C over just a few kilometres, between sub-Antarctic waters and the warmer waters of the subtropical gyres. Gough Island, previously situated south of the convergence and surrounded by cool sub-Antarctic waters, hosted a population of stem rockhopper penguins that were widespread across the sub-Antarctic islands at the time. When, 0.9 Ma, the subtropical convergence shifted southwards to a new position south of Gough Island, the island became ecologically isolated from the other sub-Antarctic islands by the warm waters now surrounding it. Studies of mitochondrial DNA have shown that the divergence between the northern rockhopper penguin (*Eudyptes moseleyi*), found on Gough Island, and the rest of the rockhopper clade occurred at the same time, strongly suggesting that the ecological isolation of Gough Island by the movement of the water masses appears to have driven the divergence and speciation of the northern rockhopper penguin (de Dinechin *et al.* 2009 and see Figure 1.1).

Oceanic regimes may also act as a barrier to gene flow in the coastal gentoo penguin (*Pygoscelis papua*) which breeds across islands of the sub-Antarctic. The gentoo penguin is found both above and below the Polar Front, which separates the relatively warm waters of the southern Atlantic Ocean from the cold waters of the Antarctic circumpolar current (ACC). The Polar Front has been found to limit dispersal across a range of taxa (Rogers 2012; Shaw *et al.* 2004; Thornhill *et al.* 2008) and appears to prevent dispersal between gentoo penguin colonies, with populations either side being differentiated at both nuclear (Levy *et al.* (2016) and see chapter 5) and mitochondrial markers (Clucas *et al.* 2014).

In this thesis, the genetic variation within four pelagic species of penguin was compared, alongside the coastal gentoo penguin. These species differ in their foraging ranges during the breeding and non-breeding seasons (reviewed by Ratcliffe & Trathan 2011). The gentoo penguin, as a coastal seabird, has the most restricted range, mostly foraging in inshore waters throughout its lifetime. Adult birds forage very close to colonies during the breeding season, seldom travelling more than 40km from the colony. During the non-breeding season they are rarely seen more than

50km offshore, although they do travel around the coastline within archipelagos (Clausen & Pütz 2003; Masello *et al.* 2010). However, rare sightings of individuals ~2000 km from the nearest colony suggest that they are capable of dispersing further (Enticott 1986; Voisin 1979). Chinstrap penguins (*Pygoscelis antarctica*) have a similarly restricted foraging range during the breeding season, generally foraging within 60km of their colonies (Lynnes *et al.* 2002), but are known to disperse much more widely during the non-breeding season, having been tracked up to 1500km from their colonies (Trivelpiece *et al.* 2007; Wilson *et al.* 1998), with one individual from Bouvet Island travelling 1863 km to the South Sandwich Islands to moult (Biuw *et al.* 2010) and another vagrant observed on Gough Island more than 3000km away (Enticott 1986). The final species in the *Pygoscelis* genus, the Adélie penguin (*Pygoscelis adeliae*), ranges much more widely than its congeners during the breeding season, with individuals tracked foraging up to 270 km from their colonies during incubation (Clarke *et al.* 2006). The Adélie, like the chinstrap penguin, also disperses widely during the non-breeding season, to forage at the Antarctic pack-ice edge over 1000 km from some colonies (Ballard *et al.* 2010; Dunn *et al.* 2011).

The other two pelagic species compared in this thesis were also congeners: the emperor (*Aptenodytes forsteri*) and king (*Aptenodytes patagonicus*) penguins. The emperor penguin mostly forages up to 150 km from its colonies during the breeding season (Zimmer *et al.* 2008), although some females disperse up to 500 km whilst males are incubating eggs (Wienecke & Robertson 1997). Foraging range may be constrained by having to cross large stretches of sea ice on foot (Wienecke & Robertson 1997), and, where ice is extensive, emperor penguins rely upon gaps in the ice formed by tides cracks and polynyas (areas of ocean kept free of ice by winds or upwelling of warm waters) to forage (Kirkwood & Robertson 1997). During the non-breeding season, individuals have been tracked travelling up to 1400 km from their colonies (Kooyman *et al.* 2000; Wienecke *et al.* 2004). King penguins appear to be more dispersive than emperor penguins during the breeding season, which can be largely attributed to their strong preference for foraging at frontal systems, and almost exclusively at the Polar Front (Bost *et al.* 2009). The distance of the front from the colony determines their foraging range, which can be up to 600 km, and varies from year to year (Bost *et al.* 2015; Scheffer *et al.* 2012). During the non-breeding season, they are also highly dispersive, travelling south to the marginal ice zone up to 1800 km away (Bost *et al.* 2004; Pütz 2002).

An important aspect of seabird life history, which has largely gone unstudied due to logistical constraints, is the effect of juvenile dispersal on population structure. The majority of seabirds are long-lived and have delayed onset of breeding, with juveniles not being sighted at colonies for a number of years before they return to begin breeding (Coulson 2001). Tracking juvenile seabirds is challenging; battery-life constrains the deployment of loggers that automatically transmit

positions, as they do not last long enough to capture juvenile dispersal, and the chances of retrieving multi-year loggers such as geolocators are very low. We therefore have very little data on the juvenile dispersal of these species, although it is likely to be a major mechanism in facilitating gene flow by allowing individuals to prospect other colonies. Sightings of vagrant emperor penguins at South Georgia, over 1500 km from the nearest colony, were mostly juvenile birds (Enticott 1986), and of the few juvenile emperor penguins that have been tracked, most seem to travel north to the Polar Front over 2000 km away (Kooyman *et al.* 1996), with one travelling over 7000 km during the first months after fledging (Thiebot *et al.* 2013). The juvenile range of king penguins has not been studied by tracking, however banding studies suggest juveniles visit other archipelagos more often than adults birds, sometimes visiting colonies up to 5,600 km away (Weimerskirch *et al.* 1985). Fledgling Adélie penguins were tracked from Béchervaise Island in East Antarctica and ranged between 500 km and 1,900 km away from the colony before transmissions ended (Clarke *et al.* 2003). There is no data on the juvenile movements of either chinstrap or gentoo penguins. Juvenile chinstrap penguins are likely to have a large juvenile range based on their dispersal as adults, whereas gentoo penguins are likely to have a very limited juvenile dispersal range. This would also be supported by the age of first breeding, which in gentoo penguins is two to three years, and ranges from three to eight years in the other species (Williams 1995).

The distribution of colonies within each species may also partially determine their population structure. Those species with a more continuous breeding distribution may show less population structure compared to the more patchily distributed species (Wright 1943). If dispersal is distance limited and colonies are distributed linearly, such as emperor penguin colonies around the coastline of Antarctica (Fretwell *et al.* 2012), individuals may disperse between adjacent colonies according to the stepping stone model (Kimura & Weiss 1964), leading to isolation by distance (Wright 1943). Of the species studied here, the emperor penguin has the most continuous distribution (Figure 1.2) due to its ability to breed around the coast of Antarctica on fast ice (ice that is held “fast” either to the coastline, icebergs or the ice fronts of glaciers and ice shelves). The Adélie penguin is constrained to breeding on ice-free ground, where pebbles are available for nest building, and within walking distance of the ice-edge. Therefore, whilst its distribution is still somewhat regular around the coastline of Antarctica, some gaps in its distribution do occur, for example in the Weddell Sea and Dronning Maud Land where glaciation is extensive (Lynch & LaRue 2014). It is also found breeding on the islands of the Scotia Arc and so its range is divided into regions of continuity and patchiness (Figure 1.2). Gentoo, chinstrap and king penguins all breed on islands and archipelagos, and hence have highly patchy breeding distributions. The chinstrap penguin is largely restricted to the islands of the Scotia Arc, whilst both the gentoo and

king penguins are distributed across the sub-Antarctic islands which encircle Antarctica, and so have the patchiest distributions out of any of the species (Borboroglu & Boersma 2013)(Figure 1.2).

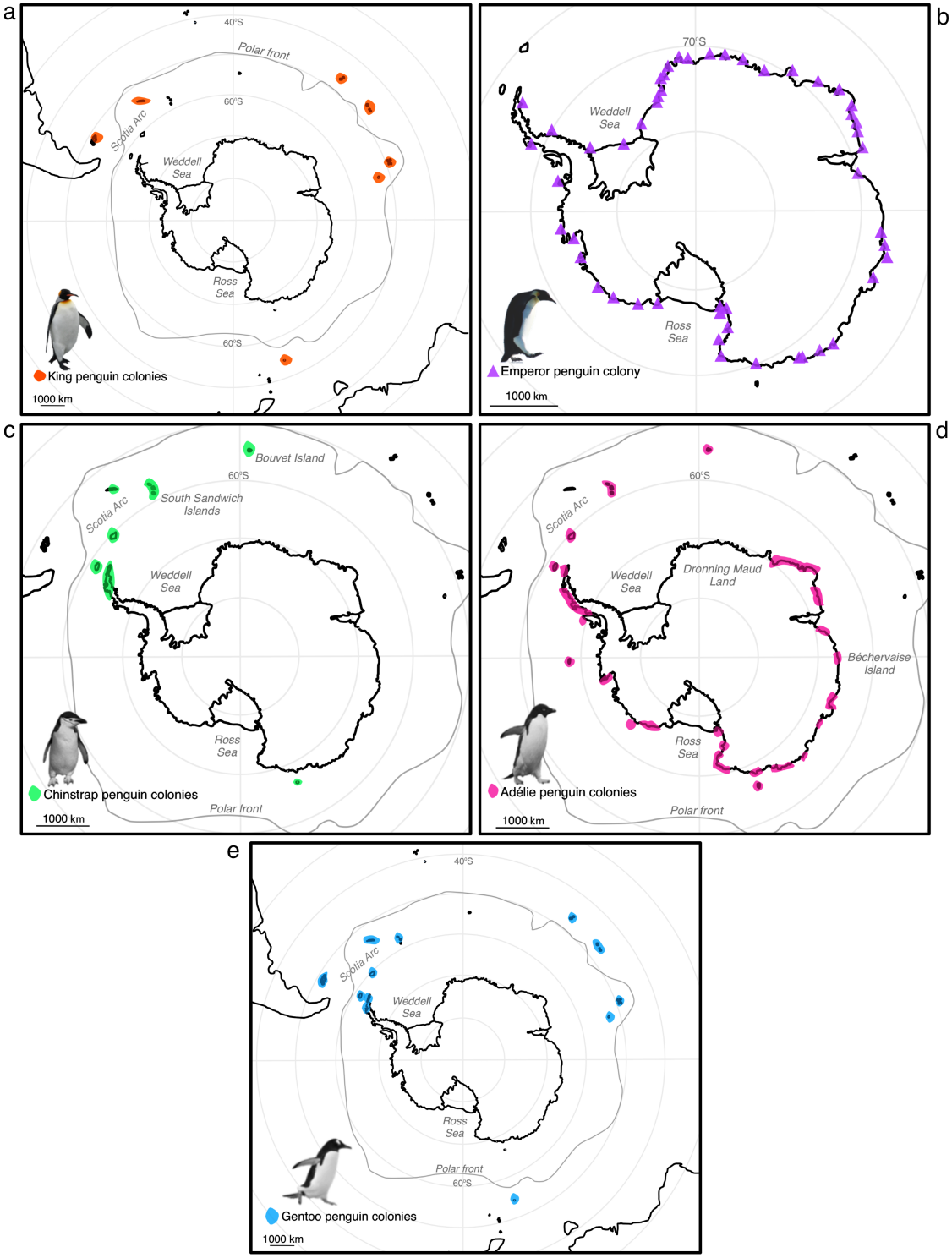


Figure 1.2. Breeding distributions of the penguins studied in this thesis.

1.4 Genetic markers used to assess population structure and demographic history

1.4.1 Mitochondrial DNA

The haploid mitochondrial genome is inherited through the maternal line, meaning its effective population size is one quarter the size of the nuclear genome. As a result, it has been suggested to be a powerful tool in recovering the ancestral relationships among taxa that diverged recently (Moore & DeFilippis 1997; Zink & Barrowclough 2008). This is because the mitochondrial gene tree is more likely to be congruent with the species tree than nuclear genes, which may suffer from incomplete lineage sorting due to larger effective population size (Moore 1995). Therefore, in the time before next generation sequencing (NGS) could generate large amounts of DNA sequence data, many studies relied on mitochondrial sequences to study species and sub-species relationships, as well as higher level taxonomy (Zink & Barrowclough 2008).

The rapid evolutionary rate of the mitochondrial control region compared to more conserved mitochondrial genes (Baker & Marshall 1997), coupled with the small effective population size and putatively neutral evolution makes it a good marker for recovering the recent demographic history of populations and population structure. The signals of population bottlenecks associated with isolation in glacial refugia, along with other population processes associated with palaeoclimatic events, such as leading edge colonisation and secondary contact, have been uncovered from the mitochondrial DNA of a range of taxa (e.g. lesser snow geese (Quinn 1992), elephant seals (de Bruyn *et al.* 2014)). Since the introduction of "skyline" methods that describe the change in effective population size through time using coalescent theory (Drummond *et al.* 2005; Heled & Drummond 2008; Pybus *et al.* 2000), the mitochondrial control region found even greater popularity (e.g. (Campos *et al.* 2010; Shapiro *et al.* 2004)). However, the maternal inheritance of the mitochondrial genome means that it may not reflect processes that are severely sex-biased, such as male-dominated dispersal (Zink & Barrowclough 2008).

1.4.2 Restriction site associated DNA sequencing and its application to population genomics

Restriction site associated DNA sequencing (RADSeq) (Baird *et al.* 2008) can be used to identify hundreds of thousands of single nucleotide polymorphisms (SNPs) in species without advanced genomic resources, such as a reference genome. There are now a number of variations on the original method (reviewed by Andrews *et al.* (2016)) but all rely on the use of restriction enzymes to randomly target regions of the genome for sequencing. This allows a subset of the genome to

be sequenced across tens to hundreds of individuals, and has therefore been adopted as the chosen method for conducting population genomic research in wild populations. Restriction sites can occur in both coding and non-coding regions of the genome, and will often be conserved across closely related taxa, and so the sequence data generated can be used to answer a variety of ecological and evolutionary questions.

Using thousands of unlinked SNPs, estimates of genome-wide processes can be made with a high degree of precision and accuracy, such as estimating inbreeding coefficients (Li *et al.* 2011), effective population sizes (Locke *et al.* 2011) and migration rates (Allendorf *et al.* 2010) (but see chapter 3 for a discussion of the limitations of the methods currently available for estimating migration rates with unphased SNP data). This method should allow fine-scale population structure to be discerned, and neutral SNP datasets to be identified through the removal of putatively selected loci, which can have a confounding effect on estimates of population differentiation (Allendorf *et al.* 2010). Similarly, non-neutral loci can dramatically affect the branch lengths of coalescent based trees (Landry *et al.* 2002), and so the identification and removal of them from genomic datasets can allow for accurate estimates of demographic parameters, such as population size changes, as well as the correct branch lengths on the underlying tree. However methods to estimate demographic history from SNP datasets are still in their infancy, and are highly computationally demanding (Bouckaert *et al.* 2014). Finally, the abundance of loci provided by RADSeq also allows for genome scans to identify locus-specific process, such as local adaptation and introgression, when a reference genome is available (Hohenlohe *et al.* 2010; Nadeau *et al.* 2014).

1.5 Thesis and chapter aims

The aim of this thesis was to identify the evolutionary and ecological processes that have determined the intraspecific genetic variation of penguins in the Southern Ocean, with a view to understanding how these processes may contribute to speciation.

Chapter 2 focuses on the glacial history of Antarctica and how increased sea ice negatively impacted the emperor penguin. This was assessed using mitochondrial DNA isolated from modern and ancient emperor penguin samples from around the continent. The published article from *Global Change Biology* is included in Appendix A.

Chapter 3 presents the population structure of king penguins across their range, and identifies the role of dispersal in maintaining connectivity between king penguin populations. This chapter uses genome-wide SNPs identified through RADSeq and mitochondrial DNA. The published article from *Ecology & Evolution* is included in Appendix B.

Chapter 4 uses SNPs identified through RADSeq data to uncover metapopulation structure within emperor penguins and critically contrasts this with a previous study that failed to find such a result. The chapter argues for greater recognition of study limitations and the need for more rigorous scientific methods when the results are likely to be used for conservation management of threatened species. This chapter is in review at *Nature Communications*.

Chapter 5 identifies the processes that have determined the population structure of king, emperor, chinstrap, Adélie and gentoo penguins across their range, using RADSeq data. This chapter uses the data presented in chapters 3 and 4 for king and emperor penguins, but also includes novel analysis of gentoo, chinstrap and Adélie penguin population structure. The study benefits from a comparative framework where population structure is compared among closely related species with overlapping distributions, and across a range of habitats.

Chapter 2: Too much of a good thing: sea ice extent may have forced emperor penguins into refugia during the last glacial maximum[1]

2.1 Abstract

The relationship between population structure and demographic history is critical to understanding microevolution and for predicting the resilience of species to environmental change. Using mitochondrial DNA from extant colonies and radiocarbon-dated subfossils, we present the first microevolutionary analysis of emperor penguins (*Aptenodytes forsteri*) and show their population trends throughout the last glacial maximum (LGM, 19.5 – 16 kya) and during the subsequent period of warming and sea ice retreat. We found evidence for three mitochondrial clades within emperor penguins, suggesting that they were isolated within three glacial refugia during the LGM. One of these clades has remained largely isolated within the Ross Sea, while the two other clades have intermixed around the coast of Antarctica from Adélie Land to the Weddell Sea. The differentiation of the Ross Sea population has been preserved despite rapid population growth and opportunities for migration. Low effective population sizes during the LGM, followed by a rapid expansion around the beginning of the Holocene, suggest that an optimum set of sea ice conditions exist for emperor penguins, corresponding to available foraging area.

2.2 Introduction

Genetic data both from modern and subfossil samples, palaeo-ecological niche modelling, and fossil evidence have become vital tools for reconstructing demographic histories (e.g. woolly mammoths (*Mammuthus primigenius*)(Nogués-Bravo *et al.* 2008) and lions (*Panthera leo*)(Barnett *et al.* 2014)). Indeed such studies have shown that species' patterns of genetic diversity and distribution have varied dramatically under different climatic regimes (Carstens & Richards 2007). Climatic shifts have been one of the major drivers of species' range shifts, fluctuations in abundance, species extinctions and also in the formation of genetically distinct populations (Hewitt 1996). As climate change and habitat degradation potentially take us into the sixth mass

[1] This chapter is presented in the form that it was published: Younger JL, Clucas GV, Kooyman G, Wienecke B, Rogers AD, Trathan PN, Hart T, Miller KJ. Too much of a good thing: sea ice extent may have forced emperor penguins into refugia during the last glacial maximum. *Glob Change Biol.* 2015; 21(6):2215-26 (appended).

extinction (Barnosky *et al.* 2011), it is critical that we understand how species have coped with change in the past to be able to assess their likely responses and resilience to future climate change (Hoelzel 2010).

Emperor penguins (*Aptenodytes forsteri*) are an iconic Antarctic species whose population genetic structure has not been studied to date. We know little about dispersal among colonies or how historical climate change may have affected their range and abundance. Thus, we have limited capacity to predict how these birds may fare in the future. Projections for continent-wide declines of emperor penguins have been made based on the demographic responses of the Pointe Géologie colony to changes in sea ice conditions (Ainley *et al.* 2010; Barbraud & Weimerskirch 2001; Jenouvrier *et al.* 2009; Jenouvrier *et al.* 2012; Jenouvrier *et al.* 2014). However, decadal monitoring data is only available for this single site out of 46 known emperor penguin colonies; as such, the climate change responses of emperor penguins across their entire distribution and over millennial timescales are currently unknown (Ainley *et al.* 2010).

Emperor penguins are highly reliant on sea ice throughout most of their breeding cycle, and mating and incubation takes place on land-fast sea ice in most of the known colonies (Fretwell *et al.* 2012; Fretwell *et al.* 2014). During the breeding season, emperor penguins feed on prey that is also sea ice dependent (Gales *et al.* 1990). Significant areas of open water exist year-round within the Antarctic sea ice zone in the form of leads and polynyas (Zwally *et al.* (1985) and references therein). These areas are often important in providing emperor penguins access to their underwater foraging habitat when the fast ice extends far from their colonies (Dewasmes *et al.* 1980). Polynya formation is driven by either upwelling of Circumpolar Deep Water or by the outflow of katabatic winds that push sea ice away from the coastline (Martin 2001). Polynyas are associated with enhanced primary production, as the reduction in sea ice volume facilitates an earlier spring melting of sea ice and a coincident earlier start in photosynthetic primary productivity (Martin 2001). Some polynyas are permanent features of the sea ice zone and create areas of hyper-productivity, such as the Ross Sea polynya (Smith & Gordon 1997), whilst most are smaller and ephemeral features depending on wind stresses and currents.

Changes in the extent and duration of sea ice around Antarctica show highly regionalized trends with some areas increasing or remaining stable while others are decreasing (Vaughan *et al.* 2013; Zwally *et al.* 2002); this has an effect on the population dynamics of emperor penguins as both positive and negative sea ice anomalies can result in negative population growth rates at the local scale (Ainley *et al.* 2010; Barbraud *et al.* 2011; Jenouvrier *et al.* 2014; Massom *et al.* 2009). Despite uncertainties over the rate and extent of ice loss that will occur around Antarctica, all climate models project a reduction in the extent and duration of Antarctic sea ice by the end of

the century (Collins *et al.* 2013). As sea ice declines we might expect emperor penguins to be disadvantaged by a lack of breeding habitat (Jenouvrier *et al.* 2014), unless they have the capacity to alter their preferred choice of breeding site or their current range by colonising new areas. Recent studies have shown more plasticity than expected in the locations of breeding colonies; satellite imagery suggests that colonies where the fast ice is inadequate at the onset of the breeding season relocated or partially relocated onto ice shelves or ice-bergs (Fretwell *et al.* 2014; LaRue *et al.* 2015). However, as sea ice declines, emperor penguins may also have to contend with altered prey availability and face new threats from predators, as changing conditions differentially affect species at other trophic levels (Trathan *et al.* 2011).

During the last glacial maximum (LGM, 19.5 – 16 kya), the winter sea ice extent was approximately double the present day values, and seasonal variation in sea ice extent is thought to have been greater (Gersonde *et al.* 2005). It is unclear how this would have affected emperor penguins. Thatje *et al.* (2008) suggested that they may have migrated with the sea ice to lower latitudes, staying within energetic migration thresholds of the ice edge, and could have maintained breeding populations around Antarctica by foraging in the marginal ice zone at the sea ice edge. Alternatively, they could have remained associated with polynyas. Sediment cores suggest the existence of LGM polynyas in several locations, including the north-western Ross Sea, the south-eastern Weddell Sea off Dronning Maud Land, and the north-western Weddell Sea (Brambati *et al.* 2002; Mackensen *et al.* 1994; Smith *et al.* 2010; Thatje *et al.* 2008). In either case, reductions in overall primary productivity within what is today's seasonal sea ice zone (Domack *et al.* 1998; Kohfeld *et al.* 2005) would likely have been detrimental to emperor penguin populations (Ainley *et al.* 2010).

Little is known about the level of natal philopatry or migration among emperor penguin colonies. Understanding philopatry is particularly important in light of population models that suggest that emperor penguins may be declining as a result of local climatic shifts (Jenouvrier *et al.* 2009; Jenouvrier *et al.* 2014). High emigration rates are conceivable amongst emperor penguin colonies; satellite tracking has shown that they travel thousands of kilometres on their juvenile journeys, often passing other colonies (Kooyman *et al.* 1996; Thiebot *et al.* 2013; Wienecke *et al.* 2010). Generally, philopatry is high amongst penguins (Dehnhard *et al.* 2014; Saraux *et al.* 2011), but population structure is absent in many species (e.g. Chinstrap penguins (*Pygoscelis antarctica*)(Clucas *et al.* 2014)) as even low levels of migration can be sufficient to homogenise populations (Hartl & Clark 1997).

We analysed the population structure among eight extant emperor penguin colonies (Figure 2.1) using mitochondrial DNA sequences, and inferred population trajectories during and since the

LGM using a combination of ancient and modern DNA sequences in a Bayesian coalescent framework (Drummond *et al.* 2005). This method reconstructs past changes in abundance by estimating the genealogy from sequence data, and co-estimating the effective population size at different points in time, where the effective population size is the number of individuals that contribute offspring to the descendant generation (Pybus *et al.* 2000). We aimed to: 1) investigate how emperor penguin populations were affected by sea ice conditions during and following the LGM; and 2) to test the hypothesis that emperor penguins comprise one panmictic population as a result of the high dispersal of individuals after fledging, and the lack of obvious ecological barriers to dispersal around the Antarctic coastline.

2.3 Materials and methods

2.3.1 Sample collection

Skin tissue of dead emperor penguins was collected from Halley Bay (see Figure 2.1 for all sample locations) in November 2012 and transported frozen to the UK, where it was transferred to 90% ethanol and stored at -20°C. Blood samples were collected from Gould Bay in December 2013 and transported to the UK at ambient temperature in RNAlater (Life Technologies), and then stored at -20°C. Shed feathers were collected from the Ross Sea between 2010 and 2012, and were transported and stored at -20°C. Shed feathers were collected at least 10 meters apart to minimize sampling the same bird. Pectoral muscle biopsies were collected from dead chicks at Fold Island in September 2010, from Pointe Géologie in December 2010 and from Amanda Bay in December 2012 and 2013. Biopsies were immediately placed in 90% ethanol and stored at -20°C. Whole dead chicks were collected from Auster in September and October in 1993 and 1994 and transported and stored at -20°C. Bones from the subfossil remains of three penguins were collected at Club Lake in January 2013 and stored at -80°C. Club Lake is an ice-free area in the Vestfold Hills which is currently unoccupied by penguins. The nearest extant colony is Amanda Bay, 95 km away.

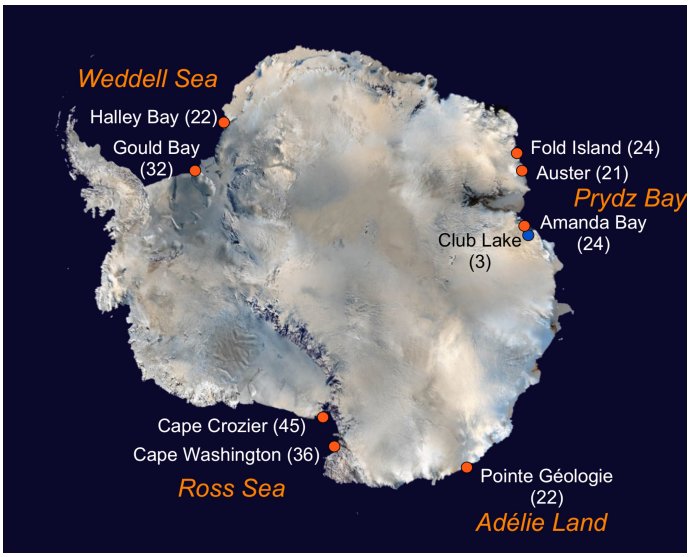


Figure 2.1. Sample locations of emperor penguin colonies. The number of sequences obtained from each location is shown in brackets. Red points indicate the origin of modern samples, and the blue point indicates the origin of the subfossil samples.

Where blood samples were taken, one handler seized the upper body with both hands and restrained the flippers, with the bird's head placed under the arm of the handler to prevent biting and minimize stress (Le Maho *et al.* 1992). The second handler took blood from the brachial vein using a 25G needle and 1 mL syringe. Total restraint time was generally two to three, but occasionally four, minutes. The bird was then released at the edge of the colony. Sampling was conducted under permits from the UK Foreign and Commonwealth Office, the US National Science Foundation and the Australian Antarctic Division. Each of these permits was issued following independent ethical review of the sampling. All sampling was carried out in accordance with UK Home Office guidelines and also received ethical approval from the University of Oxford, British Antarctic Survey and Australian Antarctic Division. The radiocarbon ages, expressed here as years BP (i.e. before 1950), of the Club Lake remains were determined using accelerated mass spectrometry by GNS Science Rafter Radiocarbon National Isotope Centre, New Zealand. The apparent ages were corrected for the marine-carbon reservoir effect (Gordon & Harkness 1992) using the calibration program Calib7.0 (St Ui & Reimer 1993).

2.3.2 DNA extraction, amplification and sequencing

Genomic DNA (gDNA) was extracted from modern samples with the QIAGEN DNeasy Blood and Tissue Kit. The manufacturer's protocols for blood and tissue samples were followed with the following modifications to the digestion step: for blood samples 30 µL of proteinase-K was used and the digestion time was 3 h; for tissue samples 40 µL proteinase-K and an additional 10 µL 1 M

dithiothreitol (skin samples only) was used with an incubation time of 32 h. All samples were treated with either 1 µL RNase A (QIAGEN) or 1 µL Riboshredder (Epicentre) according to the manufacturers' instructions. DNA was eluted in 100 µL of elution buffer following an incubation of 5 – 20 min. For subfossil samples ~50 mg of bone was decalcified in 0.5 M EDTA/0.001% Triton X100 at 56°C for 48 h and then extracted using a standard phenol chloroform protocol with ethanol precipitation and a final elution volume of 30 µL. The subfossil samples were extracted in a physically isolated laboratory which had not been used previously for avian samples to minimise the risk of contamination. The mitochondrial hypervariable region (HVR) and cytochrome B (CytB) were sequenced in all modern and ancient DNA samples. HVR is a rapidly evolving region of the mitochondrial genome, and so is suitable for investigations of recent demographic history, whilst CytB is a conserved gene and can hence give information about longer-term demographic history (Baker & Marshall 1997). HVR was amplified in all modern samples using primers F-0225 and R-INR (all primer sequences can be found in Supplementary Table 2.3). The reaction mix consisted of 7.5 µL of PCR Master Mix (QIAGEN), 0.2 µM of each primer, and 5 -10 ng gDNA, made up to 15 µL with ddH2O. Thermocycling conditions were: 94°C for 3 min; 35 cycles of 94°C for 30 s, 59.5°C for 45 s, 72°C for 1 min; followed by an extension period of 72°C for 10 min. Occasionally, double bands were apparent when the PCR product was visualized by electrophoresis. For these individuals, the shorter 755 bp band was extracted from the gel and purified using QIAGEN or Promega gel extraction kits following the manufacturer's instructions. For Fold Island, Amanda Bay, Auster and Pointe Géologie colonies, CytB was amplified using primers B1 (Baker *et al.* 2006; Kocher *et al.* 1989) and B6 (Baker *et al.* 2006) with a reaction mix consisting of 7.5 µL of GoTaq Green Master Mix (Promega), 0.2 µM of each primer, and 5-10 ng gDNA, made up to 15 µL with ddH2O. Thermocycling conditions were: 95°C for 1 min; 35 cycles of 95°C for 20 s, 52°C for 40 s, 72°C for 50 s; then 72°C for 5 min. For the Cape Washington, Cape Crozier, Gould Bay and Halley Bay samples, primers CytB-F1 and CytB-R1 were used with a reaction mix consisting of 7.5 µL of PCR Master Mix (QIAGEN), 0.2 µM of each primer, and 5 -10 ng gDNA, made up to 15 µL with ddH2O. Thermocycling conditions were: 94°C for 3 mins; 35 cycles of 94°C for 45 s, 60°C for 45 s, 72°C for 1 min; then 72°C for 10 min. For the subfossil samples we designed novel, species-specific primers (Supplementary Table 2.3) to amplify short (<150bp) overlapping fragments in order to improve the success rate of amplification from degraded DNA. The reaction mix consisted of 7.5 µL of AmpliTaq Gold 360 Master Mix (Life Technologies), 0.2 µM of each primer, and 25-50 ng gDNA, made up to 15 µL with ddH2O. Thermocycling conditions were: 95°C for 10 min; 42 cycles of 95°C for 20 s, Tm (primer) for 20 s, 72°C for 20 s; 72°C for 5 min. PCR products for Fold Island, Amanda Bay, Auster, Pointe Géologie and the subfossil samples were bi-directionally sequenced by the Australian Genome Research Facility (AGRF) via the Sanger sequencing method using the PCR primer pairs. PCR products for Gould Bay, Halley Bay, Cape Washington and Cape Crozier

were sequenced using the Sanger method by Macrogen Europe. The reverse primer for the HVR and the forward primer for CytB were used to sequence each product twice, as these were found to work best in the sequencing reaction. Geneious v5.5.9 was used for alignment. A high number of heteroplasmic sites were found in the HVR (Supplementary Figure 2.5) as has been observed in Adélie penguins (Millar *et al*. 2008) and these were re-scored manually according to IUPAC ambiguity codes. No heteroplasmic sites were recorded in the CytB sequences.

2.3.3 Data analysis – summary statistics and population structure

Arlequin v3.5 (Excoffier & Lischer 2010) was used to calculate summary statistics for HVR, CytB, and concatenated HVR and CytB. jModeltest (Posada 2009; Posada & Buckley 2004) was used to estimate the best substitution model for each dataset, and then the following corrections for calculating genetic distances were implemented in Arlequin: HVR - Tamura correction with a gamma distribution for rate heterogeneity with $\alpha = 0.016$; CytB – Tamura correction; concatenated – Tamura correction with a gamma distribution for rate heterogeneity with $\alpha = 0.109$ (Tamura 1992). Arlequin was also used to calculate pairwise genetic distances (θ_{ST}) between colonies and perform analyses of molecular variation (AMOVA) on the concatenated sequences with the Tamura & Nei correction. Network v4.612 (Fluxus Technology Ltd.) was used to draw haplotype networks. In these networks, sites with ambiguity codes were converted to 'N' before the network was calculated.

2.3.4 Data analysis – demographic histories

Bayesian phylogenetic analyses and demographic reconstructions were performed using BEAST v1.8 (Drummond *et al.* 2012). The dataset was partitioned into HVR and CytB, with a nucleotide substitution model of HKY (Hasegawa *et al.* 1985) with four gamma categories for HVR and TN93 (Tamura & Nei 1993) for CytB. We set the option to use ambiguous states to "true", such that sites with an ambiguity code, such as an 'R', were treated as an 'A' or a 'G', rather than a missing site. We used the coalescent Extended Bayesian Skyline Plot tree prior (Heled & Drummond 2008) with a strict molecular clock. For molecular clock calibration, the HVR substitution rate prior was specified as a normal distribution around a mean value of 0.55 substitutions/site/Myr (SD = 0.15), to reflect the substitution rate of the HVR in Adélie penguins (*Pygoscelis adeliae*) (Millar *et al.* 2008). In the absence of a published substitution rate for CytB in penguins we used a uniform prior of $5x10^{-4}$ to $5x10^{-1}$ substitutions/site/Myr with a starting value of $2x10^{-2}$ (Weir & Schluter 2008). The corrected radiocarbon ages of the Club Lake samples were input as tip dates, for additional calibration of the molecular clock. Based on these initial priors, substitution rates for our dataset were estimated during the analysis. The posterior distributions of substitution rates,

phylogenetic trees and effective population size through time were generated using the Markov chain Monte Carlo (MCMC) sampling procedure, implemented in BEAST, which was run for 120 million generations with samples drawn every 6000 steps and the first 10% discarded as burn-in. Tracer v.1.5 was used to check effective sample size (ESS) values to confirm convergence with all values >200. Three independent BEAST analyses were performed to ensure reproducibility of the posterior distribution. The population size parameter of the demographic model (Ne*tau) was converted to N_{ef} by dividing the parameter by 14 years, which is the estimated generation length of emperor penguins (Forcada & Trathan 2009; Jenouvrier *et al.* 2005). Phylogenetic trees were visualised using FigTree v1.4.

2.4 Results

2.4.1 Present day population structure

We sequenced 226 individuals from eight colonies (Figure 2.1) plus three subfossil birds whose ages ranged from 643 – 881 years BP (after correction for marine reservoir effect). We sequenced 629 bp of the mitochondrial hypervariable region (HVR) and 867 bp of cytochrome b (CytB) from each individual (GenBank accession numbers KP644787 - KP645015 and KP640645 - KP640873, respectively). Genetic diversity was extremely high for the HVR, with 220 haplotypes recorded out of the 229 individuals sequenced; the mean number of pairwise differences between haplotypes was 20.62 ± 9.14 (Table 2.1). Genetic diversity was much lower for CytB, with just 59 unique haplotypes recorded.

Our results show a high level of gene flow among all the EAWS colonies (East Antarctica including Adélie Land, and the Weddell Sea) and between the two Ross Sea colonies (Table 2.2), but little exchange between the EAWS and Ross Sea colonies (pairwise θ_{ST} values range from 0.213 to 0.617, Table 2.2). When colonies are grouped into two populations (Ross Sea and EAWS), a high proportion (17.7%) of the genetic variation is explained by the difference between the groups, and there is strong and significant genetic differentiation between them (AMOVA, $F_{ST} = 0.196$, $p < 0.001$). This pattern is also evident from haplotype networks (Supplementary Figure 2.6and Supplementary Figure 2.7), which show that Ross Sea individuals tend to be closely related, whilst sequences from EAWS colonies tend to cluster independently from the Ross Sea haplotypes. However, some Ross Sea sequences are found across the network, and vice versa. This could indicate low-level gene flow between the Ross Sea and EAWS.

Table 2.1. Summary statistics by geographic and genetic region. n = number of individuals; N_H = unique haplotypes; N_P = polymorphic loci; H = haplotype diversity; d_x = mean number of pairwise differences between sequences within geographic region; significance is indicated for Tajima's D and Fu's F_S test statistic where * denotes $p < 0.05$, ** denotes $p < 0.01$, *** denotes $p < 0.001$.

Geographic Region	Genetic Region	n	N_H	N_P	H	d_x	Tajima's D	Fu's F_S
All sequences	HVR + CytB	229	222	205	0.999 ± 0.000	23.12 ± 10.21	-1.03	-23.63**
Ross Sea	HVR + CytB	81	80	124	0.999 ± 0.002	18.49 ± 8.28	-0.903	-22.84***
East Antarctic & Weddell Sea	HVR + CytB	148	145	171	0.999 ± 0.001	22.15 ± 9.81	-0.930	-23.80**
All sequences	HVR	229	220	164	0.999 ± 0.001	20.62 ± 9.14	-0.835	-23.67**
Ross Sea	HVR	81	76	109	0.997 ± 0.003	16.81 ± 7.56	-0.836	-17.21*
East Antarctic & Weddell Sea	HVR	148	144	138	0.999 ± 0.001	19.58 ± 8.71	-0.758	-23.85**
All sequences	CytB	229	59	41	0.864 ± 0.016	2.94 ± 1.54	-1.651*	-21.79***
Ross Sea	CytB	81	26	15	0.876 ± 0.028	1.99 ± 1.14	-0.979	-3.43
East Antarctic & Weddell Sea	CytB	148	41	33	0.797 ± 0.031	2.96 ± 1.56	-1.482*	-12.92***

Table 2.2. Pairwise genetic differentiation between colonies. Pairwise θ_{ST}'s are presented below the diagonal, and associated p-values above the diagonal. Significance is indicated by bold text, where * denotes $p < 0.05$, ** denotes $p < 0.01$, *** denotes $p < 0.001$.

	Gould Bay	Halley Bay	Fold Island	Auster	Amanda Bay	Pointe Géologie	Cape Washington	Cape Crozier
Gould Bay		0.596	0.731	0.560	0.186	0.129	**0.000**	**0.006**
Halley Bay	-0.027		0.566	0.798	0.708	0.301	**0.002**	**0.008**
Fold Island	-0.050	-0.032		0.515	0.323	0.462	**0.000**	**0.007**
Auster	-0.026	-0.086	-0.027		0.595	0.797	**0.000**	**0.000**
Amanda Bay	0.055	-0.058	0.014	-0.038		0.576	**0.000**	**0.000**
Pointe Géologie	0.091	0.029	-0.012	-0.085	-0.033		**0.000**	**0.000**
Cape Washington	**0.355*** **	**0.440** **	**0.468*** **	**0.567*** **	**0.617*** **	**0.596*** **		0.509
Cape Crozier	**0.213** **	**0.266** **	**0.256** **	**0.432*** **	**0.447*** **	**0.428*** **	-0.011	

2.4.2 Population history with respect to climate change

There is evidence of past population expansion in emperor penguins across Antarctica as indicated by our Extended Bayesian Skyline Plots (EBSPs) (Figure 2.2). An almost nine-fold increase in abundance of the EAWS population commenced approximately 12 kya. The Ross Sea population expanded three-fold from approximately 9.5 kya. Superimposing expansion signals over the estimated temperature derived from ice cores (Figure 2.2c), it is clear that population expansion followed the end of the LGM. Tajima's *D* and Fu's F_S statistics provide further support for an expansion of both populations (Table 2.1).

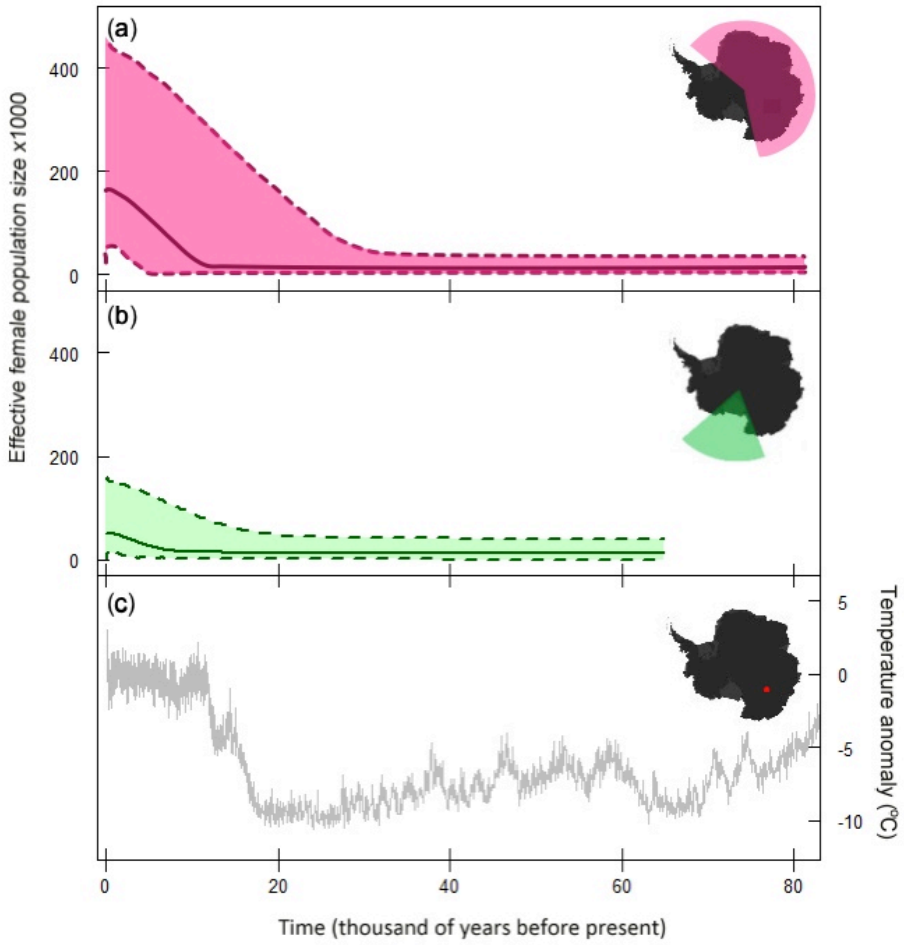


Figure 2.2. Extended Bayesian skyline plots showing the change in effective female population size (N_{ef}). Solid lines show the median estimate; dotted lines show the 95% highest posterior density interval. a) EAWS colonies; b) Ross Sea colonies; c) the Antarctic temperature anomaly (the difference from the average of the last 1000 years) as estimated from the EPICA Dome C ice core (Jouzel & Masson-Delmotte 2007), with the ice core location indicated in red.

Our phylogenetic analyses indicate three highly supported clades (Figure 2.3), which diverged during the Late Pleistocene (97 kya, 95% HPD: 50−154 kya). One of these clades is comprised predominantly of Ross Sea penguins, whereas the other two are dominated by EAWS individuals. There was no evidence that heteroplasmy was limited to a single clade or colony, as it was widespread in all three clades and all colonies (Supplementary Figure 2.8).

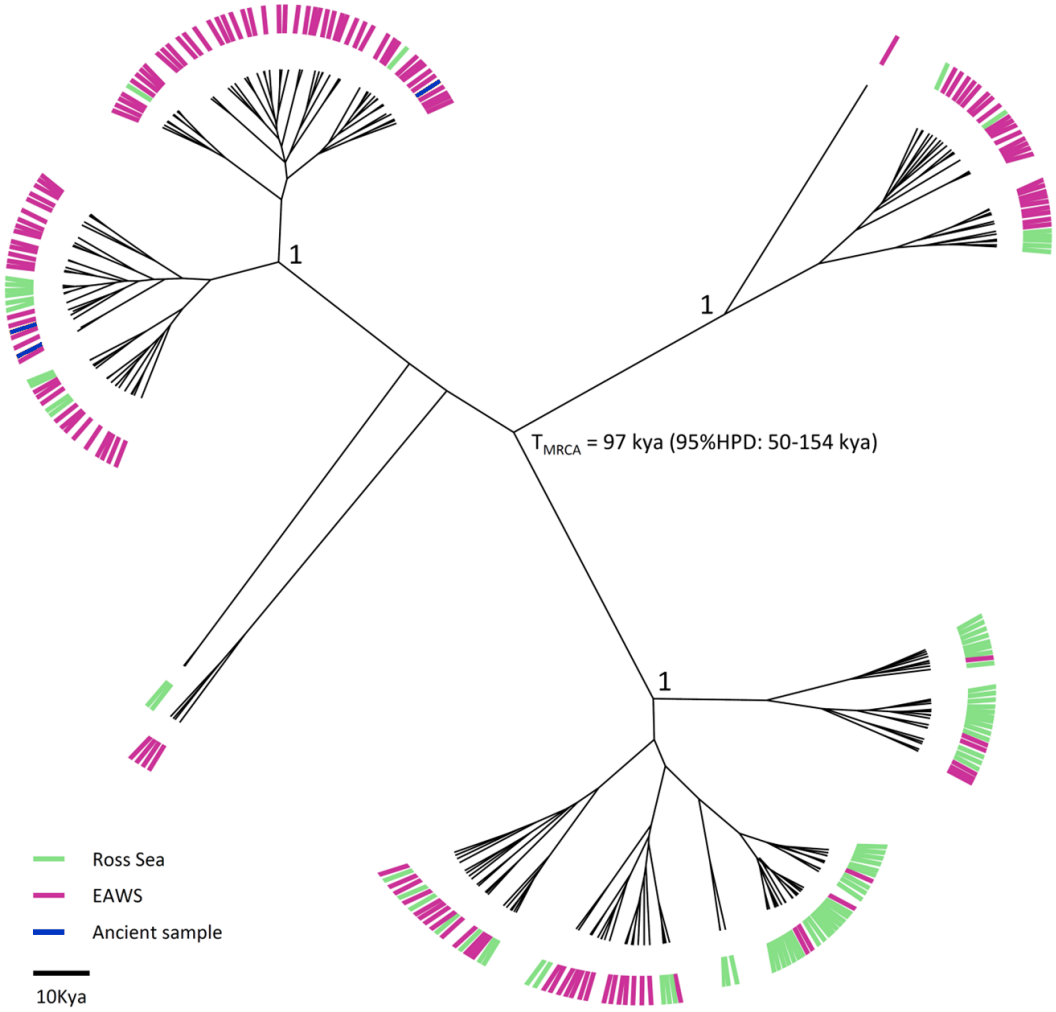


Figure 2.3. Phylogenetic relationships among individuals. Magenta - EAWS individuals; green - Ross Sea individuals; blue – subfossil individuals. The posterior probabilities are shown for the major, strongly supported clades.

2.5 Discussion

This first analysis of emperor penguin population structure shows colonies within the Ross Sea are genetically distinct from other Antarctic colonies, whereas those from the rest of the continent and spanning up to 8000 km of coastline are panmictic (Table 2.1). The admixture of the EAWS emperor penguins supports our hypothesis of limited population structure and indicates a very large dispersal range for the species. Given our genetic evidence of extensive mixing across Antarctica, the unique structure in the Ross Sea emperor penguins is surprising, and interestingly the same pattern was reported for the sympatric Adélie penguin (Ritchie *et al.* 2004), providing further evidence that the Ross Sea has a unique evolutionary history.

The existence of distinct penguin populations in the Ross Sea is puzzling. There are neither geographic nor oceanographic barriers isolating the Ross Sea from the rest of Antarctica. Furthermore, the relative distance between the Ross Sea and other colonies does not adequately explain its isolation as, for example, the Pointe Géologie colony is approximately 5600 km closer to the Ross Sea colonies than to those in the Weddell Sea (Figure 2.1). Emperor penguins are known for their extraordinary migrations; satellite tracking showed that juveniles can travel >7000 km in eight months (Thiebot *et al.* 2013). These observations support our genetic results for the EAWS region and indicate juvenile emperor penguins could comfortably traverse the 1800 km between Pointe Géologie and the Ross Sea colonies. There are also no clear habitat, environmental or foraging differences between the Ross Sea colonies and those located elsewhere (Budd 1961; Smith *et al.* 2012), except that Ross Sea colonies are located closer to the ice edge, and are therefore potentially more resilient to increases in sea ice. We suggest that the divergence of emperor penguins into two populations is historical in origin.

There are three ancestral lineages within modern emperor penguins, providing evidence that populations were isolated in the past (Figure 2.3) and diverged through microevolutionary processes, such as selection or genetic drift, which occur more rapidly in small, isolated populations (Hewitt 2000). One of these lineages is mostly limited to the Ross Sea, indicating that the isolation of this region has persisted through time. Indeed, emperor penguins occupying the Ross Sea may have become so differentiated that interbreeding with the EAWS penguins occurs at very low rates, perhaps because of genetic, behavioural (Templeton 1981) or cultural incompatibilities, such as the timing of breeding or the development of regional dialects (de Dinechin *et al.* 2012; Jouventin & Aubin 2002; MacDougall-Shackleton & MacDougall-Shackleton 2001).

Emperor penguins use complex display calls to recognise their mates and offspring (Robisson *et al.* 1993). Vocalisation is known to be an important part of the courtship process for most

penguins (Richdale 1945; Waas *et al.* 2000). Interestingly, royal penguins (*Eudyptes schlegeli*) respond more strongly to calls from their own colony members than to calls originating from different colonies, suggesting differences in dialect (Waas *et al.* 2000). Differences in vocalisations have also been found among gentoo penguin (*Pygoscelis papua*) populations (de Dinechin *et al.* 2012). If dialects become too different, then courtship may be inhibited, thereby limiting interbreeding. This has been observed in passerine birds, in which genetically distinct groups have unique mating songs (MacDougall-Shackleton & MacDougall-Shackleton 2001). Emperor penguin vocalisation patterns have only been recorded at Pointe Géologie (Robisson *et al.* 1993), but our hypothesis could be explored in the future by comparing vocalisations of emperor penguins from the Ross Sea with those of other colonies.

Although the isolation and differentiation of the Ross Sea emperor penguins has persisted, the other two historical lineages show no geographic bias and have now mixed to form one EAWS population. Incomplete mixing of ancestral lineages is typical of species that have survived the Pleistocene ice-ages in multiple refugia (Hewitt 1996). Our EBSPs indicate that both the EAWS and Ross Sea populations had reduced effective population sizes during the LGM (Figure 2.2). Thus, contrary to a hypothesis that emperor penguins would benefit from glaciation as a result of reduced competition with other predators (Thatje *et al.* 2008), it seems that they, like other Antarctic and sub-Antarctic penguin species (Clucas *et al.* 2014; Ritchie *et al.* 2004; Trucchi *et al.* 2014), were adversely affected by the LGM.

We propose that both the reduced abundance and divergence into three lineages were linked to breeding and foraging habitat availability. Today emperor penguins have a circumpolar distribution with suitable habitat spanning the entire continent (Fretwell & Trathan 2009). However, Antarctica during the LGM looked very different than the continent we know today (Figure 2.4). Most of the continental shelf was covered by ice as a result of both the extension of ice-sheets and thick, perennial sea ice, which reduced productivity south of the modern-day Polar Front drastically (Anderson *et al.* 2002; Anderson *et al.* 2009; Domack *et al.* 1998; Gersonde *et al.* 2005; Kohfeld *et al.* 2005)(Samuel Jaccard *pers. comm*.). We suggest that the increased sea ice extent would have severely restricted the foraging habitat available for emperor penguins and, coupled with lower primary production, could have resulted in a scarcity of prey resources. Additionally, air temperatures were approximately 13°C colder than the present day (Jouzel *et al.* 2007), which may have been near the penguins' lower limit of temperature tolerance (Le Maho *et al.* 1976), potentially impacting both breeding success and adult survival.

The extent and duration of sea ice are important factors in the breeding success of emperor penguins (Massom *et al.* 2009). Emperor penguins require stable fast ice to breed, but they have

to traverse the sea ice to establish colonies in autumn and to forage in winter and spring. The distances between the colonies and potential foraging areas can influence breeding success where the fast ice extent is variable (Massom *et al.* 2009), but not in locations where the extent is relatively stable (Robertson *et al.* 2014). We therefore expect that if the winter sea ice extent was substantially greater in the LGM, or if the timing of sea ice retreat was altered, that this would have made some of the extant colony locations energetically untenable during the LGM.

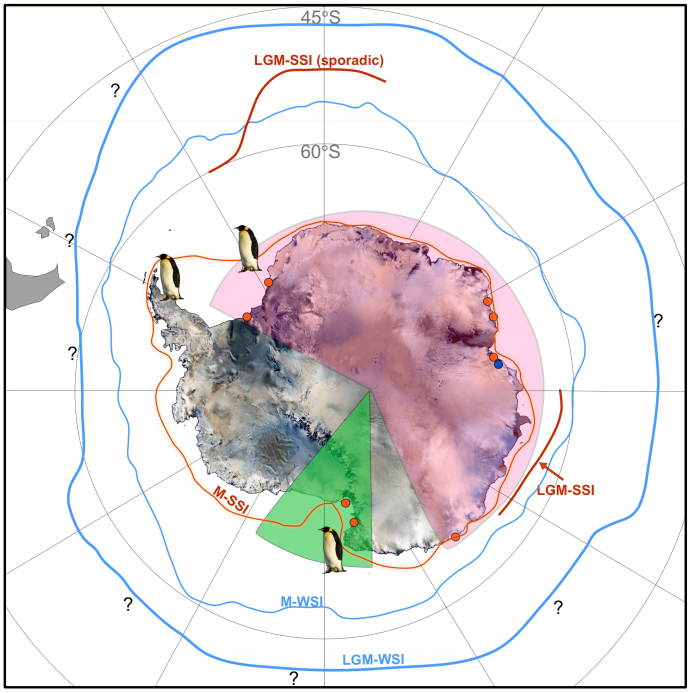


Figure 2.4. Schematic of contemporary population structure and reconstruction of historical conditions. Sampled colonies are indicated by dots, as in Figure 1. The magenta and green shading indicates population structure as estimated from this study. Lines represent the sea ice edge, as in Gersonde *et al.* (2005). M-SSI = modern summer sea ice edge; LGM-SSI = LGM summer sea ice edge; M-WSI = modern winter sea ice edge; LGM-WSI = LGM winter sea ice edge, ? = insufficient data to reconstruct the sea ice edge. Penguins represent hypothesised locations of polynya refugia. Emperor penguin picture: © Samuel Blanc / www.sblanc.com.

During the LGM, the summer sea ice extent was similar to that which we observe today, whereas the winter extent was roughly doubled (Gersonde *et al.* 2005). Colonies may have been located close to the continent so that the ice remained stable throughout the breeding season, but this would have required adults to walk immense distances to reach foraging areas during winter and spring while provisioning the chick. In that case, the chicks would receive fewer meals and be less likely to survive. The present distribution of colonies close to land (Fretwell & Trathan 2009) suggests that fast ice proximate to land provides a more stable platform than near the fast ice

edge. Also, stable ice close to the coast occurs in predictable locations that might be important for colony establishment and cohesion. Colonies further away from the coast may therefore be difficult to maintain. Our discovery of three distinct lineages provides evidence against a straightforward, latitudinal range shift in line with the sea ice edge, and suggests that emperor penguins may have survived the LGM in three suitably situated, geographically isolated refugia.

Emperor penguin refugia during the LGM may have been linked to the presence of polynyas. Several extant emperor penguin colonies are located near polynyas, which may be utilised for foraging during the winter (Croxall *et al.* 2002). Polynyas acted as "hot spots" of primary productivity during the LGM, supporting marine life and flying seabirds (Thatje *et al.* 2008). Sediment cores in the north-western Ross Sea indicate open water polynya conditions throughout the LGM (Brambati *et al.* 2002; Thatje *et al.* 2008) and this polynya could have sustained a refuge population until the Ross Sea began to clear of ice (Figure 2.4). By 9.6 kya most of the northern Ross Sea was open water (Licht & Andrews 2002). The retreating sea ice and increased upwelling during deglaciation increased productivity in the Ross Sea (Anderson *et al.* 2002) and likely increased the foraging habitat and prey availability to emperor penguins, therefore we hypothesise that these factors drove the three-fold expansion of emperor penguins in this region around this time (Figure 2.2). The LGM polynya may have also supported Adélie penguins, accounting for the existence of a distinct Ross Sea clade as previously observed for this species (Ritchie *et al.* 2004).

Another polynya was located in the south-eastern Weddell Sea off Dronning Maud Land (DML) (Mackensen *et al.* 1994; Thatje *et al.* 2008)) (Figure 2.4). Colonies of snow petrels (*Pagrodroma nivea*) were present in DML throughout the LGM, associated with this polynya (Wand & Hermichen 2005), and it may have also provided a refuge for emperor penguins. There is evidence from sediment cores for a third LGM polynya, located in the north-western Weddell Sea (Smith *et al.* 2010) (Figure 2.4); this would be consistent with our third emperor penguin refuge, given that the refuge is likely to be more proximate to DML than the Ross Sea, since the two refugial lineages mixed post-glacially while the Ross Sea lineage remained distinct.

We propose that two refuge populations that were isolated in the Weddell Sea expanded their range into Prydz Bay and Adélie Land and merged during the retreat of the East Antarctic ice sheet 14 – 7 kya (Mackintosh *et al.* 2011). At this time, the onset of more favourable environmental conditions could have resulted in the dramatic, nine-fold increase in abundance shown here (Figure 2.2). A seasonal sea ice cycle was established in Prydz Bay approximately 10.4 kya (Barbara *et al.* 2010), opening up foraging habitat and coinciding with high levels of primary productivity (e.g. (Anderson *et al.* 2009; Sedwick *et al.* 2001)). In Adélie Land, primary productivity

and the duration of the ice-free season increased from 9 kya (Denis *et al.* 2009a; Denis *et al.* 2009b). This new habitat could have facilitated the range expansion of the EAWS lineages.

It should be noted that the timing of the abundance increase of emperor penguins does not coincide exactly with the end of the LGM (Figure 2.2). We hypothesise that it is not the temperature change itself, but rather the subsequent change in sea ice conditions and primary productivity that are most likely to affect emperor penguins. Indeed, it has been proposed that there is an optimal level of sea ice at the large temporal/spatial scale for emperor penguins, which roughly corresponds to current conditions (Ainley *et al.* 2010). Therefore, the greater sea ice extent of the LGM was most likely sub-optimal for emperor penguin populations. The end of the LGM is measured when temperatures began to increase (19 – 16 kya). However deglaciation, during which ice-sheets and sea ice retreated and primary productivity increased, occurred slowly over an extended time period (ca. 17 – 11 kya)(Anderson *et al.* 2009). These events occurred later in the Ross Sea than in East Antarctica, and our results support the hypothesis that ice-sheet and sea ice retreat and increasing primary productivity were the main factors controlling emperor penguin abundance, as the Ross Sea emperor penguin population expanded later than the EAWS population (Figure 2.2). Furthermore, emperor penguins produce only one chick per year and take approximately five years to reach sexual maturity (Jenouvrier *et al.* 2005), so any abundance increase would be initially slow.

Our hypothesis of three refugial populations of emperor penguins during the LGM could be tested using a higher density of genetic markers. This would allow for the investigation of clinal variation in genetic diversity arising from founder effects as new areas were colonized following the expansion from refugia after the LGM (Hewitt 1996). It should be noted that our present study is based on mitochondrial DNA and therefore represents dispersal patterns of females only, but nonetheless supports a plausible explanation for past and present microevolutionary processes in emperor penguins. The next step should be to verify these findings using nuclear markers to account for male-mediated gene flow.

In this continent-wide study of microevolution in an Antarctic penguin we suggest that past climatic changes have greatly impacted emperor penguin populations. As conditions became more favourable after the LGM, their global population expanded and the populations from the Weddell Sea and East Antarctic intermixed to form one large, panmictic population. Interestingly, the isolation of the Ross Sea emperor penguins has persisted until today. The reasons for this isolation remain unknown, but we suggest that separate management plans are required for the Ross Sea and EAWS populations. By conserving the full spectrum of genetic variation and, in

particular, all phylogeographic lineages, the evolutionary potential of the species can be maximised (D'Amen *et al.* 2013).

Our study suggests that emperor penguins have shown important historic responses to past climate shifts and their population increase post-LGM was remarkable. However, the projected rate of temperature increase over the next century is an order of magnitude greater than that following the LGM (Collins *et al.* 2013; Masson-Delmotte *et al.* 2013; Shakun *et al.* 2012). At present, emperor penguins become heat stressed around 0°C, so may exist near the upper limits of their physiological tolerance (Wienecke, pers obs). Whether the resilience demonstrated in the past of this highly cold-adapted species will enable it to adapt to projected climate change remains to be seen, as rising temperatures will alter its breeding grounds and foraging space more rapidly than in the past.

2.6 Supplementary Information

Figure 2.5 Chromatograms showing heteroplasmic sites in emperor penguin HVR sequences. Heteroplasmic sites are indicated by the orange arrows; a) an individual with one heteroplasmic site, b) an individual with eight heteroplasmic sites.

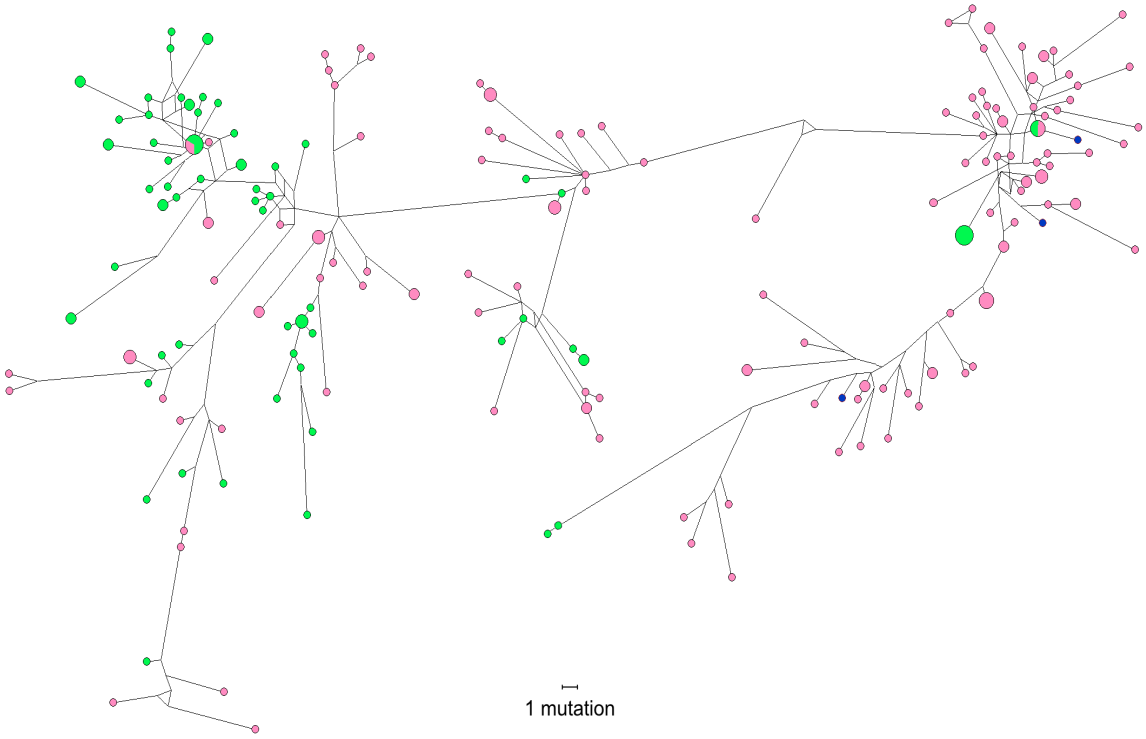


Figure 2.6. Haplotype network of phylogenetic relationships among all HVR sequences. Magenta – EAWS colonies; green – Ross Sea colonies; blue – subfossil samples; the size of the circle indicates the relative frequency of the haplotype.

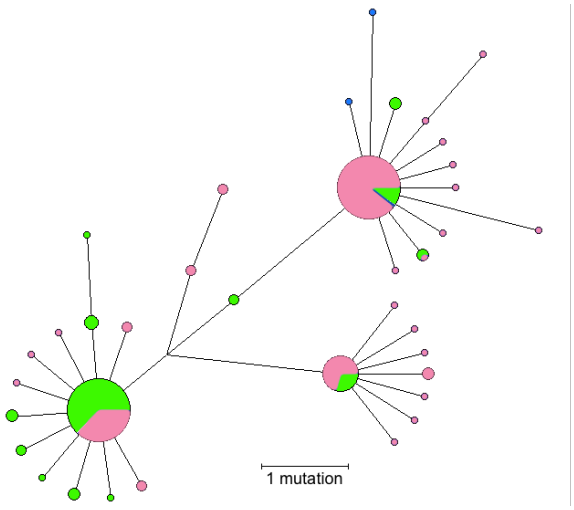


Figure 2.7. Haplotype network of phylogenetic relationships among all CytB sequences. Magenta – EAWS colonies; green – Ross Sea colonies; blue – subfossil samples; the size of the circle indicates the relative frequency of the haplotype.

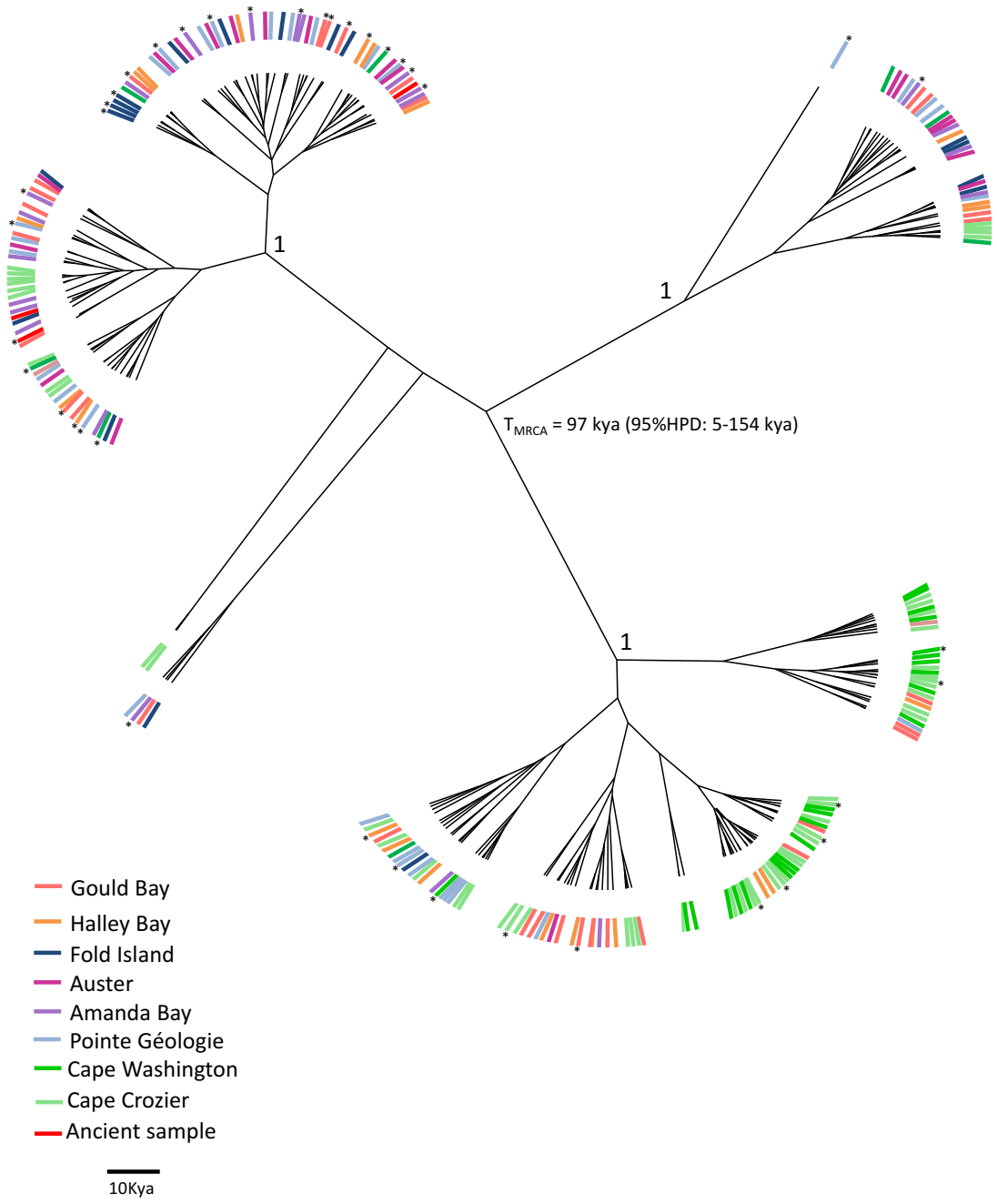


Figure 2.8. Phylogenetic relationships among individuals. The individuals are coloured according to the colony where they were sampled, and individuals which did not have any heteroplasmic sites in the HVR region are marked with asterisks.

Table 2.3. Primer Sequences.

Primer name	Primer sequence 5’ – 3’
F-0225	GGAACCTCCCAAAGAGTACCA
RINR	CCAACCAGATGTATCGGTGA
HVR-R1	TGAAAGTATTGCTTTACGTATCCTT
HVR-F2	AAGGATACGTAAAGCAATACTTTC
HVR-R2	AGGAGTAATTGTTGAGTACATGACA
HVR-F3	TCATGTACTCAACAATTACTCCTG
HVR-R3	TCACGTGAGAAGACCGACTAA
HVR-F4	ATCTCCTGAGGCGCTAGCTT
B1	CCATCCAACATCTCAGCATGATGAAA
B6	CCATCCAACATCTCAGCATGATGAAA
CytB-F1	ACTGCAGACACAACCCTAGC
CytB-R1	GGGAAGAGGATCAGGAGGGT
CytB-R1	AATGATGCTCCGTTTGCATGTAGGTT
CytB-F2	ACACATGCCGAAACGTACAG
CytB-R2	GTAGCCTACGAAGGCGGTTG
CytB-F3	GAAACCTGAAACACAGGCATT
CytB-R3	CGGGTTAATGTGGGGTTGT
CytB-F4	CTCAGCCATCCCTTACATTG
CytB-R4	TTGTGGAGTAGTAGGGGTGGA
CytB-F5	CAAATAACCCACTGGGCATC
CytB-R5	TCATTCTGGTTTGATGTGTGG
CytB-F6	CCAGCAAACCCACTAGTCAC
CytB-R6	GGGCTCAGAATAGGAGTTGG
CytB-F7	ATAGCTTTCCGCCCTCTCT

Chapter 3: Dispersal in the sub-Antarctic: King penguins show remarkably little population genetic differentiation across their range[2]

3.1 Abstract

Background: Seabirds are important components of marine ecosystems, both as predators and as indicators of ecological change, being conspicuous and sensitive to changes in prey abundance. To determine whether fluctuations in population sizes are localised or indicative of large-scale ecosystem change, we must first understand population structure and dispersal. King penguins are long-lived seabirds that occupy a niche across the sub-Antarctic zone close to the Polar Front. Colonies have very different histories of exploitation, population recovery, and expansion.

Results: We investigated the genetic population structure and patterns of colonisation of king penguins across their current range using a dataset of 5,154 putatively unlinked, high-coverage single nucleotide polymorphisms generated via restriction site associated DNA sequencing (RADSeq). Despite breeding at a small number of discrete, geographically separate sites, we find only very slight genetic differentiation among colonies separated by thousands of kilometres of open-ocean, suggesting migration among islands and archipelagos may be common. Our results show that the South Georgia population is slightly differentiated from all other colonies and suggest that the recently founded Falkland Island colony is likely to have been established by migrants from the distant Crozet Islands rather than nearby colonies on South Georgia, possibly as a result of density-dependent processes.

Conclusions: The observed subtle differentiation among king penguin colonies must be considered in future conservation planning and monitoring of the species, and demographic models that attempt to forecast extinction risk in response to large-scale climate change must take into account migration. It is possible that migration could buffer king penguins against some of the impacts of climate change where colonies appear panmictic, although it is unlikely to protect them completely given the widespread physical changes projected for their Southern Ocean foraging grounds. Overall, large-scale population genetic studies of marine predators

[2] This chapter is presented in the form that it was published: Clucas GV, Younger JL, Kao D, Rogers AD, Handley J, Miller GD, Jouventin P, Nolan P, Gharbi K, Miller KJ, Hart T. Dispersal in the sub-Antarctic: king penguins show remarkably little population genetic differentiation across their range. BMC Evol Biol. 2016; 16(1):211 (appended).

across the Southern Ocean are revealing more interconnection and migration than previously supposed.

3.2 Background

Understanding the patterns and mechanisms of population structure is essential for successful species conservation (Manel *et al.* 2003). For example, species with a high degree of population differentiation and limited dispersal among colonies may have a reduced ability to respond to unfavourable local environmental conditions (Walther *et al.* 2002) and may lose a large portion of their total genetic variation if local populations are lost or reduced (Friesen *et al.* 2007a). Accurate data regarding the geographic boundaries of breeding populations and the degree of genetic exchange among them are therefore essential for species risk assessments and conservation planning, including to mitigate the effects of climate change. However, the extent of differentiation among natural populations of seabirds is difficult to predict and has been shown to vary widely among taxa (Friesen 2015; Friesen *et al.* 2007a). In general, seabirds are philopatric, with adults returning to natal sites to breed (Coulson 2001), and this behaviour can be an isolating mechanism that acts as a barrier to gene flow. Seabirds that breed at high latitudes, such as the polar regions, or that have large foraging ranges, are thought to be the least likely to have differentiated populations as a result of recent range expansions and retained ancestral variation (Friesen *et al.* 2007a).

King penguins (*Aptenodytes patagonicus*) are thought to be vulnerable to climate change impacts via changes in their prey distribution in the future (Bost *et al.* 2015; Le Bohec *et al.* 2008) and an understanding of their population structure is required to accurately model these impacts and make inferences about observed changes in population size. King penguins congregate in large breeding colonies on coastal ice-free ground on sub-Antarctic islands between 45°S and 55°S (Bost *et al.* 2013) (Figure 3.1). Numbers have been increasing across their range over the past several decades (Bost *et al.* 2013; Delord *et al.* 2004; Rounsevell & Copson 1982; Van Den Hoff *et al.* 2009), following historic anthropogenic exploitation during the late 19th to early 20th centuries during which they were slaughtered en masse for the blubber oil industry (Conroy & White 1973). The global population of king penguins is now conservatively estimated at 1.6 million breeding pairs and still increasing (Bost *et al.* 2013).

Owing to their large and growing population size across most of their range, king penguins are currently listed as being of Least Concern on the IUCN's Red List of Threatened Species (IUCN 2016) although there have been concerns that harvest may have resulted in a population bottleneck that would have reduced genetic variation and hence their adaptive capacity.

Furthermore, king penguins will face new challenges in the coming decades as climate change alters their marine foraging habitat. The most immediate threat posed by climate change to king penguin populations is the southward shift of the Polar Front and deepening of the thermocline; both secondary to warming of the Southern Ocean's surface waters (Bost *et al.* 2015). King penguins forage almost exclusively at the Polar Front during the summer breeding season (Charrassin & Bost 2001; Scheffer *et al.* 2012; Sokolov *et al.* 2006), as a result of the combination of predictably high prey abundance and ideal diving conditions that they find at the front (Bost *et al.* 2009; Jouventin *et al.* 1994). As sea surface temperatures increase with climate change, the position of the Polar Front is shifting to the south, and this is predicted to double the king penguin's travelling distance to their preferred foraging grounds by 2100 (Péron *et al.* 2012). The coincident deepening of the thermocline means the penguins must also dive deeper to reach their prey (Bost *et al.* 2015). A study at the Crozet Islands has already demonstrated the impact that warming waters can have on king penguin numbers, with a population decline of 34% associated with reduced adult survival during an anomalously warm year in 1997 (Bost *et al.* 2015). In light of the potential threats to king penguin populations, accurate data regarding their population structure are needed (Bost *et al.* 2013). Specifically, to monitor population sizes in relation to environmental impacts we must first understand what constitutes a genetic breeding population of king penguins.

There have been no studies of genetic population structure of king penguins across their breeding range to date. A decade-long study at one colony in the Crozet Islands found that 77% of juvenile king penguins returned to their natal colony (Saraux *et al.* 2011). This suggests that the species is largely philopatric, however, even low numbers of dispersing individuals could be sufficient to homogenise populations (Wright 1969). King penguins possess a remarkable mobility, regularly conducting round-trips in excess of 3,200 km from breeding colonies to forage in Antarctic waters during the winter months (Charrassin & Bost 2001). However, the average distance between the pairs of breeding sites in our study is 6,500 km and the colonies are distributed longitudinally, whereas most of the king penguin's foraging movement is latitudinal (Baylis *et al.* 2015; Scheffer *et al.* 2012). This suggests that frequent dispersal among breeding sites should be unlikely. In spite of this, incidences of long-distance dispersal have been documented, with birds tagged on the Crozet Islands resighted resting or moulting at Marion Island (900 km away) (Gartshore *et al.* 1988), Kerguelen Island (1,500 km away), Macquarie Island (5,600 km away) (Weimerskirch *et al.* 1985) and Heard Island (1,740 km away) (Woehler 1989). It should also be noted that any genetic differentiation that arose during the founding of colonies would be expected to persist for a very long time (i.e. thousands of generations) in a species with such a large effective population size and rapid population growth rate (Boileau *et al.* 1992). Previous studies of population structure in

other penguin species revealed a remarkable lack of differentiation across thousands of kilometres, including in emperor penguins (*Aptenodytes forsteri*) (Cristofari *et al.* 2016; Younger *et al.* 2015b) and Adélie penguins (*Pygoscelis adeliae*) (Ritchie *et al.* 2004; Roeder *et al.* 2001). This is in contrast to gentoo penguins (*Pygoscelis papua*) (Levy *et al.* 2016) and chinstrap penguins (*Pygoscelis antarctica*) (Freer *et al.* 2015), which demonstrated moderate to low genetic differentiation across similar distances. Both emperor and Adélie penguins have almost continuous circumpolar distributions (Fretwell *et al.* 2012; Lynch & LaRue 2014) that may facilitate migration, whereas king penguin colonies are scattered distantly across the sub-Antarctic (Figure 3.1).

Overall, king penguins are a highly mobile marine species with huge potential for dispersal; however, genetic divergence among colonies may exist as a result of nonphysical barriers, such as philopatry, local adaptation or isolation by colonisation (Orsini *et al.* 2013). We therefore hypothesised that breeding colonies on different archipelagos would constitute genetically distinct populations. To test this hypothesis we generated a dataset of more than 5,000 unlinked single nucleotide polymorphisms (SNPs) using restriction site associated DNA sequencing (RADSeq) (Baird *et al.* 2008) for king penguins from four colonies spread across their range (Figure 3.1). We aimed to identify population structure, as well as distinct phylogenetic lineages that may have been associated with past glacial refugia. Previous studies have shown that king penguin numbers were much reduced during the last ice age (Trucchi *et al.* 2014), and the species' range may have been contracted into refugia at unknown locations (Younger *et al.* 2016b). Finally, we aimed to test the hypothesis that the recently founded colony at the Falkland Islands (Pistorius *et al.* 2012) was established via migration from nearby South Georgia (Figure 3.1).

3.3 Methods

3.3.1 Sampling

Blood was collected from 16 king penguins at each of: Volunteer Point on the Falkland Islands (Feb 2014), Fortuna Bay on South Georgia (Dec 2012), Baie du Marin on Possession Island in the Crozet Islands (Dec 2003 – Jan 2004) and Sandy Bay on Macquarie Island (Dec 2005 – Jan 2006) (Figure 3.1). To prevent biting and minimize stress during handling (Le Maho *et al.* 1992), king penguins were either seized with both hands and the flippers were restrained with the head placed under the arm of the handler, or they were wrapped in cushioned material to cover the head and prevent movement. A second handler took 1 ml blood from the brachial or ulnar vein using a 25G or 23G needle and 1 mL syringe, after cleaning the area with an alcohol swab. Total restraint time was generally two to three minutes. All field activities were conducted under

permits from the Falkland Islands Government, the Government of South Georgia and the South Sandwich Islands and the Tasmanian Parks Department, and also received ethical approval from the University of Oxford, the University of Western Australia, the Auburn University Institutional Animal Care and Use Committee and the Institut Polaire P. E. Victor. Blood samples were transported to the UK at ambient temperature in RNAlater (Life Technologies) or in Queen's Lysis buffer (http://aou.org/committees/collections/recipes_dna_buffers.php), and stored at -20 °C or -80 °C until extraction.

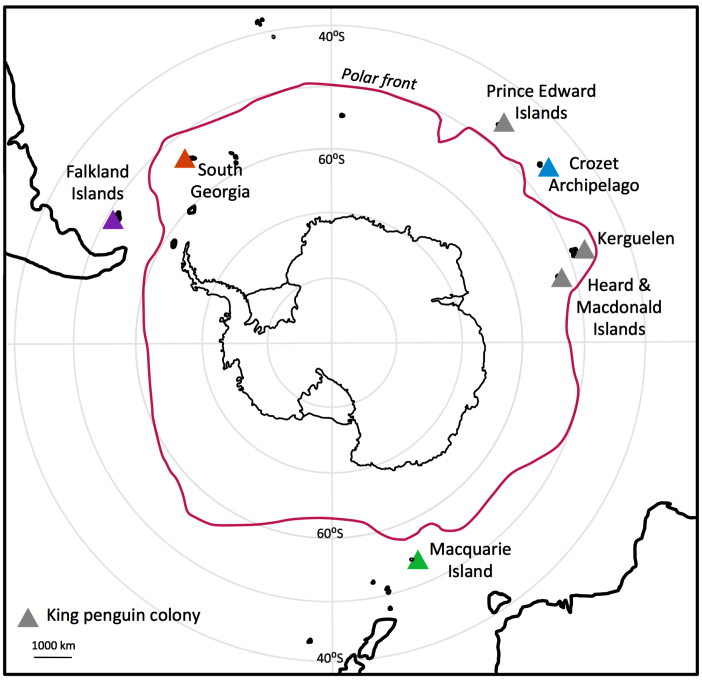


Figure 3.1. King penguin colony locations. Triangles indicate known king penguin colonies, with coloured triangles indicating the four colonies sampled for this study.

3.3.2 Sequencing

DNA was extracted from the 64 blood samples using a QIAGEN DNeasy Blood and Tissue Kit following the manufacturer's protocol, but modified to include 40 µL proteinase K at the digestion step and with the incubation time extended to 3 hrs. The samples were treated with 1 µL Riboshredder (Epicentre) to reduce RNA contamination. DNA concentration was measured with a Qubit (ThermoFisher Scientific) and high molecular weight was confirmed on a 1% gel. We sequenced the mitochondrial hypervariable region (HVR; 620 base pairs; GenBank accessions: KX857217-KX857259) because this marker has revealed phylogeographic patterns within other penguin species (Clucas *et al.* 2014; Ritchie *et al.* 2004; Younger *et al.* 2015b). The HVR was amplified in all samples using primers F-0225 (5'-GGAACCTCCCAAAGAGTACCA) and R-INR (5'-

CCAACCAGATGTATCGGTGA) (Younger *et al.* 2015b). PCR products were sequenced using the Sanger method by Macrogen Europe. Geneious v5.5.9 was used for alignment.

We employed RADSeq to generate a dataset of genome wide SNPs to assess population structure among the king penguin colonies. RAD libraries were prepared using the SbfI restriction enzyme, which was chosen because it produces a large number of RAD loci in king penguins (Trucchi *et al.* 2014). RADSeq for all individuals was performed at the Edinburgh Genomics Facility, University of Edinburgh (https://genomics.ed.ac.uk/) as described in Gonen *et al.* (2014) after Etter *et al.* (2011). Briefly, 250 ng of DNA per individual was digested with SbfI-HF (NEB), followed by ligation to barcoded P1 adapters. The uniquely barcoded individuals were pooled into multiplexed libraries, and each library sheared into fragments of □300—400 bp. Fragments were size selected using gel electrophoresis. The libraries were blunt ended (NEB Quick Blunting Kit) and A-tailed prior to ligation with P2 adapters (IDT). Enrichment PCR was performed to increase yield, followed by product purification with Ampure beads. The pooled, enriched libraries were checked for size and quantity using Qubit and a qPCR assay. Each library was then sequenced in a lane of the Illumina HiSeq 2500 using 125 base paired-end reads in high output mode (v4 chemistry).

