The Molecular Ecology of the Gentoo Penguin (*Pygoscelis papua*)

Hila Levy

A thesis submitted for the degree of

Doctor of Philosophy

University of Oxford

Department of Zoology

St John's College

Trinity Term 2020

Declaration

I declare that this thesis was composed by myself and that the work contained herein is my own except where explicitly stated in the text. Each data chapter contains an “Author Contributions” section. The work has not been submitted for any degree or professional qualification except as specified.

Hila Levy

Table of Contents

Declaration 2

Thesis Abstract 4

Acknowledgments 5

Chapter 1. Introduction 9

Section I. The Gentoo Penguin as a Host Species 43

Chapter 2. Population structure and phylogeography of the Gentoo Penguin (*Pygoscelis papua*) across the Scotia Arc 44

Chapter 3. Evidence of Pathogen-Induced Immunogenetic Selection across the Large Geographic Range of a Wild Seabird 90

Section II. Gentoo Penguin Pathogens and Microbes 145

Chapter 4. Review of Pathogens in Pygoscelid Penguins 146

Chapter 5. Identification and distribution of novel circular Rep-encoding ssDNA viruses and circular molecules in four Penguin species in South Georgia and the Antarctic Peninsula 214

Chapter 6. Identification of Circovirus Genome in a Chinstrap Penguin (*Pygoscelis antarcticus*) and Adélie Penguin (*Pygoscelis adeliae*) on the Antarctic Peninsula 246

Chapter 7. Gastrointestinal Bacterial Microbiota of Four Penguin Species and Blue-Eyed Shags across the Scotia Arc 267

Chapter 8. Screening of 16S Profiles for Potentially Pathogenic Bacteria in Four Penguin Species and a Flying Seabird in the Scotia Arc and Antarctica 312

Chapter 9. Eukaryotic Gastrointestinal Parasites of Gentoo Penguins and Sympatric Seabirds 351

Chapter 10. Discussion, Future Directions, and Concluding Remarks 404

Appendix A1 - Supplementary Data for Chapter 2 430

Appendix A2 - Supplementary Data for Chapter 3 435

Appendix A3 – Supplementary Data for Chapter 5 460

Appendix A4 – Supplementary Data for Chapter 6 468

Appendix A5 - Supplementary Data for Chapter 7 472

Appendix A6 - Supplementary Information for Chapter 8 487

Appendix A7 – Supplementary Information for Chapter 9 502

Thesis Abstract

Gentoo Penguins (*Pygoscelis papua*) are marine mesopredators facing threats from anthropogenic environmental change, habitat loss, fisheries pressure, pollution, direct disturbance at colonies, and introduced pathogens. These philopatric seabirds breed across the largest geographic and latitudinal range of any penguin taxon. Given the large biophysical range of Gentoo Penguin habitats, evolutionary theory posits that localised biotic and abiotic interactions would place differential pressures on distinct populations of the species, affecting their genotypes and phenotypes. The host-pathogen interface is one such biotic interaction known to shape evolutionary trajectories. This thesis employs molecular techniques to examine the role of geographical variation on the genetic diversity of (I) the penguin as a host and (II) its enteric viruses, bacteria, and eukaryotic parasites. In this work, I uncovered strong host phylogeographic and population genetic structure according to neutral, mitochondrial, and immune gene (Toll-like receptor, TLR) markers. I also present experimentally-supported evidence of positive selection on *TLR5*, which points to localised pathogen-driven adaptation. Given the limited understanding of the pathogens circulating in modern Gentoo Penguin habitats, I also undertook parallel sequencing-based screening surveys to report novel viral, bacterial, protozoan and helminthic sequences associated with these seabirds and their sympatric species. Alongside undescribed diversity, I found evidence of greater-than-expected connectivity among the enteric microbes of spatially separate and genetically disparate Gentoo Penguins, as well as Pygoscelid-wide microbe commonalities. This research demonstrates the importance of studying species complexes and biotic interaction at the large scale, particularly in the face of emerging threats and the challenges inherent to wildlife disease monitoring in these remote habitats.

Acknowledgments

This thesis was made possible by many collaborators, teammates, friends, and family, near and far, who accompanied me on this journey. I wish I could write a full chapter of thanks.

I am first thankful to my supervisors, Prof Adrian Smith, Dr Tom Hart, and Prof Alex Rogers for taking a chance on me, supporting my application to return to science, and guiding me along a path of discovery. This programme was not without its challenges, including the loss of our lab facilities, eviction to a new site, split operations, and finally, a global pandemic, but I appreciate them sticking with me to completion. I have learned more than I ever thought I would want to know about penguins, poo, pathogens, parasites (and people), and I thank them for taking the time to be patient along the way. I appreciate the chats, questions, and critiques that strengthened this work.

I also thank the staff of the Department of Zoology, in particular Heather Green, Phil Smith, Lynne Bradley, Kate Southey, and the finance and orders teams for their support—research does not occur in a vacuum. I am grateful to the Clarendon Fund, St John's College, the U.S. Department of Veterans Affairs, grants shared by collaborators, and Penguin Lifelines for financing portions of this dream, and for the logistics support provided by the Royal Air Force and Quark Expeditions. The staff and crew aboard the Ocean Diamond and Ocean Endeavour made me feel at home during fieldwork on the waves of the Drake Passage and amid the sea ice, even though I often smelled of guano. Thank you to Tom, Caitlin Black, Gemma Clucas, Casey Youngflesh, Mo Lynch, Rachael Herman, Mike Polito, Heather Lynch, and many others on parallel expeditions for getting intimately acquainted with aromatic penguin faeces in the name of science.

I am overwhelmingly grateful to my key collaborators Steven Fiddaman, Gemma Clucas, and Stephen Preston, who know what an undertaking it is to conduct research that involves field,

lab, and computational aspects. I thank them for being helpful and responsive through the iterative scientific process, and for commentary on portions of this thesis. I could not have gotten here without the assistance or commiseration in the lab of Patrik Flammer, Frances Colles, Rob Dixon, Kirstie McCullough, Sarah Stiffel, Abby Bailey, Cat Head, Vanessa Lovenburg, Nicolai Roterman and Michelle Taylor. I am thankful for the opportunity to interact with brilliant scientists in both Adrian and Alex's lab groups, and for the camaraderie built with officemates like Jesse Van der Grient, Oli Ashford, Fiona Jones, Bétina Frinault, and Christina Hunt before and after the Tinbergen eviction. I feel lucky to have had a talented MSc cohort a decade ago who still remain in touch and supported me as the last to undertake this big step—thank you Annette Fayet, Hannah MacGregor, James Sciberras, and Graham Prescott for keeping me focused.

There are many contributors to my research whom I have never met face-to-face, but who provided samples, sequence data, or helpful input to my work. I also want to thank Juliana Vianna, Andrés Barbosa, and Arvind Varsani for taking me under their wing from afar to make stronger, more inclusive research products. The field of science is better when people act as generously as they have.

I thank my friends around the world for being both great listeners and vocal cheerleaders through the highs and lows. Thank you to Calla Glavin, Miles and Caroline Hidalgo, Joe and Megan Chapa, Mary Eaton, Megan Musilli, and the Rhodes House staff for making my time in Oxford special, especially while my family grew. My military teammates in Air Force Special Operations Command inspire me every day, and I have been honoured to serve with them. I am extraordinarily thankful to Tiffany Curran, Melissa Prince, John Leger, Bryant Benefiel, Brandon Broussard, Erik Harris, Kristina Belcourt, and Oscar Rondón for being my circle of trust. To the commanders I have been lucky enough to work for over these years, I am in awe of the support provided to me, my family, and my career progression as 'just' a part-time reservist, and in particular, to Lt. Col. Zak Blom for pushing me to be more reflective and find my true potential.

I would not have been on this path at all if not for the teachers of my youth at St. John's School (Patricia Del Toro, Gian Toyos, Adele Mouakad, and more), faculty at the U.S. Air Force Academy (Derek Varble, David Hale, Daniel Atchley, Edward Westermann, Ann Reagan, John Putnam, and countless others), and supervisors and mentors throughout my military career. Thank you for allowing me to bloom where planted.

In the years I spent in the pursuit of pathogens, I lost friends to other battles. The loss of my friend Ian Helms during my research left an enormous void, and I hope I can make him and his family proud. I also honour the untimely losses of my friends Stephen "Cajun" Del Bagno, David Albandoz, and the crew of RICAN68. To the people of Puerto Rico who suffered so much at the hands of both natural and manmade disasters in recent years, my heart felt the burden of being away from my "patria", and I commit to working toward a brighter future.

To my parents Ramón and Rachel, who surely thought it was crazy to raise me as a city girl from a tropical island only to find me in the iciest and most remote place I could find, I am eternally grateful. They empowered me to never settle, to work hard, and to believe I could do not just something, but anything. At an early age, I remember telling them I wanted to be a zoologist… on the moon. I suppose Antarctica was close enough, but I'll still shoot for the stars. To my brother Ron, sister Meyrav, cousins and grandparents, I am sorry I spent so many years far away, thank you for being there. To my in-laws Belinda, Stephen, Rex, and Brenda, I extend my thanks for hosting me for extended periods while buried in my laptop working, and supporting me keeping their son and grandchildren an ocean away.

My husband Brad deserves more than I will ever be able to give in return for his patience and understanding throughout this stage of our lives. I dragged him abroad for longer than he bargained for and more hurdles than either of us could have expected. I had good reason to quit many times, but he kept me going, helped me juggle countless responsibilities, and stepped up

when it counted for me to prioritise finishing this work. His love and steady provision of ice cream fuelled these analyses.

To the penguins, I apologise—you were only poked and prodded because I care. I leave this work for other scientists, explorers, and policymakers to build on as a means to preserve the breathtakingly stunning Antarctic wilderness. Before I embarked on this endeavour, the world was a place where wildlife diseases rarely featured in people's minds. I submit these pages in a forever-changed world, where it is clearer that the health of our environment is inextricably linked to our own health. I hope light will come out of the darkness.

This thesis is dedicated to my beloved children, Asher, Aviva, and Ariella. It was no easy task to become a mother in the midst of the chaos, but they provide me with an enormous sense of purpose. My wish for them is endless curiosity and a lifelong search for truth, armed with data and a passion for the natural world, where we are mere guests.

Chapter 1. Introduction

The Role of Geographical Variability in Evolution

In his seminal work, *Systematics and the Origin of Species: from the Viewpoint of a Zoologist*, the evolutionary biologist Ernst Mayr explains that, "Every single locality on the Earth, even the smallest geographic district, is somewhat different from other neighbouring ones in the nature of its environment" (Mayr 1942, p. 34). This observation, though somewhat obvious, is of particular importance in the context of evolutionary theory. Every unique "environment" is composed of abiotic factors such as climate, soil, topography, and hydrographic features, along with biotic factors like vegetation, animal life, and parasites. Collectively, Mayr tells us, "we have a formidable array of selective agencies which will shape each population of a species to fit into the particular geographical and ecological niche in which it has been placed." (Mayr 1942, p. 34)

The effect of geographical variation on wildlife is a fundamental question in understanding how a population or species came to have its present characteristics. No longer restricted to Darwin's finches, this question is well-studied in diverse taxa. However, since the second half of the twentieth century, the advent of continually improving molecular techniques has refined our abilities to discern the impacts of environments on organismal genotypes and phenotypes. The field known as "Molecular Ecology"—a follow-on from "Ecological Genetics" as a term developed by E.B. Ford (1964)—applies these genetic techniques to address questions related to populations in wild environments.

By means of example, molecular ecology has given us greater insights into "environments" from as small as the guts of invertebrates, to global scales (Bourne *et al.* 2013; Reusch & Wood 2007). Biologists can now measure the impact of abiotic and biotic factors on diversity, dispersal, phylogenies, coevolution, and the delimitation of species and conservation management units. With a range of applications to conservation biology in a rapidly changing world, molecular ecology

provides insight into how geographical variation drives local adaptation, divergence, reproductive isolation, and therefore speciation.

In this thesis, I use an integrated set of molecular techniques to examine the role of geographical variability in shaping the ecology of a widely distributed seabird species complex in the Southern Ocean, the Gentoo Penguin, *Pygoscelis papua*. As I will illustrate in this introduction, the Gentoo Penguin is uniquely suited to serve as a wild exemplar of key evolutionary processes like local adaptation, natural selection, and incipient speciation thanks to its diversity of island and marine habitats, life history, and biotic interactions.

Penguins as Adaptive Models

Penguins (Sphenisciformes) are a unique order of flightless, diving seabirds inhabiting diverse coastal and island environments in the Southern Hemisphere. Penguins have a rich fossil record extending to >60 Mya, with over 50 documented extinct species (Cole *et al.* 2019b; Ksepka *et al.* 2012; Slack *et al.* 2006). There are between 18-20 extant species of penguin across 6 genera (*Aptenodytes*, *Pygoscelis*, *Eudyptula*, *Spheniscus*, *Eudyptes*, and *Megadyptes*), with debate surrounding the taxonomy of certain recently diverged lineages (Clements *et al.* 2019; Pan *et al.* 2019). Today's penguins breed along cold water currents as far north as the Equator (Galápagos Penguins, *Spheniscus mendiculus*), on continental coastlines of Africa (African Penguin, *Spheniscus demersus*), South America (Magellanic Penguin, *Spheniscus magellanicus*; Humboldt Penguin, *Spheniscus humboldti*), and Australia (Little Penguin, *Eudyptula minor*), the majority of Southern Ocean island groups (multiple species in Genus: *Eudyptes*) and as far south as 77° S, with the ice-dwelling Emperor Penguins (*Aptenodytes forsteri*) of Antarctica (Fretwell & Trathan 2009; Sclater 1888).

Previous investigations into the pattern of modern penguin distribution proposed Antarctic cooling and changes in the Antarctic Circumpolar Current as contributing to radiations of penguin ancestors out of Antarctica northward (Acosta Hospitaleche *et al.* 2013; Baker *et al.* 2006).

However, recent molecular and fossil evidence supports a secondary colonisation hypothesis, whereby today's Antarctic penguin taxa evolved from sub-Antarctic colonisers that moved south after the formation of permanent ice sheets (Cole *et al.* 2019a; Gavryushkina *et al.* 2017; Ksepka *et al.* 2006; Pan *et al.* 2019). The rich diversity of sub-Antarctic penguins, on the other hand, seems to have an even more recent origin, closely tied to the geological emergence of islands in the Plio-Pleistocene (Cole *et al.* 2019b). The processes of speciation and functional adaptation to the varied environments penguins inhabit make penguins of key interest to palaeontologists, molecular ecologists, and evolutionary biologists alike.

Thanks to one of the most extensive fossil records of any group of birds (Ksepka *et al.* 2012), we know that over macroevolutionary scales, penguins have attained their current morphology, physiology, and many of their behaviours in response to abiotic factors linked to cold temperatures and their marine environment. For example, to cope with frigid polar winters, penguins possess thick skin and fat stores, and can adjust their metabolic rates in extremely low temperatures (Le Maho 1977, 2005). Emperor Penguins are known to form huddles as protection from extreme cold during their winter breeding cycle (Gilbert *et al.* 2008), and Gentoo Penguin chicks tend to form aggregations during wet weather periods (Black *et al.* 2016). Penguins have adopted a variety of nesting strategies to keep their eggs and chicks warm and dry: keeping them in an abdominal fold above their feet in the case of Emperor and King Penguins (Genus: *Aptenodytes*), building rock nests to filter snowmelt for other Antarctic species (Genus: *Pygoscelis*), or using insulated burrows or elevated grasses in more temperate environments (Genera: *Eudyptes, Eudyptula*, *Megadyptes,* and *Spheniscus*). Recently, it has been reported that penguin presence, with resulting guano production, melts snow, enabling the start of the breeding season in Pygoscelid penguins (University of Oxford 2015).

Additional well-known adaptations are known to also involve biotic interactions. As marine mesopredators, predator-prey dynamics have played a major evolutionary role. The ancient

ancestors of modern penguins are thought to have lost their ability to fly as a biomechanical trade-off for becoming efficient swimmers and divers in pursuit of available marine prey (Elliott *et al.* 2013). Modern penguins are expert divers, with many species capable of extended deep dives thanks to physiological adaptations to pressure and cold water. These adaptations are widespread; in bones, organs, blood, and metabolic rates (Gilbert *et al.* 2008; Handrich *et al.* 1997; Ksepka *et al.* 2015). Foraging strategies are also thought to have affected neuroanatomical structures in penguin lineages based on sensory adaptations to waterborne prey (Proffitt *et al.* 2016).

Penguins are also able to adjust their metabolism to short-term and seasonal changes in prey availability (Bevan *et al.* 2002). The catastrophic moult experienced by all penguin species in a fasting state during the austral summer ensures minimal heat loss during colder months and minimal loss of foraging efficiency during the short breeding season by altering food metabolism to conserve energy (Groscolas 1986; Le Maho 1977). Penguins with abundant prey supplies digest their food relatively rapidly, in sharp contrast to the energy conservation techniques they can employ during the weeks-long moult fast or during the breeding cycle, when fasts can last up to 115 days in the extreme case of the Emperor Penguin (Gilbert *et al.* 2008; Groscolas & Robin 2001).

The Gentoo Penguin: A Wild Model of Evolution in Action

In this macroevolutionary context of environmental adaptation, one particular penguin merits attention. Unlike other species with a more restrictive range or habitat type, the Gentoo Penguin, *Pygoscelis papua*, has retained a distribution that is circumpolar, and spans from 46° S to 67° S latitude (see **Figure 1**). At first glance, it seems to be as suited to breeding on the sand dunes of the Falkland Islands/Islas Malvinas in the Atlantic as to the grasses of Indian Ocean islands and the rocky, snowy coastlines of the Western Antarctic Peninsula. In other seabird species, such a distribution might be explained by migration and/or prolonged connectivity. In fact, the sister species *Pygoscelis adeliae* (Adélie Penguin) and *Pygoscelis antarcticus* (Chinstrap Penguin) are known to travel long distances and exhibit metapopulation dynamics across the majority of their

ranges (Clucas *et al.* 2014; Clucas *et al.* 2018; Freer *et al.* 2015; Mura-Jornet *et al.* 2018; Roeder *et al.* 2001). However, Gentoo Penguins exhibit high natal philopatry, only dispersing to forage at short, inshore ranges (typically < 50 km) from their colonies (Ratcliffe & Trathan 2012).

Figure 1. Extent of Gentoo Penguin (*Pygoscelis papua*) breeding range (highlighted in orange).

The archipelagos and landmasses dotted with Gentoo Penguin colonies around the Southern Ocean each have unique geological and climatic histories. For example, major breeding areas like the Falkland Islands split off from larger continental landmasses during the breakup of

the Gondwanan supercontinent hundreds of millions of years ago, with South Georgia surfacing from the South American plate during tectonic collisions circa 10 million years ago (Carter *et al.* 2014). At around that time (~10 Ma), uplifted oceanic crust formed Macquarie Island in the Pacific Ocean (Varne & Rubenbach 1978). Still other breeding locations are products of recent volcanism, like Marion Island in the Indian Ocean, which is very young in geological terms at less than half a million years old (Boelhouwers *et al.* 2007; Quilty 2007). Volcanic activity has shaped penguin habitats as recently as this century, with significant volcanic-tectonic disruption on Deception Island, South Shetland Islands in 2014-15 (Almendros *et al.* 2018) and several eruptions of at least four volcanoes on the South Sandwich Islands between 2000-2016 (Global Volcanism Program 2013). Shaped by multiple periods of repeated glaciation and geological change, the range of habitats of Gentoo Penguins today serves as a natural laboratory for understanding the biology of climate change, wildlife colonisation, and range expansion across populations.

The need for research is pressing, not least because the Gentoo Penguin faces complex threats from anthropogenic environmental change, habitat loss, fisheries pressure, and pollution. The global population was last estimated at 387,000 breeding pairs in 2013 (Lynch 2013). In 2018, Gentoo Penguins were upgraded to “Least Concern” status from “Near Threatened” by the International Union for the Conservation of Nature Red List (BirdLife International 2018). This change is heavily influenced by the distribution of penguins in areas where their populations have been increasing, namely the South Atlantic (Falkland Islands) and Antarctic Peninsula, where a majority of the species breeds (see **Table 1**).

Table 1. Estimates of Gentoo Penguin (*Pygoscelis papua*) population size (number of breeding pairs, ordered by size), based on best available census data for each major breeding area, with census year and reference information. Table adapted and updated from Lynch (2013).

Breeding Area	**Census Population (*n* = breeding pairs)**	**Year**	**Reference**
Falkland Is./Malvinas	132,321	2010	Baylis *et al.* (2013)
South Georgia	98,867	1996	Trathan *et al.* (1996)
South Shetland Is. and Western Antarctic Peninsula	94,571	Various	Lynch (2013)
Kerguelen Is.	30,000-40,000	1988	Weimerskirch *et al.* (1989)
Heard and McDonald Is.	16,574	1987	Woehler (1993)
South Orkney Is.	10,760	2013	Lynch (2013)
Crozet Is.	9,000	1994	Jouventin (1994)
Macquarie Is.	3,800	Unknown	Holmes *et al.* (2006)
South Sandwich Is.	1,572	1996/97 and 1998	Convey *et al.* (2009)
Marion Is.	600	2012/13	Crawford *et al.* (2014)
Prince Edward Is.	40	2008/09	Crawford *et al.* (2009)
Martillo Is., Tierra del Fuego	12	2006/07	Ghys *et al.* (2008)

Population trends have been difficult to study, as they fluctuate greatly from year-to-year, and over a 10-12 year time scale. Sharp decreases followed by stable periods or increases have been reported in the Falkland Islands (Baylis *et al.* 2013), while colonisation events and 1100% increases were noted in some parts of the Antarctic Peninsula (Lynch 2013). Nonetheless, smaller

populations from Indian Ocean colonies on Marion Island, Heard Island, and Kerguelen appear to be in decline (Crawford *et al.* 2003; Crawford *et al.* 2014; Lescroël *et al.* 2009; Lescroël & Bost 2006). Modern technology and increased field effort have been helping monitor these changes (Borowicz *et al.* 2018), but the causes of population increases and declines are multifactorial (Lynch *et al.* 2012). Uneven trends have sparked several lines of research worldwide into matters associated with sea ice extent, invasive species, rising temperatures, predation, food availability, migration, disease, and stress linked to human visitation. Each of these factors—biotic and abiotic—vary across the Gentoo Penguin's range.

This brings us back to the original evolutionary question at the core of this work: how has geographical variability—and specifically biotic pressures like parasitism or lower trophic prey availability—affected this wild seabird's genotypes and phenotypes? What factors are most likely to affect our conservation management of these iconic birds in the future? We can use molecular techniques to attempt to examine this on three levels: (1) the Gentoo Penguin itself, (2) the parasites it hosts, and (3) other biota it interacts with as commensals, prey, or sympatric megafauna.

Gentoo Penguin Population Structure

To properly investigate the Gentoo Penguin as a central figure in its environments, we must consider its intraspecific interactions and population dynamics. The population structure of a species provides us with a baseline understanding of how much interaction occurs between individuals of different colonies, in the form of gene flow. Gene flow and a lack of genetic differentiation imply consistent intermixing and mating between individuals from different sites, likely facilitated by migration or a lack of site fidelity. On the other hand, strong genetic differentiation, whether due to the behavioural nature of a species or population, or to the existence of natural and physical barriers to movement, is an indicator that a population is isolated and is on a separate evolutionary trajectory (Wright 1949).

The Antarctic Polar Front, the boundary between Antarctic and sub-Antarctic waters formed by the world's strongest current, the Antarctic Circumpolar Current, has been demonstrated to be one significant barrier to seabird and pelagic organism dispersal and gene flow in the Southern Ocean (Clucas *et al.* 2016; Fraser *et al.* 2012; Friesen 2015; Friesen *et al.* 2007; Munro & Burg 2017; Pertierra *et al.* 2020). The Gentoo Penguin's range extends both north of the Polar Front (in the Falkland Islands, Tierra del Fuego, Marion Island, the Crozet Archipelago, and Macquarie Island) and south of it (in South Georgia, the South Sandwich Islands, South Orkney Islands, South Shetland Islands, and the Western Antarctic Peninsula). This hydrological barrier could therefore be important in shaping the genetic differentiation of Gentoo Penguins. It would also be important to consider when evaluating the biota present on either side of the Polar Front and whether these share similar, if any, patterns of differentiation.

Natural historians and ornithologists have long been interested in classifying island bird species in the vein of Darwin's observations of physical differentiation. First described as *Aptenodytes papua* by J.R. Forster on James Cook's second voyage in 1781 while in the Falkland Islands, naturalists slowly came to realise that the "same" birds were found in remote parts of the Southern Ocean. The Gentoo Penguin has been subjected to several twentieth-century studies on its appearance and behaviour, with debates regarding the taxonomy of the species persisting since that time (Murphy 1936, 1947; Stonehouse 1970).

There are currently two internationally-recognised morphological sub-species of Gentoo Penguin, the "Northern Gentoo", *P. papua papua*, and the "Southern Gentoo", *P. papua ellsworthi*, though many scientists suggested the existence of a third Indo-Pacific sub-species based on morphology (Mathews 1927) and acoustic data (de Dinechin *et al.* 2012; Dobson & Jouventin 2017). As shown in **Figure 2**, taken from its type description, *P. papua ellsworthi* is considered to differ "from *P. papua papua* of the Falkland Islands, in being smaller, and in having a shorter bill, more slender terminally and with a notable reduced culminicorn" (Murphy 1947).

Figure 2. Type description of two recognised sub-species, with *Pygoscelis papua papua* from the Falkland Islands on the left (topotype) and *Pygoscelis papua ellsworthi* from Deception Island, South Shetland Islands on the right (type). Drawing by Alexander Seidel, as appears in Murphy (1947).

While Clucas *et al.* (2014) and Levy *et al.* (2016) were able to resolve distinctions north and south of the Polar Front in the South Atlantic and Scotia Arc using mitochondrial DNA and microsatellite markers (see **Chapter 2**), further study using microsatellites by Vianna *et al.* (2017), RAD-Seq by Clucas *et al.* (2018) and ddRAD-Seq by Pertierra *et al.* (2020) also sampled Indian Ocean sites. By incorporating more of the species' circumpolar range, these recent studies further expanded our knowledge of *P. papua* as a species complex, and suggest there are at least four Evolutionarily Significant Units (ESUs) or clades of Gentoo Penguins. Molecular techniques support a revision of the two sub-species concept in favour of an expanded taxonomy that takes into account the underlying genetic diversity of the multiple clades. To facilitate further reference to subspecies and clades, **Figure 3** illustrates the historical progression of Gentoo Penguin taxonomy.

Figure 3. Visual guide to taxonomic classifications of the Gentoo Penguin over time. Figure 3a-e displays shifts in subspecies and clade designations (yellow = *Pygoscelis papua papua*, orange = *P. papua ellsworthi*, pink and purple depict proposed Eastern groupings, explained in each map).

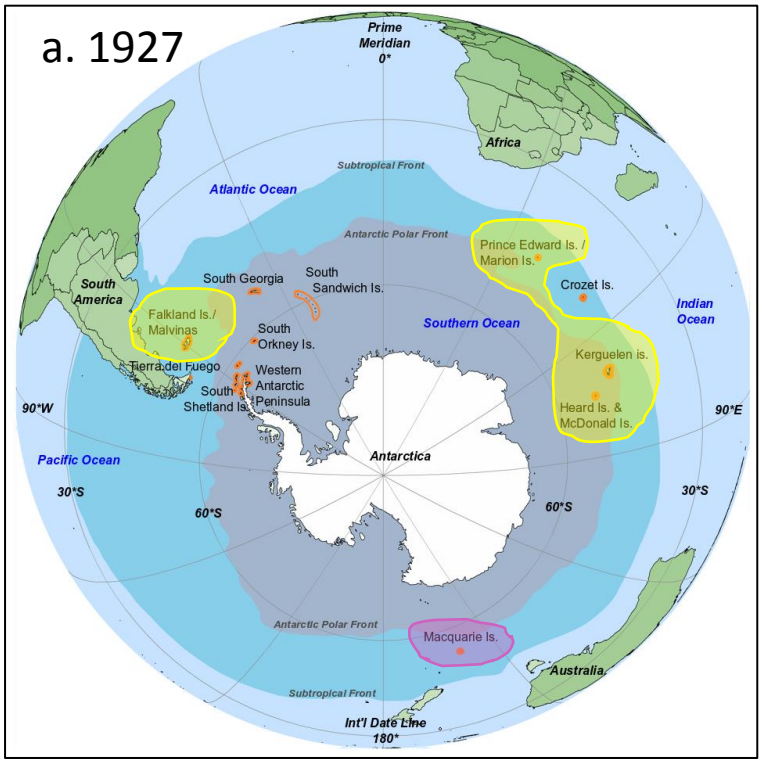


Figure 3a.: First distinctions

Mathews (1927) lists two morphological subspecies:

- *Pygoscelis papua papua* (in yellow: Falkland Is., Marion Is., Kerguelen, and Heard Is.)

- *P. papua taeniata* (in pink: Macquarie Is.)

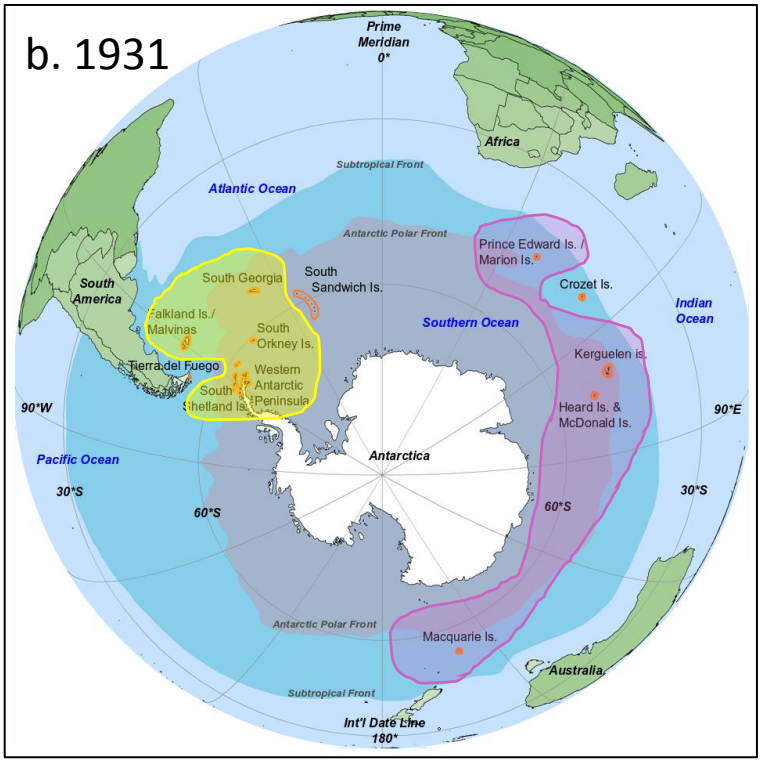


Figure 3b.: Early Morphology Leads to Debate

Peters (1931) proposes a different distribution by morphology:

- *P. papua papua* as Atlantic subspecies with expanded range (in yellow: Falkland Is., South Orkney Is., South Georgia, South Shetlands, and Antarctica)

- *P. papua taeniata* as Eastern subspecies (in pink: Marion Is., Kerguelen, Heard Is. and Macquarie Is.)

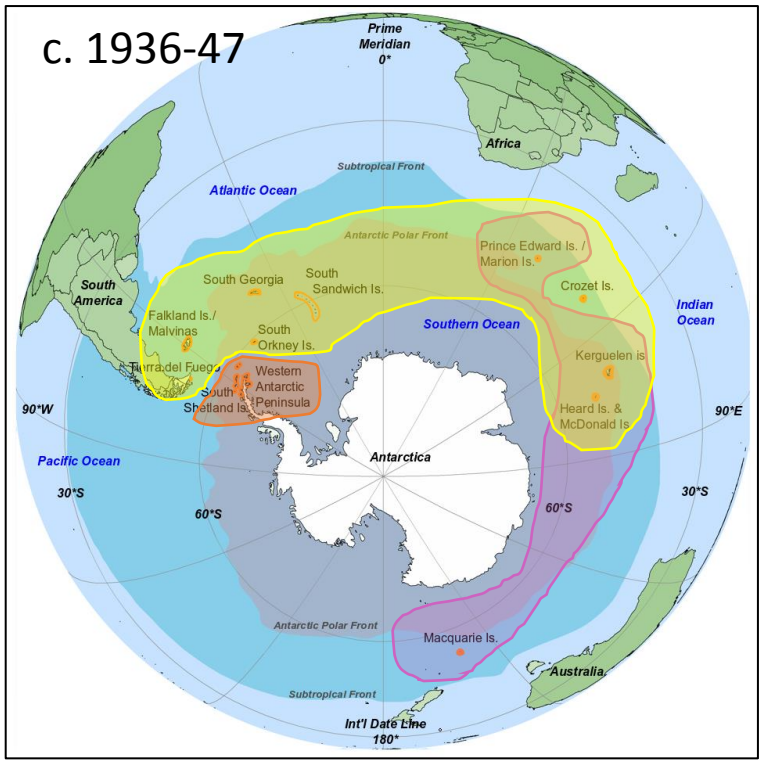


Figure 3c.: A Southern Subspecies is Named

Murphy (1936) introduces concept of Southern subspecies in first major description of Gentoo Penguin biology

Murphy (1947) classifies Southern subspecies as *P. papua ellsworthi* distinct from *P. papua papua* based on morphology, breeding on South Shetland Is. and Western Antarctic Peninsula (in orange on map); does not discuss Eastern classifications

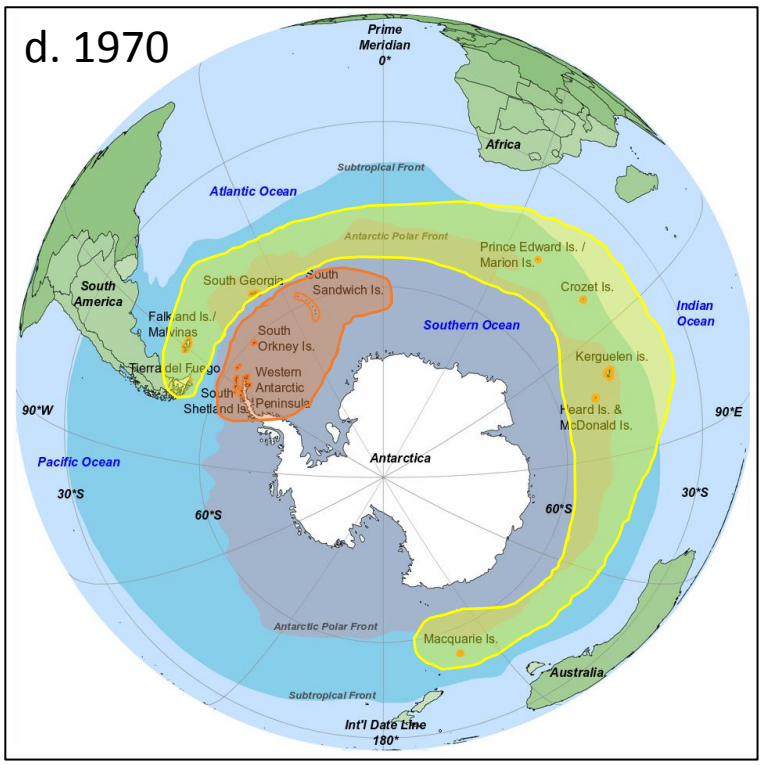


Figure 3d.: Two Morphological Subspecies Prevail

Stonehouse (1970) uses morphology to draw distinct line at 60° S latitude between "Northern" subspecies *P. papua papua* and "Southern" subspecies *P. papua ellsworthi* (in orange on map)

Note: This morphological description remains the current internationally-recognised taxonomy of Gentoo Penguins.

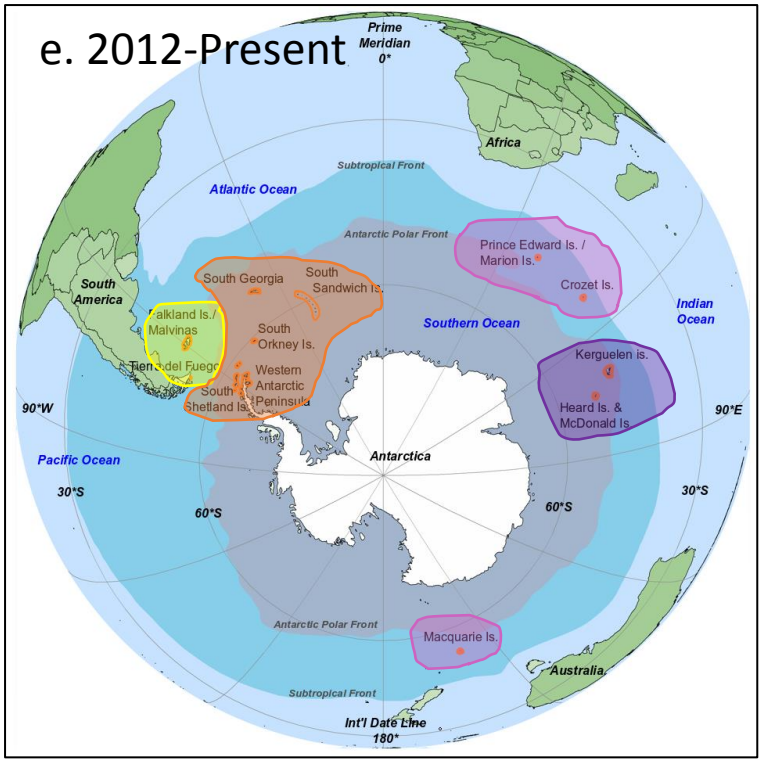


Figure 3e. The Molecular Era—Building Toward a Four Clade Consensus:

de Dinechin *et al.* (2012) propose putative Indian Ocean subspecies at Kerguelen, distinct from Falkland Islands and Antarctic Peninsula populations based on morphology, acoustics, and mtDNA.

Clucas *et al.* (2014) and Levy *et al.* (2016) use mtDNA and microsatellites, respectively, to support two Atlantic clades, but suggest Gentoo Penguins South Georgia cluster more closely genetically to the Southern subspecies (in orange).

Vianna *et al.* (2017) propose four genetic clusters, two in the Atlantic, North (yellow) and South (orange) of the Polar Front, along with separate clusters in Crozet Is. (pink) and Kerguelen (purple), supported by RAD-Seq in Clucas *et al.* (2018), who also found differentiation of South Georgia within the Southern subspecies.

Pertierra *et al*. (2020) expand sampling, with ddRAD and mtDNA supporting the above four clades and showing Indian Ocean clades also split along the Polar Front, with a North Indian Ocean clade (pink: Marion Is./Crozet Is./Macquarie Is.) and South Indian Ocean clade (purple: Kerguelen and Heard Is.).

These subspecies debates are not simply theoretical, as they have practical implications for conservation management. Population genetic studies, with their ability to identify the genotypes beneath cryptic morphological traits are therefore useful tools to clarify the taxonomy and designation of management units. Any differentiation observed in a host species could also provide a backdrop for examining other comparative questions of interest regarding host-parasite coevolution across the geographical range of *P. papua*.

Parasitism

"Parasites" can be broadly defined as organisms that rely on a "host" in a mutual, but non-symbiotic relationship, to the detriment of the host organism. In their landmark set of papers in 1979, Robert M. May and Roy M. Anderson distinguished two classes of parasites, "microparasites" and "macroparasites". Microparasites include viruses, bacteria, and protozoans, which are small in size, reproduce quickly (generally within the host), and can induce immunity to reinfection in those hosts. Macroparasites include parasitic helminths and arthropods, which can have longer generation times, and tend to multiply outside the host or within the host (Anderson & May 1979; May & Anderson 1979).

The study of host-parasite interactions is built on the foundations of the studies of the behaviour, ecology, population dynamics, and physiology of both the host and the parasite, closely linked to the fundamentals of the theory of coevolution. True "coevolution" is defined as "an evolutionary change in a trait of the individuals in one population in response to a trait of the individuals of a second population, followed by an evolutionary response by the second population to the change in the first" (Janzen 1980).

The short life cycles of parasites, in general, allow them to accumulate beneficial mutations rapidly. Coevolution can result in a so-called escalating "arms race", in which parasites evolve to enable them to survive and multiply, while hosts battle to eliminate the parasites (or at least survive and reproduce in spite of their presence) (Anderson & May 1982; Morse 1994). Arms races

tend to be influenced by directional selection, maintained by frequency-dependent or density-dependent selection, typically seen in cases where host-parasite interactions are governed by multiple genes (Berenbaum & Zangerl 1998). Alternatively, though more rarely, when resistance is caused by the action of single genes, stable cycling can result (Seger & Antonovics 1988).

The Effects of Parasitism on Host Individuals

By definition, parasitism implies a negative effect on the host organism. Despite the complexity of the interaction on a molecular and biochemical scale, potential impacts to fitness can be quantified in experimental and observational studies. Studies have observed that high parasite loads can decrease host breeding success (Hudson 1986; Knowles *et al.* 2010a), alter life history traits (Agnew *et al.* 2000), and regulate populations (Albon *et al.* 2002), often by shifting efforts towards survival and immunity over reproduction (Knowles *et al.* 2010b; Lachish *et al.* 2011; Lochmiller & Deerenberg 2000). Parasitic interactions can become infinitely more complex, entailing the presence of multiple parasites simultaneously, and even behavioural manipulation to make hosts more vulnerable to predators (Lafferty & Morris 1996; Seppälä *et al.* 2004). Finally, parasitism can take the ultimate toll on an individual by causing mortal disease—with the extreme potential to limit the viability of a population or species (McCallum & Dobson 1995; Smith *et al.* 2009).

Parasitism and its burden are context-dependent. The baseline condition of hosts is an important variable in understanding the level of impact parasite burdens can inflict (Beldomenico *et al.* 2009). Food shortages can be determining factors for whether a given burden is sub-clinical, disease-inducing, or fatal (Gulland 1992). Many epizootics are only noted once the event has reached extreme levels, whether due to rapid scavenging and decay, or to the isolation of many wild populations away from scientific monitoring (McCallum & Dobson 1995).

An important distinction must be made between the so-called state of “disease” and that of sub-clinical infection. For example, in avian species, blood parasites such as *Haemoproteus* and *Plasmodium* infect hosts, resulting in an acute phase in which the highest costs to fitness occur,

including mortality. If infection does not lead to disease, or survival ensues, a chronic state can persist. Chronic (long-term) infections are common, with relatively lower levels of parasites, and may have relatively few costs to fitness due to host-acquired immunity (Lachish *et al.* 2011).

The Effects of Parasitism on Host Populations

Parasites are considered a major selective force within the animal kingdom (May & McLean 2007). They are competitors for host resources, and can therefore be a threat to the survival and persistence of a population or species (Smith *et al.* 2009; Smith *et al.* 2006). Mass mortality events and population declines have been attributed to infectious disease in various animal species, including epizootics of rabies in the endangered Ethiopian wolf (Randall *et al.* 2004; Sillero-Zubiri *et al.* 1996), canine distemper virus in seals (Forsyth *et al.* 1998; Kennedy *et al.* 2000; Mamaev *et al.* 1995) and lions (Roelke-Parker *et al.* 1996), bacterial *Pasteurella multocida* in endangered Saiga antelope (Kock *et al.* 2018), and plague in black-footed ferrets, prairie dogs, and other rodents (Perry & Fetherston 1997), among numerous other examples across vertebrate taxa (MacPhee & Greenwood 2013; Smith *et al.* 2006; Wikelski *et al.* 2004).

Deadly epizootics typically lead to population fluctuations, and are risk factors for population decline and local extinctions. In extreme cases though, parasites may be ultimately responsible for local and global species extinction (MacPhee & Greenwood 2013) as in the case of avian malaria in island-dwelling birds (LaPointe *et al.* 2012; Warner 1968), chytridiomycosis in amphibians (Cheng *et al.* 2011; Daszak *et al.* 1999; O'Hanlon *et al.* 2018; Rachowicz *et al.* 2006; Skerratt *et al.* 2007), and potentially, trypanosomiasis in the rodents of Christmas Island (Wyatt *et al.* 2008).

Compounding the effect of condition on disease state in individuals, environmental and anthropogenic changes that fragment a habitat or force shifts in the range of a host can also contribute to increased fitness costs and mortality. For example, in 1988, a major outbreak of phocine distemper virus occurred in North Sea seals, with a strain unique to Harp seals from the

Gulf of Labrador in the Western Atlantic. Investigations into the outbreak posit that the collapse of the cod fishery off the Canadian coast (or less likely, ocean warming) led to a range shift in Harp seals, which then dispersed East with the currents and spread the virus across the genus to a naïve population of North Sea seals with devastating effects (Heide-Jorgensen *et al.* 1992).

When a pathogen is shared by multiple host species, parasite-mediated (apparent) competition can also shape community structure of host species (Holt & Lawton 1994; Hudson & Greenman 1998). Host-specificity is important to investigate in order to understand how long a parasite and its host have been associated (Clayton & Moore 1997). Being limited to a single host tends to indicate that the parasite has been associated with this host for much of its evolutionary history, though the correlation cannot itself provide support of a coevolutionary process (Janzen 1980; Nuismer *et al.* 2010).

The ability of a parasite to thrive within several host species suggests it is a generalist, or has developed an ability to host shift, typically a process associated with a change in parasite virulence (Bull 1994; Toft & Karter 1990). Sympatry of host species is common among rapidly diverging lineages, which essentially provides a larger total population of hosts for generalist parasites, and an opportunity for specialist parasites to host shift and evolve in new hosts, potentially into separate parasite lineages (Ricklefs *et al.* 2004; Zietara *et al.* 2002).

Parasite diversity is also strongly influenced by diet. Parasites of prey may be ingested and thus move up trophic levels or have differing intermediate hosts. For example, the parasitic infection trichinosis in humans is caused by ingesting *Trichinella* tapeworm cysts found in raw or undercooked meat. Similarly, parasites of crustaceans, fish, molluscs and other marine prey are also found in seabirds (Galaktionov 2017; Khan *et al.* 2019). Parasites may also serve as prey themselves. Research into food webs has shown that parasite consumption may not be strictly accidental and may represent an important trophic link of biomass transfer (Johnson *et al.* 2010; Lafferty *et al.* 2006). Host dietary intake can therefore directly reduce parasitism or alter the gut microbiome. In turn, numerous studies in model animal systems have demonstrated that the

gastrointestinal microbiome can reduce or increase parasite resistance and susceptibility to disease (Harris *et al.* 2019).

Because of the various means by which parasites affect hosts, small shifts in individual fitness will ultimately lead to differential survival or changes in breeding success. This means that parasites lead to changes in genotypic frequencies in the host population. As part of the host-parasite arms race of developing defences and counter-defences, diversification of populations, and even speciation, is facilitated over time (Nunn *et al.* 2004). In fact, this coevolutionary process can lead to very high levels of genetic diversity within and between host and parasite populations, selecting for different host traits and even rare genotypes (Burdon & Thrall 1999; Carius *et al.* 2001; Dybdahl & Lively 1998).

The Penguin Immune System

The primary means of defence against parasites is a host's immune system—the collective of molecules, cells, tissues, and organs involved in fighting disease. Because of the important role of pathogen responses in evolution, I will briefly describe some of the primary pathogen-facing components of the avian immune system that are not covered elsewhere within the chapters.

The avian immune system shares some similarities with other vertebrates, though certain features are unique to birds, and can vary between extant bird lineages. All vertebrates have both an innate and adaptive component to their immune systems, which work together to recognise pathogens and eliminate them (Zimmerman *et al.* 2010).

The innate immune system is the "first line of defence" against pathogens and cell damage. It includes physical and chemical barriers like skin, mucosal epithelia, and gastric or respiratory secretions that can help a host avoid infection. It also includes cellular components (phagocytes like macrophages and dendritic cells) that recognise features common to many pathogens (Kaiser 2010). Pattern-recognition receptors (PRRs) are proteins expressed by innate immune cells that recognise conserved molecular signatures known as microbe-associated molecular patterns

(MAMPs) found in bacteria, viruses, fungi, and protozoans, and damage-associated molecular patterns (DAMPs) found in damaged host cells (Kawasaki & Kawai 2014).

PRRs are encoded in the germ-line and expressed in all cells of a given type, independently of the host's history of pathogen encounter (Akira *et al.* 2006). PRRs are the first to recognise a pathogenic threat to the host and activate signalling pathways to induce local inflammation, as well as recruiting adaptive immune effector cells to the site of infection. Some of the best-characterised PRRs are the Toll-like receptors (TLRs), which are a group of MAMP-recognising cell surface and endosomal receptors (Kawai & Akira 2006). The TLR family of proteins is well-conserved across vertebrates, though some TLR-coding genes have been duplicated or pseudogenised in different lineages, with birds typically having ten TLRs (Boyd *et al.* 2007; Velová *et al.* 2018). These receptor proteins recognise MAMPs like bacterial lipopolysaccharides (TLR4), flagellin (TLR5), double-stranded RNA (TLR3) and single-stranded RNA (TLR7), among other ligands (Brownlie & Allan 2011). The clear relationship between TLRs and the pathogens they recognise makes them of particular interest in the study of host-parasite coevolution and natural selection.

The adaptive immune system, on the other hand, can target an exponentially broader repertoire of antigens than the ten MAMPs recognised by TLRs. It also retains an immunological "memory" to facilitate recognition of previously encountered pathogens. Adaptive immunity (humoral and cell-mediated) is primarily governed by lymphocytes like B-cells and T-cells. In humoral immunity, B-lymphocytes, which originate and mature in the Bursa of Fabricius in birds, produce immunoglobulin molecules (IgA, IgM, IgY) present in the blood or gut as antibodies to identify and neutralise specific extracellular antigens (Kaiser 2010). Avian antibody repertoires are achieved through a unique process of gene conversion (Davison 2014). Crucially, recombination and somatic hypermutation provide avian hosts with orders of magnitude more adaptive immune receptors than the few germline-encoded recognition receptors of the innate immune system.

T-cells are the main effectors of adaptive immunity and mature in the avian thymus. CD4+ helper T-cell repertoires recognise peptide antigens on the pathogen surface to release cytokines that trigger complex pathways to break down infected cells (Degen *et al.* 2005). Specific cytokine pathways interface with innate (e.g. inflammation, macrophages) and adaptive components (CD8+ cytotoxic T-cells, eosinophils, and B-cells) to clear the pathogen (Degen *et al.* 2005) and regulate immune response (Kaiser 2010). T-cell receptors rely upon major histocompatibility complex (MHC) molecules to bind and present peptide fragments for the immune response (Smith & Göbel 2014). The MHC genetic locus in birds is much smaller than mammalian counterparts (chicken MHC is approximately 20 times smaller than human MHC, Kaufman *et al.* (1999)). This has facilitated population-level studies of MHC genetic diversity, including in some penguin species (Bollmer *et al.* 2007; Kikkawa *et al.* 2009; Knafler *et al.* 2012; Sallaberry-Pincheira *et al.* 2016).

Gentoo Penguins are likely to share many elements of the immune system as other birds, though there may be some differences that will emerge as new genomic data is studied in greater depth. Currently, Gentoo Penguins are known to at least share the same conserved anatomical features (Watts *et al.* 2009), immunoglobulins (Han *et al.* 2017), and TLRs (unpublished data). In this thesis, I build upon the limited knowledge of penguin immunity and disease by evaluating the genetic signatures of evolutionary selection over time in certain TLR genes to look at the past, while seeking to identify pathogens in the present that may shape future evolution of penguin populations.

Thesis Outline

This thesis aims to investigate whether geographic variation influences patterns of genetic diversity in Gentoo Penguins and their microbes and parasites. The work is presented in two sections: (I) The Gentoo Penguin as a Host Species and (II) Gentoo Penguin Pathogens and Microbes. Community-based genetic surveys of population samples allow us to measure host and

microbial diversity, while phylogenetic and statistical methods allow us to evaluate differential patterns across archipelagos, and across the Antarctic Polar Front.

In **Section I** (Host), **Chapter 2** elucidates the baseline population structure of Gentoo Penguins in the South Atlantic and Antarctic using microsatellite markers and mitochondrial DNA, while **Chapter 3**, uses population immunogenetics to identify patterns of differentiation in three bacterial- and viral-recognising innate immune loci, the genes coding for Toll-like receptors 4, 5, and 7, across a wider geographical range that also includes Indian Ocean breeding sites.

Based on the results of the immunogenetic study, which found evidence of natural selection in two bacterial pathogen-recognition proteins (TLR4 and TLR5), I subsequently attempt to characterise the potential pathogens present in Gentoo Penguins and their sympatric, congeneric species, the Chinstrap Penguin and Adélie Penguin (collectively, Pygoscelids). **Section II** (Gentoo Penguin Pathogens and Microbes) therefore begins with an overall review of known pathogens in Pygoscelid penguins (**Chapter 4**). Following that, and starting with the smallest microbes, **Chapters 5 and 6** highlight 99 novel viruses and viral-like circular molecules detected in penguins south of the Polar Front. **Chapters 7 and 8** delve into the bacterial microbiota of the penguin gastrointestinal tract, exploring potential pathogens that might reside among commensal bacteria, followed by a microscopic and molecular examination of eukaryotic gut parasites in **Chapter 9**.

Finally, bringing together knowledge of these interspecific biotic interactions, I look across the South Atlantic and Antarctic range of the species complex to examine how current patterns of microbe and pathogen distribution could shape the evolutionary trajectory of Gentoo Penguins in a changing world.

More focused introductions of each topic, with their relevant literature, are contained in **Chapters 2-8**.

References

Acosta Hospitaleche C, Reguero M, Scarano A (2013) Main pathways in the evolution of the Paleogene Antarctic Sphenisciformes. *Journal of South American Earth Sciences* **43**, 101-111.

Agnew P, J CK, Michalakis Y (2000) Host life history responses to parasitism. *Microbes Infect* **2**, 891-896.

Akira S, Uematsu S, Takeuchi O (2006) Pathogen Recognition and Innate Immunity. *Cell* **124**, 783-801.

Albon SD, Stien A, Irvine RJ*, et al.* (2002) The role of parasites in the dynamics of a reindeer population. *Proc Biol Sci* **269**, 1625-1632.

Almendros J, Carmona E, Jiménez V, Díaz-Moreno A, Lorenzo F (2018) Volcano-Tectonic Activity at Deception Island Volcano Following a Seismic Swarm in the Bransfield Rift (2014–2015). *Geophysical Research Letters* **45**, 4788-4798.

Anderson RM, May RM (1979) Population Biology of Infectious-Diseases .1. *Nature* **280**, 361-367.

Anderson RM, May RM (1982) Coevolution of hosts and parasites. *Parasitology* **85**, 411-426.

Baker AJ, Pereira SL, Haddrath OP, Edge KA (2006) Multiple gene evidence for expansion of extant penguins out of Antarctica due to global cooling. *Proc Biol Sci* **273**, 11-17.

Baylis AMM, Crofts S, Wolfaardt AC (2013) Population trends of Gentoo Penguins *Pygoscelis papua* breeding at the Falkland Islands. *Marine Ornithology* **41**, 1-5.

Beldomenico PM, Telfer S, Lukomski L*, et al.* (2009) Host condition and individual risk of cowpox virus infection in natural animal populations: cause or effect? *Epidemiol Infect* **137**, 1295-1301.

Berenbaum MR, Zangerl AR (1998) Chemical phenotype matching between a plant and its insect herbivore. *Proc Natl Acad Sci U S A* **95**, 13743-13748.

Bevan RM, Butler PJ, Woakes AJ, Boyd IL (2002) The energetics of Gentoo Penguins, *Pygoscelis papua*, during the breeding season. *Functional Ecology* **16**, 175-190.

BirdLife International (2018) *Pygoscelis papua*. *The IUCN Red List of Threatened Species* **2018**, e.T22697755A132600694.

Black C, Collen B, Johnston D, Hart T (2016) Why Huddle? Ecological Drivers of Chick Aggregations in Gentoo Penguins, *Pygoscelis papua*, across Latitudes. *PLoS One* **11**, e0145676.

Boelhouwers JC, Meiklejohn KI, Holness SD, Hedding DW (2007) Geology, geomorphology and climate change. In: *The Prince Edward Islands: Land-sea interactions in a changing ecosystem* (eds. Chown SL, Froneman PW), pp. 65-96. SUN PRESS, Stellenbosch.

Bollmer JL, Vargas FH, Parker PG (2007) Low MHC variation in the endangered Galápagos penguin (Spheniscus mendiculus). *Immunogenetics* **59**, 593-602.

Borowicz A, McDowall P, Youngflesh C*, et al.* (2018) Multi-modal survey of Adelie penguin mega-colonies reveals the Danger Islands as a seabird hotspot. *Sci Rep* **8**, 3926.

Bourne DG, Dennis PG, Uthicke S*, et al.* (2013) Coral reef invertebrate microbiomes correlate with the presence of photosymbionts. *The ISME Journal* **7**, 1452-1458.

Boyd A, Philbin VJ, Smith AL (2007) Conserved and distinct aspects of the avian Toll-like receptor (TLR) system: implications for transmission and control of bird-borne zoonoses. *Biochemical Society Transactions* **35**, 1504-1507.

Brownlie R, Allan B (2011) Avian toll-like receptors. *Cell and Tissue Research* **343**, 121-130.

Bull JJ (1994) Perspective: Virulence. *Evolution* **48**, 1423-1437.

Burdon JJ, Thrall PH (1999) Spatial and Temporal Patterns in Coevolving Plant and Pathogen Associations. *The American Naturalist* **153**, S15-S33.

Carius HJ, Little TJ, Ebert D (2001) Genetic variation in a host-parasite association: potential for coevolution and frequency-dependent selection. *Evolution* **55**, 1136-1145.

Carter A, Curtis M, Schwanethal J (2014) Cenozoic tectonic history of the South Georgia microcontinent and potential as a barrier to Pacific-Atlantic through flow. *Geology* **42**, 299-302.

Cheng TL, Rovito SM, Wake DB, Vredenburg VT (2011) Coincident mass extirpation of neotropical amphibians with the emergence of the infectious fungal pathogen Batrachochytrium dendrobatidis. *Proceedings of the National Academy of Sciences* **108**, 9502-9507.

Clayton DH, Moore J (1997) *Host-parasite Evolution: General Principles and Avian Models* Oxford University Press.

Clements JF, Schulenberg TS, Iliff MJ*, et al.* (2019) The eBird/Clements Checklist of Birds of the World: v2019.

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

Clucas GV, Younger JL, Kao D*, et al.* (2018) Comparative population genomics reveals key barriers to dispersal in Southern Ocean penguins. *Mol Ecol* **27**, 4680-4697.

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 Evol Biol* **16**, 211.

Cole TL, Dutoit L, Dussex N*, et al.* (2019a) Receding ice drove parallel expansions in Southern Ocean penguins. *Proc Natl Acad Sci U S A*.

Cole TL, Ksepka DT, Mitchell KJ*, et al.* (2019b) Mitogenomes Uncover Extinct Penguin Taxa and Reveal Island Formation as a Key Driver of Speciation. *Molecular Biology and Evolution* **36**, 784-797.

Convey P, Morton A, Poncet J (2009) Survey of marine birds and mammals of the South Sandwich Islands. *Polar Record* **35**, 107-124.

Crawford RJM, Cooper J, Du Toit M*, et al.* (2003) Population and Breeding of the Gentoo Penguin *Pygoscelis papua* at Marion Island, 1994/95 – 2002/03. *African Journal of Marine Science* **25**, 463-474.

Crawford RJM, Dyer BM, Upfold L, Makhado AB (2014) Congruent, decreasing trends of gentoo penguins and Crozet shags at sub-Antarctic Marion Island suggest food limitation through common environmental forcing. *African Journal of Marine Science* **36**, 225-231.

Crawford RJM, Whittington PA, Upfold L*, et al.* (2009) Recent trends in numbers of four species of penguins at the Prince Edward Islands. *African Journal of Marine Science* **31**, 419-426.

Daszak P, Berger L, Cunningham AA*, et al.* (1999) Emerging infectious diseases and amphibian population declines. *Emerging Infectious Diseases* **5**, 735-748.

Davison F (2014) The Importance of the Avian Immune System and its Unique Features. In: *Avian Immunology (Second Edition)* (eds. Schat KA, Kaspers B, Kaiser P), pp. 1-9. Academic Press, Boston.

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

Degen WGJ, Daal Nv, Rothwell L, Kaiser P, Schijns VEJC (2005) Th1/Th2 polarization by viral and helminth infection in birds. *Veterinary Microbiology* **105**, 163-167.

Dobson FS, Jouventin P (2017) Why penguins communicate : the evolution of visual and vocal signals, p. 1 online resource. Academic Press,, Amsterdam.

Dybdahl MF, Lively CM (1998) Host-Parasite Coevolution: Evidence for Rare Advantage and Time-Lagged Selection in a Natural Population. *Evolution* **52**, 1057-1066.

Elliott KH, Ricklefs RE, Gaston AJ*, et al.* (2013) High flight costs, but low dive costs, in auks support the biomechanical hypothesis for flightlessness in penguins. *Proceedings of the National Academy of Sciences* **110**, 9380-9384.

Ford EB (1964) *Ecological genetics* Methuen ; Wiley, London ; New York.

Forsyth MA, Kennedy S, Wilson S, Eybatov T, Barrett T (1998) Canine distemper virus in a Caspian seal. *Vet Rec* **143**, 662-664.

Fraser CI, Nikula R, Ruzzante DE, Waters JM (2012) Poleward bound: biological impacts of Southern Hemisphere glaciation. *Trends in Ecology & Evolution* **27**, 462-471.

Freer J, Mable B, Clucas G*, et al.* (2015) Limited genetic differentiation among chinstrap penguin (*Pygoscelis antarctica*) colonies in the Scotia Arc and Western Antarctic Peninsula.

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

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

Friesen VL, Burg TM, McCoy KD (2007) Mechanisms of population differentiation in seabirds. *Mol Ecol* **16**, 1765-1785.

Galaktionov KV (2017) Patterns and processes influencing helminth parasites of Arctic coastal communities during climate change. *Journal of Helminthology* **91**, 387-408.

Gavryushkina A, Heath TA, Ksepka DT*, et al.* (2017) Bayesian Total-Evidence Dating Reveals the Recent Crown Radiation of Penguins. *Systematic Biology* **66**, 57-73.

Ghys MI, Rey AR, Schiavini A (2008) Population Trend and Breeding Biology of Gentoo Penguin in Martillo Island, Tierra Del Fuego, Argentina. *Waterbirds* **31**, 625-631, 627.

Gilbert C, Blanc S, Le Maho Y, Ancel A (2008) Energy saving processes in huddling emperor penguins: from experiments to theory. *Journal of Experimental Biology* **211**, 1-8.

Global Volcanism Program (2013) Volcanoes of the World v. 4.8.7 (ed. Venzke E). Smithsonian Institution.

Groscolas R (1986) Changes in body mass, body temperature and plasma fuel levels during the natural breeding fast in male and female emperor penguinsAptenodytes forsteri. *Journal of Comparative Physiology B* **156**, 521-527.

Groscolas R, Robin J-P (2001) Long-term fasting and re-feeding in penguins. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology* **128**, 643-653.

Gulland FMD (1992) The role of nematode parasites in Soay sheep (Ovis aries L.) mortality during a population crash. *Parasitology* **105**, 493-503.

Han B, Li Y, Han H*, et al.* (2017) Three IgH isotypes, IgM, IgA and IgY are expressed in Gentoo penguin and zebra finch. *PLoS One* **12**, e0173334.

Handrich Y, Bevan RM, Charrassin JB*, et al.* (1997) Hypothermia in foraging king penguins. *Nature* **388**, 64-67.

Harris EV, de Roode JC, Gerardo NM (2019) Diet–microbiome–disease: Investigating diet's influence on infectious disease resistance through alteration of the gut microbiome. *PLOS Pathogens* **15**, e1007891.

Heide-Jorgensen M-P, Harkonen T, Dietz R, Thompson PM (1992) Retrospective of the 1988 European seal epizootic. *Diseases of Aquatic Organisms* **13**, 37-62.

Holmes ND, Giese M, Achurch H, Robinson S, Kriwoken LK (2006) Behaviour and breeding success of gentoo penguins *Pygoscelis papua* in areas of low and high human activity. *Polar Biology* **29**, 399-412.

Holt RD, Lawton JH (1994) The Ecological Consequences of Shared Natural Enemies. *Annual Review of Ecology and Systematics* **25**, 495-520.

Hudson P, Greenman J (1998) Competition mediated by parasites: biological and theoretical progress. *Trends in Ecology & Evolution* **13**, 387-390.

Hudson PJ (1986) The Effect of a Parasitic Nematode on the Breeding Production of Red Grouse. *Journal of Animal Ecology* **55**, 85-92.

Janzen DH (1980) When is it coevolution? *Evolution* **34**, 611-612.

Johnson PTJ, Dobson A, Lafferty KD*, et al.* (2010) When parasites become prey: ecological and epidemiological significance of eating parasites. *Trends in Ecology & Evolution* **25**, 362-371.

Jouventin P (1994) Les populations d'oiseaux marins des T.A.A.F.: résumé de 20 années de recherche. *Alauda* **62**, 44-47.

Kaiser P (2010) Advances in avian immunology—prospects for disease control: a review. *Avian Pathology* **39**, 309-324.

Kaufman J, Milne S, Göbel TWF*, et al.* (1999) The chicken B locus is a minimal essential major histocompatibility complex. *Nature* **401**, 923-925.

Kawai T, Akira S (2006) TLR signaling. *Cell Death & Differentiation* **13**, 816-825.

Kawasaki T, Kawai T (2014) Toll-Like Receptor Signaling Pathways. *Frontiers in Immunology* **5**.

Kennedy S, Kuiken T, Jepson PD*, et al.* (2000) Mass die-Off of Caspian seals caused by canine distemper virus. *Emerg Infect Dis* **6**, 637-639.

Khan JS, Provencher JF, Forbes MR*, et al.* (2019) Parasites of seabirds: A survey of effects and ecological implications. In: *Advances in Marine Biology* (ed. Sheppard C), pp. 1-50. Academic Press.

Kikkawa EF, Tsuda TT, Sumiyama D*, et al.* (2009) Trans-species polymorphism of the Mhc class II DRB-like gene in banded penguins (genus Spheniscus). *Immunogenetics* **61**, 341-352.

Knafler GJ, Clark JA, Boersma PD, Bouzat JL (2012) MHC diversity and mate choice in the magellanic penguin, Spheniscus magellanicus. *J Hered* **103**, 759-768.

Knowles SC, Palinauskas V, Sheldon BC (2010a) Chronic malaria infections increase family inequalities and reduce parental fitness: experimental evidence from a wild bird population. *J Evol Biol* **23**, 557-569.

Knowles SC, Wood MJ, Sheldon BC (2010b) Context-dependent effects of parental effort on malaria infection in a wild bird population, and their role in reproductive trade-offs. *Oecologia* **164**, 87-97.

Kock RA, Orynbayev M, Robinson S*, et al.* (2018) Saigas on the brink: Multidisciplinary analysis of the factors influencing mass mortality events. *Science Advances* **4**, eaao2314.

Ksepka DT, Bertelli S, Giannini NP (2006) The phylogeny of the living and fossil Sphenisciformes (penguins). *Cladistics* **22**, 412-441.

Ksepka DT, Fordyce RE, Ando T, Jones CM (2012) New fossil penguins (Aves, Sphenisciformes) from the Oligocene of New Zealand reveal the skeletal plan of stem penguins. *Journal of Vertebrate Paleontology* **32**, 235-254.

Ksepka DT, Werning S, Sclafani M, Boles ZM (2015) Bone histology in extant and fossil penguins (Aves: Sphenisciformes). *Journal of anatomy* **227**, 611-630.

Lachish S, Knowles SC, Alves R, Wood MJ, Sheldon BC (2011) Fitness effects of endemic malaria infections in a wild bird population: the importance of ecological structure. *J Anim Ecol* **80**, 1196-1206.

Lafferty KD, Dobson AP, Kuris AM (2006) Parasites dominate food web links. *Proceedings of the National Academy of Sciences* **103**, 11211-11216.

Lafferty KD, Morris AK (1996) Altered Behavior of Parasitized Killifish Increases Susceptibility to Predation by Bird Final Hosts. *Ecology* **77**, 1390-1397.

LaPointe DA, Atkinson CT, Samuel MD (2012) Ecology and conservation biology of avian malaria. *Annals of the New York Academy of Sciences* **1249**, 211-226.

Le Maho Y (1977) The Emperor Penguin: A Strategy to Live and Breed in the Cold: Morphology, physiology, ecology, and behavior distinguish the polar emperor penguin from other penguin species, particularly from its close relative, the king penguin. *American Scientist* **65**, 680-693.

Le Maho Y (2005) How penguins cope with cold and lack of food. *Alauda* **73**, 207-209.

Lescroël A, Bajzak C, Bost C-A (2009) Breeding ecology of the gentoo penguin *Pygoscelis papua* at Kerguelen Archipelago. *Polar Biology* **32**, 1495.

Lescroël A, Bost C-A (2006) Recent decrease in gentoo penguin populations at Iles Kerguelen. *Antarctic Science* **18**, 171-174.

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**, 1834-1853.

Lochmiller RL, Deerenberg C (2000) Trade-offs in evolutionary immunology: just what is the cost of immunity? *Oikos* **88**, 87-98.

Lynch H, White R, Black A, Naveen R (2012) Detection, differentiation, and abundance estimation of penguin species by high-resolution satellite imagery. *Polar Biology* **35**, 963-968.

Lynch HJ (2013) Gentoo Penguin (*Pygoscelis papua*). In: *Penguins: Natural History and Conservation* (eds. Borboroglu PG, Boersma PD), pp. 73-88. University of Washington Press, Seattle, WA.

MacPhee RDE, Greenwood AD (2013) Infectious Disease, Endangerment, and Extinction. *International Journal of Evolutionary Biology* **2013**, 571939.

Mamaev LV, Denikina NN, Belikov SI*, et al.* (1995) Characterisation of morbilliviruses isolated from Lake Baikal seals (Phoca sibirica). *Vet Microbiol* **44**, 251-259.

Mathews GM (1927) *Systema avium Australasianarum: A systematic list of the birds of the Australasian region* British Ornithological Union, London.

May RM, Anderson RM (1979) Population Biology of Infectious-Diseases .2. *Nature* **280**, 455-461.

May RM, McLean AR (2007) *Theoretical ecology : principles and applications*, 3rd edn. Oxford University Press, Oxford ; New York.

Mayr E (1942) *Systematics and the Origin of Species: from the Viewpoint of a Zoologist* Columbia University Press, New York.

McCallum H, Dobson A (1995) Detecting disease and parasite threats to endangered species and ecosystems. *Trends in Ecology & Evolution* **10**, 190-194.

Morse SS (1994) *The Evolutionary biology of viruses* Raven Press, New York.

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

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

Murphy RC (1936) *Oceanic Birds of South America* MacMillan, New York.

Murphy RC (1947) A new zonal race of the Gentoo Penguin. *Auk* **64**, 454-455.

Nuismer SL, Gomulkiewicz R, Ridenhour BJ (2010) When is correlation coevolution? *Am Nat* **175**, 525-537.

Nunn CL, Altizer S, Sechrest W*, et al.* (2004) Parasites and the Evolutionary Diversification of Primate Clades. *The American Naturalist* **164**, S90-S103.

O'Hanlon SJ, Rieux A, Farrer RA*, et al.* (2018) Recent Asian origin of chytrid fungi causing global amphibian declines. *Science* **360**, 621-627.

Pan H, Cole TL, Bi X*, et al.* (2019) High-coverage genomes to elucidate the evolution of penguins. *Gigascience* **8**.

Perry RD, Fetherston JD (1997) Yersinia pestis--etiologic agent of plague. *Clin Microbiol Rev* **10**, 35-66.

Pertierra LR, Segovia NI, Noll D*, et al.* (2020) Cryptic speciation in gentoo penguins is driven by geographic isolation and regional marine conditions: Unforeseen vulnerabilities to global change. *Diversity and Distributions* **26**, 958-975.

Peters JL (1931) *Check-List of Birds of the World* Harvard University Press, Cambridge, MA.

Proffitt JV, Clarke JA, Scofield RP (2016) Novel insights into early neuroanatomical evolution in penguins from the oldest described penguin brain endocast. *Journal of anatomy* **229**, 228-238.

Quilty PG (2007) Origin and evolution of the sub-Antarctic islands: the foundation. *Papers and Proceedings of the Royal Society of Tasmania* **141**, 35-58.

Rachowicz LJ, Knapp RA, Morgan JAT*, et al.* (2006) Emerging Infectious Disease as a Proximate Cause of Amphibian Mass Mortality. *Ecology* **87**, 1671-1683.

Randall DA, Williams SD, Kuzmin IV*, et al.* (2004) Rabies in Endangered Ethiopian Wolves. *Emerging Infectious Diseases* **10**, 2214-2217.

Ratcliffe N, Trathan P (2012) A review of the diet and at-sea distribution of penguins breeding within the CAMLR Convention Area. *CCAMLR Science* **19**, 75-114.

Reusch TBH, Wood TE (2007) Molecular ecology of global change. *Mol Ecol* **16**, 3973-3992.

Ricklefs RE, Fallon SM, Bermingham E (2004) Evolutionary Relationships, Cospeciation, and Host Switching in Avian Malaria Parasites. *Systematic Biology* **53**, 111-119.

Roeder AD, Marshall RK, Mitchelson AJ*, et al.* (2001) Gene flow on the ice: genetic differentiation among Adelie penguin colonies around Antarctica. *Mol Ecol* **10**, 1645-1656.

Roelke-Parker ME, Munson L, Packer C*, et al.* (1996) A canine distemper virus epidemic in Serengeti lions (Panthera leo). *Nature* **379**, 441-445.

Sallaberry-Pincheira N, Gonzalez-Acuna D, Padilla P*, et al.* (2016) Contrasting patterns of selection between MHC I and II across populations of Humboldt and Magellanic penguins. *Ecol Evol* **6**, 7498-7510.

Sclater PL (1888) Notes on the Emperor Penguin (Aptenodytes. forsteri). *Ibis* **30**, 325-334.

Seger J, Antonovics J (1988) Dynamics of Some Simple Host-Parasite Models with More than Two Genotypes in Each Species [and Discussion]. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **319**, 541-555.

Seppälä O, Karvonen A, Tellervo Valtonen E (2004) Parasite-induced change in host behaviour and susceptibility to predation in an eye fluke–fish interaction. *Animal Behaviour* **68**, 257-263.

Sillero-Zubiri C, King AA, Macdonald DW (1996) Rabies and mortality in Ethiopian wolves (Canis simensis). *J Wildl Dis* **32**, 80-86.

Skerratt LF, Berger L, Speare R*, et al.* (2007) Spread of Chytridiomycosis Has Caused the Rapid Global Decline and Extinction of Frogs. *Ecohealth* **4**, 125.

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.

Smith AL, Göbel TW (2014) Avian T Cells: Antigen Recognition and Lineages. In: *Avian Immunology (Second Edition)* (eds. Schat KA, Kaspers B, Kaiser P), pp. 91-102. Academic Press, Boston.

Smith KF, Acevedo-Whitehouse K, Pedersen AB (2009) The role of infectious diseases in biological conservation. *Animal Conservation* **12**, 1-12.

Smith KF, Sax DF, Lafferty KD (2006) Evidence for the role of infectious disease in species extinction and endangerment. *Conserv Biol* **20**, 1349-1357.

Stonehouse B (1970) Geographic Variation in Gentoo Penguins *Pygoscelis papua*. *Ibis* **112**, 52-57.

Toft CA, Karter AJ (1990) Parasite-host coevolution. *Trends in Ecology & Evolution* **5**, 326-329.

Trathan PN, Daunt FJH, Murphy EJ (1996) *South Georgia: An Ecological Atlas* British Antarctic Survey, Cambridge.

University of Oxford (2015) *Secret life of penguins revealed for World Penguin Day.* http://www.ox.ac.uk/news/2015-04-25-secret-life-penguins-revealed-world-penguin-day

Varne R, Rubenbach MJ (1978) Geology of Macquarie Island and Its Relationship to Oceanic Crust. In: *Antarctica Oceanology II: The Australian–New Zealand Sector* (ed. Hayes DE), pp. 251-266.

Velová H, Gutowska-Ding MW, Burt DW, Vinkler M (2018) Toll-like receptor evolution in birds: gene duplication, pseudogenisation and diversifying selection. *Mol Biol Evol*.

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

Warner RE (1968) The Role of Introduced Diseases in the Extinction of the Endemic Hawaiian Avifauna. *The Condor* **70**, 101-120.

Watts JM, Miller GD, Shellam GR (2009) Infectious Bursal Disease Virus and Antarctic Birds. In: *Health of Antarctic Wildlife: A Challenge for Science and Policy* (eds. Kerry KR, Riddle M), pp. 95-105. Springer Berlin Heidelberg, Berlin, Heidelberg.

Weimerskirch H, Zotier R, Jouventin P (1989) The Avifauna of the Kerguelen Islands. *Emu* **89**, 15-29.

Wikelski M, Foufopoulos J, Vargas H, Snell H (2004) Galápagos Birds and Diseases: Invasive Pathogens as Threats for Island Species. *Ecology and Society* **9**.

Woehler EJ (1993) *The Distribution and Abundance of Antarctic and Subantarctic Penguins* Scientific Committe on Antarctic Research, Cambridge.

Wright S (1949) The Genetical Structure of Populations. *Annals of Eugenics* **15**, 323-354.

Wyatt KB, Campos PF, Gilbert MTP*, et al.* (2008) Historical Mammal Extinction on Christmas Island (Indian Ocean) Correlates with Introduced Infectious Disease. *PLoS One* **3**, e3602.

Zietara MS, Lumme J, Poulin R (2002) Speciation by host switch and adaptive radiation in a fish parasite genus Gyrodactylus (Monogenea, Gyrodactylidae). *Evolution* **56**, 2445-2458.

Zimmerman LM, Vogel LA, Bowden RM (2010) Understanding the vertebrate immune system: insights from the reptilian perspective. *The Journal of Experimental Biology* **213**, 661-671.

Section I. The Gentoo Penguin as a Host Species

Foreword

In this section, the population genetics of *Pygoscelis papua* populations are considered at both fine and large geographic scales, starting with mitochondrial DNA and microsatellites as neutral genetic markers in **Chapter 2**, then followed by evaluations of the structure and diversity of functional immune genes, three Toll-like receptors, in **Chapter 3.** This section aims to showcase the evolutionary history of this species complex prior to exploring its current pathogen challenges.

These two chapters are reproduced here from their published form. Author contribution sections are found at the end of each chapter. *These authors are joint first authors.

- **Chapter 2**
    - **Levy, H.***, Clucas, G.V.*, Rogers, A.D., Leaché, A.D., Ciborowski, K.L., Polito, M.J., Lynch, H.J., Dunn, M.J. and Hart, T. (2016), Population structure and phylogeography of the Gentoo Penguin (*Pygoscelis papua*) across the Scotia Arc. *Ecology and Evolution*, 6: 1834-1853. https://doi.org/10.1002/ece3.1929
- **Chapter 3**
    - **Levy, H.***, Fiddaman, S.R.,* Vianna, J.A., Noll, D., Clucas, G.V., Sidhu, J.K.H., Polito, M.J., Bost, C.A., Phillips, R.A., Crofts, S., Miller, G.D., Pistorius, P., Bonnadonna, F., Le Bohec, C., Barbosa, A., Trathan, P., Raya Rey, A., Frantz, L.A.F., Hart, T., and Smith, A.L. (2020), Evidence of Pathogen-Induced Immunogenetic Selection across the Large Geographic Range of a Wild Seabird. *Molecular Biology and Evolution*, 6: 1708-1726. https://doi.org/10.1093/molbev/msaa040

Chapter 2. 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] and 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, OH 45433-7765 USA
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, WA 98195-1800, USA
5. Department of Biology, Woodland Road, Bristol, BS8 1UG, UK
6. Department of Oceanography and Coastal Sciences, Louisiana State University, Baton Rouge, LA 70803 USA
7. Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY 11794 USA
8. British Antarctic Survey, High Cross, Madingley Road, Cambridge, CB3 0ET, UK

†Author for correspondence:

Hila Levy: hila.levy@zoo.ox.ac.uk

* These authors have provided equal contribution and share first authorship.

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

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 Gentoo Penguins expanded from a possible glacial refuge in the centre 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. Palaeoecological 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 **Figure 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 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 Penguin (*Pygoscelis papua papua*) breeds on the Falkland Islands, whilst the Southern Gentoo Penguin (*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).

Figure 4. Adult Gentoo Penguins (*Pygoscelis papua*) guard their chicks on the nest in the Western Antarctic Peninsula. (Photo: Hila Levy).

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 vagrant 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 moulted 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.

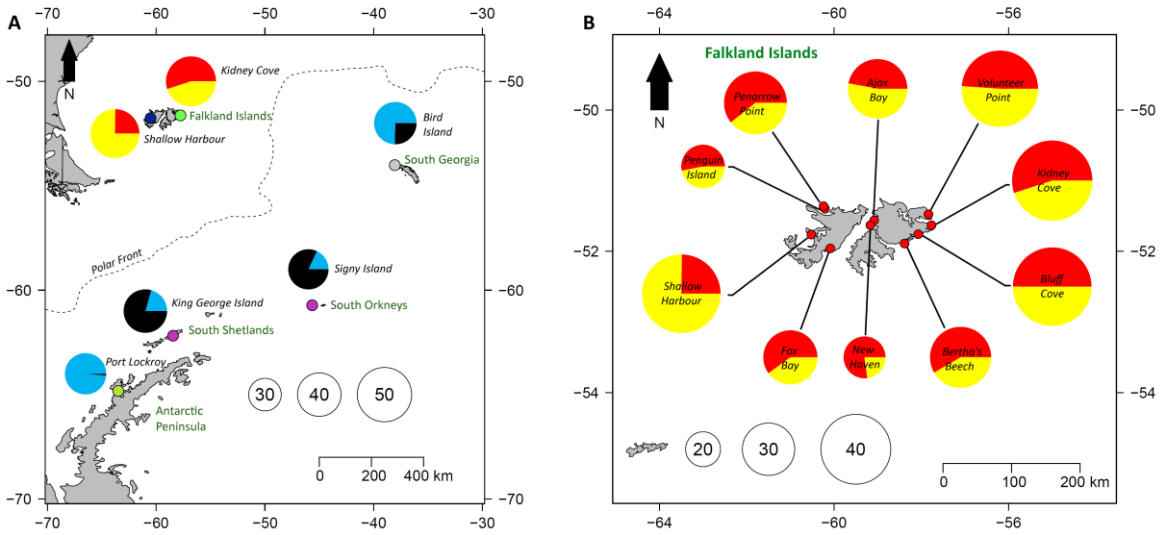


Figure 2. Population structure (pie charts) and locations of Gentoo Penguin (*Pygoscelis papua*) breeding sites [different colours in (A), red circles in (B)] sampled in this study across the Scotia Arc. Pie chart colours denote different genetic clusters as identified by STRUCTURE, and the size of each 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 **Figure 4B**). Location markers are coloured 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 **Figure 4B**).

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 moulted 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 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 Penguin 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 Analyser. 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.

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

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} (Weir and Cockerham 1984) and we used the SGoF+ method (Carvajal-Rodríguez 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} 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 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 α = 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 analysed 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 modelled 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, Figure 3**). Locus-by-locus diversity measures for each sampling site are shown in **Supplementary Table S1**.

Table 2. Genetic diversity of Gentoo Penguins (*Pygoscelis papua*) at fourteen breeding sites across the Scotia Arc. 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. See **Supplementary Table 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

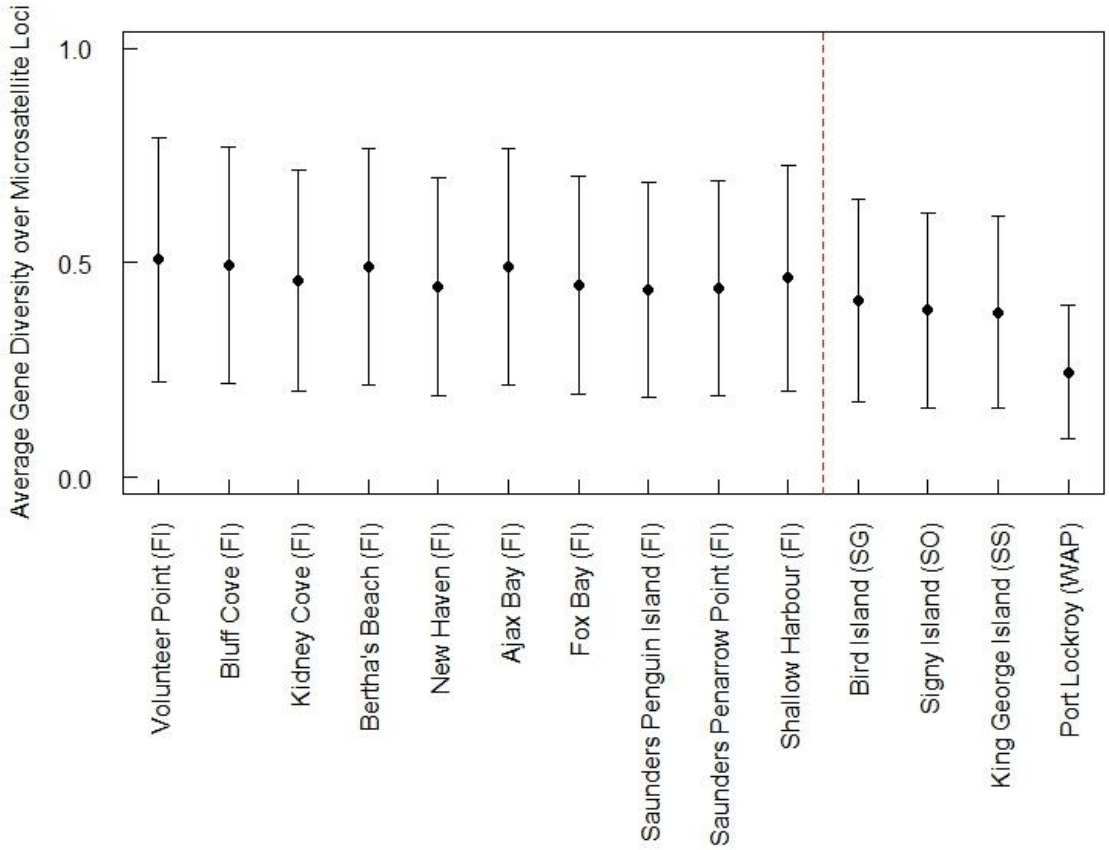


Figure 3. Genetic diversity of Gentoo Penguins (*Pygoscelis papua*) at fourteen breeding sites across the Scotia Arc. Mean observed (H_O) heterozygosity is shown, along with standard deviation (SD), over eight microsatellite loci for all individuals. FI = Falkland Islands, SG = South Georgia, SO = South 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**.

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 behaviour, 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 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 Gentoo Penguins, 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 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. FI = Falkland Islands, SG = South Georgia, SO = South Orkney Islands, SS = South Shetland Islands, and WAP = Western Antarctic Peninsula.

	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)**	*

Table 4. Analysis of Molecular Variance (AMOVA) of microsatellite data from Gentoo Penguin populations of the Scotia Arc, when grouped by varying assignment criteria.

Grouping Criteria	Within population % variation, F_{ST} (*P*-value)	Among population % variation, F_{SC} (*P*-value)	Among groups % variation, F_{CT} (*P*-value)
<u>14 populations of the Scotia Arc</u>			
1 group	88.28, 0.11721 (<0.001)		
2 groups by sub-species (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, 0.38218711 (<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)
<u>10 populations of the Falkland Islands</u>			
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. *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.

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 sub-species / all Falkland Islands)	101	40	25	0.9228 (0.0157)	0.00906 (0.00533)
P. p. ellsworthii (southern sub-species)	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)

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.

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 (**Figure 4**), except under the No Admixture model for correlated allele frequencies, with no *a priori* location data, when *K* = 4 was favoured. The two-population split most strongly coincided with the Northern and Southern Gentoo Penguin 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**.

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

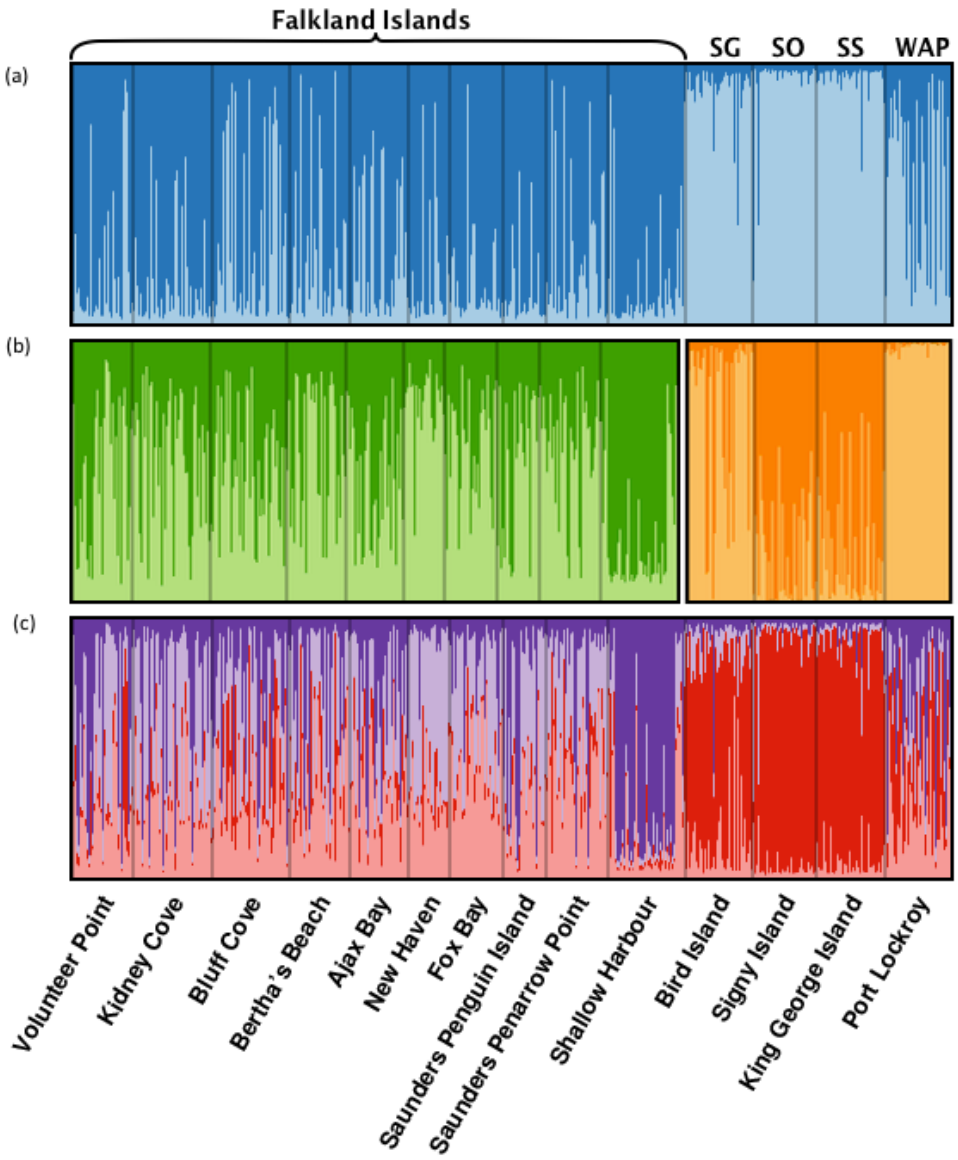


Figure 4. Plots of assignment probabilities from STRUCTURE showing the posterior probability of assigning each individual to each of the inferred clusters. Each individual is represented by a vertical bar, and the colours refer to the different clusters. All plots were generated from 10 runs using the 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 analysed 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 Gentoo Penguins. SG = South Georgia, SO = South Orkney Islands, SS = South Shetland Islands, and WAP = Western Antarctic Peninsula.

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 Gentoo Penguins (*P. papua ellsworthii*), with all Falkland individuals grouping together, as a separate clade from individuals

south of the Polar Front. The Southern Gentoo Penguins 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 **Supplementary Figure 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.

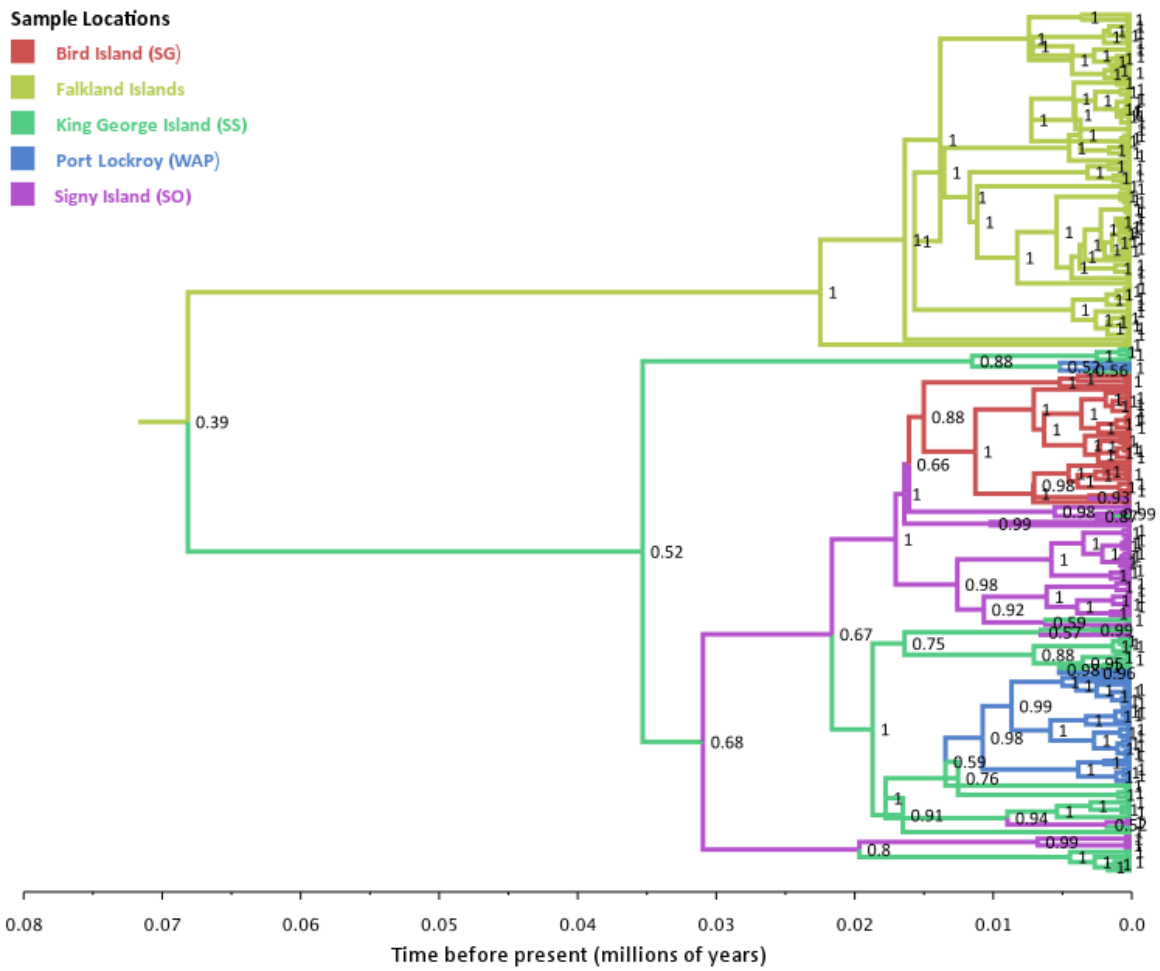


Figure 5. Maximum clade credibility tree derived from mtDNA showing the origin and differentiation of *Pygoscelis papua* lineages north (Falkland Islands, light green above) and south of the Polar Front (all other colours and locations). Node colours represent the most likely location of each ancestral 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.