3.3.3 Bioinformatics

FastQC was used to assess read quality and check for adapter contamination. We used *process_radtags* within the Stacks pipeline v1.35 (Catchen *et al.* 2011; Catchen *et al.* 2013) to de-multiplex, trim and clean reads. We then truncated reads to 113 bp to exclude the four terminal bases in order to avoid poor sequence quality. We excluded read pairs in which either read had uncalled bases, a low quality score and/or a barcode or cut-site with more than one mismatch. The remaining paired reads were aligned to the emperor penguin reference genome (http://gigadb.org/dataset/100005) using *bwa-mem* (Li 2013). We prevented terminal alignments by enforcing a clipping penalty of 100. Reads with more than five mismatches, multiple alignments and/or more than two indels were removed using a custom python script (filter.py, available online: 10.5061/dryad.7c0q8). We removed PCR duplicates with Picardtools (http://broadinstitute.github.io/picard).

We used the Stacks pipeline (*pstacks – cstacks – sstacks – rxstacks – cstacks – sstacks - populations*) to prepare a dataset of putatively unlinked, filtered SNPs from the RAD reads, following many of the suggestions outlined in the framework of Benestan *et al.* (2016). In *pstacks* we selected a minimum stack depth of six reads mapping to the same location and used the bounded SNP model with a significance level of α = 0.05, an upper bound of 0.1 and a lower bound of 0.0041 (corresponding to the highest sequencing error rate recorded by phiX spikes in

the sequencing lanes). All 64 individuals were used to build the catalog in *cstacks*. In *rxstacks* we removed confounded loci (those with a biologically implausible number of haplotypes, such as from repetitive sequences or paralogous loci) with a conservative confidence limit of 0.25. Also in *rxstacks*, we removed excess haplotypes as well as any loci with a mean log likelihood < -10. Further filtering was conducted in the *populations* module. We removed SNPs with a minor allele frequency (MAF) < 0.01 because these are likely to be the result of sequencing errors. We also removed loci with a heterozygosity > 0.5, as these could be paralogs (Benestan *et al.* 2016). A single SNP per RADtag was chosen at random in order to remove tightly linked SNPs from the dataset. We also specified that a locus must be present in all colonies to be included in the final dataset, as well as genotyped in at least 80% of individuals from each colony. We then removed any SNPs with a mean coverage exceeding 100X using vcftools v0.1.13 (Danecek *et al.* 2011) to avoid SNPs from repetitive regions of the genome (Supplementary Figure 3.5). We also removed SNPs that were out of Hardy Weinberg equilibrium (HWE) in > 50% of the colonies when $p < 0.01$ using the *adegenet* package in R (Jombart 2008; Jombart & Ahmed 2011) and vcftools. Finally, PGDSpider v2.0.8.2 (Lischer & Excoffier 2012a) was used to convert the vcf file into other formats for subsequent analyses.

3.3.4 Outlier loci detection

We investigated whether SNPs were potentially under selection before proceeding with population genetic analyses, because loci under either directional or balancing selection violate the assumption of neutrality that is a caveat of most population genetic methods. We used a Bayesian F_{ST} outlier test as implemented in BayeScan 2.1 (Foll & Gaggiotti 2008) to identify loci to be discarded from the neutral dataset. BayeScan has been shown to have good power for detecting loci genuinely under selection under a range of demographic scenarios, but with an accompanying high false-positive rate (Lotterhos & Whitlock 2014). Given that our reason for testing for outlier loci is to obtain a truly neutral dataset, we are not concerned by the high false-positive rate in this case. We set the prior odds of neutrality parameter at five, which refers to the probability that a given locus in the dataset is under selection (i.e. for every five loci one is under selection). This prior was chosen as we aimed to remove all loci that could possibly be under selection. We deemed q-values of < 0.1 to be a significant result, meaning that for a dataset of 100 F_{ST} outliers we can expect ten of these to be false-positive neutral loci (Lotterhos & Whitlock 2014; Storey & Tibshirani 2003).

3.3.5 Contemporary population structure

The genetic structuring among king penguin colonies was assessed using several different methods. Firstly, the Weir and Cockerham (1984) unbiased estimator of F_{ST} was calculated between all pairs of colonies using Genodive v2.0b27 (Meirmans & Van Tienderen 2004). The hypothesis of departure from panmixia was tested with 5,000 random permutations of the data to determine the statistical significance of each pairwise F_{ST} value between colonies, with the significance level adjusted for multiple testing using Sequential Goodness of Fit (SGoF+) (Carvajal-Rodriguez & de Uña-Alvarez 2011).

To identify the number of genetic populations ("clusters") among the 64 individuals, we used the *find.clusters K*-means clustering algorithm within the *adegenet* package (Jombart 2008; Jombart & Ahmed 2011), retaining all principal components. We also used a Bayesian clustering approach with a Markov chain Monte Carlo (MCMC) sampling procedure within *Structure* v2.3.4 (Pritchard *et al.* 2000). The analysis estimated the membership coefficient of each individual to each of the inferred clusters, effectively assigning individuals to genetic populations. We used the admixture model with correlated allele frequencies, because our pairwise F_{ST} results suggest that it is highly likely that these colonies have experienced admixture in the past and/or are still exchanging migrants. Models were run both with and without location priors to reflect the colony that each individual was sampled at, to detect subtle versus strong population structure. We conducted an initial run to infer the value of lambda, using a setting of $K = 1$ and an MCMC length of 100,000 generations (with the first 50,000 discarded as burn-in), allowing lambda to vary. The value of lambda was then fixed at 0.39 for subsequent analyses. *K* values (the number of inferred clusters) from one to four were tested, with each value of *K* run a total of ten times from different random seeds. Each analysis was run for 150,000 generations with the first 50,000 discarded as burn-in. *structure harvester* web v0.6.94 (Earl 2012) was used to compare *K* values using the Evanno method (Evanno *et al.* 2005) and prepare files for CLUMPP (Jakobsson & Rosenberg 2007). Replicate runs for each value of *K* were aligned using CLUMPP to check for multimodality, and the membership coefficients of each individual to each cluster were visualised with DISTRUCT v1.1 (Rosenberg 2004).

Discriminant Analysis of Principal Components (DAPC) (Jombart *et al.* 2010) can be used to describe clusters in genetic data by creating synthetic variables (discriminant functions) that maximise variance among groups whilst minimising variance within groups. DAPC was run when individuals were grouped by colony of origin and when individuals were grouped by the genetic clusters found in our other analyses, for comparison. These groups were (1) South Georgia and (2)

the Falkland Islands, Crozet and Macquarie. The optimal number of principal components (PCs) to retain in each analysis was determined by the average of 20 runs of the function *optim.a.score*.

We conducted individual-based population assignment tests, in which an assignment algorithm attempts to assign the individuals in the test set to their population of origin (Paetkau *et al.* 1995). Individuals were grouped into the two genetic populations we described above. As assignment tests can be sensitive to uneven sample numbers, we randomly sampled 16 individuals from the larger population to match the size of the South Georgia population. Each group was divided into a *training* set and a *hold-out* set and we identified the most informative SNPs for colony assignment using the *training* set in TRES v1.0 (Kavakiotis *et al.* 2015). We used the Informative for Assignment test (I_n) to identify ancestry informative markers (AIMs), as I_n has been shown to be the most powerful method for estimating ancestry proportions (Ding *et al.* 2011a). For population assignment tests it is recommended to trial different numbers of SNPs, therefore, we exported the top 100, 200, 500, 1,000 and 2,000 most informative SNPs. These SNP datasets were used to assign the *hold-out* set of individuals to their populations of origin within Genodive. If the minor allele was not sampled in either population (i.e. its frequency was zero) the frequency was replaced with 0.005 as recommended by Paetkau *et al.* (2004). We used the likelihood that the individual comes from the population it was sampled in (L_h) as the test statistic and a Monte Carlo test with 10,000 generations to estimate the null distribution of likelihood values. The threshold value was defined for each population based on the null distribution, at $\alpha = 0.05$.

3.3.6 Past population patterns

We used a species tree approach, as implemented in SNAPP (Bryant *et al.* 2012) within BEAST v2.4.0 (Bouckaert *et al.* 2014), to estimate the evolutionary relationships and order of splitting among the geographically isolated colonies to determine whether any of the colonies may have been glacial refugia in the past, as well as the source of the new Falkland Islands colony. SNAPP uses a coalescent method to infer species trees from unlinked biallelic markers, such as SNPs. SNAPP is highly computationally demanding and analysis of the full dataset of individuals was implausible. We therefore selected two representative individuals from each colony (i.e. four haplotypes) for analysis, and to ensure consistency of the posterior we ran the analysis twice with different randomly selected colony representatives. Any SNPs that were no longer polymorphic within the reduced datasets were removed from analysis, leaving datasets of 2,668 and 2,626 SNPs. The mutation rates (u and v) were calculated from the data, rather than estimated as part of the MCMC. We ran the MCMC for 5 million generations with a burn-in of 10%. This was more than sufficient for convergence, with Tracer v1.6 (Rambaut & Drummond 2007) indicating ESSs > 4000. The likelihood plots were also visually inspected for convergence. The Bayesian method

results in not a single topology, but a posterior distribution of the possible topologies; we used DensiTree v2.0.1 (Bouckaert 2010) to visualise the entire posterior distributions of trees as a cloudogram, excluding a 10% burn-in.

We used RAxML v8.2.7 (Stamatakis 2014) to infer maximum likelihood (ML) phylogenies among the full dataset of king penguin individuals. We applied an ascertainment bias correction to the likelihood calculations, as recommended when using SNPs to account for the lack of invariant sites (Leaché *et al.* 2015). For the ascertainment correction to function, all invariant sites must be removed. In practice, this means that an alignment site consisting of only heterozygotes and homozygotes for a single allele (e.g. an alignment site that is only Rs and As with no Gs) is considered potentially invariant by RAxML and must be removed. We filtered out such sites using the Phrynomics R script (https://rstudio.stat.washington.edu/shiny/phrynomics/). After this filtering step 1,727 SNPs remained in the dataset. We conducted a rapid bootstrap analysis and search for the best-scoring maximum likelihood tree in a single program run using the MRE-based bootstopping criterion (Pattengale *et al.* 2010) to ascertain when sufficient bootstrap replicates had been generated. All searches were conducted under the GTRGAMMA nucleotide substitution model. We also conducted a ML search on the HVR sequences, because HVR has been shown to resolve distinct phylogenetic lineages within Adélie penguins (Clucas *et al.* 2014; Ritchie *et al.* 2004; Younger *et al.* 2015a), emperor penguins (Younger *et al.* 2015b) and gentoo penguins (Clucas *et al.* 2014; Levy *et al.* 2016). We used the same search protocol as for the SNP dataset, but without an ascertainment bias correction. Finally, we constructed a median-joining haplotype network for the HVR sequences using PopArt (http://popart.otago.ac.nz).

3.4 Results

3.4.1 Genotyping

The 64 king penguin samples yielded 6.27 — 55.9 million unpaired reads per individual, with an average of 15.7 million reads per individual. On average, 97.3% of reads per individual passed the quality filters in *process_radtags* and, of these, an average of 97.7% successfully aligned to the emperor penguin reference genome. After specifying a minimum stack depth of six, a total of 34,171 RAD-tags remained, containing 35,766 SNPs. Our SNP filtering protocols resulted in a final dataset of 5,154 SNPs (available online 10.5061/dryad.7c0q8) for use in subsequent analyses. Of these we detected no loci that were putatively under selection (BayeScan output available online: 10.5061/dryad.7c0q8) and none that were out of HWE in > 50% of colonies. There were no notable differences in genetic diversity measures (number of private alleles, expected heterozygosity, observed heterozygosity or nucleotide diversity) among colonies (Table 3.1).

Table 3.1. Genetic diversity measures by colony. Number of individuals (N), number of private alleles (private alleles), expected heterozygosity (H_E), observed heterozygosity (H_O) and nucleotide diversity (π), variance (var) and standard error (StdErr)

	N	**Private alleles**	H_E **(mean)**	H_E **(var)**	H_E **(StdErr)**	H_O **(mean)**	H_O **(var)**	H_O **(StdErr)**	**π (mean)**	**π (var)**	**π (StdErr)**
Falkland Islands	16	148	0.1179	0.0175	0.0018	0.1107	0.0170	0.0018	0.1219	0.0187	0.0019
South Georgia	16	147	0.1161	0.0174	0.0018	0.1066	0.0161	0.0018	0.1200	0.0185	0.0019
Crozet	16	117	0.1178	0.0177	0.0019	0.1151	0.0183	0.0019	0.1217	0.0189	0.0019
Macquarie	16	180	0.1187	0.0178	0.0019	0.1115	0.0175	0.0018	0.1225	0.0189	0.0019

3.4.2 Genetic populations of king penguins

We conducted multiple analyses of population assignment and delimitation to identify the number and geographic boundaries of distinct genetic populations among the four sampled king penguin colonies. The optimal number of clusters among the 64 individuals were $K = 3$ and $K = 2$ for *structure* analyses with and without location priors, respectively, as determined by the Evanno method. However, the highest posterior mean log probability of the data for both scenarios (i.e. with and without the sampling location specified as a prior) was at $K = 1$. The rate of change in log probability (deltaK) is not defined at K=1, and so the Evanno method is unable to determine whether this is actually the true value of K. This suggests that the signal for multiple clusters is weak. Inspection of the individual assignment plots (Figure 3.2) showed that three clusters explain the majority of the subtle structure. The Falkland Islands and Crozet Islands cluster together, whereas the Macquarie Island and South Georgia colonies appear differentiated. The *K*-means clustering algorithm was unable to distinguish these clusters as the lowest value of the BIC, which indicates the optimal clustering solution was found at $K = 1$.

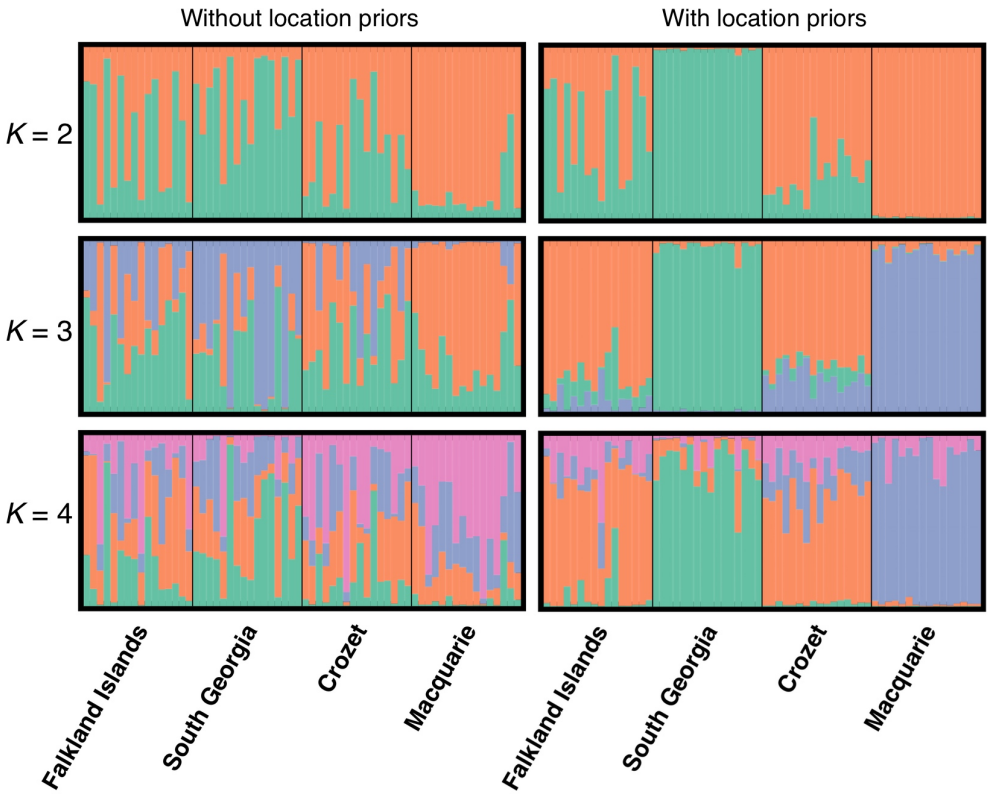


Figure 3.2. Population assignment of individuals by Bayesian clustering with the program *Structure*. Membership coefficients for each individual are shown by vertical bars with the clusters represented by colours. The Evanno method selected $K = 2$ when no location prior was used and $K = 3$ when a location prior was used. When $K = 3$ the three clusters correspond to 1) the Falkland Islands and Crozet colonies, 2) the South Georgia colony, and 3) the Macquarie Island colony.

Our measures of pairwise F_{ST} (Table 2.1) indicate that the Crozet and Falkland Islands colonies are not differentiated from one another (F_{ST} = -0.001), and that Macquarie and Crozet Islands are not significantly differentiated from each other (F_{ST} = 0.001). All other pairs of populations are statistically significantly differentiated after SGoF+ correction for multiple tests, however, the values of F_{ST} are very small (0.003 — 0.005), indicating only subtle genetic differences between these pairs of colonies. Therefore there are at least two slightly differentiated genetic populations among the sampled colonies: (1) the South Georgia population and (2) a population including the Falkland Islands, Crozet and Macquarie.

Table 3.2. Pairwise genetic differentiation (F_{ST}) between pairs of colonies. Results that are significantly different from zero at the α = 0.05 level, following SGoF+ correction, are indicated with asterisks.

	Falkland Islands	**South Georgia**	**Crozet**
South Georgia	0.003*		
Crozet	-0.001	0.003*	
Macquarie	0.003*	0.005*	0.001

DAPC was unable to distinguish among the four sampled colonies or between the two slightly differentiated populations, with the distribution of individuals overlapping in both scenarios (Figure 3.3). For the individual-based population assignment tests, the 100 SNP dataset was found to be best at assigning the test set of individuals back to their population of origin. However, the test performed poorly, with only seven individuals assigned correctly out of the 16 individuals in the test dataset. Given that there were only two possible populations of origin, this is slightly worse than assigning individuals to colonies at random. This again suggests that there is very little differentiation among the king penguin colonies.

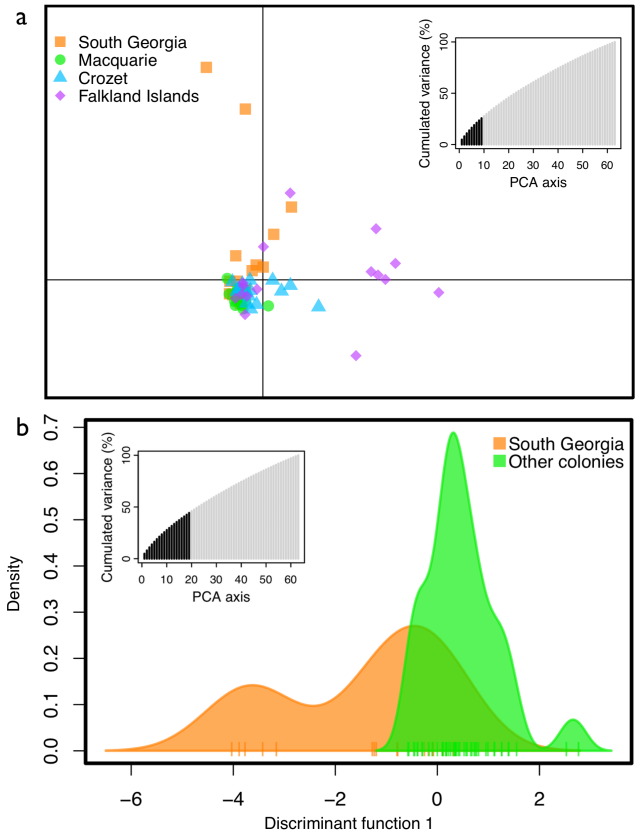


Figure 3.3. Discriminant analysis of principal components. Individuals are grouped by a) their colony of origin and b) the two genetic clusters identified by other analyses. The retained PCs are shown in black on the inset graphs.

Overall, our analyses of population structure among the four king penguin colonies have yielded some surprising results. Despite separation of thousands of kilometres, there is very little genetic differentiation among these colonies. The South Georgia population was subtly differentiated from all other colonies, and the Macquarie population was further very subtly differentiated from some colonies by a subset of our analyses. It is particularly interesting that the Falkland Islands colony is genetically indistinguishable from the Crozet Islands colony, despite a separation of *ca*. 7,500 km, whereas the nearby South Georgia colony is differentiated; based on our results it seems most likely that the Falkland Islands colony was founded by individuals from the Crozet Islands, rather than nearby South Georgia, even though there seems to be no obvious biological explanation for why this might be so.

3.4.3 Phylogeography

We attempted to ascertain the branching structure among colonies using the species tree approach implemented in SNAPP. We have presented the full posterior distribution of trees in order to highlight the uncertainty in the topology (Figure 3.4). The majority of the topologies support the grouping of the Falkland and Crozet Islands colonies (Figure 3.4), congruent with our *Structure* and pairwise F_{ST} results. However, aside from this one clade, the rest of the branching structure among the colonies is unresolved.

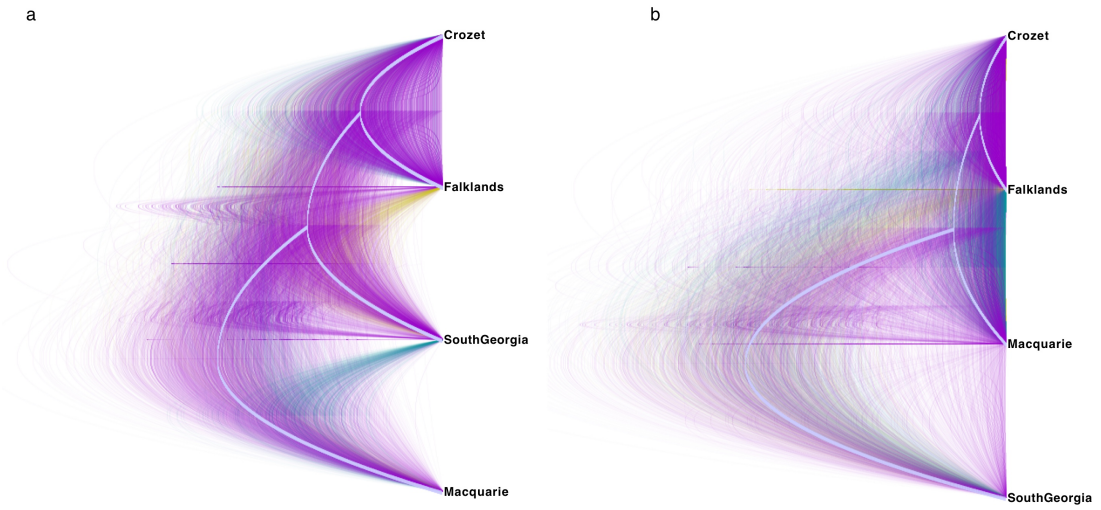


Figure 3.4. Evolutionary relationships among colonies. The full posterior distributions of trees from the SNAPP analyses, excluding a 10% burn-in, are shown. The colours represent the different topologies; purple is the most highly supported, teal is the next most supported, and gold is the least supported. The consensus tree is shown in grey. (a) and (b) are the outcomes of the two different analyses with different randomly selected representative individuals.

We constructed maximum likelihood phylogenies for the full set of individuals using both HVR and the dataset of SNPs in order to determine if there are any strongly supported phylogenetic lineages that are not necessarily affiliated with the contemporary colony sites. The MRE bootstopping-criterion was satisfied by 550 and 800 bootstraps for the SNP and HVR searches, respectively. The best-scoring likelihood and majority rule extended consensus trees for the SNP dataset had very low support across the entire topology, with only a single node having a branch support value > 50 (topology not shown). The HVR topology did not show any more resolution, with 75% of nodes in the tree having branch support values < 50 and no evidence of any well-supported phylogenetic lineages (topology not shown). A median joining network of the haplotypes of the mitochondrial HVR also showed no clear phylogeographic pattern and no evidence of ancestral haplotypes (Supplementary Figure 3.6). Overall, there are no distinct lineages among king penguins, no remnant signatures of refugia and no evidence for the order of colonisation of the islands.

3.5 Discussion

In the first study of king penguin global population structure we found very low levels of population differentiation across the species' entire distribution, despite using 5,154 SNPs distributed throughout the genome. Penguins from the Crozet Islands were not genetically differentiated from those 7,450 km west on the Falkland Islands, nor those 7,100 km east on Macquarie Island. There was very low, yet statistically significant, genetic differentiation between the colony on South Georgia and all other colonies, including the Falkland Islands located only 1,400 km to the northeast. Our phylogeographic analyses showed no evidence of distinct king penguin lineages.

The lack of genetic differentiation across such vast distances is surprising given that king penguin colonies are sparsely distributed across the Southern Ocean. There are very few locations that support king penguin breeding between the archipelagos we have sampled; the only other colonies are in the Indian Ocean sector close to the Crozet Islands (Figure 3.1). Therefore, there are very few "stepping stones" between colonies and the lack of differentiation between Crozet and Macquarie suggests that migration is not distance-limited.

There are two alternative explanations for the observed low levels of genetic differentiation among king penguin colonies. Firstly, it could be the result of frequent migration of individuals among these isolated archipelagos. In this scenario, dispersing individuals must also be recruited into the breeding population upon arrival, if they are to contribute to the gene flow that is maintaining near genetic homogeneity of king penguins. Alternatively, all extant colonies may

share a common ancestral population and insufficient time has passed for them to diverge, even if they are now isolated. Despite the large geographic distances separating them, there is a growing body of evidence to suggest that king penguin colonies do exchange migrants (Gartshore *et al.* 1988; Weimerskirch *et al.* 1985; Woehler 1989); we therefore consider the former hypothesis, that migration is maintaining gene flow among populations, to be the most likely explanation for the genetic similarity found here.

The recent formation of new colonies at Volunteer Point on the Falkland Islands (Pistorius *et al.* 2012), Possession Island in the Crozet Islands (Delord *et al.* 2004) and on Macquarie Island (Van Den Hoff *et al.* 2009) provides direct evidence that some individuals will breed away from their natal colony. A handful of individuals banded as breeders have also been observed breeding at non-natal colonies within the Crozet Islands (Bost, C. A. *pers. comm.*). Furthermore, the rate of population growth at Possession Island over the past several decades has been too great to have been maintained by intrinsic recruitment alone; therefore, the population growth must be partially attributable to immigration (Delord *et al.* 2004). Small numbers of king penguins, and in particular juveniles, have been observed at colonies up to 5,600 km from their natal colonies (Gartshore *et al.* 1988; Weimerskirch *et al.* 1985; Weimerskirch *et al.* 1992; Woehler 1989). This suggests that king penguins probably prospect other colonies and breeding habitats, including those far from their natal colony, and this may occur most often before they begin to breed. This prospecting behaviour may facilitate emigration when conditions at the natal colony are less favourable than those found elsewhere.

Previous studies have shown that seabirds with large foraging ranges or those that disperse widely in the non-breeding season are least likely to show genetic differentiation among colonies (Friesen *et al.* 2007a). During the summer breeding season, king penguin foraging trips typically last days to weeks and can cover hundreds to thousands of kilometres (Ratcliffe & Trathan 2012). During the winter, king penguins rarely provision their chicks, and so adults are not restricted to central-place foraging. These winter foraging trips often take them over 1,500 km away from their colonies to the marginal ice zone around Antarctica, and journeys in excess of 10,000 km have been recorded, although there is no evidence for foraging range overlap among breeding colonies thus far (Pütz 2002; Pütz *et al.* 1999). The few juveniles that have been tracked after fledging dispersed widely in their first six months, probably bringing them into contact with individuals from other colonies (Pütz *et al.* 2014). Therefore juvenile dispersal and possibly also foraging range overlap during the non-breeding season appears to facilitate gene flow in king penguins, as it does across a variety of seabirds (Friesen 2015), but without more data on the winter dispersal of king penguins it is difficult to determine the relative importance of these mechanisms.

It is unclear whether the observed low level of genetic differentiation is maintained by consistent background levels of migration, or whether episodic periods of higher migration have occurred, or both. Abiotic factors such as glacial expansion and retreat, landslides, erosion, flooding, volcanic activity or other such catastrophic events (Van Den Hoff *et al.* 2009) could result in periods of increased emigration, whilst large-scale climatic anomalies that affect the proximity of oceanic fronts and prey availability to colonies (Bost *et al.* 2015) could also increase the emigration rate if adults perceive the habitat quality to have declined. The harvesting of king penguins during the late 19th and early 20th century could have temporarily increased emigration rates, if individuals emigrated to less disturbed colonies. Biotic factors could also play a role, as emigration may be favoured when colonies reach carrying capacity and/or density-dependent factors limit population growth, such as competition for food and nest sites, predation and pathogen load (Delord *et al.* 2004). The colony at Lusitania Bay on Macquarie Island is thought to have reached carrying capacity in 1975 when all available breeding habitat was occupied and individuals were forced to spill over to other colonies (Rounsevell & Copson 1982; Van Den Hoff *et al.* 2009). Two large colonies, Petite Manchotière and Jardin Japonais, on Possession Island in the Crozet Islands are also believed to have reached carrying capacity in the late 1980s, with all areas free of vegetation being occupied (Delord *et al.* 2004). As these colonies approached carrying capacity, the formation of the two new colonies on Possession Island in 1979 and 1986 could have been the direct result of these large colonies spilling over, with individuals emigrating rather than competing for nest spaces at their natal colonies. This could also account for the colonization of the Falkland Islands in the late 1970s. We found no evidence for genetic differentiation between the Falkland Islands and the Crozet Islands, and the colonies grouped together in our species tree analysis. Therefore it seems likely that individuals from the Crozet Islands, possibly forced to emigrate due to competition for space at their natal colonies, founded the population at the Falkland Islands. This finding was somewhat unexpected given the 7,450 km between the populations, and the relative proximity of the South Georgia population just 1,400 km away. Furthermore, the observation of an individual that was banded as a chick in South Georgia but was later found breeding in the Falkland Islands (Otley *et al.* 2007) would also tend to suggest that the Falkland Island population would have been founded by immigrants from South Georgia. However, our genetic results indicate that there has been a higher rate of immigration from the Crozet Islands than from South Georgia.

The difference in the oceanic regime experienced by king penguins at South Georgia could explain why this colony was genetically differentiated from all other colonies (Friesen 2015). South Georgia lies to the south of the Polar Front, whilst all other studied colonies lie to the north, and thus birds at South Georgia experience colder oceanic and air temperatures and a more krill-

dominated food web. The different ecological conditions either side of the Polar Front appear to act as a barrier to gene flow in many species (Rogers 2012), including gentoo penguins (Clucas *et al.* 2014), although this effect appears much weaker in king penguins.

While it would be useful to be able to determine the actual migration rates among the colonies studied here, the very low levels of genetic differentiation preclude the calculation of accurate estimates. Hence, whether the colonies are demographically linked or should be considered as separate management units cannot be determined (Palsbøll *et al.* 2007). Furthermore, there is currently no generalized framework for determining the level of migration necessary to maintain demographic linkage (Waples & Gaggiotti 2006). BayesAss (Wilson & Rannala 2003), which is typically used to determine recent directional migration rates between populations (gene flow occurring over the last few generations), has been found to be unreliable when F_{ST} values are less than 0.05 (i.e. an order of magnitude greater than observed among king penguins) (Faubet *et al.* 2007). Methods to estimate migration based on *F*-statistics are also unreliable because the assumptions of the island model (Wright 1931) that relates F_{ST} to the number of migrants entering a population (Nm) are usually violated in natural systems, limiting the amount of quantitative information about migration that can be gained from F-statistics (Whitlock & McCauley 1999). Finally, coalescent methods, such as Migrate-n (Beerli 2001), which estimate migration over evolutionary timescales, are also likely to be inaccurate when population differentiation is low and only a small number of loci can be used because of massive computational demands (Cristofari *et al.* 2016). Coalescent methods also rely on an estimate of the mutation rate for the specific loci used in the analysis, to translate the mutation-scaled migration rate into an estimate of the number of migrants entering a population, and accurate mutation rates are difficult to estimate for RAD loci (Harvey & Brumfield 2015; Harvey *et al.* 2016).

The lack of phylogenetic signal or mitochondrial lineages suggests that small populations of king penguins have not been isolated from one another in their recent history. Some colonies went through rapid declines when king penguins were harvested for their blubber. For example, the Macquarie Island colony was reduced from hundreds of thousands of birds to about 3,000 (Van Den Hoff *et al.* 2009). These rapid declines, although extreme demographically, were unlikely to have caused a genetic bottleneck resulting in lineage divergence, as they were neither severe enough nor lasted long enough for significant genetic drift to have taken place. Certainly there is no signature of recent genetic bottlenecks in our data. Furthermore, if the harvesting also caused a pulse of increased emigration and gene flow, then genetic diversity is unlikely to have been affected. Indeed, the Macquarie Island population appears to have retained genetic diversity throughout the period of harvesting, as demonstrated by a comparison of ancient, pre-harvest genetic diversity to the modern population (Heupink *et al.* 2012). The king penguin population at

La Baie du Marin colony on the Crozet Islands was much smaller during the last glacial maximum (LGM), and then rapidly increased in size following Holocene warming (Trucchi *et al.* 2014). LGM conditions appear to have isolated refugial populations of Adélie (Clucas *et al.* 2014; Ritchie *et al.* 2004; Younger *et al.* 2015a), emperor (Younger *et al.* 2015b) and gentoo penguins (Clucas *et al.* 2014; Levy *et al.* 2016) in ice age refugia, resulting in distinct mitochondrial lineages. Our results do not support this for king penguins, although distinct lineages could exist outside of the colonies we sampled. The single mitochondrial lineage found here suggests that gene flow between populations of king penguins was maintained during the LGM even if their population sizes were reduced, and their tendency to disperse probably allowed this. Interestingly, the emperor penguin, the sister-species to king penguins in the *Aptenodytes* genus, also has remarkable dispersal abilities, exhibiting very low levels of genetic differentiation around its global range (Cristofari *et al.* 2016; Younger *et al.* 2015b), similar to Adélie penguins (Clucas *et al.* 2014; Roeder *et al.* 2001; Younger *et al.* 2015a). Yet we see distinct mitochondrial lineages in the emperor penguin, with origins dated to the last ice age (Younger *et al.* 2015b), that are not apparent in king penguins. We propose that the sub-Antarctic distribution of king penguins may explain this contrast. Many of the sub-Antarctic islands king penguins breed on have been heavily glaciated (Hall 2004), reducing available breeding area, but the increased sea ice extent during glacial periods (Gersonde *et al.* 2005) would probably not have created barriers to king penguin migration as it did not extend as far north as the king penguin's sub-Antarctic range.

3.6 Conclusions

Our study has revealed an unexpectedly low level of genetic differentiation among king penguin colonies spanning thousands of kilometres of the Southern Ocean, with some colonies separated by more than 7,000 km showing no significant genetic divergence. On the other hand, the South Georgia colony does appear to be subtly differentiated from all other studied colonies, despite it lying in close proximity to the Falkland Island colony.

The very low level of genetic differentiation we have shown among king penguin colonies needs to be considered in management to mitigate future climate change impacts on the species. Colonies within the same archipelago are highly likely to be panmictic and demographically linked, and thus monitoring of king penguins should be considered at the archipelago level, rather than at the colony level. The subtle differentiation we found between some archipelagos, and our inability to determine whether migration is consistent or episodic, cautions against the assumption that colonies are demographically linked globally. Therefore, as a precaution, we recommend that populations at the archipelago level are managed as separate units. Given the relatively few archipelagos that host king penguins, and that climate change effects will be

heterogeneous across their range, declines at any of these locations should be considered as significant and would hinder the recovery of the species, even if a loss of genetic diversity would not occur.

Demographic models that attempt to forecast extinction risk in response to large-scale climate change must also take into account migration. Recently, Tavecchia *et al.* (2016) showed that migration can decouple the relationship between population growth rates and climate variables, such that even if demographic rates are sensitive to climate-driven variations, this does not necessarily result in climate-driven population changes when immigration of new individuals occurs. Migration could therefore buffer king penguins against their forecasted risk of extinction under climate change (Le Bohec *et al.* 2008) although it may not protect them completely (Bost *et al.* 2015).

3.7 Supplementary Information

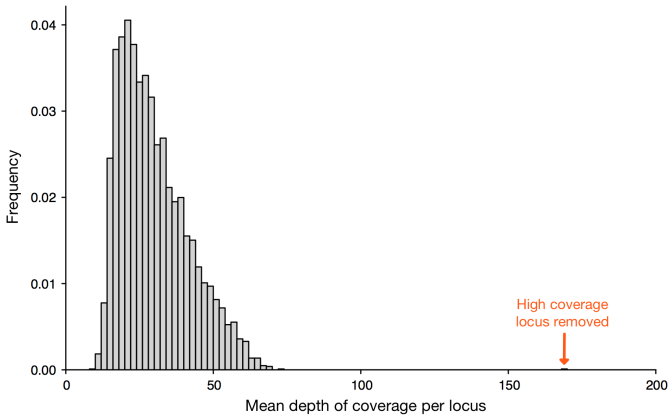


Figure 3.5. Histogram showing mean depth of coverage per king penguin SNP locus.

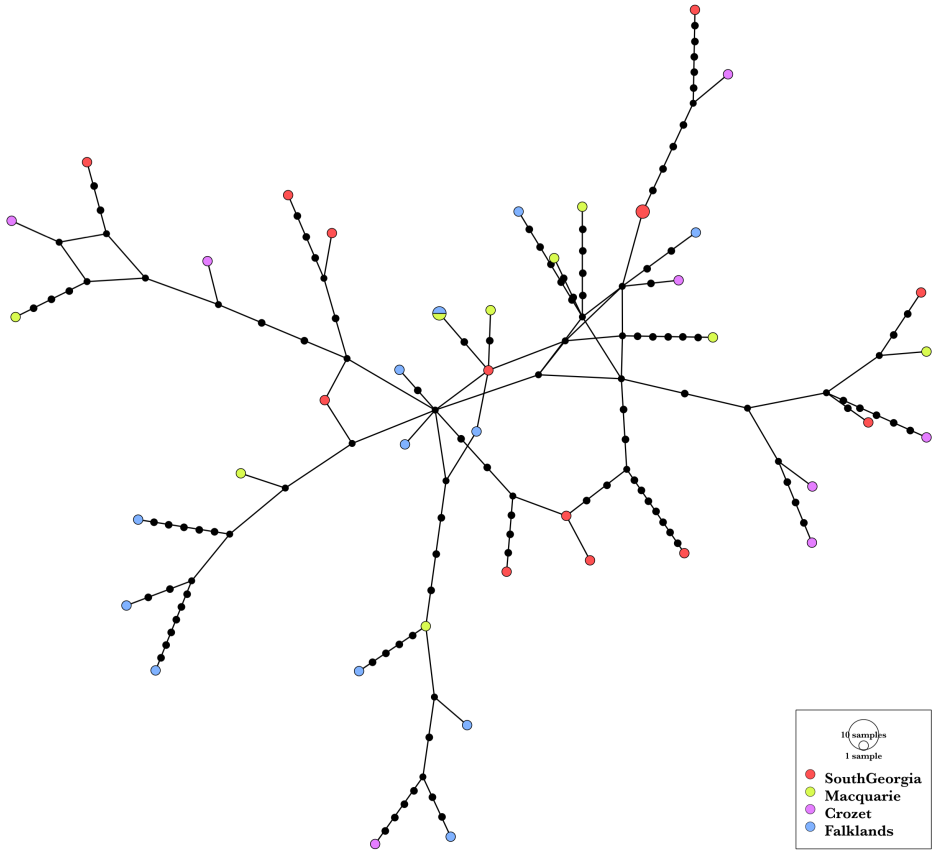


Figure 3.6 Median-joining haplotype network of king penguin HVR sequences.

Chapter 4: Recognising metapopulation structure in emperor penguins is critical for their future conservation[3]

4.1 Abstract

Emperor penguin populations may be facing extinction as a result of climate change-associated changes in sea ice. Understanding the boundaries of breeding populations in threatened species is of great importance, because unrecognised dispersal barriers can hinder conservation efforts and lead to erroneous estimates of extinction risk. Previous studies of emperor penguin dispersal have differed in their conclusions, particularly whether the Ross Sea, a major stronghold for the species, is isolated or not. We investigated emperor penguin population structure using genome-wide data, including a robust sample of Ross Sea individuals. We find that emperor penguins are comprised of at least four metapopulations connected by low levels of dispersal. The Ross Sea is clearly a distinct metapopulation, and thus we have highlighted the need for robust sample sizes and careful scrutiny of genomic data using multiple methods. Rigorous scientific methods are crucial, particularly when the future of a species is in the balance.

4.2 Introduction

Emperor penguins (*Aptenodytes forsteri*) are likely under threat (Ainley *et al.* 2010; Jenouvrier *et al.* 2009; Jenouvrier *et al.* 2014) from changing Antarctic sea ice conditions (Collins *et al.* 2013; Vaughan *et al.* 2013). Emperor penguins form breeding colonies on sea ice at the majority of their known colony locations (Fretwell *et al.* 2012) and have been shown to be sensitive to fluctuations in sea ice extent and seasonal duration (Ainley *et al.* 2010; Fretwell *et al.* 2014; Trathan *et al.* 2011). In light of this threat, risk assessments for the species under future climate change scenarios have begun (Jenouvrier *et al.* 2009; Jenouvrier *et al.* 2014). Jenouvrier *et al.* (2014) predicted continent-wide population trends of emperor penguins, based on a sea ice dependent demographic model paired with projected changes in local (colony-specific) sea ice conditions. Their findings were dire, with predicted declines of >50% at two thirds of the colonies examined by the year 2100, concordant with a minimum of 19% global decline in emperor penguin numbers (Jenouvrier *et al.* 2014).

[3] This chapter is presented as it was submitted to *Nature Communications*, where it is in review.

A major limitation of the forecasting study by Jenouvrier *et al.* (2014) was that each colony was modelled as an isolated breeding unit, with no exchange of individuals among locations. The assumption of isolation is likely to have led to erroneous projections, because dispersal is known to decouple correlations between population trajectories and climatic variables even where local vital rates are climate dependent (Tavecchia *et al.* 2016). In addition to altering the relationship between demographic rates and local climate, dispersal can increase the resilience of a species in several key ways. Firstly, immigrants can promote population stability by compensating for low survival or birth rates of natives (Lowe & Allendorf 2010), secondly, dispersal facilitates gene flow among breeding sites, replenishing the gene pool of a population with new, potentially adaptive alleles and, finally, dispersal enables range shifts (Walther *et al.* 2002). Range shifts may be particularly relevant for emperor penguins in the short term, as there is some evidence they may shift their colony sites and establish new colonies as local sea ice conditions become unfavourable (Ancel *et al.* 2014; LaRue *et al.* 2015). It is therefore crucial that connectivity among colony sites is incorporated into modelling studies and into risk assessments for the species. Delimiting the geographic boundaries of breeding populations is also essential for accurate monitoring of population trajectories and for implementing meaningful management plans (Funk *et al.* 2012; Palsbøll *et al.* 2007).

Recent evidence suggests that emperor penguin colonies are not as demographically isolated as previously thought. LaRue *et al.* (2015) gave evidence for six instances of either colony relocation or establishment of a new colony over five years after traditional breeding grounds were lost, as well as evidence for fluctuations in colony sizes explainable by temporary emigration. In addition, mobility of individual emperor penguins is known to be high, with juveniles shown to travel more than 7,000 km in just eight months (Thiebot *et al.* 2013), often in the vicinity of many different colonies (Kooyman *et al.* 1996; Wienecke *et al.* 2010). Given these observations, the likelihood of genetic exchange among geographically isolated colonies seems high.

A study of variation in emperor penguin mitochondrial DNA (mtDNA) found no statistically significant genetic differentiation across *ca.* 8,000 km of coastline, from the Adélie Land Coast to the Weddell Sea (Younger *et al.* 2015b). This genetic homogeneity could be the result of frequent dispersal of individuals among colonies that are successfully recruited into the breeding population upon arrival. However, given the nature of the genetic marker used, it is also possible that this pattern reflects that all extant colonies are derived from a common ancestral population, and insufficient time has passed for these colonies to diverge genetically, even if they are not currently exchanging migrants. Younger *et al.* (2015b) also found that penguins from the Ross Sea were genetically differentiated from those elsewhere on the continent, suggesting there are at least two genetically distinct populations of emperor penguins. A caveat of that study was that it

was based on mitochondrial DNA and therefore represents the genetic structure of a single, uniparentally-inherited, locus.

Multiple independent loci from across the genome should allow for a more sensitive investigation of genetic differentiation among colonies, to tease apart contemporary vs. historical connections. This approach was used by Cristofari *et al.* (2016) in a follow-up study of emperor penguin population structure, and the authors concluded that emperor penguins are not genetically differentiated among colonies and that the species consists of one single, panmictic population around Antarctica. Unfortunately, because of the small sample size in the Ross Sea, Cristofari *et al.* (2016) could not adequately explore the differentiation reported in Younger *et al.* (2015b). This represents a major limitation, because the Ross Sea contains genetically distinct populations of both emperor (Younger *et al.* 2015b) and Adélie (Ritchie *et al.* 2004) penguins, acted as an ice age refugium for both Antarctic penguin species (Ritchie *et al.* 2004; Younger *et al.* 2015b) and is currently home to the world's largest breeding colonies of both emperor (Fretwell *et al.* 2012) and Adélie (Lynch & LaRue 2014) penguins. Furthermore, it is the only region with a predicted stable or increasing population of emperor penguins (Jenouvrier *et al.* 2014). Finally, Cristofari *et al.* (2016) did not include colonies in West Antarctica, the region of Antarctica experiencing the most dramatic reductions in sea ice (Stammerjohn *et al.* 2012; Vaughan *et al.* 2013). As a result, their conclusion that "as a single genetic population, emperor penguins will respond to climate change through a unified evolutionary trajectory" (Cristofari *et al.* 2016) may be premature.

It is common practice to use genetics to delimit breeding populations and assess population connectivity (Lowe & Allendorf 2010), particularly for taxa such as emperor penguins, for which direct monitoring of dispersal is a logistical challenge. However, interpreting geographic patterns of genetic differentiation is difficult (Lowe & Allendorf 2010). It is crucial in such studies to recognise the distinctions between genetic and demographic connectivity (Lowe & Allendorf 2010; Palsbøll *et al.* 2007), and between ecological and evolutionary populations (Waples & Gaggiotti 2006). For populations to be demographically connected, their demographic rates, such as population growth rate, survival and birth rate, must be affected by immigration or emigration (Gilpin & Ilkka 1997; Lowe & Allendorf 2010). Genetic connectivity, on the other hand, depends solely on the number of dispersal events among populations (Lowe & Allendorf 2010). While this may seem a subtle distinction, it is an important one from a management perspective, because even very large dispersal rates do not necessarily indicate demographic linkage and these factors will ultimately affect a population's response to future change.

In this study we further investigate genetic connectivity among emperor penguin colonies around Antarctica using a robust sampling design including the Ross Sea. To achieve this, we expanded

our previous mitochondrial DNA analyses (Younger *et al.* 2015b) with a dataset of 4,596 high coverage genome-wide single nucleotide polymorphisms (SNPs) generated using restriction site associated DNA sequencing (RADSeq)(Baird *et al.* 2008). In contrast to the other recently published SNP study (Cristofari *et al.* 2016), our SNP data confirmed the existence of genetic differences between the Ross Sea and other Antarctic populations. Furthermore, we conclude that subtle yet significant genetic differences revealed among colonies are indicative of multiple metapopulations rather than a single panmictic population of emperor penguins. Recognising the likely frustration with which decision makers might view two apparently similar studies that provide contrasting conclusions about genetic structure in an iconic species, we advocate applying the precautionary principle such that, in the absence of definitive evidence for panmixia, the breeding populations identified here should be considered as separate units for management.

4.3 Results

4.3.1 Genotyping

A total of 110 individuals from eight emperor penguin colonies around Antarctica were genotyped for this study. To assess genetic differentiation at both continental and regional scales, two colonies from each of three major geographic regions (the Ross Sea, Weddell Sea and Prydz Bay) were included, along with an additional two colonies in East Antarctica (Figure 4.1). The individuals were sequenced across eight libraries along with samples of other species. The average number of reads per library was 14 million (range 11 – 17 million), 97% of which were retained after filtering for reads of low quality, adapter contamination, ambiguous barcodes and ambiguous RAD cut-sites. After reads were aligned to the emperor penguin reference genome, Stacks (Catchen *et al.* 2013) identified 423,479 SNPs, of which we retained 4,596 after following standard filtering steps (Benestan *et al.* 2016) (Table 4.1). The coverage of these SNPs ranged from 11 – 56X, with a mean of 33X. The number of individuals successfully genotyped at each colony varied between 10 and 16 (Table 4.2). Genetic diversity indices (expected and observed heterozygosity, and nucleotide diversity) were similar across all colonies (Table 4.2).

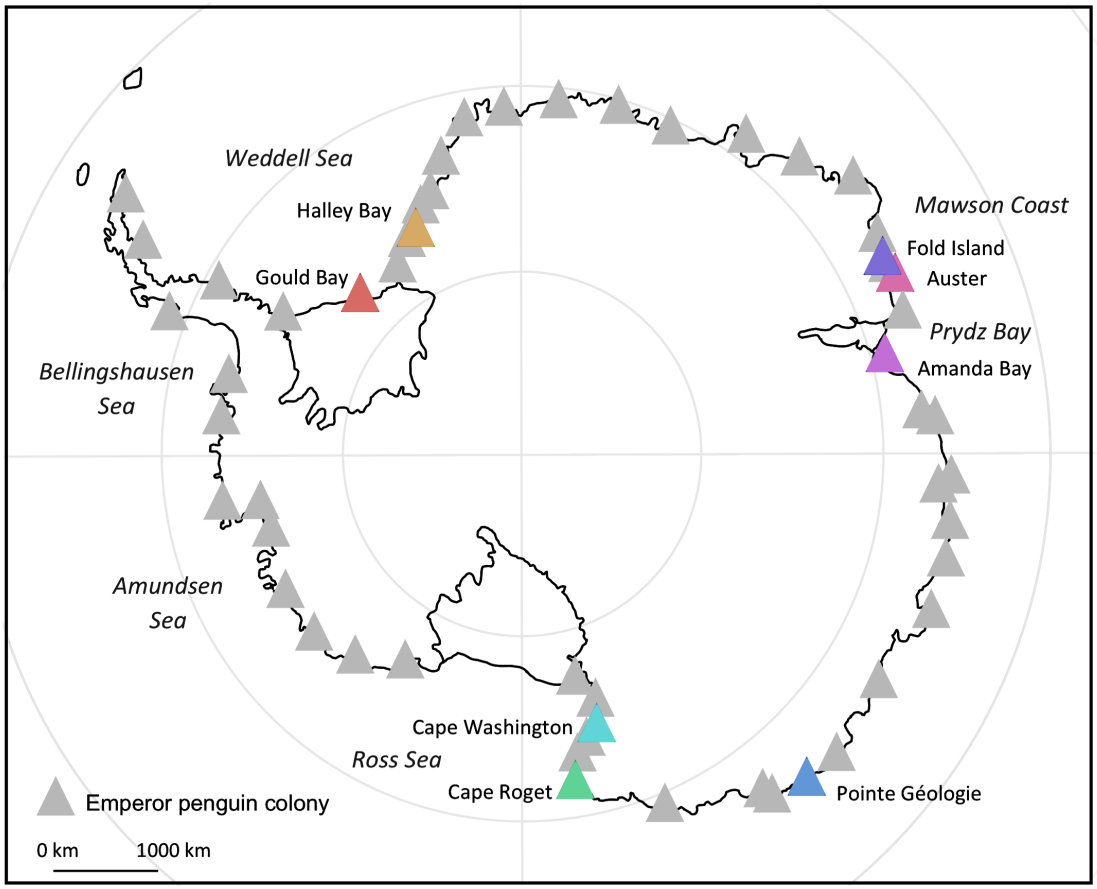


Figure 4.1. Map showing all known emperor penguin colonies around Antarctica (Fretwell *et al.* 2012). The colonies sampled in this study are coloured and labelled.

Table 4.1. The number of SNPs retained after applying each of the filters.

Filter	SNPs retained
Initial stacks catalog	423,479
After running rxstacks	64,320
Minor allele frequency > 0.01	30,441
Heterozygosity < 0.5	30,125
Genotyped in > 80% individuals per population	9,857
Single SNP per RADtag	4,600
Putatively neutral	4,596

Table 4.2. Genetic diversity indices for each colony, based on the dataset of 4,596 SNPs (variant sites only) retained after filtering. N – number of individuals successfully sequenced, H_E – expected heterozygosity; H_O – observed heterozygosity; π – nucleotide diversity; var – variance; stdErr – standard error of the mean.

	N	Private Alleles	H_E (mean)	H_E (var)	H_E (stdErr)	H_O (mean)	H_O (var)	H_O (stdErr)	π (mean)	π (var)	π (stdErr)
Gould Bay	13	7	0.1218	0.0178	0.0020	0.1133	0.0176	0.002	0.1261	0.0191	0.0020
Halley Bay	13	6	0.1223	0.0182	0.0020	0.1187	0.0192	0.002	0.1265	0.0194	0.0021
Fold Island	16	11	0.1231	0.0177	0.0020	0.1112	0.0164	0.0019	0.1271	0.0189	0.0020
Auster	16	5	0.1223	0.0173	0.0019	0.1189	0.0177	0.002	0.1264	0.0185	0.0020
Amanda Bay	16	9	0.1243	0.0173	0.0019	0.1171	0.0171	0.0019	0.1283	0.0184	0.0020
Pointe Géologie	15	6	0.1227	0.0180	0.0020	0.1131	0.0174	0.0019	0.1269	0.0192	0.0020
Cape Roget	10	1	0.1151	0.0196	0.0021	0.1071	0.0201	0.0021	0.1220	0.0220	0.0022
Cape Washington	11	1	0.1174	0.0186	0.0020	0.1134	0.0199	0.0021	0.1232	0.0205	0.0021

4.3.2 Clustering of populations and individuals

The AMOVA-based *K*-means clustering of populations (Meirmans 2012) and the Bayesian clustering of individuals performed with *Structure* (Pritchard *et al.* 2000) both separated colonies into genetic populations. However, both methods yielded several possibilities for the number of clusters, suggesting that differentiation among emperor penguin colonies is subtle and/or hierarchical. The two different *K*-means summary statistics for assessing the number of clusters yielded different results: $K = 3$ with the pseudo-F statistic and $K = 6$ with the Bayesian Inference Criterion (BIC). The *Structure* analysis did not converge on a single result, with different replicate runs arriving at distinct solutions. This was a result of genuine multimodality between runs, as opposed to label switching (Jakobsson & Rosenberg 2007) (Supplementary Figure 4.9), which can occur when population differentiation is subtle and the MCMC therefore becomes trapped in local optima. For *Structure*, the highest posterior mean log-likelihood was achieved at $K = 1$ both with and without location priors (Supplementary Figure 4.10). The alternative method for assessing the optimal number of clusters in *Structure*, the Evanno method (Evanno *et al.* 2005), selected $K = 3$ with location priors and $K = 6$ without priors, consistent with the *K*-means pseudo-F statistic and BIC results, respectively. It should be noted that the *Structure* Evanno method does not allow $K = 1$ to be tested. For the three population scenario, suggested by both the *K*-means pseudo-F statistic and the *Structure* Evanno method with location priors, the following populations were defined: 1) the Ross Sea population (Cape Roget and Cape Washington colonies), 2) the Mawson Coast population (Fold Island and Auster colonies), and 3) all other colonies (Pointe Géologie, Amanda Bay, Gould Bay and Halley Bay). Under the three-cluster

scenario 32.2% of the total variance was explained (r^2 = 0.322). Upon visual inspection of the *Structure* plots for *K* = 2 to *K* = 6 (Figure 4.2), it is apparent that *K* = 4 is the finest level of genetic structure that can be discerned. At *K* = 5 and *K* = 6 there is no further resolution of population structure. In every scenario from *K* = 2 to *K* = 6 the Ross Sea population is clearly genetically distinct from all other colonies (Figure 4.2).

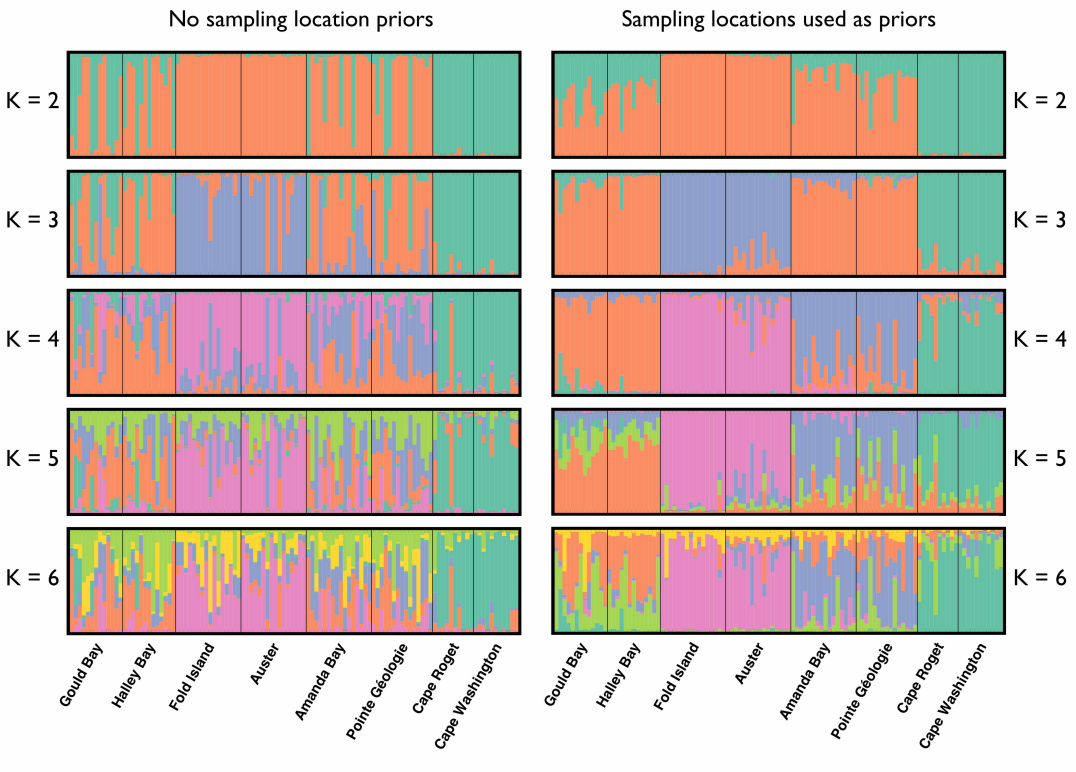


Figure 4.2. *Structure* results showing population structure from K = 2 to K = 6. The results are averaged for the ten replicate runs that were performed at each value of K. Each individual is a vertical bar, with the colours showing the proportion of ancestry assigned to each of the clusters.

Individuals (as opposed to populations) could not be separated into genetically distinct groups by a successive *K*-means procedure (find.clusters from the *adegenet* (Jombart & Ahmed 2011) package in R) as the minimum value for the BIC was at *K* = 1. This could be because the first few principal components of the principal components analysis (PCA) used to transform the data before *K*-means clustering were unable to explain much of the variation (Figure 4.3, inset). However Discriminant Analysis of Principal Components (DAPC)(Jombart *et al.* 2010) distinguished the Ross Sea colonies when individuals were grouped by their colony of origin and when individuals were grouped by geographic region (Figure 4.3).

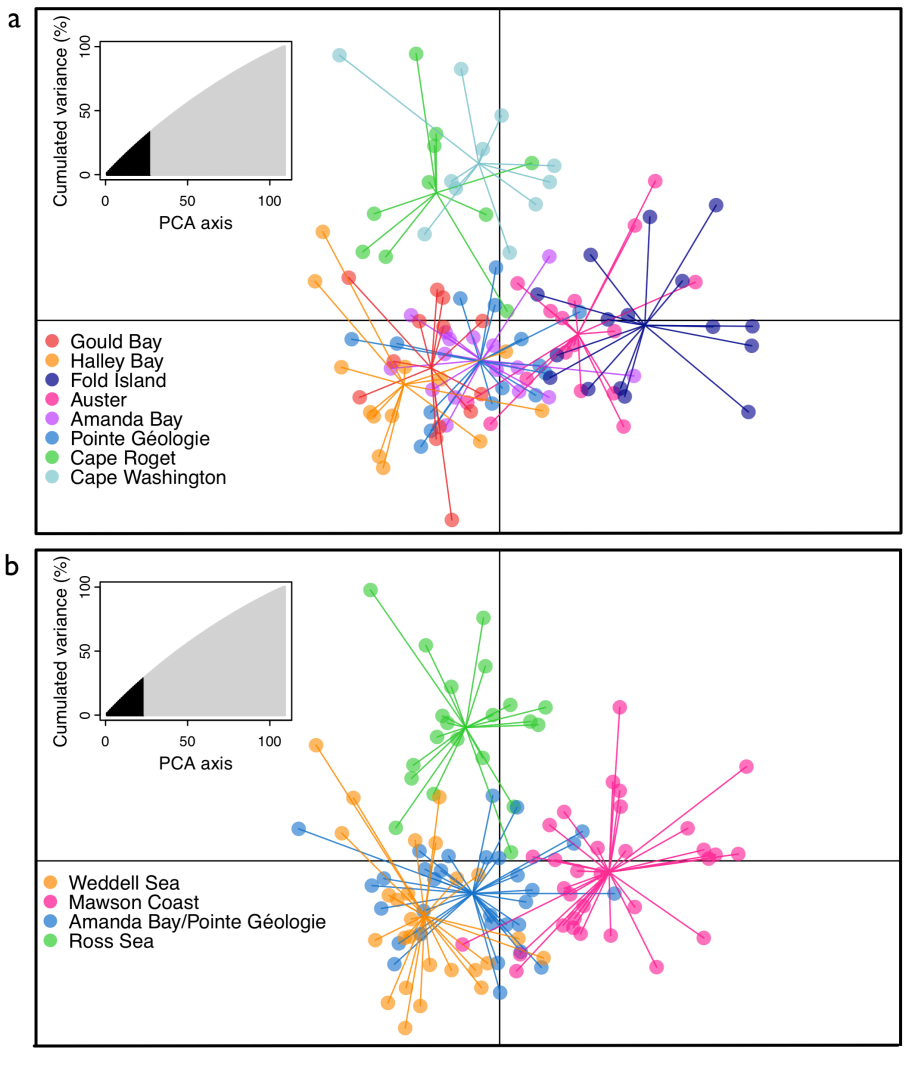


Figure 4.3. Results from Discriminant Analysis of Principal Components. Individuals were grouped by a) colony and b) major geographic region. The numbers of retained principal components are shown in black on the inset graphs and were a) 27 and b) 23.

Overall, our clustering results suggest the existence of at least four genetic populations of emperor penguins, each consisting of at least two sampled breeding colonies. The Ross Sea is the most differentiated population overall, followed by the Mawson Coast population, then the Weddell Sea and Amanda Bay/Pointe Géologie populations, which show the least amount of differentiation from one another. The selection of $K = 1$ as the most likely number of clusters by some of the methods indicates that differentiation among these four genetic populations is subtle.

4.3.3 Population differentiation

Pairwise genetic differentiation, as measured by Weir & Cockerham's F_{ST} estimator (Weir & Cockerham 1984), was statistically significant for 17 out of 28 pairs of colonies (Figure 4.4). The values of F_{ST} were small (range 0 – 0.006), consistent with the clustering results that indicated subtle differentiation. It should be noted that small values of F_{ST} are expected for datasets of biallelic SNPs with low minor allele frequencies, and in such cases small F_{ST} should not be interpreted as a lack of differentiation (Jakobsson *et al.* 2013). In this study we selected SNPs with minor allele frequencies > 0.01, averaging 0.078 across all SNPs, which is the likely cause of the small, yet statistically significant, F_{ST} values observed.

Of note was that within geographic regions (Weddell Sea, Mawson Coast and Ross Sea) there was no genetic differentiation (Figure 4.4). And despite the *ca*. 3,200 km between them, there was no differentiation between the two East Antarctic populations of Amanda Bay and Point Géologie (F_{ST} = -0.001, p = 0.813), also consistent with the clustering results. At large spatial scales there was significant genetic differentiation corresponding with the major geographic regions e.g. between colonies in the Ross Sea, colonies in the Weddell Sea, colonies from the Mawson Coast and colonies from East Antarctica (Figure 4.4). Every comparison in a pairwise F_{ST} analysis that grouped colonies into these four geographic regions was statistically significant, with F_{ST} values ranging from 0.002 to 0.005 (Figure 4.5). The Ross Sea was the most differentiated from all other populations. Overall, our pairwise F_{ST} analyses further support the conclusion that there are four genetic populations of emperor penguins, corresponding with major geographical regions.

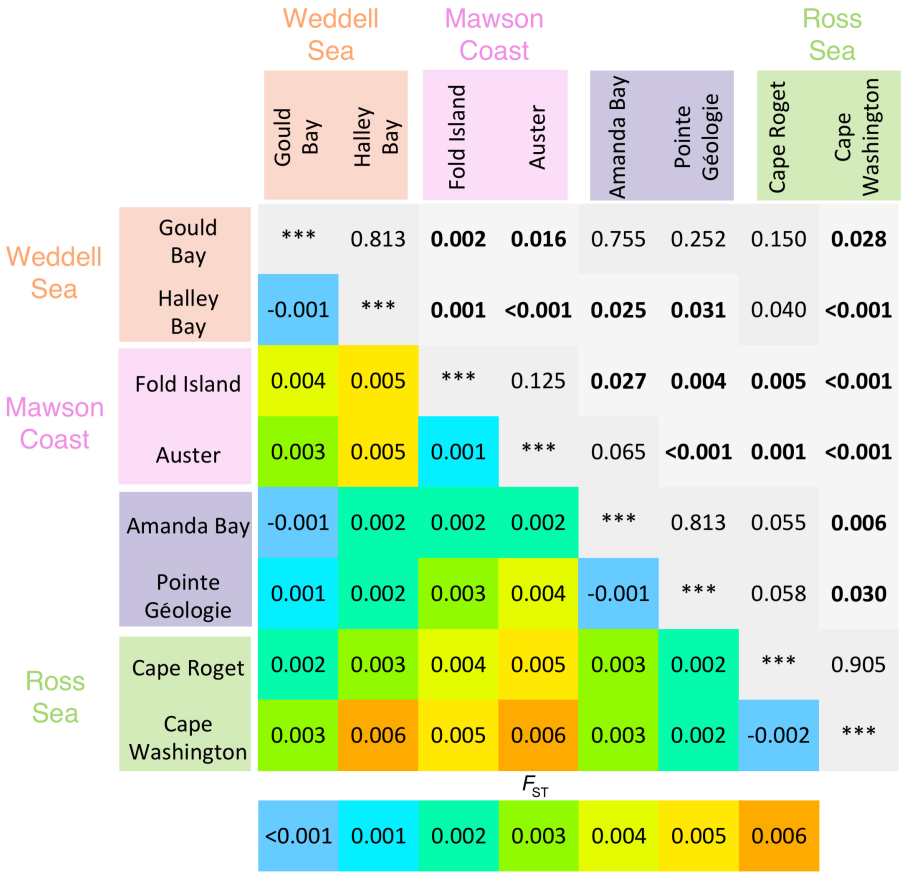


Figure 4.4. Genetic differentiation between all pairs of emperor penguin colonies. F_{ST} values are shown below the diagonal, with associated *p*-values above the diagonal. Significant *p*-values after SGoF+ correction for multiple tests are shown in bold.

	Weddell Sea	Mawson Coast	Amanda Bay & Pointe Géologie	Ross Sea
Weddell Sea	***	**<0.001**	**0.004**	**<0.001**
Mawson Coast	0.004	***	**<0.001**	**<0.001**
Amanda Bay & Pointe Géologie	0.002	0.003	***	**<0.001**
Ross Sea	0.005	0.005	0.004	***

Figure 4.5. Genetic differentiation between all major geographic regions. F_{ST} values are shown below the diagonal, with associated *p*-values above the diagonal.

4.3.4 Phylogeography

The posterior distribution of trees resulting from both of our SNAPP (Bryant *et al.* 2012) species-tree analyses formed a diffuse cloud (one is shown in Figure 4.6) indicating that there was no single, well-supported topology. The lack of a well-supported topology prevents us from determining the evolutionary relationships among the colonies, and whether, for example, any colonies were formed by a group of founders from another population. This further reflects the lack of strong genetic differentiation among colonies, but highlights the usefulness of the Bayesian approach implemented in BEAST 2 (Bouckaert *et al.* 2014), as the uncertainty over the tree topology would not have been obvious with a maximum likelihood approach.

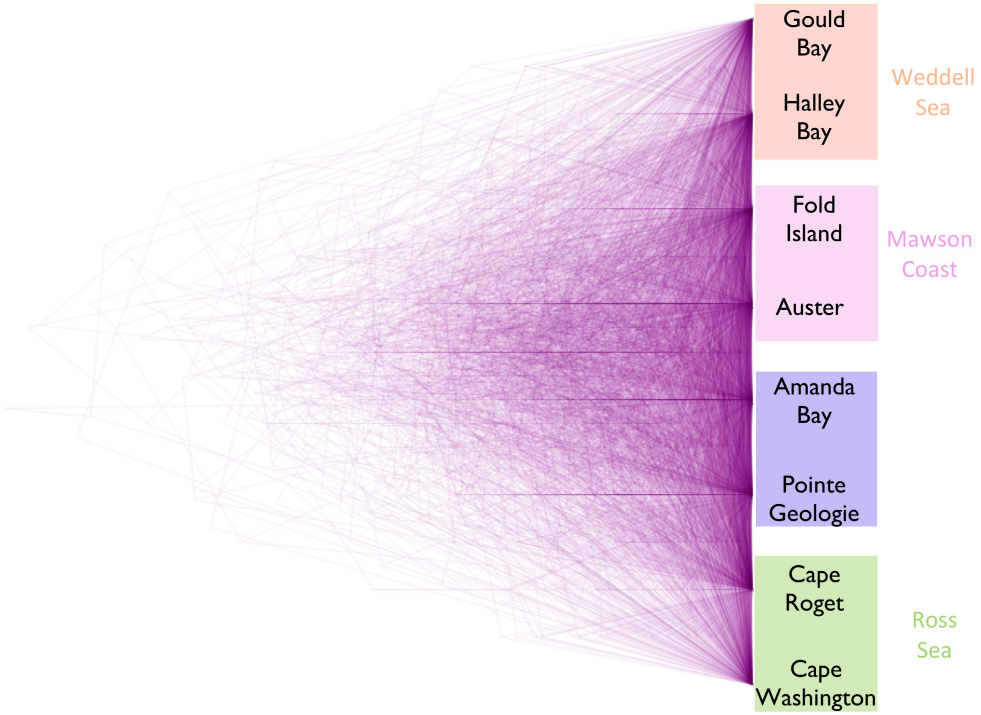


Figure 4.6. Cloudogram of the posterior distribution of trees resulting from our SNAPP species-tree analysis. There is no single, well-supported topology and so the phylogenies form a diffuse cloud.

4.4 Discussion

Our findings of low, yet statistically significant, genetic differentiation among emperor penguin colonies spanning more than 10,000 km of the Antarctic coastline refutes claims of panmixia in this species and provides evidence that emperor penguins in Antarctica are composed of at least four distinct metapopulations. These metapopulations are variously connected through spatial and temporal gene flow, whilst colonies within each appear to be panmictic because of high rates of migration.

Interestingly, some clustering analyses struggled to detect more than a single genetic cluster of emperor penguins, and if these results were not scrutinised further it might be easy to conclude that there is one globally panmictic population. However, scrutiny of data is crucial, because clustering methods are known to perform poorly at estimating the number of genetic populations when gene flow is high (Waples & Gaggiotti 2006), which we reasonably expect may be the case for emperor penguins (LaRue *et al.* 2015; Younger *et al.* 2015b). By examining multiple values of *K* we were able to ascertain a clear pattern of genetic structure among emperor penguin colonies that is further supported by our pairwise F_{ST} comparisons and DAPC. The four genetic populations of emperor penguins within our sampling range are the Ross Sea, Mawson Coast, Weddell Sea and East Antarctic Coast (Figure 4.7c). Over short distances, i.e. less than 600 km, colonies appear to be genetically connected. The two colonies in the Weddell Sea – Gould Bay and Halley Bay, located 550 km apart – are not genetically differentiated. Likewise, the two colonies in the Ross Sea, located 300 km apart, are genetically indistinguishable, as are the Fold Island and Auster colonies of the Mawson Coast, which are separated by only 190 km, as well as Amanda Bay and Auster, located 600 km apart. Where colonies are separated by more than 600 km the patterns of genetic differentiation are less predictable; for example, Amanda Bay penguins are not genetically differentiated from those *ca*. 3,200 km away at Pointe Géologie, whereas they are significantly differentiated from the Fold Island colony just 790 km away. We therefore urge that assumptions should not be made about the connectivity of emperor penguin colonies based on geographic proximity alone.

In our first study of emperor penguin genetic connectivity, based on mtDNA sequences, we found that penguins in the Ross Sea were distinct from those in East Antarctica and the Weddell Sea (Younger *et al.* 2015b), a finding that is supported here by genome-wide SNPs. However, the SNPs revealed further population subdivision within the East Antarctic and Weddell Sea regions, with three distinct genetic populations resolved among the six colonies studied, which were grouped as a single genetic population using mitochondrial DNA alone (Figure 4.7a,c). This indicates that large SNP datasets are superior to mtDNA for detecting fine-scale genetic differentiation among

colonies, as expected (Baird *et al.* 2008). Analyses of genetic connectivity within other penguin species, including Adélies (Clucas *et al.* 2014; Ritchie *et al.* 2004; Younger *et al.* 2015a) and chinstraps (Clucas *et al.* 2014; Freer *et al.* 2015) using either mtDNA or microsatellites, resolved similar levels of genetic differentiation as found in the mtDNA analysis of emperor penguins (Younger *et al.* 2015b). It is therefore possible that other penguin species have higher levels of genetic differentiation than reported to date, which could be detected using genome-wide SNPs.

A recent study by Cristofari *et al.* (2016) of emperor penguin genetic connectivity also used SNPs and concluded that emperor penguins are "a fully panmictic species" (Cristofari *et al.* 2016) (Figure 4.7b). Genetic or demographic connectivity can be described as a spectrum ranging from panmixia, in which mating is random (i.e. equally likely) amongst all pairs of individuals across the colonies studied, to complete isolation, in which there is no genetic exchange among colonies (Figure 4.8). Whilst Cristofari *et al.* (2016) proposed panmixia, Jenouvrier *et al.* (2014) used complete isolation to forecast population trends for the species under future climate change scenarios. Our genetic data, both here and in Younger *et al.* (2015b), support an intermediate scenario, as would be expected for most species in the natural world (Waples & Gaggiotti 2006). Specifically, the emperor penguin species is comprised of multiple metapopulations with some degree of connectivity among metapopulations and very high connectivity between sub-populations within each (Figure 4.8b,c).

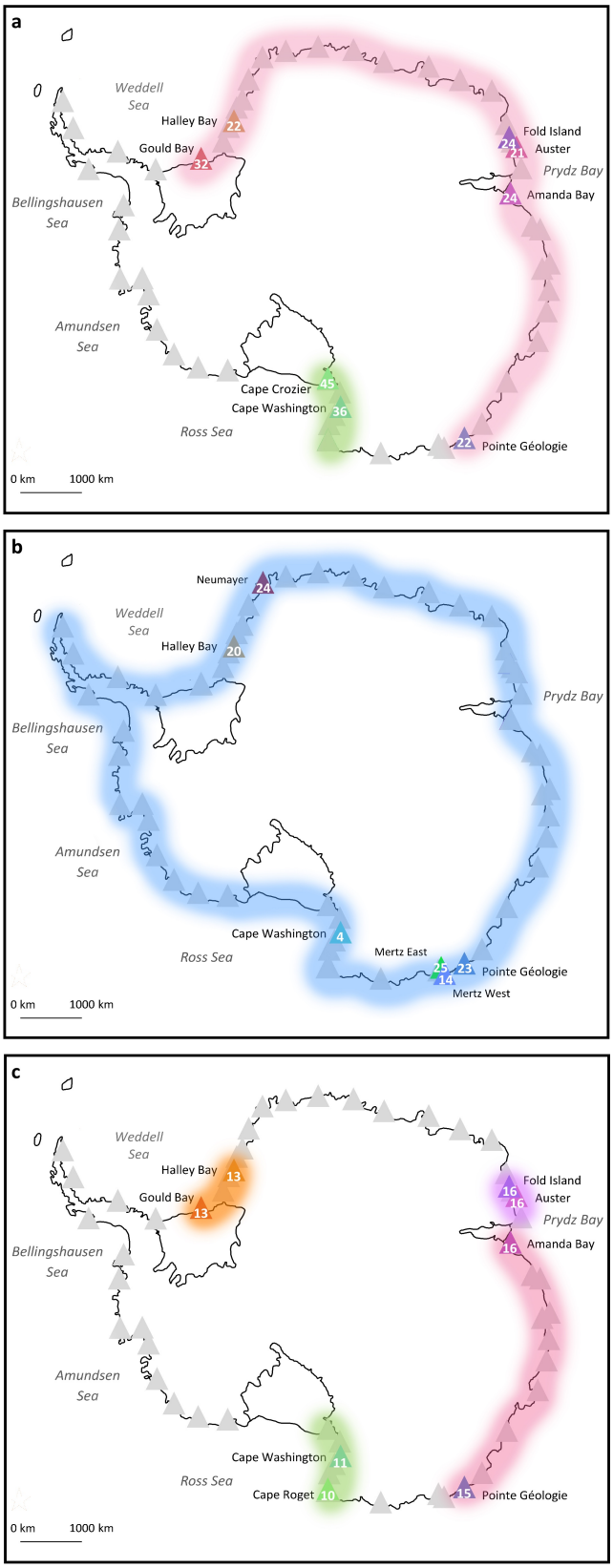


Figure 4.7. Emperor penguin population structures as determined by a) mitochondrial DNA (Younger *et al.* 2015b), b) SNP study of Cristofari *et al.* (2016), and c) this study. Coloured triangles indicate the colonies included in each study, with the number of individuals sampled at each colony indicated in white.

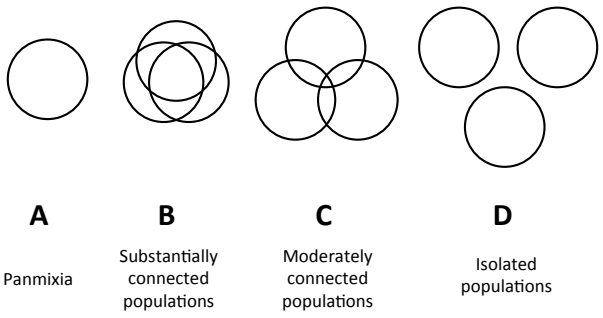


Figure 4.8. The population structure spectrum, after Waples and Gaggiotti (2006). Each circle represents an emperor penguin colony, with each group of circles representing varying levels of genetic connectivity among colonies; A is panmixia, B is substantial connectivity, C is moderate connectivity, D is complete independence.

The difference between our results and that of Cristofari *et al.* (2016) warrants interrogation. How might two similar data sets be so differently interpreted? We attest that the coalescent-based comparison by Cristofari *et al.* (2016) of different population structure scenarios using Migrate-n (Beerli & Palczewski 2010) with model ranking based on Bayes factors rejected the panmixia model and gave "clear support for a full-migration model with very high gene flow" (Cristofari *et al.* 2016), i.e. discrete populations with some degree of connectivity (Figure 8b,c). This is the equivalent to our results. Secondly, one of the authors' main arguments for panmixia of emperor penguins was the "very low" values of F_{ST} between pairs of colonies, ranging from 0.0080 between the proximate East and West Mertz colonies, to 0.0240 between Cape Washington and Pointe Géologie (Cristofari *et al.* 2016). However, as discussed above, biallelic SNP datasets with rare minor alleles are expected to yield small F_{ST} values (Jakobsson *et al.* 2013), and these are not evidence of panmixia. Cristofari *et al.* (2016) did not report the significance levels of their pairwise F_{ST} values, nor the *p*-value associated with their AMOVA test, so it is unknown whether the reported values represent differentiation or homogeneity. As a comparison, we observed F_{ST} values as low as 0.002 (between Pointe Géologie and Halley Bay, Amanda Bay and Halley Bay, and Amanda Bay and Fold Island) that represented statistically significant genetic differentiation at the $\alpha = 0.05$ level after SGoF+ correction for multiple tests. To place these findings in a broader context, a recent study by Toews *et al.* (2016) compared two species of parulid warblers, *Vermivora chrysoptera* and *Vermivora cyanoptera*, and found an F_{ST} value of 0.0045 based on 11.4 million SNPs. This F_{ST} value, a measure of differentiation between two distinct bird species (Toews *et al.* 2016), is smaller than the smallest F_{ST} value reported by Cristofari *et al.* (2016) for emperor penguins.

A key finding of our study is the genetic differentiation between the Ross Sea and colonies elsewhere. This confirms the results of our mtDNA study (Younger *et al.* 2015b), which found a

high degree of genetic differentiation between the Ross Sea colonies and those elsewhere in East Antarctica (Figure 4.7). The Ross Sea supports the highest density of breeding emperor penguins with an estimated 67,554 pairs in 2009, representing 28% of the global population (Fretwell *et al.* 2012). This abundance may be attributable to the presence of the Ross Sea polynya, which supports the most spatially expansive and predictable phytoplankton bloom in Antarctica, meaning that prey is particularly abundant in the Ross Sea compared to other regions of Antarctica (Smith & Gordon 1997). This could make dispersal of penguins out of the Ross Sea unfavourable, whilst potential immigrants may face steep competition for resources where colonies are already densely occupied. The genetic differentiation of the Ross Sea emperor penguins highlights the importance of the region in Antarctica. The abundance of emperor penguins in the Ross Sea is matched by the greatest density of Adélie penguins, estimated at 1.2 million breeding pairs (Lynch & LaRue 2014), and the existence of a highly divergent "Ross Sea" lineage in Adélie penguins (Ritchie *et al.* 2004). The region is also predicted to support growing populations of emperor and Adélie penguins under climate change (Cimino *et al.* 2016; Jenouvrier *et al.* 2014) and may have acted as a refuge for emperor penguins during the last glacial maximum (Younger *et al.* 2015b). The overall indication is that this region has been an important breeding habitat for penguins for tens of thousands of years and harbours unique genetic diversity that must be conserved to maximise the evolutionary and adaptive potential (D'Amen *et al.* 2013) of both species. The October 2016 announcement of a Ross Sea marine protected area by the Commission for the Conservation of Antarctic Marine Living Resources (CCAMLR) is a promising development that will protect the ecosystem from the expansion of krill and Antarctic toothfish fisheries for the next 35 years. However, long-term protection of the ecosystem is necessary, as it may become an important refuge for penguins under future climate change.

We note that our original finding of the genetic differentiation of the Ross Sea based on mtDNA (Younger *et al.* 2015b) has been criticised as being an erroneous result caused by pseudo-replication of haplotypes (Cristofari *et al.* 2016). We find that Cristofari *et al.* (2016)'s re-analysis and interpretation of our mtDNA data is invalid (see the Supplementary Information for details). They replaced a suggested false positive with a false negative result, because of low sample size and little power to detect weak differentiation. Based on their SNP dataset, Cristofari *et al.* (2016) reported that Ross Sea emperor penguins were panmictic with all other emperor penguins across Antarctica (Figure 4.7b). However, their study included only four individuals from the Ross Sea, a sample size that precludes any reasonable analysis of population structure. We have now conclusively validated the genetic differentiation of the Ross Sea based on robustly called genome-wide SNPs from blood samples rather than shed feathers, and recommend that the Ross

Sea emperor penguins be conservatively considered as a distinct management unit (Funk *et al.* 2012).

In order to conduct meaningful species risk assessments (Jenouvrier *et al.* 2014; Tavecchia *et al.* 2016), population monitoring and effective conservation strategies (Funk *et al.* 2012; Palsbøll *et al.* 2007), we would ideally translate our knowledge of the genetic patterns among colonies into a measure of demographic connectivity, defined as the extent to which a population's vital rates and growth are influenced by dispersal (Lowe & Allendorf 2010). To interpret the observed genetic patterns, we must consider two alternative scenarios. Firstly, the lack of genetic differentiation between some colonies could be the result of contemporary dispersal of individuals among breeding sites, suggesting that demographic linkage is a possibility. However, it is also possible that the observed genetic similarities could be the result of shared ancestry among now isolated colonies, which have simply not been isolated for a sufficient time to diverge genetically. Given the growing body of evidence that emperor penguins disperse widely as juveniles (Kooyman *et al.* 1996; Thiebot *et al.* 2013; Wienecke *et al.* 2010), exhibit dynamism with respect to their colony locations (Ancel *et al.* 2014; LaRue *et al.* 2015), and may temporarily emigrate (LaRue *et al.* 2015), we consider the former scenario, that migration is maintaining gene flow among populations, to be the most likely explanation for the genetic similarity found here.

An accurate estimate of the rate of migration among colonies would substantially improve our understanding of emperor penguin demographic connectivity. Unfortunately, the low differentiation observed among colonies precludes the estimation of migration rates with any degree of accuracy. BayesAss, which is typically used to estimate rates of migration over the past few generations to estimate contemporary demographic connectivity (Wilson & Rannala 2003), is unreliable where F_{ST} values are less than 0.05 (Faubet *et al.* 2007), a full order of magnitude greater than observed among emperor penguin colonies. Estimation of migration rates using F-statistics is also unreliable, because the assumptions underlying the island model (Wright 1931) that relates F_{ST} to the number of migrants are almost always violated in natural systems (Whitlock & McCauley 1999). Coalescent methods, such as Migrate-n (Beerli 2001), are used to infer rates of migration over evolutionary timescales, making them inappropriate for estimating contemporary demographic connectivity. The coalescent approach is also likely to be inaccurate when population differentiation is low and only a small number of loci can be used because of massive computational demands. It is also unclear what effect different SNP calling pipelines may have on estimates. Low frequency SNPs are believed to be important for estimating migration rates (P. Beerli *pers. comm.*), but are hard to distinguish from sequencing errors, meaning SNP data may not be suitable for this type of analysis. Furthermore, the coalescent method is contingent on accurate estimation of evolutionary rates for the loci in question, in order to convert the

mutation-scaled migration rate into the number of migrants, and it is not possible to accurately estimate evolutionary rates for RAD loci (Harvey & Brumfield 2015; Harvey *et al.* 2016). To compound these issues further, unsampled 'ghost' populations that exchange migrants with the studied populations have serious confounding effects on the estimation of migration rates, specifically, populations appear to be exchanging migrants when they are not (Beerli 2004; Slatkin 2005). This is almost certainly the case for emperor penguins, for which there are many unsampled colonies that lie between the colonies included in all the genetic studies to date (Figure 4.7). We note that Cristofari *et al.* (2016) reported migration rates among emperor penguin colonies based on coalescent methods, however, these estimates are likely to be affected by many of the issues described above and, at any rate, only describe migration over evolutionary timescales, rather than contemporary connectivity of colonies. Even if it were possible to generate an accurate estimate of the number of emperor penguin migrants, as yet there is no generalized framework for determining the level of migration necessary to maintain demographic linkage (Waples & Gaggiotti 2006) and even high rates of dispersal do not guarantee demographic interdependence (Lowe & Allendorf 2010). Therefore, it is currently impossible to determine the extent to which emperor penguin colonies are demographically linked (Palsbøll *et al.* 2007). Future studies could address this gap by combining genetic methods with data on movement behaviour, perhaps from capture-mark-recapture methods, local demographic rates and measures of the reproductive success of both immigrants and residents (Lowe & Allendorf 2010; Tavecchia *et al.* 2016). With these caveats acknowledged, we tentatively conclude that emperor penguin colonies are most likely not demographically isolated, nor are they likely to be linked continent-wide. Within our dataset we have identified four genetic populations of emperor penguins (Figure 4.7c) among eight study colonies, and future population forecasts (Jenouvrier *et al.* 2014) and monitoring schemes should take this structure into account.

It is important to note that there are many emperor penguin colonies that have not been included in genetic studies to date. Including these colonies in future work may aid in understanding the true links among emperor penguin populations. It is possible that some unsampled colonies may be isolated by unidentified dispersal barriers (as observed for the Ross Sea), or may have acted as refugia in the past, or contain pockets of unique genetic variation and evolutionary potential (D'Amen *et al.* 2013).

Here we have assessed the genetic population structure among emperor penguin colonies and synthesised the available information on the patterns of connectivity among colonies. The difference between complete panmixia and distinct metapopulations with gene flow may seem like a subtle distinction, but it is a distinction that has important implications for our understanding of emperor penguin population dynamics, behaviour, ecology, adaptability, range-

shift potential, population trajectories and, most importantly, risk assessments and conservation planning. As a research community it is crucial that we clearly delimit the extent of our knowledge, and that the subtleties and limitations of population genetic studies are fully communicated in our scientific reporting, so that managers and conservation planners can make properly informed decisions. We find evidence for distinct metapopulations within emperor penguins, which are connected by some level of migration. Until demographic connectivity can be determined, a precautionary approach would warrant that these metapopulations are considered separate management units, in particular the Ross Sea metapopulation, which is the most distinct.

4.5 Methods

4.5.1 Sampling, DNA extraction and RADSeq

Muscle biopsies from the pectoral region of chick carcasses were collected at Fold Island in 2010 and Pointe Géologie in 2010, and from Amanda Bay in 2012 and 2013 (see Figure 4.1 for colony locations). Skin biopsies from the foot or back were collected from dead adults and chicks at Halley Bay in 2012. Whole chick carcasses were collected at Auster in 1993 and 1994. Blood samples were collected from Gould Bay in 2013 and from the Ross Sea colonies in 1992 and 1993. For detailed methods of blood sampling see Younger *et al.* (2015b). Sampling was conducted under permits from the UK Foreign and Commonwealth Office, the US National Science Foundation and the Australian Antarctic Division. Ethical approval of each of these permits was granted by the permitting institution and additional ethical approvals were received from the University of Oxford and British Antarctic Survey.

Genomic DNA (gDNA) was extracted from 10 – 16 individuals per colony using QIAGEN DNEasy blood and tissue kits with the following modifications to the digestion step: 30 µL proteinase K was added to blood samples and the incubation time was extended to 3 hrs; 40 µL proteinase K was added to tissue samples (plus an additional 10 µL 1 M dithiothreitol for skin samples from the foot) and the incubation step was extended to 32 hrs. RNA contamination was reduced by treating samples with 1 µL RNase A (QIAGEN) or 1 µL Riboshredder (Epicentre). DNA concentration was measured with a Qubit (ThermoFisher Scientific). The presence of high molecular weight DNA was confirmed on a 1% gel and DNA contamination was measured using a Nanodrop (ThermoFisher Scientific).

RADseq, including library preparation, was performed by Edinburgh Genomics, University of Edinburgh (https://genomics.ed.ac.uk/). The method used follows Gonen *et al.* (2014) after Etter

et al. (2011). For each individual, 250 ng of gDNA was digested with the SbfI-HF (NEB) restriction enzyme, followed by ligation to barcoded P1 adapters. The individually barcoded samples were then pooled into multiplexed libraries, and sheared into fragments of 300-400 bp. Gel electrophoresis was used to size-select fragments. The NEB Quick Blunting Kit was used to blunt the libraries, followed by A-tailing and ligation to P2 adapters (IDT). Yields were increased via enrichment PCR, followed by purification with Ampure beads. The enriched libraries were checked for size and concentration using Qubit and a qPCR assay. The libraries were sequenced using 125 base paired-end reads on an Illumina HiSeq 2500 in high output mode (v4 chemistry).

4.5.2 Bioinformatics, SNP calling and filtering

Overall read quality and the presence of adapters were assessed using FastQC. Reads were demultiplexed, trimmed and cleaned using process_radtags from the Stacks software pipeline v1.35 (Catchen *et al.* 2011; Catchen *et al.* 2013). The reads were truncated to 113 bp to exclude four terminal bases as sequence quality can decrease at the end of the read. Read pairs where either of the pair had a low quality score, uncalled bases and/or a barcode or cut-site with more than one mismatch were excluded from further analysis. The remaining paired reads were then aligned to the emperor penguin reference genome (http://gigadb.org/dataset/100005) using bwa-mem (Li 2013). Terminal alignments were prevented using a clipping penalty of 100. Reads with multiple alignments, more than 5 mismatches and/or more than 2 indels (Catchen *et al.* 2013) were removed using a custom python script (filter.py available from the Dryad Digital Repository http://dx.doi.org/10.5061/dryad.7c0q8). PCR duplicates were removed using Picardtools (http://broadinstitute.github.io/picard).

SNP calling was performed using the Stacks pipeline (*pstacks – cstacks – sstacks – rxstacks – cstacks – sstacks – populations*) with the following options: in *pstacks* we specified a minimum depth of six reads mapping to the same location (-m 6) and employed the bounded SNP model with a significance level of $\alpha = 0.05$, an upper bound of 0.1 and a lower bound of 0.0041 (the highest sequencing error rate recorded by phiX spikes in the sequencing lanes, --model_type bounded --bound_low 0.0041 --bound_high 0.1 --alpha 0.05); all individuals were used to build the catalogue in *cstacks*; in *rxstacks* confounded loci (loci with too many haplotypes to be biologically possible such as from repetitive regions or paralogous loci) were removed with a conservative confidence limit of 0.25 (--conf_lim 0.25), excess haplotypes were removed from individuals (--prune_haplo), and loci with a mean log likelihood < -10 were removed from further analysis (--ln_lim -10). In the *populations* module, we removed any SNPs with a minor allele frequency (MAF) < 0.01, which are likely to be the result of sequencing errors (--min_maf 0.01); loci with a heterozygosity > 0.5, which could be the result of paralogous sections of the genome

being merged into a single locus (--max_obs_het 0.5); a single SNP per RADtag was chosen at random, to remove tightly linked SNPs from the dataset (--write_random_snp); a locus must have been present in all populations to be included in the final dataset (-p 8); and a locus must have been genotyped in at least 80% of individuals per population to be included (-r 0.8). We further checked that no SNPs had a mean coverage greater than 100X using vcftools v0.1.13 (Danecek *et al.* 2011), to avoid SNPs from repetitive regions of the genome, and checked that none were out of Hardy Weinberg equilibrium (HWE) in more than 50% of the populations when $p < 0.01$ using the *adegenet* (Jombart & Ahmed 2011) package in R and vcftools. We used PGDSpider v2.0.8.2 (Lischer & Excoffier 2012b) to convert between file types for use in population genomics analyses.

4.5.3 Outlier detection

Loci under directional or balancing selection violate the assumption of neutrality that is used in most population genetic analyses. We used BayeScan v2.1 (Foll & Gaggiotti 2008), which employs a Bayesian F_{ST} outlier test, to identify loci that might be under selection and hence discard them. It has been demonstrated (Lotterhos & Whitlock 2014) that BayeScan is powerful at detecting loci genuinely under selection in a range of demographic scenarios, but with an accompanying high false-positive rate. As our aim was to achieve a neutral set of loci, a high false positive rate was not a concern. We conservatively set the prior odds of neutrality parameter to five (for every five loci one is expected to be under selection) to ensure all loci under selection were detected. We deemed q-values of < 0.1 to be a significant result, so we can expect one in ten loci to be a false-positive neutral locus that is wrongly discarded (Lotterhos & Whitlock 2014; Storey & Tibshirani 2003).

4.5.4 Clustering of individuals and populations

We visualized population structure with a Discriminant Analysis of Principal Components (DAPC) (Jombart *et al.* 2010) using the *adegenet* (Jombart & Ahmed 2011) package in R. This method can benefit from groups being defined by successive *K*-means clustering (*find.clusters* function), to avoid *a priori* assignment of individuals to groups based on sampling locations. However, the optimal number of clusters suggested by *K*-means was one, and so we ran DAPC on the individuals when they were grouped by colony and by geographic region (Ross Sea, Mawson Coast, Weddell Sea, and Amanda Bay/Pointe Géologie). In each case, the optimal number of principal components to retain was found by running *optim.a.score* 20 times and taking the average. As well as clustering individuals, we clustered populations using the Analysis of Molecular Variance (AMOVA) based method of *K*-means clustering in Genodive (Meirmans 2012) using simulated annealing. We compared results with the BIC and pseudo-F statistic (Caliński & Harabasz 1974).

The BIC is useful for detecting whether there is any structure ($K > 1$ rather than $K = 1$) because the pseudo-F statistic cannot be defined for $K = 1$, but the pseudo-F statistic is known to perform better under random-mating (Meirmans 2012). Finally, individuals were clustered using the Bayesian clustering algorithm in the program *Structure* v2.3.4 (Pritchard *et al.* 2000) which estimates the proportion of ancestry each individual shares with ancestral populations, effectively showing their assignment probability to each cluster. From the results of our population differentiation analysis, which showed that differentiation was low, the admixture model with correlated allele frequencies was deemed the most appropriate model. We ran the program initially with $K = 1$, for 100,000 generations, discarding the first 50,000 as burnin, allowing lambda to vary (inferlamda = 1). From this initial run lambda was estimated to be 0.40, so we set lambda to 0.40 for all subsequent runs. We then varied K from one to eight, running each value of K ten times with a random seed, for 150,000 generations, discarding the first 50,000 generations of each run as burnin. We did this both with and without sampling locations supplied as prior information. Supplying sampling locations can be useful for detecting structure when it is weak. Structure Harvester web v0.6.94 (Earl 2012) was used to compare runs with different values of K using the Evanno method (Evanno *et al.* 2005) and to prepare files for CLUMPP (Jakobsson & Rosenberg 2007). We used CLUMPP to check for multimodality between the replicate runs at each value of K and to average across the ten replicates. The individual assignments to ancestral populations were plotted using distruct v1.1 (Rosenberg 2004).

4.5.5 Population differentiation

We calculated the Weir and Cockerham (1984) unbiased estimator of F_{ST} between all pairs of colonies, and between the geographic regions (Ross Sea, Mawson Coast, Weddell Sea and Amanda Bay/Pointe Géologie), using Genodive v2.0b27 (Meirmans & Van Tienderen 2004). Significance was estimated using 5,000 permutations of the data, and the significance level was adjusted for multiple testing using Sequential Goodness of Fit (SGoF+)(Carvajal-Rodriguez & de Uña-Alvarez 2011).

4.5.6 Phylogeography

To estimate the evolutionary relationships among the colonies and the order in which they were founded, we used a species tree approach in the SNAPP add-on (Bryant *et al.* 2012) to BEAST v.2.4.0 (Bouckaert *et al.* 2014). SNAPP uses a coalescent approach with unlinked biallelic markers (like SNPs) to infer species trees. It is highly computationally demanding, hence we selected two random individuals (i.e. four haplotypes) per colony to include in the analysis, and we repeated the analysis twice, with different individuals, to ensure the results were replicable. Within the

reduced dataset, any loci that were no longer polymorphic were removed, leaving datasets of 3,221 and 3,237 SNPs. We calculated the mutation rates (*u* and *v*) from the dataset rather than estimating them as part of the MCMC. The MCMC was run for 5 million generations with the first 10% of generations discarded as burnin. We checked for convergence using Tracer v1.6 (Rambaut & Drummond 2007) and found ESSs > 4,000, which is more than sufficient. We also inspected the posterior distributions of parameters for convergence. The advantage of using a Bayesian method as implemented in BEAST is that the uncertainty in the parameter of interest, here the tree, is estimated via the posterior distribution. We used DensiTree v2.0.1 to visualize the entire posterior distribution of trees as a cloudogram.

4.6 Supplementary Information

Cristofari *et al.* (2016) suggested that we had sampled multiple shed feathers from the same bird in the Ross Sea. Our mtDNA dataset contained a large number of heteroplasmic sites in the hypervariable region that we called as ambiguous sites (Younger *et al.* 2015b). Cristofari *et al.* (2016) claimed that "no reliable and controlled model can account for diploid sites in mitochondrial DNA" and hence re-analysed our data masking all sites with ambiguous base calls, shortening the sequence length by over 100 base pairs and removing many of the most variant sites from the alignment. Millar *et al.* (2008) reported heteroplasmic sites in the Adélie penguin mtDNA hypervariable region, which resulted from mutations in the germline that persisted in a heteroplasmic state across generations. That is, more than one mitochondrial genome was inherited from mother to offspring. Therefore, by masking all ambiguous sites in the emperor penguin hypervariable region Cristofari et al., (2016) excluded the most highly variable (and hence most informative) sites, and it is therefore unsurprising that they then recorded a high percentage of duplicate haplotypes. Notably, following the masking procedure, repetitive haplotypes were also recorded in colonies where pectoral biopsies were taken directly from carcasses, rather than shed feathers. For example, one haplotype was observed for seven Pointe Géologie penguins (that were pectorally biopsied), one Cape Crozier penguin and two Cape Washington penguins. We find that their re-analysis and interpretation of our data is invalid.

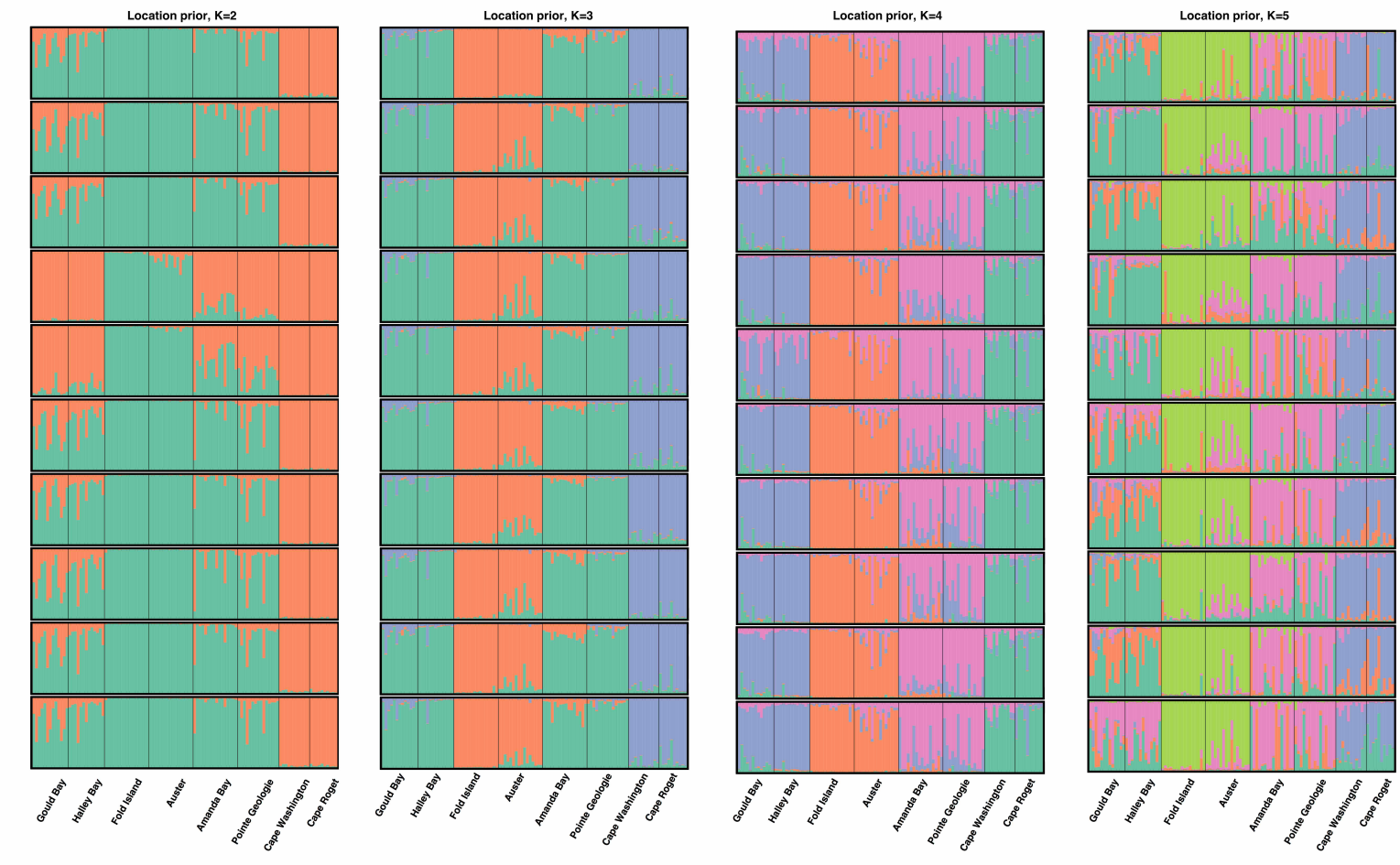


Figure 4.9. Results from ten independent runs of the Bayesian clustering algorithm *Structure* for K = 2 to K = 5 using sampling locations as prior information. The discrepancies between the replicate runs show that genuine multimodality was an issue, which often occurs when population differentiation is subtle.

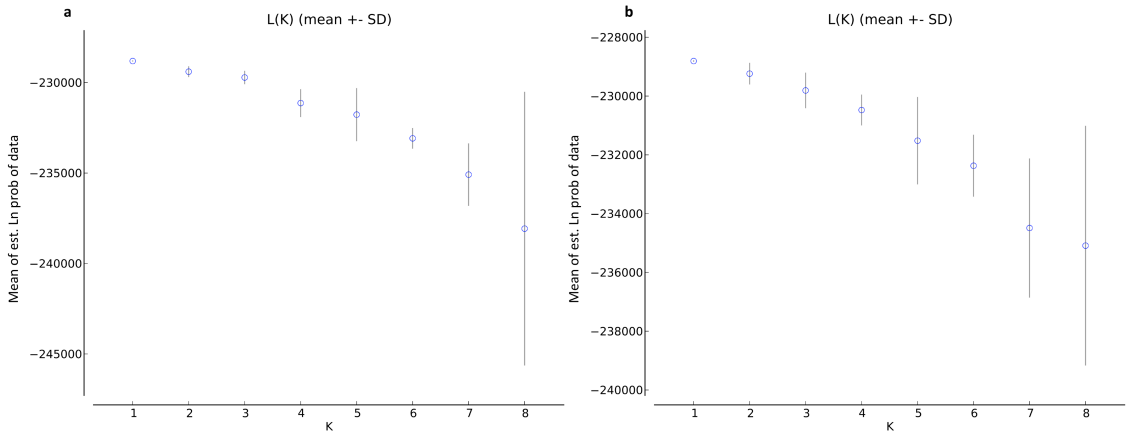


Figure 4.10. Estimated log probability of the data Pr($X|K$) for each value of K after repeated *Structure* runs with (a) sampling locations used as prior information and (b) without sampling locations used as priors. The highest log probability of the data is achieved at $K = 1$ in both cases.

Chapter 5: Comparative population genomics in Antarctica: ecological & evolutionary factors driving patterns of intraspecific genetic variation in Antarctic and sub-Antarctic penguins[4]

5.1 Abstract

Climate change is predicted to have damaging impacts on Antarctic penguin populations, although it is expected to have heterogeneous effects around the continent. So far, forecasts have failed to factor-in penguins' potential to track changes in habitat through dispersal. The key to understanding the likely evolutionary responses of species to large-scale habitat change lies in their patterns of intraspecific genetic variation, which can inform us of the boundaries to breeding populations and barriers to dispersal. Using comparative population genomics among penguins varying in their life-history traits and occupying habitats that vary temporally and spatially in both availability and quality, we identify the factors that have shaped intraspecific genetic variation. We present the population genomic structure of five penguin species, assessed using robust SNP datasets generated through RADSeq. We find that the pelagic lifestyle of emperor, king, chinstrap and Adélie penguins facilitates gene flow over thousands of kilometres. In contrast, the coastal lifestyle of the gentoo penguin restricts gene flow and leads to high levels of genetic differentiation between colonies, resulting in ongoing allopatric speciation requiring taxonomic revision. We also find evidence that natal philopatry may restrict gene flow between gentoo penguin colonies and that the Polar Front acts as an additional barrier to dispersal in both gentoo and king penguins. We find little evidence that the relative ephemerality of habitat or temporal variation in habitat accessibility and prey availability is responsible for the patterns of intraspecific genetic variation across the five species. Instead, their different ecological niches in terms of a pelagic or coastal lifestyle appear to govern their dispersal and hence evolutionary trajectories.

[4] This chapter is formatted for the journal *Proceedings of the National Academy of Sciences of the United States of America*.

5.2 Introduction

During the last 60 million years periods of global warming and cooling have had dramatic impacts on penguin evolution (Clarke *et al.* 2007; Slack *et al.* 2006; Zachos *et al.* 2001). Extant penguins have persisted in Antarctica throughout multiple glacial cycles, particularly in the last 2.58 million years, and appear to be resilient in the face of change. However, the Anthropocene (Lewis & Maslin 2015) will be characterized by changes outside of natural variability (IPCC 2013) in the terrestrial and marine realms, the cryosphere, and the atmosphere, contributing to the sixth mass extinction event (Barnosky *et al.* 2011). As we enter this new epoch, the evolutionary pressures acting on species will change. Fossil studies have elucidated the evolution of penguin species and clades over time, but the key to understanding their evolutionary responses currently lies in investigations of patterns of genetic variation, and the factors that impede or drive dispersal.

The dispersal of individuals has profound effects on populations. From an ecological perspective, dispersal can determine the distribution and abundance of individuals among populations, rescue declining populations from extinction and (re)colonise patches of suitable habitat when they become available. From a genetic perspective, dispersal can facilitate gene flow, and thus affect the evolutionary trajectory of species through, for example, the dispersal of adaptive genes or by preventing populations from diverging through genetic drift. For successful biodiversity conservation, understanding dispersal and the mechanisms behind dispersal decisions in response to habitat variability is key (Doerr *et al.* 2011). This is especially true in light of climate change, which may increase habitat variability in the future (IPCC 2013). Studying dispersal not only allows us to distinguish breeding populations and define management units (Moritz 1994), therefore informing the scale at which population monitoring should be conducted, but will also allow accurate predictions of local or global extinction risk in response to habitat change, as dispersal can decouple the relationship between population growth rates and demographic parameters that are sensitive to climate (Tavecchia *et al.* 2016). Finally, an understanding of dispersal can allow effective conservation strategies to be implemented, by predicting the impact of the intervention using well-defined population limits (e.g. Robertson and Gemmell (2004)).

Many factors are known to influence patterns of dispersal in seabirds (Friesen 2015; Friesen *et al.* 2007a) and several of these may be relevant to Southern Ocean penguins. Oceanographic fronts, which form the boundaries between water masses that vary in their physical and hence biological characteristics, are known to act as barriers to dispersal in some seabirds (Friesen 2015). Fronts may impede gene flow by simply deterring dispersal, or by reducing the fitness of immigrants that are not adapted to the foreign conditions (Friesen 2015). The at-sea ranges of seabirds during both the breeding and non-breeding seasons, as well as the juvenile phase, are also known to

contribute to the dispersal patterns of seabirds (Friesen 2015; Friesen *et al.* 2007a). Seabirds with large at-sea ranges have greater opportunity to come into contact with individuals from other colonies or to prospect other breeding sites, both of which may facilitate dispersal. Conversely, seabirds that stay in close proximity to their colonies throughout the year and forage inshore are far less likely to disperse among colonies than those that range widely (Friesen *et al.* 2007a). The availability of quality breeding habitat has great potential to influence dispersal patterns. Breeding habitat that is temporally variable in either quality or availability should select for phenotypic plasticity and dispersal (Johnson & Gaines 1990). Conversely, stable breeding habitat may lead to philopatry and local adaptation, and hence low dispersal. Geographically continuous breeding habitat should also facilitate dispersal, whereas highly disjunct breeding distributions, such as those spanning isolated archipelagos, could be expected to impede dispersal. Because the Southern Ocean is a highly heterogeneous environment both spatially and temporally, breeding habitat quality, continuity and ephemerality are therefore likely to play a large role in the patterns of dispersal in penguins.

Theoretical models have shown that when habitat quality is temporally variable, selection favours dispersal, as it allows individuals to track changing resources (Johnson & Gaines 1990), even if there is a cost to dispersal (Levin *et al.* 1984). Individuals within dispersive species, occupying temporally variable habitat, are still faced with the decision of when to disperse. One solution is to base their decision on previous reproductive success, such that successfully raising offspring promotes philopatry; the so-called win-stay:lose-switch rule, often seen amongst birds (Shields et al 1988, Schmidt 2004, Gavin & Bollinger 1988, Bollinger & Gavin 1989, Haas 1998, Blancher & Robertson 1985). Temporal variability in patch quality would therefore lead to episodic dispersal when patch quality declines, and periods of philopatric behaviour when habitat quality remains high.

Temporal variability in habitat is characteristic of the niches occupied by penguins breeding at high-latitudes around Antarctica. For example, the minimum summer sea ice extent in Antarctica has varied by as much as 30% around the 1981 - 2010 mean, since satellite records began (NSIDC, https://nsidc.org/). Inter-annual to multi-decadal variability in the climate dramatically affects the marine ecosystem via variability in sea surface temperatures and sea ice extent (Jones *et al.* 2016), which leads to fluctuations in primary productivity (Arrigo *et al.* 1998; Comiso *et al.* 1993; Smith Jr *et al.* 1988). Variability arises from the Southern Annular mode (SAM), the El Nino/ Southern Oscillation (ENSO) (Ding *et al.* 2011b; Thompson & Wallace 2000) and regional variability such as the Amundsen-Sea Low (Holland & Kwok 2012). In addition, interconnections between changes in the atmospheric circulation, oceanic circulation, and sea ice extent create feedback loops (Yuan 2004) that can lead to large-scale perturbations in the abiotic environment

from year to year. Local-scale variability in fast-ice extent (sea ice which is held "fast" to shorelines, icebergs, or ice fronts) is also driven by wind direction and the frequency of storms, the distribution of grounded icebergs, and the extent of grounded ice sheets and floating ice shelves (Massom *et al.* 2010), which are dynamic themselves. Polynyas, regions of the ocean which are kept free of sea ice due to wind or the upwelling of warm water, also vary inter- and intra-annually depending on circulation patterns and the distribution of sea ice and icebergs surrounding them (Massom *et al.* 2001). Therefore, the distribution of sea ice versus open ocean around Antarctica shows a high degree of inter-annual variability.

Changes in sea ice extent and thickness have been implicated in the formation, movement and extinction of colonies of both emperor (*Aptenodytes forsteri*) and Adélie penguins (*Pygoscelis adeliae*) over the last few decades (LaRue *et al.* 2013; LaRue *et al.* 2015; Lynch & LaRue 2014; Southwell & Emmerson 2013; Southwell *et al.* 2016; Trathan *et al.* 2011) and over millenia (Baroni & Orombelli 1994; Emslie & Patterson 2007; Younger *et al.* 2015a; Younger *et al.* 2015b). Both species breed around the coastline of Antarctica (Figure 5.1) and are sensitive to changes in sea ice extent, yet they have very different ecologies: the Adélie penguin breeds during the summer on ice-free bare rock, whilst the emperor penguin breeds during the winter, mainly on fast-ice. Variability in sea ice extent and primary productivity leads to fluctuations in krill (*Euphausia superba* and *Euphausia crystallorophias*) and Antarctic silverfish (*Plueragramma antarctica*) (Atkinson *et al.* 2004; Ducklow *et al.* 2007; Loeb *et al.* 1997), which are key prey items. Little is known about the impact of changes in prey biomass on emperor penguins, although starvation was suggested as the cause of one emperor penguin colony decline during a period of high sea surface temperatures (Barbraud & Weimerskirch 2001). Sea ice associated declines in krill biomass has been shown to affect foraging effort, breeding success, juvenile recruitment and population growth rates in Adélie penguins (Fraser & Hofmann 2003; Hinke *et al.* 2007; Lynch *et al.* 2012; Lynnes *et al.* 2004; Wilson *et al.* 2001). Sea ice extent can also affect breeding success in both species by determining the distance they must travel from their colonies to open water, with years with very large sea ice extents depressing breeding success (Massom *et al.* 2009; Shepherd *et al.* 2005).

Emperor penguins rely on a delicate optimal sea ice extent where they have sufficient access to the ocean to forage, yet the ice is stable enough to breed upon (Ainley *et al.* 2010; Massom *et al.* 2009). If fast-ice forms too late or its distribution alters, colonies have been observed to relocate, possibly to sub-optimal locations (Ancel *et al.* 2014; Fretwell *et al.* 2014; LaRue *et al.* 2015); whilst if it breaks out too early, chicks may not survive (Barbraud & Weimerskirch 2001). Therefore, variability in sea ice extent as a result of inter-annual to multi-decadal changes in atmospheric circulation creates a high degree of temporal variability in habitat quality for both Adélie and

emperor penguins, but especially so for emperor penguins. Under climate change, habitat quality is expected to decline significantly for both species, but with regional heterogeneity (Cimino *et al.* 2016; Jenouvrier *et al.* 2014).

Variability in high-latitude atmospheric circulation also affects the islands in the Scotia Arc and sub-Antarctic (Holton *et al.* 1989; Murphy *et al.* 2007). These islands can be divided into the "maritime-Antarctic" islands, which include the islands of the Scotia Arc as far north as the South Sandwich Islands, and the "sub-Antarctic" islands which all lie above 55 °S near to the Polar Front (Figure 5.1). Inter-annual variability in climate occurs in all regions of the Southern Ocean, as anomalies are propagated around the system (Holton *et al.* 1989; Murphy *et al.* 2007), but the amplitude of the variability diminishes away from the Scotia Arc and at lower latitudes, particularly above the Polar Front (Inchausti *et al.* 2003; Nevoux *et al.* 2010; Smith 2002). As a result, the amplitude of the temporal variation in habitat quality for penguins also varies spatially. For example, years with late sea ice break-out around southern maritime Antarctic islands affected the numbers and reproductive success of breeding chinstrap penguins (Lishman 1985; Rombolá *et al.* 2003; Trathan *et al.* 1996), but this may not be an issue on islands further north where ice always breaks out before breeding is initiated. In addition, penguins on these islands breed on ice-free ground, so the habitat is predictable, and the window of suitable conditions that defines the breeding season is longer. Therefore breeding phenology is less constrained at lower latitudes (Black 2016) allowing the king penguin (*Aptenodytes patagonicus*), for example, to lay over four months (Bost *et al.* 2013). The combination of a less constrained breeding season and more predictable habitat means that disturbance due to environmental variability is likely diminished, and temporal variability in breeding success likely becomes more dominated by changes in prey availability (Bost *et al.* 2015; Reid & Croxall 2001).

The different distributions of Adélie, chinstrap (*Pygoscelis antarctica)*, gentoo (*Pygoscelis papua*) and king penguins across the maritime- and sub-Antarctic (Figure 5.1) reflect their different ecologies. The *Pygoscelis* penguins (gentoos, chinstraps and Adélies) all require ice-free ground to nest on, but have different tolerances to sea ice (Ainley *et al.* 1983; Williams 1995). For example, the Adélie penguin, as mentioned, is a high-latitude specialist closely associated with sea ice, and only occurs as far north as the maximum extent of winter sea ice in the maritime-Antarctic and on Bouvet Island. The chinstrap penguin does not tolerate dense sea ice, only breeding north of 64 °S on the Antarctic Peninsula, and with the largest colonies found in the maritime-Antarctic region of the Scotia Arc (Trivelpiece & Trivelpiece 2013). The gentoo penguin, in the broadest sense not taking putative subspecies into account, breeds on nearly all maritime- and sub-Antarctic islands, and in addition on the temperate Falkland Islands (Islas Malvinas), and shows a high degree of flexibility in its diet and habitat requirements, although it is a strictly coastal forager, unlike all the

other species studied here (Borboroglu & Boersma 2013). Finally, the king penguin, congener of the emperor penguin, is a sub-Antarctic specialist. Similar to the emperor penguin, individuals incubate their eggs and chicks on their feet and do not build nests, but king penguins require ice-free ground to breed upon (Bost *et al.* 2013). The king penguin breeds on sub-Antarctic islands close to the Polar Front, which forms the main foraging site for most colonies. It has a varied diet of fish, squid and krill and as a result may be less sensitive to variation in krill availability than other species (Reid & Croxall 2001).

Antarctic and sub-Antarctic penguins present a significant logistical challenge for studying dispersal, as the vast majority of colonies are in remote locations, yet they are a major component of the marine ecosystem, likely representing *ca.* 80% of the avian biomass in the Southern Ocean (Williams 1995). Banding studies initially suggested a high degree of philopatry in many species (Ainley *et al.* 1983; Williams 1995) and thus forecasts of extinction risk have not considered the potential buffering effect of dispersal (Jenouvrier *et al.* 2009; Jenouvrier *et al.* 2014). However recent genetic analyses (Clucas *et al.* 2016; Younger *et al.* 2015b), observations of colony movements and fluctuations in colony size have indicated that dispersal may be common (Freer *et al.* 2015; LaRue *et al.* 2015; Roeder *et al.* 2001). Climate change poses a great threat to Antarctic and sub-Antarctic penguins through declines in habitat quality and prey availability (Ainley *et al.* 2010; Bost *et al.* 2015; Cimino *et al.* 2016; Clucas *et al.* 2014; Jenouvrier *et al.* 2014; Lynch *et al.* 2012). Determining the mechanisms governing their dispersal could lead to more accurate estimates of extinction risk, and identify effective management solutions for their conservation.

Through comparative population genomics, we can test the relative importance of different ecological and evolutionary factors in determining population genetic differentiation and dispersal in these Antarctic and sub-Antarctic penguins. The study design covered almost the entire range of five penguin species that are distributed across a range of latitudes and habitats. The species included were the sister species of the king and emperor penguins from the *Aptenodytes* genus, and the Adélie, chinstrap, and gentoo penguins from the *Pygoscelis* genus, and so we can also identify evolutionary constraints on dispersal. Although dispersal in these species has been studied previously, this is the first study to use identical genomic techniques across all species and cover such a wide spatial scale.

If temporal variability in habitat were the main determinant of dispersal, combined with a relatively continuous distribution that reduces costs to dispersal, then we would expect to see the most gene flow between emperor penguin colonies occupying the unstable fast-ice habitat around Antarctica. Adélie penguins occupy a slightly less variable niche and have a slightly more disjointed distribution, and so may be expected to be slightly less dispersive and more philopatric.

King and gentoo penguins in the sub-Antarctic occupy the most predictable niche with colonies separated by thousands of kilometres of open-ocean, and so the temporal stability combined with expected high costs to dispersal would suggest philopatry would be selected for, which could eventually lead to allopatric speciation. The chinstrap penguin may be intermediate as it occupies the semi-variable maritime Antarctic zone. Alternatively, if foraging range and non-breeding distribution were the main determinants of dispersal, then the coastal lifestyle of the gentoo penguin might limit dispersal in comparison to the other more pelagic species. To answer this question we generated robust single nucleotide polymorphism (SNP) datasets to study intraspecific genetic variation, using 374 individuals from 32 colonies in total, covering almost the entire range of all five species.

5.3 Results

5.3.1 Genotyping

We generated RADseq data from 376 individuals across five species of penguin, from a total of 32 different colonies around Antarctica and the sub-Antarctic (Figure 5.1). Up to 16 individuals were sampled per colony, and for those colonies on islands or archipelagos we sampled a single colony from the region. To ease comparisons between species, we refer to them by the island or archipelago name, rather than colony name (full details are available in Supplementary Table 5.1 and Supplementary Table 5.2). On average, we generated 11.6 million reads per individual (range = 981,992–63,386,692), with 97.1% retained after quality control. Alignment to the emperor or Adélie penguin reference genomes, SNP calling, and SNP filtering generated the following high-coverage, reliably-called SNP datasets: 5,154 neutral SNPs for king penguins with no outlier loci identified; 4,596 neutral SNPs for emperor penguins after four F_{ST} outlier loci were removed; 12,921 neutral SNPs for chinstrap penguins with no outlier loci identified; 3,872 neutral SNPs for Adélie penguins with no outlier loci identified; and 10,560 SNPs for gentoo penguins with 452 SNPs identified as outliers, removal of which left us with 10,108 SNPs in the neutral gentoo SNP dataset. Given the large number of outliers recovered within the gentoo penguin dataset, we investigated whether loci putatively under selection may be driving patterns of genetic differentiation in gentoo penguins by comparing pairwise F_{ST} values among colonies calculated using both the neutral and total SNP datasets (Supplementary Table 5.7 and Supplementary Table 5.8). The differentiation patterns were the same; therefore we proceeded with the neutral SNP dataset for most analyses, unless mentioned otherwise.

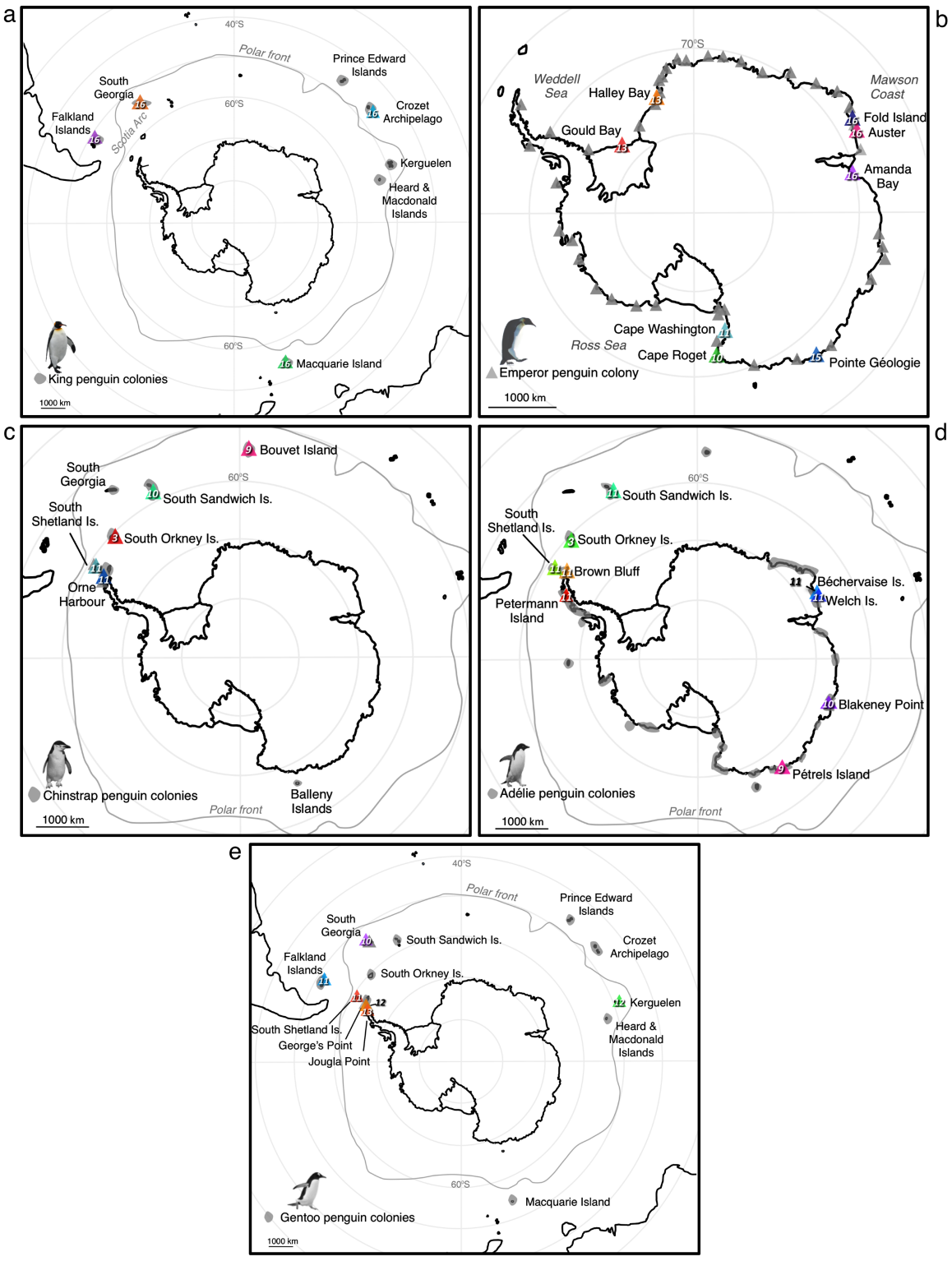


Figure 5.1. Species distributions and sampled colony locations. The species' distribution is shown in grey, and the sampled colonies are coloured (to match Figure 5.2) for: a) king penguins, b) emperor penguins, c) chinstrap penguins, d) Adélie penguins, and e) gentoo penguins. The number of individuals sampled at each colony is also indicated.

5.3.2 King penguins

In total, 64 king penguins from the Falkland Islands, South Georgia, the Crozet Islands and Macquarie Island were genotyped (Figure 5.1a). Genetic differentiation among the sampled colonies was subtle, despite colonies being distributed across discrete archipelagos separated by thousands of kilometres of open-ocean (Clucas *et al.* 2016). The successive *K*-means clustering algorithm was unable to distinguish clusters within the 64 king penguin individuals (the minimum value of the BIC occurred at $K = 1$). DAPC was also unable to clearly distinguish genetic clusters when individuals were grouped by their colony of origin, although the Falkland Islands and South Georgia did appear divergent from one another (Figure 5.2a). *Structure* was better able to distinguish the subtle genetic differentiation among colonies (Figure 5.3a). The highest mean log probability was achieved at $K = 1$ with and without location priors (Supplementary Figure 5.5), and when location priors were not used, differentiation was difficult to determine (Supplementary Figure 5.6). However, use of location priors and scrutiny of analyses for $K > 1$ primarily differentiated South Georgia when $K = 2$ (Supplementary Figure 5.7), and further differentiated Macquarie Island when $K = 3$ (not shown) or $K = 4$ (Figure 5.3a). This pattern was also supported by the pairwise F_{ST} values between colonies: all comparisons involving South Georgia were significant, and Macquarie Island was also significantly genetically distinct, except when compared to the Crozet Islands (Supplementary Table 5.3). Therefore, there appeared to be two or three subtly differentiated populations among the sampled king penguin colonies: South Georgia formed its own subtly differentiated population, and within the remaining colonies Macquarie Island was slightly differentiated in a subset of our analyses. Surprisingly, the Falkland Islands were genetically differentiated from nearby South Georgia, *ca.* 1,400 km away, but indistinguishable from the Crozet Islands *ca.* 7,500 km away. Therefore, it seems most likely that individuals from the Crozet Islands, rather than from the more proximate colonies on South Georgia, founded the recently formed colony on the Falkland Islands. Genetic diversity did not differ at any of the four sampled colonies (Supplementary Table 5.2), an unsurprising finding given that none of the colonies appeared to be genetically isolated.

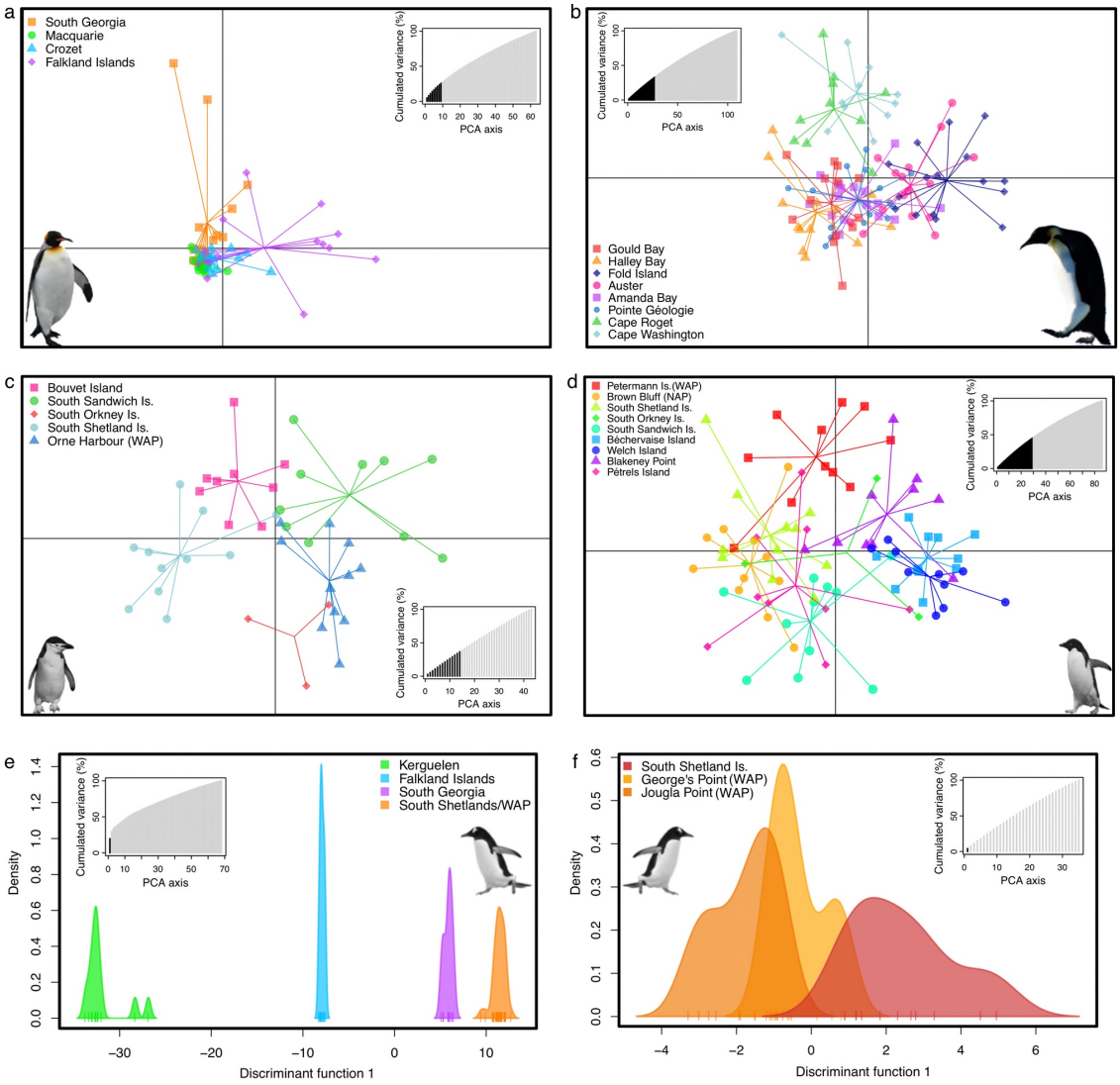


Figure 5.2. Discriminant analysis of principal components. Individuals are grouped by colony in: a) king penguins, b) emperor penguins, c) chinstrap penguins, and d) Adélie penguins. Gentoo penguins were grouped by the 4 clusters suggested by *K*-means clustering in (e), and the southern gentoo penguins from the south Shetland islands and western Antarctic Peninsula (WAP) were grouped by colony in (f). The number of principal components retained and the variance explained is shown on the inset graphs. NAP = northern Antarctic Peninsula.

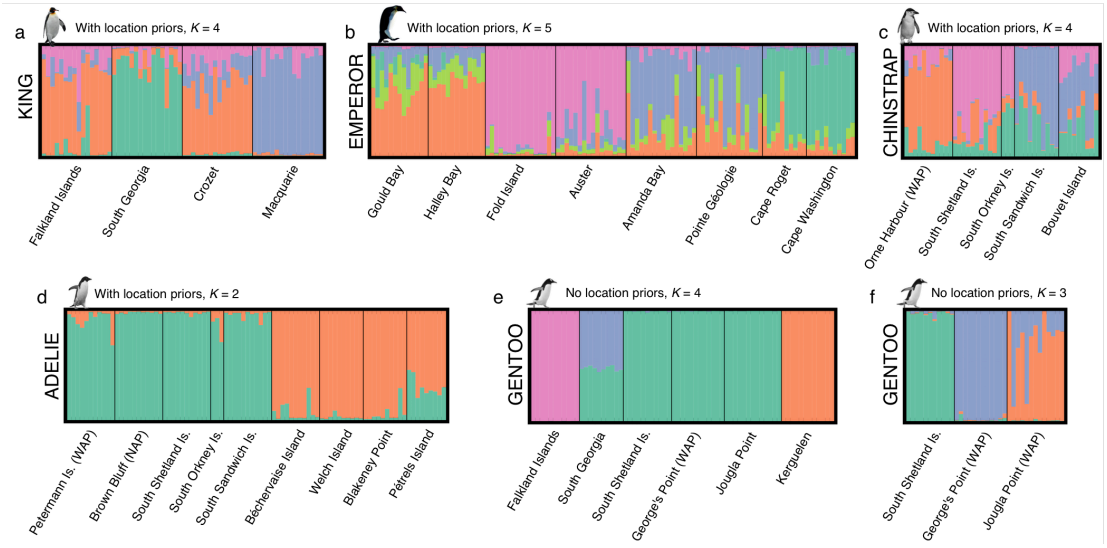


Figure 5.3. Individual assignments based on *Structure* analysis. A vertical bar represents each individual and the colours indicate the percentage of its ancestry assigned to each cluster. The results are the average of ten replicate analyses with the number of clusters and the use of location priors indicated on each panel for: a) king penguins, b) emperor penguins, c) chinstrap penguins, d) Adélie penguins, e) gentoo penguins, and f) southern gentoo penguins from the south Shetland islands and western Antarctic Peninsula (WAP). NAP = northern Antarctic Peninsula.

5.3.3 Emperor penguins

In total, 110 emperor penguins from eight colonies around Antarctica were genotyped. We included two colonies from each of the Ross Sea, Weddell Sea and the Mawson Coast, along with two additional colonies in East Antarctica (Figure 5.1b). We found evidence for four subtly differentiated metapopulations among the sampled colonies, with the Ross Sea metapopulation the most differentiated (Younger *et al.* submitted), consistent with findings from mitochondrial DNA that the Ross Sea is highly genetically distinct (Younger *et al.* 2015b). Successive *K*-means clustering could not detect genetic clusters among the 110 individuals (the minimum value for the BIC occurred at $K = 1$), however, pairwise F_{ST} values between colonies were statistically significant for 17 out of 28 comparisons (Supplementary Table 5.4). When individuals were grouped by their colony of origin, DAPC determined that the Ross Sea colonies (Cape Roget and Cape Washington) were well differentiated, and the Mawson Coast colonies of Fold Island and Auster were also subtly differentiated from other colonies (Figure 5.2b). The Bayesian clustering analysis performed in *Structure* also supported this clustering scenario. The highest posterior probability was achieved when $K = 1$, yet clustering was apparent both with and without location priors when $K > 1$. When location priors were used, colonies clustered by geographic region, with clusters consisting of the two Ross Sea colonies, the Mawson Coast colonies, the Weddell Sea colonies, and Amanda Bay and Point Géologie clustered together when $K > 4$ (Figure 5.3b). The Weddell Sea

colonies, Amanda Bay and Pointe Géologie were not distinguished from one another at $K = 3$ with location priors (Supplementary Figure 5.8), nor without location priors when $K \geq 4$ (Supplementary Figure 5.9), suggesting they are only slightly distinct. Genetic diversity did not differ substantially among colonies. Thus there appeared to be four subtly differentiated metapopulations among the eight colonies we sampled, with the Ross Sea metapopulation being the most distinct, and the colonies within each metapopulation panmictic.

5.3.4 Chinstrap penguins

We sampled 44 chinstrap penguins from five colonies around the Scotia Arc and Antarctic Peninsula, covering all the major strongholds of the species. We had robust sample numbers from Bouvet Island, the South Sandwich Islands, the South Shetland Islands and Orne Harbour (West Antarctic Peninsula) (Figure 5.1c). At the South Orkney Islands we only sampled three individuals and so our conclusions regarding that colony should be considered preliminary. Overall, population differentiation in chinstrap penguins was extremely low; only three of the ten pairwise F_{ST} values were significant (Supplementary Table 5.5), successive *K*-means clustering could not detect any clusters, the Bayesian clustering of individuals performed by *Structure* had the highest mean posterior probability when $K = 1$, and *Structure* could not discern any clusters when location priors were not used (Supplementary Figure 5.10). However, the use of location priors detected subtle structure; when $K = 3$ (not shown) and $K = 4$ Orne Harbour on the West Antarctic Peninsula was subtly differentiated, the South Sandwich Islands and Bouvet Island appeared to cluster together, as did the South Orkney Islands and South Shetland Islands (Figure 5.3c). DAPC differentiated all colonies from one another (Figure 5.2c), but given the very low levels of differentiation detected by our *Structure* and pairwise F_{ST} analyses, the distinct DAPC clusters may be a result of over-fitting, despite using *optim.a.score* to determine how many principal components (PCs) should be retained. Retaining fewer PCs in the DAPC created a plot that closely matched the subtle differentiation observed in the *Structure* and pairwise F_{ST} analyses (Supplementary Figure 5.11).

5.3.5 Adélie penguins

In total, 87 Adélie penguins from nine colonies around Antarctica were genotyped. Between nine and 11 individuals were sampled per colony, except in the South Orkney Islands where only three individuals were sampled, thus our results for that colony should be considered preliminary. The colonies can be divided into the "western colonies" around the Antarctic Peninsula and Scotia Arc, and the "eastern colonies" along the Mawson Coast and East Antarctica. The subtle genetic differentiation we uncovered within Adélie penguins was coincident with this geographic division.

Successive *K*-means clustering could not detect any genetic differentiation and the highest posterior mean log likelihood was achieved at $K = 1$, both with and without location priors in *Structure*. However, when location priors were used, a distinction between western and eastern colonies was apparent when $K \geq 2$ (Figure 5.3d). Without location priors this signal was lost (Supplementary Figure 5.12), suggesting that differentiation was subtle despite the western and eastern colonies being separated by more than 4000 km. The subtlety of genetic differentiation between these regions was supported by pairwise F_{ST} values among colonies, where only seven of 36 comparisons were significant, six of which were between eastern and western colonies (Supplementary Table 5.6). Interestingly, Pétrels Island, the furthest east of the eastern colonies, was the least differentiated from the western colonies in the *Structure* analysis (Figure 5.3d). This pattern was also observed in the DAPC, where Pétrels Island individuals were observed to cluster with the western colonies, which were otherwise marginally differentiated from the eastern colonies (Figure 5.2d). Petermann Island, on the western Antarctic Peninsula, was subtly differentiated from all other colonies by DAPC, and this was also reflected in the *Structure* analysis for the western colonies using location priors (Supplementary Figure 5.13). Within eastern colonies, no further differentiation was apparent in a *Structure* analysis on those colonies with location priors at $K = 2$ (Supplementary Figure 5.14), although when $K = 3$ two individuals from Blakeney Point did seem genetically distinct and could be migrants from a divergent population elsewhere in Antarctica (Supplementary Figure 5.15). Overall, the western and eastern colonies appeared to be subtly differentiated from one another, and within the western colonies, Petermann Island was very subtly differentiated from the other colonies by some analyses. There were no differences in the genetic diversity among the colonies (Supplementary Table 5.2).

5.3.6 Gentoo penguins

For gentoo penguins, we designed our sampling scheme to include the two currently recognised subspecies of gentoo penguins; the northern gentoo (the nominate subspecies, *Pygoscelis papua papua*, (Forster 1781)) which is formally distributed north of 60°S, and the southern gentoo (*Pygoscelis papua ellsworthii*), formally distributed on the Antarctic Peninsula and maritime Antarctic islands south of 60°S (Clements *et al.* 2016; Martínez *et al.* 2016; Murphy 1947; Stonehouse 1970); as well as the putative Indian Ocean subspecies proposed by de Dinechin *et al.* (2012), which is still formally regarded as *P. p. papua* (Clements *et al.* 2016; Martínez *et al.* 2016).

In total, we genotyped 69 gentoo penguins from six colonies (Figure 5.1e and see Supplementary Table 5.1 for colony details). We included one colony from the Falkland Islands (northern gentoos); one colony from the South Shetland Islands and two colonies from the western Antarctic Peninsula (southern gentoos); one colony from Kerguelen in the Indian Ocean (currently

designated as northern gentoos but putatively belonging to the proposed Indian Ocean subspecies); and one colony from South Georgia, which are currently defined as northern gentoos (Martínez *et al.* 2016), but most likely require taxonomic revision because evidence from morphology (de Dinechin *et al.* 2012), mitochondrial DNA and microsatellites (Levy *et al.* 2016) suggest that South Georgia gentoos should belong to the southern subspecies rather than the northern subspecies.

All of our analyses showed four distinct groupings of gentoo penguins. Both the maximum likelihood phylogeny (Figure 5.4a) and the coalescent based species-tree (Figure 5.4b) gave 100% support for four clades, corresponding to 1) the Falkland Islands (northern gentoos), 2) Kerguelen (Indian Ocean gentoos), 3) South Shetland Islands and western Antarctic Peninsula (southern gentoos), and 4) South Georgia. Both phylogenies showed Kerguelen to be the most distantly related clade, and both placed South Georgia and the southern gentoos as reciprocally monophyletic sister groups. The same pattern was evident from our DAPC, which only required one PC to explain over 15% of the variance and again differentiated the four clades completely (Figure 5.2e), with the Indian Ocean clade being the most distinct and the South Georgia and southern gentoo clades being most closely related. *Structure* also completely differentiated the Falkland Islands and Kerguelen populations from all other colonies, with no evidence of admixture among those three groups when $K = 4$, even when location priors were not used (Figure 5.3e). *Structure* also clearly differentiated South Georgia from the southern gentoo colonies when Kerguelen and Falkland Islands were removed from the analysis; the maximum posterior log likelihood occurred at $K = 2$ and individuals from South Georgia were fully assigned to a separate cluster, even when no location priors were used (Supplementary Figure 5.16). Pairwise F_{ST} values across the four clades ranged 0.127 to 0.298 and were all highly significant ($p < 0.001$, Supplementary Table 5.7). Consistent with our other analyses, the pairwise F_{ST} comparisons showed the Indian Ocean clade to be most divergent, and South Georgia to be most closely related to the southern gentoo colonies. These F_{ST} values were two orders of magnitude greater than those observed in any of the other penguin species in this study, even though geographic distances between colonies were similar. The genetic diversity of the four clades were significantly different, with the Indian Ocean subspecies showing much greater genetic diversity than the northern and southern gentoo penguins.

Our findings overwhelmingly support the proposal of de Dinechin *et al.* (2012) for an Indian Ocean subspecies, with the gentoo penguins at Kerguelen clearly the most distinct among the colonies we sampled. Our data also support the existing classification of a northern gentoo subspecies, however, contrary to current taxonomic limits (Martínez *et al.* 2016), we found that the birds on South Georgia are more closely related to the southern subspecies than the northern, a

conclusion that is supported by previous studies of morphology (de Dinechin *et al.* 2012), mitochondrial DNA (Clucas *et al.* 2014) and microsatellites (Levy *et al.* 2016). We recommend formal taxonomic revision of the boundary between northern and southern gentoo penguins to reflect this.

The three southern gentoo colonies on the South Shetland Islands and western Antarctic Peninsula were all significantly differentiated by pairwise F_{ST}, with F_{ST} values exceeding those observed for all other species (range = 0.009 – 0.017, Supplementary Table 5.7; all other species range = 0 – 0.008, Supplementary Table 5.3 – Supplementary Table 5.6). This differentiation was clearly discernible in a *Structure* analysis for the three colonies without location priors (Figure 5.3f). Given the geographic proximity of these colonies (George's Point and Jougla Point are separated by *ca*. 50 km at sea, and the South Shetland Island colony is a further 350 to 400 km away, respectively) this level of genetic differentiation is in stark contrast to the other penguin species. The genetic diversity of the three colonies did not differ, and DAPC could not fully differentiate them from one another (Figure 5.2f).

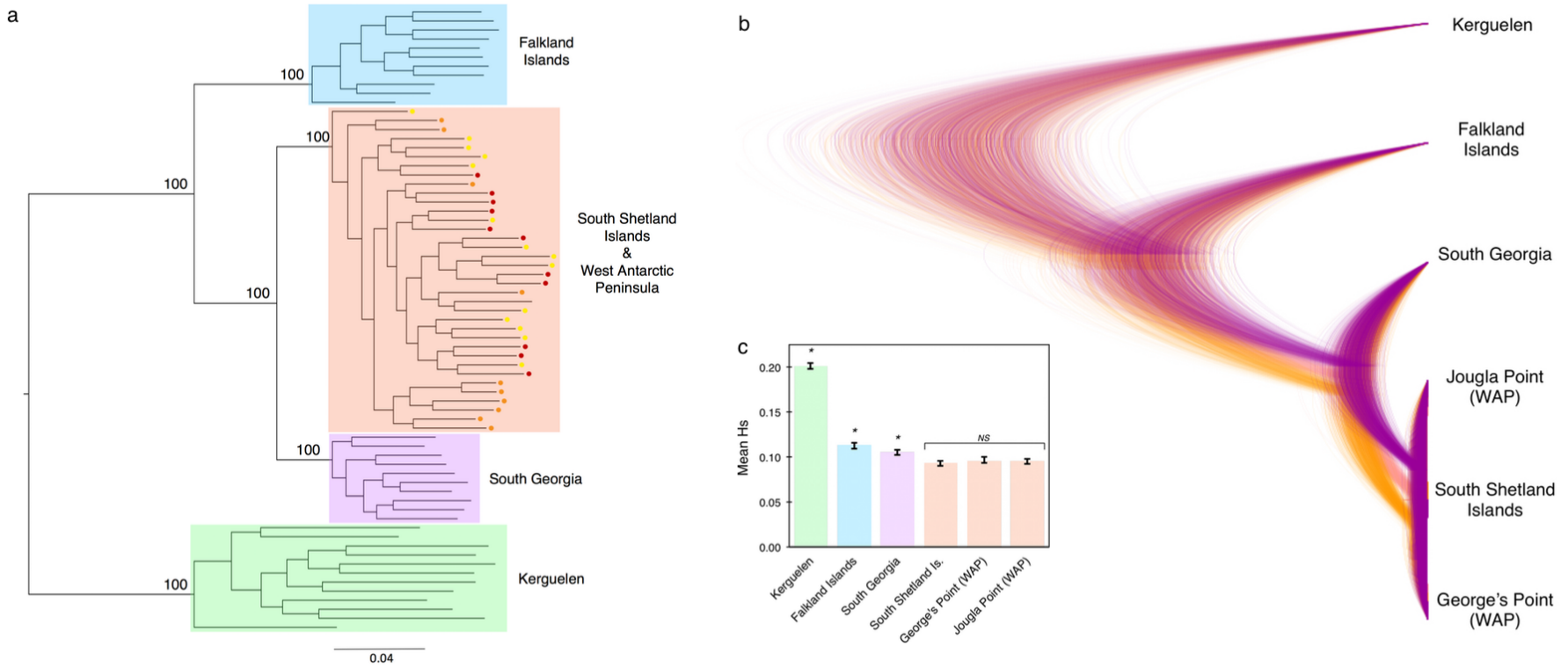


Figure 5.4. Phylogenetics and genetic diversity of gentoo penguins. a) Best scoring maximum likelihood tree from RAxML with bootstrap support for well-supported (>85%) bipartitions. The colony of origin for the individuals in the South Shetland and West Antarctic Peninsula clade are marked by the coloured circles at the tips of the tree: red – South Shetland Islands, yellow – George's Point (WAP), orange – Jougla Point (WAP). None of the nodes within this clade had bootstrap support >50, and so the branching pattern cannot be meaningfully interpreted within the clade. b) Cloudogram of the posterior distribution of trees resulting from our SNAPP species-tree analysis. The magenta topologies were the most highly supported; the orange topologies were less well supported. c) Mean expected heterozygosity per colony with 95% confidence intervals. Significant differences are indicated by asterisks, NS = not significantly different, WAP = western Antarctic Peninsula.

5.4 Discussion

5.4.1 Factors influencing patterns of genetic variation in penguins

To identify key factors that influence dispersal in penguins we compared patterns of intraspecific genetic differentiation within five closely related species in the *Aptenodytes* and *Pygoscelis* genera across their global distributions using robust datasets of genome-wide SNPs. Four out of five species (the king, emperor, chinstrap and Adélie penguins) showed evidence of gene flow over thousands of kilometres, whereas the fifth species, the gentoo penguin, had remarkably high levels of intraspecific genetic differentiation consistent with on-going allopatric speciation. Many factors have been previously identified as drivers of dispersal patterns in seabirds (Friesen 2015; Friesen *et al.* 2007a) and we will discuss our genomic results with respect to the most relevant of these for Southern Ocean penguins.

5.4.1.1 At-sea range

The most important factor in determining patterns of intraspecific genetic variation in these species appears to be their at-sea range. The four species for which we found evidence of dispersal over large spatial scales are all considered pelagic, spending at least a portion of their life history in the open ocean far from their colonies. Adélie penguins travel up to 270 km and 1,000 km away from colonies during the breeding and non-breeding seasons, respectively (Ballard *et al.* 2010; Clarke *et al.* 2006; Dunn *et al.* 2011), and juveniles have been recorded over 1,900 km away from the natal colony (Clarke *et al.* 2003). Adult emperor penguins show similar patterns, undertaking breeding and non-breeding foraging journeys of up to 500 km and 1,400 km, respectively (Kooyman *et al.* 2000; Kooyman *et al.* 2004; Wienecke *et al.* 2004; Wienecke & Robertson 1997). Juvenile emperor penguins are incredibly mobile and have been recorded travelling in excess of 7,000 km in just eight months (Thiebot *et al.* 2013), with individuals documented in the vicinity of other breeding colonies (Kooyman *et al.* 1996; Wienecke *et al.* 2010). Chinstrap penguins are less wide-ranging during the breeding season, generally keeping within 60 km of colonies, however, their non-breeding range is similar to that of the other pelagic species with some individuals dispersing up to 1,800 km (Lynnes *et al.* 2002; Trivelpiece *et al.* 2007; Wilson *et al.* 1998). Seabirds that forage preferentially at upwelling zones are thought to be particularly dispersive, because they must regularly adjust their foraging distributions to track oceanographic features (Taylor *et al.* 2011). This may apply to king penguins which direct their foraging trips toward frontal zones, the locations of which vary from year to year (Bost *et al.* 2009; Bost *et al.* 2015). In the winter months king penguins travel up to 1,800 km to forage in the

marginal ice zone (Charrassin & Bost 2001) and juveniles have been observed at breeding colonies up to 5,600 km from their natal colonies (Weimerskirch *et al.* 1985). This wide-ranging behaviour in the pelagic penguins is likely to facilitate their dispersal, as evidenced by overall low genetic differentiation within all the pelagic species.

On the other hand, gentoo penguins have a coastal lifestyle and forage inshore on a variety of prey (Borboroglu & Boersma 2013; Ratcliffe & Trathan 2012), rather than making long journeys to exploit specific prey resources. Gentoo penguins generally forage within 40 km of colonies during the breeding season and are rarely seen more than 50 km offshore during the non-breeding season, although they may travel around within archipelagos and visit other colonies (Clausen & Pütz 2003; Masello *et al.* 2010; Ratcliffe & Trathan 2012). There is currently no data on the juvenile journeys of gentoo penguins, but based on the at-sea distribution of adults they are unlikely to be wide-ranging. This tendency to stay close to home appears to act as a barrier to dispersal in gentoo penguins by reducing mixing with individuals from other colonies, contributing to the high degree of genetic differentiation we recorded. It should be noted, however, that genetic differentiation was observed between gentoo penguin colonies separated by less than 50 km. This finding cannot be explained solely by the gentoo penguin's coastal lifestyle, because the species is known to visit other colonies within this range (Ratcliffe & Trathan 2012). Given the very small spatial scale over which population differentiation was observed, it is possible that natal philopatry also plays a role in limiting gene flow of gentoo penguins. Natal philopatry is thought to be common among seabirds (Coulson 2001) and has been identified as a barrier to gene flow in other species, although it usually acts in combination with other isolating mechanisms (Friesen 2015).

Despite this possible natal philopatry, paradoxically the gentoo penguin's range has been expanding southwards along the western Antarctic Peninsula as sea ice in the region has declined (Lynch *et al.* 2012). The local scale genetic differentiation found here would suggest that continued immigration into newly founded colonies could be low, and high rates of breeding success and recruitment may explain rapid growth rates after establishment (Trathan *et al.* 2008). Sampling at the newly established colonies could determine this.

5.4.1.2 Oceanographic fronts

Both gentoo and king penguins have breeding distributions spanning the Polar Front (Figure 5.1a,e), which we found may be a barrier to dispersal in both species. King penguins from South Georgia, which is the only breeding population situated south of the front that we sampled (Figure 5.1a), were the most genetically divergent in our dataset. Furthermore, king penguins from the Falkland Islands were genetically indistinguishable from those at Crozet *ca*. 7,500 km

away but situated on the same side of the Polar Front, whereas they were differentiated from those at South Georgia, only 1,400 km away but on the opposite side of the front. A similar pattern was evident in gentoo penguins. Our study included one gentoo penguin colony north of the Polar Front, at the Falkland Islands, whereas the other colonies were all distributed south of the front (Figure 5.1e). The Falkland Islands were genetically divergent from all other colonies. Compellingly, the colonies on the South Shetland Islands and Antarctic Peninsula were more closely related to the South Georgia colony, which is on the same side of the front, than they are to the Falkland Islands, which is more proximate but on the opposite side of the front. This suggests that, in addition to their coastal lifestyle, the Polar Front is a barrier to gentoo penguin dispersal.

5.4.1.3 Breeding habitat quality, continuity and ephemerality

Emperor penguins have a relatively continuous distribution around Antarctica (Figure 5.1b; (Fretwell & Trathan 2009)) and their fast-ice breeding habitat is highly ephemeral, leading to changes in their breeding distribution over years (Ancel *et al.* 2014; Fretwell *et al.* 2014; LaRue *et al.* 2015; Trathan *et al.* 2011) and millennia (Younger *et al.* 2016a; Younger *et al.* 2015b). The low levels of genetic differentiation among emperor penguin colonies likely reflect the need for breeding location flexibility. The case of the Adélie penguin is similar, in that its breeding habitat is somewhat ephemeral, being periodically blocked by sea ice, and is somewhat continuously distributed, but with several large discontinuities coinciding with extensive glaciation (Figure 5.1d). We found that Adélie penguins were subtly genetically differentiated on either side of a gap of several thousand kilometres in their breeding distribution (Figure 5.1d), suggesting that the break in breeding habitat impedes gene flow. In regions where Adélies are continuously distributed there was no evidence of genetic divergence over thousands of kilometres, indicative of dispersal consistent with the ephemerality of the breeding habitat and facilitated by its continuity.

While the sub-Antarctic breeding habitat of the king and northern colonies of gentoo penguins experiences variability in climate, in general it is far more stable than the Antarctic habitat of the Adélie and emperor penguins, being present and accessible year-round. The chinstrap penguin occupies a somewhat intermediate habitat in the maritime Antarctic, which is subject to variability in sea ice that may occasionally limit access. The king, gentoo, and chinstrap penguins also have patchy distributions, with breeding sites situated on archipelagos (Figure 1a,c,e). The patchiness and relative stability would both suggest that philopatry may be selected for in these species, however we find that not to be the case, except in the gentoo penguin, for which other dispersal barriers have already been noted. The large at-sea distributions of chinstrap and king

penguins appear to facilitate gene flow to such a degree that the dispersal barriers posed by their patchy distributions are overcome. Occasional large-scale disruptions in breeding habitat may cause pulses of dispersal in some species. This is most likely for the chinstrap penguin, which has large populations on the highly volcanically active South Sandwich Islands (Holdgate & Baker 1979).

5.4.1.4 Implications for modelling studies

Regardless of whether dispersal is continuous or episodic, our results show that pelagic penguins can, and do, disperse among colonies separated by thousands of kilometres. However, dispersal has so far been discounted by studies that attempt to model local extinction risk for pelagic penguins (Adélie penguins (Cimino *et al.* 2016); emperor penguins (Jenouvrier *et al.* 2009; Jenouvrier *et al.* 2014)). While there is great value in estimating the likelihood of local or global species extinctions, such studies must include estimates of dispersal to have any hope of accuracy. Dispersal decouples the relationship between local climate and demographic rates (Tavecchia *et al.* 2016), facilitates range shifts, furnishes populations with potentially adaptive genetic variants (Walther *et al.* 2002), and can bolster population stability by compensating for low birth rates and/or survival (Lowe & Allendorf 2010). The findings of modelling studies that discount dispersal should be treated with great caution, and we urge that future models incorporate the dispersal patterns that we have outlined here.

5.4.2 Cryptic speciation in gentoo penguins

The degree of genetic divergence we have found within gentoo penguins points to a need for taxonomic revision to recognise on-going allopatric speciation. While there has been previous suggestion of the need for revision based on mitochondrial DNA (Clucas *et al.* 2014; de Dinechin *et al.* 2012) and morphometrics (de Dinechin *et al.* 2012), we have now confirmed cryptic speciation within gentoo penguins using a robust dataset of genome-wide SNPs. The currently recognised taxonomy of gentoo penguins is for two subspecies, the northern (*Pygoscelis papua papua*) distributed north of 60°S, and the southern (*Pygoscelis papua ellsworthii*) distributed on the Antarctic Peninsula and maritime Antarctic islands south of 60°S (Clements *et al.* 2016; Forster 1781; Martínez *et al.* 2016; Murphy 1947). The taxonomy should be revised to include a third subspecies breeding at Kerguelen, which is the most divergent group of gentoo penguins. We also suggest that the boundary of southern gentoos be revised to include South Georgia. In light of these results, there is also a need to characterise gentoo penguins breeding at other archipelagos, and particularly the distant Macquarie Island, which is likely to harbour more cryptic diversity as a result of its location far from any other colonies. Accurate species boundaries and the recognition

of cryptic species are crucial for the conservation of biodiversity, particularly in light of the challenges that will face Southern Ocean biota in the Anthropocene. The three subspecies of gentoo penguins are on separate evolutionary trajectories. By conserving their full spectrum of genetic variation the evolutionary and adaptive potential of the gentoo penguin can be maximised.

5.4.3 Concluding remarks

Understanding the mechanisms behind patterns of species dispersal has never been more important. Climate change is dramatically altering the marine environment, particularly in the Southern Ocean (Orr *et al.* 2005; Vaughan *et al.* 2003), leading to changes in habitat availability, quality and ephemerality, as well as shifting oceanographic features (Constable *et al.* 2014). Understanding the current barriers to penguin dispersal is essential for enabling greater accuracy in forecasting how these species might respond to changes in their environment, and for implementing ecologically meaningful conservation strategies. Here we have shown that oceanographic fronts and natal philopatry play roles as barriers to dispersal in some penguins. Pelagic species of penguin maintain gene flow over extraordinary distances, in excess of 7,500 km in some cases, whereas penguins with a coastal lifestyle are non-dispersive, leading to ongoing allopatric speciation. Our study highlights that the different ecological niches of these closely related taxa have led them toward different evolutionary trajectories.

5.5 Methods

5.5.1 Sampling and sequencing

Blood or tissue samples were collected from up to 16 individuals per colony across a large part of the range of each species (Figure 5.1). Colony names, collection dates and tissue types are provided in Supplementary Table 5.1. Further details of the tissues collected from Adélie penguins at Béchervaise Island, Welch Island, Blakeney Point and Pétrels Island can be found in (Younger *et al.* 2015a). Details of the tissue samples collected from emperor penguins at Halley Bay, Fold Island, Auster, Amanda Bay and Pointe Géologie can be found in Younger *et al.* (2015a). All other samples were blood samples. To take blood, penguins were held with the flippers restrained and the head placed under the arm of the handler, or they were wrapped in cushioned material covering the head and preventing movement, to minimize stress during handling (Le Maho *et al.* 1992). A second handler took up to 1 mL blood from the brachial, intertarsal or jugular vein using a 25G or 23G needle and 1 mL syringe, after cleaning the area with an alcohol swab. Total restraint time was generally two to three minutes. All field activities were conducted under

appropriate permits and were subject to independent ethical review. Samples were stored in ethanol or Queen's Lysis buffer at -20 °C and transported frozen back to the UK or Australia, or in RNAlater (Life Technologies) and transported at ambient temperature and frozen at -20 °C upon arrival.

DNA was extracted from blood and tissue samples using QIAGEN DNeasy Blood and Tissue kits. The digestion step was modified to include 40 µL proteinase K and extended to 3 hrs for blood samples. Details of the modifications made to the protocols for tissue samples are available in Younger *et al.* (2015a) and Younger *et al.* (2015b). All samples were treated with 1 µL Riboshredder (Epicentre) to reduce RNA contamination and DNA was visualized on a 1% agarose gel to confirm high molecular weight DNA was present. DNA concentration and purity was measured on a Qubit and Nanodrop (Thermofisher Scientific), respectively.

To identify genome-wide SNPs for each species, we used standard RADSeq (Baird *et al.* 2008) with individual barcoding and the Sbf1 restriction enzyme. The NERC Biomolecular Analysis Facility at Edinburgh Genomics (https://genomics.ed.ac.uk) performed the library preparation and sequencing as described by Gonen *et al.* (2014) following Etter *et al.* (2011). In short, 250 ng of DNA per individual was digested with Sbf1-HF (NEB) and then ligated to barcoded P1 adapters. Individuals were multiplexed into 18 libraries consisting of 19–23 barcoded individuals, and were sheared into fragments of < 300–400 bp. Size selection was performed by gel electrophoresis. Libraries were blunt ended (NEB Quick Blunting Kit) and A-tailed before P2 adapters (IDT) were ligated. Enrichment PCR and purification with Ampure beads was performed before libraries were checked for size and quantity using Qubit and qPCR assays. Each library was then sequenced in a lane of the Illumina HiSeq 2500 using 125 base paired-end reads in high output mode (v4 chemistry).

5.5.2 Bioinformatics

Read quality was assessed with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Demultiplexing, removal of reads with adapter contamination and trimming to 113 bp was performed with *process_radtags* from the Stacks pipeline v1.35 (Catchen *et al.* 2011; Catchen *et al.* 2013). We also used *process_radtags* to remove any read pairs with uncalled bases, a low quality score and/or a barcode or cut-site with more than one mismatch. King and emperor penguin reads were aligned to the emperor penguin reference genome (http://gigadb.org/dataset/100005) whilst Adélie, chinstrap and gentoo penguin reads were aligned to the Adélie penguin reference genome (http://gigadb.org/dataset/100006) using *bwa-mem* (Li 2013). Terminal alignments were

prevented by enforcing a clipping penalty of 100 and reads with more than five mismatches, multiple alignments and/or more than two indels (Catchen *et al.* 2013) were removed using a custom python script (filter.py, available from 10.5061/dryad.7c0q8). The effect of these thresholds on the number of uniquely and multi-mapped reads generated when mapping to a species-specific (emperor to emperor) or closely related (king to emperor) reference genome was investigated and the results showed it was valid to use the same settings in both mapping scenarios (see Supplementary Information). PCR duplicates were removed with Picardtools (http://broadinstitute.github.io/picard).

We called and filtered SNPs separately for each species using the Stacks pipeline, following many of the suggestions outlined in Benestan *et al.* (2016). All of the settings and filters were applied in the same way for each species. We ran the modules of the pipeline separately: *pstacks* – *cstacks* – *sstacks* – *rxstacks* – *cstacks* – *sstacks* – *populations*. Briefly, in *pstacks* we required a minimum stack depth of six reads mapping to the same location and used the bounded SNP model (significance level of $\alpha = 0.05$, upper bound = 0.1, lower bound = 0.00041 corresponding to the highest sequencing error rate recoded by phiX spikes). In *cstacks* we used all the individuals in each species to build the catalogue of loci. In *rxstacks* we removed confounded loci with a conservative confidence limit of 0.25, we removed excess haplotypes from individuals and also removed any loci with a mean log likelihood < -10. We filtered SNPs in the *populations* module to retain a single random SNP per RAD-tag, remove any SNPs with a minor allele frequency (MAF) < 0.01 or heterozygosity > 0.5, remove any loci that were not present in all populations and only retain SNPs that were genotyped in at least 80% of individuals per population. In the *adegenet* package (Jombart 2008; Jombart & Ahmed 2011) in R (Team 2013) we calculated whether SNPs were in Hardy Weinberg equilibrium (HWE) in each population and used vcftools v0.1.13 (Danecek *et al.* 2011) to calculate mean coverage for each SNP. SNPs were removed if they were out of HWE in > 50% of the colonies or had a mean coverage greater than twice the standard deviation away from the mean for the species. PGDSpider v2.0.8.2 (Lischer & Excoffier 2012b) was used to convert the vcf file into other formats for further analyses.

5.5.3 Outlier loci detection

We identified SNPs that were potentially under selection in each species using the F_{ST} outlier method in BayeScan v2.1 (Foll & Gaggiotti 2008). These loci were removed prior to coalescent-based or population genetic analyses that assume loci are evolving neutrally. With this aim, the high false-positive rate associated with BayeScan (Lotterhos & Whitlock 2014) was not a concern, and its power to detect loci genuinely under selection under a range of demographic scenarios was advantageous (Lotterhos & Whitlock 2014). We set a conservative prior on the odds of

neutrality (for every five loci our prior expectation is that one is under selection) to identify all loci that could potentially be under selection. We deemed q-values < 0.1 to be significant, meaning that one in ten loci identified was expected to be a false-positive neutral locus (Lotterhos & Whitlock 2014; Storey & Tibshirani 2003). SNPs that were identified to be putatively under selection were removed using vcftools. We refer to the remaining SNPs as "neutral SNP datasets", and those with the full complement before outliers were removed as the "total SNP dataset", although these definitions are applicable only to gentoo penguins, as these were the only species in which a large number of outliers were identified.

5.5.4 Contemporary population structure and summary statistics

The number of private alleles in each colony was calculated with the *populations* module in Stacks, using the total SNP datasets for each species. We calculated the observed (H_O) and expected (H_S) heterozygosity for each colony with the neutral SNP datasets using Genodive v2.0b27 (Meirmans & Van Tienderen 2004).

We used Genodive to calculate the Weir and Cockerham unbiased F_{ST} estimator (Weir & Cockerham 1984) between all pairs of colonies within each species. This measure has been shown to be robust to small sample sizes when F_{ST} is low (Willing *et al.* 2012). Significance was calculated using 10,000 permutations of the data, and corrected for multiple tests using the sequential goodness of fit method (SGoF+) (Carvajal-Rodriguez & de Uña-Alvarez 2011). For gentoo colonies we measured F_{ST} using both the neutral SNP dataset and the total SNP dataset.