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 (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 Gentoo Penguins 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 Gentoo Penguins 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 centre 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 Penguin'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 analysed 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 Penguin (*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 Rockhopper Penguins. 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 flow 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, behavioural, 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 inter-annual 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 ± 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 Gentoo Penguins 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.

Ethics Statement

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.

Acknowledgments

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.

Author Contributions

HL and TH designed the study and performed field work in the Falkland Islands. HL and GVC and equally performed lab work, analysed the data, and wrote the paper. ADL assisted with genetic data analysis and Bayesian methodology. TH and ADR managed the project and provided lab facilities. KLC provided consumables and advice on microsatellite analysis. MJP and MJD provided samples from King George Island and Signy Island. All authors participated in drafting and revising the manuscript. All authors gave final approval on this manuscript. *These authors are joint first authors.

Conflict of Interest

None declared.

Data Accessibility

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. Microsatellite genotyping data, along with GPS coordinates for sample sites are available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.84c78.

References

Ahmed S, Hart T, Dawson DA*, et al.* (2009) Isolation and characterization of macaroni penguin (*Eudyptes chrysolophus*) microsatellite loci and their utility in other penguin species (Spheniscidae, AVES). *Molecular Ecology Resources* **9**, 1530-1535.

Ainley D, Russell J, Jenouvrier S*, et al.* (2010) Antarctic penguin response to habitat change as Earth's troposphere reaches 2 degrees C above preindustrial levels. *Ecological Monographs* **80**, 49-66.

Akçakaya HR, Mills G, Doncaster CP (2007) The role of metapopulations in conservation. In: *Key Topics in Conservation* (eds. Macdonald DW, Service K), pp. 64-84. Blackwell, Oxford, UK.

Akst EP, Boersma PD, Fleischer RC (2002) A comparison of genetic diversity between the Galapagos Penguin and the Magellanic Penguin. *Conservation Genetics* **3**, 375-383.

Allcock AL, Strugnell JM (2012) Southern Ocean diversity: new paradigms from molecular ecology. *Trends Ecol Evol* **27**, 520-528.

Barnes DKA, Hodgson DA, Convey P, Allen CS, Clarke A (2006) Incursion and excursion of Antarctic biota: past, present and future. *Global Ecology and Biogeography* **15**, 121-142.

Baylis AMM, Crofts S, Wolfaardt AC (2013) Population trends of Gentoo Penguins *Pygoscelis papua* breeding at the Falkland Islands. *Marine Ornithology* **41**, 1-5.

Bennie J, Hodgson JA, Lawson CR, et al. (2013) Range expansion through fragmented landscapes under a variable climate. *Ecology Letters* **16**, 921-929.

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*. *Molecular Ecology Notes* **7**, 425-427.

Black, CE (2016) Phenology of *Pygoscelis* penguins: A review of current and historic dates. *Polar Biology*.

Boersma PD (2008) Penguins as Marine Sentinels. *BioScience*. **58**, 597-607.

Boessenkool S, King TM, Seddon PJ, Waters JM (2008) Isolation and characterization of microsatellite loci from the yellow-eyed penguin (*Megadyptes antipodes*). *Molecular Ecology Resources* **8**, 1043-1045.

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*). *Molecular Ecology* **18**, 2390-2400.

Bost CA, Jouventin P (1991) The breeding performance of the Gentoo Penguin *Pygoscelis papua* at the northern edge of its range. *Ibis* **133**, 14-25.

Brookfield JFY (1996) A simple new method for estimating null allele frequency from heterozygote deficiency. *Molecular Ecology* **5**, 453-455.

Brown RM, Jordan WC (2009) Characterization of polymorphic microsatellite loci from Round Island Petrels (*Pterodroma arminjoniana*) and their utility in other seabird species. *Journal of Ornithology* **150**, 925-929.

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(9)**, e24700.

Casanovas P, Naveen R, Forrest S, Poncet J, Lynch H (2015) A comprehensive coastal seabird survey maps out the front lines of ecological change on the western Antarctic Peninsula. *Polar Biology* **38**, 927-940.

Chapuis MP, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution* **24**, 621-631.

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

Chown SL, Clarke A, Fraser CI*, et al.* (2015) The changing form of Antarctic biodiversity. *Nature* **522**, 431-438.

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.

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

Cook AJ, Vaughan DG (2010) Overview of areal changes of the ice shelves on the Antarctic Peninsula over the past 50 years. *Cryosphere* **4**, 77-98.

Convey P, Barnes DKA, Griffiths HJ, Grant SM, Linse K, Thomas DN (2012) Biogeography and regional classifications of Antarctica. In: *Antarctic Ecosystems: An extreme environment in a changing world* (ed. Rogers AD, Johnston NM, Murphy EJ, Clarke A), pp. 469-491. Blackwell Publishing Ltd.

Croxall JP, McCann TS, Prince PA, Rothery P (1988) Reproductive Performance of Seabirds and Seals at South Georgia and Signy Island, South Orkney Islands, 1976–1987: Implications for Southern Ocean Monitoring Studies. In: *Antarctic Ocean and Resources Variability* (ed. Sahrhage D), pp. 261-285. Springer Berlin Heidelberg.

Croxall JP, Trathan PN, Murphy EJ (2002) Environmental Change and Antarctic Seabird Populations. *Science* **297**, 1510-1514.

Dann P, Sidhu LA, Jessop R*, et al.* (2014) Effects of flipper bands and injected transponders on the survival of adult Little Penguins *Eudyptula minor*. *Ibis* **156**, 73-83.

Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* **9**, 772.

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

de Dinechin M, Dobson FS, Zehtindjiev P*, et al.* (2012) The biogeography of Gentoo Penguins (P*ygoscelis papua*). *Canadian Journal of Zoology-Revue Canadienne De Zoologie* **90**, 352-360.

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* **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.

Dugger KM, Ballard G, Ainley DG, Lyver POB, Schine C (2014) Adélie penguins coping with environmental change: results from a natural experiment at the edge of their breeding range. *Frontiers in Ecology and Evolution* **2**.

Earl DA, Vonholdt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4**, 359-361.

Emslie SD (2001) Radiocarbon dates from abandoned penguin colonies in the Antarctic Peninsula region. *Antarctic Science* **13**, 289-295.

Emslie SD, Coats L, Licht K (2007) A 45,000 yr record of Adélie penguins and climate change in the Ross Sea, Antarctica. *Geology* **35**, 61-64.

Enticott JW (1986) Distribution of penguins at sea in the South-eastern Atlantic and Southwestern Indian Oceans. *Cormorant* **13**, 114-142.

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, Laval G, Schneider S (2005) Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics* **1**, 47-50.

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.

Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes* **7**, 574-578.

Ferreira MAR, Suchard MA (2008) Bayesian analysis of elapsed times in continuous-time Markov chains. *Canadian Journal of Statistics* **36**, 355-368.

Fox AJ, Vaughan DG (2005) The retreat of Jones Ice Shelf, Antarctic Peninsula. *Journal of Glaciology* **51**, 555-560.

Fraser CI, Nikula R, Spencer HG, Waters JM (2009) Kelp genes reveal effects of subantarctic sea ice during the Last Glacial Maximum. *Proceedings of the National Academy of Sciences* **106(9)**, 3249-3253.

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 Biology* **11**, 525-531.

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

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.

Froget G, Gautier-Clerc M, Le Maho Y, Handrich Y (1998) Is Penguin banding harmless? *Polar Biology* **20**, 409-413.

Gain L (1913) The Penguins of the Antarctic Regions. In: *Smithsonian Report for 1912*, pp. 475-482. Smithsonian, Washington.

Gauthier–Clerc M, Gendner J-P, Ribic CA*, et al.* (2004) *Long–term effects of flipper bands on penguins*.

Ghys MI, Rey AR, Schiavini A (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, Estoup A, Mortier F, Cosson JF (2005a) A spatial statistical model for landscape genetics. *Genetics* **170**, 1261-1280.

Guillot G, Mortier F, Estoup A (2005b) Geneland: a computer package for landscape genetics. *Molecular Ecology Notes* **5**, 712-715.

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

Guo SW, Thompson EA (1992) Performing the Exact Test of Hardy-Weinberg Proportion for Multiple Alleles. *Biometrics* **48**, 361-372.

Heide-Jorgensen M-P, Harkonen T, Dietz R, Thompson PM (1992) Retrospective of the 1988 European seal epizootic. *Diseases of Aquatic Organisms* **13**, 37-62.

Hersteinsson P, MacDonald DW (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 JR, Todd FS, Rumboll M, Swartz B (1979) Pelagic birds in the South Atlantic Ocean and at South Georgia in the austral autumn. *Le Gerfaut* **69**, 13-27.

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, Stroeve J*, et al.* (2014) Projected continent-wide declines of the emperor penguin under climate change. *Nature Climate Change*.

LaRue MA, Kooyman G, Lynch HJ, Fretwell P (2015) Emigration in emperor penguins: implications for interpretation of long-term studies. *Ecography* **38**, 114-120.

Laws RM (1984) *Antarctic ecology* Academic Press, London.

Lemey P, Rambaut A, Drummond AJ, Suchard MA (2009) Bayesian phylogeography finds its roots. *PLoS Comput Biol* **5**, e1000520.

Lescroel A, Bajzak C, Bost CA (2009) Breeding ecology of the Gentoo Penguin *Pygoscelis papua* at Kerguelen Archipelago. *Polar Biology* **32**, 1495-1505.

Lescroel A, Bost CA (2005) Foraging under contrasting oceanographic conditions: the Gentoo Penguin at Kerguelen Archipelago. *Marine Ecology-Progress Series* **302**, 245-261.

Lynch H, Fagan W, Naveen R (2010) Population trends and reproductive success at a frequently visited penguin colony on the western Antarctic Peninsula. *Polar Biology* **33**, 493-503.

Lynch HJ, Naveen R, Fagan WF (2008) Censuses of penguin, blue-eyed shag (*Phalacrocorax atriceps*) and southern giant petrel (*Macronectes giganteus*) populations on the Antarctic Peninsula. *Marine Ornithology* **36**, 83-97.

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.

Mayewski PA, Meredith MP, Summerhayes CP*, et al.* (2009) State of the Antarctic and Southern Ocean climate system. *Reviews of Geophysics* **47**, RG1003.

Mayr E (1942) *Systematics and the Origin of Species, from the Viewpoint of a Zoologist* Harvard University Press.

Millar CD, Dodd A, Anderson J*, 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. February 2006. Falklands Conservation, Stanley, Falkland Islands.

Oksanen J, Blanchet FG, Kindt R*, et al.* (2015) Vegan: Community Ecology Package.

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. *Mol Ecol* **13**, 55-65.

Parmesan C, Ryrholm N, Stefanescu C, et al. (1999) Poleward shifts in geographical ranges of butterfly species associated with regional warming. *Nature* **399**, 579-583.

Peña M F, Poulin E, Dantas GPM*, et al.* (2014) Have Historical Climate Changes Affected Gentoo Penguin (*Pygoscelis papua*) Populations in Antarctica? *PLoS ONE* **9**, e95375.

Pistorius PA, Huin N, Crofts S (2010) Population change and resilience in Gentoo Penguins *Pygoscelis papua* at the Falkland Islands. *Marine Ornithology* **38**, 49-53.

Posada D (2008) jModelTest: phylogenetic model averaging. *Mol Biol Evol* **25**, 1253-1256.

Pritchard J, Falush D, Stephens M (2002) Inference of population structure in recently admixed populations. *American Journal of Human Genetics* **71**, 177-177.

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

Rambaut A, Suchard MA, Xie D, Drummond AJ (2014) Tracer v1.6.

Ratcliffe N, Trathan P (2011) A review of the diet and at-sea distribution of penguins breeding within the CAMLR Convention Area. *CCAMLR Science* **18**, 75-114.

Rignot E, Casassa G, Gogineni S*, et al.* (2005) Recent ice loss from the Fleming and other glaciers, Wordie Bay, West Antarctic Peninsula. *Geophysical Research Letters* **32**.

Roeder AD, Marshall RK, Mitchelson AJ*, et al.* (2001) Gene flow on the ice: genetic differentiation among Adelie penguin colonies around Antarctica. *Molecular Ecology* **10**, 1645-1656.

Roeder AD, Ritchie PA, Lambert DM (2002) New DNA markers for penguins. *Conservation Genetics* **3**, 341-344.

Rogers AD (2012) Evolution and Biodiversity of Antarctic Organisms. In: *Antarctic Ecosystems*, pp. 417-467. John Wiley & Sons, Ltd.

Rogers AD, Murphy EJ, Johnston NM, Clarke A (2007) Introduction. Antarctic ecology: from genes to ecosystems. Part 2. Evolution, diversity and functional ecology. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **362**, 2187-2189.

Saraux C, Le Bohec C, Durant JM*, et al.* (2011) Reliability of flipper-banded penguins as indicators of climate change. *Nature* **469**, 203-206.

Schlosser JA, Dubach JM, Garner TWJ*, et al.* (2009) Evidence for gene flow differs from observed dispersal patterns in the Humboldt penguin, *Spheniscus humboldti*. *Conservation Genetics* **10**, 839-849.

Slatkin M, Excoffier L (1996) Testing for linkage disequilibrium in genotypic data using the expectation-maximization algorithm. *Heredity* **76**, 377-383.

Smouse PE, Long JC, Sokal RR (1986) Multiple-Regression and Correlation Extensions of the Mantel Test of Matrix Correspondence. *Systematic Zoology* **35**, 627-632.

Stonehouse B (1970) Geographic Variation in Gentoo Penguins *Pygoscelis papua*. *Ibis* **112**, 52-&.

Strugnell JM, Watts PC, Smith PJ, Allcock AL (2012) Persistent genetic signatures of historic climatic events in an Antarctic octopus. *Molecular Ecology* **21**, 2775-2787.

Sugden DE, Clapperton CM (1977) The maximum ice extent on island groups in the Scotia Sea, Antarctica. *Quaternary Research* **7(2)**, 268-282.

Tanton JL, Reid K, Croxall JP, Trathan PN (2004) Winter distribution and behaviour of Gentoo Penguins *Pygoscelis papua* at South Georgia. *Polar Biology* **27**, 299-303.

Thomas CD, Lennon JJ (1999) Birds extend their ranges northwards. *Nature* **399**, 213-213.

Thurston MH (1982) Ornithological observations in the South Atlantic Ocean and Weddell Sea 1959-64. *British Antarctic Survey Bulletin* **55**, 77-103.

Trathan PN, Forcada J, Atkinson R, Downie RH, 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, Forcada J, Murphy EJ (2007) Environmental forcing and Southern Ocean marine predator populations: effects of climate change and variability. *Philosophical Transactions of the Royal Society B-Biological Sciences* **362**, 2351-2365.

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**.

Trivelpiece WZ, Hinke JT, Miller AK*, et al.* (2011) Variability in krill biomass links harvesting and climate warming to penguin population changes in Antarctica. *Proceedings of the National Academy of Sciences*.

Trivelpiece WZ, Trivelpiece SG, Volkman NJ (1987) Ecological Segregation of Adelie, Gentoo, and Chinstrap Penguins at King George Island, Antarctica. *Ecology* **68**, 351-361.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4**, 535-538.

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 aux îles Tristan da Cunha et Gough. *Alauda* **47**, 245-280.

Weir BS, Cockerham CC (1984) Estimating F-Statistics for the Analysis of Population-Structure. *Evolution* **38**, 1358-1370.

White RW, Gillon KW, Black AD, Reid JB (2002) The distribution of seabirds and marine mammals in Falkland Islands waters, Peterborough, UK.

Williams TD (1995) *The penguins: Spheniscidae* Oxford University Press, Oxford ; New York.

Williams TD, Croxall JP (1991) The Gentoo Penguin as a candidate species for the CCAMLR Ecosystem Monitoring Program. In: *Selected Scientific Papers*, pp. 483-488. CCAMLR, Hobart, Australia.

Williams TD, Rodwell S (1992) Annual Variation in Return Rate, Mate and Nest-Site Fidelity in Breeding Gentoo and Macaroni Penguins. *The Condor* **94**, 636-645.

Wilson RP, Alvarrez B, Latorre L*, et al.* (1998) The movements of Gentoo Penguins *Pygoscelis papua* from Ardley Island, Antarctica. *Polar Biology* **19**, 407-413.

Woehler EJ, Cooper J, Croxall JP*, et al.* (2001) A Statistical Assessment of the Status and Trends of Antarctic and Sub-Antarctic Seabirds; Montana Workshop Report. Montana Workshop Report.

Xu W, Ouyang Z, Viña A*, et al.* (2006) Designing a conservation plan for protecting the habitat for giant pandas in the Qionglai mountain range, China. *Diversity and Distributions* **12**, 610-619.

Zwally HJ, Giovinetto MB, Li J*, et al.* (2005) Mass changes of the Greenland and Antarctic ice sheets and shelves and contributions to sea-level rise: 1992–2002. *Journal of Glaciology* **51**, 509-527.

Chapter 3. Evidence of Pathogen-Induced Immunogenetic Selection across the Large Geographic Range of a Wild Seabird

Hila Levy[1]*, Steven R. Fiddaman[1]*, Juliana A. Vianna[2], Daly Noll[2,3], Gemma V. Clucas[4,5], Jasmine K. H. Sidhu[1], Michael J. Polito[6], Charles A. Bost[7], Richard A. Phillips[8], Sarah Crofts[9], Gary D. Miller[10], Pierre Pistorius[11], Francesco Bonnadonna[12], Céline Le Bohec[13,14], Andrés A. Barbosa[15], Phil Trathan[8], Andrea Raya Rey[16,17,18], Laurent A.F. Frantz[19], Tom Hart[1], Adrian L. Smith[1]†

1. Department of Zoology, University of Oxford, South Parks Road, Oxford, OX1 3PS, United Kingdom
2. Pontificia Universidad Católica de Chile, Departamento de Ecosistemas y Medio Ambiente, Vicuña Mackenna 4860, Macul, Santiago, Chile
3. Instituto de Ecología y Biodiversidad, Universidad de Chile, Departamento de Ciencias Ecológicas, Santiago, Chile
4. Cornell Atkinson Center for a Sustainable Future, Cornell University, Ithaca, New York 14850, USA
5. Cornell Lab of Ornithology, Cornell University, Ithaca, New York 14850, USA
6. Department of Oceanography and Coastal Sciences, Louisiana State University, Baton Rouge, Louisiana 70803, USA
7. Centre d'Etudes Biologiques de Chizé (CEBC), UMR 7372 du CNRS-Université de La Rochelle, Villiers-en-Bois, 79630, France
8. British Antarctic Survey, High Cross, Madingley Road, Cambridge, CB3 0ET, United Kingdom
9. Falklands Conservation, PO Box 26, Stanley, Falkland Islands, FIQQ 1ZZ, United Kingdom
10. Microbiology and Immunology, PALM, University of Western Australia, Crawley, WA, 6009, Australia

11. DST/NRF Centre of Excellence at the Percy FitzPatrick Institute for African Ornithology, Department of Zoology, Nelson Mandela University, Port Elizabeth, 6031, South Africa
12. CEFE UMR 5175, CNRS, Université de Montpellier, Université Paul-Valéry Montpellier, EPHE, 1919 route de Mende, 34293 Montpellier cedex 5, France
13. Université de Strasbourg, CNRS, IPHC UMR 7178, 23 rue Becquerel, F-67000 Strasbourg, France
14. Centre Scientifique de Monaco, Département de Biologie Polaire, 8 quai Antoine 1er, MC 98000 Monaco, Principality of Monaco
15. Museo Nacional de Ciencias Naturales, Departamento de Ecología Evolutiva, CSIC, C/José Gutiérrez Abascal, 2, 28006 Madrid, Spain
16. Centro Austral de Investigaciones Científicas – Consejo Nacional de Investigaciones Científicas y Técnicas (CADIC-CONICET), Bernardo Houssay 200, Ushuaia, Tierra del Fuego, Argentina
17. Instituto de Ciencias Polares, Ambiente y Recursos Naturales, Universidad Nacional de Tierra del Fuego, Yrigoyen 879, Ushuaia, Argentina
18. Wildlife Conservation Society, Amenábar 1595, Office 19, C1426AKC CABA, Buenos Aires, Argentina
19. School of Biological and Chemical Sciences, Fogg Building, Mile End Rd, Bethnal Green, London E1 4DQ, United Kingdom

†Author for Correspondence:

Adrian L. Smith: adrian.smith@zoo.ox.ac.uk

* These authors have provided equal contribution and share first authorship.

Keywords: immunogenetics, positive selection, Toll-like receptors, pathogen-mediated selection, Antarctica and Southern Ocean, Gentoo Penguin

Abstract

Over evolutionary time, pathogen challenge shapes the immune phenotype of the host to better respond to an incipient threat. The extent and direction of this selection pressure depend on the local pathogen composition, which is in turn determined by biotic and abiotic features of the environment. However, little is known about adaptation to local pathogen threats in wild animals. The Gentoo Penguin (*Pygoscelis papua*) is a species complex that lends itself to the study of immune adaptation because of its circumpolar distribution over a large latitudinal range, with little or no admixture between different clades. In this study, we examine the diversity in a key family of innate immune genes—the Toll-like receptors (TLRs)—across the range of the Gentoo Penguin. The three TLRs that we investigated present varying levels of diversity, with *TLR4* and *TLR5* greatly exceeding the diversity of *TLR7*. We present evidence of positive selection in *TLR4* and *TLR5*, which points to pathogen-driven adaptation to the local pathogen milieu. Finally, we demonstrate that two positively selected cosegregating sites in *TLR5* are sufficient to alter the responsiveness of the receptor to its bacterial ligand, flagellin. Taken together, these results suggest that Gentoo Penguins have experienced distinct pathogen-driven selection pressures in different environments, which may be important given the role of the Gentoo Penguin as a sentinel species in some of the world's most rapidly changing environments.

Introduction

All organisms are challenged by pathogens in their surrounding environments, but it is clear that the pathogen pressure can vary by location. Similarly to free-living metazoans, a latitudinal species richness gradient has been identified in several parasitic and pathogenic taxa, which may be driven by temperature and other abiotic and biotic factors (Rohde and Heap 1998; Guernier, et al. 2004; Dionne, et al. 2007). Given this gradient in pathogen pressure, it follows that natural selection on the host will favour distinct immune phenotypes in different environments, as suggested by MHC II genetic diversity patterns in Humboldt penguins

associated with higher pathogen diversity in lower latitudes (Sallaberry-Pincheira, et al. 2015; Sallaberry-Pincheira, et al. 2016). In our study, we sought to test the hypothesis that pathogen-driven selection can drive distinct patterns of host immune system genotype and phenotype, using the Gentoo Penguin (*Pygoscelis papua* ssp.) as a model species complex.

The Gentoo Penguin complex (Vianna, et al. 2017; Clucas, et al. 2018) is ideally suited for investigating pathogen-driven selection on the immune system. Firstly, it has a circumpolar range, spanning the largest latitudinal range of any penguin species, between 46-66° S, with breeding colonies in most of the Southern Ocean's sub-Antarctic islands, as well as the islands off Tierra del Fuego in South America, South Georgia, the Scotia Arc, and the Western Antarctic Peninsula (Stonehouse 1970). Secondly, population monitoring of the species has shown it to be growing at the southern end of its range (Lynch, et al. 2012), with highly fluctuating changes over time in colonies in the South Atlantic and Indian Oceans (Lescroel and Bost 2006; Trathan, et al. 2007). Thirdly, the Gentoo Penguin is a highly philopatric seabird known to remain close to its breeding colonies year round (Trivelpiece, et al. 1987; Wilson, et al. 1998; Clausen and Putz 2003; Hinke, et al. 2017), limiting gene flow across breeding regions (Levy, et al. 2016; Vianna, et al. 2017; Clucas, et al. 2018).

Furthermore, across its range, Gentoo Penguins overlap with (and occasionally co-occur in mixed colonies with) King (*Aptenodytes patagonicus*), Magellanic (*Spheniscus magellanicus*), Macaroni (*Eudyptes chrysolophus*), and Southern Rockhopper penguins (*Eudyptes chrysocome*) in sub-Antarctic colonies, as well as congeneric Adélie (*P. adeliae*) and Chinstrap Penguins (*P. antarcticus*) in its Antarctic range. Gentoo Penguin colonies are also frequented by a number of flying birds with vast ranges, including albatrosses and petrels (Order: Procellariiformes), as well as predator-scavengers like skuas (genus *Stercorarius*) and sheathbills (genus *Chionis*) that could introduce and/or spread novel avian pathogens. Levels of human interaction also vary across the range, from permanent settlements with livestock present near colonies in the Falkland/Malvinas

Islands, to seasonal or year-round scientific research stations in Sub-Antarctic and Antarctic colonies, and an increasing presence of Antarctic tourism. Differences in sympatric interactions with other species across the range of the Gentoo is likely to result in different pathogen challenges and therefore different selective pressures.

To investigate genetic diversity across the immune system, many immunogenetic studies on penguins have focused on the major histocompatibility complex (MHC; Tsuda, et al. 2001; Bollmer, et al. 2007; Knafler, et al. 2012; Sallaberry-Pincheira, et al. 2016). Increasingly, however, the Toll-like receptors (TLRs) are recognised as important monogenic determinants of disease resistance phenotypes, and are therefore important operands for natural selection (Grueber, et al. 2014). Toll-like receptors are the best-studied family of pattern-recognition receptors in the vertebrate innate immune system, representing the front line of detection of pathogen challenge (Kawai and Akira 2006). TLRs respond to highly conserved microbe-associated molecular patterns (MAMPs) that are structurally conserved in large groups of pathogens. Upon binding of a MAMP, TLRs undergo dimerization and initiate an intracellular signalling cascade that culminates in the production of anti-pathogen effector molecules (Akira, et al. 2001; Botos, et al. 2011).

Vertebrates have six major families of TLRs which are typically conserved across evolutionary time to retain specificity for a particular MAMP or family of MAMPs. In most avian species, there are ten recognized Toll-like receptors (Roach, et al. 2005; Boyd, et al. 2007; Brownlie and Allan 2011). Of these, TLR4 and TLR5 respond to the bacterial agonists lipopolysaccharide and flagellin, respectively, while TLR7 responds to single-stranded RNA of viruses in the endosomal compartment (Chow, et al. 1999; Gewirtz, et al. 2001; Lund, et al. 2004).

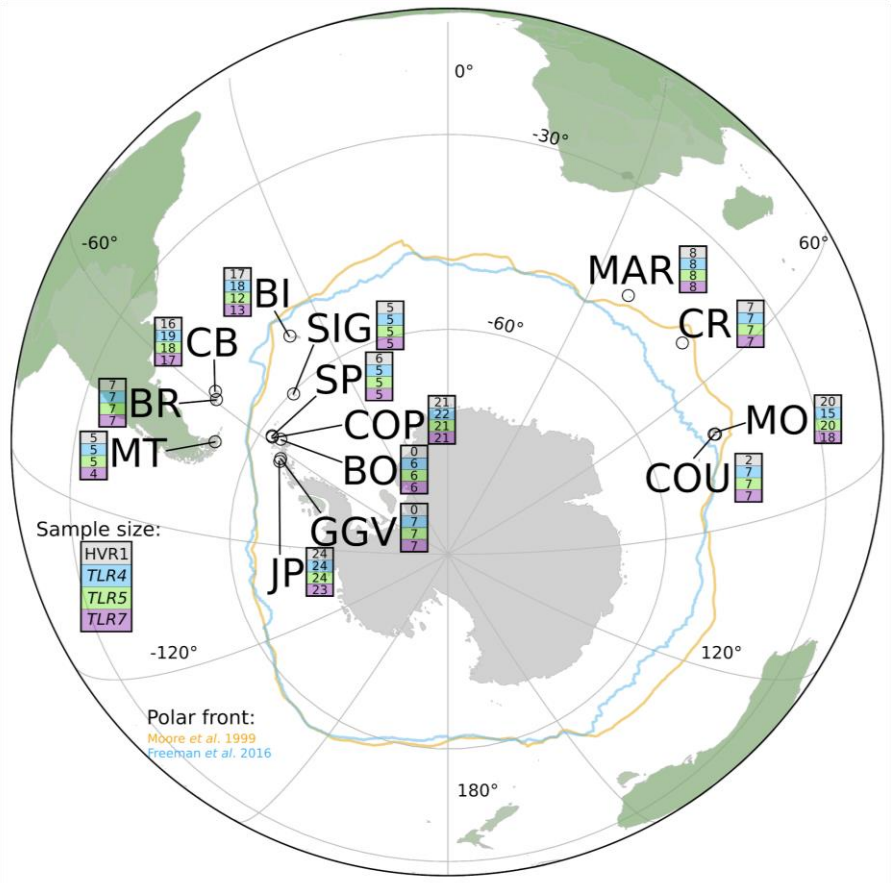


Figure 1. Locations and per-locus sample sizes for each sampled Gentoo Penguin colony in the Southern Ocean and Antarctica. Solid coloured lines represent the reported position of the Polar Front, based on the analyses of Freeman et al. (2016) and Moore et al. (1999). Depending on the analysis, the Kerguelen and Crozet Islands can lie just north or just south of the Polar Front (Moore, et al. 1999; Freeman, et al. 2016). (CR = Crozet Island; MAR = Marion Island; COU = Courbet Peninsula, Kerguelen; MO = Pointe du Morne, Kerguelen; CB = Cow Bay, Falkland/Malvinas Islands; BR = Bull Roads, Falkland/Malvinas Islands; BI = Bird Island, South Georgia; MT = Martillo Island, Tierra del Fuego; SIG = Signy Island, South Orkney Islands; COP = Copacabana (Admiralty Bay), King George Island, South Shetland Islands; SP = Stranger Point, King George Island, South Shetland Islands; BO = Bernardo O'Higgins Base, Western Antarctic Peninsula; GGV = Gabriel González Videla Base, Western Antarctic Peninsula; JP = Jougla Point, Western Antarctic Peninsula).

To investigate TLR diversity across the range of the Gentoo Penguin complex, we sequenced the full coding sequences of *TLR4*, *TLR5* and *TLR7*, as opposed to targeted portions of certain exons as in previous studies (Dalton, et al. 2016a). These three genes represent bacterial- and viral-sensing Toll-like receptors that are present in almost all vertebrates. Samples (n = 155) were obtained from a broad geographic range across the range of the species (**Figure 1**), representing the largest geospatial scale of any immunogenetics study outside of humans. We describe patterns of diversity in TLRs that have a clear spatial component, and provide evidence that some of the diversity in TLR4 and TLR5 is driven by positive selection between different locations. We also demonstrate that two of the positively selected residues in TLR5 yield a phenotypic difference in the response of the receptor to flagellin, providing further evidence that Gentoo Penguins have experienced differential pathogen-driven selection pressures in different environments.

Results

Amplification of TLR genes in the Gentoo Penguin

Through successful amplification by PCR or whole-genome sequencing, we were able to confirm that Gentoo Penguins have clear homologs of *TLR4*, *TLR5*, and *TLR7*, finding no evidence of gene loss or pseudogenization, as has been reported in other avian lineages for *TLR5* (Velová, et al. 2018) or in African penguins for *TLR7* (Dalton, et al. 2016a).

The length of the *P. papua TLR4* coding sequence matches the longest reported length (2550 bp/849 aa) in other bird species. For *TLR5*, there is a start codon that yields an open reading frame in line with the length of previously published *TLR5* sequences (2589 bp/862 aa; Velová, et al 2018), but the ORF continues upstream of the putative start codon, yielding a complete ORF that is 2643 bp/880 aa, which is 54 bp longer than other reported avian *TLR5* sequences. Both the longer and shorter ORF respond to flagellin in our *in vitro* system (data not shown), suggesting both could be functional *in vivo*. The length of the *TLR7* coding exon, at 3126 bp/1042 aa falls within the reported range of avian coding sequence lengths.

Diversity Indices and Population Differentiation

Mitochondrial Hypervariable Region (HVR1)

All colonies with more than two sampled individuals presented high levels of mitochondrial hypervariable region 1 haplotype diversity (H_d = 0.60-1.00; **Figure 2** and **Supplementary Table S1**) as obtained in DnaSP 6.12.10 (Rozas, et al. 2017). Differentiation between colonies at this locus was significant and in line with previous data for this species (see **Supplementary Table S2D**), with four clearly differentiated clades obtained through population-level analyses in Arlequin v3.5.1.3 (Excoffier and Lischer 2010): (1) a southern clade consisting of colonies South of the Polar Front in South Georgia, the South Orkneys, the South Shetlands, and Western Antarctic Peninsula; (2) a South American/Falklands/Malvinas clade, (3) a Kerguelen clade and (4) a North Indian Ocean clade (Marion and Crozet Islands).

Our BEAST 2 phylogenetic analysis of HVR1 showed support for a division within the Gentoo Penguin complex occurring approximately 3.36 Mya (1.72-4.88 Mya), when the North Indian Ocean clade diverges from all others. The Kerguelen lineage diverges from the Atlantic lineages 2.36 Mya (1.14-3.51 Mya), and the populations North and South of the Polar Front within the Atlantic Ocean appear to diverge 1.19 Mya (0.51-1.75 Mya; **Figure 3)**. Within each clade, no clear site-specific mtDNA structure was noted.

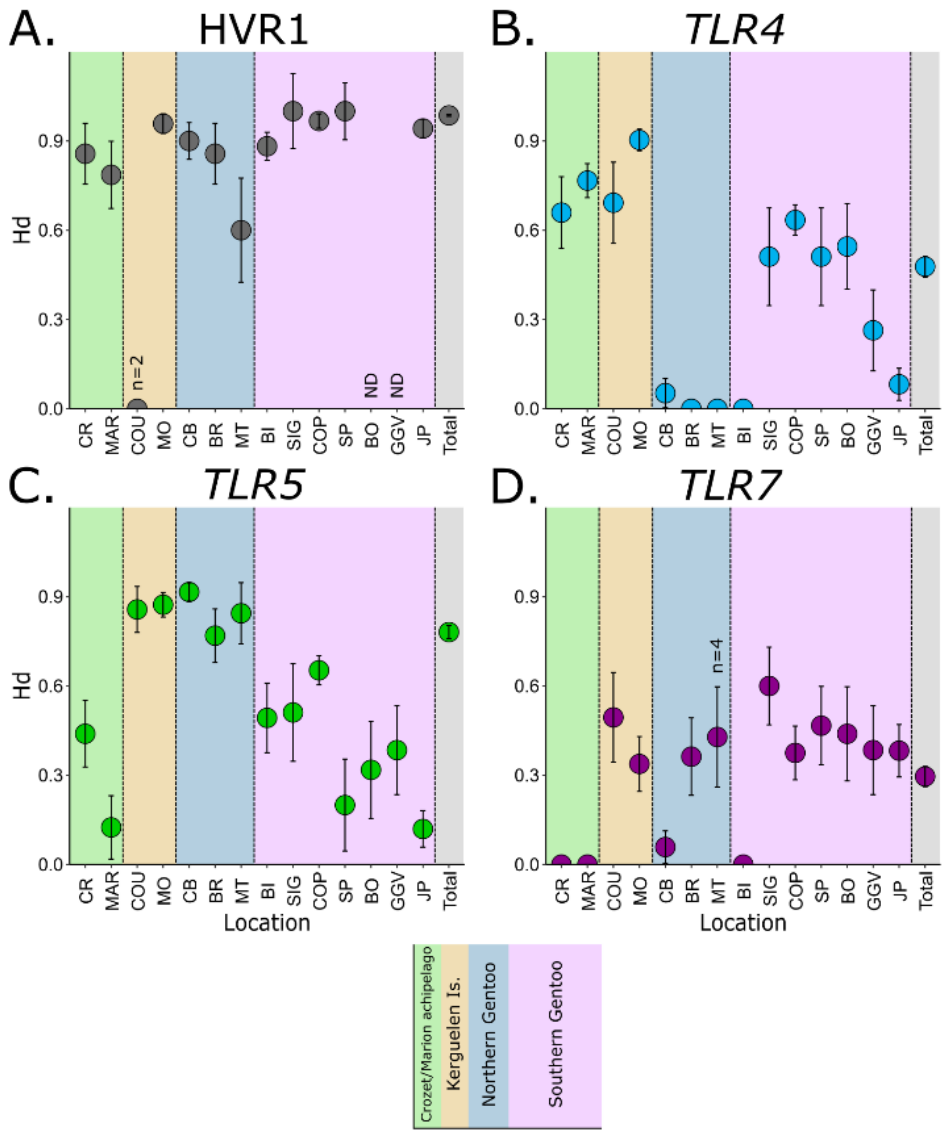


Figure 2. Haplotype diversity (H_d) for the hypervariable mitochondrial control region (HVR1; A) and three TLR loci (B, C, D), for all sampling sites for which data was available. All sites had n ≥ 5 unless otherwise indicated (ND = no data). Location abbreviations are the same as in previous figure.

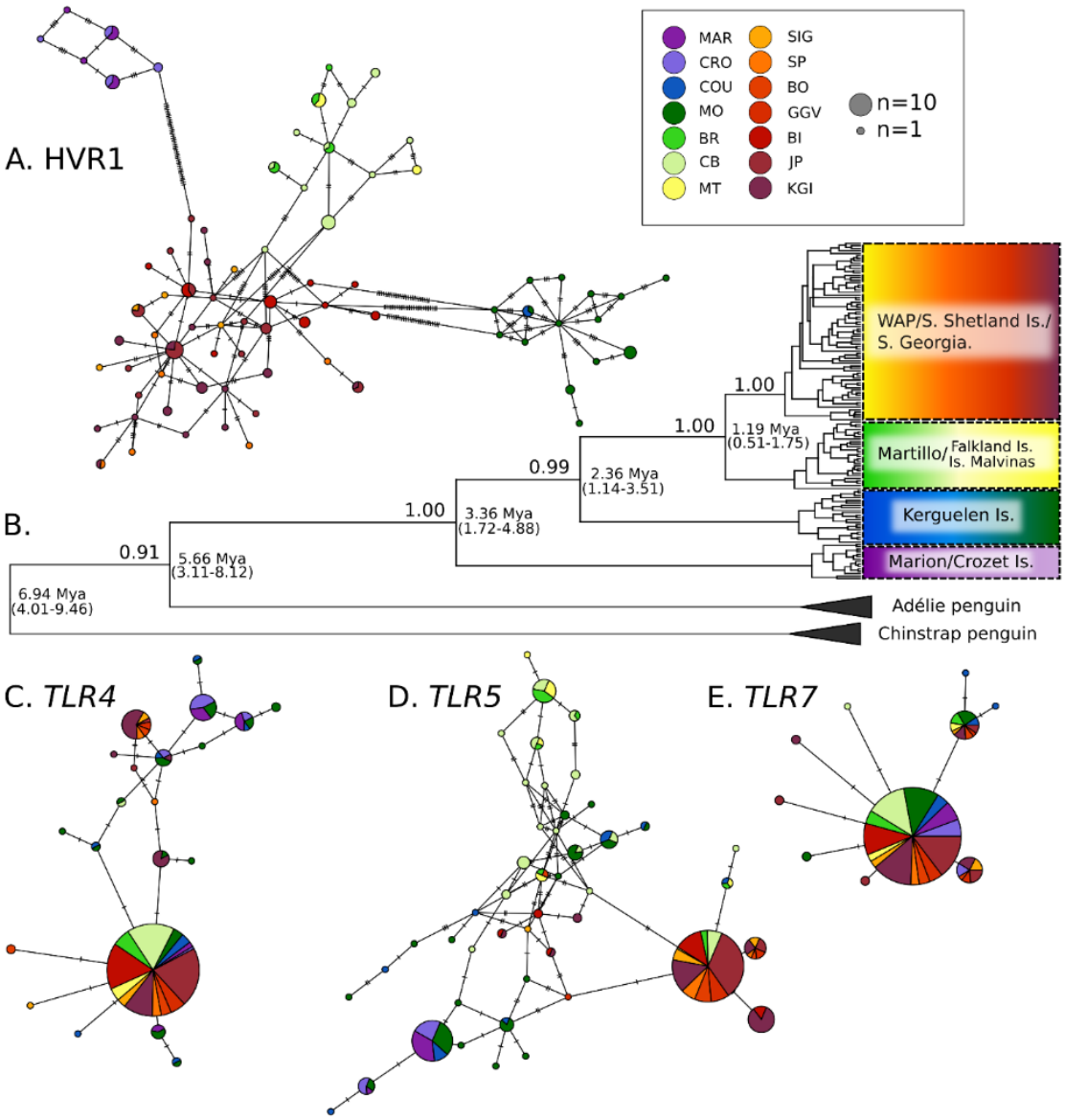


Figure 3. Minimum spanning haplotype networks for mtDNA HVR1 (A), *TLR4* (C), *TLR5* (D), and *TLR7* (E) for Gentoo Penguin colonies, along with mtDNA HVR1 maximum clade credibility tree (B), using congeneric penguin species as outgroups. Location abbreviations are the same as in previous figures. For minimum spanning haplotype networks, pie charts represent single haplotypes, while segment size refers to the contribution of individual sampled sites to the proportion of overall haplotype frequency. Size of pie charts reflects the number of individual birds with the observed haplotype. Dashes on connecting lines each denote one nucleotide change.

TLR4

For *TLR4*, 13 polymorphic sites were identified, with a total of 21 distinct phased haplotypes coding for 9 unique protein variants (**Supplementary Table S1**). Indian Ocean colonies (CR, MAR, COU, and MO) presented the highest levels of haplotype diversity (H_d = 0.66-0.90), while the South American colony in Tierra del Fuego (MT) and the Falkland/Malvinas Islands (CB and BR) had very low diversity (H_d = 0.00-0.05; **Figure 2**). Among Gentoo Penguin colonies south of the Polar Front, diversity was highest in the central colonies of SIG, COP, SP, and BO (H_d = 0.51-0.63), decreasing sharply in BI in South Georgia (H_d = 0.00), as well as southward down the Western Antarctic Peninsula (GGV and JP; H_d = 0.08-0.26).

Comparing pairwise genetic distances (F_{ST} and Φ_{ST}) for in Arlequin *TLR4* (**Figure 5** and **Supplementary Table S2A**), the Indian Ocean populations of Crozet and Marion differed significantly ($F_{ST} > 0.3$, $p < 0.01$) or near-significantly (corrected $p \sim 0.01$) from all other colonies in the Atlantic. Crozet and Marion did not differ significantly from each other (F_{ST} = 0.001, p = 0.40), while some haplotypes were shared between CR/MAR and the Kerguelen Island colonies of COU and MO, yielding some non-significant pairings among the four Indian Ocean colonies. In the Atlantic, one haplotype, also present in the Indian Ocean in lower frequencies, dominated across all the colonies (**Figure 4**). The central sites in the Scotia Arc (COP/SP on King George Island, Signy Island, and BO on the northern end of the Western Antarctic Peninsula) exhibited greater diversity than other Atlantic sites and contained private alleles. The overall pattern of significance for F_{ST} and Φ_{ST} comparisons can be seen in **Figure 5**. Not surprisingly, colonies within the same island group that are in close proximity to each other (COU/MO, CB/BR, COP/SP) did not differ significantly, despite variations in sample sizes.

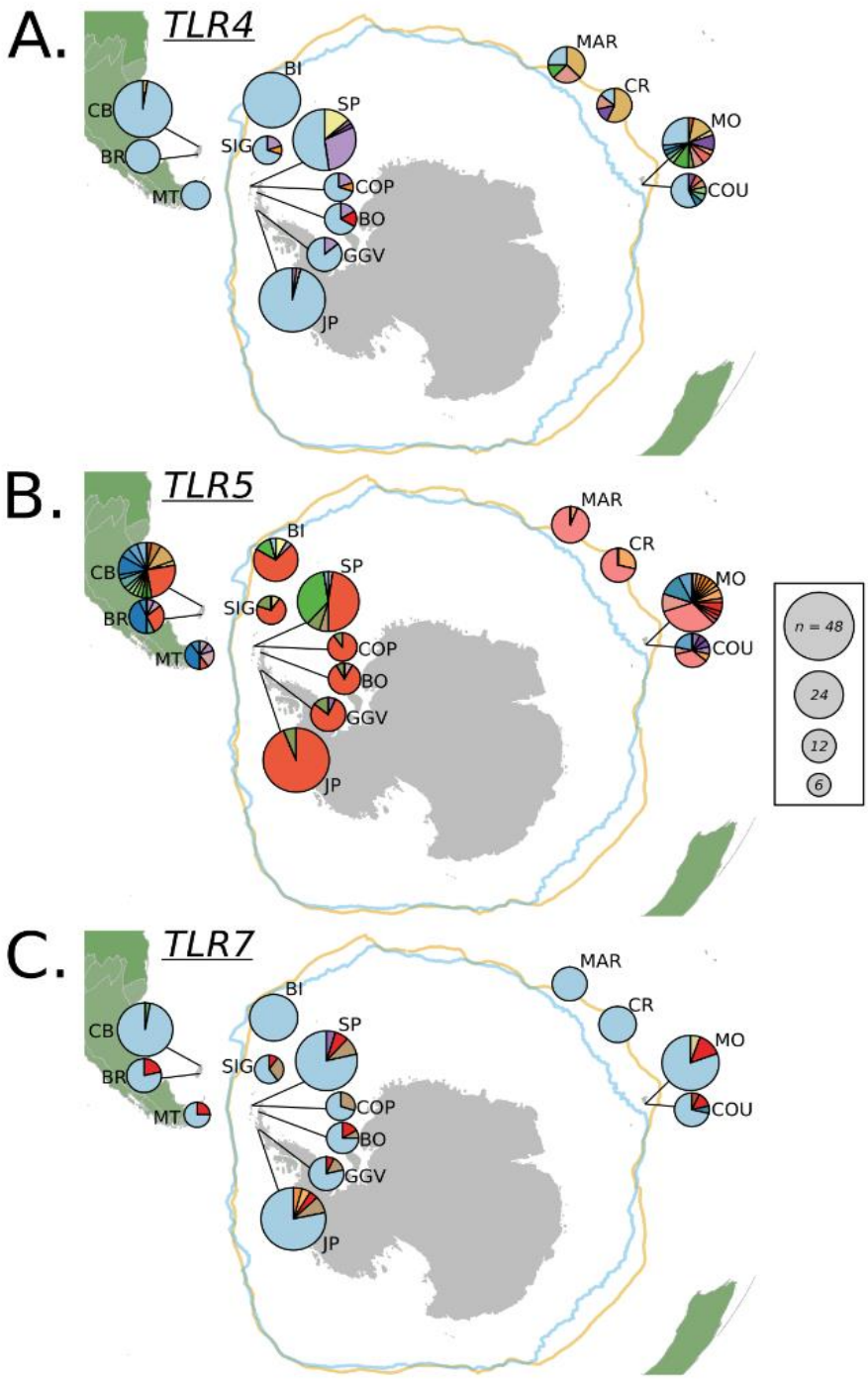


Figure 4. Haplotype diversity across Gentoo Penguin sample populations for (A) *TLR4*, (B) *TLR5*, and (C) *TLR7*. For each locus, different colours represent unique haplotypes and each segment size reflects the proportion of birds in each location with that haplotype. Overall size of the pie chart reflects the number of birds sampled in each location. Location abbreviations are the same as in previous figures.

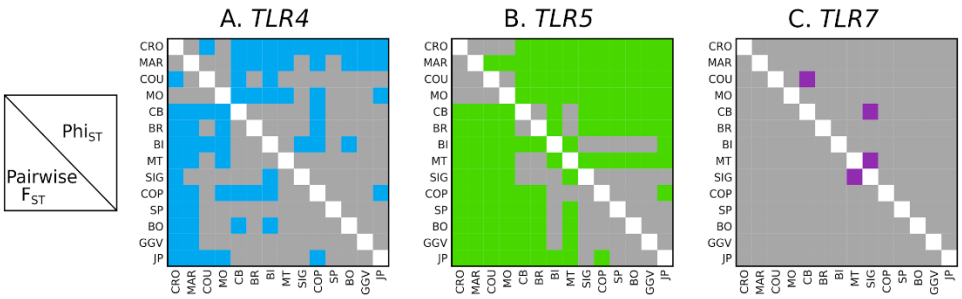


Figure 5. Visualization of pairings of Gentoo Penguin breeding colonies, where pairwise F_{ST} values (below diagonal) or Φ_{ST} values (above diagonal) with significance $p < 0.01$ after correction for multiple tests using SGOF+ (Carvajal-Rodriguez and de Uña-Alvarez 2011) are shown in colour for (A) *TLR4*, (B) *TLR5*, and (C) *TLR7*. Location abbreviations are the same as in previous figures.

Hierarchical population structure was detected for *TLR4* across Gentoo Penguin colonies using Arlequin (AMOVA, global $F_{ST} = 0.32$, $p < 0.0001$). The proportion of variation resulting from differences among groups was 24.01% ($F_{CT} = 0.24$, $p = 0.003$) when colonies were placed into four groups, coinciding with the four mtDNA clades: (1) Marion/Crozet archipelagos; (2) Kerguelen Is.; (3) Falkland/Malvinas and Tierra del Fuego; and (4) south of the Polar Front in the Scotia Arc and Antarctic Peninsula. However, F_{CT} increased to 28.19% ($p = 0.007$) when Bird Island was grouped separately from other Southern Gentoo Penguin colonies, a pattern also suggested in genomic-level analyses (Clucas, et al. 2018).

TLR5

TLR5 was the most diverse TLR locus analysed. Twenty polymorphic sites were identified, with a total of 46 distinct phased haplotypes coding for 32 unique protein variants (**Supplementary Table S1**). Five Gentoo Penguin colonies north of the Polar Front (COU/MO in Kerguelen, CB/BR in Falklands/Malvinas and MT in Tierra del Fuego) exhibited the highest diversity measures (6-17 haplotypes, $H_d = 0.77$-0.92, **Figure 2**), despite differences in sample size. This is unexpected given that Martillo Island is the smallest known population of this species (N_c =

12 breeding pairs, Ghys, et al. 2008), yet still maintained high diversity (5 unique protein variants in a sample of n = 5) at this locus. Interestingly, Crozet and Marion Island colonies exhibited substantially lower genetic variation at this locus with only two haplotypes (H_d = 0.13-0.44), though these were shared with the Kerguelen colonies (H_d = 0.86-0.87; **Figure 4**). Southern colonies exhibited moderate diversity (H_d = 0.51-0.65), with the exception of SP in the South Shetland Islands and the Western Antarctic Peninsula at the edge of the range, with only two or three haplotypes (H_d = 0.12-0.38). Strikingly, Atlantic colonies south of the Polar Front were dominated by one haplotype found less frequently (10-28%) in the northern Atlantic colonies and completely absent from all Indian Ocean colonies. The second-most prevalent haplotype in southern colonies was completely absent from all northern colonies.

Pairwise F_{ST} and Φ_{ST} values obtained in Arlequin revealed significant clustering by clade in terms of genetic distance at the *TLR5* locus (**Figure 5** and **Supplementary Table S2B**). Within three of the four clades mentioned above, there were no significant differences (CR/MAR, COU/MO, and CB/BR/MT). The only significantly differentiated within-clade pairs lay in Gentoo Penguin colonies south of the Polar Front, where Jougla Point differed significantly from COP (highest diversity among Southern Gentoo Penguins) and BI (at the opposite geographic edge of this clade's range). All other colonies differed significantly from all colonies outside their clade (F_{ST} range 0.1-0.88, $p < 0.01$), with CR/COU being near-significant (F_{ST} = 0.13, p = 0.02). This was reflected in AMOVA results, where a four-clade grouping presented a proportion of variation resulting from differences among groups of 32.70% (F_{CT} = 0.327, $p < 0.0001$), and isolating Jougla Point increased the F_{CT} to 56.56% ($p < 0.0001$).

TLR7

TLR7 was the least diverse TLR locus, with 9 polymorphic sites, 10 phased haplotypes, and 8 unique protein variants present across the study colonies (**Figure 2** and **Supplementary Table S1**). One haplotype was predominant in all colonies (frequency of 70-100%) and 7 of the 10

haplotypes were private alleles, only found in single colonies (**Figure 4**). In the Indian Ocean, Kerguelen colonies (COU/MO) had relatively more diversity (H_d = 0.34-0.50) than the Crozet and Marion Island colonies (H_d = 0.00), which had only one haplotype. In the Atlantic, South Georgia (BI) had only one haplotype (H_d = 0.00), while all northern Atlantic colonies (CB/BR/MT) and SP in the South Shetland Islands presented two haplotypes (H_d = 0.06-0.47). Other southern colonies contained 3-4 haplotypes (H_d = 0.38-0.60), while the southernmost colony at Jougla Point (JP) on the Western Antarctic Peninsula exhibited the most unique haplotypes (H = 5; H_d = 0.38) and contained two private alleles. Unsurprisingly, only 1 of 91 pairwise comparisons between colonies were significant in terms of F_{ST} and 3/91 for Φ_{ST} ($p < 0.01$), with no pattern to this differentiation (**Figure 5** and **Supplementary Table S2C**). Different AMOVA groupings yielded an among-group variation (F_{CT}) no higher than 2.85% (four-clade grouping), further highlighting the lack of structure in this locus.

Effect of Population Size on Diversity

Census population size was not significantly correlated to TLR haplotype diversity (See **Supplementary Figure S1** and **Supplementary Tables S3 and S4**: *TLR4* p = 0.067; *TLR5* p = 0.75; *TLR7* p = 0.64).

Isolation by Distance

Significant isolation by distance, using shortest distances by sea between colonies in a Mantel's test, was detected in both *TLR4* (r = 0.515, p = 0.001) and *TLR5* (r = 0.593, p = 0.001; **Supplementary Table S5A**). mtDNA HVR1 was less strongly correlated to isolation by distance (r = 0.312), though marginally significant (p = 0.015). On the other hand, the lack of diversity and structure in the *TLR7* locus yielded no significant correlation across the range of Gentoo Penguin colonies sampled.

Analysis of Positive Selection

We investigated the *P. papua TLR4*, *TLR5* and *TLR7* genes for evidence of positive selection, which could be an indicator of adaptation to local pathogen environments across the

natural range of the species. Neutrality tests (Tajima's *D* and Fu's *Fs*) did not yield observable patterns of significant deviation from neutrality across the full length of these genes (**Supplementary Table S1**). Using a codon-specific approach, the site models in the *codeml* package of programs in PAML v4.9 (Yang 1997; Yang 2007) were employed to test for signatures of positive selection in *P. papua* TLR loci. Non-synonymous sites observed and analysed are graphically depicted in their relative positions on the proteins in **Figure 6**.

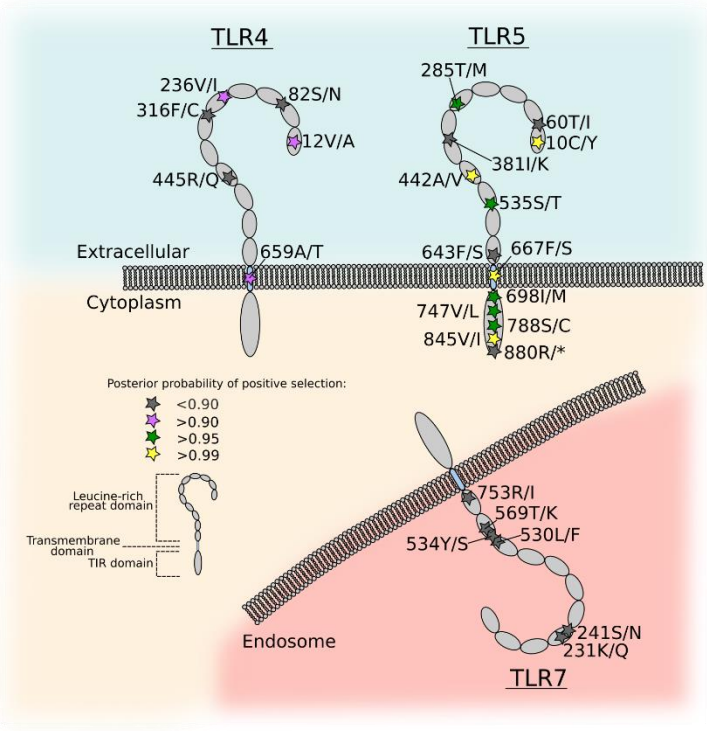


Figure 6. Positions of polymorphic and positively selected sites in *P. papua TLR4*, *5* and *7*. Schematic diagrams of TLRs in the extracellular (TLR4 and 5) and endosomal (TLR7) compartments show positions of amino acid variants resulting from non-synonymous nucleotide substitutions. Variant positions are marked with stars and are colored according to the likelihood of positive selection as determined by the M2a model in the *codeml* program in PAML.

As expected, the majority of codons (*TLR4*, 98.8%; *TLR5*, 98.4%) were predicted to be under purifying selection with the ratio of non-synonymous to synonymous substitutions (*dN*/*dS*) being < 1. Interestingly, 1.2% (*TLR4*) or 1.6% (*TLR5*) of codons in the alignment were found to be

positively selected using M2a, and similar frequencies, 1.2% (*TLR4*) or 1.7% (*TLR5*), were found using M8. For *TLR7*, 99.9% of sites were found to be under purifying selection, while the remaining 0.1% were predicted to be under neutral selection. We investigated whether models that permit positive selection were a significantly better fit to the multiple alignments than models where $dN/dS \not> 1$ by performing likelihood ratio tests between pairs of models. For *TLR4* and *TLR5*, all model comparisons (M1a *vs.* M2a, M7 *vs.* M8, and M8a *vs.* M8) significantly favoured the positive selection model compared to the neutral model (*TLR4*, $p \leq 0.017$; *TLR5*, $p \leq 7.2\times10^{-23}$ for all comparisons), indicating that *P. papua TLR4* and *TLR5* have likely undergone positive selection. In contrast, for *TLR7*, the data were not a significantly better fit to the positive selection model compared to the neutral model ($p = 1$) so the null hypothesis of codons being negatively- and neutrally-selected was not rejected.

For *TLR4* and *TLR5*, we then used the Bayes Empirical Bayes (BEB) algorithm to infer the posterior probability that a particular codon has experienced positive selection. For *TLR4*, three codons were predicted to have undergone positive selection at posterior probability of >0.90 under model M2a (**Supplementary Tables S6A-C**). All but one of the *TLR4* polymorphic residues (12, 82, 236, 316 and 445) were located in the extracellular (LRR) domain, of which two were positively selected (12 and 236). The final positively selected site (659) was located in the transmembrane domain (**Figure 6**). Of the three selected sites, one (12V/A) is a relatively non-conservative change (see **Supplementary Table S7** for amino acid distance metrics). One site in TLR4 (12) has previously been found to be under positive selection in birds (Velová, et al. 2018), while the remaining two (236 and 659) are novel selected sites in penguins (**Supplementary Table S8**).

For TLR5, there were 13 amino acid variants, of which nine were predicted to have undergone positive selection at posterior probability of >0.90 in both M2a and M8 (**Supplementary Table S6D-F**). Of these, four were located in the extracellular domain (10, 285,

442 and 535), one was in the transmembrane domain (667), and four were located in the TIR (intracellular) domain (698, 747, 788 and 845; **Figure 6**). Two sites (442 and 698) have previously been reported as being positively selected in other birds (Grueber, et al. 2014; Velová, et al. 2018), while the remaining seven sites are novel selected sites in penguins. Interestingly, one site in the TLR5 extracellular domain (285) is adjacent to two residues known to be important for flagellin binding in Interface B (Yoon, et al. 2012; Song, et al. 2017), and so could be important for ligand preference (**Supplementary Table S8**; **Supplementary Figure S2**). Four positively selected sites in TLR5 were non-conservative changes (10C/Y, 442A/V, 667S/F and 788S/C). It is interesting to note that the non-conservative TLR5 667F/S polymorphism is located in the transmembrane domain, a region that is typically constrained by the physiochemical requirement to embed in the cell membrane. Furthermore, the homology-based methods of amino acid substitution consequence prediction SIFT (Ng and Henikoff, 2003) and PolyPhen-2 (Adzhubei, et al. 2013) both predict the S667F change to be of high functional consequence (SIFT score = 0.01; PolyPhen-2 score = 0.998; **Supplementary Table S9**). Transmembrane integrity is predicted to remain intact, despite the non-conservative polymorphism, and the Phobius tool (Käll, et al. 2004) predicts the transmembrane domain is unchanged by the polymorphism. While the vast majority (57/64, 89% of available sequences) of avian TLR5 have a serine in this position in the transmembrane domain, only one other bird - the Northern Fulmar (*Fulmarus glacialis*) - has a phenylalanine in this position (**Supplementary Figure S3**), indicating there likely to have been a functional consequence to the S667F change in the Gentoo Penguin.

Functional Analysis of Selected TLR5 Residues

While *in silico* methods can be useful indicators of protein residues under selection, functional study is the only means to isolate the selected phenotype and its relevance. In order to assess whether key selected sites identified in the positive selection analyses have functional consequences, we developed an *in vitro* assay using transient expression of TLRs in a reporter cell line.

Since extracellular domain polymorphisms in TLRs are likely to give rise to preferences in ligand type (Faber, et al. 2018), we focused on TMD/TIR domain polymorphisms that are likely to give rise to differential signalling in response to the same agonist. Given that five *TLR5* polymorphisms were located in the TMD/TIR domain, we tested the two polymorphisms with the highest posterior probability of selection from the PAML analysis – residues 667 (TMD) and 845 (TIR). Polymorphisms at these positions segregate well with geographical location: birds from Crozet and Marion were all homozygous for the 667S/845I haplotype, while 87.3% (n = 71) of birds from colonies South of the Polar Front were homozygous for the derived haplotype, 667F/845V (**Figure 7**). In the Kerguelen Islands, the ancestral 667S/845I haplotype predominated, but variants at both positions were present at lower frequencies. South American/ Falklands/ Malvinas colonies were the most diverse at the positions of interest, where 46.7% (n = 14) of birds were heterozygous at one or both of the sites. Overall, however, polymorphisms at these loci tended to co-segregate: 72.4% (n = 110) of birds were homozygote for either the ancestral (667S/845I) or derived (667F/845V) haplotypes.

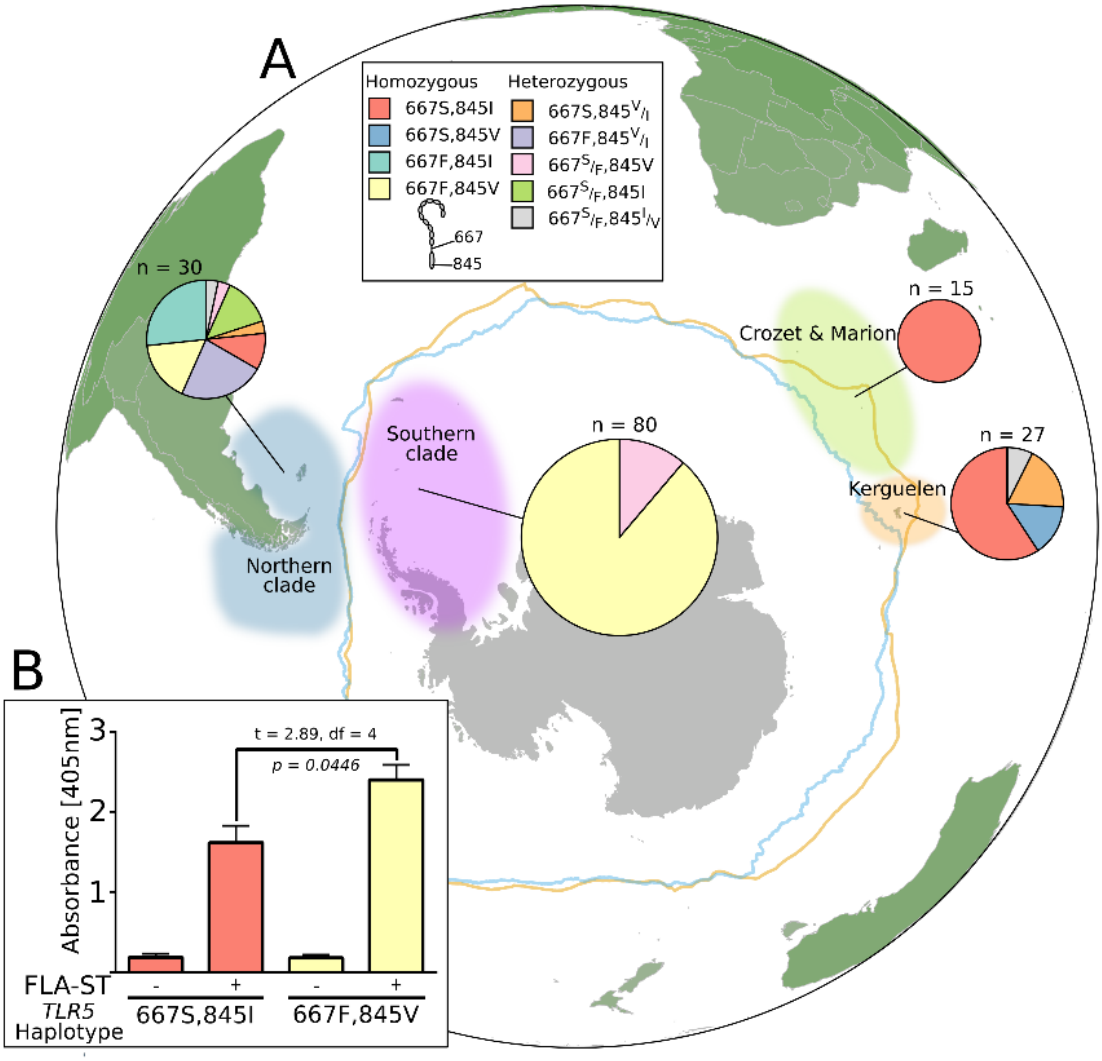


Figure 7. Distribution and functional differences of two selected amino acid residues in TLR5. (A) Mapped areas represent the four clades of *P. papua* as determined by HVR1 analysis. Pie charts represent the allele proportions at the two selected residues in TLR5, and chart area is proportional to the number of samples. (B) Bar chart displays NF-kB response of the two genotypes following stimulation with FLA-ST.

Given the strong tendency for the alleles at these positions to co-segregate at the extremes of the range of *P. papua*, and the fact that these were the TIR/TMD polymorphisms with the highest likelihood of positive selection, the functional consequences of altering both residues together were investigated. FLAG-tagged constructs of both of the TLR5 TIR/TMD variants with the same LRR domain were transiently expressed in *TLR5*$^{-/-}$ HEK-Blue™ Null1 NF-κB reporter cells and protein expression levels were normalised using an anti-FLAG ELISA. Cells expressing either constructs were then treated with *Salmonella enterica* serovar Typhimurium-derived flagellin, or PBS control, and the NF- κB response was measured. Cells expressing either construct responded to flagellin, demonstrating that the *P. papua* TIR domain interacts efficiently with human adapter molecules. Interestingly, there was a marked enhancement (~1.5 fold, p = 0.04) of the flagellin response in the variant that was predominantly found in the Southern Gentoo compared to the variant found in the Indian Ocean clades, suggesting that the derived haplotype 667F/845V has enhanced signalling capability compared to the ancestral genotype 667S/845I (**Figure 7**). These data provide further evidence that *P. papua TLR5* has undergone positive selection for different immune capabilities during the expansion of the species below the Polar Front towards the Antarctic Peninsula.

Discussion

All vertebrates are subject to challenge by a plethora of pathogens that can exert strong selective pressures on host populations. The innate immune system, and in particular the Toll-like receptors, is responsible for both recognizing, and responding to, a pathogen threat by inducing inflammation and priming the adaptive immune response. To investigate TLR adaptation in Gentoo Penguins, we sequenced the entire coding regions of *TLR4, TLR5* and *TLR7*, which recognize products from bacterial and viral pathogens. Multiple individuals were sequenced from colonies at the extremes of the species' range (~8000 km between the most distant colonies), providing an extensive geospatial component to our analysis. We found spatially-associated patterns of diversity in the TLRs, although greater diversity was observed in *TLR4* and *TLR5*

compared to *TLR7*. Furthermore, clear evidence of positive selection in both *TLR4* and *TLR5* was identified, which was further reinforced by the demonstration that two of the TLR5 TMD/TIR domain polymorphisms are sufficient to alter the magnitude of responsiveness to flagellin. To our knowledge, no other TLR study outside of humans has supported predictions of positive selection with confirmation of functional polymorphism.

Patterns of Diversification and Selection

Most studies of TLR genetic diversity in wild populations have investigated small, bottlenecked, and/or endangered avian populations (Grueber, et al. 2012; Grueber, et al. 2013; Hartmann, et al. 2014; González-Quevedo, et al. 2015; Dalton, et al. 2016a; Dalton, et al. 2016b). In these populations, drift, rather than selection, is suspected to have been the main driver of sampled diversity due to recent bottlenecks or pronounced founder effects. Several studies have documented TLR diversity in domesticated animals such as pigs (Darfour-Oduro, et al. 2016), cows (Novak, et al. 2019), and chickens (Świderská, et al. 2018), but large-scale studies on whole-gene TLR variation in a wild population are lacking. Although study design can dramatically affect the diversity detected in different studies it is noteworthy that with Gentoo Penguins *TLR5* exhibited higher diversity than *TLR4* whereas with domestic chicken breeds (n=110, 25 breeds; Świderská, et al. 2018) and grey partridge (n=10; Vinkler, et al. 2015) *TLR4* was more diverse than *TLR5* (**Supplementary Table S10**). The diversity of *TLR7* (compared with *TLR4* and/or *TLR5*) was relatively low in Gentoo Penguins, domestic chicken breeds and grey partridges (**Supplementary Table S10**).

To provide an internal reference for TLR diversity, we sequenced the mitochondrial hypervariable region (HVR1) as a neutral marker in the same individuals. In line with previous studies, we found evidence of at least four deeply divergent clades in *P. papua* based on HVR1 sequence (Vianna, et al. 2017; Clucas, et al. 2018; Pertierra, et al. 2020). These more recent analyses support a revision, first proposed by de Dinechin et al. (2012), of the previously-accepted two subspecies model that was based on morphological characteristics (Stonehouse

1970). These four clades are likely to have much greater divergence (millions of years) than what would be expected at the intraspecific level. We found evidence of differentiation according to this underlying population structure in *TLR4* and *TLR5* which further supports the argument for taxonomic revision of the species, with particular focus on the classification of colonies in South Georgia and the Indian Ocean. Conversely, *TLR7* was highly conserved across the species range and is clearly not subject to the same selection pressures as *TLR4* and *TLR5*. Overall, our study highlights that the genetic differentiation across Gentoo Penguin clades is not just driven by drift, but by clear population-specific adaptations to the environment.

Diversity has been widely reported to vary between different families of TLRs, particularly comparing extracellular and intracellular TLRs. Some authors have reported that TLRs that respond to viral ligands are more likely to be under purifying selection, at least in mammals (Barreiro, et al. 2009; Wlasiuk and Nachman 2010; Wang, et al. 2016; Kloch, et al. 2018), although the pattern may not be consistent in birds, with *TLR3* displaying the greatest number of non-synonymous variants of the four TLRs tested in different chicken breeds (Świderská, et al. 2018), and *TLR7* diversity in the house finch (*Carpodacus mexicanus*) far exceeding that of *TLR4* and *TLR5* (Alcaide and Edwards 2011). Consistent with the pattern observed in mammals (and also the lesser kestrel, *Falco naumanni*; Alcaide and Edwards 2011), we observed overall nucleotide diversity measurements that were several times higher for the extracellular/bacterial TLRs *4* and *5* (*TLR4*: $6.1 \times 10^{-4} \pm 0.5 \times 10^{-4}$; *TLR5*: $14.8 \times 10^{-4} \pm 0.6 \times 10^{-4}$) compared to the intracellular/viral *TLR7* ($1.0 \times 10^{-4} \pm 0.1 \times 10^{-4}$), indicating strong purifying selection for maintenance of function in *TLR7*. In addition, we found no evidence of any codons under selection in *TLR7*, compared to three and nine sites in *TLR4* and *TLR5*, respectively, similar to the pattern of positively selected residues reported in several avian species (Alcaide and Edwards 2011).

Within TLR sequences, levels of variation are not uniformly distributed across the domains of the receptor. TLRs are type I integral membrane glycoproteins with highly conserved

architecture across large phylogenetic distances (Botos, et al. 2011). Typical TLR structure comprises an N-terminal extracellular (or intraluminal for intracellular TLRs) leucine-rich repeat (LRR) domain for ligand binding, a single transmembrane helix, and a C-terminal cytoplasmic signalling (Toll/interleukin-1 receptor, TIR) domain interacting with intracellular adapter proteins (Bell, et al. 2003). The leucine-rich repeat domain directly binds microbe-derived ligands in all known vertebrate TLRs, with the exception of TLR4 recognition of lipopolysaccharide via an accessory molecule, myeloid differentiation factor, MD-2 (Park, et al. 2009). Since the LRR domain represents the interface between host and pathogen, and pathogens exhibit variable MAMPs to evade detection (Andersen-Nissen, et al. 2005), there is often an excess of diversity in the LRR domain compared to the TIR domain (Świderská, et al. 2018; Velová, et al. 2018). In contrast, TIR domains interact with adapter proteins such as MyD88 (myeloid differentiation primary-response protein 88) which are shared between several TLRs, although a MyD88-independent pathway also facilitates TLR3 and TLR4 signalling (Akira and Takeda 2004). Unsurprisingly, TIR domains were found to be much more highly conserved than their corresponding extracellular domains in a study of 366 vertebrate TLRs from 96 species with the exception of *TLR10* (Mikami, et al. 2012). Within species, the same trend is evident: in a study of *TLR3*, *4*, *5* and *7* diversity across domestic chicken breeds, only three of the 46 non-synonymous polymorphisms (two in *chTLR3* and one in *chTLR7*) were located in the TIR domain (Świderská, et al. 2018).

In line with previous evidence for the asymmetric distribution of polymorphisms in TLR domains, we identified an excess of polymorphisms in the LRR domain compared to the TIR domain in two of the three TLRs studied (*TLR4*: 8 LRR vs. 0 TIR; *TLR7*: 6 LRR vs. 0 TIR). Interestingly, *P. papua TLR5* contained a greater number of TIR domain polymorphisms than would be expected from other species (11 LRR vs. 7 TIR), particularly given the LRR is over three-times the length of the TIR domain. Of the seven *TLR5* TIR domain polymorphisms, five were non-

synonymous substitutions, suggesting that the TIR domain of *TLR5* has been under selection to modulate signalling intensity.

Somewhat surprisingly, we also identified non-synonymous polymorphic sites in the transmembrane domains of both *TLR4* (659 Ala/Thr) and *TLR5* (667 Ser/Phe). The transmembrane domain is an uncommon location for TLR polymorphisms, presumably because the region is highly constrained by chemical and functional requirements. As such, the effects of polymorphisms in this region are often large. For instance, the human *TLR1* 602S variant is associated with disrupted cell surface localization of the receptor but is protective against pathology associated with leprosy. It is also noteworthy that the Gentoo Penguin TLR5 transmembrane polymorphic site (667) identified in this study is highly conserved elsewhere in avian phylogeny. Of the other birds with published TLR5 sequences, displayed in the alignment (**Supplementary Figure S3**), 57 (89%) have a serine (ancestral *P. papua* genotype) at this position in the transmembrane domain, and only one other bird – the Northern Fulmar (*Fulmarus glacialis*) – has a phenylalanine residue (derived *P. papua* genotype). The high conservation of serine at this position in the protein points to a widespread pressure for maintenance of function across avian phylogeny, and provides more evidence of a positively selected residue with functional consequences.

Functional polymorphisms in *TLR5* support positive selection

We identified a number of positively selected codons in both *TLR4* and *TLR5*, making both of these receptors candidates for further functional investigation. Polymorphisms in TLR LRR domains have the potential to yield preferences for subtly different microbial ligands, such as LPS or flagellins from different bacterial species (Nahori, et al. 2005; Faber, et al. 2018). However, very limited data are available regarding which pathogens are present in the environments of each of the Gentoo clades, and therefore elucidation of any differences in ligand preference will require further study. We did, identify one positively selected site in TLR5 (285) that is adjacent to two important resides for flagellin binding in Interface B (Yoon, et al. 2012; Song, et al. 2017).

This site would be a good candidate for functional investigation of changes in flagellin ligand preferences, but this would be difficult in the absence of known flagellin variants in candidate *P. papua* pathogens. Instead, we chose to investigate TIR and transmembrane domain polymorphisms for functional consequences because these can yield signalling intensity differences in response to the same agonist (Faber, et al. 2018). Given that the TIR domain of TLR4 did not show any non-synonymous polymorphisms, we focused on the TLR5 TIR/transmembrane domain, and in particular the two residues with the strongest signature of positive selection (667 and 845). Site 667 was likely to be of significant functional consequences because of its transmembrane location, non-conservative amino acid change (serine to phenylalanine) and both SIFT and PolyPhen-2 predicting the change to be of high importance.

The Gentoo Penguin is reported to have undergone a circumpolar expansion, with ancestral populations in the Indian Ocean seeding northern populations that expanded into the Atlantic, and further expansions south of the Polar Front and to the West Antarctic Peninsula – the southernmost extreme of the range (de Dinechin, et al. 2012; Peña, et al. 2014). It is interesting to note that one of the ancestral Indian Ocean clades (Marion and Crozet archipelagos) is completely dominated by birds of the 667S/845I genotype, while the most derived clade of Gentoo Penguins south of the Polar Front are almost entirely dominated by the 667F/845V genotype. These data may reflect an incipient selective sweep of the 667F/845V genotype in Southern Gentoo Penguins.

An alternative explanation for the reduction in diversity at residues 667 and 845 could be genetic bottlenecking during the expansion of *P. papua* south of the Polar Front. However, evidence from neutral markers in this study and a previous study (Levy, et al. 2016) reveal that neutral variation is maintained in the Southern Gentoo Penguin colonies at levels comparable with the northern Atlantic Gentoo Penguin clade up to the southernmost extreme of the range. In *TLR5*, we also saw no correlation between census population size and haplotype diversity.

These findings support the hypothesis that a selective sweep, rather than a bottlenecking event, is responsible for the near-fixation of the *TLR5* 667F/845V haplotype in the Southern Gentoo Penguin. Perhaps more importantly, the finding that the 667F/845V haplotype has enhanced signalling capability provides a functional basis for selection of this TLR5 haplotype.

Potential drivers of selection

Toll-like receptors, like other genes of the immune system, are subject to competing types of selection. Balancing selection works to maintain diversity at a population level in response to the diversity of pathogens in the environment, as was recently proposed in TLRs of the bank vole (*Myodes glareolus*; Kloch, et al. 2018). In contrast, purifying selection may predominate (to retain key functionality), which has been described in large-scale studies of human TLRs in different ethnic backgrounds (Mukherjee, et al. 2014). Finally, positive selection may promote the fixation of novel variants that confer a fitness advantage in the response to pathogens. In the present study, we found evidence of positive selection in *TLR4* and *TLR5*, which likely indicates that the pathogen composition differs substantially between distinct locations in the Gentoo Penguin's range.

Spatial heterogeneity in the profile of pathogens that afflict Gentoo Penguins would be a key driver for the patterns of selection identified in the TLR variants. Latitudinal species diversity gradients have been described for pathogens (and their hosts) (Rohde and Heap, 1998; Guégan, et al. 2008, Dionne, et al. 2007), which might suggest fewer pathogens in Antarctic species. However, a diverse range of pathogens are found in these environments (discussed below). Moreover, within the Gentoo range there are diverse biotic and abiotic characters that exhibit spatial variation (Trathan, et al. 2007, Barbosa, et al. 2009, Barbosa, et al. 2013, Lamont, et al. 2018; Chown, et al. 2015) and these factors will affect the transmission of pathogens. Indeed, the regionalized selection of TLR alleles in different sectors of the Gentoo range support the premise that different challenges are more prevalent or pathogenic in different populations.

The dense colonial conditions and ubiquitous guano (faeces) that characterise Gentoo Penguin habitats provide ideal conditions for the transmission of a wide range of pathogens transmitted by direct contact or faeces. Furthermore, penguins as a group are known to be highly susceptible to a variety of infectious diseases, including, avian cholera (Jaeger, et al. 2018), avian pox (Kane, et al. 2012), avian malaria (Fix, et al. 1988; Grilo, et al. 2016) and aspergillosis (Flach, et al. 1990). A number of infection associated mass mortality events have been documented in both wild and captive penguin populations (Grimaldi, et al. 2015). However, little is known about the pathogens that exist in sub-Antarctic and Antarctic regions, their prevalence, or their fitness costs on penguin populations. Limited data are available on the prevalence of diseases in penguin populations (Clarke and Kerry 2000; Barbosa and Palacios 2009; Woods, et al. 2009; Grimaldi, et al. 2015), and most studies rely upon short notes, observations, and case reports closely tied to obvious signs of disease and mass mortality in well-studied and highly visited penguin colonies. Studies that survey the environmental and host microbiomes to characterise pathogen presence in polar regions remain limited to sites near major polar research stations, have small sample sizes, and/or do not cover large spatial and temporal ranges (Zdanowski, et al. 2004; Dewar, et al. 2013; Fan, et al. 2013; Ma, et al. 2013; Dewar, et al. 2014).

The presence of Gram-negative bacteria exhibiting both lipopolysaccharides and flagella, including *Campylobacter*, *Escherichia*, *Salmonella* and others, has been demonstrated in Gentoo Penguin colonies (Dimitrov, et al. 2009; Bonnedahl 2011; Barbosa, et al. 2013; González-Acuna, et al. 2013; García-Peña, et al. 2017). However, it is not known whether any of these (or other bacterial pathogens) vary across the Gentoo Penguin's range, or may have played a role in the selection of Gentoo Penguin *TLR4* or *TLR5* variants.

Studies of single-stranded RNA viruses (which would typically be recognised by TLR7) are similarly lacking in Gentoo Penguins. Though single-stranded RNA viruses, including the causative agents of Newcastle disease virus and avian influenza, have occasionally been detected in

Pygoscelis penguins through immunological assays and direct isolation (Morgan and Westbury 1988; Wallensten, et al. 2006; Neira, et al. 2017; Olivares, et al. 2019; Wille, et al. 2019), the fitness consequences of viral infection on penguin populations are unknown. We know of only one case report from Signy Island where evidence of a puffinosis outbreak (normally caused by Coronavirus) was described in Gentoo Penguin chicks (Mac Donald and Conroy 1971). The viral drivers behind the strong purifying selection we observed in *TLR7* are unknown, but it could be that the ssRNA viruses that affect Gentoo Penguins are less diverse across the species range than flagellated Gram-negative bacteria.

Sympatric interactions with a diversity of migratory flying birds and their parasites may be important contributors to pathogen diversity in Gentoo Penguin colonies. Birds such as albatrosses, petrels, shearwaters, sheathbills, shags, gulls, terns, and skuas are often observed in close proximity to Gentoo Penguin colonies and there are 46 species recognised in Antarctica alone (Lepage, et al. 2014). There is some evidence that ectoparasites and blood parasites are transmitted between co-occurring bird species (Barbosa et al. 2011; Levin, et al. 2013). It is plausible that cross-species transmission events are important in cross-colony transmission and structuring the profile of pathogens afflicting particular Gentoo Penguins.

It remains unclear why the two functionally tested TMD/TIR residues in TLR5 would confer increased responsiveness to flagellin in Gentoo Penguins south of the Polar Front. TLR signalling must be tightly controlled and aberrant TLR-induced inflammation can lead to immune pathology, toxic shock syndrome and death. It is unsurprising, therefore, that TLR polymorphisms have been described that confer reduced sensitivity to their agonist and a state of tolerance. For instance, the replacement of a highly conserved proline residue by a histidine at position 712 of TLR4 confers endotoxin resistance in certain strains of mice (Qureshi, et al. 1999). It could be that the exposure to (or diversity of) pathogens is decreased for the Southern Gentoo Penguin clade compared to other clades of Gentoo Penguins, and thus individuals can tolerate enhanced

signalling to a prevailing infection. Alternatively, the enhanced signalling could be a manifestation of the Southern Gentoo Penguin adapting to a particular pathogen that is present in the West Antarctic Peninsula and absent elsewhere. The finding of adaptive changes in the Gentoo Penguin immune system necessitates a much better understanding of the pathogen threats faced by Gentoo Penguins in order for their significance to be realized.

Concluding remarks

This wide-ranging immunogenetic study of Toll-like receptors in wild Gentoo Penguins reveals differential selection and adaptation to local pathogen pressure. While the drivers behind the observed patterns of diversity and selection remain unclear in the context of currently available data, it is clear that the Gentoo Penguin has undergone adaptation to local pathogen assemblages across its range.

Infectious disease threats to penguins are likely to become ever more severe in the coming decades given the rapidly changing polar climate (Mayewski 2012; Lynch, et al. 2012). There is also evidence of reverse zoonosis of enteric bacteria being transmitted from humans to sea bird species in Antarctica (Cerda-Cuellar, et al. 2019), which could further increase transmission, especially in light of increasing tourist and scientific research program presence in Antarctica. Although the Gentoo Penguin is not currently one of the 13 out of 18 penguin species with a conservation state of threatened or near-threatened, certain sub-Antarctic populations have experienced sharp declines (Crawford, et al. 2003; Lescroel and Bost 2006; Crawford, et al. 2009; Crawford, et al. 2014). Consequently, the vulnerability of pathogen-naïve populations of penguins should not be underestimated, nor should the importance of the Gentoo Penguin as a sentinel species in the Southern Ocean (Carpenter-Kling, et al. 2019).

Our findings have important implications for the conservation of not just Gentoo Penguins, but also many other vertebrate species, both in the wild and in captivity. Until now, most efforts to genetically delineate conservation units have relied mostly on neutral markers.

The ever-increasing availability of genomic data allows targeted analysis of pathogen-recognition and other immune genes to assess whether different populations possess specific functional adaptations to their environments and should therefore be conserved separately. The approach used here, together with pathogen discovery and surveillance systems, could better define conservation units in species that occupy varied habitats and ecological niches in order to focus resources on potentially susceptible populations.

Methods

Sample Collection

This study used 155 blood samples from Gentoo Penguins, previously obtained in the framework of other projects. Samples were collected between 1999-2017 at the 14 sites shown in **Figure 1** (details in **Supplementary Table S11**). 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 (Lemaho, 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 or capillary, after cleaning the area with an alcohol swab. Total restraint time was generally two to three minutes. The animal was then released at the edge of the colony and observed to ensure it returned to its normal behaviour. Blood was stored in 95% ethanol or Queen's Lysis Buffer at -20 °C for transport at room temperature and subsequent storage at -20 °C upon arrival. All blood samples were imported under the appropriate animal by-product import licenses.

Sampling was conducted under permits from each site's territorial government or governing agency. These permits for animal handling were issued following independent institutional ethical review of the sampling protocols, in accordance with Scientific Committee for Antarctic Research (SCAR) guidelines.

DNA extraction

DNA for samples from MO, CB, BI, COP, and JP was extracted from blood samples using QIAGEN DNeasy Blood and Tissue kits. The digestion step was modified to include 40 µL proteinase K and extended to 3 hours for blood samples. Details of the modifications made to the protocols for tissue samples are available in (Younger, Emmerson, et al. 2015; Younger, Clucas, et al. 2015). All these 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. These samples are stored at the University of Oxford for future analysis. DNA from all other sampling sites was isolated using a modified salt protocol (Aljanabi and Martinez 1997), with details in Vianna, et al. (2017), stored at the Pontificia Universidad Católica de Chile for future analysis.

TLR Genotyping

Primers for *TLR4*, *5*, and *7* coding regions were designed using the primer design feature based on Primer3 2.3.7 (Untergasser, et al. 2012) in Geneious v.11.0 (Biomatters, http://www.geneious.com). Reference coding sequences for primer design were derived from the congeneric Adélie Penguin (*Pygoscelis adeliae*) reference genome (Genbank accession JMFP01000000) and unpublished Gentoo Penguin genomic data. *TLR4* amplifications for samples from MO, CB, BI, COP, and JP, were conducted in 12 µL volumes (9 µL Qiagen *Taq* PCR Master Mix, 2 µL 10 µM forward and reverse primer mix, and 1 µL of template DNA diluted 1:100). *TLR5* and *TLR7* amplifications for these samples were conducted in 25 µL volumes containing 5 µL 5X Phusion High Fidelity (HF) Buffer (New England Biolabs, UK), 0.5 µL 10 mM dNTPs, 1.25 µL of 10 µM forward primer, 1.25 µL of 10 µM reverse primer, 2 µL of template DNA diluted 1:100, 0.25 µL of Phusion Hot Start Flex DNA Polymerase (New England Biolabs, UK), and 14.75 µL nuclease-free water. One GC-rich region in *TLR7* required the use of Phusion GC buffer, rather than HF buffer for amplification. PCR products were visualized on a 1% agarose gel stained with SYBR Safe. The primers and PCR reaction conditions are fully detailed in the Supplementary

Information (**Supplementary Table S12**). PCR products were sequenced using Macrogen Europe's EZ-Seq (*TLR4*) or Eco-Seq (*TLR5* and *TLR7*) services (http://www.macrogen.com, Netherlands) for purification and Sanger sequencing, using the same PCR primers for sequencing.

For all remaining sampling sites, which underwent whole genome sequencing, a total of 100 ng of genomic DNA was fragmented to an average of 350 base pairs to construct paired-end libraries using the Illumina TruSeq Nano kit with the included indexed adapter and barcode. A total of six PCR cycles were used for enrichment, purified with Sample Purification beads, quantified using a Qubit fluorometer and then sequenced to ~20x coverage with 150 paired-end reads using an Illumina HiSeq X platform at MedGenome (USA).

Sequences for each TLR coding region were assembled, edited, and aligned using Geneious v.11.0. For *TLR4*, which has multiple exons along the coding region, exon sequences were extracted and concatenated for further analysis. Heterozygous sites and single nucleotide polymorphisms (SNPs) were detected by visually examining chromatograms. In cases of doubt, resequencing was accomplished so that only high-quality reads from multiple sequencing runs were called as SNPs. All heterozygous sites also had homozygous individuals within the dataset, and each gene had at least one haplotype homozygous across the full length of the gene. All alleles were verified using a combination of multiple independent Sanger sequencing runs and where available, the whole genome sequencing data. The International Union of Pure and Applied Chemistry (IUPAC) code for degenerate nucleotides was used to label heterozygous positions.

mtDNA Genotyping

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 (Clucas, et al. 2014) for samples from MO, CB, BI, COP, and JP. Amplifications were conducted in 25 µL volumes according to the manufacturer's instructions, using Phusion Hot Start Flex DNA Polymerase (New England Biolabs, UK) and 2 µL of template DNA diluted 1:100. Amplifications

involved a two-step PCR, with an initial cycle of 98 °C for 30 seconds, 40 cycles of 98 °C for 10 seconds and 72 °C for 20 seconds, followed by a 10-minute extension at 72 °C. PCR products were visualized on a 1% agarose gel stained with SYBR Safe. PCR fragments were purified using an ethanol/sodium acetate precipitation, and sequencing was performed using the Applied Biosystems BigDye Terminator v3.1 sequencing kit (Applied Biosystems) with the same PCR primers as sequencing primers.

Published mtDNA HVR1 sequences for the samples from BI (GenBank accessions KJ646314-KJ646330, n = 16) and COP (KJ646361-KJ646382, n = 21) were included in the analysis for the relevant individuals. mtDNA data was not available for the individuals from two Antarctic sites. Though other individuals from those sites have sequence data available, we only included data from individuals sequenced for both TLRs and HVR1.

Individual mtDNA fragments from all remaining sites (CR, MAR, COU, BR, MT, SIG, SP) were amplified using the primers tRNAGlu and AH530 (Roeder, et al. 2002). These reactions included 0.4 uM of each primer, 1.5 mM 1X of PCR reaction buffer, $MgCl_2$, 200 uM of each dNTP, and 1U of Taq Polymerase Platinum (Invitrogen) in a two-phase touchdown program (Korbie and Mattick 2008): (1) 10 minutes at 95 °C, and 11 cycles of 95 °C for 15 seconds; a touchdown with an annealing temperature of 60 °C–50 ºC for 30 seconds, with one cycle per 1 °C interval, and 72 °C for 45 seconds; (2) 35 amplification cycles at 95 °C for 15 seconds, 50 °C for 30 seconds, and 72 °C for 45 seconds; and a final extension period of 30 minutes at 72 °C. The purification of these PCR products and sequencing was carried out by Macrogen using an ABI PRISM 3730XL.

Only overlapping segments of the HVR1 sequences common to all samples were used by aligning and editing with Geneious v11.0. Consensus sequences for the resulting 273 bp region of interest were extracted for analysis.

Population-Level Analyses

Haplotypes were inferred for each of the diploid TLR loci using the PHASE algorithm (Stephens, et al. 2001) implemented in DnaSP 6.12.10 (Rozas, et al. 2017) with 10,000 iterations and 1,000 burn-in iterations. Phased haplotype data was used as input to determine standard genetic diversity measures of each population, including number of polymorphic sites, haplotypes, haplotype diversity, and nucleotide diversity, using DnaSP 6.12.10 and Arlequin v3.5.1.3 (Excoffier and Lischer 2010). Minimum spanning haplotype networks were constructed and visualized using PopART 1.7 (Bandelt, et al. 1999; Leigh and Bryant 2015) for each locus. DnaSP was also used to identify synonymous and non-synonymous polymorphic sites and frequencies. Arlequin was used to calculate Tajima's *D* (Tajima 1989), and Fu's *Fs* (Fu 1997). FSTAT v.2.9.3 (Goudet 1995) was used to calculate allelic richness, taking into account differences in sample size.

Because TLR nucleotide diversity has been observed to have a correlation to population size in some island bird populations (Gilroy, et al. 2017), we evaluated the relationship between Gentoo Penguin census population sizes (N_c) and haplotype diversity (H_d) at each locus. Gentoo Penguins are philopatric, but also show evidence of admixture within island groups and adjacent coastlines (Levy, et al. 2016 and Vianna, et al. 2017). For this reason, the census population sizes selected for the analysis were numbers of breeding pairs from the most recent available surveys of each archipelago or region (in the case of the South Shetland Islands and Western Antarctic Peninsula). The N_c size and source survey reference is available in **Supplementary Table S3**. Spearman's rank correlations and *p* values were calculated for each diversity-size comparison.