We used two different clustering methods for each species: the Bayesian clustering algorithm employed by the program *Structure* v2.3.4 (Pritchard *et al.* 2000), and Discriminant Analysis of Principal Components (DAPC) (Jombart *et al.* 2010). *Structure* uses a Bayesian clustering approach with a Markov Chain Monte Carlo (MCMC) sampling procedure, which results in estimates of the membership coefficients of each individual to each of the inferred clusters, effectively identifying genetic populations and then assigning individuals to those populations. For all taxa, we used the admixture model with correlated allele frequencies and ran the model both with and without supplying sampling locations as priors, to detect subtle versus strong population structure. In each case, we first ran the model for 100,000 generations, discarding the first 50,000 as burnin, setting *K* (the number of clusters) to one but allowing lambda to vary in order to estimate the species-specific value of lambda to use. For subsequent runs, the species-specific value of lambda was set and the number of clusters was allowed to vary from *K* = 1 to *K* = N, where N was the number of colonies sampled for that species. Each analysis was run for 150,000 generations, discarding the first 50,000 as burnin, and repeated ten times from a different random seed. We used *structure*

harvester web v0.6.94 (Earl 2012) to compare replicates and prepare files for CLUMPP (Jakobsson & Rosenberg 2007), which aligns the results from replicate runs of *Structure* to check for multimodality, and calculates the average membership coefficients of each individual to each cluster, ready for visualisation with DISTRUCT v1.1 (Rosenberg 2004). Mostly, we did not use the results from the Evanno method for estimating the "true" number of clusters in the data (Evanno *et al.* 2005), as it is not defined for $K = 1$ and it was often hard to find biological meaning in the results. This is not unexpected, as the Evanno method has been shown to perform poorly for scenarios of moderate to low genetic differentiation (Waples & Gaggiotti 2006). Instead we discuss our results for multiple values of K for each taxon, to fully understand the levels of structure in the data.

The second clustering analysis that we used, DAPC, can be used to describe genetic clusters by creating synthetic variables (discriminant functions) that maximize the variance among clusters whilst minimizing the variance within them. When genetic differentiation is moderate to strong, individuals can be assigned to clusters using successive K-means clustering (the *find.clusters* function in *adegenet* (Jombart 2008; Jombart & Ahmed 2011) before DAPC, thus negating the *a priori* assignment of individuals to groups determined by their sampling location. However, in all species other than gentoo penguins, the successive K-means clustering suggested $K = 1$ was most likely (the Bayesian Inference Criterion was at its minimum at $K = 1$) and so DAPC was performed when individuals were grouped by their colony of origin. The number of principal components to retain was determined by averaging the results from 20 runs of *optima.a.score.*

5.5.5 Phylogenetics of gentoo penguins

To investigate the phylogeographic relationships among gentoo penguin colonies we used the coalescent species tree approach implemented in the SNAPP package (Bryant *et al.* 2012) in BEAST v.2.4.0 (Bouckaert *et al.* 2014). SNAPP infers species trees from unlinked biallelic markers, such as SNPs. The method calculates species tree likelihoods directly from the data by estimating the probability of allele frequency change across nodes, thus avoiding the necessity of finding and combining individual gene trees. Nevertheless, the method is highly computationally demanding, therefore we selected two random individuals (i.e. four haplotypes) per colony to include in the analysis, and repeated the analysis twice with different individuals to ensure reproducibility. The neutral dataset was used, and loci that were no longer polymorphic in the reduced set of individuals were removed, leaving 6,868 and 6,754 SNPs. The mutation rates (u and v) were calculated from the data rather than estimated as part of the MCMC. The MCMCs were run for three million generations with the first 10% discarded as burnin. We monitored the traces for convergence using Tracer v1.6 (Rambaut & Drummond 2007) and when ESSs for all parameters

were large (> 300) and the traces had reached stationarity we concluded the analyses. DensiTree v2.0.1 was used to visualize the posterior distributions of topologies as cloudograms, hence allowing for a clear depiction of uncertainty in the topology.

To investigate whether described (Stonehouse 1970) and putative (de Dinechin *et al.* 2012) subspecies of gentoo penguins are reciprocally monophyletic, and therefore taxonomically valid, we used RAxML v8.2.7 (Stamatakis 2014) to infer maximum likelihood phylogenies among the full complement of gentoo penguin individuals using the total SNP dataset. An ascertainment bias correction was applied to the likelihood calculations, as recommended when using SNPs to account for the lack of invariant sites (Leaché *et al.* 2015). When using an ascertainment bias correction all potentially invariant sites must be removed from the dataset. An alignment site consisting of only heterozygotes and homozygotes for a single allele (e.g. Rs and As with no Gs) is considered potentially invariant by RAxML, therefore we filtered out such sites using the Phrynomics R script (https://rstudio.stat.washington.edu/shiny/phrynomics/). After this filtering step 5,871 SNPs remained in the dataset. We conducted 20 independent maximum likelihood tree inferences and then drew bootstrap supports from 1,000 replicates onto the best scoring topology. All searches were conducted under the GTRGAMMA nucleotide substitution model.

5.6 Supplementary information

Table 5.1. Sampling details for all species and colonies included in the study.

Colony	Collection seasons	Tissue type
Kings		
Falkland Islands – Volunteer Pt.	2013 – 2014	Blood
South Georgia – Fortuna Bay	2011 – 2012	Blood
Crozet Archipelago – Baie du Marin	2003 – 2004	Blood
Macquarie Island – Sandy Bay	2005 – 2006	Blood
Emperors		
Weddell Sea – Halley Bay	2012 – 2013	Tissue
Weddell Sea – Gould Bay	2013 – 2014	Blood
Fold Island	2010 – 2011	Tissue
Auster	1993 – 1994	Tissue
Amanda Bay	2012 – 2013	Tissue
Pointe Géologie	2010 – 2011	Tissue
Ross Sea – Cape Roget	1992 – 1993	Blood
Ross Sea – Cape Washington	1992 – 1993	Blood
Chinstraps		
Bouvet Island – Nyrøysa	1997 – 1998	Blood
South Sandwich Islands – Hewison Point, Thule Island	2010 – 2011	Blood
South Orkneys – Signy Island	2006 – 2007	Blood
South Shetlands – Admiralty Bay	2010 – 2011	Blood
Orne Harbour (WAP)	2013 – 2014	Blood
Adélies		
South Sandwich Islands – Bellingshausen Island	2010 – 2011	Blood
South Orkneys – Signy Island	2006 – 2007	Blood
South Shetlands – Admiralty Bay	2010 – 2011	Blood
Brown Bluff (NAP)	1997 - 2000	Blood
Petermann Is. (WAP)	2013 – 2014	Blood
Béchervaise Island	2012 – 2014	Tissue
Welch Island	2012 – 2014	Tissue
Blakeney Point	2012 – 2014	Tissue
Pétrels Island	2012 – 2014	Tissue
Gentoos		
Kerguelen – Morne	2013 – 2014	Blood
Falkland Islands – Cow Bay	2013 – 2014	Blood
South Georgia – Bird Island	2005 – 2006	Blood
South Shetlands – Admiralty Bay	2010 – 2011	Blood
George's Point (WAP)	2013 – 2014	Blood
Jougla Point (WAP)	2013 – 2014	Blood

Table 5.2. Genetic diversity per population, based on variant (SNP) sites. H_S = expected heterozygosity per population, H_O = observed heterozygosity per population.

	N individuals	*N* private alleles	H_S (mean)	H_S (std dev)	H_O (mean)	H_O (std dev)
Kings	*64*					
Falkland Islands	16	148 (2.9%)	0.122	0.002	0.111	0.002
South Georgia	16	147 (2.9%)	0.120	0.002	0.107	0.002
Crozet Archipelago	16	117 (2.3%)	0.122	0.002	0.115	0.002
Macquarie Island	16	180 (3.5%)	0.123	0.002	0.111	0.002
Emperors	*110*					
Halley Bay	13	6 (0.13 %)	0.127	0.002	0.117	0.002
Gould Bay	13	7 (0.15 %)	0.126	0.002	0.111	0.002
Fold Island	16	11 (0.24 %)	0.128	0.002	0.111	0.002
Auster	16	5 (0.11 %)	0.126	0.002	0.119	0.002
Amanda Bay	16	9 (0.20 %)	0.128	0.002	0.117	0.002
Pointe Géologie	15	6 (0.13 %)	0.127	0.002	0.113	0.002
Cape Roget	10	1 (0.02 %)	0.123	0.002	0.107	0.002
Cape Washington	11	1 (0.02 %)	0.124	0.002	0.113	0.002
Chinstraps	*44*					
Bouvet Island	9	1063 (8.2 %)	0.127	0.001	0.122	0.001
South Sandwich Islands	10	1126 (8.7 %)	0.123	0.001	0.115	0.001
South Orkney Islands	3	325 (2.5 %)	0.124	0.002	0.114	0.002
South Shetlands	11	1264 (9.8 %)	0.126	0.001	0.120	0.001
Orne Harbour (WAP)	11	1245 (9.6 %)	0.125	0.001	0.118	0.001
Adélies	*87*					
South Sandwich Islands	11	23 (0.59 %)	0.101	0.002	0.096	0.002
South Orkney Islands	3	1 (0.03 %)	0.098	0.003	0.090	0.003
South Shetlands	11	23 (0.59 %)	0.102	0.002	0.098	0.002
Brown Bluff (NAP)	11	32 (0.82 %)	0.103	0.002	0.097	0.002
Petermann Is. (WAP)	11	35 (0.90 %)	0.103	0.002	0.099	0.002
Bechervaise Island	11	42 (1.1 %)	0.106	0.002	0.102	0.002
Welch Island	11	24 (0.62%)	0.099	0.002	0.093	0.002
Blakeney Point	10	32 (0.83 %)	0.101	0.002	0.094	0.002
Pétrels Island	9	35 (0.90 %)	0.104	0.002	0.100	0.002
Gentoos	*69*					
Kerguelen	12	3653 (35 %)	0.201	0.002	0.189	0.002
Falkland Islands	11	744 (7.0 %)	0.113	0.002	0.109	0.002
South Georgia	10	333 (3.2 %)	0.105	0.002	0.100	0.002
South Shetlands	11	61 (0.58 %)	0.093	0.002	0.089	0.002
George's Point (WAP)	12	61 (0.58 %)	0.096	0.002	0.092	0.002
Jougla Point (WAP)	13	67 (0.63 %)	0.095	0.002	0.088	0.002

Table 5.3. Pairwise F_{ST} values between all king penguin colonies. F_{ST} values are below the diagonal with *p* values above. Bold values were significant at the α = 0.05 level after SGoF+ correction for multiple tests. This data also appears in Table 3.2

	Falkland Islands	South Georgia	Crozet	Macquarie
Falkland Islands	***	**0.002**	0.829	**0.008**
South Georgia	0.003	***	**0.009**	**<0.001**
Crozet	-0.001	0.003	***	0.126
Macquarie	0.003	0.005	0.001	***

Table 5.4. Pairwise F_{ST} values between all emperor penguin colonies. F_{ST} values are below the diagonal with *p* values above. Bold values were significant at the α = 0.05 level after SGoF+ correction for multiple tests. This data also appears in Figure 4.4.

	Gould Bay	Halley Bay	Fold Island	Auster	Amanda Bay	Pointe Géologie	Cape Roget	Cape Washington
Gould Bay	***	0.813	**0.002**	**0.016**	0.755	0.252	0.150	**0.028**
Halley Bay	-0.001	***	**0.001**	**<0.001**	**0.025**	**0.031**	0.040	**<0.001**
Fold Island	**0.004**	**0.005**	***	0.125	**0.027**	**0.004**	**0.005**	**<0.001**
Auster	**0.003**	**0.005**	0.001	***	0.065	**<0.001**	**0.001**	**<0.001**
Amanda Bay	-0.001	**0.002**	**0.002**	0.002	***	0.813	0.055	**0.006**
Pointe Géologie	0.001	**0.002**	**0.003**	**0.004**	-0.001	***	0.058	**0.030**
Cape Roget	0.002	0.003	**0.004**	**0.005**	0.003	0.002	***	0.905
Cape Washington	**0.003**	**0.006**	**0.005**	**0.006**	**0.003**	**0.002**	-0.002	***

Table 5.5. Pairwise F_{ST} values between all chinstrap penguin colonies. F_{ST} values are below the diagonal with *p* values above. Bold values were significant at the α = 0.05 level after SGoF+ correction for multiple tests.

	South Sandwich Islands	South Orkney Islands	South Shetland Islands	Orne Harbour (WAP)	Bouvet Island
South Sandwich Islands	***	0.147	**<0.001**	0.013	0.528
South Orkney Islands	0.003	***	0.565	0.250	0.457
South Shetland Islands	**0.005**	0.000	***	**0.008**	0.585
Orne Harbour (WAP)	0.003	0.002	**0.002**	***	**0.011**
Bouvet Island	0.000	0.000	0.000	**0.002**	***

Table 5.6. Pairwise F_{ST} values between all Adélie penguin colonies. F_{ST} values are below the diagonal with *p* values above. Bold values were significant at the α = 0.05 level after SGoF+ correction for multiple tests.

	South Sandwich Islands	South Orkney Islands	South Shetland Islands	Brown Bluff (NAP)	Petermann Island	Bechervaise Island	Welch Island	Blakeney Point	Pétrels Island
South Sandwich Islands	***	0.695	0.188	0.709	0.07	0.057	0.366	**0.006**	0.494
South Orkney Islands	-0.003	***	0.58	0.506	0.631	0.823	0.949	0.478	0.715
South Shetland Islands	0.002	-0.001	***	0.078	0.186	**0.045**	**0.008**	**0.045**	0.156
Brown Bluff (NAP)	-0.001	0.000	0.003	***	**0.048**	**0.026**	**0.01**	0.091	0.614
Petermann Island	0.003	-0.002	0.002	**0.003**	***	0.094	0.063	0.074	0.179
Bechervaise Island	0.003	-0.005	**0.004**	**0.004**	0.003	***	0.978	0.458	0.318
Welch Island	0.001	-0.008	**0.005**	**0.005**	0.003	-0.004	***	0.475	0.54
Blakeney Point	**0.008**	0.000	**0.005**	0.004	0.005	0.000	0.000	***	0.084
Pétrels Island	0.000	-0.002	0.002	-0.001	0.002	0.001	0.000	0.005	***

Table 5.7. Pairwise F_{ST} values between all gentoo penguin colonies using the neutral SNP dataset. F_{ST} values are below the diagonal with *p* values above. No correction for multiple tests was performed as the range of the p-values was too small.

	Falkland Islands	South Georgia	South Shetland Is.	George's Point (WAP)	Jougla Point (WAP)	Kerguelen
Falkland Islands	***	<0.001	<0.001	<0.001	<0.001	<0.001
South Georgia	0.247	***	<0.001	<0.001	<0.001	<0.001
South Shetland Is.	0.277	0.131	***	<0.001	<0.001	<0.001
George's Point (WAP)	0.274	0.127	0.014	***	<0.001	<0.001
Jougla Point (WAP)	0.274	0.133	0.017	0.009	***	<0.001
Kerguelen	0.260	0.265	0.291	0.293	0.298	***

Table 5.8. Pairwise F_{ST} values between all gentoo penguin colonies using the total SNP dataset. F_{ST} values are below the diagonal with *p* values above. No correction for multiple tests was performed as the range of the p-values was too small.

	Falkland Islands	South Georgia	South Shetland Is.	George's Point (WAP)	Jougla Point (WAP)	Kerguelen
Falkland Islands	***	<0.001	<0.001	<0.001	<0.001	<0.001
South Georgia	0.232	***	<0.001	<0.001	<0.001	<0.001
South Shetland Is.	0.259	0.12	***	<0.001	<0.001	<0.001
George's Point (WAP)	0.256	0.117	0.011	***	<0.001	<0.001
Jougla Point (WAP)	0.256	0.122	0.015	0.007	***	<0.001
Kerguelen	0.261	0.270	0.294	0.297	0.301	***

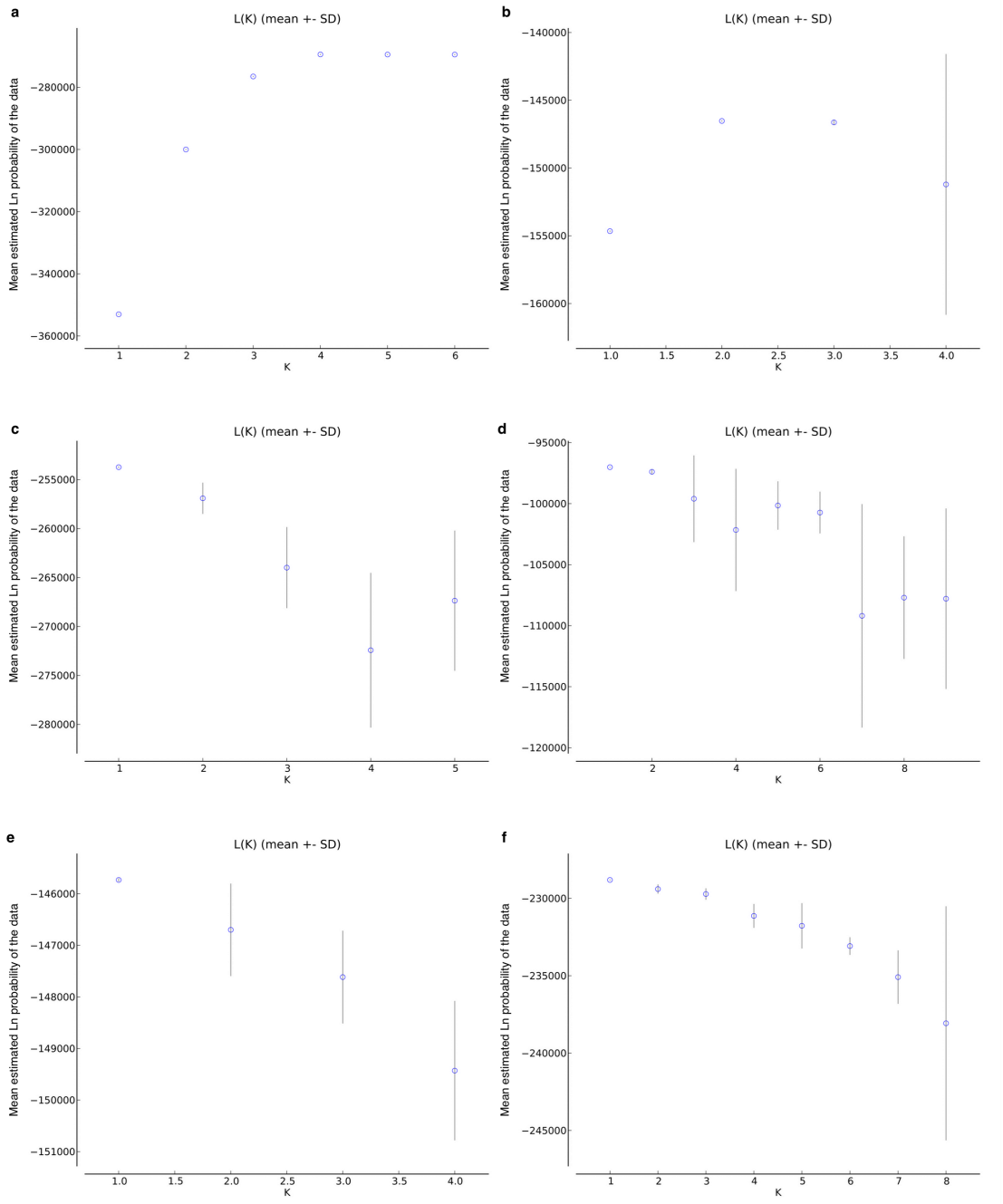


Figure 5.5. Estimated log probability of the data Pr($X|K$) for each value of K after repeated *Structure* runs. Results are presented for (a) all gentoo penguins without location priors; (b) southern gentoo penguins without location priors; (c) chinstrap penguins with location priors; (d) Adélie penguins with location priors; (e) king penguins with location priors; and (f) emperor penguins with location priors.

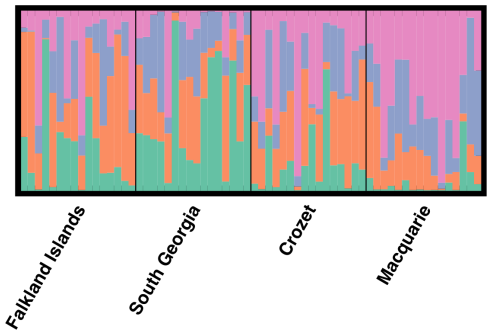


Figure 5.6. *Structure* plot for king penguin colonies without location priors and *K* = 4. Population structure was not visible when location priors were not used.

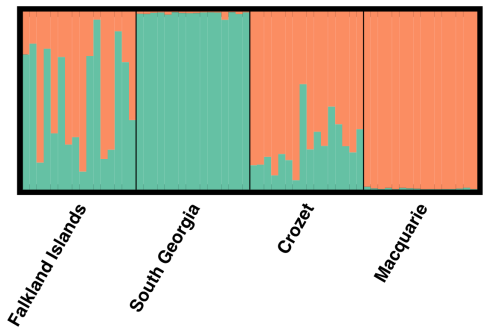


Figure 5.7. *Structure* plot for king penguin colonies with location priors and *K* = 2. The differentiation of South Georgia and Macquarie from one another and the other colonies became visible when location priors were used.

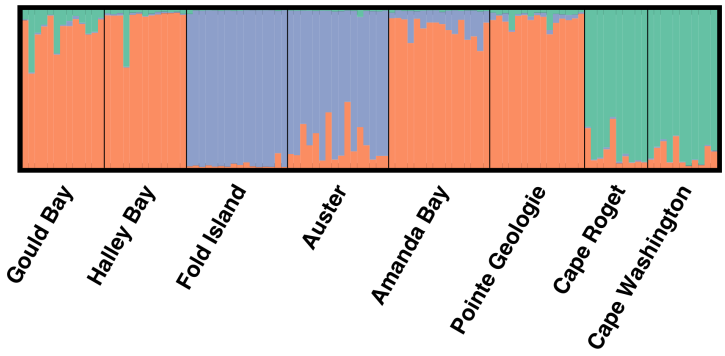


Figure 5.8. *Structure* plot for emperor penguin colonies with location priors and *K* = 3. The Ross Sea colonies and Mawson Coast colonies were differentiated (blue and green clusters, respectively) but the Weddell Sea colonies, Amanda Bay and Pointe Géologie appear admixed when *K* = 3.

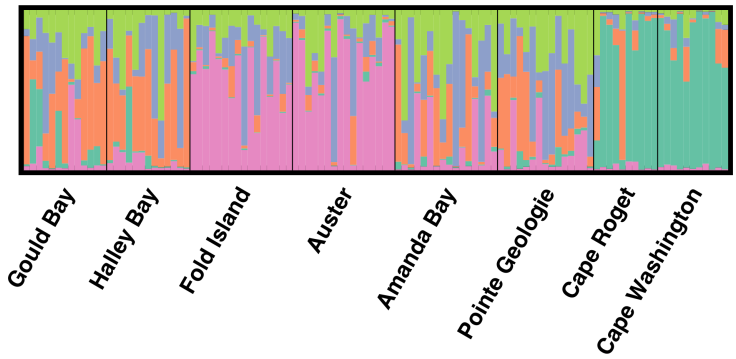


Figure 5.9. *Structure* plot for emperor penguin colonies without location priors and *K* = 5. The Ross Sea colonies and Mawson Coast colonies were differentiated (green and pink clusters, respectively), but the Weddell Sea colonies, Amanda Bay and Pointe Géologie appear admixed when location priors were not used.

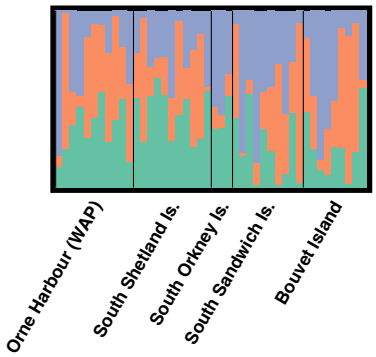


Figure 5.10. *Structure* plot for chinstrap penguin colonies without location priors and *K* = 3. Population structure was not discernible when location priors were not used.

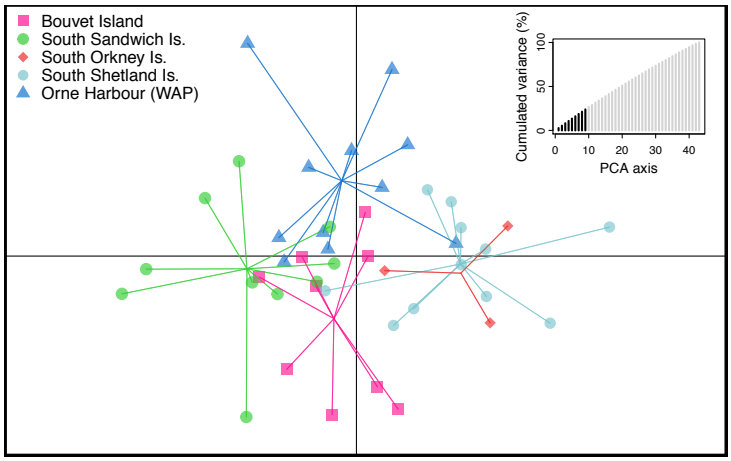


Figure 5.11. DAPC for chinstrap penguins with 9 PCs retained. The differentiation of the chinstrap penguins resembles that found with the program Structure when fewer principal components are retained. The South Sandwich Islands and Bouvet Island appear closely related (bottom left), the South Shetland and South Orkney Islands cannot be distinguished (bottom right), and Orne Harbour is marginally distinct (top).

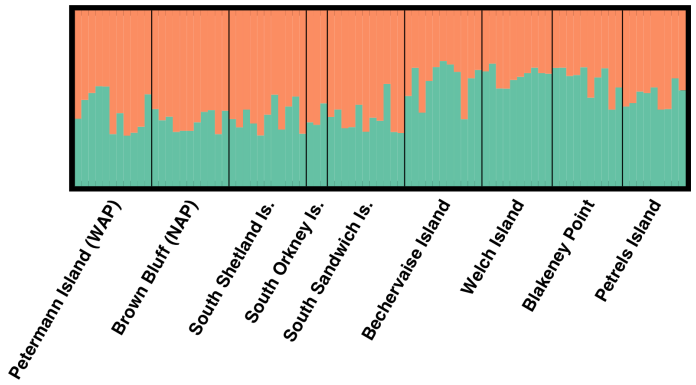


Figure 5.12. *Structure* plot for Adélie penguin colonies without location priors and *K* = 2. The differentiation of western and eastern colonies was much less distinct when location priors were not used.

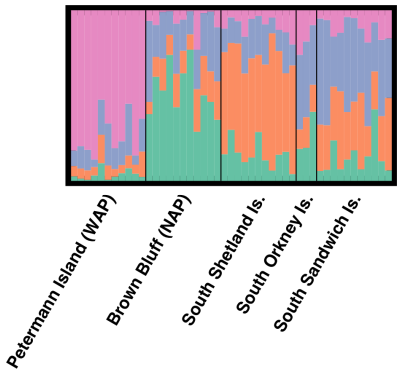


Figure 5.13. *Structure* plot for western Adélie penguin colonies with location priors and *K* = 4. When location priors were used, subtle differentiation was apparent between Petermann Island on the Antarctic Peninsula and all other colonies. Brown Bluff also appeared to be subtly differentiated from other colonies.

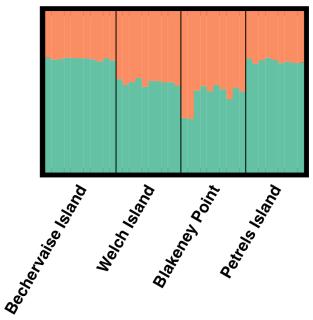


Figure 5.14. *Structure* plot for eastern Adélie penguin colonies with location priors and *K* = 2. There was no significant population differentiation between the eastern colonies even when using location priors.

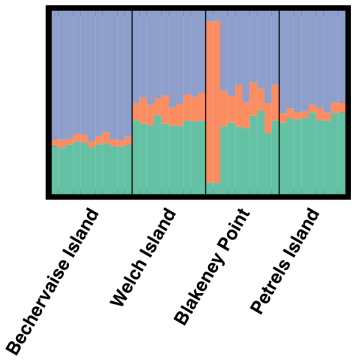


Figure 5.15. *Structure* plot for eastern Adélie penguin colonies with location priors and *K* = 3. Two individuals sampled at Blakeney Point were somewhat genetically differentiated and could be migrants from elsewhere.

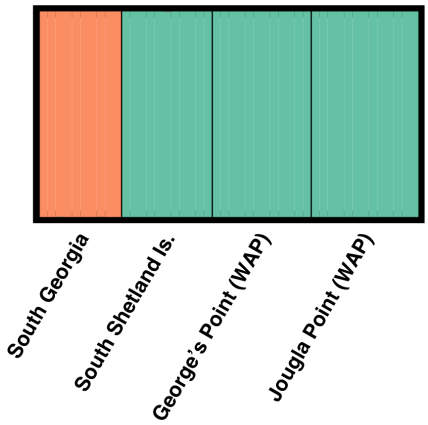


Figure 5.16. *Structure* plot for southern gentoo penguin colonies without location priors and *K* = 2. The strong genetic differentiation of South Georgia is apparent.

Investigating the effect of mismatch and indel thresholds on mapping success when using a species-specific and closely related reference genome

The question of whether the same settings should be used for short read mapping to a species-specific reference genome or a closely related reference genome ultimately comes down to whether the difference between the species is greater than or less than the differences observed between paralogs in the same species. When mapping to a species-specific genome, mapping parameters must be strict enough to prevent reads from mapping to multiple paralogous positions throughout the genome, whilst on the other hand, when mapping to a closely-related reference genome, parameters must be relaxed enough to allow mapping despite some evolutionary divergence of the genomes, whilst still preventing multiple mapping to paralogous regions. Using mapping parameters that are too strict, could result in reads only mapping successfully to conserved regions of the genome and many reads being discarded. Whilst the question of whether we have only mapped reads to conserved regions of the genome is impossible to answer without a reference genome for the species in question, we can look at the influence of mismatch thresholds on the proportion of reads that are uniquely mapped or multi-mapped.

To investigate this, a short analysis of mapping success was performed with the emperor and king penguin data. The top 10 individual libraries (ranked by the number of reads) from emperors and kings were mapped to the emperor penguin reference genome with *bwa-mem*, as in the methods section for this chapter, however a very low mismatch and indel penalty was given, to allow for more mismatches and indels during the mapping process. In the original analysis, reads were discarded after mapping if they had more than five mismatches and/or two indels, and they were also discarded if they were multi-mapped after applying these thresholds. Using this new dataset of permissively mapped reads, we can investigate the effect of changing the mismatch/indel filtering thresholds on how many reads are still multi-mapped.

Table 5.9. Results from permissively mapping reads to the emperor penguin reference genome.

Library	Total reads	Multi-mapped reads	Percentage of multi-mapped reads
Emperors			
EPDDU014	27351704	1334543	4.9
EPDDU012	15383125	799037	5.2
EPAUS011	14736201	736267	5.0
EPDDU007	14317452	731897	5.1
EPWIS021	13468658	716062	5.3
EPWIS011	14071328	692383	4.9
EPDDU001	12445886	640390	5.1
EPDDU008	11117490	590260	5.3
EPCWS003	11493272	578281	5.0
EPDDU009	10577672	527107	5.0
Kings			
KPMQ044	22294619	1267101	5.7
KPMQ037	13645277	713936	5.2
KPMQ049	12855355	626965	4.9
KPMQ016	12301320	599422	4.9
KPUVF012	10920369	545221	5.0
KPMQ017	10482993	519007	5.0
KPMQ023	9098848	506115	5.6
KPFORT009	10825891	494401	4.6
KPVP020	9293605	454754	4.9
KPFORT005	9056371	420626	4.6

After permissive mapping, around 5% of reads were multi-mapped in both species (Supplementary Table 5.9). Supplementary Figure 5.17 shows the proportion of reads which were considered as multi-mapped, and which then became uniquely mapped by specifying a mismatch/indel threshold. For example, if a read mapped to locus A with one mismatch and locus B with five mismatches, then a mismatch threshold of eight would retain both mapping locations and the read would be discarded, however if the mismatch threshold is set to three, then the locus will appear to be uniquely mapped to locus A and will be retained. Thus, more reads appear to be uniquely mapped the lower the threshold is set.

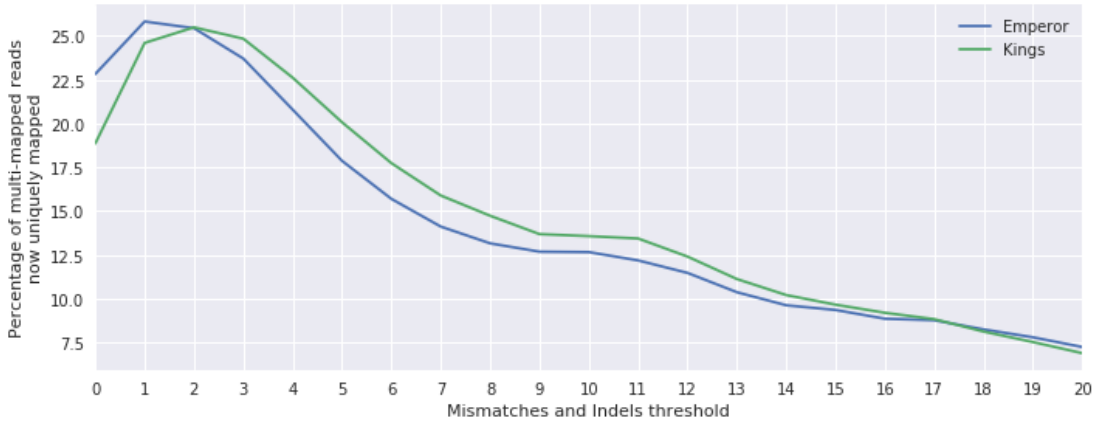


Figure 5.17. The influence of mismatch/indel thresholds on mapping for emperor and king penguins. The percentage of multi-mapped reads which became uniquely mapped after applying a mismatch/indel threshold is shown to be highly similar for emperor and king penguins.

Supplementary Figure 5.17 shows that both species have a very similar distribution of the number of mismatches and indels found in multi-mapped reads, and thus using differential thresholds for the two species is unlikely to have yielded many benefits. Since this graph is looking at the proportion of reads that were originally multi-mapped, which was around 5% per individual, using differential thresholds for emperors and kings would have resulted in a change of less than 1% in the number of reads that were considered as uniquely mapped and thus retained for SNP calling.

Furthermore, looking at just the reads which were uniquely mapped after the permissive mapping used in this analysis, we can see that the frequency of mismatches between reads and the reference genome is very similar between emperors and kings (Supplementary Figure 5.18). More emperor reads are exact matches to the reference genome than kings (as expected) and king reads have, on average, more mismatches, but very few uniquely mapping reads have more than five mismatches in either species despite using permissive mapping in this analysis. Thus changing our mismatch parameter to allow reads with more than five mismatches in our original analysis would have resulted in a negligible increase in the number of uniquely mapping reads retained.

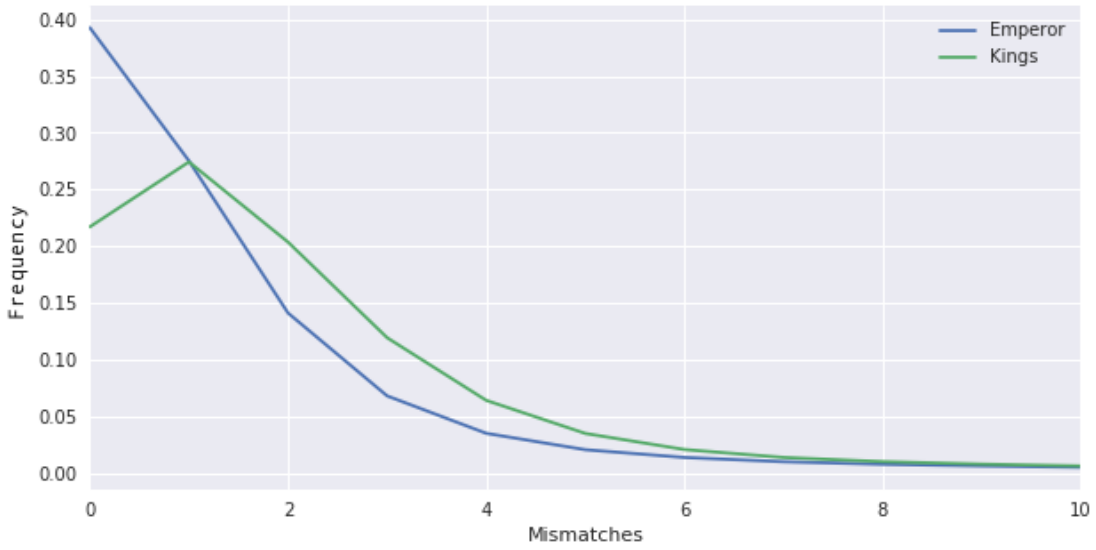


Figure 5.18. The number of mismatches in the uniquely mapped reads for emperor and king penguins.

Finally, looking at the number of indels in the uniquely mapping reads, in both species over 80% of reads had no indels, and the majority of the remaining reads had just one indel, with a very small percentage having two or more indels (Supplementary Figure 5.19). Thus, increasing our indel threshold above two would not have affected the number of reads that were retained whether we were mapping to a species-specific or closely related reference genome.

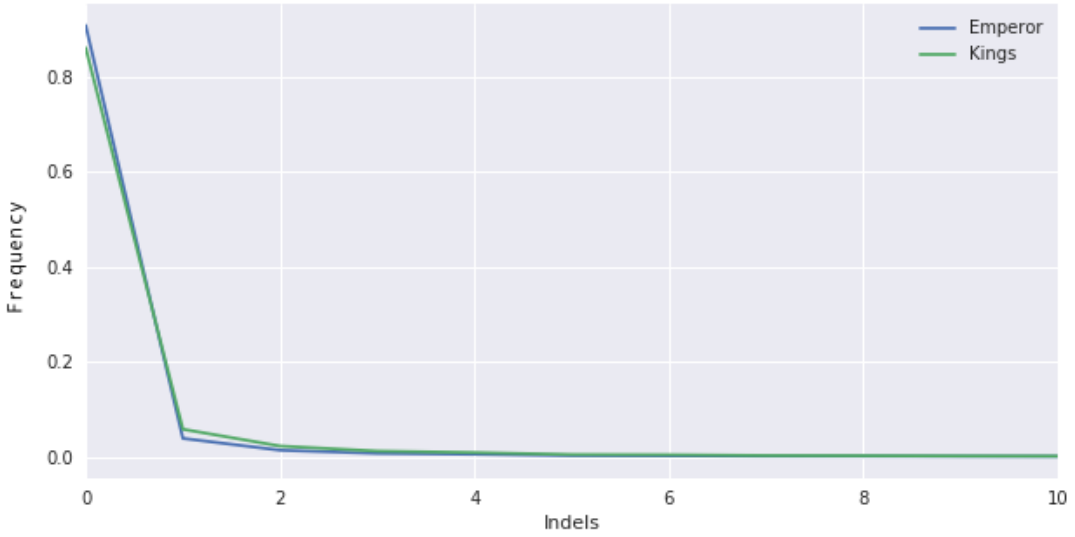


Figure 5.19. The number of indels in the uniquely mapped reads for emperor and king penguins.

Overall, this analysis shows that the number of uniquely mapping reads and the number of mismatches and indels in uniquely mapped reads was very similar whether mapping to a species-specific reference genome or a closely-related reference genome. Using more permissive mismatch/indel thresholds in the king penguin would have retained fewer than 1% more reads as uniquely mapping, and would thus have been very unlikely to affect SNP calling or any of our downstream analysis. The thresholds of 5 mismatches and/or 2 indels used in our analysis and that of Catchen *et al.* (2013) seems highly appropriate to retain the majority of uniquely mapped reads whether mapping to a species-specific or closely related reference genome.

Chapter 6: Conclusions

Each chapter, which is presented in manuscript format, is accompanied by a discussion. Here I briefly bring together the main themes that I have investigated and expand on those not covered in previous chapters. The aim of the thesis was to identify those factors that have created the patterns of intraspecific genetic variation in five species of penguin from Antarctica and the sub-Antarctic. I have investigated both the effect of historical glaciation on the habitat that penguins occupied and, using a comparative framework, the effects of at-sea range, habitat variability, habitat continuity, and oceanic fronts in determining population genetic differentiation.

6.1 Glacial history

I show in chapter 2 that emperor penguin populations around Antarctica were adversely affected by the last glacial period (110 – 12 kya). The total effective population size during the LGM was 11-fold smaller than the present, and the three mitochondrial lineages shared a most recent common ancestor during this time, suggesting that three small and isolated populations survived in ice age refugia, possibly associated with polynyas. Population expansion occurred in line with retreating ice and with increasing primary productivity 12.5 – 9 kya. One of the refugial populations was likely located in the Ross Sea, where the majority of the "Ross Sea" haplotypes were recovered. The other two lineages were found mixed throughout colonies in East Antarctica and the Weddell Sea, suggesting a high degree of post-glacial connectivity between populations.

This is mirrored by an almost identical pattern in Adélie penguins; Lambert *et al.* (2002) and Ritchie *et al.* (2004) found two mitochondrial lineages in Adélie penguins, one of which is largely restricted to the Ross Sea, and another lineage which is widespread around Antarctica (Clucas *et al.* 2014; Younger *et al.* 2015a). The two lineages share a most recent common ancestor during the last glacial period, similar to emperor penguins. Adélie penguin numbers were also much reduced during the last glacial period (Clucas *et al.* 2014; Ritchie *et al.* 2004; Younger *et al.* 2015a) and have undergone significant expansions in line with post-glacial ice retreat.

Therefore the impact of the last glacial period on the two fully Antarctic penguins appears to have been similar: both had much reduced population sizes, with only two Adélie and three emperor penguin populations surviving. The Ross Sea appears to have been an important refuge for both species, possibly due to a persistent polynya that supported primary productivity in the north-western Ross Sea (Brambati *et al.* 2002; Thatje *et al.* 2008). Habitat availability and primary productivity appears to have limited both species during the LGM and for a few thousand years

afterwards, as populations did not begin to expand until ice sheets had retreated to close to their current extents. This is likely because glaciation reduced the amount of ice-free ground available for Adélie penguins and current sea ice extents are thought to be near optimal for emperor penguins (Ainley *et al.* 2010).

In chapter 3 I found no evidence for distinct mitochondrial lineages in king penguins. This suggests that during the last glacial period, although king penguin population sizes were likely reduced (Trucchi *et al.* 2014), continued migration of individuals among populations likely prevented genetic divergence. Previous work showed a similar pattern in chinstrap penguins (Clucas *et al.* 2014), with chinstrap population size being much smaller during the last glacial period, with no evidence of distinct mitochondrial lineages.

It is impossible to estimate how many populations of chinstrap and king penguins were exchanging migrants during the LGM for a number of reasons. Firstly, the glacial history of most maritime and sub-Antarctic islands is not well known, although there is evidence that some were not fully ice-covered (Hodgson *et al.* 2014). Nevertheless, estimating the amount of habitat that was available to penguins remains difficult. Sea level during the LGM was 120 – 135m lower than at the present (Clark & Mix 2002), and so the remains of ancient colonies are likely to have been submerged by post-glacial sea level rise. Finally, the high level of contemporary migration between colonies may have eroded the signals of ancestral populations (Petit *et al.* 2003). Populations close to the location of ancestral refuges are expected to have higher diversity than post-glacially founded colonies as a result of founder effects (Hewitt 2000), but all extant populations of chinstrap and king penguins sampled had similar genetic diversity, likely as a result of high contemporary gene flow (chapter 5).

The preference of chinstrap and king penguins for breeding in habitats with little to no sea ice, respectively, likely explains their lack of distinct mitochondrial lineages. Chinstrap and king penguins may have shifted their ranges to the edge or beyond the extent of LGM sea ice, perhaps remaining on ice-free sub-Antarctic islands or colonising areas around the coast of South America, New Zealand and the Patagonian shelf, which may have been exposed by falling sea levels (Thatje *et al.* 2008). Breeding above the extent of LGM sea ice would have allowed them to continue to disperse widely during the non-breeding season and therefore maintain gene flow with any other remaining colonies.

Glacial conditions also appear to have affected the gentoo penguin, with the biggest impacts on southern gentoo penguins as a result of their higher latitudinal distribution. Given the high genetic diversity of Kerguelen and Falkland Island gentoo penguin populations (chapter 5, Figure 5.4), it seems highly likely that these populations survived the last glacial period *in situ*, and

indeed, both Kerguelen and the Falkland Islands are not thought to have been extensively glaciated (Hodgson *et al.* 2014). As further evidence to this hypothesis, previous work showed that the Falkland Island gentoo penguin population has been relatively stable since *ca.* 25 kya, with no evidence for a genetic bottleneck (Clucas *et al.* 2014). The distance between Kerguelen and the Falkland Islands, and the coastal life-style of the gentoo penguin would appear to explain the incipient allopatric speciation occurring between these populations. The timing of the divergence has not been investigated and may correspond to changing climatic conditions.

The lower genetic diversities of the southern gentoo penguin populations on South Georgia and in the maritime Antarctic (chapter 5, Figure 5.4) are very likely due to the impact of glaciation in these higher latitude regions, with habitat only recently becoming available for colonisation. Indeed, southern gentoo penguins went through a population expansion as habitat became available after the LGM (Clucas *et al.* 2014). Currently, it is not possible to determine the exact sequence of colonisation of these regions. One possibility is that a small population of southern gentoos from the Falkland Islands established in the region of the South Shetland or South Orkney Islands during or just after the LGM, and then expanded to colonise South Georgia to the north and the Antarctic Peninsula to the south, as suggested in my previous work using mitochondrial DNA (Clucas *et al.* 2014; Levy *et al.* 2016) (Figure 6.1a). The slightly higher genetic diversity of the South Georgia population relative to colonies further south could be due to a low level of gene flow from the Falkland Islands into South Georgia. Alternatively, the South Georgia population could have been founded during or just after the LGM by migrants from the Falkland Islands, and migrants from South Georgia subsequently colonized the islands of the maritime Antarctic and Antarctic Peninsula after ice retreated (Figure 6.1b). If South Georgia were the ancestral population among southern gentoos then this would explain the higher genetic diversity recorded there, but this is incongruent with the mitochondrial tree of Levy *et al.* (2016). This second scenario appears to be the most likely; South Georgia was not heavily glaciated during the LGM and was beyond the northern extent of LGM summer sea ice, probably making it available for colonisation before the maritime Antarctic islands (Gersonde *et al.* 2005; Hodgson *et al.* 2014), and in addition, the mitochondrial tree of Levy *et al.* (2016) was not well supported.

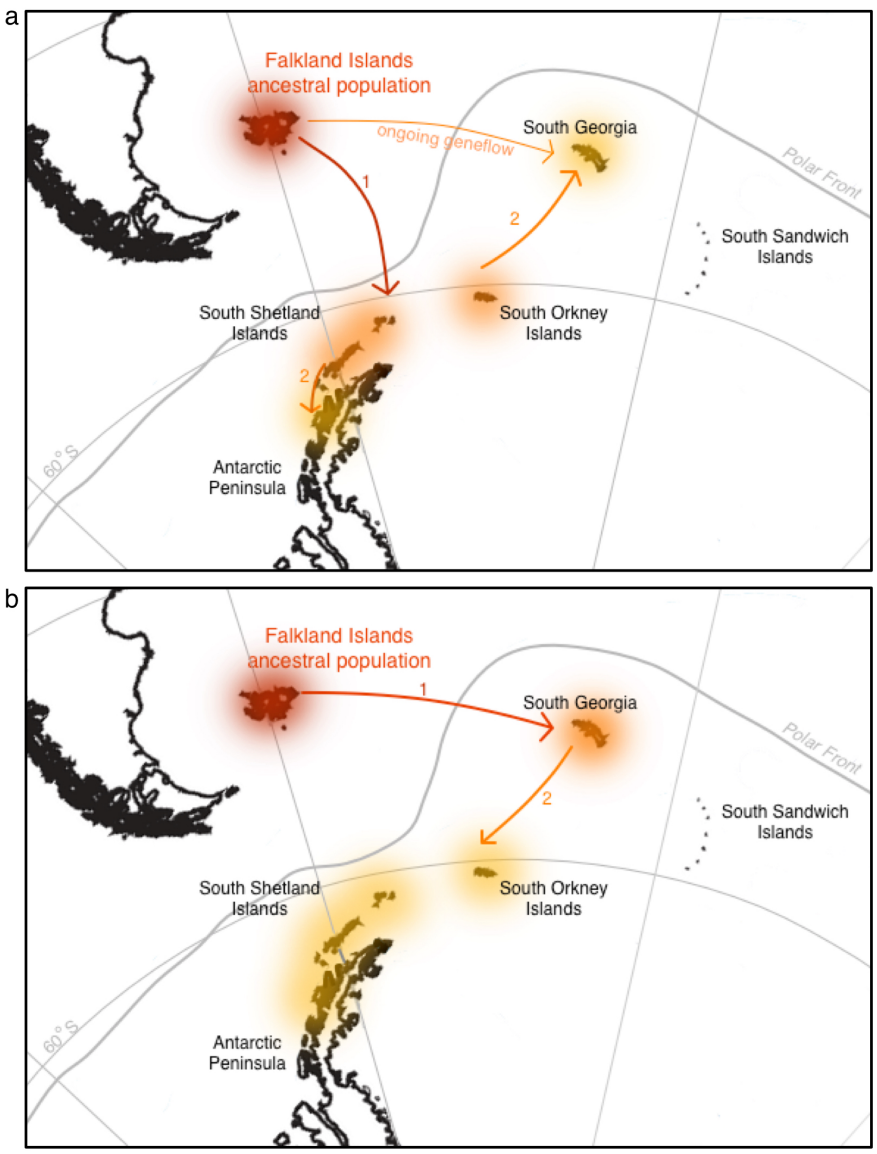


Figure 6.1. Possible routes of colonisation of the maritime Antarctic by southern gentoo penguins. In scenario (a) migrants from the ancestral population in the Falkland Islands cross the Polar Front to colonise either the South Shetland or South Orkney Islands, and then expand to the north and south from there, as suggested by Levy *et al.* (2016). In scenario (b) migrants from the ancestral population in the Falkland Islands cross the Polar Front to colonise South Georgia, and then expand to the southwest.

In summary, all Antarctic and sub-Antarctic penguins studied were adversely affected by reduced habitat availability during the last glacial period and have undergone expansions during the Holocene. Adélie and emperor penguins, with their high latitudinal distributions and reliance on sea ice, were isolated within glacial refugia, resulting in the evolution of distinct mitochondrial lineages (Clucas *et al.* 2014; Ritchie *et al.* 2004; Younger *et al.* 2015a; Younger *et al.* 2015b). Chinstrap and king penguin populations may have maintained connectivity due to their higher latitudinal distribution, although their numbers were also reduced (Clucas *et al.* 2014; Clucas *et al.*

2016; Trucchi *et al.* 2014). Northern gentoo penguins on the Falkland Islands appear to have been little affected, but southern gentoo penguins went through significant post-glacial expansion as new habitat became available (Clucas *et al.* 2014; Levy *et al.* 2016). An avenue for further investigation would be to determine the responses of the other sub-Antarctic species such as macaroni and royal penguins, and southern and eastern rockhopper penguins, which also inhabit the sub-Antarctic and are highly dispersive (Borboroglu & Boersma 2013). They may too have gone through population size reductions due to loss of habitat during the LGM but may have maintained connectivity between populations, similar to chinstrap and king penguins.

6.2 Dispersal

In chapters 3 – 5 I reviewed the factors determining the dispersal of individuals among populations of emperor, king, gentoo, chinstrap, and Adélie penguins. The at-sea range of the species, that is, whether it is a pelagic or coastal species, was the main determinant of intraspecific genetic variation. At-sea range is an internal (behavioural) barrier to gene flow, being the result of the foraging niche of the species. The pelagic foraging niches of emperor, king, chinstrap and Adélie penguins, and their large at-sea ranges appear to facilitate gene flow across thousands of kilometres, resulting in very low levels of population differentiation within these species. In comparison, the coastal life-style and inshore foraging niche of the gentoo penguin restricts gene flow among gentoo penguin populations separated by open ocean, leading to high levels of intraspecific genetic variation. The Kerguelen population of gentoo penguins is highly divergent, and the differentiation of it from the Falkland Islands population may be the result of a dispersal-vicariance event. The genetic differentiation of the Kerguelen population should be recognised by taxonomic revision, elevating the population to sub-species level at least.

Other barriers to gene flow identified here were the Polar Front, an external barrier, and natal philopatry, a behavioural barrier. Spatial and temporal habitat availability did not explain the patterns of genetic differentiation, as the king penguin, which inhabits the most predictable and least continuous habitat, was found to be at least as dispersive as the emperor penguin, inhabiting the more continuous but ephemeral fast-ice habitat around Antarctica.

The Polar Front appears to restrict gene flow in both gentoo and king penguins which reside above and below the Polar Front. It has been recognised as a barrier to dispersal across a range of species (Rogers 2012; Shaw *et al.* 2004; Thornhill *et al.* 2008), possibly as a result of the different ecological conditions, such as temperature, salinity and species composition, on either side. There was no evidence in this study of local adaptation in king penguins either side of the front, as no outlier loci could be identified. However our SNP dataset only included 5,154 SNPs, and genomic

islands of adaptation under high gene flow may be narrow (Feder *et al.* 2012). Therefore a higher density of SNPs may be necessary to identify signals of adaptation with gene flow in king penguins. The same is true for emperor, chinstrap and Adélie penguins, in which gene flow is high and few or no outlier loci could be identified in this study. In gentoo penguins, with little migration between populations, 452 outlier loci were identified and warrant further investigation. Local adaptation could be an additional isolating mechanism between gentoo penguin populations that has so far not been studied. Indeed, gentoo penguins occupy a very wide range of habitats, from the temperate Falkland Islands to the maritime Antarctic climate on the Antarctic Peninsula. Local adaptation would appear highly likely given the range of conditions they experience. Natal philopatry may also explain some of the genetic differentiation within gentoo penguins, as colonies less than 50km apart from one another and not separated by large stretches of open water were weakly genetically differentiated.

Interestingly, the high levels of contemporary gene flow in emperor penguins are likely eroding the effects of glacial history on emperor penguin population structure. In chapter 2, I found high levels of population differentiation between the Ross Sea colonies and all other colonies using mitochondrial DNA. Yet in chapter 4, using SNPs, I found a much weaker signal of genetic differentiation of the Ross Sea emperor penguins. This suggests that the populations that originally underwent divergence in glacial refugia have come into secondary contact since the ice retreated, and have inter-bred. The signal of the glacial refuge has been preserved in mitochondrial DNA as it is a non-recombining locus, and so the "Ross Sea" haplotypes still occur at a higher frequency within the Ross Sea populations. However the SNPs, which are recombinant, appear to have been mostly homogenised across populations through the migration of individuals into and out of the Ross Sea. It would be interesting to investigate Ross Sea Adélie penguin population structure using SNPs, to see if the erosion of the glacial signal is occurring in Adélie penguins as well. Given their highly dispersive behaviour, it would seem likely.

In this thesis I have used both mitochondrial and nuclear genetic loci (SNPs) to identify the processes governing intraspecific population differentiation in Antarctic and sub-Antarctic penguins. This approach has shown that: (1) different genetic loci are useful to capture processes occurring at different time-scales; that (2) a comparative framework with closely related and sympatric species allows the mechanisms determining population structure to be identified; and that (3) dispersal within these species is mainly determined by their at-sea range. It is hoped that the high degree of dispersal and metapopulation boundaries within emperor, chinstrap, king and Adélie penguins will be taken into account when designing monitoring programmes for these species, as a high degree of immigration and emigration is likely to skew colony responses to habitat variability. Likewise, dispersal should be taken into account when studies attempt to

forecast these species' responses to climate change, as their flexibility and potential to shift their ranges has likely been underestimated to date. Climate change poses a grave risk to all species (Bost *et al.* 2015; Jenouvrier *et al.* 2014; Lynch *et al.* 2012), but with a better understanding of their dispersal behaviours, better monitoring and conservation action plans should be possible.

Appendices

Appendix A

Global Change Biology

Global Change Biology (2015), doi: 10.1111/gcb.12882

Too much of a good thing: sea ice extent may have forced emperor penguins into refugia during the last glacial maximum

JANE L. YOUNGER[1,2*], GEMMA V. CLUCAS[3,4*], GERALD KOOYMAN[5], BARBARA WIENECKE[6], ALEX D. ROGERS[4], PHILIP N. TRATHAN[7], TOM HART[4†] and KAREN J. MILLER[8,9†]

[1]*Institute for Marine and Antarctic Studies, University of Tasmania, Private Bag 129, Hobart 7001 Tas., Australia,* [2]*Australian School of Advanced Medicine, Macquarie University, 2 Technology Place, Sydney 2109 NSW, Australia,* [3]*Ocean and Earth Sciences, University of Southampton Waterfront Campus, Southampton SO14 3ZH, UK,* [4]*Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK,* [5]*Scripps Institution of Oceanography, University of California San Diego, San Diego CA, USA,* [6]*Australian Antarctic Division, 203 Channel Highway, Kingston 7050 Tas., Australia,* [7]*British Antarctic Survey, High Cross, Madingley Road, Cambridge CB3 0ET, UK,* [8]*Australian Institute of Marine Science, The UWA Oceans Institute, 35 Stirling Highway, Crawley, WA 6009, Australia,* [9]*School of Biological Sciences, University of Tasmania, Private Bag 5, Hobart 7001 Tas., Australia*

Abstract

The relationship between population structure and demographic history is critical to understanding microevolution and for predicting the resilience of species to environmental change. Using mitochondrial DNA from extant colonies and radiocarbon-dated subfossils, we present the first microevolutionary analysis of emperor penguins (*Aptenodytes forsteri*) and show their population trends throughout the last glacial maximum (LGM, 19.5–16 kya) and during the subsequent period of warming and sea ice retreat. We found evidence for three mitochondrial clades within emperor penguins, suggesting that they were isolated within three glacial refugia during the LGM. One of these clades has remained largely isolated within the Ross Sea, while the two other clades have intermixed around the coast of Antarctica from Adélie Land to the Weddell Sea. The differentiation of the Ross Sea population has been preserved despite rapid population growth and opportunities for migration. Low effective population sizes during the LGM, followed by a rapid expansion around the beginning of the Holocene, suggest that an optimum set of sea ice conditions exist for emperor penguins, corresponding to available foraging area.

Keywords: Antarctica, *Aptenodytes forsteri*, climate change ecology, molecular ecology, paleoecology, phylogeography, polynya, Ross Sea

Received 24 July 2014; revised version received 13 November 2014 and accepted 30 December 2014

Introduction

Genetic data from both modern and subfossil samples, paleoecological niche modeling, and fossil evidence have become vital tools for reconstructing demographic histories (e.g., woolly mammoths (*Mammuthus primigenius*) (Nogués-Bravo *et al*., 2008) and lions (*Panthera leo*) (Barnett *et al*., 2014)). Indeed, such studies have shown that species' patterns of genetic diversity and distribution have varied dramatically under different climatic regimes (Carstens & Richards, 2007). Climatic shifts have been one of the major drivers of species' range shifts, fluctuations in abundance, species extinctions, and also in the formation of genetically distinct populations (Hewitt, 1996). As climate change and habitat degradation potentially take us into the 6th mass extinction (Barnosky *et al*., 2011), it is critical that we understand how species have coped with change in the past to be able to assess their likely responses and resilience to future climate change (Hoelzel, 2010).

Emperor penguins (*Aptenodytes forsteri*) are an iconic Antarctic species whose population genetic structure has not been studied to date. We know little about dispersal among colonies or how historical climate change may have affected their range and abundance. Thus, we have limited capacity to predict how these birds may fare in the future. Projections for continent-wide declines of emperor penguins have been made based on the demographic responses of the Pointe Géologie colony to changes in sea ice conditions (Barbraud & Weimerskirch, 2001; Jenouvrier *et al*., 2009, 2012, 2014;

Correspondence: Jane Younger, tel. +61 439 869 364, e-mail: Jane.Younger@utas.edu.au; Gemma Clucas, tel. +44 1865 281991, e-mail: gemma.clucas@noc.soton.ac.uk

*These authors have contributed equally to the manuscript and are joint first and corresponding authors.

†These authors have contributed equally to the manuscript and are joint last authors.

Ainley *et al.*, 2010). However, decadal monitoring data are only available for this single site out of 46 known emperor penguin colonies; as such, the climate change responses of emperor penguins across their entire distribution and over millennial timescales are currently unknown (Ainley *et al.*, 2010).

Emperor penguins are highly reliant on sea ice throughout most of their breeding cycle, and mating and incubation takes place on land-fast sea ice in most of the known colonies (Fretwell *et al.*, 2012, 2014). During the breeding season, emperor penguins feed on prey that is also sea ice dependent (Gales *et al.*, 1990). Significant areas of open water exist year-round within the Antarctic sea ice zone in the form of leads and polynyas (Zwally *et al.*, 1985 and references therein). These areas are often important in providing emperor penguins access to their underwater foraging habitat when the fast ice extends far from their colonies (Dewasmes *et al.*, 1980). Polynya formation is driven by either upwelling of circumpolar deep water or by the outflow of katabatic winds that push sea ice away from the coastline (Martin, 2001). Polynyas are associated with enhanced primary production, as the reduction in sea ice volume facilitates an earlier spring melting of sea ice and a coincident earlier start in photosynthetic primary productivity (Martin, 2001). Some polynyas are permanent features of the sea ice zone and create areas of hyperproductivity, such as the Ross Sea polynya (Smith & Gordon, 1997), while most are smaller and ephemeral features depending on wind stresses and currents.

Changes in the extent and duration of sea ice around Antarctica show highly regionalized trends with some areas increasing or remaining stable while others are decreasing (Zwally *et al.*, 2002; Vaughan *et al.*, 2013); this has an effect on the population dynamics of emperor penguins as both positive and negative sea ice anomalies can result in negative population growth rates at the local scale (Massom *et al.*, 2009; Ainley *et al.*, 2010; Barbraud *et al.*, 2011; Jenouvrier *et al.*, 2014). Despite uncertainties over the rate and extent of ice loss that will occur around Antarctica, all climate models project a reduction in the extent and duration of Antarctic sea ice by the end of the century (Collins *et al.*, 2013). As sea ice declines, we might expect emperor penguins to be disadvantaged by a lack of breeding habitat (Jenouvrier *et al.*, 2014), unless they have the capacity to alter their preferred choice of breeding site or their range by colonizing new areas. Recent studies have shown more plasticity than expected in the locations of breeding colonies; satellite imagery suggests that colonies where the fast ice is inadequate at the onset of the breeding season relocated or partially relocated onto ice shelves or icebergs (Fretwell *et al.*, 2014; LaRue *et al.*, 2014). However, as sea ice declines, emperor penguins may also have to contend with altered prey availability and face new threats from predators as changing conditions differentially affect species at other trophic levels (Trathan *et al.*, 2011).

During the last glacial maximum (LGM, 19.5–16 kya), the winter sea ice extent was approximately double the present-day values, and seasonal variation in sea ice extent is thought to have been greater (Gersonde *et al.*, 2005). It is unclear how this would have affected emperor penguins. Thatje *et al.* (2008) suggested that they may have migrated with the sea ice to lower latitudes, staying within energetic migration thresholds of the ice edge, and could have maintained breeding populations around Antarctica by foraging in the marginal ice zone at the sea ice edge. Alternatively, they could have remained associated with polynyas. Sediment cores suggest the existence of LGM polynyas in several locations, including the northwestern Ross Sea, the southeastern Weddell Sea off Dronning Maud Land, and the northwestern Weddell Sea (Mackensen *et al.*, 1994; Brambati *et al.*, 2002; Thatje *et al.*, 2008; Smith *et al.*, 2010). In either case, reductions in overall primary productivity within what is today's seasonal sea ice zone (Domack *et al.*, 1998; Kohfeld *et al.*, 2005) would likely have been detrimental to emperor penguin populations (Ainley *et al.*, 2010).

Little is known about the level of natal philopatry or migration among emperor penguin colonies. Understanding philopatry is particularly important in light of population models that suggest that emperor penguins may be declining as a result of local climatic shifts (Jenouvrier *et al.*, 2009, 2014). High emigration rates are conceivable among emperor penguin colonies; satellite tracking has shown that they travel thousands of kilometers on their juvenile journeys, often passing other colonies (e.g., Kooyman *et al.*, 1996; Wienecke *et al.*, 2010; Thiebot *et al.*, 2013). Generally, philopatry is high among penguins (Saraux *et al.*, 2011; Dehnhard *et al.*, 2014), but population structure is absent in many species (e.g., Chinstrap penguins (*Pygoscelis antarctica*) (Clucas *et al.*, 2014)) as even low levels of migration can be sufficient to homogenize populations (Hartl & Clark, 1997).

We analysed the population structure among eight extant emperor penguin colonies (Fig. 1) using mitochondrial DNA sequences and inferred population trajectories during and since the LGM using a combination of ancient and modern DNA sequences in a Bayesian coalescent framework (Drummond *et al.*, 2005). This method reconstructs past changes in abundance by estimating the genealogy from sequence data and co-estimating the effective population size at different points in time, where the effective population size is the number of individuals that contribute offspring

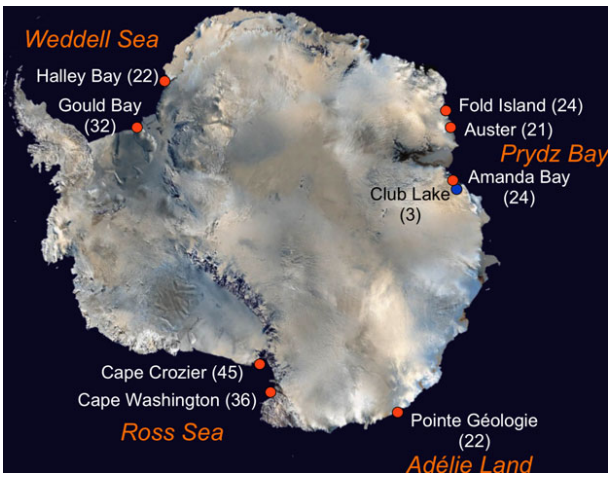


Fig. 1 Sample locations of emperor penguin colonies. The number of sequences obtained from each location is shown in brackets. Red points indicate the origin of modern samples, and the blue point indicates the origin of the subfossil samples.

to the descendant generation (Pybus *et al.*, 2000). We aimed to: (1) investigate how emperor penguin populations were affected by sea ice conditions during and following the LGM; and (2) to test the hypothesis that emperor penguins comprise one panmictic population as a result of the high dispersal of individuals after fledging and the lack of obvious ecological barriers to dispersal around the Antarctic coastline.

Materials and methods

Sample collection

Skin tissue of dead emperor penguins was collected from Halley Bay (see Fig. 1 for all sample locations) in November 2012 and transported frozen to the United Kingdom, where it was transferred to 90% ethanol and stored at −20 °C. Blood samples were collected from Gould Bay in December 2013 and transported to the United Kingdom at ambient temperature in RNAlater (Life Technologies, Carlsbad, CA, USA) and then stored at −20 °C. Shed feathers were collected from the Ross Sea between 2010 and 2012 and were transported and stored at −20 °C. Shed feathers were collected at least 10 m apart to minimize sampling the same bird. Pectoral muscle biopsies were collected from dead chicks at Fold Island in September 2010, from Pointe Géologie in December 2010 and from Amanda Bay in December 2012 and 2013. Biopsies were immediately placed in 90% ethanol and stored at −20 °C. Whole dead chicks were collected from Auster in September and October in 1993 and 1994 and transported and stored at −20 °C. Bones from the subfossil remains of three penguins were collected at Club Lake in January 2013 and stored at −80 °C. Club Lake is an ice-free area in the Vestfold Hills which is currently unoccupied by penguins. The nearest extant colony is Amanda Bay, 95 km away.

Where blood samples were taken, one handler seized the upper body with both hands and restrained the flippers, with the bird's head placed under the arm of the handler to prevent biting and minimize stress (Le Maho *et al.*, 1992). The second handler took blood from the brachial vein using a 25-G needle and 1-mL syringe. Total restraint time was generally two to three, but occasionally four, minutes. The bird was then released at the edge of the colony. Sampling was conducted under permits from the UK Foreign and Commonwealth Office, the US National Science Foundation, and the Australian Antarctic Division. Each of these permits was issued following independent ethical review of the sampling. All sampling was carried out in accordance with UK Home Office guidelines and also received ethical approval from the University of Oxford, British Antarctic Survey, and Australian Antarctic Division. The radiocarbon ages, expressed here as years BP (i.e., before 1950), of the Club Lake remains were determined using accelerated mass spectrometry by GNS Science Rafter Radiocarbon National Isotope Centre, New Zealand. The apparent ages were corrected for the marine-carbon reservoir effect (Gordon & Harkness, 1992) using the calibration program CALIB 7.0 (St Ui & Reimer, 1993).

DNA extraction, amplification, and sequencing

Genomic DNA (gDNA) was extracted from modern samples with the QIAGEN DNeasy blood and tissue kit. The manufacturer's protocols for blood and tissue samples were followed with the following modifications to the digestion step: for blood samples, 30 μL of proteinase K was used and the digestion time was 3 h; for tissue samples, 40 μL of proteinase K and an additional 10 μL of 1 M dithiothreitol (skin samples only) was used with an incubation time of 32 h. All samples were treated with either 1 μL RNase A (QIAGEN Venlo, Limburg, the Netherlands) or 1 μL Riboshredder (Epicentre) according to the manufacturers' instructions. DNA was eluted in 100 μL of elution buffer following an incubation of 5–20 min. For subfossil samples, ~50 mg of bone was decalcified in 0.5 M EDTA/0.001% Triton X100 at 56 °C for 48 h and then extracted using a standard phenol–chloroform protocol with ethanol precipitation and a final elution volume of 30 μL. The subfossil samples were extracted in a physically isolated laboratory which had not been used previously for avian samples to minimize the risk of contamination. The mitochondrial hypervariable region (HVR) and cytochrome B (CytB) were sequenced in all modern and ancient DNA samples. HVR is a rapidly evolving region of the mitochondrial genome and so is suitable for investigations of recent demographic history, while CytB is a conserved gene and can hence give information about longer-term demographic history (Baker & Marshall, 1997). HVR was amplified in all modern samples using primers F-0225 and R-INR (all primer sequences can be found in Table S1). The reaction mix consisted of 7.5 μL of PCR Master Mix (QIAGEN), 0.2 μM of each primer, and 5–10 ng gDNA, made up to 15 μL with ddH_2O. Thermocycling conditions were as follows: 94 °C for 3 min, 35 cycles of 94 °C for 30 s, 59.5 °C for 45 s, 72 °C for 1 min, followed by an extension period of 72 °C for 10 min. Occasionally, double bands were

apparent when the PCR product was visualized by electrophoresis. For these individuals, the shorter 755-bp band was extracted from the gel and purified using QIAGEN or Promega gel extraction kits following the manufacturer's instructions. For Fold Island, Amanda Bay, Auster, and Pointe Géologie colonies, CytB was amplified using primers B1 (Kocher *et al*., 1989; Baker *et al*., 2006) and B6 (Baker *et al*., 2006) with a reaction mix consisting of 7.5 μL of GoTaq Green Master Mix (Promega, Madison, WI, USA), 0.2 μM of each primer, and 5–10 ng gDNA, made up to 15 μL with ddH_2O. Thermocycling conditions were as follows: 95 °C for 1 min, 35 cycles of 95 °C for 20 s, 52 °C for 40 s, 72 °C for 50 s, then 72 °C for 5 min. For the Cape Washington, Cape Crozier, Gould Bay, and Halley Bay samples, primers CytB-F1 and CytB-R1 were used with a reaction mix consisting of 7.5 μL of PCR Master Mix (QIAGEN), 0.2 μM of each primer, and 5–10 ng gDNA, made up to 15 μL with ddH_2O. Thermocycling conditions were as follows: 94 °C for 3 min, 35 cycles of 94 °C for 45 s, 60 °C for 45 s, 72 °C for 1 min, then 72 °C for 10 min. For the subfossil samples, we designed novel, species-specific primers (Table S1) to amplify short (<150 bp) overlapping fragments to improve the success rate of amplification from degraded DNA. The reaction mix consisted of 7.5 μL of AmpliTaq Gold 360 Master Mix (Life Technologies), 0.2 μM of each primer, and 25–50 ng gDNA, made up to 15 μL with ddH_2O. Thermocycling conditions were as follows: 95 °C for 10 min, 42 cycles of 95 °C for 20 s, T_m (primer) for 20 s, 72 °C for 20 s, 72 °C for 5 min. PCR products for Fold Island, Amanda Bay, Auster, Pointe Géologie and the subfossil samples were bidirectionally sequenced by the Australian Genome Research Facility (AGRF) via the Sanger sequencing method using the PCR primer pairs. PCR products for Gould Bay, Halley Bay, Cape Washington, and Cape Crozier were sequenced using the Sanger method by Macrogen Europe. The reverse primer for the HVR and the forward primer for CytB were used to sequence each product twice, as these were found to work best in the sequencing reaction. GENEIOUS v5.5.9 was used for alignment. A high number of heteroplasmic sites were found in the HVR, and these were rescored manually according to IUPAC ambiguity codes. No heteroplasmic sites were recorded in the CytB sequences.

Data analysis – summary statistics and population structure

ARLEQUIN v3.5 (Excoffier & Lischer, 2010) was used to calculate summary statistics for HVR, CytB, and concatenated HVR and CytB. jModeltest (Posada & Buckley, 2004) was used to estimate the best substitution model for each dataset, and then the following corrections for calculating genetic distances were implemented in Arlequin: HVR – Tamura correction with a gamma distribution for rate heterogeneity with $\alpha = 0.016$; CytB – Tamura correction; concatenated – Tamura correction with a gamma distribution for rate heterogeneity with $\alpha = 0.109$ (Tamura, 1992). Arlequin was also used to calculate pairwise genetic distances (θ_{ST}) between colonies and perform analyses of molecular variation (AMOVA) on the concatenated sequences with the Tamura & Nei correction. NETWORK v4.612 (Fluxus Technology Ltd., Suffolk, UK) was used to draw haplotype networks.

Data analysis – demographic histories

Bayesian phylogenetic analyses and demographic reconstructions were performed using BEAST v1.8 (Drummond *et al*., 2012). The dataset was partitioned into HVR and CytB, with a nucleotide substitution model of HKY (Hasegawa *et al*., 1985) with four gamma categories for HVR and TN93 (Tamura & Nei, 1993) for CytB, with ambiguous states permitted. We used the coalescent extended Bayesian skyline plot tree prior (Heled & Drummond, 2008) with a strict molecular clock. For molecular clock calibration, the HVR substitution rate prior was specified as a normal distribution around a mean value of 0.55 substitutions/site/Myr (SD = 0.15), to reflect the substitution rate of the HVR in Adélie penguins (*Pygoscelis adeliae*) (Millar *et al*., 2008). In the absence of a published substitution rate for CytB in penguins, we used a uniform prior of 5×10^{-4} to 5×10^{-1} substitutions/site/Myr with a starting value of 2×10^{-2} (Weir & Schluter, 2008). The corrected radiocarbon ages of the Club Lake samples were input as tip dates for additional calibration of the molecular clock. Based on these initial priors, substitution rates for our dataset were estimated during the analysis. The posterior distributions of substitution rates, phylogenetic trees, and effective population size through time were generated using the Markov chain Monte Carlo (MCMC) sampling procedure, implemented in BEAST, which was run for 120 million generations with samples drawn every 6000 steps and the first 10% discarded as burn-in. TRACER v.1.5 was used to check effective sample size (ESS) values to confirm convergence with all values >200. Three independent BEAST analyses were performed to ensure reproducibility of the posterior distribution. The population size parameter of the demographic model (N_e**tau*) was converted to N_{ef} by dividing the parameter by 14 years, which is the estimated generation length of emperor penguins (Jenouvrier *et al*., 2005; Forcada & Trathan, 2009). Phylogenetic trees were visualized using FIGTREE v1.4.

Results

Present-day population structure

We sequenced 226 individuals from eight colonies (Fig. 1) plus three subfossil birds whose ages ranged from 643 to 881 BP (after correction for marine reservoir effect). We sequenced 629 bp of the mitochondrial hypervariable region (HVR) and 867 bp of cytochrome *b* (CytB) from each individual (GenBank accession numbers KP644787-KP645015 and KP640645-KP640873, respectively). Genetic diversity was extremely high for the HVR, with 220 haplotypes recorded out of the 229 individuals sequenced; the mean number of pairwise differences between haplotypes was 20.62 ± 9.14 (Table 1). Genetic diversity was much lower for CytB, with just 59 unique haplotypes recorded.

Table 1 Summary statistics by geographic and genetic region. n = number of individuals; N_H = unique haplotypes; N_P = polymorphic loci; H = haplotype diversity; π = mean number of pairwise differences between sequences; significance is indicated for Tajima's D and Fu's F_S test statistic, where * denotes $P < 0.05$, ** denotes $P < 0.01$, and ***denotes $P < 0.001$

Geographic Region	Genetic Region	n	N_H	N_P	H	π	Tajima's D	Fu's F_S
All sequences	HVR + CytB	229	222	205	0.999 ± 0.000	23.12 ± 10.21	−1.03	−23.63**
Ross Sea	HVR + CytB	81	80	124	0.999 ± 0.002	18.49 ± 8.28	−0.903	−22.84***
East Antarctic & Weddell Sea	HVR + CytB	148	145	171	0.999 ± 0.001	22.15 ± 9.81	−0.930	−23.80**
All sequences	HVR	229	220	164	0.999 ± 0.001	20.62 ± 9.14	−0.835	−23.67**
Ross Sea	HVR	81	76	109	0.997 ± 0.003	16.81 ± 7.56	−0.836	−17.21*
East Antarctic & Weddell Sea	HVR	148	144	138	0.999 ± 0.001	19.58 ± 8.71	−0.758	−23.85**
All sequences	CytB	229	59	41	0.864 ± 0.016	2.94 ± 1.54	−1.651*	−21.79***
Ross Sea	CytB	81	26	15	0.876 ± 0.028	1.99 ± 1.14	−0.979	−3.43
East Antarctic & Weddell Sea	CytB	148	41	33	0.797 ± 0.031	2.96 ± 1.56	−1.482*	−12.92***

Our results show a high level of gene flow among all the EAWS colonies (East Antarctica including Adélie Land, and the Weddell Sea) and between the two Ross Sea colonies (Table 2), but little exchange between the EAWS and Ross Sea colonies (pairwise θ_{ST} values range from 0.213 to 0.617, Table 2). When colonies are grouped into two populations (Ross Sea and EAWS), a high proportion (17.7%) of the genetic variation is explained by the difference between the groups, and there is strong and significant genetic differentiation between them (AMOVA, $F_{ST} = 0.196$, $P < 0.001$). This pattern is also evident from haplotype networks (Figs S1 and S2), which show that Ross Sea individuals tend to be closely related, while sequences from EAWS colonies tend to cluster independently from the Ross Sea haplotypes. However, some Ross Sea sequences are found across the network and vice versa. This could indicate low-level gene flow between the Ross Sea and EAWS.

Population history with respect to climate change

There is evidence of past population expansion in emperor penguins across Antarctica as indicated by our extended Bayesian skyline plots (EBSPs) (Fig. 2). An almost ninefold increase in abundance of the EAWS population commenced approximately 12 kya. The Ross Sea population expanded threefold from approximately 9.5 kya. Superimposing expansion signals over the estimated temperature derived from ice cores (Fig. 2c), it is clear that population expansion followed the end of the LGM. Tajima's D and Fu's F_S statistics provide further support for an expansion of both populations (Table 1).

Our phylogenetic analyses indicate three highly supported clades (Fig. 3), which diverged during the Late Pleistocene (97 kya, 95% HPD: 50−154 kya). One of these clades is comprised predominantly of Ross Sea penguins, whereas the other two are dominated by EAWS individuals.

Discussion

This first analysis of emperor penguin population structure shows colonies within the Ross Sea are genetically distinct from other Antarctic colonies, whereas those from the rest of the continent and spanning up to 8000 km of coastline are panmictic (Table 2). The admixture of the EAWS emperor penguins supports our hypothesis of limited population structure and

Table 2 Pairwise genetic differentiation between colonies. Pairwise θ_{ST} is presented below the diagonal and associated P-values above the diagonal. Significance is indicated by bold text, where * denotes $P < 0.05$, ** denotes $P < 0.01$, and *** denotes $P < 0.001$

	Gould Bay	Halley Bay	Fold Island	Auster	Amanda Bay	Pointe Géologie	Cape Washington	Cape Crozier
Gould Bay		0.596	0.731	0.560	0.186	0.129	**0.000**	**0.006**
Halley Bay	−0.027		0.566	0.798	0.708	0.301	**0.002**	**0.008**
Fold Island	−0.050	−0.032		0.515	0.323	0.462	**0.000**	**0.007**
Auster	−0.026	−0.086	−0.027		0.595	0.797	**0.000**	**0.000**
Amanda Bay	0.055	−0.058	0.014	−0.038		0.576	**0.000**	**0.000**
Pointe Géologie	0.091	0.029	−0.012	−0.085	−0.033		**0.000**	**0.000**
Cape Washington	**0.355*** **	**0.440** **	**0.468*** **	**0.567*** **	**0.617*** **	**0.596*** **		0.509
Cape Crozier	**0.213** **	**0.266** **	**0.256** **	**0.432*** **	**0.447*** **	**0.428*** **	−0.011	

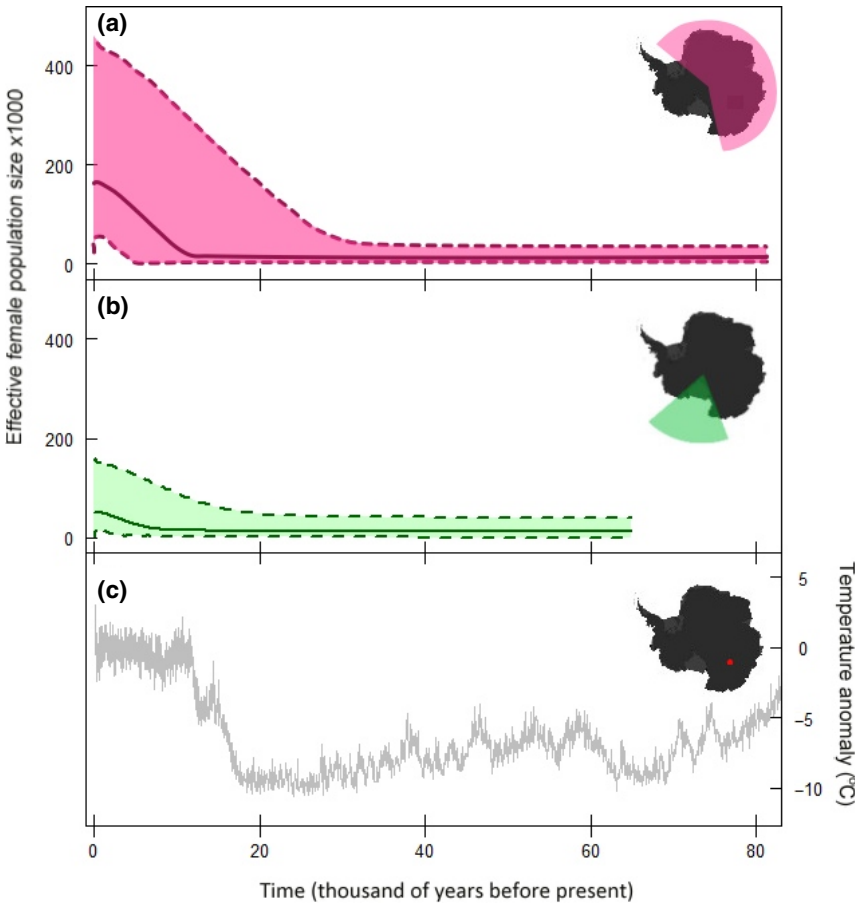


Fig. 2 Extended Bayesian skyline plots showing the change in effective female population size (Ne_f). Solid lines show the median estimate; dotted lines show the 95% highest posterior density interval. (a) EAWS colonies; (b) Ross Sea colonies; and (c) the Antarctic temperature anomaly (the difference from the average of the last 1000 years) as estimated from the EPICA Dome C ice core (Jouzel *et al.*, 2007), with the ice core location indicated in red.

indicates a very large dispersal range for the species. Given our genetic evidence of extensive mixing across Antarctica, the unique structure in the Ross Sea emperor penguins is surprising, and interestingly, the same pattern was reported for the sympatric Adélie penguin (Ritchie *et al.*, 2004), providing further evidence that the Ross Sea has a unique evolutionary history.

The existence of distinct penguin populations in the Ross Sea is puzzling. There are neither geographic nor oceanographic barriers isolating the Ross Sea from the rest of Antarctica. Furthermore, the relative distance between the Ross Sea and other colonies does not adequately explain its isolation as, for example, the Pointe Géologie colony is approximately 5600 km closer to the Ross Sea colonies than to those in the Weddell Sea (Fig. 1). Emperor penguins are known for their extraordinary migrations; satellite tracking showed that juveniles can travel >7000 km in eight months (Thiebot *et al.*, 2013). These observations support our genetic results for the EAWS region and indicate juvenile emperor penguins could comfortably traverse the 1800 km between Pointe Géologie and the Ross Sea colonies. There are also no clear habitat, environmental, or foraging differences between the Ross Sea colonies and those located elsewhere (Budd, 1961; Smith *et al.*, 2012), except that Ross Sea colonies are located closer to the ice edge, and are therefore potentially more resilient to increases in sea ice. We suggest that the divergence of emperor penguins into two populations is historical in origin.

There are three ancestral lineages within modern emperor penguins, providing evidence that populations were isolated in the past (Fig. 3) and diverged through microevolutionary processes, such as selection or genetic drift, which occur more rapidly in small, isolated populations (Hewitt, 2000). One of these lineages is mostly limited to the Ross Sea, indicating that the isolation of this region has persisted through time. Indeed, emperor penguins occupying the Ross Sea may have

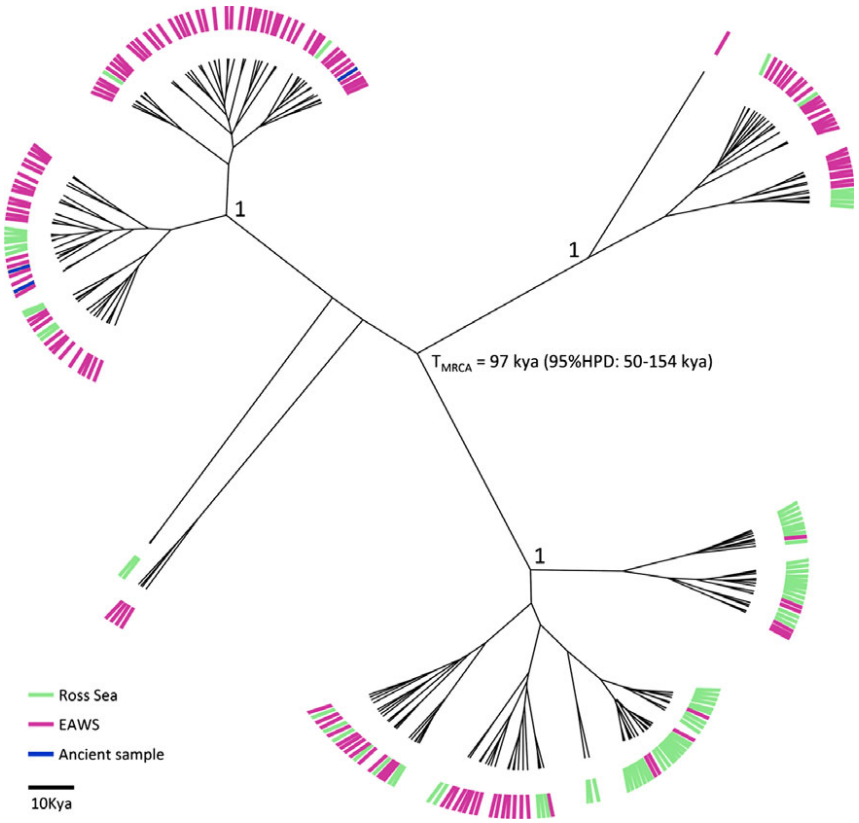


Fig. 3 Phylogenetic relationships among individuals. Magenta – EAWS individuals; green – Ross Sea individuals; blue – subfossil individuals. The posterior probabilities are shown for the major, strongly supported clades.

become so differentiated that interbreeding with the EAWS penguins occurs at very low rates, perhaps because of genetic, behavioral (Templeton, 1981), or cultural incompatibilities, such as the timing of breeding or the development of regional dialects (Macdougall-Shackleton & Macdougall-Shackleton, 2001; Jouventin & Aubin, 2002; de Dinechin *et al*., 2012).

Emperor penguins use complex display calls to recognize their mates and offspring (Robisson *et al*., 1993). Vocalization is known to be an important part of the courtship process for most penguins (Richdale, 1944; Waas *et al*., 2000). Interestingly, royal penguins (*Eudyptes schlegeli*) respond more strongly to calls from their own colony members than to calls originating from different colonies, suggesting differences in dialect (Waas *et al*., 2000). Differences in vocalizations have also been found among gentoo penguin (*Pygoscelis papua*) populations (de Dinechin *et al*., 2012). If dialects become too different, then courtship may be inhibited, thereby limiting interbreeding. This has been observed in passerine birds, in which genetically distinct groups have unique mating songs (Macdougall-Shackleton & Macdougall-Shackleton, 2001). Emperor penguin vocalization patterns have only been recorded at Pointe Géologie (Robisson *et al*., 1993), but our hypothesis could be explored in the future by comparing vocalizations of emperor penguins from the Ross Sea with those of other colonies.

Although the isolation and differentiation of the Ross Sea emperor penguins have persisted, the other two historical lineages show no geographic bias and have now hybridized to form one EAWS population. Incomplete mixing of ancestral lineages is typical of species that have survived the Pleistocene ice ages in multiple refugia (Hewitt, 1996). Our EBSPs indicate that both the EAWS and Ross Sea populations had reduced effective population sizes during the LGM (Fig. 2). Thus, contrary to a hypothesis that emperor penguins would benefit from glaciation as a result of reduced competition with other predators (Thatje *et al*., 2008), it seems that they, like other Antarctic and sub-Antarctic penguin species (Ritchie *et al*., 2004; Clucas *et al*., 2014; Trucchi *et al*., 2014), were adversely affected by the LGM.

We propose that both the reduced abundance and divergence into three lineages were linked to breeding and foraging habitat availability. Today emperor penguins have a circumpolar distribution with suitable habitat spanning the entire continent (Fretwell & Trathan, 2009). However, Antarctica during the LGM

looked very different than the continent we know today (Fig. 4). Most of the continental shelf was covered by ice as a result of both the extension of ice sheets and thick, perennial sea ice, which reduced productivity south of the modern-day polar front drastically (Anderson *et al.*, 2002, 2009; Domack *et al.*, 1998; Gersonde *et al.*, 2005; Kohfeld *et al.*, 2005; S. Jaccard, personal communications). We suggest that the increased sea ice extent would have severely restricted the foraging habitat available for emperor penguins and, coupled with lower primary production, could have resulted in a scarcity of prey resources. Additionally, air temperatures were approximately 13 °C colder than the present day (Jouzel *et al.*, 2007), which may have been near the penguins' lower limit of temperature tolerance (Le Maho *et al.*, 1976), potentially impacting both breeding success and adult survival.

The extent and duration of sea ice are important factors in the breeding success of emperor penguins (Massom *et al.*, 2009). Emperor penguins require stable fast ice to breed, but they have to traverse the sea ice to establish colonies in autumn and to forage in winter and spring. The distances between the colonies and potential foraging areas can influence breeding success where the fast ice extent is variable (Massom *et al.*, 2009), but not in locations where the extent is relatively stable (Robertson *et al.*, 2013). We therefore expect that if the winter sea ice extent was substantially greater in the LGM, or if the timing of sea ice retreat was altered, that this would have made some of the extant colony locations energetically untenable during the LGM.

During the LGM, the summer sea ice extent was similar to what we observe today, whereas the winter extent was roughly doubled (Gersonde *et al.*, 2005). Colonies may have been located close to the continent so that the ice remained stable throughout the breeding season, but this would have required adults to walk immense distances to reach foraging areas during winter and spring while provisioning the chick. In that case, the chicks would receive fewer meals and be less likely to

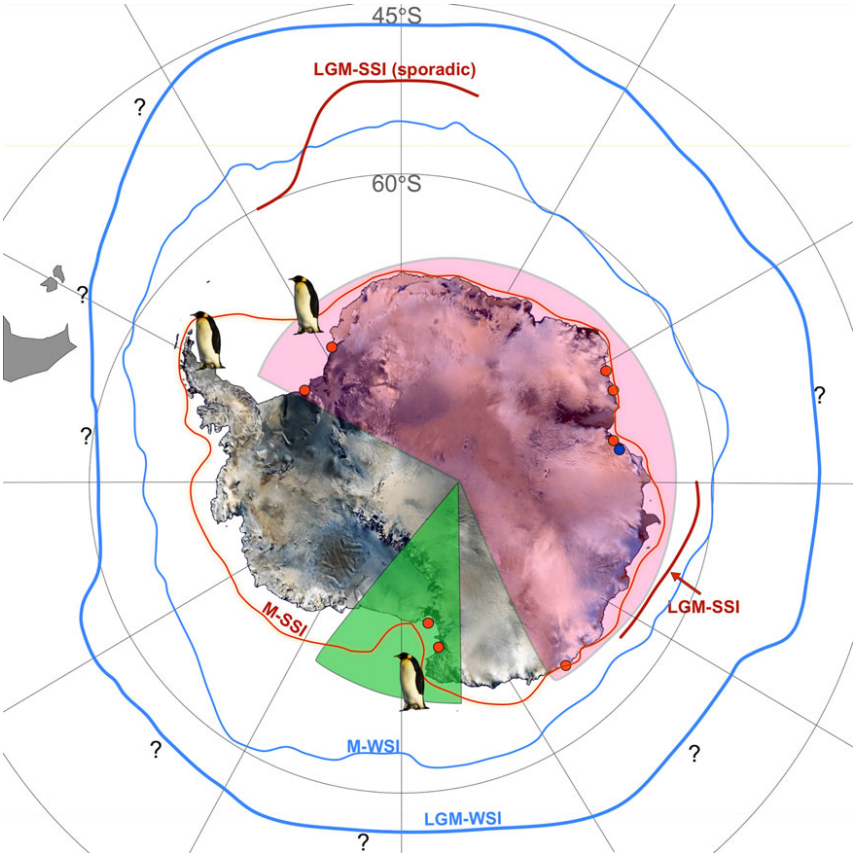


Fig. 4 Schematic of contemporary population structure and reconstruction of historical conditions. Sampled colonies are indicated by dots, as in Fig. 1. The magenta and green shading indicates population structure as estimated from this study. Lines represent the sea ice edge, as in Gersonde *et al.*, 2005. M-SSI = modern summer sea ice edge; LGM-SSI = LGM summer sea ice edge; M-WSI = modern winter sea ice edge; LGM-WSI = LGM winter sea ice edge; ? = insufficient data to reconstruct the sea ice edge. Penguins represent hypothesized locations of polynya refugia. Emperor penguin picture: © Samuel Blanc/www.sblanc.com.

survive. The present distribution of colonies close to land (Fretwell & Trathan, 2009) suggests that fast ice proximate to land provides a more stable platform than near the fast ice edge. Also, stable ice close to the coast occurs in predictable locations that might be important for colony establishment and cohesion. Colonies further away from the coast may therefore be difficult to maintain. Our discovery of three distinct lineages provides evidence against a straightforward, latitudinal range shift in line with the sea ice edge and suggests that emperor penguins may have survived the LGM in three suitably situated, geographically isolated refugia.

Emperor penguin refugia during the LGM may have been linked to the presence of polynyas. Several extant emperor penguin colonies are located near polynyas, which may be utilized for foraging during the winter (Croxall *et al*., 2002). Polynyas acted as 'hot spots' of primary productivity during the LGM, supporting marine life and flying seabirds (Thatje *et al*., 2008). Sediment cores in the northwestern Ross Sea indicate open water polynya conditions throughout the LGM (Brambati *et al*., 2002; Thatje *et al*., 2008) and this polynya could have sustained a refuge population until the Ross Sea began to clear of ice (Fig. 4). By 9.6 kya, most of the northern Ross Sea was open water (Licht & Andrews, 2002). The retreating sea ice and increased upwelling during deglaciation increased productivity in the Ross Sea (Anderson *et al*., 2009) and likely increased the foraging habitat and prey available to emperor penguins; therefore, we hypothesize that these factors drove the threefold expansion of emperor penguins in this region around this time (Fig. 2). The LGM polynya may have also supported Adélie penguins, accounting for the existence of a distinct Ross Sea clade as previously observed for this species (Ritchie *et al*., 2004).

Another polynya was located in the southeastern Weddell Sea off Dronning Maud Land (DML) (Mackensen *et al*., 1994; Thatje *et al*., 2008) (Fig. 4). Colonies of snow petrels (*Pagodroma nivea*) were present in DML throughout the LGM, associated with this polynya (Wand & Hermichen, 2005), and it may have also provided a refuge for emperor penguins. There is evidence from sediment cores for a third LGM polynya, located in the northwestern Weddell Sea (Smith *et al*., 2010) (Fig. 4); this would be consistent with our third emperor penguin refuge, given that the refuge is likely to be more proximate to DML than the Ross Sea, as the two refugial lineages hybridized postglacially, while the Ross Sea lineage remained distinct.

We propose that two refuge populations that were isolated in the Weddell Sea expanded their range into Prydz Bay and Adélie Land and merged during the retreat of the East Antarctic ice sheet 14–7 kya (Mackintosh *et al*., 2011). At this time, the onset of more favorable environmental conditions could have resulted in the dramatic, ninefold increase in abundance shown here (Fig. 2). A seasonal sea ice cycle was established in Prydz Bay approximately 10.4 kya (Barbara *et al*., 2010), opening up foraging habitat and coinciding with high levels of primary productivity (e.g., Anderson *et al*., 2009; Sedwick *et al*., 2001). In Adélie Land, primary productivity and the duration of the ice-free season increased from 9 kya (Denis *et al*., 2009a,b). This new habitat could have facilitated the range expansion of the EAWS lineages.

It should be noted that the timing of the abundance increase of emperor penguins does not coincide exactly with the end of the LGM (Fig. 2). We hypothesize that it is not the temperature change itself, but rather the subsequent change in sea ice conditions and primary productivity that are most likely to affect emperor penguins. Indeed, it has been proposed that there is an optimal level of sea ice at the large temporal/spatial scale for emperor penguins, which roughly corresponds to current conditions (Ainley *et al*., 2010). Therefore, the greater sea ice extent of the LGM was most likely suboptimal for emperor penguin populations. The end of the LGM is measured when temperatures began to increase (19–16 kya). However, deglaciation, during which ice sheets and sea ice retreated and primary productivity increased, occurred slowly over an extended time period (ca. 17–11 kya) (Anderson *et al*., 2009). These events occurred later in the Ross Sea than in East Antarctica, and our results support the hypothesis that ice sheet and sea ice retreat and increasing primary productivity were the main factors controlling emperor penguin abundance, as the Ross Sea emperor penguin population expanded later than the EAWS population (Fig. 2). Furthermore, emperor penguins produce only one chick per year and take approximately 5 years to reach sexual maturity (Jenouvrier *et al*., 2005), so any abundance increase would be initially slow.

Our hypothesis of three refugial populations of emperor penguins during the LGM could be tested using a higher density of genetic markers. This would allow for the investigation of clinal variation in genetic diversity arising from founder effects as new areas were colonized following the expansion from refugia after the LGM (Hewitt, 1996). It should be noted that our present study is based on mitochondrial DNA and therefore represents dispersal patterns of females only, but nonetheless supports a plausible explanation for past and present microevolutionary processes in emperor penguins. The next step should be to verify these findings using nuclear markers to account for male-mediated gene flow.

In this continent-wide study of microevolution in an Antarctic penguin, we suggest that past climatic

changes have greatly impacted emperor penguin populations. As conditions became more favorable after the LGM, their global population expanded and the populations from the Weddell Sea and East Antarctica intermixed to form one large, panmictic population. Interestingly, the isolation of the Ross Sea emperor penguins has persisted until today. The reasons for this isolation remain unknown, but we suggest that separate management plans are required for the Ross Sea and EAWS populations. By conserving the full spectrum of genetic variation and, in particular, all phylogeographic lineages, the evolutionary potential of the species can be maximized (D'Amen *et al.*, 2013).

Our study suggests that emperor penguins have shown important historical responses to past climate shifts and their population increase post-LGM was remarkable. However, the projected rate of temperature increase over the next century is an order of magnitude greater than that following the LGM (Shakun *et al.*, 2012; Collins *et al.*, 2013; Masson-Delmotte *et al.*, 2013). At present, emperor penguins become heat stressed around 0 °C, so may exist near the upper limits of their physiological tolerance (B. Wienecke, personal observations). Whether the resilience demonstrated in the past of this highly cold-adapted species will enable it to adapt to projected climate change remains to be seen, as rising temperatures will alter its breeding grounds and foraging space more rapidly than in the past.

Acknowledgements

We gratefully acknowledge the assistance of Dick Filby, Adventure Network International and Hannah McKeand in sample collection and logistics at Gould Bay, Françoise Amélineau for samples from Pointe Géologie, James Bumac for sample collection at Fold Island, British Antarctic Survey staff at Halley Bay, Roger Kirkwood for sample collection at Auster, and Paul Ponganis and the NSF for facilitating feather collection in the Ross Sea (NSF 09-44220). We thank John van den Hoff for his insights regarding Antarctic refugia, Gareth Dyke for commenting on earlier drafts of the manuscript, and Samuel Jaccard for advice regarding LGM primary productivity estimates. We are also grateful to Simon Foote for providing key infrastructure and to Patrick Lelliott for his assistance in the laboratory. This project was funded by Australian Antarctic Science Project 4184 (KM, BW, JY), ANZ Trustees Holsworth Wildlife Research Endowment (JY), NERC and Penguin Lifelines (GVC), British Antarctic Survey Ecosystems Programme (PNT) and the Darwin Initiative (TH).

References

Ainley D, Russel J, Jenouvrier S, Woehler E, Lyver PO'B, Fraser WR, Kooyman G (2010) Antarctic penguin response to habitat change as Earth's troposphere reaches 2°C above preindustrial levels. *Ecological Monographs*, **80**, 49–66.

Anderson JB, Shipp SS, Lowe AL, Wellner JS, Mosola AB (2002) The Antarctic ice sheet during the Last Glacial Maximum and its subsequent retreat history: a review. *Quaternary Science Reviews*, **21**, 49–70.

Anderson R, Ali S, Bradtmiller L, Nielsen S, Fleisher M, Anderson B, Burckle L (2009) Wind-driven upwelling in the Southern Ocean and the deglacial rise in atmospheric CO_2. *Science*, **323**, 1443–1448.

Baker A, Marshall H (1997) Mitochondrial control region sequences as tools for understanding evolution. In: *Avian molecular evolution and systematics* (ed Mindell DP), pp. 51–82. Academic Press, NewYork.

Baker AJ, Pereira SL, Haddrath OP, Edge K-A (2006) Multiple gene evidence for expansion of extant penguins out of Antarctica due to global cooling. *Proceedings of the Royal Society B: Biological Sciences*, **273**, 11–17.

Barbara L, Crosta X, Massé G, Ther O (2010) Deglacial environments in eastern Prydz Bay, East Antarctica. *Quaternary Science Reviews*, **29**, 2731–2740.

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

Barbraud C, Gavrilo M, Mizin Y, Weimerskirch H (2011) Comparison of emperor penguin declines between Pointe Géologie and Haswell Island over the past 50 years. *Antarctic Science*, **23** , 461–468.

Barnett R, Yamaguchi N, Shapiro B *et al.* (2014) Revealing the maternal demographic history of *Panthera leo* using ancient DNA and a spatially explicit genealogical analysis. *BMC Evolutionary Biology*, **14**, 70.

Barnosky AD, Matzke N, Tomiya S *et al.* (2011) Has the Earth's sixth mass extinction already arrived? *Nature*, **471**, 51–57.

Brambati A, Melis R, Quaia T, Salvi G (2002) Late Quaternary climatic changes in the Ross Sea area, Antarctica. *Antarctica at the Close of the Millennium. Royal Society New Zealand, Wellington*, **35**, 359–364.

Budd GM (1961) The biotopes of emperor penguin rookeries. *Emu*, **61**, 171–189.

Carstens BC, Richards CL (2007) Integrating coalescent and ecological niche modeling in comparative phylogeography. *Evolution*, **61**, 1439–1454.

Clucas GV, Dunn MJ, Dyke G *et al.* (2014) A reversal of fortunes: climate change 'winners' and 'losers' in Antarctic Peninsula penguins. *Scientific Reports*, **4**, 5024.

Collins M, Knutti R, Arblaster J *et al.* (2013) Long-term climate change: projections, commitments and irreversibility. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM), pp. 1029–1136. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Croxall JP, Trathan PN, Murphy EJ (2002) Environmental change and Antarctic seabird populations. *Science*, **297**, 1510–1514.

D'Amen M, Zimmermann NE, Pearman PB (2013) Conservation of phylogeographic lineages under climate change. *Global Ecology and Biogeography*, **22**, 93–104.

Dehnhard N, Poisbleau M, Demongin L, Ludynia K, Quillfeldt P (2014) High juvenile annual survival probabilities in southern rockhopper penguins *Eudyptes chrysocome* are independent of individual fledging traits. *Ibis*, **156**, 548–560.

Denis D, Crosta X, Schmidt S *et al.* (2009a) Holocene glacier and deep water dynamics, Adélie Land region, East Antarctica. *Quaternary Science Reviews*, **28**, 1291–1303.

Denis D, Crosta X, Schmidt S *et al.* (2009b) Holocene productivity changes off Adélie Land (East Antarctica). *Paleoceanography*, **24**, PA3207.

Dewasmes G, Le Maho Y, Cornet A, Groscolas R (1980) Resting metabolic rate and cost of locomotion in long-term fasting Emperor Penguins. *Journal of Applied Physiology*, **49**, 888–896.

de Dinechin M, Dobson FS, Zehtindjiev P *et al.* (2012) The biogeography of gentoo penguins (*Pygoscelis papua*). *Canadian Journal of Zoology*, **90**, 352–360.

Domack E, O'brien P, Harris P, Taylor F, Quilty PG, Santis LD, Raker B (1998) Late Quaternary sediment facies in Prydz Bay, East Antarctica and their relationship to glacial advance onto the continental shelf. *Antarctic Science*, **10**, 236–246.

Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution*, **22**, 1185–1192.

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, **29**, 1969–1973.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.

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

Fretwell PT, Trathan PN (2009) Penguins from space: faecal stains reveal the location of emperor penguin colonies. *Global Ecology and Biogeography*, **18**, 543–552.

Fretwell PT, Larue MA, Morin P *et al.* (2012) An emperor penguin population estimate: the first global, synoptic survey of a species from space. *PLoS ONE*, **7**, e33751.

Fretwell PT, Trathan PN, Wienecke B, Kooyman GL (2014) Emperor penguins breeding on iceshelves. *PLoS ONE*, **9**, e85285.

Gales N, Klages N, Williams R, Woehler E (1990) The diet of the emperor penguin, *Aptenodytes forsteri*, in Amanda Bay, Princess Elizabeth Land, Antarctica. *Antarctic Science*, **2**, 23–28.

Gersonde R, Crosta X, Abelmann A, Armand L (2005) 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. *Quaternary Science Reviews*, **24**, 869–896.

Gordon JE, Harkness DD (1992) Magnitude and geographic variation of the radiocarbon content in Antarctic marine life: implications for reservoir corrections in radiocarbon dating. *Quaternary Science Reviews*, **11**, 697–708.

Hartl DL, Clark AG (1997) *Principles of Population Genetics*. Sinauer Associates, Sunderland, Massachusetts.

Hasegawa M, Kishino H, Yano T-A (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, **22**, 160–174.

Heled J, Drummond AJ (2008) Bayesian inference of population size history from multiple loci. *BMC Evolutionary Biology*, **8**, 289.

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, **58**, 247–276.

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

Hoelzel AR (2010) Looking backwards to look forwards: conservation genetics in a changing world. *Conservation Genetics*, **11**, 655–660.

Jenouvrier S, Weimerskirch H, Barbraud C, Park YH, Cazelles B (2005) Evidence of a shift in the cyclicity of Antarctic seabird dynamics linked to climate. *Proceedings of the Royal Society B: Biological Sciences*, **272**, 887–895.

Jenouvrier S, Caswell H, Barbraud C, Holland M, Stroeve J, Weimerskirch H (2009) Demographic models and IPCC climate projections predict the decline of an emperor penguin population. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 1844–1847.

Jenouvrier S, Holland M, Stroeve J, Barbraud C, Weimerskirch H, Serreze M, Caswell H (2012) Effects of climate change on an emperor penguin population: analysis of coupled demographic and climate models. *Global Change Biology*, **18**, 2756–2770.

Jenouvrier S, Holland M, Stroeve J, Serreze M, Barbraud C, Weimerskirch H, Caswell H (2014) Projected continent-wide declines of the emperor penguin under climate change. *Nature Climate Change*, **4**, 715–718.

Jouventin P, Aubin T (2002) Acoustic systems are adapted to breeding ecologies: individual recognition in nesting penguins. *Animal Behaviour*, **64**, 747–757.

Jouzel J, Masson-Delmotte V, Cattani O *et al*. (2007) Orbital and millennial Antarctic climate variability over the past 800,000 years. *Science*, **317**, 793–796.

Kocher TD, Thomas WK, Meyer A, Edwards SV, Pääbo S, Villablanca FX, Wilson AC (1989) Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proceedings of the National Academy of Sciences of the United States of America*, **86**, 6196–6200.

Kohfeld KE, Le Quéré C, Harrison SP, Anderson RF (2005) Role of marine biology in glacial-interglacial CO_2 cycles. *Science*, **308**, 74–78.

Kooyman GL, Kooyman TG, Horning M, Kooyman CA (1996) Penguin dispersal after fledging. *Nature*, **383**, 397.

LaRue M, Kooyman G, Lynch H, Fretwell P (2014) Emigration in emperor penguins: implications for interpretation of long-term studies. *Ecography*, **37**, 001–007.

Le Maho Y, Delclitte P, Chatonnet J (1976) Thermoregulation in Fasting Emperor Penguins Under Natural Conditions. *American Journal of Physiology*, **231**, 913–922.

Le Maho Y, Karmann H, Briot D *et al*. (1992) Stress in birds due to routine handling and a technique to avoid it. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, **263**, R775–R781.

Licht KJ, Andrews JT (2002) The 14C record of late Pleistocene ice advance and retreat in the central Ross Sea, Antarctica. *Arctic, Antarctic, and Alpine Research*, **34**, 324–333.

Macdougall-Shackleton EA, Macdougall-Shackleton SA (2001) Cultural and genetic evolution in mountain white-crowned sparrows: song dialects are associated with population structure. *Evolution*, **55**, 2568–2575.

Mackensen A, Grobe H, Hubberten H-W, Kuhn G (1994) Benthic foraminiferal assemblages and the d13C-signal in the Atlantic sector of the Southern Ocean: glacial-to-interglacial contrasts. In: *Carbon Cycling in the Glacial Ocean: Constraints on the Ocean's Role in Global Change* (eds Zahn R, Pederson TF, Kaminski MA, Labeyrie L), pp. 105–144. Springer, Berlin, Heidelberg, Germany.

Mackintosh A, Golledge N, Domack E *et al*. (2011) Retreat of the East Antarctic ice sheet during the last glacial termination. *Nature Geoscience*, **4**, 195–202.

Martin S (2001) Polynyas. In: *Encyclopedia of Ocean Sciences*, 2nd edn (ed. Steele JH), pp. 540–545. Academic Press, Oxford, UK.

Massom RA, Hill K, Barbraud C, Adams N, Ancel A, Emmerson L, Pook MJ (2009) Fast ice distribution in Adélie Land, East Antarctica: interannual variability and implications for emperor penguins, *Aptenodytes forsteri*. *Marine Ecology Progress Series*, **374**, 243–257.

Masson-Delmotte V, Schulz M, Abe-Ouchi A *et al*. (2013) Information from paleoclimate archives. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM), pp. 383–464. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Millar CD, Dodd A, Anderson J *et al*. (2008) Mutation and evolutionary rates in Adélie penguins from the Antarctic. *PLoS Genetics*, **4**, e1000209.

Nogués-Bravo D, Rodríguez J, Hortal J, Batra P, Araújo MB (2008) Climate change, humans, and the extinction of the woolly mammoth. *PLoS Biology*, **6**, e79.

Posada D, Buckley TR (2004) Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Systematic Biology*, **53**, 793–808.

Pybus OG, Rambaut A, Harvey PH (2000) An integrated framework for the inference of viral population history from reconstructed genealogies. *Genetics*, **155**, 1429–1437.

Richdale LE (1944) Courtship and allied behaviour in penguins. *Emu*, **44**, 305–319.

Ritchie PA, Millar CD, Gibb GC, Baroni C, Lambert DM (2004) Ancient DNA enables timing of the Pleistocene origin and Holocene expansion of two Adélie penguin lineages in Antarctica. *Molecular Biology and Evolution*, **21**, 240–248.

Robertson G, Wienecke B, Emmerson L, Fraser AD (2013) Long-term trends in the population size and breeding success of emperor penguins at the Taylor Glacier colony, Antarctica. *Polar Biology*, **37**, 251–259.

Robisson P, Aubin T, Bremond J-C (1993) Individuality in the voice of the emperor penguin *Aptenodytes forsteri*: adaptation to a noisy environment. *Ethology*, **94**, 279–290.

Saraux C, Viblanc VA, Hanuise N, Le Maho Y, Le Bohec C (2011) Effects of individual pre-fledging traits and environmental conditions on return patterns in juvenile king penguins. *PLoS ONE*, **6**, e20407.

Sedwick PN, Harris PT, Robertson LG, Mcmurtry GM, Cremer MD, Robinson P (2001) Holocene sediment records from the continental shelf of Mac. Robertson Land. *East Antarctica*. *Paleoceanography*, **16**, 212–225.

Shakun JD, Clark PU, He F *et al*. (2012) Global warming preceded by increasing carbon dioxide concentrations during the last deglaciation. *Nature*, **484**, 49–54.

Smith WO, Gordon LI (1997) Hyperproductivity of the Ross Sea (Antarctica) polynya during austral spring. *Geophysical Research Letters*, **24** , 233–236.

Smith JA, Hillenbrand C-D, Pudsey CJ, Allen CS, Graham AG (2010) The presence of polynyas in the Weddell Sea during the last glacial period with implications for the reconstruction of sea-ice limits and ice sheet history. *Earth and Planetary Science Letters*, **296**, 287–298.

Smith WO, Ainley DG, Cattaneo-Vietti R, Hofmann EE (2012) The Ross Sea continental shelf: regional biogeochemical cycles, trophic interactions, and potential future changes. In: *Antarctic Ecosystems: An Extreme Environment in a Changing World* (eds Rogers AD, Johnston NM, Murphy EJ, Clarke A), pp. 213–242. Wiley-Blackwell, Hoboken, NJ.

St Ui M, Reimer J (1993) Extended 14C data base and revised CALIB 3.0 14C age calibration program. *Radiocarbon*, **35**, 215–230.

Tamura K (1992) Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases. *Molecular Biology and Evolution*, **9**, 678–687.

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, **10**, 512–526.

Templeton AR (1981) Mechanisms of speciation – a population genetic approach. *Annual Review of Ecology and Systematics*, **12**, 23–48.

Thatje S, Hillenbrand CD, Mackensen A, Larter R (2008) Life hung by a thread: endurance of Antarctic fauna in glacial periods. *Ecology*, **89**, 682–692.

Thiebot J-B, Lescroël A, Barbraud C, Bost C-A (2013) Three-dimensional use of marine habitats by juvenile emperor penguins *Aptenodytes forsteri* during post-natal dispersal. *Antarctic Science*, **25**, 536–544.

Trathan PN, Fretwell PT, Stonehouse B (2011) First recorded loss of an emperor penguin colony in the recent period of Antarctic regional warming: implications for other colonies. *PLoS ONE*, **6**, e14738.

Trucchi E, Gratton P, Whittington JD, Cristofari R, Le Maho Y, Stenseth NC, Le Bohec C (2014) King penguin demography since the last glaciation inferred from genome-wide data. *Proceedings of the Royal Society B: Biological Sciences*, **281**, 20140528.

Vaughan DG, Comiso JC, Allison I *et al*. (2013) Observations: cryosphere. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker TF,

Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM), pp. 317–382. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Waas JR, Caulfield M, Colgan PW, Boag PT (2000) Colony sound facilitates sexual and agonistic activities in royal penguins. *Animal Behaviour*, **60**, 77–84.

Wand U, Hermichen W-D (2005) Late Quaternary ice level changes in central Dronning Maud Land, East Antarctica, as inferred by 14C ages of mumiyo deposits in snow petrel colonies. *Geol Jahrbuch*, **B97**, 237–254.

Weir JT, Schluter D (2008) Calibrating the avian molecular clock. *Molecular Ecology*, **17**, 2321–2328.

Wienecke B, Raymond B, Robertson G (2010) Maiden journey of fledgling emperor penguins from the Mawson Coast, East Antarctica. *Marine Ecology Progress Series*, **410**, 269–282.

Zwally HJ, Comiso JC, Gordon AL (1985) Antarctic offshore leads and polynyas and oceanographic effects. In: *Oceanology of the Antarctic Continental Shelf* (ed. Jacobs SS), pp. 203–226. American Geophysical Union, Washington, D. C..

Zwally HJ, Comiso JC, Parkinson CL, Cavalieri DJ, Gloersen P (2002) Variability of Antarctic sea ice 1979–1998. *Journal of Geophysical Research: Oceans (1978–2012)*, **107**, 9–19.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. Haplotype network of phylogenetic relationships among all HVR sequences. Magenta – EAWS colonies; green – Ross Sea colonies; blue – subfossil samples; the size of the circle indicates the relative frequency of the haplotype.
Figure S2. Haplotype network of phylogenetic relationships among all CytB sequences. Magenta – EAWS colonies; green – Ross Sea colonies; blue – subfossil samples; the size of the circle indicates the relative frequency of the haplotype.
Table S1. Primer sequences.

Appendix B

Clucas et al. BMC Evolutionary Biology (2016) 16:211
DOI 10.1186/s12862-016-0784-z

BMC Evolutionary Biology

RESEARCH ARTICLE Open Access

Dispersal in the sub-Antarctic: king penguins show remarkably little population genetic differentiation across their range

Gemma V. Clucas[1,2*], Jane L. Younger[1,3*], Damian Kao[1], Alex D. Rogers[1], Jonathan Handley[4], Gary D. Miller[5], Pierre Jouventin[6], Paul Nolan[7], Karim Gharbi[8], Karen J. Miller[9] and Tom Hart[1]

Abstract

Background: Seabirds are important components of marine ecosystems, both as predators and as indicators of ecological change, being conspicuous and sensitive to changes in prey abundance. To determine whether fluctuations in population sizes are localised or indicative of large-scale ecosystem change, we must first understand population structure and dispersal. King penguins are long-lived seabirds that occupy a niche across the sub-Antarctic zone close to the Polar Front. Colonies have very different histories of exploitation, population recovery, and expansion.

Results: We investigated the genetic population structure and patterns of colonisation of king penguins across their current range using a dataset of 5154 unlinked, high-coverage single nucleotide polymorphisms generated via restriction site associated DNA sequencing (RADSeq). Despite breeding at a small number of discrete, geographically separate sites, we find only very slight genetic differentiation among colonies separated by thousands of kilometers of open-ocean, suggesting migration among islands and archipelagos may be common. Our results show that the South Georgia population is slightly differentiated from all other colonies and suggest that the recently founded Falkland Island colony is likely to have been established by migrants from the distant Crozet Islands rather than nearby colonies on South Georgia, possibly as a result of density-dependent processes.

Conclusions: The observed subtle differentiation among king penguin colonies must be considered in future conservation planning and monitoring of the species, and demographic models that attempt to forecast extinction risk in response to large-scale climate change must take into account migration. It is possible that migration could buffer king penguins against some of the impacts of climate change where colonies appear panmictic, although it is unlikely to protect them completely given the widespread physical changes projected for their Southern Ocean foraging grounds. Overall, large-scale population genetic studies of marine predators across the Southern Ocean are revealing more interconnection and migration than previously supposed.

Keywords: Southern Ocean, Seabirds, Molecular ecology, *Aptenodytes patagonicus*, Dispersal, Genetic homogeneity, RAD-Seq, Colonisation, Gene flow

* Correspondence: gemma.clucas@noc.soton.ac.uk; jyounger@luc.edu
Gemma V. Clucas and Jane L. Younger share first-authorship
Karen J. Miller and Tom Hart share last-authorship
[1]Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK
Full list of author information is available at the end of the article

Background

Understanding the patterns and mechanisms of population structure is essential for successful species conservation [1]. For example, species with a high degree of population differentiation and limited dispersal among colonies may have a reduced ability to respond to unfavorable local environmental conditions [2] and may lose a large portion of their total genetic variation if local populations are lost or reduced [3]. Accurate data regarding the geographic boundaries of breeding populations and the degree of genetic exchange among them are therefore essential for species risk assessments and conservation planning, including to mitigate the effects of climate change. However, the extent of differentiation among natural populations of seabirds is difficult to predict and has been shown to vary widely among taxa [3, 4]. In general, seabirds are philopatric, with adults returning to natal sites to breed [5], and this behavior can be an isolating mechanism that acts as a barrier to gene flow. Seabirds that have large foraging ranges, or that breed at high latitudes, such as the polar regions, are thought to be the least likely to have differentiated populations as a result of recent range expansions and retained ancestral variation [3].

King penguins (*Aptenodytes patagonicus*) are thought to be vulnerable to climate change impacts in the future [6, 7] and an understanding of their population structure is required to accurately model these impacts and make inferences about observed changes in population size. King penguins congregate in large breeding colonies on coastal ice-free ground on sub-Antarctic islands between 45 ° and 55 ° south [8] (Fig. 1). Numbers have been increasing across their range over the past several decades [8–11], following historic anthropogenic exploitation during the late 19th to early 20th centuries when they were slaughtered en masse for the blubber oil industry [12]. The global population of king penguins is now conservatively estimated at 1.6 million breeding pairs and still increasing [8].

Owing to their large and growing population size across most of their range, king penguins are currently listed as being of Least Concern on the IUCN's Red List of Threatened Species [13] although there have been concerns that harvest may have resulted in a population

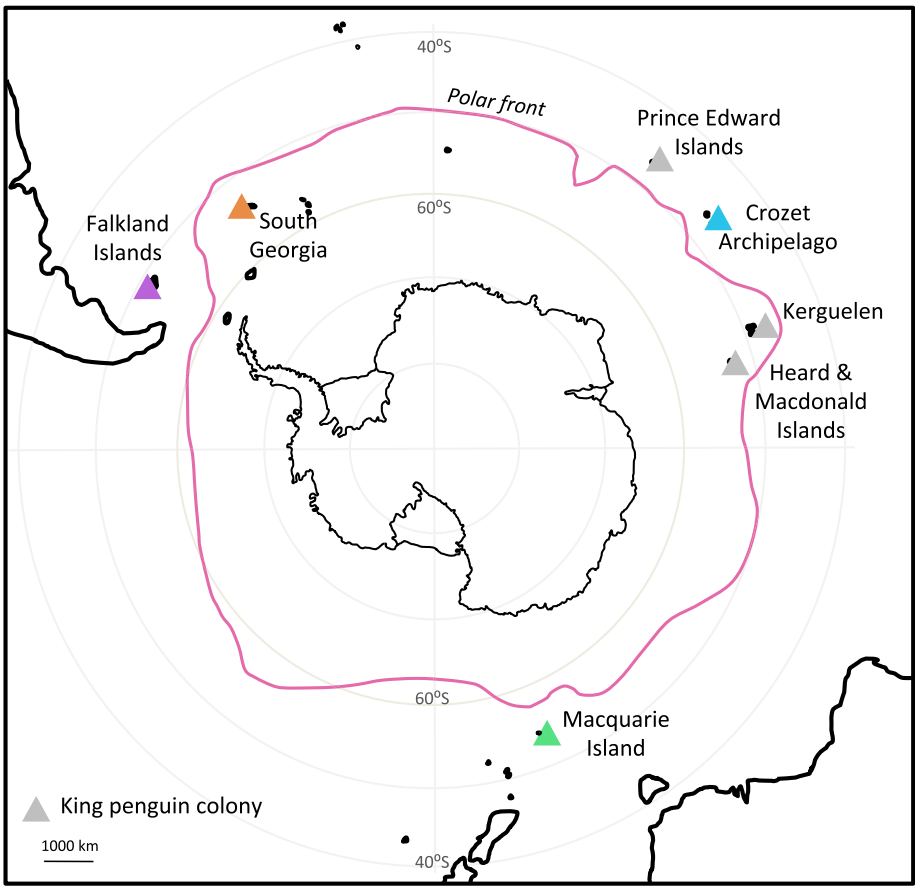


Fig. 1 King penguin colony locations. *Triangles* indicate known king penguin colonies, with coloured triangles indicating the four colonies sampled for this study

bottleneck that would have reduced genetic variation and hence their adaptive capacity. Furthermore, king penguins will face new challenges in the coming decades as climate change alters their marine foraging habitat. The most immediate threat posed by climate change to king penguin populations is the southward shift of the Polar Front and deepening of the thermocline; both secondary to warming of the Southern Ocean's surface waters [7]. King penguins forage almost exclusively at the Polar Front during the summer breeding season [14–16], as a result of the combination of predictably high prey abundance and ideal diving conditions that they find at the front [17, 18]. As sea surface temperatures increase with climate change, the position of the Polar Front is shifting to the south, and this is predicted to double the king penguin's travelling distance to their preferred foraging grounds by 2100 [19]. The coincident deepening of the thermocline means the penguins must also dive deeper to reach their prey [7]. A study at the Crozet Islands has already demonstrated the impact that warming waters can have on king penguin numbers, with a population decline of 34 % associated with an anomalously warm year in 1997 [7]. In light of the potential threats to king penguin populations, accurate data regarding their population structure are needed [8]. Specifically, to monitor population sizes in relation to environmental impacts we must first understand what constitutes a genetic breeding population of king penguins.

There have been no studies of genetic population structure of king penguins across their breeding range to date. A decade-long study at one colony in the Crozet Islands found that 77 % of juvenile king penguins returned to their natal colony [20]. This suggests that the species is largely philopatric, however, even low numbers of dispersing individuals could be sufficient to homogenise populations [21]. King penguins possess a remarkable mobility, regularly conducting round-trips in excess of 3200 km from breeding colonies to forage in Antarctic waters during the winter months [15]. However, the average distance between the pairs of breeding sites in our study is 6500 km and the colonies are distributed longitudinally, whereas most of the king penguin's foraging movement is latitudinal [16, 22]. This suggests that frequent dispersal among breeding sites should be unlikely. In spite of this, incidences of long-distance dispersal have been documented, with birds tagged on the Crozet Islands resighted resting or molting at Marion Island (900 km away) [23], Kerguelen Island (1500 km away), Macquarie Island (5600 km away) [24] and Heard Island (1740 km away) [25]. It should also be noted that any genetic differentiation that arose during the founding of colonies would be expected to persist for a very long time (i.e. thousands of generations) in a species with such a large effective population size and rapid population growth rate [26]. Previous studies of population structure in other penguin species revealed a remarkable lack of differentiation across thousands of kilometers, including in emperor penguins (*Aptenodytes forsteri*) [27, 28] and Adélie penguins (*Pygoscelis adeliae*) [29, 30]. This is in contrast to gentoo penguins (*Pygoscelis papua*) [31] and chinstrap penguins (*Pygoscelis antarctica*) [32], which demonstrated moderate to low genetic differentiation across similar distances. Both emperor and Adélie penguins have almost continuous circumpolar distributions [33, 34] that may faciliate migration, whereas king penguin colonies are scattered distantly across the sub-Antarctic (Fig. 1).

Overall, king penguins are a highly mobile marine species with huge potential for dispersal; however, genetic divergence among colonies may exist as a result of non-physical barriers, such as philopatry, local adaptation or isolation by colonisation [35]. We therefore hypothesised that breeding colonies on different archipelagos would constitute genetically distinct populations. To test this hypothesis we generated a dataset of more than 5000 unlinked single nucleotide polymorphisms (SNPs) using restriction site associated DNA sequencing (RADSeq) [36] for king penguins from four colonies spread across their range (Fig. 1). We aimed to identify population structure, as well as distinct phylogenetic lineages that may have been associated with past glacial refugia. Previous studies have shown that king penguin numbers were much reduced during the last ice age [37], and the species' range may have been contracted into refugia at unknown locations [38]. Finally, we aimed to test the hypothesis that the recently founded colony at the Falkland Islands [39] was established via migration from nearby South Georgia (Fig. 1). Throughout we use the term 'dispersal' to refer to individual movements away from the natal colony and 'migration' to refer to an individual breeding at a different colony from its natal colony.

Methods

Sampling

Blood was collected from 16 king penguins at each of: Volunteer Point on the Falkland Islands (Feb 2014), Fortuna Bay on South Georgia (Dec 2012), Baie du Marin on Possession Island in the Crozet Islands (Dec 2003–Jan 2004) and Sandy Bay on Macquarie Island (Dec 2005–Jan 2006) (Fig. 1). To prevent biting and minimize stress during handling [40], king penguins were either seized with both hands and the flippers were restrained with the head placed under the arm of the handler, or they were wrapped in cushioned material to cover the head and prevent movement. A second handler took 1 mL blood from the brachial or ulnar vein using a 25G or 23G needle and 1 mL syringe, after cleaning the area with an alcohol swab. Total restraint time was generally two to

three minutes. All field activities were conducted under permits from the Falkland Islands Government, the Government of South Georgia and the South Sandwich Islands and the Tasmanian Parks Department, and also received ethical approval from the University of Oxford, the University of Western Australia, the Auburn University Institutional Animal Care and Use Committee and the Institut Polaire P. E. Victor. Blood samples were transported to the UK at ambient temperature in RNAlater (Life Technologies) or in Queen's Lysis buffer, and stored at −20 °C or −80 °C until extraction.

Sequencing

DNA was extracted from the 64 blood samples using a QIAGEN DNeasy Blood and Tissue Kit following the manufacturer's protocol, but modified to include 40 μL proteinase K at the digestion step and with the incubation time extended to 3 h. The samples were treated with 1 μL Riboshredder (Epicentre) to reduce RNA contamination. DNA concentration was measured with a Qubit (ThermoFisher Scientific) and high molecular weight was confirmed on a 1 % gel. We sequenced the mitochondrial hypervariable region (HVR; 620 base pairs; GenBank accessions: KX857217-KX857259) because this marker has revealed phylogeographic patterns within other penguin species [28, 29, 41]. The HVR was amplified in all samples using primers F-0225 (5'-GGAACCTCCC AAAGAGTACCA) and R-INR (5'-CCAACCAGATGT ATCGGTGA) [28]. PCR products were sequenced using the Sanger method by Macrogen Europe. Geneious v5.5.9 was used for alignment.

We employed RADSeq to generate a dataset of genome-wide SNPs to assess population structure among the king penguin colonies. RAD libraries were prepared using the SbfI restriction enzyme, which was chosen because it produces a large number of RAD loci in king penguins [37]. RADSeq for all individuals was performed at the Edinburgh Genomics Facility, University of Edinburgh (https://genomics.ed.ac.uk/) as described in Gonen et al. [42] after Etter et al. [43]. Briefly, 250 ng of DNA per individual was digested with SbfI-HF (NEB), followed by ligation to barcoded P1 adapters. The uniquely barcoded individuals were pooled into multiplexed libraries, and each library sheared into fragments of ~ 300—400 bp. Fragments were size selected using gel electrophoresis. The libraries were blunt ended (NEB Quick Blunting Kit) and A-tailed prior to ligation with P2 adapters (IDT). Enrichment PCR was performed to increase yield, followed by product purification with Ampure beads. The pooled, enriched libraries were checked for size and quantity using Qubit and a qPCR assay. Each library was then sequenced in a lane of the Illumina HiSeq 2500 using 125 base paired-end reads in high output mode (v4 chemistry).

Bioinformatics

FastQC was used to assess read quality and check for adapter contamination. We used *process_radtags* within the Stacks pipeline v1.35 [44, 45] to de-multiplex, trim and clean reads. We then truncated reads to 113 bp to exclude the four terminal bases in order to avoid poor sequence quality. We excluded read pairs in which either read had uncalled bases, a low quality score and/or a barcode or cut-site with more than one mismatch. The remaining paired reads were aligned to the emperor penguin reference genome (http://gigadb.org/dataset/100005) using *bwa-mem* [46]. We prevented terminal alignments by enforcing a clipping penalty of 100. Reads with more than five mismatches, multiple alignments and/or more than two indels were removed using a custom python script (filter.py, available online [47]). We removed PCR duplicates with Picardtools (http://broadinstitute.github.io/picard).

We used the Stacks pipeline (*pstacks – cstacks – sstacks – rxstacks – cstacks – sstacks - populations*) to prepare a dataset of unlinked, filtered SNPs from the RAD reads, following many of the suggestions outlined in the framework of Benestan et al. [48]. In *pstacks* we selected a minimum stack depth of six reads mapping to the same location and used the bounded SNP model with a significance level of $\alpha = 0.05$, an upper bound of 0.1 and a lower bound of 0.0041 (corresponding to the highest sequencing error rate recorded by phiX spikes in the sequencing lanes). All 64 individuals were used to build the catalog in *cstacks*. In *rxstacks* we removed confounded loci (those with a biologically implausible number of haplotypes, such as from repetitive sequences or paralogous loci) with a conservative confidence limit of 0.25. Also in *rxstacks*, we removed excess haplotypes from individuals as well as any loci with a mean log likelihood < −10. Further filtering was conducted in the *populations* module. We removed SNPs with a minor allele frequency (MAF) < 0.01 because these are likely to be the result of sequencing errors. We also removed loci with a heterozygosity > 0.5, as these could be paralogs [48]. A single SNP per RAD-tag was chosen at random in order to remove tightly linked SNPs from the dataset. We also specified that a locus must be present in all colonies to be included in the final dataset, as well as genotyped in at least 80 % of individuals from each colony. We then removed any SNPs with a mean coverage exceeding 100X using vcftools v0.1.13 [49] to avoid SNPs from repetitive regions of the genome. We also removed SNPs that were out of Hardy Weinberg equilibrium (HWE) in > 50 % of the colonies when $p < 0.01$ using the *adegenet* package in R [50, 51] and vcftools. Finally, PGDSpider v2.0.8.2 [52] was used to convert the vcf file into other formats for subsequent analyses.

Outlier loci detection

We investigated whether SNPs were potentially under selection before proceeding with population genetic analyses,

because loci under either directional or balancing selection violate the assumption of neutrality that is a caveat of most population genetic methods. We used a Bayesian F_{ST} outlier test as implemented in BayeScan 2.1 [53] to identify loci to be discarded from the neutral dataset. BayeScan has been shown to have good power for detecting loci genuinely under selection under a range of demographic scenarios, but with an accompanying high false-positive rate [54]. Given that our reason for testing for outlier loci was to obtain a truly neutral dataset, we are not concerned by the high false-positive rate in this case. We set the prior odds of neutrality parameter at five, which refers to the probability that a given locus in the dataset is under selection (i.e. for every five loci one is under selection). This prior was chosen as we aimed to remove all loci that could possibly be under selection. We deemed q-values of < 0.1 to be a significant result, meaning that for a dataset of 100 F_{ST} outliers we can expect ten of these to be false-positive neutral loci [54, 55].

Contemporary population structure

The genetic structuring among king penguin colonies was assessed using several different methods. Firstly, the Weir and Cockerham [56] unbiased estimator of F_{ST} was calculated between all pairs of colonies using Genodive v2.0b27 [57]. The hypothesis of departure from panmixia was tested with 5000 random permutations of the data to determine the statistical significance of each pairwise F_{ST} value between colonies, with the significance level adjusted for multiple testing using Sequential Goodness of Fit (SGoF+) [58].

To identify the number of genetic populations ("clusters") among the 64 individuals, we used the *find.clusters* *K*-means clustering algorithm within the *adegenet* package [50, 51], retaining all principal components. We also used a Bayesian clustering approach with a Markov chain Monte Carlo (MCMC) sampling procedure within *structure* v2.3.4 [59]. The analysis estimated the membership coefficient of each individual to each of the inferred clusters, effectively assigning individuals to genetic populations. We used the admixture model with correlated allele frequencies, because our pairwise F_{ST} results suggested that it is highly likely that these colonies have experienced admixture in the past and/or are still exchanging migrants. Models were run both with and without location priors to reflect the colony that each individual was sampled at, to detect subtle versus strong population structure. We conducted an initial run to infer the value of lambda, using a setting of $K = 1$ and an MCMC length of 100,000 generations (with the first 50,000 discarded as burn-in), allowing lambda to vary. The value of lambda was then fixed at 0.39 for subsequent analyses. K values (the number of inferred clusters) from one to four were tested, with each value of K run a total of ten times from different random seeds. Each analysis was run for 150,000 generations with the first 50,000 discarded as burn-in. *structure harvester* web v0.6.94 [60] was used to compare K values using the Evanno method [61] and prepare files for CLUMPP [62]. Replicate runs for each value of K were aligned using CLUMPP to check for multimodality, and the membership coefficients of each individual to each cluster were visualised with DISTRUCT v1.1 [63].

Discriminant Analysis of Principal Components (DAPC) [64] can be used to describe clusters in genetic data by creating synthetic variables (discriminant functions) that maximise variance among groups whilst minimising variance within groups. DAPC was run when individuals were grouped by colony of origin and when individuals were grouped by the genetic clusters found in our other analyses, for comparison. These groups were (1) South Georgia and (2) the Falkland Islands, Crozet and Macquarie. The optimal number of principal components (PCs) to retain in each analysis was determined by the average of 20 runs of the function *optim.a.score*.

We conducted individual-based population assignment tests, in which an assignment algorithm attempts to assign the individuals in the test set to their population of origin [65]. Individuals were grouped into the two genetic populations we described above. As assignment tests can be sensitive to uneven sample numbers, we randomly sampled 16 individuals from the larger population to match the size of the South Georgia population. Each group was divided into a *training* set and a *hold-out* set and we identified the most informative SNPs for colony assignment using the *training* set in TRES v1.0 [66]. We used the Informative for Assignment test (I_n) to identify ancestry informative markers (AIMs), as I_n has been shown to be the most powerful method for estimating ancestry proportions [67]. For population assignment tests it is recommended to trial different numbers of SNPs, therefore, we exported the top 100, 200, 500, 1000 and 2000 most informative SNPs. These SNP datasets were used to assign the *hold-out* set of individuals to their populations of origin within Genodive. If the minor allele was not sampled in either population (i.e. its frequency was zero) the frequency was replaced with 0.005 as recommended by Paetkau et al. [68]. We used the likelihood that the individual comes from the population it was sampled in (L_h) as the test statistic and a Monte Carlo test with 10,000 generations to estimate the null distribution of likelihood values. The threshold value was defined for each population based on the null distribution, at $\alpha = 0.05$.

Past population patterns

We used a species tree approach, as implemented in SNAPP [69] within BEAST v2.4.0 [70], to estimate the evolutionary relationships and order of splitting among

the geographically isolated colonies to determine whether any of the colonies may have been glacial refugia in the past, as well as the source of the new Falkland Islands colony. SNAPP uses a coalescent method to infer species trees from unlinked biallelic markers, such as SNPs. SNAPP is highly computationally demanding and analysis of the full dataset of individuals was implausible. We therefore selected two representative individuals from each colony (i.e. four haplotypes) for analysis, and to ensure consistency of the posterior we ran the analysis twice with different randomly-selected colony representatives. Any SNPs that were no longer polymorphic within the reduced datasets were removed from analysis, leaving datasets of 2668 and 2626 SNPs. The mutation rates (u and v) were calculated from the data, rather than estimated as part of the MCMC. We ran the MCMC for 5 million generations with a burn-in of 10 %. This was more than sufficient for convergence, with Tracer v1.6 [71] indicating ESSs > 4000. The likelihood plots were also visually inspected for convergence. The Bayesian method results in not a single topology, but a posterior distribution of the possible topologies; we used DensiTree v2.0.1 [72] to visualise the entire posterior distributions of trees as a cloudogram, excluding a 10 % burn-in.

We used RAxML v8.2.7 [73] to infer maximum likelihood (ML) phylogenies among the full dataset of king penguin individuals. We applied an ascertainment bias correction to the likelihood calculations, as recommended when using SNPs to account for the lack of invariant sites [74]. For the ascertainment correction to function, all invariant sites must be removed. In practice, this means that an alignment site consisting of only heterozygotes and homozygotes for a single allele (e.g. an alignment site that is only Rs and As with no Gs) is considered potentially invariant by RAxML and must be removed. We filtered out such sites using the Phrynomics R script (https://rstudio.stat.washington.edu/shiny/phrynomics/). After this filtering step 1727 SNPs remained in the dataset. We conducted a rapid bootstrap analysis and search for the best-scoring maximum likelihood tree in a single program run using the MRE-based bootstopping criterion [75] to ascertain when sufficient bootstrap replicates had been generated. All searches were conducted under the GTRGAMMA nucleotide substitution model. We also conducted a ML search on the HVR sequences, because HVR has been shown to resolve distinct phylogenetic lineages within Adélie penguins [29, 41, 76], emperor penguins [28] and gentoo penguins [31, 41]. We used the same search protocol as for the SNP dataset, but without an ascertainment bias correction. Finally, we constructed a median-joining haplotype network for the HVR sequences using PopArt (http://popart.otago.ac.nz).

Results

Genotyping

The 64 king penguin samples yielded 6.27–55.9 million unpaired reads per individual, with an average of 15.7 million reads per individual. On average, 97.3 % of reads per individual passed the quality filters in *process_radtags* and, of these, an average of 97.7 % successfully aligned to the emperor penguin reference genome. After specifying a minimum stack depth of six, a total of 34,171 RAD-tags remained, containing 35,766 SNPs. Our SNP filtering protocols resulted in a final dataset of 5154 SNPs [47] for use in subsequent analyses. Of these we detected no loci that were putatively under selection (BayeScan output available online [47]) and none that were out of HWE in > 50 % of colonies. There were no notable differences in genetic diversity measures (number of private alleles, expected heterozygosity, observed heterozygosity or nucleotide diversity) among colonies (Table 1).

Genetic populations of king penguins

We conducted multiple analyses of population assignment and delimitation to identify the number and geographic boundaries of distinct genetic populations among the four sampled king penguin colonies. The optimal number of clusters among the 64 individuals were $K = 3$ and $K = 2$ for *structure* analyses with and without location priors, respectively, as determined by the Evanno method. However, the highest posterior mean log probability of the data for both scenarios (i.e. with and without the sampling location specified as a prior) was at $K = 1$. The rate of change in log probability (delta$K$) is not defined at $K = 1$, and so the Evanno method is unable to determine whether this is actually the true value of K. This suggests that the signal for multiple clusters is weak. Inspection of the individual assignment plots (Fig. 2) showed that three clusters

Table 1 Genetic diversity measures by colony, based on variant (SNP) sites only.

	N private alleles	H_E (mean)	H_E (Var)	H_E (StdErr)	H_O (mean)	H_O (Var)	H_O (StdErr)	π (mean)	π (Var)	π (StdErr)
Falklands	148	0.1179	0.0175	0.0018	0.1107	0.0170	0.0018	0.1219	0.0187	0.0019
South Georgia	147	0.1161	0.0174	0.0018	0.1066	0.0161	0.0018	0.1200	0.0185	0.0019
Crozet	117	0.1178	0.0177	0.0019	0.1151	0.0183	0.0019	0.1217	0.0189	0.0019
Macquarie	180	0.1187	0.0178	0.0019	0.1115	0.0175	0.0018	0.1225	0.0189	0.0019

Number of private alleles, expected heterozygosity (H_E), observed heterozygosity (H_O) and nucleotide diversity (π)

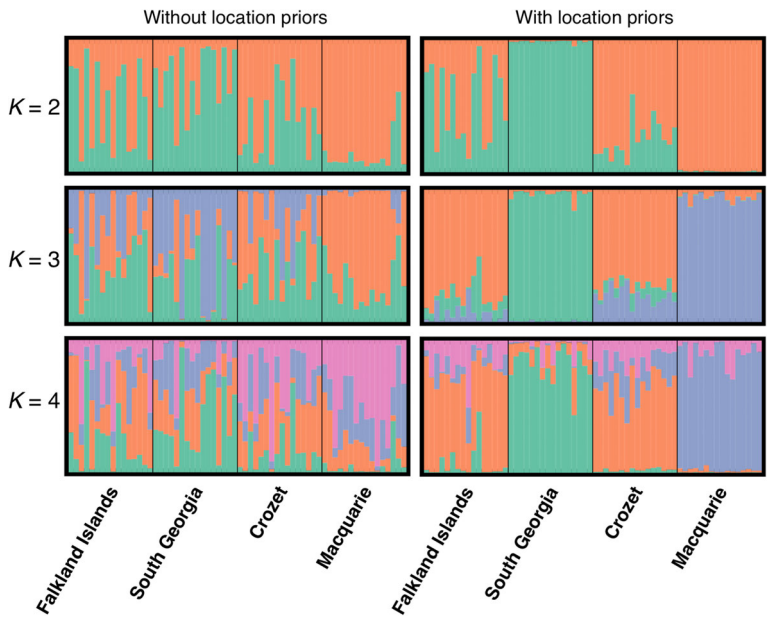


Fig. 2 Population assignment of individuals by Bayesian clustering in *structure*. Membership coefficients for each individual are shown by vertical bars with the clusters represented by colours. The Evanno method selected $K = 2$ when no location prior was used and $K = 3$ when a location prior was used. When $K = 3$ the three clusters correspond to 1) the Falkland Islands and Crozet colonies, 2) the South Georgia colony, and 3) the Macquarie Island colony

explain the majority of the subtle structure. The Falkland Islands and Crozet Islands cluster together, whereas the Macquarie Island and South Georgia colonies appear differentiated. The *K*-means clustering algorithm was unable to distinguish these clusters as the lowest value of the BIC, which indicates the optimal clustering solution, was found at $K = 1$.

Our measures of pairwise F_{ST} (Table 2) indicate that the Crozet and Falkland Islands colonies are not differentiated from one another ($F_{ST} = -0.001$), and that Macquarie and Crozet Islands are not significantly differentiated from each other ($F_{ST} = 0.001$). All other pairs of populations are statistically significantly differentiated after SGoF+ correction for multiple tests, however, the values of F_{ST} are very small (0.003–0.005), indicating only subtle genetic differences between these pairs of colonies. Therefore there are at least two slightly differentiated genetic populations among the sampled colonies: (1) the South Georgia population and (2) a population including the Falkland Islands, Crozet and Macquarie.

Table 2 Pairwise genetic differentiation (F_{ST}) between pairs of colonies

	Falkland Islands	South Georgia	Crozet
South Georgia	0.003*		
Crozet	−0.001	0.003*	
Macquarie	0.003*	0.005*	0.001

Results that are significantly different from zero at the α = 0.05 level, following SGoF+ correction, are indicated with asterisks

DAPC was unable to distinguish among the four sampled colonies or between the two slightly differentiated populations, with the distribution of individuals overlapping in both scenarios (Fig. 3). For the individual-based population assignment tests, the 100 SNP dataset was found to be best at assigning the test set of individuals back to their population of origin. However, the test performed poorly, with only seven individuals assigned correctly out of the 16 individuals in the test dataset. Given that there were only two possible populations of origin, this is slightly worse than assigning individuals to colonies at random. This again suggests that there is very little differentiation among the king penguin colonies.

Overall, our analyses of population structure among the four king penguin colonies have yielded some surprising results. Despite separation of thousands of kilometers, there is very little genetic differentiation among these colonies. The South Georgia population was subtly differentiated from all other colonies, and the Macquarie population was further very subtly differentiated from some colonies by a subset of our analyses. It is particularly interesting that the Falkland Islands colony is genetically indistinguishable from the Crozet Islands colony, despite a separation of *ca*. 7500 km,

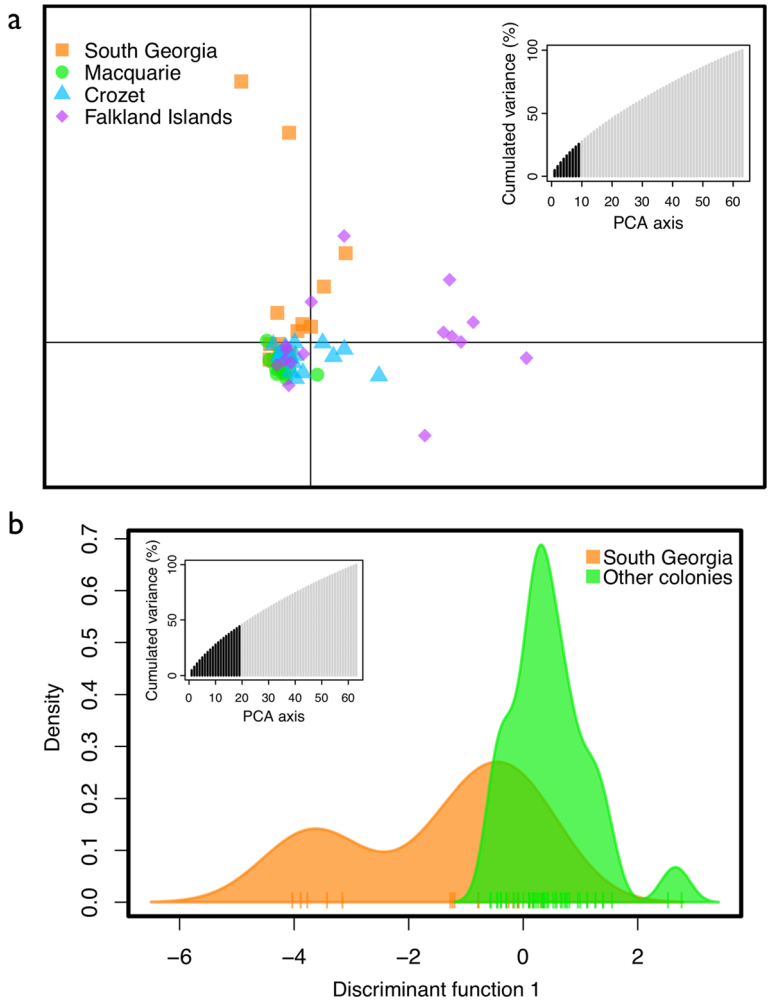


Fig. 3 Discriminant analysis of principal components. Individuals are grouped by **a**) their colony of origin and **b**) the two genetic clusters identified by other analyses. The retained PCs are shown in *black* on the inset graphs

whereas the nearby South Georgia colony is differentiated; based on our results it seems most likely that the Falkland Islands colony was founded by individuals from the Crozet Islands, rather than nearby South Georgia, even though there seems to be no obvious biological explanation for why this might be so.

Phylogeography

We attempted to ascertain the branching structure among colonies using the species tree approach implemented in SNAPP. We have presented the full posterior distribution of trees in order to highlight the uncertainty in the topology (Fig. 4). The majority of the topologies support the grouping of the Falkland and Crozet Islands colonies (Fig. 4), congruent with our *structure* and pairwise F_{ST} results. However, aside from this one clade, the rest of the branching structure among the colonies is unresolved.

We constructed maximum likelihood phylogenies for the full set of individuals using both HVR and the dataset of SNPs in order to determine if there are any strongly supported phylogenetic lineages that are not necessarily affiliated with the contemporary colony sites. The MRE bootstopping-criterion was satisfied by 550 and 800 bootstraps for the SNP and HVR searches, respectively. The best-scoring likelihood and majority rule extended consensus trees for the SNP dataset had very low support across the entire topology, with only a single node having a branch support value > 50 (topology not shown). The HVR topology did not show any more resolution, with 75 % of nodes in the tree having branch support values < 50 and no evidence of any well-supported phylogenetic lineages (topology not shown). A median joining network of the haplotypes of the mitochondrial HVR also showed no clear phylogeographic pattern and no evidence of ancestral haplotypes

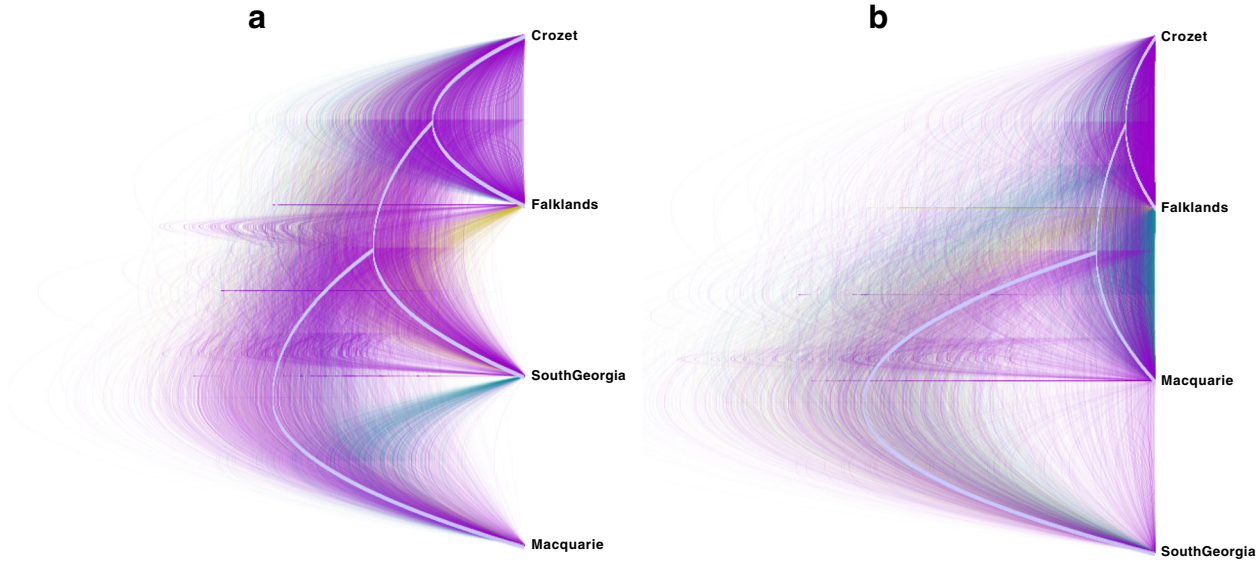


Fig. 4 Evolutionary relationships among colonies. The full posterior distributions of trees from the SNAPP analyses, excluding a 10 % burn-in, are shown. The colours represent the different topologies; *purple* is the most highly supported, *teal* is the next most supported, and *gold* is the least supported. The consensus tree is shown in *grey*. (**a**) and (**b**) are the outcomes of the two different analyses with different randomly selected representative individuals

(Additional file 1: Figure S1). Overall, there are no distinct lineages among king penguins, no remnant signatures of refugia and no evidence for the order of colonisation of the islands.

Discussion

In the first study of king penguin global population structure we found very low levels of population differentiation across the species' entire distribution, despite using 5154 SNPs distributed throughout the genome. Penguins from the Crozet Islands were not genetically differentiated from those 7450 km west on the Falkland Islands, nor those 7100 km east on Macquarie Island. There was very low, yet statistically significant, genetic differentiation between the colony on South Georgia and all other colonies, including the Falkland Islands located only 1400 km to the northwest. Our phylogeographic analyses showed no evidence of distinct king penguin lineages.

The lack of genetic differentiation across such vast distances is surprising given that king penguin colonies are sparsely distributed across the Southern Ocean. There are very few locations that support king penguin breeding between the archipelagos we have sampled; the only other colonies are in the Indian Ocean sector close to the Crozet Islands (Fig. 1). Therefore, there are very few "stepping stones" between colonies and the lack of differentiation between Crozet and Macquarie suggests that migration is not distance-limited.

There are two alternative explanations for the observed low levels of genetic differentiation among king penguin colonies. Firstly, it could be the result of frequent migration of individuals among these isolated archipelagos. In this scenario, dispersing individuals must also be recruited into the breeding population upon arrival, if they are to contribute to the gene flow that is maintaining near genetic homogeneity of king penguins. Alternatively, all extant colonies may share a common ancestral population and insufficient time has passed for them to diverge, even if they are now isolated. Despite the large geographic distances separating them, there is a growing body of evidence to suggest that king penguin colonies do exchange migrants [23–25]; we therefore consider the former hypothesis, that migration is maintaining gene flow among populations, to be the most likely explanation for the genetic similarity found here.

The recent formation of new colonies at Volunteer Point on the Falkland Islands [39], Possession Island in the Crozet Islands [11] and on Macquarie Island [9] provides direct evidence that some individuals will breed away from their natal colony. A handful of individuals banded as breeders have also been observed breeding at non-natal colonies within the Crozet Islands (Bost, C. A. *pers comm*). Furthermore, the rate of population growth at Possession Island over the past several decades has been too great to have been maintained by intrinsic recruitment alone; therefore, the population growth must be partially attributable to immigration [11]. Small numbers of king penguins, and in particular juveniles, have been observed at colonies up to 5600 km from their natal colonies [23–25, 77]. This suggests that king

penguins probably prospect other colonies and breeding habitats, including those far from their natal colony, and this may occur most often before they begin to breed. This prospecting behavior may facilitate emigration when conditions at the natal colony are less favorable than those found elsewhere.

Previous studies have shown that seabirds with large foraging ranges or those that disperse widely in the non-breeding season are least likely to show genetic differentiation among colonies [3]. During the summer breeding season, king penguin foraging trips typically last days to weeks and can cover hundreds to thousands of kilometers [78]. During the winter, king penguins rarely provision their chicks, and so adults are not restricted to central-place foraging. These winter foraging trips often take them over 1500 km away from their colonies to the marginal ice zone around Antarctica, and journeys in excess of 10,000 km have been recorded, although there is no evidence for foraging range overlap among breeding colonies thus far [79, 80]. The few juveniles that have been tracked after fledging dispersed widely in their first 6 months, probably bringing them into contact with individuals from other colonies [81]. Therefore juvenile dispersal and possibly also foraging range overlap during the non-breeding season appears to facilitate gene flow in king penguins, as it does in a variety of seabirds [4], but without more data on the winter dispersal of king penguins it is difficult to determine the relative importance of these mechanisms.

It is unclear whether the observed low level of genetic differentiation is maintained by consistent background levels of migration, or whether episodic periods of higher migration have occurred, or both. Abiotic factors such as glacial expansion and retreat, landslides, erosion, flooding, volcanic activity or other such catastrophic events [9] could result in periods of increased emigration, whilst large-scale climatic anomalies that affect the proximity of oceanic fronts and prey availability to colonies [7] could also increase the emigration rate if adults perceive the habitat quality to have declined. The harvesting of king penguins during the late 19th and early 20th century could have temporarily increased emigration rates, if individuals emigrated to less disturbed colonies. Biotic factors could also play a role, as emigration may be favored when colonies reach carrying capacity and/or density-dependent factors limit population growth, such as competition for food and nest sites, predation and pathogen load [11]. The colony at Lusitania Bay on Macquarie Island is thought to have reached carrying capacity in 1975 when all available breeding habitat was occupied and individuals were forced to spill over to other colonies [9, 10]. Two large colonies, Petite Manchotière and Jardin Japonais, on Possession Island in the Crozet Islands are also believed to have reached carrying capacity in the late 1980s, with all areas free of vegetation being occupied [11]. As these colonies approached carrying capacity, the formation of the two new colonies on Possession Island in 1979 and 1986 could have been the direct result of these large colonies spilling over, with individuals emigrating rather than competing for nest spaces at their natal colonies. This could also account for the colonization of the Falkland Islands in the late 1970s. We found no evidence for genetic differentiation between the Falkland Islands and the Crozet Islands, and the colonies grouped together in our species tree analysis. Therefore it seems likely that individuals from the Crozet Islands, possibly forced to emigrate due to competition for space at their natal colonies, founded the population at the Falkland Islands. This finding was somewhat unexpected given the 7450 km between the populations, and the relative proximity of the South Georgia population just 1400 km away. Furthermore, the observation of an individual that was banded as a chick in South Georgia but was later found breeding in the Falkland Islands [82] would also tend to suggest that the Falkland Island population would have been founded by immigrants from South Georgia. However, our genetic results indicate that there has been a higher rate of immigration from the Crozet Islands than from South Georgia.

The difference in the oceanic regime experienced by king penguins at South Georgia could explain why this colony was genetically differentiated from all other colonies [4]. South Georgia lies to the south of the Polar Front, whilst all other studied colonies lie to the north, and thus birds at South Georgia experience colder oceanic and air temperatures and a more krill-dominated food web. The different ecological conditions either side of the Polar Front appear to act as a barrier to gene flow in many species [83], including gentoo penguins [41], although this effect appears much weaker in king penguins.

While it would be useful to be able to determine the actual migration rates among the colonies studied here, the very low levels of genetic differentiation preclude the calculation of accurate estimates. Hence, whether the colonies are demographically linked or should be considered as separate management units cannot be determined [84]. Furthermore, there is currently no generalized framework for determining the level of migration necessary to maintain demographic linkage [85]. BayesAss [86], which is typically used to determine recent directional migration rates between populations (gene flow occurring over the last few generations), has been found to be unreliable when F_{ST} values are less than 0.05 (i.e. an order of magnitude greater than observed among king penguins) [87]. Methods to estimate migration based on *F*-statistics are also unreliable because the assumptions of the island model [88] that relates F_{ST} to the number of migrants

entering a population (Nm) are usually violated in natural systems, limiting the amount of quantitative information about migration that can be gained from *F*-statistics [89]. Finally, coalescent methods, such as Migrate-n [90], which estimate migration over evolutionary timescales, are also likely to be inaccurate when population differentiation is low and only a small number of loci can be used because of massive computational demands [27]. Coalescent methods also rely on an estimate of the mutation rate for the specific loci used in the analysis, to translate the mutation-scaled migration rate into an estimate of the number of migrants entering a population, and accurate mutation rates are difficult to estimate for RAD loci [91, 92].

The lack of phylogenetic signal or mitochondrial lineages suggests that small populations of king penguins have not been isolated from one another in their recent history. Some colonies went through rapid declines when king penguins were harvested for their blubber. For example, the Macquarie Island colony was reduced from hundreds of thousands of birds to about 3000 [9]. These rapid declines, although extreme demographically, were unlikely to have caused a genetic bottleneck resulting in lineage divergence, as they were neither severe enough nor lasted long enough for significant genetic drift to have taken place. Certainly there is no signature of recent genetic bottlenecks in our data. Furthermore, if the harvesting also caused a pulse of increased emigration and gene flow, then genetic diversity is unlikely to have been affected. Indeed, the Macquarie Island population appears to have retained genetic diversity throughout the period of harvesting, as demonstrated by a comparison of ancient, pre-harvest genetic diversity to the modern population [93]. The king penguin population at La Baie du Marin colony on the Crozet Islands was much smaller during the last glacial maximum (LGM), and then rapidly increased in size following Holocene warming [37]. LGM conditions appear to have isolated refugial populations of Adélie [29, 41, 76], emperor [28] and gentoo penguins [31, 41] in ice age refugia, resulting in distinct mitochondrial lineages. Our results do not support this for king penguins, although distinct lineages could exist outside of the colonies we sampled. The single mitochondrial lineage found here suggests that gene flow between populations of king penguins was maintained during the LGM even if their population sizes were reduced, and their tendency to disperse probably allowed this. Interestingly, the emperor penguin, the sister-species to king penguins in the *Aptenodytes* genus, also has remarkable dispersal abilities, exhibiting very low levels of genetic differentiation around its global range [27, 28], similar to Adélie penguins [30, 41, 76]. Yet we see distinct mitochondrial lineages in the emperor penguin, with origins dated to the last ice age [28], that are not apparent in king penguins. We propose that the sub-Antarctic distribution of king penguins may explain this contrast. Many of the sub-Antarctic islands king penguins breed on have been heavily glaciated [94], reducing available breeding area, but the increased sea ice extent during glacial periods [95] would probably not have created barriers to king penguin migration as it did not extend as far north as the king penguin's sub-Antarctic range.

Conclusions

Our study has revealed an unexpectedly low level of genetic differentiation among king penguin colonies spanning thousands of kilometers of the Southern Ocean, with some colonies separated by more than 7000 km showing no significant genetic divergence. On the other hand, the South Georgia colony does appear to be subtly differentiated from all other studied colonies, despite it lying in close proximity to the Falkland Island colony.

The very low level of genetic differentiation we have shown among king penguin colonies needs to be considered in management plans to mitigate future climate change impacts on the species. Colonies within the same archipelago are highly likely to be panmictic and demographically linked, and thus monitoring of king penguins should be considered at the archipelago level, rather than at the colony level. The subtle differentiation we found between some archipelagos, and our inability to determine whether migration is consistent or episodic, cautions against the assumption that colonies are demographically linked globally. Therefore, as a precaution, we recommend that populations at the archipelago level are managed as separate units. Given the relatively few archipelagos that host king penguins, and that climate change effects will be heterogeneous across their range, declines at any of these locations should be considered as significant and would hinder the recovery of the species, even if a loss of genetic diversity would not occur.

Demographic models that attempt to forecast extinction risk in response to large-scale climate change must also take into account migration. Recently, Tavecchia et al. (2016) showed that migration can decouple the relationship between population growth rates and climate variables, such that even if demographic rates are sensitive to climate-driven variations, this does not necessarily result in climate-driven population changes when immigration of new individuals occurs [96]. Migration could therefore buffer king penguins against their forecasted risk of extinction under climate change [6] although it may not protect them completely [7].

Additional file

Additional file 1: Figure S1. Median-joining haplotype network of king penguin HVR sequences. (PDF 281 kb)

Acknowledgements
We are grateful to Jan Cheek and the warden of Johnson's Harbour, Derek Petterson, for access to the Volunteer Point colony, and Andy Black, Sarah Crofts, Judy Horsburgh and Marion Nicolaus for their help in sample collection. We would like to thank Edinburgh Genomics staff for RAD library preparation and sequencing and gratefully acknowledge the use of the University of Oxford Advanced Research Computing (ARC) facility in carrying out this work (http://dx.doi.org/10.5281/zenodo.22558).

Funding
This work was supported by Australian Antarctic Science Programme grant 4184 (JY, KM), a NERC PhD studentship (GC), an Endeavor Research Fellowship (JY), US National Science Foundation OPP-0128913 (PN), the NERC Biomolecular Analysis Facility (GC, JY, TH), the Australian Antarctic Division (GM), Tasmanian Parks and Wildlife project 2555 (GM), the Charities Advisory Trust (TH) and public donations (TH). The funders were not involved in the design of the study, the analysis, or interpretation of the data.

Availability of data and materials
The mitochondrial DNA sequences supporting the results of this article are available in the GenBank repository, accession numbers: KX857217-KX857259. The Illumina short reads are available from the NCBI sequence read archive, study accession number: SRP089928. Our final SNP dataset, BayeScan output and custom python script (filter.py) are available from the Dryad Digital Repository http://dx.doi.org/10.5061/dryad.7c0q8 [47].

Authors' contributions
GC carried out molecular laboratory work, analyzed the data, interpreted the data, drafted the manuscript and participated in conceiving and designing the study. JY analyzed the data, interpreted the data, drafted the manuscript and participated in conceiving and designing the study. DK analyzed the data. JH, GM, PN, and PJ collected samples. KG participated in conceiving and designing the study. KM participated in conceiving and designing the study and in interpretation of the data. TH collected samples, and participated in interpreting the data and conceiving and designing the study. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Consent for publication
Not applicable.

Ethics approval and consent to participate
Ethical approval was provided by: the University of Oxford for sampling in the Falkland Islands and South Georgia; the University of Western Australia for sampling on Macquarie Island; and the Auburn University Institutional Animal Care and Use Committee and the Institut Polaire P. E. Victor for samples collected in the Crozet Archipelago.

Author details
[1]Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK. [2]Ocean & Earth Sciences, University of Southampton Waterfront Campus, European Way, Southampton SO14 3ZH, UK. [3]Department of Biology, Loyola University Chicago, 1032 W. Sheridan Road, Chicago, IL 60660, USA. [4]DST/NRF Centre of Excellence, Percy Fitzpatrick Institute of African Ornithology, Department of Zoology, Nelson Mandela Metropolitan University, South Campus, Port Elizabeth 6031, South Africa. [5]Microbiology and Immunology, PALM, University of Western Australia, Crawley, WA 6009, Australia. [6]Centre National de la Recherche Scientifique, Centre d'Ecologie Fonctionnelle et Evolutive, UMR 5175 du CNRS, 1919 route de Mende, F-34293 Montpellier Cedex 5, France. [7]Department of Biology, The Citadel, 171 Moultrie St, Charleston, SC 29409, USA. [8]Edinburgh Genomics, Ashworth Laboratories, University of Edinburgh, Edinburgh EH9 3JT, UK. [9]Australian Institute of Marine Science, The UWA Oceans Institute, 35 Stirling Highway, Crawley, WA 6009, Australia.