For population differentiation comparisons, Arlequin was used to calculate pairwise F_{ST} distances based on haplotype frequencies (Weir and Cockerham 1984), and pairwise Φ_{ST}s for the TLR and mtDNA sequences (Excoffier, et al. 1992). FindModel (Posada and Crandall 1998) was used to find the best fit substitution model for use in Arlequin. Φ_{ST} calculations for the TLR loci were obtained using the Tamura and Nei substitution model (Tamura and Nei 1993), while

mtDNA Φ_{ST} analysis was carried out using the Kimura 2-Parameter model (Kimura 1980) with a gamma of 0.27. Analysis of molecular variance (AMOVA) was used to compute hierarchical F-statistics, with 10,000 permutations, to evaluate likely patterns of genetic structure, seeking to identify the population grouping that maximized the among-group variation (F_{CT}) and minimized the variation among colonies within groups (F_{SC}) (Excoffier, et al. 1992). Significance of overall and pairwise genetic distances were computed using 1,000,000 permutations. We used the SGoF+ method (Carvajal-Rodriguez and de Una-Alvarez 2011) within the Myriads software (Carvajal-Rodriguez 2018) to correct for multiple tests.

To test for isolation-by-distance, shortest distances by sea (during summer ice extent), in km, were computed between each sampling location, using Google Earth v7.3.2.5776 (**Supplementary Table S5B**), which were then related to pairwise F_{ST} in a Mantel test, implemented in Arlequin.

Population Divergence and Phylogeography

Phylogenetic reconstruction and estimates of divergence time were carried out using BEAST 2.5.2 (Bouckaert, et al. 2019). The evolutionary model for mtDNA analysis was selected using jModelTest v. 2.1.10 (Darriba, et al. 2012), testing 88 candidate models and selecting the best fit using the Bayes Information Criterion (BIC). All 88 models were within the 100% confidence interval, with HKY+G selected for further divergence analyses (Hasegawa, et al. 1985). A total of 15 Adélie Penguin (*P. adeliae*) and 15 Chinstrap Penguin (*P. antarcticus*) mitochondrial HVR1 GenBank sequences (Clucas, et al. 2014), aligned and cropped to the equivalent size of the Gentoo sequences to avoid bias, were included in the analysis as outgroups.

The most recent common ancestor prior was set for the *Pygoscelis* genus at 7.6 Mya (Subramanian, et al. 2013), derived from the fossil calibration for *Pygoscelis grandis* (Walsh and Suarez 2006), with a normal distribution, and standard deviation (σ) of 1.3 Mya. A strict molecular clock with a starting prior of 1.0 and a Yule speciation process for branching rates, with

uniform priors for birth and clock rates of 1.0 was applied. Four independent runs of 30 million MCMC chains were performed, logging parameters every 3,000 steps. The four independent runs were then combined using LogCombiner v.2.5.2 and assessed for convergence within Tracer v.1.7.1 (Rambaut, et al. 2018). All parameters converged with ESS values greater than 6,000. A maximum clade credibility tree was then generated using TreeAnnotator v.2.5.2 (part of the BEAST software distribution) and visualized in FigTree v.1.4.3.

Selection Analyses

Phylogenetic inference was carried out on phased sequence data for each TLR locus using RAxML-NG using a GTR substitution matrix (Kozlov, et al. 2019). To detect selection, maximum likelihood analysis of ratios of non-synonymous to synonymous nucleotide substitutions (dN/dS; ω) was performed with the *codeml* package of programs in PAML v. 4.9 (Yang 1997; Yang 2007). Various models were fitted to the multiple alignments: M1a (neutral model; two site classes: $0 < \omega_0 < 1$ and $\omega_1 = 1$); M2a (positive selection; three site classes: $0 < \omega_0 < 1$, $\omega_1 = 1$ and $\omega_2 > 1$); M7 (neutral model; values of ω fit to a beta distribution where $\omega > 1$ disallowed); M8 (positive selection; similar to M7 but with an additional codon class of $\omega > 1$); M8a (neutral model; similar to M8 but with a fixed codon class at $\omega = 1$). Likelihood ratio tests were performed on pairs of models to assess whether models allowing positively selected codons gave a significantly better fit to the data than neutral models (model comparisons were M1a vs. M2a, M7 vs. M8, and M8a vs. M8). In situations where the null hypothesis of neutral codon evolution could be rejected ($p < 0.05$), the posterior probability of codons under selection in M2a and M8 were inferred using the Bayes Empirical Bayes algorithm (Yang, et al. 2005).

In Silico Prediction of Polymorphism Functional Consequences

Physiochemical distances between amino acid variants were assessed using distance matrices provided by several authors (Sneath, 1966; Epstein, 1967; Grantham, 1974; Miyata, 1979; Urbina, et al. 2006; **Supplementary Table S7**). Predicted functional consequences of amino acid substitutions were assessed using the homology-based tools SIFT (Ng and Henikoff, 2003)

and PolyPhen-2 (Adzhubei, et al. 2013), using online servers (https://sift.bii.a-star.edu.sg/ and http://genetics.bwh.harvard.edu/pph2/; both accessed December 2019). Transmembrane domain positions were predicted using Phobius (Käll, et al. 2004; http://phobius.sbc.su.se/; accessed December 2019).

Functional Analysis of TLR5 Genotype Expression in CRISPR-Cas9 edited HEK-Blue Cells

In order to functionally assess positively selected *TLR5* polymorphisms *in vitro*, two full-length *TLR5* sequences were synthesised including the two signalling domain polymorphisms that had the strongest signature of selection (GBlocks, IDT). Synthetic genes were cloned using the Gibson assembly method (Gibson, et al. 2009) into the p3XFLAG-CMV™-14 expression vector (Sigma) which incorporates a 3x-FLAG sequence on the C-terminus of the expressed construct. Insert-containing vector was purified using the ZymoPURE II plasmid Maxiprep with the optional endotoxin-removal step (Zymo). Both constructs were transiently expressed using TransIT®-2020 (Mirus Bio) in custom HEK-Blue™ Null1 cells (InvivoGen) that had undergone genome editing using the CRISPR-Cas9 technique to disrupt endogenous human *TLR5*. Cells expressing Gentoo *TLR5* constructs were challenged with *Salmonella* Typhimurium-derived flagellin (FLA-ST; InvivoGen) at 100 ng/mL and incubated for 24 h. Cell supernatants were harvested and NF-κB activity was assessed by measuring the absorbance at 405 nm on a FLUOstar® Omega microplate reader (BMG Labtech) following the addition of p-nitrophenyl phosphate substrate, according to the manufacturer's instructions (SIGMAFAST™, Sigma). Expression levels were monitored by subjecting cell lysates to a direct anti-FLAG ELISA. Cell lysates were harvested in ice-cold RIPA buffer (ThermoFisher) and proteins were immobilised on high-bind ELISA plates (VWR) using coating buffer (BioLegend) overnight at 4 °C. Wells were blocked using StartingBlock™ (PBS) blocking buffer (ThermoFisher) for 1 h and then incubated with mouse monoclonal anti-FLAG M2 antibody (Sigma, 1:1000) at 37 °C for 1 h, followed by incubation with goat anti-mouse IgG-HRP (ThermoFisher, 1:10000) for 1 h at 37 °C. Reactive protein amount was then assessed by the addition of 3,3',5,5'-tetramethylbenzidine substrate and measurement of absorbance at 650 nm.

Expression data were then used to normalise signalling data. Statistical differences between means were determined by a two-tailed Student's t-test, and statistical significance was considered to be $p < 0.05$. Transfections were carried out in three independent wells per condition, and the experiment was conducted on at least three independent occasions.

Acknowledgments

Financial support for this study was provided by an Oxford Clarendon Fund scholarship for HL and from the Biotechnology and Biological Sciences Research Council (BBSRC) [grant number BB/M011224/1] for SRF. Sample collection was funded in part by CONICYT PIA ACT172065 GAB and Spanish Ministry of Science projects CGL2007-60369 and CTM2015-64720. Logistic and field costs for sampling at the Crozet and Kerguelen archipelagos were supported by the Institut Polaire Français Paul-Emile Victor (IPEV: Programme 137 – C.L.B. and Programme 354 – F.B., respectively) with additional support from the Laboratoire International Associé 647 'BioSensib' (CSM/CNRS-University of Strasbourg). We appreciate the hospitality of all of our sampling logistics providers, including W. Trivelpiece and the U.S. Antarctic Marine Living Resource Program, the Argentinean Antarctic Station Carlini and the transportation by the Spanish Polar ship Las Palmas. We also thank A. D. Rogers for providing laboratory resources and guiding the early phases of this study.

Author Contributions

HL and SRF conducted all analyses and wrote the paper, guided by TH and ALS who designed the study. HL, JAV, DN, GVC, MJP, CAB, RAP, SC, GDM, PP, FB, CLB, AAB, PT, ARR, and TH performed fieldwork and contributed samples. HL, GVC, and DN extracted all DNA. HL, SRF, and JKHS performed all PCR amplifications. JAV and DN prepared and provided whole genome sequence data. SRF and JKHS performed all cell culture, cloning, and expression assays. JAV, DN and LAFF provided analytical support with phylogenetic analyses. All authors discussed the study results and implications and contributed to editing the manuscript.

Conflict of Interest

None declared.

Data Accessibility

All sequence data generated in this study were deposited in GenBank under accession numbers MN394222–MN394376 (TLR4), MN313018–MN313169 (TLR5), MN312870–MN313017 (TLR7), and MN566362–MN566421 (mitochondrial control region, HVR1).

References

Adzhubei I, Jordan DM, Sunyaev SR. 2013. Predicting functional effect of human missense mutations using PolyPhen-2. Current Protocols in Human Genetics 76(1)7.20.1-7.20.41.

Akira S, Takeda K. 2004. Toll-like receptor signalling. Nature Reviews Immunology 4:499-511.

Akira S, Takeda K, Kaisho T. 2001. Toll-like receptors: critical proteins linking innate and acquired immunity. Nature Immunology 2:675-680.

Alcaide M, Edwards SV. 2011. Molecular Evolution of the Toll-Like Receptor Multigene Family in Birds. Molecular Biology and Evolution 28:1703-1715.

Aljanabi SM, Martinez I. 1997. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. Nucleic Acids Research 25:4692-4693.

Andersen-Nissen E, Smith KD, Strobe KL, Barrett SLR, Cookson BT, Logan SM, Aderem A. 2005. Evasion of Toll-like receptor 5 by flagellated bacteria. Proceedings of the National Academy of Sciences of the United States of America 102:9247-9252.

Bandelt HJ, Forster P, Rohl A. 1999. Median-joining networks for inferring intraspecific phylogenies. Molecular Biology and Evolution 16:37-48.

Barbosa A, Benzal J, Vidal V, D'Amico V, Coria N, Diaz J, Motas M, Palacios MJ, Cuervo JJ, Ortiz J, et al. 2011. Seabird ticks (Ixodes uriae) distribution along the Antarctic Peninsula. Polar Biology 34:1621-1624.

Barbosa A, De Mas E, Benzal J, Diaz JI, Motas M, Jerez S, Pertierra L, Benayas J, Justel A, Lauzurica P, et al. 2013. Pollution and physiological variability in gentoo penguins at two rookeries with different levels of human visitation. Antarctic Science 25:329-338.

Barbosa A, Palacios MJ. 2009. Health of Antarctic birds: a review of their parasites, pathogens and diseases. Polar Biology 32:1095-1115.

Barreiro LB, Ben-Ali M, Quach H, Laval G, Patin E, Pickrell JK, Bouchier C, Tichit M, Neyrolles O, Gicquel B, et al. 2009. Evolutionary Dynamics of Human Toll-Like Receptors and Their Different Contributions to Host Defense. Plos Genetics 5.

Bell JK, Mullen GED, Leifer CA, Mazzoni A, Davies DR, Segal DM. 2003. Leucine-rich repeats and pathogen recognition in Toll-like receptors. Trends in Immunology 24:528-533.

Bollmer JL, Vargas FH, Parker PG. 2007. Low MHC variation in the endangered Galapagos penguin (Spheniscus mendiculus). Immunogenetics 59:593-602.

Bonnedahl J. 2011. Antibiotic Resistance in Enterobacteriaceae Isolated from Wild Birds. In. Uppsala University.

Botos I, Segal DM, Davies DR. 2011. The Structural Biology of Toll-like Receptors. Structure 19:447-459.

Bouckaert R, Vaughan TG, Barido-Sottani J, Duchene S, Fourment M, Gavryushkina A, Heled J, Jones G, Kuhnert D, De Maio N, et al. 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. Plos Computational Biology 15.

Boyd A, Philbin VJ, Smith AL. 2007. Conserved and distinct aspects of the avian Toll-like receptor (TLR) system: implications for transmission and control of bird-borne zoonoses. Biochemical Society Transactions 35:1504-1507.

Brownlie R, Allan B. 2011. Avian toll-like receptors. Cell and Tissue Research 343:121-130.

Carpenter-Kling T, Handley JM, Connan M, Crawford RJM, Makhado AB, Dyer BM, Froneman W, Lamont T, Wolfaardt AC, Landman M, et al. 2019. Gentoo Penguins as sentinels of climate change at the sub-Antarctic Prince Edward Archipelago, Southern Ocean. Ecological Indicators 101:163-172.

Carvajal-Rodriguez A. 2018. Myriads: P-value-based multiple testing correction. Bioinformatics 34:1043-1045.

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.

Cerda-Cuellar M, More E, Ayats T, Aguilera M, Munoz-Gonzalez S, Antilles N, Ryan PG, Gonzalez-Solis J. 2019. Do humans spread zoonotic enteric bacteria in Antarctic? Science of the Total Environment 654:190-196.

Chow JC, Young DW, Golenbock DT, Christ WJ, Gusovsky F. 1999. Toll-like receptor-4 mediates lipopolysaccharide-induced signal transduction. Journal of Biological Chemistry 274.

Chown S, Clarke A, Fraser C, Cary SC, Moon KL, McGeoch MA. 2015. The changing form of Antarctic biodiversity. Nature 522(7557):431–438.

Clarke J, Kerry K. 2000. Diseases and parasites of penguins. Penguin Conservation 13:5-24.

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

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. Scientific Reports 4.

Clucas GV, Younger JL, Kao D, Emmerson L, Southwell C, Wienecke B, Rogers AD, Bost C-A, Miller GD, Polito MJ, et al. 2018. Comparative population genomics reveals key barriers to dispersal in Southern Ocean penguins. Molecular Ecology 27:4680-4697.

Crawford RJM, Cooper J, Du Toit M, Greyling MD, Hanise B, Holness CL, Keith DG, Nel JL, Petersen SL, Spencer K, et al. 2003. Population and breeding of the gentoo penguin *Pygoscelis papua* at Marion Island, 1994/95-2002/03. African Journal of Marine Science 25:463-474.

Crawford RJM, Dyer BM, Upfold L, Makhado AB. 2014. Congruent, decreasing trends of gentoo penguins and Crozet shags at sub-Antarctic Marion Island suggest food limitation through common environmental forcing. African Journal of Marine Science 36:225-231.

Crawford RJM, Whittington PA, Upfold L, Ryan PG, Petersen SL, Dyer BM, Cooper J. 2009. Recent trends in numbers of four species of penguins at the Prince Edward Islands. African Journal of Marine Science 31:419-426.

Dalton DL, Vermaak E, Roelofse M, Kotze A. 2016a. Diversity in the Toll-Like Receptor Genes of the African Penguin (Spheniscus demersus). Plos One 11.

Dalton DL, Vermaak E, Smit-Robinson HA, Kotze A. 2016b. Lack of diversity at innate immunity Toll-like receptor genes in the Critically Endangered White-winged Flufftail (Sarothrura ayresi). Scientific Reports 6.

Darfour-Oduro KA, Megens H-J, Roca A, Groenen MAM, Schook LB. 2016. Evidence for adaptation of porcine Toll-like receptors. Immunogenetics 68:179-189.

Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. Nature Methods 9:772-772.

de Dinechin M, Dobson FS, Zehtindjiev P, Metcheva R, Couchoux C, Martin A, Quillfeldt P, Jouventin P. 2012. The biogeography of Gentoo Penguins (*Pygoscelis papua*). Canadian Journal of Zoology-Revue Canadienne De Zoologie 90:352-360.

Dewar ML, Arnould JPY, Dann P, Trathan P, Groscolas R, Smith S. 2013. Interspecific variations in the gastrointestinal microbiota in penguins. Microbiologyopen 2:195-204.

Dewar ML, Arnould JPY, Krause L, Dann P, Smith SC. 2014. Interspecific variations in the faecal microbiota of Procellariiform seabirds. Fems Microbiology Ecology 89:47-55.

Dimitrov K, Metcheva R, Kenarova A. 2009. Salmonella presence - an indicator of direct and indirect human impact on Gentoo in Antarctica. Biotechnology & Biotechnological Equipment 23:246-249.

Dionne M, Miller KM, Dodson JJ, Caron F, Bernatchez L. 2007. Clinal variation in MHC diversity with temperature: evidence for the role of host-pathogen interaction on local adaptation in Atlantic salmon. Evolution 61:2154-2164.

Epstein CJ. 1967. Non-randomness of ammo-acid changes in the evolution of homologous proteins. Nature 215(5099):355-359.

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.

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.

Faber E, Tedin K, Speidel Y, Brinkmann MM, Josenhans C. 2018. Functional expression of TLR5 of different vertebrate species and diversification in intestinal pathogen recognition. Scientific Reports 8.

Fan J, Li L, Han J, Ming H, Li J, Na G, Chen J. 2013. Diversity and structure of bacterial communities in Fildes Peninsula, King George Island. Polar Biology 36:1385-1399.

Fix AS, Waterhouse C, Greiner EC, Stoskopf MK. 1988. Plasmodium relictum as a cause of avian malaria in wild-caught magellanic penguins (Spheniscus magellanicus). Journal of Wildlife Diseases 24:610-619.

Flach EJ, Stevenson MF, Henderson GM. 1990. Aspergillosis in Gentoo penguins (*Pygoscelis papua*) at Edinburgh Zoo, 1964 to 1988. Veterinary Record 126:81-85.

Freeman NM, Lovenduski NS, Gent PR. 2016. Temporal variability in the Antarctic Polar Front (2002-2014). Journal of Geophysical Research-Oceans 121:7263-7276.

Fu YX. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147:915-925.

García-Peña FJ, Llorente MT, Serrano T, Ruano MJ, Belliure J, Benzal J, Herrera-Leon S, Vidal V, D'Amico V, Perez-Boto D, et al. 2017. Isolation of Campylobacter spp. from Three Species of Antarctic Penguins in Different Geographic Locations. Ecohealth 14:78-87.

Gewirtz AT, Navas TA, Lyons S, Godowski PJ, Madara JL. 2001. Cutting edge: Bacterial flagellin activates basolaterally expressed TLR5 to induce epithelial proinflammatory gene expression. Journal of Immunology 167:1882-1885.

Ghys MI, Raya Rey A, Schiavini A. 2008. Population trend and breeding biology of Gentoo penguin in Martillo Island, Tierra Del Fuego, Argentina. Waterbirds 31(4):625-631.

Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, III, Smith HO. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nature Methods 6:343-U341.

Gilroy DL, van Oosterhout C, Komdeur J, Richardson, DS. 2017. Toll-like receptor variation in the bottlenecked population of the endangered Seychelles warbler. Anim Conserv 20:235-250.

González-Acuna D, Hernandez J, Moreno L, Herrmann B, Palma R, Latorre A, Medina-Vogel G, Kinsella MJ, Martin N, Araya K, et al. 2013. Health evaluation of wild gentoo penguins (*Pygoscelis papua*) in the Antarctic Peninsula. Polar Biology 36:1749-1760.

Grantham R. 1974. Amino acid difference formula to help explain protein evolution. Science 185(4154:862-864)

González-Quevedo C, Spurgin LG, Carlos Illera J, Richardson DS. 2015. Drift, not selection, shapes toll-like receptor variation among oceanic island populations. Molecular Ecology 24:5852-5863.

Goudet J. 1995. FSTAT (Version 1.2): A computer program to calculate F-statistics. Journal of Heredity 86:485-486.

Grilo ML, Vanstreels RET, Wallace R, Garcia-Parraga D, Braga EM, Chitty J, Catao-Dias JL, Madeira de Carvalho LM. 2016. Malaria in penguins - current perceptions. Avian Pathology 45:393-407.

Grimaldi WW, Seddon PJ, Lyver POB, Nakagawa S, Tompkins DM. 2015. Infectious diseases of Antarctic penguins: current status and future threats. Polar Biology 38:591-606.

Grueber CE, Wallis GP, Jamieson IG. 2014. Episodic Positive Selection in the Evolution of Avian Toll-Like Receptor Innate Immunity Genes. Plos One 9.

Grueber CE, Wallis GP, Jamieson IG. 2013. Genetic drift outweighs natural selection at toll-like receptor (TLR) immunity loci in a re-introduced population of a threatened species. Molecular Ecology 22:4470-4482.

Grueber CE, Wallis GP, King TM, Jamieson IG. 2012. Variation at Innate Immunity Toll-Like Receptor Genes in a Bottlenecked Population of a New Zealand Robin. Plos One 7.

Guégan JF, Prugnolle F, Thomas F. 2007. Global spatial patterns of infectious diseases and human evolution. In: Stearns SC, Koella JC, editors. Evolution in health and disease. Oxford Scholarship Online. DOI: 10.1093/acprof:oso/9780199207466.001.0001 .

Guernier V, Hochberg ME, Guégan JF. 2004. Ecology drives the worldwide distribution of human diseases. PLoS Biol 2:e141.

Hartmann SA, Schaefer HM, Segelbacher G. 2014. Genetic depletion at adaptive but not neutral loci in an endangered bird species. Molecular Ecology 23:5712-5725.

Hasegawa M, Kishino H, Yano TA. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial-DNA. Journal of Molecular Evolution 22:160-174.

Hinke JT, Cossio AM, Goebel ME, Reiss CS, Trivelpiece WZ, Watters GM. 2017. Identifying Risk: Concurrent Overlap of the Antarctic Krill Fishery with Krill-Dependent Predators in the Scotia Sea. Plos One 12.

Jaeger A, Lebarbenchon C, Bourret V, Bastien M, Lagadec E, Thiebot J-B, Boulinier T, Delord K, Barbraud C, Marteau C, et al. 2018. Avian cholera outbreaks threaten seabird species on Amsterdam Island. Plos One 13.

Käll L, Krogh A, Sonnhammer ELL. 2004. A Combined Transmembrane Topology and Signal Peptide Prediction Method. Journal of Molecular Biolog, 338(5):1027-1036.

Kane OJ, Uhart MM, Rago V, Pereda AJ, Smith JR, Van Buren A, Clark JA, Boersma PD. 2012. Avian Pox in Magellanic Penguins (Spheniscus magellanicus). Journal of Wildlife Diseases 48:790-794.

Kawai T, Akira S. 2006. TLR signaling. Cell Death and Differentiation 13:816-825.

Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution 16:111-120.

Kloch A, Wenzel MA, Laetsch DR, Michalski O, Welc-Faleciak R, Piertney SB. 2018. Signatures of balancing selection in Toll-like receptor (TLRs) genes novel insights from a free-living rodent. Scientific Reports 8.

Knafler GJ, Clark JA, Boersma PD, Bouzat JL. 2012. MHC Diversity and Mate Choice in the Magellanic Penguin, Spheniscus magellanicus. Journal of Heredity 103:759-768.

Korbie DJ, Mattick JS. 2008. Touchdown PCR for increased specificity and sensitivity in PCR amplification. Nature Protocols 3:1452-1456.

Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAxML-NG: A fast, scalable, and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics (Oxford, England).

Leigh JW, Bryant D. 2015. POPART: full-feature software for haplotype network construction. Methods in Ecology and Evolution 6:1110-1116.

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. American Journal of Physiology 263:R775-R781.

Lepage D, Vaidya G, Guralnick R. 2014. Avibase - a database system for managing and organizing taxonomic concepts. Zookeys:117-135.

Lescroel A, Bost CA. 2006. Recent decrease in gentoo penguin populations at Iles Kerguelen. Antarctic Science 18:171-174.

Levin II, Zwiers P, Deem SL, Geest EA, Higashiguchi JM, Iezhova TA, Jimenez-Uzcategui G, Kim DH, Morton JP, Perlut NG, et al. 2013. Multiple Lineages of Avian Malaria Parasites

(Plasmodium) in the Galapagos Islands and Evidence for Arrival via Migratory Birds. Conservation Biology 27:1366-1377.

Levy H, Clucas GV, Rogers AD, Leache AD, Ciborowski KL, Polito MJ, Lynch HJ, Dunn MJ, Hart T. 2016. Population structure and phylogeography of the Gentoo Penguin (*Pygoscelis papua*) across the Scotia Arc. Ecology and Evolution 6:1834-1853.

Lund JM, Alexopoulou L, Sato A, Karow M, Adams NC, Gale NW, Iwasaki A, Flavell RA. 2004. Recognition of single-stranded RNA viruses by Toll-like receptor 7. Proceedings of the National Academy of Sciences of the United States of America 101:5598-5603.

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.

Ma D, Zhu R, Ding W, Shen C, Chu H, Lin X. 2013. Ex-situ enzyme activity and bacterial community diversity through soil depth profiles in penguin and seal colonies on Vestfold Hills, East Antarctica. Polar Biology 36:1347-1361.

Mac Donald JW, Conroy JWH. 1971. Virus disease resembling puffinosis in the gentoo penguin (*Pygoscelis papua*) on signy island south orkney islands. British Antarctic Survey Bulletin:80-82.

Mayewski P. 2012. State of the Antarctic climate system. Excerpts from SASOCS (Mayewski et al. 2009). Anales del Instituto de la Patagonia 40:25-30.

Mikami T, Miyashita H, Takatsuka S, Kuroki Y, Matsushima N. 2012. Molecular evolution of vertebrate Toll-like receptors: Evolutionary rate difference between their leucine-rich repeats and their TIR domains. Gene 503:235-243.

Miyata T. 1979. Two types of amino acid substitution in protein evolution. Journal of Molecular Evolution 12(3):219-236

Moore JK, Abbott MR, Richman JG. 1999. Location and dynamics of the Antarctic Polar Front from satellite sea surface temperature data. Journal of Geophysical Research-Oceans 104:3059-3073.

Morgan IR, Westbury HA. 1988. Studies of viruses in penguins in the Vestfold Hills. Hydrobiologia 165:263-269.

Mukherjee S, Ganguli D, Majumder PP. 2014. Global Footprints of Purifying Selection on Toll-Like Receptor Genes Primarily Associated with Response to Bacterial Infections in Humans. Genome Biology and Evolution 6:551-558.

Nahori MA, Fournie-Amazouz E, Que-Gewirth NS, Balloy V, Chignard M, Raetz CRH, Saint Girons I, Werts C. 2005. Differential TLR recognition of leptospiral lipid A and lipopolysaccharide in muine and human cells. Journal of Immunology 175:6022-6031.

Neira V, Tapia R, Verdugo C, Barriga G, Mor S, Ng TFF, García V, Del Río J, Rodrigues P, Briceño C, et al. 2017. Novel Avulaviruses in Penguins, Antarctica. Emerg Infect Dis 23:1212-1214.

Ng PC, Henikoff S. 2003. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Research 31(13):3812-3814.

Novak K, Bjelka M, Samake K, Valcikova T. 2019. Potential of TLR-gene diversity in Czech indigenous cattle for resistance breeding as revealed by hybrid sequencing. Archives Animal Breeding 62:477-490.

Olivares F, Tapia R, Gálvez C, Meza F, Barriga GP, Borras-Chavez R, Mena-Vasquez J, Medina RA, Neira V. 2019. Novel penguin Avian avulaviruses 17, 18 and 19 are widely distributed in the Antarctic Peninsula. Transbound Emerg Dis.

Park BS, Song DH, Kim HM, Choi B-S, Lee H, Lee J-O. 2009. The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex. Nature 458:1191-U1130.

Peña M F, Poulin E, Dantas GPM, González-Acuña D, Petry MV, Vianna JA. 2014. Have Historical Climate Changes Affected Gentoo Penguin (*Pygoscelis papua*) Populations in Antarctica? Plos One 9.

Pertierra L SN, Martínez P, Barbosa A, Raya Rey A, Pistorius P, Polanowski A, Bonadonna F, Le Bohec C, Bi K, et al. 2020. Cryptic speciation in gentoo penguins is driven by geographic isolation and regional marine conditions: Unforeseen vulnerabilities to global change. Diversity and Distributions. 26:958-975.

Posada D, Crandall KA. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14:817-818.

Qureshi ST, Lariviere L, Leveque G, Clermont S, Moore KJ, Gros P, Malo D. 1999. Endotoxin-tolerant mice have mutations in toll-like receptor 4 (Tlr4). Journal of Experimental Medicine 189.

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Systematic Biology 67:901-904.

Roach JC, Glusman G, Rowen L, Kaur A, Purcell MK, Smith KD, Hood LE, Aderem A. 2005. The evolution of vertebrate Toll-like receptors. Proc Natl Acad Sci U S A 102:9577-9582.

Roeder AD, Ritchie PA, Lambert DM. 2002. New DNA markers for penguins. Conservation Genetics 3:341-344.

Rohde K, Heap M. 1998. Latitudinal differences in species and community richness and in community structure of metazoan endo- and ectoparasites of marine teleost fish. Int J Parasitol 28:461-474.

Rozas J, Ferrer-Mata A, Sanchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sanchez-Gracia A. 2017. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. Molecular Biology and Evolution 34:3299-3302.

Sallaberry-Pincheira N, González-Acuña D, Herrera-Tello Y, Dantas GP, Luna-Jorquera G, Frere E, Valdés-Velasquez A, Simeone A, Vianna JA. 2015. Molecular Epidemiology of Avian Malaria in Wild Breeding Colonies of Humboldt and Magellanic Penguins in South America. Ecohealth 12:267-277.

Sallaberry-Pincheira N, González-Acuña D, Padilla P, Dantas GPM, Luna-Jorquera G, Frere E, Valdés-Velásquez A, Vianna JA. 2016. Contrasting patterns of selection between MHC I and II across populations of Humboldt and Magellanic penguins. Ecol Evol 6:7498-7510.

Sneath PHA. 1966. Relations between chemical structure and biological activity in peptides. Journal of Theoretical Biology 12(2):157-195.

Song WS, Jeon YJ, Namgung B, Hong M, Yoon SI. 2017. A conserved TLR5 binding and activation hot spot on flagellin. Scientific Reports 7: 40878

Stephens M, Smith NJ, Donnelly P. 2001. A new statistical method for haplotype reconstruction from population data. American Journal of Human Genetics 68:978-989.

Stonehouse B. 1970. Geographic Variation in Gentoo Penguins *Pygoscelis papua*. Ibis 112:52-57.

Subramanian S, Beans-Picon G, Swaminathan SK, Millar CD, Lambert DM. 2013. Evidence for a recent origin of penguins. Biology Letters 9.

Świderská Z, Smidova A, Buchtova L, Bryjova A, Fabianova A, Munclinger P, Vinkler M. 2018. Avian Toll-like receptor allelic diversity far exceeds human polymorphism: an insight from domestic chicken breeds. Scientific Reports 8.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585-595.

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.

Trathan PN, Forcada J, Murphy EJ. 2007. Environmental forcing and Southern Ocean marine predator populations: effects of climate change and variability. Philosophical Transactions of the Royal Society B-Biological Sciences 362:2351-2365.

Trivelpiece WZ, Trivelpiece SG, Volkman NJ. 1987. Ecological segregation of Adelie, Gentoo, and Chinstrap penguins at King George Island, Antarctica. Ecology 68:351-361.

Tsuda TT, Tsuda M, Naruse T, Kawata H, Ando A, Shiina T, Fukuda M, Kurita M, LeMaho I, Kulski JK, et al. 2001. Phylogenetic analysis of penguin (Spheniscidae) species based on sequence variation in MHC class II genes. Immunogenetics 53:712-716.

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. 2012. Primer3-new capabilities and interfaces. Nucleic Acids Research 40.

Urbina D, Tang B, Higgs PG. 2006. The response of amino acid frequencies to directional mutation pressure in mitochondrial genome sequences is related to the physical properties of the amino acids and to the structure of the genetic code. Journal of Molecular Evolution 62(3):340-361.

Velová H, Gutowska-Ding MW, Burt DW, Vinkler M. 2018. Toll-Like Receptor Evolution in Birds: Gene Duplication, Pseudogenization, and Diversifying Selection. Molecular Biology and Evolution 35:2170-2184.

Vianna JA, Noll D, Dantas GPM, Petry MV, Barbosa A, Gonzalez-Acuna D, Le Bohec C, Bonadonna F, Poulin E. 2017. Marked phylogeographic structure of Gentoo penguin reveals an ongoing diversification process along the Southern Ocean. Molecular Phylogenetics and Evolution 107:486-498.

Vinkler, M., Bainová, H., Bryjová, A. Tomášek, O, Albrecht, T, Bryja, J. 2015. Characterisation of Toll-like receptors 4, 5 and 7 and their genetic variation in the grey partridge. Genetica 143(1):101-112.

Wallensten A, Munster VJ, Osterhaus A, Waldenstrom J, Bonnedahl J, Broman T, Fouchier RAM, Olsen B. 2006. Mounting evidence for the presence of influenza A virus in the avifauna of the Antarctic region. Antarctic Science 18:353-356.

Walsh SA, Suarez ME. 2006. New penguin remains from the Pliocene of northern Chile. Historical Biology 18:119-130.

Wang J, Zhang Z, Liu J, Zhao J, Yin D. 2016. Ectodomain Architecture Affects Sequence and Functional Evolution of Vertebrate Toll-like Receptors. Scientific Reports 6.

Weir BS, Cockerham CC. 1984. Estimating F-Statistics for the Analysis of Population Structure. Evolution 38:1358-1370.

Wille M, Aban M, Wang J, Moore N, Shan S, Marshall J, González-Acuña D, Vijaykrishna D, Butler J, Hall RJ, et al. 2019. Antarctic Penguins as Reservoirs of Diversity for Avian Avulaviruses. J Virol 93.

Wilson RP, Alvarrez B, Latorre L, Adelung D, Culik B, Bannasch R. 1998. The movements of gentoo penguins *Pygoscelis papua* from Ardley Island, Antarctica. Polar Biology 19:407-413.

Wlasiuk G, Nachman MW. 2010. Adaptation and Constraint at Toll-Like Receptors in Primates. Molecular Biology and Evolution 27:2172-2186.

Woods R, Jones HI, Watts J, Miller GD, Shellam GR. 2009. Diseases of Antarctic seabirds. Health of Antarctic wildlife: a challenge for science and policy.:35-55.

Yang ZH. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586-1591.

Yang ZH. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Computer Applications in the Biosciences 13:555-556.

Yang ZH, Wong WSW, Nielsen R. 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. Mol Biol Evol 22:1107-1118.

Yoon SI, Kurnasov O, Natarajan V, Hong M, Gudkov AV, Osterman AL, Wilson IA. 2012. Structural basis of TLR5-flagellin recognition and signaling. Science 335(6070):859-864.

Younger J, Emmerson L, Southwell C, Lelliott P, Miller K. 2015. Proliferation of East Antarctic Adelie penguins in response to historical deglaciation. Bmc Evolutionary Biology 15.

Younger JL, Clucas GV, Kooyman G, Wienecke B, Rogers AD, Trathan PN, Hart T, Miller KJ. 2015. 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.

Zdanowski MK, Weglenski P, Golik P, Sasin JM, Borsuk P, Zmuda MJ, Stankovic A. 2004. Bacterial diversity in Adelie penguin, *Pygoscelis adeliae*, guano: molecular and morpho-physiological approaches. Fems Microbiology Ecology 50:163-173.

Section II. Gentoo Penguin Pathogens and Microbes

Foreword

After considering the host, *Pygoscelis papua*, in Section I it is important to now consider the pathogen profile. The aim of this section (**Chapters 4-9**) is to review the current status of the field and to populate gaps in understanding of the enteric bacterial, viral, and eukaryotic parasite diversity of Gentoo Penguins. These chapters evaluate the entire genus *Pygoscelis* as hosts, given that many pathogens may be shared, a factor that may be important where the range of these three species overlap in the South Atlantic and Antarctic sectors of the Gentoo Penguin's range. In certain portions, sympatric King Penguins (*Aptenodytes patagonicus*, **Chapters 5, 7, and 8**) and Blue-Eyed Shags (*Phalacrocorax atriceps bransfieldiensis*, **Chapter 7-9**) are also included in analyses.

Chapters 5 and **6** are reproduced here in manuscript form based on their submission and acceptance, respectively, for publication:

- **Chapter 5**
    - **Levy, H.**, Fontenele, R.S., Harding, C., Suazo, C., Kraberger, S., Schmidlin, K., Djurhuus, A., Black, C.E., Hart, T., Smith, A.L., and Varsani, A. Identification and distribution of novel circular Rep-encoding ssDNA viruses and circular molecules in four Penguin species in South Georgia and the Antarctic Peninsula. (In Review, *Viruses*).
- **Chapter 6**
    - **Levy, H.**, Fiddaman, S.R., Djurhuus, A., Black, C.E., Kraberger, S., Smith, A.L., Hart, T., and Varsani, A. (2020), Identification of Circovirus Genome in a Chinstrap Penguin (*Pygoscelis antarcticus*) and Adélie Penguin (*Pygoscelis adeliae*) on the Antarctic Peninsula. *Viruses*, 12(8): 858. https://doi.org/10.3390/v12080858

Chapter 4. Review of Pathogens in Pygoscelid Penguins

Hila Levy[1], Tom Hart[1], Adrian L. Smith[1]

1. Department of Zoology, University of Oxford, South Parks Road, Oxford, OX1 3SZ, United Kingdom

Keywords: penguins, wildlife disease, pathogens, viruses, Antarctica, *Pygoscelis*

Abstract

Populations of the Antarctic and Sub-Antarctic penguins in the genus *Pygoscelis* face stressors related to natural biotic and abiotic conditions that may be complicated by anthropogenic factors and climate change. The presence of endemic or introduced pathogens in this environment could play major roles in shaping penguin populations. The Antarctic ecosystem retains a diversity of bacteria, viruses, and macroparasites that have evolved to parasitise seabird hosts, yet remain largely understudied. In this review, we consider the state of our knowledge on pathogens of Pygoscelid penguins across their range. The diversity of sources include case reports, mortality events, culture-based and sequence based pathogen surveys. Having considered the historical and modern literature we highlight areas where knowledge of infectious disease in penguin populations could be improved to better serve future challenges in penguin management and conservation.

Introduction

Parasitism and Disease in Penguins

Penguins (Order: Sphenisciformes) are colonial seabirds found across the Southern Hemisphere. There are 18 recognised species in six genera that form part of the avian family Spheniscidae, although the taxonomy of the Gentoo Penguin and Macaroni/Royal Penguin species

complexes are disputed (Pan *et al.* 2019). Members of the genus *Pygoscelis*, consisting of Adélie Penguins (*P. adeliae*), Chinstrap Penguins (*P. antarcticus*) and Gentoo Penguins (*P. papua*) have varied distributions extending onto the Antarctic continent, where they make up a large part of the avian biomass in the region, and serve a critical ecosystem role as marine mesopredators. As sentinels of changes to the sensitive environments which they inhabit (Ainley 2002), penguin population dynamics are frequently used to measure the impacts of anthropogenic factors such as fishing, pollution, and global climate change (Clucas *et al.* 2014; Trathan *et al.* 2015; Trivelpiece *et al.* 2011). Nonetheless, infectious disease remains an important factor to consider in the conservation of these species, and in monitoring disruptions to the larger Antarctic food web.

Pygoscelid penguins inhabit a large latitudinal range, from the coldest reaches of Antarctica at 77° S (*P. adeliae*), across the entire Antarctic Peninsula (all *Pygoscelis* spp.), to remote sub-Antarctic Islands as far north as 46° S (*P. papua*). This extent across varied marine and coastal environments makes the genus special among flightless birds and seabirds. Like all penguins, these marine predators forage in cold ocean waters, but come ashore to mate, lay eggs, and rear their chick(s).

Pygoscelis spp. nest closely together on ice-free rocky shores of sub-Antarctic and Antarctic coastlines. Though the annual breeding season varies, the entire breeding cycle is typically completed within a 6-8 week period (Black 2016). A limited ice-free area creates dense patches of nests during this season. This density may hold the key to understanding the spread of disease among Pygoscelid penguins. Considering the limited availability of suitable habitat for these species—particularly in Antarctica, where only 0.18% of the continent is ice-free (Burton-Johnson *et al.* 2016)—increased density may lead to increased competition for mates, prey, and nesting space. In addition, dense colonial conditions are likely to foster transmission of respiratory and faecal-oral transmitted diseases (Nash *et al.* 2015).

Importantly, penguins defecate frequently, and forcefully, in the same areas in which they nest and reside (Meyer-Rochow & Gal 2003). As a result of a high-throughput digestive system,

penguin faeces, known as guano, coats most surfaces in and around penguin colonies (see **Figure 1**). Collectively referred to along with shed feathers, egg shells, and dead birds as "ornithogenic material" (Ugolini 1972), this obvious feature of seabird habitats cannot be overlooked when assessing potential sources of infection and persistence of parasites in the environment. Many pathogens employ faecal-oral transmission routes to move from host-to-host (Nash *et al.* 2015). During nest-building, Pygoscelids use their beaks to build nests and move or gift stones in mating rituals (see **Figure 2**), leaving an opportunity for faecal-oral transmission. Pathogens may also be passed from adults to young during chick provisioning, carried out via regurgitation.

Figure 1. Penguins defecate around their nests in ice-free areas, accumulating faecal material (guano) in the vicinity of their land-dwelling environment. The guano is seen here in white, green, and pink colours, changing based on diet composition. Chicks are incubated directly beneath their parents, in contact with rocky nests and often, guano. Photo: Hila Levy, 2014.

Figure 2. Penguins use their beaks to manipulate food and build their nests. Here, a Gentoo Penguin (*Pygoscelis papua*) is seen grasping a pebble while incubating two of its chicks. Faecal-oral transmission of pathogens is increasingly likely due to the amount of ground-based materials like shed feathers and stones handled by penguins while on land. Photo: Hila Levy, 2014.

Pygoscelid penguins share their habitats with flying birds that feed and/or breed near Antarctic and sub-Antarctic penguin colonies. These birds include 46 species of albatrosses, petrels, prions, shearwaters, sheathbills, shags, gulls, terns, and skuas in Antarctica alone (Lepage 2015). Many of these flying birds are migratory, exhibiting extensive ranges that span thousands of kilometres. Certain seabird species, including the Arctic tern (*Sterna paradisaea*), also migrate between the Arctic and Antarctic, with circumpolar and bipolar distributions of their habitats and routes that extend between 59,500 – 81,600 km (Egevang *et al.* 2010). These birds represent potential sources of pathogen introduction, interspecies transmission, and transfer of pathogens between penguin colonies. Indeed, introductions of ectoparasites and blood parasites are known to occur in this way (Levin *et al.* 2013).

In addition, Pygoscelids occasionally share their environments with pinnipeds like the Weddell seal, Leopard seal, Crabeater seal, Ross seal, Southern elephant seal, and Antarctic fur seal that also come ashore to mate and nurse their offspring. Proximity between penguin nesting sites and pinniped breeding and resting areas vary across sites, but interactions with marine mammals or their faeces do occur during penguin transit to and from the sea to feed. The presence of pinnipeds, as with flying birds, can also potential sources of cross-species transmission (Chan *et al.* 2015).

Finally, the Antarctic and sub-Antarctic environments contain arthropod vectors of transmissible diseases. The seabird tick *Ixodes uriae* is present on sub-Antarctic islands as well as the northern reaches of the Antarctic Peninsula (Barbosa *et al.* 2011). This ectoparasite is capable of transmitting the Lyme disease causative spirochetes of the genus *Borrelia*, as well as numerous arboviruses (St. George *et al.* 1985). Transhemispheric exchange of Lyme disease is further facilitated by the movements of migratory birds in conjunction with the presence of local nest-dwelling ticks that persist in the environment (Olsen *et al.* 1995).

A state of chronic sub-clinical infection is common among wild species (Lachish *et al.* 2011). Hence it can be extremely difficult to detect the fitness costs to the host if examining a bird outside the acute phase of disease. Attempting to tease out the fitness costs would require large sample sizes and long-term data monitoring of multiple behavioural, physiological, and morphological traits of both infected and uninfected penguins. Though new technologies like satellite imaging, unmanned aerial vehicles, automated cameras, weighbridges and even penguin-shaped rovers can help monitor populations from a distance (Borowicz *et al.* 2018; Fretwell & Trathan 2009; Green *et al.* 2006; Le Maho *et al.* 2014; Ropert-Coudert & Wilson 2005), these do not provide much insight into the health status or physiological states of penguins (barring large-scale mortality events) without some level of individual tracking.

Pygoscelid Penguin Disease and Mortality Events

Many gaps exist in our knowledge of the presence, prevalence, and pathogenicity of microorganisms in sub-Antarctic and Antarctic regions. There are a small number of comprehensive reviews of Antarctic diseases available in the literature (Barbosa & Palacios 2009; Clarke & Kerry 2000; Diaz *et al.* 2017; Grimaldi *et al.* 2015a; Smeele *et al.* 2018; Woods *et al.* 2009). For the most part, these reviews rely upon short notes, observations, and case reports closely tied to obvious signs of disease and mass mortality in well-studied and highly visited penguin colonies (disease-linked mortality events are summarised in **Table 1**). A greater number of case reports within veterinary literature include observations of disease and mortality in captive penguins, and these are not covered in depth in this review, unless the captive penguins originated from the wild (Brandao *et al.* 2014; Clay 1967; Clay & Moreby 1970; Wright 1954).

Only a small number of pathogens are reported as associated with mortality events, namely the causative agent of avian cholera, the bacterium *Pasteurella moltocida*, as well as the avian pox virus, both of which may have other avian reservoirs, including chickens in the Falkland Islands, or flying birds like kelp gulls, petrels and skuas (Leotta *et al.* 2006b; Leotta *et al.* 2003; Munro 2006). The small number of reports of Pygoscelid deaths likely represents a gross underestimate, as only those cases near human population areas or inhabited research stations are regularly reported.

Table 2. Pygoscelid penguin mortality events with potential pathogen associations as reported in literature.

Location of Mortality Event	**Region**	**Year**	**Host Species**	**No. birds affected**	**Putative Pathogen**	**Report Reference**
West Point Is.	FI	1962	Gentoo	~250 birds, 30 dead	Avian pox virus	(Munro 2006)
Signy Is.	SO	1969	Gentoo	"Several hundred" dead chicks, 94/100 live chicks and 8/30 adults symptomatic	Uncharacter-ised puffinosis virus	(Mac Donald & Conroy 1971)
Hope Bay	NEAP	2000-01	Adélie	2 dead adults	Subcutaneous clostridial infection with *Clostridium sporogenes* and *C. cadaveris*	(Nievas *et al.* 2007)
Hope Bay	NEAP	2001	Adélie	40 adults, 5 chicks dead or dying	*Pasteurella multocida gallicida* A 1	(Leotta *et al.* 2006b)
Cooper Bay	SG	2004	Chinstrap	"Several hundred"	*Pasteurella multocida*	(Government of South Georgia and the South Sandwich Islands 2005)
Multiple Is.	FI	2006	Gentoo	>455 dead across at least 5 colonies	Avian pox virus	(Munro 2006)
Multiple Is.	FI	2008	Gentoo	Unknown total (10-60% dead in some colonies)	Avian pox virus	(Pistorius 2009)
Cooper Bay	SG	2010	Chinstrap	Unknown	*Pasteurella multocida*	(Government of South Georgia and the South Sandwich Islands 2013)

Human Impact on Penguin Disease

One of the major concerns with the human research presence in Antarctica—the only means of reporting disease occurrence—is the potential that humans, including scientists, are themselves a part of pathogen pollution of the ecosystem (Cerdà-Cuéllar *et al.* 2019; Frenot *et al.* 2005). The anthropogenic introduction of new microbes, pathogens or arthropod vectors, into the Antarctic is an ever-growing threat (Grimaldi *et al.* 2015a; Jabour 2009). The scientific bodies that govern Antarctic research have repeatedly raised concerns regarding such introductions, formulating working groups, consultative committees, and recommendations such as those within the 1992 Protocol on Environmental Protection to the Antarctic Treaty, also known as the Antarctic-Environmental Protocol or the Madrid Protocol (Barbosa & Palacios 2009; Grimaldi *et al.* 2011; Scientific Committee for Antarctic Research (SCAR) 2010). Tourism operators who belong to the International Association of Antarctica Tour Operators (IAATO) were among the first to generate biosecurity guidelines for non-scientific visitors (IAATO 2011), though persons visiting the continent outside this framework are not bound by such guidelines. Over 55,000 tourists visited Antarctica with IAATO-affiliated tours in 2018-19 alone (IAATO 2019).

Unfortunately, the damage that may have already been done is only recently coming to light with evidence of accidental introductions of non-native seeds, fungi and insects into the region (Chown *et al.* 2015; Hughes *et al.* 2010; Whinam *et al.* 2005), as well as infectious agents. For example, Newcastle Disease Virus, Infectious Bursal Disease Virus, and human/poultry-associated *Campylobacter* genotypes have been reported in the region, likely introduced by importation of poultry products and produce to research stations (Cerdà-Cuéllar *et al.* 2019; Gauthier-Clerc *et al.* 2002; Hughes *et al.* 2011; Retamal *et al.* 2017). Wider surveys of potential pathogens in more remote (and less visited) colonies may help to shed light on whether these pathogens represent local anthropomorphic introductions or represent part of the natural (or current) pathogen profile. Moreover, detailed molecular analysis can be used to understand the

most likely sources and transmission dynamics of these microbes once they enter the penguin population.

Bacterial Pathogen Research in Pygoscelids

In the early days of microbiological Antarctic research, and even in some recent publications, studies rely on indirect detection methods for microparasites, such as serology and culture, or laboratory techniques such as microscopy (blood and faecal smears, reliant on parasite morphology for identification). More recently, PCR-based detection techniques have come into greater use, which allow the verification of the presence of the microorganism itself in the sample at a given point in time, whereas serology supports evidence of previous exposure. One of the primary differences in approaches is the fact that serology can result in false positives, due to cross-reactivity, while molecular techniques confirm matches based upon sequence detection and provide greater potential resolution to identify strains of the agent in question, depending on the molecular marker chosen.

Most DNA-based bacterial pathogen identification studies in the Antarctic tend to be localised, using small sample sizes due to the logistical, financial, and practical complexities of accessing multiple sites within one austral summer. There are few studies that contrast bacterial community structures across penguin species and sites, and fewer that use individually assignable samples as opposed to non-individualised guano or soil samples from penguin colonies. **Table 2** summarises studies that endeavour to characterise bacterial communities from penguin guts, faeces, or penguin-associated soils.

Bacterial community studies that span multiple sites tend to remain within one regional sector, and have low numbers of samples, while studies with larger sampling efforts over one or more research seasons are generally constrained to locations near polar research stations (Aislabie *et al.* 2009a; Aislabie *et al.* 2009b). Although some try to contrast multiple species with sample sizes of n = 6-13 per species (Dewar *et al.* 2013a; Dewar *et al.* 2014; Lee *et al.* 2019; Yew *et al.* 2017), there have not been analyses of the microbial communities of a single species across its

range. One attempt at a range of sites in Adélie Penguins looked at faecal microbial flora, though only at sites within East Antarctica, not evaluating Adélie colonies in the Western Antarctic Peninsula or further north in the South Shetlands and South Sandwich Islands (Banks et al. 2009). Chinstrap Penguins have only been sampled directly in three studies: one of stomach ingesta (n = 6) on Signy Island, South Orkney Islands (Yew *et al.* 2017), and two of cloacal swabs in the South Shetland Islands (Lee *et al.* 2019), of which one sampled 53 individuals (27 adults and 26 chicks), at a single site (Barbosa *et al.* 2016). Despite their larger range, Gentoo Penguins are poorly represented, with two cloacal swab studies conducted, both from single sites (Dewar *et al.* 2013b; Lee *et al.* 2019). Overall sampling effort, particularly for guano and soil-derived studies, remains concentrated in the South Shetland Islands (King George Is., Livingston Is., and Ardley Is.) and East Antarctica (Ross Is.), where research station density is highest. No studies have been carried out in the Falkland Islands or the Western Antarctic Peninsula.

These community studies can be useful in characterising the overall bacterial diversity of Pygoscelid penguin gastrointestinal tracts and habitats, but they were not focused on identifying pathogenic bacteria. Instead, a number of other studies have focused on isolating and characterising putative bacterial pathogens through differential culture techniques, accompanied by the use of PCR-based techniques since 2005. Studies reporting positive indications of bacteria pertaining to genera of putative pathogenicity in Pygoscelid penguins are summarised in **Table 3**.

Table 2. Details of penguin-associated microbial community studies in the literature that use at least one sequence-based technique, indicating sample type, aims, study sites, sample sizes and species, and sequence techniques for each reference. Culture-only techniques are excluded.

Sample Type	Penguin-Associated Microbial Community Study	Study Aim	Location(s) (Region)	Sample Details (*n*)	Sequence Technique
Stomach ingesta	Yew *et al.* (2017)	Compare inter- and intra-specific variations in stomach bacterial communities of two adjacent penguin species	Signy Island (South Orkneys)	Adélie stomach ingesta (6) Chinstrap stomach ingesta (6)	16S V4 / Illumina MiSeq
Cloacal swabs	Banks *et al.* (2009)	Investigate distribution and acquisition of faecal microbiota in Adélie Penguins	Cape Adare/Cape Hallett/Cape Wheatstone/Cotter Cliffs in Victoria Land (East Antarctica) and Cape Royds and Cape Crozier on Ross Is. (East Antarctica)	Cloacal swabs (12) from 1-3 Adélie Penguins in each of 6 colonies for ARISA (Automated ribosomal intergenic spacer analysis); (1 swab per colony = 6 for 16S cloning)	Cultured 16S clones and ARISA / Megabace 500 capillary sequencer
	Dewar *et al.* (2013b)	Study baseline interspecific variation in gastrointestinal microbiota of four penguin species	Bird Island (South Georgia); Baie du Marine, Possession Is. (Crozet Archipelago); Phillip Is. (Australia)	Cloacal swabs from: King Penguins (12); Gentoo Penguins (12); Macaroni Penguins (12); Little Penguins (12)	16S V2-V3 / Roche 454 Pyrosequencing

	Dewar *et al.* (2014)	Examine impact of fasting/moult on gastrointestinal microbiota of two penguin species	Bird Island (South Georgia) Phillip Is. (Australia)	Cloacal swabs from: King Penguins (12) Little Penguins (9)	16S V2-V3 / Roche 454 Pyrosequencing
	Barbosa *et al.* (2016)	Examine age-related differences in gastrointestinal microbiota	Deception Is. (South Shetlands)	Cloacal swabs from Chinstrap Penguins (53 total: 27 adults and 26 chicks)	16S V1-V3 / Roche 454 Pyrosequencing
	Dewar *et al.* (2017)	Characterise gastrointestinal microbiota throughout development	Phillip Is. (Australia)	Cloacal swabs from Little penguin (9) chicks sampled every two weeks during development	16S V2-V3 / Roche 454 Pyrosequencing
	Lee *et al.* (2019)	Examine change in faecal penguin microbiota during the moult-fast period	King George Is. (South Shetlands)	Cloacal swabs from Chinstrap (12) and Gentoo Penguins (13)	16S V3-V4 / Illumina MiSeq
Guano (faeces)	Zdanowski *et al.* (2004)	Evaluate culturable bacterial diversity in guano over 42-day decomposition	King George Is. (South Shetlands)	"Hundreds of samples"/>0.5 kg of Adélie Penguin guano, analysing 60 subsamples over time	Morpho-physiological tests vs. 16S Sanger (ABI PRISM 377)

					sequencing of single isolates
	Grzesiak *et al.* (2020)	Analysis of changes in bacterial communities in ornithogenic soils across one penguin breeding season	King George Is. (South Shetlands)	Adélie Penguin guano (9; 3 bags of 0.5 kg each, taken monthly for 3 months)	16S V3-V4 / Illumina MiSeq
Environmental (ornithogenic soils or water)	Aislabie *et al.* (2009a)	Characterise bacterial composition of ornithogenic soils compared to mineral soils	Cape Hallett, Victoria Land and Cape Bird, Ross Is. (East Antarctica)	Soil samples (9) from 4 sites in current or former Adélie Penguin colonies	16S clone libraries and restriction fragment length polymorphism (RFLP) analysis / Sanger sequencing
	Aislabie *et al.* (2009b)	Characterise culturable microbial community in groundwater underlying ornithogenic soil	Cape Hallett, Victoria Land (East Antarctica)	Groundwater samples (6) near Adélie Penguin colony	Culture, 16S clone libraries and restriction fragment length polymorphism (RFLP) analysis /

					Sanger sequencing
	Kim *et al.* (2012)	Compare bacterial community compositions in soil with varying levels of penguin guano	King George Is. (South Shetlands)	Soil samples (5) near former and active Gentoo and Chinstrap Penguin colonies	16S / Roche 454 Pyrosequencing
	Fan *et al.* (2013)	Examine microbiota of lacustrine sediment, marine sediment, penguin ornithogenic sediments, and soils using culture-dependent and culture-independent methods	King George Is. (South Shetlands)	13 soil/sediment samples of which: penguin ornithogenic sediments (2), marine sediment (1), lake sediment (2)	PCR-DGGE with 16S clone libraries
	Ma *et al.* (2013)	Investigate bacterial community structure in penguin and seal colony soils	Vestfold Hills / Davis Station (East Antarctica)	4 soil cores (2 penguin colonies, 2 seal colonies)	PCR-DGGE with 16S clone libraries
	Teixeira *et al.* (2013)	Assess influence of plants and birds on soil bacterial communities	King George Is. (South Shetlands)	Soil samples in triplicate from 6 sites (18) of which 3 sites were penguin colonies (Adélie / Chinstrap / Gentoo)	16S Microarray / PhyloChip

	Guo *et al.* (2018)	Assess effect of penguin faeces on soil microbiomes	Ardley Is. (South Shetlands)	Soil samples (4) in and around Adélie Penguin colony	16S V3-V4 / Illumina MiSeq
	Romaniuk *et al.* (2018)	Analyse diversity of heavy metal-resistant bacteria in Antarctic soils with varying anthropogenic and zoogenic influence	King George Is. (South Shetlands)	Soil samples (3) including one from Adélie Penguin colony	Culture, 16S and four heavy metal resistance genes via Sanger sequencing / ABI 377
	Ramirez-Fernandez *et al.* (2019)	Assess influence of animal concentrations on soil bacterial communities	Livingston Is. (South Shetlands)	Soil samples in triplicate from 6 sites (18) with different vertebrate presence (2 penguin sites: Chinstrap and Gentoo Penguin colonies)	16S V1-V3 / Roche 454 Pyrosequencing

Table 3. Studies reporting presence of bacteria of potentially pathogenic genera in Pygoscelid penguin hosts, by host species and location. Regions: CWAP = Central Western Antarctic Peninsula; EA = East Antarctica; FI = Falkland Islands/Malvinas; KERG = Kerguelen Islands; MACQ = Macquarie Island; NEAP = Northeast Antarctic Peninsula; NWAP = Northwestern Antarctic Peninsula; SG = South Georgia; SH = South Shetland Islands; SO = South Orkney Islands; SWAP = Southwestern Antarctic Peninsula. Detection Methods: AFLP = Amplified Fragment Length Polymorphism; ERIC-PCR = Enterobacterial Repetitive Intergenic Consensus Sequence-based Polymerase Chain Reaction; MALDI TOF MS = Matrix-Assisted Laser-Desorption/Ionization Time-Of-Flight Mass Spectrometry; MLST = Multilocus Sequence Typing; NGS = Next Generation Sequencing; PCR = Polymerase Chain Reaction; PFGE = Pulsed-Field Gel Electrophoresis; RAPD-PCR = Random Amplified Polymorphic DNA Polymerase Chain Reaction.

Potential Bacterial Pathogen	**Host Species**	**Location**	**Region**	**Detection Method (Reference)**
Achromobacter sp.	Gentoo	King George Is.	SH	Culture (Jorge *et al.* 2002)
Actinomyces sp.	Gentoo	King George Is.	SH	Culture (Jorge *et al.* 2002)
Bacillus spp. (*B. cereus, circulans, megatherium, sphaericus*)	Gentoo	King George Is.	SH	Culture (Jorge *et al.* 2002)
Campylobacter sp.	Gentoo	Bernardo O'Higgins Base	NWAP	Culture, RNA PCR (González-Acuña *et al.* 2013)
		Waterboat Pt.	CWAP	Culture, RNA PCR (González-Acuña *et al.* 2013)
		Ardley Is.	SH	Culture, RNA PCR (González-Acuña *et al.* 2013)

Campylobacter lari (inclusive of unknown, *lari* or *concheus* subspecies)	Adélie	Stranger Pt.	SH	Culture, PFGE, MLST (García-Peña *et al.* 2017)
		Paulet Is.	NEAP	Culture, 16S rDNA PCR (Bonnedahl *et al.* 2005)
		Hope Bay	NEAP	Culture (Leotta *et al.* 2006a)
		Avian Is.	SWAP	Culture, PFGE, MLST (García-Peña *et al.* 2017)
	Chinstrap	King George Is.	SH	Culture, 16S rDNA PCR, MLST, MALDI-TOF (Johansson *et al.* 2018)
		Deception Is.	SH	Culture, PFGE, MLST (García-Peña *et al.* 2017)
		Kopaitik Is.	NWAP	Culture, 16S rDNA PCR, MLST, MALDI-TOF (Johansson *et al.* 2018)
		Rongé Is.	CWAP	Culture, PFGE, MLST (García-Peña *et al.* 2017)
	Gentoo	Ardley Is.	SH	Culture, 16S rDNA PCR, MLST, MALDI-TOF (Johansson *et al.* 2018)
		King George Is.	SH	Culture, PFGE, MLST (García-Peña *et al.* 2017)
		Livingston Is.	SH	Culture, 16S rDNA PCR (Bonnedahl *et al.* 2005); Culture (Barbosa *et al.* 2013); Culture, PFGE, MLST (García-Peña *et al.* 2017) Culture, PCR, PFGE, MLST (Cerdà-Cuéllar *et al.* 2019)
		Bernardo O'Higgins Base	NWAP	Culture, 16S rDNA PCR, MLST, MALDI-TOF (Johansson *et al.* 2018)
		Kopaitik Is.	NWAP	Culture, 16S rDNA PCR, MLST, MALDI-TOF (Johansson *et al.* 2018)
		Waterboat Pt.	CWAP	Culture, 16S rDNA PCR, MLST, MALDI-TOF (Johansson *et al.* 2018)
		Paradise Bay	CWAP	Culture, PFGE, MLST (García-Peña *et al.* 2017)
		Cierva Cove	CWAP	Culture, PFGE, MLST (García-Peña *et al.* 2017)
		Neko Harbour	CWAP	Culture, 16S rDNA PCR (Bonnedahl *et al.* 2005)
Campylobacter subantarcticus	Gentoo	Bird Island	SG	Culture, 16S rDNA PCR, AFLP (Debruyne *et al.* 2010)

Chlamydia sp.	Adélie	Mawson Station	EA	Antibodies (Moore & Cameron 1969)
	Gentoo	Macquarie Is.	MACQ	Antibodies (Moore & Cameron 1969)
Chlamydia psittaci	Adélie	Ross Is.	EA	Antibodies (Sladen 1962)
Clostridium cadaveris	Adélie	Hope Bay	NEAP	Culture, Histopathology, PCR (Nievas *et al.* 2007)
Clostridium sporogenes	Adélie	Hope Bay	NEAP	Culture, Histopathology, PCR (Nievas *et al.* 2007)
Corynebacterium sp.	Chinstrap	King George Is.	SH	Culture (Jorge *et al.* 2002)
Edwardsiella tarda	Adélie	King George Is.	SH	Culture (Leotta *et al.* 2009)
		Hope Bay	NEAP	Culture (Leotta *et al.* 2009)
	Chinstrap	King George Is.	SH	Culture (Leotta *et al.* 2009)
	Gentoo	King George Is.	SH	Culture (Leotta *et al.* 2009)
		Hope Bay	NEAP	Culture (Leotta *et al.* 2009)
Enterococcus faecalis	Adélie	Warriner Is.	EA	Culture, whole genome NGS (Spence *et al.* 2014)
	Chinstrap	King George Is.	SH	Culture (Jorge *et al.* 2002)
Escherichia coli	Adélie	Hope Bay	NEAP	Culture, PCR (Nievas *et al.* 2007)
	Chinstrap	King George Is.	SH	Culture (Jorge *et al.* 2002)
	Gentoo	Ardley Is.	SH	Culture, RNA PCR (González-Acuña *et al.* 2013)
		Bernardo O'Higgins Base	NWAP	Culture, RNA PCR (González-Acuña *et al.* 2013)
		Waterboat Pt.	CWAP	Culture, RNA PCR (González-Acuña *et al.* 2013)
		Neko Harbour	CWAP	Culture, 16S rDNA PCR (Bonnedahl 2011; Bonnedahl *et al.* 2005)
Escherichia intermedia	Gentoo	King George Is.	SH	Culture (Jorge *et al.* 2002)

Micrococcus nishinomiyaensis	Chinstrap	King George Is.	SH	Culture (Jorge *et al.* 2002)
Micrococcus varians	Chinstrap	King George Is.	SH	Culture (Jorge *et al.* 2002)
Pasteurella multocida	Adélie	Hope Bay	NEAP	Culture, PFGE, ERIC-PCR (Leotta *et al.* 2006b)
	Chinstrap	Cooper Bay	SG	Symptoms (Government of South Georgia and the South Sandwich Islands 2005, 2013)
Pseudoramibacter alactolyticus	Chinstrap	King George Is.	SH	Culture [Reported as *Eubacterium alactolyticum*] (Jorge *et al.* 2002)
Salmonella enterica subsp. *enterica* (serovars in Ref column)	Adélie	Cape Denison	EA	Culture [Serovars: Enteriditis PT4, PT8, PT23] (Iveson *et al.* 2009)
		Ross Is.	EA	Culture [Serovars: Blockley, Infantis, Johannesburg, Panama, Typhimurium] (Oelke & Steiniger 1973)
		Hope Bay	NEAP	Culture, PFGE, RAPD-PCR [Serovar: Enteritidis] (Vigo *et al.* 2011)
	Gentoo	Bird Is.	SG	Culture, PFGE [Serovars: Enteriditis PT4/PT4-like/PT35, Havana] (Olsen *et al.* 1996; Palmgren *et al.* 2000)
		King George Is.	SH	Culture [Serovars: Enteriditis, Typhimurium] (Dimitrov *et al.* 2009)
		Bernardo O'Higgins Base	NWAP	Culture, PCR, antibiotic resistance and virulence typing [Serovar: Enteritidis] (Retamal *et al.* 2017)
Staphylococcus sp.	Adélie	Hope Bay	NEAP	Culture, PCR (Nievas *et al.* 2007)
Staphylococcus delphini	Adélie	James Ross Is.	NEAP	Culture, PCR, MLST, MALDI TOF MS, Whole genome NGS (Vrbovska *et al.* 2020)
Staphylococcus saprophyticus	Chinstrap	King George Is.	SH	Culture (Jorge *et al.* 2002)

Potentially Pathogenic Bacteria in Pygoscelids

While certain strains of bacteria listed in **Table 3** belong to genera routinely associated with disease in other vertebrates, their pathogenicity in penguins cannot be ascertained. Some, like *Actinomyces*, *Bacillus*, *Corynebacterium*, *Escherichia*, and *Micrococcus* are likely to be commensals that function as opportunistic pathogens in cases of compromised immunity, host stress, nutritional deficiency, or injury. Other bacteria have only recently been associated with rare infectious disease in humans, such as certain *Achromobacter* subspecies linked to pneumonia, meningitis, endocarditis and peritonitis (Awadh *et al.* 2017) and *Pseudoramibacter alactolyticus* found in endodontic infections (Siqueira & Rôças 2003). Their identification in a sample by culture or DNA-based analysis does not signify a pathogenic relationship with the host, merely the presence of such bacteria.

Our understanding of the pathology of bacterial infection in birds largely derives from research with poultry, where *Clostridium, Enterococcus*, *Escherichia coli*, *Mycoplasma*, *Salmonella, Staphylococcus aureus*, and *Streptococcus* bacteria can result in clinical symptoms, often due to secondary infection or septicaemia (Glisson 1998; McNamee *et al.* 2000; Merck 2016). These bacteria have been shown to cause symptoms in several wild species including pigeons, gulls, pheasants, and passerines (Benskin *et al.* 2009), and could reasonably be expected to act similarly in penguins. Some strains of *Clostridium* can cause acute outbreaks in poultry and *Clostridium cadaveris* and *C. sporogenes* were considered to be the causative agent in the death of at least two Adélie Penguins on the Antarctic Peninsula, with *Escherichia coli* and *Staphylococcus* also isolated from those birds (Nievas *et al.* 2007).

Salmonella enterica and *Campylobacter* spp. are well-researched taxa as a result of their economic importance in poultry flocks and burdens as human food-borne illnesses tied to poultry consumption. Salmonellosis due to host-adapted *S. enterica* subsp. *enterica* serovars Pullorum, Gallinarum can cause high mortality in chickens and turkeys of any age, while serovars Typhimurium and Enteritidis can cause clinical signs particularly in young chicks, and are commonly

associated with human food poisoning (Beal *et al.* 2004; Proux *et al.* 2002). *S. enterica* subsp. *enterica* serovar Typhimurium is also known to cause disease, reduce egg hatching probability, and lower chick survival rates in wild bird species (Benskin *et al.* 2009).

In contrast, *Campylobacter* spp. (mainly *C. jejuni*, and occasionally *C. coli* or *C. lari*) are important causes of human gastroenteritis, but are often relatively apathogenic in chickens although there is evidence of gastrointestinal damage, and diarrhoea in some cases (Humphrey *et al.* 2014). Wild birds carry diverse sequence types of *Campylobacter* sp. (Griekspoor *et al.* 2013; Waldenström *et al.* 2002) and contribute to some human infections (Cody *et al.* 2015). In Pygoscelid penguins, *Campylobacter* has been detected in the South Shetland Islands and Antarctic Peninsula (Barbosa *et al.* 2013; Bonnedahl *et al.* 2005; Cerdà-Cuéllar *et al.* 2019; García-Peña *et al.* 2017; González-Acuña *et al.* 2013; Johansson *et al.* 2018; Leotta *et al.* 2006a) although the clinical consequence of these infections are unknown.

A few bacterial pathogens are known to have high lethality in penguins, including the causative agent of avian cholera, *Pasteurella multocida*. This Gram-negative bacterium has a long history of outbreaks in domestic and wild birds, with at least 47 different avian families affected (Samuel *et al.* 2007). *P. multocida* mortality events are well-studied in waterbirds and wetlands, but is less well-documented in seabirds (Friend 2015). It is suspected to have caused several major mortality events in Antarctic flying birds (Leotta *et al.* 2003; Montalti *et al.* 1996; Parmelee *et al.* 1979) and in penguin breeding sites worldwide (Crawford *et al.* 1992; de Lisle *et al.* 1990; Kaschula & Truter 1951; Waller & Underhill 2007). In Pygoscelid penguins, three avian cholera events have been reported (see **Table 1**), linked to the deaths of Adélie Penguins, Chinstrap Penguins, kelp gulls, skuas, and other birds in South Georgia and the Antarctic Peninsula (Government of South Georgia and the South Sandwich Islands 2005, 2013; Leotta *et al.* 2006b).

Psittacosis, or avian chlamydiosis, is a serious illness in birds caused by *Chlamydia psittaci* that can be transmitted to humans and other mammals (Vanrompay *et al.* 1995). Strains of *C. psittaci* vary in virulence, and studies in poultry reveal a range of effects, with some infected birds

asymptomatic, while others can experience respiratory infection, diarrhoea, polyuria, conjunctivitis, and death (Andersen & Vanrompay 2000). Early studies to investigate chlamydiosis in Antarctic birds used serological techniques, and found antibody reactivity in Adélie Penguins, Gentoo Penguins, Rockhopper Penguins, a Royal Penguin (*Eudyptes schlegeli*), and sheathbills (Moore & Cameron 1969; Sladen 1962), though wild isolation of the bacterium has only been reported from a single Emperor Penguin chick (Cameron 1968). Outbreaks of *C. psittaci* in captive settings have been fatal to Magellanic Penguins (*Spheniscus magellanicus*) on at least two occasions (Jencek *et al.* 2012; Melton *et al.* 2006), so this bacterium can be presumed pathogenic in wild penguins.

Viral Research in Pygoscelids

As is typical of such remote and logistically-intensive investigations, current research on viral infections in Antarctica also faces challenges related to detection techniques and site sampling bias. Most recently reviewed by Smeele *et al.* (2018), viruses associated with Antarctic wildlife have largely been reported through serological techniques since the late 1970s, with a recent boom of novel identifications in seabirds and pinnipeds in the last decade, when high-throughput sequencing became more common. Antarctic environmental virology from soil (Adriaenssens *et al.* 2017; Zablocki *et al.* 2014a; Zablocki *et al.* 2014b), lake (Aguirre de Cárcer *et al.* 2016; López-Bueno *et al.* 2009; Rastrojo & Alcamí 2018; Yau *et al.* 2011; Yau & Seth-Pasricha 2019; Zawar-Reza *et al.* 2014), marine (Brum *et al.* 2015; Gong *et al.* 2018; Miranda *et al.* 2016), and ice-associated samples (Sommers *et al.* 2019) using high-throughput sequencing techniques has yielded an immense trove of new RNA and DNA viruses previously unknown to science, indicating the substantial viral diversity that could be associated with penguins and their environments. **Table 4** summarises studies reporting virus-related findings in Pygoscelid penguins.

Table 4. Studies indicating observations of viruses in Pygoscelid penguin hosts, by host species and location. Regions: CWAP = Central Western Antarctic Peninsula; EA = East Antarctica; FI = Falkland Islands/Malvinas; NEAP = Northeast Antarctic Peninsula; NWAP = Northwestern Antarctic Peninsula; SH = South Shetland Islands; SO = South Orkney Islands; SWAP = Southwestern Antarctic Peninsula. Detection Methods: ELISA = Enzyme-Linked Immunosorbent Assay; Ha = Haemagglutination Assay; HI = Haemagglutination Inhibition Assay; IDD = Immuno-Double-Diffusion; NGS = Next Generation Sequencing; PCR = Polymerase Chain Reaction; qPCR = Quantitative Polymerase Chain Reaction; RT-PCR = Reverse-Transcriptase Polymerase Chain Reaction; SDS-PAGE = Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis; VNT = Viral Neutralisation Test.

Virus Family / Genus	**Virus Species or Name**	**Host Species**	**Location**	**Region**	**Detection Method (Reference)**
Adenoviridae / *Siadenovirus*	*Penguin siadenovirus A* [Chinstrap Penguin adenovirus / Gentoo Penguin adenovirus]	Adélie	Ardley Is.	SH	PCR (Lee *et al.* 2016)
		Chinstrap	King George Is.	SH	PCR (Lee *et al.* 2014)
			Ardley Is.	SH	PCR, whole genome sequencing (Lee *et al.* 2016)
		Gentoo	Ardley Is.	SH	PCR, whole genome sequencing (Lee *et al.* 2016)
Astroviridae / Avastrovirus	Penguin astrovirus	Adélie	Ross Is.	EA	RT-PCR, NGS (Grimaldi *et al.* 2015b)
		Chinstrap	King George Is.	SH	RNA-Seq / NGS (Wille *et al.* 2020)
Birnaviridae / *Avibirnavirus*	*Infectious bursal disease virus* (IBDV)	Adélie	Mawson Station, Davis	EA	VNT, Antibodies (Gardner *et al.* 1997; Watts *et al.* 2009)

			Station, Terra Nova Station		
Calicivirdae / Unassigned	Wilkes Virus	Adélie	Kopaitik Is.	NWAP	RNA-Seq / NGS (Wille *et al.* 2020)
Circoviridae / *Circovirus*	Penguin circovirus	Adélie	Booth Is.	CWAP	Genomic NGS, cloned and Sanger sequenced (**Chapter 6**)
			Ross Is.	EA	Genomic NGS, cloned and Sanger sequenced (Morandini *et al.* 2019)
		Chinstrap	Booth Is.	CWAP	Genomic NGS, cloned and Sanger sequenced (**Chapter 6**)
Coronaviridae / *Deltacoronavirus*	Deltacoronavirus	Gentoo	Kopaitik Is.	NWAP	RNA-Seq / NGS (Wille *et al.* 2020)
Coronaviridae [?] / Unknown	Uncharacterised "Puffinosis-like" virus	Gentoo	Signy Is.	SO	Symptoms, histology (Mac Donald & Conroy 1971)
Herpesviridae / *Alphaherpesvirus*	Oates Virus	Chinstrap	King George Is.	SH	RNA-Seq / NGS (Wille *et al.* 2020)
Nairoviridae / *Orthonairovirus*	Taggert Virus	Chinstrap	Kopaitik Is.	NWAP	RNA-Seq / NGS (Wille *et al.* 2020)
Orthomyxoviridae / *Alphainfluenzavirus*	*Influenza A virus*	Adélie	King George Is.	SH	Viral culture, PCR, Antibodies (Hurt *et al.* 2014)
			Hope Bay	NEAP	Antigen-specific tests, HI (Baumeister *et al.* 2004)

			Bernardo O'Higgins Base	NWAP	Viral culture, PCR, Antibodies (Hurt *et al.* 2014)
			Peterson Is.	EA	HI, Immunodiffusion (Morgan & Westbury 1981)
			Ross Is.	EA	HI, Neuraminidase inhibition (Austin & Webster 1993)
		Chinstrap	Aitcho Is.	SH	RT-PCR, Symptoms (Barriga *et al.* 2016)
			King George Is.	SH	Antigen-specific tests, HI (Baumeister *et al.* 2004) ; RT-PCR, NGS, ELISA using nucleoprotein, Ha (Hurt *et al.* 2016) ; RNA-Seq / NGS (Wille *et al.* 2020)
		Gentoo	Aitcho Is.	SH	RT-PCR (Barriga *et al.* 2016)
			King George Is.	SH	Antigen-specific tests, HI (Baumeister *et al.* 2004)
			Bird Island	SG	Antibodies, ELISA (Wallensten *et al.* 2006)
Papillomaviridae / *Treisepsilonpapillomavirus*	*Treisepsilonpapillomavirus 1* (Pygoscelis adeliae papillomavirus)	Adélie	Ross Is.	EA	Genomic NGS, cloned and Sanger sequenced (Van Doorslaer *et al.* 2017; Varsani *et al.* 2014)

Paramyxoviridae / *Orthoavulavirus*	Unspecified Avian Paramyxoviruses	Adélie	Wilkes Base, Casey Stn., Peterson Is., D'Urville Stn.	EA	Virus isolation, Morphology, Immunodiffusion, Ha, HI (Morgan & Westbury 1981); Antibodies, HI, IDD, SDS-PAGE (Alexander *et al.* 1989)
			Vestfold Hills (Davis Stn.)	EA	Virus isolation, morphology, Ha test (Morgan & Westbury 1988)
			Ross Is.	EA	Indirect ELISA, electron microscopy (Austin & Webster 1993); HI, IDD, Antibodies, SDS-PAGE (Alexander *et al.* 1989)
			Mawson Stn.	EA	HI, IDD, Antibodies, SDS-PAGE (Alexander *et al.* 1989)
	Avian orthoavulavirus 1 (Avian paramyxovirus 1 [or Newcastle disease virus, NDV])	Adélie	Wilkes Base	EA	Virus isolation, Immunodiffusion, Ha, HI (Morgan & Westbury 1981)
			King George Is.	SH	Virus culture, RT-PCR and real-time PCR, Ha test, Sanger sequencing (Thomazelli *et al.* 2010)
	Avian orthoavulavirus 10-like	Adélie	King George Is.	SH	Virus isolation, RT-PCR, NGS (Wille *et al.* 2019)
	Avian orthoavulavirus 17 (Antarctic penguin virus A)	Adélie	Anvers Is.	CWAP	HI (Olivares *et al.* 2019)
			Kopaitik Is.	NWAP	RNA-Seq / NGS (Wille *et al.* 2020)

		Avian Is.	SWAP	HI (Olivares *et al.* 2019)
		King George Is.	SH	Virus isolation, RT-PCR, NGS (Wille *et al.* 2019); RNA-Seq/NGS (Wille *et al.* 2020)
	Chinstrap	Livingston Is.	SH	
	Gentoo	Ardley Is.	SH	HI (Olivares *et al.* 2019)
		Livingston Is.	SH	HI (Olivares *et al.* 2019)
		Kopaitik Is.	NWAP	HI (Olivares *et al.* 2019)
				Ha, HI, real-time qPCR, Sanger sequencing (Neira *et al.* 2017)
Avian orthoavulavirus 18 (Antarctic penguin virus B)	Adélie	Kopaitik Is.	NWAP	Virus isolation, RT-PCR, NGS (Wille *et al.* 2019)
	Chinstrap	Livingston Is.	SH	HI (Olivares *et al.* 2019)
	Gentoo	Livingston Is.	SH	HI (Olivares *et al.* 2019)
Avian orthoavulavirus 19 (Antarctic penguin virus C)	Adélie	King George Is.	SH	Virus isolation, RT-PCR, NGS (Wille *et al.* 2019)
		Kopaitik Is.	NWAP	Virus isolation, RT-PCR, NGS (Wille *et al.* 2019)
		Anvers Is.	CWAP	HI (Olivares *et al.* 2019)
	Chinstrap	Livingston Is.	SH	HI (Olivares *et al.* 2019)
		Kopaitik Is.	NWAP	Virus isolation, RT-PCR, NGS (Wille *et al.* 2019)

		Gentoo	Livingston Is. Doumer Is., Paradise Bay	SH CWAP	HI (Olivares *et al.* 2019) HI (Olivares *et al.* 2019)
Picobirnaviridae / Picobirnavirus	Hilary Virus	Adélie	Kopaitik Is.	NWAP	RNA-Seq / NGS (Wille *et al.* 2020)
Picornaviridae / *Megrivirus*	*Megrivirus E* (Penguin megrivirus)	Adélie	King George Is.	SH	NGS, whole genome sequencing (Yinda *et al.* 2019)
Picornaviridae / Unassigned	Amundsen Virus	Chinstrap	Kopaitik Is.	NWAP	RNA-Seq / NGS (Wille *et al.* 2020)
	Pingu Virus	Gentoo	King George Is.	SH	Virus isolation, NGS whole genome sequencing (de Souza *et al.* 2019)
	Ross Virus	Chinstrap	King George Is.	SH	RNA-Seq / NGS (Wille *et al.* 2020)
	Scott Virus	Chinstrap	King George Is.	SH	RNA-Seq / NGS (Wille *et al.* 2020)
	Shirase Virus	Adélie	King George Is. Kopaitik Is.	SH NWAP	RNA-Seq / NGS (Wille *et al.* 2020) RNA-Seq / NGS (Wille *et al.* 2020)
	Weddell Virus	Chinstrap	Kopaitik Is.	NWAP	RNA-Seq / NGS (Wille *et al.* 2020)
Polyomaviridae / *Gammapolyomavirus*	*Pygoscelis adeliae polyomavirus 1* [Adelie Penguin polyomavirus]	Adélie	Ross Is.	EA	Genomic NGS, cloned and Sanger sequenced (Varsani *et al.* 2015)
Poxviridae/ *Avipoxvirus*	Penguinpox virus	Gentoo	Multiple Islands	FI	Histology, Microscopy (Munro 2006); Symptoms (Pistorius 2009)
Reoviridae / Rotavirus	Mawson Virus	Adélie	King George Is.	SH	RNA-Seq / NGS (Wille *et al.* 2020)
	Shackleton Virus	Gentoo	Kopaitik Is.	NWAP	RNA-Seq / NGS (Wille *et al.* 2020)

As in other wild bird species, early interest in penguin virology was focused on zoonotic or commercially important viruses, namely the agents of avian influenza, infectious bursal disease, and Newcastle disease (Krauss *et al.* 1963; Morgan & Westbury 1981; Morgan *et al.* 1985; Morgan *et al.* 1981; Shellam & Mosbauer 2008). In Pygoscelids, viral investigations were concentrated on Macquarie Island Gentoo Penguins (negative findings, Morgan & Westbury 1981) and Adélie Penguins from East Antarctica until 2004. Since then, sampling has expanded to all three Pygoselids, with an increase in studies published from South Shetland Island sites and the Antarctic Peninsula. One study sampled Gentoo Penguins from South Georgia in search of influenza (Wallensten *et al.* 2006). To date, the only reports of viral infection from the South Orkneys or Falkland Islands were related to key mortality events, the 1969 die-off of hundreds of Gentoo Penguin chicks due to a "puffinosis-like" virus on Signy Island (Mac Donald & Conroy 1971), and widespread outbreaks of avian pox in the Falkland Islands in 2006 and 2007 (Munro 2006; Pistorius 2009).

Influenza A antibodies have been detected in all three Pygoscelids, along with South Polar Skuas, and Southern Giant Petrels (*Macronectes giganteus*) (Austin & Webster 1993; Baumeister *et al.* 2004; Miller *et al.* 2008; Morgan & Westbury 1981), but it was not until 2013 that H11N2 influenza A viruses were isolated from Adélie Penguins on the Antarctic Peninsula, and H5N5 virus from a Chinstrap Penguin in the South Shetland Islands (Hurt *et al.* 2016; Hurt *et al.* 2014), where H5N5 was again found in this population of Chinstrap Penguins in 2015 (Wille *et al.* 2020). Influenza viruses are thought to be widespread and prevalent among Antarctic birds, though the pathogenicity of these viruses remains unknown. Molecular evidence suggests penguins are reservoirs for long-term, undetected circulation of Influenza A viruses (Hurt *et al.* 2016; Wille *et al.* 2020).

The avibirnavirus *Infectious bursal disease virus* (IBDV) is, on the other hand, considered pathogenic in all birds. IBDV attacks the Bursa of Fabricius, a key site of B cell development in birds, and infection with IBDV can lead to substantial immunosuppression. Anti-IBDV antibodies have

been detected in Adélie and Emperor Penguin colonies near several East Antarctic research stations, leading to suspicion that it may have been introduced to the region by human activity (Gardner *et al.* 1997). Subsequent research, and detection of anti-IBDV antibodies in other wild populations, suggests that IBDV may be widespread in Antarctica (Watts *et al.* 2009), though the virus has not been isolated or sequenced from a penguin or Antarctic flying bird.

There is increasing evidence based on prevalence and molecular phylogenetics that Antarctic penguins also act as central reservoirs of avulaviruses (formerly avian paramyxoviruses) (Wille *et al.* 2019; Wille *et al.* 2020). Newcastle disease virus (*Avian orthoavulavirus 1* in Family *Paramyxoviridae*), has been isolated from Adélie Penguins in the South Shetland Islands, where seropositivity to antibodies was 33.3% (*n* = 100). Previous antibody tests in East Antarctica also showed presence of antibodies to this virus and other avian paramyxoviruses in Adélie Penguins (Alexander *et al.* 1989; Austin & Webster 1993; Morgan & Westbury 1981, 1988). Interestingly, Gentoo Penguins from Macquarie Island were sero-negative when sympatric Rockhopper Penguins (*Eudyptes chrysocome*) and Royal Penguins (*Eudyptes schlegeli*) were sero-positive (Morgan *et al.* 1981). Despite the detection of Newcastle disease virus in penguins, the pathogenic consequences of infection are unclear. Recent studies on orthoavulavirus in the South Shetland Islands and Antarctic Peninsula have yielded a number of novel viruses (*Avian orthoavulavirus 17, 18, 19*) with subsequent antibody detection across the range of the three Pygoscelids (Neira *et al.* 2017; Olivares *et al.* 2019; Wille *et al.* 2019) and viral sequence isolation in Adélie Penguins in the South Shetland Islands and Antarctic Peninsula (Wille *et al.* 2020). When transferred to chickens, these novel viruses established infection but did not cause clinical signs (Wille *et al.* 2019).

Very recently, sequencing-based techniques have allowed the characterisation of over a hundred new RNA viruses retrieved from Pygoscelid penguins. In one study alone, Wille *et al.* (2020) identified 13 novel species of likely penguin-associated RNA viruses and 82 likely diet- or microbiome-associated RNA viral species from cloacal swabs obtained from the South Shetland Islands and Antarctic Peninsula. Viruses identified included the first confirmed recovery in penguins

of a coronavirus (*Deltacoronavirus*) in Gentoo Penguins, which clustered with other described avian viruses in that genus, and the newly-named Wilkes Virus in the genus *Nacovirus* (*Calciviridae*), a taxon dominated by avian viruses. Two novel rotaviruses in the *Reoviridae* family: Shackleton Virus in Gentoo Penguins and Mawson Virus in Adélie Penguins were also sequenced, along with Hilary Virus, a picobirnavirus (*Picobirnaviridae*) in a clade with both avian and mammalian hosts (Wille *et al.* 2020). An avastrovirus (*Astroviridae*) was also recovered from Chinstrap Penguins, with 88% similarity to a short viral fragment previously found in Ross Sea Adélie Penguins (Grimaldi *et al.* 2015b), suggesting it could be penguin-specific (Wille *et al.* 2020).

In the family *Picornaviridae*, seven new species have been identified in Pygoscelids from the South Shetland Islands or Antarctic Peninsula: Penguin Megrivirus (*Megrivirus E*) in Adélie Penguins (Yinda *et al.* 2019) and six unassigned picornaviruses: Pingu Virus in Gentoo Penguins (de Souza *et al.* 2019), the Sincivirus-like Shirase Virus in Adélie Penguins, the Picornavirus B-like Weddell Virus and three Hepatovirus-like viruses (Ross, Scott, and Amundsen Viruses) in Chinstrap Penguins (Wille *et al.* 2020). As a family, picornaviruses can cause a range of diseases in mammals, including polio, hepatitis, and meningitis, but are also commonly asymptomatic or associated with mild respiratory or digestive symptoms (Zell *et al.* 2017). The effects of these novel penguin picornaviruses on their hosts are unknown.

Additionally, several novel DNA viruses in viral families associated with vertebrate disease have been sequenced, though again, pathogenicity is unknown in penguins. These include an adenovirus (*Penguin siadenovirus A*) found in all three Pygoscelids in the South Shetland Islands (Lee *et al.* 2014; Lee *et al.* 2016), two novel papillomavirus sequences (Varsani *et al.* 2014) and a polyomavirus isolated from Adélie Penguins on Ross Island (Varsani *et al.* 2015). Another 40 novel ssDNA viruses and 57 viral-like circular molecules have also been detected from Pygoscelids hosts at sites on the South Shetland Islands and Western Antarctic Peninsula (see **Chapter 5**).

Among DNA viruses with potential pathogenicity, a herpesvirus (Oates Virus) in genus *Alphaherpesvirus* was sequenced from Chinstrap Penguins in the South Shetland Islands, marking

a first detection in the Antarctic or in Pygoscelid penguins (Wille *et al.* 2020). This Chinstrap alphaherpesvirus was closely related to two contagious pathogenic avian viruses, *Gallid alphaherpesvirus 1*, which causes avian infectious laryngotracheitis in poultry (Carter & Wise 2006) and *Psittacid alphaherpesvirus 1*, which causes acute Pacheco's disease in parrots (Thureen & Keeler 2006). The novel Oates Virus is more distantly related to *Gallid alphaherpesvirus 2*, the causative agent of the neoplastic Marek's disease, as well as the Spheniscid herpesviruses isolated from rescued Magellanic Penguins (*Spheniscus magellanicus*) suffering from an acute respiratory disease in a rehabilitation facility in Brazil, and sequenced from asymptomatic wild penguins in Argentina (Niemeyer *et al.* 2017). Alphaherpesvirus has also been suspected as the aetiological agent of disease in captive outbreaks of respiratory disease in African Penguins, *S. demersus* (Kincaid et al. 1988; Parsons et al. 2015).

In recent years, a feather loss syndrome has been reported in Antarctic penguins (Barbosa *et al.* 2014; Grimaldi *et al.* 2015b). The aetiological agent has not yet been confirmed, despite several studies, but the recent isolation of a novel circovirus in Adélie Penguins (Morandini *et al.* 2019), including one individual exhibiting these symptoms, may be relevant. Circoviruses have been implicated in similar symptoms in psittacine birds, notably beak and feather disease virus (Pass & Perry 1984). The same species of Penguin Circovirus has subsequently been found in an Adélie Penguin and Chinstrap Penguin on the Antarctic Peninsula (see **Chapter 6**).

Eukaryotic Endoparasites of Pygoscelids

Eukaryotic parasite discovery presents its own challenges in penguins. Blood parasites, such as *Plasmodium* and *Babesia* can only be detected through blood sampling, which is more invasive than obtaining cloacal swabs or faecal specimens from a seabird colony. Haemoparasites can be detected directly on smears, or antibodies can indicate previous exposure to the pathogen. Gastrointestinal endoparasites, including the acanthocephalans, cestodes, nematodes, trematodes, and coccidian protozoa, are typically shed in faeces in the form of oocysts (coccidians), eggs or free-living worms. Definitive identification of coccidians (e.g. *Eimeria* spp.) requires good

quality imaging of sporulated oocysts or in the case of helminths, eggs or larval worms which is not always practical in faecal samples. Molecular techniques can help overcome this problem by detecting parasite DNA. Necropsies and histology tend to yield the most detailed understanding of helminth presence and parasite load in the gastrointestinal tracts of penguins, though such procedures are only ethical to perform after encountering a dead bird. Visual identification of parasitic worms in fatal cases help us understand the diversity of gastrointestinal parasites in a particular site or species, but since numbers are generally low cannot be used to estimate prevalence.

Table 5 summarises studies reporting eukaryotic endoparasite presence in Pygoscelids, and the reader is also directed towards Diaz *et al.* (2017) for a focussed discussion of Antarctic penguin macroparasites. Overall, much research on Antarctic seabird endoparasites continues to be performed as either a single-site (Cielecka *et al.* 1992) or single-species (Diaz *et al.* 2016; Diaz *et al.* 2013) study. These often rely on opportunistic collection of penguin cadavers (Fredes *et al.* 2006; Martin *et al.* 2016), while others have used faecal samples or both approaches (Fredes *et al.* 2007b; González-Acuña *et al.* 2013).

One interesting and highly relevant study experimentally examined the impact of gastrointestinal helminths on Chinstrap Penguin chick growth rates on Deception Island (South Shetlands). In the study, 22 broods of two chicks were identified and one sibling was treated with antihelminthics whilst the other remained untreated. After 15 days, treated chicks had a higher body mass than untreated siblings, implying a negative effect of helminths on penguin growth in early life stages (Palacios *et al.* 2012), which is likely to impact chick survival (Moreno *et al.* 1999).

Table 5. Studies reporting presence of eukaryotic endoparasites in Pygoscelid penguin hosts, by parasite phylum, host species and location. Regions: BOUV = Bouvet Island (South Atlantic); CRO = Crozet Islands; CWAP = Central Western Antarctic Peninsula; EA = East Antarctica; FI = Falkland Islands/Malvinas; H&M = Heard Island and McDonald Islands; KERG = Kerguelen Islands; MACQ = Macquarie Island; NEAP = Northeast Antarctic Peninsula; NWAP = Northwestern Antarctic Peninsula; SG = South Georgia; SH = South Shetland Islands; SO = South Orkney Islands; SWAP = Southwestern Antarctic Peninsula. Detection Methods: PCR = Polymerase Chain Reaction.

Endoparasite	**Host Species**	**Location**	**Region**	**Detection Method (Reference)**
Acanthocephala				
Corynosoma sp.	Chinstrap	Deception Is.	SH	Microscopy (Vidal *et al.* 2012)
	Gentoo	Macquarie Is.	MACQ	Microscopy (Edmonds 1954)
Corynosoma bullosum	Gentoo	King George Is., Elephant Is.	SH	Microscopy (Diaz *et al.* 2013; Hoberg 1986)
Corynosoma hamanni	Adélie	[Unknown]	Antarctica	Microscopy (Holloway & Bier 1967)
	Gentoo	King George Is.	SH	Microscopy (Diaz *et al.* 2013)
Corynosoma pseudohamanni	Adélie	[Unknown]	Antarctica	Microscopy (Zdzitowiecki 1991)
	Chinstrap	Livingston Is.	SH	Microscopy (Dimitrova *et al.* 1996)
Corynosoma shackletoni	Gentoo	King George Is.	SH	Microscopy (Diaz *et al.* 2013; Hoberg 1986)
		Bernardo O'Higgins Base	NWAP	Microscopy (González-Acuña *et al.* 2013)
		Waterboat Pt.	CWAP	Microscopy (González-Acuña *et al.* 2013)

Apicomplexa				
Babesia sp.	Chinstrap	Deception Is.	SH	Microscopy, PCR and Sanger Sequencing (Montero *et al.* 2016)
Cryptosporidium sp.	Adélie	Ardley Is.	SH	Microscopy (Fredes *et al.* 2007b)
	Gentoo	Ardley Is.	SH	Microscopy (Fredes *et al.* 2008a)
Eimeria pygosceli, *Eimeria* sp.	Adélie	Livingston Is., King George Is.	SH	Microscopy (Golemansky 2003, 2011)
	Chinstrap	Livingston Is.	SH	Microscopy (Golemansky 2003, 2011)
	Gentoo	Livingston Is., King George Is.	SH	Microscopy (Golemansky 2003, 2008, 2011)
Isospora sp.	Adélie	Livingston Is., King George Is.	SH	Microscopy (Golemansky 2003, 2011)
	Chinstrap	Livingston Is.	SH	Microscopy (Golemansky 2003, 2011)
	Gentoo	Livingston Is., King George Is.	SH	Microscopy (Golemansky 2003, 2008, 2011)
Plasmodium relictum	Gentoo	Kerguelen/Crozet Is.	KERG/CRO	Antibodies (Graczyk *et al.* 1995)
Cestoda				
Diphyllobothrium sp.	Gentoo	Ardley Is.	SH	Microscopy (González-Acuña *et al.* 2013)
Parorchites zederi	Adélie	King George Is.	SH	Microscopy (Cielecka *et al.* 1992; Diaz *et al.* 2016); Histology (Martin *et al.* 2016)
		Avian Is.	SWAP	Microscopy (Diaz *et al.* 2016); Histology (Martin *et al.* 2016)
	Chinstrap	King George Is.	SH	Microscopy (Cielecka *et al.* 1992)
		Deception Is.	SH	

				Microscopy (Vidal *et al.* 2012); Histology (Martin *et al.* 2016)
	Gentoo	King George Is.	SH	Microscopy (Cielecka *et al.* 1992; Diaz *et al.* 2013)
		Livingston Is.	SH	Microscopy (Georgiev *et al.* 1996); Histology (Martin *et al.* 2016)
		Bernardo O'Higgins Base	NWAP	Microscopy (González-Acuña *et al.* 2013)
		Rongé Island,	CWAP	Histology (Martin *et al.* 2016)
		Waterboat Pt.	CWAP	Microscopy (González-Acuña *et al.* 2013)
Tetrabothrius sp.	Adélie	Ardley Is.	SH	Microscopy (Fredes *et al.* 2008b)
		Avian Is.	SWAP	Microscopy (Diaz *et al.* 2016)
	Gentoo	Ardley Is.	SH	Microscopy (Fredes *et al.* 2006)
		Paradise Harbor	CWAP	Microscopy (Fredes *et al.* 2007a)
Tetrabothrius joubini	Chinstrap	King George Is.	SH	Microscopy (Cielecka *et al.* 1992)
		Livingston Is.	SH	Microscopy (Georgiev *et al.* 1996)
Tetrabothrius pauliani	Adélie	King George Is.	SH	Microscopy (Cielecka *et al.* 1992)
	Chinstrap	Bouvet Is.	BOUV	Microscopy (Andersen & Lysfjord 1982)
		King George Is.	SH	Microscopy (Cielecka *et al.* 1992)
		Livingston Is.	SH	Microscopy (Georgiev *et al.* 1996)
		Deception Is.	SH	Microscopy (Vidal *et al.* 2012)
	Gentoo	King George Is.	SH	Microscopy (Diaz *et al.* 2013)

		Kerguelen Is.	KERG	Microscopy (Prudhoe 1969)
Tetrabothrius wrighti	Adélie	Ross Is.	EA	Microscopy (Leiper & Atkinson 1914)
Unknown Cyclophyllid	Adélie	Ardley Is.	SH	Microscopy (Fredes *et al.* 2008b)
	Gentoo	Ardley Is.	SH	Microscopy (Fredes *et al.* 2006)
Nematoda				
Contracaecum sp.	Gentoo	Ardley Is.	SH	Microscopy (Fredes *et al.* 2006)
		Paradise Harbor	CWAP	Microscopy (Fredes *et al.* 2007a)
Contracaecum antarcticum (*osculatum*)	Adélie	[Unknown]	EA	Microscopy (Mawson 1953)
Contracaecum heardi	Gentoo	Heard Is.	H&M	Microscopy (Mawson 1953)
Stegophorus macronectes	Adélie	King George Is.	SH	Microscopy (Diaz *et al.* 2016; Zdzitowiecki & Drozdz 1980)
		Hope Bay	NEAP	Microscopy (Diaz *et al.* 2016)
	Chinstrap	Deception Is.	SH	Microscopy (Vidal *et al.* 2012); PCR and rDNA Characterisation (Vidal *et al.* 2016)
	Gentoo	Ardley Is.	SH	Microscopy (González-Acuña *et al.* 2013)
		King George Is.	SH	Microscopy (Diaz *et al.* 2013)
		Bernardo O'Higgins Base	NWAP	Microscopy (González-Acuña *et al.* 2013)
		Waterboat Pt.	CWAP	Microscopy (González-Acuña *et al.* 2013)
		Macquarie Is.	MACQ	Microscopy (Mawson 1953)
Streptocara sp.	Adélie	Ardley Is.	SH	Microscopy (Fredes *et al.* 2008b)

	Gentoo	Ardley Is.	SH	Microscopy (Fredes *et al.* 2006)
Stomachus sp.	Gentoo	Macquarie Is.	MACQ	Microscopy (Mawson 1953)
Tetrameres sp.	Adélie	King George Is.	SH	Microscopy (Diaz *et al.* 2016)
Tetrameres wetzeli	Gentoo	King George Is.	SH	Microscopy (Diaz *et al.* 2013)
Trematoda				
Renicola sloanei	Chinstrap,	United Kingdom (zoo)	Captive	Histology, microscopy (Wright 1954)
	Gentoo	United Kingdom (Edinburgh Zoo)	Captive	Histology, microscopy (Wright 1954)

Overall, eukaryotic endoparasite diversity is low in Pygoscelid penguins. Only one haemoparasite has been isolated in the wild, *Babesia* from Chinstrap Penguins on Deception Island (Montero *et al.* 2016), and antibodies to the malaria parasite *Plasmodium relictum* were detected in one study of Gentoo Penguins on Crozet and Kerguelen (Graczyk *et al.* 1995). The effect of *Babesia* is not yet known, but avian malaria is known to decimate most penguin species when infected in captive settings (Bak *et al.* 1984; Vanstreels *et al.* 2014).

Gastrointestinal parasite diversity of Pygoscelids is considered low, even relative to other penguin genera and seabirds (Diaz *et al.* 2017), likely due to the harsh abiotic conditions of Pygoscelid habitats that impede certain parasites from completing their life cycles, as well as host reliance on a narrow range of marine prey species found in the "wasp-waisted" Antarctic food web (Atkinson *et al.* 2014). Varied diets and complex food webs would facilitate exposure to more parasite species through consumption of intermediate hosts (Hoberg 1996). Wild Pygoscelid gut parasites include the apicomplexan genera *Cryptosporidium*, *Eimeria*, and *Isospora*; five nematode genera (*Contracaecum*, *Stegophorus*, *Streptocara*, *Stomachus*, and *Tetrameres*); at least three cestode genera (*Diphyllobothrium*, *Parorchites*, and *Tetrabothrius*); and one genus of acanthocephalans (*Corynosoma*). A single trematode species has been reported in Pygoscelids, though only in captive settings (Wright 1954), and has not been observed in the wild. Increased research effort has helped build a picture of gastrointestinal parasite prevalence across Pygoscelids, indicating that certain helminths are common in all species (namely *Parorchites zederi*, *Tetrabothrius* sp., and *Stegophorus macronectes*, see **Table 5** for references), as well as the proposed coccidian species *Eimeria pygosceli* (Golemansky 2011).

Molecular techniques have not yet been widely adopted in the study of parasites, with only the rDNA of *Stegophorus macronectes* reported (Vidal *et al.* 2016). Although DNA-based approaches have many advantages in identifying parasites that may not be readily visible in faecal samples from live penguins, they rely heavily on good preliminary characterisation of each parasite in sequence databases. For this reason, aside from continuing to examine parasite loads,

prevalence, and expansion to new regions and species, Antarctic parasitological research could benefit from more microscopy studies and the establishment of a suite of DNA based tools for determination of species identity as well as detection. These approaches should be extended to other marine taxa like fish and crustaceans, whose parasites are poorly described, but may be important within the often complex parasite life cycles that can involve multiple hosts.

Ectoparasites and Arthropod-Borne Pathogens in Pygoscelid Penguins

As might be expected in such harsh environments, the diversity of known ectoparasites on Pygoscelids is low, restricted to fewer than eight arthropod species (see **Table 6**). Insects, let alone parasitic ones, are rarely observed in the Antarctic, though the chewing lice *Austrogoniodes antarcticus* and *A. bifasciatus* have been described on Adélie Penguins in East Antarctica (Banks *et al.* 2006; Harrison 1937). Other species, *Austrogoniodes keleri* and *A. macquariensis* have only been identified on Pygoscelids in captive settings of imported penguins, but are otherwise known to occur on other penguin hosts in sub-Antarctic islands (Clay 1967; Clay & Moreby 1970). *A. gressitti* appears to have the widest wild range among these lice, found in South Georgia on Chinstrap Penguins and Gentoo Penguins, and elsewhere in the South Atlantic and Antarctic Peninsula on Gentoo Penguins (Banks *et al.* 2006; Clay & Moreby 1970; González-Acuña *et al.* 2013; Murray *et al.* 1991). The bird flea *Parapsyllus longicornis* is found in sub-Antarctic regions, where it has been associated with other penguin species, but only reported once on a Gentoo Penguin in the Falkland Islands (De Meillon *et al.* 1952).

Table 6. Studies reporting presence of ectoparasites on Pygoscelid penguin hosts or in nests, by arthropod order, host species and location. Regions: CWAP = Central Western Antarctic Peninsula; EA = East Antarctica; FI = Falkland Islands/Malvinas; MACQ = Macquarie Island; NWAP = Northwestern Antarctic Peninsula; SG = South Georgia; SH = South Shetland Islands.

Ectoparasite (Arthropoda)	Host Species	Location	Region	Detection Method (Reference)
Order Ixodida (Ticks)				
Ixodes uriae	Adélie	King George Is.	SH	Visual identification (Barbosa *et al.* 2011)
		Anvers/Humble/Cormorant Is.	CWAP	Visual identification (Benoit *et al.* 2009)
		Petermann Is.	CWAP	Visual identification (Lynch *et al.* 2010)
	Chinstrap	Deception Is.	SH	Visual identification (Barbosa *et al.* 2011)
		Livingston Is.	SH	Visual identification (Barbosa *et al.* 2011)
		Rongé Is.	CWAP	Visual identification (Barbosa *et al.* 2011)
	Gentoo	Ardley Is.	SH	Visual identification (González-Acuña *et al.* 2013)
		Livingston Is.	SH	Visual identification (Barbosa *et al.* 2011)
		Rongé Is.,	CWAP	Visual identification (Barbosa *et al.* 2011)
		Doumer Is.	CWAP	Visual identification (Casanueva & Moyano 2000)
		Waterboat Pt.	CWAP	Visual identification (González-Acuña *et al.* 2013)
Order Phthiraptera (Lice)				
Austrogoniodes antarcticus	Adélie	Ross Is.	EA	Visual identification (Banks *et al.* 2006; Harrison 1937)

Austrogoniodes bifasciatus	Adélie	Ross Is.	EA	Visual identification (Clay 1967)
Austrogoniodes gressitti	Chinstrap	South Georgia	SG	Visual identification (Clay & Moreby 1970)
	Gentoo	South Georgia	SG	Visual identification (Clay & Moreby 1970; Murray *et al.*
		Sea Lion Is.	FI	1991)
		Ardley Is.	SH	Visual identification (Banks *et al.* 2006)
		Bernardo O'Higgins Base	NWAP	Visual identification (González-Acuña *et al.* 2013)
		Waterboat Pt.	CWAP	Visual identification (González-Acuña *et al.* 2013)
		Anvers Is.	CWAP	Visual identification (González-Acuña *et al.* 2013)
				Visual identification (Murray *et al.* 1991)
Austrogoniodes keleri	Gentoo	London Zoo, from Falkland Is.	Captive	Visual identification (Clay 1967)
Austrogoniodes	Chinstrap	London Zoo, from South	Captive	Visual identification (Clay & Moreby 1970)
macquariensis	Gentoo	Georgia	Captive	Visual identification (Clay 1967)
		London Zoo, from Falkland Is.		
*Naubates prioni**	Gentoo	Macquarie Is.	MACQ	Visual identification in Watson (1967) *considered straggler or contaminant (Palma *et al.* 2002; Palma & Pilgrim 2002)
Order Siphonaptera (Fleas)				
Parapsyllus longicornis	Gentoo	Falkland Is.	FI	Visual identification (De Meillon *et al.* 1952)

The most prevalent ectoparasite among Pygoscelid breeding areas is the hard-bodied tick *Ixodes uriae*, which is widely associated with seabirds worldwide, including in the Arctic (McCoy *et al.* 2005). Heavy infestations of ticks as blood-suckers may affect chick survival in extreme cases, as is known to occur in King Penguins (Gauthier-Clerc *et al.* 1998; Mangin *et al.* 2003), though ticks are normally of greater concern due to their role as disease vectors (Muñoz-Leal & González-Acuña 2015; Wille *et al.* 2020). In its natural range on sub-Antarctic islands, *Ixodes uriae* has been known to infect penguins with various pathogens, to include *Borrelia* spirochaetes, *Babesia* haemoparasites, and several arboviruses in the families *Nairoviridae*, *Phleboviridae*, *Reoviridae* and *Flaviviridae* (Earle *et al.* 1993; Major *et al.* 2009), though studies have not yet detected these in sympatric Gentoo Penguins. While exploring the RNA virome of Pygoscelids and *I. uriae* ticks in their nests, Wille *et al.* (202) found nine (8 novel) tick-borne viruses, two of which are thought to have the potential to infect penguins. Taggert Virus, a nairovirus in the genus *Orthonairovirus* (order Bunyavirales) among those identified in penguin-associated ticks on Macquarie Island (Doherty *et al.* 1975), was also found in Chinstrap Penguins on Kopaitik Island, and is closely related to Crimean-Congo Haemorrhagic fever virus, a known pathogen. The novel Fennes Virus found in ticks from Pygoscelid colonies clusters genetically in the genus *Coltivirus* (*Reoviridae*), which also includes Colorado tick fever virus, other tick-borne pathogens and a viral species from African bats (Wille *et al.* 2020).

DNA of *Borrelia burgdorferi* sensu lato, the causative bacterial agent of Lyme disease in humans, has previously been detected in an *I. uriae* tick parasitizing a King Penguin on the Crozet archipelago (Schramm *et al.* 2014), and a "rickettsia-like organism" (bacterium) was isolated from *I. uriae* on Rockhopper and Macaroni penguins from Mayes Island, Kerguelen (Chastel *et al.* 1993). Babesiosis and borreliosis have been implicated in illness and potentially death, in African Penguins and vagrant King Penguins (Parsons *et al.* 2018; Parsons *et al.* 2016; Yabsley *et al.* 2012). In Pygoscelid penguins, *Ixodes uriae* is considered the only suitable vector for transmitting the

Babesia blood parasites found recently in healthy-looking Chinstrap Penguins on Deception Island (Montero *et al.* 2016).

In other sympatric seabirds, including penguins in sub-Antarctic regions, *Ixodes* populations typically exhibit strong host preference and genetic structure at the local level (McCoy *et al.* 2013). The exception to this segregation has been in closely-related species that present similar host environments for the ticks, such as Macaroni penguins (*Eudyptes chrysocome*) and Rockhopper Penguins (*Eudyptes chrysolophus*) on Kerguelen (McCoy *et al.* 2005). Population genetic studies of *Ixodes* ticks from the South Shetland Islands and Antarctic Peninsula have revealed that there is relatively high genetic diversity among ticks in this region, and no genetic structure was found between tick populations exploiting different Pygoscelids (McCoy *et al.* 2012). The wide distribution of these ticks across major Pygoscelid breeding areas likely means that greater sampling effort will yield further reports of tick-borne pathogens. These are likely to be found not just in Chinstrap Penguins, but in the congeneric species as well. Further study of host-*Ixodes* interactions and the epidemiology of its potential pathogens is needed to understand any health or fitness effects on Pygoscelid penguins.

Conclusion and Future Directions

Changes in migratory patterns of birds, and the potential for pathogen range shifts or importation are increasingly likely as warming temperatures affect the Antarctic Peninsula and human visitation increases (Frenot *et al.* 2005; Hughes *et al.* 2011; IAATO 2019; Whinam *et al.* 2005). Penguins play an essential role in their diverse Southern Ocean habitats, making them useful sentinels for detecting disruptions in their environments (Ainley 2002). In light of threats from shifting ice coverage, competition with krill fisheries, pollution, and anthropogenic disturbance, penguin population fluctuations are an indicator of overall ecosystem health (Cimino *et al.* 2016; Jenouvrier *et al.* 2014; Trathan *et al.* 2015). The current state of pathogen presence in this rapidly changing environment merits widespread investigation alongside the stresses

experienced by penguins and other Antarctic biota. With high levels of endemism in Antarctica and the Sub-Antarctic islands, studying the introduction, presence, and spread of disease-causing pathogens provides clues for understanding the population trends seen in penguin populations.

The substantial impact of disease introduction to bird species endemic to islands is well-known. The endangered penguin population in the Galápagos Islands (*Spheniscus mendiculus*) has been closely monitored for disease, with approximately 1,200 individuals remaining (Boersma *et al.* 2013). The steep decline of African Penguins (*Spheniscus demersus*) since the 1980s into endangered status by 2010 has been astonishing, but is largely explained by decreased prey availability, oil spills, and feral predators. In light of these threats, however, mortality events and the spread of disease are likely contributing to decreases in breeding success and population declines (Crawford *et al.* 1992; Crawford *et al.* 2006; Crawford & Dyer 1995; Parsons & Underhill 2005).

In spite of these areas of veterinary and ecological focus for threatened species, the interest in understanding penguin pathogens is relatively recent and remains limited in the context of sub-Antarctic and Antarctic species, perhaps due to the relatively large numbers, wide-ranging distribution, and conservation designations of members of the genera *Aptenodytes* (King Penguin = IUCN Least Concern; Emperor Penguins = Near Threatened) and *Pygoscelis* (Adélie, Chinstrap, and Gentoo Penguins: Least Concern) compared to other penguin taxa that are considered more vulnerable or endangered.

For Pygoscelids, important contributions to the field have come from several multi-site assessments in the South Shetland Islands and Antarctic Peninsula Region (Barbosa *et al.* 2011; Diaz *et al.* 2016; García-Peña *et al.* 2017; Olivares *et al.* 2019) and from long-term studies in East Antarctica, but many of these remain restricted to a single pathogen or pathogen type (bacteria, viruses, gastrointestinal parasites, or ectoparasites). However, there are many gaps in our understanding of pathogen prevalence in geographic areas like the Falkland Islands, South

Georgia, Marion Island, Îles Crozet, and Îles Kerguelen, all of which are frequented by biologists. Some portions of the breeding range have never been assessed for pathogen presence, such as the South Sandwich Islands, or have not been reported in the literature for several decades, such as Macquarie Island, Heard Island and McDonald Islands. A baseline of investigation across a large swath of Antarctic seabird ranges is critical to achieving the recommendations established for disease monitoring in the area and ensuring the preservation of the fragile ecosystem. With this knowledge, we can then address questions of whether these parasites appear to be host species- or genus-specific, are restricted in their geographic occurrence, or might be shared across larger extents of host ranges to inform conservation management.

In the last decade, molecular techniques involving metagenomics and high-throughput sequencing have helped overcome some of the challenges to understanding pathogen presence and diversity in the Antarctic and sub-Antarctic. However, an integrated approach to characterising viruses, bacteria, and macroparasites is still underutilised. Further knowledge of the virulence and epidemiology of these parasites is required to determine the role they play in Antarctic and sub-Antarctic environments. Only then can we posit the relationship between the host and its macro- and microparasites, building an understanding of how these are related, and how they have evolved over time. Equipped with the answers to these questions, we can begin to integrate ecological data to ask questions about the cost of parasitism for these hosts and the future viability of host populations in order to better protect life in the planet's shrinking cryosphere.

Acknowledgments

We thank Arvind Varsani for advising on the taxonomy and nomenclature of viruses.

Author Contributions

HL conducted the literature review and wrote the paper under the guidance of ALS and TH, who provided feedback on its content.

References

Adriaenssens EM, Kramer R, Van Goethem MW*, et al.* (2017) Environmental drivers of viral community composition in Antarctic soils identified by viromics. *Microbiome* **5**, 83.

Aguirre de Cárcer D, López-Bueno A, Alonso-Lobo JM, Quesada A, Alcamí A (2016) Metagenomic analysis of lacustrine viral diversity along a latitudinal transect of the Antarctic Peninsula. *Fems Microbiology Ecology* **92**.

Ainley DG (2002) *The Adélie penguin: bellweather of climate change* Columbia University Press, New York, NY.

Aislabie J, Jordan S, Ayton J*, et al.* (2009a) Bacterial diversity associated with ornithogenic soil of the Ross Sea region, Antarctica. *Can J Microbiol* **55**, 21-36.

Aislabie J, Ryburn J, Sarmah A (2009b) Culturable microbes in shallow groundwater underlying ornithogenic soil of Cape Hallett, Antarctica. *Can J Microbiol* **55**, 12-20.

Alexander DJ, Manvell RJ, Collins MS*, et al.* (1989) Characterization of Paramyxoviruses isolated from penguins in Antarctica and Sub-Antarctica during 1976-1979. *Archives of Virology* **109**, 135-143.

Andersen AA, Vanrompay D (2000) Avian chlamydiosis. *Rev Sci Tech* **19**, 396-404.

Andersen KI, Lysfjord S (1982) The Functional Morphology of the Scolex of 2 *Tetrabothrius Rudolphi* Species (Cestoda, Tetrabothriidae) from Penguins. *Zeitschrift Fur Parasitenkunde-Parasitology Research* **67**, 299-307.

Atkinson A, Hill SL, Barange M*, et al.* (2014) Sardine cycles, krill declines, and locust plagues: revisiting 'wasp-waist' food webs. *Trends in Ecology & Evolution* **29**, 309-316.

Austin FJ, Webster RG (1993) Evidence of ortho- and paramyxoviruses in fauna from Antarctica. *J Wildl Dis* **29**, 568-571.

Awadh H, Mansour M, Aqtash O, Shweihat Y (2017) Pneumonia due to a Rare Pathogen: *Achromobacter xylosoxidans*, Subspecies denitrificans. *Case reports in infectious diseases* **2017**, 3969682-3969682.

Bak UB, Park JC, Lim YJ (1984) An Outbreak of Malaria in Penguins at the Farmland Zoo Korea. *Korean Journal of Parasitology* **22**, 267-272.

Banks JC, Cary SC, Hogg ID (2009) The phylogeography of Adelie penguin faecal flora. *Environ Microbiol* **11**, 577-588.

Banks JC, Palma RL, Paterson AM (2006) Cophylogenetic relationships between penguins and their chewing lice. *Journal of Evolutionary Biology* **19**, 156-166.

Barbosa A, Balague V, Valera F*, et al.* (2016) Age-Related Differences in the Gastrointestinal Microbiota of Chinstrap Penguins (*Pygoscelis antarctica*). *PLoS One* **11**, e0153215.

Barbosa A, Benzal J, Vidal V*, et al.* (2011) Seabird ticks (*Ixodes uriae*) distribution along the Antarctic Peninsula. *Polar Biology* **34**, 1621-1624.

Barbosa A, Colominas-Ciuró R, Coria N*, et al.* (2014) First record of feather-loss disorder in Antarctic penguins. *Antarctic Science* **27**, 69-70.

Barbosa A, De Mas E, Benzal J*, et al.* (2013) Pollution and physiological variability in gentoo penguins at two rookeries with different levels of human visitation. *Antarctic Science* **25**, 329-338.

Barbosa A, Palacios MJ (2009) Health of Antarctic birds: a review of their parasites, pathogens and diseases. *Polar Biology* **32**, 1095-1115.

Barriga GP, Boric-Bargetto D, San Martin MC*, et al.* (2016) Avian Influenza Virus H5 Strain with North American and Eurasian Lineage Genes in an Antarctic Penguin. *Emerg Infect Dis* **22**, 2221-2223.

Baumeister E, Leotta G, Pontoriero A*, et al.* (2004) Serological evidences of influenza A virus infection in Antarctica migratory birds. *International Congress Series* **1263**, 737-740.

Beal RK, Wigley P, Powers C*, et al.* (2004) Age at primary infection with *Salmonella enterica* serovar Typhimurium in the chicken influences persistence of infection and subsequent immunity to re-challenge. *Vet Immunol Immunopathol* **100**, 151-164.

Benoit JB, Lopez-Martinez G, Elnitsky MA, Lee RE, Jr., Denlinger DL (2009) Increase in feeding by the tick, *Ixodes uriae*, on Adelie penguins during a prolonged summer. *Antarctic Science* **21**, 151-152.

Benskin CMH, Wilson K, Jones K, Hartley IR (2009) Bacterial pathogens in wild birds: a review of the frequency and effects of infection. *Biological Reviews* **84**, 349-373.

Black CE (2016) A comprehensive review of the phenology of *Pygoscelis* penguins. *Polar Biology*, 1-28.

Boersma PD, Steinfurth A, Merlen G*, et al.* (2013) Galápagos Penguin (*Spheniscus mendiculus*). In: *Penguins: Natural History and Conservation* (eds. Borboroglu PG, Boersma PD), pp. 286-302. University of Washington Press, Seattle, WA.

Bonnedahl J (2011) *Antibiotic Resistance in Enterobacteriaceae Isolated from Wild Birds*, Uppsala University.

Bonnedahl J, Broman T, Waldenstrom J*, et al.* (2005) In search of human-associated bacterial pathogens in Antarctic wildlife: report from six penguin colonies regularly visited by tourists. *Ambio* **34**, 430-432.

Borowicz A, McDowall P, Youngflesh C*, et al.* (2018) Multi-modal survey of Adelie penguin mega-colonies reveals the Danger Islands as a seabird hotspot. *Sci Rep* **8**, 3926.

Brandao ML, Moreira J, Luque JL (2014) Checklist of Platyhelminthes, Acanthocephala, Nematoda and Arthropoda parasitizing penguins of the world. *Check List* **10**, 562-573.

Brum JR, Ignacio-Espinoza JC, Roux S*, et al.* (2015) Patterns and ecological drivers of ocean viral communities. *Science* **348**, 1261498.

Burton-Johnson A, Black M, Fretwell P, Kaluza-Gilbert J (2016) An automated methodology for differentiating rock from snow, clouds and sea in Antarctica from Landsat 8 imagery: A new rock outcrop map and area estimation for the entire Antarctic continent. *The Cryosphere* **10**, 1665-1677.

Cameron AS (1968) The The isolation of a psittacosis-lymphogranuloma venerum (PL) agent from an Emperor Penguin (*Aptenodytes forsteri*) chick. *Australian Journal of Experimental Biology and Medical Science* **46**, 647-649.

Carter GR, Wise DJ (2006) Herpesviridae. In: *Concise Review of Veterinary Virology* (eds. Carter GR, Wise DJ, De Miguel NA). International Veterinary Information Service, USA.

Casanueva ME, Moyano HI (2000) *Ixodes uriae* White 1852 (Acari: Ixodidae) asociado con el pinguino papua, *Pygoscelis papua* (Forster), en la peninsula antartica. *Gayana* **64**, 235-236.

Cerdà-Cuéllar M, Moré E, Ayats T*, et al.* (2019) Do humans spread zoonotic enteric bacteria in Antarctica? *Science of the Total Environment* **654**, 190-196.

Chan JF-W, To KK-W, Chen H, Yuen K-Y (2015) Cross-species transmission and emergence of novel viruses from birds. *Current Opinion in Virology* **10**, 63-69.

Chastel C, Demazure M, Chastel O*, et al.* (1993) A Rickettsia-like Organism from *Ixodes uriae* Ticks Collected on the Kerguelen Islands (French Sub-Antarctic Territories). *Acta Virologica* **37**, 11-20.

Chown SL, Clarke A, Fraser CI*, et al.* (2015) The changing form of Antarctic biodiversity. *Nature* **522**, 431-438.

Cielecka D, Wojciechowska A, Zdzitowiecki K (1992) Cestodes from penguins on King George Island (South Shetlands, Antarctica). *Acta Parasitologica Polonica* **37**, 65-72.

Cimino MA, Lynch HJ, Saba VS, Oliver MJ (2016) Projected asymmetric response of Adelie penguins to Antarctic climate change. *Sci Rep* **6**, 28785.

Clarke J, Kerry K (2000) Diseases and parasites of penguins. *Penguin Conservation* **13**, 5-24.

Clay T (1967) Mallophaga (Biting Lice) and Anoplura (Sucking Lice) Part I: *Austrogoniodes* (Mallophaga) Parasitic on Penguins (Sphenisciformes). In: *Entomology of Antarctica* (ed. Gressitt JL), pp. 149-155. American Geophysical Union.

Clay T, Moreby C (1970) Mallophaga and Anoplura of Subantarctic islands. *Pacific Insects Monograph* **23**, 216-220.

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

Cody AJ, McCarthy ND, Bray JE*, et al.* (2015) Wild bird-associated *Campylobacter jejuni* isolates are a consistent source of human disease, in Oxfordshire, United Kingdom. *Environmental Microbiology Reports* **7**, 782-788.

Crawford RJM, Allwright DM, Heyl CW (1992) High Mortality of Cape Cormorants (*Phalacrocorax capensis*) off Western South Africa in 1991 caused by *Pasteurella multocida*. *Colonial Waterbirds* **15**, 236-238.

Crawford RJM, Barham PJ, Underhill LG*, et al.* (2006) The influence of food availability on breeding success of African penguins *Spheniscus demersus* at Robben Island, South Africa. *Biological Conservation* **132**, 119-125.

Crawford RJM, Dyer BM (1995) Responses by four seabird species to a fluctuating availability of Cape Anchovy *Engraulis capensis* off south Africa. *Ibis* **137**, 329-339.

de Lisle GW, Stanislawek WL, Moors PJ (1990) *Pasteurella multocida* infections in rockhopper penguins (*Eudyptes chrysocome*) from Campbell Island, New Zealand. *J Wildl Dis* **26**, 283-285.

De Meillon B, Zumpt F, Expeditions ANAR (1952) *The Fleas of Sea Birds in the Southern Ocean* Antarctic Division, Department of External Affairs.

de Souza WM, Fumagalli MJ, Martin MC*, et al.* (2019) Pingu virus: A new picornavirus in penguins from Antarctica. *Virus Evol* **5**, vez047.

Debruyne L, Broman T, Bergstrom S*, et al.* (2010) *Campylobacter subantarcticus* sp nov., isolated from birds in the sub-Antarctic region. *International Journal of Systematic and Evolutionary Microbiology* **60**, 815-819.

Dewar ML, Arnould JP, Dann P*, et al.* (2013a) Interspecific variations in the gastrointestinal microbiota in penguins. *Microbiologyopen* **2**, 195-204.

Dewar ML, Arnould JPY, Allnutt TR*, et al.* (2017) Microbiota of little penguins and short-tailed shearwaters during development. *PLoS One* **12**, e0183117.

Dewar ML, Arnould JPY, Dann P*, et al.* (2013b) Interspecific variations in the gastrointestinal microbiota in penguins. *Microbiologyopen* **2**, 195-204.

Dewar ML, Arnould JPY, Krause L*, et al.* (2014) Influence of Fasting during Moult on the Faecal Microbiota of Penguins. *PLoS One* **9**.

Diaz JI, Fusaro B, Longarzo L*, et al.* (2016) Gastrointestinal helminths of Adélie penguins (*Pygoscelis adeliae*) from Antarctica. *Polar Research* **35**.

Diaz JI, Fusaro B, Longarzo L*, et al.* (2013) Gastrointestinal helminths of Gentoo penguins (*Pygoscelis papua*) from Stranger Point, 25 de Mayo/King George Island, Antarctica. *Parasitol Res* **112**, 1877-1881.

Diaz JI, Fusaro B, Vidal V*, et al.* (2017) Macroparasites in Antarctic Penguins. In: *Biodiversity and Evolution of Parasitic Life in the Southern Ocean* (eds. Klimpel S, Kuhn T, Mehlhorn H), pp. 183-204. Springer International Publishing, Cham.

Dimitrov K, Metcheva R, Kenarova A (2009) *Salmonella* presence - an indicator of direct and indirect human impact on gentoo in Antarctica. *Biotechnology & Biotechnological Equipment* **23**, 246-249.

Dimitrova ZM, Chipev NH, Georgiev BB (1996) Record of *Corynosoma pseudohamanni* Zdzitowiecki, 1984 (Acanthocephala, Polymorphidae) in birds at Livingston Island, South Shetlands, with a review of Antarctic avian acanthocephalans. *Bulgarian Antarctic Research Life Sciences* **1**, 102-110.

Doherty RL, Carley JG, Murray MD*, et al.* (1975) Isolation of arboviruses (Kemerovo group, Sakhalin group) from *Ixodes uriae* collected at Macquarie Island, Southern ocean. *The American Journal of Tropical Medicine and Hygiene* **24**, 521-526.

Earle RA, Huchzermeyer FW, Bennett GF, Brossy JJ (1993) *Babesia peircei* sp. nov. from the Jackass Penguin. *South African Journal of Zoology* **28**, 88-90.

Edmonds SJ (1954) Acanthocephala collected by the Australian National Antarctic Research Expedition on Heard Island and Macquarie Island during 1948-50. *Transactions of the Royal Society of South Australia* **78**, 141-144.

Egevang C, Stenhouse IJ, Phillips RA*, et al.* (2010) Tracking of Arctic terns *Sterna paradisaea* reveals longest animal migration. *Proceedings of the National Academy of Sciences* **107**, 2078-2081.

Fan J, Li L, Han J*, et al.* (2013) Diversity and structure of bacterial communities in Fildes Peninsula, King George Island. *Polar Biology* **36**, 1385-1399.

Fredes F, Diaz A, Raffo E, Munoz P (2008a) *Cryptosporidium* spp. oocysts detected using acid-fast stain in faeces of gentoo penguins (*Pygoscelis papua*) in Antarctica. *Antarctic Science* **20**, 495-496.

Fredes F, Madariaga C, Raffo E*, et al.* (2007a) Gastrointestinal parasite fauna of gentoo penguins (*Pygoscelis papua*) from the Peninsula Munita, Bahia Paraiso, Antarctica. *Antarctic Science* **19**, 93-94.

Fredes F, Raffo E, Muñoz P (2007b) Short Note: First report of *Cryptosporidium* spp. oocysts in stool of Adélie penguin from the Antarctic using acid-fast stain. *Antarctic Science* **19**, 437-438.

Fredes F, Raffo E, Munoz P, Herrera M (2006) Gastrointestinal parasite fauna of nine gentoo penguin chicks (*Pygoscelis papua*) found dead in an Antarctic territory protected zone. *Parasitologia Latinoamericana* **61**, 179-182.

Fredes F, Raffo E, Munoz P, Herrera M, Godoy C (2008b) Gastrointestinal parasite fauna in Adelia penguin (*Pygoscelis adeliae*) in an Antarctic territory protected zone. *Parasitologia Latinoamericana* **63**, 64-68.

Frenot Y, Chown SL, Whinam J*, et al.* (2005) Biological invasions in the Antarctic: extent, impacts and implications. *Biological Reviews* **80**, 45-72.

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

Friend M (2015) Avian cholera. In: *Field manual of wildlife diseases: U.S. Geological Survey Techniques and Methods 15* (eds. Franson JC, Friend M, Gibbs SEJ, Wild MA), pp. 76-92. U.S. Geological Survey, Madison, WI.

García-Peña FJ, Llorente MT, Serrano T*, et al.* (2017) Isolation of *Campylobacter* spp. from Three Species of Antarctic Penguins in Different Geographic Locations. *Ecohealth* **14**, 78-87.

Gardner H, Kerry K, Riddle M, Brouwer S, Gleeson L (1997) Poultry virus infection in Antarctic penguins. *Nature* **387**, 245-245.

Gauthier-Clerc M, Clerquin Y, Handrich Y (1998) Hyperinfestation by ticks *Ixodes uriae*: A possible cause of death in adult King Penguins, a long-lived seabird. *Colonial Waterbirds* **21**, 229-233.

Gauthier-Clerc M, Eterradossi N, Toquin D*, et al.* (2002) Serological survey of the king penguin, *Aptenodytes patagonicus* in Crozet Archipelago for antibodies to infectious bursal disease, influenza A and Newcastle disease viruses. *Polar Biology* **25**, 316-319.

Georgiev BB, Vasileva GP, Chipev NH, Dimitrova ZM (1996) Cestodes of seabirds at Livingston Island, South Shetlands. *Bulgarian Antarctic Research Life Sciences* **1**, 111-127.

Glisson JR (1998) Bacterial respiratory disease of poultry. *Poultry Science* **77**, 1139-1142.

Golemansky V (2003) *Eimeria pygosceli* n. sp. (Coccidia: Eimeriidae) from the penguins (Pygoscelidae) of the Livingston Island (the Antarctic). *Acta Zoologica Bulgarica* **55**, 3-8.

Golemansky V (2008) Diversity and prevalence of the coccidians (Apicomplexa : Eucoccidia) of Gentoo penguins (*Pygoscelis papua* FORSTER, 1781) from King George Island (South Shetland Archipelago, the Antarctic). *Acta Zoologica Bulgarica* **60**, 71-75.

Golemansky V (2011) Coccidian parasites (Apicomplexa) of penguins (*Pygoscelis* ssp.) from Livingston Island and King George Island, the Antarctic. *Polish Polar Research* **32**, 263-268.

Gong Z, Liang Y, Wang M*, et al.* (2018) Viral Diversity and Its Relationship With Environmental Factors at the Surface and Deep Sea of Prydz Bay, Antarctica. *Frontiers in Microbiology* **9**, 2981-2981.

González-Acuña D, Hernández J, Moreno L*, et al.* (2013) Health evaluation of wild gentoo penguins (*Pygoscelis papua*) in the Antarctic Peninsula. *Polar Biology* **36**, 1749-1760.

Government of South Georgia and the South Sandwich Islands (2005) South Georgia Government Update.

Government of South Georgia and the South Sandwich Islands (2013) Biosecurity Protocol 2013.

Graczyk TK, Cranfield MR, Brossy JJ*, et al.* (1995) Detection of Avian Malaria Infections in Wild and Captive Penguins. *Journal of the Helminthological Society of Washington* **62**, 135-141.

Green CJ, Trathan PN, Preston M (2006) A new automated logging gateway to study the demographics of macaroni penguins (*Eudyptes chrysolophus*) at Bird Island, South Georgia: testing the reliability of the system using radio telemetry. *Polar Biology* **29**, 1003-1010.

Griekspoor P, Colles FM, McCarthy ND*, et al.* (2013) Marked host specificity and lack of phylogeographic population structure of *Campylobacter jejuni* in wild birds. *Mol Ecol* **22**, 1463-1472.

Grimaldi W, Jabour J, Woehler EJ (2011) Considerations for minimising the spread of infectious disease in Antarctic seabirds and seals. *Polar Record* **47**, 56-66.

Grimaldi W, Seddon P, Lyver PB, Nakagawa S, Tompkins D (2015a) Infectious diseases of Antarctic penguins: current status and future threats. *Polar Biology* **38**, 591-606.

Grimaldi WW, Hall RJ, White DD*, et al.* (2015b) First report of a feather loss condition in Adelie penguins (*Pygoscelis adeliae*) on Ross Island, Antarctica, and a preliminary investigation of its cause. *Emu - Austral Ornithology* **115**, 185-189.

Grzesiak J, Kaczynska A, Gawor J*, et al.* (2020) A smelly business: Microbiology of Adelie penguin guano (Point Thomas rookery, Antarctica). *Sci Total Environ* **714**, 136714.

Guo Y, Wang N, Li G*, et al.* (2018) Direct and Indirect Effects of Penguin Feces on Microbiomes in Antarctic Ornithogenic Soils. *Front Microbiol* **9**, 552.

Harrison LJS (1937) Mallophaga and Siphunculata. Scientific Reports, Australasian Antarctic Expedition 1911-14, series C - zoology and botany 2, pp. 1-47.

Hoberg EP (1986) Aspects of Ecology and Biogeography of Acanthocephala in Antarctic Seabirds. *Annales De Parasitologie Humaine Et Comparee* **61**, 199-214.

Hoberg EP (1996) Faunal Diversity among Avian Parasite Assemblages: The Interaction of History, Ecology, and Biogeography in Marine Systems. *Bulletin of the Scandinavian Society of Parasitology* **6**, 65-89.

Holloway HL, Jr., Bier JW (1967) Notes on the host specificity of *Corynosoma hamanni* (Linstow, 1892). *Bulletin of the Wildlife Disease Association* **3**, 76-78.

Hughes KA, Convey P, Maslen NR, Smith RIL (2010) Accidental transfer of non-native soil organisms into Antarctica on construction vehicles. *Biological Invasions* **12**, 875-891.

Hughes KA, Lee JE, Tsujimoto M*, et al.* (2011) Food for thought: Risks of non-native species transfer to the Antarctic region with fresh produce. *Biological Conservation* **144**, 1682-1689.

Humphrey S, Chaloner G, Kemmett K*, et al.* (2014) *Campylobacter jejuni* Is Not Merely a Commensal in Commercial Broiler Chickens and Affects Bird Welfare. *Mbio* **5**, e01364-01314.

Hurt AC, Su YCF, Aban M*, et al.* (2016) Evidence for the Introduction, Reassortment, and Persistence of Diverse Influenza A Viruses in Antarctica. *J Virol* **90**, 9674-9682.

Hurt AC, Vijaykrishna D, Butler J*, et al.* (2014) Detection of Evolutionarily Distinct Avian Influenza A Viruses in Antarctica. *Mbio* **5**.

IAATO (2011) *Guidelines for Visitors to the Antarctic.* IAATO, Newport, RI. http://www.iaato.org/visitor-guidelines and http://iaato.org/en_GB/dont-pack-a-pest

IAATO (2019) *Tourism Statistics Download.* International Association of Antarctica Tour Operators. https://iaato.org/tourism-statistics-327mnsyd

Iveson JB, Shellam GR, Bradshaw SD*, et al.* (2009) *Salmonella* infections in Antarctic fauna and island populations of wildlife exposed to human activities in coastal areas of Australia. *Epidemiology and Infection* **137**, 858-870.

Jabour J (2009) National Antarctic programs and their impact on the environment. In: *Health of Antarctic Wildlife: A Challenge for Science and Policy* (eds. Kerry K, Riddle M), pp. 211-229. Springer, London.

Jencek JE, Beaufrere H, Tully TN, Jr*., et al.* (2012) An Outbreak of *Chlamydophila psittaci* in an Outdoor Colony of Magellanic Penguins (Spheniscus magellanicus). *Journal of Avian Medicine and Surgery* **26**, 225-231.

Jenouvrier S, Holland M, Stroeve J*, et al.* (2014) Projected continent-wide declines of the emperor penguin under climate change. *Nature Clim. Change* **4**, 715-718.

Johansson H, Ellström P, Artursson K*, et al.* (2018) Characterization of *Campylobacter* spp. isolated from wild birds in the Antarctic and Sub-Antarctic. *PLoS One* **13**, e0206502.

Jorge MC, Najle R, Montalti D (2002) Cloacal bacterial flora on Antarctic birds. *Rivista Italiana di Ornitologia* **71**, 163-169.

Kaschula VR, Truter DE (1951) Fowl cholera in sea gulls on Dassen Island. *Journal of South African Veterinary Medical Association* **22**, 191-192.

Kim OS, Chae N, Lim HS*, et al.* (2012) Bacterial diversity in ornithogenic soils compared to mineral soils on King George Island, Antarctica. *J Microbiol* **50**, 1081-1085.

Kincaid AL, Bunton TE, Cranfield M (1988) Herpesvirus-like Infection in Black-Footed Penguins (*Spheniscus demersus*). *Journal of Wildlife Diseases* **24**, 173-175.

Krauss H, Paulick C, Huchzermeyer F, Gylstorff I (1963) Newcastle disease a king penguin (*Aptenodytes patachonica*). *Deutsche tierarztliche Wochenschrift* **70**, 307-309.

Lachish S, Knowles SC, Alves R, Wood MJ, Sheldon BC (2011) Fitness effects of endemic malaria infections in a wild bird population: the importance of ecological structure. *J Anim Ecol* **80**, 1196-1206.

Le Maho Y, Whittington JD, Hanuise N*, et al.* (2014) Rovers minimize human disturbance in research on wild animals. *Nature Methods* **11**, 1242-1244.

Lee SY, Kim JH, Park YM*, et al.* (2014) A novel adenovirus in Chinstrap penguins (*Pygoscelis antarctica*) in Antarctica. *Viruses* **6**, 2052-2061.

Lee SY, Kim JH, Seo TK*, et al.* (2016) Genetic and Molecular Epidemiological Characterization of a Novel Adenovirus in Antarctic Penguins Collected between 2008 and 2013. *PLoS One* **11**, e0157032.

Lee WY, Cho H, Kim M*, et al.* (2019) Faecal microbiota changes associated with the moult fast in chinstrap and gentoo penguins. *PLoS One* **14**, e0216565.

Leiper RT, Atkinson EL (1914) Helminths of the British Antarctic Expedition 1910-1913. *Proceedings of the Zoological Society of London* **1**, 222-226.

Leotta G, Vigo G, Giacoboni G (2006a) Isolation of *Campylobacter lari* from seabirds in Hope Bay, Antarctica. *Polish Polar Research* **27**, 303-308.

Leotta GA, Chinen I, Vigo GB, Pecoraro M, Rivas M (2006b) Outbreaks of avian cholera in Hope Bay, Antarctica. *Journal of Wildlife Diseases* **42**, 259-270.

Leotta GA, Pineyro P, Serena S, Vigo GB (2009) Prevalence of *Edwardsiella tarda* in Antarctic wildlife. *Polar Biology* **32**, 809-812.

Leotta GA, Rivas M, Chinen I*, et al.* (2003) Avian cholera in a southern giant petrel (*Macronectes giganteus*) from Antarctica. *Journal of Wildlife Diseases* **39**, 732-735.

Lepage D (2015) Avibase: The World Bird Database.

Levin II, Zwiers P, Deem SL*, et al.* (2013) Multiple Lineages of Avian Malaria Parasites (*Plasmodium*) in the Galapagos Islands and Evidence for Arrival via Migratory Birds. *Conservation Biology* **27**, 1366-1377.

López-Bueno A, Tamames J, Velázquez D*, et al.* (2009) High Diversity of the Viral Community from an Antarctic Lake. *Science* **326**, 858-861.

Lynch H, Fagan W, Naveen R (2010) Population trends and reproductive success at a frequently visited penguin colony on the western Antarctic Peninsula. *Polar Biology* **33**, 493-503.

Ma D, Zhu R, Ding W*, et al.* (2013) Ex-situ enzyme activity and bacterial community diversity through soil depth profiles in penguin and seal colonies on Vestfold Hills, East Antarctica. *Polar Biology* **36**, 1347-1361.

Mac Donald JW, Conroy JWH (1971) Virus Disease Resembling Puffinosis in the Gentoo Penguin *Pygoscelis papua* on Signy Island South Orkney Islands. *British Antarctic Survey Bulletin*, 80-82.

Major L, Linn ML, Slade RW*, et al.* (2009) Ticks associated with macquarie island penguins carry arboviruses from four genera. *PLoS One* **4**, e4375.

Mangin S, Gauthier-Clerc M, Frenot Y, Gendner JP, Le Maho Y (2003) Ticks *Ixodes uriae* and the breeding performance of a colonial seabird, king penguin Aptenodytes patagonicus. *Journal of Avian Biology* **34**, 30-34.

Martin MA, Ortiz JM, Seva J*, et al.* (2016) Mode of Attachment and Pathology Caused by *Parorchites Zederi* in Three Species of Penguins: *Pygoscelis Papua, Pygoscelis Adeliae*, and *Pygoscelis Antarctica* in Antarctica. *J Wildl Dis* **52**, 568-575.

Mawson PM (1953) Parasitic Nematoda collected by the Australian National Antarctic Research Expedition: Heard Island and Macquarie Island, 1948–1951. *Parasitology* **43**, 291-297.

McCoy KD, Beis P, Barbosa A*, et al.* (2012) Population genetic structure and colonisation of the western Antarctic Peninsula by the seabird tick *Ixodes uriae*. *Marine Ecology Progress Series* **459**, 109-120.

McCoy KD, Chapuis E, Tirard C*, et al.* (2005) Recurrent evolution of host-specialized races in a globally distributed parasite. *Proceedings of the Royal Society B-Biological Sciences* **272**, 2389-2395.

McCoy KD, Leger E, Dietrich M (2013) Host specialization in ticks and transmission of tick-borne diseases: a review. *Front Cell Infect Microbiol* **3**, 57.

McNamee PT, Smyth JA, Smyth JA (2000) Bacterial chondronecrosis with osteomyelitis ('femoral head necrosis') of broiler chickens: A review. *Avian Pathology* **29**, 253-270.

Melton B, McCloskey V, Orlando R (2006) Management and treatment of avian chlamydiosis in a captive Magellanic penguin (*Spheniscus magellanicus*) colony at the San Francisco Zoo. *Animal Keepers' Forum* **33**, 533-540.

Merck (2016) *The Merck Veterinary Manual, 11th Edition* Merck & Co., Inc., Whitehouse Station, NJ.

Meyer-Rochow VB, Gal J (2003) Pressures produced when penguins pooh - calculations on avian defaecation. *Polar Biology* **27**, 56-58.

Miller GD, Watts JM, Shellam GR (2008) Viral antibodies in south polar skuas around Davis Station, Antarctica. *Antarctic Science* **20**, 455-461.

Miranda JA, Culley AI, Schvarcz CR, Steward GF (2016) RNA viruses as major contributors to Antarctic virioplankton. *Environ Microbiol* **18**, 3714-3727.

Montalti D, Coria NR, Curtosi A (1996) Unusual deaths of Subantarctic Skuas *Catharacta antarctica* at Hope Bay, Antarctica. *Marine Ornithology* **24**, 39-40.

Montero E, Gonzalez LM, Chaparro A*, et al.* (2016) First record of *Babesia* sp. in Antarctic penguins. *Ticks Tick Borne Dis* **7**, 498-501.

Moore BW, Cameron AS (1969) Research Note: Chlamydia Antibodies in Antarctic Fauna. *Avian Diseases* **13**, 681-684.

Morandini V, Dugger KM, Ballard G*, et al.* (2019) Identification of a Novel Adelie Penguin Circovirus at Cape Crozier (Ross Island, Antarctica). *Viruses* **11**.

Moreno E, Barbosa A, de León A, Fargallo JA (1999) Phenotypic selection on morphology at independence in the Chinstrap penguin *Pygoscelis antarctica*.

Morgan IR, Westbury HA (1981) Virological studies of Adelie Penguins (*Pygoscelis adeliae*) in Antarctica. *Avian Dis* **25**, 1019-1026.

Morgan IR, Westbury HA (1988) Studies of viruses in penguins in the Vestfold Hills. *Hydrobiologia* **165**, 263-269.

Morgan IR, Westbury HA, Campbell J (1985) Viral infections of little blue penguins (*Eudyptula minor*) along the southern coast of Australia. *J Wildl Dis* **21**, 193-198.

Morgan IR, Westbury HA, Caple IW, Campbell J (1981) A survey of virus infection in sub-antarctic penguins on Macquarie Island, Southern Ocean. *Aust Vet J* **57**, 333-335.

Muñoz-Leal S, González-Acuña D (2015) The tick *Ixodes uriae* (Acari: Ixodidae): Hosts, geographical distribution, and vector roles. *Ticks and Tick-Borne Diseases* **6**, 843-868.

Munro G (2006) Outbreak of avian pox virus in gentoo penguins in the Falklands. February 2006. Falklands Conservation, Stanley, Falkland Islands.

Murray MD, Palma RL, Pilgrim RLC (1991) Ectoparasites of Australian, New Zealand and Antarctic Birds. Appendix I. In: *Handbook of Australian, New Zealand and Antarctic Birds, Volume I, Part A* (eds. Marchant S, Higgins PJ). Oxford University Press, Melbourne.

Nash AA, Dalziel RG, Fitzgerald JR (2015) Attachment to and Entry of Microorganisms into the Body. In: *Mims' Pathogenesis of Infectious Disease (Sixth Edition)* (eds. Nash AA, Dalziel RG, Fitzgerald JR), pp. 9-49. Academic Press, Boston.

Neira V, Tapia R, Verdugo C*, et al.* (2017) Novel Avulaviruses in Penguins, Antarctica. *Emerging Infectious Diseases* **23**, 1212-1214.

Niemeyer C, Favero CM, Shivaprasad HL*, et al.* (2017) Genetically diverse herpesviruses in South American Atlantic coast seabirds. *PLoS One* **12**, e0178811-e0178811.

Nievas VF, Leotta GA, Vigo GB (2007) Subcutaneous clostridial infection in Adelie penguins in Hope Bay, Antarctica. *Polar Biology* **30**, 249-252.

Oelke H, Steiniger F (1973) Salmonella in Adelie penguins (*Pygoscelis adeliae*) and South Polar Skuas (*Catharacta maccormicki*) on Ross Island, Antarctica. *Avian Dis* **17**, 568-573.

Olivares F, Tapia R, Galvez C*, et al.* (2019) Novel penguin Avian avulaviruses 17, 18 and 19 are widely distributed in the Antarctic Peninsula. *Transbound Emerg Dis* **66**, 2227-2232.

Olsen B, Bergstrom S, McCafferty DJ, Sellin M, Wistrom J (1996) *Salmonella enteritidis* in Antarctica: Zoonosis in man or humanosis in penguins? *Lancet* **348**, 1319-1320.

Olsen B, Duffy DC, Jaenson TGT*, et al.* (1995) Transhemispheric Exchange of Lyme Disease Spirochetes by Seabirds. *Journal of Clinical Microbiology* **33**, 3270-3274.

Palacios MJ, Valera F, Barbosa A (2012) Experimental assessment of the effects of gastrointestinal parasites on offspring quality in chinstrap penguins (*Pygoscelis antarctica*). *Parasitology* **139**, 819-824.

Palma RL, Horning DS, Expeditions ANAR (2002) *The lice (Insecta: Phthiraptera) from Macquarie Island* Australian Antarctic Division.

Palma RL, Pilgrim RLC (2002) A revision of the genus *Naubates* (Insecta: Phthiraptera: Philopteridae). *Journal of the Royal Society of New Zealand* **32**, 7-60.

Palmgren H, McCafferty D, Aspan A*, et al.* (2000) *Salmonella* in sub-Antarctica: low heterogeneity in salmonella serotypes in South Georgian seals and birds. *Epidemiology and Infection* **125**, 257-262.

Pan H, Cole TL, Bi X*, et al.* (2019) High-coverage genomes to elucidate the evolution of penguins. *Gigascience* **8**.

Parmelee DF, Maxson SJ, Bernstein NP (1979) Fowl cholera outbreak among brown skuas at Palmer Station. *Antarctic Journal of the United States*, 168-169.

Parsons NJ, Gous TA, Cranfield MR*, et al.* (2018) Novel vagrant records and occurrence of vector-borne pathogens in King Penguins (*Aptenodytes patagonicus*) in South Africa. *Polar Biology* **41**, 79-86.

Parsons NJ, Gous TA, Schaefer AM, Vanstreels RE (2016) Health evaluation of African penguins *Spheniscus demersus* in southern Africa. *Onderstepoort J Vet Res* **83**, e1-e13.

Parsons NJ, Gous TA, van Wilpe E, Strauss V, Vanstreels RE (2015) Herpesvirus-like respiratory infection in African penguins *Spheniscus demersus* admitted to a rehabilitation centre. *Dis Aquat Organ* **116**, 149-155.

Parsons NJ, Underhill LG (2005) Oiled and injured African penguins *Spheniscus demersus* and other seabirds admitted for rehabilitation in the Western Cape, South Africa, 2001 and 2002. *African Journal of Marine Science* **27**, 289-296.

Pass DA, Perry RA (1984) The pathology of psittacine beak and feather disease. *Aust Vet J* **61**, 69-74.

Pistorius PA (2009) Falkland Island Seabird Monitoring Programme. Annual Report 2008/2009. Falklands Conservation, Stanley, Falkland Islands.

Proux K, Humbert F, Jouy E*, et al.* (2002) Improvements required for the detection of *Salmonella Pullorum* and *Gallinarum*. *Canadian journal of veterinary research = Revue canadienne de recherche veterinaire* **66**, 151-157.

Prudhoe S (1969) *Cestodes from Fish, Birds and Whales* Mawson Institute for Antarctic Research, The University of Adelaide.

Ramirez-Fernandez L, Trefault N, Caru M, Orlando J (2019) Seabird and pinniped shape soil bacterial communities of their settlements in Cape Shirreff, Antarctica. *PLoS One* **14**, e0209887.

Rastrojo A, Alcamí A (2018) Chapter Two - Viruses in Polar Lake and Soil Ecosystems. In: *Advances in Virus Research* (ed. Malmstrom CM), pp. 39-54. Academic Press.

Retamal P, Llanos-Soto S, Salas LM*, et al.* (2017) Isolation of drug-resistant *Salmonella enterica* serovar enteritidis strains in gentoo penguins from Antarctica. *Polar Biology* **40**, 2531-2536.

Romaniuk K, Ciok A, Decewicz P*, et al.* (2018) Insight into heavy metal resistome of soil psychrotolerant bacteria originating from King George Island (Antarctica). *Polar Biology* **41**, 1319-1333.

Ropert-Coudert Y, Wilson RP (2005) Trends and perspectives in animal-attached remote sensing. *Frontiers in Ecology and the Environment* **3**, 437-444.

Samuel MD, Botzler RG, Wobeser GA (2007) *Avian cholera*.

Schramm F, Gauthier-Clerc M, Fournier J-C*, et al.* (2014) First detection of *Borrelia burgdorferi* sensu lato DNA in king penguins (*Aptenodytes patagonicus* halli). *Ticks and Tick-Borne Diseases* **5**, 939-942.

Scientific Committee for Antarctic Research (SCAR) (2010) Preliminary Results from the International Polar Year Programme: Aliens in Antarctica (ed. Antarctic Treaty Consultative Meeting XXXIII CfEPCX, Working Paper 04. 3-14 May, 2010), Punta del Este, Uruguay.

Shellam G, Mosbauer A (2008) Status of Disease in the Emperor Penguins of Auster Rookery. *Australian Antarctic Data Centre*.

Siqueira JF, Jr., Rôças IN (2003) *Pseudoramibacter alactolyticus* in Primary Endodontic Infections. *Journal of Endodontics* **29**, 735-738.

Sladen WJL (1962) Studies on respiratory pathogens in Antarctica. *Polar Record* **11**, 318.

Smeele ZE, Ainley DG, Varsani A (2018) Viruses associated with Antarctic wildlife: From serology based detection to identification of genomes using high throughput sequencing. *Virus Res* **243**, 91-105.

Sommers P, Fontenele RS, Kringen T*, et al.* (2019) Single-Stranded DNA Viruses in Antarctic Cryoconite Holes. *Viruses* **11**, 1022.

Spence RJ, Pavasovic A, Smith JJ, Prentis PJ (2014) Draft Genome Sequence of *Enterococcus faecalis* Strain PF3, Isolated from Adelie Penguin Feces from Antarctica. *Genome announcements* **2**.

St. George TD, Doherty RL, Carley JG*, et al.* (1985) The Isolation of Arboviruses including a new Flavivirus and a new Bunyavirus from *Ixodes* (Ceratixodes) *uriae* (Ixodoidea: Ixodidae) Collected at Macquarie Island, Australia, 1975–1979. *The American Journal of Tropical Medicine and Hygiene* **34**, 406-412.

Teixeira LC, Yeargeau E, Balieiro FC*, et al.* (2013) Plant and bird presence strongly influences the microbial communities in soils of Admiralty Bay, Maritime Antarctica. *PLoS One* **8**, e66109.

Thomazelli LM, Araujo J, Oliveira DB*, et al.* (2010) Newcastle disease virus in penguins from King George Island on the Antarctic region. *Veterinary Microbiology* **146**, 155-160.

Thureen DR, Keeler CL (2006) Psittacid Herpesvirus 1 and Infectious Laryngotracheitis Virus: Comparative Genome Sequence Analysis of Two Avian Alphaherpesviruses. *Journal of Virology* **80**, 7863.

Trathan PN, Garcia-Borboroglu P, Boersma D*, et al.* (2015) Pollution, habitat loss, fishing, and climate change as critical threats to penguins. *Conservation Biology* **29**, 31-41.

Trivelpiece WZ, Hinke JT, Miller AK*, et al.* (2011) Variability in krill biomass links harvesting and climate warming to penguin population changes in Antarctica. *Proc Natl Acad Sci U S A* **108**, 7625-7628.

Ugolini FC (1972) Ornithogenic Soils of Antarctica. In: *Antarctic Terrestrial Biology* (ed. Llano GA), pp. 181-193. American Geophysical Union, Washington, D.C.

Van Doorslaer K, Ruoppolo V, Schmidt A*, et al.* (2017) Unique genome organization of non-mammalian papillomaviruses provides insights into the evolution of viral early proteins. *Virus Evol* **3**, vex027.

Vanrompay D, Ducatelle R, Haesebrouck F (1995) *Chlamydia psittaci* infections: a review with emphasis on avian chlamydiosis. *Vet Microbiol* **45**, 93-119.

Vanstreels RE, Kolesnikovas CK, Sandri S*, et al.* (2014) Outbreak of avian malaria associated to multiple species of *Plasmodium* in magellanic penguins undergoing rehabilitation in southern Brazil. *PLoS One* **9**, e94994.

Varsani A, Kraberger S, Jennings S*, et al.* (2014) A novel papillomavirus in Adelie penguin (*Pygoscelis adeliae*) faeces sampled at the Cape Crozier colony, Antarctica. *J Gen Virol* **95**, 1352-1365.

Varsani A, Porzig EL, Jennings S*, et al.* (2015) Identification of an avian polyomavirus associated with Adelie penguins (*Pygoscelis adeliae*). *J Gen Virol* **96**, 851-857.

Vidal V, Ortiz J, Diaz JI*, et al.* (2012) Gastrointestinal parasites in Chinstrap Penguins from Deception Island, South Shetlands, Antarctica. *Parasitology Research* **111**, 723-727.

Vidal V, Ortiz J, Diaz JI*, et al.* (2016) Morphological, molecular and phylogenetic analyses of the spirurid nematode *Stegophorus macronectes* (Johnston & Mawson, 1942). *J Helminthol* **90**, 214-222.

Vigo GB, Leotta GA, Ines Caffer M*, et al.* (2011) Isolation and characterization of *Salmonella enterica* from Antarctic wildlife. *Polar Biology* **34**, 675-681.

Vrbovska V, Sedlacek I, Zeman M*, et al.* (2020) Characterization of *Staphylococcus intermedius* Group Isolates Associated with Animals from Antarctica and Emended Description of *Staphylococcus delphini*. *Microorganisms* **8**.

Waldenström J, Broman T, Carlsson I*, et al.* (2002) Prevalence of *Campylobacter jejuni*, *Campylobacter lari*, and *Campylobacter coli* in different ecological guilds and taxa of migrating birds. *Applied and Environmental Microbiology* **68**, 5911-5917.

Wallensten A, Munster VJ, Osterhaus ADME*, et al.* (2006) Mounting evidence for the presence of influenza A virus in the avifauna of the Antarctic region. *Antarctic Science* **18**, 353-356.

Waller LJ, Underhill LG (2007) Management of avian cholera *Pasteurella multocida* outbreaks on Dyer Island, South Africa, 2002-2005. *African Journal of Marine Science* **29**, 105-111.

Watson KC (1967) *The terrestrial arthropoda of Macquarie Island. ANARE Scientific Reports, series B(1) Zoology* Australian Antarctic Division Department of External Affairs, Melbourne.

Watts JM, Miller GD, Shellam GR (2009) *Infectious bursal disease virus and Antarctic birds*.

Whinam J, Chilcott N, Bergstrom DM (2005) Subantarctic hitchhikers: expeditioners as vectors for the introduction of alien organisms. *Biological Conservation* **121**, 207-219.

Wille M, Aban M, Wang J*, et al.* (2019) Antarctic Penguins as Reservoirs of Diversity for Avian Avulaviruses. *J Virol* **93**.

Wille M, Harvey E, Shi M*, et al.* (2020) Sustained RNA virome diversity in Antarctic penguins and their ticks. *The ISME Journal*.

Woods R, Jones HI, Watts J, Miller GD, Shellam GR (2009) Diseases of Antarctic Seabirds. In: *Health of Antarctic wildlife: a challenge for science and policy* (eds. Kerry KR, Riddle M). Springer, London.

Wright CA (1954) Trematodes of the genus *Renicola* from birds in British zoos, with descriptions of two new species. *Proceedings of the Zoological Society of London* **124**, 51-61.

Yabsley MJ, Parsons NJ, Horne EC, Shock BC, Purdee M (2012) Novel relapsing fever *Borrelia* detected in African penguins (*Spheniscus demersus*) admitted to two rehabilitation centers in South Africa. *Parasitol Res* **110**, 1125-1130.

Yau S, Lauro FM, DeMaere MZ*, et al.* (2011) Virophage control of antarctic algal host-virus dynamics. *Proc Natl Acad Sci U S A* **108**, 6163-6168.

Yau S, Seth-Pasricha M (2019) Viruses of Polar Aquatic Environments. *Viruses* **11**, 189.

Yew WC, Pearce DA, Dunn MJ, Samah AA, Convey P (2017) Bacterial community composition in Adélie (*Pygoscelis adeliae*) and Chinstrap (*Pygoscelis antarctica*) Penguin stomach contents from Signy Island, South Orkney Islands. *Polar Biology* **40**, 2517-2530.

Yinda CK, Esefeld J, Peter HU, Matthijnssens J, Zell R (2019) Penguin megrivirus, a novel picornavirus from an Adelie penguin (*Pygoscelis adeliae*). *Arch Virol* **164**, 2887-2890.

Zablocki O, van Zyl L, Adriaenssens EM*, et al.* (2014a) High-Level Diversity of Tailed Phages, Eukaryote-Associated Viruses, and Virophage-Like Elements in the Metaviromes of Antarctic Soils. *Applied and Environmental Microbiology* **80**, 6888-6897.

Zablocki O, van Zyl L, Adriaenssens EM*, et al.* (2014b) Niche-dependent genetic diversity in Antarctic metaviromes. *Bacteriophage* **4**, e980125.

Zawar-Reza P, Arguello-Astorga GR, Kraberger S*, et al.* (2014) Diverse small circular single-stranded DNA viruses identified in a freshwater pond on the McMurdo Ice Shelf (Antarctica). *Infect Genet Evol* **26**, 132-138.

Zdanowski MK, Weglenski P, Golik P*, et al.* (2004) Bacterial diversity in Adelie penguin, *Pygoscelis adeliae*, guano: molecular and morpho-physiological approaches. *FEMS Microbiol Ecol* **50**, 163-173.

Zdzitowiecki K (1991) Antarctic Acanthocephala. In: *Synopses of the Antarctic Benthos. Volume III* (eds. Wagele JW, Sieg J), pp. 1-116. Koeltz Scientific Books, Koenigstein.

Zdzitowiecki K, Drozdz J (1980) Redescription of *Stegophorus macronectes* and Description of *Stegophorus acrotowskii* new species Nematoda spirurida from birds of South Shetlands the Antarctic. *Acta Parasitologica Polonica* **27**, 205-212.

Zell R, Delwart E, Gorbalenya AE*, et al.* (2017) ICTV Virus Taxonomy Profile: Picornaviridae. *J Gen Virol* **98**, 2421-2422.

Chapter 5. Identification and distribution of novel circular Rep-encoding ssDNA viruses and circular molecules in four Penguin species in South Georgia and the Antarctic Peninsula

Hila Levy[1], Rafaela S. Fontenele[2], Ciara Harding[2], Crystal Suazo[2], Simona Kraberger[2], Kara Schmidlin[2], Anni Djurhuus[3], Caitlin E. Black[4], Tom Hart[1], Adrian L. Smith[1*], and Arvind Varsani[2,5*]

1. Department of Zoology, University of Oxford, South Parks Road, Oxford, OX1 3SZ, United Kingdom
2. The Biodesign Center for Fundamental and Applied Microbiomics, Center for Evolution and Medicine, School of Life Sciences, Arizona State University, Tempe, AZ 85287-5001, USA
3. Faculty of Science and Technology, University of the Faroe Islands, Vestarabryggja 15, FO-100 Tórshavn, Faroe Islands
4. Department of Evolutionary Biology and Environmental Studies, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland
5. Structural Biology Research Unit, Department of Integrative Biomedical Sciences, University of Cape Town, Observatory, Cape Town 7701, South Africa

*Authors for Correspondence:

Adrian L. Smith: adrian.smith@zoo.ox.ac.uk

Arvind Varsani: arvind.varsani@asu.edu

Keywords: CRESS, circular molecules, penguins, viruses, Antarctica, *Pygoscelis antarcticus*, *Pygoscelis adeliae*, *Pygoscelis papua*

Abstract

There is growing interest in uncovering the viral diversity present in wild animal species. The remote Antarctic region is home to a wealth of uncovered microbial diversity, some of which is associated with its megafauna, including penguin species, the dominant avian biota. Penguins interface with a number of other biota in their roles as marine mesopredators, and several species overlap in their ranges and habitats. To characterize the DNA virome of these environmental sentinel species, cloacal swabs (n = 95) were obtained from King Penguins in South Georgia, and congeneric Adélie Penguins, Chinstrap Penguins, and Gentoo Penguins across the South Shetland Islands and Antarctic Peninsula. Using a combination of high-throughput sequencing, abutting primers-based PCR recovery of circular genomic elements, cloning and Sanger sequencing, we detected 97 novel sequences comprising 40 ssDNA viral genomes and 57 viral-like circular molecules from 45 individual penguins. We present their occurrence and prevalence patterns, with Chinstrap Penguins harbouring the highest number of new sequences. The novel Antarctic viruses identified appear to be host-specific, while one circular molecule was shared between sympatric Chinstrap and Gentoo Penguins. We also report viral genotype sharing between three adult-chick pairs, one in each Pygoscelid species. Sequence similarity network approaches coupled with Maximum likelihood phylogenies of the clusters indicate the 40 novel viral genomes do not fall within any known viral families, and likely fall within the recently established phylum *Cressdnaviricota* based on their replication-associated protein sequences. Similarly, 83 capsid protein sequences encoded by the viruses or viral-like circular molecules identified in this study do not cluster with any of those encoded by classified viral groups. Further research is warranted to expand knowledge of the Antarctic DNA virome and would help elucidate the importance of viral-like molecules in vertebrate host evolution.

Introduction

Our knowledge of global viral diversity remains limited, with only a small fraction of viruses known to affect humans and wildlife having been properly described or studied. In the case of Antarctic associated viruses, only a speck of the proverbial iceberg of viral diversity has surfaced in modern research. The Antarctic Polar Front, the marine boundary between Antarctic and sub-Antarctic waters formed in the Eocene, has played a part in the evolutionary divergence and high levels of endemism present in the Southern Ocean (Cole *et al.* 2019; Gavryushkina *et al.* 2017). Penguins (Order: Sphenisciformes) are diving seabird mesopredators distributed across the islands and continental landmass of the Southern Ocean who form an important part of the avian biomass in the region.

Our understanding of diversity in the vertebrate hosts of the Antarctic region remains limited, with early research using serological techniques typically focusing on well-known avian and poultry pathogenic viruses associated with influenza, Newcastle Disease, and infectious bursal disease (Austin & Webster 1993; Gardner *et al.* 1997; Morgan & Westbury 1981; Watts *et al.* 2009) or distemper viruses likely transferred from sled dogs to Antarctic pinnipeds (Bengtson *et al.* 1991; McFarlane 2009). A small number of symptomatic events linked to viruses, such as avian pox and "puffinosis-like" disease have also been recorded in case reports (Mac Donald & Conroy 1971; Munro 2006; Pistorius 2009). Nonetheless, our knowledge of existing Antarctic vertebrate viruses beyond these reports remained constrained until the advent of high-throughput sequencing techniques, uncovering a growing number of viral genomes in the last decade, most recently reviewed by Smeele *et al.* (2018) and complemented with identification of RNA viruses in penguins and their ectoparasites (Wille *et al.* 2020).

Recent high-throughput sequencing metagenomic analysis of soil (Adriaenssens *et al.* 2017; Zablocki *et al.* 2014a, b), lacustrine (Aguirre de Carcer *et al.* 2016; Lopez-Bueno *et al.* 2009; Rastrojo & Alcamí 2018; Yau *et al.* 2011; Yau & Seth-Pasricha 2019; Zawar-Reza *et al.* 2014),

marine (Brum *et al.* 2015; Gong *et al.* 2018; Miranda *et al.* 2016), and ice-associated samples (Sommers *et al.* 2019) in Antarctica have revealed robust viral profiles of genomes with novel characteristics. With an aim to uncover further viral diversity associated with penguins south of the Polar Front, we undertook a study to investigate the diversity of DNA viruses and viral-like circular molecules in four penguin species: the Adélie Penguin (*Pygoscelis adeliae*), Chinstrap Penguin (*P. antarcticus*), Gentoo Penguin (*P. papua*), and King Penguin (*Aptenodytes patagonicus*).

King Penguins are primarily sub-Antarctic seabirds with a circumpolar distribution on islands near the Polar Front. Adélie Penguins are considered true Antarctic, ice-loving birds, breeding along the entire Antarctic coast and southern islands of the Scotia Arc. Chinstrap Penguins, on the other hand, are restricted to South Georgia and the South Sandwich Islands, South Orkney Islands, South Shetland Islands and Western Antarctic Peninsula. The widest-ranging penguin species is the Gentoo Penguin, with a circumpolar distribution that spans 46°-66° S, breeding north and south of the Polar Front on Sub-Antarctic Islands and throughout the entire Scotia Arc and down the Western Antarctic Peninsula. Some of the four species therefore overlap in portions of their range and provide an opportunity to investigate virome patterns across sites and in sympatric breeders. We also evaluated whether viral sequences are shared between parent-chick pairs of Pygoscelid penguins.

Materials and Methods

Field sampling

Between 22 December 2015 and 17 January 2016, a total of 95 cloacal swab samples were obtained from penguins with no obvious signs of illness at eight sites on Deception Island, the South Shetland Islands, South Georgia, and the Antarctic Peninsula (**Supplementary Table S1, Supplementary Data 1**). These consisted of 17 Adélie Penguins (*Pygoscelis adeliae:* 9 adults, 8 chicks), 32 Chinstrap Penguins (*P. antarcticus*: 21 adults, 11 chicks), 26 Gentoo Penguins (*P.*

papua: 16 adults, 10 chicks), and 20 King Penguins (*Aptenodytes patagonicus*: 10 adults, 10 chicks) (**Figure 1**). Regular flocked swabs (Eswab™, Copan®, Murrieta, CA, USA) were used, stored in 1 ml of liquid Amies medium, which was stored frozen aside from transport with cold packs to the United Kingdom. Adult and juvenile King Penguins were sampled using a two-person handling technique as they transited within the colony. Pygoscelid penguins were sampled in transit to or from the ocean, or where applicable, while on the nest in order to confirm chick-adult parentage, minimising disturbance as much as possible and ensuring parents returned to their chicks.

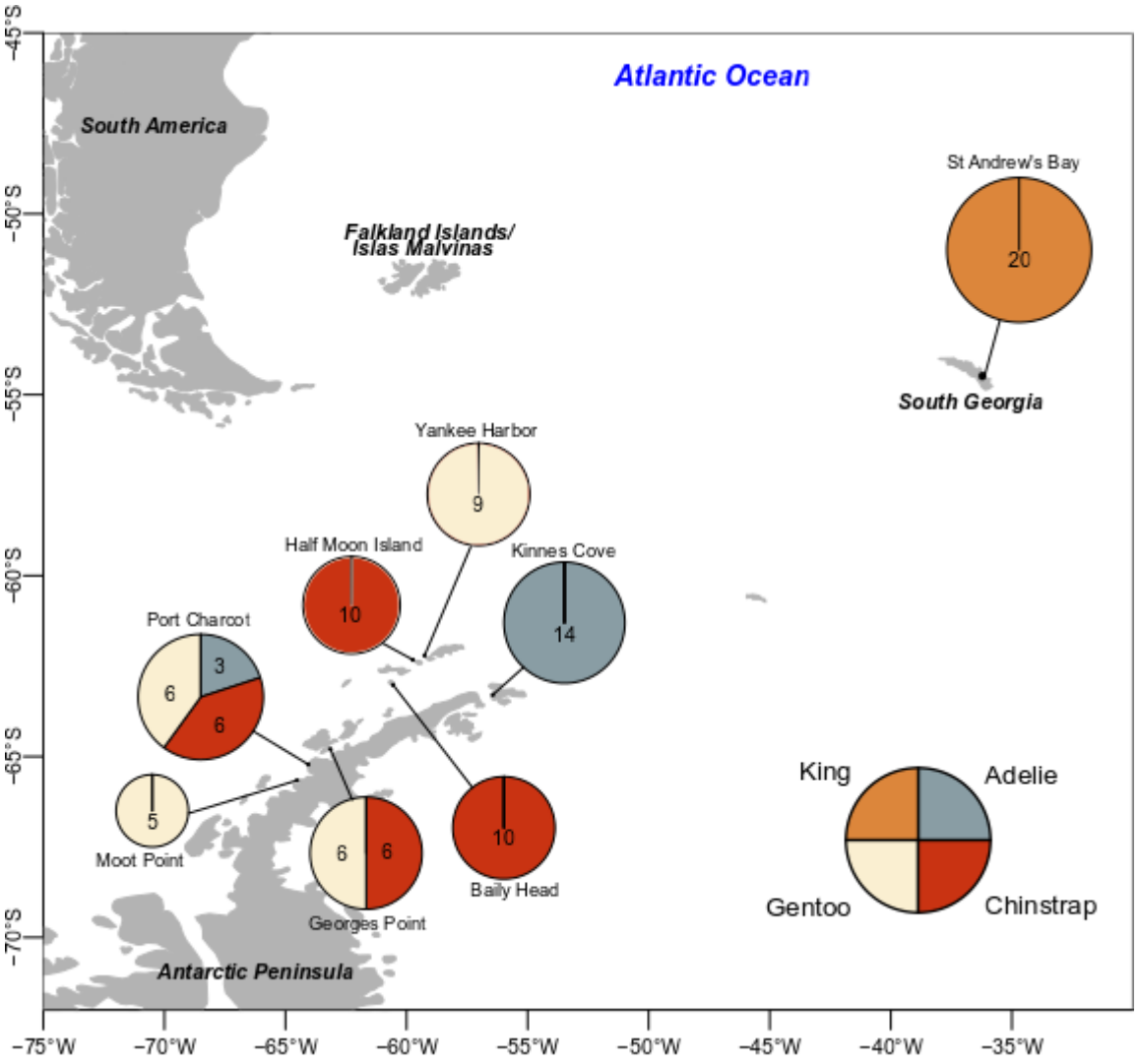


Figure 1: Sampling sites of the four-penguin species based on their breeding colonies in South Georgia and the Antarctic Peninsula. Numbers in the pie chart indicate the number of individual samples per species.

Sampling sites are shown in **Figure 1** and details are found in **Supplementary Table S1** and **Supplementary Data 1**. All sampling carried out in the Antarctic Treaty Area was performed in accordance with United Kingdom Home Office guidelines, under permits granted by the United Kingdom Foreign and Commonwealth Office (permits S3 04/2014 and S7 03/2014) and in South Georgia under permits granted by the Government of South Georgia and the South Sandwich Islands (RAP-2015-018), with ethical approval from the University of Oxford Animal Welfare and Ethical Review Board. All samples were transported in accordance with applicable export permits and United Kingdom Department for Environment, Food and Rural Affairs import permits.

Viral extraction and high-throughput sequencing analysis

Viral DNA was extracted from 200 µl of the swab suspension after vortexing, using the High Pure Viral Nucleic Acid Kit (Roche Diagnostics, Indianapolis, IN, USA). Following viral DNA extraction, circular DNA was preferentially amplified by rolling circle amplification (RCA) using the TempliPhi™ 100 Amplification Kit (GE Healthcare, Chicago, IL, USA). An aliquot of the resulting RCA-amplified DNA (5 µl) was pooled based on animal species per site and used to generate Illumina sequencing libraries using the Nextera DNA Flex Library Prep Kit (Illumina Inc, San Diego, CA, USA). The libraries were sequenced on an Illumina 4000 sequencer (2 × 100 bp library) and the resulting paired-end raw reads were trimmed using Trimmomatic (Bolger *et al.* 2014). The trimmed reads were *de novo* assembled using metaSPAdes v 3.12.0 (Bankevich *et al.* 2012). All assembled contigs > 500 nt were analyzed against a viral RefSeq (O'Leary *et al.* 2016) protein database using BLASTx (Altschul *et al.* 1990).

Recovery and sequencing of viral genomes and viral-like circular elements

Twenty-eight viral-like contigs were identified and these were used to design abutting primer pairs (**Supplementary Table S2**) for the recovery of complete viral genomes by PCR. The specific primers pairs, together with KAPA HiFi HotStart DNA Polymerase (Kapa Biosystems, Wilmington, MA, USA) were used to screen and amplify the virus genomes and viral-like elements

from each sample, using the thermal cycling protocol recommended by the manufacturer with an annealing temperature of 60°C and 0.5 µl of the RCA product. The resulting amplicons were resolved on a 0.7% agarose gel, excised, purified and cloned into pJET1.2 plasmid (ThermoFisher, Waltham, MA, USA). The recombinant plasmids were Sanger-sequenced by primer walking at Macrogen Inc. (Seoul, Korea) and contigs assembled using Geneious Prime (Kearse *et al.* 2012).

Bioinformatic analyses of recovered viral genomes and viral-like elements

The open reading frames (ORFs) in the viral genomes and viral-like elements were identified using ORFfinder (https://www.ncbi.nlm.nih.gov/orffinder/) and annotated using Geneious Prime (Kearse *et al.* 2012). All pairwise identities (genome-wide and protein) were determined using SDTv1.2 (Muhire *et al.* 2014).

A dataset of the Replication-associated proteins (Rep) encoded by the members of the recently established phylum *Cressdnaviricota* (Krupovic *et al.* 2020) i.e., alphasatellites, circoviruses, geminiviruses, genomoviruses, nanoviruses, redondoviruses and smacoviruses was assembled with sequences available in GenBank. This dataset was clustered with a 0.9 sequence identity cut-off using CD-HIT (Huang *et al.* 2010) and a representative from each cluster was used to build a Rep-CRESS dataset together with all the Rep sequences of unclassified CRESS DNA viruses and plasmids analyzed in Kazlauskas *et al.* (2019). The Rep-CRESS dataset together with 54 Reps from this study was used to generate a sequence similarity network (SSN) analysis using EST-EFI (Gerlt 2017; Zallot *et al.* 2018) with a minimum similarity score of 60. The SSN was visualized in an organic layout with Cytoscape V3.7.1 (Shannon *et al.* 2003).

Similarly, a capsid protein (CP) dataset of the CPs encoded by circoviruses, geminiviruses, genomoviruses, nanoviruses, redondoviruses and smacoviruses was assembled and clustered with a 0.9 sequence identity cut-off using CD-HIT (Huang *et al.* 2010). Representative sequences from each cluster were assembled together with the RNA virus CP sequences (since they have similarities to some CRESS DNA virus CPs) of tombusviruses (n=54; derived from RefSeq),

albetoviruses (n=3; derived from RefSeq), all CPs of unclassified CRESS DNA viruses, and 84 CPs from this study to build a CP-CRESS-alberto-tombus dataset. The CP-CRESS-alberto-tombus dataset was used to generate a sequence similarity network (SSN) analysis using EST-EFI (Gerlt 2017; Zallot *et al.* 2018) with a minimum similarity score of 10. The SSN was visualized in an organic layout with Cytoscape V3.7.1 (Shannon *et al.* 2003).

Sequences within a network cluster that included sequences from this study were extracted and aligned using MAFFT (Katoh & Standley 2013, 2016). The alignments were trimmed using TrimAl (Capella-Gutierrez *et al.* 2009) with gappy option and the resulting alignments were used to determine the best amino acid substitution model using ProtTest (Darriba *et al.* 2011) and infer Maximum likelihood (ML) phylogenetic trees with aLRT branch support. All trees were mid-point rooted and branches with <0.8 aLRT support were collapsed using TreeGraph2 (Stover & Muller 2010). The trees were visualized using iTOL v4 (Letunic & Bork 2019).

Species-level sequence alignments of Antarctic viruses (AntVs) and Antarctic circular molecules (AntCMs) identified in this study were used to identify any evidence of recombination using RDP4 v.4.97 (Martin, Murrell, Golden, Khoosal, & Muhire, 2015) with default settings. Only events that were detected by three or more recombination detection methods implemented in RDP v4.97 with p-values < 0.05 were accepted as credible.

Results and Discussion

Identification of viruses and viral-like molecules from cloacal swabs

We identified 28 contigs >500nt from the *de novo* assembled sequences that had similarity to viral sequences. Using abutting primers designed from the *de novo* assembled contigs, we amplified, cloned, and Sanger-sequenced 40 viral genomes (with detectable Rep and CP-coding ORFs) and 57 viral-like elements. The 40 viral genomes (2322-2729 nt; GenBank accession numbers MT196222-MT196223, MT196247-MT196250, MT196252-MT196253, MT196261, MT196279, MT196289-MT196318) do not fall within any known viral families, and

based on their Rep sequences would be part of the expanding unclassified circular Rep-encoding single stranded (CRESS) DNA viruses within the recently established phylum *Cressdnaviricota* (Krupovic *et al.* 2020). The phylum *Cressdnaviricota* includes seven viral families (*Bacilladnaviridae*, *Circoviridae*, *Geminiviridae*, *Genomoviridae*, *Nanoviridae*, *Redondoviridae* and *Smacoviridae*), the Reps of the satellite nucleic acids in the family *Alphasatellitidae*, and a suite of viral groups that have been loosely labelled CRESS DNA viruses. The CRESS DNA viruses that cannot be classified into the eight families above, have been identified primarily through viral metagenomic approaches from a variety of samples including animal tissues, as highlighted in the recent report by Tisza *et al.* (2020), fecal samples (Blinkova *et al.* 2010; Ng *et al.* 2015; Steel *et al.* 2016) and environmental samples including plant leaves (Kraberger *et al.* 2015b), soil (Reavy *et al.* 2015), wastewater (Kraberger *et al.* 2015a; Pearson *et al.* 2016; Phan *et al.* 2015), seawater (Labonte & Suttle 2013), freshwater (Dayaram *et al.* 2016) and sea spray (Geoghegan *et al.* 2018). In Antarctica, CRESS DNA viruses have been identified from cryoconite samples from the dry valley (Sommers *et al.* 2019), Adélie Penguin fecal samples (Levy *et al.* 2020; Morandini *et al.* 2019), lake samples (Lopez-Bueno *et al.* 2009) and a pond sample (Zawar-Reza *et al.* 2014).

Of 57 circular viral-like elements identified in this study, one viral-like element (1982 nt; MT196283) encodes a CP and a partial Rep (missing the N terminus which is the endonuclease domain) and two additional unknown proteins, while eight (1243-2181 nt; MT196267-MT196271, MT196283, MT196286-MT196287) encode only a Rep. Six of the circular viral-like elements (2646-4800nt; MT196226, MT196234-MT196236, MT196251, MT196288) encode Reps more similar to plasmid Reps. Finally,43 viral-like elements (1759-2274 nt; MT196224-MT196225, MT196227-MT196233, MT196235-MT196254, MT19660, MT196262- MT196266, MT196272-MT196278, MT196280-MT196282, MT196284-MT196285) encode a CP and an unknown protein **(Figure 2)**.

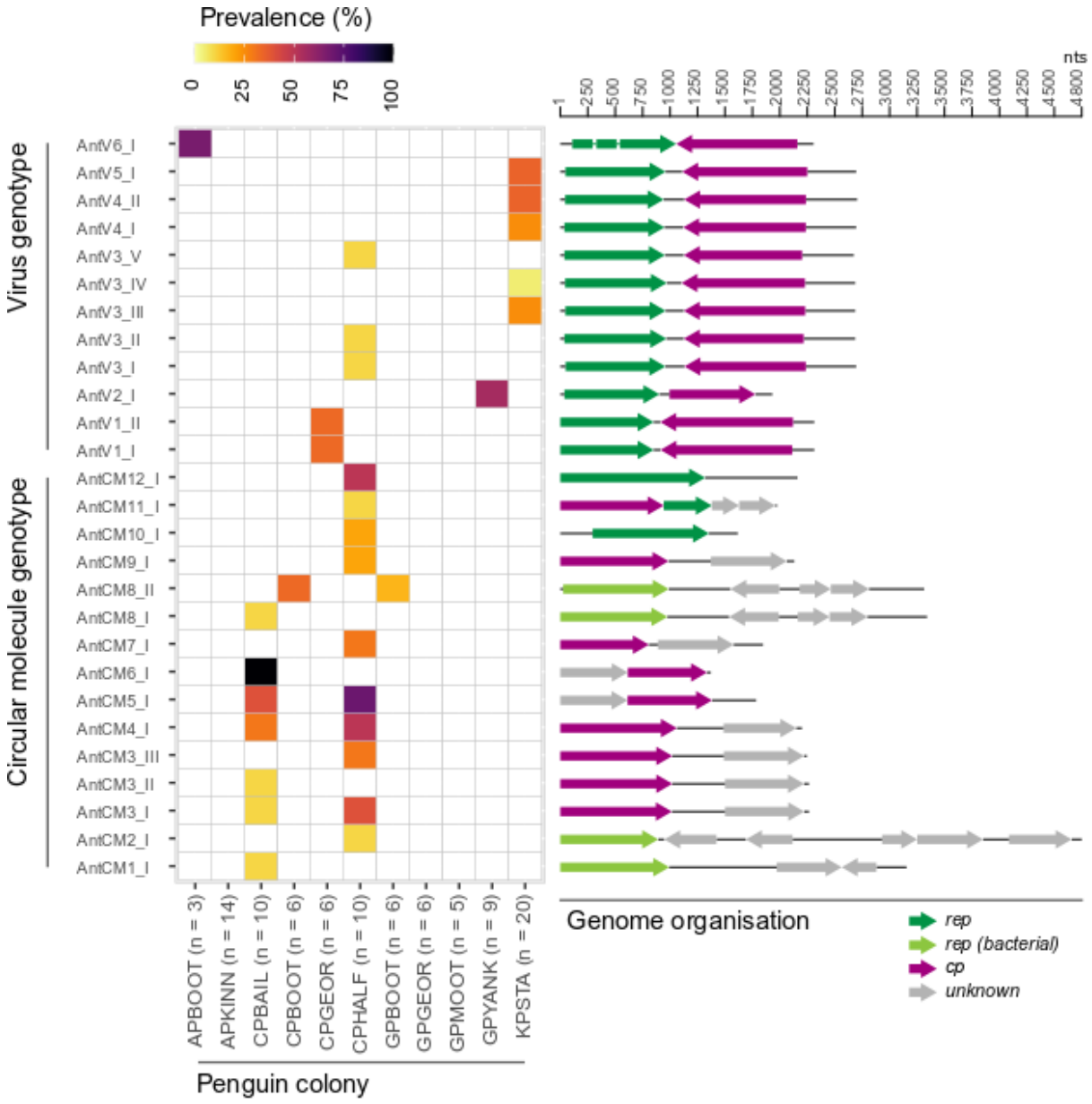


Figure 2: Heat map illustrating the prevalence of virus and circular molecule genotypes with the linearized genome organization shown on the right. For the purpose of this study the viruses and circular molecules have been classified as unique species based on a cutoff threshold of 80% genome-wide pairwise identity and genotypes based on 98% genome-wide pairwise identity. The penguin colony key based on location is STA = St Andrews Bay (South Georgia), KINN = Kinnes Cove (Northern Antarctic Peninsula), BOOT = Booth Island (Western Antarctic Peninsula), BAIL = Baily Head (Deception Island), GEOR = Georges Point (Western Antarctic Peninsula), HALF = Half Moon Island (South Shetland Islands), MOOT = Moot Point (Western Antarctic Peninsula), YANK = Yankee Harbor (South Shetland Islands).

For the purpose of this study, we adopted an 80% genome-wide pairwise identity cutoff (similar to other CRESS DNA viruses) as a putative species demarcation cutoff and 98% identity for genotypes (**Supplementary Data 2**). Thus, 40 viral genomes represent six tentative species (labelled as Antarctic viruses, AntV 1-6) and 12 genotypes **(Table 1)**. The circular molecules (labelled as Antarctic circular molecules, AntCM 1-12) can be split into 12 species and 15 genotypes **(Table 1)**.

A preliminary BLASTn based analysis revealed that only AntCM3, -4, -5, -6, -7, -9, 10 and -12 have hits with >20% sequence coverage to viral-like sequences identified in sea water samples from Saanich Inlet (BC, Canada) (Labonte & Suttle 2013), fish and abalone tissue (Tisza *et al.* 2020). The results of this are summarized in **Table 2**. JX904537 does not have annotated ORFs thus is not included in the coding sequence analysis in this study.

Distribution of viruses and viral-like molecules across sampling sites and penguin species

Of 95 swab samples tested in this study, the 97 viral genomes and circular molecules were recovered from 45 individuals (2/17 Adélie Penguins, 24/32 Chinstrap Penguin, 6/26 Gentoo Penguins, and 13/20 King Penguins) with at least one AntV or AntCM (**Table 1**; **Figure 2**).

The 40 AntVs were recovered across the four penguin species **(Table 1)**. From the 57 AntCMs recovered, eight sequences that encode Reps were recovered only from Chinstrap Penguins while the other 49 sequences were recovered from Chinstrap Penguins and one Gentoo Penguin (**Table 1; Figure 2**).