Received: 20 July 2016 Accepted: 30 September 2016
Published online: 13 October 2016

References

1. Manel S, Schwartz MK, Luikart G, Taberlet P. Landscape genetics: combining landscape ecology and population genetics. TREE. 2003;18(4):189–97.
2. Walther G-R, Post E, Convey P, Menzel A, Parmesan C, Beebee TJ, Fromentin J-M, Hoegh-Guldberg O, Bairlein F. Ecological responses to recent climate change. Nature. 2002;416(6879):389–95.
3. Friesen VL, Burg TM, McCoy KD. Mechanisms of population differentiation in seabirds. Mol Ecol. 2007;16(9):1765–85.
4. Friesen VL. Speciation in seabirds: why are there so many species… why aren't there more? J Ornithol. 2015;156(1):27–39.
5. Coulson J. Colonial breeding in seabirds. In: Schreiber EA, Burger J, editors. Biology of marine birds. Boca Raton: CRC Press; 2002. p. 87–113.
6. Le Bohec C, Durant JM, Gauthier-Clerc M, Stenseth NC, Park Y-H, Pradel R, Gremillet D, Gendner J-P, Le Maho Y. King penguin population threatened by Southern Ocean warming. PNAS. 2008;105(7):2493–7.
7. Bost CA, Cotté C, Terray P, Barbraud C, Bon C, Delord K, Gimenez O, Handrich Y, Naito Y, Guinet C. Large-scale climatic anomalies affect marine predator foraging behaviour and demography. Nat Commun. 2015;6:8820.
8. Bost C, Delord K, Barbraud C, Cotté C, Péron C, Weimerskirch H. King Penguin. In: Borboroglu PG, Boersma PD, editors. Penguins: Natural History and Conservation. Seattle: University of Washington Press; 2013. p. 7–19.
9. Van Den Hoff J, McMahon CR, Field I. Tipping back the balance: recolonization of the Macquarie Island isthmus by king penguins (Aptenodytes patagonicus) following extermination for human gain. Antarc Sci. 2009;21(03):237–41.
10. Rounsevell D, Copson G. Growth rate and recovery of a king penguin, aptenodytes patagonicus, population after exploitation. Wildlife Res. 1982; 9(3):519–25.
11. Delord K, Barbraud C, Weimerskirch H. Long-term trends in the population size of king penguins at Crozet archipelago: environmental variability and density dependence? Polar Biol. 2004;27(12):793–800.
12. Conroy J, White M. The breeding status of the king penguin (aptenodytes patagonica). Bull Brit Antarct Survey. 1973;32:31–40.
13. IUCN. The IUCN Red list of threatened species. 2016. http://www.iucnredlist.org. Accessed 31 May 2016.
14. Sokolov S, Rintoul SR, Wienecke B. Tracking the Polar Front south of New Zealand using penguin dive data. Deep-Sea Res Pt 1. 2006; 53(4):591–607
15. Charrassin J-B, Bost C-A. Utilisation of the oceanic habitat by king penguins over the annual cycle. Mar Ecol Prog Ser. 2001;221:285–98.
16. Scheffer A, Bost C-A, Trathan PN. Frontal zones, temperature gradient and depth characterize the foraging habitat of king penguins at South Georgia. Mar Ecol Prog Ser. 2012;465:281–97.
17. Bost C-A, Cotté C, Bailleul F, Cherel Y, Charrassin J-B, Guinet C, Ainley DG, Weimerskirch H. The importance of oceanographic fronts to marine birds and mammals of the southern oceans. J Mar Syst. 2009;78(3):363–76.
18. Jouventin P, Capdeville D, Cuenot-Chaillet F, Boiteau C. Exploitation of pelagic resources by a non-flying seabird: satellite tracking of the king penguin throughout the breeding cycle. Mar Ecol Prog Ser. 1994;106:11–9.
19. Péron C, Weimerskirch H, Bost C-A. Projected poleward shift of king penguins' (aptenodytes patagonicus) foraging range at the Crozet Islands, southern Indian ocean. P Roy Soc Lond B Biol. 2012;279(1738): 2515–23.
20. Saraux C, Viblanc VA, Hanuise N, Le Maho Y, Le Bohec C. Effects of individual Pre-fledging traits and environmental conditions on return patterns in juvenile king penguins. PLoS One. 2011;6(6):e20407.
21. Wright S. Evolution and the genetics of populations: Vol. 2. The theory of gene frequencies. Chicago: University of Chicago Press; 1969.
22. Baylis AM, Orben RA, Pistorius P, Brickle P, Staniland I, Ratcliffe N. Winter foraging site fidelity of king penguins breeding at the Falkland Islands. Mar Biol. 2015;162(1):99–110.
23. Gartshore N, Cooper J, Hunter S. Bird ringing at Marion and Prince Edward Islands, 1982–1987; with an analysis of movements since 1951. S Afr J Antarc Res. 1988;18:23–9.
24. Weimerskirch H, Jouventin P, Mougin J, Stahl J, Van BM. Banding recoveries and the dispersal of seabirds breeding in French Austral and Antarctic Territories. Emu. 1985;85(1):22–33.
25. Woehler EJ. Resighting and recoveries of banded seabirds at Heard Island, 1985–1988. Corella. 1989;13:38–40.
26. Boileau MG, Hebert PD, Schwartz SS. Non-equilibrium gene frequency divergence: persistent founder effects in natural populations. J Evol Biol. 1992;5(1):25–39.

27. Cristofari R, Bertorelle G, Ancel A, Benazzo A, Le Maho Y, Ponganis PJ, Stenseth NC, Trathan PN, Whittington JD, Zanetti E. Full circumpolar migration ensures evolutionary unity in the emperor penguin. Nat Commun. 2016;7:11842.
28. Younger JL, Clucas GV, Kooyman G, Wienecke B, Rogers AD, Trathan PN, Hart T, Miller KJ. Too much of a good thing: sea ice extent may have forced emperor penguins into refugia during the last glacial maximum. Glob Change Biol. 2015;21(6):2215–26.
29. Ritchie PA, Millar CD, Gibb GC, Baroni C, Lambert DM. Ancient DNA enables timing of the Pleistocene origin and Holocene expansion of two adélie penguin lineages in Antarctica. Mol Biol Evol. 2004;21(2):240–8.
30. Roeder AD, Marshall RK, Mitchelson AJ, Visagathilagar T, Ritchie PA, Love DR, Pakai TJ, McPartlan HC, Murray ND, Robinson NA. Gene flow on the ice: genetic differentiation among adélie penguin colonies around Antarctica. Mol Ecol. 2001;10(7):1645–56.
31. Levy H, Clucas GV, Rogers AD, Leaché AD, Ciborowski KL, Polito MJ, Lynch HJ, Dunn MJ, Hart T. Population structure and phylogeography of the gentoo penguin (pygoscelis Papua) across the Scotia Arc. Ecol Evol. 2016;6(6):1834.
32. Freer JJ, Mable BK, Clucas G, Rogers AD, Polito MJ, Dunn M, Naveen R, Levy H, Hart T. Limited genetic differentiation among chinstrap penguin (pygoscelis Antarctica) colonies in the Scotia Arc and western Antarctic Peninsula. Polar Biol. 2015;38(9):1493–502.
33. Fretwell PT, LaRue MA, Morin P, Kooyman GL, Wienecke B, Ratcliffe N, Fox AJ, Fleming AH, Porter C, Trathan PN. An emperor penguin population estimate: the first global, synoptic survey of a species from space. PLoS One. 2012;7(4):e33751.
34. Lynch H, LaRue M. First global census of the adélie penguin. Auk. 2014; 131(4):457–66.
35. Orsini L, Vanoverbeke J, Swillen I, Mergeay J, Meester L. Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. Mol Ecol. 2013;22(24):5983–99.
36. Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One. 2008;3(10):e3376.
37. Trucchi E, Gratton P, Whittington JD, Cristofari R, Le Maho Y, Stenseth NC, Le Bohec C. King penguin demography since the last glaciation inferred from genome-wide data. Proc R Soc B. 2014;281(1787):20140528.
38. Younger JL, Emmerson LM, Miller KJ. The influence of historical climate changes on Southern Ocean marine predator populations: a comparative analysis. Glob Change Biol. 2016;22(2):474–93.
39. Pistorius PA, Baylis A, Crofts S, Pütz K. Population development and historical occurrence of king penguins at the Falkland Islands. Antarc Sci. 2012;24(05):435–40.
40. Le Maho Y, Karmann H, Briot D, Handrich Y, Robin JP, Mioskowski E, Cherel Y, Farni J. Stress in birds due to routine handling and a technique to avoid it. Am J Physiol. 1992;263(4):R775–R81.
41. Clucas GV, Dunn MJ, Dyke G, Emslie SD, Levy H, Naveen R, Polito MJ, Pybus OG, Rogers AD, Hart T. A reversal of fortunes: climate change 'winners' and 'losers' in Antarctic Peninsula penguins. Sci Rep. 2014;4:5024.
42. Gonen S, Lowe NR, Cezard T, Gharbi K, Bishop SC, Houston RD. Linkage maps of the Atlantic salmon (Salmo salar) genome derived from RAD sequencing. BMC Genom. 2014;15(1):1.
43. Etter PD, Bassham S, Hohenlohe PA, Johnson EA, Cresko WA. SNP discovery and genotyping for evolutionary genetics using RAD sequencing. In: Orgogozo V, Rockman MV, editors. Molecular methods for evolutionary genetics. New York: Springer; 2011. p. 157–78.
44. Catchen J, Amores A, Hohenlohe PA, Cresko WA, Postlethwait JH. Stacks: building and genotyping loci de novo from short-read sequences. G3: genes, genomes. Genetics. 2011;1:171–82.
45. Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. Stacks: an analysis tool set for population genomics. Mol Ecol. 2013;22(11):3124–40.
46. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:13033997. 2013.
47. Clucas GV, Younger JL, Kao D, Rogers AD, Handley J, Miller GD, Jouventin P, Nolan P, Gharbi K, Miller KJ, Hart T. Data from: Dispersal in the sub-Antarctic: King penguins show remarkably little population genetic differentiation across their range. Data Dryad Repository. 2016. http://dx.doi.org/10.5061/dryad.7c0q8
48. Benestan LM, Ferchaud AL, Hohenlohe PA, Garner BA, Naylor GJ, Baums IB, Schwartz MK, Kelley JL, Luikart G. Conservation genomics of natural and managed populations: building a conceptual and practical framework. Mol Ecol. 2016;25(13):2967–77.
49. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST. The variant call format and VCFtools. Bioinformatics. 2011;27(15):2156–8.
50. Jombart T, Ahmed I. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics. 2011;27(21):3070–1.
51. Jombart T. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics. 2008;24(11):1403–5.
52. Lischer H, Excoffier L. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics. 2012;28(2):298–9.
53. Foll M, Gaggiotti O. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics. 2008;180(2):977–93.
54. Lotterhos KE, Whitlock MC. Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. Mol Ecol. 2014; 23(9):2178–92.
55. Storey JD, Tibshirani R. Statistical significance for genomewide studies. PNAS. 2003;100(16):9440–5.
56. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. Evolution. 1984;38(No. 6):1358–70.
57. Meirmans PG, Van Tienderen PH. GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. Mol Ecol Notes. 2004;4(4):792–4.
58. Carvajal-Rodriguez A, de Uña-Alvarez J. Assessing significance in high-throughput experiments by sequential goodness of fit and q-value estimation. PLoS One. 2011;6(9):e24700.
59. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000;155:945–59.
60. Earl DA. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour. 2012;4(2):359–61.
61. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol. 2005;14(8):2611–20.
62. Jakobsson M, Rosenberg NA. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics. 2007;23(14):1801–6.
63. Rosenberg NA. DISTRUCT: a program for the graphical display of population structure. Mol Ecol Notes. 2004;4(1):137–8.
64. Jombart T, Devillard S, Balloux F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet. 2010;11(1):1.
65. Paetkau D, Calvert W, Stirling I, Strobeck C. Microsatellite analysis of population structure in Canadian polar bears. Mol Ecol. 1995;4(3):347–54.
66. Kavakiotis I, Triantafyllidis A, Ntelidou D, Alexandri P, Megens H-J, Crooijmans RP, Groenen MA, Tsoumakas G, Vlahavas I. TRES: identification of discriminatory and informative SNPs from population genomic data. J Hered. 2015;106(5):672–6.
67. Ding L, Wiener H, Abebe T, Altaye M, Go RC, Kercsmar C, Grabowski G, Martin LJ, Hershey GKK, Chakorborty R. Comparison of measures of marker informativeness for ancestry and admixture mapping. BMC Genom. 2011;12(1):622.
68. Paetkau D, Slade R, Burden M, Estoup A. Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. Mol Ecol. 2004;13(1):55–65.
69. Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, RoyChoudhury A. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. Mol Biol Evol. 2012;29(8):1917–32.
70. Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard MA, Rambaut A, Drummond AJ. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput Biol. 2014;10(4):e1003537.
71. Rambaut A, Suchard MA, Xie D, Drummond AJ. Tracer v1.6. 2014. http://beast.bio.ed.ac.uk/Tracer
72. Bouckaert RR. DensiTree: making sense of sets of phylogenetic trees. Bioinformatics. 2010;26(10):1372–3.
73. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312–3.
74. Leaché AD, Banbury BL, Felsenstein J, de Oca AN-M, Stamatakis A. Short tree, long tree, right tree, wrong tree: new acquisition bias corrections for inferring SNP phylogenies. Syst Biol. 2015;64(6):1032–47.
75. Pattengale ND, Alipour M, Bininda-Emonds OR, Moret BM, Stamatakis A. How many bootstrap replicates are necessary? J Comput Biol. 2010;17(3): 337–54.

76. Younger J, Emmerson L, Southwell C, Lelliott P, Miller K. Proliferation of East Antarctic adélie penguins in response to historical deglaciation. BMC Evol Biol. 2015;15(1):1–11.
77. Weimerskirch H, Stahl J, Jouventin P. The breeding biology and population dynamics of king penguins aptenodytes patagonica on the Crozet Islands. Ibis. 1992;134(2):107–17.
78. Ratcliffe N, Trathan P. A review of the diet and at-sea distribution of penguins breeding within the CAMLR convention area. CCAMLR Sci. 2011;18:75–114.
79. Pütz K, Ropert-Coudert Y, Charrassin J, Wilson R. Foraging areas of king penguins aptenodytes patagonicus breeding at possession Island, southern Indian Ocean. Mar Ornithol. 1999;27:77–84.
80. Pütz K. Spatial and temporal variability in the foraging areas of breeding king penguins. Condor. 2002;104(3):528–38.
81. Pütz K, Trathan PN, Pedrana J, Collins MA, Poncet S, Lüthi B. Post-fledging dispersal of king penguins (aptenodytes patagonicus) from two breeding sites in the South Atlantic. PLoS One. 2014;9(5):e97164.
82. Otley H, Clausen A, Christie D, Huin N, Pütz K. Breeding patterns of king penguins on the Falkland Islands. Emu. 2007;107(2):156–64.
83. Rogers AD. Evolution and biodiversity of Antarctic organisms: a molecular perspective. In: Rogers AD, Johnston NM, Murphy EJ, Clarke A, editors. Antarctic ecosystems: an extreme environment in a changing world. Oxford: Wiley-Blackwell; 2012. p. 417–67.
84. Palsbøll PJ, Berube M, Allendorf FW. Identification of management units using population genetic data. TREE. 2007;22(1):11–6.
85. Waples RS, Gaggiotti O. INVITED REVIEW: what is a population? an empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. Mol Ecol. 2006;15(6):1419–39.
86. Wilson GA, Rannala B. Bayesian inference of recent migration rates using multilocus genotypes. Genetics. 2003;163(3):1177–91.
87. Faubet P, Waples RS, Gaggiotti OE. Evaluating the performance of a multilocus Bayesian method for the estimation of migration rates. Mol Ecol. 2007;16(6):1149–66.
88. Wright S. Evolution in mendelian populations. Genetics. 1931;16(2):97.
89. Whitlock MC, McCauley DE. Indirect measures of gene flow and migration: FST ≠ 1/(4 Nm + 1). Heredity. 1999;82(2):117–25.
90. Beerli P. How to use MIGRATE or why are Markov chain Monte Carlo programs difficult to use. In: Bertorelle G, Bruford MW, Hauffe HC, Rizzoli A, Vernesi C, editors. Population genetics for animal conservation. Cambridge: Cambridge University Press; 2001. p. 42–79.
91. Harvey MG, Brumfield RT. Genomic variation in a widespread neotropical bird (xenops minutus) reveals divergence, population expansion, and gene flow. Mol Phylogenet Evol. 2015;83:305–16.
92. Harvey MG, Smith BT, Glenn TC, Faircloth BC, Brumfield RT. Sequence capture versus restriction site associated DNA sequencing for shallow systematics. Syst Biol. 2016;65(5):910–24. doi:10.1093/sysbio/syw036.
93. Heupink TH, van den Hoff J, Lambert DM. King penguin population on Macquarie Island recovers ancient DNA diversity after heavy exploitation in historic times. Biol Lett. 2012;8(4):586–9.
94. Hall KJ. Quaternary glaciations of the Sub-Antarctic Islands. In: Ehlers J, Gibbard PL, editors. Quaternary glaciations - extent and chronology part III. Amsterdam: Elsevier; 2004. p. 339–45.
95. 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. 2005;24(7–9):869–96.
96. Tavecchia G, Tenan S, Pradel R, Igual JM, Genovart M, Oro D. Climate-driven vital rates do not always mean climate-driven population. Glob Change Biol. 2016. doi:10.1111/gcb.13330

Appendix C

SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:
MOLECULAR ECOLOGY
POPULATION GENETICS
MOLECULAR EVOLUTION
CLIMATE-CHANGE ECOLOGY

Received
17 December 2013

Accepted
25 March 2014

Published
12 June 2014

Correspondence and requests for materials should be addressed to G.V.C. (gemma.clucas@noc.soton.ac.uk)

A reversal of fortunes: climate change 'winners' and 'losers' in Antarctic Peninsula penguins

Gemma V. Clucas[1,2], Michael J. Dunn[3], Gareth Dyke[1], Steven D. Emslie[4], Hila Levy[2,5], Ron Naveen[6], Michael J. Polito[7], Oliver G. Pybus[2], Alex D. Rogers[2] & Tom Hart[2]

[1]Ocean and Earth Sciences, National Oceanography Centre, University of Southampton Waterfront Campus, European Way, Southampton, SO14 3ZH, UK, [2]Department of Zoology, South Parks Road, Oxford, OX1 3PS, UK, [3]British Antarctic Survey, High Cross, Madingley Road, Cambridge, CB3 0ET, UK, [4]Department of Biology and Marine Biology, University of North Carolina Wilmington, Wilmington, NC, 28403, USA, [5]USAF, Air Force Institute of Technology, 2950 Hobson Way, WPAFB, OH 45433-7765, [6]Oceanites Inc, PO Box 15259, Chevy Chase, MD 20825, USA, [7]Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA, 02543, USA.

Climate change is a major threat to global biodiversity. Antarctic ecosystems are no exception. Investigating past species responses to climatic events can distinguish natural from anthropogenic impacts. Climate change produces 'winners', species that benefit from these events and 'losers', species that decline or become extinct. Using molecular techniques, we assess the demographic history and population structure of *Pygoscelis* penguins in the Scotia Arc related to climate warming after the Last Glacial Maximum (LGM). All three pygoscelid penguins responded positively to post-LGM warming by expanding from glacial refugia, with those breeding at higher latitudes expanding most. Northern (*Pygoscelis papua papua*) and Southern (*Pygoscelis papua ellsworthii*) gentoo sub-species likely diverged during the LGM. Comparing historical responses with the literature on current trends, we see Southern gentoo penguins are responding to current warming as they did during post-LGM warming, expanding their range southwards. Conversely, Adélie and chinstrap penguins are experiencing a 'reversal of fortunes' as they are now declining in the Antarctic Peninsula, the opposite of their response to post-LGM warming. This suggests current climate warming has decoupled historic population responses in the Antarctic Peninsula, favoring generalist gentoo penguins as climate change 'winners', while Adélie and chinstrap penguins have become climate change 'losers'.

Climate warming around the western Antarctic Peninsula and in west Antarctica is amongst the fastest observed anywhere on Earth[1–3]. Changes in species' phenology, ranges and abundances have occurred over the past few decades[4–7] but predicting these responses is complex as they occur at all trophic levels alongside changes in the abiotic environment. For example, the extent and duration of sea ice in the region is declining[8] with correlated reductions in Antarctic krill[6] (*Euphausia superba*), the main prey item for most meso- and top-predators in the Antarctic ecosystem. However the Antarctic climate has oscillated dramatically throughout the last 50 million years (Myr). The rate of current warming is highly unusual but not unprecedented for the Holocene period[9] and the Pleistocene has been characterized by large-scale oscillations in global climate such as the 100,000 year cycles of ice ages[10]. During glacial periods, the Antarctic and sub-Antarctic ice sheets expanded and permanent sea ice was much more extensive[11].

Understanding how past climate change has affected populations is critical for distinguishing between natural and anthropogenic impacts, especially in polar regions[12]. It can also help predict probable responses to future climate change - "looking backwards to look forwards"[13]. Molecular techniques allow us to identify major events in the evolutionary and demographic history of species and populations, thus revealing how climatic events have shaped the distribution and abundance of species through time. As such, it is possible to identify species as either climate change 'winners', with populations that remain stable or expand during these events, or climate change 'losers': species that decline in abundance and distribution or become extinct. Molecular techniques can also map the distribution of biodiversity at the sub-specific level. The maintenance of genetic diversity underpins conservation genetics[14], and is a key priority of the Convention on Biological Diversity, albeit seldom measured[15].

Here we use molecular techniques to characterize the demographic history and population structure of the *Pygoscelis* penguins breeding around the Antarctic Peninsula and Scotia Arc. We specifically investigate how climate change associated with the end of the last glacial period affected *Pygoscelis* penguin populations, and we

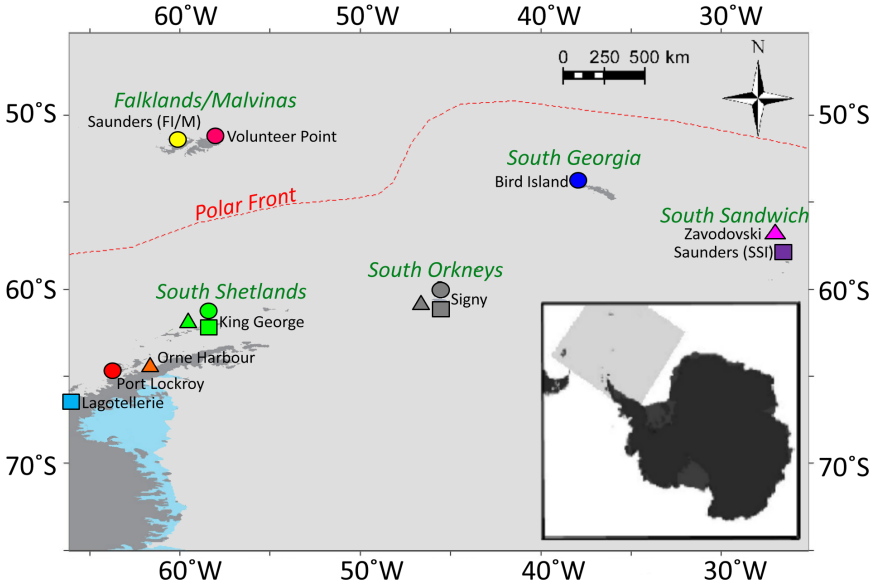


Figure 1 | **Sample locations across the Scotia Arc.** Insert shows the location of the map relative to the Antarctic continent and South America. Gentoo penguin sample locations are shown with circles, chinstrap penguin colonies with triangles and Adélie penguin colonies with squares. Each sample location is coloured independently, and is consistent with Figure 2. The archipelago names are given in green. The map was produced by TH with help from Dr. Heather Lynch using ArcGIS and modified in ArcSoft® PhotoStudio.

then compare these results with analyses from the literature about their responses to current climate change. Adélie (*Pygoscelis adéliae*), chinstrap (*P. antarctica*) and gentoo (*P. papua*) penguins are sympatric in this region, with overlapping breeding colonies in some areas and all three species show high levels of breeding site fidelity[16]. In this region, Adélie and chinstrap penguins have a diet dominated by Antarctic krill during the breeding season, whilst gentoo penguins have a more variable diet feeding on varying proportions of krill, fish and small amounts of squid[17]. *Pygoscelis* penguins are important meso-predators in the marine food web and thus are sensitive indicators of environmental change, already showing responses to current climate warming[5,18]. Indeed, in one colony in the Ross Sea, Adélie penguins appear to be climate change 'winners' as warming is creating more nesting habitat as glacial ice fields retreat[19].

Results

Population structure. We sequenced a fragment of the hypervariable region of the mitochondrial control region (HVR1) from colonies of each species spanning their entire latitudinal ranges and main breeding sites around the Antarctic Peninsula and Scotia Arc (Fig. 1). We sequenced a 316 base pair (bp) fragment from 249 gentoo penguins, a 465 bp fragment from 166 chinstrap penguins and a 601 bp fragment from 122 Adélie penguins (Table 1 and Supplementary Table S1 online; GenBank accession numbers: KJ646026-KJ646562). Although the length of the fragment sequenced in gentoo penguins was short compared to most studies of avian phylogeography, the proportion of variable sites was very high, giving sufficient information content for this study.

We detected significant population structure in gentoo penguins ($\Phi_{ST} = 0.62$, $p = 0.000$) with all colonies being genetically differentiated from one another (Supplementary Table S2 online) and showing isolation by distance ($r = 0.63$, $p = 0.003$). There was weak but significant population structure in chinstrap penguins ($\Phi_{ST} = 0.027$, $p = 0.002$) with just one colony, Zavadovski, showing genetic differentiation from the others (Supplementary Table S3 online) with no isolation by distance ($r = 0.84$, $p = 0.084$). We detected no population structure in Adélie penguins ($\Phi_{ST} = 0.007$, $p = 0.07$, Supplementary Table S4 online) despite sequencing the longest HVR1 fragment in this species, thus making our results robust to variations in fragment length.

There was significant hierarchical population structure within gentoo penguins: when colonies were grouped into Falkland Island colonies versus all other colonies, 68.9% of the genetic variation was explained by the difference between these groups (AMOVA, among groups variation = 68.9%). The haplotype network for gentoo penguins (Fig. 2A) shows that there are two distinct monophyletic lineages that do not overlap geographically. One monophyletic lineage is found in the Falkland Islands and the other corresponds to haplotypes found in colonies south of the Polar Front. These two gentoo penguin lineages have previously been classified into two subspecies based on morphological differences: Northern gentoos

Table 1 | mtDNA diversity and neutrality test results for each species and sub-species

	n	N_H	N_P	H (SD)	π (SD)	Fu's F_S	Tajima's D
Gentoo penguin	249	110	58	0.981 (0.003)	0.023 (0.012)	**−24.51*****	−0.726
P. p. papua	91	40	22	0.955 (0.009)	0.008 (0.005)	**−26.79*****	−1.222
P. p. ellsworthii	158	70	48	0.984 (0.003)	0.012 (0.007)	**−25.79*****	**−1.647***
Chinstrap penguin	166	116	46	0.987 (0.004)	0.006 (0.004)	**−26.36*****	**−1.895****
Adélie penguin	122	115	128	0.999 (0.001)	0.016 (0.008)	**−24.49*****	**−1.980****

n, number of individuals sequenced; N_H, number of haplotypes; N_P, number of polymorphic sites; H, haplotype diversity; π, nucleotide diversity; SD, standard deviation.
*denotes significance at $\alpha = 0.05$;
**denotes significance at $\alpha = 0.01$;
***denotes significance at $\alpha = 0.001$.

(*Pygoscelis papua papua*) in the Falkland Islands and Southern gentoos (*Pygoscelis papua ellsworthii*) further south[20,21].

Using a Bayesian coalescent framework implemented in BEAST we estimated the time to the most recent common ancestor of Northern and Southern gentoo penguins. To calibrate the genealogy we used the rate of molecular evolution of the HVR1 region estimated for Adélie[22] and Northern rockhopper penguins (*Eudyptes moseleyi*)[23]. These calibrations date the divergence to have occurred 25 kyr ago (95% HPD = 11–42 kya) and 44 kyr ago (95% HPD = 30–59 kya), respectively.

Demographic histories with respect to climate. All three species have undergone demographic expansions during the Holocene, as demonstrated by their star-shaped haplotype networks (Fig. 2), unimodal mismatch distributions (Supplementary Fig. S1 online) and significantly negative values of Fu's F_S statistic (Table 1). However the extent of their demographic expansions appears to have been mediated by their latitudinal distributions. This mediation is reflected in the values of Tajima's *D* test statistics, which become more negative the further south the species is distributed (Table 1). Bayesian skyline plots, which show the effective female population sizes over time, also show this latitudinal pattern (Fig. 3). Northern gentoos, breeding the furthest north, have expanded the least, Southern gentoos which breed slightly further south have expanded to a greater extent and chinstraps and Adélies, breeding the furthest south, have expanded the most. Figure 3 (bottom panel) also plots Antarctic temperature anomalies for the past 30 kyr. The population expansions all occur following the climatic warming that occurred after the last LGM, suggesting that the populations were expanding out of glacial refugia.

Discussion

The differences in the degrees of population structure in these species may be explained by their different dispersal behaviors in the austral winter. Gentoo penguins are resident at or near their colonies in winter whilst chinstrap and Adélie penguins are more dispersive, often travelling hundreds to thousands of kilometers in the winter to forage at the pack ice edge[16,17]. Winter dispersal has been shown to be an important determinant of population structure in seabirds[24] and the patterns observed here are in agreement with the majority of seabirds studied thus far; those that are residents in winter show higher levels of population structure than more migratory species. It is important to note here that we are investigating population structure at the regional, not the local scale. At the local scale, where colonies are separated by tens of kilometers or are within the same archipelago, we would not expect to find population structure as members of all three of these species have been observed visiting nearby colonies at low rates[19,25–27] and chinstrap penguins lack population structure at these scales[28]. Changes in the local conditions, such as increased sea ice or ice-bergs which block access to colonies, has been found to increase the chances of individual Adélie penguins visiting nearby colonies[25]. However it seems like the long migrations undertaken by chinstrap and Adélie penguins during the winter must facilitate gene flow at the regional level, whilst the lack of long migrations in gentoo penguins creates the population structure we have observed.

The Polar Front acts as a barrier to gene flow in many diverse marine taxa[29] because of the abrupt change in ecological conditions that it represents: from the relatively warm waters of the southern Atlantic, Pacific or Indian Ocean to the cold waters of the Antarctic Circumpolar Current. The Polar Front may also act as a barrier to

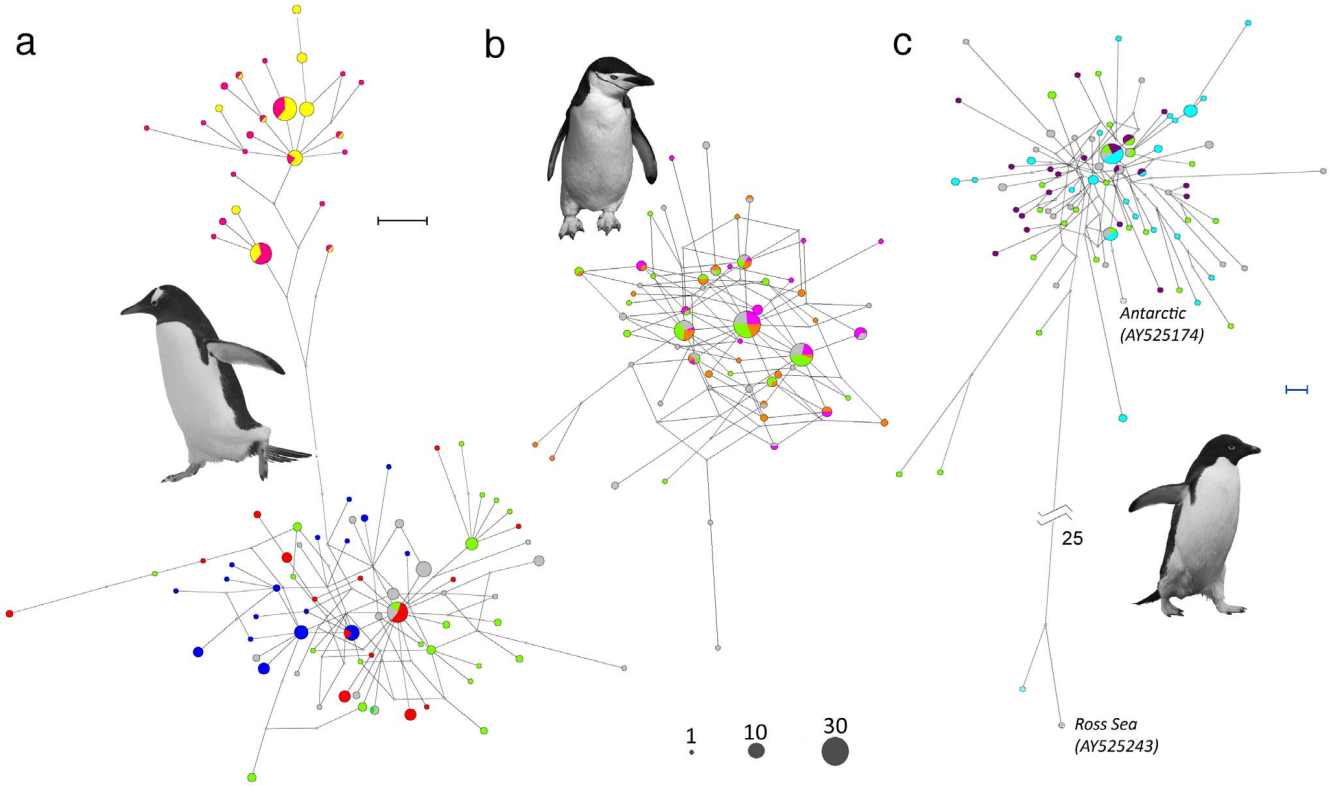


Figure 2 | Median-joining haplotype networks for gentoo (a), chinstrap (b) and Adélie penguins (c). The area of each pie chart represents the number of haplotypes as shown by the scale at the bottom. Star contraction has been applied to the chinstrap penguin haplotype network and so some of the terminal nodes are not displayed. The representatives of the "Ross Sea" and "Antarctic" lineages (with GenBank accession numbers) are indicated on the Adélie network. Colours represent the populations where the haplotype was sampled, according to symbols on Figure 1. Black scale bar shows one mutation in gentoo and chinstrap penguins; blue scale bar shows one mutation in Adélie penguins; broken line shows 25 mutational steps. Photographic images belong to TH.

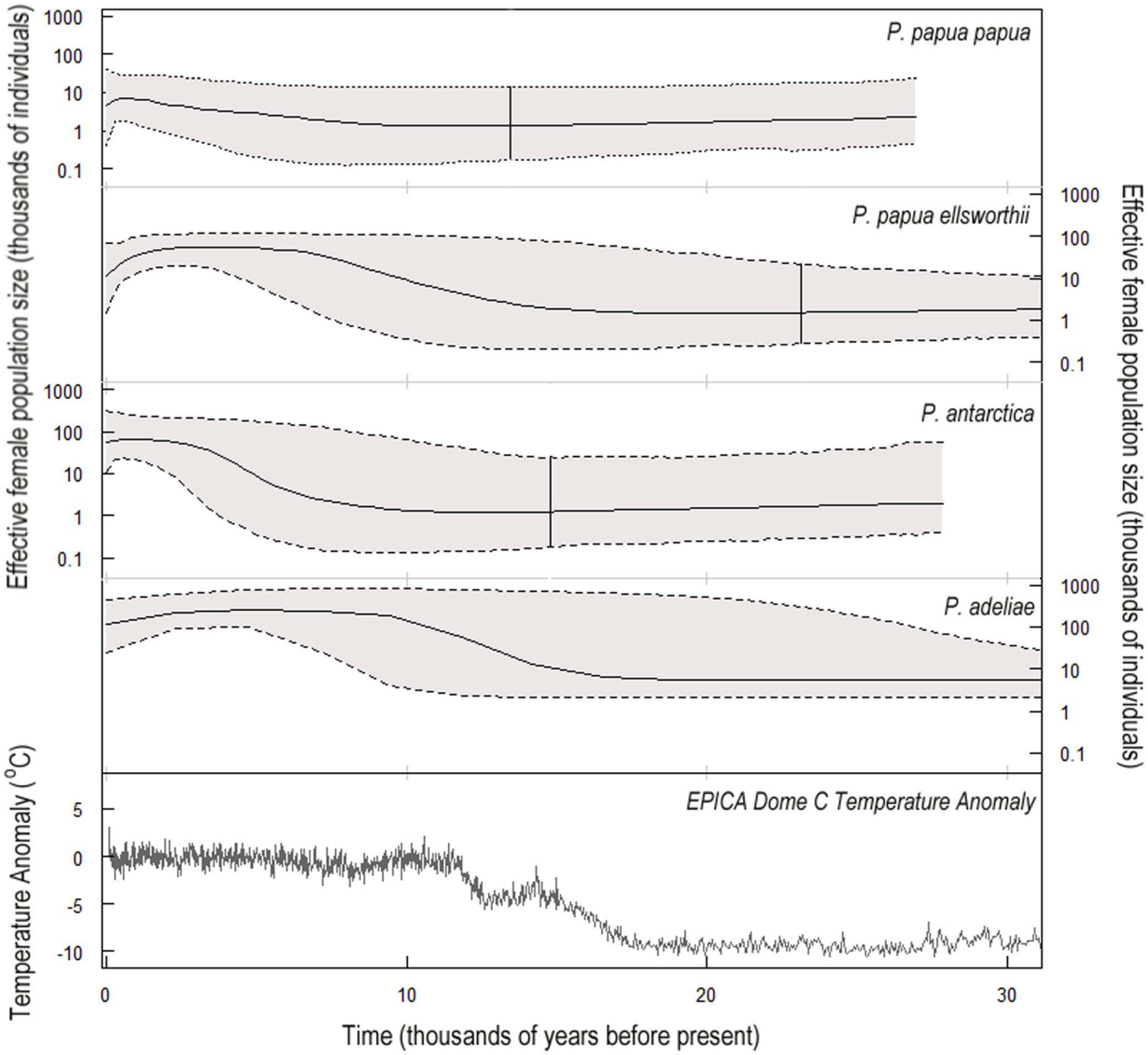


Figure 3 | Bayesian skyline plots showing the change in effective female population size for each species and sub-species. Solid lines show the median estimate; dotted lines show the 95% highest posterior density interval. Solid vertical line shows the mean tMRCA for the population, whilst the projection is made to the upper limit of the 95% highest posterior density interval of the tMRCA. The bottom panel shows the Antarctic temperature anomaly (the difference from the average of the last 1000 years) as estimated from the EPICA Dome C ice core[54].

gene flow in gentoo penguins as shown by the monophyletic lineages observed either side of it. Using rates of molecular evolution as calibrations, we estimate that the two sub-species diverged from one another between 11 and 59 kya. This means the divergence most likely occurred during the last glacial period or just after it. Lineage divergence may have occurred for two reasons. Firstly, the populations could have been isolated from one another within different glacial refugia. Long-term isolation of populations from one another results in genetic differentiation through genetic drift, and this generates distinct genetic lineages[30]. Secondly, they may have diverged from one another following glacial retreat as more habitat became available. Southern gentoos may have migrated out of a single gentoo refuge to colonize areas south of the Polar Front as the ice retreated. Lineage divergence during the last glacial period is also evident in Adélie penguins in the Ross Sea. Ritchie and colleagues[31] identified two mitochondrial DNA lineages of Adélie penguins which diverged from one another approximately 75 kya, and they suggest that limited breeding opportunities during the last glaciation separated the two lineages from one another in glacial refugia. Thus climate change in Antarctica appears to have been a strong driver of micro-evolutionary change.

Past climate change has also had a serious effect on the population sizes of the *Pygoscelis* penguins. We have shown strong evidence that these species were expanding southwards out of glacial refugia as the ice retreated after the last glacial maximum (LGM, ca 19.5–16 kya[32]), as those species which breed at higher latitudes were able to recolonize the most habitat as it became available. All three species require ice-free ground to build their nests on as well as open water in the vicinity, allowing them to travel to and from their foraging grounds during the breeding season. During the LGM, Antarctica was encircled by 100% more winter sea ice than today[32] and although summer sea ice extents are largely unknown, permanent and thick sea ice most likely surrounded the entire continent. A few polynyas persisted, but these probably would not have supported penguins

(ref[33], Sven Thatje pers. comm. 2012). Thus all the *Pygoscelis* penguins would have been forced to move further north onto islands and other land-masses that remained unglaciated and free from permanent sea ice, or were exposed by the lower sea level. It is interesting to note that sea levels were 120 m lower at the LGM compared to today[11], and so ice-age colony sites are now probably submerged.

As the climate warmed after the LGM (Figure 3, bottom panel), the extent and duration of winter sea ice declined and ice shelves retreated, allowing *Pygoscelis* penguins to expand as more habitats became available to the south. This impact is similar to current climate change, which is also reducing the extent and duration of winter sea ice around the western Antarctic Peninsula[8]. This current warming is benefitting Southern gentoo penguins, as they are expanding their range southwards and increasing in number, especially at their more southerly colonies[5]. This mirrors the pattern we detected in response to warming after the LGM: Southern gentoos expanded more than Northern gentoos. However chinstrap and Adélie penguins in the Antarctic Peninsula are currently in decline. The reasons for these declines are debated but the abundance of these two species appears to be closely linked to the availability[18] and recruitment[27] of Antarctic krill, their main prey. Adélie and chinstrap penguins showed population increases during the first part of the 20th century when climatic conditions were favorable for krill and the harvesting of marine mammals reduced competition between penguins and other krill predators[34–36]. This has been followed by chinstrap and Adélie population declines, when sea ice reductions resulted in krill declines[18]. Declines in Adélie penguins may have been exacerbated by declines in Antarctic silverfish, which are also a component of Adélie penguin diets and require sea ice for protection during larval phases[7,37]. However, others argue that krill stocks are sufficient for Adélie penguins, and there are suggestions that other factors such as snow accumulation and increased melt-water run-off are responsible for declining breeding success[38,39]. There is also evidence that in the southern sector of the Antarctic Peninsula some Adélie penguin colonies are increasing, whilst others are decreasing. Differences in population dynamics over relatively small spatial scales in this region mean that identifying a trend is difficult. However, overall it seems that climate warming is no longer benefitting all three *Pygoscelis* penguins in the Antarctic Peninsula by creating more suitable breeding habitat as it did after the LGM, but it is only benefitting the more opportunistic and generalist gentoo penguin, whose diverse and flexible foraging niche[40,41] likely make this species relatively less sensitive to declines in krill.

This 'reversal of fortunes' for two former climate change 'winners' has resulted from anthropogenic impacts outside the range of natural variation that has occurred in the past. Rapid warming trends in the Antarctic Peninsula over the past 50 years has led to decreased sea ice, loss of winter habitat, and a reduction in krill stocks that is negatively affecting Adélie and chinstrap penguins, but not gentoo penguins[5,18], which apparently are not as reliant on krill[17]. While we know of no other examples of 'reversal in fortunes' as documented here, we expect many more will be identified as global warming proceeds and biodiversity declines.

Methods

Sample collection. Shed penguin feathers were collected from Volunteer Point and Saunders Island in May 2010 and from Port Lockroy, Orne Harbour and Lagotellerie over three field seasons from 2009 to 2012. When collecting shed feathers, 80–125 molted penguin body and tail feathers were collected with feathers being collected at least 2 meters apart to minimize the chance of obtaining duplicate samples from an individual. Feathers were stored dry at ambient temperature until extraction.

Where direct samples were taken (Bird, Zavodovski, Saunders (SSI), King George and Signy Islands), birds were seized with both hands by the upper body and the flippers were restrained by the same handler. The head was placed under the arm of the handler to stop the bird biting in accordance with the literature on minimizing stress in restrained penguins[42,43]. The second person plucked two feathers from this bird's lower back or took blood samples. Where taken, bloods were from the brachial vein using a 25 G needle and syringe, and were immediately stored in 95% ethanol at ambient temperature. The animal was then released at the edge of the colony. It is possible to pluck feathers with a minimum of stress within 30 seconds, but usually no longer than 2 minutes. Blood samples usually take 2–3 minutes of restraint. Only 40 gentoo penguin adult blood samples were used in this study, previously obtained from Bird Island, South Georgia by researchers from the British Antarctic Survey. All other direct samples were plucked feathers.

Direct sampling was conducted under permits from the Falkland Islands Environmental Planning Department, The Government of South Georgia and the South Sandwich Islands, the UK Foreign and Commonwealth Office, and the US National Science Foundation. Each of these permits was issued following independent ethical review of the sampling. There are no legal restrictions covering research on animals in South Georgia or Antarctica. However, all sampling was carried out in accordance with UK Home Office guidelines and received ethical approval from the University of Oxford, the Zoological Society of London and the University of North Carolina, Wilmington. Blood sampling at Bird Island received ethical approval from the British Antarctic Survey.

DNA extraction and amplification. Feathers were prepared by finely slicing the proximal 3 mm of the feather calamus and any attached tissue for DNA extraction. Where tail feathers were available, the calamus was further sliced open and 2 mm of the blood capillary was sampled in addition to the proximal end of the calamus. Genomic DNA (gDNA) was extracted from feather fragments and blood using DNeasy Blood and Tissue Kits (http://www.qiagen.com/) according to the manufacturer's instructions for animal tissue, with the following modification to the incubation step for feather samples: 40 µl of proteinase K and 180 µl buffer ATL was added to the tissue and incubation was extended to 48 hours at 56°C.

The hypervariable region 1 (HVR-1) of the mtDNA genome was amplified from chinstrap penguin gDNA using the primers L-tRNA^{glu} and H-A650 (ref. 31,44). The primer AP1STR (5′-CCACCCTATACATACAATTCCCCTCCC-3′) was designed using Primer3 (http://primer3.wi.mit.edu/) from sequences published on GenBank to amplify the Adélie HVR-1 region paired with H-A650. The primers GPPAIR3F (5′-TTCACGTGAGGAGCCCGACCA-3′) and GPPAIR3R (5′-CTCAGGGCTAAACGGGAACTCTGC-3′) were designed in the same way to amplify the gentoo HVR-1 region. The PCR reaction mix consisted of 7.5 µl Qiagen Taq PCR Master Mix, 2 nM primers, approximately 10 ng of Adélie or gentoo gDNA, or 1 ng chinstrap gDNA, made up to a final volume of 15 µl with Milli-Q water. The thermocycling conditions for Adélie and chinstrap reactions were: 94°C for 3 minutes; 40 cycles of 94°C for 45 seconds, 52.5°C for 45 seconds and 72°C for 1 minute; followed by a 10 minute extension period at 72°C. The thermocycling conditions for the gentoo penguin amplifications were the same but the annealing temperature was raised to 54°C.

PCR products were purified and sequenced in both directions using the EZ-Seq service offered by Macrogen Europe (http://www.macrogen.com/). The same primers from the PCR amplification were used for sequencing, with the exception that chinstrap PCR products were sequenced with H-A650 and CPSEQLHS2 (5′-TTAGGGTTGTTATTGTACTCTGGA-3′). CPSEQLHS2 was designed using Primer3 from the sequences generated with H-A650, because L-tRNA^{glu} was found to be problematic when used in the chinstrap sequencing reaction.

Geneious Basic v5.6.4, created by Biomatters (http://www.geneious.com), was used to align forward and reverse sequences and extract a consensus sequence. When two fluorescent signals were observed at a single base position, as a result of heteroplasmy[22], these sites were treated as missing data.

Data analysis. Arlequin v3.5 (ref. 45) was used to calculate standard molecular diversity indices and pairwise Φ_{ST}s, to perform Mantel tests for isolation by distance, analyses of molecular variance (AMOVAs), neutrality tests and to calculate mismatch distributions. Molecular diversity measures and molecular distances were calculated where possible with the Tamura correction for unequal base frequencies and a gamma distribution model of substitution rate heterogeneity among sites. The shape parameter (α) of the gamma distribution was 0.102, 0.01 and 0.125 for gentoos, chinstraps and Adélies, respectively, as calculated in jModelTest v0.1.1 (ref. 46,47). Pairwise Φ_{ST}s were calculated between all colonies within species and significance was determined using 10,000 permutations of haplotypes between colonies, followed by the Bonferroni correction for multiple comparisons. Mantel tests for isolation by distance were performed by calculating the shortest at sea route between colonies using Google Earth v6.1 (http://earth.google.co.uk). Significance was determined through 10,000 permutations of the data. AMOVAs were used to look for hierarchical population structure. Population structures tested in gentoo penguins were: (A) no grouping of populations; (B) populations divided into Falkland Island populations (*P. p. papua*) and non-Falkland Island populations (*P. p. ellsworthii*); and (C) populations divided into *P. p. papua*, Bird Island gentoos and all other *P. p. ellsworthii*. Given the high degree of divergence found between *P. p. papua* and *P. p. ellsworthii*, AMOVAs were repeated for the *P. p. ellsworthii* populations with the following structures: (A) no population groupings; and (B) populations divided into Bird Island gentoos versus all other gentoos. Population structures tested in chinstrap penguins were: (A) no grouping of populations; and (B) populations divided into South Sandwich Island chinstraps (Zavodovzki) versus all other chinstraps. Structures tested in Adélie penguins were: (A) no grouping of populations; (B) populations grouped into South Sandwich Islands (Saunders SSI) versus Antarctic populations; and (C) populations grouped into South Sandwich Islands (Saunders SSI), northern Antarctic populations (King George Island and Signy Island) and southern Antarctic populations (Lagotellerie). 95% confidence intervals were calculated using 5,000 bootstrap replicates. Tajima's D and Fu's F_S statistics were calculated for the entire

species or sub-species (Table 1) and for each individual colony (Supplementary Table S1).

Median joining haplotype networks were drawn in Network v4.6.1.0 (http://www.fluxus-engineering.com). Because of the complexity of the chinstrap haplotype network, the star contraction option was used with a maximum star radius of five, to remove some of the terminal branches from the network. Members of the two Adélie penguin lineages identified previously were taken from GenBank (accession numbers AY525423 and AY525174) and included in the Adélie haplotype network for comparison.

To date the divergence of *P. p. papua* and *P. p. ellsworthii*, the time to their most recent common ancestor (T_{MRCA}) was estimated using the Bayesian MCMC approach implemented in BEAST v1.7.2 (ref. 48). A demographic model of exponential population growth was used and jModelTest combined with Bayes factor (BF) analysis was used to select the K2P + Γ substitution model (after all models with invariant sites were excluded). When the Bayesian Inference Criterion implemented in jModelTest supported more than one substitution model, model support was evaluated using Bayes factors[49]. The marginal likelihoods for the Bayes factor calculations were estimated under each model using both the path sampling (PS) and stepping stone sampling (SS) methods implemented in BEAST using 100 million generations and a burn-in of 10%. Statistical support was then evaluated using log (BF) using both PS and SS methods as per Kass & Raftery[50]. In the case of log (BF) < 1, which does not constitute support for either model, the simpler substitution model was selected for the main analysis. The genealogy was calibrated using two different rates of molecular evolution under a strict molecular clock model: the rate of 0.86 substitutions/site/Myr was implemented as a normally distributed prior with a standard deviation of 0.2 to represent the uncertainty in this estimate[22], whilst the rate of 0.35 s/s/Myr was implemented as a fixed rate as the uncertainty in this estimate was not published[23]. The MCMC chain was run for 100 million generations, sampling every 10,000 generations, with a burn-in of 10%. MCMC output was visualised in Tracer v1.5 (http://tree.bio.ed.ac.uk/software/tracer/) to check for convergence and mixing, and all effective sample sizes (ESSs) were >300. Each run was repeated at least four times to check for stationarity. To estimate the T_{MRCA}, the results from at least three runs were combined in Tracer.

Bayesian skyline plots (BSPs)[51], created using BEAST and Tracer, were used to investigate the historical population size of *P. p. papua*, *P. p. ellsworthii*, chinstrap and Adélie penguins. In all analyses, rate constancy between branches could not be rejected (the coefficient of variation under the uncorrelated relaxed molecular clock was not different from zero) and so strict clocks were used. All trees were calibrated with the rate of molecular evolution from Adélie penguin pedigrees as a normal prior (mean = 0.55 s/s/Myr, SD = 0.15)[22]. Weak (uniform) priors were specified for population size, with an upper limit of 1×10^{10} for the each population size parameter. jModelTest with the Bayesian Inference Criterion and Bayes factors were used to select the most appropriate substitution model for each dataset, as above. The models selected were the TN93 + G (ref. 52) model for Adélies and *P. p. ellsworthii* and the K80 + G (ref. 53) model for chinstraps and *P. p. papua*. The MCMC chain length and evaluation was described as above and the results from two or three runs were combined to make sure all ESSs were >300. A generation time of 8 years was assumed for all species. Where the shape and 95% highest posterior density (HPD) of the BSPs suggested that a constant population size through time was plausible, demographic models of exponential growth versus constant population size were compared using Bayes factors, keeping all other settings in the MCMC run the same. In all cases, the exponential growth model received greater support than a constant size model.

1. Smith, R. C. *et al.* Marine ecosystem sensitivity to climate change. *BioScience* **49**, 393–404 (1999).
2. Vaughan, D. G. *et al.* Recent rapid regional climate warming on the Antarctic Peninsula. *Climatic Change* **60**, 243–274 (2003).
3. Bromwich, D. H. *et al.* Central West Antarctica among the most rapidly warming regions on Earth. *Nat Geosci* **6**, 139–145 (2013).
4. Lynch, H. J., Fagan, W. F., Naveen, R., Trivelpiece, S. G. & Trivelpiece, W. Z. Differential advancement of breeding phenology in response to climate may alter staggered breeding among sympatric pygoscelid penguins. *Mar Ecol Prog Ser* **454**, 135–145 (2012).
5. Lynch, H. J., Naveen, R., Trathan, P. N. & Fagan, W. F. Spatially integrated assessment reveals widespread changes in penguin populations on the Antarctic Peninsula. *Ecology* **93**, 1367–1377 (2012).
6. Atkinson, A., Siegel, V., Pakhomov, E. & Rothery, P. Long-term decline in krill stock and increase in salps within the Southern Ocean. *Nature* **432**, 100–103 (2004).
7. Ducklow, H. W. *et al.* Marine pelagic ecosystems: the west Antarctic Peninsula. *Philos Trans Roy Soc B* **362**, 67–94 (2007).
8. Smith, R. C. & Stammerjohn, S. E. Variations of surface air temperature and sea-ice extent in the western Antarctic Peninsula region. *Ann Glaciol* **33**, 493–500 (2001).
9. Mulvaney, R. *et al.* Recent Antarctic Peninsula warming relative to Holocene climate and ice-shelf history. *Nature* **489**, 141–144 (2012).
10. Shackleton, N. J. The 100,000-year ice-age cycle identified and found to lag temperature, carbon dioxide, and orbital eccentricity. *Science* **289**, 1897–1902 (2000).
11. Williams, M., Dunkerley, D., De Decker, P., Kershaw, P. & Chappell, J. *Quaternary Environments.* (Arnold, London, 1998).
12. Sun, L. G. *et al.* Vertebrate records in polar sediments: Biological responses to past climate change and human activities. *Earth-Sci Rev* **126**, 147–155 (2013).
13. Hoelzel, A. R. Looking backwards to look forwards: conservation genetics in a changing world. *Conserv Genet* **11**, 655–660 (2010).
14. Frankel, O. *Conservation and Evolution.* (Cambridge University Press, 1981).
15. Laikre, L. *et al.* Neglect of genetic diversity in implementation of the convention on biological diversity. *Conserv Biol* **24**, 86–88 (2010).
16. Borboroglu, P. B. & Boersma, P. D. *Penguins: Natural History and Conservation.* (University of Washington Press, 2013).
17. Ratcliffe, N. & Trathan, P. A review of the diet and at-sea distribution of penguins breeding within the CAMLR Convention Area. *CCAMLR Science* **18**, 75–114 (2011).
18. Trivelpiece, W. Z. *et al.* Variability in krill biomass links harvesting and climate warming to penguin population changes in Antarctica. *PNAS* **108**, 7625–7628 (2011).
19. LaRue, M. A. *et al.* Climate change winners: receding ice fields facilitate colony expansion and altered dynamics in an Adelie penguin metopopulation. *PLoS One* **8**, e60568 (2013).
20. Stonehouse, B. Geographic variation in Gentoo penguins *Pygoscelis papua*. *Ibis* **112**, 52–57 (1970).
21. de Dinechin, M. *et al.* The biogeography of Gentoo Penguins (*Pygoscelis papua*). *Can J Zool* **90**, 352–360 (2012).
22. Millar, C. D. *et al.* Mutation and evolutionary rates in Adélie penguins from the Antarctic. *PLoS Genet* **4**, e1000209 (2008).
23. de Dinechin, M., Ottvall, R., Quillfeldt, P. & Jouventin, P. Speciation chronology of rockhopper penguins inferred from molecular, geological and palaeoceanograhic data. *J Biogeog* **36**, 693–702 (2009).
24. Friesen, V. L., Burg, T. M. & McCoy, K. D. Mechanisms of population differentiation in seabirds. *Mol Ecol* **16**, 1765–1785 (2007).
25. Shepherd, L. D. *et al.* Microevolution and mega-icebergs in the Antarctic. *PNAS* **102**, 16717–16722 (2005).
26. Clausen, A. & Pütz, K. Winter diet and foraging range of gentoo penguins (Pygoscelis papua) from Kidney Cove, Falkland Islands. *Polar Biol* **26**, 32–40 (2003).
27. Hinke, J. T., Salwicka, K., Trivelpiece, S. G., Watters, G. M. & Trivelpiece, W. Z. Divergent responses of Pygoscelis penguins reveal a common environmental driver. *Oecologia* **153**, 845–855 (2007).
28. Korczak-Abshire, M., Chwedorzewska, K. J., Wąsowicz, P. & Bednarek, P. T. Genetic structure of declining chinstrap penguin (Pygoscelis antarcticus) populations from South Shetland Islands (Antarctica). *Polar Biol* **35**, 1681–1689 (2012).
29. Rogers, A. D. *Antarctic Ecosystems: an Extreme Environment in a Changing World* Alex, D. Rogers, Nadine, M. Johnston, Eugene, J. Murphy & Andrew, Clarke (eds) 417–467 (Wiley-Blackwell, 2012).
30. Hewitt, G. The genetic legacy of the Quaternary ice ages. *Nature* **405**, 907–913 (2000).
31. Ritchie, P. A., Millar, C. D., Gibb, G. C., Baroni, C. & Lambert, D. M. Ancient DNA enables timing of the Pleistocene origin and Holocene expansion of two Adélie penguin lineages in Antarctica. *Mol Biol Evol* **21**, 240–248 (2004).
32. 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).
33. Thatje, S., Hillenbrand, C. D., Mackensen, A. & Larter, R. Life hung by a thread: endurance of Antarctic fauna in glacial periods. *Ecology* **89**, 682–692 (2008).
34. Sladen, W. *Biologie Antarctique* Carrick, R., Holdgate, M. W. & Prevost, J. (eds) 359–365 (Hermann, Paris, 1964).
35. Laws, R. Seals and whales of the Southern Ocean. *Philos T Roy Soc B* **279**, 81–96 (1977).
36. Laws, R. M. The ecology of the Southern Ocean. *Am Sci* **73**, 26–40 (1985).
37. Schofield, O. *et al.* How do polar marine ecosystems respond to rapid climate change? *Science* **328**, 1520–1523 (2010).
38. Sailley, S. F. *et al.* Carbon fluxes and pelagic ecosystem dynamics near two western Antarctic Peninsula Adélie penguin colonies: an inverse model approach. *Mar Ecol Prog Ser* **492**, 253–272 (2013).
39. Chapman, E. W., Hofmann, E. E., Patterson, D. L., Ribic, C. A. & Fraser, W. R. Marine and terrestrial factors affecting Adélie-penguin Pygoscelis adeliae chick growth and recruitment off the western Antarctic Peninsula. *Mar Ecol Prog Ser* **436**, 273–289 (2011).
40. Miller, A. K., Karnovsky, N. J. & Trivelpiece, W. Z. Flexible foraging strategies of gentoo penguins Pygoscelis papua over 5 years in the South Shetland Islands, Antarctica. *Mar Biol* **156**, 2527–2537 (2009).
41. Miller, A. K., Kappes, M. A., Trivelpiece, S. G. & Trivelpiece, W. Z. Foraging-Niche Separation of Breeding Gentoo and Chinstrap Penguins, South Shetland Islands, Antarctica. *The Condor* **112**, (2010).
42. Le Maho, Y. *et al.* Stress in birds due to routine handling and a technique to avoid it. *Am J Physiol* **263**, R775–R781 (1992).
43. Wilson, R. P. A method for restraining penguins. *Mar Ornithol* **25**, 72–73 (1997).
44. Roeder, A. D., Ritchie, P. A. & Lambert, D. M. New DNA markers for penguins. *Conserv Genet* **3**, 341–344 (2002).

45. Excoffier, L. & Lischer, H. E. L. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* **10**, 564–567 (2010).
46. Posada, D. Selection of models of DNA evolution with jModelTest. *Methods Mol Biol* **537**, 112 (2009).
47. Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**, 696–704 (2003).
48. Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* **29**, 1969–1973 (2012).
49. Baele, G. *et al*. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol Biol Evol* **29**, 2157–2167 (2012).
50. Kass, R. E. & Raftery, A. E. Bayes factors. *J Am Stat Assoc* **90**, 773–795 (1995).
51. Drummond, A. J., Rambaut, A., Shapiro, B. & Pybus, O. G. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol* **22**, 1185–1192 (2005).
52. Tamura, K. & Nei, M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* **10**, 512–526 (1993).
53. Kimura, M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* **16**, 111–120 (1980).
54. Jouzel, J. *et al*. EPICA Dome C Ice Core 800 KYr Deuterium Data and Temperature Estimates. *IGBP PAGES/World Data Center for Paleoclimatology Data Contribution Series* # 2007-091. NOAA/NCDC Paleoclimatology Program, Boulder CO, USA (2007).

Acknowledgments

We thank the Zoological Society of London, Quark Expeditions, Exodus Travels ltd., Oceanites, the Holly Hill Charitable Trust, the Charities Advisory Trust and an U.S. National Science Foundation (NSF) Office of Polar Programs grant (ANT-0739575) for funding. Thanks to Dr. Richard Phillips, Dr. Kevin Hughes, Jérôme Poncet, Dr. Wayne Trivelpiece, the U.S. Antarctic Marine Living Resource Program, and the crew of the Golden Fleece for logistical and sample collection support. Thanks to Dr. Kate Ciborowski, Dr. Andrew Hitchcock and Dr. John Gittins for advice in the lab, and Dr. Simon Ho and Dr. Mike Lee for advice about BEAST.

Author contributions

G.V.C. performed all lab and analytical work, and wrote the paper. T.H., S.D.E., M.J.D., M.J.P., H.L. and R.N. conducted field work and provided samples. O.G.P. provided analytical support with Bayesian coalescent inferences. T.H. designed the study and prepared all figures. G.D. and A.D.R. were also involved in study design. All authors discussed the study results and implications and contributed to editing the manuscript.

Additional information

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Clucas, G.V. *et al*. A reversal of fortunes: climate change 'winners' and 'losers' in Antarctic Peninsula penguins. *Sci. Rep.* **4**, 5024; DOI:10.1038/srep05024 (2014).

Appendix D

Ecology and Evolution

Open Access

Population structure and phylogeography of the Gentoo Penguin (*Pygoscelis papua*) across the Scotia Arc

Hila Levy[1,2,†], Gemma V. Clucas[1,3,†], Alex D. Rogers[1], Adam D. Leaché[4], Kate L. Ciborowski[5], Michael J. Polito[6], Heather J. Lynch[7], Michael J. Dunn[8] & Tom Hart[1]

[1]Department of Zoology, University of Oxford, South Parks Road, Oxford, OX1 3PS, UK
[2]USAF, Air Force Institute of Technology, 2950 Hobson Way, WPAFB, Ohio 45433-7765
[3]Ocean and Earth Sciences, University of Southampton, Waterfront Campus, European Way, Southampton, SO14 3ZH, UK
[4]Department of Biology and Burke Museum of Natural History and Culture, University of Washington, Box 351800, Seattle, Washington 98195-1800
[5]Department of Biology, University of Bristol, Woodland Road, Bristol, BS8 1UG, UK
[6]Department of Oceanography and Coastal Sciences, Louisiana State University, Baton Rouge, Louisiana 70803
[7]Department of Ecology and Evolution, Stony Brook University, Stony Brook, New York 11794
[8]British Antarctic Survey, High Cross, Madingley Road, Cambridge, CB3 0ET, UK

Keywords
Bayesian phylogeography, dispersal, microsatellites, penguins, population genetics, *Pygoscelis papua*.

Correspondence
Hila Levy, Department of Zoology, University of Oxford, South Parks Road, Oxford, OX1 3PS, UK.
Tel: +44 (0)1865 281329;
E-mail: hila.levy@zoo.ox.ac.uk

Funding Information
This study was funded by the Falkland Islands Government Environmental Planning Department, the Charities Advisory Trust, the John Ray Trust, Sigma Xi, and the Rhodes Trust. We thank British International Helicopters Ltd in the Falkland Islands, the US Air Force, and Royal Air Force for their support of this project.

Received: 23 October 2015; Revised: 8 December 2015; Accepted: 10 December 2015

doi: 10.1002/ece3.1929

[†]These authors have provided equal contribution and share first authorship.

Abstract

Climate change, fisheries' pressure on penguin prey, and direct human disturbance of wildlife have all been implicated in causing large shifts in the abundance and distribution of penguins in the Southern Ocean. Without mark-recapture studies, understanding how colonies form and, by extension, how ranges shift is challenging. Genetic studies, particularly focused on newly established colonies, provide a snapshot of colonization and can reveal the extent to which shifts in abundance and occupancy result from changes in demographic rates (e.g., reproduction and survival) or migration among suitable patches of habitat. Here, we describe the population structure of a colonial seabird breeding across a large latitudinal range in the Southern Ocean. Using multilocus microsatellite genotype data from 510 Gentoo penguin (*Pygoscelis papua*) individuals from 14 colonies along the Scotia Arc and Antarctic Peninsula, together with mitochondrial DNA data, we find strong genetic differentiation between colonies north and south of the Polar Front, that coincides geographically with the taxonomic boundary separating the subspecies *P. p. papua* and *P. p. ellsworthii*. Using a discrete Bayesian phylogeographic approach, we show that southern Gentoos expanded from a possible glacial refuge in the center of their current range, colonizing regions to the north and south through rare, long-distance dispersal. Our findings show that this dispersal is important for new colony foundation and range expansion in a seabird species that ordinarily exhibits high levels of natal philopatry, though persistent oceanographic features serve as barriers to movement.

Introduction

Crucial to the study of evolution is characterizing population differentiation and understanding the mechanisms that disrupt gene flow. Population differentiation is the first step toward reproductive isolation in the classic model of allopatric speciation (Mayr 1942) and creates opportunities for local adaptation. Extrinsic barriers to

dispersal are common in terrestrial environments (e.g., mountain ranges, rivers), but less obvious in the marine environment, where many taxa are highly mobile and physical barriers may only be manifested as changes in hydrography or fronts between water masses. Intrinsic factors, such as breeding asynchrony or natal philopatry, may therefore be expected to be more important in disrupting gene flow in the marine environment than in terrestrial systems. Seabirds provide useful models to investigate these mechanisms. They are tied to breeding locations annually, making them relatively easy to sample, yet they range widely during the nonbreeding season, allowing the comparison of extrinsic versus intrinsic factors in generating population differentiation. Characterizing population structure and dispersal is also critical to understanding population dynamics. This is particularly pertinent in regions undergoing rapid climate change, where large perturbations in population sizes can be expected. Whether these changes are driven by changes in demographic rates, such as survival and breeding success, or dispersal and range shifts, can further inform our understanding of the drivers of population differentiation and evolution.

The Antarctic Peninsula and archipelagos lying in the Scotia Sea in the South Atlantic, better known as the Scotia Arc, are showing marked physical and ecological changes and may be experiencing some of the most rapid climate change on the planet (Vaughan et al. 2003; Fox and Vaughan 2005; Rignot et al. 2005; Ducklow et al. 2007; Mayewski et al. 2009; Cook and Vaughan 2010). Increases in sea surface temperature and changes in the extent and seasonal timing of sea ice coverage in this region (Zwally et al. 2005; Ainley et al. 2010; Lynch et al. 2012) are thought to affect predators through changes in the abundance of krill (*Euphausia superba*) (Trivelpiece et al. 2011) and breeding habitat (Fretwell and Trathan 2009; Trathan et al. 2011). The role of penguins in the Antarctic ecosystem is crucial, as they make up a large part of the avian biomass in the region, and serve as marine mesopredators. As sentinels of changes to the sensitive environments that they inhabit, penguin population dynamics are frequently used to measure the impacts of anthropogenic factors such as fishing, pollution, and global climate change (Boersma 2008). Penguins are thought to be useful indicator species because they are mobile, long-lived predators with demographic rates that correlate strongly with environmental conditions and integrate the effects of physical and biological variability in the Antarctic environment over large temporal and spatial scales (Fraser et al. 1992; Trathan et al. 2007; Trivelpiece et al. 2011). Questions regarding gene flow and demography are important components to scientific understanding of such environmental changes. Studies have made attempts to predict the responses of penguins to such changes (Jenouvrier et al. 2009; Ainley et al. 2010), but without a mechanistic understanding of their responses and the factors that may limit their dispersal, we may be unable to make accurate predictions. Paleoecological evidence from ancient penguin colonies suggests that colonization and extinction of breeding populations have long been a part of their metapopulation dynamics (Emslie 2001; Emslie et al. 2007) and so delimiting these metapopulations using genetic techniques has been identified as a priority for future research (Chown et al. 2015). Penguins in the Scotia Arc have already shown significant changes in population sizes and ranges in concert with climate change (Woehler et al. 2001; Croxall et al. 2002; Lynch et al. 2012). This is probably not a new phenomenon, as the Antarctic ecosystem has undergone multiple fluctuations in temperature and ice extent over geological timescales. We should therefore expect penguin populations to be relatively flexible in responding to environmental fluctuations. At present, Gentoo penguin (*Pygoscelis papua*, depicted in Fig. 1) populations are increasing and moving south (Lynch et al. 2012), while Adélie and Chinstrap populations are declining in the region north of Marguerite Bay (Casanovas et al. 2015), possibly because of differential survival rather than reproductive success (Lynch et al. 2010). Understanding the genetic structure of these populations is important for interpreting whether changes in population size result from changes in local survival and recruitment or, alternatively, migration. While the southward progression of Gentoo penguins is now well documented (Fraser et al. 1992; Lynch et al. 2012), the genetic origin of new populations remains unstudied. Moreover, the observed

Figure 1. Adult Gentoo penguins (*Pygoscelis papua*) guard their chicks on the nest in the Western Antarctic Peninsula. (Photo: Hila Levy).

establishment of a colony at the edge of this species' range, at Port Lockroy on the Antarctic Peninsula circa 1985 (Trathan et al. 2008), affords us the opportunity to investigate where migrants originate, and whether migration continues to play a part in the new colony's dynamics post-colonization. Answering this question could reveal the scale at which Gentoo penguins may exist in metapopulations, and thus the scales at which demographic models and conservation efforts should be targeted. In light of broader effects of climate change and competition for prey on faunal range expansion, understanding mechanisms of colonization is particularly important in determining the best locations for networks of protected areas to maintain population viability and genetic variation (Xu et al. 2006; Akçakaya et al. 2007). Competition with fisheries for prey such as krill and fish is concerning to marine biologists and conservationists. Competition for prey has been known to drive faunal range expansions in marine species such as grey seals (Heide-Jorgensen et al. 1992) and terrestrial polar species (Hersteinsson and MacDonald 1992). Additionally, shifts in environmental conditions can drive bird and insect species to expand their geographic ranges (Parmesan et al. 1999; Thomas and Lennon 1999; Bennie et al. 2013). In the case of Gentoo penguins, both rapid climate change and fisheries' pressures are of concern in the Western Antarctic Peninsula, where rising temperatures and growing krill fisheries are in place. Designing an effective network of protected areas would require knowledge of patterns of colonization and movement for the species. We therefore seek to identify the population structure of Gentoo penguins across their latitudinal range and identify where the founders of a new population originated.

Gentoo penguins reside both above and below the Antarctic Polar Front, which separates the temperate climate of the Falkland Islands from the sub-Antarctic climate of South Georgia. Further south, the climate of the South Orkney and South Shetland Islands and the western side of the Antarctic Peninsula is commonly described as maritime Antarctic (Laws 1984). There are two recognized subspecies of Gentoo penguin, and a third subspecies, recently suggested but not yet described, in the Indian Ocean (de Dinechin et al. 2012). The Northern Gentoo (*Pygoscelis papua papua*) breeds on the Falkland Islands, whilst the Southern Gentoo (*P. papua ellsworthii*) is found breeding in the Southern Ocean on South Georgia, the South Sandwich Islands, the South Orkney and South Shetland Islands and the Western Antarctic Peninsula (Stonehouse 1970; de Dinechin et al. 2012).

While they have a large breeding distribution, Gentoo penguins also exhibit high levels of natal philopatry and tend to remain within the same archipelagos year-round (Stonehouse 1970; Tanton et al. 2004; Ratcliffe and Trathan 2011). However, determining the limits of the species' foraging range over a lifetime is difficult because of ethical, cost, and technical challenges associated with long-term tag or transponder use (Froget et al. 1998; Gauthier–Clerc et al. 2004; Saraux et al. 2011; Dann et al. 2014). Previous studies have indicated that Gentoo penguins have a maximum observed dispersal of 276 km during the nonbreeding season in the Falkland Islands (Clausen and Pütz 2003) and 268 km in the South Shetland Islands (Wilson et al. 1998). When foraging to provide for their chicks in the summer brood and crèche phases, they tend to stay within 50 km (Williams and Rodwell 1992; Ratcliffe and Trathan 2011) of their breeding colony. Gentoo penguins are rarely observed far out at sea (Jehl et al. 1979; Thurston 1982; White et al. 2002), although individuals have been observed as far as 2000 km from the nearest potential breeding point (Voisin 1979; Enticott 1986), indicating that long-distance dispersal is physically possible even if its impact on genetic structuring is poorly understood. Having discrete nonbreeding habitats with large stretches of ocean between archipelagos, this species is likely to show considerable population genetic differentiation (Friesen et al. 2007a).

Mitochondrial DNA studies have revealed significant population structure in Gentoo penguins previously (de Dinechin et al. 2012; Clucas et al. 2014; Peña et al. 2014), but additional studies of population genetic structure using multiple loci are likely to play an important role in our understanding of species' responses to environmental stressors in the polar regions by demonstrating fine-scale levels of connectivity and dispersal barriers. For example, recent population models of Emperor penguin declines against climate change (Jenouvrier et al. 2009, 2014) have assumed populations had limited dispersal, although more recent evidence from satellites, used in high-resolution censuses of seabird populations, suggests that colonies are more fluid than had been previously believed (Trathan et al. 2011; LaRue et al. 2015). Although the impact of these movement patterns on long-term population projections is unknown, it is clear that even nominally site-faithful species can have complex spatial dynamics that could influence our interpretation of mark-recapture studies or our understanding of habitat suitability (Dugger et al. 2014). With poor resolution on the forces that have led to such wide-ranging distribution of Gentoo penguins worldwide, we set out to use genetic markers to delineate the extent to which philopatry and oceanic barriers shape the population structure of these seabirds.

Using microsatellites taken from a range of previous studies, along with sequences of the mitochondrial hypervariable control region, we determine the population

structure of Gentoo penguins across the Polar Front and around the Scotia Arc. Using fine-scale sampling within the Falkland Island archipelago, and large-scale sampling across the Scotia Arc region, we are able to describe the population structure in the region, while assessing the origin of a recent founder population and describe how areas were colonized postglacially. We further interpret these results in the context of recent shifts in population size and range expansion in concert with environmental change across the Scotia Arc, and implications for conservation policy.

Materials and Methods

Population sampling

We collected DNA samples using a mixture of shed feathers, plucked feathers, and blood samples from locations shown in Figure 2. We collected shed feathers from 10 Falkland Island colonies (Volunteer Point, Kidney Cove, Bluff Cove, Bertha's Beach, Ajax Bay, New Haven, Fox Bay, Saunders Penguin Island, Saunders Penarrow Point, and Shallow Harbour) in May 2010 and from Port Lockroy on the Antarctic Peninsula in February 2010. At each of these sites, we took 80–125 molted penguin body and tail (retrix) feathers from nesting sites. We picked feathers from at least 2 m apart to minimize the chance of obtaining duplicate samples from an individual. All colonies sampled are known to contain at least 300 breeding pairs. We also excluded one of any pair of samples with identical mitochondrial and microsatellite genotypes. We plucked feathers from breeding birds at King George Island in the South Shetland Islands and Signy Island in the South Orkneys, and genotyped blood samples taken from adult birds in a previous study on Bird Island, South Georgia. Blood samples were drawn from the brachial vein using a 25G needle and syringe, and stored in 95% ethanol at −20°C. Feather samples were individually stored dry at ambient temperature until extraction.

DNA extraction

Between 48 and 54 feathers from each colony were selected for extraction based on feather cleanliness and appearance. Tail feathers (retrices) were prioritized over molted body feathers, as they contain a blood supply (Williams 1995). For all feathers, approximately 5 mm of the proximal end (calamus) of the feather was finely chopped using a sterile razor blade, and deposited into a 1.5-mL sterile microcentrifuge tube with sterile tweezers. In the case of tail feathers, the retrix was first cut down to a length of approximately 10 cm, cutting away the shaft (rachis) and leaving the superior and inferior umbilicus intact. Heavy-duty sterile scissors were then

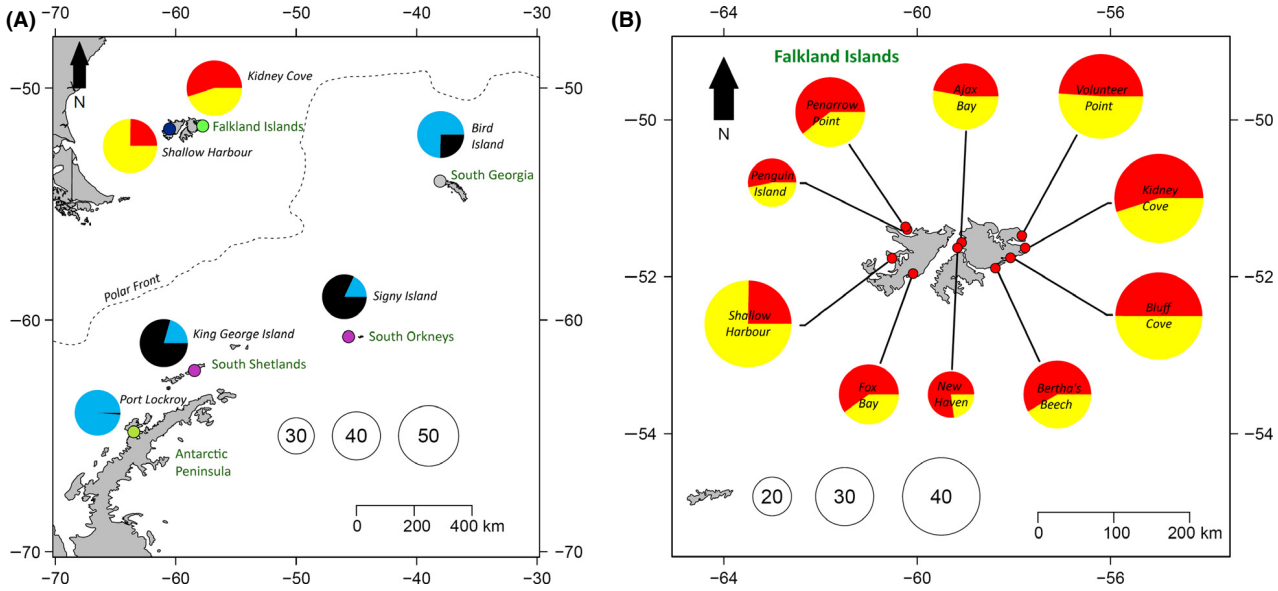


Figure 2. Population structure (pie charts) and locations of Gentoo penguin (*Pygoscelis papua*) breeding sites [different colors in (A), red circles in (B)] sampled in this study across the Scotia Arc. Pie chart colors denote different genetic clusters as identified by STRUCTURE, and the size of each pie slice shows the average probability of assignment of individuals to the cluster. The size of each pie chart is scaled according to the number of individuals sampled at each colony (clear circles at the bottom of each plot for scale). (A) The results of structural analysis for the whole of the Scotia Arc, which was conducted as two separate analyses (see Fig. 4B). Location markers are colored according to the output of GENELAND, which found five population clusters. (B) The results from STRUCTURE for the Falkland Islands, which identified two populations (see Fig. 4B).

used to cut the feather lengthwise and expose the inner pulp and blood supply. Once a dry blood vessel was found, it was scraped out of the feather casing and diced with a scalpel, and then transferred to the microcentrifuge tube with the feather tip. DNA was then extracted using the DNeasy Blood & Tissue Kit (Qiagen, Crawley, West Sussex, UK) according to the manufacturer's instructions for animal tissue, with the following user modification for the lysis step: 180 μL of lysis Buffer ATL and 30 μL of Proteinase K were added to the microcentrifuge tube, which was vortexed and then incubated for an extended period of 24–72 h at 56°C in a shaking incubator. Blood samples obtained from Bird Island, South Georgia, were extracted using the DNeasy Blood & Tissue Kit (Qiagen, Crawley, West Sussex, UK) according to the manufacturer's instructions for animal blood. The extracted DNA was suspended in 200 μL of elution Buffer AE and stored at −20°C.

Molecular data

Polymorphic microsatellite loci have been characterized for Galápagos penguins (*Spheniscus mendiculus*), Magellanic penguins (*Spheniscus magellanicus*) (Akst et al. 2002), Humboldt penguins (*Spheniscus humboldti*) (Schlosser et al. 2009), Little penguins (*Eudyptula minor*) (Billing et al. 2007), yellow-eyed penguins (*Megadyptes antipodes*) (Boessenkool et al. 2008), Macaroni penguins (*Eudyptes chrysolophus*) (Ahmed et al. 2009), and closely related Adélie penguins (*Pygoscelis adeliae*), some of which have been tested for cross-amplification in small numbers of Gentoo penguins (Roeder et al. 2001; Ahmed et al. 2009; Schlosser et al. 2009). Because of the lack of species-specific primers in Gentoo penguins, a subset of 14 Gentoo individuals from seven colonies were screened for amplification and polymorphism using PCR primers developed for 34 microsatellite loci in other species of penguins, as well as two loci developed for petrels (Brown and Jordan 2009) (see Table 1). Amplifications were conducted in 8.5-μL volumes containing 4 μL 2× Multiplex PCR Master Mix (Qiagen, Crawley, West Sussex, UK), 2 μM of each primer, and 2.5 μL of template DNA. Amplifications involved an initial cycle of 95°C for 15 min, 40 cycles of 94°C for 30 sec, 60°C for 90 sec, and 72°C for 45 sec, followed by a 10-min extension at 72°C. A volume of 1 μL of diluted PCR product (1:90) was then suspended in 9 μL of HiDi Formamide (Applied Biosystems, Foster City, CA) with 0.3 μL of GeneScan 500 LIZ size standard, and visualized on an ABI 3100 Genetic Analyzer. All but one marker amplified, although not all did so consistently. Twenty-nine markers could be scored, with only eight loci being sufficiently variable and reliable for large-scale analyses (after testing for linkage disequilibrium and null alleles): Ech030, Ech036, Ech050, Ech065, Ech071, Ech091 (Akst et al. 2002; Ahmed et al. 2009), Emm4 (Billing et al. 2007), and RM3 (Roeder et al. 2002). These eight markers were grouped into three multiplexes, as detailed in Table 1. Peaks were then scored for length using GeneMapper v4.0 (Applied Biosystems, Foster City, CA). In cases where peaks could not be easily scored, or there was any doubt to the identity of an allele, the sample underwent amplification and genotyping in singleplex reactions two to five additional times to minimize scoring errors.

For mitochondrial DNA, the hypervariable region of the mitochondrial control region (HVR1), also known as Domain I, was amplified using the primers GPPAIR3F and GPPAIR3R, as described in Clucas et al. (2014). Ten new mtDNA sequences were included plus 249 sequences from Clucas et al. (2014) with a total n of 259. Mitochondrial sequences were visualized using Geneious Basic v5.6.4 (Biomatters, http://www.geneious.com). Forward and reverse sequences were aligned and a consensus multiple sequence alignment was generated in Geneious and exported for further analysis.

Genetic diversity

Micro-checker (Van Oosterhout et al. 2004) was used to test for genotyping errors resulting from null alleles, large allele dropout, and stutter. Standard indices of genetic variability, including observed and expected heterozygosities (H_O and H_E, respectively) and number of alleles, were quantified for each colony at each locus using Arlequin v3.5.1.2 (Excoffier et al. 2005). Linkage disequilibrium was tested using likelihood ratio tests with 10,000 permutations (Slatkin and Excoffier 1996). Expectations for Hardy–Weinberg equilibrium were estimated for each locus and for all loci using exact tests with 1,000,000 steps (Guo and Thompson 1992).

For microsatellites, Arlequin was used to estimate pairwise F_{ST}'s (Weir and Cockerham 1984) and we used the SGoF+ method (Carvajal-Rodriguez and de Uña-Alvarez 2011) to correct for multiple hypothesis testing, using the modal method for π_0 estimation and a significance level of 0.05. Arlequin was used to calculate a global F_{ST} using analysis of molecular variance (AMOVA). Hierarchical F-statistics were then calculated to search for genetic structure and find the population grouping that maximized the among-group variation (F_{CT}) and minimized the variation among populations within groups (F_{SC}) (Excoffier et al. 1992). Significance of both overall and pairwise F_{ST}'s was computed using 1,000,000 permutations. The frequency of null alleles was estimated according to Brookfield (Brookfield 1996), and FreeNA (Chapuis and Estoup 2007) was used to determine

Table 1. Microsatellite primer sequences used for analysis with multiplex set information, number of alleles (A), and allelic sizes and range (bp), for $n = 510$ individually genotyped Gentoo penguins (*Pygoscelis papua*).