Table 1: Table of occurrence of viral-like circular molecules (AntCM) and viruses (AntV) across sampled host species and sites. Total sample size n (Adults/Chicks) shown in header, while number of individuals in which a given genotype was detected (Adults/Chicks) shown within table. Cases where a single individual had two sequences of a single genotype are indicated with an asterisk (*). Regions: CWAP = Central Western Antarctic Peninsula, NEAP = Northeast Antarctic Peninsula, SG = South Georgia, SH = South Shetland Islands.

		Penguin Species										
		Adélie Penguin		Chinstrap Penguin				Gentoo Penguin				King Penguin
	Site, Region	Kinnes Cove, NEAP	Booth Is., CWAP	Half Moon Is., SH	Baily Head, SH	Georges Point, CWAP	Booth Is., CWAP	Yankee Harbor, SH	Georges Point, CWAP	Booth Is., CWAP	Moot Point, CWAP	St. Andrew's Bay, SG
Viral Molecule	GenBank Accession(s)	n=14 adults=7 chicks=7	n=3 adults=2 chicks=1	n=10 adults=10 chicks=0	n=10 adults=5 chicks=5	n=6 adults=3 chicks=3	n=6 adults=3 chicks=3	n=9 adults=6 chicks=3	n=6 adults=3 chicks=3	n=6 adults=3 chicks=3	n=5 adults=4 chicks=1	n=20 adults=10 chicks=10
AntCM_1_I	MT196226				1/0							
AntCM_2_I	MT196251			1/0								
AntCM_3_I	MT196224, MT196272- MT196276			4/0	1/0							
AntCM_3_II	MT196225				1/0							
AntCM_3_III	MT196275, MT196277- MT196278			3/0								
AntCM_4_I	MT196227- MT196229, MT196262- MT196266			5/0	3/0							
AntCM_5_I	MT196230- MT196233, MT196254- MT196260			7/0	4/0							
AntCM_6_I	MT196235- MT196244				5/5							
AntCM_7_I	MT196280- MT196282			3/0								
AntCM_8_I	MT196234				1/0							
AntCM_8_II	MT196245- MT196246, MT196288						2*/0			1/0		
AntCM_9_I	MT196284- MT196285			2/0								
AntCM_10_I	MT196286- MT196287			2/0								
AntCM_11_I	MT196283			1/0								
AntCM_12_I	MT196267- MT196271			5/0								
AntV_1_I	MT196247- MT196248					2/0						
AntV_1_II	MT196249- MT196250					1/1						
AntV_2_I	MT196289- MT196293							4/1				
AntV_3_I	MT196252			1/0								
AntV_3_II	MT196253 MT196261			1*/0 1*/0								
AntV_3_III	MT196294- MT196295, MT196297- MT196299											1/4
AntV_3_IV	MT196296											0/1
AntV_3_V	MT196279			1/0								
AntV_4_I	MT1962											5/0
AntV_4_II	MT196301- MT196207											1/6
AntV_5_I	MT196312- MT196218											5/2
AntV_6_I	MT196222- MT196223		1/1									

Table 2: Summary of the BLASTn result of the representative sequences of AntCMs and AntVs.

Virus / circular molecule	Top BLASTn hit accession	Source	Query Cover	E-value	Identity
AntCM1	-		-	-	-
AntCM2	-		-	-	-
AntCM3	MH648984	Seabass tissue	86%	4.00E-167	73.86%
AntCM4	JX904537	Sea water (Saanich Inlet)	48%	0	79.50%
AntCM5	MH510272	Seabass tissue	43%	4.00E-84	69.79%
AntCM6	MH616824	Abalone tissue	24%	2.00E-42	72.02%
AntCM7	MH616694	Abalone tissue	41%	3.00E-110	72.54%
AntCM8	-		-	-	-
AntCM9	JX904537	Sea water (Saanich Inlet)	87%	0.00E+00	88.53%
AntCM10	JX904559	Sea water (Saanich Inlet)	53%	1.00E-76	68.88%
AntCM11	-		-	-	-
AntCM12	MH649197	Red snapper tissue	71%	8.00E-170	70.75%
AntV1	-		-	-	-
AntV2	-		-	-	-
AntV3	-		-	-	-
AntV4	-		-	-	-
AntV5	-		-	-	-
AntV6	-		-	-	-

AntCMs (12 species/15 genotypes), were found in 19/32 individual Chinstrap Penguins sampled and in a single Gentoo Penguin (**Table 1; Figure 2**). Seven AntCM genotypes were found in multiple individuals, with 3 genotypes shared across the two Chinstrap colonies sampled in the South Shetland Islands. AntCV6 genotype I was found in 100% (n = 10) of Chinstrap Penguins shared between Chinstrap Penguin parents and their chicks at Baily Head, Deception Island. One set of AntCM genotypes with a bacterial *rep* (AntCM8 genotype I and AntCM8 genotype II) at the edge of the species threshold (78.8-79.1% genome-wide identity) were found at two sites 290 km apart (Booth Island, Western Antarctic Peninsula and Deception Island, South Shetland Islands). AntCM8 genotype II was unique in this study as being the only genotype found in two distinct host species: an adult Gentoo Penguin and an adult Chinstrap Penguin (two sequences) nesting at the same colony on Booth Island, with >99.9% shared genome-wide identity. The adult Chinstrap Penguins was recently found to also carry a penguin circovirus (Levy *et al.* 2020). Three species (each as a single genotype) of Antarctic circular molecules that encode Reps (AntCM10-12) were

identified in this sample set, all from a single Chinstrap colony in the South Shetland Islands, where two genotypes, those which only contain a *rep* gene (*rep*: 963-1335 nts), were found in multiple individuals. The other genotype, AntCM11, found in a single Chinstrap Penguin, contained a *cp* gene (960 nts), an unknown ORF, and a partial *rep* (459 nts).

Among the AntVs (six species/12 genotypes), each genotype was only detected in a single host species, with five only found in Chinstrap Penguins, one in Adélie Penguins, one in Gentoo Penguins, and five in King Penguins. Each Antarctic virus sequence was only found at a single site, though 8/12 genotypes were found in multiple individuals at those sites (**Figure 2**). AntV3 has five genotypes, two of which (genotype III and IV) were only found in King Penguins from South Georgia, sharing 82.5-94.5% genome-wide identity with the three genotypes (genotype I, II and V) found only in Chinstrap Penguins from Deception Island **(Supplementary Data 1)**, sites separated by approximately ~1700 km of the Southern Ocean (**Figure 1**, **Supplementary Table S1**). This suggests that this virus is broadly distributed across the Scotia Arc.

Cases of parent-chick viral genotype sharing were found in each Pygoscelid species. AntV1 genotype II was found in one Chinstrap Penguin adult/chick pair at Georges Point (Western Antarctic Peninsula), whereas AntV6 was shared in one Adélie Penguin adult/chick pair on Booth Island (Western Antarctic Peninsula). AntV2 was identified in five Gentoo Penguin individuals from Yankee Harbor in the South Shetland Islands, including an adult/chick pair (**Figure 2, Table 1,** and **Supplementary Data 1**). King Penguin chicks had already fledged during sampling, so parentage could not be determined in this sample set. In those cases where penguin parents and their chicks both share the same viruses future studies have to be conducted to observe if those were vertically or horizontally transmitted.

Co-occurrence of these novel AntVs and AntCMs was noted in 14/32 Chinstrap Penguins (2-6 sequences per individual), while 11/20 King Penguins carried two or more viral sequences (maximum of 3 sequences in one chick) (**Figure 2, Supplementary Table S1**). At this time, we are

unable to discern whether the presence or co-occurrence of one or more AntVs or AntCMs indicates that these infect penguins, cause decreased fitness in the hosts, or are penguin-associated through trophic transfer.

Analyses of the replication-associated protein

The hallmarks of the Rep protein are the endonuclease and helicase domains that are involved in initiation of replication in ssDNA viruses, and to some extent in bacterial plasmids (reviewed in Kazlauskas *et al.* (2019) and Rosario *et al.* (2012)). The Reps encoded by the AntVs (n=40) have the rolling circle replication (RCR) endonuclease domain and the super family 3 (SF3) helicase motifs (**Table 3**). 13 of the 14 and AntCMs also have the RCR domain and SF3 motifs. AntCM11 which has a partial Rep is missing the RCR domain and the Arg finger domain (**Table 3**).

Since we identified six AntCMs with Reps that are most closely related to those of plasmids, we included a suite of plasmids grouped by Kazlauskas *et al.* (2019) for clustering the Reps using a sequence similarity network (SSN) based approach. We have found in the past (Fontenele *et al.* 2019; Orton *et al.* 2020; Sommers *et al.* 2019) that a SSN analysis using EST-EFI (Gerlt 2017; Zallot *et al.* 2018) with a network threshold of 60 allows for reasonable family level clusters for families in the phylum *Cressdnaviricota* (Krupovic *et al.* 2020). Based on the sequence similarity network analysis of the Rep amino acid sequences (**Figure 3**), none of the AntV Reps cluster with known viral sequences. A vast majority of the AntV Reps (n=33) are part of a large cluster of Reps (n=213). The 33 AntV Reps within this cluster are distributed within 3 clades in the ML phylogenetic tree (RtREV+I+G+F amino acid substitution model; **Figure 3**). One of the clades (Clade I in Figure 3) includes Reps of AntCM10, the second clade (Clade II in **Figure 3**) includes Reps of AntV6. The Reps of AntV3, -4 and -5 are represented in Clade III (**Figure 3**) forming a well-supported clade and share >72 % amino acid identity. The Reps of AntCV3, -4 and-5 share <60% amino acid identity with other Reps in this cluster. AntCV2 Reps, which share 100% amino acid identity, cluster with a Rep of Copepod LaCopCV (JF912805) sampled in Tampa Bay (Florida, USA)

(Dunlap *et al.* 2013) sharing 46.8% amino acid identity. AntCV1 Reps cluster together: those that are genotype I share 94.4-94.7% amino acid identity with genotype II (**Figure 3**). AntCM12 Reps cluster together sharing >97.3% amino acid identity. The AntCM11 with a partial Rep is a singleton sharing 44.1% amino acid identity to a Rep of a virus (MN928918) from a cloacal swab of an Indian blue peafowl (*Pavo cristatus*).

Table 3: Summary of the RCR endonuclease motifs (Motif I, II, II) and SF3 helicase motifs (Walker A, Walker B, and Motif C) and Arg finder motif of the Reps identified in this study.

Virus / Circular molecule	Genotype	Isolate	Accession #	Motif I	Motif II	Motif III	Walker A	Walker B	Motif C	Arg finger
AntV1	I	CPGEORsw001Ad	MT196247	CVTINN	HVQM	WADQEYCFK	GPAGCGKNKLTT	LMHEY	VFTTN	AFWRRI
AntV1	I	CPGEORsw002Ad	MT196248	CVTINN	HVQM	WADQEYCFK	GPAGCGKNKLTT	LMHEY	VFTTN	AFWRRI
AntV1	II	CPGEORsw003Ad	MT196249	CVTINN	HVQM	WADQEYCFK	GPAGCGKNKLTT	LMHEY	VFTTN	AFWRRI
AntV1	II	CPGEORsw003Ch	MT196250	CVTINN	HVQM	WADQEYCFK	GPAGCGKNKLTT	LMHEY	VFTTN	AFWRRI
AntV2	I	GPYANKsw001Ad	MT196289	SITYNN	HFQC	MAAFKYCQK	GVPGSGKSYQAR	IIEEM	IVTSN	AIKRRF
AntV2	I	GPYANKsw001Ch	MT196290	SITYNN	HFQC	MAAFKYCQK	GVPGSGKSYQAR	IIEEM	IVTSN	AIKRRF
AntV2	I	GPYANKsw002Ad	MT196291	SITYNN	HFQC	MAAFKYCQK	GVPGSGKSYQAR	IIEEM	IVTSN	AIKRRF
AntV2	I	GPYANKsw003Ad	MT196292	SITYNN	HFQC	MAAFKYCQK	GVPGSGKSYQAR	IIEEM	IVTSN	AIKRRF
AntV2	I	GPYANKsw004Ad	MT196293	SITYNN	HFQC	MAAFKYCQK	GVPGSGKSYQAR	IIEEM	IVTSN	AIKRRF
AntV3	I	CPHALFsw004Ad	MT196252	CFTLNN	HWQG	QQAIDYCKK	GETGTGKSRKLW	ALEEW	IVTSN	PLKRRF
AntV3	II	CPHALFsw005Ad	MT196253	CFTLNN	HWQG	TQAIDYCKK	GETGTGKSRKLW	ALEEW	IVTSN	PLKRRF
AntV3	II	CPHALFsw005Ad	MT196261	CFTLNN	HWQG	TQAIDYCKK	GETGTGKSRKLW	ALEEW	IVTSN	PLKRRF
AntV3	III	KPSTAsw010Ad	MT196294	CFTLNN	HWQG	QQAIDYCKK	GETGTGKSRKLW	ALEEW	IVTSN	PLKRRF
AntV3	III	KPSTAsw014Ch	MT196295	CFTLNN	HWQG	QQAIDYCKK	GETGTGKSRKLW	ALEEW	IVTSN	PLKRRF
AntV3	III	KPSTAsw016Ch	MT196297	CFTLNN	HWQG	QQAIDYCKK	GETGTGKSRKLW	ALEEW	IVTSN	PLKRRF
AntV3	III	KPSTAsw018Ch	MT196298	CFTLNN	HWQG	QQAIDYCKK	GETGTGKSRKLW	ALEEW	IVTSN	PLKRRF
AntV3	III	KPSTAsw019Ch	MT196299	CFTLNN	HWQG	QQAIDYCKK	GETGTGKSRKLW	ALEEW	IVTSN	PLKRRF
AntV3	IV	KPSTAsw015Ch	MT196296	CFTLNN	HWQG	QQAIDYCKK	GETGTGKSRKLW	ALEEW	IVTSN	PLKRRF
AntV3	V	CPHALFsw003Ad	MT196279	VFTINN	HWQG	KQAIDYCKK	GATGTGKSRKLW	ALEEW	IVTSN	PLKRRF
AntV4	I	KPSTAsw001Ad	MT196300	CFTINN	HWQG	QQAIDYCKK	GGTGTGKSRKLW	AIEEW	IVTSN	PLKRRF
AntV4	I	KPSTAsw002Ad	MT196308	CFTINN	HWQG	QQAIDYCKK	GGTGTGKSRKLW	AIEEW	IVTSN	PLKRRF
AntV4	I	KPSTAsw008Ad	MT196309	CFTINN	HWQG	QQAIDYCKK	GGTGTGKSRKLW	AIEEW	IVTSN	PLKRRF
AntV4	I	KPSTAsw009Ad	MT196310	CFTINN	HWQG	QQAIDYCKK	GGTGTGKSRKLW	AIEEW	IVTSN	PLKRRF
AntV4	I	KPSTAsw010Ad	MT196311	CFTINN	HWQG	QQAIDYCKK	GGTGTGKSRKLW	AIEEW	IVTSN	PLKRRF
AntV4	II	KPSTAsw004Ad	MT196301	CFTINN	HWQG	QQAIDYCKK	GETGTGKSRKLW	AIEEW	IVTSN	PLKRRF
AntV4	II	KPSTAsw014Ch	MT196302	CFTINN	HWQG	QQAIDYCKK	GETGTGKSRKLW	AIEEW	IVTSN	PLKRRF
AntV4	II	KPSTAsw015Ch	MT196303	CFTINN	HWQG	QQAIDYCKK	GETGTGKSRKLW	AIEEW	IVTSN	PLKRRF
AntV4	II	KPSTAsw016Ch	MT196304	CFTINN	HWQG	QQAIDYCKK	GETGTGKSRKLW	AIEEW	IVTSN	PLKRRF
AntV4	II	KPSTAsw017Ch	MT196305	CFTINN	HWQG	QQAIDYCKK	GETGTGKSRKLW	AIEEW	IVTSN	PLKRRF
AntV4	II	KPSTAsw018Ch	MT196306	CFTINN	HWQG	QQAIDYCKK	GETGTGKSRKLW	AIEEW	IVTSN	PLKRRF
AntV4	II	KPSTAsw019Ch	MT196307	CFTINN	HWQG	QQAIDYCKK	GETGTGKSRKLW	AIEEW	IVTSN	PLKRRF
AntV5	I	KPSTAsw001Ad	MT196312	CFTLNN	HWQG	KQAIDYCKK	GATGTGKSQKLW	AIEEW	IVTSN	PLKRRF
AntV5	I	KPSTAsw002Ad	MT196313	CFTLNN	HWQG	KQAIDYCKK	GATGTGKSQKLW	AIEEW	IVTSN	PLKRRF
AntV5	I	KPSTAsw004Ad	MT196314	CFTLNN	HWQG	KQAIDYCKK	GATGTGKSQKLW	AIEEW	IVTSN	PLKRRF
AntV5	I	KPSTAsw008Ad	MT196315	CFTLNN	HWQG	KQAIDYCKK	GATGTGKSQKLW	AIEEW	IVTSN	PLKRRF
AntV5	I	KPSTAsw009Ad	MT196316	CFTLNN	HWQG	KQAIDYCKK	GATGTGKSQKLW	AIEEW	IVTSN	PLKRRF
AntV5	I	KPSTAsw018Ch	MT196317	CFTLNN	HWQG	KQAIDYCKK	GATGTGKSQKLW	AIEEW	IVTSN	PLKRRF
AntV5	I	KPSTAsw020Ch	MT196318	CFTLNN	HWQG	KQAIDYCKK	GATGTGKSQKLW	AIEEW	IVTSN	PLKRRF
AntV6	I	APBOOTsw001Ad	MT196222	VFTWNN	HLQG	QECDKYCRK	GESGCGKTRAVH	LLDDI	IVTSQ	ALLRRF
AntV6	I	APBOOTsw001Ch	MT196223	VFTWNN	HLQG	QECDKYCRK	GESGCGKTRAVH	LLDDI	IVTSQ	ALLRRF
AntCM1	I	CPBAILsw002Ad	MT196226	IVTLPA	HWQI	QQAYEYVTK	GPTRVGKSRNVV	VLDEF	WVVSN	AFEARF
AntCM2	I	CPHALFsw001Ad	MT196251	LLTLPA	HWQL	QQVYDYVTK	GPTGVGKTRHIY	VLDEF	WIVSN	AFRRRL
AntCM8	I	CPBAILsw001Ad	MT196234	FLTISA	HYHA	AKSIKYLKK	GAAGGGKSTLAR	VYDEF	IIVSN	AFKRRI
AntCM8	II	CPBOOTsw002Ad	MT196245	FITISG	HYHI	HDVLKYIQK	GVPGGGKSTLAR	IFDEF	IIASN	AFKRRI
AntCM8	II	CPBOOTsw002Ad	MT196246	FITISG	HYHI	HDVLKYIQK	GVPGGGKSTLAR	IFDEF	IIASN	AFKRRI
AntCM8	II	GPBOOTsw003Ad	MT196288	FITISG	HYHI	HDVLKYIQK	GVPGGGKSTLAR	IFDEF	IIASN	AFKRRI
AntCM10	I	CPHALFsw008Ad	MT196286	MFTINN	HYQG	KQAVDYVSK	GKPGTGKSHQAR	ILDDL	VVTSN	AINRRF
AntCM10	I	CPHALFsw010Ad	MT196287	MFTINN	HYQG	KQAVDYVSK	GKPGTGKSHQVR	ILDDL	VVTSN	AINRRF
AntCM11	I	CPHALFsw005Ad	MT196283	-	-	-	DYEGNKGKSWLSR	IIDVP	MVITN	-
AntCM12	I	CPHALFsw003Ad	MT196267	FITIPQ	HYHI	NAVITYIKK	GKPNSGKSFMFN	CWDIP	IVFAN	LPQGRT
AntCM12	I	CPHALFsw004Ad	MT196268	FITIPQ	HYHI	PAVITYIKK	GRPNSGKSFMFN	CWDIP	IVFAN	LPKGRT
AntCM12	I	CPHALFsw008Ad	MT196269	FITIPQ	HYHI	PAVITYIKK	GRPNSGKSFMFN	CWDIP	IVFAN	LPKGRT
AntCM12	I	CPHALFsw009Ad	MT196270	FITIPQ	HYHI	PAVITYIKK	GRPNSGKSFMFN	CWDIP	IVFAN	LPKGRT
AntCM12	I	CPHALFsw010Ad	MT196271	FITIPQ	HYHI	NAVITYIKK	GKPNSGKSFMFN	CWDIP	IVFAN	LPKGRT

The plasmid-like Reps of AntCM1, -2 and -8 do not cluster with the plasmids (pCRESS) grouped by Kazlauskas *et al.* (2019) and form two clusters, one of two Reps (AntCM1 and -2) sharing 44% pairwise amino acid identity and one of four Reps with AntCM8 genotype I Reps sharing 100% identity and collectively they share 59.6% identity with that of AntCM8 genotype II **(Figure 3)**.

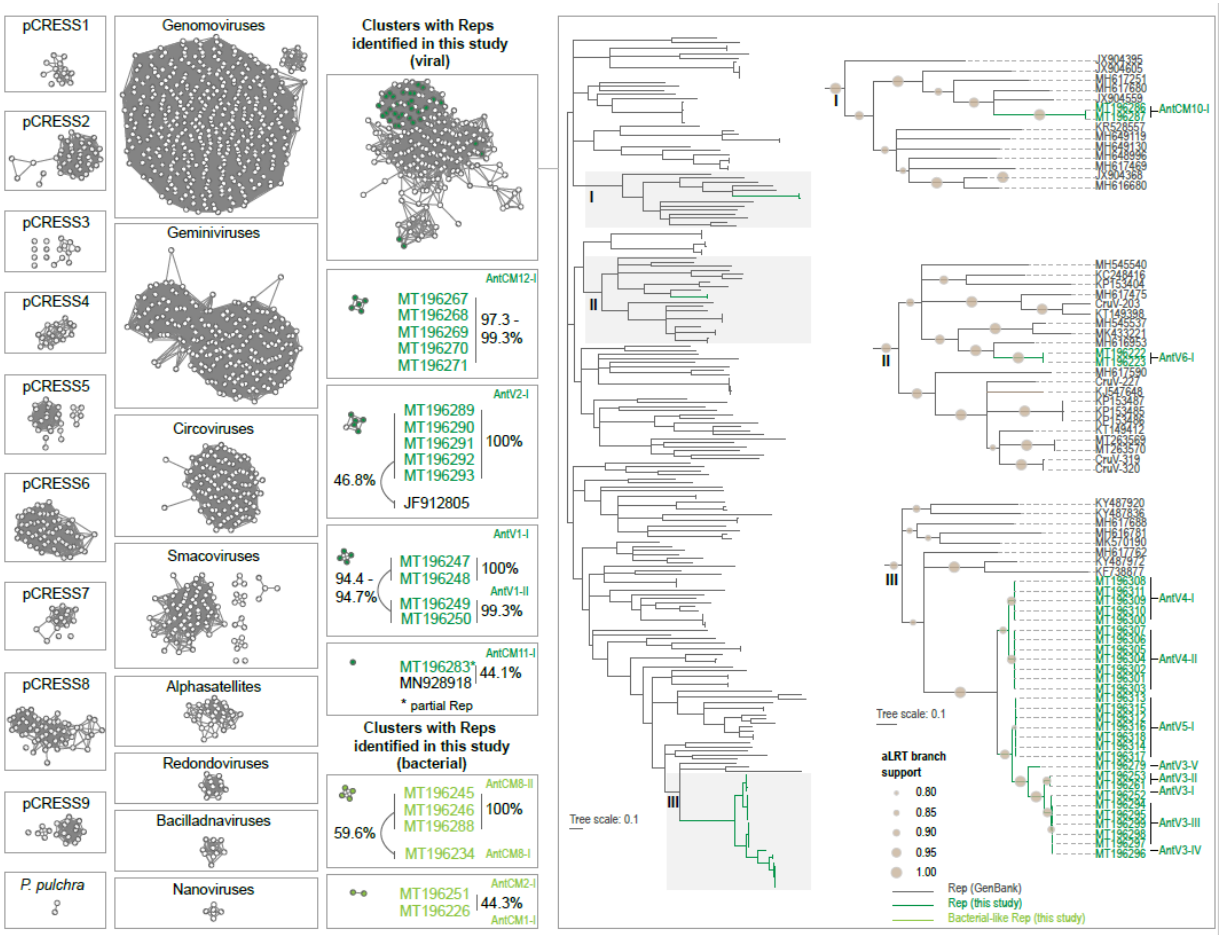


Figure 3: Sequence similarity network (SSN) analysis and Maximum Likelihood (ML) phylogenetic tree of the Rep amino acid sequences identified in this study compared with those of known viruses in classified families and plasmids identified by Kazlauskas *et al.* (2019). The ML phylogenetic tree based on the SSN is mid-point rooted and branches with aLRT support <0.8 have been collapsed. The Reps identified in this study are highlighted in green and relevant clades of the tree are shown in expanded view. Sequences with taxa names CruV-203, -227, -319 and -320 are from metagenomic derived sequences reported in de la Higuera *et al.* (2020) and do not have accession #s assigned. The level of branch support (>0.8) is depicted as different sized shaded circles.

Analyses of the capsid protein

In general, for viruses in the phylum *Cressdnaviricota*, the CPs are more diverse than the Reps. Furthermore, the CPs of geminiviruses share homology to those of albetoviruses (linear ssRNA(+) genomes) (Zhang *et al.* 2001) and a group of unclassified cressdnavirids have CPs that are homologous to those of tombusviruses (linear ssRNA(+) genomes) (de la Higuera *et al.* 2020; Diemer & Stedman 2012; Krupovic *et al.* 2015; Roux *et al.* 2013).

The SSN analysis of the CPs reveals that AntCM11 CP clusters with those of smacoviruses and smacovirus-like CPs (n=175) sharing <33 % identity (**Figure 4**). In a ML phylogenetic tree (amino acid substitution model LG+I+G+F), AntCM11 CP is part of a well-supported clade (Clade I, **Figure 4**). The remaining 83 CPs identified in this study cluster with none of the classified viral groups.

The CPs of AntCM3, -4 and -9 are part of a cluster of 57 CPs sharing 46-57% amino acid identity. In the ML phylogenetic tree (amino acid substitution model rtRev+I+G+F) these CPs are part of a well-supported clade (clade II; **Figure 4**). The CP of AntCM3 is most closely related (sharing 76-79% amino acid identity) to that of MH648984 identified in a seabass tissue (Tisza *et al.* 2020). AntCM4 CP shares ~49% amino acid identity with the CP of MH649117 from a seabass tissue, whereas AntCM9 shares 69-77% amino acid identity with MH648943 (seabass tissue), MH616644 (red snapper tissue) and KX246262 (sea cucumber sample) (Jackson *et al.* 2016; Tisza *et al.* 2020).

The CPs of AntV3, -4, -5 and -6 are part of a cluster of 77 CPs but the CPs of AntV6 (Clade III) are part of a distinct clade in the ML phylogenetic tree (amino acid substitution model LG+I+G+F) that is different from the one of AntV3, -4 and -5 (**Figure 4**). AntCV6 CPs share ~40% amino acid identity with the CP of MH616953 from rainbow trout tissue (Tisza *et al.* 2020). The CPs of AntV3, -4, and -5 share >61% amino acid identity and less that 40% identity with other CPs within the cluster.

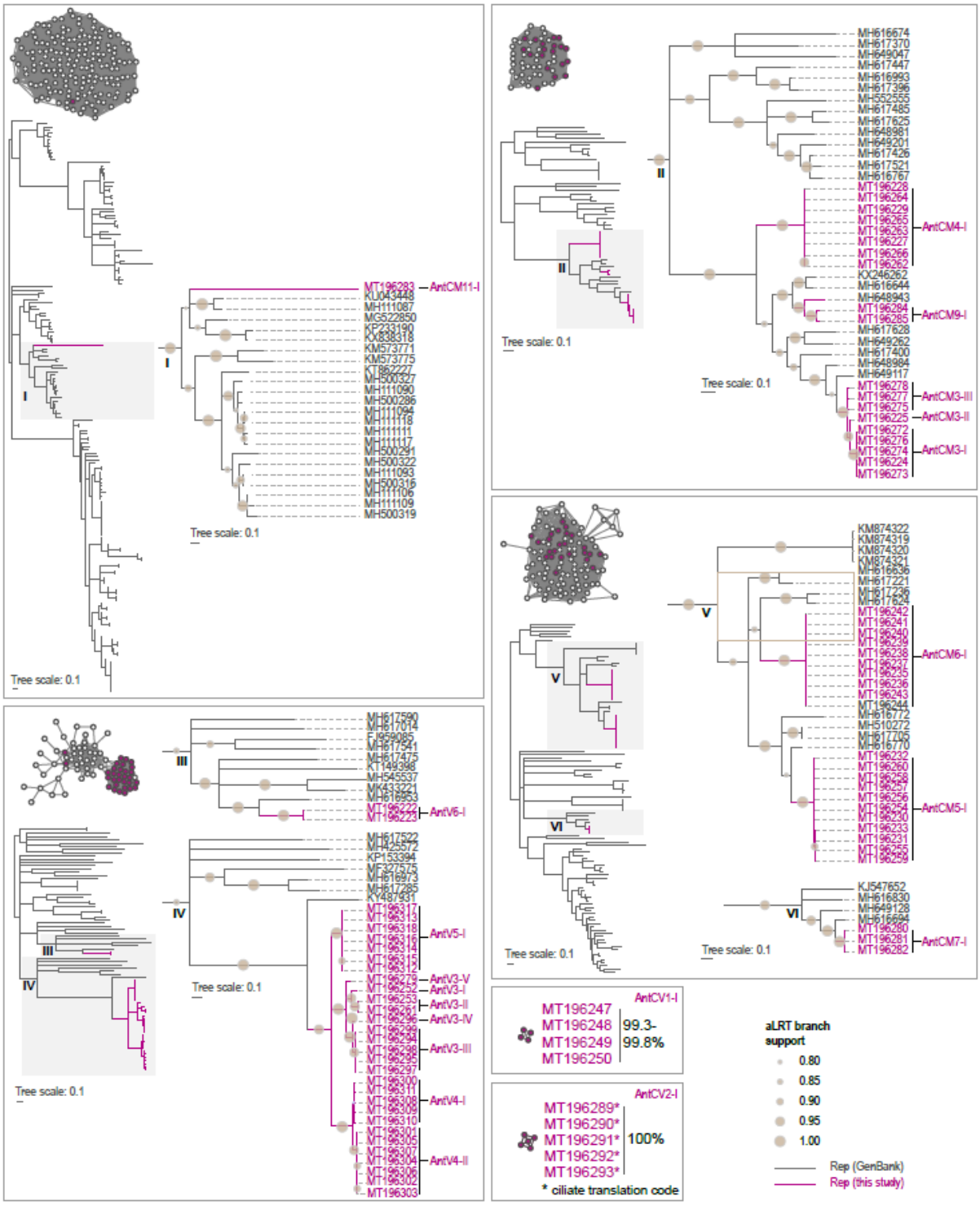


Figure 4: Sequence similarity network (SSN) analysis and Maximum Likelihood (ML) phylogenetic tree of the CP amino acid sequences that fall within specific networks that encompass sequences identified in this study (highlighted in purple). The ML phylogenetic trees based on the SSNs are mid-point rooted and branches with aLRT support <0.8 have been collapsed. The level of branch support (>0.8) is depicted as different sized shaded circles.

The CPs of AntCM5, -6 and -7 are part of the same cluster composed of 108 CPs and form two distinct clades in the ML phylogenetic tree (amino acid substitution model LG+I+G+F) with AntCM5 and -6 CPs being part of the same clade (clade V; **Figure 4**) sharing ~40% amino acid identity. AntCM5 CP share the highest amino acid identity (~73%) with the CPs of MH510272 and MH617705 both from seabass tissue (Tisza *et al.* 2020). On the other hand, the CPs of AntCM6 share the highest amino acid identity (~46%) with the CPs of MH617236 and MH617624 identified in abalone and red snapper tissue, respectively (Tisza *et al.* 2020). When compared to other CPs in this cluster, CPs of AntCM7 are part of a smaller clade (clade VI; **Figure 4**) sharing highest amino acid identity 77.6% with the CP of MH616694 and MH617624 of a viral sequence identified in abalone tissue (Tisza *et al.* 2020).

AntV1 and -2 CPs form distinct clusters each sharing 99.3-99.8 and 100% pairwise identity, respectively (**Figure 4**). The CPs of AntV2 (from translation using the ciliate translation table) share highest identity (35%) with CP-like element MK012509 from crucian carp tissue whereas those of AntCV1 share ~27% with a CP-like element of MH648877 from seabass tissue (Tisza *et al.* 2020). Although the CPs from this study have a high degree of sequence diversity it is interesting to note they are more closely related to other CPs of viruses identified in the marine ecosystem in which penguins play important roles.

Recombination analysis

The AntVs and AntCMs were aligned using MAFFT (Katoh & Standley 2016) at a species level and these were used to identify evidence of recombination using RDP4 v.4.97 (Martin, Murrell, Golden, Khoosal, & Muhire, 2015). We identified evidence of recombination in genomes of AntV3 and -4, and AntCM3 (**Figure 5**). For AntV3, two recombination events were identified, one spanning ~46% of the genome, with genotype I being the recombinant, with the recombinant region derived from genotypes III and IV. The AntV3 genotype III has a 198 nt recombinant region from the non-coding region of an unknown / unsampled viral sequence. Five variants of AntV4

genotype II have a ~290 nt recombinant region in the non-coding region derived from AntV4 genotype I. AntCM3 genotype II has two recombinant regions, both derived from genotype III and in total they account for 46% of the sequence. The second smaller recombinant region of ~180 nt is found in two AntCM3 genotype I variants as well (**Figure 5**). Evidence of recombination between these viruses and circular molecules indicates host commonality.

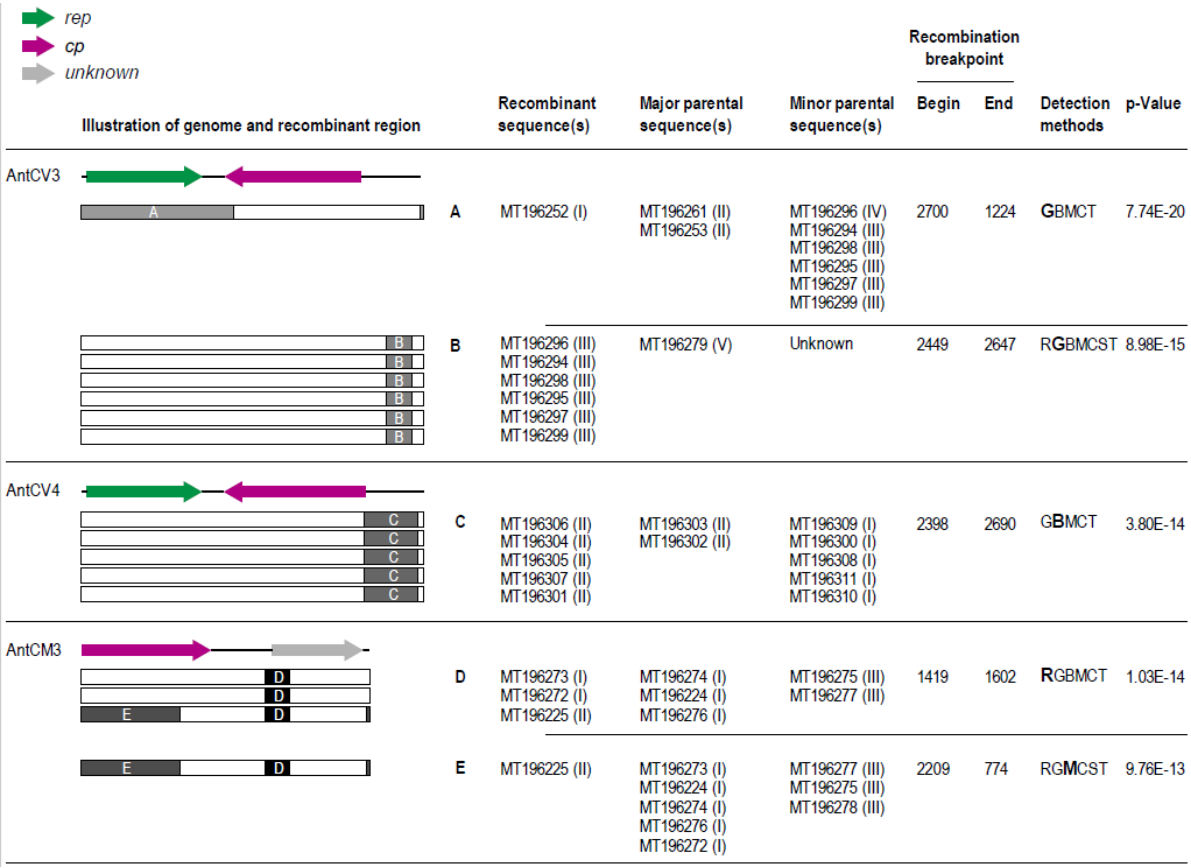


Illustration of genome and recombinant region		Recombinant sequence(s)	Major parental sequence(s)	Minor parental sequence(s)	Recombination breakpoint Begin	Recombination breakpoint End	Detection methods	p-Value
	A	MT196252 (I)	MT196261 (II) MT196253 (II)	MT196296 (IV) MT196294 (III) MT196298 (III) MT196295 (III) MT196297 (III) MT196299 (III)	2700	1224	**G**BMCT	7.74E-20
	B	MT196296 (III) MT196294 (III) MT196298 (III) MT196295 (III) MT196297 (III) MT196299 (III)	MT196279 (V)	Unknown	2449	2647	R**G**BMCST	8.98E-15
	C	MT196306 (II) MT196304 (II) MT196305 (II) MT196307 (II) MT196301 (II)	MT196303 (II) MT196302 (II)	MT196309 (I) MT196300 (I) MT196308 (I) MT196311 (I) MT196310 (I)	2398	2690	G**B**MCT	3.80E-14
	D	MT196273 (I) MT196272 (I) MT196225 (II)	MT196274 (I) MT196224 (I) MT196276 (I)	MT196275 (III) MT196277 (III)	1419	1602	R**G**BMCT	1.03E-14
	E	MT196225 (II)	MT196273 (I) MT196224 (I) MT196274 (I) MT196276 (I) MT196272 (I)	MT196277 (III) MT196275 (III) MT196278 (III)	2209	774	RG**M**CST	9.76E-13

Figure 5: Summary of the recombination detected by RDP4. The methods used to detect recombination are RDP (R) GENCONV (G), BOOTSCAN (B), MAXCHI (M), CHIMERA (C), SISCAN (S) and 3SEQ (T). The method with the highest p-value for each recombination event is bolded. For each genome, a graphic on the left indicates the recombination event with its breakpoint location within the genome. Roman numerals in parenthesis after accession numbers indicate genotype.

Concluding remarks

The overall knowledge of penguin-associated viruses is limited, and thus it is not surprising that penguin disease understanding also remains limited to observations of mass mortality events, case reports, and veterinary knowledge obtained from captive animals, reviewed in (Barbosa & Palacios 2009; Clarke & Kerry 2000; Grimaldi *et al.* 2015; Smeele *et al.* 2018; Woods *et al.* 2009). Advances in high-throughput sequencing have shown that penguins are associated with a number of (pathogenic or non-pathogenic) viruses not previously detected through serological means (Lee *et al.* 2014; Lee *et al.* 2016; Morandini *et al.* 2019; Neira *et al.* 2017; Olivares *et al.* 2019; Van Doorslaer *et al.* 2017; Varsani *et al.* 2014; Varsani *et al.* 2015; Wille *et al.* 2019). Studies such as these to directly detect nucleic acid presence in animal hosts add to our baseline knowledge, which still remains constrained by sampling effort.

In this study, the viruses and circular molecules that we identify are novel and thus we are unable to distinguish whether penguins serve as the hosts for replication of these viruses, as they may enter the penguin gastrointestinal tract through prey consumption, indirectly through the consumption of phytoplankton or zooplankton by penguin prey, or through ingestion of ocean water. Based on the CP network analysis, it does appear that some of the CPs identified in this study have homologues in sequences derived from various marine fish. Further investigation into viral presence in other mesopredators, such as flying seabirds and pinnipeds, along with cetaceans, may help elucidate whether relationships exist among these molecules. Coupled with long-term monitoring studies of individual animals and sites in the rapidly changing Antarctic, detection of potential disease-causing agents will inform conservation management and biosecurity.

Acknowledgements

We thank Quark Expeditions for logistical and financial support provided to H.L., T.H., A.D. and C.E.B. The molecular work described in this study was supported by the Center of

Evolution and Medicine Venture Fund (Center of Evolution and Medicine, Arizona State University, U.S.A.) grant awarded to A.V. H.L.'s work was supported by a University of Oxford Clarendon Fund Scholarship.

Author Contributions

HL, AV, AD, TH and ALS developed and designed the study. TH obtained permits and secured logistics. HL, CEB, and TH collected all samples. HL extracted viral DNA. AD and ALS managed the collaboration between the University of Oxford and Arizona State University, where RSF, CH, CZ, SK, and KS screened all samples, performed lab work, and cloned viruses and circular molecules. AV performed all bioinformatics, annotations, network analyses, maximum likelihood trees and recombination analyses (Figures 2b-5). HL analysed prevalence data (Figures 1-2a). HL and AV drafted the manuscript and generated figures. All authors participated in discussion of results, reviewing drafts and developed the final manuscript.

Ethics Statement

Animal handling protocols followed Scientific Committee on Antarctic Research Code of Conduct on the use of Animals for Scientific Purposes in Antarctica. All sampling in the Antarctic Treaty Area was carried out in accordance with U.K. Home Office guidelines, under permits granted by the U.K. Foreign and Commonwealth Office (permits S3 34/2015 & S7/S9 35-2015) and in South Georgia under permits granted by the Government of South Georgia and the South Sandwich Islands (RAP-2015-018) to T.H., with ethical approval from the University of Oxford Animal Welfare and Ethical Review Board.

Conflict of Interest Statement

The authors declare no conflict of interest.

Data Availability Statement

Sequences for the viruses and circular molecules in this study are deposited in GenBank under accession numbers MT196222-196318.

References

Adriaenssens EM, Kramer R, Van Goethem MW*, et al.* (2017) Environmental drivers of viral community composition in Antarctic soils identified by viromics. *Microbiome* **5**, 83.

Aguirre de Carcer D, Lopez-Bueno A, Alonso-Lobo JM, Quesada A, Alcami A (2016) Metagenomic analysis of lacustrine viral diversity along a latitudinal transect of the Antarctic Peninsula. *FEMS Microbiol Ecol* **92**, fiw074.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* **215**, 403-410.

Austin FJ, Webster RG (1993) Evidence of ortho- and paramyxoviruses in fauna from Antarctica. *J Wildl Dis* **29**, 568-571.

Bankevich A, Nurk S, Antipov D*, et al.* (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19**, 455-477.

Barbosa A, Palacios MJ (2009) Health of Antarctic birds: a review of their parasites, pathogens and diseases. *Polar Biol* **32**, 1095.

Bengtson JL, Boveng P, Franzen U*, et al.* (1991) Antibodies to Canine-Distemper Virus in Antarctic Seals. *Marine Mammal Science* **7**, 85-87.

Blinkova O, Victoria J, Li Y*, et al.* (2010) Novel circular DNA viruses in stool samples of wild-living chimpanzees. *J Gen Virol* **91**, 74-86.

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120.

Brum JR, Ignacio-Espinoza JC, Roux S*, et al.* (2015) Ocean plankton. Patterns and ecological drivers of ocean viral communities. *Science* **348**, 1261498.

Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972-1973.

Clarke J, Kerry K (2000) Diseases and parasites of penguins. *Penguin Conservation* **13**, 5-24.

Cole TL, Dutoit L, Dussex N*, et al.* (2019) Receding ice drove parallel expansions in Southern Ocean penguins. *Proc Natl Acad Sci U S A*.

Darriba D, Taboada GL, Doallo R, Posada D (2011) ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* **27**, 1164-1165.

Dayaram A, Galatowitsch ML, Arguello-Astorga GR*, et al.* (2016) Diverse circular replication-associated protein encoding viruses circulating in invertebrates within a lake ecosystem. *Infect Genet Evol* **39**, 304-316.

de la Higuera I, Kasun GW, Torrance EL*, et al.* (2020) Unveiling Crucivirus Diversity by Mining Metagenomic Data. *bioRxiv*, 2020.2002.2027.967703.

Diemer GS, Stedman KM (2012) A novel virus genome discovered in an extreme environment suggests recombination between unrelated groups of RNA and DNA viruses. *Biol Direct* **7**, 13.

Dunlap DS, Ng TF, Rosario K*, et al.* (2013) Molecular and microscopic evidence of viruses in marine copepods. *Proc Natl Acad Sci U S A* **110**, 1375-1380.

Fontenele RS, Lacorte C, Lamas NS*, et al.* (2019) Single Stranded DNA Viruses Associated with Capybara Faeces Sampled in Brazil. *Viruses* **11**.

Gardner H, Kerry K, Riddle M, Brouwer S, Gleeson L (1997) Poultry virus infection in Antarctic penguins. *Nature* **387**, 245-245.

Gavryushkina A, Heath TA, Ksepka DT*, et al.* (2017) Bayesian Total-Evidence Dating Reveals the Recent Crown Radiation of Penguins. *Systematic Biology* **66**, 57-73.

Geoghegan JL, Pirotta V, Harvey E*, et al.* (2018) Virological Sampling of Inaccessible Wildlife with Drones. *Viruses* **10**.

Gerlt JA (2017) Genomic Enzymology: Web Tools for Leveraging Protein Family Sequence-Function Space and Genome Context to Discover Novel Functions. *Biochemistry* **56**, 4293-4308.

Gong Z, Liang Y, Wang M*, et al.* (2018) Viral Diversity and Its Relationship With Environmental Factors at the Surface and Deep Sea of Prydz Bay, Antarctica. *Frontiers in microbiology* **9**, 2981.

Grimaldi WW, Seddon PJ, Lyver PO, Nakagawa S, Tompkins DM (2015) Infectious diseases of Antarctic penguins: current status and future threats. *Polar Biology* **38**, 591-606.

Huang Y, Niu B, Gao Y, Fu L, Li W (2010) CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* **26**, 680-682.

Jackson EW, Bistolas KS, Button JB, Hewson I (2016) Novel Circular Single-Stranded DNA Viruses among an Asteroid, Echinoid and Holothurian (Phylum: Echinodermata). *PLoS One* **11**, e0166093.

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution* **30**, 772-780.

Katoh K, Standley DM (2016) A simple method to control over-alignment in the MAFFT multiple sequence alignment program. *Bioinformatics* **32**, 1933-1942.

Kazlauskas D, Varsani A, Koonin EV, Krupovic M (2019) Multiple origins of prokaryotic and eukaryotic single-stranded DNA viruses from bacterial and archaeal plasmids. *Nat Commun* **10**, 3425.

Kearse M, Moir R, Wilson A*, et al.* (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647-1649.

Kraberger S, Arguello-Astorga GR, Greenfield LG*, et al.* (2015a) Characterisation of a diverse range of circular replication-associated protein encoding DNA viruses recovered from a sewage treatment oxidation pond. *Infect Genet Evol* **31**, 73-86.

Kraberger S, Farkas K, Bernardo P*, et al.* (2015b) Identification of novel Bromus- and Trifolium-associated circular DNA viruses. *Arch Virol* **160**, 1303-1311.

Krupovic M, Varsani A, Kazlauskas D*, et al.* (2020) Cressdnaviricota: a virus phylum unifying 7 families of Rep-encoding viruses with single-stranded, circular DNA genomes. *J Virol*.

Krupovic M, Zhi N, Li J*, et al.* (2015) Multiple layers of chimerism in a single-stranded DNA virus discovered by deep sequencing. *Genome Biol Evol* **7**, 993-1001.

Labonte JM, Suttle CA (2013) Previously unknown and highly divergent ssDNA viruses populate the oceans. *ISME J* **7**, 2169-2177.

Lee SY, Kim JH, Park YM*, et al.* (2014) A novel adenovirus in Chinstrap penguins (*Pygoscelis antarctica*) in Antarctica. *Viruses* **6**, 2052-2061.

Lee SY, Kim JH, Seo TK*, et al.* (2016) Genetic and Molecular Epidemiological Characterization of a Novel Adenovirus in Antarctic Penguins Collected between 2008 and 2013. *PLoS One* **11**, e0157032.

Letunic I, Bork P (2019) Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* **47**, W256-W259.

Lopez-Bueno A, Tamames J, Velazquez D*, et al.* (2009) High diversity of the viral community from an Antarctic lake. *Science* **326**, 858-861.

Mac Donald JW, Conroy JWH (1971) Virus Disease Resembling Puffinosis in the Gentoo Penguin *Pygoscelis papua* on Signy Island South Orkney Islands. *British Antarctic Survey Bulletin*, 80-82.

McFarlane RA (2009) Health Assessment and Diseases of the Weddell seal, *Leptonochotes weddelli,* in Vestfold Hills, East Antarctica. In: *Health of Antarctic Wildlife* (eds. Kerry KR, Riddle M), pp. 139-166. Springer Berlin Heidelberg, Berlin, Heidelberg.

Miranda JA, Culley AI, Schvarcz CR, Steward GF (2016) RNA viruses as major contributors to Antarctic virioplankton. *Environ Microbiol* **18**, 3714-3727.

Morandini V, Dugger KM, Ballard G*, et al.* (2019) Identification of a Novel Adelie Penguin Circovirus at Cape Crozier (Ross Island, Antarctica). *Viruses* **11**.

Morgan IR, Westbury HA (1981) Virological studies of Adelie Penguins (*Pygoscelis adeliae*) in Antarctica. *Avian Dis* **25**, 1019-1026.

Muhire BM, Varsani A, Martin DP (2014) SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. *PLoS One* **9**, e108277.

Munro G (2006) Outbreak of avian pox virus in gentoo penguins in the Falklands. February 2006. Falklands Conservation, Stanley, Falkland Islands.

Neira V, Tapia R, Verdugo C*, et al.* (2017) Novel Avulaviruses in Penguins, Antarctica. *Emerg Infect Dis* **23**, 1212-1214.

Ng TF, Zhang W, Sachsenroder J*, et al.* (2015) A diverse group of small circular ssDNA viral genomes in human and non-human primate stools. *Virus Evol* **1**, vev017.

O'Leary NA, Wright MW, Brister JR*, et al.* (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research* **44**, D733-D745.

Olivares F, Tapia R, Galvez C*, et al.* (2019) Novel penguin Avian avulaviruses 17, 18 and 19 are widely distributed in the Antarctic Peninsula. *Transbound Emerg Dis* **66**, 2227-2232.

Orton JP, Morales M, Fontenele RS*, et al.* (2020) Virus Discovery in Desert Tortoise Fecal Samples: Novel Circular Single-Stranded DNA Viruses. *Viruses* **12**.

Pearson VM, Caudle SB, Rokyta DR (2016) Viral recombination blurs taxonomic lines: examination of single-stranded DNA viruses in a wastewater treatment plant. *PeerJ* **4**, e2585.

Phan TG, Mori D, Deng X*, et al.* (2015) Small circular single stranded DNA viral genomes in unexplained cases of human encephalitis, diarrhea, and in untreated sewage. *Virology* **482**, 98-104.

Pistorius PA (2009) Falkland Island Seabird Monitoring Programme. Annual Report 2008/2009. Falklands Conservation, Stanley, Falkland Islands.

Rastrojo A, Alcamí A (2018) Chapter Two - Viruses in Polar Lake and Soil Ecosystems. In: *Advances in Virus Research* (ed. Malmstrom CM), pp. 39-54. Academic Press.

Reavy B, Swanson MM, Cock PJ*, et al.* (2015) Distinct circular single-stranded DNA viruses exist in different soil types. *Appl Environ Microbiol* **81**, 3934-3945.

Rosario K, Duffy S, Breitbart M (2012) A field guide to eukaryotic circular single-stranded DNA viruses: insights gained from metagenomics. *Arch Virol* **157**, 1851-1871.

Roux S, Enault F, Bronner G*, et al.* (2013) Chimeric viruses blur the borders between the major groups of eukaryotic single-stranded DNA viruses. *Nat Commun* **4**, 2700.

Shannon P, Markiel A, Ozier O*, et al.* (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504.

Smeele ZE, Ainley DG, Varsani A (2018) Viruses associated with Antarctic wildlife: From serology based detection to identification of genomes using high throughput sequencing. *Virus Res* **243**, 91-105.

Sommers P, Fontenele RS, Kringen T*, et al.* (2019) Single-Stranded DNA Viruses in Antarctic Cryoconite Holes. *Viruses* **11**, 1022.

Steel O, Kraberger S, Sikorski A*, et al.* (2016) Circular replication-associated protein encoding DNA viruses identified in the faecal matter of various animals in New Zealand. *Infect Genet Evol* **43**, 151-164.

Stover BC, Muller KF (2010) TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. *Bmc Bioinformatics* **11**, 7.

Tisza MJ, Pastrana DV, Welch NL*, et al.* (2020) Discovery of several thousand highly diverse circular DNA viruses. *eLife* **9**.

Van Doorslaer K, Ruoppolo V, Schmidt A*, et al.* (2017) Unique genome organization of non-mammalian papillomaviruses provides insights into the evolution of viral early proteins. *Virus Evol* **3**, vex027.

Varsani A, Kraberger S, Jennings S*, et al.* (2014) A novel papillomavirus in Adelie penguin (*Pygoscelis adeliae*) faeces sampled at the Cape Crozier colony, Antarctica. *J Gen Virol* **95**, 1352-1365.

Varsani A, Porzig EL, Jennings S*, et al.* (2015) Identification of an avian polyomavirus associated with Adelie penguins (*Pygoscelis adeliae*). *J Gen Virol* **96**, 851-857.

Watts JM, Miller GD, Shellam GR (2009) *Infectious bursal disease virus and Antarctic birds*.

Wille M, Aban M, Wang J*, et al.* (2019) Antarctic Penguins as Reservoirs of Diversity for Avian Avulaviruses. *J Virol* **93**.

Wille M, Harvey E, Shi M*, et al.* (2020) Sustained RNA virome diversity in Antarctic penguins and their ticks. *ISME J*.

Woods R, Jones HI, Watts J, Miller GD, Shellam GR (2009) Diseases of Antarctic Seabirds. In: *Health of Antarctic wildlife: a challenge for science and policy* (eds. Kerry KR, Riddle M). Springer, London.

Yau S, Lauro FM, DeMaere MZ*, et al.* (2011) Virophage control of antarctic algal host-virus dynamics. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 6163-6168.

Yau S, Seth-Pasricha M (2019) Viruses of Polar Aquatic Environments. *Viruses* **11**, 189.

Zablocki O, van Zyl L, Adriaenssens EM*, et al.* (2014a) High-level diversity of tailed phages, eukaryote-associated viruses, and virophage-like elements in the metaviromes of antarctic soils. *Appl Environ Microbiol* **80**, 6888-6897.

Zablocki O, van Zyl L, Adriaenssens EM*, et al.* (2014b) Niche-dependent genetic diversity in Antarctic metaviromes. *Bacteriophage* **4**, e980125.

Zallot R, Oberg NO, Gerlt JA (2018) 'Democratized' genomic enzymology web tools for functional assignment. *Curr Opin Chem Biol* **47**, 77-85.

Zawar-Reza P, Arguello-Astorga GR, Kraberger S*, et al.* (2014) Diverse small circular single-stranded DNA viruses identified in a freshwater pond on the McMurdo Ice Shelf (Antarctica). *Infect Genet Evol* **26**, 132-138.

Zhang W, Olson NH, Baker TS*, et al.* (2001) Structure of the Maize streak virus geminate particle. *Virology* **279**, 471-477.

Chapter 6. Identification of Circovirus Genome in a Chinstrap Penguin (*Pygoscelis antarcticus*) and Adélie Penguin (*Pygoscelis adeliae*) on the Antarctic Peninsula

Hila Levy[1], Steven R. Fiddaman[1], Anni Djurhuus[2], Caitlin E. Black[3], Simona Kraberger[4], Adrian L. Smith[1*], Tom Hart[1] and Arvind Varsani[4,5*]

1. Department of Zoology, University of Oxford, South Parks Road, Oxford, OX1 3SZ, United Kingdom
2. Faculty of Science and Technology, University of the Faroe Islands, Vestarabryggja 15, FO-100 Tórshavn, Faroe Islands
3. Department of Evolutionary Biology and Environmental Studies, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland
4. The Biodesign Center for Fundamental and Applied Microbiomics, Center for Evolution and Medicine, School of Life Sciences, Arizona State University, Tempe, AZ 85287-5001, USA
5. Structural Biology Research Unit, Department of Integrative Biomedical Sciences, University of Cape Town, Observatory, Cape Town 7701, South Africa

*Authors for Correspondence:

Adrian L. Smith: adrian.smith@zoo.ox.ac.uk

Arvind Varsani: arvind.varsani@asu.edu

Keywords: circovirus, penguins, Antarctica, *Pygoscelis antarcticus*, *Pygoscelis adeliae*

Abstract

Circoviruses infect a variety of animal species and have small (~1.8-2.2 kb) circular single-stranded DNA genomes. Recently a penguin circovirus (PenCV) was identified associated with an Adélie Penguin with feather disorder and in the cloacal swabs of three asymptomatic Adélie Penguins at Cape Crozier, Antarctica. A total of 75 cloacal swab samples obtained from adults and chicks of three species of penguin (Genus: *Pygoscelis*) from seven Antarctic breeding colonies (South Shetland Islands and Western Antarctic Peninsula) in the 2015-16 breeding season were screened for the PenCV. We identified new variants of PenCV in one Adélie Penguin (*Pygoscelis adeliae*) and one Chinstrap Penguin (*Pygoscelis antarcticus*) from Port Charcot, Booth Island, Southwest Antarctic Peninsula, a site home to all three species of Pygoscelid penguins. These two PenCV genomes (length of 1986 nucleotides) share >99% genome-wide nucleotide identity with each other, and share ~87% genome-wide nucleotide identity with the PenCV sequences described from Adélie Penguins at Cape Crozier ~4400 km away in East Antarctica. We did not find any evidence of recombination among PenCV sequences. This is the first report of PenCV in Chinstrap Penguins, and the first detection outside of Ross Island, East Antarctica. Given the limited knowledge on Antarctic animal viral diversity, future samples from Antarctic wildlife should be screened for these and other viruses to determine the prevalence and potential impact of viral infections.

Introduction

Growing interest and access to the Antarctic continent has both increased human-induced challenges to Antarctic wildlife, while expanding our knowledge of the region's endemic biota. However, our knowledge of pathogen presence and prevalence in even the continent's most iconic group of animals, penguins (Order: Sphenisciformes), remains limited. To date, most studies of Antarctic penguin viral pathogens have focused on colonies located near fixed scientific research stations, which can monitor sites seasonally or in some cases, year-round, over several

years (Barbosa & Palacios, 2009; Grimaldi, Seddon, Lyver, Nakagawa, & Tompkins, 2015; Morandini et al., 2019; Neira et al., 2017; Olivares et al., 2019; Smeele, Ainley, & Varsani, 2018; Wille et al., 2019; Yinda, Esefeld, Peter, Matthijnssens, & Zell, 2019).

Literature concerning penguin disease primarily consists of observations and case reports of pathologies or mass mortality events in these well-studied colonies. Exploratory studies to determine sub-clinical prevalence of viral pathogens are more recent, progressing from immunoassays, histopathology, plate culture, and microscopic techniques (Alexander et al., 1989; Austin & Webster, 1993; Baumeister et al., 2004; Gardner, Kerry, Riddle, Brouwer, & Gleeson, 1997; Mac Donald & Conroy, 1971; Morgan & Westbury, 1981, 1988; Morgan, Westbury, Caple, & Campbell, 1981; Wallensten et al., 2006) to PCR- and sequencing-based methods to survey host viromes and identify particular viruses (Hurt et al., 2014; Lee et al., 2014; Neira et al., 2017; Thomazelli et al., 2010; Varsani et al., 2014; Varsani et al., 2015). Though these methodological advances have broadened our knowledge of the viral profile present in the Antarctic environment, more analyses are required to both document new viruses and understand viral prevalence.

A recent study of one of the world's largest penguin colonies, the Cape Crozier Adélie Penguin (*Pygoscelis adeliae*) colony on Ross Island in East Antarctica, identified a novel circovirus associated with Adélie Penguins (Morandini et al., 2019). Circoviruses (family: *Circoviridae*; genus: *Circovirus*) are circular single stranded DNA viruses that infect a variety of avian, fish, mammalian and reptilian species. Their genome encodes two genes in an ambisense orientation, the replication-associated protein (*rep*) in the virion sense and the capsid protein (*cp*) in the complementary sense (Breitbart et al., 2017). The penguin circovirus (PenCV) was first noted in the guano of a chick exhibiting feather disorder in the 2018-19 breeding season (Morandini et al., 2019). Subsequent analysis of cloacal swabs from the 2014-15 breeding season also identified three other individuals with this virus sharing >99% shared genome-wide identity (Morandini et

al., 2019). Feather disorder in penguins was first observed in an Emperor Penguin (*Aptenodytes forsteri*) chick at Cape Washington, near the Ross Sea in 1996 (Morandini et al., 2019) with subsequent observations in African Penguin (*Spheniscus demersus*) and Magellanic Penguin (*Spheniscus magellanicus*) chicks (Kane et al., 2010). Adélie Penguin chicks with the disorder have been reported three times in the literature: (1) at Esperanza Bay (Hope Bay), Antarctic Peninsula in 2013-14 (Andrés Barbosa et al., 2014), (2) at all three colonies on Ross Island in 2011-12 (Grimaldi et al., 2015), and (3) the recent case at Cape Crozier, Ross Island in 2018-19 (Morandini et al., 2019). A number of viruses have been found to infect Adélie Penguins (reviewed in Smeele et al. (2018)), though the etiological agent of the feather disorder has not been confirmed. The recent identification of the PenCV circovirus in an affected individual is intriguing.

Chinstrap Penguins (*Pygoscelis antarcticus*) are closely related to Adélie Penguins (estimated divergence 3.47 mya [95% HPD: 1.68-5.27] (Gavryushkina et al., 2017). The southernmost portions of the Chinstrap Penguin range overlap with the northernmost reaches of the Adélie Penguin in the South Sandwich Islands, the South Orkney Island, South Shetland Islands, and down the Western Antarctic Peninsula to 64°S. Where the species overlap, colonies tend to be either entirely single-species, or have species-segregated aggregations within the larger colonies. Among the larger, mixed colonies, it is rare to have more than two penguin species breeding. One interesting exception is the Port Charcot colony on Booth Island, Western Antarctic Peninsula, which is home to all three species in the genus *Pygoscelis* (Lynch, Naveen, & Fagan, 2008). Predominated by the Gentoo Penguin (*Pygoscelis papua*), with approximately 1300-1800 nests, it is also home to 4-10 nests of Adélie Penguins and 16-25 nests of Chinstrap Penguins (Humphries et al., 2017). The three species nest in very close proximity to each other, which make it a site of interest in exploring host-pathogen interactions in sympatric species. In this paper, we report the molecular detection of sequences of Adélie Penguin circovirus (PenCV) in both an Adélie Penguin and Chinstrap Penguin at Port Charcot, representing the first detection

of this virus in the Antarctic Peninsula, 4400 km away from its first reported location at Cape Crozier, and the first detection in a Chinstrap Penguin.

Materials and Methods

Between 27 December 2015-17 January 2016, a total of 75 cloacal swab samples were obtained from penguins in the genus *Pygoscelis* at seven sites in the South Shetland Islands and the Antarctic Peninsula (**Figure 1**) as part of a DNA virome study. These consisted of 17 Adélie Penguins (9 adults, 8 chicks), 32 Chinstrap Penguins (21 adults, 11 chicks), and 26 Gentoo Penguins (16 adults, 10 chicks). Regular flocked swabs (Eswab™, Copan®, Murrieta, CA, USA) were used, stored in 1 ml of liquid Amies medium, which was stored frozen aside from transport with cold packs under relevant import permits to the United Kingdom. Adult birds were swabbed at the nest when their cloacas were positioned accessibly (facing out), such that the sample could be obtained in under 30 seconds with minimal disturbance or restraint. In cases of adult-chick paired sampling, the adult was swabbed first, and then chicks were retrieved by hand from the nest, swabbed, and immediately returned to the nest in under one minute. All sampling in Antarctica was conducted under permits S3 34/2015 & S7/S9 35-2015, with ethical approval from the University of Oxford Animal Welfare and Ethical Review Board.

Viral DNA was extracted using 200 µl of the swab suspension after vortexing, using the High Pure Viral Nucleic Acid Kit (Roche Diagnostics, Indianapolis, IN, USA). The virus was preferentially amplified for circular DNA by rolling circle amplification (RCA) using the TempliPhi™ 100 Amplification Kit (GE Healthcare, Chicago, IL, USA). The resulting RCA-amplified DNA (5 µl) was pooled based on animal species per site and used to generate Illumina sequencing libraries using the Nextera DNA Flex Library Prep Kit (Illumina Inc, San Diego, CA, USA). The libraries were sequenced on an Illumina4000 sequencer (2 × 100 bp library) and the resulting paired-end raw reads were trimmed using Trimmomatic (Bolger, Lohse, & Usadel, 2014) and then *de novo* assembled using metaSPAdes v 3.12.0 (Bankevich et al., 2012). Assembled contigs > 500 nts were

analyzed against a viral RefSeq (O'Leary et al., 2016) protein database using BLASTx (Altschul, Gish, Miller, Myers, & Lipman, 1990). We identified two contigs (1508 and 1859 nts) in two libraries from Booth Island (one each for Adélie and Chinstrap Penguins) that had high similarity (>85%) to the recently identified penguin circovirus (Morandini et al., 2019). Based on the sequences of the contigs, we designed a pair of abutting primers (5'-TGAAAGCATGGAGAACTCTGTATAATAAAGT-3' and 5'- GCGTAATCATTTAATTCGTTCTCGTCATCT-3')

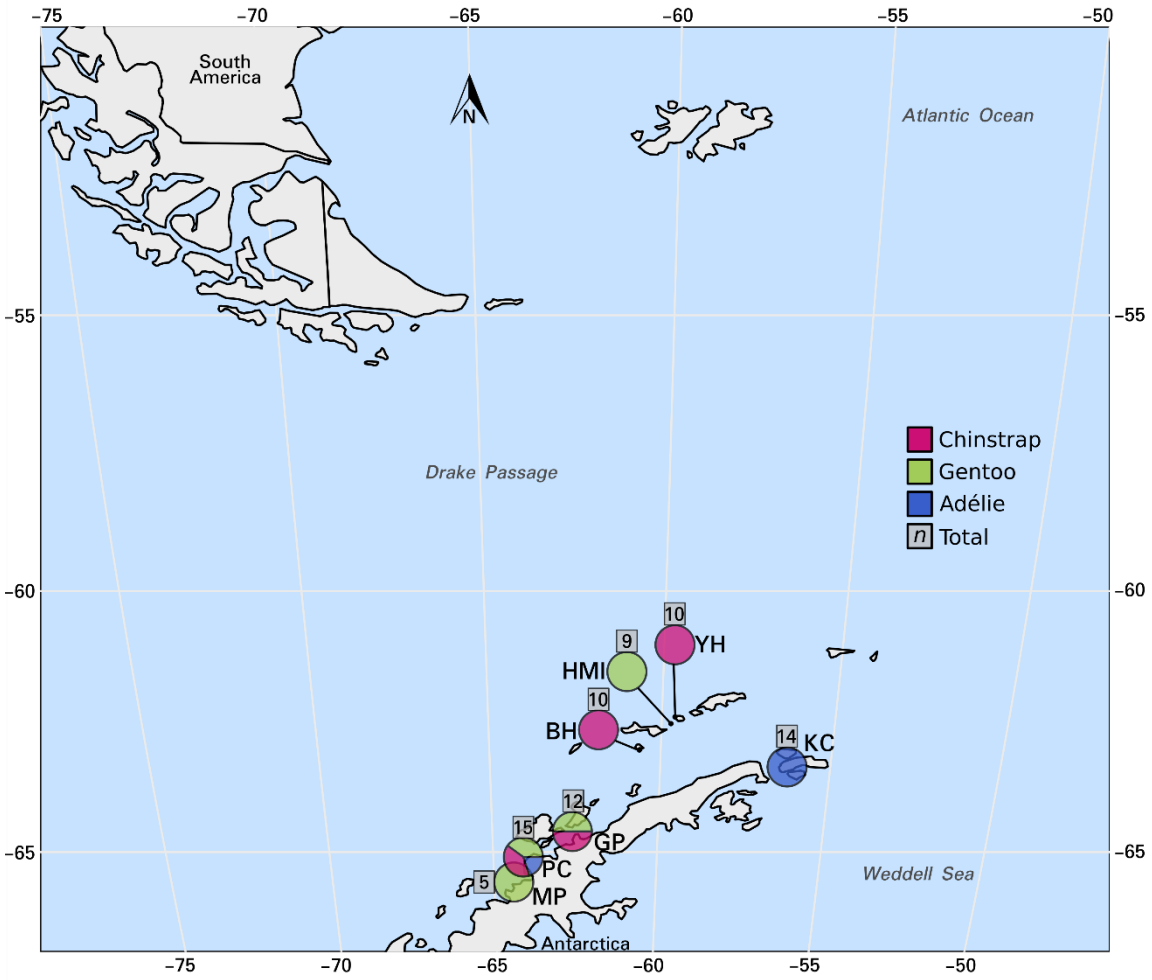


Figure 1: Map of *Pygoscelis* species penguin colonies sampled along the South Shetland Islands and Western Antarctic Peninsula, with pie chart colors representing the species sampled at each site. The number above each pie chart reflects the total number of swabs obtained at the site. Site abbreviations: HMI = Half Moon Island; YH = Yankee Harbor; Baily Head (BH); Kinnes Cove (KC); Georges Point (GP); Port Charcot (PC); Moot Point (MP) as in **Table 1**.

to screen all the individual samples (n=75) and recover the full genomes. The genomes were recovered with this primer pair by PCR using 0.5 µl of the RCA product as a template with KAPA HiFi HotStart DNA Polymerase (Kapa Biosystems, Wilmington, MA, USA) in 20 µl reactions. The amplicons were resolved on a 0.7% agarose gel, excised, purified and cloned into pJET1.2 plasmid (ThermoFisher, Waltham, MA, USA). The recombinant plasmids were Sanger-sequenced by primer walking at Macrogen Inc. (Seoul, Korea) and contigs assembled using Geneious Prime (Kearse et al., 2012).

The genomes of the circoviruses identified in this study were aligned with the four penguin circoviruses (GenBank MN164703- MN164706) and representative genomes of other avian circoviruses using MUSCLE (Edgar, 2004). The alignment was used to infer a Maximum-likelihood phylogenetic tree with 1000 bootstrap replicates using PhyML (Guindon et al., 2010) with GTR+I+G nucleotide substitution model (identified as the best fit model using jModelTest (Posada, 2008). Similarly, the Rep and CP amino acid sequences encoded by the circoviruses identified in this study were aligned with those of the four penguin circoviruses and representative avian circoviruses. The resulting alignments were used to infer Maximum-likelihood phylogenetic trees with 1000 bootstrap replicates using PhyML (Guindon et al., 2010) using the substitution models rtREV+G+I+F for CP and WAG+G for Rep determined as best fit models using ProtTest (Darriba, Taboada, Doallo, & Posada, 2011). For all phylogenetic trees, branches with <60% branch support were collapsed using TreeGraph2 (Stöver & Müller, 2010) and the trees were rooted with circoviruses sequences of duck circovirus (DuCV), goose circovirus (GoCV) and swan circovirus (SwCV).

All pairwise identities (genomes and amino acids) were determined using SDT v1.2 (Muhire, Varsani, & Martin, 2014). We aligned the PenCV sequences (two from this study and four reported by Morandini et al., 2019) and analyzed these for recombination using RDP4 v.4.97 (Martin, Murrell, Golden, Khoosal, & Muhire, 2015).

In order to test for signatures of positive selection, the *codeml* program in PAML v4.9 (Yang, 2007) was used. Two sets of model comparisons were used, (i) the M1/M2a comparison, where positive selection is disallowed and allowed, respectively, and (ii) the M7/M8 comparison, where dN/dS is modeled using a beta distribution and positive selection is disallowed and allowed respectively. The Bayes Empirical Bayes algorithm was then used to identify particular sites under positive selection (Yang, Wong, & Nielsen, 2005).

Results and Discussion

Of 75 swabs tested in this study, two from Port Charcot, Booth Island, Southwest Antarctic Peninsula presented penguin circovirus (PenCV) sequences, found in one Adélie Penguin adult (1/2 sampled at the site) and one Chinstrap Penguin adult (1/3 sampled) (GenBank accession numbers MN877414 -MN877415). As the circovirus species delineation threshold is 80% genome-wide pairwise identity (Rosario et al., 2017), this is still considered to be the same species of PenCV recently identified, found only in Adélie Penguins in East Antarctica (Morandini et al., 2019). No other sites sampled, and no chicks or Gentoo Penguins at this site yielded any circovirus sequences, though this absence may not be indicative of true prevalence due to the small sample size in this study (**Table 1**).

Though the Booth Island sample size is small, there were only 6 Adélie Penguin chicks and 25 Chinstrap Penguin chicks counted at this colony that year (Humphries et al., 2017), and so this sample set may be considered representative. Gentoo Penguins, on the other hand, had 1023 chicks that year at the site (Humphries et al., 2017), and further sampling would be needed to determine whether this virus might be shared in that species. The absence of virus in the chicks of infected adults could indicate that infection occurs during the non-breeding season elsewhere, and/or that this virus is not vertically transmitted or was undetectable in our samples. It is impossible to understand whether this is native or an introduction.

Table 3. Sampling site details and penguin circovirus (PenCV) viral sequence prevalence in adults and chicks. Colony names and subareas follow the Antarctic Site Inventory naming convention (Naveen et al., 2009).

Colony and Antarctic Site Inventory Subarea	**Coordinates**	**Penguin Species (Individuals Sampled)**	**PenCV Identification in Adults Sampled**	**PenCV Identification in Chicks Sampled**
Half Moon Island (HMI), South Shetland Islands	60˚36'S, 59˚55'W	Chinstrap (10)	0/10	-
Yankee Harbor (YH), Greenwich Island, South Shetland Islands	62˚32'S, 59˚47'W	Gentoo (9)	0/6	0/3
Baily Head (BH), Deception Island, South Shetland Islands	62°58'S, 60°30'W	Chinstrap (10)	0/5	0/5
Kinnes Cove (KC, Madder Cliff), Joinville Island, Northeast Antarctic Peninsula	63°18'S, 56°29'W	Adélie (14)	0/7	0/7
Georges Point (GP), Ronge Island, Northwest Antarctic Peninsula	64°40'S, 62°40'W	Chinstrap (6) Gentoo (6)	0/3 0/3	0/3 0/3
Port Charcot (PC), Booth Island, Southwest Antarctic Peninsula	65°05'S, 64°00'W	Adélie (3) Chinstrap (6) Gentoo (6)	**1/2** **1/3** 0/3	0/1 0/3 0/3
Moot Point (MP), Southwest Antarctic Peninsula	65°12'S, 64°06'W	Gentoo (5)	0/4	0/1
Total by Species		**Adélie (17)** **Chinstrap (32)** **Gentoo (26)**	**1/9** **1/21** **0/16**	**0/8** **0/11** **0/10**

Both genomes recovered from our samples were 1986 nucleotides in length, with capsid protein (*cp*; 726 nt) and replication-associated protein (*rep*; 870 nt) genes matching the reported length of PenCV genes. The two sequences share 99.6% genome-wide nucleotide identity with each other, and share 87.1-87.8% genome-wide nucleotide identity with the PenCV sequences recently discovered ~4400 km away from three adult and one Adélie Penguin chick (**Figure 2**). The circovirus sequences reported in this manuscript are approximately 12% divergent to those reported in Morandini et al. (2019) with all PenCV represented in a monophyletic clade based upon whole genome or encoded protein specific analyses (**Figure 2**; **Supplementary Data 1**). Interestingly, the PenCV reported in this manuscript sampled from Adélie and Chinstrap Penguin (*Pygoscelis antarcticus*) cloacal swabs are more similar to each other than the PenCV from Adélie Penguins from Cape Crozier on Ross Island. Hence, the sequence differences between Booth Island PenCV appears to be more related to location than a consequence of host species adaptation. In psittacine birds, different host species have been known to share circoviruses (beak and feather disease virus) with >98% shared genome-wide identity (Julian et al., 2013). There was no evidence of recombination between PenCVs sequenced at different locations. The two proteins encoded by PenCV share 99.6% (CP) and 99.3% (Rep) amino acid sequence identity between the two sequences, and 90.9-91.7% and 95.8% CP and Rep amino acid identity, respectively, with the Cape Crozier sequences (**Figure 2; Supplementary Data 1**).

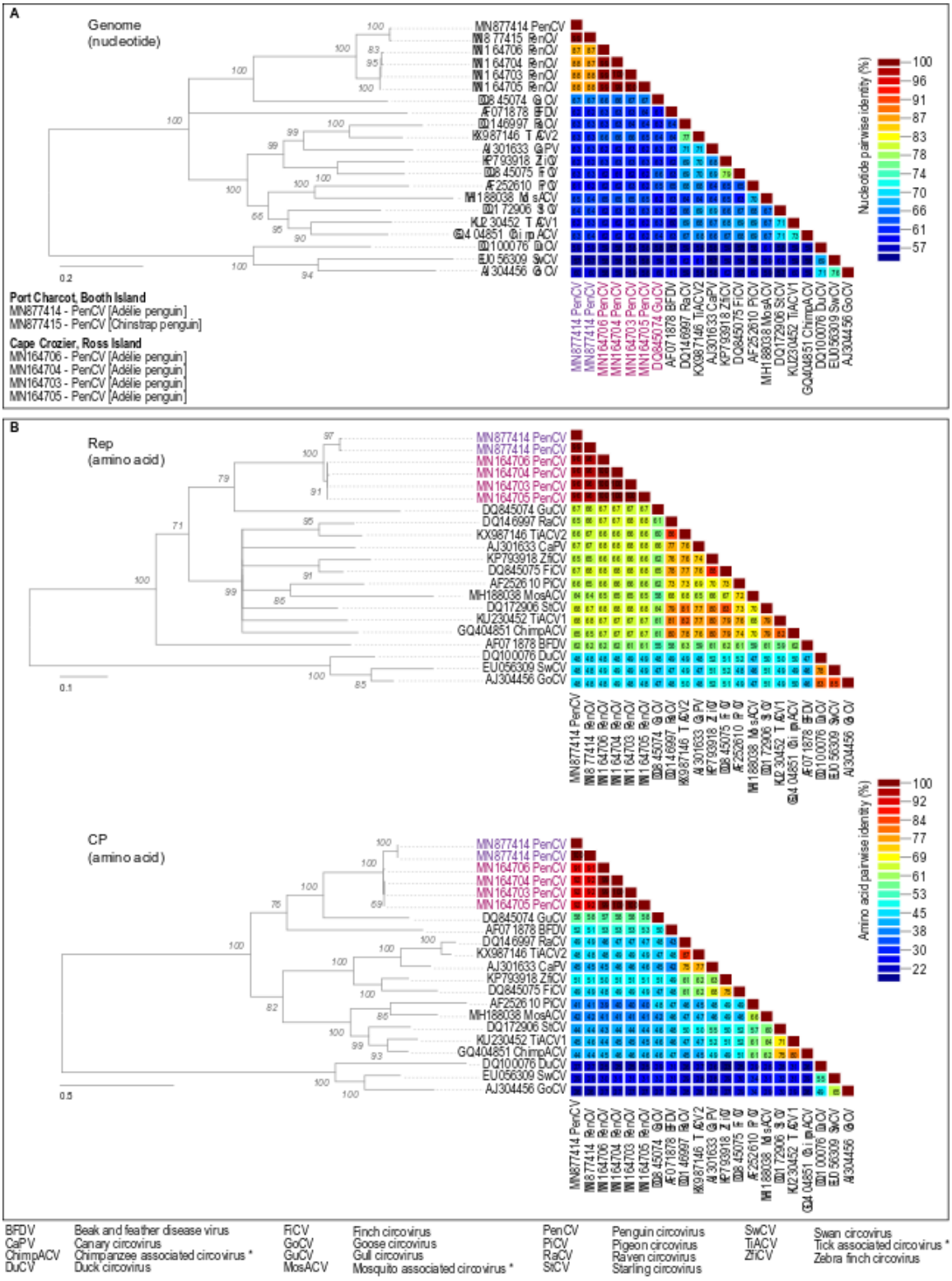


Figure 2: A. Maximum-likelihood phylogenetic tree based on the genomic sequences of PenCVs and other avian circoviruses with a color-coded nucleotide pairwise identity matrix. **B**. Maximum-likelihood phylogenetic trees of the Rep and CP amino acid sequences encoded by PenCVs with corresponding color-coded amino acid pairwise identity matrixes. Branches with <60% support

have been collapsed and the trees are rooted with sequences of duck circovirus (DuCV), goose circovirus (GoCV) and swan circovirus (SwCV). *Viruses likely infect avian species but they have been identified in fecal samples and insect vectors.

The addition of these new PenCV sequences continues to support their distinct categorization as a separate species from other known avian circoviruses, clustering most closely with gull circoviruses (GuCV) from *Larus* spp., with ~67% genome-wide nucleotide identity. There are no known circovirus sequences identified in other penguin species or seabirds in the next-closest related order Procellariformes. Similar clustering is seen for amino acid sequence-based phylogenies, with PenCVs sharing 66-67% Rep amino acid identity and 57-58% CP amino acid identity with GuCV. PenCV clusters with other avian circoviruses identified in parrot (beak and feather disease virus, BFDV), raven (RaCV), canary (CaCV), zebra finch (ZfiCV), finch (FiCV), pigeon (PiCV), and starling (StCV), sharing 62-64% amino acid identity, with their Reps and CPs sharing 61-68% and 39-53% amino acid identity, respectively.

As might be expected the level of shared amino acid identity between different circoviruses sequenced from different hosts was much higher for the Rep than for the CP protein (**Figure 2B; Supplementary Data 1**). It is interesting to note the very high level of identity within the PenCV sequences from Adélie and Chinstrap Penguins from Booth Island even within the CP (>99.6% amino acid identity). Within the PenCV clade, PAML analysis of the *rep* gene showed no evidence of positive selection (M1/M2a model, $p = 0.868$; M7/M8 model $p = 0.462$; **Supplementary Data 2**). In contrast, *cp* tended towards significance for the M7/M8 comparison ($p = 0.102$) and one codon was determined to have undergone positive selection using the BEB algorithm (site 201, posterior probability 0.973; **Supplementary Data 2**). Further increasing the power of this analysis with more penguin circovirus representatives may give a clearer indication

of the sites that may be functionally important for endonuclease or helicase function in case of the Rep or receptor bind in the case of the CP.

At the time of sampling, there was no evidence of physical deformities or feather anomalies, therefore the manuscript simply reports the identification of new PenCV and does not link it to any clinical presentation. Throughout the sampling season (2015-16), visits to approximately 100 different Pygoscelid penguin colonies in the South Shetland Islands and Antarctic Peninsula yielded no observations of feather disorder, although one was noted in a Rockhopper Penguin in the Falkland Islands. Previous sequencing of PenCV in a chick exhibiting feather disorder in Cape Crozier merits further exploration since feather abnormalities and immunosuppression in other orders of birds—Psittaciformes, Passeriformes, Anseriformes, Columbiformes, Struthioniformes, and Charadriiformes—have been correlated to avian circovirus infections (Smyth et al., 2006; Todd, 2000; Woods & Latimer, 2000).

In conclusion, the investigation of prevalence of circoviruses and other potential viral pathogens should be expanded across the range of all penguin species, particularly given their vulnerability to environmental stressors and the rapid pace of abiotic changes to their habitats (Cimino, Lynch, Saba, & Oliver, 2016; Clucas et al., 2014).

Acknowledgements

We thank Quark Expeditions for logistical and financial support provided to H.L., T.H., A.D. and C.E.B. The molecular work described in this study was supported by the Center of Evolution and Medicine Venture Fund (Center of Evolution and Medicine, Arizona State University, U.S.A.) grant awarded to A.V. H.L.'s work was supported by a University of Oxford Clarendon Fund Scholarship. SRF was supported by funding from the Biotechnology and Biological Sciences Research Council (BBSRC) under grant number BB/M011224/1.

Author Contributions

HL, AV, AD, TH and ALS developed and designed the study. TH obtained permits and secured logistics. HL, CEB, and TH collected all samples. HL extracted viral DNA. AV and HL analysed data and drafted the manuscript. AD and ALS managed the collaboration between the University of Oxford and Arizona State University, where SK and AV conducted amplification, cloning and sequencing. SRF conducted PAML analysis. All authors participated in discussion of results, reviewing drafts and developed the final manuscript.

Ethics Statement

Animal handling protocols followed Scientific Committee on Antarctic Research Code of Conduct on the use of Animals for Scientific Purposes in Antarctica. All sampling was conducted under permits S3 34/2015 and S7 35-2015 granted by the U.K. Foreign and Commonwealth Office to T.H., with ethical approval from the University of Oxford Animal Welfare and Ethical Review Board.

Conflict of Interest Statement

The authors declare no conflict of interest.

Data Availability Statement

Sequences for the viruses in this study are deposited in GenBank under accession numbers MN877414 -MN877415.

References

Alexander, D. J., Manvell, R. J., Collins, M. S., Brockman, S. J., Westbury, H. A., Morgan, I., & Austin, F. J. (1989). Characterization of Paramyxoviruses isolated from penguins in Antarctica and Sub-Antarctica during 1976-1979. *Archives of Virology, 109*(1-2), 135-143. doi:10.1007/bf01310525

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology, 215*(3), 403-410. doi:https://doi.org/10.1016/S0022-2836(05)80360-2

Austin, F. J., & Webster, R. G. (1993). Evidence of Orthomyxoviruses and Paramyxoviruses in fauna from Antarctica. *Journal of Wildlife Diseases, 29*(4), 568-571.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., . . . Pevzner, P. A. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology : a journal of computational molecular cell biology, 19*(5), 455-477. doi:10.1089/cmb.2012.0021

Barbosa, A., Colominas-Ciuró, R., Coria, N., Centurión, M., Sandler, R., Negri, A., & Santos, M. (2014). First record of feather-loss disorder in Antarctic penguins. *Antarctic Science, 27*(1), 69-70. doi:10.1017/S0954102014000467

Barbosa, A., & Jose Palacios, M. (2009). Health of Antarctic birds: a review of their parasites, pathogens and diseases. *Polar Biology, 32*(8), 1095-1115. doi:10.1007/s00300-009-0640-3

Baumeister, E., Leotta, G., Pontoriero, A., Campos, A., Montalti, D., Vigo, G., . . . Savy, V. (2004). Serological evidences of influenza A virus infection in Antarctica migratory birds. *International Congress Series, 1263*, 737-740. doi:http://dx.doi.org/10.1016/j.ics.2004.02.099

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England), 30*(15), 2114-2120. doi:10.1093/bioinformatics/btu170

Breitbart, M., Delwart, E., Rosario, K., Segalés, J., Varsani, A., & Ictv Report, C. (2017). ICTV Virus Taxonomy Profile: Circoviridae. *The Journal of general virology, 98*(8), 1997-1998. doi:10.1099/jgv.0.000871

Cimino, M. A., Lynch, H. J., Saba, V. S., & Oliver, M. J. (2016). Projected asymmetric response of Adélie penguins to Antarctic climate change. *Sci Rep, 6*(1), 28785. doi:10.1038/srep28785

Clucas, G. V., Dunn, M. J., Dyke, G., Emslie, S. D., Levy, H., Naveen, R., . . . Hart, T. (2014). A reversal of fortunes: climate change 'winners' and 'losers' in Antarctic Peninsula penguins. *Sci Rep, 4*, 5024. doi:10.1038/srep05024

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2011). ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics (Oxford, England), 27*(8), 1164-1165. doi:10.1093/bioinformatics/btr088

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research, 32*(5), 1792-1797. doi:10.1093/nar/gkh340

Gardner, H., Kerry, K., Riddle, M., Brouwer, S., & Gleeson, L. (1997). Poultry virus infection in Antarctic penguins. *Nature, 387*(6630), 245-245. doi:10.1038/387245a0

Gavryushkina, A., Heath, T. A., Ksepka, D. T., Stadler, T., Welch, D., & Drummond, A. J. (2017). Bayesian Total-Evidence Dating Reveals the Recent Crown Radiation of Penguins. *Systematic Biology, 66*(1), 57-73. doi:10.1093/sysbio/syw060

Grimaldi, W., Seddon, P., Lyver, P. B., Nakagawa, S., & Tompkins, D. (2015). Infectious diseases of Antarctic penguins: current status and future threats. *Polar Biology, 38*(5), 591-606. doi:10.1007/s00300-014-1632-5

Grimaldi, W. W., Hall, R. J., White, D. D., Wang, J., Massaro, M., & Tompkins, D. M. (2015). First report of a feather loss condition in Adelie penguins (Pygoscelis adeliae) on Ross Island,

Antarctica, and a preliminary investigation of its cause. *Emu - Austral Ornithology, 115*(2), 185-189. doi:10.1071/MU14068

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology, 59*(3), 307-321. doi:10.1093/sysbio/syq010

Humphries, G. R. W., Naveen, R., Schwaller, M., Che-Castaldo, C., McDowall, P., Schrimpf, M., & Lynch, H. J. (2017). Mapping Application for Penguin Populations and Projected Dynamics (MAPPPD): data and tools for dynamic management and decision support. *Polar Record, 53*(2), 160-166. doi:10.1017/S0032247417000055

Hurt, A. C., Vijaykrishna, D., Butler, J., Baas, C., Maurer-Stroh, S., Carolina Silva-de-la-Fuente, M., . . . Gonzalez-Acuna, D. (2014). Detection of Evolutionarily Distinct Avian Influenza A Viruses in Antarctica. *Mbio, 5*(3). doi:10.1128/mBio.01098-14

Julian, L., Piasecki, T., Chrząstek, K., Walters, M., Muhire, B., Harkins, G. W., . . . Varsani, A. (2013). Extensive recombination detected among beak and feather disease virus isolates from breeding facilities in Poland. *Journal of General Virology*, 94(5), 1086-1095. doi:10.1099/vir.0.050179-0

Kane, O. J., Smith, J. R., Boersma, P. D., Parsons, N. J., Strauss, V., Garcia-Borboroglu, P., & Villanueva, C. (2010). Feather-loss Disorder in African and Magellanic Penguins. *Waterbirds, 33*(3), 415-421. doi:10.1675/063.033.0321

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., . . . Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics (Oxford, England), 28*(12), 1647-1649. doi:10.1093/bioinformatics/bts199

Lee, S.-Y., Kim, J.-H., Park, Y. M., Shin, O. S., Kim, H., Choi, H.-G., & Song, J.-W. (2014). A Novel Adenovirus in Chinstrap Penguins (Pygoscelis antarctica) in Antarctica. *Viruses-Basel, 6*(5), 2052-2061. doi:10.3390/v6052052

Lynch, H. J., Naveen, R., & Fagan, W. F. (2008). Censuses of penguin, Blue-eyed Shag *Phalacrocorax atriceps* and Southern Giant Petrel *Macronectes giganteus* populations on the Antarctic Peninsula, 2001-2007. *Marine Ornithology, 36*(2), 83-97.

Mac Donald, J. W., & Conroy, J. W. H. (1971). Virus Disease Resembling Puffinosis in the Gentoo Penguin *Pygoscelis papua* on Signy Island South Orkney Islands. *British Antarctic Survey Bulletin*(26), 80-82.

Martin, D. P., Murrell, B., Golden, M., Khoosal, A., & Muhire, B. (2015). RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evolution, 1*(1). doi:10.1093/ve/vev003

Morandini, V., Dugger, K. M., Ballard, G., Elrod, M., Schmidt, A., Ruoppolo, V., . . . Varsani, A. (2019). Identification of a Novel Adélie Penguin Circovirus at Cape Crozier (Ross Island, Antarctica). *Viruses, 11*(12), 1088.

Morgan, I. R., & Westbury, H. A. (1981). Virological studies of Adélie penguins (*Pygoscelis adeliae*) in Antarctica. *Avian Diseases, 25*(4), 1019-1026. doi:10.2307/1590077

Morgan, I. R., & Westbury, H. A. (1988). Studies of viruses in penguins in the Vestfold Hills. *Hydrobiologia, 165*, 263-269. doi:10.1007/bf00025595

Morgan, I. R., Westbury, H. A., Caple, I. W., & Campbell, J. (1981). A survey of virus-infection in Sub-Antarctic penguins on Macquarie Island, Southern Ocean. *Australian Veterinary Journal, 57*(7), 333-335. doi:10.1111/j.1751-0813.1981.tb05839.x

Muhire, B. M., Varsani, A., & Martin, D. P. (2014). SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. *PLoS One, 9*(9), e108277-e108277. doi:10.1371/journal.pone.0108277

Naveen, R., Forrest, S. C., Dagit, R. G., Blight, L. K., Trivelpiece, W. Z., & Trivelpiece, S. G. (2009). Censuses of penguin, blue-eyed shag, and southern giant petrel populations in the Antarctic Peninsula region, 1994–2000. *Polar Record, 36*(199), 323-334. doi:10.1017/S0032247400016818

Neira, V., Tapia, R., Verdugo, C., Barriga, G., Mor, S., Ng, T. F. F., . . . González-Acuña, D. (2017). Novel Avulaviruses in Penguins, Antarctica. *Emerging Infectious Diseases, 23*(7), 1212-1214. doi:10.3201/eid2307.170054

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., . . . Pruitt, K. D. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research, 44*(D1), D733-D745. doi:10.1093/nar/gkv1189

Olivares, F., Tapia, R., Gálvez, C., Meza, F., Barriga, G. P., Borras-Chavez, R., . . . Neira, V. (2019). Novel penguin Avian avulaviruses 17, 18 and 19 are widely distributed in the Antarctic Peninsula. *Transboundary and Emerging Diseases, 66*(6), 2227-2232. doi:10.1111/tbed.13309

Posada, D. (2008). jModelTest: phylogenetic model averaging. *Mol Biol Evol, 25*(7), 1253-1256. doi:10.1093/molbev/msn083

Rosario, K., Breitbart, M., Harrach, B., Segalés, J., Delwart, E., Biagini, P., & Varsani, A. (2017). Revisiting the taxonomy of the family Circoviridae: establishment of the genus Cyclovirus and removal of the genus Gyrovirus. *Archives of Virology, 162*(5), 1447-1463. doi:10.1007/s00705-017-3247-y

Smeele, Z. E., Ainley, D. G., & Varsani, A. (2018). Viruses associated with Antarctic wildlife: From serology based detection to identification of genomes using high throughput sequencing. *Virus Research, 243*, 91-105. doi:https://doi.org/10.1016/j.virusres.2017.10.017

Smyth, J. A., Todd, D., Scott, A., Beckett, A., Twentyman, C. M., Bröjer, C., . . . Gavier-Widen, D. (2006). Identification of circovirus infection in three species of gull. *Veterinary Record, 159*(7), 212-214. doi:10.1136/vr.159.7.212

Stöver, B. C., & Müller, K. F. (2010). TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. *Bmc Bioinformatics, 11*, 7-7. doi:10.1186/1471-2105-11-7

Thomazelli, L. M., Araujo, J., Oliveira, D. B., Sanfilippo, L., Ferreira, C. S., Brentano, L., . . . Durigon, E. L. (2010). Newcastle disease virus in penguins from King George Island on the Antarctic region. *Veterinary Microbiology, 146*(1-2), 155-160. doi:10.1016/j.vetmic.2010.05.006

Todd, D. (2000). Circoviruses: Immunosuppressive threats to avian species: A review. *Avian Pathology, 29*(5), 373-394. doi:10.1080/030794500750047126

Varsani, A., Kraberger, S., Jennings, S., Porzig, E. L., Julian, L., Massaro, M., . . . Ainley, D. G. (2014). A novel papillomavirus in Adélie penguin (Pygoscelis adeliae) faeces sampled at the Cape Crozier colony, Antarctica. *Journal of General Virology, 95*(6), 1352-1365. doi:doi:10.1099/vir.0.064436-0

Varsani, A., Porzig, E. L., Jennings, S., Kraberger, S., Farkas, K., Julian, L., . . . Ainley, D. G. (2015). Identification of an avian polyomavirus associated with Adelie penguins (Pygoscelis adeliae). *J Gen Virol, 96*(Pt 4), 851-857. doi:10.1099/vir.0.000038

Wallensten, A., Munster, V. J., Osterhaus, A. D. M. E., Waldenstrom, J., Bonnedahl, J., Broman, T., . . . Olsen, B. (2006). Mounting evidence for the presence of influenza A virus in the

avifauna of the Antarctic region. *Antarctic Science, 18*(3), 353-356. doi:10.1017/s095410200600040x

Wille, M., Aban, M., Wang, J., Moore, N., Shan, S., Marshall, J., . . . Hurt, A. C. (2019). Antarctic Penguins as Reservoirs of Diversity for Avian Avulaviruses. *Journal of Virology, 93*(11), e00271-00219. doi:10.1128/JVI.00271-19

Woods, L. W., & Latimer, K. S. (2000). Circovirus Infection of Nonpsittacine Birds. *Journal of Avian Medicine and Surgery, 14*(3), 154-163, 110.

Yang, Z. (2007). PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and Evolution, 24*(8), 1586-1591. doi:10.1093/molbev/msm088

Yang, Z., Wong, W. S. W., & Nielsen, R. (2005). Bayes Empirical Bayes Inference of Amino Acid Sites Under Positive Selection. *Molecular Biology and Evolution, 22*(4), 1107-1118. doi:10.1093/molbev/msi097

Yinda, C. K., Esefeld, J., Peter, H. U., Matthijnssens, J., & Zell, R. (2019). Penguin megrivirus, a novel picornavirus from an Adélie penguin (Pygoscelis adeliae). *Archives of Virology, 164*(11), 2887-2890. doi:10.1007/s00705-019-04404-9

Chapter 7. Gastrointestinal Bacterial Microbiota of Four Penguin Species and Blue-Eyed Shags across the Scotia Arc

Hila Levy[1], Stephen G. Preston[1], Tom Hart[1], Adrian L. Smith[1]

1. Department of Zoology, University of Oxford, South Parks Road, Oxford, OX1 3SZ, United Kingdom

Keywords: penguins, wildlife disease, pathogens, 16S rRNA, Antarctica, *Pygoscelis*

Abstract

The bacterial microbiota of animal gastrointestinal tracts are diverse communities shaped by complex biotic and abiotic factors. Within the growing research effort devoted to characterising the relationships between hosts and their gastrointestinal microbiota, Sub-Antarctic and Antarctic seabirds remain under-represented or under-sampled. In this study, 16S rRNA gene parallel sequencing is employed to profile the bacterial microflora of four penguin species and a flying seabird with different extents of range overlap in the Scotia Arc and Antarctic Peninsula: Adélie Penguins (*Pygoscelis adeliae*), Chinstrap Penguins (*P. antarcticus*), Gentoo Penguins (*P. papua*), King Penguins (*Aptenodytes patagonicus*), and Blue-Eyed Shags (*Phalacrocorax atriceps bransfieldiensis*). Over 6.6 million reads of 11,383 unique amplicon sequence variants were obtained from 461 samples (*n* = 374 guano samples, *n* = 87 cloacal swabs) from 29 breeding sites. By characterising the main bacterial constituents and diversity of each avian host, site, and geographic region, we find connectivity in microbial sharing across sympatric hosts, with implications for wildlife disease epidemiology. In the widely distributed Gentoo Penguin, a species complex with strong philopatry and population genetic structure, we see little differentiation across long distances, but a signal of inter-annual variation that reflects responsiveness to local abiotic or biotic environmental conditions. These findings have

implications for understanding baseline interactions in sentinel marine species currently facing multiple stressors.

Introduction

The gastrointestinal tract of an animal host forms an ecosystem within an ecosystem (Miller *et al.* 2018). It is a habitat itself—home to an internal, complex assembly of bacteria, viruses, and fungi (Tasnim *et al.* 2017)—while interacting with the host's external surroundings, through ingested prey, water, abiotic elements and potentially, contaminants. Globally, a growing body of research has expanded our knowledge of the role of gastrointestinal microbiota in vertebrate host physiology, health and disease under varying conditions (Grond *et al.* 2018; Harris *et al.* 2019). Despite increased attention to the microbiome's importance, human and mammalian studies predominate relative to avian research, and characterisations of wild bird species remain sparse compared to domestic or captive taxa (Bahrndorff *et al.* 2016; Benskin *et al.* 2010; Grond *et al.* 2018; Waite & Taylor 2015; Zhu *et al.* 2002).

Seabirds comprise the majority of avian biomass in the Antarctic and Sub-Antarctic (Williams 1995), where their host biology makes characterisation of their microbiomes of interest. Antarctic seabird species have a shared evolutionary history in Clade Aequorlitornithes, sensu Prum *et al.* (2015), (Cole *et al.* 2019; Gavryushkina *et al.* 2017), prey upon resources within the same marine food web, and form dense terrestrial breeding colonies that may comprise multiple host species in close proximity (Humphries *et al.* 2017; Miller *et al.* 2010; Pickett *et al.* 2018; Tarroux *et al.* 2018; Trivelpiece *et al.* 1987).

Despite these unique ecological characteristics, few microbiome studies have examined the baseline microbiota for Antarctic seabird species. In Antarctic penguins, stomach ingesta, cloacal swabs, or faeces have been obtained for several research purposes: to characterise microbiota from different portions of the gastrointestinal tract (Yew *et al.* 2017), conduct intra-species comparisons across sites (Banks *et al.* 2009), perform inter-species comparisons at shared

(Yew *et al.* 2017) or distinct sites (Dewar *et al.* 2013), look for age-related differences within a species (Barbosa *et al.* 2016), and follow microbial changes across the annual moult fast (Dewar *et al.* 2014; Lee *et al.* 2019). [For a table summarising penguin-associated microbiome studies and their sample sizes, see **Chapter 4**.]

The challenges associated with wild sampling in such environments have often restricted microbiome studies to a single site or to a small number of samples across sites. While the microbiomes of at least some King Penguins (*Aptenodytes patagonicus*) (Dewar *et al.* 2013; Dewar *et al.* 2014; Thouzeau *et al.* 2003), Adélie Penguins (*Pygoscelis adeliae*) (Banks *et al.* 2009; Yew *et al.* 2017), Chinstrap Penguins (*P. antarcticus*) (Barbosa *et al.* 2016; Lee *et al.* 2019; Yew *et al.* 2017), and Gentoo Penguins (*P. papua*) (Dewar *et al.* 2013; Lee *et al.* 2019), have been characterised, spatial gaps remain in each of their studied ranges. On the other hand, perhaps due to their smaller populations and difficulty accessing nests, no literature fully characterising the microbiome of Antarctic flying birds has been noted, though some South Polar Skuas (*Stercorarius maccormicki*), Brown Skuas (*Stercorarius antarcticus*), Southern Giant Petrels (*Macronectes giganteus*), gulls (genus: *Larus*), and cormorant species (genus: *Phalacrocorax*) have been sampled occasionally for bacterial culture techniques or post-mortem in search of pathogens (Bonnedahl *et al.* 2005; Leotta *et al.* 2001; Leotta *et al.* 2006; Woods *et al.* 2009).

Gentoo Penguins present the widest spatial range of any penguin species. Their breeding sites are circumpolar and span 20° of latitude from Sub-Antarctic islands in the Atlantic, Indian, and Pacific Oceans, and throughout the Scotia Arc onto the Antarctic Peninsula (Lynch 2013). Depending on the location, in the Sub-Antarctic (sub-species *P. papua papua*), they breed sympatrically with as many as four other penguin species at once, including Magellanic Penguins (*Spheniscus magellanicus*), Southern Rockhopper Penguins (*Eudyptes chrysocome*), Royal Penguins (*E. schlegeli*), Macaroni Penguins (*E. chrysolophus*), congeneric Chinstrap Penguins, and King Penguins who have a similarly circumpolar sub-Antarctic extent (Lynch 2013; Pütz *et al.*

2001), and often overlap with at least one other penguin species at a site. In their Antarctic range, Gentoo Penguins (sub-species *P. papua ellsworthii*) commonly share breeding sites with Chinstrap Penguins, a species restricted to the Atlantic sector of the Southern Ocean, and Adélie Penguins, whose range encircles the Antarctic Continent (Lynch & LaRue 2014; Lynch *et al.* 2012). Penguins also overlap or coexist with seals and nesting flying birds in their colonies, including numerous albatrosses, petrels, storm petrels, and shearwaters (Order Procellariiformes); skuas, sheathbills, gulls, and terns (Order Charadriiformes); and cormorants (shags, Order Suliformes) (Clements *et al.* 2019; Lynch *et al.* 2008; Naveen *et al.* 2009).

Gentoo Penguins are an example of a species with seemingly high environmental plasticity combined with strong population genetic structure, natal site philopatry, and locally-restricted movement year-round (Clucas *et al.* 2014; Clucas *et al.* 2018; Levy *et al.* 2016; Pertierra *et al.* 2020; Tanton *et al.* 2004; Trivelpiece *et al.* 1987; Vianna *et al.* 2017). As such, this species complex presents an opportunity to investigate its host enteric diversity. In this study, we used opportunistic sampling to characterise the Gentoo Penguin microbiome through 16S rRNA next-generation sequencing.

We aim to assess whether the Gentoo Penguin's spatially-demarcated population structure in the Scotia Arc mirrors that of its bacterial microbiota. In parallel, by sampling an additional three penguin species (King, Adélie and Chinstrap Penguins) and the Blue-Eyed Shag (*Phalacrocorax atriceps bransfieldiensis*), we aim to evaluate whether sympatrically breeding hosts share microbiota, while expanding the baseline knowledge of enteric bacteria in the Antarctic region.

Materials and Methods

Sample Collection

Across four austral summers (2014/15, 2015/16, 2016/17, and 2017/18), 461 samples (*n* = 374 guano samples, *n* = 87 cloacal swabs) were collected for microbiome analysis from seabird

breeding colonies across the Scotia Arc and Antarctic Peninsula, from 29 unique sites. Samples were collected from four penguin species: Adélie Penguins (3 sites, *n* = 23), Chinstrap Penguins (4 sites, *n* = 56), Gentoo Penguins (25 sites, *n* = 364), King Penguins (1 site, *n* = 20) and a small number of Blue-Eyed Shags (*Phalacrocorax atriceps bransfieldiensis*, 2 sites, *n* = 11).

Cloacal swab samples were obtained from penguins at eight sites in the Falkland Islands, South Shetland Islands, South Georgia, and the Antarctic Peninsula between December 2015 and January 2016. The 87 swabs consisted of 17 Adélie Penguins (*Pygoscelis adeliae:* 9 adults, 8 chicks), 22 Chinstrap Penguins (*P. antarcticus*: 11 adults, 11 chicks), 28 Gentoo Penguins (*P. papua*: 17 adults, 10 chicks), and 20 King Penguins (*Aptenodytes patagonicus*: 10 adults, 10 chicks). Regular flocked swabs (Eswab™, Copan®, Murrieta, CA, USA) were used, stored in 1 ml of Amies liquid transport medium (maintenance medium comprising inorganic phosphate buffer, calcium and magnesium salts, and sodium chloride with a reduced environment due to the presence of sodium thioglycollate), which was stored frozen aside from transport with cold packs to the United Kingdom. Pygoscelid penguins were sampled while on the nest in order to confirm chick-adult parentage. Adult birds were swabbed at the nest when their cloacas were positioned accessibly (facing out), such that the sample could be obtained in under 30 seconds with minimal disturbance or restraint. In cases of adult-chick paired sampling, the adult was swabbed first, and then chicks were retrieved by hand from the nest, swabbed, and immediately returned to the nest in under one minute. Adult and juvenile king penguins were sampled using a two-person handling technique, captured by hand as they transited within the colony, and released in under one minute.

To collect faecal material, approximately 1.15 g (range: 0.1214 – 3.2825 g) of fresh guano was scooped into a 15 mL plastic conical centrifuge tube (Corning®, USA), with an additional ~0.3 g (range: 0 – 0.776 g) were placed into RNALater® preservative solution in a 1.5 mL microcentrifuge (Eppendorf®) tube. Sampling was distributed throughout the extent of the

colony to reduce bias, though fresh, more voluminous excreta of any colour were collected as a preference over dried material. Tubes were sealed and labelled, and stored at 4° C until shipping. Samples spent between 60-90 days in transport at room temperature prior to arriving at the laboratory, where they were catalogued and stored at -20° C.

Sampling sites are shown in **Figure 1** and details are found in **Table 1**. All sampling in the Antarctic Treaty Area was carried out in accordance with United Kingdom Home Office guidelines, under permits granted by the United Kingdom Foreign and Commonwealth Office (permits: S3-23/2013, S7 28/2013, S3 04/2014, S7 03/2014, 33/2016, 34/2016, 37/2017), in South Georgia under permits granted by the Government of South Georgia and the South Sandwich Islands (SCI-2014-017, RAP-2015-018, RAP-2016-035 and RAP-2017-034), and in the Falkland Islands under Falkland Islands Government Environmental Planning Department permits (R16.2014, R15 2014, and R04 2017), with ethical approval from the University of Oxford Animal Welfare and Ethical Review Board. All samples were transported in accordance with applicable export permits and United Kingdom Department for Environment, Food and Rural Affairs import permits.

DNA Extraction

For cloacal swabs, total genomic DNA was isolated from 200 µl of the swab suspension after vortexing, using a MoBio PowerSoil® DNA Extraction kit as per the manufacturer's instructions while onboard the expedition ship. Extracted genomic DNA was stored at -20°C and shipped on cold packs. For DNA extraction from guano, the MoBio PowerSoil® DNA Extraction kit (after 2016, rebranded as the Qiagen QIAamp® PowerFecal® DNA Kit) was used with approximately 0.1 g of starting material. The extraction kit protocol was modified such that the initial lysis step was extended to 12-18 hours at 65°C on a heated shaker block, and the subsequent homogenisation step was carried out on an MP Biomedicals™ FastPrep-24™ ("Marine

Sediment" Program: 2 cycles of 40 seconds at 5.5 m/s with a 5 minute pause between cycles). All other steps were per protocol and extracted DNA was stored at -20°C.

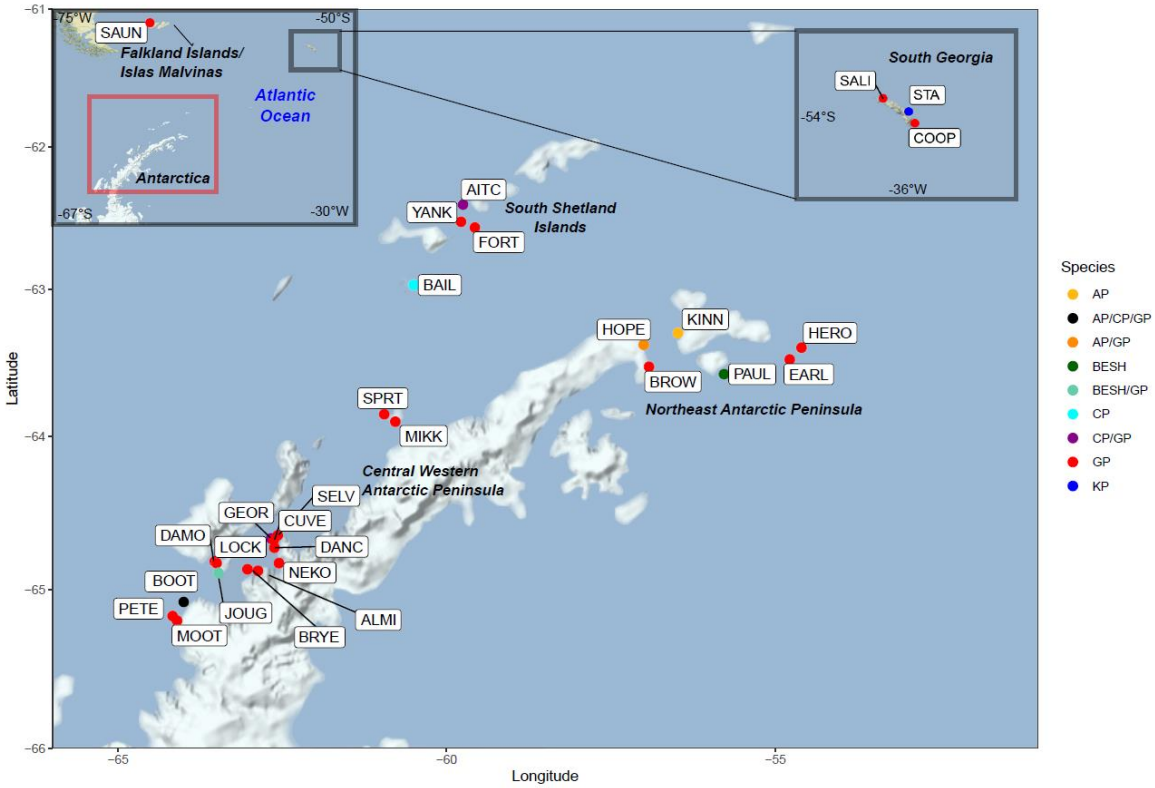


Figure 5. Sampling sites in the Scotia Arc (Falkland Islands, South Georgia, and Antarctic Peninsula). Species sampled are designated by colour in the legend. AP = Adélie Penguin, BESH = Blue-Eyed Shag, CP = Chinstrap Penguin, GP = Gentoo Penguin, KP = King Penguin. Colony names are expanded in Table 1.

16S rDNA Amplification and Sequencing

The V3-V4 region of the prokaryotic 16S ribosomal RNA (16S rRNA) gene was amplified using the primers 16S_338F (ACTCCTACGGGAGGCAGCAGT) (Lane 1991) and 16S_806R (GGACTACHVGGGTWTCTAAT) (Caporaso *et al.* 2011). The forward primer had a unique 7 nt 'barcode' tag attached for amplification of each sample to allow pooling of PCR products prior to library preparation. DNA from each guano sample was amplified in triplicate with different barcodes, to ensure that primer barcode bias did not affect amplification success, and to avoid

run-specific biases. PCR was carried out in 25 µL reactions using the high fidelity Phusion Hot Start Flex DNA polymerase enzyme (New England Biolabs, UK, M0535) according to manufacturer's recommendations, with 2 µL of faecal DNA diluted 1:10 in water to minimise inhibitors. Thermal cycling conditions were 98°C for 1 min; followed by 40 cycles of 98°C for 15 seconds, 60°C for 15 seconds, 72°C for 5 seconds; with a final extension of 72°C for 30 seconds. A water control was included for every 48 samples.

PCR products were visualised on a 1% agarose gel. A subset of all reactions was quantified using the Qubit 1X dsDNA High-Sensitivity Assay Kit (Invitrogen, UK, Q33231) using a Qubit Fluorometer, normalising DNA quantities from each reaction prior to pooling. Up to 96 PCR reactions from different samples with individually tagged 16S primers were combined and cleaned using the MinElute PCR purification kit (Qiagen, UK, 28004). Library preparation was performed using the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® (New England Biolabs, UK, E7103S) with according to the manufacturer's instructions, with no size selection. NEBNext Multiplex Oligos for Illumina (New England Biolabs, UK, E7335S and E7500S) indexing primers were used in order to run multiple libraries on the MiSeq machine in parallel. TapeStation (Agilent Genomics) and qPCR (NEBNext® Library Quant Kit for Illumina®, E7630S, New England Biolabs, UK) analyses were used to confirm expected template size and success of the library preparation before loading onto the MiSeq along with 5% phiX at the University of Oxford, Department of Zoology. Sequencing was split across four MiSeq runs, all using the 600-cycle MiSeq Reagent kit v3 (Illumina, UK), giving 300 nucleotide paired reads that could be merged with sufficient overlap to cover the ~468 bp amplicons (with the barcode sequence).

Sequencing Analysis and Taxonomic Classification

A custom Python script was used to demultiplex each library according to barcode, and to remove terminal primer sequences. This script made use of the NumPy v1.16.6 (Oliphant 2006) and python-Levenshtein PyPI v0.12.0 packages. Any reads still containing a copy of either primer

sequence after this trimming step were identified using cutadapt v1.15 (Martin 2011) and excluded from further analyses, on the assumption that they represent PCR artefacts (e.g. concatamers).

The package dada2 v1.12.1 (Callahan *et al.* 2016) was used in R v3.6.2 (The R Foundation, https://cran.r-project.org/) to process the sequencing data into counts of "Amplicon Sequence Variants" (ASVs), predictions of the actual biological sequences present. Processing with dada2 consisted of several stages:

1. Trimming sequences at the 3' end to decrease overall error rate (Illumina sequencing typically sees a drop off in sequence quality towards the end of a read) and filtering to remove sequences which had a high expected error rate even after trimming. The fastq_eestats2 command in USEARCH v10.0.240 (Edgar 2013) was used to investigate the effect of different trimming length/expected error filtering thresholds on yield. Read 1 of each read pair was trimmed to 270 bp, while the more error-prone read 2 was trimmed to 230, 240 or 270 bp, depending on the error rate characteristics of the run. In all cases, the remaining sequence is long enough to provide sufficient overlap for merging of read pairs. Reads were filtered to remove read pairs with more than either 1 or 2 expected errors in read 1 (again, depending on the characteristics of the run) or with more than 2 expected errors in read 2.
2. Inference of ASVs. Because the data were generated over five MiSeq runs, inference was performed in two stages, according to the method suggested by the author of the dada2 package [https://github.com/benjjneb/dada2/issues/479]. This is necessary because each run has a unique error-rate profile, and so requires its own error model. The approach is to perform inference on each run separately, then to rerun the inference for each using the ASVs predicted in the initial inference in *any* of the MiSeq runs as "priors". This should allow the two runs to converge towards a shared set of ASVs, reducing the

risk of a given biological sequence being described by two very similar, but distinct, ASVs in the data from different MiSeq runs, while also maximising the sensitivity of the algorithm for rare sequences. For the first round of inference runs, dada2 was set to pool all samples from each run for inference, to maximise sensitivity to rare variants (pool = TRUE). ASVs which were identified in at least 10 PCR reactions across the two runs were defined as priors for the second inference runs, for which dada2 was set to infer ASVs from each sample separately (pool = FALSE), though of course with reference to the supplied priors (this mimics the operation of dada2 in pool = 'pseudo' mode, while allowing for combined processing of two MiSeq runs.

3. Merging read pairs, using the mergePairs function.
4. Pooling of read counts of ASVs with their reverse complement where present. Because the v1.12.1 dada2 does not have this functionality (it has subsequently been added), a custom R script was used. The script uses the Biostrings package v2.54.0 (Pagès *et al.* 2019) for sequence handling and the DECIPHER package v2.14.0 (Wright 2016) to put all sequences into the same orientation.
5. Removal of ASVs which are likely to represent PCR chimeras, using the removeBimeraDenovo function with method = 'consensus'.

Reads from triplicate PCRs were combined to produce a single set of ASV counts per sample using a custom R script, utilising the data.table package (v1.12.8).

Taxonomy was assigned to each ASV using the Ribosomal Database Project (RDP) naïve Bayesian classifier method (Wang *et al.* 2007) implemented in dada2 as the assignTaxonomy() function. To ensure the robustness to database differences, classification was carried out using two publicly available training sets:

1. RefSeq-RDP16S_v3_May2018.fa.gz, obtained from https://zenodo.org/record/3266798 and consisting of the NCBI RefSeq 16S rRNA database (release 86) supplemented with extra sequences from the RDP database, compiled on 14 May 2018.
2. Silva_nr_v138_train_set.fa.gz, obtained from https://zenodo.org/record/3731176 and consisting of the Silva SSU 138 database of 16S rRNA (Quast *et al.* 2012).

For some analyses, ASVs were binned according to taxonomy by combining ASVs assigned the same genus into an OTU, and combining those not assigned to genera into an OTU if they shared 99% sequence similarity. To provide greater taxonomic detail about unidentified ASVs or OTUs and to stop removal of these during analyses that agglomerate to a given taxonomic level in downstream analysis, the taxonomy table was fully annotated to genus level using higher level assignments (e.g. ASV_13 was named "Family_Peptostreptococcaceae" at the genus level).

With reads representing only a fraction of the 16S gene, and of the larger bacterial genome, species-level assignment was not attempted across the full dataset (though see **Chapter 8** for pathogen-specific analysis below the genus level). Searching, parsing of BLAST results, and selection of taxonomic rank were performed using custom Python and R scripts. The R scripts used the data.table package v1.12.8 (Dowle & Srinivasan 2019) for data handling and the kmer package (v1.1.2) for OTU clustering. The resulting abundance tables, taxonomic classifications for binned (OTU) and unbinned (ASV) data, and sample metadata were loaded into R for analysis and plotting using phyloseq 1.30.0 (McMurdie & Holmes 2013) and ggplot2 (Wickham 2016).

Run-specific bias verification

To determine whether data could be analysed together, we produced an NMDS plot in phyloseq using the Bray-Curtis distance matrix to visualise variation across MiSeq runs (see **Supplementary Figure S1**). One particular MiSeq run, which happened to contain all swab samples collected in 2016 (MS15), clustered entirely separately from the remaining (guano)

samples. Though this could have been attributable to differences in the collection technique itself, our water controls also clustered by MiSeq run, leading us to believe that there were run-specific biases that would require these samples to be analysed as a subset apart from the remaining (guano) samples. Because only a single Gentoo Penguin cloacal swab from the Falkland Islands exists outside this set (seen in **Supplementary Figure S1** with two replicates run on MS33 and one on MS39 and verified as an outlier in swab-only NMDS plots, not shown), that sample was excluded from further swab-only analyses. Subsequent analyses were performed separately for "swab" and "guano" samples.

Microbiome community analyses

We used two different approaches to visualise differences in community composition between samples from a given site-host species-sampling season, for both the swab data set and guano data set. Ordinations were performed using phyloseq in R, employing both a count-based approach with the Bray-Curtis dissimilarity, and a compositional approach taking into account the phylogenetic relationships between sequences, the Phylogenetic Isometric Log-Ratio Transform (PhILR) with the philr R package (Silverman *et al.* 2017). Visual interpretation of these ordinations aided us in determining whether particular subsets (colonies, species, etc.) should also be analysed in greater depth. To determine the microbiome variation attributable to variables of host species, colony, season, region, or age (where applicable), we conducted a permutational ANOVA (PERMANOVA) using the adonis function in the vegan (Oksanen *et al.* 2019) package.

For each species, we sought to determine the core microbiome. Traditionally, a "core microbiome" consists of those microorganisms present in 100% of all hosts with given characteristics. We calculated the core microbiome of individual samples using a detection threshold of 0.001% and a prevalence threshold of 99% (i.e. a given OTU must be present in 99% of individuals with a relative abundance of at least 0.001%) in the microbiome R package v 1.8.0 (Lahti & Shetty 2017). Methodological differences meant that this analysis was performed on

subset groups within a given host species and sample type (i.e. Adélie Penguin swabs, Blue-Eyed Shag guano, Chinstrap Penguin swabs, Gentoo Penguin swabs, Gentoo Penguin guano, and King Penguin swabs).

Table 4. Table of guano and swab sampling regime for microbial analysis, with swab samples highlighted in **bold**. Species: Adélie Penguin (AP), Blue-Eyed Shag (BESH), Chinstrap Penguin (CP), Gentoo Penguin (GP), King Penguin (KP). Regions: Falkland/Malvinas Islands (FI), South Georgia (SG), South Shetland Islands (SH), Northeast Antarctic Peninsula (NEAP), Central Western Antarctic Peninsula (CWAP) as classified by the Antarctic Site Inventory (ASI).

Colony (ASI Site Code)	Region	Latitude	Longitude	Species	Samples (n) per Species and Breeding Season				
					2014/2015	2015/2016	2016/2017	2017/2018	Total (All Years)
Saunders Neck (SAUN)	FI	-51.31	-60.24	GP	10			20 (**1** swab)	30
Salisbury Plain (SALI)	SG	-54.05	-37.35	GP	13		15	21	49
St Andrew's Bay (STA)	SG	-54.44	-36.11	KP		**20**			20
Cooper Bay (COOP)	SG	-54.78	-35.81	GP	10				10
Barrientos Island (AITC)	SH	-62.41	-59.75	CP				20	20
				GP				20	20
Yankee Harbor (YANK)	SH	-62.53	-59.78	GP		**10**			10
Fort Point (FORT)	SH	-62.57	-59.57	GP				6	6
Baily Head (BAIL)	SH	-62.97	-60.5	CP		**10**			10
Kinnes Cove (KINN)	NEAP	-63.3	-56.48	AP		**14**			14
Hope Bay (HOPE)	NEAP	-63.38	-57.00	AP				6	6

				GP				26	26
Heroina Island (HERO)	NEAP	-63.40	-54.60	GP		10			10
Earle Island (EARL)	NEAP	-63.48	-54.78	GP		10			10
Brown Bluff (BROW)	NEAP	-63.53	-56.92	GP	10			20	30
Paulet Island (PAUL)	NEAP	-63.58	-55.78	BESH				6	6
Spert Island (SPRT)	CWAP	-63.85	-60.95	GP				1	1
Mikkelsen Harbor (MIKK)	CWAP	-63.90	-60.78	GP	10			20	30
Selvick Cove (SELV)	CWAP	-64.65	-62.57	GP				2	2
Georges Point (GEOR)	CWAP	-64.67	-62.67	CP		**6**			6
				GP		**6**			6
Cuverville Island (CUVE)	CWAP	-64.68	-62.63	GP				9	9
Danco Island (DANC)	CWAP	-64.73	-62.62	GP				6	6
Damoy Point (DAMO)	CWAP	-64.82	-63.53	GP				3	3
Jougla Point (JOUG)	CWAP	-64.83	-63.5	BESH				5	5
				GP				23	23
Neko Harbor (NEKO)	CWAP	-64.83	-62.55	GP				10	10
Port Lockroy (LOCK)	CWAP	-64.83	-63.5	GP				3	3
Bryde Island East (BRYE)	CWAP	-64.87	-63.03	GP				24	24

Almirante Brown Station (ALMI)	CWAP	-64.88	-62.87	GP				5	5
Booth Island (BOOT)	CWAP	-65.08	-64	AP		**3**			3
				CP		**6**			6
				GP		**6**		20	26
Petermann Island (PETE)	CWAP	-65.17	-64.17	GP				5	5
Moot Point (MOOT)	CWAP	-65.2	-64.1	GP		**5**		6	11
Totals (Species)					53 (GP)	106 (17 AP, 22 CP, 47 GP, 20 KP)	15 (GP)	287 (6 AP, 11 BESH, 20 CP, 250 GP)	461 (23 AP, 11 BESH, 42 CP, 365 GP, 20 KP)

Results and Discussion

16S rRNA Sequencing

A total of 6,764,059 reads were retrieved for 15,043 ASVs. After removing those that were not classified as Bacteria, 6,635,543 "included" reads and 11,383 ASVs remained (~1.9% removed). There were 497 classified genera in the included data set. 1,234,277 of original reads could not be classified at the Genus level (18.2%). Collectively, 2,421 OTUs were identified. For each host species and sample type, the total, minimum, maximum, and median reads per sample from the included data set are shown in **Table 2**, along with the number of unique ASVs and OTUs (classified genera or >99% sequence similarity). The mean number of included reads per sample was 14,394 across the study (range: 1,287-52,300), with a mean of 275 ASVs observed per sample (range: 65-733).

Inter-Specific Comparison of Bacterial Composition

Taxonomic analysis revealed the presence of 26 classified phyla in the gut microbiota of the data set. Of these, 14 phyla represented at least 2% of the relative abundance in at least one sample, signifying that 12 classified phyla were rarely encountered. For each species, the number of phyla present in at least one sample at greater than 2% abundance was: 7 in Adélie Penguins (6 distinct phyla in swabs, 4 in guano); 8 in Chinstrap Penguins (7 in swabs, 6 in guano); 12 in Gentoo Penguins (7 in swabs, 11 in guano); 8 in King Penguin swabs, and 6 in Blue-Eyed Shag guano. Of the full suite of classified phyla, four were uniquely found at single sites (Caldisericota, Hydrogenedentes, Latescibacterota, and Nitrospirota), with another three found in 3 or fewer colonies (Abditibacteriota, Dependentiae, Fibrobacterota). Guano samples consistently presented higher relative abundances of Fusobacteria than swabs, while higher proportions of Bacteroidota were obtained from swab samples relative to guano (**Figure 2** for comparisons).

Table 5. Sequencing and taxonomic classification details by host species and sample type after removal of non-bacterial sequences. ASVs (Amplicon Sequence Variants) depict unique sequences recovered from within the dataset, while OTUs (Operational Taxonomic Units) represent the number of taxa present after grouping by 99% sequence similarity (in the case of unclassified genera) or genus-level assignment. The Shannon-Weiner Index (H') listed represents merged samples for each host species/sample type.

Species (*n*) and Sample Type	**Total Reads**	**Min-Max Reads per Sample (Median)**	**Unique ASVs**	**Unique OTUs**	**Shannon Index (H')**
Adélie Penguin (23)	329,554	6,732-22,261 (15,298)	2,684	802	4.856
Swab (17)	219,371	6,732-19,544 (12,290)	2,021	629	4.479
Guano (6)	110,183	14,393-22,261 (18,187)	791	243	3.826
Blue-Eyed Shag (11)	274,151	13,885-34,370 (25,616)	1,888	564	5.291
Guano (11)					
Chinstrap Penguin (42)	605,844	3,124-33,314 (12,350)	3,792	1,227	4.798
Swab (22)	177,054	3,124-14,451 (7,248)	2,039	689	4.474
Guano (20)	428,790	13,279-33,314 (20,353)	2,013	654	4.154
Gentoo Penguin (365)	5,372,116	1,288-52,723 (12,015)	12,488	4,179	5.038
Swab (28)	234,926	2,974-18,620 (8,332)	2,524	896	4.660
Guano (337)	5,137,190	1,288-52,723 (12,544)	10,960	3,616	4.930
King Penguin (20)	182,393	4,377-15,967 (8,526)	1,850	725	4.537
Swab (20)					
Total (461)	6,635,543	1,287-52,300 (11,861)	11,383	2,421	

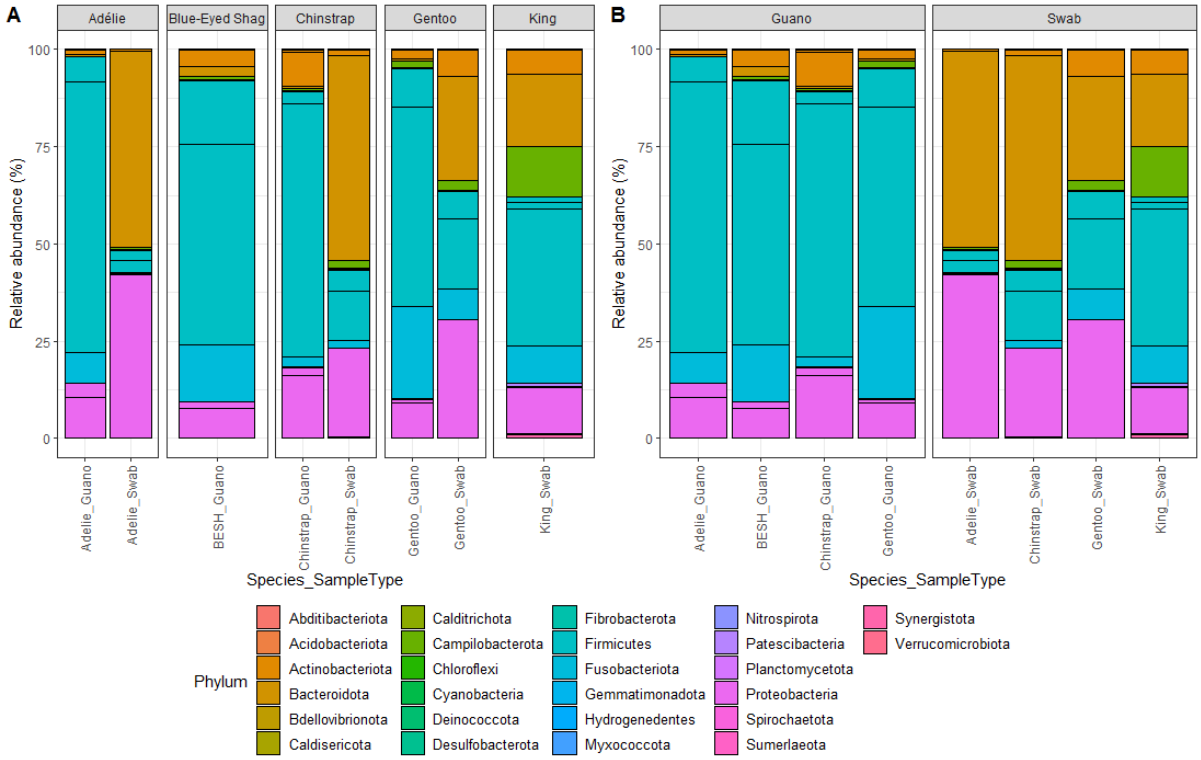


Figure 2. Relative abundance of bacterial phyla identified in the faecal and cloacal swab samples of four penguin species and Blue-Eyed Shags, with the same data depicted by (A) host species and (B) sample type, for comparison.

Diversity indices were calculated for total microbial composition of each host species and sample type. Swab samples among host species had similar combined Shannon indices (H', see **Table 2**), with Gentoo Penguins highest (H' = 4.66) followed by King Penguins (H' = 4.54), Adélie Penguins (H' = 4.48) and Chinstrap Penguins (H' = 4.47). Guano samples presented higher combined Shannon indices, with Blue-Eyed Shags highest (H' = 5.29), followed by Gentoo Penguins (H' = 4.93), Chinstrap Penguins (H' = 4.15) and Adélie Penguins (H' = 3.82). These numbers are higher than previously reported Shannon indices for Chinstrap Penguins at a single site in the South Shetland Islands, H' = 2.97 (Barbosa *et al.* 2016), even though our study had fewer Chinstrap Penguin reads, but might be explained by our inclusion of additional sites. Conversely, in King Penguins, where we only sampled one site, our diversity was higher than a study of two sites (Bird Island, South Georgia and Crozet Is. in the Indian

Ocean), with H' = 2.9 (Dewar *et al.* 2013), though we included more samples and reads. Gentoo Penguin diversity is more difficult to compare, as we had millions more reads and hundreds more samples than any previous study of this species, though diversity was previously reported as H' = 2.4 from four individuals on Bird Island (Dewar *et al.* 2013).

Attempts at ordination of the guano data set in fewer than 4 dimensions using NMDS with Bray-Curtis dissimilarity did not yield productive visualisations due to high stress >0.19. Grouping by host species and individual sampling sites for a given host in a given year (single species-site-year) groups were significant ($p = 0.001$), though PERMDISP values were also significant ($p = 0.001$) for these terms. Bacterial community variation among individuals within groups therefore seems to account for most of the variation in the guano data set. PCoA using a phylogenetically-informed PhILR transform also supports the minimal variation present by host in the guano data set (see **Figure 3**).

The higher inter-individual (within group) variation is expected in the context of the multifactorial complexity of the host microbiome. In a large-scale study of human populations ($n = 3948$), Falony *et al.* (2016) tested 503 metadata variables for covariation with the microbiome found 69 factors were significantly correlated with overall community variation, yet only explained 1.5-14.74% of genus abundance variation. This highlights the importance of considering the animal gut microbiome as a community system with numerous factors influencing its composition, not all of which are clearly understood.

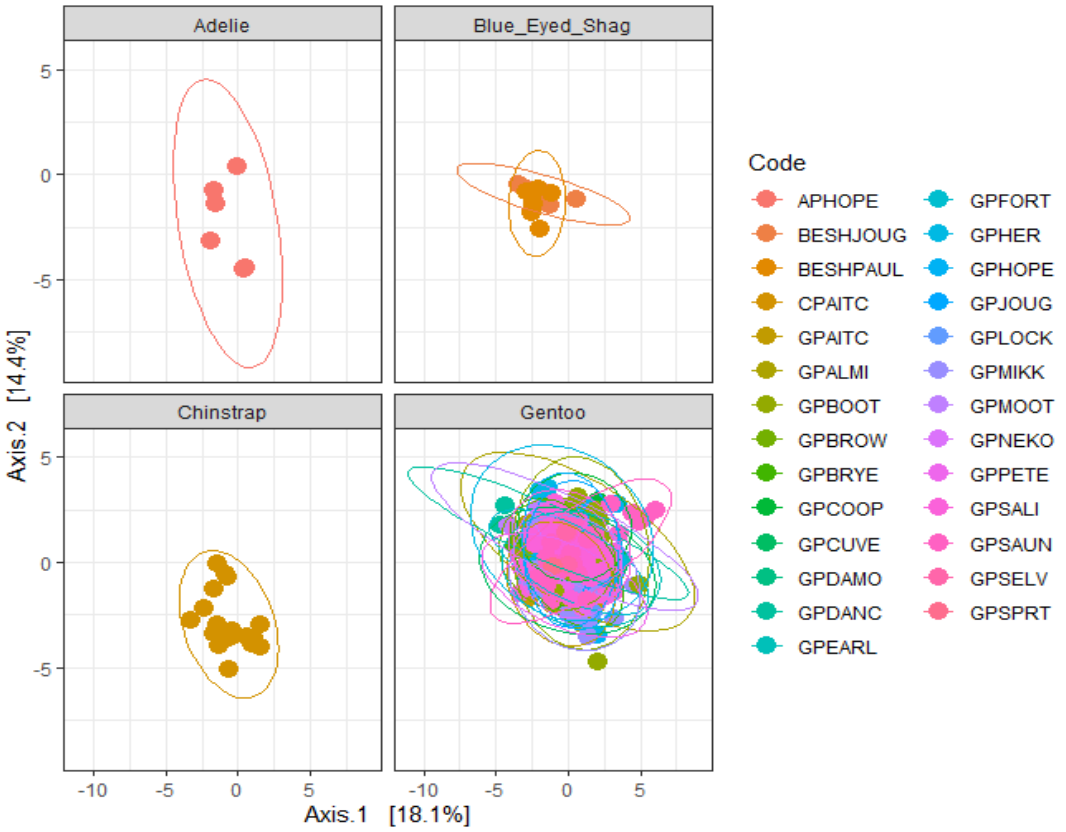


Figure 3. Principal coordinate analysis (PCoA) using the PhILR transform of the total microbiome of faecal samples (guano) obtained from three penguin species (AP = Adélie Penguin, CP = Chinstrap Penguin, GP = Gentoo Penguin) and Blue-Eyed Shags (BESH) at various breeding colonies, with split plots by host species. The four-letter Antarctic Site Inventory (ASI) code following the host species designator represents the colonies sampled, as listed in **Table 1**.

Where two species breed in physical proximity, local variation (Colony identity) has a greater effect than host species identity, allowing for cross-species commonalities in a given location. **Figure 4** shows that subtle differences can be noted within a host species when taking location into account in a broader context of biotic interaction. Colony accounts for 18.1% of differences (PERMANOVA using Bray-Curtis distance: F 7.9498, R^2 0.18088, p = 0.001, with beta dispersion non-significant p = 0.059) among mixed host microbial communities,

while Host Species accounts for 9.78% of differences (but does not meet PERMANOVA assumptions due to significant beta dispersion of the variance, p = 0.001).

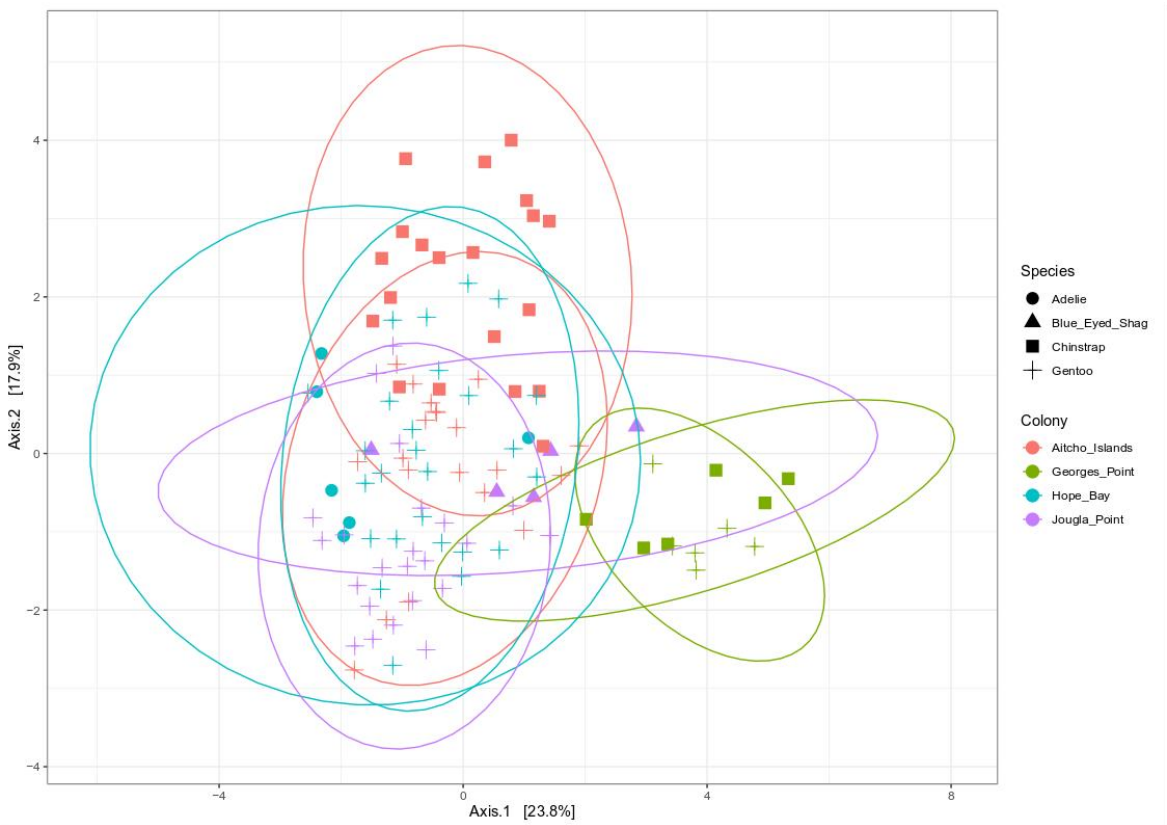


Figure 4. Principal coordinate analysis (PCoA) using the PhILR transform of the total microbiome of faecal samples (guano) obtained from mixed-host colonies, where Gentoo Penguins breed with either Adélie Penguins (Hope Bay, Northeast Antarctic Peninsula), Chinstrap Penguins (Aitcho Islands, South Shetland Islands and Georges Point, Central Western Antarctic Peninsula), or Blue-Eyed Shags (Jougla Point, Central Western Antarctic Peninsula). Colours represent each site, and shapes indicate host species. Ellipses group points of a given species at a single colony.

One potential explanation for the limited between-species differences in the guano data set is that a common diet has shaped the optimum composition for suitable digestion and energy acquisition within a given inshore environment. In **Figure 4**, we see that at the

class level, there are conserved components across species, and in some cases, like Georges Point (Central Western Antarctic Peninsula), Chinstrap Penguins and Gentoo Penguins at the same colony are more similar to each other in relative abundance than to other sites Chinstrap Penguins on Barrientos/Aitcho Is. (South Shetlands) or Gentoo Penguins even at Jougla Point, which is within the same Antarctic Peninsula region. This supports hypotheses of a localised adaptive gut response, though patterns are not consistent within a geographic region.

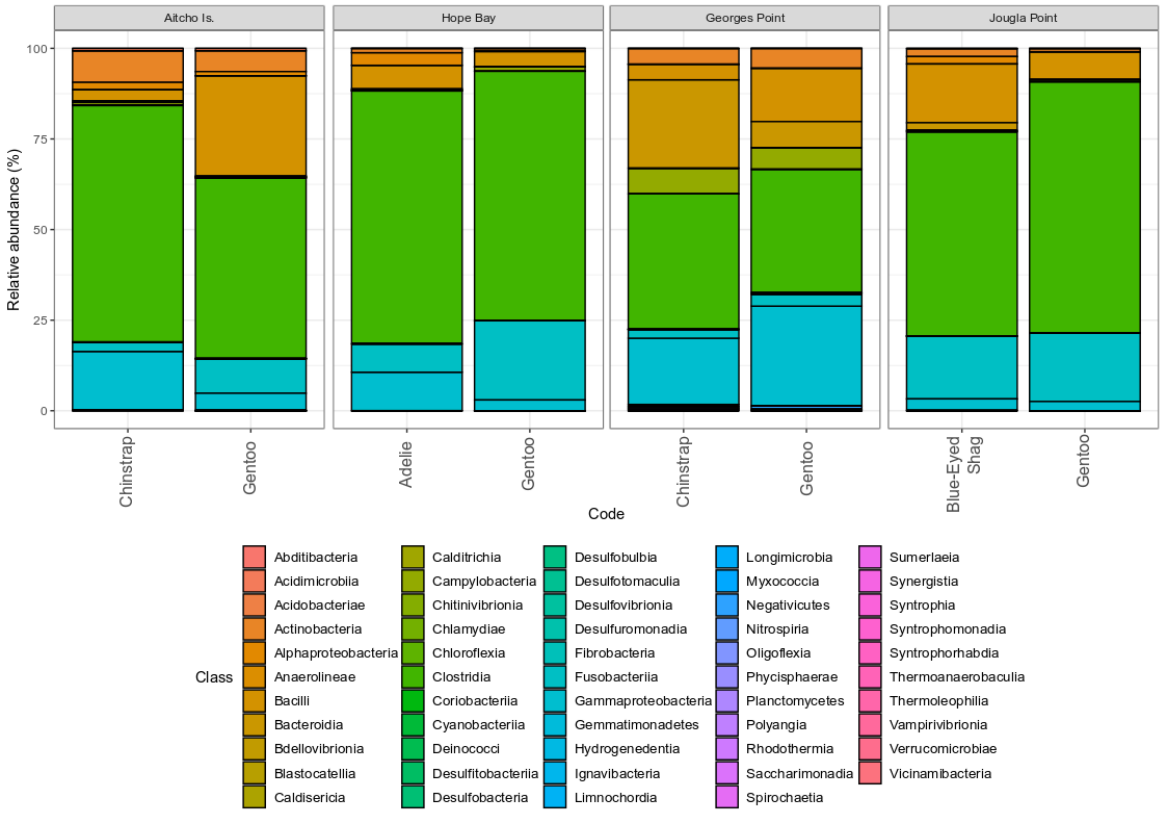


Figure 5. Relative abundance of bacterial classes identified in the faecal (guano) samples of four colonies shared by both Gentoo Penguins and another seabird host: Barrientos Is., Aitcho Is. in the South Shetland Islands (Chinstrap and Gentoo Penguins); Hope Bay, Northeast Antarctic Peninsula (Adélie and Gentoo Penguins); Georges Point, Central Western Antarctic Peninsula (Chinstrap and Gentoo Penguins); and Jougla Point, Central Western Antarctic Peninsula (Blue-Eyed Shags and Gentoo Penguins).

Niche partitioning is known to occur between Pygoscelid penguins breeding at the same site in terms of foraging times, distances, and depths (Miller *et al.* 2010; Trivelpiece *et al.* 1987; Wilson 2010), yet they largely consume *Euphausia superba*, Antarctic Krill (Pickett *et al.* 2018; Volkman *et al.* 1980). In years where krill availability is limited or krill size is less optimal, Pygoscelid penguins are known to prey-switch to greater proportions of fish, at an energetic foraging cost (Miller & Trivelpiece 2008), though the three species are not equally plastic in their responses to such shifts (McMahon *et al.* 2019; Miller *et al.* 2009; Polito *et al.* 2015). The diets of Blue-Eyed Shags are poorly studied in the Antarctic region we sampled, though our unpublished data indicate they include a limited diversity of fish as well as crustaceans. Elsewhere, in South Georgia, regurgitated Blue-Eyed Shag stomach contents have mainly consisted of nototheniid fish, cephalopods, and crustaceans (Wanless *et al.* 1992). Further study of Blue-Eyed Shag ecology and microbial sharing across seabird hosts on the Antarctic Peninsula is required, given some of the commonalities we found in their microbiomes.

Among swab sample microbiomes, Pygoscelid penguins present total microbiomes that are more similar within regions across species than between species alone. This seems to indicate that sympatry and interactions in a shared environment promote microbial connectivity across closely related hosts, and might also be linked to localised prey availability in that given year. The geographic proximity of the South Shetland Islands and Northeast Antarctic Peninsula appears to facilitate microbiome similarities between Pygoscelid hosts in those areas, while these three hosts group together in the Central Western Antarctic Peninsula region (see **Figure 6**).

King Penguins (sampled only in South Georgia) cluster distinctly from Pygoscelid hosts, though in our study design we cannot compare their microbiomes to any Pygoscelid penguins

on South Georgia to evaluate the role of shared environment relative to host identity within that region. The microbiomes of each of the host species are discussed in greater detail below.

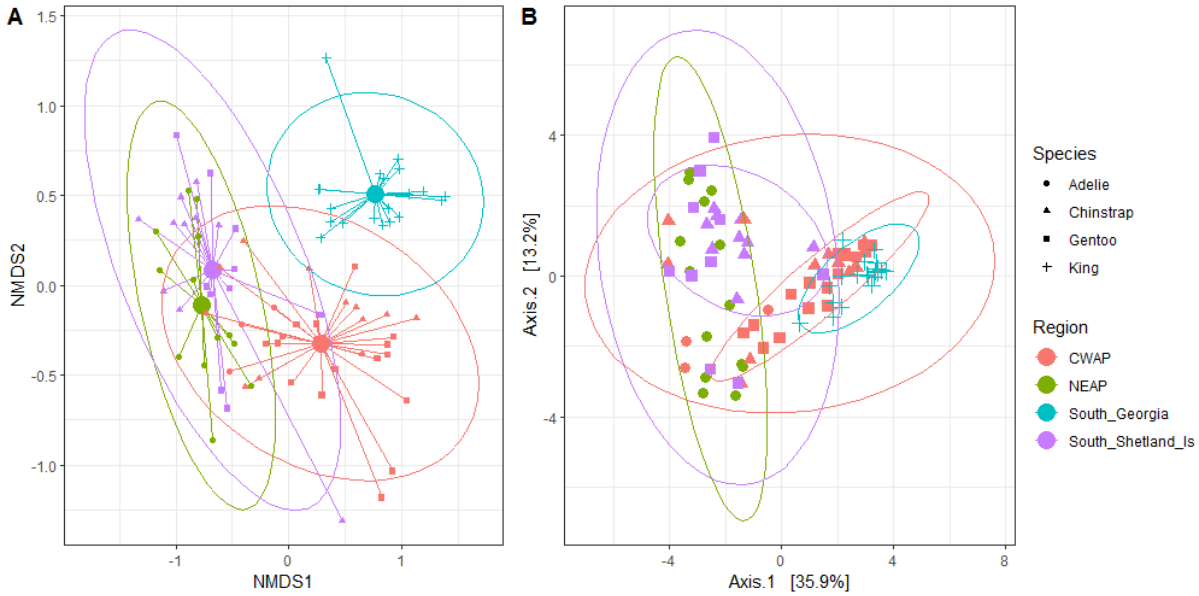


Figure 6. Ordination plots of the total microbiome of swab samples of four penguin species using (A) NMDS by Bray-Curtis dissimilarity with centroids displayed as larger circles and (B) PCoA with PhILR transformation, coloured by region: Central Western Antarctic Peninsula (CWAP), Northeast Antarctic Peninsula (NEAP), South Georgia, and South Shetland Islands. Shapes represent host penguin species.

Gentoo Penguins

Gentoo Penguins formed the bulk of samples in this study, and therefore constituted the highest number of reads, ASVs, and OTUs retrieved. The total microbiome of Gentoo Penguins, combined across samples of both cloacal swab and faecal types, is dominated by the phylum Firmicutes (61.6%), with Fusobacteria (19.3%) and Proteobacteria (12.2%) making notable contributions (**Figure 2** and **Figure 7**). This is in contrast to the dominance of Fusobacteria (55%) in a previous study of Gentoo Penguins (n = 4) from Bird Island, South Georgia (Dewar *et al.* 2013), though the three top phyla remain consistent. Among these

guano samples, these three top phyla compose >81% of bacterial relative abundance per site in a given year (mean 94.8% per colony), while among swabs, they range from 51.6%-78.6% of relative abundance per colony. Swab samples instead feature greater contributions of Bacteroidota and Actinobacterota, with swabs from Yankee Harbour in the South Shetland Islands distinct in their microbial composition (45.0% Bacteroidota, 38.6% Proteobacteria, 12.1% Firmicutes, 2.0% Actinobacteria and <1% Firmicutes). The differentiation between swab colonies (and therefore between the South Shetland Islands and more southerly colonies in the CWAP) is evident in the swab ordination plots (**Supplementary Figure S3**), where Yankee Harbour clusters separately by both NMDS (stress 0.129) and PCoA using a PhILR transformation.

We aimed to determine whether significant structuring of microbial communities existed between colonies and regions, mirroring host phylogeography in the Scotia Arc. The Gentoo Penguin guano data set presented the broadest sampling range with which to examine this question, with colonies in the Falkland Islands, South Georgia, South Shetland Islands and two Antarctic Peninsula regions sampled, and some across multiple years.

We hypothesised that given differences in both host population genotypes (Clucas *et al.* 2018) and bacterial recognising immune gene (Toll-like receptor) diversity (Levy *et al.* 2020), that strong differentiation would be observable between the Falkland Islands (located north of the Polar Front, a strong barrier to gene flow), and other geographic regions to the south. It has recently been reported that the abiotic marine ecological niches (climate variables) of each Gentoo Penguin habitat are distinct and non-overlapping, a contributing factor to the divergent trajectory of northern and southern host clades in the Atlantic and Antarctic Sector (Pertierra *et al.* 2020). We also expected that on top of host genetic differences, the distinctive host environments would lead to robust differences in microbial community composition.

Instead, among these guano samples, we found that variation between individuals within a colony was greater than between group variation (consistently >75% for all tested variables). Gentoo Penguins appear to have relatively conserved bacterial microbiomes. Dispersion across the data set was too significant for colony, region, and site-year, so PERMANOVA assumptions were not met.

The only variable that significantly accounts for differences between Gentoo Penguin groups is sampling season (year), although this only accounts for 1.2% of variation in the data set (PERMANOVA using Bray-Curtis distance: F 4.2678, R^2 0.01258, p = 0.001). Nonetheless, this fits with the differences apparent in relative abundances (**Figure 7**) and ordinations (**Figure 8**) for the few sites that were sampled two or three times in the course of the study (Dec 2014-Jan 2018), particularly for Saunders Neck, Falkland Islands/Malvinas (GPSAUN), Salisbury Plain, South Georgia (GPSALI), and Brown Bluff, Northeast Antarctic Peninsula (GPBROW), while the southern site of Mikkelsen Harbor did not experience notable changes between those years. Gentoo Penguin colonies are known to have unstable population dynamics, with inter-annual fluctuations in breeding success, diet, and phenology (Baylis *et al.* 2012; Black *et al.* 2018; Black 2016; Coria *et al.* 2000; Dunn *et al.* 2019; Juáres *et al.* 2013; Miller *et al.* 2009; Pütz *et al.* 2001; Williams 1990; Xavier *et al.* 2017). While we are not able to directly link the year-to-year microbiome differences seen in our study to environmental variables, factors such as air and water temperature, precipitation, the location of the Polar Front, and prey availability may play a role in shaping microbial community shifts. Future studies that evaluate the influence of diet on the microbiome more directly (analyses conducted on the same samples or individuals) may help clarify the role of prey availability on individual and colony-wide microbial community ecology, especially considering the foraging plasticity of Gentoo Penguins.

Figure 7. Relative abundance of bacterial phyla (A) and classes (B) in Gentoo Penguin samples obtained at a given colony in a given year, grouped by sample type (Guano or Swab) and geographic region: Falkland/Malvinas Islands (FI), South Georgia (SG), South Shetland Islands (SH), Northeast Antarctic Peninsula (NEAP), and Central Western Antarctic Peninsula (CWAP).

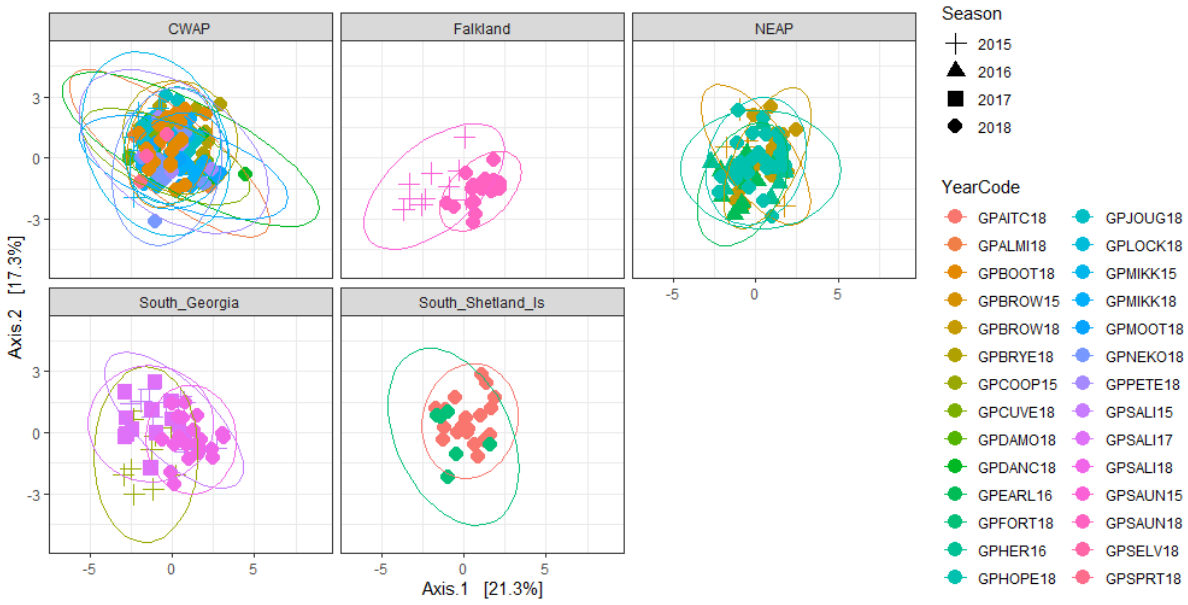


Figure 8. Principal Coordinates Analyses (PCoA) using the PhILR transformation of the total microbiome in Gentoo Penguin guano samples, shown by geographic region, with colours depicting single site-year combinations and shapes representing sampling season.

Adélie Penguins

For Adélie Penguins, the most abundant Phyla were Bacteroidota, Firmicutes, and Proteobacteria (**Figure 2**, **Supplementary Figure S4**). Combined relative abundance of all Adélie Penguin samples (across sample types) was 37.9% Proteobacteria, 33.2% Bacteroidota, and 24.6% Firmicutes. The relative abundance of these three leading phyla varied substantially between sites and sample types, with guano samples from Hope Bay (n = 6) dominated by Firmicutes (Class Clostridia), while Adélie Penguin swabs from Booth Island (n = 3) and Kinnes Cove (n = 14) were either Bacteroidota (Class Bacteroidia) or Proteobacteria (Class Gammaproteobacteria)-dominated depending on the individual. Colony did not significantly explain variation between Adélie Penguin swabs (PERMANOVA with Bray-Curtis distance: p = 0.126, using PhILR transform p = 0.42).

Individual variation in this host seems to fall in line with the microbial composition reported from cloacal swab samples across Ross Sea colonies of Adélie Penguins in a previous study by Banks *et al.* (2009). In that study, the microbiota of single individuals from six colonies were cloned, cultured, and sequenced, resulting in widely varied proportions of Bacteroidetes, Proteobacteria, Actinobacteria, and Firmicutes. Though our methods are not directly comparable, collectively, the limited sample sizes and the inter-individual differences in Adélie Penguin bacterial microbiomes seen in our dataset and in Banks *et al.* (2009) should encourage further studies of this host species in a systematic sampling effort across its circumpolar range.

Blue-Eyed Shags

Blue-Eyed Shag bacterial communities were generally dominated by Firmicutes (overall relative abundance of 68.2%) with 13.1% Fusobacteria, 9.6% Proteobacteria, and 5.3% Actinobacteria (**Figure 2**), though one sample from each colony presented higher relative abundances of Fusobacteria (44% and 75%), see **Supplementary Figure S5**.

Ordinations using both non-metric multidimensional scaling (NMDS, Bray-Curtis dissimilarity, stress 0.107) and Principal coordinate analysis (PCoA) with a phylogenetic isometric log-ratio (PhILR) transformation show similar patterns of sample clustering by colony (**Supplementary Figure S6**). PERMANOVA analyses revealed that Blue-Eyed Shag colony differences were significant and accounted for 18.6% of variation in the case of Bray-Curtis dissimilarity (F 2.0578, R^2 0.18609, p = 0.009) and 20.7% of variation in the PhILR compositional model (F 2.3436, R^2 0.2066, p = 0.02). These two sites are approximately 450 km apart by coastal/marine routing (shortest flying distance ~400 km). Since this is the first community-based evaluation of Antarctic Shag microbiota to our knowledge, further sampling is needed to determine whether community similarities are present along a finer geographic scale, and to evaluate whether host phylogeny (sub-species) and host population genetic

structure (currently unknown for Antarctica) correlates with microbial community composition in shags.

Chinstrap Penguins

Chinstrap Penguin communities as a whole (all sample types) were composed of 44.0% Firmicutes, 24.7% Bacteroidota, 20.0% Proteobacteria, and 6.4% Actinobacteria (**Figure 2**). There were substantial differences between sample types (see **Supplementary Figure S6**), in that Bacteroidota (Class Bacteroidia), which was dominant in many swab samples, did not represent more than 2% of relative abundance in any Chinstrap Penguin guano samples (see **Supplementary Figure S7**). Instead, Firmicutes (Class Clostridia) were generally dominant in Chinstrap Penguin guano samples, with some exceptions where Actinobacteria or Proteobacteria were more abundant. Colonies where swab samples were obtained did not differ significantly in their alpha diversity measures, though PERMANOVA revealed that beta diversity differed significantly (p=0.002) among colonies, with the colony term accounting for 24.3% of variation under Bray-Curtis dissimilarity (F 3.0492, R^2 0.24298, p = 0.002) and 27.8% using the PhILR transformation (F 3.651, R^2 0.27762, p = 0.002). Ordination plots are shown in **Supplementary Figure S8**.

A previous study of 53 Chinstrap Penguin cloacal swabs carried out on Vapour Col, Deception Island found that while adults and chicks shared the same bacterial phyla, significant differences in evenness were present between adults and chicks, with chick microbiota dominated primarily by Firmicutes (Class Clostridia) with low representation of Bacteroidota and Proteobacteria than adults (Barbosa *et al.* 2016). This finding contrasts with the results of our sampling of adult/chick pairs from Baily Head, Deception Island, located under 10 km away, where age did not explain differences in composition (PERMANOVA by Bray-Curtis, p = 0.20). Additionally, Firmicutes represented 60% of all reads at Vapour Col, but represented only 5.7% of total reads for our samples from Baily Head, where Firmicutes was

present at abundances of 2.1-13.6% per individual. Aside from expected inter-individual variation, these differences could be accounted for in multiple ways, given that Barbosa *et al.* (2016) conducted sampling in 2008-2009 (inter-annual differences), using different primers (V1-V3 region) and sequencing platform (Roche 454 GS FLX Titanium). The chicks they sampled were 7 days old, whereas the age at sampling of our chicks was unknown, but may be similar or younger, given that other nests had eggs and hatching events on the sampling date.

King Penguins

The microbiota of King Penguins structure distinctly from all other host species sampled (see **Figure 6**). The leading phyla by total relative abundance were Firmicutes (38.4%), Bacteroidota (18.3%), Proteobacteria (13.2%), Campilobacterota (9.2%), Actinobacteriota (7.9%), and Fusobacteria (7.1%), see **Supplementary Figure S10**. Campilobacterota are less prevalent overall in other host species, where they make up less than 1.3% of total overall relative abundance per species. Only a single Gentoo Penguin colony in the Falkland Islands had higher relative abundance of this phylum (17.8%) during one sampling year, dropping to less than 1% in a subsequent sampling season. Other sites in South Georgia had very low abundance of this phylum, which may indicate that King Penguins in South Georgia are either normally host to these bacteria at relatively consistent abundances across individuals, or this is a group whose abundance in a host fluctuates significantly over time, rather than being exclusively tied to milder climate regions.

There were no statistically significant differences in taxa richness or evenness between King Penguin adults and chicks by Welch's two-tailed t-test (Observed alpha diversity: t = -0.84816, df = 11.734, p = 0.413; Chao1 index: t = -0.97334, df = 15.168, p = 0.346; Shannon index: t = -1.0261, df = 17.994, p = 0.3185). However, age does lead to significant structuring of bacterial abundance profiles by PERMANOVA (Bray-Curtis: F 2.597, R^2 = 0.12609, p = 0.001; PhILR F 4.6486, R^2 = 0.20525, p = 0.001), with ordinations by Bray-Curtis

NMDS (stress 0.112) and PCoA using PhILR transformation shown in **Supplementary Figure S11**.

Core Microbiome

The core microbiome of King Penguins at a single site in South Georgia was the most taxa-rich, with 17 genera present in 100% of individuals (**Supplementary Figure S12A**). 39 genera were present in at least 50% of individuals. The core microbiome of Adélie Penguin swabs consisted of 10 genera present in 100% of individuals (**Supplementary Figure S12B**), while that of Chinstrap Penguin swabs consisted of 9 genera (all common to the Adélie Penguin swab core microbiome, **Supplementary Figure S12C**). Eight genera were shared across King, Adélie, and Chinstrap Penguin core microbiomes (*Flavobacterium, Chryseobacterium, Campylobacter, Carnobacterium,* a genus in the Family Peptostreptococcales-Tissierellales*, Fusobacterium, Halomonas,* and *Psychrobacter*). Gentoo Penguin swab samples had a less diverse core microbiome, on the other hand, with only three genera present in ≥ 99% of samples: *Chryseobacterium*, *Fusobacterium*, and *Psychrobacter*, all of which were common to the other species' swab samples (**Supplementary Figure S12D**). As these data sets included both chicks and adults, it is interesting to note the high diversity retained in the swab samples, even across age groups, indicating that these core taxa are retained across life stages.

The core microbiome of Gentoo Penguin guano was more restricted, containing just two genera: sequences in the genus *Clostridium* sensu stricto 1 and an OTU in the Family Peptostreptococcaceae of undetermined genus (**Supplementary Figure S12E**). Taking a broader look at taxa with widespread prevalence in Gentoo Penguins, swab samples were still more diverse, with 27 genera present across 50% of Gentoo Penguin swabs, with only 15 genera present in more than 50% of Gentoo Penguin guano samples. The prevalent microbiome in Gentoo Penguin guano did not have many similarities with swab samples.

Across the two types of Gentoo Penguin samples, only 7 of the genera were shared across at least 50% of samples of each type, with *Clostridium* sensu stricto 1 ranking far less prevalently in swabs (57% vs. 99% in guano). The core microbiome of Blue-Eyed Shag guano had similarities to Gentoo Penguin guano in that it only contained three genera, and had *Clostridium* sensu stricto 1 in common (**Supplementary Figure S12F**).

The narrow findings in Gentoo Penguin guano may be due to the large sample size of this data set taken across a wider geographic range than any other species, limiting the signal of commonality across this host. However, this may be more reflective of the true core microbiome of the Atlantic sector of Gentoo Penguins than for other species, which were analysed in relatively low numbers in a fraction of their full range. The true core may therefore be much narrower than that detected in this data set. We would encourage further research of how geospatial range affects the core in widespread species like King Penguins, or penguins that disperse in the non-breeding season. Other studies in migratory birds have indicated that environmental differences between breeding and wintering areas can alter microbial community beta diversity and abundance, as well as pathogenic bacteria occurrence, though the core microbiome remains unchanged (Wu *et al.* 2018). Further study would enable baseline understanding of what "normal" variation is expected, as a proxy for indications of environmental disturbance (Bahrndorff *et al.* 2016).

Conclusion

There is debate as to whether a shared environment, diet, or a host's genotype is more influential in shaping vertebrate gut microbiomes (Muegge *et al.* 2011; Rothschild *et al.* 2018; Spor *et al.* 2011; Youngblut *et al.* 2019). Among studies that incorporate wild avian study species, host phylogeny and host identity appear to be strongly predictive of community structure (Waite & Taylor 2014), even in cases where a host species is experiencing habitat change (San Juan *et al.* 2020) or is critically endangered (Waite *et al.* 2012). Penguins and

flying seabirds of the Sub-Antarctic and Antarctic are under threat from human disturbance, pollution, and fisheries competition, in addition to rapid climate change (Cimino *et al.* 2016; Trathan *et al.* 2015; Trivelpiece *et al.* 2011). Adélie, Chinstrap, Gentoo, and King Penguins all face challenges to their varied habitats and ecological niches as a result (Lynch *et al.* 2010; Lynch *et al.* 2012), and their sentinel status as marine mesopredators may allow us to explore individual, population, and species-level responses to change through their microbial communities.

In this study we expand knowledge of the constituents of Sub-Antarctic and Antarctic seabird microbiota, and find that localised interactions between host species may facilitate cross-host microbiome commonalities. Within the geospatially dispersed Gentoo Penguin, differences across regions are less apparent, with individual differences exceeding between-population structures. Despite strong host genetic differentiation and limited admixture between archipelagos, inter-annual variation appears to play a more important role in structuring local population-level responses than host genotype. This localised intra- and inter-specific variation has implications for conservation management of Important Bird Areas in the Antarctic and for epidemiological monitoring in a threatened ecosystem.

Acknowledgements

We thank Quark Expeditions, Cheeseman's Ecology Safaris, Oceanwide Expeditions, and J. Poncet for their assistance with fieldwork logistics. We thank Anni Djurhuus, Caitlin E. Black, Gemma V. Clucas, Michael J. Polito, Heather J. Lynch, Michael B. Schrimpf, Rachael Herman, Maureen Lynch, and Casey Youngflesh for assisting with sample collection. We thank Kirstie McLoughlin, Patrik Flammer, Rob Dixon, Sophie Hedges, Frances Colles, for sharing MiSeq runs and assisting with the sequencing workflow, and Amanda Williams and Rebecca Busby for overseeing sequencing runs. Molecular work was supported by Penguin Lifelines, the U.S. Department of Veteran's Affairs G.I. Bill, and Stony Brook University. ALS received

funding from the BBSRC and Bertarelli Foundation. HL's work was supported by a University of Oxford Clarendon Fund Scholarship.

Author Contributions

HL, TH, and ALS developed and designed the study. TH obtained permits and secured logistics. HL and TH collected samples in addition to collaborators listed in acknowledgments. HL extracted all DNA and performed all PCR amplifications, quality control, and pooling. HL and SGP performed Illumina MiSeq library preparation, quality control, and sequencing with technical support from collaborators listed in acknowledgments. SGP processed all raw sequence data and created custom scripts for demultiplexing, pre-processing reads, and performing automated taxonomic assignment. HL conducted all analyses and wrote the paper with feedback from all authors.

Ethics Statement

Animal handling protocols followed Scientific Committee on Antarctic Research Code of Conduct on the use of Animals for Scientific Purposes in Antarctica. All sampling in the Antarctic Treaty Area was carried out in accordance with United Kingdom Home Office guidelines, under permits granted by the United Kingdom Foreign and Commonwealth Office (permits: S3-23/2013, S7 28/2013, S3 04/2014, S7 03/2014, 33/2016, 34/2016, 37/2017), in South Georgia under permits granted by the Government of South Georgia and the South Sandwich Islands (SCI-2014-017, RAP-2015-018, RAP-2016-035 and RAP-2017-034), and in the Falkland Islands under Falkland Islands Government Environmental Planning Department permits (R16.2014, R15 2014, and R04 2017), with ethical approval from the University of Oxford Animal Welfare and Ethical Review Board.

Conflict of Interest Statement

The authors declare no conflict of interest.

Data Availability Statement

Custom scripts, genetic sequences, sample metadata, abundance tables, and taxonomy assignment tables will be made available prior to publication on a public repository.

References

Bahrndorff S, Alemu T, Alemneh T, Lund Nielsen J (2016) The Microbiome of Animals: Implications for Conservation Biology. *International Journal of Genomics* **2016**, 5304028.

Banks JC, Cary SC, Hogg ID (2009) The phylogeography of Adelie penguin faecal flora. *Environ Microbiol* **11**, 577-588.

Barbosa A, Balague V, Valera F*, et al.* (2016) Age-Related Differences in the Gastrointestinal Microbiota of Chinstrap Penguins (*Pygoscelis antarctica*). *PLoS One* **11**, e0153215.

Baylis AMM, Zuur AF, Brickle P, Pistorius PA (2012) Climate as a driver of population variability in breeding Gentoo Penguins *Pygoscelis papua* at the Falkland Islands. *Ibis* **154**, 30-41.

Benskin CMH, Rhodes G, Pickup RW, Wilson K, Hartley IR (2010) Diversity and temporal stability of bacterial communities in a model passerine bird, the zebra finch. *Mol Ecol* **19**, 5531-5544.

Black C, Collen B, Lunn D*, et al.* (2018) Time-lapse cameras reveal latitude and season influence breeding phenology durations in penguins. *Ecol Evol* **8**, 8286-8296.

Black CE (2016) A comprehensive review of the phenology of *Pygoscelis* penguins. *Polar Biology*, 1-28.

Bonnedahl J, Broman T, Waldenstrom J*, et al.* (2005) In search of human-associated bacterial pathogens in Antarctic wildlife: report from six penguin colonies regularly visited by tourists. *Ambio* **34**, 430-432.

Callahan BJ, McMurdie PJ, Rosen MJ*, et al.* (2016) DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods* **13**, 581-583.

Caporaso JG, Lauber CL, Walters WA*, et al.* (2011) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences* **108**, 4516-4522.

Cimino MA, Lynch HJ, Saba VS, Oliver MJ (2016) Projected asymmetric response of Adelie penguins to Antarctic climate change. *Sci Rep* **6**, 28785.

Clements JF, Schulenberg TS, Iliff MJ*, et al.* (2019) The eBird/Clements Checklist of Birds of the World: v2019.

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

Clucas GV, Younger JL, Kao D*, et al.* (2018) Comparative population genomics reveals key barriers to dispersal in Southern Ocean penguins. *Mol Ecol* **27**, 4680-4697.

Cole TL, Dutoit L, Dussex N*, et al.* (2019) Receding ice drove parallel expansions in Southern Ocean penguins. *Proc Natl Acad Sci U S A*.

Coria N, Libertelli M, Casaux R, Darrieu C (2000) Inter-Annual Variation in the Autumn Diet of the Gentoo Penguin at Laurie Island, Antarctica. *Waterbirds: The International Journal of Waterbird Biology* **23**, 511-517.

Dewar ML, Arnould JPY, Dann P*, et al.* (2013) Interspecific variations in the gastrointestinal microbiota in penguins. *Microbiologyopen* **2**, 195-204.

Dewar ML, Arnould JPY, Krause L*, et al.* (2014) Influence of Fasting during Moult on the Faecal Microbiota of Penguins. *PLoS One* **9**.

Dowle M, Srinivasan A (2019) data.table: Extension of `data.frame`. R package version 1.12.8.

Dunn MJ, Forcada J, Jackson JA*, et al.* (2019) A long-term study of gentoo penguin (*Pygoscelis papua*) population trends at a major Antarctic tourist site, Goudier Island, Port Lockroy. *Biodiversity and Conservation* **28**, 37-53.

Edgar RC (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods* **10**, 996-998.

Falony G, Joossens M, Vieira-Silva S*, et al.* (2016) Population-level analysis of gut microbiome variation. *Science* **352**, 560-564.

Gavryushkina A, Heath TA, Ksepka DT*, et al.* (2017) Bayesian Total-Evidence Dating Reveals the Recent Crown Radiation of Penguins. *Systematic Biology* **66**, 57-73.

Grond K, Sandercock BK, Jumpponen A, Zeglin LH (2018) The avian gut microbiota: community, physiology and function in wild birds. *Journal of Avian Biology* **49**, e01788.

Harris EV, de Roode JC, Gerardo NM (2019) Diet–microbiome–disease: Investigating diet's influence on infectious disease resistance through alteration of the gut microbiome. *PLOS Pathogens* **15**, e1007891.

Humphries GRW, Naveen R, Schwaller M*, et al.* (2017) Mapping Application for Penguin Populations and Projected Dynamics (MAPPPD): data and tools for dynamic management and decision support. *Polar Record* **53**, 160-166.

Juáres MA, Santos MM, Negrete J*, et al.* (2013) Better late than never? Interannual and seasonal variability in breeding chronology of gentoo penguins at Stranger Point, Antarctica. *Polar Research* **32**, 18448.

Lahti L, Shetty S (2017) Microbiome R package version 1.8.0.

Lane DJ (1991) 16S/23S rRNA sequencing. In: *Nucleic acid techniques in bacterial systematics* (eds. Stackebrandt E, Goodfellow M), pp. 115-175. Wiley, Chichester, United Kingdom.

Lee WY, Cho H, Kim M*, et al.* (2019) Faecal microbiota changes associated with the moult fast in chinstrap and gentoo penguins. *PLoS One* **14**, e0216565.

Leotta G, Cerda R, Coria NR, Montalti D (2001) Preliminary studies on some avian diseases in Antarctic birds. *Polish Polar Research* **22**, 227-231.

Leotta GA, Chinen I, Vigo GB, Pecoraro M, Rivas M (2006) Outbreaks of avian cholera in Hope Bay, Antarctica. *Journal of Wildlife Diseases* **42**, 259-270.

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**, 1834-1853.

Levy H, Fiddaman SR, Vianna JA*, et al.* (2020) Evidence of pathogen-induced immunogenetic selection across the large geographic range of a wild seabird. *Mol Biol Evol* **6**, 1708-1726.

Lynch HJ (2013) Gentoo Penguin (*Pygoscelis papua*). In: *Penguins: Natural History and Conservation* (eds. Borboroglu PG, Boersma PD), pp. 73-88. University of Washington Press, Seattle, WA.

Lynch HJ, Fagan WF, Naveen R (2010) Population trends and reproductive success at a frequently visited penguin colony on the western Antarctic Peninsula. *Polar Biology* **33**, 493-503.

Lynch HJ, LaRue MA (2014) First global census of the Adélie Penguin. *The Auk* **131**, 457-466.

Lynch HJ, Naveen R, Fagan WF (2008) Censuses of penguin, Blue-eyed Shag *Phalacrocorax atriceps* and Southern Giant Petrel *Macronectes giganteus* populations on the Antarctic Peninsula, 2001-2007. *Marine Ornithology* **36**, 83-97.

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.

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011* **17**, 3.

McMahon KW, Michelson CI, Hart T*, et al.* (2019) Divergent trophic responses of sympatric penguin species to historic anthropogenic exploitation and recent climate change. *Proc Natl Acad Sci U S A* **116**, 25721-25727.

McMurdie PJ, Holmes S (2013) phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* **8**, e61217.

Miller AK, Kappes MA, Trivelpiece SG, Trivelpiece WZ (2010) Foraging-Niche Separation of Breeding Gentoo and Chinstrap Penguins, South Shetland Islands, Antarctica. *The Condor* **112**, 683-695.

Miller AK, Karnovsky NJ, Trivelpiece WZ (2009) Flexible foraging strategies of gentoo penguins *Pygoscelis papua* over 5 years in the South Shetland Islands, Antarctica. *Marine Biology* **156**, 2527-2537.

Miller AK, Trivelpiece WZ (2008) Chinstrap penguins alter foraging and diving behavior in response to the size of their principle prey, Antarctic krill. *Marine Biology* **154**, 201-208.

Miller ET, Svanbäck R, Bohannan BJM (2018) Microbiomes as Metacommunities: Understanding Host-Associated Microbes through Metacommunity Ecology. *Trends in Ecology & Evolution* **33**, 926-935.

Muegge BD, Kuczynski J, Knights D*, et al.* (2011) Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science (New York, N.Y.)* **332**, 970-974.

Naveen R, Forrest SC, Dagit RG*, et al.* (2009) Censuses of penguin, blue-eyed shag, and southern giant petrel populations in the Antarctic Peninsula region, 1994–2000. *Polar Record* **36**, 323-334.

Oksanen J, Blanchet FG, Friendly M*, et al.* (2019) Vegan: Community Ecology Package. R package version 2.5-6. https://CRAN.R-project.org/package=vegan.

Oliphant TE (2006) *A guide to NumPy (Vol. 1)* Trelgol Publishing, USA.

Pagès H, Aboyoun P, Gentleman R, DebRoy S (2019) Biostrings: Efficient manipulation of biological strings.

Pertierra LR, Segovia NI, Noll D*, et al.* (2020) Cryptic speciation in gentoo penguins is driven by geographic isolation and regional marine conditions: Unforeseen vulnerabilities to global change. *Diversity and Distributions* **26**, 958-975.

Pickett EP, Fraser WR, Patterson-Fraser DL*, et al.* (2018) Spatial niche partitioning may promote coexistence of *Pygoscelis* penguins as climate-induced sympatry occurs. *Ecol Evol* **8**, 9764-9778.

Polito MJ, Trivelpiece WZ, Patterson WP*, et al.* (2015) Contrasting specialist and generalist patterns facilitate foraging niche partitioning in sympatric populations of *Pygoscelis* penguins. *Marine Ecology Progress Series* **519**, 221-237.

Prum RO, Berv JS, Dornburg A*, et al.* (2015) A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature* **526**, 569-573.

Pütz K, Ingham RJ, Smith JG, Croxall JP (2001) Population trends, breeding success and diet composition of gentoo *Pygoscelis papua*, magellanic *Spheniscus magellanicus* and rockhopper *Eudyptes chrysocome* penguins in the Falkland Islands. A review. *Polar Biology* **24**, 793-807.

Quast C, Pruesse E, Yilmaz P*, et al.* (2012) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* **41**, D590-D596.

Rothschild D, Weissbrod O, Barkan E*, et al.* (2018) Environment dominates over host genetics in shaping human gut microbiota. *Nature* **555**, 210-215.

San Juan PA, Hendershot JN, Daily GC, Fukami T (2020) Land-use change has host-specific influences on avian gut microbiomes. *The ISME Journal* **14**, 318-321.

Silverman JD, Washburne AD, Mukherjee S, David LA (2017) A phylogenetic transform enhances analysis of compositional microbiota data. *eLife* **6**, e21887.

Spor A, Koren O, Ley R (2011) Unravelling the effects of the environment and host genotype on the gut microbiome. *Nature Reviews Microbiology* **9**, 279-290.

Tanton JL, Reid K, Croxall JP, Trathan PN (2004) Winter distribution and behaviour of gentoo penguins *Pygoscelis papua* at South Georgia. *Polar Biology* **27**, 299-303.

Tarroux A, Lydersen C, Trathan PN, Kovacs KM (2018) Temporal variation in trophic relationships among three congeneric penguin species breeding in sympatry. *Ecol Evol* **8**, 3660-3674.

Tasnim N, Abulizi N, Pither J, Hart MM, Gibson DL (2017) Linking the Gut Microbial Ecosystem with the Environment: Does Gut Health Depend on Where We Live? *Frontiers in Microbiology* **8**.

Thouzeau C, Froget G, Monteil H, Le Maho Y, Harf-Monteil C (2003) Evidence of stress in bacteria associated with long-term preservation of food in the stomach of incubating king penguins (Aptenodytes patagonicus). *Polar Biology* **26**, 115-123.

Trathan PN, Garcia-Borboroglu P, Boersma D*, et al.* (2015) Pollution, habitat loss, fishing, and climate change as critical threats to penguins. *Conservation Biology* **29**, 31-41.

Trivelpiece WZ, Hinke JT, Miller AK*, et al.* (2011) Variability in krill biomass links harvesting and climate warming to penguin population changes in Antarctica. *Proc Natl Acad Sci U S A* **108**, 7625-7628.

Trivelpiece WZ, Trivelpiece SG, Volkman NJ (1987) Ecological Segregation of Adelie, Gentoo, and Chinstrap Penguins at King George Island, Antarctica. *Ecology* **68**, 351-361.

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

Volkman NJ, Presler P, Trivelpiece W (1980) Diets of Pygoscelid Penguins at King George Island, Antarctica. *The Condor* **82**, 373-378.

Waite DW, Deines P, Taylor MW (2012) Gut Microbiome of the Critically Endangered New Zealand Parrot, the Kakapo (*Strigops habroptilus*). *PLoS One* **7**, e35803.

Waite DW, Taylor MW (2014) Characterizing the avian gut microbiota: membership, driving influences, and potential function. *Frontiers in Microbiology* **5**.

Waite DW, Taylor MW (2015) Exploring the avian gut microbiota: current trends and future directions. *Frontiers in Microbiology* **6**, 673-673.

Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology* **73**, 5261-5267.

Wanless S, Harris MP, Morris JA (1992) Diving behaviour and diet of the blue-eyed shag at South Georgia. *Polar Biology* **12**, 713-719.

Wickham H (2016) *ggplot2: Elegant Graphics for Data Analysis* Springer-Verlag, New York.

Williams TD (1990) Annual variation in breeding biology of gentoo penguins, *Pygoscelis papua*, at Bird Island, South Georgia. *Journal of Zoology* **222**, 247-258.

Williams TD (1995) *The penguins : Spheniscidae* Oxford University Press, Oxford ; New York.

Wilson RP (2010) Resource partitioning and niche hyper-volume overlap in free-living Pygoscelid penguins. *Functional Ecology* **24**, 646-657.

Woods R, Jones HI, Watts J, Miller GD, Shellam GR (2009) Diseases of Antarctic Seabirds. In: *Health of Antarctic wildlife: a challenge for science and policy* (eds. Kerry KR, Riddle M). Springer, London.

Wright ES (2016) Using DECIPHER v2.0 to Analyze Big Biological Sequence Data in R. *The R Journal* **8**, 352-359.

Wu Y, Yang Y, Cao L*, et al.* (2018) Habitat environments impacted the gut microbiome of long-distance migratory swan geese but central species conserved. *Sci Rep* **8**, 13314.

Xavier JC, Trathan PN, Ceia FR*, et al.* (2017) Sexual and individual foraging segregation in Gentoo penguins *Pygoscelis papua* from the Southern Ocean during an abnormal winter. *PLoS One* **12**, e0174850.

Yew WC, Pearce DA, Dunn MJ, Samah AA, Convey P (2017) Bacterial community composition in Adélie (*Pygoscelis adeliae*) and Chinstrap (*Pygoscelis antarctica*) Penguin stomach contents from Signy Island, South Orkney Islands. *Polar Biology* **40**, 2517-2530.

Youngblut ND, Reischer GH, Walters W*, et al.* (2019) Host diet and evolutionary history explain different aspects of gut microbiome diversity among vertebrate clades. *Nature communications* **10**, 2200-2200.

Zhu XY, Zhong T, Pandya Y, Joerger RD (2002) 16S rRNA-Based Analysis of Microbiota from the Cecum of Broiler Chickens. *Applied and Environmental Microbiology* **68**, 124-137.

-

Chapter 8. Screening of 16S Profiles for Potentially Pathogenic Bacteria in Four Penguin Species and a Flying Seabird in the Scotia Arc and Antarctica

Hila Levy[1], Stephen G. Preston[1], Tom Hart[1], Adrian L. Smith[1]

1. Department of Zoology, University of Oxford, South Parks Road, Oxford, OX1 3SZ, United Kingdom

Keywords: penguins, wildlife disease, pathogens, 16S rRNA, Antarctica, *Pygoscelis*

Abstract

Bacterial infection is a significant cause of disease and mortality in wild seabirds and other Sub-Antarctic and Antarctic wildlife. There is growing concern of human-mediated pathogen introduction into the Antarctic region. However, knowledge of the presence of potentially pathogenic bacteria in the normal gut microflora of seabirds is currently limited. In this study, we screened 461 samples (87 cloacal swabs and 374 faecal samples) from Adélie (n = 23), Chinstrap (n = 56), Gentoo (n = 364), and King (n = 20) Penguins, as well as Blue-Eyed Shags (n = 11) for genera of pathogenic interest, using parallel sequencing of 16S rRNA profiles. In this molecular survey, we encountered 1,584 unique amplicon sequence variants assigned to 29 genera of potential pathogenic interest. Among these, we detected highly abundant and widespread *Clostridium*, *Corynebacterium*, *Helicobacter*, *Pasteurella*, *Pseudomonas*, and *Staphylococcus* in the bacterial microflora. Among other genera, we also identified sequence variants attributable to *Brucella*, *Chlamydia*, *Haemophilus*, *Klebsiella*, *Neisseria*, *Providencia* and *Vibrio*, some of which form the first reports of such bacterial sequences in a given host species or in the Antarctic region. We propose the use of molecular screening for pathogens as a tool to be integrated with culture, multilocus sequence typing, or whole-genome sequencing for further bacterial characterisation of this understudied microbial ecosystem.

Introduction

Bacterial pathogens are known to cause significant clinical disease and mortality in wild birds (Benskin *et al.* 2009). Within the Antarctic region, penguins breed ashore in dense colonies that are fertile ground for transmission of enteric bacterial infections. Though likely under-reported due to the remoteness of these areas, mortality events attributable to bacterial infection have been described in penguins and flying seabirds in the Scotia Arc. At Hope Bay, Antarctica, two Adélie Penguins (*Pygoscelis adeliae*) died of subcutaneous *Clostridium sporogenes* and *Clostridium cadaveris* infection (Nievas *et al.* 2007), and a devastating die-off due to avian cholera associated with *Pasteurella multocida* killed over 86 birds, including 45 Adélie Penguins (Leotta *et al.* 2006b). *P. multocida* has also been linked to two substantial outbreaks on South Georgia, where "several hundred" Chinstrap Penguins (*Pygoscelis antarcticus*) perished (Government of South Georgia and the South Sandwich Islands 2005, 2013).