Locus	Genbank/EMBL accession number and source	Repeat motif	Primer sequence (5′–3′) and fluoro label	Multiplex Set	A	Allele sizes observed (bp) ($n = 510$)	Obs. allele size range (bp)
Ech030	FM878361 (Ahmed et al. 2009)	$(CTAT)_{14}$	F: [HEX]-TGACGCCGCAGGGACTTC R: GCTCAGCTCTTGCTCACAGTTTCAG	1	15	264, 266, 268, 270, 272, 274, 276, 280, 284, 288, 292, 296, 300, 304, 308	264–308
Ech036	FM878367 (Ahmed et al. 2009)	$(GT)_{18}$	F: [HEX]-GAGAGGGTTCAGAATGACATCACG R: GTCCATGGGAGCAGACCTGAG	1	3	176, 177, 178	176–178
Ech050	FM878381 (Ahmed et al. 2009)	$(AC)_{12}$	F: [HEX]-TGTCCAAGTCAGCAAAGCATCC R: CGTCTGCTGGCTGGTGAGAG	1	4	320, 322, 324, 326	320–326
Ech065	FM878396 (Ahmed et al. 2009)	$(GT)_{12}$	F: [FAM]-TGACATGTATGGGGAGGAAAGGTT R: ACACTGGGCCTGTGGGAAAA	1	3	146, 152, 154	146–154
Ech071	FM878402 (Ahmed et al. 2009)	$(CTCAT)_{14}$	F: [HEX]-CAGCCCACCGGTCTCTTACAG R: TGCAATGGTCTCTTCAGGAGATG	2	11	192, 197, 202, 207, 212, 217, 222, 227, 232, 233, 237	192–237
Ech091	FM878422 (Ahmed et al. 2009)	$(GT)_{9}$	F: [FAM]-TCCGCAGTTCACGAGGAGTC R: ACAAGCCCTCTGCCTGTCTTG	2	9	406, 410, 412, 414, 416, 418, 420, 422, 431	406–431
Emm4	DQ837732 (Billing et al. 2007)	$(CT)_{12}$	F: [FAM]-GGGAGGGCCTAACAAACTAC R: TTAGATGCCTGGTCATTTGG	2	7	237, 239, 241, 243, 245, 247, 249	237–249
RM3	AF289546 (Roeder et al. 2002)	$(CA)_{10}$	F: [FAM]-AATCAGGCTCCAAGGTCAGT R: ATGCAAGTGACACAAAGGCT	3	7	208, 211, 213, 215, 217, 219, 221	208–221

whether null alleles were biasing estimates of population differentiation.

For the mtDNA, we calculated standard molecular diversity indices and pairwise Φ_{ST}s in Arlequin. Molecular diversity measures and molecular distances were calculated with the Tamura and Nei substitution model and a gamma distribution (with $\alpha = 0.066$) for rate heterogeneity among sites, as calculated in jModelTest 0.1.1 (Guindon and Gascuel 2003; Posada 2008). Pairwise Φ_{ST}s were calculated between all colonies and significance was determined using 10,000 permutations of haplotypes between colonies.

Isolation by distance

To test for isolation by distance among microsatellite loci, the shortest geographic distance by sea was calculated using Google Earth Pro (Google, Version 7.1.5.1557), and linearized estimates of F_{ST} were tested for correlation with distance using Mantel's test (Smouse et al. 1986) in R with the vegan package (Oksanen et al. 2015). Statistical significance of correlation coefficients was estimated using 10,000 permutations.

To test for isolation by distance for mitochondrial data, the correlation between these same geographic distances and pairwise Φ_{ST}s was calculated using Mantel's test with 10,000 permutations in Arlequin.

Population structure

We explored two approaches to derive population structure from multilocus microsatellite data. First, population structure was analyzed using STRUCTURE (Pritchard et al. 2000). We compared analyses that assumed correlated and uncorrelated allele frequencies, both with and without treating sampling locations as a priori information (Pritchard et al. 2000, 2002). For admixture model conditions, α was allowed to vary. The program was run with a burn-in of 10,000 iterations, followed by 1,000,000 MCMC steps. Each value of K (number of populations) between 1 and 14 was run 10 times, and significance was calculated from the posterior probabilities (Pritchard et al. 2002; Evanno et al. 2005; Falush et al. 2007). The most likely value of K was determined using the delta K values from Structure Harvester (Earl and Vonholdt 2012).

Secondly, to visualize population assignment in a spatial context, we used the GENELAND package within R (Guillot et al. 2005a,b; Guillot 2008). This program incorporates GPS data for each individual (set for each breeding colony sampled) and multilocus genotype data to estimate the number of populations and the geographic boundaries between the inferred clusters. We set the number of populations from 1 to 14, varying the initial population (prior) from 1 to 14 for 1,000,000 MCMC iterations using the spatial model, testing both the correlated and uncorrelated allele frequency models.

In addition, in order to verify the presence of any confounding signal from subspecies differentiation, to test for hierarchical population structure, and to detect fine-scale structure in a highly sampled geographic area, all analyses were repeated for the 10 Falkland Island colonies alone, and for colonies south of the Polar Front.

Bayesian phylogeography

We estimated the ancestral locations of Gentoo penguins using a Bayesian discrete phylogeographic approach (Lemey et al. 2009) with BEAST v1.8.1 (Drummond et al. 2012). We used the mtDNA data (HVR1 region, 320 bp) for 259 penguins. To select an appropriate model of nucleotide substitution, jModelTest v2.1.6 was used (Darriba et al. 2012). We evaluated the likelihood scores for 24 substitution models, and then used the Bayesian information criterion to select the model. There were two models in the 95% confidence interval (K80 + I+G and HKY + I + G), and we used the HKY + I + G in subsequent Bayesian phylogeographic analyses. We assigned each penguin to one of five island populations: Bird Island, South Georgia ($n = 38$); Falklands ($n = 101$); King George, South Shetland Islands ($n = 41$); Port Lockroy, Antarctic Peninsula ($n = 37$); and Signy Island, South Orkney Islands ($n = 42$). We modeled island location as a discrete trait using a symmetric substitution model with the Bayesian stochastic search variable selection (BSSVS) procedure, and we reconstructed ancestral states for all ancestors. We set the clock model for the mtDNA data to a strict molecular clock. We used a coalescent tree prior with constant population size and used a normally distributed prior for the mtDNA clock rate with a mean of 0.55 and standard deviation of 0.15, based on previous calculations for the mitochondrial mutation rate in the sister species *Pygoscelis adeliae* (Millar et al. 2008). As our focus is the tree topology and the locations of ancestral populations, and not the time to the most recent common ancestor, we show a single mutation rate. However, see Clucas et al. (2014) for a greater discussion on the node ages assessed using multiple rates. The prior for locations used the approximate continuous time Markov chain rate reference prior (Ferreira and Suchard 2008). We ran the analysis for 10 million generations, sampling states every 10,000 steps. We repeated the analysis four times, checked for convergence in Tracer (Rambaut et al. 2014), and then combined the four runs using LogCombiner. We obtained a maximum clade credibility tree (MCC tree) using Tree Annotator v1.8.1.

Population assignment

Finally, we used the microsatellite data to assign individuals to populations to determine whether there were any recent migrants within the populations that we had sampled. Assignment tests were run in Genodive v2.0b27 (Meirmans & Van Tienderen 2004). Allele frequencies that were found to be equal to zero were replaced with 0.005; 50,000 permutations of the Monte Carlo test were performed to determine the null distribution of likelihood values and the significance threshold was chosen to be 0.002 (Paetkau et al. 2004). The test statistic used was the Home Likelihood (L_h), as we had not sampled all possible source locations for migrants.

Results

Genetic diversity

A total of 510 individuals were genotyped across the 14 colonies using eight microsatellite loci. These loci had between three and 15 alleles each (Table 1). Only individuals where 100% of loci could be scored were included in the analysis. None of the markers were found to be under linkage disequilibrium or to consistently deviate from Hardy–Weinberg equilibrium across colonies. Port Lockroy exhibited a slightly lower average gene diversity compared with the other colonies (H_O, Table 2, Fig. 3). Locus-by-locus diversity measures for each sampling site are shown in Appendix S1.

Pairwise F_{ST} values and associated *P*-values are shown in Table 3. With the exception of Shallow Harbour, colonies within the Falkland Islands exhibit little genetic differentiation from one another. Although several pairs of Falkland Island colonies have significant pairwise F_{ST} values, these are all <0.05 and are several orders of magnitude lower than the differentiation seen with colonies outside of the Falklands. This significance may be attributable to natal behavior, but it should be noted that these are also the colonies with the lowest sample sizes, and which coincidentally could have been affected by demographic events such as a 2006 epidemic of avian pox on the Falklands (Munro 2006). Notably, Shallow Harbour shows consistent, moderate genetic differentiation from all colonies within the Falkland Islands (F_{ST} range = 0.036–0.141). All colonies within the Falkland

Table 2. Genetic diversity of Gentoo penguins (*Pygoscelis papua*) at 14 breeding sites across the Scotia Arc. See Appendix S1 for diversity indices at each locus.

Colony	*n*	H_E	SD	H_O	SD
Volunteer Point (FI)	35	0.50725	0.23398	0.39286	0.20015
Kidney Cove (FI)	46	0.45813	0.24199	0.41304	0.26472
Bluff Cove (FI)	45	0.49482	0.22322	0.48611	0.25405
Bertha's Beach (FI)	35	0.56190	0.16796	0.47347	0.17211
Ajax Bay (FI)	34	0.49166	0.23030	0.40074	0.24199
New Haven (FI)	24	0.50722	0.23709	0.45238	0.21168
Fox Bay (FI)	31	0.44738	0.24771	0.40726	0.22544
Saunders Penguin Island (FI)	25	0.49983	0.20295	0.36000	0.29029
Saunders Penarrow Point (FI)	36	0.50419	0.19574	0.42063	0.23484
Shallow Harbour (FI)	45	0.53098	0.15775	0.46667	0.23518
Bird Island (S. Georgia)	39	0.41146	0.26591	0.46795	0.29989
Signy Island (S. Orkney Is.)	37	0.38898	0.27999	0.32432	0.25559
King George Island (S. Shetland Is.)	40	0.43856	0.25379	0.46071	0.26687
Port Lockroy (Western Antarctic Peninsula)	38	0.28000	0.26133	0.28195	0.27504

Mean observed (H_O) and expected (H_E) heterozygosity is shown, along with standard deviation (SD), over eight microsatellite loci for all individuals (*n*). FI, Falkland Islands.

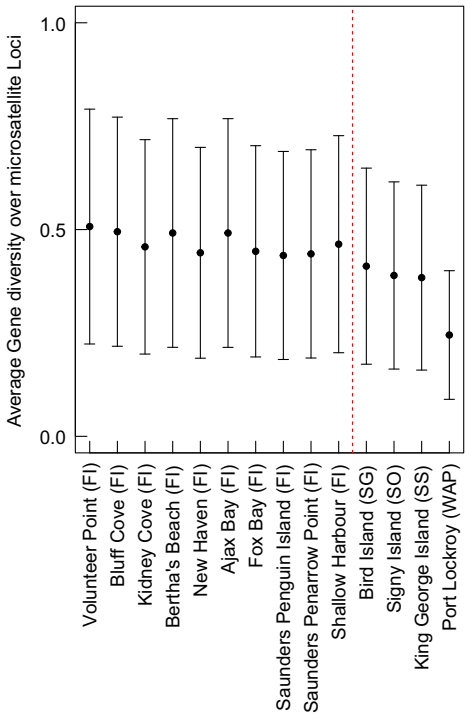


Figure 3. Genetic diversity of Gentoo penguins (*Pygoscelis papua*) at 14 breeding sites across the Scotia Arc. Mean observed (H_O) heterozygosity is shown, along with standard deviation (SD), over eight microsatellite loci for all individuals. FI = Falkland Islands, SG = South Georgia, SO = South Orkney Islands, SS = South Shetland Islands, and WAP = Western Antarctic Peninsula. Colonies north and south of the Polar Front are shown divided by the dotted line. See also Table 2.

Table 3. Pairwise F_{ST} values (and associated *P*-values) among Gentoo penguin breeding sites across the Scotia Arc based on microsatellite data (below diagonal), with comparisons that are significant after correction for multiple tests (using SGoF+) shown in bold. Pairwise geographic distances (most direct distance by sea, in km) used for Mantel's test, derived from Google Earth Pro, are shown above the diagonal.

	Volunteer Point (FI)	Kidney Cove (FI)	Bluff Cove (FI)	Bertha's Beach (FI)	Ajax Bay (FI)	New Haven (FI)	Fox Bay (FI)	Saunders Penguin Is. (FI)	Saunders Penarrow Point (FI)	Shallow Harbour (FI)	Bird Island (SG)	Signy Is. (SO)	King George Is. (SS)	Port Lockroy (WAP)
Volunteer Point (FI)	*	22.88	60.03	79.63	133.62	149.61	214.95	188.95	199.22	286.40	1362.08	1284.8	1207.09	1549.12
Kidney Cove (FI)	0.00864 (0.12613)	*	40.98	61.10	154.22	169.32	217.00	209.32	219.59	306.31	1349.59	1267.10	1188.79	1531.27
Bluff Cove (FI)	0.00000 (0.55450)	0.01114 (0.02762)	*	29.81	188.78	227.52	191.12	243.50	253.77	295.05	1364.64	1264.49	1171.01	1507.72
Bertha's Beach (FI)	0.00170 (0.49837)	0.01696 **(0.01564)**	0.00406 (0.24740)	*	212.99	204.02	168.91	266.56	276.83	277.87	1379.19	1263.76	1157.17	1490.56
Ajax Bay (FI)	0.00485 (0.33700)	0.02013 **(0.00693)**	0.00861 (0.09920)	0.00051 (0.55589)	*	36.15	107.33	110.52	120.71	208.92	1489.86	1402.85	1234.93	1543.97
New Haven (FI)	0.02874 **(0.00950)**	0.01841 (0.02426)	0.03672 **(0.00010)**	0.02398 (0.01634)	0.04772 **(0.00020)**	*	80.73	126.40	136.67	208.95	1507.14	1372.96	1204.72	1512.64
Fox Bay (FI)	0.01438 (0.06376)	0.00873 (0.111217)	0.00742 (0.10969)	0.00588 (0.25205)	0.00638 (0.25661)	0.02265 (0.01970)	*	197.00	211.26	194.13	1500.29	1325.39	1158.84	1463.16
Saunders Penguin Is. (FI)	0.01006 (0.21958)	0.00805 (0.19364)	0.01053 (0.10563)	0.02851 **(0.00921)**	0.02646 (0.01762)	0.04244 **(0.00287)**	0.01611 (0.08346)	*	10.27	114.81	1543.66	1490.08	1287.54	1577.65
Saunders Penarrow Point (FI)	0.01726 (0.03148)	0.01288 (0.4227)	0.01131 (0.04406)	0.01409 (0.6702)	0.01879 (0.02109)	0.02777 **(0.00842)**	0.00441 (0.31621)	0.02028 (0.04138)	*	104.96	1553.94	1498.44	1295.11	1584.72
Shallow Harbour (FI)	0.03601 **(0.00010)**	0.06855 **(0.00000)**	0.04329 **(0.00000)**	0.06189 **(0.00000)**	0.04396 **(0.00000)**	0.14086 **(0.00000)**	0.07488 **(0.00000)**	0.05963 **(0.00000)**	0.09126 **(0.00000)**	*	1640.38	1411.02	1207.17	1496.91
Bird Island (SG)	0.13304 **(0.00000)**	0.15886 **(0.00000)**	0.11542 **(0.00000)**	0.10636 **(0.00000**)	0.10258 **(0.00000**)	0.14969 **(0.00000)**	0.13988 **(0.00000)**	0.19602 **(0.00000)**	0.12443 **(0.00000)**	0.20027 **(0.00000)**	*	920.20	1476.18	1892.16
Signy Is. (SO)	0.20248 **(0.00000)**	0.22583 **(0.00000)**	0.20145 **(0.00000)**	0.19315 **(0.00000)**	0.17446 **(0.00000)**	0.22409 **(0.00000)**	0.22898 **(0.00000)**	0.28239 **(0.00000)**	0.20945 **(0.00000)**	0.26736 **(0.00000)**	0.04503 **(0.00000)**	*	686.69	1051.78
King George Is. (SS)	0.19482 **(0.00000)**	0.22237 **(0.00000)**	0.19489 **(0.00000)**	0.18340 **(0.00000)**	0.16216 **(0.00000)**	0.23258 **(0.00000)**	0.22415 **(0.00000)**	0.27731 **(0.00000)**	0.20088 **(0.00000)**	0.25087 **(0.00000)**	0.04903 **(0.00000)**	0.00793 (0.08178)	*	434.24
Port Lockroy (WAP)	0.21842 **(0.00000)**	0.26091 **(0.00000)**	0.19490 **(0.00000)**	0.17618 **(0.00000)**	0.20139 **(0.00000)**	0.27919 **(0.00000)**	0.23031 **(0.00000)**	0.31456 **(0.00000)**	0.22934 **(0.00000)**	0.25102 **(0.00000)**	0.15845 **(0.00000)**	0.27013 **(0.00000)**	0.26951 **(0.00000)**	*

FI, Falkland Islands; SG, South Georgia; SO, South Orkney Islands; SS, South Shetland Islands; WAP, Western Antarctic Peninsula.
*Indicates colonies crossed pairwise with other colonies, which yields no distances

Islands are strongly and consistently differentiated from those south of the Polar Front (F_{ST} range = 0.103–0.315), with strong differentiation from Signy and King George Island (F_{ST} range = 0.162–0.282) and from the most distant colony, Port Lockroy (F_{ST} range = 0.176–0.315). Of those colonies south of the Polar Front, all are strongly differentiated from one another apart from King George and Signy Island, which are not significantly differentiated from one another (F_{ST} = 0.008). Relative to the other Southern Gentoos, Bird Island (South Georgia) is genetically closer to, but still retains moderate to strong differentiation from, the Falkland Island colonies (F_{ST} range = 0.106–0.200).

AMOVA indicated the presence of hierarchical population structure across the Scotia Arc (global F_{ST} = 0.117, $P < 0.001$). The proportion of variation resulting from differences among groups was maximized at 16.38% ($P < 0.001$) when populations were placed into four groups: (1) Falkland Island colonies; (2) Bird Island, South Georgia; (3) Signy Island, South Orkneys and King George Island, South Shetlands; and (4) Port Lockroy, Antarctic Peninsula. Explained among-group variation remained high at 15.66% ($P < 0.001$) when split into five groups, with Signy Island (South Orkneys) and King George Island (South Shetlands) split into separate groups, and at 15.47% ($P < 0.001$) when Shallow Harbour (Falkland Islands) was isolated from the remaining Falkland Island colonies, as part of a five-group hierarchy, as shown in Table 4.

Within the Falkland Islands, very weak differentiation was present (global F_{ST} = 0.027, $P < 0.001$), with almost all variation explained by separating Shallow Harbour from the rest of the colonies (F_{CT} = 5.26%), although this was not significant (P = 0.100).

Haplotypic and nucleotide diversity measures for the mitochondrial DNA data are depicted in Table 5.

Table 4. Analysis of Molecular Variance (AMOVA) of microsatellite data from Gentoo penguin populations of the Scotia Arc, when grouped by varying assignment criteria. The bold values indicate the grouping that maximizes among-group variation.

Grouping criteria	Within-population % variation, F_{ST} (*P*-value)	Among-population % variation, F_{SC} (*P*-value)	Among-group % variation, F_{CT} (*P*-value)
14 populations of the Scotia Arc			
1 group	88.28, 0.11721 (<0.001)		
2 groups by subspecies (Falkland Islands, Rest of Scotia Arc)	81.70, 0.18301 (<0.001)	4.62, 0.05355 (<0.001)	13.68, 0.13678 (0.001)
2 groups (Falkland Islands and South Georgia, Rest of Scotia Arc)	81.91, 0.18092 (<0.001)	6.51, 0.07359 (<0.001)	11.59, 0.11585 (0.005)
2 groups (Falkland Islands and Antarctic Peninsula, South Georgia and South Orkneys and South Shetlands)	80.43, 0.19571 (<0.001)	5.24, 0.06120 (<0.001)	14.33, 0.14328 (0.002)
3 groups (Falkland Islands, South Georgia, Rest of Scotia Arc)	82.77, 0.17225 (<0.001)	4.94, 0.05636 (<0.001)	12.28, 0.12282 (0.001)
3 groups (Falkland Islands, South Georgia and Antarctic Peninsula, South Shetlands and South Orkneys)	81.61, 0.18391 (<0.001)	3.24, 0.03822 (<0.001)	15.15, 0.15147 (<0.001)
4 groups (Falklands and Antarctic Peninsula excluding Shallow Harbour, Shallow Harbour, South Georgia, South Shetlands and South Orkneys)	83.51, 0.16492 (<0.001)	4.88, 0.05518 (<0.001)	11.61, 0.11615 (0.002)
4 groups (Falkland Islands, South Georgia, South Shetlands and South Orkneys, Antarctic Peninsula)	**81.29, 0.382187 (<0.001)**	**2.33, 0.02786 (<0.001)**	**16.38, 0.16382 (<0.001)**
5 groups (Falkland Islands, South Georgia, South Shetlands, South Orkneys, Antarctic Peninsula)	81.76, 0.18236 (<0.001)	2.57, 0.03048 (<0.001)	15.66, 0.15665 (0.001)
5 groups (Shallow Harbour, Rest of Falklands, South Georgia, South Shetlands and South Orkneys, Antarctic Peninsula)	83.26, 0.16737 (<0.001)	1.26, 0.01494 (<0.001)	15.47, 0.15474 (<0.001)
10 populations of the Falkland Islands			
1 group	97.26, 0.02739 (<0.001)		
2 groups (Shallow Harbour, Rest of Falkland Islands)	93.41, 0.06594 (<0.001)	1.33, 0.01407 (<0.001)	5.26, 0.05261 (0.100)
2 groups (East Falkland, West Falkland including Saunders Island)	97.21, 0.02785 (<0.001)	2.68, 0.02680 (<0.001)	0.11, 0.00109 (0.403)
3 groups by body of water (East-facing Falkland, Falkland Sound, West-facing Falkland)	97.26, 0.02739 (<0.001)	2.74, 0.02738 (<0.001)	0.00, 0.00001 (0.491)
3 groups by island (East Falkland, West Falkland, Saunders Island)	97.03, 0.02969 (<0.001)	2.36, 0.02376 (<0.001)	0.61, 0.00607 (0.246)

Table 5. mtDNA diversity measures for each island grouping.

Grouping	*n*	N_H	N_P	*H* (SD)	π (SD)
Gentoo penguin (all)	259	115	58	0.9800 (0.0031)	0.02404 (0.01245)
P. p. papua (northern subspecies/all Falkland Islands)	101	40	25	0.9228 (0.0157)	0.00906 (0.00533)
P. p. ellsworthii (southern subspecies)	158	75	48	0.9776 (0.0047)	0.01353 (0.00746)
Bird Island (South Georgia)	38	19	18	0.9346 (0.0216)	0.00877 (0.00527)
Signy Island (S. Orkney Is.)	42	19	23	0.9338 (0.0203)	0.01041 (0.00607)
King George Island (S. Shetland Is.)	41	23	26	0.9598 (0.0150)	0.01699 (0.00929)
Port Lockroy (Western Antarctic Peninsula)	37	18	19	0.9099 (0.0300)	0.01166 (0.00670)

n, number of individuals sequenced; N_H, number of haplotypes; N_P, number of polymorphic sites; *H*, haplotype diversity; π, nucleotide diversity; SD, standard deviation.

Isolation by distance

Mantel's test detected significant isolation by distance in both the microsatellite ($r = 0.841$, $P < 0.001$) and mitochondrial ($r = 0.679$, $P < 0.001$) regions, as might be expected over this geographic range.

Table 6. Summary of inferred number of populations (*K*) resulting from STRUCTURE analysis, changing model assumptions for presence of admixture, use of the LOCPRIOR setting for a priori location assignment, correlated and independent allele frequencies.

Admixture or No admixture model	LOCPRIOR	Correlated or independent allele frequency model	No. of populations (*K*) inferred by Evanno method
All samples			
Admixture	Yes	Correlated	2
Admixture	Yes	Independent	2
Admixture	No	Correlated	2
Admixture	No	Independent	2
No Admixture	Yes	Correlated	2
No Admixture	Yes	Independent	2
No Admixture	No	Correlated	4
No Admixture	No	Independent	2
Falklands Only			
Admixture	Yes	Correlated	2
Admixture	Yes	Independent	2
Admixture	No	Correlated	2
Admixture	No	Independent	3
No Admixture	Yes	Correlated	2
No Admixture	Yes	Independent	2
No Admixture	No	Correlated	2
No Admixture	No	Independent	3
South of Polar Front			
Admixture	Yes	Correlated	2
Admixture	Yes	Independent	2
Admixture	No	Correlated	3
Admixture	No	Independent	3
No Admixture	Yes	Correlated	2
No Admixture	Yes	Independent	2
No Admixture	No	Correlated	2
No Admixture	No	Independent	2

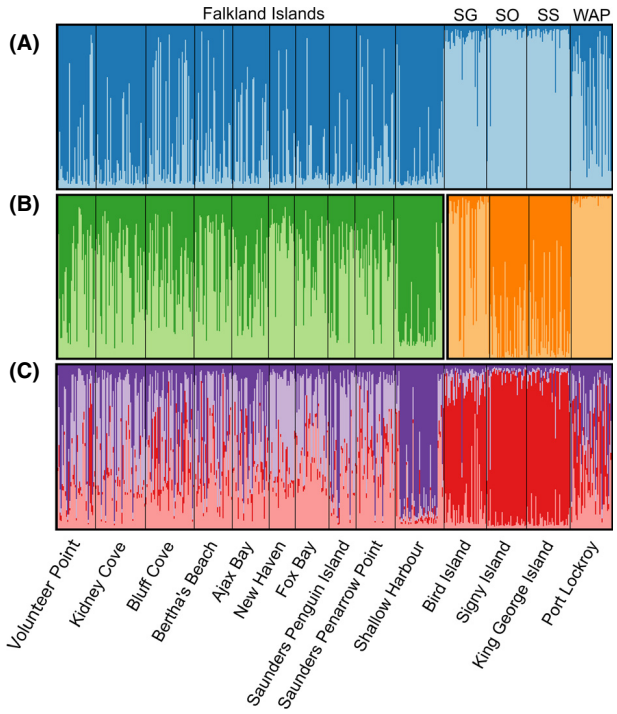


Figure 4. Plots of assignment probabilities from STRUCTURE showing the posterior probability of assigning each individual to each of the inferred clusters. Each individual is represented by a vertical bar, and the colors refer to the different clusters. All plots were generated from 10 runs using the admixture model with correlated allele frequencies. No location information was supplied for these runs. (A) $K = 2$ was the most likely number of clusters when all colonies were included, which clearly delineates the difference between the northern and southern subspecies of Gentoo penguin. (B) When we analyzed the northern and southern subspecies separately, $K = 2$ was most likely for each subset. (C) For illustrative purposes, we present the results from all colonies when $K = 4$, which clearly shows the differentiation of Shallow Harbour from the other Falkland Island colonies, and the difference between Northern and Southern Gentoos. SG = South Georgia, SO = South Orkney Islands, SS = South Shetland Islands, and WAP = Western Antarctic Peninsula.

Population structure

The summary of results for the number of populations (K) inferred from STRUCTURE using Evanno's method (2005) is depicted in Table 6. The modal population when all samples were included was $K = 2$ (Fig. 4), except under the No Admixture model for correlated allele frequencies, with no a priori location data, when $K = 4$ was favored. The two-population split most strongly coincided with the Northern and Southern Gentoo subspecies, with the 10 Falkland Island colonies as one population, and the four southerly colonies as another. STRUCTURE tends to detect the uppermost level of hierarchical structure in a population (Evanno et al. 2005), which may explain why $K = 2$ was frequently reported, despite our suspicions of additional underlying structure. To delve further into the next level of hierarchical structure, we performed analyses separately on each of these two groups, which further elucidated two or three populations within the Falkland Island samples, and two or three populations within the island groups south of the Polar Front (for a total $K = 4$–6), depending on model assumptions. Taking into account the AMOVA results, the plots for individual assignments when $K = 4$ are displayed in Figure 4C.

GENELAND's calculation of the number of populations was strongly influenced by whether or not the correlated or uncorrelated model of allele frequencies was employed. The improved MCMC algorithm within Guillot (2008), which revisits the correlated model, seems to best explain the biological reality of our sample populations, where weak differentiation exists across most individuals, but strong differentiation is also present. The summary of GENELAND results is presented in Table 7. The resulting clustering pattern supports individual assignment to five distinct populations: (1) Shallow Harbour, Falkland Islands; (2) remaining Falkland Island colonies; (3) Bird Island, South Georgia; (4) King George Island, South Shetlands and Signy Island, South Orkneys; and (5) Port Lockroy, Antarctic Peninsula.

Table 7. Summary of inferred number of populations (K) resulting from GENELAND analysis of microsatellite data, based on both the correlated and uncorrelated models for each given prior.

Prior K	Inferred K for correlated model	Inferred K for uncorrelated model
1	5	2
2	5	2
3	5	2
4	5	2
5	5	2
6	5	2
7	5	3
8	5	2
9	5	3
10	5	2
11	5	3
12	5	2
13	5	3
14	5	3

Bayesian phylogeography and population assignment

The maximum clade credibility tree resulting from discrete Bayesian phylogeographic analysis is depicted in Figure 5. This mitochondrial DNA tree corroborates the strong differentiation between Northern (*P. papua papua*) and Southern Gentoos (*P. papua ellsworthii*), with all Falkland individuals grouping together, as a separate clade from individuals south of the Polar Front. The Southern Gentoos seem to have radiated from a population that was in the vicinity of King George Island in the South Shetland Islands ($P = 0.52$), or possibly Signy Island in the South Orkney Islands ($P = 0.38$). From here, migrants appear to have dispersed north to South Georgia (Bird Island), with another portion of migrants moving southward from King George Island (South Shetlands) to the Antarctic Peninsula (Port Lockroy). Both the Bird Island and Port Lockroy individuals cluster to form well-defined clades on the MCC tree, with the exception of three individuals from Port Lockroy. This suggests that populations in both South Georgia and the Antarctic Peninsula were established by a single or a small number of migration events and that ongoing gene flow has been low. As this tree is based on a single mitochondrial marker, we do not attempt to date the nodes on the tree, as more genetic loci would be necessary to draw reliable conclusions, but 95% highest posterior densities (HPDs) for node heights can be seen in Appendix S2.

Results from the population assignment tests also suggest that ongoing gene flow between those colonies south of the Polar Front has been low. None of the individuals were identified as migrants at Bird Island, King George Island, Signy Island, or Port Lockroy. However within the Falkland Islands, six individuals were identified as recent migrants between the colonies of the Falkland Islands.

Discussion

This study has revealed that both intrinsic and extrinsic factors are important in determining gene flow in Gentoo penguins. The Polar Front, an extrinsic barrier, appears to be the most important determinant of genetic differentiation across the Scotia Arc, rather than a tendency toward natal philopatry and year-round residency near colonies

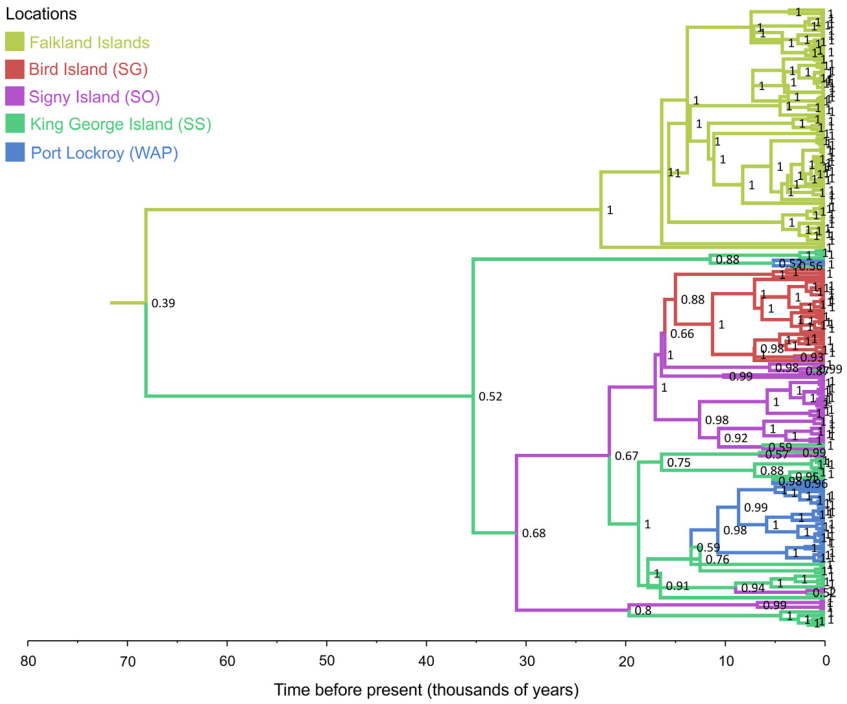


Figure 5. Maximum clade credibility tree derived from mtDNA showing the origin and differentiation of *Pygoscelis papua* lineages north (Falkland Islands, light green above) and south of the Polar Front (all other colors and locations). Node colors represent the most likely location of each ancestral node, whilst node labels show the level of support for each location. SG = South Georgia, SS = South Shetland Islands, WAP = Western Antarctic Peninsula, and SO = South Orkney Islands.

(intrinsic factors). Using both nuclear and mitochondrial markers, we have shown significant population structure in Gentoo penguins at the regional scale and explored the levels of admixture present at a finer scale within the species. The high degree of population differentiation either side of the Polar Front is consistent with similar evidence from biometric and acoustic analyses (de Dinechin et al. 2012) and mitochondrial DNA (Clucas et al. 2014), supporting the existence of two subspecies in the Scotia Arc region. There are also substantial differences in the timing of breeding between the subspecies and indeed between colonies within subspecies. Gentoo penguins have the greatest annual variation in phenology of Pygoscelid penguins. Egg laying can start as early as October in the Falkland Islands and Argentina, tends to begin in early November in the South Orkneys and South Shetlands, and extends into late November and December at Port Lockroy on the Antarctic Peninsula (Black 2015). The oceanographic barrier posed by the Polar Front, along with differences in timing of breeding (Black 2015; Bost and Jouventin 1991; Trivelpiece et al. 1987), has probably prevented noticeable admixture between Gentoo subspecies since their estimated divergence during the last glacial period (Clucas et al. 2014). The very weak genetic differentiation of populations within an archipelago, but significant population differentiation between all archipelagos, indicates that internal recruitment and survival processes probably determine population dynamics at the archipelago level, whilst a lack of long-distance dispersal helps to maintain genetic differentiation among colonies in this philopatric seabird.

In addition, discrete Bayesian phylogeographic methods have allowed us to use mitochondrial DNA to investigate colonization patterns across island groups south of the Polar Convergence. Despite being a sex-biased marker, the mitochondrial MCC tree shows a signal of historical radiation of Southern Gentoos from a population in the vicinity of the South Shetland or South Orkney Islands. This agrees with results from our previous work that suggested Southern Gentoos had expanded postglacially to colonize new habitat as it became available (Clucas et al. 2014). Migrants appear to have dispersed north to South Georgia, and south to the Antarctic Peninsula from this center of origin, during a limited number of migration events. The significant genetic differentiation between these regions in both the microsatellite and mitochondrial markers and the population assignment tests performed using the microsatellite data suggests that there is little ongoing gene flow to the northern and southern extremes of the Southern Gentoo's range, highlighting the role of recruitment in governing population dynamics.

It is not possible to determine whether the population in the region of the South Orkney and South Shetland Islands was a refugial population during the last glaciation or whether it was colonized soon after the Last Glacial Maximum. More molecular markers, including nuclear genes, and precise calibrations would be needed to accurately date the time to the most recent common

ancestor, and hence the likely date of population splitting. Furthermore, it is also not possible to assess the likelihood of habitat being available for a glacial refuge. Bathymetry around the South Orkney and South Shetland Islands suggests that they were extensively glaciated during the last glacial period (Sugden and Clapperton 1977), but whether the coastlines were fully ice-bound is difficult to determine, as the ancient coastlines are now submerged because of rising postglacial sea levels. The extent of historical summer sea ice in the Scotia Sea region is not well understood (Gersonde et al. 2005, Fraser et al. 2009) and so the availability of habitat for a glacial refuge is currently unclear.

Gentoo penguins are spatially segregated during summer because they are tied to their breeding sites; additional evidence suggests that they overwinter close to their summer breeding grounds and are also spatially segregated in winter (Clausen and Pütz 2003; Tanton et al. 2004; Lescroel and Bost 2005; Ghys et al. 2008; Lescroel et al. 2009). Our finding of genetic differentiation is consistent with reviews of population structure in seabirds that report that resource partitioning between populations of the same species needs to occur year-round if it is to lead to population differentiation (Friesen et al. 2007a,b).

Some of the highest levels of population differentiation that we observed were between the Falkland Island colonies and those south of the Polar Front. This is unsurprising, given that morphological and genetic differentiation has previously been used to classify these populations into two subspecies (de Dinechin et al. 2012; Stonehouse 1970). It is notable that many genetic studies that have analyzed populations across the South Atlantic and in the Southern Ocean have detected significant genetic differentiation of populations lying either side of the Polar Front (Allcock and Strugnell 2012; Rogers 2012; Strugnell et al. 2012), but none so far in avian populations, although evidence has been found for differentiation between sub-Antarctic and mainland New Zealand populations of yellow-eyed penguins (Boessenkool et al. 2009). The Polar Front acts as a significant barrier to dispersal of taxa from the Antarctic to areas further north and vice versa. The reasons for this will vary by taxa, but probably reflect the marked gradient in physical conditions (mainly temperature and salinity) that extend to the seafloor, creating different biogeographic realms either side of the frontal region (Convey et al. 2012). The Polar Front has remained in its position between the Falkland Islands and South Georgia throughout the glacial history of the Antarctic (Sugden and Clapperton 1977). Habitat preferences and philopatry probably explain how this acts as a barrier to the dispersal of Gentoo penguins across the Polar Front. Ocean temperatures and prey availability differ greatly between the Falkland Islands and South Georgia (Ratcliffe and Trathan 2011), and so local adaptation could be maintaining the separation of the two subspecies. Historical factors, such as the fragmentation of populations caused by the advance and retreat of ice sheets during past glaciations and changes in the location and strength of the Antarctic Circumpolar Current, will also have played a role in creating population differentiation across many taxa (Barnes et al. 2006; Rogers et al. 2007; Strugnell et al. 2012; Chown et al. 2015). Ocean barriers, such as the Subtropical Convergence (also known as the Sub-Antarctic Front), have obstructed gene flow in other species of penguins, including the Northern and Southern Rockhopper (*Eudyptes moseleyi* and *E. chrysocome*, respectively) (de Dinechin et al. 2009). Although currently classified as subspecies, the strong genetic differences in both nuclear and mitochondrial DNA of Northern and Southern Gentoo penguins may be indicative of incipient speciation similar to that of Rockhoppers. In a recent review, Friesen (2015) found that differences in ocean regimes, like that above and below the Polar Front, were amongst the most important factors in restricting gene for many seabirds. This disruption of gene flow has been significant enough to lead to speciation, resulting in sister-species which occur in adjacent ocean regimes (see Friesen 2015 for a full review). Further investigation of temporal, behavioral, and spatial barriers to breeding, as well as measures of adaptation to local environments, would be needed to delve into further taxonomic elucidation of the two Gentoo groups.

While there is clear and strong differentiation between populations above and below the Polar Front, some of the most ecologically relevant differentiation exists within the Southern Ocean. Bird Island (South Georgia) emerges as a distinct population, being most closely related to the population that comprises King George Island (South Shetland Islands) and Signy Island (South Orkney Islands). Port Lockroy on the Western Antarctic Peninsula also emerges as a distinct population, again being most closely related to King George and Signy Island. It also has lower genetic diversity than any of the other populations, which is expected because of its recent establishment (*c.* 1985) (Trathan et al. 2008). Records from Charcot's expedition to the Antarctic Peninsula in 1909 suggest that there have been occasional breeders at or near Port Lockroy for at least 80 years prior to colony establishment (Charcot and Walsh 1911; Gain 1913), with established colonies observed within 35 km (Charcot and Walsh 1911). However, the area around Port Lockroy was observed to empty of breeding Gentoo penguins immediately prior to 1985 (Trathan et al. 2008).

The establishment of new colonies at the southern end of the Gentoo penguin's breeding distribution in concert with recent climate change has been interpreted as the

result of local dispersal from range-edge populations to newly suitable breeding sites just beyond (Lynch et al. 2008, 2012). Such dispersal is likely to be rare. While rapid population growth at some new Gentoo penguin colonies suggests an extended period of continuous immigration (Lynch *Unpublished data*), the maintenance of a founder population at Port Lockroy is suggestive of a single immigration event involving immigrants from outside the immediate vicinity. To have a strong founder signal, a small group of founders from the same population must have established the colony and then internal recruitment, rather than continued immigration, must have been the main driver of population growth. Populations residing along the northern portions of the Antarctic Peninsula, between Port Lockroy and King George Island, were not sampled in this study, and therefore we cannot attest to the strength of this founder effect or a potential ghost gradient that could exist in this unsampled spatial arena. However, the clustering of the majority of the Port Lockroy individuals into a clade on the mitochondrial DNA tree and the lack of detected migrants can rule out regular immigration into the Port Lockroy colony from areas outside of the Antarctic Peninsula. Future fine-scale sampling of colonies between Port Lockroy and King George Island, as well as at additional newly established populations south of Port Lockroy, may provide additional information on the frequency and history of dispersal at smaller spatial scales.

The large population north of the Polar Front in the Falkland Islands also merits discussion. Indices of genetic diversity can assist in the elucidation of a population's recent demographic history. Overall, we see that penguins within the Falkland archipelago are interbreeding at sufficient levels to maintain similar allele frequencies and levels of overall genetic diversity across colonies. Effective population and census population size are a key component to this interbreeding. The population size of Gentoo penguins is known to have large temporal variability and interannual variation across the species range, partially accounted for by breeding abstention and deferral or low breeding success during times of low food availability (Croxall et al. 1988; Williams and Croxall 1991). This has meant that the population of Gentoo penguins in the Falkland Islands has fluctuated widely. In 1995, the estimate was 64,426 breeding pairs, followed by an increase in 2000–113,571 (79% increase). However, in November 2002, a harmful algae bloom and associated paralytic shellfish poisoning affected certain Western Falkland penguin colonies, causing mass mortality (Huin 2003). Gentoo penguin counts declined by 2005–65,857 (42% decline) following this event. Between 2005 and 2010, the population nearly doubled to 132,321 $\pm$ 2288 breeding pairs (95% increase) in the most recent comprehensive census in 2010 (Pistorius et al. 2010; Baylis et al. 2013).

Shallow Harbour, a small, west-facing colony located on West Falkland, presents some of the most interesting signals of fine-scale differentiation. Analytic techniques differed in their ability to find significance in the patterns of allele frequencies in this particular colony relative to the rest of the Falkland Islands. AMOVA did not show increased levels of among-group variation when Shallow Harbour was separated from the remaining Falkland Island colonies (5 groups vs. 4 groups), although pairwise F_{ST} values showed significant differentiation between them. STRUCTURE visualizations show distinctions between Shallow Harbour and the other colonies. However, GENELAND analyses grouped Shallow Harbour as a separate population under the improved correlated spatial model. This could indicate that the colony underwent a recent demographic change, linked with the mass-mortality event, which led to a deviation from Hardy–Weinberg equilibrium at four of the eight microsatellite loci assessed. Mitochondrial analysis was only performed on a very small number of individuals from this colony (primarily to discard duplicate individuals), and therefore a full assessment of mitochondrial diversity within Shallow Harbour is not possible with our data.

Conclusions

The population genetic structure of Gentoo penguins in the Scotia Arc coincides with the oceanographic barrier presented by the Polar Front, with additional population genetic structure across the Antarctic Peninsula and on the Falkland Islands. The Polar Front appears to act as an extrinsic barrier to gene flow, even in this highly mobile seabird. We also detected a genetic signal of radiation among Southern Gentoos from King George Island, which has led to a southward founder effect at Port Lockroy. Long-distance dispersal and colonization events appear to be rare in this species. These patterns indicate that recruitment and survival strongly influence population dynamics and that intrinsic factors such as philopatry and a tendency to remain near the breeding colony year-round have resulted in population differentiation around the Scotia Arc. Furthermore, our findings highlight how understanding patterns of genetic diversity can help identify the demographic mechanisms influencing recent population trends in Southern Ocean predators in a time of rapid environmental change.

Acknowledgments

Sampling was conducted under permits from the Falkland Islands Environmental Planning Department, The

Government of South Georgia and the South Sandwich Islands, the UK Foreign and Commonwealth Office, and the US National Science Foundation. This study was funded by the Falkland Islands Government Environmental Planning Department, the Charities Advisory Trust, the John Ray Trust, Sigma Xi, and the Rhodes Trust. Samples from Bird Island were provided by R. Phillips, and samples from Port Lockroy were collected with the assistance of A. Malaos and Quark Expeditions Ltd. W. Trivelpiece and the U.S. Antarctic Marine Living Resource Program provided logistical support at King George Island. We thank British International Helicopters Ltd in the Falkland Islands, the US Air Force, and Royal Air Force for their support of this project. We are grateful for landowner permission and support in the Falkland Islands, and we thank A. Tierney, K. and H. Kilmartin, K. Marsh, G., D. Dixon, A. Black, D. and S. Pole-Evans, P. Brickle, I. Strange, N. Rendell, S. Poncet, M. Collins, and D. Briggs for assistance with fieldwork and sample collection. D. Gottelli, C. Ariani, and I. Helms helped with genotyping and analysis.

Conflict of Interest

None declared.

Data Accessibility

New mitochondrial DNA sequences will be deposited on GenBank prior to publication. Microsatellite genotyping data, along with GPS coordinates for sample sites, will be deposited on Dryad. The ten new mitochondrial sequences are available from Genbank with accession numbers KU527128 – KU527137. The sequences taken from our previous study can be accessed via Genbank accession numbers KJ646314 – KJ646562.

References

Ahmed, S., T. Hart, D. A. Dawson, G. J. Horsburgh, P. N. Trathan, and A. D. Rogers. 2009. Isolation and characterization of macaroni penguin (*Eudyptes chrysolophus*) microsatellite loci and their utility in other penguin species (Spheniscidae, AVES). Mol. Ecol. Resour. 9:1530–1535.

Ainley, D., J. Russell, S. Jenouvrier, et al. 2010. Antarctic penguin response to habitat change as Earth's troposphere reaches 2 degrees C above preindustrial levels. Ecol. Monogr. 80:49–66.

Akçakaya, H. R., G. Mills, and C. P. Doncaster. 2007. The role of metapopulations in conservation. Pp. 64–84 *in* D. W. Macdonald and K. Service, eds. Key topics in conservation. Blackwell, Oxford, UK.

Akst, E. P., P. D. Boersma, and R. C. Fleischer. 2002. A comparison of genetic diversity between the Galapagos Penguin and the Magellanic Penguin. Conserv. Genet. 3:375–383.

Allcock, A. L., and J. M. Strugnell. 2012. Southern Ocean diversity: new paradigms from molecular ecology. Trends Ecol. Evol. 27:520–528.

Barnes, D. K. A., D. A. Hodgson, P. Convey, C. S. Allen, and A. Clarke. 2006. Incursion and excursion of Antarctic biota: past, present and future. Glob. Ecol. Biogeogr. 15:121–142.

Baylis, A. M. M., S. Crofts, and A. C. Wolfaardt. 2013. Population trends of Gentoo Penguins *Pygoscelis papua* breeding at the Falkland Islands. Mar. Ornithol. 41:1–5.

Bennie, J., J. A. Hodgson, C. R. Lawson, C. T. R. Holloway, D. B. Roy, T. Brereton, et al. 2013. Range expansion through fragmented landscapes under a variable climate. Ecol. Lett. 16:921–929.

Billing, T. M., P. J. Guay, A. J. Peucker, and R. A. Mulder. 2007. PRIMER NOTE: isolation and characterization of polymorphic microsatellite loci for the study of paternity and population structure in the little penguin *Eudyptula minor*. Mol. Ecol. Notes 7:425–427.

Black, C. E. 2015. A comprehensive review of the phenology of *Pygoscelis* penguins. Polar Biol. DOI: 10.1007/s00300-015-1807-8.

Boersma, P. D. 2008. Penguins as marine sentinels. Bioscience 58:597–607.

Boessenkool, S., T. M. King, P. J. Seddon, and J. M. Waters. 2008. Isolation and characterization of microsatellite loci from the yellow-eyed penguin (*Megadyptes antipodes*). Mol. Ecol. Resour. 8:1043–1045.

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

Bost, C. A., and P. Jouventin. 1991. The breeding performance of the Gentoo Penguin *Pygoscelis papua* at the northern edge of its range. The Ibis 133:14–25.

Brookfield, J. F. Y. 1996. A simple new method for estimating null allele frequency from heterozygote deficiency. Mol. Ecol. 5:453–455.

Brown, R. M., and W. C. Jordan. 2009. Characterization of polymorphic microsatellite loci from Round Island Petrels (*Pterodroma arminjoniana*) and their utility in other seabird species. J. Ornithol. 150:925–929.

Carvajal-Rodriguez, A., and J. de Uña-Alvarez. 2011. Assessing significance in high-throughput experiments by sequential goodness of fit and q-value estimation. PLoS ONE 6:e24700.

Casanovas, P., R. Naveen, S. Forrest, J. Poncet, and H. Lynch. 2015. A comprehensive coastal seabird survey maps out the front lines of ecological change on the western Antarctic Peninsula. Polar Biol. 38:927–940.

Chapuis, M. P., and A. Estoup. 2007. Microsatellite null alleles and estimation of population differentiation. Mol. Biol. Evol. 24:621–631.

Charcot, J., and P. Walsh. 1911. The voyage of the 'Why Not?' in the Antarctic. Hodder and Stoughton, New York and London.

Chown, S. L., A. Clarke, C. I. Fraser, S. C. Cary, K. L. Moon, M. A. McGeoch. 2015. The changing form of Antarctic biodiversity. Nature 522:431–438.

Clausen, A., and K. Pütz. 2003. Winter diet and foraging range of gentoo penguins (*Pygoscelis papua*) from Kidney Cove, Falkland Islands. Polar Biol. 26:32–40.

Clucas, G. V., M. J. Dunn, G. Dyke, S. D. Emslie, H. Levy, R. Naveen, et al. 2014. A reversal of fortunes: climate change 'winners' and 'losers' in Antarctic Peninsula penguins. Sci. Rep. 4:e5024.

Convey, P., D. K. A. Barnes, H. J. Griffiths, S. M. Grant, K. Linse, and D. N. Thomas. 2012. Biogeography and regional classifications of Antarctica. Pp. 469–491. *in* A. D. Rogers, N. M. Johnston, E. J. Murphy and A. Clarke, eds. Antarctic Ecosystems: an extreme environment in a changing world. Blackwell Publishing Ltd, Oxford, UK.

Cook, A. J., and D. G. Vaughan. 2010. Overview of areal changes of the ice shelves on the Antarctic Peninsula over the past 50 years. Cryosphere 4:77–98.

Croxall, J. P., T. S. McCann, P. A. Prince, and P. Rothery. 1988. Reproductive Performance of Seabirds and Seals at South Georgia and Signy Island, South Orkney Islands, 1976–1987: Implications for Southern Ocean Monitoring Studies. Pp. 261–285 *in* D. Sahrhage, ed. Antarctic Ocean and resources variability. Springer, Berlin, Heidelberg.

Croxall, J. P., P. N. Trathan, and E. J. Murphy. 2002. Environmental Change and Antarctic Seabird Populations. Science 297:1510–1514.

Dann, P., L. A. Sidhu, R. Jessop, L. Renwick, M. Healy, B. Dettmann, et al. 2014. Effects of flipper bands and injected transponders on the survival of adult Little Penguins *Eudyptula minor*. The Ibis 156:73–83.

Darriba, D., G. L. Taboada, R. Doallo, and D. Posada. 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods 9:772.

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

de Dinechin, M., F. S. Dobson, P. Zehtindjiev, R. Metcheva, C. Couchoux, A. Martin, et al. 2012. The biogeography of Gentoo Penguins (P*ygoscelis papua*). Can. J. Zool. 90:352–360.

Drummond, A. J., M. A. Suchard, D. Xie, and A. Rambaut. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29:1969–1973.

Ducklow, H. W., K. Baker, D. G. Martinson, L. B. Quetin, R. M. Ross, R. C. Smith, et al. 2007. Marine pelagic ecosystems: the West Antarctic Peninsula. Philos. Trans. R. Soc. B Biol. Sci. 362:67–94.

Dugger, K. M., G. Ballard, D. G. Ainley, P. O. B. Lyver, and C. Schine. 2014. Adélie penguins coping with environmental change: results from a natural experiment at the edge of their breeding range. Front. Ecol. Evol. 2:68.

Earl, D. A., and B. M. Vonholdt. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. 4:359–361.

Emslie, S. D. 2001. Radiocarbon dates from abandoned penguin colonies in the Antarctic Peninsula region. Antarct. Sci. 13:289–295.

Emslie, S. D., L. Coats, and K. Licht. 2007. A 45,000 yr record of Adélie penguins and climate change in the Ross Sea, Antarctica. Geology 35:61–64.

Enticott, J. W. 1986. Distribution of penguins at sea in the South-eastern Atlantic and Southwestern Indian Oceans. Cormorant 13:114–142.

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. Mol. Ecol. 14:2611–2620.

Excoffier, L., P. E. Smouse, and J. M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes – application to human mitochondrial-DNA restriction data. Genetics 131:479–491.

Excoffier, L., G. Laval, and S. Schneider. 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol. Bioinform. 1:47–50.

Falush, D., M. Stephens, and J. K. Pritchard. 2007. Inference of population structure using multilocus genotype data: dominant markers and null alleles. Mol. Ecol. Notes 7: 574–578.

Ferreira, M. A. R., and M. A. Suchard. 2008. Bayesian analysis of elapsed times in continuous-time Markov chains. Can. J. Stat. 36:355–368.

Fox, A. J., and D. G. Vaughan. 2005. The retreat of Jones Ice Shelf, Antarctic Peninsula. J. Glaciol. 51:555–560.

Fraser, W. R., W. Z. Trivelpiece, D. G. Ainley, and S. G. Trivelpiece. 1992. Increases in Antarctic penguin populations - reduced competition with whales or a loss of sea ice due to environmental warming. Polar Biol. 11:525–531.

Fraser, C. I., R. Nikula, H. G. Spencer, and J. M. Waters. 2009. Kelp genes reveal effects of subantarctic sea ice during the Last Glacial Maximum. Proc. Natl Acad. Sci. 106:3249–3253.

Fretwell, P. T., and P. N. Trathan. 2009. Penguins from space: faecal stains reveal the location of emperor penguin colonies. Glob. Ecol. Biogeogr. 18:543–552.

Friesen, V. L., T. M. Burg, and K. D. Mccoy. 2007a. Mechanisms of population differentiation in seabirds. Mol. Ecol. 16:1765–1785.

Friesen, V. L., A. L. Smith, E. Gomez-Diaz, M. Bolton, R. W. Furness, J. González-Solis, and L. R. Monteiro. 2007b.

Sympatric speciation by allochrony in a seabird. Proc. Natl Acad. Sci. USA 104:18589–18594.

Friesen, V. L. 2015. Speciation in seabirds: why are there so many species... and why aren't there more? J. Ornith. 156:27–39.

Froget, G., M. Gautier-Clerc, Y. Le Maho, and Y. Handrich. 1998. Is Penguin banding harmless? Polar Biol. 20:409–413.

Gain, L. 1913. The Penguins of the Antarctic Regions. Pp. 475–482 *in* Smithsonian Report for 1912. Smithsonian, Washington.

Gauthier–Clerc, M., J.-P. Gendner, C. A. Ribic, W. R. Fraser, E. J. Woehler, S. Descamps, et al. 2004. Long-term effects of flipper bands on penguins. Proc. Roy. Soc. B Biol. Sci. 271: S423–S426.

Gersonde, R., X. Crosta, A. Abelmann, and L. Armand. 2005. 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.

Ghys, M. I., A. R. Rey, and A. Schiavini. 2008. Population trend and breeding biology of gentoo penguin in martillo Island, Tierra Del Fuego, Argentina. Waterbirds 31:625–631.

Guillot, G. 2008. Inference of structure in subdivided populations at low levels of genetic differentiation—the correlated allele frequencies model revisited. Bioinformatics 24:2222–2228.

Guillot, G., A. Estoup, F. Mortier, and J. F. Cosson. 2005a. A spatial statistical model for landscape genetics. Genetics 170:1261–1280.

Guillot, G., F. Mortier, and A. Estoup. 2005b. Geneland: a computer package for landscape genetics. Mol. Ecol. Notes 5:712–715.

Guindon, S., and O. Gascuel. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52:696–704.

Guo, S. W., and E. A. Thompson. 1992. Performing the exact test of hardy-Weinberg proportion for multiple alleles. Biometrics 48:361–372.

Heide-Jorgensen, M.-P., T. Harkonen, R. Dietz, and P. M. Thompson. 1992. Retrospective of the 1988 European seal epizootic. Dis. Aquat. Organ. 13:37–62.

Hersteinsson, P., and D. W. MacDonald. 1992. Interspecific competition and the geographical distribution of red and arctic foxes *Vulpes vulpes* and *Alopex lagopus*. Oikos 64:505–515.

Huin, N. 2003. Falkland Islands seabird monitoring programme annual report 2002/2003. Falklands Conservation, Stanley, Falkland Islands.

Jehl, J. R., F. S. Todd, M. Rumboll, and B. Swartz. 1979. Pelagic birds in the South Atlantic Ocean and at South Georgia in the austral autumn. Le Gerfaut 69:13–27.

Jenouvrier, S., H. Caswell, C. Barbraud, M. Holland, J. Stroeve, and H. Weimerskirch. 2009. Demographic models and IPCC climate projections predict the decline of an emperor penguin population. Proc. Natl Acad. Sci. USA 106:1844–1847.

Jenouvrier, S., M. Holland, J. Stroeve, M. Serreze, C. Barbraud, H. Weimerskirch, and H. Caswell. 2014. Projected continent-wide declines of the emperor penguin under climate change. Nat. Clim. Chang. 4:715–718.

LaRue, M. A., G. Kooyman, H. J. Lynch, and P. Fretwell. 2015. Emigration in emperor penguins: implications for interpretation of long-term studies. Ecography 38: 114–120.

Laws, R. M.. 1984. Antarctic ecology. Academic Press, London.

Lemey, P., A. Rambaut, A. J. Drummond, and M. A. Suchard. 2009. Bayesian phylogeography finds its roots. PLoS Comput. Biol. 5:e1000520.

Lescroel, A., and C. A. Bost. 2005. Foraging under contrasting oceanographic conditions: the gentoo penguin at Kerguelen Archipelago. Mar. Ecol. Prog. Ser. 302:245–261.

Lescroel, A., C. Bajzak, and C. A. Bost. 2009. Breeding ecology of the gentoo penguin *Pygoscelis papua* at Kerguelen Archipelago. Polar Biol. 32:1495–1505.

Lynch, H. J., R. Naveen, and W. F. Fagan. 2008. Censuses of penguin, blue-eyed shag (*Phalacrocorax atriceps*) and southern giant petrel (*Macronectes giganteus*) populations on the Antarctic Peninsula. Mar. Ornithol. 36:83–97.

Lynch, H., W. Fagan, and R. Naveen. 2010. Population trends and reproductive success at a frequently visited penguin colony on the western Antarctic Peninsula. Polar Biol. 33:493–503.

Lynch, H. J., R. Naveen, P. N. Trathan, and W. F. Fagan. 2012. Spatially integrated assessment reveals widespread changes in penguin populations on the Antarctic Peninsula. Ecology 93:1367–1377.

Mayewski, P. A., M. P. Meredith, C. P. Summerhayes, J. Turner, A. Worby, P. J. Barrett, et al. 2009. State of the Antarctic and Southern Ocean climate system. Rev. Geophys. 47:RG1003.

Mayr, E.. 1942. Systematics and the origin of species, from the viewpoint of a zoologist. Harvard University Press, Cambridge, MA, USA.

Meirmans, P. G., and P. H. Van Tienderen. 2004. Genotype and Genodive: two programs for the analysis of genetic diversity of asexual organisms. Mol. Ecol. Notes 4:792–794.

Millar, C. D., A. Dodd, J. Anderson, G. C. Gibb, P. A. Ritchie, C. Baroni, et al. 2008. Mutation and evolutionary rates in Adélie Penguins from the Antarctic. PLoS Genet. 4: e1000209.

Munro, G. 2006. Outbreak of avian pox virus in gentoo penguins in the Falklands. Falklands Conservation, Stanley, Falkland Islands.

Oksanen, J., F. G. Blanchet, R. Kindt, P. Legendre, P. R. Michin, R. B. O'Hara, et al. 2015. Vegan: Community Ecology Package.

Paetkau, D., R. Slade, M. Burden, and A. Estoup. 2004. Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. Mol. Ecol. 13:55–65.

Parmesan, C., N. Ryrholm, C. Stefanescu, J. K. Hill, C. D. Thomas, H. Descimon, et al. 1999. Poleward shifts in geographical ranges of butterfly species associated with regional warming. Nature 399:579–583.

Peña, M. F., E. Poulin, G. P. M. Dantas, D. González-Acuña, M. V. Petry, and J. A. Vianna. 2014. Have historical climate changes affected Gentoo Penguin (*Pygoscelis papua*) populations in Antarctica? PLoS ONE 9:e95375.

Pistorius, P. A., N. Huin, and S. Crofts. 2010. Population change and resilience in Gentoo Penguins *Pygoscelis papua* at the Falkland Islands. Mar. Ornithol. 38:49–53.

Posada, D. 2008. jModelTest: phylogenetic model averaging. Mol. Biol. Evol. 25:1253–1256.

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

Pritchard, J., D. Falush, and M. Stephens. 2002. Inference of population structure in recently admixed populations. Am. J. Hum. Genet. 71:177–177.

Rambaut, A., M. A. Suchard, D. Xie, and A. J. Drummond. 2014. Tracer v1.6.

Ratcliffe, N., and P. Trathan. 2011. A review of the diet and at-sea distribution of penguins breeding within the CAMLR Convention Area. CCAMLR Sci. 18:75–114.

Rignot, E., G. Casassa, S. Gogineni, P. Kanagaratnam, W. Krabill, H. Pritchard, et al. 2005. Recent ice loss from the Fleming and other glaciers, Wordie Bay, West Antarctic Peninsula. Geophys. Res. Lett. 32:L07502.

Roeder, A. D., R. K. Marshall, A. J. Mitchelson, T. Visagathilagar, P. A. Ritchie, D. R. Love, et al. 2001. Gene flow on the ice: genetic differentiation among Adelie penguin colonies around Antarctica. Mol. Ecol. 10: 1645–1656.

Roeder, A. D., P. A. Ritchie, and D. M. Lambert. 2002. New DNA markers for penguins. Conserv. Genet. 3:341–344.

Rogers, A. D.. 2012. Evolution and Biodiversity of Antarctic Organisms. Pp. 417–467 *in* A. D. Rogers, N. M. Johnston, E. J. Murphy, and A. Clarke, eds. Antarctic Ecosystems: an extreme environment in a changing world. Blackwell Publishing Ltd, Oxford, UK.

Rogers, A. D., E. J. Murphy, N. M. Johnston, and A. Clarke. 2007. Introduction. Antarctic ecology: from genes to ecosystems. Part 2. Evolution, diversity and functional ecology. Philos. Trans. R. Soc. Lond. B Biol. Sci. 362: 2187–2189.

Saraux, C., C. Le Bohec, J. M. Durant, V. A. Viblanc, M. Gauthier-Clerc, D. Beaune, et al. 2011. Reliability of flipper-banded penguins as indicators of climate change. Nature 469:203–206.

Schlosser, J. A., J. M. Dubach, T. W. J. Garner, B. Araya, M. Bernal, A. Simeone, et al. 2009. Evidence for gene flow differs from observed dispersal patterns in the Humboldt penguin, *Spheniscus humboldti*. Conserv. Genet. 10:839–849.

Slatkin, M., and L. Excoffier. 1996. Testing for linkage disequilibrium in genotypic data using the expectation-maximization algorithm. Heredity 76:377–383.

Smouse, P. E., J. C. Long, and R. R. Sokal. 1986. Multiple-regression and correlation extensions of the mantel test of matrix correspondence. Syst. Zool. 35:627–632.

Stonehouse, B. 1970. Geographic variation in gentoo penguins *Pygoscelis papua*. The Ibis 112:52–000.

Strugnell, J. M., P. C. Watts, P. J. Smith, and A. L. Allcock. 2012. Persistent genetic signatures of historic climatic events in an Antarctic octopus. Mol. Ecol. 21:2775–2787.

Sugden, D. E., and C. M. Clapperton. 1977. The maximum ice extent on island groups in the Scotia Sea, Antarctica. Quat. Res. 7:268–282.

Tanton, J. L., K. Reid, J. P. Croxall, and P. N. Trathan. 2004. Winter distribution and behaviour of gentoo penguins *Pygoscelis papua* at South Georgia. Polar Biol. 27:299–303.

Thomas, C. D., and J. J. Lennon. 1999. Birds extend their ranges northwards. Nature 399:213–213.

Thurston, M. H. 1982. Ornithological observations in the South Atlantic Ocean and Weddell Sea 1959-64. Br. Antarct. Sur. Bull. 55:77–103.

Trathan, P. N., J. Forcada, and E. J. Murphy. 2007. Environmental forcing and Southern Ocean marine predator populations: effects of climate change and variability. Philos. Trans. R. Soc. B Biol. Sci. 362:2351–2365.

Trathan, P. N., J. Forcada, R. Atkinson, R. H. Downie, and J. R. Shears. 2008. Population assessments of gentoo penguins (*Pygoscelis papua*) breeding at an important Antarctic tourist site, Goudier Island, Port Lockroy, Palmer Archipelago, Antarctica. Biol. Conserv. 141:3019–3028.

Trathan, P. N., P. T. Fretwell, and B. Stonehouse. 2011. First recorded loss of an emperor penguin colony in the recent period of antarctic regional warming: implications for other colonies. PLoS ONE 6:e14738.

Trivelpiece, W. Z., S. G. Trivelpiece, and N. J. Volkman. 1987. Ecological Segregation of Adelie, Gentoo, and Chinstrap Penguins at King George Island, Antarctica. Ecology 68:351–361.

Trivelpiece, W. Z., J. T. Hinke, A. K. Miller, C. S. Reiss, S. G. Trivelpiece, and G. M. Watters. 2011. Variability in krill biomass links harvesting and climate warming to penguin population changes in Antarctica. Proc. Natl Acad. Sci. 108:7625–7628.

Van Oosterhout, C., W. F. Hutchinson, D. P. M. Wills, and P. Shipley. 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Mol. Ecol. Notes 4:535–538.

Vaughan, D. G., G. J. Marshall, W. M. Connolley, C. Parkinson, R. Mulvaney, D. A. Hodgson, et al. 2003. Recent

rapid regional climate warming on the Antarctic Peninsula. Clim. Change. 60:243–274.
Voisin, J.-F. 1979. Observations ornithologiques aux îles Tristan da Cunha et Gough. Alauda 47:245–280.
Weir, B. S., and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population-structure. Evolution 38: 1358–1370.
White, R. W., K. W. Gillon, A. D. Black, and J. B. Reid. 2002. The distribution of seabirds and marine mammals in Falkland Islands waters. Joint Nature Conservation Committee, Peterborough, UK.
Williams, T. D. 1995. The penguins: Spheniscidae. Oxford University Press, Oxford, New York.
Williams, T. D., and J. P. Croxall. 1991. The gentoo penguin as a candidate species for the CCAMLR Ecosystem Monitoring Program. Pp. 483–488. *in* Selected scientific papers. CCAMLR, Hobart, Australia.
Williams, T. D., and S. Rodwell. 1992. Annual variation in return rate, mate and nest-site fidelity in breeding gentoo and macaroni penguins. Condor 94:636–645.
Wilson, R. P., B. Alvarrez, L. Latorre, D. Adelung, B. Culik, and R. Bannasch. 1998. The movements of gentoo penguins *Pygoscelis papua* from Ardley Island, Antarctica. Polar Biol. 19:407–413.
Woehler, E. J., J. Cooper, J. P. Croxall, W. R. Fraser, G. L. Kooyman, G. D. Miller, et al. 2001. A Statistical Assessment of the Status and Trends of Antarctic and Sub-Antarctic Seabirds; Scientific Committee on Antarctic Research, Cambridge, UK.
Xu, W., Z. Ouyang, A. Viña, H. Zheng, J. Liu, and Y. Xiao. 2006. Designing a conservation plan for protecting the habitat for giant pandas in the Qionglai mountain range, China. Divers. Distrib. 12:610–619.
Zwally, H. J., M. B. Giovinetto, J. Li, H. G. Cornejo, M. A. Beckley, A. C. Brenner, et al. 2005. Mass changes of the Greenland and Antarctic ice sheets and shelves and contributions to sea-level rise: 1992–2002. J. Glaciol. 51: 509–527.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1. Genetic diversity of Gentoo penguins (*Pygoscelis papua*) at fourteen breeding sites across the Scotia Arc.
Appendix S2. Maximum clade credibility tree derived from mtDNA showing the origin and differentiation of *Pygoscelis papua* lineages north (Falkland Islands, light green above) and south of the Polar Front (all other colors and locations).
Appendix S3. Graph of the mean likelihood of the number of populations $L(K)$ versus the number of populations K, resulting from STRUCTURE Harvester using data from all study colonies ($n = 14$), under the Admixture model with Correlated allele frequencies, using No Prior Location in STRUCTURE.
Appendix S4. Graph of the mean likelihood of the number of populations $L(K)$ versus the number of populations K, resulting from STRUCTURE Harvester using data from all Falkland Island colonies ($n = 10$), under the Admixture model with Correlated allele frequencies, using No Prior Location in STRUCTURE.

Appendix E

Polar Biol (2015) 38:1493–1502
DOI 10.1007/s00300-015-1711-2

ORIGINAL PAPER

Limited genetic differentiation among chinstrap penguin (*Pygoscelis antarctica*) colonies in the Scotia Arc and Western Antarctic Peninsula

Jennifer J. Freer[1] · Barbara K. Mable[1] · Gemma Clucas[2,3] · Alex D. Rogers[3] · Michael J. Polito[4] · Michael Dunn[5] · Ron Naveen[6] · Hila Levy[3] · Tom Hart[3]

Received: 23 June 2014 / Revised: 28 April 2015 / Accepted: 11 May 2015 / Published online: 26 May 2015

Abstract Long-term monitoring of seabird numbers around Antarctica has revealed that the chinstrap penguin (*Pygoscelis antarctica*) is largely declining throughout its range in the Scotia Arc. Whether archipelagos across this large area remain connected via dispersal or represent genetically isolated groups has not yet been established. The purpose of this study was to assess the level of genetic differentiation between four breeding colonies on the Western Antarctic Peninsula (WAP), South Shetland, South Orkney, and South Sandwich Islands using microsatellite-based analysis of population structure. All colonies had similar levels of genetic diversity (mean heterozygosity, $H_O = 0.583$) but colonies from the WAP and South Orkney Island had significant inbreeding coefficients. Hierarchical and pairwise *F*-statistics revealed very limited population structure in the Scotia Arc, with weak differentiation between colonies from the WAP, South Shetland and South Orkney Islands relative to the South Sandwich Islands, which are situated at least 1000 km apart from these other archipelagos. Bayesian model-based clustering methods found no evidence of significant population structuring, suggesting that whilst some isolation by distance may occur, there are no strong barriers to dispersal across this wide geographic range. No evidence of sex-biased dispersal was detected. We conclude that chinstrap penguin colonies across the Scotia Arc represent one interconnected breeding population. High levels of gene flow may be important in maintaining smaller, less stable colonies, and this status should be preserved by creating dispersal corridors throughout the Scotia Arc.

Keywords Scotia Arc · Population genetics · Microsatellite · Molecular sexing · *Pygoscelis antarctica*

Electronic supplementary material The online version of this article (doi:10.1007/s00300-015-1711-2) contains supplementary material, which is available to authorized users.

✉ Tom Hart
tom.hart@zoo.ox.ac.uk

[1] Institute of Biodiversity, Animal Health and Comparative Medicine, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow G12 8QQ, UK

[2] Ocean and Earth Sciences, University of Southampton, Waterfront Campus, European Way, Southampton SO14 3ZH, UK

[3] Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

[4] Department of Oceanography and Coastal Sciences, Louisiana State University, Baton Rouge, LA 70803, USA

[5] British Antarctic Survey, High Cross, Madingley Road, Cambridge CB3 0ET, UK

[6] Oceanites Inc, PO Box 15259, Chevy Chase, MD 20825, USA

Introduction

Penguins are a dominant component of avian biomass in the Southern Ocean and are important predators in marine ecosystems (Croxall 1987; Forcada and Trathan 2009). Population surveys have provided long-term data and show that although penguins are highly adapted to survive the Antarctic environment, they are sensitive to climatic changes (Forcada et al. 2006; Forcada and Trathan 2009; Jenouvrier et al. 2009; Ainley et al. 2010). Because they are an integral component of the Antarctic environment and are not directly targeted by humans, monitoring penguin populations can thus serve as a sensitive bio-indicator

of Antarctic ecosystem changes (Croxall et al. 2002; Forcada and Trathan 2009; Trivelpiece et al. 2011).

Previous monitoring using estimates of abundance and recruitment has revealed species-specific differences in the response of penguin populations to climate warming in the Antarctic Peninsula and Scotia Arc region (Forcada et al. 2006; Ducklow et al. 2007; Hinke et al. 2007) in the recent and long-term past (Clucas et al. 2014). For example, it is thought that Adélie penguins (*Pygoscelis adeliae*) have suffered population declines as the result of reduction in sea ice area and prey density (Trivelpiece et al. 2011; Lynch et al. 2012), whilst gentoo penguins (*Pygoscelis papua*) have expanded their range southward, as breeding sites open up with retreating ice (Fraser et al. 1992; Lynch et al. 2012). Whist at least one small breeding colony in the Western Antarctic Peninsula (WAP) has increased in size ($\sim$300 breeding pairs; (Ducklow et al. 2007)), there is a growing body of evidence indicating that chinstrap penguins (*Pygoscelis antarctica*) have largely declined across the Antarctic Peninsula and Scotia Arc in response to recent, climate-driven declines in the abundance of their primary prey, Antarctic krill (Sander et al. 2007; Forcada and Trathan 2009; Trivelpiece et al. 2011; Lynch et al. 2012). These significant declines have been suggested by data from Deception Island in the South Shetland Islands (Barbosa et al. 2012), where the total breeding population of chinstrap penguins declined by at least 50 % between 1987 (141,000–191,000 pairs; Shuford and Spear 1988) and 2011 (79,849 pairs; Naveen et al. 2012). Such population declines mean that it is important to determine how hidden population structure may influence the population dynamics of this species (Kramer et al. 2009), which could inform management strategies.

The characteristic reproductive biology of penguins, however, represents a challenge to conservation scientists. All pygoscelid penguins (gentoo, Adélie, and chinstrap) are known to exhibit natal philopatry, the tendency for individuals to breed at or near their place of origin (Korczak-Abshire et al. 2012), and a high degree of both mate (Ancel et al. 2013) and site (Williams 1995) fidelity, with partners returning to the same nest site year after year. These factors have the potential to limit connectivity between colonies, creating isolated and genetically distinct groups. Patterns of genetic structuring, however, appear to be species specific. Analyses using the hypervariable region-1 (HVR-1) of the mitochondrial control region found clear genetic differentiation between colonies of gentoo penguins, weak differentiation between chinstrap colonies, and no differentiation between Adélie colonies (Clucas et al. 2014). Such differences in population structure between species are most likely the result of different distributions in the non-breeding season (Friesen et al. 2007). Species that are residents, and stay close to their breeding grounds all year, are likely to have defined population structure (as demonstrated by gentoo penguins), whilst species that are dispersive in the non-breeding season are more likely to be unstructured (such as Adélie and chinstrap penguins).

Other studies have investigated region-wide population genetics of penguins using microsatellite markers. These are inherited from both parents and have higher rates of substitution relative to mitochondrial DNA, making them more sensitive to weak or recent population structure. Most of these studies have found high genetic diversity and a lack of genetic differentiation between colonies (Adélie, Roeder et al. 2001; Magellanic, *Spheniscus magellanicus*, Bouzat et al. 2009; Humboldt, *Spheniscus humboldti*, Schlosser et al. 2009), though breaks in genetic connectivity with distance have also been recorded (little penguin, *Eudyptula minor*, Overeem et al. 2008; yellow-eyed penguin, *Megadyptes antipodes*, Boessenkool et al. 2009). For chinstrap penguins, there has not been a region-wide investigation using microsatellite markers. Korczak-Abshire et al. (2012) found no evidence of differentiation between colonies within the South Shetland Islands based on nuclear anonymous amplified fragment length polymorphism (AFLP) markers, but this was over a very short distance of 32 km. There thus remains the possibility that population structuring exists between archipelagos. Alternatively, the chinstrap penguin's ability to migrate over distances up to 3600 km outside of their breeding season (Biuw et al. 2009) could act to homogenise breeding colonies and allow connectivity to persist over a large, patchy geographic range. Furthermore, if the recent colonisation of the WAP was made by a small number of individuals, reduced genetic diversity within these colonies can be expected.

Differences in dispersal between sexes also could obscure any fine-scale patterns of structuring. Sex-biased dispersal in chinstrap penguins is yet to be investigated, though it is a common phenomenon in migratory avian species (Scribner et al. 2001). A prominence of female-biased dispersal in birds was found by Greenwood (1980) and again by Clarke et al. (1997). However, the latter also mentions reports of male-biased natal dispersal in Adélie penguins and in yellow-eyed penguins. More recently, Bouzat et al. (2009) detected differences in the spatial genetic variance between mtDNA and microsatellite markers for the Magellanic penguin, which could also be attributed to a higher dispersal rate in males than in females. Investigating sex-biased dispersal in chinstrap penguins will not only further our understanding of any pattern of structure we detect, but may help in resolving differences between the general trend of female-biased dispersal in birds, against the indication of male-biased dispersal in penguins.

The purpose of this study was to use microsatellite markers to assess the level of genetic differentiation between colonies of chinstrap penguins on the WAP and

archipelagos within the Scotia Arc, which encompasses the South Shetland Islands, South Orkney Islands, and South Sandwich Islands (Fig. 1). Our specific objectives were to: (1) identify any underlying level of genetic structuring between colonies; (2) discriminate any differences between male and female patterns of genetic differentiation; and (3) assess the amount of genetic variation of chinstrap penguins within the colonies of our study area. In addition, we assessed the reliability of non-invasively collected shed feathers (relative to plucked feathers and blood samples) for genotyping chinstrap penguins using microsatellite markers, as there is growing support for the use of shed feathers and other non-invasive sampling methods in genetic studies (Hogan et al. 2008; Johansson et al. 2012; Vili et al. 2013).

Materials and methods

Sample collection and DNA extraction

A total of 209 samples from colonies distributed along the Scotia Arc and on the WAP (Fig. 1) were gathered between 2009 and 2012: (1) Orne Harbour (OH; $n = 48$) on the WAP; (2) Admiralty Bay, King George Island (KG; $n = 47$) on the South Shetland Islands; (3) Signy Island (SI; $n = 47$) on the South Orkney Islands; and Thule Island (TH; $n = 67$) on the South Sandwich Islands (Table 1). To assess the reliability of non-invasively versus invasively collected samples, a mixture of plucked feathers, shed feathers, and blood samples were used for DNA extraction (Table 1). Where direct samples were taken (blood and plucked feathers), birds were seized with both hands by the upper body and the flippers were restrained by the same handler. The head was placed under the arm of the handler to stop the bird biting in accordance with the literature on minimising stress in restrained penguins (Lemaho et al. 1992; Wilson 1997). The second person plucked two feathers from this bird's lower back or took blood samples (0.1 ml). Where taken, bloods were from the brachial vein using a 25G needle and syringe and were immediately stored in 95 % ethanol at ambient temperature. The animal was then released at the edge of the colony. It is possible to pluck feathers with a minimum of stress within 30 s, but usually no longer than 2 min. Blood samples usually take 2–3 min of restraint. Direct samples were only taken from breeding adults towards the edge of the colony. Shed feathers were collected from around the breeding colony every 2–5 m. At each colony, 80–125 moulted penguin body and tail (retrix) feathers were collected from nesting sites. Feathers were collected several metres apart to minimise the chance of obtaining duplicate samples from an individual. DNA from all samples were extracted using DNeasy Blood and Tissue kits (QIAGEN, Crawley, West Sussex, UK), as reported in Clucas et al. (2014).

Microsatellite genotyping

Genetic variation was assessed using 10 microsatellites (Online Resource 1) that were initially designed for other

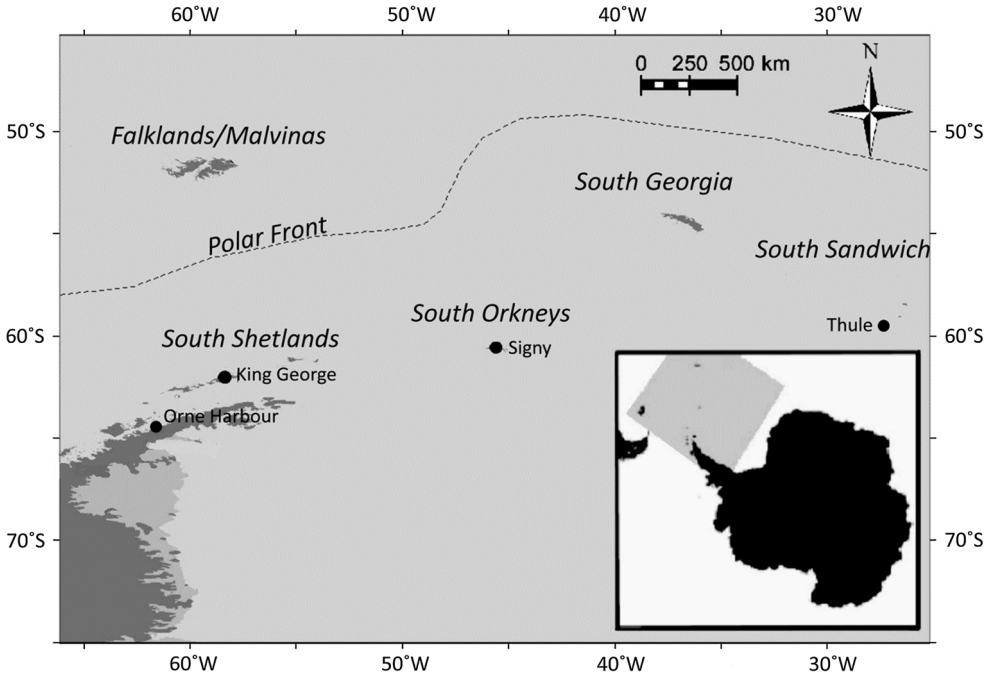


Fig. 1 Locations of chinstrap penguin (*Pygoscelis antarctica*) colonies sampled for this study across the Scotia Arc (King George, Signy, and Thule Islands) and the Western Antarctic Peninsula (Orne Harbour). The archipelago names are given in *italics*, and sample locations are indicated by *dots*. The *dotted line* shows the boundary of the Polar Front. *Inset* shows the location of the map in grey relative to the Antarctic continent in *black*

Table 1 Summary information of chinstrap penguin (*Pygoscelis antarctica*) colonies sampled

Colony name	Archipelago	Number of samples	Type of sample used
Orne Harbour (OH)	Western Antarctic Peninsula (WAP)	26	Plucked feather
		22	Shed feather
King George Island (KG)	South Shetland	47	Plucked feather
Signy Island (SI)	South Orkney	47	Plucked feather
Thule Island (TH)	South Sandwich	48	Shed feather
		19	Blood

The full name and abbreviation for each colony are given, as well as the number of individuals sampled and the tissue type used for DNA extractions (plucked feather, shed feather, or blood)

penguin species: Adélie (Roeder et al. 2002); little penguin (Billing et al. 2006); macaroni, *Eudyptes chrysolophus* (Ahmed et al. 2009). The forward primer of each marker was labelled with either 6FAM or HEX fluorescent dyes. Markers were then divided into two multiplexes: multiplex 1 included the microsatellites Ech0078, Emm4, and B3-2 and was amplified using an annealing temperature of 54 °C; multiplex 2 included the microsatellites Ech010, Ech011, Ech030, Ech050, Ech071, Ech091, and RM3 and was amplified using an annealing temperature of 60 °C. Optimal annealing temperatures were determined based on an initial temperature gradient PCR (52–62 °C).