In parallel, increasing human contact through tourism and scientific programmes has raised questions regarding the introduction of new or human-associated microbes into this sensitive wilderness (Cerdà-Cuéllar *et al.* 2019; Hernández & González-Acuña 2016; Retamal *et al.* 2017). Concerns surrounding mortality events, anthropogenic pathogen pollution, and the role of wildlife disease on population dynamics and conservation have led to increased investigations of bacterial pathogen presence in Antarctica (see reviews by Barbosa & Palacios (2009), Woods *et al.* (2009), and also tables and references within **Chapter 4**). Much of the literature focuses on breeding colonies in close proximity to research stations, and use bacterial culture techniques to identify specific pathogens of interest, occasionally complemented by molecular characterisation (Bonnedahl 2011; Bonnedahl *et al.* 2005; Bonnedahl *et al.* 2008; Debruyne *et al.* 2010; González-Acuña *et al.* 2013; Iveson *et al.* 2009; Jorge *et al.* 2002; Vigo *et al.* 2011).

In order to add to our understanding of the baseline presence of bacterial pathogens in Sub-Antarctic and Antarctic penguin colonies, and to focus future monitoring efforts, we

embarked on a molecular screening survey of four penguin species and one flying seabird across a large spatial range. Using 16S rDNA parallel sequencing, we screened for potentially pathogenic bacterial sequences in 461 samples (374 faecal samples and 87 cloacal swabs) from Blue-Eyed Shags (*Phalacrocorax atriceps bransfieldiensis*), King Penguins (*Aptenodytes patagonicus*) and Pygoscelid penguins (Adélie Penguins, Chinstrap Penguins, and Gentoo Penguins, *P. papua*) from 29 unique sites across four austral summers. We aimed to determine the presence of any potentially pathogenic bacterial genera and describe their prevalence and geographic distribution among host species.

Materials and Methods

Methods pertaining to sample collection, DNA extraction, 16S rDNA amplification and sequencing, bioinformatics post-processing and taxonomic classification are all as described in **Chapter 7**. To screen for potential bacterial pathogens from the 16S sequence results obtained, genera containing species and strains of bacteria of known clinical relevance in vertebrates were highlighted for further analysis. The starting points for designating these genera included reviews of avian bacterial pathogens and diseases of Antarctic wildlife discussed in **Chapter 4** of this thesis (Barbosa & Palacios 2009; Benskin *et al.* 2009; González-Acuña *et al.* 2013; Woods *et al.* 2009). Sequences from these pathogen-associated genera were subsetted within the phyloseq R package (McMurdie & Holmes 2013) to determine their presence, prevalence, and abundance within samples. Individual amplicon sequence variants (ASVs) were input into the online BLAST RefSeq 16S database (release 86) to determine closest matches for species assignment, where possible and to obtain % sequence similarities. For each genus, the reference 16S sequences deposited in the List of Prokaryotic names with Standing in Nomenclature (Euzéby 1997; Parte 2013, 2018) were downloaded (https://lpsn.dsmz.de/, last accessed 28 May 2020) for alignment with ASVs assigned to that genus using MAFFT v7.450 (Katoh & Standley 2013) in Geneious Prime 2020.0.3 (https://www.geneious.com). Neighbour-Joining trees were constructed in Geneious

Prime (HKY distance model) to visually examine the topology of phylogenetic relatedness to curated bacterial reference sequences.

Results and Discussion

Pathogen Presence

The broader bacterial 16S dataset contained 497 classified genera. Of these, 29 genera (1,584 ASVs, 1,736,846 reads, 25.67% of total reads) were considered as taxa linked to potential (or opportunistic) pathogenicity in vertebrates. We discuss these taxa below by bacterial phylum, with all nomenclature per the List of Prokaryotic names with Standing in Nomenclature as of May 2020 (Euzéby 1997; Parte 2013, 2018). The graphical summary is presented in **Figure 1,** with summary sequencing results presented in **Table 1**.

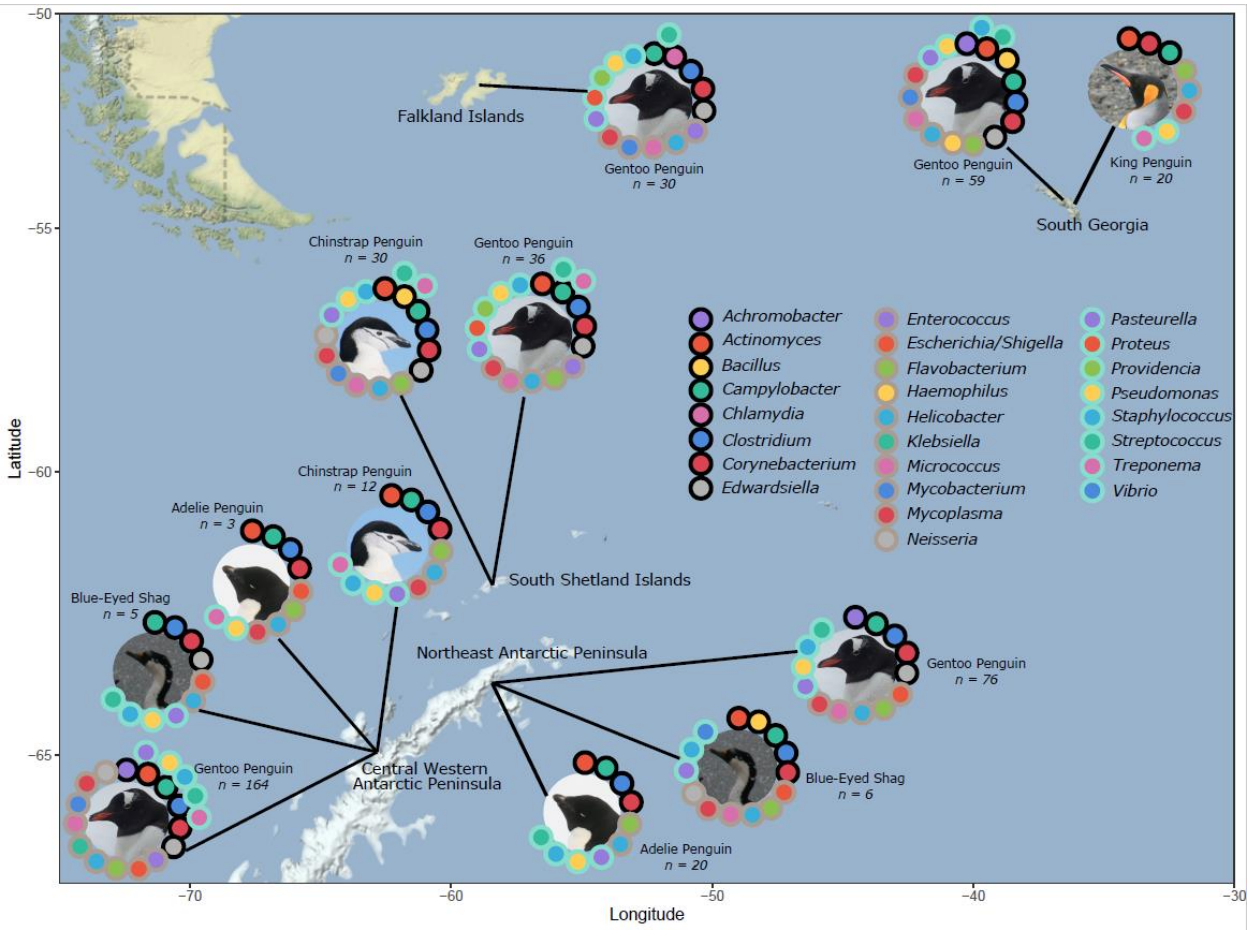


Figure 6. Map of the sampling region, with images depicting host species sampled from each geographic region. Coloured dots (clockwise from top in alphabetical order) indicate presence of

sequence variants assigned to potentially pathogenic genera at a minimum threshold of 10 reads in at least one sample from that host-region. This map excludes *Alcaligenes*, *Brucella*, and *Yersinia* sequence variants retrieved from single samples at low read numbers.

Table 6. Summary of the amplicon sequence variants (ASVs) assigned to potentially pathogenic genera in the study of four penguin species and Blue-Eyed Shags from the Scotia Arc and Antarctic Peninsula, with their total read abundance. The number of individual samples at which any number of assigned reads (>0) was retrieved is listed. 461 total samples were screened, and all contained at least one sequence variant of interest.

Bacterial Genus	**Unique ASVs**	**Total Abundance (Reads)**	**Individual Samples Detected (>0 reads)**
Achromobacter	1	185	8
Actinomyces	33	4,899	139
Alcaligenes	1	2	1
Bacillus	3	822	10
Brucella	1	4	1
Campylobacter	39	18,676	146
Chlamydia	1	22	1
Clostridium	368	878,843	451
Corynebacterium	57	20,528	237
Edwardsiella	16	10,109	138
Enterococcus	1	1,024	46
Escherichia/Shigella	2	987	25
Flavobacterium	571	222,210	246
Haemophilus	5	39	5

Helicobacter	75	114,262	429
Klebsiella	5	106	8
Micrococcus	5	7,553	122
Mycobacterium	9	1,481	50
Mycoplasma	32	6,150	188
Neisseria	5	186	5
Pasteurella	27	130,024	426
Proteus	1	530	18
Providencia	5	3,360	61
Pseudomonas	165	37,992	340
Staphylococcus	125	270,555	356
Streptococcus	15	4,826	90
Treponema	14	992	55
Vibrio	1	474	6
Yersinia	1	5	1
Totals:	**1584**	**1,736,846**	**461**

Actinobacteria

Bacteria within the phylum Actinobacteria can be found in a diversity of environmental habitats, but only a few genera are associated with pathogenicity. We detected the presence of *Actinomyces* (33 ASVs in 139 samples) and *Corynebacterium* (57 ASVs in 237 samples) in all host species and regions, as well as *Micrococcus* (5 ASVs in 122 samples from Blue-Eyed Shags, Chinstrap Penguins, and in Gentoo Penguins from all regions). Infections that result from *Actinomyces* and *Micrococcus* bacteria are rare in vertebrates, and are generally commensals only associated with opportunistic infections in immunocompromised individuals or open

wounds (Valour *et al.* 2014). *Corynebacterium* is a genus with several species of veterinary and human pathogenic interest (Oliveira *et al.* 2017), including the causative agent of diphtheria, *C. diphtheriae*, and other species that can cause zoonotic infection (Bernard 2012). The virulence factors that make coryneform bacteria of interest would not be detectable in the 16S analysis, so we cannot comment on potential pathogenicity. Of note, three of the top ten most abundant *Corynebacterium* sequence variants in our data set were 99.76-100% matches to either *C. spheniscorum* or *C. sphenisci*, originally isolated from wild Magellanic Penguins, *Spheniscus magellanicus* (Goyache *et al.* 2003a; Goyache *et al.* 2003b).

The genus *Mycobacterium*, on the other hand, contains the obligate parasitic species responsible for tuberculosis and leprosy in humans, though most other species are environmental or commensal. Though initial taxonomic assignment categorised 9 ASVs present in 50 samples from Chinstrap and Gentoo Penguin as *Mycobacterium*, a closer look at the sequence variants and their relationship to other references places them in a sister genus recently described by phylogenomic analysis, *Mycolicibacterium* (Gupta *et al.* 2018), see **Supplementary Figure S1**. This genus (synonymous with the "Fortuitum-Vaccae" clade of *Mycobacterium*) is considered non-pathogenic, with cases of opportunistic disease only described in rare immune conditions in humans (Tanaka *et al.* 2019). It is therefore likely that the Actinobacteria we detected in our study are primarily commensals, with potential for opportunistic infection, though further study and characterisation are warranted.

Bacteroidetes

Within the phylum Bacteroidetes (syn. Bacteroidota), the genus *Flavobacterium* contains known fish pathogens of commercial relevance (Chen *et al.* 2017). At least one species, *Flavobacterium meningosepticum*, has been isolated from diseased chickens, a pigeon, and a zebra finch (Vancanneyt *et al.* 1994). The genus as a whole is diverse, with 228 recognised species. In our study, we recovered 571 unique ASVs classified as *Flavobacterium*, comprising

222,210 reads (12.8% of total reads assigned to potentially pathogenic genera), across 246 samples from every host and region. While it is likely that many ASVs are strains of individual species, this finding indicates high diversity of this genus among the Sub-Antarctic and Antarctic regions sampled. Analysis of the top 20 most abundant *Flavobacterium* sequences indicated that 17 of our top sequence variants likely represent psychrophilic environmental flavobacteria (see **Supplementary Figure S2**). Nine of the most abundant taxa in this genus were closely related to *F. kingsejongi*, isolated from penguin faeces (Choi *et al.* 2018) and *F. phocarum* from soil amid a seal habitat (Zhou *et al.* 2018), both from King George Island (South Shetland Islands). Another six of the sequence variants fall within a clade with *F. antarcticum* (Yi *et al.* 2005), *F. weaverense* and *F. segetis* from soil on King George Island (Yi & Chun 2006) and *F. tegetincola* from sea ice (McCammon & Bowman 2000). Another falls within a separate clade with high similarity to two Antarctic environmental samples (Králová *et al.* 2019), while another had high similarity and phylogeny to psychrophilic species isolated from the Arctic, *F. hiemivividum* (Chaudhary et al. 2020), and isolates from Chinese glaciers (Zhang *et al.* 2006; Zhang *et al.* 2019). These potentially psychrophilic taxa were highly abundant in certain colonies South of the Polar Front, particularly around the South Shetland Islands, but absent from the Falkland Islands, where temperatures are warmer, further supporting their categorisation as environmental bacteria.

Of interest, however, was the fact that the other three most abundant sequence variants of *Flavobacterium* clustered among species isolated from both environmental samples such as ponds (Kim *et al.* 2009) and freshwater streams in Asia (Kim *et al.* 2012), as well as from diseased farmed salmonid fish in Chile (Kämpfer *et al.* 2012), and sister to a clade containing the majority of recognised wild and farmed fish pathogens (Loch & Faisal 2015; Zamora *et al.* 2013). These sequences were found across a number of sites on the Antarctic Peninsula in Pygoscelid penguins, and also in low abundances among King Penguins from South Georgia. Without further analysis, biochemical characterisation, or genomic sequencing, we cannot assert whether these represent fish pathogens trophically transmitted into the penguin gastrointestinal tract or are

otherwise environmental bacteria. Though flavobacterial infections have not been reported in seabirds, there is much to be learned about the ecology of these bacteria in the Antarctic and in the megafauna that may be associated with certain species.

Chlamydiae

The phylum Chlamydiae consists of obligate intracellular bacteria, the most notable of which are the pathogens in the genera *Chlamydia* and *Chlamydiophila* that are attributed to human trachoma, chlamydia, and pneumonia, as well as the zoonotic avian disease psittacosis (Everett *et al.* 1999; Harkinezhad *et al.* 2009). *Chlamydia* was found only in a single sample (22 reads out of 35,015) from a Gentoo Penguin on the Falkland Islands. A neighbour-joining tree built with the sequence and all known *Chlamydia* reference sequences using a *Parachlamydia* outgroup revealed an independent branch nested within other *Chlamydia* species, but the closest match was *C. pecorum* at ~94% similarity (see **Supplementary Figure S3**), which may indicate a new species or strain meriting culture and further characterisation.

Firmicutes

Firmicutes are mostly Gram-positive bacteria, and include anaerobic Clostridia as well as the Bacilli. This phylum was dominant among guano samples in our data set, and within it we chose to analyse reads assigned to the genera *Bacillus*, *Clostridium*, *Enterococcus*, *Mycoplasma*, *Staphylococcus*, and *Streptococcus*. The rarest Firmicutes genus was *Bacillus*, from which we recovered 3 ASVs in ten samples (1 Adélie Penguin, 2 Blue-Eyed Shags, 1 Chinstrap Penguin, and 6 Gentoo Penguins). The most abundant sequence variant, recovered from the Blue-Eyed Shags (ASV_627) was abundant and clustered with *B. licheniformis* (100% BLAST hit), which is commonly found in soil and the bird feathers of ground-dwelling and aquatic birds. The next most-abundant ASV, found in a Gentoo Penguin in South Georgia was a 100% match to *B. idriensis*, recovered from both clinical human samples and from environmental samples (Ponpandian *et al.* 2019). Certain strains of these two bacterial species have been considered

pathogenic to immunocompromised (Haydushka *et al.* 2012) or newborn humans (Ko *et al.* 2006), though their importance in seabirds is unknown. *Bacillus* sp. have been detected previously in culture from Gentoo Penguins on King George Island, characterised as non-pathogenic *B. cereus*, *B. circulans*, *B. megatherium*, and *B. sphaericus* (Jorge et al. 2002).

We also recovered one sequence variant of *Enterococcus*, a 100% match to *Enterococcus faecalis* strains, from 46 samples. Of those, 6 samples had appreciable read numbers (>10), all in Gentoo Penguins, from the Falkland Islands (penguin carcass swab from SAUN, n = 1), South Shetland Islands (AITC n = 3), and Central Western Antarctic Peninsula (BOOT n = 1 and GEOR n = 1). *Enterococcus faecalis* has been reported once previously (designated as *Streptococcus faecalis* in source report) in culture from a Chinstrap Penguin on King George Island (Jorge *et al.* 2002). This bacterium is part of the normal gastrointestinal microbiota in humans and birds, but is considered an opportunistic pathogen capable of causing septicaemia, endocarditis, and other clinical symptoms in avian hosts. *E. faecalis* is genotypically diverse, with over 600 sequence types, approximately 90 of which are associated with avian disease (Jørgensen *et al.* 2017; Stępień-Pyśniak *et al.* 2018). Future research should involve whole genome sequencing or multilocus sequence typing (MLST) for strain determination.

The mostly prevalent bacterial genus in our data set was *Clostridium*, for which the taxonomy is somewhat debated. Microbiome characterisation showed that *Clostridium sensu stricto I* represents part of the "core" microbiota of both Gentoo Penguins and Blue-Eyed Shags (see **Chapter 7**). Reads classified as *Clostridium sensu stricto* represented 50.6% of all reads assigned to potentially pathogenic genera. We recovered 368 ASVs from 451 samples (all host species, all regions). Given that Clostridia are generally anaerobic commensals, our interest was to determine whether the pathogenic species *C. cadaveris, C. perfringens*, *C. sporogenes* or *C. tetani* might be present *C. cadaveris* and *C. sporogenes* were both implicated in a subcutaneous infection leading to the deaths of two adult Adélie Penguins in Hope Bay, Antarctica in the 2000-

01 breeding season (Nievas *et al.* 2007). The most abundant *Clostridium* sequence variant in our data set was a 100% match to *C. perfringens*, abundant (mean: 243, median: 27, max: 3361 reads) and prevalent across all hosts and regions. Numerous other sequences (133 ASVs) were ≥ 98.0% similar to *C. perfringens* in our data set. A 100% match to *C. cadaveris* was retrieved from a sequence variant found in two Gentoo Penguins, one from the South Shetland Islands, and another from a swab obtained from a carcass found on the Falkland Islands. Though we cannot determine whether the bacteria we sequenced were pathogenic or commensal, this finding shows a high prevalence of Clostridia across the Scotia Arc and Antarctica, some of which may be capable of acting as opportunistic seabird pathogens.

Certain *Mycoplasma* species are causative agents of respiratory disease, membrane inflammation, and other lesions in birds (Stipkovits 1979). We recovered 32 ASVs assigned to *Mycoplasma* from 188 samples across the full sampling range, though 45 samples had appreciable (>10) read numbers for consideration, including an Adélie Penguin (BOOT), Blue-Eyed Shag (PAUL), six Chinstrap Penguins (BAIL n = 1 and BOOT n = 5), six King Penguins (STA) and 30 Gentoo Penguins (all regions). None of the ASVs clustered within the clade containing *Mycoplasma gallisepticum* (see **Supplementary Figure S4**)*,* an important bacterial pathogen in poultry that has been detected in at least 36 avian species from 6 orders, but not yet reported in penguins (Sawicka *et al.* 2020). Many of the ASVs clustered as unique clades within *Mycoplasma* sister to avian or other vertebrate mycoplasmas and merit further characterisation, while four ASVs grouped closely with *M. sphenisci*, which was first isolated from an African penguin (*Spheniscus demersus*) in an aquarium setting (Frasca *et al.* 2005), and also from healthy-appearing Adélie Penguins in Antarctica and King Penguins from South Georgia (Banks *et al.* 2009; Dewar *et al.* 2013).

Staphylococcus (125 ASVs in 356 samples) was highly prevalent and diverse, found in guano samples (rarely recovered from cloacal swabs) from Pygoscelid hosts and Blue-Eyed Shags

across geographic regions. *Staphylococcus* represented 15.6% of bacterial reads of pathogenic interest, with the top ten most abundant ASVs present across 280 samples. These were found in greatest abundances in Gentoo Penguin colonies in South Georgia and the Falkland Islands. *Staphylococcus* forms part of the normal microflora of many bird species, but can lead to staphylococcosis in localised lesions when skin or mucous membranes have been ruptured, or to more severe systemic infections, particularly in immunocompromised birds (Merck Veterinary Manual, 2016). Many strains are aetiological agents of disease in birds and other vertebrates, particularly *S. aureus* (Hermans et al. 2000), though further sequencing and gene typing would be needed to determine pathogenicity of our sampled bacteria.

Streptococcus (15 ASVs) was recovered from 90 samples in Pygoscelid hosts and one Blue-Eyed Shag, with very low abundances found in one King Penguin swab. The sequences (3 ASVs) detected in the Blue-Eyed Shag from Jougla Point, Central Western Antarctic Peninsula formed a unique clade most closely related to *S. phocae*, a species that has been isolated from a post-mortem investigations of a number of marine mammal species (Taurisano *et al.* 2018) and is considered an emerging pathogen of farmed salmonids in Chile (Romalde *et al.* 2008). The more abundant sequences found in Pygoscelids (9 ASVs) clustered within a clade that includes species in the *S. mitis* group, known commensals in human oral microflora that can serve as opportunistic pathogens (Abranches *et al.* 2018). *Streptococcus* is part of the normal gastrointestinal microbiota of birds, but can act as an opportunistic pathogen, leading to localised infection in lesions, sub-acute or chronic symptoms or acute septicaemia (Merck Veterinary Manual, 2016). None of our sequences fell within the groups of known avian pathogens by 16S taxonomic analysis.

Proteobacteria

The phylum Proteobacteria is major phylum of Gram-negative bacteria that includes many pathogenic genera of importance in vertebrates. This phylum featured prominently in the

overall bacterial communities of the seabirds sampled, and we recovered reads from 17 genera of interest, some of which were highly prevalent (*Helicobacter*, n = 429, and *Pasteurella*, n = 426), while others were very rare (*Brucella*, *Vibrio* and *Yersinia* in single individuals). We present our findings by bacterial family.

Alcaligenaceae

The genus *Alcaligenes* had been investigated as potentially pathogenic, as *Alcaligenes faecalis* has been associated with various clinical infections in humans (Tena *et al.* 2015). However, the 2 reads recovered from a single Blue-Eyed Shag from Paulet Island had a 100% sequence identity match to *Alcaligenes aquatilis*, a bacterial species first isolated from estuarine and salt-marsh sediments in Europe and the Southeastern United States, which is not recognised as pathogenic (Van Trappen *et al.* 2005).

A single sequence variant assigned to the genus *Achromobacter* was recovered from the dataset, found in one Chinstrap Penguin and seven Gentoo Penguins. *Achromobacter* reads were found in samples spanning from South Georgia (Salisbury Plain), the South Shetland Islands (Aitcho Island, in both species), the Northeast Antarctic Peninsula (Heroina Is. and Brown Bluff) and the Central Western Antarctic Peninsula (Mikkelsen Harbor). The sequence variant had a 99.77% match and nested most closely to *Achromobacter xylosoxidans*, which has been implicated in rare clinical infections in humans (Awadh *et al.* 2017). The pathogenicity of this current strain is unknown.

Brucellaceae

A small number of *Brucella* reads were found in a single Gentoo Penguin scat from South Georgia, with species assignment not possible from this locus due to 100% matches to *B. papionis*, *B. ovis*, *B. melitensis*, and *B. canis*, suggesting further sequencing would be needed to refine the classification of these reads. The role of avian species as reservoirs or mechanical vectors for *Brucella* transmission is controversial, though antibodies have been found in several

avian families of birds, with rare instances of direct bacterial isolation (Wareth *et al.* 2020). Conversely, *B. pinnipedialis* and *B. ceti* are commonly found in marine mammals, with antibodies (but not isolation) having been detected in Antarctic fur seals (*Arctocephalus gazella*) and Weddell's seals (*Leptonychotes weddellii*) and Southern Elephant Seals (*Mirounga leonina*) (Abalos *et al.* 2009; Jensen *et al.* 2013). Given the uncertainty surrounding the strain of *Brucella* recovered, it is possible that the scat sample had traces of pinniped faeces present on the substrate or tracked into the colony by penguins. Alternatively, should this recovery represent true presence, anti-*Brucella* seropositivity studies would be merited in Sub-Antarctic penguins.

Campylobacteraceae

Campylobacter spp. play an important role as human pathogens associated with poultry consumption, and their presence in wild avian hosts is of research interest to understand bacterial reservoirs of pathogen transmission and evolution (Cody *et al.* 2015; Griekspoor *et al.* 2013; Hughes *et al.* 2009; Waldenström *et al.* 2002). We detected 39 ASVs assigned to *Campylobacter*, found in all regions and all host species. After eliminating the rarest taxa (<10 total reads), we analysed 27 ASVs to determine their taxonomic placement within the genus and any similarities to known species (see **Supplementary Figure S5**). We were somewhat surprised to see that many of our sequences form new clades or populate less-speciose clades isolated from other birds. Using a *Helicobacter pylori* sequence as an outgroup, six ASVs formed two clades completely separate from all known *Campylobacter* sequences in our neighbour-joining tree, but still within the genus, showing approximately 92.5-94.8% shared 16S fragment identity with *C. helveticus* as the top BLAST hit. Another 12 ASVs clustered (~99% similarity) with *C. canadensis*, isolated from captive whooping cranes in Canada, which until now has formed its own distinct clade in the genus (Inglis *et al.* 2007). This grouping included several ASVs that were unique to Blue-Eyed Shags, in which *Campylobacter* has not been detected previously. These results suggest it is likely that there is undescribed *Campylobacter* diversity in sub-Antarctic and Antarctic birds.

One sequence (ASV_357) was a 99.5% match to both *C. jejuni* subspecies *jejuni* and *doylei*. This ASV was recovered in low abundances from the three Pygoscelids, but was highly abundant among seven King Penguin chicks from South Georgia. We cannot determine the likely pathogenicity of this strain from our marker, though *C. jejuni* is often associated with gastroenteritis in humans (Parker *et al.* 2007) and some strains may be associated with diarrhoea and gut damage in chickens (Humphrey *et al.* 2014). However, generally *Campylobacter* is not considered particularly pathogenic in birds, including commercial poultry that are the source of most infection in humans (Colles *et al.* 2015). Three additional ASVs clustered closely with *Campylobacter* species associated with shellfish (*C. armoricus*, *C. peloridis*) and marine mammals (*C. insulaenigrae*), some of which have also been found to cause clinical presentations of enteritis and septicaemia in humans (Boukerb *et al.* 2019; Chua *et al.* 2007; Debruyne *et al.* 2009; Foster *et al.* 2004). We expected to find *C. lari* or *C. subantarcticus* in our dataset, given that these have been previously isolated from the sampling area (Debruyne *et al.* 2010; Johansson *et al.* 2018), but the three most closely related ASVs to the *C. lari* group (~99.5% identity) were unique to Blue-Eyed Shags from Paulet Island in the Northeast Antarctic Peninsula.

In the Antarctic region, *Campylobacter* is widely studied to track bacterial strains that may have been introduced through human activity such as poultry product importation or sewage discharge from ships and research stations (Bonnedahl *et al.* 2005; Cerdà-Cuéllar *et al.* 2019). In Pygoscelid penguins, *Campylobacter* has been detected in the South Shetland Islands and Antarctic Peninsula (Barbosa *et al.* 2013; Bonnedahl *et al.* 2005; García-Peña *et al.* 2017; González-Acuña *et al.* 2013; Johansson *et al.* 2018; Leotta *et al.* 2006a) although the clinical consequence of these infections are unknown. This genus has been cultured and sequenced from flying seabirds (Giant Petrel, Kelp Gull, Brown Skua, and Snowy Sheathbill) on South Georgia (Johansson *et al.* 2018), but we report the first detection in King Penguins there. Many questions remain regarding the ecology of *Campylobacter* among sub-Antarctic and Antarctic seabirds, including potential pathogenicity of strains and cross-species transmission. The use of culture

techniques, multilocus sequence typing, targeted gene sequencing or even whole-genome sequencing would be useful to address some of these questions.

Enterobacteriaceae

There were no sequences assigned to the genus *Salmonella* in the 461 samples. *Salmonella enterica* serovars have cultured and sequenced from Gentoo Penguins in South Georgia and the South Shetland Islands, and in Adélie and Gentoo Penguins from Antarctica in other studies (Dimitrov *et al.* 2009; Iveson *et al.* 2009; Oelke & Steiniger 1973; Olsen *et al.* 1996; Retamal *et al.* 2017; Vigo *et al.* 2011).

Two sequence variants were assigned to the designation *Escherichia/Shigella*. These were found in all host species (n = 25), but discounting those with rare abundance (<10 reads), one ASV primarily came from an Adélie Penguin (BOOT), five Antarctic Blue-Eyed Shags (JOUG n = 4 and PAUL n = 1), and Gentoo Penguins from Antarctic Peninsula sites (EARL in the Northeast, JOUG and MIKK in the Central Western region), while the other ASV was found only in a Blue-Eyed Shag from Paulet Island in the Northeast Antarctic Peninsula. These two sequence variants were 100% matches to multiple species in this clade (see **Supplementary Figure S6**), *Escherichia albertii*, *E. coli*, *E. fergusonii*, and *Shigella flexneri*, indicating that our target fragment is insufficient for fine-scale strain or species distinction. These enterobacteria can be commensal or pathogenic in birds, depending on bacterial virulence factors, environmental conditions, and host predisposition (Dho-Moulin & Fairbrother 1999). Other studies have isolated *E. coli* in Pygoscelids in the South Shetland Islands and Antarctica (Bonnedahl 2011; Bonnedahl *et al.* 2005; González-Acuña *et al.* 2013; Nievas *et al.* 2007), but sequencing, biochemical tests, and genotyping would be required for more precise characterisation of our samples.

Klebsiella sequence variants (5 ASVs, 106 reads) were recovered from 8 Gentoo Penguins from 7 colonies (6 in the Central Western Antarctic Peninsula and 1 on South Georgia). These sequence variants were most closely related to *K. pneumoniae* and *K. variicola*, both of which are

considered opportunistic pathogens in humans and can be easily confused in culture (Rodríguez-Medina *et al.* 2019). *K. pneumoniae* is associated with infection and mortality in Sub-Antarctic sea lion pups (Castinel *et al.* 2007; Roe *et al.* 2015), as well as multiple other mammalian and avian species (Brisse & Duijkeren 2005; Davies *et al.* 2016). Environmental isolates have been proposed as potential pathogen reservoirs (Struve & Krogfelt 2004), though only one report of isolation from Sub-Antarctic soil (Pinpimai *et al.* 2018) and one from Antarctic seawater have been noted (Mohammed *et al.* 2018). *Klebsiella* has rarely been reported in penguins, with *K. pneumoniae* cultured once from a rescued Magellanic penguin stranded in Brazil (Eckhardt De Pontes *et al.* 2007) and sequenced from a Yellow-Eyed penguin (and two Sub-Antarctic Skuas) on Enderby Island, New Zealand (Pinpimai *et al.* 2018). This is the first report of *Klebsiella* detection from an Antarctic host species. The wide range of its presence suggests it may be more prevalent in these environments than previously thought.

Hafniaceae

Edwardsiella sequences (16 ASVs, 1 OTU) were detected in 138 samples (2 Blue-Eyed Shags, Chinstrap Penguins and 133 Gentoo Penguins) from 20 sites. 15 of the 16 sequence variants represented 99.3-100% matches to *Edwardsiella tarda* (the remaining ASV was a 96% match but only present as <0.1% of *Edwardsiella* reads). In one Gentoo Penguin from South Georgia, *Edwardsiella* represented 32.33% of all recovered reads from that individual. *E. tarda* is a known pathogen of fish, birds, amphibians, reptiles, mammals, and humans, associated with wound infections, meningitis, and haemorrhagic enteritis (Janda & Abbott 1993; White *et al.* 1973). A common problem for aquaculture, it is generally associated with aquatic animals or food-borne infections obtained by eating fish or seafood (Abbott & Janda 2006; Davies *et al.* 2018). It can also be spread through avian faecal contamination (Berg & Anderson 1972). Disease and mortality events linked to *E. tarda* have been reported in penguins (King, Magellanic, and Rockhopper Penguins) in captive or rehabilitative settings (Campos *et al.* 2013; Cook & Tappe 1985; Nimmervoll *et al.* 2011). Isolation of this bacteria from wild penguins has been reported in

Adélie, Gentoo, King, Macaroni, Royal, and Rockhopper Penguins (Woods *et al.* 2009), and was cultured from samples from all three Pygoscelids in Hope Bay (NEAP) and King George Island in the South Shetland Islands (Leotta *et al.* 2009), suggesting it might be common in Antarctic species. The current findings would support an extensive range and prevalence of *E. tarda* in the marine food web, with Antarctic seabirds as incidental hosts of this bacterium.

Helicobacteraceae

Helicobacter sequences (75 ASVs) were among the most prevalent in our data set, with reads recovered from 429 samples in all host species and regions. Among the top 10 most abundant sequences, which were found in 285 penguin samples (all four species), we noted that some sequences were structured by host and region. One sequence (ASV_261) was primarily in Chinstrap Penguins, another was found in Adélie and King Penguins, and three were largely restricted to King Penguins (over 10x more abundant than in all other species: ASV_609, ASV_282, and ASV_384). The most abundant sequence in our dataset (ASV_11) was restricted to Gentoo Penguins, but found across all geographic regions. The Falkland Islands (Gentoo Penguins) had the highest abundances of *Helicobacter*, and had four dominant Gentoo Penguin-unique sequences, two of which were unique to the Falklands (ASV_221, ASV_298). Nine of the ten most abundant ASVs clustered similarly (>99%) into two sister clades (see **Supplementary Figure S7**), one of which contained *H. cholecystus*, first isolated from a diseased hamster gallbladder (Franklin *et al.* 1996), and *H. pametensis* from bird and swine faeces (Dewhirst *et al.* 1994). These sequence variants are of unknown pathogenicity, and were 98.5% similar to *H. pullorum*, which causes enteritis in farm-raised birds (Javed *et al.* 2017). The other abundant ASV had 99.2% similarity to *H. anseris* isolated from Canada Geese (Fox *et al.* 2006). Though we cannot determine the role of these bacteria in seabird hosts, penguins should now be considered as potential wild reservoirs.

Morganellaceae

A single sequence variant assigned to the genus *Proteus* was recovered from an Adélie Penguin and several Gentoo Penguins, though the highest abundance was in a Falkland Islands Gentoo Penguin swab sample obtained from a carcass. *Proteus* spp. include some opportunistic human pathogens attributed to urinary tract and wound infections that are also found in the gastrointestinal tract of birds and other vertebrates (Armbruster & Mobley 2012). Their relationship with avian hosts is undetermined, and likely commensal, but wild birds are thought to contribute to transmission of these bacteria to domestic animals and humans (Drzewiecka 2016). We also report 5 ASVs belonging to the genus *Providencia*, found in 61 samples (all Pygoscelid species and one Blue-Eyed Shag), though generally at very low abundances aside from the Gentoo Penguins from the South Shetland Islands and a substantially high abundance of all 5 ASVs in the Gentoo Penguin swab sample from the Falkland Islands that also contained a *Proteus* sequence variant. *Providencia* is also associated with opportunistic and nosocomial urinary tract infections as well as diarrhoea in children and travellers (Actor 2012). Although the veterinary significance of these two bacterial genera is unknown, we report these findings as an expansion of their known geographic range in animal hosts and as first reports in wild penguins.

Neisseriaceae

Among *Neisseria* sequences, 5 ASVs were recovered, each from a different individual sample (1 Blue-Eyed Shag from Paulet Is., 1 Gentoo Penguin from South Georgia, 1 Gentoo Penguin from Mikkelsen Harbor, and both 1 Chinstrap Penguin and 1 Gentoo Penguin from Barrientos Is.). These ASVs clustered together and each presented 99-100% matches to strains of *Neisseria* generally considered to be commensal and non-pathogenic (*N. perflava*, *N. macacae*, *N. cinerea*, and *N. mucosa*). To distinguish species, biochemical characterisation in culture is likely required. In humans and primates, multiple case reports exist of sometimes fatal bacterial infections associated with these commensal strains acting as opportunistic pathogens (Humbert

& Christodoulides 2020; Johnson 1983; Vecten *et al.* 2017). Among numerous animal hosts, *Neisseria* have been detected in avian hosts as well as marine mammals like dolphins and the California sea lion (Liu *et al.* 2015). In penguins, two microbiome studies detected *Neisseria* in Gentoo Penguins, King Penguins, and Macaroni Penguins from the Sub-Antarctic, and in Little Penguins in Australia (Dewar *et al.* 2013; Dewar *et al.* 2014). The new data extends detection of *Neisseria* to Chinstrap Penguins and Blue-Eyed Shags, as well as geographically to the Antarctic region.

Pasteurellaceae

Discounting samples with very low read numbers (<10), *Pasteurella* was found in 271 Pygoscelid penguins and 9 Blue-Eyed Shags, and was recovered from all geographic regions. The 27 ASVs grouped together, as a distinct clade from all other *Pasteurella* (see **Supplementary Figure S8**). The closest similarity was to *P. testudinis* (~95%). Though *Pasteurella* is of clear interest because of its association with several fowl cholera mortality events in penguins and Antarctic seabirds, we found no sequences similar to *P. multocida*, the typical causative agent of the highly lethal avian disease. It is likely that the recovered sequences belong to one or more species of undescribed *Pasteurella*-related bacteria though further sequencing and culture would be needed to confirm this.

Five sequence variants with a 99.5-100% match to *Haemophilus parainfluenzae* were detected in 1 Chinstrap Penguin and 4 Gentoo Penguins, with unique ASVs in each sample, at 3 sites in South Georgia, the Falkland Islands, and South Shetland Islands (Barrientos Island). *Haemophilus* species are known pathogens in birds, though they typically act opportunistically as secondary invaders in respiratory infections of multiple avian hosts (Gerlach 1994). *Haemophilus* (unknown species) in penguins has only been reported in Little Penguins in Australia (Dewar *et al.* 2013). *Haemophilus* has not been reported in any Antarctic animal hosts previously, though human researchers overwintering at a base on Signy Island (South Orkney Is.) presented

widespread infection with *H. parainfluenzae* in 1990 (Hobson *et al.* 1995; Kerr *et al.* 1993). Sampling of penguin respiratory tracts would likely yield more informative data on the prevalence of *Haemophilus* in the wild.

Pseudomonadaceae

Pseudomonas bacteria are widely found in nature, but certain strains of *P. aeruginosa* can be an opportunistic avian respiratory and systemic pathogen (Benskin *et al.* 2009). We recovered 165 ASVs assigned to this genus from 240 samples, in every host species and colony. Considering the top 20 *Pseudomonas* sequence variants by abundance (which were found across the four penguin species), as well as the two most abundant Blue-Eyed Shag sequences, we found a diversity of *Pseudomonas* bacteria sequences that appear to be most genetically similar to varied environmental bacteria isolated from seawater, terrestrial environments, or anaerobic sludge (see **Supplementary Figure S9**). We also identified ASVs related to *P. deceptionensis*, a psychrotolerant species first isolated on Deception Island (Carrión et al. 2011), that we recovered from two sites on the Northeast Antarctic Peninsula, A highly abundant sequence variant from South Georgia, the Falkland Islands, and Mikkelsen Harbor (Central Western Antarctic Peninsula) was most similar to *P. versuta*, first isolated further south on the Antarctic Peninsula (See-Too *et al.* 2017). None of the most abundant sequences grouped near *P. aeruginosa*, a well-documented potential pathogen.

Vibrionaceae

One *Vibrio* sequence variant was recovered in a Blue-Eyed Shag from Paulet Island on the Northeast Antarctic Peninsula (469 reads, 1.3% of reads for that sample), with just one read found in three Chinstrap Penguins from AITC and one in each of two Gentoo Penguins from two sites in the Central Western Antarctic Peninsula, and may therefore be technical spillover. However, the number of reads present in the Paulet Island Blue-Eyed Shag are sufficient to be considered an accurate result, representing the first report of *Vibrio* in Blue-Eyed Shags and the

first report of detection in Antarctic birds. Species resolution was not possible, as the ASV was a 100% match in this 16S region to *V. cholerae*, *V. cholerae* biovar albensis, and *V. mimicus*. These species are associated with cholera, gastroenteritis, wound infection, and occasionally bacteraemia in humans, though they are free-living in aquatic ecosystems and have been found in multiple marine prey organisms (Guardiola-Avila *et al.* 2016). There is increasing evidence that waterbirds are vectors for *Vibrio* transfer, ingesting them in copepods and fish, and dispersing them in faeces during migration (Laviad-Shitrit *et al.* 2019). *V. cholerae* has been detected in wild Great Cormorants (*Phalacrocorax carbo*) and experimental studies have shown that food chain transfer is a likely mechanism for bacterial dispersal (Laviad -Shitrit *et al.* 2017). *Vibrio* spp. have not been isolated in wild penguins before, though a case was reported of vibrionic septicaemia in a captive King Penguin (Fraser 1961). Further investigation is required in this region, as the only previous detection of a *Vibrio* sp. in Antarctica was reported from an endolithic community in East Antarctica (Colwell *et al.* 1989).

Yersiniaceae

A single *Yersinia* sequence variant was recovered from a swab of a Chinstrap Penguin chick on Deception Island. This ASV had >99% 16S sequence similarity to six *Yersinia* species (*Y. aldovae, Y. aleksiciae, Y. bercovieri, Y. intermedia, Y. mollaretii,* and *Y. pekkanenji*), so species assignment was not possible. Though some *Yersinia* strains are highly virulent, biotyping in culture and/or plasmid sequencing would be required to understand whether this particular penguin strain is pathogenic. Migratory birds are known to carry non-pathogenic and pathogenic *Yersinia* strains (Niskanen *et al.* 2003), and a strain of *Y. rohdei* was isolated in Sub-Antarctic South Georgia from two kelp gulls and a feral reindeer (hosts that have since been eradicated from the island) in 2012 (Bonnedahl *et al.* 2018). However, less is known about the presence and prevalence of this bacterial genus in wild penguins or in the Antarctic biome.

Spirochaetes

Among spirochaetes of pathogenic interest, no *Borrelia* sequences were recovered, in line with our expectations for non-blood samples. However, we did retrieve 14 ASVs in 55 samples (all four penguin species) assigned to another pathogen-associated genus, *Treponema* (which contains the syphilis-causing human pathogen *T. pallidum*). Gastrointestinal spirochaetes have been identified in a number of birds and mammals, with some thought to form part of the normal flora (Stoutenburg *et al.* 1995), while other species formerly known as *Treponema* but now classified as *Brachyspira* (including *B. hyodysenteriae*, *B. pilosicoli* and other strongly haemolytic strains), can be pathogenic in birds (Le Roy *et al.* 2015), swine, and humans (Mappley *et al.* 2014) with waterbirds as reservoir hosts (Hampson *et al.* 2015). *Brachyspira* sequences have recently been found in snowy sheathbills from South Georgia and Antarctica, kelp geese from the Falkland Islands and Tierra del Fuego, and brown skuas from the Falkland Islands and South Shetland Islands (Jansson *et al.* 2015). Within a tree built with sequences from across both genera, our sequence variants seem to form a separate clade well within *Treponema* (see **Supplementary Figure S10**), but we cannot infer whether these bacteria are potentially pathogenic or not.

Concluding Remarks

Molecular screening techniques using the well-accepted reference target of 16S rDNA can aid in focusing bacterial pathogen research from environmental or animal-derived samples. In this study, we detected bacteria belonging to 29 genera of pathogenic interest to vertebrates. Among these, there were high abundances and prevalence of *Clostridium*, *Corynebacterium*, *Helicobacter*, *Pasteurella*, *Pseudomonas*, and *Staphylococcus*, among other bacteria that likely form part of the normal gastrointestinal flora of Adélie, Chinstrap, Gentoo, and King Penguins, and Blue-Eyed Shags. Sequence variants attributable to *Brucella*, *Chlamydia*, *Haemophilus*, *Klebsiella*, *Neisseria*, *Providencia*, *Vibrio*, and *Yersinia* were also identified, some of which form

the first reports of these bacteria in the Antarctic region or in a given host. Single-target parallel sequencing is limited in its ability to inform us of the likely pathogenicity, type strain, or ecology of many of these bacteria. For this reason, future work should consider molecular screening in conjunction with culture for biochemical characterisation, as well as multilocus sequence typing or whole genome sequencing of key samples and taxa of interest. Continued research into the bacterial microflora of Antarctic wildlife is important to conservation as a proxy for understanding human-mediated impacts and biological introductions, as well as for identifying zoonotic threats.

Acknowledgements

We thank Quark Expeditions, Cheeseman's Ecology Safaris, Oceanwide Expeditions, and J. Poncet for their assistance with fieldwork logistics. We thank Anni Djurhuus, Caitlin E. Black, Gemma V. Clucas, Michael J. Polito, Heather J. Lynch, Michael B. Schrimpf, Rachael Herman, Maureen Lynch, and Casey Youngflesh for assisting with sample collection. We thank Kirstie McLoughlin, Patrik Flammer, Rob Dixon, Sophie Hedges, Frances Colles, for sharing MiSeq runs and assisting with the sequencing workflow, and Amanda Williams and Rebecca Busby for overseeing sequencing runs. We thank Steven R. Fiddaman for assistance with formatting Figure 1. Molecular work was supported by Penguin Lifelines, the U.S. Department of Veteran's Affairs G.I. Bill, and Stony Brook University. ALS received funding from the BBSRC and Bertarelli Foundation. HL's work was supported by a University of Oxford Clarendon Fund Scholarship.

Author Contributions

HL, TH, and ALS developed and designed the study. TH obtained permits and secured logistics. HL and TH collected samples in addition to collaborators listed in acknowledgments. HL extracted all DNA and performed all PCR amplifications, quality control, and pooling. HL and SGP performed Illumina MiSeq library preparation, quality control, and sequencing with technical support from collaborators listed in acknowledgments. SGP processed all raw sequence data and

created custom scripts for demultiplexing, pre-processing reads, and performing automated taxonomic assignment. HL conducted all analyses and wrote the paper with feedback from all authors.

Ethics Statement

Animal handling protocols followed Scientific Committee on Antarctic Research Code of Conduct on the use of Animals for Scientific Purposes in Antarctica. All sampling in the Antarctic Treaty Area was carried out in accordance with United Kingdom Home Office guidelines, under permits granted by the United Kingdom Foreign and Commonwealth Office (permits: S3-23/2013, S7 28/2013, S3 04/2014, S7 03/2014, 33/2016, 34/2016, 37/2017), in South Georgia under permits granted by the Government of South Georgia and the South Sandwich Islands (SCI-2014-017, RAP-2015-018, RAP-2016-035 and RAP-2017-034), and in the Falkland Islands under Falkland Islands Government Environmental Planning Department permits (R16.2014, R15 2014, and R04 2017), with ethical approval from the University of Oxford Animal Welfare and Ethical Review Board.

Conflict of Interest Statement

The authors declare no conflict of interest.

Data Availability Statement

Custom scripts, genetic sequences, sample metadata, abundance tables, and taxonomy assignment tables will be made available prior to publication on a public repository.

References

Abalos P, Retamal P, Blank O, Torres D, Valdenegro V (2009) *Brucella* infection in marine mammals in Antarctica. *Veterinary Record* **164**, 250-250.

Abbott SL, Janda JM (2006) The Genus Edwardsiella. In: *The Prokaryotes: Volume 6: Proteobacteria: Gamma Subclass* (eds. Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E), pp. 72-89. Springer New York, New York, NY.

Abranches J, Zeng L, Kajfasz JK*, et al.* (2018) Biology of Oral Streptococci. *Microbiol Spectr* **6**.

Actor JK (2012) Clinical Bacteriology. In: *Elsevier's Integrated Review Immunology and Microbiology (Second Edition)* (ed. Actor JK), pp. 105-120. W.B. Saunders, Philadelphia.

Armbruster CE, Mobley HLT (2012) Merging mythology and morphology: the multifaceted lifestyle of *Proteus mirabilis*. *Nature reviews. Microbiology* **10**, 743-754.

Awadh H, Mansour M, Aqtash O, Shweihat Y (2017) Pneumonia due to a Rare Pathogen: *Achromobacter xylosoxidans*, Subspecies denitrificans. *Case reports in infectious diseases* **2017**, 3969682-3969682.

Banks JC, Cary SC, Hogg ID (2009) The phylogeography of Adelie penguin faecal flora. *Environ Microbiol* **11**, 577-588.

Barbosa A, De Mas E, Benzal J*, et al.* (2013) Pollution and physiological variability in gentoo penguins at two rookeries with different levels of human visitation. *Antarctic Science* **25**, 329-338.

Barbosa A, Palacios MJ (2009) Health of Antarctic birds: a review of their parasites, pathogens and diseases. *Polar Biology* **32**, 1095-1115.

Benskin CMH, Wilson K, Jones K, Hartley IR (2009) Bacterial pathogens in wild birds: a review of the frequency and effects of infection. *Biological Reviews* **84**, 349-373.

Berg RW, Anderson AW (1972) Salmonellae and *Edwardsiella tarda* in gull feces: a source of contamination in fish processing plants. *Appl Microbiol* **24**, 501-503.

Bernard K (2012) The Genus *Corynebacterium* and Other Medically Relevant Coryneform-Like Bacteria. *Journal of Clinical Microbiology* **50**, 3152-3158.

Bonnedahl J (2011) *Antibiotic Resistance in Enterobacteriaceae Isolated from Wild Birds*, Uppsala University.

Bonnedahl J, Berg C, Andersson DI*, et al.* (2018) Occurrence of *Yersinia rohdei* among feral reindeer (*Rangifer t. tarandus*) and kelp gulls (*Larus dominicanus*) on the Sub-Antarctic island South Georgia. *Infection Ecology & Epidemiology* **8**, 1517582.

Bonnedahl J, Broman T, Waldenstrom J*, et al.* (2005) In search of human-associated bacterial pathogens in Antarctic wildlife: report from six penguin colonies regularly visited by tourists. *Ambio* **34**, 430-432.

Bonnedahl J, Olsen B, Waldenstrom J*, et al.* (2008) Antibiotic susceptibility of faecal bacteria in Antarctic penguins. *Polar Biology* **31**, 759-763.

Boukerb AM, Penny C, Serghine J*, et al.* (2019) *Campylobacter armoricus* sp. nov., a novel member of the *Campylobacter lari* group isolated from surface water and stools from humans with enteric infection. *International Journal of Systematic and Evolutionary Microbiology* **69**, 3969-3979.

Brisse S, Duijkeren Ev (2005) Identification and antimicrobial susceptibility of 100 *Klebsiella* animal clinical isolates. *Veterinary Microbiology* **105**, 307-312.

Campos SDE, Pereira AM, Maia ALPS*, et al.* (2013) Identification of septicemia by *Edwardsiella tarda* in *Spheniscus magellanicus* (Sphenisciformes: Spheniscidae) - case report. *Revista Brasileira de Medicina Veterinária* **35**, 224-228.

Carrión O, Miñana-Galbis D, Montes MJ, Mercadé E (2011) *Pseudomonas deceptionensis* sp. nov., a psychrotolerant bacterium from the Antarctic. *International Journal of Systematic and Evolutionary Microbiology* **61**, 2401-2405.

Castinel A, Duignan PJ, Pomroy WE*, et al.* (2007) Neonatal mortality in New Zealand sea lions (*Phocarctos hookeri*) at Sandy Bay, Enderby Island, Auckland Islands from 1998 to 2005. *Journal of Wildlife Diseases* **43**, 461-474.

Cerdà-Cuéllar M, Moré E, Ayats T*, et al.* (2019) Do humans spread zoonotic enteric bacteria in Antarctica? *Science of the Total Environment* **654**, 190-196.

Chaudhary DK, Dahal RH, Kim D-U, Kim J (2020) *Flavobacterium sandaracinum* sp. nov., *Flavobacterium caseinilyticum* sp. nov., and *Flavobacterium hiemivividum* sp. nov., novel psychrophilic bacteria isolated from Arctic soil. *International Journal of Systematic and Evolutionary Microbiology* **70**, 2269-2280.

Chen S, Blom J, Loch TP, Faisal M, Walker ED (2017) The Emerging Fish Pathogen *Flavobacterium spartansii* Isolated from Chinook Salmon: Comparative Genome Analysis and Molecular Manipulation. *Frontiers in Microbiology* **8**.

Choi JY, Kim JH, Lee PC (2018) *Flavobacterium kingsejongi* sp. nov., a carotenoid-producing species isolated from Antarctic penguin faeces. *Int J Syst Evol Microbiol* **68**, 911-916.

Chua K, Gürtler V, Montgomery J*, et al.* (2007) *Campylobacter insulaenigrae* causing septicaemia and enteritis. *Journal of Medical Microbiology* **56**, 1565-1567.

Cody AJ, McCarthy ND, Bray JE*, et al.* (2015) Wild bird-associated *Campylobacter jejuni* isolates are a consistent source of human disease, in Oxfordshire, United Kingdom. *Environmental Microbiology Reports* **7**, 782-788.

Colles FM, McCarthy ND, Bliss CM, Layton R, Maiden MCJ (2015) The long-term dynamics of *Campylobacter* colonizing a free-range broiler breeder flock: an observational study. *Environ Microbiol* **17**, 938-946.

Colwell RR, Macdonell MT, Swartz D (1989) Identification of an Antarctic Endolithic Microorganism by 5S rRNA Sequence Analysis. *Systematic and Applied Microbiology* **11**, 182-186.

Cook RA, Tappe JP (1985) Chronic enteritis associated with *Edwardsiella tarda* infection in Rockhopper penguins. *J Am Vet Med Assoc* **187**, 1219-1220.

Davies YM, Cunha MPV, Oliveira MGX*, et al.* (2016) Virulence and antimicrobial resistance of *Klebsiella pneumoniae* isolated from passerine and psittacine birds. *Avian Pathology* **45**, 194-201.

Davies YM, Xavier de Oliveira MG, Paulo Vieira Cunha M*, et al.* (2018) *Edwardsiella tarda* outbreak affecting fishes and aquatic birds in Brazil. *The veterinary quarterly* **38**, 99-105.

Debruyne L, Broman T, Bergstrom S*, et al.* (2010) *Campylobacter subantarcticus* sp nov., isolated from birds in the sub-Antarctic region. *International Journal of Systematic and Evolutionary Microbiology* **60**, 815-819.

Debruyne L, On SLW, De Brandt E, Vandamme P (2009) Novel *Campylobacter lari*-like bacteria from humans and molluscs: description of *Campylobacter peloridis* sp. nov., *Campylobacter lari* subsp. concheus subsp. nov. and *Campylobacter lari* subsp. lari subsp. nov. *International Journal of Systematic and Evolutionary Microbiology* **59**, 1126-1132.

Dewar ML, Arnould JPY, Dann P*, et al.* (2013) Interspecific variations in the gastrointestinal microbiota in penguins. *Microbiologyopen* **2**, 195-204.

Dewar ML, Arnould JPY, Krause L*, et al.* (2014) Influence of Fasting during Moult on the Faecal Microbiota of Penguins. *PLoS One* **9**.

Dewhirst FE, Seymour C, Fraser GJ, Paster BJ, Fox JG (1994) Phylogeny of *Helicobacter* Isolates from Bird and Swine Feces and Description of *Helicobacter pametensis* sp. nov. *International Journal of Systematic and Evolutionary Microbiology* **44**, 553-560.

Dho-Moulin M, Fairbrother JM (1999) Avian pathogenic *Escherichia coli* (APEC). *Vet Res* **30**, 299-316.

Dimitrov K, Metcheva R, Kenarova A (2009) Salmonella presence - an indicator of direct and indirect human impact on gentoo in Antarctica. *Biotechnology & Biotechnological Equipment* **23**, 246-249.

Drzewiecka D (2016) Significance and Roles of *Proteus* spp. Bacteria in Natural Environments. *Microbial Ecology* **72**, 741-758.

Eckhardt De Pontes LA, Paes Petrucci M, Araujo Barbosa L*, et al.* (2007) Aspergillosis and candidiasis in rescued penguins of Magalhaes (*Spheniscus magellanicus* Foster, 1781) in

the coast of Espirito Santo, Brazil. *Revista Universidade Rural Serie Ciencias da Vida* **27**, 57-59.

Euzéby JP (1997) List of Bacterial Names with Standing in Nomenclature: a Folder Available on the Internet. *International Journal of Systematic and Evolutionary Microbiology* **47**, 590-592.

Everett KDE, Bush RM, Andersen AA (1999) Emended description of the order Chlamydiales, proposal of Parachlamydiaceae fam. nov. and Simkaniaceae fam. nov., each containing one monotypic genus, revised taxonomy of the family Chlamydiaceae, including a new genus and five new species, and standards for the identification of organisms. *International Journal of Systematic and Evolutionary Microbiology* **49**, 415-440.

Foster G, Holmes B, Steigerwalt AG*, et al.* (2004) *Campylobacter insulaenigrae* sp. nov., isolated from marine mammals. *International Journal of Systematic and Evolutionary Microbiology* **54**, 2369-2373.

Fox JG, Taylor NS, Howe S*, et al.* (2006) *Helicobacter anseris* sp. nov. and *Helicobacter brantae* sp. nov., isolated from feces of resident Canada geese in the greater Boston area. *Applied and Environmental Microbiology* **72**, 4633-4637.

Franklin CL, Beckwith CS, Livingston RS*, et al.* (1996) Isolation of a novel *Helicobacter* species, *Helicobacter cholecystus* sp. nov., from the gallbladders of Syrian hamsters with cholangiofibrosis and centrilobular pancreatitis. *J Clin Microbiol* **34**, 2952-2958.

Frasca S, Weber ES, Urquhart H*, et al.* (2005) Isolation and characterization of *Mycoplasma sphenisci* sp nov from the choana of an aquarium-reared jackass penguin (*Spheniscus demersus*). *Journal of Clinical Microbiology* **43**, 2976-2979.

Fraser G (1961) Isolation of an Anaerobic *Vibrio* from a King Penguin (*Aptenodytes longirostris*). *Avian Diseases* **5**, 243-250.

García-Peña FJ, Llorente MT, Serrano T*, et al.* (2017) Isolation of *Campylobacter* spp. from Three Species of Antarctic Penguins in Different Geographic Locations. *Ecohealth* **14**, 78-87.

Gerlach H (1994) Bacteria. In: *Avian Medicine: Principles anf Application* (eds. Ritchie BW, Harrison GJ, Harrison LR). Wingers Publishing Inc., Lake Worth, FL, USA.

González-Acuña D, Hernández J, Moreno L*, et al.* (2013) Health evaluation of wild gentoo penguins (*Pygoscelis papua*) in the Antarctic Peninsula. *Polar Biology* **36**, 1749-1760.

Government of South Georgia and the South Sandwich Islands (2005) South Georgia Government Update.

Government of South Georgia and the South Sandwich Islands (2013) Biosecurity Protocol 2013.

Goyache J, Ballesteros C, Vela AI*, et al.* (2003a) *Corynebacterium sphenisci* sp. nov., isolated from wild penguins. *International Journal of Systematic and Evolutionary Microbiology* **53**, 1009-1012.

Goyache J, Vela AI, Collins MD*, et al.* (2003b) *Corynebacterium spheniscorum* sp. nov., isolated from the cloacae of wild penguins. *International Journal of Systematic and Evolutionary Microbiology* **53**, 43-46.

Griekspoor P, Colles FM, McCarthy ND*, et al.* (2013) Marked host specificity and lack of phylogeographic population structure of *Campylobacter jejuni* in wild birds. *Mol Ecol* **22**, 1463-1472.

Guardiola-Avila I, Acedo-Felix E, Sifuentes-Romero I*, et al.* (2016) Molecular and Genomic Characterization of *Vibrio mimicus* Isolated from a Frozen Shrimp Processing Facility in Mexico. *PLoS One* **11**, e0144885-e0144885.

Gupta RS, Lo B, Son J (2018) Phylogenomics and Comparative Genomic Studies Robustly Support Division of the Genus *Mycobacterium* into an Emended Genus *Mycobacterium* and Four Novel Genera. *Frontiers in Microbiology* **9**.

Hampson DJ, La T, Phillips ND (2015) Emergence of *Brachyspira* species and strains: reinforcing the need for surveillance. *Porcine health management* **1**, 8-8.

Harkinezhad T, Geens T, Vanrompay D (2009) *Chlamydophila psittaci* infections in birds: A review with emphasis on zoonotic consequences. *Veterinary Microbiology* **135**, 68-77.

Haydushka IA, Markova N, Kirina V, Atanassova M (2012) Recurrent sepsis due to *Bacillus licheniformis*. *Journal of global infectious diseases* **4**, 82-83.

Hermans K, Devriese LA, De Herdt P, Godard C, Haesebrouck F (2000) *Staphylococcus aureus* infections in psittacine birds. *Avian Pathol* **29**, 411-415.

Hernández J, González-Acuña D (2016) Anthropogenic antibiotic resistance genes mobilization to the polar regions. *Infect Ecol Epidemiol* **6**, 32112.

Hobson RP, Williams A, Rawal K, Pennington TH, Forbes KJ (1995) Incidence and spread of *Haemophilus influenzae* on an Antarctic base determined using the polymerase chain reaction. *Epidemiology and Infection* **114**, 93-103.

Hughes LA, Bennett M, Coffey P*, et al.* (2009) Molecular Epidemiology and Characterization of *Campylobacter* spp. Isolated from Wild Bird Populations in Northern England. *Applied and Environmental Microbiology* **75**, 3007-3015.

Humbert MV, Christodoulides M (2020) Atypical, Yet Not Infrequent, Infections with *Neisseria* Species. *Pathogens* **9**, 10.

Humphrey S, Chaloner G, Kemmett K*, et al.* (2014) *Campylobacter jejuni* Is Not Merely a Commensal in Commercial Broiler Chickens and Affects Bird Welfare. *Mbio* **5**, e01364-01314.

Inglis GD, Hoar BM, Whiteside DP, Morck DW (2007) *Campylobacter canadensis* sp. nov., from captive whooping cranes in Canada. *International Journal of Systematic and Evolutionary Microbiology* **57**, 2636-2644.

Iveson JB, Shellam GR, Bradshaw SD*, et al.* (2009) *Salmonella* infections in Antarctic fauna and island populations of wildlife exposed to human activities in coastal areas of Australia. *Epidemiology and Infection* **137**, 858-870.

Janda JM, Abbott SL (1993) Infections associated with the genus *Edwardsiella*: the role of *Edwardsiella tarda* in human disease. *Clinical Infectious Diseases* **17**, 742-748.

Jansson DS, Mushtaq M, Johansson K-E*, et al.* (2015) Intestinal spirochaetes (genus *Brachyspira*) colonise wild birds in the southern Atlantic region and Antarctica. *Infection Ecology & Epidemiology* **5**, 29296.

Javed S, Gul F, Javed K, Bokhari H (2017) *Helicobacter pullorum*: An Emerging Zoonotic Pathogen. *Frontiers in Microbiology* **8**, 604-604.

Jensen SK, Nymo IH, Forcada J, Hall A, Godfroid J (2013) *Brucella* antibody seroprevalence in Antarctic seals (*Arctocephalus gazella, Leptonychotes weddellii* and *Mirounga leonina*). *Diseases of Aquatic Organisms* **105**, 175-181.

Johansson H, Ellström P, Artursson K*, et al.* (2018) Characterization of *Campylobacter* spp. isolated from wild birds in the Antarctic and Sub-Antarctic. *PLoS One* **13**, e0206502.

Johnson AP (1983) The pathogenic potential of commensal species of *Neisseria*. *Journal of clinical pathology* **36**, 213-223.

Jorge MC, Najle R, Montalti D (2002) Cloacal bacterial flora on Antarctic birds. *Rivista Italiana di Ornitologia* **71**, 163-169.

Jørgensen SL, Poulsen LL, Thorndal L*, et al.* (2017) Characterization of *Enterococcus faecalis* isolated from the cloaca of 'fancy breeds' and confined chickens. *Journal of Applied Microbiology* **122**, 1149-1158.

Kämpfer P, Lodders N, Martin K, Avendaño-Herrera R (2012) *Flavobacterium chilense* sp. nov. and *Flavobacterium araucananum* sp. nov., isolated from farmed salmonid fish. *International Journal of Systematic and Evolutionary Microbiology* **62**, 1402-1408.

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**, 772-780.

Kerr GR, Forbes KJ, Williams A, Pennington TH (1993) An analysis of the diversity of *Haemophilus parainfluenzae* in the adult human respiratory tract by genomic DNA fingerprinting. *Epidemiology and Infection* **111**, 89-98.

Kim J-H, Kim K-Y, Cha C-J (2009) *Flavobacterium chungangense* sp. nov., isolated from a freshwater lake. *International Journal of Systematic and Evolutionary Microbiology* **59**, 1754-1758.

Kim J-J, Kanaya E, Weon H-Y*, et al.* (2012) *Flavobacterium compostarboris* sp. nov., isolated from leaf-and-branch compost, and emended descriptions of *Flavobacterium hercynium*, *Flavobacterium resistens* and *Flavobacterium johnsoniae*. *International Journal of Systematic and Evolutionary Microbiology* **62**, 2018-2024.

Ko KS, Oh WS, Lee MY*, et al.* (2006) *Bacillus infantis* sp. nov. and *Bacillus idriensis* sp. nov., isolated from a patient with neonatal sepsis. *Int J Syst Evol Microbiol* **56**, 2541-2544.

Králová S, Busse HJ, Švec P*, et al.* (2019) *Flavobacterium circumlabens* sp. nov. and *Flavobacterium cupreum* sp. nov., two psychrotrophic species isolated from Antarctic environmental samples. *Syst Appl Microbiol* **42**, 291-301.

Laviad-Shitrit S, Izhaki I, Halpern M (2019) Accumulating evidence suggests that some waterbird species are potential vectors of *Vibrio cholerae*. *PLOS Pathogens* **15**, e1007814-e1007814.

Laviad -Shitrit S, Lev-Ari T, Katzir G*, et al.* (2017) Great cormorants (*Phalacrocorax carbo*) as potential vectors for the dispersal of *Vibrio cholerae*. *Sci Rep* **7**, 7973.

Le Roy CI, Mappley LJ, La Ragione RM, Woodward MJ, Claus SP (2015) *Brachyspira pilosicoli*-induced avian intestinal spirochaetosis. *Microbial ecology in health and disease* **26**, 28853-28853.

Leotta G, Vigo G, Giacoboni G (2006a) Isolation of *Campylobacter lari* from seabirds in Hope Bay, Antarctica. *Polish Polar Research* **27**, 303-308.

Leotta GA, Chinen I, Vigo GB, Pecoraro M, Rivas M (2006b) Outbreaks of avian cholera in Hope Bay, Antarctica. *Journal of Wildlife Diseases* **42**, 259-270.

Leotta GA, Pineyro P, Serena S, Vigo GB (2009) Prevalence of *Edwardsiella tarda* in Antarctic wildlife. *Polar Biology* **32**, 809-812.

Liu G, Tang CM, Exley RM (2015) Non-pathogenic *Neisseria*: members of an abundant, multi-habitat, diverse genus. *Microbiology* **161**, 1297-1312.

Loch TP, Faisal M (2015) Emerging flavobacterial infections in fish: A review. *Journal of Advanced Research* **6**, 283-300.

Mappley LJ, La Ragione RM, Woodward MJ (2014) *Brachyspira* and its role in avian intestinal spirochaetosis. *Vet Microbiol* **168**, 245-260.

McCammon SA, Bowman JP (2000) Taxonomy of Antarctic *Flavobacterium* species: description of *Flavobacterium gillisiae* sp. nov., *Flavobacterium tegetincola* sp. nov., and *Flavobacterium xanthum* sp. nov., nom. rev. and reclassification of [*Flavobacterium] salegens* as *Salegentibacter salegens* gen. nov., comb. nov. *International Journal of Systematic and Evolutionary Microbiology* **50**, 1055-1063.

McMurdie PJ, Holmes S (2013) phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* **8**, e61217.

Merck (2016) *The Merck Veterinary Manual, 11th Edition* Merck & Co., Inc., Whitehouse Station, NJ.

Mohammed A, Abdul-Wahab MF, Hashim M*, et al.* (2018) Biohydrogen Production by Antarctic Psychrotolerant *Klebsiella* sp. ABZ11. *Polish Journal of Microbiology* **67**, 283-290.

Nievas VF, Leotta GA, Vigo GB (2007) Subcutaneous clostridial infection in Adelie penguins in Hope Bay, Antarctica. *Polar Biology* **30**, 249-252.

Nimmervoll H, Wenker C, Robert N, Albini S (2011) Septicaemia caused by *Edwardsiella tarda* and *Plesiomonas shigelloides* in captive penguin chicks. *Schweizer Archiv Fur Tierheilkunde* **153**, 117-121.

Niskanen T, Waldenström J, Fredriksson-Ahomaa M, Olsen B, Korkeala H (2003) virF-positive *Yersinia pseudotuberculosis* and *Yersinia enterocolitica* found in migratory birds in Sweden. *Applied and Environmental Microbiology* **69**, 4670-4675.

Oelke H, Steiniger F (1973) *Salmonella* in Adelie penguins (Pygoscelis adeliae) and South Polar Skuas (Catharacta maccormicki) on Ross Island, Antarctica. *Avian Dis* **17**, 568-573.

Oliveira A, Oliveira LC, Aburjaile F*, et al.* (2017) Insight of Genus *Corynebacterium*: Ascertaining the Role of Pathogenic and Non-pathogenic Species. *Frontiers in Microbiology* **8**, 1937-1937.

Olsen B, Bergstrom S, McCafferty DJ, Sellin M, Wistrom J (1996) *Salmonella enteritidis* in Antarctica: Zoonosis in man or humanosis in penguins? *Lancet* **348**, 1319-1320.

Parker CT, Miller WG, Horn ST, Lastovica AJ (2007) Common genomic features of *Campylobacter jejuni* subsp. doylei strains distinguish them from *C. jejuni* subsp. jejuni. *BMC Microbiology* **7**, 50.

Parte AC (2013) LPSN—list of prokaryotic names with standing in nomenclature. *Nucleic Acids Research* **42**, D613-D616.

Parte AC (2018) LPSN – List of Prokaryotic names with Standing in Nomenclature (bacterio.net), 20 years on. *International Journal of Systematic and Evolutionary Microbiology* **68**, 1825-1829.

Pinpimai K, Roe WD, Biggs PJ, Dittmer KE, Michael SA (2018) Draft Whole-Genome Sequences of Five *Klebsiella pneumoniae* Isolates from the Subantarctic Islands of New Zealand. *Microbiology resource announcements* **7**, e01328-01318.

Ponpandian LN, Rim SO, Shanmugam G*, et al.* (2019) Phylogenetic characterization of bacterial endophytes from four *Pinus* species and their nematicidal activity against the pine wood nematode. *Sci Rep* **9**, 12457.

Retamal P, Llanos-Soto S, Salas LM*, et al.* (2017) Isolation of drug-resistant *Salmonella enterica* serovar enteritidis strains in gentoo penguins from Antarctica. *Polar Biology* **40**, 2531-2536.

Rodríguez-Medina N, Barrios-Camacho H, Duran-Bedolla J, Garza-Ramos U (2019) *Klebsiella variicola*: an emerging pathogen in humans. *Emerging microbes & infections* **8**, 973-988.

Roe WD, Rogers L, Pinpimai K*, et al.* (2015) Septicaemia and meningitis caused by infection of New Zealand sea lion pups with a hypermucoviscous strain of *Klebsiella pneumoniae*. *Veterinary Microbiology* **176**, 301-308.

Romalde JL, Ravelo C, Valdés I*, et al.* (2008) *Streptococcus phocae*, an emerging pathogen for salmonid culture. *Veterinary Microbiology* **130**, 198-207.

Sawicka A, Durkalec M, Tomczyk G, Kursa O (2020) Occurrence of *Mycoplasma gallisepticum* in wild birds: A systematic review and meta-analysis. *PLoS One* **15**, e0231545.

See-Too WS, Salazar S, Ee R*, et al.* (2017) *Pseudomonas versuta* sp. nov., isolated from Antarctic soil. *Systematic and Applied Microbiology* **40**, 191-198.

Stępień-Pyśniak D, Hauschild T, Nowaczek A, Marek A, Dec M (2018) Wild Birds as a Potential Source of Known and Novel Multilocus Sequence Types of Antibiotic-Resistant *Enterococcus faecalis*. *J Wildl Dis* **54**, 219-228.

Stipkovits L (1979) The pathogenicity of avian mycoplasmas. *Zentralbl Bakteriol Orig A* **245**, 171-183.

Stoutenburg JE, Swayne DE, Hoepf TM, Wack R, Kramer L (1995) Frequency of Intestinal Spirochetes in Birds, Rats, and Pigs from a Zoologic Collection and Private Rhea Farms in Ohio. *Journal of Zoo and Wildlife Medicine* **26**, 272-278.

Struve C, Krogfelt KA (2004) Pathogenic potential of environmental *Klebsiella pneumoniae* isolates. *Environ Microbiol* **6**, 584-590.

Tanaka S, Hoshino Y, Sakagami T*, et al.* (2019) Pathogenicity of *Mycolicibacterium phlei*, a non-pathogenic nontuberculous mycobacterium in an immunocompetent host carrying anti-interferon gamma autoantibodies: a case report. *BMC Infectious Diseases* **19**, 454.

Taurisano ND, Butler BP, Stone D*, et al.* (2018) *Streptococcus phocae* in Marine Mammals of Northeastern Pacific and Arctic Canada: A Retrospective Analysis of 85 Postmortem Investigations. *Journal of Wildlife Diseases* **54**, 101-111.

Tena D, Fernández C, Lago M, R. (2015) *Alcaligenes faecalis*: an Unusual Cause of Skin and Soft Tissue Infection. *Japanese Journal of Infectious Diseases* **68**, 128-130.

Valour F, Sénéchal A, Dupieux C*, et al.* (2014) Actinomycosis: etiology, clinical features, diagnosis, treatment, and management. *Infection and drug resistance* **7**, 183-197.

Van Trappen S, Tan T-L, Samyn E, Vandamme P (2005) *Alcaligenes aquatilis* sp. nov., a novel bacterium from sediments of the Weser Estuary, Germany, and a salt marsh on Shem Creek in Charleston Harbor, USA. *International Journal of Systematic and Evolutionary Microbiology* **55**, 2571-2575.

Vancanneyt M, Segers P, Hauben L*, et al.* (1994) *Flavobacterium meningosepticum*, a pathogen in birds. *Journal of Clinical Microbiology* **32**, 2398-2403.

Vecten M, Martel H, Casalta J-P*, et al.* (2017) Fatal *Neisseria macacae* infective endocarditis: first report. *Infection* **45**, 369-371.

Vigo GB, Leotta GA, Ines Caffer M*, et al.* (2011) Isolation and characterization of *Salmonella enterica* from Antarctic wildlife. *Polar Biology* **34**, 675-681.

Waldenström J, Broman T, Carlsson I*, et al.* (2002) Prevalence of *Campylobacter jejuni, Campylobacter lari,* and *Campylobacter coli* in different ecological guilds and taxa of migrating birds. *Applied and Environmental Microbiology* **68**, 5911-5917.

Wareth G, Kheimar A, Neubauer H, Melzer F (2020) Susceptibility of Avian Species to *Brucella* Infection: A Hypothesis-Driven Study. *Pathogens* **9**.

White FH, Simpson CF, Williams LE, Jr. (1973) Isolation of *Edwardsiella tarda* from aquatic animal species and surface waters in Florida. *J Wildl Dis* **9**, 204-208.

Woods R, Jones HI, Watts J, Miller GD, Shellam GR (2009) Diseases of Antarctic Seabirds. In: *Health of Antarctic wildlife: a challenge for science and policy* (eds. Kerry KR, Riddle M). Springer, London.

Yi H, Chun J (2006) *Flavobacterium weaverense* sp. nov. and *Flavobacterium segetis* sp. nov., novel psychrophiles isolated from the Antarctic. *International Journal of Systematic and Evolutionary Microbiology* **56**, 1239-1244.

Yi H, Oh HM, Lee JH, Kim SJ, Chun J (2005) *Flavobacterium antarcticum* sp. nov., a novel psychrotolerant bacterium isolated from the Antarctic. *Int J Syst Evol Microbiol* **55**, 637-641.

Zamora L, Fernández-Garayzábal JF, Sánchez-Porro C*, et al.* (2013) *Flavobacterium plurextorum* sp. nov. Isolated from Farmed Rainbow Trout (Oncorhynchus mykiss). *PLoS One* **8**, e67741.

Zhang D-C, Wang H-X, Liu H-C, Dong X-Z, Zhou P-J (2006) *Flavobacterium glaciei* sp. nov., a novel psychrophilic bacterium isolated from the China No.1 glacier. *International Journal of Systematic and Evolutionary Microbiology* **56**, 2921-2925.

Zhang G-Q, Liu Q, Liu H-C, Zhou Y-G, Xin Y-H (2019) *Flavobacterium ranwuense* sp. nov., isolated from glacier. *International Journal of Systematic and Evolutionary Microbiology* **69**, 3812-3817.

Zhou M-Y, Zhang Y-J, Zhang X-Y*, et al.* (2018) *Flavobacterium phocarum* sp. nov., isolated from soils of a seal habitat in Antarctica. *International Journal of Systematic and Evolutionary Microbiology* **68**, 536-541.

Chapter 9. Eukaryotic Gastrointestinal Parasites of Gentoo Penguins and Sympatric Seabirds

Hila Levy[1], Stephen G. Preston[1], Patrik G. Flammer[1], Tom Hart[1], Adrian L. Smith[1]

1. Department of Zoology, University of Oxford, South Parks Road, Oxford, OX1 3SZ, United Kingdom

Keywords: penguins, parasites, 18S, *Eimeria*, protozoa, nematodes, Antarctica, *Pygoscelis*

Abstract

Parasite communities are influenced by their host's ecology and phylogeny. Knowledge of eukaryotic parasite diversity is limited among Antarctic seabirds, who breed in dense colonies and consume a variety of vertebrate and invertebrate marine prey. In this study, we use microscopy and 18S ribosomal small subunit parallel sequencing of faecal samples to determine parasite presence and prevalence in a wide-ranging but non-migratory host, the Gentoo Penguin, *Pygoscelis papua*, alongside the congeneric Adélie Penguin *P. adeliae* and Chinstrap Penguin, *P. antarcticus*, and a sympatric flying seabird, the Blue-Eyed Shag (*Phalacrocorax atriceps bransfieldiensis*). We found 205 unique sequences assigned to parasitic taxa within the phyla Apicomplexa, Acanthocephala, Nematoda, and Platyhelminthes. We present a description of the prevalence and geographic distribution of endoparasites in each host, while uncovering a similar numbers of taxa that are either host-specific and geographically limited or broadly distributed and shared across the genus *Pygoscelis* or beyond. Our study lends support to the idea that Antarctic microfaunal diversity is underestimated and greater effort is needed to fully describe and characterise parasites through genetic and morphological means.

Introduction

By definition, parasites are symbionts that draw resources away from their hosts, negatively affecting them in a variety of ways (Anderson & May 1982). Endoparasites of vertebrate gastrointestinal tracts can affect host condition, cause clinical symptomatology, and decrease reproductive fitness (Poulin 2007). Though captive and experimental conditions allow these effects to be directly linked, our understanding of parasitic impacts in wild species is more limited. In the Antarctic and sub-Antarctic, explorations of parasite diversity, abundance, and ecosystem interactions have been limited (Palacios *et al.* 2012). Though megafauna are comparatively well documented in their diversity and distributions, microscopic biodiversity continues to be uncovered (Chown *et al.* 2015).

In seabirds, parasite burdens can affect host resiliency to stress, environmental change and pollution (Khan *et al.* 2019). In light of increasingly documented anthropogenic impacts in the Southern Ocean from fisheries, contamination, visitation, and invasive species introductions (Barbosa 2011; Cimino *et al.* 2016; Constable *et al.* 2014; Lescroel *et al.* 2014; Lynch *et al.* 2010; Trathan *et al.* 2015; Whinam *et al.* 2005), the role of parasitism on avian hosts invites a survey of the spatiotemporal interactions between host assemblages and parasites. In particular, we seek to determine whether parasites are shared across sympatric species or are host-specific, and consider their modes of transmission.

Penguins are diving seabirds in the Order Sphenisciformes that make up more than 90% of avian biomass in the Southern Ocean (Williams 1995). They are marine mesopredators, primarily feeding on fish, krill, and squid, among other minor prey items (Polito *et al.* 2011). Present on the islands of the South Atlantic Ocean's Scotia Arc and the Antarctic Peninsula, the genus *Pygoscelis* contains three related penguin species: Adélie Penguins (*P. adeliae*), Chinstrap Penguins (*P. antarcticus*), and Gentoo Penguins (*P. papua*). Gentoo Penguins present strong population genetic structure (Clucas *et al.* 2014; Clucas *et al.* 2018; Levy *et al.* 2016), and are non-

migratory, remaining close to their colonies year-round, while Adélie Penguins and Chinstrap Penguins disperse further afield in the non-breeding season, with high levels of gene flow and host connectivity (Clucas *et al.* 2014; Clucas *et al.* 2018; Freer *et al.* 2015; Roeder *et al.* 2001). These penguins come ashore annually to breed, nesting in densely packed colonies that sometimes shelter two or more of these host species, along with flying seabirds like albatrosses, petrels, gulls, and cormorants, as well as scavenging skuas and sheathbills (Clements *et al.* 2019). Hence there are two main opportunities for gastrointestinal parasites to transfer between penguins (or to be introduced to penguins), via faecal oral or food-associated transmission.

Eukaryotic gastrointestinal parasites have been documented in Pygoscelid penguins using necropsies for gross examination, along with microscopy of histological tissues and faeces. Commonly detected eukaryotic gut parasites in these wild hosts include apicomplexans (*Cryptosporidium* spp., *Eimeria* spp., and *Isospora* spp.), acanthocephalans (*Corynosoma* spp.), cestodes (*Diphyllobothrium* spp., *Parorchites zederi*, and *Tetrabothrius* spp.), and nematodes (*Contracaecum* spp., *Stegophorus macronectes*, *Streptocara* spp., and *Tetrameres* spp.) (Diaz *et al.* 2017).

Molecular characterisation of Pygoscelid penguin eukaryotic parasites has only been performed on individual parasites retrieved from tissue (the gut nematode *Stegophorus macronectes*) or blood (a haemoparasite in the genus *Babesia*) taken from Chinstrap Penguins on Deception Island (Montero *et al.* 2016; Vidal *et al.* 2016). On the other hand, studies targeting fragments of the eukaryotic 18S ribosomal DNA gene in faecal samples have amplified signals of gut parasite presence as a byproduct of intended seabird dietary analyses in albatrosses and Adélie Penguins (Jarman *et al.* 2013; McInnes *et al.* 2017; McInnes *et al.* 2016).

We sought to test whether eukaryotic parasites differ between species and regions and whether co-occurrence of sympatric host species influences the parasite load. Using the dual approach of microscopy in faeces (154 Pygoscelid penguin scats) and eukaryotic 18S sequencing,

we characterised the eukaryotic gut parasite communities of 436 Gentoo Penguins sampled across a large geographic range, alongside a smaller set of sympatric seabirds, Adélie Penguins (n = 6), Chinstrap Penguins (n = 20), and Blue-Eyed Shags (*Phalacrocorax atriceps bransfieldiensis* synonymously known as *Leucocarbo bransfieldiensis*, n = 18).

Materials and Methods

Sample Collection

Across four austral summers (2014/15, 2015/16, 2016/17, and 2017/18), 581 samples (580 guano and 1 cloacal swab) were collected from seabird breeding colonies across the Scotia Arc and Antarctic Peninsula, from 30 unique sites. Guano was collected from four species: Adélie Penguins (3 sites, *n* = 26), Chinstrap Penguins (4 site, *n* = 61), Gentoo Penguins (24 sites, *n* = 476), and Blue-Eyed Shags (3 sites, n = 18). To collect faecal material, approximately 1.15 g (range: 0.1214 – 3.2825 g) of fresh guano was scooped into a 15 mL plastic conical centrifuge tube (Corning®, USA), 5 mL screw cap cryotube, or 1.5 mL microcentrifuge (Eppendorf®) tube. Samples from the 2015/16 season were stored in 2.5% w/v potassium dichromate solution (as a prelude to microscopic examination), which inhibits bacterial and fungal growth, and some samples from the 2014/15 season were stored with no preservative (due to logistical constraints, see **Table 1**). For all other samples, approximately ~0.3 g (range: 0 – 0.776 g) was placed into RNAlater® preservative solution in a 1.5 mL microcentrifuge (Eppendorf®) tube. Sampling was distributed throughout the extent of the colony to reduce bias, though fresh, more voluminous excreta of any colour were collected as a preference over dried material. Tubes were sealed and labelled, and stored at 4° C until shipping. Samples spent between 60-90 days in transport at room temperature prior to arriving at the laboratory, where they were catalogued and stored at -20° C for RNAlater samples and those with no preservative, or room temperature for potassium dichromate samples. In addition, a single Gentoo Penguin cloacal swab (Eswab™ regular flocked swab in 1 ml of Amies liquid transport medium, Copan®, Murrieta, CA, USA) was obtained from a

carcass found at Saunders Neck, Falkland Islands/Malvinas, stored frozen aside from transport on cold packs.

Sampling sites are shown in **Figure 1** and details are found in **Table 1**. All sampling in the Antarctic Treaty Area was carried out in accordance with United Kingdom Home Office guidelines, under permits granted by the United Kingdom Foreign and Commonwealth Office (permits: S3-23/2013, S7 28/2013, S3 04/2014, S7 03/2014, 33/2016, 34/2016, 37/2017), in South Georgia under permits granted by the Government of South Georgia and the South Sandwich Islands (SCI-2014-017, RAP-2015-018, RAP-2016-035 and RAP-2017-034), and in the Falkland Islands under Falkland Islands Government Environmental Planning Department permits (R16.2014, R15 2014, and R04 2017), with ethical approval from the University of Oxford Animal Welfare and Ethical Review Board. All samples were transported in accordance with applicable export permits and United Kingdom Department for Environment, Food and Rural Affairs import permits.

DNA Extraction

To extract DNA from guano, the MoBio PowerSoil® DNA Extraction Kit (after 2016, rebranded as the Qiagen QIAamp® PowerFecal® DNA Kit) was used with approximately 0.1 g of starting material, weighed and drained from its preservative. DNA from the cloacal swab was extracted from 200 µL of the swab suspension as starting material, after vortexing. Potassium dichromate samples were washed three times by adding 1 mL of MilliQ water, mixing by inversion and centrifugation for 1 minute at 13,400 rpm.

The extraction kit protocol was modified such that the initial lysis step was extended to 12-18 hours at 65°C on a heated shaker block, and the subsequent homogenisation (bead-beating) step was carried out on an MP Biomedicals™ FastPrep-24™ ("Marine Sediment" Program: 2 cycles of 40 seconds at 5.5 m/s with a 5 minute pause between cycles) instead of a vortex. All other steps were per protocol and eluted DNA was stored at -20°C.

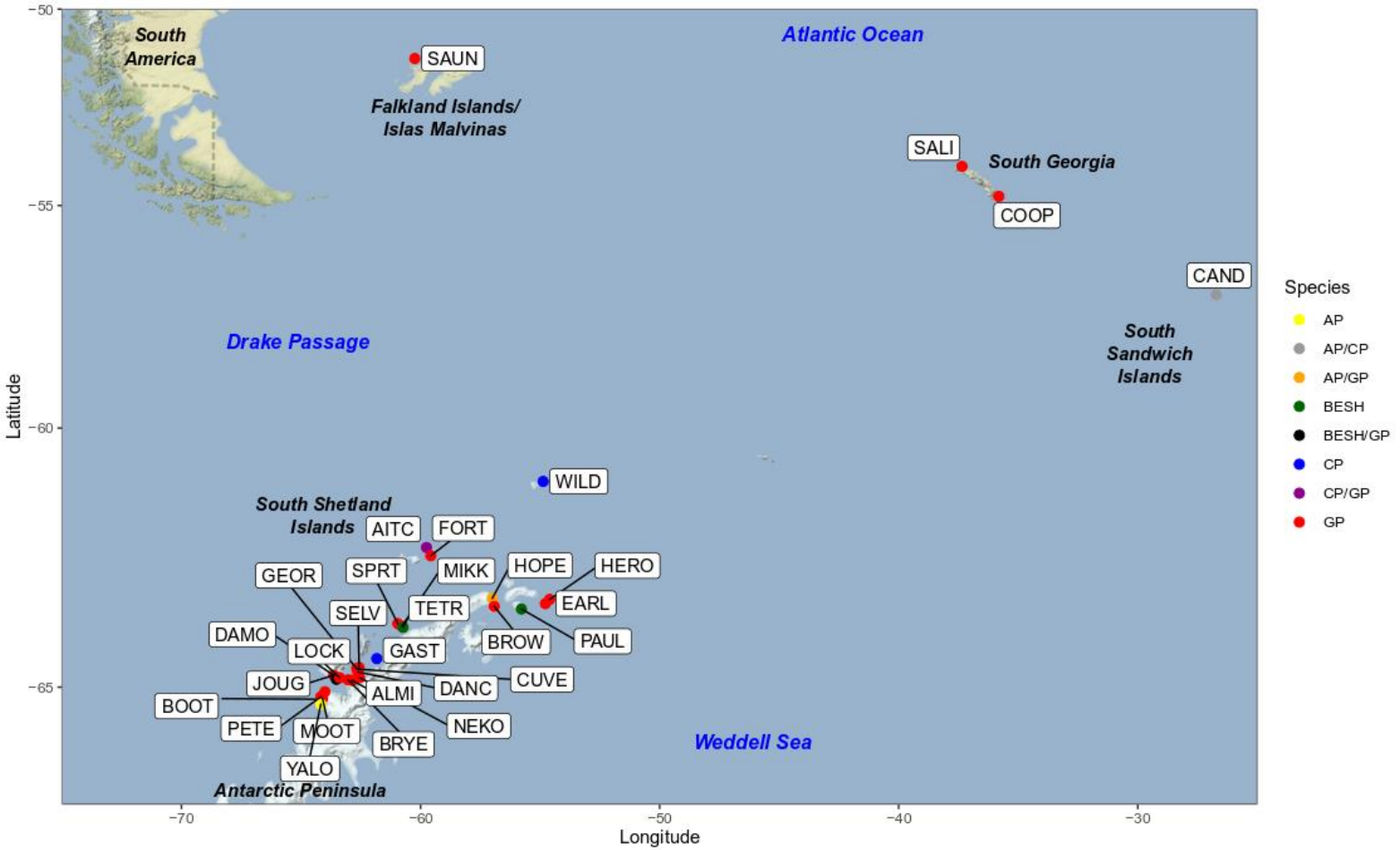


Figure 7. Sampling sites in the Scotia Arc (Falkland Islands/Islas Malvinas, South Georgia and the South Sandwich Islands, and the Western Antarctic Peninsula). Species sampled are designated by colour in the legend. AP = Adélie Penguin, BESH = Blue-Eyed Shag, CP = Chinstrap Penguin, GP = Gentoo Penguin. Colony names are expanded in Table 1.

Microscopy

Pygoscelid penguin faecal samples collected in 2014/15 (n = 154, underlined in **Table 1**) were analysed microscopically for parasite presence. Approximately 0.1 g of raw guano was placed into a 1.5 mL microcentrifuge tube. The tube was weighed empty, and then full, to obtain the amount of raw material transferred. 100 µL of double distilled water was added to the tube, which was vortexed and left overnight at room temperature. At least 12 hours later, 900 µL of saturated NaCl solution was added to the tube, which was briefly vortexed to allow clumps to come into contact with the salt solution. The tubes were left to stand for ten minutes to allow parasites to float to the surface. A disposable Pasteur pipette was used to draw up to 1 ml of solution from the surface into a McMaster counting chamber, which was allowed to stand for an additional ten minutes to allow further flotation of parasites. Using low power, each sample was screened in the counting chamber for the presence of coccidians, helminths, or other items of microscopic interest using a Nikon Eclipse E400 microscope with a 10× lens (Nikon UK, Kingston-Upon-Thames, UK). Parasite counts were logged. Photographs were recorded on a QImaging MP5.0 RTV camera (QImaging, Surrey, Canada) whenever a helminth, egg, or oocyst was encountered. For the few samples that had too many oocysts to count, one grid segment of the chamber was counted and then multiplied by the number of grids to obtain an approximate oocyst count. The total number of oocysts or helminths were divided by the starting weight of the sample to obtain oocysts per gram of faeces and helminths per gram of faeces.

Samples from subsequent sampling seasons that presented high abundances of coccidian parasite reads following molecular analysis (methods below) were analysed using light microscopy to attempt morphological characterisation. These samples used a McMaster counting chamber for flotation and imaging at low power (10x). For photography at higher power (20x or 40x), a drop of vortexed guano material was visualised directly on a slide with a coverslip.

Table 7. Table of guano sampling regime for microscopic (underlined) and molecular (**bolded**) parasite analysis. Samples were stored in RNAlater with the following exceptions: 2015/2016 samples indicated in *italics* were stored in potassium dichromate, samples from HERO and EARL marked with an asterisk (*) were divided and stored in both media, and 2014/15 samples with superscript cross (†) were stored without preservative (microscopy only). Species: Adélie Penguin (AP), Blue-Eyed Shag (BESH), Chinstrap Penguin (CP), Gentoo Penguin (GP). Regions north to south: Falkland/Malvinas Islands (FI), South Georgia (SG), South Sandwich Islands (SSI), Elephant Island (EI), South Shetland Islands (SH), Northeast Antarctic Peninsula (NEAP), Central Western Antarctic Peninsula (CWAP) as classified by the Antarctic Site Inventory (ASI).