Multiplex PCR was carried out using Qiagen Multiplex PCR Mastermix, and conditions were as recommended by the manufacturers (Qiagen Multiplex PCR Kit, Qiagen Ltd.), with the exception of a primer concentration of 0.5 μM. PCR's were run using the following protocol: 95 °C for 15 min; 40 cycles of 94 °C for 30 s; 54 or 60 °C for 90 s; and 72 °C for 45 s, followed by a final step at 72 °C for 10 min. All reactions were carried out on a MJ Research PTC-200 thermal cycler (MJ Research Inc., Quebec, Canada). PCR products were sent to the DNA Sequences and Services at the University of Dundee (Dundee, UK) for genotyping using an ABI 3730 automated sequencer. At Dundee, PCR products were diluted to 1:60 using HiDi Formamide (Applied Biosystems, UK), run with a ROX 400HD size standard, and visualised on a 3730 Genetic Analyser (Applied Biosystems). Peak calling and allele designation were performed using GeneMapper v4.0 (Applied Biosystems).

Molecular sexing

The sex of each individual was determined using a molecular sexing technique. A well-established primer pair for non-ratite bird sexing from the Chromosome-helicase-DNA binding-1 (CHD1) gene region was used; 2550F [5′-GTTACTGATTCGTCTACGAGA-3′ (Fridolfsson and Ellegren 1999)] and 2757R [5′-AATTCCCCTTTTATTGA TCCA TC-3′ (R. Griffiths unpubl. data)]. The following PCR conditions were used: a final volume of 10 μl consisting of 0.8 μM of 2550F and 2757R primers, 200 μM of each dNTP, 2μM of 25 mM of $MgCl_2$, 2μM (5×) of GoTAQ Flexi buffer (Promega, UK), 0.35 units of GoTaq polymerase (Promega), and 100 μM of target DNA. PCR's were run using the following protocol: 94 °C for 2 min; 29 cycles of 94 °C for 45 s; 49 °C for 1 min; and 72 °C for 1 min, followed by 49 °C for 2 min and a final step at 72 °C for 10 min. All reactions were carried out on a TGradient96 Biometra thermal cycler (Biometra GmBH, Goettingen, Germany). PCR products were separated using 2 % agarose gels with 1XTBE buffer and visualised by ethidium bromide staining.

Marker quality control

For microsatellite products, MICRO-CHECKER (Van Oosterhout et al. 2004) was used to detect the presence of null alleles, large-allele drop out and scoring errors. If such issues were detected, genotypes were re-checked on GeneMapper and allele calls modified if there was obvious ambiguity in scoring that could explain the deviations from expected patterns of observed heterozygosity. MICRO-CHECKER was then rerun and if null alleles were still detected, further analyses were run with and without the affected loci.

A linkage disequilibrium test was carried out using ARLEQUIN v.3.5.1.2 (Excoffier and Lischer 2010) to test for significant non-random association of alleles between loci. Significance was determined using Chi-squared tests, and the significance of p values was adjusted using a Bonferroni correction to account for multiple tests (Rice 1989). If linkage disequilibrium was found, one of the affected markers was dropped from further analyses.

Descriptive statistics

Microsatellite Analyser (Dieringer and Schlotterer 2003) was used to compute descriptive statistics for each marker and colony, respectively, including the number of alleles

(N), observed and expected heterozygosity (H_O and H_E), and allele counts and frequencies. The inbreeding coefficient (F_{IS}) was computed for each marker and colony using ARLEQUIN. In order to detect deviations from Hardy–Weinberg equilibrium (HWE), ARLEQUIN was used to perform an exact test of HWE for each marker within each colony. Significance of deviations was computed using 10^6 Markov Chain steps and 10^5 dememorisation steps, and significance was adjusted for multiple tests using a Bonferroni correction (Rice 1989).

Population genetic analyses

Using ARLEQUIN, the spatial genetic structure of populations was analysed using both pairwise F_{ST}'s and Analysis of Molecular Variance (AMOVA). Pairwise F_{ST} measures the extent of genetic differentiation between each subpopulation (Weir and Cockerham 1984), in this case defined as the four colonies. AMOVA partitions the genetic variation into hierarchical levels of spatial separation and gives a value for overall F_{ST}, F_{IS}, and F_{IT} (Excoffier et al. 1992). For this study, levels of separation were defined as variation among colonies, variation within colonies, and variation within individuals. The significance of pairwise F_{ST}'s and AMOVA was computed using 10,000 permutations.

To determine whether genetic distances (pairwise F_{ST} values) were correlated with the degree of geographic separation (distance in km) of colonies, Mantel tests, as implemented in ARLEQUIN, were used to test for significant isolation by distance. The statistical significances of correlation coefficients between the two matrices were estimated using a permutation procedure with 10,000 permutations. Furthermore, a regression line was fitted to a scatterplot of geographic and F_{ST} distances.

The software *STRUCTURE* v2.3.2 (Pritchard et al. 2000) was used to infer whether there is population structure between colonies both with and without defining populations a priori. To do this, it uses a model-based clustering method; individuals are assigned probabilistically to populations (K) using Bayesian Markov Chain Monte Carlo (MCMC) simulations. A burn-in length of 50,000 steps and an MCMC run length of 10^6 steps were used. All simulations were run using the admixture ancestry model with and without a priori location information in combination with the correlated allele frequency model.

The most likely number of clusters was inferred by averaging the log-likelihood values from each independent run ($n = 10$) over each value of K assumed (1–4); both the maximum likelihood and the variance between runs are informative about the value of K that best explains the data. The ΔK method of Evanno et al. (2005) was also applied to judge the most likely values of K. In the analysis of each cluster, plots of the mean log-likelihood L(K) and ΔK were generated by Structure Harvester (Earl and Vonholdt 2012).

Sex-biased dispersal

Tests for sex-biased dispersal were conducted in FSTAT v.2.9.3.2 (Goudet 2002). Five tests based on differences in the inbreeding coefficient (F_{IS}), fixation index (F_{ST}), degree of relatedness, mean assignment indices (*mAIc*), and variance of the assignment indices (*vAIc*) between the philopatric and dispersing groups were utilised (Goudet et al. 2002). In principle, unequal levels of gene flow between males and females would lead to a Wahlund effect and heterozygote deficiency, resulting in a higher F_{IS} and correspondingly lower F_{ST} and relatedness values in the most dispersing sex (Goudet 2002). Furthermore, as the assignment index statistic indicates the probability of a genotype occurring in a population, it may be expected that the most dispersing sex displays negative values of *mAIc* (as the distribution is centred on zero) and a corresponding increase in *vAIc*.

Results

Microsatellite genotyping and molecular sexing

Genotyping results varied depending on the type of tissue sample used. Of the 120 plucked feather samples analysed, 103 gave clear and repeatable allele peaks. Similarly, all 19 blood samples from TH gave allele peaks that were easily scored. Of the 70 shed feather samples analysed, only 15 out of 48 from TH and none out of 22 from OH gave usable allele peaks. In comparison with the direct samples, shed feather genotypes gave higher numbers of null alleles and linkage disequilibrium of loci, with all 8 loci showing significant linkage disequilibrium and all except Emm4 having null alleles. All shed feathers were thus dropped from further analyses, leaving a final sample size of 122 individuals (see Table 2 for final sample size per colony). Using the CHD1 sexing primers, all of these 122 individuals gave clear amplification products resulting in 62 males and 60 females (Table 2). Since only direct samples were used, no exclusions of duplicated genotypes were made.

Marker quality control

Significant homozygote excess, indicative of null alleles, was found for markers B3-2 and Ech008 at KG and markers Ech008, Ech030, and RM3 at OH. No markers were in homozygote excess for SI or blood samples from TH. There was significant linkage at SI between markers

Table 2 Descriptive statistics for each colony averaged over the 8 loci (markers Ech008 and B32 were excluded due to null alleles), indicating the total sample size (N), the number of males (N_M) and females (N_F), number of alleles per locus (N_A), observed (H_O) and expected (H_E) heterozygosity, and inbreeding coefficient (F_{IS})

Colony	N	N_M, N_F	N_A	H_E	H_O	F_{IS}
OH	23	9, 14	6	0.64	0.52	**0.112**
KG	43	21, 22	6	0.65	0.66	−0.037
SI	37	22, 15	7	0.66	0.57	**0.096**
TH	19	10, 9	5	0.62	0.58	0.019

F_{IS} values in bold are significantly different than zero ($p < 0.05$)

Table 3 Summary of pairwise geographic (km) and genetic (F_{ST}) differences between colonies

Colony 1	Colony 2	Distance (km)	F_{ST}
OH	KG	364.27	0.000
OH	SI	994.13	0.001
KG	SI	706.92	0.006
OH	TH	1902.58	**0.075**
KG	TH	1690.12	**0.086**
SI	TH	1012.76	**0.065**

Bold values are significantly different from zero after Bonferroni correction ($p < 0.008$)

Table 4 Results of the five tests of sex-biased dispersal and their corresponding p value, as well as the number (N) of males and females used in the analyses

Sex	N	F_{IS}	F_{ST}	Relatedness	*mAIC*	*vAIc*
Male	62	0.053	0.039	0.072	0.290	6.092
Female	60	0.096	0.039	0.069	−0.281	8.072
p value		0.417	0.968	0.876	0.246	0.382

B-32, Ech008, and Emm4 after Bonferroni correction for multiple tests. All other colonies had no significant linkage between markers. Because of the combined issues of null alleles and linkage disequilibrium, markers B-32 and Ech008 were dropped from further analyses, leaving a final number of eight microsatellite markers. Markers Ech010 and Ech031 at KG and markers Ech030 and RM3 at OH departed significantly from HWE (see Online Resource 2 for linkage disequilibrium and HWE results).

Descriptive statistics

Mean number of alleles per locus (N_A) ranged from two in ECH050 and ECH091 to 11 in ECH030. F_{IS} values for all loci were low and insignificant (mean $F_{IS} = 0.08 \pm 0.09$) (Online Resource 3). Summarising statistics within each colony (Table 2) highlighted that levels of heterozygosity were similar across colonies (mean $H_E = 0.643$, mean $H_O = 0.583$). Reductions between expected and observed heterozygosity were largest for SI and OH, and these were found to have significant F_{IS} values. This is in spite of SI having the largest N_A (7), an indicator of allelic diversity, though N_A may be dependent on the sample size of the colony.

Population genetic analyses

For pairwise F_{ST} analyses, there was a small but significant difference in allele frequencies between TH and all other colonies ($p < 0.001$; Table 3). Hierarchical AMOVA showed significant variation between colonies (3 %, $p < 0.001$) and significant variation within colonies (3.7 %, $p = 0.042$), although genetic variation within individuals was greatest (93.3 %, $p = 0.002$).

Using a Mantel's test, no significant isolation by distance was detected yet linear regression analysis indicated a significant positive relationship between geographic and genetic distance (slope = 0.00005, $R^2 = 0.69$, $p = 0.039$).

When interpreting the *STRUCTURE* analyses, the modal log-likelihood of the data, L(K), yielded higher values at $K = 1$ [L(K) = −2962.34] than all other values of K, suggesting the true value of K was equal to 1 with no population structure detectable (Online Resource 4). When K was set to 2, on average, genotypes from each colony had roughly equal proportion of membership to each of the two clusters (KG = 0.47:0.53; OH = 0.47:0.53; SI = 0.48:0.52; TH = 0.40:0.6), further indicating that colonies cannot be clustered separately based on their genotype frequencies.

Sex-biased dispersal

When comparing male and female genotypes, females gave a higher F_{IS}, negative *mAIc,* and a higher *vAIc* (Table 4). However, F_{ST} values showed no difference between sexes and none of the five tests were found to be significant.

Discussion

DNA quality and non-invasive sampling

The quality of DNA, and thus the reliability of the results generated, depended greatly on the type of tissue used for genetic analyses. However, the need for high-quality DNA must be balanced against the level of impact that some forms of tissue sampling may incur; thus, there has been growing support for the use of shed feathers and other non-invasive sampling methods in genetic studies (Hogan et al. 2008; Johansson et al. 2012; Vili et al. 2013). The majority

of shed feather samples were unsuccessful in amplifying multilocus genotypes and those that did were incomplete, failed to meet marker quality control tests, and could not be included in further analyses. For example, shed feather genotypes had a high rate of null alleles and a significant departure from HWE (HWE is an assumption for the linkage disequilibrium tests; Excoffier and Slatkin 1998), which may explain the increase in the non-random association of alleles between loci for these data compared to plucked feather and blood samples.

Consequently, for chinstrap penguins in our study, the cost and ethical advantages of using non-invasive sampling for microsatellite genotyping and molecular sexing are outweighed by their inability to yield consistently high data quality and reliability. Although shed feathers often have sufficient quality and quantity of DNA for PCR-based sexing or DNA sequencing of high copy number regions (such as mtDNA), they suffer from the same problems as other non-invasive methods (e.g. faecal samples) for microsatellite-based genotyping. This is because of the increased chance of allelic drop out using samples with low- or poor-quality DNA (Taberlet and Luikart 1999). Problems with reliability add to the other drawback of shed feather samples in that they are "anonymous"—not being able to include information about the individual such as age, reproductive status and exact breeding location reduces inferences that more invasive sampling allows. We would not recommend shed feathers to obtain reliable genotypes for microsatellite markers. However, if a PCR product can be obtained, then sequence-based approaches are still feasible (Clucas et al. 2014).

Regional population connectivity

All population analyses agreed that there is high connectivity between King George Island, Orne Harbour, and Signy Island breeding colonies, complementing previous genetic work that found no differentiation within archipelagos (Korczak-Abshire et al. 2012). The limited genetic structure detected can be explained by: (1) large effective population sizes of chinstrap colonies; (2) a recent population expansion; and (3) population intermixing through non-breeding dispersal. Large population sizes would prevent differentiation in neutral markers because of greater stability of genotype frequencies, as has been found for Adélie (Roeder et al. 2001) and Magellanic penguins (Bouzat et al. 2009). A recent regional population expansion of chinstrap penguins breeding in the South Shetland Islands in the early twentieth century has taken place, caused by competitive release following whale and seal harvesting and favourable environmental conditions (Emslie et al. 1998). This offers a further explanation for the limited structuring found in our study, since the relatively short time scale over which this population expansion took place would potentially limit the degree of genetic differentiation. Given the high degree of mate (Ancel et al. 2013) and site (Williams 1995) fidelity recorded for chinstrap penguins, natal dispersal and dispersal of non-breeding individuals could be driving the pattern of connectivity we detected. The distribution of bird species in the non-breeding season has an important influence on population structure (Friesen et al. 2007), and chinstrap penguins are known to disperse thousands of kilometres from their breeding grounds (Wilson et al. 1998; Trivelpiece et al. 2007; Biuw et al. 2009).

Analyses based on pairwise F_{ST} suggested that the South Sandwich Islands (Thule) may be weakly differentiated from the other Scotia Sea archipelagos sampled: the South Orkneys (Signy), South Shetlands (King George), and the WAP (Orne Harbour). Separation between the South Sandwich Islands and the other archipelagos was also found by Clucas et al. (2014) based on the mitochondrial hypervariable region-1 (HVR-1), using individuals sampled from Zavadovski Island in the South Sandwich archipelago, only 350 km from Thule. Yet, the F_{ST} values we report are very low and a lack of strong structure was found using model-based clustering. Whilst this could reflect allele sharing as a result of shared ancestry, it may also suggest long-distance dispersal is frequent enough to sustain a level of connectivity between distant islands that is sufficient to prevent substantial genetic differentiation. This scenario supports tracking data, which recorded chinstrap penguins moving large distances from the South Shetlands towards the South Sandwich Islands (1600 km) (Wilson et al. 1998), and South Orkney Islands (1300 km) (Trivelpiece et al. 2007) and between Bouvet and the South Sandwich Islands (3600 km) (Biuw et al. 2009).

Patterns of genetic variation within colonies

All populations showed similar and substantial levels of genetic diversity based on H_O and N_A results. This result is somewhat unexpected when examined in the context of changes in chinstrap penguin population dynamics over the past 100 years. The over-harvesting and near extinction of krill-eating seals and baleen whales between the late eighteenth century and twentieth century, as well as favourable climatic conditions provided a surplus of krill available to penguins (Emslie and Patterson 2007) and is thought to have led to increases in chinstrap penguin populations in the early to mid-twentieth century (Laws 1977; Fraser et al. 1992; Trivelpiece et al. 2011). However, recent increases in temperature and reductions in sea ice over the past 30 years are now thought to be impacting marine ecosystems in the Scotia Arc/Antarctic Peninsula region, leading to declines in both krill (Atkinson et al.

2004; Meredith and King 2005) and chinstrap penguin populations (Trivelpiece et al. 2011; Lynch et al. 2012). If the expansion of chinstrap penguins in the early twentieth century was driven by only a subset of individuals, a founder event and reduced genetic diversity within regions might have been expected, especially in a small breeding colony like Orne Harbour, near the southern range limit of this species (Poncet and Poncet 1987). However, as similarly high levels of genetic diversity were found across regions, we suggest there has been enough gene flow from other areas, and from a large number of individuals, to sustain diversity even in smaller colonies.

Significant differences between expected and observed heterozygosity are indicative of non-random mating, as was detected for Signy and Orne Harbour colonies. This finding may result from differences in population sizes both within regions and at the specific breeding sites sampled in our study. Orne Harbour has a population of around 500 pairs (Aguirre and Acero 1995), and the total populations of chinstrap penguins in the WAP are estimated at 71,970 pairs (Trivelpiece and Trivelpiece 2013). The breeding population studied on Signy Island is around 19,500 pairs (M Dunn unpubl. data), and the total populations of chinstrap penguins in South Orkney Islands are estimated at 405,600 pairs (Trivelpiece and Trivelpiece 2013). Thus, the total breeding population of chinstrap penguins at these two sites and regions is smaller than those found at Admiralty Bay, King George Island (around 2000 pairs; Sander et al. 2007; Trivelpiece et al. 2011) in the South Shetland Islands (986,440 pairs; Trivelpiece and Trivelpiece 2013) and Thule Island (45,000 pairs; Convey et al. 1997), in the South Sandwich Islands (1,285,000 pairs; (Trivelpiece and Trivelpiece 2013). The low but significant level of inbreeding signatures found at Orne Harbour and Signy Island thus may be a function of their relatively smaller population sizes.

Sex-biased dispersal

This study is the first to compare the connectivity of male and female chinstrap penguins throughout the Scotia Arc. Although we report test values consistent with female bias dispersal (higher F_{IS}, *vAIc* and a negative *mAIc* in females), none of these tests were significant, and so we conclude from our data, a lack of evidence of sex-biased dispersal in these penguin colonies. This result differs from previous studies that have discussed the potential of male-biased natal dispersal in penguins (Clarke et al. 1997; Bouzat et al. 2009). Long-distance dispersal, as is the case here, is more likely to be aimed at colonising empty patches than to avoid inbreeding, in which case dispersal is less likely to be sex specific (Goudet et al. 2002).

Conclusion

Overall, our results suggest no strong barriers to dispersal among populations of chinstrap penguins in the Western Antarctic Peninsula and surrounding archipelagos of the Scotia Arc. Colonies of chinstrap penguins across this region represent a single, diverse population, and this high connectivity could be important in maintaining the persistence of smaller colonies such as Orne Harbour. We suggest that conservation efforts should be focused at preserving this status by the formation of dispersal corridors among archipelagos. As a step in this direction, the Convention for the Conservation of Antarctic Marine Living Resources (CCAMLR) designated the first high seas marine protected area off the southern shelf of the South Orkney Islands in 2009, in the hope that this will be the first in a network of protected areas to preserve the biodiversity of the Scotia Sea (CCAMLR 2009). This region may well represent a migration corridor between the Scotia Arc populations. CCAMLR and its member states should prioritise the development of this network. For chinstrap penguins to benefit, it should encompass both breeding and foraging habitats and should allow for their high dispersal ability across the entire region.

Acknowledgments The authors would like to thank Aileen Adam and Elizabeth Kilbride from the University of Glasgow for their help in the laboratory, especially during initial DNA screening and marker optimisation processes. Sampling was conducted under permits from the Government of South Georgia and the South Sandwich Islands, the UK Foreign and Commonwealth Office and the US National Science Foundation. Fieldwork on the South Sandwich Islands would not have been possible without the knowledge and assistance of Jérôme Poncet and the crew of the Golden Fleece. Fieldwork in the Antarctic Peninsula was carried out with the assistance of Quark Expeditions Ltd, and W. Trivelpiece and the US Antarctic Marine Living Resource Program provided logistical support for sampling at King George Island. This work was funded by the Charities Advisory Trust, the Holly Hill Charitable Trust, and the Zoological Society of London (TH).

References

Aguirre CA, Acero JM (1995) Distribution and abundance of birds in the Errera Channel, Antarctic Peninsula, during the 1992/93 breeding season. Mar Ornithol 23:129–134

Ahmed S, Hart T, Dawson DA, Horsburgh GJ, Trathan PN, Rogers AD (2009) Isolation and characterization of macaroni penguin (*Eudyptes chrysolophus*) microsatellite loci and their utility in other penguin species (Spheniscidae, AVES). Mol Ecol Resour 9:1530–1535

Ainley D, Russell J, Jenouvrier S, Woehler E, Lyver PO, Fraser WR, Kooyman GL (2010) Antarctic penguin response to habitat change as Earth's troposphere reaches 2 degrees C above preindustrial levels. Ecol Monogr 80:49–66

Ancel A, Beaulieu M, Gilbert C (2013) The different breeding strategies of penguins: a review. C R Biol 336:1–12

Atkinson A, Siegel V, Pakhomov E, Rothery P (2004) Long-term decline in krill stock and increase in salps within the Southern Ocean. Nature 432:100–103

Barbosa A, Benzal J, De Leon A, Moreno J (2012) Population decline of chinstrap penguins (*Pygoscelis antarctica*) on Deception Island, South Shetlands, Antarctica. Polar Biol 35:1453–1457

Billing TM, Guay PJ, Peucker AJ, Mulder RA (2006) PRIMER NOTE: isolation and characterization of polymorphic microsatellite loci for the study of paternity and population structure in the little penguin *Eudyptula minor*. Mol Ecol Notes 7:425–427

Biuw M, Lydersen C, Nico De Bruyn PJ, Arriola A, Hofmeyr GGJ, Kritzinger P, Kovacs KM (2009) Long-range migration of a chinstrap penguin from Bouvetøya to Montagu Island, South Sandwich Islands. Antarct Sci 22:157–162

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

Bouzat JL, Walker BG, Boersma PD (2009) Regional genetic structure in the Magellanic penguin (*Spheniscus magellanicus*) suggests metapopulation dynamics. Auk 126:326–334

CCAMLR (2009) Protection of the South Orkney Islands southern shelf. Conservation measure 91-03. Commission report

Clarke AL, Saether BE, Roskaft E (1997) Sex biases in avian dispersal: a reappraisal. Oikos 79:429–438

Clucas GV, Dunn MJ, Dyke G, Emslie SD, Levy H, Naveen R, Polito MJ, Pybus OG, Rogers AD, Hart T (2014) A reversal of fortunes: climate change 'winners' and 'losers' in Antarctic Peninsula penguins. Sci Rep 4:5024

Convey P, Morton A, Poncet J (1997) Survey of marine birds and mammals of the South Sandwich Islands. Pol Rec 35:107–124

Croxall JP (1987) The status and conservation of Antarctic seals and seabirds—a review. Environ Int 13:55–70

Croxall JP, Trathan PN, Murphy EJ (2002) Environmental change and Antarctic seabird populations. Science 297:1510–1514

Dieringer D, Schlotterer C (2003) MICROSATELLITE ANALYSER (MSA): a platform independent analysis tool for large microsatellite data sets. Mol Ecol Notes 3:167–169

Ducklow HW, Baker K, Martinson DG, Quetin LB, Ross RM, Smith RC, Stammerjohn SE, Vernet M, Fraser W (2007) Marine pelagic ecosystems: the West Antarctic Peninsula. Philos Trans R Soc B Biol Sci 362:67–94

Earl DA, Vonholdt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359–361

Emslie SD, Patterson WP (2007) Abrupt recent shift in delta C-13 and delta N-15 values in Adelie penguin eggshell in Antarctica. Proc Natl Acad Sci USA 104:11666–11669

Emslie SD, Fraser W, Smith RC, Walker W (1998) Abandoned penguin colonies and environmental change in the Palmer Station area, Anvers Island, Antarctic Peninsula. Antarct Sci 10:257–268

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

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10:564–567

Excoffier L, Slatkin M (1998) Incorporating genotypes of relatives into a test of linkage disequilibrium. Am J Hum Genet 62:171–180

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes—application to human mitochondrial-DNA restriction data. Genetics 131:479–491

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

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

Fraser WR, Trivelpiece WZ, Ainley DG, Trivelpiece SG (1992) Increases in Antarctic penguin populations—reduced competition with whales or a loss of sea ice due to environmental warming. Polar Biol 11:525–531

Fridolfsson AK, Ellegren H (1999) A simple and universal method for molecular sexing of non-ratite birds. J Avian Biol 30:116–121

Friesen VL, Burg TM, Mccoy KD (2007) Mechanisms of population differentiation in seabirds. Mol Ecol 16:1765–1785

Goudet J (2002) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). http://www2.unil.ch/popgen/softwares/fstat.htm. Accessed 21/01/15

Goudet J, Perrin N, Waser P (2002) Tests for sex-biased dispersal using bi-parentally inherited genetic markers. Mol Ecol 11:1103–1114

Greenwood PJ (1980) Mating systems, philopatry and dispersal in birds and mammals. Anim Behav 28:1140–1162

Hinke JT, Salwicka K, Trivelpiece SG, Watters GM, Trivelpiece WZ (2007) Divergent responses of Pygoscelis penguins reveal a common environmental driver. Oecologia 153:845–855

Hogan FE, Cooke R, Burridge CP, Normano JA (2008) Optimizing the use of shed feathers for genetic analysis. Mol Ecol Resour 8:561–567

Jenouvrier S, Caswell H, Barbraud C, Holland M, Str Ve J, Weimerskirch H (2009) From the cover: demographic models and IPCC climate projections predict the decline of an emperor penguin population. Proc Natl Acad Sci USA 106:1844–1847

Johansson MP, Mcmahon BJ, Hoglund J, Segelbacher G (2012) Amplification success of multilocus genotypes from feathers found in the field compared with feathers obtained from shot birds. Ibis 154:15–20

Korczak-Abshire M, Chwedorzewska KJ, Wasowicz P, Bednarek PT (2012) Genetic structure of declining chinstrap penguin (*Pygoscelis antarcticus*) populations from South Shetland Islands (Antarctica). Polar Biol 35:1681–1689

Kramer AM, Dennis B, Liebhold AM, Drake JM (2009) The evidence for allee effects. Popul Ecol 51:341–354

Laws RM (1977) Seals and whales of the Southern Ocean. Philos Trans R Soc B Biol Sci 279:81–96

Lemaho Y, Karmann H, Briot D, Handrich Y, Robin JP, Mioskowski E, Cherel Y, Farni J (1992) Stress in birds due to routine handling and a technique to avoid it. Am J Physiol 263:R775–R781

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

Meredith MP, King JC (2005) Rapid climate change in the ocean west of the Antarctic Peninsula during the second half of the 20th century. Geophys Res Lett 32:L19604

Naveen R, Lynch HJ, Forrest S, Mueller T, Polito M (2012) First direct, site-wide penguin survey at Deception Island, Antarctica, suggests significant declines in breeding chinstrap penguins. Polar Biol 35:1879–1888

Overeem RL, Peucker AJ, Austin CM, Dann P, Burridge CP (2008) Contrasting genetic structuring between colonies of the World's smallest penguin, *Eudyptula minor* (Aves: Spheniscidae). Conserv Genet 9:893–905

Poncet S, Poncet J (1987) Censuses of Penguin populations of the Antarctic Peninsula, 1983–87. Br Antarct Surv B 77:109–129

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

Rice WR (1989) Analyzing tables of statistical tests. Evolution 43:223–225

Roeder AD, Marshall RK, Mitchelson AJ, Visagathilagar T, Ritchie PA, Love DR, Pakai TJ, Mcpartlan HC, Murray ND, Robinson NA, Kerry KR, Lambert DM (2001) Gene flow on the ice: genetic differentiation among Adelie penguin colonies around Antarctica. Mol Ecol 10:1645–1656

Roeder AD, Ritchie PA, Lambert DM (2002) New DNA markers for penguins. Conserv Genet 3:341–344

Sander M, Balbao TC, Polito MJ, Costa ES, Carneiro APB (2007) Recent decrease in chinstrap penguin (*Pygoscelis antarctica*) populations at two of Admiralty Bay's islets on King George Island, South Shetland Islands, Antarctica. Polar Biol 30:659–661

Schlosser JA, Dubach JM, Garner TWJ, Araya B, Bernal M, Simeone A, Smith KA, Wallace RS (2009) Evidence for gene flow differs from observed dispersal patterns in the Humboldt penguin, *Spheniscus humboldti*. Conserv Genet 10:839–849

Scribner KT, Petersen MR, Fields RL, Talbot SL, Pearce JM, Chesser RK (2001) Sex-biased gene flow in spectacled eiders (anatidae): inferences from molecular markers with contrasting modes of inheritance. Evolution 55:2105–2115

Shuford WD, Spear LB (1988) Surveys of breeding chinstrap penguins in the South Shetland Islands, Antarctica. Br Antarct Surv B 81:19–30

Taberlet P, Luikart G (1999) Non-invasive genetic sampling and individual identification. Biol J Linn Soc 68:41–55

Trivelpiece W, Trivelpiece S (2013) Chinstrap Penguin (*Pygoscelis antarctica*). In: Borboroglu PG, Boersma PD (eds) Penguins: Natural history and conservation. University of Washington Press, Seattle, pp 59–72

Trivelpiece WZ, Buckelew S, Reiss C, Trivelpiece SG (2007) The winter distribution of chinstrap penguins from two breeding sites in the South Shetland Islands of Antarctica. Polar Biol 30:1231–1237

Trivelpiece WZ, Hinke JT, Miller AK, Reiss CS, Trivelpiece SG, Watters GM (2011) Variability in krill biomass links harvesting and climate warming to penguin population changes in Antarctica. Proc Natl Acad Sci USA 108:7625–7628

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes 4:535–538

Vili N, Nemeshazi E, Kovacs S, Horvath M, Kalmar L, Szabo K (2013) Factors affecting DNA quality in feathers used for non-invasive sampling. J Ornithol 154:587–595

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

Williams TD (1995) The Penguins. Oxford University Press, Oxford

Wilson RP (1997) A method for restraining penguins. Mar Ornithol 25:72–73

Wilson RP, Culik BM, Kosiorek P, Adelung D (1998) The over-winter movements of a chinstrap penguin (*Pygoscelis antarctica*). Polar Rec 34:107–112

Bibliography

Abbott CL, Double MC (2003) Genetic structure, conservation genetics and evidence of speciation by range expansion in shy and white-capped albatrosses. *Molecular Ecology* **12**, 2953-2962.

Ainley D, Russell J, Jenouvrier S*, et al.* (2010) Antarctic penguin response to habitat change as Earth's troposphere reaches 2° C above preindustrial levels. *Ecological Monographs* **80**, 49-66.

Ainley DG, LeResche RE, Sladen WJ (1983) *Breeding biology of the Adelie penguin* University of California Press, Berkeley.

Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. *Nature Reviews Genetics* **11**, 697-709.

Ancel A, Cristofari R, Fretwell PT*, et al.* (2014) Emperors in hiding: When ice-breakers and satellites complement each other in Antarctic exploration. *PLoS One* **9**, e100404.

Anderson JB, Shipp SS, Lowe AL, Wellner JS, Mosola AB (2002) The Antarctic Ice Sheet during the Last Glacial Maximum and its subsequent retreat history: a review. *Quaternary Science Reviews* **21**, 49-70.

Anderson R, Ali S, Bradtmiller L*, et al.* (2009) Wind-driven upwelling in the Southern Ocean and the deglacial rise in atmospheric CO2. *Science* **323**, 1443-1448.

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics* **17**, 81-92.

Arrigo KR, Weiss AM, Smith WO (1998) Physical forcing of phytoplankton dynamics in the southwestern Ross Sea. *Journal of Geophysical Research: Oceans* **103**, 1007-1021.

Atkinson A, Siegel V, Pakhomov E, Rothery P (2004) Long-term decline in krill stock and increase in salps within the Southern Ocean. *Nature* **432**, 100-103.

Avise JC, Arnold J, Ball RM*, et al.* (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual review of ecology and systematics* **18**, 489-522.

Avise JC, Nelson WS, Bowen BW, Walker D (2000) Phylogeography of colonially nesting seabirds, with special reference to global matrilineal patterns in the sooty tern (Sterna fuscata). *Molecular Ecology* **9**, 1783-1792.

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

Baker A, Marshall H (1997) Mitochondrial control region sequences as tools for understanding evolution. In: *Avian molecular evolution and systematics* (ed. Mindell DP), pp. 51-82. Academic Press, San Diego.

Baker AJ, Pereira SL, Haddrath OP, Edge KA (2006) Multiple gene evidence for expansion of extant penguins out of Antarctica due to global cooling. *Proceedings of the Royal Society B: Biological Sciences* **273**, 11-17.

Ballard G, Toniolo V, Ainley DG*, et al.* (2010) Responding to climate change: Adélie penguins confront astronomical and ocean boundaries. *Ecology* **91**, 2056-2069.

Barbara L, Crosta X, Massé G, Ther O (2010) Deglacial environments in eastern Prydz Bay, East Antarctica. *Quaternary Science Reviews* **29**, 2731-2740.

Barbraud C, Gavrilo M, Mizin Y, Weimerskirch H (2011) Comparison of emperor penguin declines between Pointe Géologie and Haswell Island over the past 50 years. *Antarctic Science* **23**, 461-468.

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

Barnett R, Yamaguchi N, Shapiro B*, et al.* (2014) Revealing the maternal demographic history of Panthera leo using ancient DNA and a spatially explicit genealogical analysis. *BMC evolutionary biology* **14**, 70.

Barnosky AD, Matzke N, Tomiya S*, et al.* (2011) Has the Earth's sixth mass extinction already arrived? *Nature* **471**, 51-57.

Baroni C, Orombelli G (1994) Abandoned penguin rookeries as Holocene paleoclimatic indicators in Antarctica. *Geology* **22**, 23.

Baylis AM, Orben RA, Pistorius P*, et al.* (2015) Winter foraging site fidelity of king penguins breeding at the Falkland Islands. *Marine biology* **162**, 99-110.

Beerli P (2001) How to use MIGRATE or why are Markov chain Monte Carlo programs difficult to use. In: *Population genetics for animal conservation* (eds. Bertorelle G, Bruford MW, Hauffe HC, Annapaolo R, Vernesi C), pp. 42-79. Cambridge University Press, Cambridge.

Beerli P (2004) Effect of unsampled populations on the estimation of population sizes and migration rates between sampled populations. *Molecular Ecology* **13**, 827-836.

Beerli P, Palczewski M (2010) Unified framework to evaluate panmixia and migration direction among multiple sampling locations. *Genetics* **185**, 313-326.

Benestan LM, Ferchaud AL, Hohenlohe PA*, et al.* (2016) Conservation genomics of natural and managed populations: building a conceptual and practical framework. *Molecular Ecology*.

Bernardi G (2000) Barriers to gene flow in Embiotoca jacksoni, a marine fish lacking a pelagic larval stage. *Evolution* **54**, 226-237.

Bertelli S, Giannini NP (2005) A phylogeny of extant penguins (Aves: Sphenisciformes) combining morphology and mitochondrial sequences. *Cladistics* **21**, 209-239.

Biuw M, Lydersen C, Nico de Bruyn P*, et al.* (2010) Long-range migration of a chinstrap penguin from Bouvetøya to Montagu Island, South Sandwich Islands. *Antarctic Science* **22**, 157-162.

Black CE (2016) A comprehensive review of the phenology of Pygoscelis penguins. *Polar Biology* **39**, 405-432.

Boileau MG, Hebert PD, Schwartz SS (1992) Non-equilibrium gene frequency divergence: persistent founder effects in natural populations. *Journal of Evolutionary Biology* **5**, 25-39.

Borboroglu PB, Boersma PD (2013) *Penguins: Natural History and Conservation* University of Washington Press, Washington.

Bost C, Delord K, Barbraud C*, et al.* (2013) King Penguin. In: *Penguins Natural History and Conservation* (eds. Borboroglu PB, Boersma PD), pp. 7 - 21. University of Washington Press, Seattle.

Bost C-A, Charrassin J-B, Clerquin Y, Ropert-Coudert Y, Le Maho Y (2004) Exploitation of distant marginal ice zones by king penguins during winter. *Marine Ecology Progress Series* **283**, 293-297.

Bost C-A, Cotté C, Bailleul F*, et al.* (2009) The importance of oceanographic fronts to marine birds and mammals of the southern oceans. *Journal of Marine Systems* **78**, 363-376.

Bost CA, Cotté C, Terray P*, et al.* (2015) Large-scale climatic anomalies affect marine predator foraging behaviour and demography. *Nature communications* **6**, 8820.

Bouckaert R, Heled J, Kühnert D*, et al.* (2014) BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Comput Biol* **10**, e1003537.

Bouckaert RR (2010) DensiTree: making sense of sets of phylogenetic trees. *Bioinformatics* **26**, 1372-1373.

Brambati A, Melis R, Quaia T, Salvi G (2002) Late Quaternary climatic changes in the Ross Sea area, Antarctica.

Brown RM, Nichols RA, Faulkes CG*, et al.* (2010) Range expansion and hybridization in Round Island petrels (Pterodroma spp.): evidence from microsatellite genotypes. *Molecular Ecology* **19**, 3157-3170.

Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, RoyChoudhury A (2012) Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution*, mss086.

Budd GM (1961) The biotopes of emperor penguin rookeries. *Emu* **61**, 171-189.

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

Caliński T, Harabasz J (1974) A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods* **3**, 1-27.

Campos PF, Willerslev E, Sher A*, et al.* (2010) Ancient DNA analyses exclude humans as the driving force behind late Pleistocene musk ox (Ovibos moschatus) population dynamics. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 5675-5680.

Carstens BC, Richards CL (2007) Integrating coalescent and ecological niche modeling in comparative phylogeography. *Evolution* **61**, 1439-1454.

Carvajal-Rodriguez A, de Uña-Alvarez J (2011) Assessing significance in high-throughput experiments by sequential goodness of fit and q-value estimation. *PLoS One* **6**, e24700.

Catchen J, Amores A, Hohenlohe PA, Cresko WA, Postlethwait JH (2011) Stacks: building and genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics* **1**, 171 - 182.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. *Molecular Ecology* **22**, 3124-3140.

Charrassin J-B, Bost C-A (2001) Utilisation of the oceanic habitat by king penguins over the annual cycle. *Marine Ecology Progress Series* **221**, 285-298.

Cimino MA, Lynch HJ, Saba VS, Oliver MJ (2016) Projected asymmetric response of Adelie penguins to Antarctic climate change. *Scientific Reports* **6**, 28785.

Clark PU, Mix AC (2002) Ice sheets and sea level of the Last Glacial Maximum. *Quaternary Science Reviews* **21**, 1-7.

Clarke J, Emmerson LM, Otahal P (2006) Environmental conditions and life history constraints determine foraging range in breeding Adélie penguins. *Marine Ecology Progress Series* **310**, 247-261.

Clarke J, Kerry K, Fowler C*, et al.* (2003) Post-fledging and winter migration of Adélie penguins *Pygoscelis adeliae* in the Mawson region of East Antarctica. *Marine Ecology Progress Series* **248**, 267-278.

Clarke JA, Ksepka DT, Stucchi M*, et al.* (2007) Paleogene equatorial penguins challenge the proposed relationship between biogeography, diversity, and Cenozoic climate change. *Proceedings of the National Academy of Sciences* **104**, 11545-11550.

Clausen A, Pütz K (2003) Winter diet and foraging range of gentoo penguins (Pygoscelis papua) from Kidney Cove, Falkland Islands. *Polar Biology* **26**, 32-40.

Clements JF, Schulenberg TS, Iliff MJ*, et al.* (2016) The eBird/Clements checklist of birds of the world: v2016.

Clucas GV, Dunn MJ, Dyke G*, et al.* (2014) A reversal of fortunes: climate change 'winners' and 'losers' in Antarctic Peninsula penguins. *Scientific Reports* **4**, 5024.

Clucas GV, Younger JL, Kao D*, et al.* (2016) Dispersal in the sub-Antarctic: king penguins show remarkably little population genetic differentiation across their range. *BMC evolutionary biology* **16**, 211.

Collins M, Knutti R, Arblaster J*, et al.* (2013) Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. *Long-term Clim. Chang. Proj. Commitments Irreversibility, Cambridge Univ. Press. Cambridge, UK, New York*.

Comiso J, McClain C, Sullivan C, Ryan J, Leonard C (1993) Coastal Zone Color Scanner pigment concentrations in the Southern Ocean and relationships to geophysical surface features. *Journal of Geophysical Research: Oceans* **98**, 2419-2451.

Conroy J, White M (1973) The breeding status of the king penguin (Aptenodytes patagonica). *Bulletin of the British Antarctic Survey*, 31-40.

Constable AJ, Melbourne-Thomas J, Corney SP*, et al.* (2014) Climate change and Southern Ocean ecosystems I: How changes in physical habitats directly affect marine biota. *Global Change Biology* **20**, 3004-3025.

Coulson JC (2001) Colonial Breeding in Seabirds. In: *Biology of marine birds* (eds. Schreiber EA, Burger J), pp. 87 - 113. CRC Press, Boca Raton.

Cristofari R, Bertorelle G, Ancel A*, et al.* (2016) Full circumpolar migration ensures evolutionary unity in the Emperor penguin. *Nature communications* **7**, 11842.

Croxall JP, Trathan PN, Murphy EJ (2002) Environmental change and Antarctic seabird populations. *Science* **297**, 1510-1514.

D'Amen M, Zimmermann NE, Pearman PB (2013) Conservation of phylogeographic lineages under climate change. *Global ecology and biogeography* **22**, 93-104.

Danecek P, Auton A, Abecasis G*, et al.* (2011) The variant call format and VCFtools. *Bioinformatics* **27**, 2156-2158.

de Bruyn M, Pinsky ML, Hall B*, et al.* (2014) Rapid increase in southern elephant seal genetic diversity after a founder event. *Proceedings of the Royal Society B: Biological Sciences* **281**, 20133078.

de Dinechin M, Dobson FS, Zehtindjiev P*, et al.* (2012) The biogeography of Gentoo Penguins (*Pygoscelis papua*). *Canadian Journal of Zoology* **90**, 352-360.

de Dinechin M, Ottvall R, Quillfeldt P, Jouventin P (2009) Speciation chronology of rockhopper penguins inferred from molecular, geological and palaeoceanograhic data. *Journal of Biogeography* **36**, 693-702.

Dehnhard N, Poisbleau M, Demongin L, Ludynia K, Quillfeldt P (2014) High juvenile annual survival probabilities in Southern Rockhopper Penguins Eudyptes chrysocome are independent of individual fledging traits. *Ibis* **156**, 548-560.

Delord K, Barbraud C, Weimerskirch H (2004) Long-term trends in the population size of king penguins at Crozet archipelago: environmental variability and density dependence? *Polar Biology* **27**, 793-800.

Denis D, Crosta X, Schmidt S*, et al.* (2009a) Holocene glacier and deep water dynamics, Adélie Land region, East Antarctica. *Quaternary Science Reviews* **28**, 1291-1303.

Denis D, Crosta X, Schmidt S*, et al.* (2009b) Holocene productivity changes off Adélie Land (East Antarctica). *Paleoceanography* **24**.

Dewasmes G, Le Maho Y, Cornet A, Groscolas R (1980) Resting metabolic rate and cost of locomotion in long-term fasting emperor penguins. *Journal of Applied Physiology* **49**, 888-896.

Ding L, Wiener H, Abebe T*, et al.* (2011a) Comparison of measures of marker informativeness for ancestry and admixture mapping. *Bmc Genomics* **12**, 622.

Ding Q, Steig EJ, Battisti DS, Küttel M (2011b) Winter warming in West Antarctica caused by central tropical Pacific warming. *Nature Geoscience* **4**, 398-403.

Dobzhansky T (1937) *Genetics and the Origin of Species* Columbia University Press, New York.

Doerr VA, Barrett T, Doerr ED (2011) Connectivity, dispersal behaviour and conservation under climate change: a response to Hodgson et al. *Journal of Applied Ecology* **48**, 143-147.

Domack E, O'Brien P, Harris P*, et al.* (1998) Late Quaternary sediment facies in Prydz Bay, East Antarctica and their relationship to glacial advance onto the continental shelf. *Antarctic Science* **10**, 236-246.

Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution* **22**, 1185-1192.

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* **29**, 1969-1973.

Ducklow HW, Baker K, Martinson DG*, et al.* (2007) Marine pelagic ecosystems: the west Antarctic Peninsula. *Philosophical Transactions of the Royal Society B: Biological Sciences* **362**, 67-94.

Dunn M, Silk J, Trathan P (2011) Post-breeding dispersal of Adélie penguins (Pygoscelis adeliae) nesting at Signy Island, South Orkney Islands. *Polar Biology* **34**, 205-214.

Earl DA (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation genetics resources* **4**, 359-361.

Ehrlich PR, Raven PH (1969) Differentiation of Population: Gene Flow Seems to be Less Important Than the Neo-darwinians Thought.

Emslie SD, Patterson WP (2007) Abrupt recent shift in delta 13C and delta 15N values in Adelie penguin eggshell in Antarctica. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 11666-11669.

Enticott J (1986) Distribution of penguins at sea in the Southeastern Atlantic and Southwestern Indian Oceans. *Cormorant* **13**, 118-142.

Etter PD, Bassham S, Hohenlohe PA, Johnson EA, Cresko WA (2011) SNP discovery and genotyping for evolutionary genetics using RAD sequencing. In: *Molecular methods for evolutionary genetics* (eds. Orgogozo V, Rockman MV), pp. 157-178. Springer, New York.

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

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**, 564-567.

Faubet P, Waples RS, Gaggiotti OE (2007) Evaluating the performance of a multilocus Bayesian method for the estimation of migration rates. *Molecular Ecology* **16**, 1149-1166.

Feder JL, Egan SP, Nosil P (2012) The genomics of speciation-with-gene-flow. *Trends in Genetics* **28**, 342-350.

Fisher R (1930) *The genetical theory of natural selection* Clarendon, Oxford.

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* **180**, 977-993.

Forcada J, Trathan PN (2009) Penguin responses to climate change in the Southern Ocean. *Global Change Biology* **15**, 1618-1630.

Forster J-R (1781) *Indische Zoologie oder systematische Beschreibungen seltener und unbekannter Thiere aus Indien* Johann Jacob Gebauer.

Fraser WR, Hofmann EE (2003) A predator¹s perspective on causal links between climate change, physical forcing and ecosystem response. *Marine Ecology Progress Series* **265**, 1-15.

Freer JJ, Mable BK, Clucas G*, et al.* (2015) Limited genetic differentiation among chinstrap penguin (Pygoscelis antarctica) colonies in the Scotia Arc and Western Antarctic Peninsula. *Polar Biology* **38**, 1493-1502.

Fretwell PT, LaRue MA, Morin P*, et al.* (2012) An emperor penguin population estimate: the first global, synoptic survey of a species from space. *PLoS One* **7**, e33751.

Fretwell PT, Trathan PN (2009) Penguins from space: faecal stains reveal the location of emperor penguin colonies. *Global ecology and biogeography* **18**, 543-552.

Fretwell PT, Trathan PN, Wienecke B, Kooyman GL (2014) Emperor Penguins Breeding on Iceshelves. *PLoS One* **9**, e85285.

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

Friesen VL, Burg TM, McCoy KD (2007a) Mechanisms of population differentiation in seabirds. *Molecular Ecology* **16**, 1765-1785.

Friesen VL, Smith AL, Gomez-Diaz E*, et al.* (2007b) Sympatric speciation by allochrony in a seabird. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 18589-18594.

Funk WC, McKay JK, Hohenlohe PA, Allendorf FW (2012) Harnessing genomics for delineating conservation units. *Trends in Ecology & Evolution* **27**, 489-496.

Gales N, Klages N, Williams R, Woehler E (1990) The diet of the emperor penguin, Aptenodytes forsteri, in Amanda Bay, Princess Elizabeth Land, Antarctica. *Antarctic Science* **2**, 23-28.

Gartshore N, Cooper J, Hunter S (1988) Bird ringing at Marion and Prince Edward Islands, 1982-1987; with an analysis of movements since 1951. *South African Journal of Antractic Research* **18**, 23-29.

Gavryushkina A, Heath TA, Ksepka DT*, et al.* (2016) Bayesian total-evidence dating reveals the recent crown radiation of penguins. *Systematic Biology,* syw060*.*

Gersonde R, Crosta X, Abelmann A, Armand L (2005) 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. *Quaternary Science Reviews* **24**, 869-896.

Gilpin MEH, Ilkka A (1997) *Metapopulation biology: ecology, genetics, and evolution*.

Gómez-Díaz E, González-Solís J, Peinado M (2009) Population structure in a highly pelagic seabird, the Cory's shearwater Calonectris diomedea: an examination of genetics, morphology and ecology. *Marine Ecology Progress Series* **382**, 197-209.

Gonen S, Lowe NR, Cezard T*, et al.* (2014) Linkage maps of the Atlantic salmon (Salmo salar) genome derived from RAD sequencing. *Bmc Genomics* **15**, 1.

Gordon JE, Harkness DD (1992) Magnitude and geographic variation of the radiocarbon content in Antarctic marine life: implications for reservoir corrections in radiocarbon dating. *Quaternary Science Reviews* **11**, 697-708.

Haldane JBS (1930) A mathematical theory of natural and artificial selection.(Part VI, Isolation.) **26**, 220-230.

Hall KJ (2004) Quaternary Glaciations of the Sub-Antarctic Islands. In: *Quaternary Glaciations - Extent and Chronology Part III* (eds. Ehlers J, Gibbard PL), pp. 339 - 345. Elsevier.

Hartl DL, Clark AG (1997) *Principles of population genetics*, Third edn. Sinauer Associates Sunderland, Massachusetts.

Harvey MG, Brumfield RT (2015) Genomic variation in a widespread Neotropical bird (Xenops minutus) reveals divergence, population expansion, and gene flow. *Molecular Phylogenetics and Evolution* **83**, 305-316.

Harvey MG, Smith BT, Glenn TC, Faircloth BC, Brumfield RT (2016) Sequence Capture Versus Restriction Site Associated DNA Sequencing for Shallow Systematics. *Systematic Biology*.

Hasegawa M, Kishino H, Yano T-a (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* **22**, 160-174.

Heled J, Drummond AJ (2008) Bayesian inference of population size history from multiple loci. *BMC evolutionary biology* **8**, 1.

Heupink TH, van den Hoff J, Lambert DM (2012) King penguin population on Macquarie Island recovers ancient DNA diversity after heavy exploitation in historic times. *Biology Letters* **8**, 586-589.

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

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* **58**, 247-276.

Hey J (2006) Recent advances in assessing gene flow between diverging populations and species. *Curr Opin Genet Dev* **16**, 592-596.

Hey J (2010) Isolation with migration models for more than two populations. *Molecular Biology and Evolution* **27**, 905-920.

Hinke JT, Salwicka K, Trivelpiece SG, Watters GM, Trivelpiece WZ (2007) Divergent responses of Pygoscelis penguins reveal a common environmental driver. *Oecologia* **153**, 845-855.

Hodgson DA, Graham AGC, Roberts SJ*, et al.* (2014) Terrestrial and submarine evidence for the extent and timing of the Last Glacial Maximum and the onset of deglaciation on the maritime-Antarctic and sub-Antarctic islands. *Quaternary Science Reviews* **100**, 137-158.

Hoelzel AR (2010) Looking backwards to look forwards: conservation genetics in a changing world. *Conservation Genetics* **11**, 655-660.

Hohenlohe PA, Bassham S, Etter PD*, et al.* (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics* **6**, e1000862.

Holdgate MW, Baker PE (1979) *The South Sandwich Islands: I. General description* British Antarctic Survey.

Holland PR, Kwok R (2012) Wind-driven trends in Antarctic sea-ice drift. *Nature Geoscience* **5**, 872-875.

Holton JR, Dmowska R, Philander SG (1989) *El Niño, La Niña, and the southern oscillation* Academic press.

Inchausti P, Guinet C, Koudil M*, et al.* (2003) Inter-annual variability in the breeding performance of seabirds in relation to oceanographic anomalies that affect the Crozet and the Kerguelen sectors of the Southern Ocean. *Journal of Avian Biology* **34**, 170-176.

IPCC (2013) Climate Change 2013: The Physical Science Basis. In: *Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* eds. Stocker TF, Qin D, Plattner G-K*, et al.*).

IUCN (2016) The IUCN Red List of Threatened Species. Version 2016-1 <http://www.iucnredlist.org>.

Jakobsson M, Edge MD, Rosenberg NA (2013) The relationship between FST and the frequency of the most frequent allele. *Genetics* **193**, 515-528.

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

Jenouvrier S, Caswell H, Barbraud C*, et al.* (2009) Demographic models and IPCC climate projections predict the decline of an emperor penguin population. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 1844-1847.

Jenouvrier S, Holland M, Strœve J*, et al.* (2012) Effects of climate change on an emperor penguin population: analysis of coupled demographic and climate models. *Global Change Biology* **18**, 2756-2770.

Jenouvrier S, Holland M, Stroeve J*, et al.* (2014) Projected continent-wide declines of the emperor penguin under climate change. *Nature Climate Change* **4**, 715-718.

Jenouvrier S, Weimerskirch H, Barbraud C, Park Y-H, Cazelles B (2005) Evidence of a shift in the cyclicity of Antarctic seabird dynamics linked to climate. *Proceedings of the Royal Society of London B: Biological Sciences* **272**, 887-895.

Johnson ML, Gaines MS (1990) Evolution of Dispersal: Theoretical Models and Empirical Tests Using Birds and Mammals. *Annual review of ecology and systematics* **21**, 449-480.

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

Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* **27**, 3070-3071.

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* **11**, 1.

Jones JM, Gille ST, Goosse H*, et al.* (2016) Assessing recent trends in high-latitude Southern Hemisphere surface climate. *Nature Climate Change* **6**, 917-926.

Jouventin P, Aubin T (2002) Acoustic systems are adapted to breeding ecologies: individual recognition in nesting penguins. *Animal Behaviour* **64**, 747-757.

Jouventin P, Capdeville D, Cuenot-Chaillet F, Boiteau C (1994) Exploitation of pelagic resources by a non-flying seabird: satellite tracking of the king penguin throughout the breeding cycle. *Marine Ecology Progress Series* **106**, 11-19.

Jouzel J, Masson-Delmotte V (2007) EPICA Dome C Ice Core 800KYr deuterium data and temperature estimates (ed. Antarctica EPflCi).

Jouzel J, Masson-Delmotte V, Cattani O*, et al.* (2007) Orbital and millennial Antarctic climate variability over the past 800,000 years. *Science* **317**, 793-796.

Kavakiotis I, Triantafyllidis A, Ntelidou D*, et al.* (2015) TRES: Identification of discriminatory and informative SNPs from population genomic data. *Journal of Heredity*, esv044.

Kerr KC, Dove CJ (2013) Delimiting shades of gray: phylogeography of the Northern Fulmar, Fulmarus glacialis. *Ecology and evolution* **3**, 1915-1930.

Kimura M, Weiss GH (1964) The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics* **49**, 561.

Kirin M, McQuillan R, Franklin CS*, et al.* (2010) Genomic runs of homozygosity record population history and consanguinity. *PLoS One* **5**, e13996.

Kirkwood R, Robertson G (1997) Seasonal change in the foraging ecology of emperor penguins on the Mawson Coast, Antarctica. *Marine Ecology Progress Series* **156**, 205-223.

Kocher TD, Thomas WK, Meyer A*, et al.* (1989) Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proceedings of the National Academy of Sciences* **86**, 6196-6200.

Kohfeld KE, Le Quéré C, Harrison SP, Anderson RF (2005) Role of marine biology in glacial-interglacial CO2 cycles. *Science* **308**, 74-78.

Kooyman G, Hunke E, Ackley S, Van Dam R, Robertson G (2000) Moult of the emperor penguin: travel, location, and habitat selection. *Marine ecology. Progress series* **204**, 269-277.

Kooyman G, Siniff D, Stirling I, Bengtson J (2004) Moult habitat, pre-and post-moult diet and post-moult travel of Ross Sea emperor penguins. *Marine Ecology Progress Series* **267**, 281-290.

Kooyman GL, Kooyman TG, Horning M, Kooyman CA (1996) Penguin dispersal after fledging. *Nature* **383**, 397.

Ksepka DT, Bertelli S, Giannini NP (2006) The phylogeny of the living and fossil Sphenisciformes (penguins). *Cladistics* **22**, 412-441.

Lambert DM, Ritchie PA, Millar CD*, et al.* (2002) Rates of evolution in ancient DNA from Adélie penguins. *Science* **295**, 2270-2273.

Landry P-A, Koskinen MT, Primmer CR (2002) Deriving Evolutionary Relationships Among Populations Using Microsatellites and (δμ) 2: All Loci Are Equal, but Some Are More Equal Than Others…. *Genetics* **161**, 1339-1347.

LaRue MA, Ainley D, Swanson M*, et al.* (2013) Climate change winners: receding ice fields facilitate colony expansion and altered dynamics in an Adelie penguin metopopulation. *PLoS One* **8**, e60568.

LaRue MA, Kooyman G, Lynch HJ, Fretwell P (2015) Emigration in emperor penguins: implications for interpretation of long-term studies. *Ecography*.

Le Bohec C, Durant JM, Gauthier-Clerc M*, et al.* (2008) King penguin population threatened by Southern Ocean warming. *Proceedings of the National Academy of Sciences* **105**, 2493-2497.

Le Maho Y, Delclitte P, Chatonnet J (1976) Thermoregulation in fasting emperor penguins under natural conditions. *American Journal of Physiology--Legacy Content* **231**, 913-922.

Le Maho Y, Karmann H, Briot D*, et al.* (1992) Stress in birds due to routine handling and a technique to avoid it. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* **263**, R775-R781.

Leaché AD, Banbury BL, Felsenstein J, de Oca AN-M, Stamatakis A (2015) Short tree, long tree, right tree, wrong tree: new acquisition bias corrections for inferring SNP phylogenies. *Systematic Biology*, 1032-1047.

Levin SA, Cohen D, Hastings A (1984) Dispersal strategies in patchy environments. *Theoretical population biology* **26**, 165-191.

Levy H, Clucas GV, Rogers AD*, et al.* (2016) Population structure and phylogeography of the Gentoo Penguin (Pygoscelis papua) across the Scotia Arc. *Ecology and evolution* **6**, 1834.

Lewis SL, Maslin MA (2015) Defining the anthropocene. *Nature* **519**, 171-180.

Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*.

Li M-H, Strandén I, Tiirikka T, Sevón-Aimonen M-L, Kantanen J (2011) A comparison of approaches to estimate the inbreeding coefficient and pairwise relatedness using genomic and pedigree data in a sheep population. *PLoS One* **6**, e26256.

Licht KJ, Andrews JT (2002) The 14 C record of Late Pleistocene ice advance and retreat in the central Ross Sea, Antarctica. *Arctic, Antarctic, and Alpine Research*, 324-333.

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

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

Lishman G (1985) The comparative breeding biology of Adélie and chinstrap penguins Pygoscelis adeliae and P. antarctica at Signy Island, South Orkney Islands. *Ibis* **127**, 84-99.

Locke DP, Hillier LW, Warren WC*, et al.* (2011) Comparative and demographic analysis of orang-utan genomes. *Nature* **469**, 529-533.

Loeb V, Siegel V, Holm-Hansen O*, et al.* (1997) Effects of sea-ice extent and krill or salp dominance on the Antarctic food web. *Nature* **387**, 897-900.

Lotterhos KE, Whitlock MC (2014) Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Molecular Ecology* **23**, 2178-2192.

Lowe WH, Allendorf FW (2010) What can genetics tell us about population connectivity? *Molecular Ecology* **19**, 3038-3051.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics* **4**, 981-994.

Lynch HJ, LaRue MA (2014) First global census of the Adélie Penguin. *The Auk* **131**, 457-466.

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

Lynnes A, Reid K, Croxall J (2004) Diet and reproductive success of Adélie and chinstrap penguins: linking response of predators to prey population dynamics. *Polar Biology* **27**, 544-554.

Lynnes A, Reid K, Croxall J, Trathan P (2002) Conflict or co-existence? Foraging distribution and competition for prey between Adélie and Chinstrap Penguins. *Marine biology* **141**, 1165-1174.

MacDougall-Shackleton EA, MacDougall-Shackleton SA (2001) Cultural and genetic evolution in mountain white-crowned sparrows: Song dialects are associated with population structure. *Evolution* **55**, 2568-2575.

Mackensen A, Grobe H, Hubberten H-W, Kuhn G (1994) Benthic foraminiferal assemblages and the δ13C-signal in the Atlantic sector of the Southern Ocean: Glacial-to-interglacial contrasts. In: *Carbon cycling in the glacial ocean: constraints on the ocean's role in global change*, pp. 105-144. Springer.

Mackintosh A, Golledge N, Domack E*, et al.* (2011) Retreat of the East Antarctic ice sheet during the last glacial termination. *Nature Geoscience* **4**, 195-202.

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution* **18**, 189-197.

Martin S (2001) Polynyas. *Encyclopedia of Ocean Sciences* **3**, 2241-2247.

Martínez I, Christie DA, Jutglar F, Garcia EFJ (2016) Gentoo Penguin (*Pygoscelis papua*). In: *Handbook of the Birds of the World Alive* (eds. del Hoyo J, Elliott A, Sargatal J, Christie DA, de Juana E). Lynx Edicions, Barcelona.

Masello JF, Mundry R, Poisbleau M*, et al.* (2010) Diving seabirds share foraging space and time within and among species. *Ecosphere* **1**, 1-28.

Massom R, Hill K, Lytle V*, et al.* (2001) Effects of regional fast-ice and iceberg distributions on the behaviour of the Mertz Glacier polynya, East Antarctica. *Annals of Glaciology* **33**, 391-398.

Massom RA, Giles AB, Fricker HA*, et al.* (2010) Examining the interaction between multi-year landfast sea ice and the Mertz Glacier Tongue, East Antarctica: Another factor in ice sheet stability? *Journal of Geophysical Research: Oceans* **115**.

Massom RA, Hill K, Barbraud C*, et al.* (2009) Fast ice distribution in Adélie Land, East Antarctica: interannual variability and implications for emperor penguins Aptenodytes forsteri. *Marine Ecology Progress Series* **374**, 243-257.

Masson-Delmotte V, Schulz M, Abe-Ouchi A*, et al.* (2013) Information from paleoclimate archives. In: *Climate change 2013: the physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change* (eds. Stocker T, Qin D, Plattner G*, et al.*), pp. 383-464. Cambridge University Press, Cambridge.

Mazet O, Rodriguez W, Grusea S, Boitard S, Chikhi L (2016) On the importance of being structured: instantaneous coalescence rates and human evolution[mdash]lessons for ancestral population size inference[quest]. *Heredity* **116**, 362-371.

Meirmans PG (2012) AMOVA-based clustering of population genetic data. *Journal of Heredity* **103**, 744-750.

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

Millar CD, Dodd A, Anderson J*, et al.* (2008) Mutation and evolutionary rates in Adélie penguins from the Antarctic. *PLoS Genetics* **4**, e1000209.

Moore WS (1995) Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. *Evolution*, 718-726.

Moore WS, DeFilippis VR (1997) The window of taxonomic resolution for phylogenies based on mitochondrial cytochrome b. In: *Avian molecular evolution and systematics* (ed. Mindell DP), pp. 83-119. Academic Press, San Diego, California.

Moritz C (1994) Defining 'Evolutionarily Significant Units' for conservation. *Trends in Ecology & Evolution* **9**, 373-374.

Murphy E, Watkins J, Trathan P*, et al.* (2007) Spatial and temporal operation of the Scotia Sea ecosystem: a review of large-scale links in a krill centred food web. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **362**, 113-148.

Murphy RC (1947) A new zonal race of the gentoo penguin. *The Auk* **64**, 454-455.

Nadeau N, Ruiz M, Salazar P*, et al.* (2014) Population genomics of parallel hybrid zones in the mimetic butterflies, H. melpomene and H. erato. *Genome research* **24**, 1316-1333.

Nevoux M, Forcada J, Barbraud C, Croxall J, Weimerskirch H (2010) Bet-hedging response to environmental variability, an intraspecific comparison. *Ecology* **91**, 2416-2427.

Nogués-Bravo D, Rodríguez J, Hortal J, Batra P, Araújo MB (2008) Climate change, humans, and the extinction of the woolly mammoth. *PLoS Biology* **6**, e79.

Nosil P (2008) Speciation with gene flow could be common. *Molecular Ecology* **17**, 2103-2106.

Orr JC, Fabry VJ, Aumont O*, et al.* (2005) Anthropogenic ocean acidification over the twenty-first century and its impact on calcifying organisms. *Nature* **437**, 681-686.

Orsini L, Vanoverbeke J, Swillen I, Mergeay J, Meester L (2013) Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. *Molecular Ecology* **22**, 5983-5999.

Otley H, Clausen A, Christie D, Huin N, Pütz K (2007) Breeding patterns of king penguins on the Falkland Islands. *Emu* **107**, 156-164.

Paetkau D, Calvert W, Stirling I, Strobeck C (1995) Microsatellite analysis of population structure in Canadian polar bears. *Molecular Ecology* **4**, 347-354.

Paetkau D, Slade R, Burden M, Estoup A (2004) Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Molecular Ecology* **13**, 55-65.

Palsbøll PJ, Berube M, Allendorf FW (2007) Identification of management units using population genetic data. *Trends in Ecology & Evolution* **22**, 11-16.

Palumbi SR (1992) Marine speciation on a small planet. *Trends in Ecology & Evolution* **7**, 114-118.

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

Pattengale ND, Alipour M, Bininda-Emonds OR, Moret BM, Stamatakis A (2010) How many bootstrap replicates are necessary? *Journal of Computational Biology* **17**, 337-354.

Péron C, Weimerskirch H, Bost C-A (2012) Projected poleward shift of king penguins'(Aptenodytes patagonicus) foraging range at the Crozet Islands, southern Indian Ocean. *Proceedings of the Royal Society of London B: Biological Sciences*, rspb.2011.2705.

Pespeni MH, Palumbi SR (2013) Signals of selection in outlier loci in a widely dispersing species across an environmental mosaic. *Molecular Ecology* **22**, 3580-3597.

Petit RJ, Aguinagalde I, de Beaulieu J-L*, et al.* (2003) Glacial refugia: hotspots but not melting pots of genetic diversity. *Science* **300**, 1563-1565.

Piertney SB, MacColl AD, Bacon PJ, Dallas JF (1998) Local genetic structure in red grouse (Lagopus lagopus scoticus): evidence from microsatellite DNA markers. *Molecular Ecology* **7**, 1645-1654.

Pistorius PA, Baylis A, Crofts S, Pütz K (2012) Population development and historical occurrence of king penguins at the Falkland Islands. *Antarctic Science* **24**, 435-440.

Pogson GH, Fevolden SE (2003) Natural selection and the genetic differentiation of coastal and Arctic populations of the Atlantic cod in northern Norway: a test involving nucleotide sequence variation at the pantophysin (PanI) locus. *Molecular Ecology* **12**, 63-74.

Posada D (2009) Selection of models of DNA evolution with jModelTest. *Methods in Molecular Biology* **537**, 112.

Posada D, Buckley TR (2004) Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. *Systematic Biology* **53**, 793-808.

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

Pütz K (2002) Spatial and temporal variability in the foraging areas of breeding king penguins. *The Condor* **104**, 528-538.

Pütz K, Ropert-Coudert Y, Charrassin J, Wilson R (1999) Foraging areas of King penguins Aptenodytes patagonicus breeding at Possession Island, southern Indian Ocean. *Marine Ornithology* **27**, 77-84.

Pütz K, Trathan PN, Pedrana J*, et al.* (2014) Post-fledging dispersal of King Penguins (Aptenodytes patagonicus) from two breeding sites in the South Atlantic. *PLoS One* **9**, e97164.

Pybus OG, Rambaut A, Harvey PH (2000) An integrated framework for the inference of viral population history from reconstructed genealogies. *Genetics* **155**, 1429-1437.

Quinn T (1992) The genetic legacy of Mother Goose–phylogeographic patterns of lesser snow goose Chen caerulescens caerulescens maternal lineages. *Molecular Ecology* **1**, 105-117.

Rambaut A, Drummond AJ (2007) Tracer v1. 4.

Ratcliffe N, Trathan P (2012) A review of the diet and at-sea distribution of penguins breeding within the CAMLR Convention Area. *CCAMLR Science* **18**, 75-114.

Reid K, Croxall JP (2001) Environmental response of upper trophic-level predators reveals a system change in an Antarctic marine ecosystem. *Proceedings of the Royal Society of London B: Biological Sciences* **268**, 377-384.

Richdale LE (1945) Courtship and allied behaviour in penguins Part 2. *Emu* **45**, 37-54.

Ritchie PA, Millar CD, Gibb GC, Baroni C, Lambert DM (2004) Ancient DNA enables timing of the Pleistocene origin and Holocene expansion of two Adélie penguin lineages in Antarctica. *Molecular Biology and Evolution* **21**, 240-248.

Robertson BC, Gemmell NJ (2004) Defining eradication units to control invasive pests. *Journal of Applied Ecology* **41**, 1042-1048.

Robertson G, Wienecke B, Emmerson L, Fraser AD (2014) Long-term trends in the population size and breeding success of emperor penguins at the Taylor Glacier colony, Antarctica. *Polar Biology* **37**, 251-259.

Robisson P, Aubin T, Bremond JC (1993) Individuality in the voice of the emperor penguin Aptenodytes forsteri: adaptation to a noisy environment. *Ethology* **94**, 279-290.

Roeder AD, Marshall RK, Mitchelson AJ*, et al.* (2001) Gene flow on the ice: genetic differentiation among Adélie penguin colonies around Antarctica. *Molecular Ecology* **10**, 1645-1656.

Rogers AD (2012) Evolution and biodiversity of Antarctic organisms: a molecular perspective. In: *Antarctic ecosystems: an extreme environment in a changing world* (eds. Rogers AD, Johnston NM, Murphy EJ, Clarke A), pp. 417-467. Wiley-Blackwell, Oxford.

Rombolá E, Marschoff E, Coria N (2003) Comparative study of the effects of the late pack-ice break-off on chinstrap and Adélie penguins' diet and reproductive success at Laurie Island, South Orkney Islands, Antarctica. *Polar Biology* **26**, 41-48.

Ronquist F (1997) Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. *Systematic Biology* **46**, 195-203.

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes* **4**, 137-138.

Rounsevell D, Copson G (1982) Growth rate and recovery of a king penguin, Aptenodytes patagonicus, population after exploitation. *Wildlife Research* **9**, 519-525.

Saraux C, Viblanc VA, Hanuise N, Le Maho Y, Le Bohec C (2011) Effects of individual pre-fledging traits and environmental conditions on return patterns in juvenile king penguins. *PLoS One* **6**, e20407.

Scheffer A, Bost C-A, Trathan PN (2012) Frontal zones, temperature gradient and depth characterize the foraging habitat of king penguins at South Georgia. *Marine Ecology Progress Series* **465**, 281-297.

Schreiber EA, Burger J (2001) *Biology of marine birds* CRC Press.

Sedwick P, Harris P, Robertson L*, et al.* (2001) Holocene sediment records from the continental shelf of MacRoberston Land, East Antarctica. *Paleoceanography* **16**, 212-225.

Seehausen O, Butlin RK, Keller I*, et al.* (2014) Genomics and the origin of species. *Nature Reviews Genetics* **15**, 176-192.

Seehausen O, Van Alphen JJ, Witte F (1997) Cichlid fish diversity threatened by eutrophication that curbs sexual selection. *Science* **277**, 1808-1811.

Shakun JD, Clark PU, He F*, et al.* (2012) Global warming preceded by increasing carbon dioxide concentrations during the last deglaciation. *Nature* **484**, 49-54.

Shapiro B, Drummond AJ, Rambaut A*, et al.* (2004) Rise and fall of the Beringian steppe bison. *Science* **306**, 1561-1565.

Shaw P, Arkhipkin A, Al-Khairulla H (2004) Genetic structuring of Patagonian toothfish populations in the Southwest Atlantic Ocean: the effect of the Antarctic Polar Front and deep-water troughs as barriers to genetic exchange. *Molecular Ecology* **13**, 3293-3303.

Shepherd LD, Millar CD, Ballard G*, et al.* (2005) Microevolution and mega-icebergs in the Antarctic. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 16717-16722.

Slack KE, Jones CM, Ando T*, et al.* (2006) Early penguin fossils, plus mitochondrial genomes, calibrate avian evolution. *Molecular Biology and Evolution* **23**, 1144-1155.

Slatkin M (1987) Gene flow and the geographic structure of natural populations. *Science* **236**, 787-792.

Slatkin M (2005) Seeing ghosts: the effect of unsampled populations on migration rates estimated for sampled populations. *Molecular Ecology* **14**, 67-73.

Smith JA, Hillenbrand C-D, Pudsey CJ, Allen CS, Graham AG (2010) The presence of polynyas in the Weddell Sea during the Last Glacial Period with implications for the reconstruction of sea-ice limits and ice sheet history. *Earth and Planetary Science Letters* **296**, 287-298.

Smith Jr W, Keene N, Comiso J (1988) Interannual variability in estimated primary productivity of the Antarctic marginal ice zone. In: *Antarctic ocean and resources variability* (ed. Sahrhage D), pp. 131-139. Springer, Berlin Heidelberg.

Smith VR (2002) Climate Change in the Sub-Antarctic: An Illustration from Marion Island. *Climatic change* **52**, 345-357.

Smith WO, Ainley DG, Cattaneo-Vietti R, Hofmann EE (2012) The Ross Sea continental shelf: regional biogeochemical cycles, trophic interactions, and potential future changes. In: *Antarctic ecosystems: an extreme environment in a changing world* (eds. Rogers AD, Johnston NM, Murphy EJ, Clarke A), pp. 213-242. Wiley-Blackwell, Hoboken, NJ.

Smith WO, Gordon LI (1997) Hyperproductivity of the Ross Sea (Antarctica) polynya during austral spring. *Geophysical Research Letters* **24**, 233-236.

Sokolov S, Rintoul SR, Wienecke B (2006) Tracking the Polar Front south of New Zealand using penguin dive data. *Deep Sea Research Part I: Oceanographic Research Papers* **53**, 591-607.

Southwell C, Emmerson L (2013) Large-scale occupancy surveys in East Antarctica discover new Adélie penguin breeding sites and reveal an expanding breeding distribution. *Antarctic Science* **25**, 531-535.

Southwell C, Emmerson L, Takahashi A*, et al.* (2016) Recent studies overestimate colonization and extinction events for Adelie Penguin breeding colonies. *The Auk* **134**, 39-50.

St Ui M, Reimer J (1993) Extended 14 C data base and revised CALIB 3.0 14 C age calibration program. *Radiocarbon* **35**, 215-230.

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

Stammerjohn S, Massom R, Rind D, Martinson D (2012) Regions of rapid sea ice change: An inter-hemispheric seasonal comparison. *Geophysical Research Letters* **39**.

Stonehouse B (1970) Geographic variation in Gentoo penguins *Pygoscelis papua*. *Ibis* **112**, 52-57.

Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences* **100**, 9440-9445.

Storfer A, Murphy M, Evans J*, et al.* (2007) Putting the 'landscape'in landscape genetics. *Heredity* **98**, 128-142.

Stucchi M (2002) Una nueva especie de Spheniscus (Aves: Spheniscidae) de la formación Pisco, Perú. *Boletín Geológico del Perú* **94**, 17-24.

Tamura K (1992) Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+ C-content biases. *Molecular Biology and Evolution* **9**, 678-687.

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution* **10**, 512-526.

Tavecchia G, Tenan S, Pradel R*, et al.* (2016) Climate-driven vital rates do not always mean climate-driven population. *Global Change Biology*.

Taylor SA, Maclagan L, Anderson DJ, Friesen VL (2011) Could specialization to cold-water upwelling systems influence gene flow and population differentiation in marine organisms? A case study using the blue-footed booby, Sula nebouxii. *Journal of Biogeography* **38**, 883-893.

Team RC (2013) R: A language and environment for statistical computing.

Templeton AR (1981) Mechanisms of speciation--a population genetic approach. *Annual review of ecology and systematics*, 23-48.

Thatje S, Hillenbrand CD, Mackensen A, Larter R (2008) Life hung by a thread: endurance of Antarctic fauna in glacial periods. *Ecology* **89**, 682-692.

Thiebot J-B, Lescroël A, Barbraud C, Bost C-A (2013) Three-dimensional use of marine habitats by juvenile emperor penguins Aptenodytes forsteri during post-natal dispersal. *Antarctic Science* **25**, 536-544.

Thompson DW, Wallace JM (2000) Annular modes in the extratropical circulation. Part I: month-to-month variability*. *Journal of Climate* **13**, 1000-1016.

Thornhill DJ, Mahon AR, Norenburg JL, Halanych KM (2008) Open-ocean barriers to dispersal: a test case with the Antarctic Polar Front and the ribbon worm Parborlasia corrugatus (Nemertea: Lineidae). *Molecular Ecology* **17**, 5104-5117.

Toews DP, Taylor SA, Vallender R*, et al.* (2016) Plumage Genes and Little Else Distinguish the Genomes of Hybridizing Warblers. *Current Biology* **26**, 2313-2318.

Trathan P, Croxall J, Murphy E (1996) Dynamics of Antarctic penguin populations in relation to inter-annual variability in sea ice distribution. *Polar Biology* **16**, 321-330.

Trathan PN, Forcada J, Atkinson R, Downie R, Shears JR (2008) Population assessments of gentoo penguins (Pygoscelis papua) breeding at an important Antarctic tourist site, Goudier Island, Port Lockroy, Palmer Archipelago, Antarctica. *Biological Conservation* **141**, 3019-3028.

Trathan PN, Fretwell PT, Stonehouse B (2011) First recorded loss of an emperor penguin colony in the recent period of Antarctic regional warming: implications for other colonies. *PLoS One* **6**, e14738.

Trivelpiece WZ, Buckelew S, Reiss C, Trivelpiece SG (2007) The winter distribution of chinstrap penguins from two breeding sites in the South Shetland Islands of Antarctica. *Polar Biology* **30**, 1231-1237.

Trivelpiece WZ, Trivelpiece SG (2013) Chinstrap Penguin. In: *Penguins Natural History and Conservation* (eds. Borboroglu PB, Boersma PD), pp. 59 - 71. University of Washington Press, Seattle.

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

Turner TL, Hahn MW, Nuzhdin SV (2005) Genomic islands of speciation in Anopheles gambiae. *PLoS Biology* **3**, e285.

Twitchett RJ (2006) The palaeoclimatology, palaeoecology and palaeoenvironmental analysis of mass extinction events. *Palaeogeography, Palaeoclimatology, Palaeoecology* **232**, 190-213.

Van Den Hoff J, McMahon CR, Field I (2009) Tipping back the balance: recolonization of the Macquarie Island isthmus by king penguins (Aptenodytes patagonicus) following extermination for human gain. *Antarctic Science* **21**, 237-241.

Vaughan D, Comiso J, Allison I*, et al.* (2013) Observations: Cryosphere Climate Change 2013: the Physical Science Basis. *et al., Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA* **1**, 1.2.

Vaughan DG, Marshall GJ, Connolley WM*, et al.* (2003) Recent rapid regional climate warming on the Antarctic Peninsula. *Climatic change* **60**, 243-274.

Voisin J-F (1979) Observations ornithologiques aus Iles Tristan da Cunha et Gough. *Alauda* **47**, 73-82.

Waas JR, Caulfield M, Colgan PW, Boag PT (2000) Colony sound facilitates sexual and agonistic activities in royal penguins. *Animal Behaviour* **60**, 77-84.

Walther G-R, Post E, Convey P*, et al.* (2002) Ecological responses to recent climate change. *Nature* **416**, 389-395.

Wand U, Hermichen W-D (2005) Late Quaternary ice level changes in central Dronning Maud Land, East Antarctica, as inferred by 14C ages of mumiyo deposits in snow petrel colonies. *Geol Jahrbuch, B97*, 237-254.

Waples RS, Gaggiotti O (2006) INVITED REVIEW: What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology* **15**, 1419-1439.

Weimerskirch H, Jouventin P, Mougin J, Stahl J, Van BM (1985) Banding recoveries and the dispersal of seabirds breeding in French Austral and Antarctic Territories. *Emu* **85**, 22-33.

Weimerskirch H, Stahl J, Jouventin P (1992) The breeding biology and population dynamics of king penguins Aptenodytes patagonica on the Crozet Islands. *Ibis* **134**, 107-117.

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

Weir J, Schluter D (2008) Calibrating the avian molecular clock. *Molecular Ecology* **17**, 2321-2328.

Whitlock MC, McCauley DE (1999) Indirect measures of gene flow and migration: FST≠ 1/(4Nm+ 1). *Heredity* **82**, 117-125.

Wienecke B, Kirkwood R, Robertson G (2004) Pre-moult foraging trips and moult locations of Emperor penguins at the Mawson Coast. *Polar Biology* **27**, 83-91.

Wienecke B, Raymond B, Robertson G (2010) Maiden journey of fledgling emperor penguins from the Mawson Coast, East Antarctica. *Marine Ecology Progress Series* **410**, 269-282.

Wienecke BC, Robertson G (1997) Foraging space of emperor penguins Aptenodytes forsteri in Antarctic shelf waters in winter. *Marine Ecology Progress Series* **159**, 249-263.

Wiley AE, Welch AJ, Ostrom PH*, et al.* (2012) Foraging segregation and genetic divergence between geographically proximate colonies of a highly mobile seabird. *Oecologia* **168**, 119-130.

Williams MAJ (1998) *Quaternary Environments*, Second edn. Edward Arnold, London.

Williams TD (1995) *The Penguins: Spheniscidae* Oxford Unversity Press.

Willing E-M, Dreyer C, van Oosterhout C (2012) Estimates of Genetic Differentiation Measured by FST Do Not Necessarily Require Large Sample Sizes When Using Many SNP Markers. *PLoS One* **7**, e42649.

Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* **163**, 1177-1191.

Wilson P, Ainley D, Nur N*, et al.* (2001) Adélie penguin population change in the pacific sector of Antarctica: relation to sea-ice extent and the Antarctic Circumpolar Current. *Marine Ecology Progress Series* **213**, 301-309.

Wilson RP, Culik BM, Kosiorek P, Adelung D (1998) The over-winter movements of a chinstrap penguin (Pygoscelis antarctica). *Polar Record* **34**, 107-112.

Woehler EJ (1989) Resighting and recoveries of banded seabirds at Heard Island, 1985–1988. *Corella* **13**, 38-40.

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

Wright S (1932) The roles of mutation, inbreeding, crossbreeding and selection in evolution **1**, 356-366.

Wright S (1940) Breeding structure of populations in relation to speciation. *The American Naturalist* **74**, 232-248.

Wright S (1943) Isolation by distance. *Genetics* **28**, 114.

Wright S (1969) Evolution and the genetics of populations, Vol. 2: The theory of gene frequencies. Univ. of Chicago Press, Chicago.

Younger J, Clucas G, Kao D*, et al.* (submitted) Recognising metapopulation structure in emperor penguins is critical for their future conservation.

Younger J, Emmerson L, Southwell C, Lelliott P, Miller K (2015a) Proliferation of East Antarctic Adélie penguins in response to historical deglaciation. *BMC evolutionary biology* **15**, 1-11.

Younger J, van den Hoff J, Wienecke B, Hindell M, Miller K (2016a) Contrasting responses to a climate regime change by sympatric, ice-dependent predators. *BMC evolutionary biology* **16**.

Younger JL, Clucas GV, Kooyman G*, et al.* (2015b) Too much of a good thing: sea ice extent may have forced emperor penguins into refugia during the last glacial maximum. *Global Change Biology* **21**, 2215-2226.

Younger JL, Emmerson LM, Miller KJ (2016b) The influence of historical climate changes on Southern Ocean marine predator populations: a comparative analysis. *Global Change Biology* **22**, 474-493.

Yuan X (2004) ENSO-related impacts on Antarctic sea ice: a synthesis of phenomenon and mechanisms. *Antarctic Science* **16**, 415-425.

Zachos J, Pagani M, Sloan L, Thomas E, Billups K (2001) Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* **292**, 686-693.

Zimmer I, Wilson RP, Gilbert C*, et al.* (2008) Foraging movements of emperor penguins at Pointe Géologie, Antarctica. *Polar Biology* **31**, 229-243.

Zink RM, Barrowclough GF (2008) Mitochondrial DNA under siege in avian phylogeography. *Molecular Ecology* **17**, 2107-2121.

Zwally HJ, Comiso J, Gordon A (1985) Antarctic offshore leads and polynyas and oceanographic effects. *Oceanology of the Antarctic Continental Shelf*, 203-226.

Zwally HJ, Comiso JC, Parkinson CL, Cavalieri DJ, Gloersen P (2002) Variability of Antarctic sea ice 1979–1998. *Journal of Geophysical Research: Oceans (1978–2012)* **107**, 9-1-9-19.