Colony (ASI Site Code)	Region	Latitude	Longitude	Species	Samples (n) per Species and Breeding Season				
					2014/2015	2015/2016	2016/2017	2017/2018	Total (All Years)
Saunders Neck (SAUN)	FI	-51.31	-60.24	GP	**10**	***10***		**20**	40
Salisbury Plain (SALI)	SG	-54.05	-37.35	GP	**13**	***10***	**15**	**21**	59
Cooper Bay (COOP)	SG	-54.78	-35.81	GP	**10**				10
Candlemas Island (CAND)	SSI	-57.09	-26.72	AP	10				10
				CP	10				10
Point Wild (WILD)	EI	-61.10	-54.87	CP	11				11
Barrientos Island (AITCH)	SH	-62.41	-59.75	CP	10†			**20**	30
				GP	10†			**20**	30
Fort Point (FORT)	SH	-62.57	-59.57	GP				**6**	6

Hope Bay (HOPE)	NEAP	-63.38	-57.00	AP				**6**	6
				GP				**26**	26
Heroina Island (HERO)	NEAP	-63.40	-54.60	GP		**10***			10
Earle Island (EARL)	NEAP	-63.48	-54.78	GP		**10***			10
Brown Bluff (BROW)	NEAP	-63.53	-56.92	GP	**10**			**20**	30
Paulet Island (PAUL)	NEAP	-63.58	-55.78	BESH				**6**	6
Spert Island (SPRT)	CWAP	-63.85	-60.95	GP				**1**	1
Mikkelsen Harbor (MIKK)	CWAP	-63.90	-60.78	GP	**10**	***10***		**20**	40
Tetrad Islands (TETR)	CWAP	-63.92	-60.73	BESH				**7**	7
Gaston Island (GAST)	CWAP	-64.49	-61.83	CP	10†				10
Selvick Cove (SELV)	CWAP	-64.65	-62.57	GP				**2**	2
Georges Point (GEOR)	CWAP	-64.67	-62.67	GP		***9***			9
Cuverville Island (CUVE)	CWAP	-64.68	-62.63	GP		***10***		**9**	19
Danco Island (DANC)	CWAP	-64.73	-62.62	GP				**6**	6
Damoy Point (DAMO)	CWAP	-64.82	-63.53	GP				**3**	3
Jougla Point (JOUG)	CWAP	-64.83	-63.5	BESH				**5**	5
				GP				**23**	23
Neko Harbor (NEKO)	CWAP	-64.83	-62.55	GP		***10***		**10**	20
Port Lockroy (LOCK)	CWAP	-64.83	-63.5	GP	10†			**3**	13

Bryde Island East (BRYE)	CWAP	-64.87	-63.03	GP		***10***		**24**	34
Almirante Brown Station (ALMI)	CWAP	-64.88	-62.87	GP	10†			**5**	15
Booth Island (BOOT)	CWAP	-65.08	-64	GP		***10***		**20**	30
Petermann Island (PETE)	CWAP	-65.17	-64.17	GP		***10***		**5**	15
Moot Point (MOOT)	CWAP	-65.2	-64.1	GP	10†	***9***		**6**	25
Yalour Islands (YALO)	CWAP	-65.23	-64.17	AP	10†				10
Totals (Species)					154 (20 AP, 41 CP, 93 GP)	118 (GP)	15 (GP)	294 (6 AP, 18 BESH, 20 CP, 250 GP)	581 (26 AP, 18 BESH, 61 CP, 476 GP)

18S rDNA Amplification and Sequencing

A ~170 bp fragment of the V7 region of the eukaryotic nuclear small subunit (SSU, 18S) ribosomal DNA gene was amplified using primers designed for this region, as described in McInnes et al. (2016): 18S_SSU3_F (GGTCTGTGATGCCCTTAGATG) and 18S_SSU3_R (GGTGTGTACAAAGGGCAGGG). The forward primer had a unique 7 nt 'barcode' tag attached for amplification of each sample to allow pooling of PCR products prior to library preparation. DNA from each guano sample was amplified in triplicate with different barcodes, to ensure that primer barcode bias did not affect amplification success, and to avoid run-specific biases. PCR was carried out in 25 µL reactions using Phusion Hot Start Flex DNA polymerase enzyme (New England Biolabs, UK, M0535) according to manufacturer's recommendations, with 2 µL of faecal DNA diluted 1:10 in water to minimise inhibitors. Thermal cycling conditions were 98°C for 2 min; followed by 40 cycles of 98°C for 5 seconds, 67°C for 20 seconds, 72°C for 20 seconds; with a final extension of 72°C for 1 minute. A water control was included for every 48 samples.

PCR products were visualised on a 1% agarose gel. A subset of all reactions was quantified using the Qubit 1X dsDNA High-Sensitivity Assay Kit (Invitrogen, UK, Q33231) using a Qubit Fluorometer, normalising DNA quantities from each reaction prior to pooling. Up to 96 PCR reactions from different samples with individually tagged 18S primers were combined and cleaned using the MinElute PCR purification kit (Qiagen, UK, 28004). Library preparation was performed using the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® (New England Biolabs, UK, E7103S) with according to the manufacturer's instructions, with no size selection. NEBNext Multiplex Oligos for Illumina (New England Biolabs, UK, E7335S and E7500S) indexing primers were used in order to run multiple libraries on the MiSeq machine in parallel. TapeStation (Agilent Genomics) and qPCR (NEBNext® Library Quant Kit for Illumina®, E7630S, New England Biolabs, UK) analyses were used to confirm expected template size and success of the library preparation before loading onto the MiSeq along with 5% phiX at the University of Oxford,

Department of Zoology. The 300-cycle MiSeq Reagent Kit v2 (Illumina, UK, MS-102-2002) was used, giving paired 150 nucleotide reads.

Bioinformatics and Taxonomic Assignment

A custom Python script was used to demultiplex each library according to barcode, and to remove terminal primer sequences. This script made use of the NumPy v1.16.6 (Oliphant 2006) and python-Levenshtein PyPl v0.12.0 packages. Any reads still containing a copy of either primer sequence after this trimming step were identified using cutadapt v1.15 (Martin 2011) and excluded from further analyses, on the assumption that they represent PCR artefacts (e.g. concatamers). The package dada2 v1.12.1 (Callahan *et al.* 2016) was used in R v3.6.2 (The R Foundation, https://cran.r-project.org/) to process the sequencing data into counts of "Amplicon Sequence Variants" (ASVs), predictions of the actual biological sequences present. Processing with dada2 consisted of several stages:

1. Trimming sequences at the 3' end to decrease overall error rate (Illumina sequencing typically sees a drop off in sequence quality towards the end of a read) and filtering to remove sequences which had a high expected error rate even after trimming. The fastq_eestats2 command in USEARCH v10.0.240 (Edgar 2013) was used to investigate the effect of different trimming length/expected error filtering thresholds on yield. Reads were trimmed to 120bp and filtered to a maximum expected error rate of 1.
2. Inference of ASVs. Because the data were generated over two MiSeq runs, inference was performed in two stages, according to the method suggested by the author of the dada2 package [https://github.com/benjjneb/dada2/issues/479]. This is necessary because each run has a unique error-rate profile, and so requires its own error model. The approach is to perform inference on each run separately, then to rerun the inference for each using the ASVs predicted in the initial inference in *any* of the MiSeq runs as "priors". This should allow the two runs to converge towards a shared set of ASVs, reducing the

risk of a given biological sequence being described by two very similar, but distinct, ASVs in the data from different MiSeq runs, while also maximising the sensitivity of the algorithm for rare sequences. For the first round of inference runs, dada2 was set to pool all samples from each run for inference, to maximise sensitivity to rare variants (pool = TRUE). ASVs which were identified in at least 6 PCR reactions across the two runs were defined as priors for the second inference runs, for which dada2 was set to infer ASVs from each sample separately (pool = FALSE), though of course with reference to the supplied priors (this mimics the operation of dada2 in pool = 'pseudo' mode, while allowing for combined processing of two MiSeq runs.

3. Merging read pairs, using the mergePairs function.
4. Pooling of read counts of ASVs with their reverse complement where present. Because the v1.12.1 dada2 does not have this functionality (it has subsequently been added), a custom R script was used. The script uses the Biostrings package v2.54.0 (Pagès *et al.* 2019) for sequence handling and the DECIPHER package v2.14.0 (Wright 2016) to put all sequences into the same orientation.
5. Removal of ASVs which are likely to represent PCR chimeras, using the removeBimeraDenovo function with method = 'consensus'.

Reads from triplicate PCRs were combined to produce a single set of ASV counts per sample. Taxonomy was allocated to all ASVs that represented >=0.15% of total reads in at least one sample by searching the NCBI BLAST SSU eukaryote BLAST database (RefSeq release 98, O'Leary *et al.* (2016)) and selecting the lowest common taxonomic rank of the top hits. Searching, parsing of BLAST results, and selection of taxonomic rank were performed using custom python and R scripts. The R scripts used the taxonomizr package v0.5.3 (Sherrill-Mix 2019) to query the NCBI taxonomy database and select lowest common ranks, and the data.table package v1.12.8 (Dowle & Srinivasan 2019) for data handling.

1. BLAST searches were performed by calling a local instance of BLAST 2.10.0+ with the following settings:
    - -task megablast
    - -perc_identity 90
    - -evalue 0.00001
2. Results were parsed to extract query and sequence ID, hit length, hit bitscore, and hit percentage identity for each hit.
3. Taxonomy was assigned in the following manner:
    - all hits of <150bp were removed
    - if there was a single top bitscore hit of 100% identity to the query sequence, taxonomy was assigned according to this hit alone
    - or if there were multiple top bitscore hits of 100% identity, taxonomy was assigned down to the level of the lowest common rank shared between the top hits, with lower taxonomic ranks being left unassigned
    - or if the top bitscore hit was of <100% identity, taxonomy was assigned down to the level of the lowest common rank shared between the top hit(s) and the hit(s) with the second highest bitscore, with lower taxonomic ranks being left unassigned
    - finally, for sequences where there were no BLAST hits, taxa were not assigned to any level.

Read assignment was manually examined to determine taxonomic groups (phyla/classes) of endoparasitic interest. Sequences assigned to these phyla and classes were parsed for further analysis and visualisation in R. **Supplementary Figure S1** presents a summary visualisation of the relative proportions of reads categorised as belonging to the host, environmental eukaryotes (e.g. phytoplankton), contaminants (e.g. mammals and fungi), prey items, parasites, and unassignable (no phylum classification). All resulting parasite sequences were aligned using

MAFFT v7.450 (Katoh & Standley 2013) in Geneious Prime 2020.0.3 (https://www.geneious.com). Neighbour-Joining trees were constructed in Geneious Prime (HKY distance model) to visually examine the topology of phylogenetic relatedness to known eukaryotic parasite reference sequences. A tree of parasite sequences from this study, abundance tables, taxonomic classifications, and sample metadata were loaded into R for analysis and plotting using phyloseq 1.30.0 (McMurdie & Holmes 2013) and ggplot2 (Wickham 2016).

Parasite community analyses

Alpha diversity was calculated in phyloseq, with differences between two groups calculated using Welch's two-sample t-test (Moser & Stevens 1992) in R. We used two different approaches to visualise differences in community composition (beta diversity) between groups for assessment of methodological differences (storage media), host species, and geographic regions. Ordinations were performed using non-metric multidimensional scaling (Minchin 1987) in phyloseq, employing both a count-based approach with the Bray-Curtis dissimilarity, and a quantitative approach that takes into account the degree of divergence, weighted UniFrac (Lozupone *et al.* 2007).

Results and Discussion

Preliminary Microscopy

Of the 154 samples visually examined for parasite presence in the 2014/15 sampling season, a total of 23 samples contained oocysts resembling *Eimeria* spp. (range: 8.19 to 36,319.62 oocysts per gram of faeces) and 53 samples contained helminths or their eggs (range: 6.27 to 416.66 helminths per gram of faeces). A total of 5 samples (4 from Gentoo Penguins, and 1 from a Chinstrap Penguin) contained both coccidians and helminths. Helminths were in poor condition, and eggs were photographed but could not be identified, even to the phylum level. The distribution of positive samples by colony and species is depicted in **Table 2** and on the maps in **Figure 2**, while distribution by host species is depicted in **Supplementary Figure S2**.

Table 2. Number of samples positive for *Eimeria* spp., helminths, or both, based on visual examination of guano samples, grouped by species per colony. Regions north to south: Falkland/Malvinas Islands (FI), South Georgia (SG), South Sandwich Islands (SSI), Elephant Island (EI), South Shetland Islands (SH), Northeast Antarctic Peninsula (NEAP), Central Western Antarctic Peninsula (CWAP).

Site	Region	Species	*n Eimeria*-positive	*n* Helminth-positive	*n* Positive for both	*n* Inspected
Saunders Neck (SAUN)	FI	Gentoo	5	5	2	10
Salisbury Plain (SALI)	SG	Gentoo	6	5	0	13
Cooper Bay (COOP)	SG	Gentoo	3	2	1	10
Candlemas Is. (CAND)	SSI	Chinstrap	0	2	0	10
		Adélie	0	2	0	10
Point Wild (WILD)	EI	Chinstrap	0	2	0	11
Barrientos Is. (AITC)	SH	Chinstrap	1	5	1	10
		Gentoo	1	0	0	10
Brown Bluff (BROW)	NEAP	Gentoo	0	3	0	10
Mikkelsen Harbor (MIKK)	CWAP	Gentoo	1	4	1	10
Gaston Island (GAST)	CWAP	Chinstrap	0	5	0	10
Port Lockroy (LOCK)	CWAP	Gentoo	0	4	0	10
Almirante Brown (ALMI)	CWAP	Gentoo	3	3	0	10
Moot Point (MOOT)	CWAP	Gentoo	1	4	0	10
Yalour Islands (YALO)	CWAP	Adélie	2	4	0	10
		Totals:	**23**	**53**	**5**	**154**

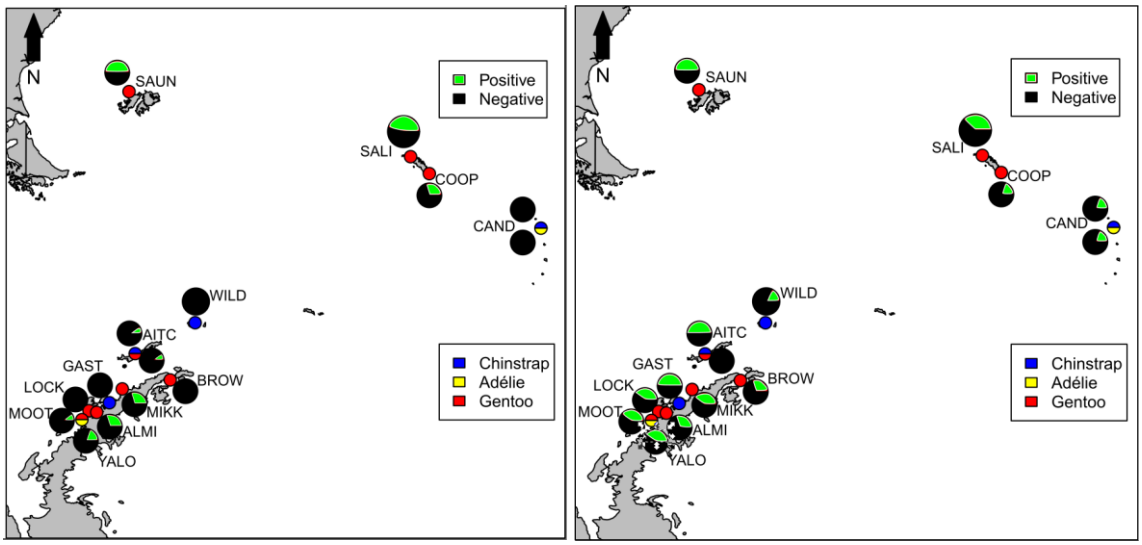


Figure 8. Maps depicting microscopy sampling locations and prevalence (% of samples with presence in pie charts) of coccidian oocysts (left) and helminths (right) detected in guano samples. Site colours represent species sampled at the site. Site names expanded in **Table 1**.

Coccidiosis, the disease resulting from intestinal infestations, typically of parasites in the genera *Eimeria*, is well-studied and prevalent among livestock and poultry, where it is commercially costly to control (Williams 1999). Several reports exist of epizootics of mortality-inducing intestinal coccidiosis in wild birds (Carpenter *et al.* 1984; Novilla & Carpenter 2004; Novilla *et al.* 1989; Windingstad *et al.* 1980). A previous mass mortality event in 1980 of Little Penguins (*Eudyptula minor*) on the Australian coast of Victoria found intestinal coccidiosis during the necropsy of 5 birds (1 adult, 2 juvenile, 2 of unknown age) out of 48 dead penguins. Coccidian schizonts were found in the small intestine of one bird. The researchers also observed unsporulated coccidian oocysts and attempt to sporulate them in laboratory conditions, but were unable to do so, and therefore did not identify or further describe the parasite species (Obendorf & McColl 1980). Experimental studies on wild birds have shown that coccidian infection causes the reduction of egg production and fertility, delayed maturation, and mortality in avian species

(Friend & Franson 1999; Yabsley 2009). As a result, there is increasing interest in characterising the prevalence and pathology of coccidians in other wild birds (Oliveira *et al.* 2017).

Microscopy can be useful for species identification, though many species cannot be distinguished using oocyst morphology alone particularly if they are not sporulated. Unsporulated oocysts were present in all three Pygoscelid species we screened visually, in all areas except for the South Sandwich Islands and Northern Antarctic Peninsula. Gentoo Penguins were most commonly sampled, and therefore represented the majority of positive samples by microscopy. As a percentage of samples from each species, Gentoo Penguins presented an *Eimeria* prevalence of 21%, compared to 10% in Adélie Penguins and 2.4% in Chinstrap Penguins.

Coccidians were first described in Pygoscelid penguins by Golemansky (2002), who described two species of *Eimeria* and one of *Isospora*, with one being fully described a new species, *Eimeria pygosceli*, from faecal samples obtained from Gentoo, Adélie, and Chinstrap Penguins inhabiting Livingston Island, in the South Shetland Islands. The species level identification of the other *Eimeria* and *Isospora* species was not possible due to lack of sporulation (Golemansky 2003). Subsequent observations in penguin hosts included findings of *Eimeria* in Galápagos Penguins (*Spheniscus mendiculus*) that could not be identified due to poor preservation and a lack of sporulation (Carrera-Játiva *et al.* 2014). Measures of the size of oocysts found (approximately 25-32 µm) appear to line up with the species description provided for *Eimeria pygosceli* by Golemansky (2003), but a match to this or any other species cannot be confirmed in our samples due to this same issue, lack of sporulation.

Microscopic evaluation and counts proved difficult and time-consuming. Moreover, they used limited faecal sample volume for a single purpose, only to find that parasites could not typically be visually identified. Sample preservation methods and insufficient sporulation of oocysts prevented full morphological characterization of coccidians according to standard methods, while helminths were fragmented or mangled and often lacked identifying features.

[See also **Appendix A7, Supplementary Photographs 1-7**]. We found that while visual identification is standard practice in parasitology, there are many limitations, which led us to adapt our subsequent sampling methods (for example, storage in potassium dichromate) and to adopt molecular characterisation where possible.

18S Sequencing and Taxonomic Classification

After filtering and pre-processing, a total of 5,021,473 reads were recovered, corresponding to 2,247 amplicon sequence variants (ASVs) present as at least 0.15% of at least one sample. After removing non-Eukaryotic sequences, 4,710,513 reads (1965 ASVs) remained. Of these, 307,992 reads in 205 ASVs corresponded to phyla or classes of endoparasitic interest, representing 6% of original reads and 6.5% of eukaryotic reads in the data set. The distribution of these sequences and reads is detailed in **Table 3**. The proportion of parasite reads was similar to the proportion of dietary reads (8%, suspected prey items) retrieved. Only three individual Gentoo Penguin samples stored in RNAlater lacked parasite reads. The proportion of reads assigned to parasitic taxa within an individual varied considerably, ranging from 0.014% to 92.04% (mean: 6.69%; median: 4.037%).

It should be noted that as a whole, of the 205 parasitic sequences analysed, many sequences could not be assigned below the Class (31 ASVs), Order (20 ASVs), or Family (137 ASVs) levels. The relative abundance of parasitic taxa within a species at a given site in a given year is depicted in **Figure 3**. Our focussed choice of endoparasitic taxa was based on published literature and previous identification of these in seabirds. It is possible, however, that there are other undescribed or poorly classified protozoans that act as endoparasites in penguins and Blue-Eyed Shags, perhaps ingested with prey or seawater. Results pertaining to each taxon are described below.

Table 3. Endoparasitic phyla and classes considered for analysis, with number of unique amplicon sequence variants and total reads for each class. Host Species indicates whether that class was detected (≥ 1 read) in Adélie Penguins (AP), Blue-Eyed Shags (BESH), Chinstrap Penguin (CP), and/or Gentoo Penguins (GP). Regions Found indicates where sequences were detected (≥ 1 read), with SG representing South Georgia, and "All" encompassing Falkland Islands/Malvinas, South Georgia, South Shetland Islands, Northeast Antarctic Peninsula, and Central Western Antarctic Peninsula.

Phylum: Class	**Unique ASVs**	**Total Reads**	**Host Species**	**Regions Found**
Acanthocephala: Palaeacanthocephala	9	13,679	AP, BESH, CP, GP	All
Apicomplexa: Conoidasida	134	183,667	AP, BESH, CP, GP	All
Nematoda: Enoplea	5	7,407	BESH, GP	All
Nematoda: Chromadorea	10	4,643	BESH, GP	All
Platyhelminthes: Cestoda	39	95,648	AP, BESH, CP, GP	All
Platyhelminthes: Trematoda	8	2,948	BESH, CP, GP	All
Total:	205	307,992		

Effect of Storage Medium

Two Gentoo Penguin colonies in close proximity (Heroina Island and Earle Island, Northeast Antarctic Peninsula) presented the opportunity to compare the effect of two guano storage media, potassium dichromate solution and RNAlater, on parasite sequence identification. Potassium dichromate, a medium commonly used in parasitology for preserving coccidian

oocysts or helminth eggs for subsequent microscopy or ex vivo maturation, yielded significantly higher alpha diversity measures by Welch's two-sample t-test (Shannon index: t = 7.5227, df = 36.107, p = 6.725e-09; Observed richness: t = 9.3334, df = 35.811, p = 3.996e-11; Chao-1: t = 3.3073, df = 55.678, p = 0.001654). Community composition differed according to storage media (see **Supplementary Figure S3**, NMDS by Bray-Curtis distance), and storage medium still accounted for differences in parasite community composition when taking into account phylogenetic relatedness of the sequences retrieved (see **Supplementary Figure S3**, NMDS by Weighted-UniFrac).

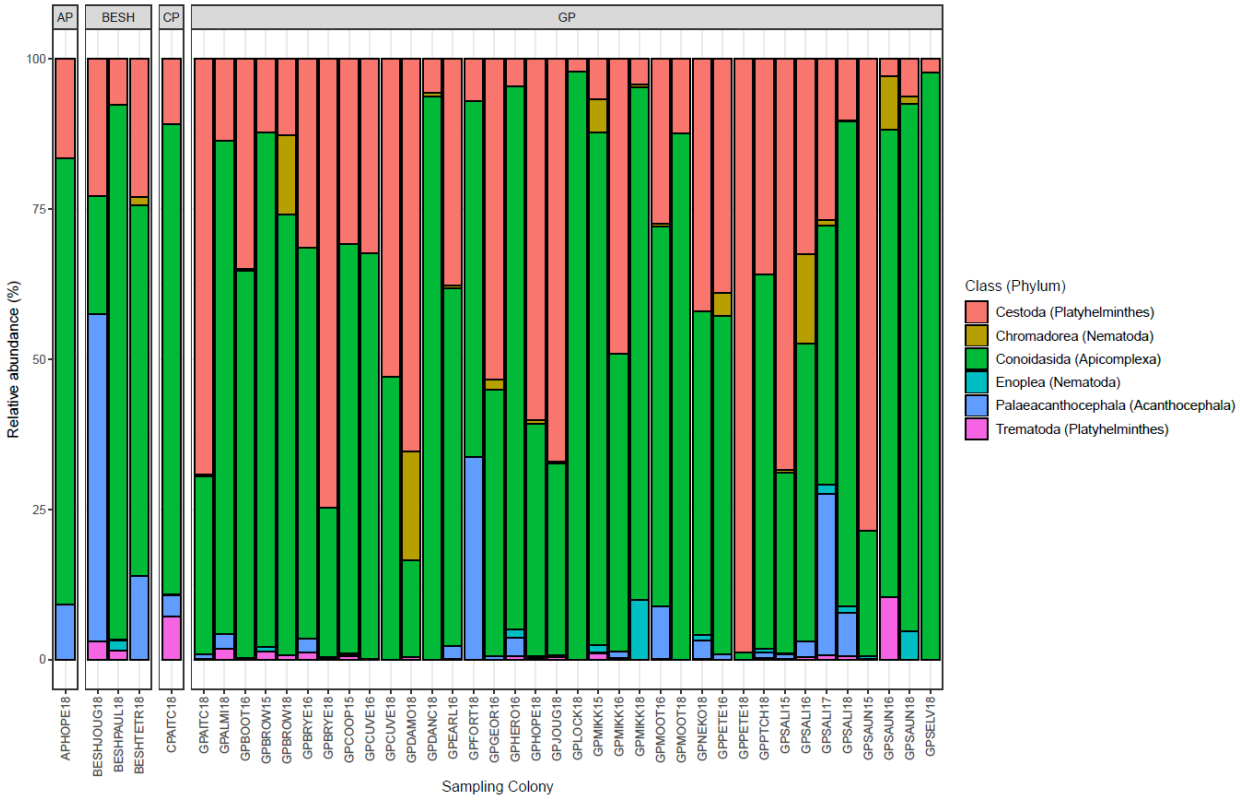


Figure 3. Relative abundance of endoparasitic classes for sequences retrieved from a given species at a particular colony in a given year. AP = Adélie Penguin, BESH = Blue-Eyed Shag, CP = Chinstrap Penguin, GP = Gentoo Penguin. Antarctic Site Inventory codes listed for each colony in Table 1, and two digit code represents sampling season.

Of 97 sequences retrieved in this comparative sample set, nematodes of the Order Rhabditida and trematode sequences (Subclass Digenea) were only found in potassium dichromate subsamples (see tree in **Supplementary Figure S4**). Conversely, RNAlater recovered sequences of the apicomplexan gregarine parasites (Family Gregarinidae) and nematodes in Family Mononchidae. Distinct sequences from coccidian parasites (Family Eimeriidae) were retrieved from both storage media. Removing extremely rare sequences (<20 total reads), 23 out of 44 remaining taxa in this sample set were detected (read threshold of >0) in both storage media. However, abundances were skewed toward one medium in every case (range 1.33x-57.5x; median 7.17x differences per sequence). Heterogeneity when dividing scats into sub-samples may account for some of these differences in unique retrievals.

In the full data set, there were a small number of taxa only found in one medium (gregarines in Family Gregarinidae and Lecudinidae, cestodes, nematodes, trematodes, and one acanthocephalan, see tree in **Supplementary Figure S5**). As most scats in this study were not paired for storage medium analysis, it is unclear how much diversity was not captured in our sequencing as a result of storage medium effects. In future studies, subsampling is recommended so that both molecular and morphological analyses can be performed on the same individual bird's endoparasites.

A single swab had been obtained from a penguin carcass found at Saunders Neck, Falkland Islands/Malvinas. This sample was analysed separately to assess whether parasites may have been linked to cause of death. Only 1% of reads from the carcass consisted of endoparasites, with no significant differences relative to guano samples from that site.

Specific Parasite Findings

Phylum Apicomplexa

The apicomplexan class Conoidasida contains two subclasses of single-celled obligate intracellular parasites, Coccidia and Gregarinasina (Adl *et al.* 2019). In our study, we recovered 134 apicomplexan sequences: 125 ASVs assigned to the coccidian family Eimeriidae, and 9 gregarine taxa.

Coccidians

Pygoscelid penguins and Blue-Eyed Shags host a greater number of sequences classified as coccidians than any other parasite group, present across the full range we sampled. We found that among the five Gentoo Penguin colonies from 2014/15 that had undergone both microscopy and sequencing (SAUN, SALI, COOP, BROW, MIKK), microscopic effort was unable to capture the 100% prevalence of coccidian sequences in all of these (n = 53) samples. Given the diversity of coccidian sequences recovered, many of which were found at low read counts and had high sequence similarity, we chose to focus our analysis on the top ten most abundant sequences within the dataset, as well as any sequences that were present in >50% of individuals in a given host species. The resulting 16 sequences are shown in a neighbour-joining tree (**Figure 4a**) using a *Cryptosporidium* outgroup to indicate their relationship to reference sequences of *Eimeria* from other vertebrate hosts, alongside their prevalence across the hosts and regions in our study, after excluding samples with under 10 reads (**Figure 4b**). Of these sequences, only one could be assigned at the genus or species level, seq143, found only in Blue-Eyed Shags (two individuals in each of the three sites), which was a 100% BLAST match to *Eimeria arnyi*. *E. arnyi* is known to be specific to snakes (Jirků *et al.* 2009) and is therefore not likely the true species recovered in our study.

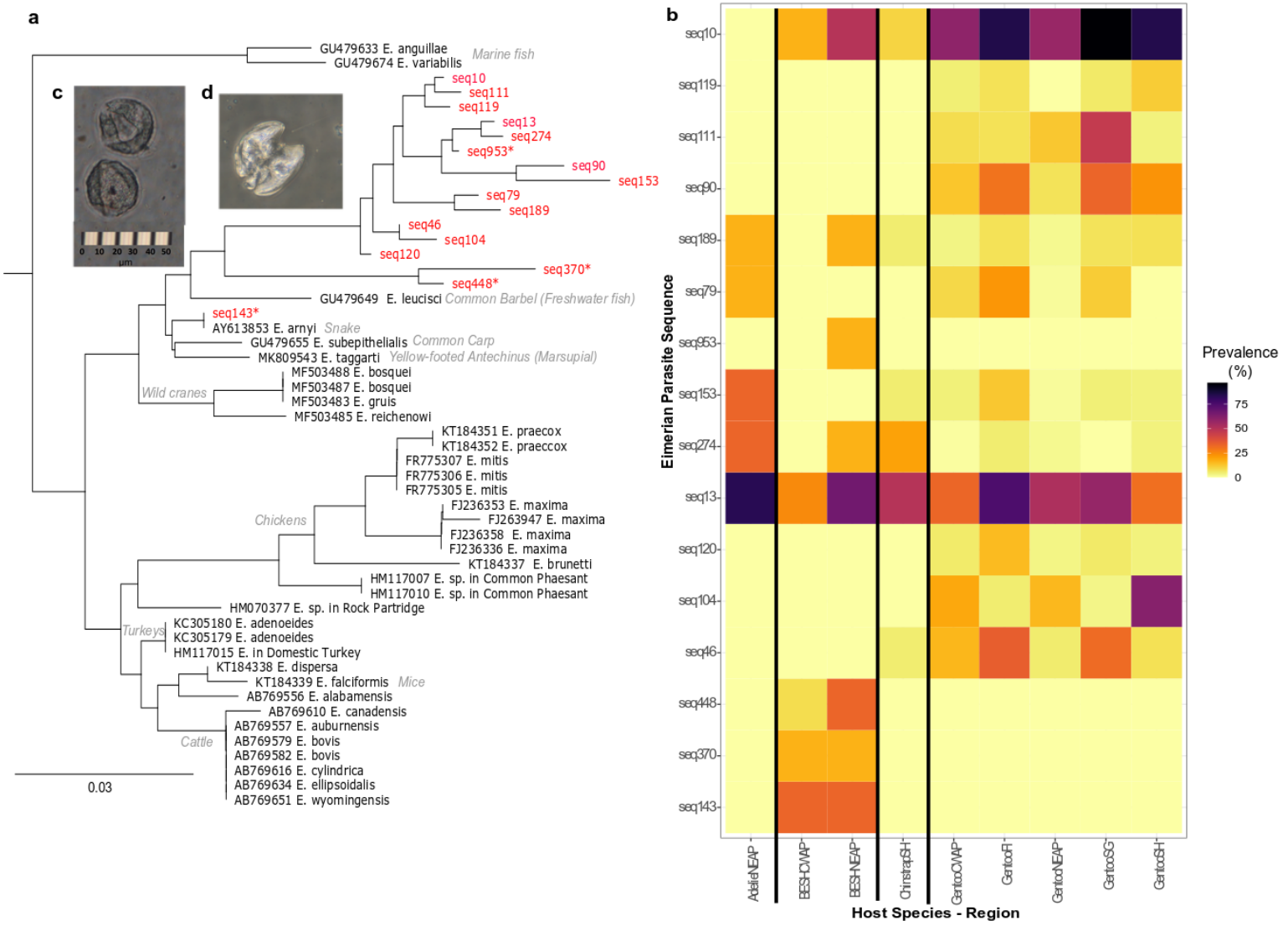

a
b
c
d
GU479633 E. anguillae
GU479674 E. variabilis
Marine fish
seq10
seq111
seq119
seq13
seq274
seq953*
seq90
seq153
seq79
seq189
seq46
seq104
seq120
seq370*
seq448*
GU479649 E. leucisci Common Barbel (Freshwater fish)
seq143*
AY613853 E. arnyi Snake
GU479655 E. subepithelialis Common Carp
MK809543 E. taggarti Yellow-footed Antechinus (Marsupial)
Wild cranes
MF503488 E. bosquei
MF503487 E. bosquei
MF503483 E. gruis
MF503485 E. reichenowi
KT184351 E. praecox
KT184352 E. praeccox
FR775307 E. mitis
FR775306 E. mitis
FR775305 E. mitis
Chickens
FJ236353 E. maxima
FJ263947 E. maxima
FJ236358 E. maxima
FJ236336 E. maxima
KT184337 E. brunetti
HM117007 E. sp. in Common Phaesant
HM117010 E. sp. in Common Phaesant
HM070377 E. sp. in Rock Partridge
Turkeys
KC305180 E. adenoeides
KC305179 E. adenoeides
HM117015 E. in Domestic Turkey
KT184338 E. dispersa
KT184339 E. falciformis Mice
AB769556 E. alabamensis
AB769610 E. canadensis
Cattle
AB769557 E. auburnensis
AB769579 E. bovis
AB769582 E. bovis
AB769616 E. cylindrica
AB769634 E. ellipsoidalis
AB769651 E. wyomingensis
0.03
0 10 20 30 40 50
µm
Eimerian Parasite Sequence
seq10
seq119
seq111
seq90
seq189
seq79
seq953
seq153
seq274
seq13
seq120
seq104
seq46
seq448
seq370
seq143
AdelieNEAP
BESHOWP
BESHNEAP
ChinstrapSH
GentooOWP
GentooR
GentooNEAP
GentooSG
GentooSH
Host Species - Region
Prevalence (%)
75
50
25
0

Figure 4. a. Neighbour-joining tree (HKY model) of 16 most abundant and most prevalent (species-wise) Eimerian parasite sequences detected in this study. Red text indicates novel 18S sequences, while reference sequences from GenBank are indicated in black text with their accession numbers. Grey italic text indicates host species for reference parasite sequences. Asterisk (*) represents novel sequences from this study found exclusively in Blue-Eyed Shags (minimum threshold of 10 reads). b. Heatmap indicating overall prevalence of each novel Eimerian sequence in the tree, grouped by host species (Adélie Penguin, Blue-Eyed Shag (BESH), Chinstrap Penguin, and Gentoo Penguin) and geographic region (CWAP = Central Western Antarctic Peninsula, FI = Falkland Islands, NEAP = Northeast Antarctic Peninsula, SG = South Georgia, SH = South Shetland Islands). c. Light microscopy image (40x) of suspected Eimerian parasites in Gentoo Penguin guano from the Falkland Islands, with scale bar. d. Light microscopy image (10x, scale bar not available) during salt flotation of suspected Eimerian parasite from Blue-Eyed Shag from the Tetrad Islands (CWAP).

Our sequences fall clearly within the genus *Eimeria* (**Figure 4a**), the most speciose group of apicomplexans, with over 1,000 described species (Ogedengbe *et al.* 2011a). *Eimeria* spp. are known for remarkable host specificity (Levine 1988). In vertebrates, aside from few exceptions, not only are Eimerians naturally limited to a narrow range of hosts, they are normally specific to particular organ systems, portions of the system, specific types of cells, and in some cases, even specific locations within cells (Marquardt 1973). Our analysis of coccidian sequences shows that the ~170 bp marker targeted is sufficient for distinguishing between *Eimeria* species within other avian hosts, such as chickens and turkeys, which are known to harbour multiple species within different portions of their gut (Blake *et al.* 2011). We found that several of our sequences presented the host-specificity we would expect of *Eimeria*. Seq370, seq448, and seq143 were found only in Blue-Eyed Shags, with two of these forming a clade separate from the other main sequences recovered. Seq953 was found in only a single Blue-Eyed Shag, but was included in our analysis due to its high individual-level relative abundance (2.7% of all reads, 19.5% of parasite reads in that sample from Paulet Island). Shag-specific sequences are denoted with an asterisk in **Figure 4a**. We also found a number of sequences (seq119, seq111, seq90, seq120, seq104) that appear to be largely restricted to Gentoo Penguins, though increased sampling of other sympatric species may be necessary to confirm this.

One surprising finding was the high prevalence and abundance of seq10 and seq13 across host species, which is not typical of *Eimeria* species. One potential explanation for this ubiquity is that these are coccidians ingested in food, but the phylogeny does not support that given how closely they align to the other sequences with clear avian hosts. A second possibility is that these sequences represent multiple Eimerian species that are in fact host-specific, but present no differentiation in this marker. Given the genetic variation seen within well-studied single species (e.g. *E. maxima*) at this marker (Ogedengbe *et al.* 2011a), this does

not seem likely. Finally, a selected number of coccidian species in the Antarctic may have evolved broader host utilisation, at the genus level (across *Pygoscelis*), or perhaps extended to other avian orders, to facilitate survival and persistence of portions of their life cycles when faced with harsh conditions. This would be outside the norm of current knowledge of this parasite, but cannot be ruled out given the limited research on the biology of *Eimeria* in Antarctic wildlife.

Cross-immunity and cross-host transmission studies have been used to identify species and their specificity under experimental conditions, but these approaches are not easily practical in the field (Marquardt 1973). There has been an increase in molecular and genomic studies of avian coccidians, though these generally focus on species of commercial interest (Heitlinger *et al.* 2014; Ogedengbe *et al.* 2011a; Ogedengbe *et al.* 2011b; Ogedengbe *et al.* 2014; Reid *et al.* 2014), and wider sequencing of coccidians in wild birds is warranted.

Microscopy of a selection of Gentoo Penguin and Blue-Eyed Shag guano samples with high coccidian read abundances did not find sporulated oocysts in sufficient numbers or in good enough condition to characterise via light microscopy alone. Examples of these are shown in **Figures 4c and 4d**. Further study using electron microscopy and more precise photographic equipment could likely aid in identification, coupled with expanded sequencing of the Internal Transcribed Spacer (ITS) region of ribosomal DNA or other genomic markers, as suggested in Schwarz *et al.* (2009).

Gregarines

While coccidians are found in nearly all animal hosts (Levine 1973), most gregarine taxa infect invertebrate gastrointestinal tracts almost exclusively (Simdyanov *et al.* 2017), aside from *Cryptosporidium*, recently re-classified under the gregarine Order Cryptogregarinorida (Adl *et al.* 2019; Rueckert *et al.* 2019). Gregarine infestations can lead to

clinical symptoms and negative effects on growth, reduced offspring production, and mortality in their hosts (Ball *et al.* 1995). Gregarines also tend to be highly host-specific, and even life-stage-specific among invertebrates, and generally use only one host organism in their life cycle (Rueckert *et al.* 2019). Reads were recovered from 9 gregarine taxa, all assignable to the species *Gregarina blattarum*, *Lankesteria cystodytae*, *Ascogregarina taiwanensis*, and *Cryptosporidium struthionis*. Among eugregarines, *G. blattarum* infects cockroaches (Clopton & Gold 1996) and *A. taiwanensis* infects mosquitoes. It is therefore likely that what we have recovered are other eugregarines infecting marine invertebrates. *A. taiwanensis* attributed sequences were found in Gentoo Penguins in the Falkland Islands, South Georgia, and the Central Western Antarctic Peninsula (MIKK), while *G. blattarum* matches were found in Adélie, Chinstrap, and Gentoo Penguins in all regions sampled aside from the Falkland Islands. *L. cystodytae*, whose sequence was retrieved from a single Gentoo Penguin in the Aitcho Islands, South Shetland Islands, is a marine eugregarine of tunicates (sea squirts, Order Ascidiacea) (Rueckert *et al.* 2015). While that particular individual did not have tunicate sequences detected, other Gentoo Penguins and Chinstrap Penguins at that site yielded ascidian reads, so this may be a feasible association.

Unidentified *Cryptosporidium* oocysts, on the other hand, have been reported in Adélie Penguin and Gentoo Penguin guano from King George Island, also in the South Shetlands (Fredes *et al.* 2008a; Fredes *et al.* 2007b). Two ASVs assigned to the species *Cryptosporidium struthionis* (an avian parasite suggested by Wood *et al.* (2013)) were found in 76 individuals (3 Blue-Eyed Shags, 4 Chinstrap Penguins, and 69 Gentoo Penguins, see **Supplementary Figure S6**), in every geographic region sampled, with highest abundances in several Falkland Island samples. These two sequences formed a separate clade from other vertebrate *Cryptosporidium* sequences, as seen in the tree in **Supplementary Figure S6**.

Phylum Acanthocephala

Acanthocephalans are parasitic, thorny-headed worms with complex life cycles. We report the presence of 9 ASVs in the Class Palaeacanthocephala in our samples, of which 8 could be classified to the Order Polymorphida, Family Polymorphidae. Phylogenetic tree building with reference sequences confirm this (see **Figure 5**), and place the outlier sequence (seq625), found in a single Blue-Eyed Shag, within the family Echinorhynchida, closely related to worms from the genus *Echinorhynchus.* Acanthocephalan sequences were found in 31 samples (at a minimum threshold of 10 reads) across all host species sampled, with one Blue-Eyed Shag (JOUG) and four Gentoo Penguins from South Georgia presenting the greatest abundances of one particular sequence, seq41 closely related to reference sequences in the genus *Corynosoma* (**Figure 5**). Sequences originated from all sample regions, though in Gentoo Penguins, Falkland Island samples presented fewer than 10 reads (not shown).

Among acanthocephalan parasites, the genus *Corynosoma* (Family Polymorphidae) has been reported in Pygoscelid penguins in the South Shetland Islands, Macquarie Island (South Pacific Ocean) and Antarctica (Diaz *et al.* 2013; Dimitrova *et al.* 1996; Edmonds 1954; González-Acuña *et al.* 2013; Hoberg 1986; Holloway & Bier 1967; Vidal *et al.* 2012; Zdzitowiecki 1991). Polymorphid worms are thought to mainly parasitise fish and aquatic birds, but to date, only reports of immature stages of *Corynosoma* have been described in penguins (Diaz *et al.* 2017). It has been postulated that some species of *Corynosoma* may be parasites of marine mammals found only incidentally in seabirds (Hoberg 1986; Vidal *et al.* 2012). While we were unable to directly classify to the genus level using reference databases for these parasites, it is plausible that some of our findings correlate with visual detections of *Corynosoma* in Antarctic seabirds.

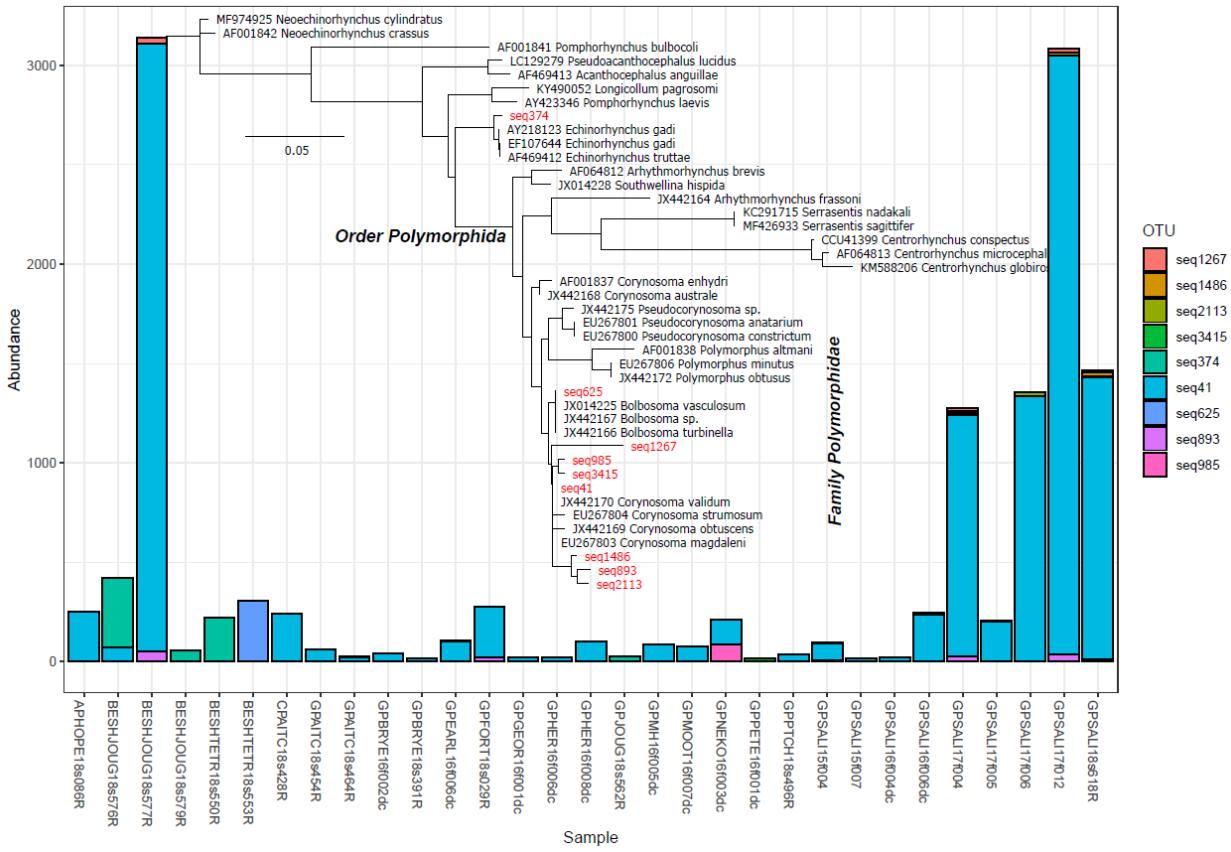


Figure 5. Sample-wise read abundance of sequences (with a minimum of 10 reads per sample) in the class Palaeoacanthocephala retrieved from the full data set, along with an overlaid phylogenetic tree (Neighbour-joining, HKY distance) of the sequences aligned with a selection of reference 18S sequences obtained from GenBank. Sample key: AP = Adélie Penguin, BESH = Blue-Eyed Shag, CP = Chinstrap Penguin, GP = Gentoo Penguin. Four-letter site names follow, as shown in Table 1, as well as the year (two digits), sample number, and storage designation (R = RNAlater, dc = Potassium dichromate).

Phylum Nematoda

Nematodes can be free-living or parasitic, and have been reported multiple times in Pygoscelid penguins across their ranges (Diaz *et al.* 2016; Diaz *et al.* 2013; Fredes *et al.* 2007a; Fredes *et al.* 2006; Fredes *et al.* 2008b; González-Acuña *et al.* 2013; Mawson 1953; Vidal *et al.* 2012). Among them, *Stegophorus macronectes* is one of the best-characterised, having been

described morphologically and sequenced (18S, 5.8S, 28S and ITS regions) by Vidal *et al.* (2016). In our study, we retrieved 15 distinct nematode ASVs in Classes Chromadorea (Orders Rhabditida, 9 ASVs and Strongylida, 1 ASV) and Enoplea (Order Mononchida, 5 ASVs), found in 57 samples (2 Blue-Eyed Shags, 55 Gentoo Penguins with at least 10 reads per sample) from all sampled regions (**Figure 6**). A few Gentoo Penguin individuals from the Falkland Islands/Malvinas, South Georgia, and Central Western Antarctic Peninsula (MIKK) had high abundances of nematode sequences. In general, hosts had either mononchid or rhabditid parasites as dominant, and rarely presented significant abundances of both. Two unique chromadorean sequences (one in Order Strongylida and a rhabditid in the Family Cephalobidae) not present in other samples were found together in one Gentoo Penguin from South Georgia (SALI).

Unfortunately, and indicative of the lack of molecular characterisation of many parasite species, only one sequence could be characterised to the genus level, *Caenorhabditis*, (often free-living but also sometimes associated with hosts in a phoretic or necromenic lifestyle albeit mostly with invertebrate hosts). This sequence was found in 2 Gentoo Penguins in South Georgia (SALI). No sequences were exact matches to *Stegophorus macronectes*, though seq1171 in one Gentoo Penguin (BROW) and one Blue-Eyed Shag (TETR) grouped phylogenetically with *Echinuria borealis*, an acuarid nematode known to parasitise waterfowl (McDonald 1969).

Microscopy of a Blue-Eyed Shag sample from the Tetrad Islands, where rhabditid nematode sequence reads were detected, yielded observations of potential nematodes, though these could not be measured or identified (see **Figure 6**). Helminth eggs were also found in several Gentoo Penguin sequences (not shown) that could not be identified, but may include nematodes.

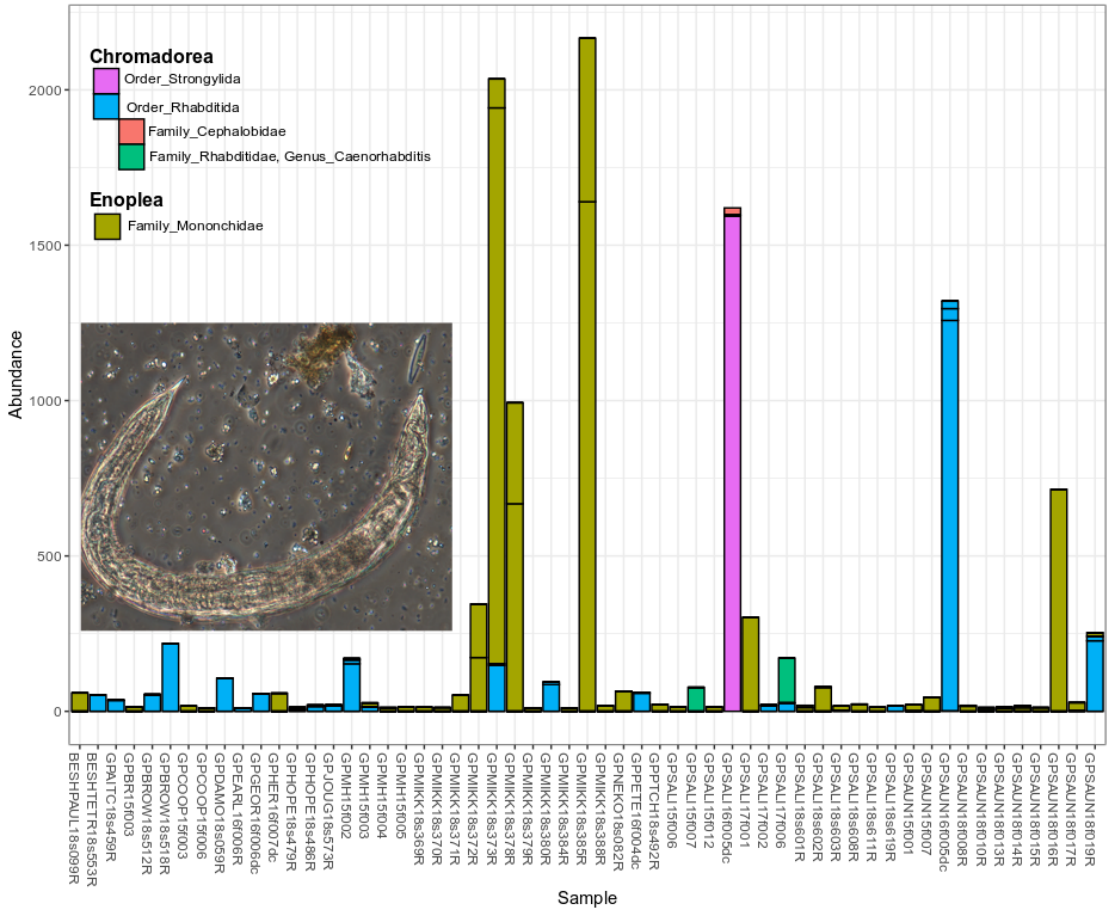


Figure 6. Sample-wise read abundance of sequences (with a minimum of 10 reads per sample) in the phylum Nematoda retrieved from the full data set, with colours by lowest taxonomic level assigned. Sample key: BESH = Blue-Eyed Shag, GP = Gentoo Penguin. Four-letter site names follow, as shown in Table 1, as well as the year (two digits), sample number, and storage designation (R = RNAlater, dc = Potassium dichromate). Inset: An example of a helminth found in faeces from a Blue-Eyed Shag from the Tetrad Islands, Central Western Antarctic Peninsula, suspected, but not confirmed, to be a nematode in the Order Rhabditida. Shown here at 20x magnification.

Phylum Platyhelminthes

Flatworms can be either free-living or parasitic. In this study, we retrieved a small number of sequences of a marine free-living triclad (Class Rhabditophora) in a single Gentoo

Penguin, which we classified as food items (whether ingested intentionally or unintentionally). Parasitic flatworms include monogeneans that are fish ectoparasites (no sequences of which were found in this study), and cestodes (tapeworms) and trematodes (flukes).

Cestodes

Two primary genera of cestodes, *Parorchites* (Order Cyclophyllidea, Family Dilepididae) and *Tetrabothrius* (Order Tetrabothriidea, Family Tetrabothriidae) have been repeatedly reported in Pygoscelid hosts, in the South Shetland Islands and Antarctic Peninsula (Cielecka *et al.* 1992; Diaz *et al.* 2013; Fredes *et al.* 2007a; Fredes *et al.* 2006; Fredes *et al.* 2008b; Georgiev *et al.* 1996; González-Acuña *et al.* 2013; Martin *et al.* 2016), East Antarctica (Leiper & Atkinson 1914), Bouvetøya Island in the South Atlantic (Andersen & Lysfjord 1982) and Iles Kerguelen in the Indian Ocean (Prudhoe 1969). The mode of attachment and pathology caused by these helminths of the gastrointestinal tract are well described in Pygoscelids (Martin *et al.* 2016), though no sequence data was available in public databases. Eggs of an unknown species of *Diphyllobothrium* and an unknown Cyclophyllid have reported on Ardley Island, South Shetland Islands (Fredes *et al.* 2006; Fredes *et al.* 2008b; González-Acuña *et al.* 2013).

Reads assigned to the Class Cestoda (39 ASVs) were highly prevalent across individuals of all host species and locations (n = 447 samples). One ASV (seq12), which could not be attributed to a specific genus, was found in 345 samples, including 330 Gentoo Penguins. This sequence was also occasionally found in high abundances, (mean abundance of 160 reads; 15 individuals with > 500 reads; max of 9,161 reads in a Falkland Islands Gentoo Penguin), and alone represents 17.9% of all parasite reads retrieved in this study. As was the case with other understudied parasite groups, however, only three ASVs matched BLAST hits below the level of order, classified as *Mesocestoides corti* (a mouse and domestic mammal parasite in the Order Cyclophyllidea occasionally found in passerine birds (Heneberg *et al.* 2019)) and

Diplogonoporus grandis (a cetacean tapeworm in the Order Diphyllobothriidea (Yamasaki *et al.* 2012)), so these attributions are likely incorrect. Microscopy revealed helminth eggs, but these could not be identified at the phylum level. The high prevalence of cestodes found in the seabirds in our study indicates the potential value in future characterisation of these gastrointestinal helminths from carcasses and faecal samples.

Trematodes

Trematodes, on the other hand, have not been reported in wild Pygoscelids, though they are known to affect penguin organs in captive settings (Wright 1954). Eight distinct trematode ASVs were recovered, (6 in Order Azygiidae with one classified to Family Hemiuridae, and one in Order Plagiorchiidae/Family Opecoelidae). Trematode sequences were recovered in appreciable numbers (>10 reads) from 28 samples (4 Blue-Eyed Shags from the Northeast and Central Western Antarctic Peninsula, 1 Chinstrap Penguin from the South Shetland Islands, and 23 Gentoo Penguins from all regions aside from the South Shetlands). The single hemiurid-assigned sequence (seq 162) was the most abundant and widespread of trematode sequences, with high relative abundances in the Chinstrap Penguin and several Falkland Island Gentoo Penguins. One sequence (seq1666) was only found at a single site, Jougla Point, Central Western Antarctic Peninsula, where it was found in both Blue-Eyed Shags and a Gentoo Penguin. Trematodes that affect vertebrates are mostly considered to have multi-host life cycles, often with an intermediate invertebrate host (Olson *et al.* 2003). Our findings suggest that trematodes in wild penguins may be underrepresented in case reports. More research is also needed to evaluate whether Antarctic seabirds serve as definitive hosts, given that our dietary data also lends support to existing knowledge that penguins consume a diversity of marine invertebrates.

Parasites across Space and Species

Aside from characterising the diversity of parasites (or parasite DNA) associated with Pygoscelid Penguins and Blue-Eyed Shags, we also considered two primary questions in our study: (1) whether parasites are shared across sympatric species, or whether they are host-specific; and (2) whether a given host species has the same parasites in different portions of its geographic range. We found that, 62 of 205 ASVs of parasite sequences, representing 87.9% of all parasite sequences, were present in more than one host (at a threshold of at least 1 read, tree shown in **Figure 7**). This is surprising given the small sample sizes of Adélie Penguins, Chinstrap Penguins, and Blue-Eyed Shags investigated in this study.

Based on our knowledge of the biology of various parasite classes, we expected to see high host specificity in certain groups, such as coccidians, which develop in a single host (monoxenous life cycle) and are directly transmitted between individuals, commonly via the faecal-oral route (Hammond 1973). Indeed, as discussed above, we observed certain coccidian sequences that appear to be specific to Blue-Eyed Shags or to Gentoo Penguins. Other coccidian sequences with greater cross-host prevalence may represent species capable of carrying out their life cycles in multiple hosts within the genus *Pygoscelis*, or within the higher waterbird clade Aequorlitornithes (sensu Prum *et al.* (2015)), which includes penguins and cormorants, in order to overcome environmental limitations.

Direct transmission can also be a feature of nematodes, some of which can infect hosts through egg ingestion, while others have more complex life cycles with multiple invertebrate and/or vertebrate intermediate hosts. Nematodes were only found in Gentoo Penguins and a small number (n = 2) of Blue-Eyed Shags, so host-specificity cannot be determined. We expect most other parasite classes discussed here (cestodes, trematodes, acanthocephalans) to be trophically transmitted, and so their presence across hosts is likely a byproduct of overlapping dietary content in the Antarctic food web.

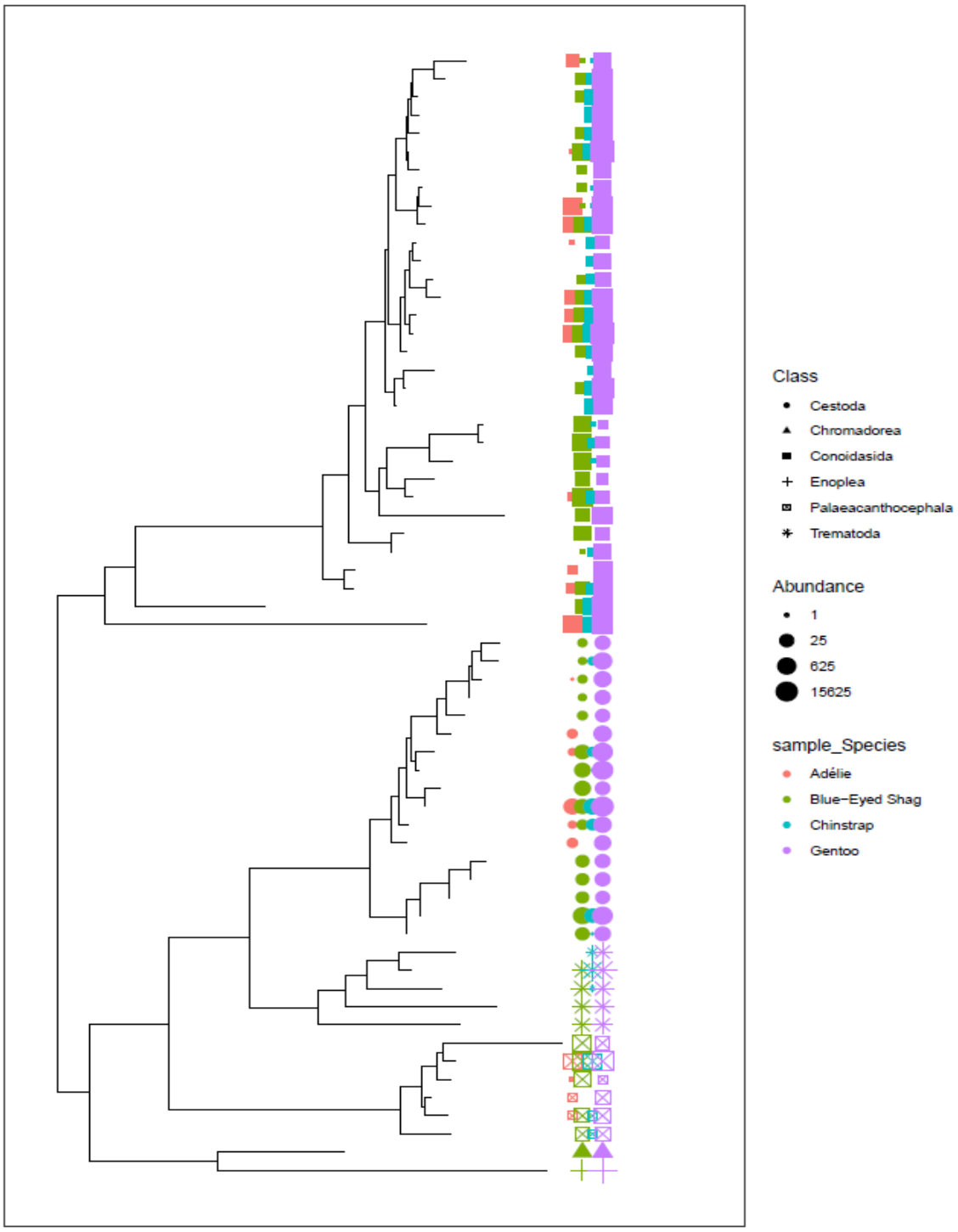


Figure 7. Phylogenetic tree (Neighbour-joining, HKY distance) of endoparasite sequences retrieved from more than one host species, with shape colour representing each host, shape size indicative of abundance (reads), and shape delineating the class-level taxonomic assignment of the shared sequence.

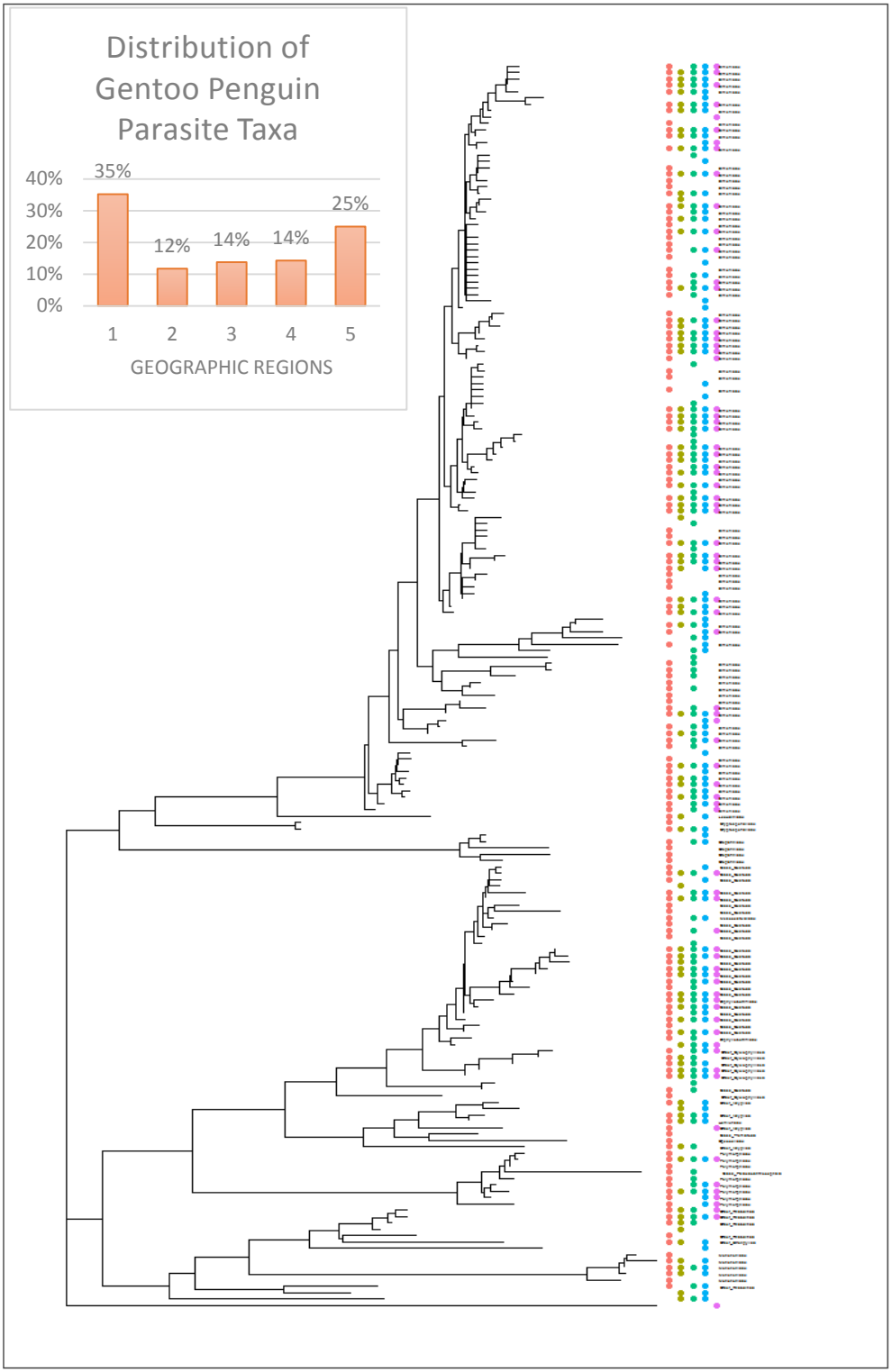


Figure 8. Distribution of Gentoo Penguin parasite taxa by prevalence across regions (inset bar graph) and phylogenetic tree (Neighbour-joining, HKY distance) of endoparasite sequences retrieved from Gentoo Penguins, with dot colour representing distinct geographic regions (as

defined by the Antarctic Site Inventory) to view sequences across regions. CWAP = Central Western Antarctic Peninsula, FI = Falkland Islands/Malvinas, NEAP = Northeast Antarctic Peninsula, SG = South Georgia, and SH = South Shetland Islands. Tip labels depict the Family level, unless the sequence could not be classified at that resolution, in which case the Phylum, Class, or Order is shown.

Among Gentoo Penguins located in different geographic regions, we expected to find high levels of segregation among the parasite communities. While quantitative statistical comparisons are not possible due differences in storage medium and read depth, we were able to examine the distribution of taxa across regions. Our study found both region-unique and geographically distributed parasite taxa. Out of 196 ASVs in Gentoo Penguins, 69 (35.2%) were unique to a single region, while the other 64.8% were found in at least 2 regions, with 25% (49/196) of sequences shared across all 5 regions (see **Figure 8**). If an OTU-based approach had been used, grouping sequences with close similarities into single taxa, it is likely that even more commonality would have been found, particularly in the numerous ASVs assigned to the family Eimeriidae. Although some connectivity would be expected within the host metapopulations of the Antarctic Peninsula and South Shetland Islands, the broad distribution of a quarter of parasite taxa in Gentoo Penguins across large ocean distances (to include South Georgia and the Falkland Islands), may indicate that parasite connectivity is maintained through lower trophic levels or sympatric migratory birds, given the absence of Gentoo Penguin gene flow and migration across the Polar Front.

Of the 69 region-unique taxa, 45 were assigned to the family Eimeriidae. Phylogenetic placement reveals the existence of multiple clades containing closely-related sequences derived from different regions, which may indicate genetic differentiation of these parasites

geographically (**Supplementary Figure S7**). Given the extrinsic barrier posed by the Polar Front in Gentoo Penguin hosts, we expected unique parasites to be present either side of the Front. Instead, only four sequences were unique to the Falkland Islands, two Eimerians, one cestode, and one rhabditid nematode. The southernmost region, the Central Western Antarctic Peninsula, was the most sampled, but also revealed a majority of the region-unique sequences in every parasite group. This lends support to the modern understanding that Antarctic biodiversity is underestimated (Chown *et al.* 2015), and adds to evidence that pathogen pressures may not necessarily decrease at higher latitudes (Wille *et al.* 2020).

Conclusion

In this study, we identified widespread prevalence of Pygoscelid penguin and Blue-Eyed Shag coccidian parasites and helminths through microscopic and molecular means. Though we were not able to classify many of these as described species, their presence indicates a diversity of endoparasites in the Antarctic region whose life cycle, stages, morphology, and phylogenies warrant investigation. Understanding parasite prevalence patterns could unlock knowledge regarding avian host evolution and population differentiation of both hosts and pathogens across the environmental gradients of these species and over time. We noted examples of host-specific behaviour in coccidians, as well as surprising levels of host-sharing of parasites and broad geographic distributions in coccidians, acanthocephalans, cestodes, nematodes, and trematodes that may be associated with trophic transmission or be maintained by migratory birds.

The characterisation of eukaryotic parasites of seabird gastrointestinal tracts has mostly been restricted to visual identification of oocysts and eggs in faeces or helminths attached to gastrointestinal organs upon post-mortem dissection. Increasingly used for dietary analyses, we have shown that 18S metabarcoding of seabird scats is a tool that can aid future investigations into endoparasitic taxa. Molecular screening of faeces provides

advantages in being non-invasive, and grants an opportunity to overcome the challenges to visual identification presented by damaged or partially digested parasites in faeces. Instead, high-throughput sequencing can highlight faecal samples with relatively high abundance of parasite genetic material to then focus microscopy or histology when faced with large numbers of samples. Though this approach is limited in that it uses only a short fragment of the small subunit of ribosomal DNA, we have found that it aids in general taxonomic placement among well-studied parasite groups, and would be useful to understand the baseline of parasite presence at a site, or in a given species. Future efforts should include longer target sequences and other targets such as the 5S, 28S, and ITS regions to refine taxonomic assignment. All of these studies should ideally be performed in conjunction with microscopy and other direct observations, with isolation for parasite types and shotgun sequencing to provide the genetic resource with which to screen future survey sample sets.

This study demonstrates the potential for faeces-targeted parasite DNA approaches to add to our understanding of parasites associated with seabirds, in particular the Pygoscelid penguins and Blue-Eyed Shag. It is important to note that the molecular signatures associated with faecal samples will include both parasites of the donor and also parasites of prey species further targeted by these marine mesopredators. The effect of parasitism on individual and population-level fitness is unknown, though environmental stressors such as climate change, pollution, competition with fisheries for prey, and human disturbance could be additive. For this reason, host-parasite relationships in seabirds are not only interesting from an evolutionary standpoint, but may be critical to understanding the potential impact of disease in conservation of these species. Molecular tools can assist in additive discoveries and further our understanding of microfaunal diversity in remote ecosystems and in wild hosts. Studies of wildlife parasitology, Antarctic and marine biodiversity are reliant upon basic exploratory

observations, and there is still an overwhelming number of taxa lacking description or taxonomic assignment.

Acknowledgements

We thank Quark Expeditions, Cheeseman's Ecology Safaris, Oceanwide Expeditions, and J. Poncet for their assistance with fieldwork logistics. We thank Anni Djurhuus, Caitlin E. Black, Gemma V. Clucas, Michael J. Polito, Heather J. Lynch, Michael B. Schrimpf, Rachael Herman, Maureen Lynch, and Casey Youngflesh for assisting with sample collection. We thank Kirstie McLoughlin, Patrik Flammer, Rob Dixon, Sophie Hedges, Frances Colles, for sharing MiSeq runs and assisting with the sequencing workflow, and Amanda Williams and Rebecca Busby for overseeing sequencing runs. Molecular work was supported by Penguin Lifelines, the U.S. Department of Veteran's Affairs G.I. Bill, and Stony Brook University. ALS received funding from the BBSRC and Bertarelli Foundation. HL's work was supported by a University of Oxford Clarendon Fund Scholarship.

Author Contributions

HL, TH, and ALS developed and designed the study. TH obtained permits and secured logistics. HL and TH collected samples in addition to collaborators listed in acknowledgments. HL and PGF extracted all DNA. HL performed all PCR amplifications, quality control, and pooling. HL and SGP performed Illumina MiSeq library preparation, quality control, and sequencing with technical support from collaborators listed in acknowledgments. SGP processed all raw sequence data and created custom scripts for demultiplexing, pre-processing reads, and performing automated taxonomic assignment. PGF and ALS assisted HL with microscopy. HL conducted all analyses and wrote the paper with feedback from all authors.

Ethics Statement

Animal handling protocols followed Scientific Committee on Antarctic Research Code of Conduct on the use of Animals for Scientific Purposes in Antarctica. All sampling in the Antarctic Treaty Area was carried out in accordance with United Kingdom Foreign and Commonwealth Office (permits: S3-23/2013, S7 28/2013, S3 04/2014, S7 03/2014, 33/2016, 34/2016, 37/2017), in South Georgia under permits granted by the Government of South Georgia and the South Sandwich Islands (SCI-2014-017, RAP-2015-018, RAP-2016-035 and RAP-2017-034), and in the Falkland Islands under Falkland Islands Government Environmental Planning Department permits (R16.2014, R15 2014, and R04 2017), with ethical approval from the University of Oxford Animal Welfare and Ethical Review Board.

Conflict of Interest Statement

The authors declare no conflict of interest.

Data Availability Statement

Custom scripts, genetic sequences, sample metadata, abundance tables, and taxonomy assignment tables will be made available prior to publication on a public repository.

References

Adl SM, Bass D, Lane CE*, et al.* (2019) Revisions to the Classification, Nomenclature, and Diversity of Eukaryotes. *Journal of Eukaryotic Microbiology* **66**, 4-119.

Andersen KI, Lysfjord S (1982) The Functional Morphology of the Scolex of 2 *Tetrabothrius Rudolphi* Species (Cestoda, Tetrabothriidae) from Penguins. *Zeitschrift Fur Parasitenkunde-Parasitology Research* **67**, 299-307.

Anderson RM, May RM (1982) Coevolution of hosts and parasites. *Parasitology* **85**, 411-426.

Ball SJ, Cunningham AA, Clarke D, Daszak P (1995) Septate Gregarines Associated with a Disease of the Hissing Cockroach *Gromphadorhina portentosa*. *Journal of Invertebrate Pathology* **65**, 311-312.

Barbosa A (2011) Climate change effects on Antarctic penguins. *Ecosistemas* **20**, 33-41.

Blake DP, Oakes R, Smith AL (2011) A genetic linkage map for the apicomplexan protozoan parasite *Eimeria maxima* and comparison with *Eimeria tenella*. *Int J Parasitol* **41**, 263-270.

Callahan BJ, McMurdie PJ, Rosen MJ*, et al.* (2016) DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods* **13**, 581-583.

Carpenter JW, Novilla MN, Fayer R, Iverson GC (1984) Disseminated visceral coccidiosis in sandhill cranes. *J Am Vet Med Assoc* **185**, 1342-1346.

Carrera-Játiva PD, Rodríguez-Hidalgo R, Sevilla C, Jiménez-Uzcátegui G (2014) Gastrointestinal parasites in the Galápagos penguin *Spheniscus mendiculus* and the Flightless cormorant *Phalacrocorax harrisi* in the Galápagos Islands. *Marine Ornithology* **42**, 77-80.

Chown SL, Clarke A, Fraser CI*, et al.* (2015) The changing form of Antarctic biodiversity. *Nature* **522**, 431-438.

Cielecka D, Wojciechowska A, Zdzitowiecki K (1992) Cestodes from penguins on King George Island (South Shetlands, Antarctica). *Acta Parasitologica Polonica* **37**, 65-72.

Cimino MA, Lynch HJ, Saba VS, Oliver MJ (2016) Projected asymmetric response of Adélie penguins to Antarctic climate change. *Sci Rep* **6**, 28785.

Clements JF, Schulenberg TS, Iliff MJ*, et al.* (2019) The eBird/Clements Checklist of Birds of the World: v2019.

Clopton RE, Gold RE (1996) Host Specificity of *Gregarina blattarum* von Siebold, 1839 (Apicomplexa: Eugregarinida) among Five Species of Domiciliary Cockroaches. *Journal of Invertebrate Pathology* **67**, 219-223.

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

Clucas GV, Younger JL, Kao D*, et al.* (2018) Comparative population genomics reveals key barriers to dispersal in Southern Ocean penguins. *Mol Ecol* **27**, 4680-4697.

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.

Diaz JI, Fusaro B, Longarzo L*, et al.* (2016) Gastrointestinal helminths of Adélie penguins (*Pygoscelis adeliae*) from Antarctica. *Polar Research* **35**.

Diaz JI, Fusaro B, Longarzo L*, et al.* (2013) Gastrointestinal helminths of Gentoo penguins (*Pygoscelis papua*) from Stranger Point, 25 de Mayo/King George Island, Antarctica. *Parasitol Res* **112**, 1877-1881.

Diaz JI, Fusaro B, Vidal V*, et al.* (2017) Macroparasites in Antarctic Penguins. In: *Biodiversity and Evolution of Parasitic Life in the Southern Ocean* (eds. Klimpel S, Kuhn T, Mehlhorn H), pp. 183-204. Springer International Publishing, Cham.

Dimitrova ZM, Chipev NH, Georgiev BB (1996) Record of *Corynosoma pseudohamanni* Zdzitowiecki, 1984 (Acanthocephala, Polymorphidae) in birds at Livingston Island, South Shetlands, with a review of Antarctic avian acanthocephalans. *Bulgarian Antarctic Research Life Sciences* **1**, 102-110.

Dowle M, Srinivasan A (2019) data.table: Extension of `data.frame`. R package version 1.12.8.

Edgar RC (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods* **10**, 996-998.

Edmonds SJ (1954) Acanthocephala collected by the Australian National Antarctic Research Expedition on Heard Island and Macquarie Island during 1948-50. *Transactions of the Royal Society of South Australia* **78**, 141-144.

Fredes F, Diaz A, Raffo E, Munoz P (2008a) Cryptosporidium spp. oocysts detected using acid-fast stain in faeces of gentoo penguins (*Pygoscelis papua*) in Antarctica. *Antarctic Science* **20**, 495-496.

Fredes F, Madariaga C, Raffo E*, et al.* (2007a) Gastrointestinal parasite fauna of gentoo penguins (*Pygoscelis papua*) from the Peninsula Munita, Bahia Paraiso, Antarctica. *Antarctic Science* **19**, 93-94.

Fredes F, Raffo E, Muñoz P (2007b) Short Note: First report of Cryptosporidium spp. oocysts in stool of Adélie penguin from the Antarctic using acid-fast stain. *Antarctic Science* **19**, 437-438.

Fredes F, Raffo E, Munoz P, Herrera M (2006) Gastrointestinal parasite fauna of nine gentoo penguin chicks (*Pygoscelis papua*) found dead in an Antarctic territory protected zone. *Parasitologia Latinoamericana* **61**, 179-182.

Fredes F, Raffo E, Munoz P, Herrera M, Godoy C (2008b) Gastrointestinal parasite fauna in Adelia penguin (*Pygoscelis adeliae*) in an Antarctic territory protected zone. *Parasitologia Latinoamericana* **63**, 64-68.

Freer J, Mable B, Clucas G*, et al.* (2015) Limited genetic differentiation among chinstrap penguin (*Pygoscelis antarctica*) colonies in the Scotia Arc and Western Antarctic Peninsula.

Friend M, Franson JC (1999) Intestinal Coccidiosis. In: *Field Manual of Wildlife Diseases: General Field Procedures and Diseases of Birds* (eds. Friend M, Franson JC), pp. 207-213. United States Geological Survey, National Wildlife Health Center, Madison, WI.

Georgiev BB, Vasileva GP, Chipev NH, Dimitrova ZM (1996) Cestodes of seabirds at Livingston Island, South Shetlands. *Bulgarian Antarctic Research Life Sciences* **1**, 111-127.

Golemansky V (2002) Coccidian parasites (Apicomplexa: Eucoccidia) of the Penguins from Livingston Island (South Shetland Islands, Antarctica). *Bulgarian Antarctic Research Life Sciences* **3**, 91-95.

Golemansky V (2003) *Eimeria pygosceli* n. sp. (Coccidia: Eimeriidae) from the penguins (Pygoscelidae) of the Livingston Island (the Antarctic). *Acta Zoologica Bulgarica* **55**, 3-8.

González-Acuña D, Hernández J, Moreno L*, et al.* (2013) Health evaluation of wild gentoo penguins (*Pygoscelis papua*) in the Antarctic Peninsula. *Polar Biology* **36**, 1749-1760.

Hammond DM (1973) Life Cycles and Development of Coccidia. In: *The Coccidia: Eimeria, Isospora, Toxoplasma, and related genera* (eds. Hammond DM, Long PL), pp. 45-79. University Park Press, Baltimore, MD.

Heitlinger E, Spork S, Lucius R, Dieterich C (2014) The genome of *Eimeria falciformis*--reduction and specialization in a single host apicomplexan parasite. *BMC Genomics* **15**, 696.

Heneberg P, Georgiev BB, Sitko J, Literák I (2019) Massive infection of a song thrush by *Mesocestoides* sp. (Cestoda) tetrathyridia that genetically match acephalic metacestodes causing lethal peritoneal larval cestodiasis in domesticated mammals. *Parasites & Vectors* **12**, 230.

Hoberg EP (1986) Aspects of Ecology and Biogeography of Acanthocephala in Antarctic Seabirds. *Annales De Parasitologie Humaine Et Comparee* **61**, 199-214.

Holloway HL, Jr., Bier JW (1967) Notes on the host specificity of *Corynosoma hamanni* (Linstow, 1892). *Bulletin of the Wildlife Disease Association* **3**, 76-78.

Jarman SN, McInnes JC, Faux C*, et al.* (2013) Adelie penguin population diet monitoring by analysis of food DNA in scats. *PLoS One* **8**, e82227.

Jirků M, Jirků M, Oborník M, Lukeš J, Modrý D (2009) A Model for Taxonomic Work on Homoxenous Coccidia: Redescription, Host Specificity, and Molecular Phylogeny of *Eimeria ranae* Dobell, 1909, with a Review of Anuran-Host *Eimeria* (Apicomplexa: Eimeriorina). *Journal of Eukaryotic Microbiology* **56**, 39-51.

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**, 772-780.

Khan JS, Provencher JF, Forbes MR*, et al.* (2019) Parasites of seabirds: A survey of effects and ecological implications. In: *Advances in Marine Biology* (ed. Sheppard C), pp. 1-50. Academic Press.

Leiper RT, Atkinson EL (1914) Helminths of the British Antarctic Expedition 1910-1913. *Proceedings of the Zoological Society of London* **1**, 222-226.

Lescroel A, Ballard G, Gremillet D, Authier M, Ainley DG (2014) Antarctic climate change: extreme events disrupt plastic phenotypic response in Adelie penguins. *PLoS One* **9**, e85291.

Levine ND (1973) Introduction, History, and Taxonomy. In: *The Coccidia: Eimeria, Isospora, Toxoplasma, and related genera* (eds. Hammond DM, Long PL), pp. 1-22. University Park Press, Baltimore, MD.

Levine ND (1988) *The Protozoan Phylum Apicomplexa : Volume 1*, 1st edn. CRC Press,, Boca Raton, FL, USA.

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**, 1834-1853.

Lozupone CA, Hamady M, Kelley ST, Knight R (2007) Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. *Applied and Environmental Microbiology* **73**, 1576-1585.

Lynch HJ, Fagan WF, Naveen R (2010) Population trends and reproductive success at a frequently visited penguin colony on the western Antarctic Peninsula. *Polar Biology* **33**, 493-503.

Marquardt WC (1973) Host and Site Specificity in the Coccidia. In: *The Coccidia: Eimeria, Isospora, Toxoplasma, and related genera* (eds. Hammond DM, Long PL). University Park Press, Baltimore, MD.

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011* **17**, 3.

Martin MA, Ortiz JM, Seva J*, et al.* (2016) Mode of Attachment and Pathology Caused by *Parorchites Zederi* in Three Species of Penguins: *Pygoscelis Papua*, *Pygoscelis Adeliae*, and *Pygoscelis Antarctica* in Antarctica. *J Wildl Dis* **52**, 568-575.

Mawson PM (1953) Parasitic Nematoda collected by the Australian National Antarctic Research Expedition: Heard Island and Macquarie Island, 1948–1951. *Parasitology* **43**, 291-297.

McDonald ME (1969) *Catalogue of helminths of waterfowl (Anatidae)* Bureau of Sport Fisheries and Wildlife, U.S. Fish and Wildlife Service, Washington, D.C.

McInnes JC, Alderman R, Deagle BE*, et al.* (2017) Optimised scat collection protocols for dietary DNA metabarcoding in vertebrates. *Methods in Ecology and Evolution* **8**, 192-202.

McInnes JC, Emmerson L, Southwell C, Faux C, Jarman SN (2016) Simultaneous DNA-based diet analysis of breeding, non-breeding and chick Adelie penguins. *R Soc Open Sci* **3**, 150443.

McMurdie PJ, Holmes S (2013) phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* **8**, e61217.

Minchin PR (1987) An evaluation of the relative robustness of techniques for ecological ordination. In: *Theory and models in vegetation science: Proceedings of Symposium, Uppsala, July 8–13, 1985* (eds. Prentice IC, van der Maarel E), pp. 89-107. Springer Netherlands, Dordrecht.

Montero E, Gonzalez LM, Chaparro A*, et al.* (2016) First record of *Babesia* sp. in Antarctic penguins. *Ticks Tick Borne Dis* **7**, 498-501.

Moser BK, Stevens GR (1992) Homogeneity of Variance in the Two-Sample Means Test. *The American Statistician* **46**, 19-21.

Novilla MN, Carpenter JW (2004) Pathology and pathogenesis of disseminated visceral coccidiosis in cranes. *Avian Pathology* **33**, 275-280.

Novilla MN, Carpenter JW, Jeffers TK, White SL (1989) Pulmonary lesions in disseminated visceral coccidiosis of sandhill and whooping cranes. *J Wildl Dis* **25**, 527-533.

O'Leary NA, Wright MW, Brister JR*, et al.* (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research* **44**, D733-D745.

Obendorf DL, McColl K (1980) Mortality in Little Penguins (*Eudyptula minor*) along the coast of Victoria, Australia. *Journal of Wildlife Diseases* **16**, 251-259.

Ogedengbe JD, Hanner RH, Barta JR (2011a) DNA barcoding identifies *Eimeria* species and contributes to the phylogenetics of coccidian parasites (Eimeriorina, Apicomplexa, Alveolata). *Int J Parasitol* **41**, 843-850.

Ogedengbe JD, Hunter DB, Barta JR (2011b) Molecular identification of *Eimeria* species infecting market-age meat chickens in commercial flocks in Ontario. *Vet Parasitol* **178**, 350-354.

Ogedengbe ME, El-Sherry S, Whale J, Barta JR (2014) Complete mitochondrial genome sequences from five *Eimeria* species (Apicomplexa; Coccidia; Eimeriidae) infecting domestic turkeys. *Parasit Vectors* **7**, 335.

Oliphant TE (2006) *A guide to NumPy (Vol. 1)* Trelgol Publishing, USA.

Oliveira PSd, Ferreira MA, Silva LMd*, et al.* (2017) Diversity and distribution of coccidia of wild birds in an Atlantic forest fragment area in southeastern Brazil. *Revista Brasileira De Parasitologia Veterinaria* **26**, 457-464.

Olson PD, Cribb TH, Tkach VV, Bray RA, Littlewood DTJ (2003) Phylogeny and classification of the Digenea (Platyhelminthes: Trematoda). *International Journal for Parasitology* **33**, 733-755.

Pagès H, Aboyoun P, Gentleman R, DebRoy S (2019) Biostrings: Efficient manipulation of biological strings.

Palacios MJ, Valera F, Barbosa A (2012) Experimental assessment of the effects of gastrointestinal parasites on offspring quality in chinstrap penguins (*Pygoscelis antarctica*). *Parasitology* **139**, 819-824.

Polito MJ, Trivelpiece WZ, Karnovsky NJ*, et al.* (2011) Integrating stomach content and stable isotope analyses to quantify the diets of pygoscelid penguins. *PLoS One* **6**, e26642.

Poulin R (2007) *Evolutionary ecology of parasites*, 2nd edn. Princeton University Press, Princeton, N.J.

Prudhoe S (1969) *Cestodes from Fish, Birds and Whales* Mawson Institute for Antarctic Research, The University of Adelaide.

Prum RO, Berv JS, Dornburg A*, et al.* (2015) A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature* **526**, 569-573.

Reid AJ, Blake DP, Ansari HR*, et al.* (2014) Genomic analysis of the causative agents of coccidiosis in domestic chickens. *Genome Res* **24**, 1676-1685.

Roeder AD, Marshall RK, Mitchelson AJ*, et al.* (2001) Gene flow on the ice: genetic differentiation among Adelie penguin colonies around Antarctica. *Mol Ecol* **10**, 1645-1656.

Rueckert S, Betts EL, Tsaousis AD (2019) The Symbiotic Spectrum: Where Do the Gregarines Fit? *Trends in Parasitology* **35**, 687-694.

Rueckert S, Wakeman KC, Jenke-Kodama H, Leander BS (2015) Molecular systematics of marine gregarine apicomplexans from Pacific tunicates, with descriptions of five novel species of *Lankesteria*. *Int J Syst Evol Microbiol* **65**, 2598-2614.

Schwarz RS, Jenkins MC, Klopp S, Miska KB (2009) Genomic analysis of *Eimeria* spp. populations in relation to performance levels of broiler chicken farms in Arkansas and North Carolina. *J Parasitol* **95**, 871-880.

Sherrill-Mix S (2019) taxonomizr: Functions to Work with NCBI Accessions and Taxonomy v. 0.5.3.

Simdyanov TG, Guillou L, Diakin AY*, et al.* (2017) A new view on the morphology and phylogeny of eugregarines suggested by the evidence from the gregarine *Ancora sagittata* (Leuckart, 1860) Labbé, 1899 (Apicomplexa: Eugregarinida). *PeerJ* **5**, e3354-e3354.

Trathan PN, Garcia-Borboroglu P, Boersma D*, et al.* (2015) Pollution, habitat loss, fishing, and climate change as critical threats to penguins. *Conservation Biology* **29**, 31-41.

Vidal V, Ortiz J, Diaz JI*, et al.* (2012) Gastrointestinal parasites in Chinstrap Penguins from Deception Island, South Shetlands, Antarctica. *Parasitology Research* **111**, 723-727.

Vidal V, Ortiz J, Diaz JI*, et al.* (2016) Morphological, molecular and phylogenetic analyses of the spirurid nematode *Stegophorus macronectes* (Johnston & Mawson, 1942). *J Helminthol* **90**, 214-222.

Whinam J, Chilcott N, Bergstrom DM (2005) Subantarctic hitchhikers: expeditioners as vectors for the introduction of alien organisms. *Biological Conservation* **121**, 207-219.

Wickham H (2016) *ggplot2: Elegant Graphics for Data Analysis* Springer-Verlag, New York.

Wille M, Harvey E, Shi M*, et al.* (2020) Sustained RNA virome diversity in Antarctic penguins and their ticks. *The ISME Journal*.

Williams RB (1999) A compartmentalised model for the estimation of the cost of coccidiosis to the world's chicken production industry. *International Journal for Parasitology* **29**, 1209-1229.

Williams TD (1995) *The penguins : Spheniscidae* Oxford University Press, Oxford ; New York.

Windingstad RM, McDonald ME, Locke LN, Kerr SM, Sinn JA (1980) Epizootic of Coccidiosis in Free-Flying Lesser Scaup. *Avian Diseases* **24**, 1044-1049.

Wood JR, Wilmshurst JM, Rawlence NJ*, et al.* (2013) A megafauna's microfauna: gastrointestinal parasites of New Zealand's extinct moa (Aves: Dinornithiformes). *PLoS One* **8**, e57315-e57315.

Wright CA (1954) Trematodes of the genus Renicola from birds in British zoos, with descriptions of two new species. *Proceedings of the Zoological Society of London* **124**, 51-61.

Wright ES (2016) Using DECIPHER v2.0 to Analyze Big Biological Sequence Data in R. *The R Journal* **8**, 352-359.

Yabsley MJ (2009) Eimeria. In: *Parasitic Diseases of Wild Birds*, pp. 162-180. Wiley-Blackwell.

Yamasaki H, Ohmae H, Kuramochi T (2012) Complete mitochondrial genomes of *Diplogonoporus balaenopterae* and *Diplogonoporus grandis* (Cestoda: Diphyllobothriidae) and clarification of their taxonomic relationships. *Parasitology International* **61**, 260-266.

Zdzitowiecki K (1991) Antarctic Acanthocephala. In: *Synopses of the Antarctic Benthos. Volume III* (eds. Wagele JW, Sieg J), pp. 1-116. Koeltz Scientific Books, Koenigstein.

Chapter 10. Discussion, Future Directions, and Concluding Remarks

The geographic distribution of the Gentoo Penguin (*Pygoscelis papua*) species complex spans a large latitudinal range across varied habitats and climates. This wide variance in biophysical environments, coupled with the distances between archipelagos with their sizable populations, invites questions regarding the evolution of this species complex alongside host pathogens and microbes. I set out with a prediction that these seabirds are excellent for testing signals of host-pathogen interaction in the wild and that geographic and behavioural barriers to gene flow would lead to divergence and differentiation in the respective genetic structure of both penguin and parasite.

Given existing knowledge of both the role of the Polar Front as a barrier to gene flow in pelagic organisms and seabirds in the Southern Ocean (Clucas *et al.* 2016; Fraser *et al.* 2012; Friesen 2015; Friesen *et al.* 2007; Munro & Burg 2017; Pertierra *et al.* 2020) and the philopatric tendencies of Gentoo Penguins within island groups (Stonehouse 1970; Tanton et al. 2004; Ratcliffe and Trathan 2011), this work sought to apply molecular techniques to uncover whether Gentoo Penguins and their enteric bacteria, viruses, and eukaryotic parasites might present parallel structure: across the Front and among archipelagos of the Scotia Arc and Antarctic Region.

In the course of the preceding data chapters, this thesis has shown that Gentoo Penguins North and South of the Polar Front in the South Atlantic are on clearly separate evolutionary trajectories based on evidence from mitochondrial, neutral and immunogenetic markers. However, molecular surveys of bacterial, viral, and eukaryotic parasite presence revealed that the geospatial picture did not fit neatly into the expected mirroring of host

structure in all cases. To shed light on this relationship, I sought to answer the following core questions:

- What microbes/pathogens/parasites are present?
- What is their geospatial distribution?
- What host-specific patterns are present?
    - Are microbes species-specific, genus-specific, or broader in their host utilisation (where possible to compare)?
    - Within a single host species, are the same microbes present across its range?

Three clear themes emerged across the course of this work. Firstly, as stated above, the Polar Front, inter-archipelago distances, and philopatry act as barriers to host gene flow in the Southern Ocean. Secondly, Pygoscelid penguins harbour more viral, bacterial, and eukaryotic parasite diversity than previously described. This diversity does not necessarily decrease when moving from milder Sub-Antarctic regions to the colder, 'pristine' Antarctic wilderness at higher latitudes. Instead, sampling effort is the limiting factor of our current knowledge of biodiversity here. Thirdly, there are commonalities in microbial sharing across seabird hosts in this sector that hint at substantial biotic connectivity in the pathogen realm, whether through direct physical encounters and habitat sharing, or through common access to the same marine resources and prey items. Within Gentoo Penguins, connectivity among microbes is noted in some parasite groups, even across the oceanic and behavioural barriers to host gene flow, such as the Polar Front. In this general discussion, I will draw together the key results from the preceding data chapters and explore the broader implications of these three themes in the context of the overall work, along with future directions for the field to explore.

Host Spatial Divergence: Peering through Time across Ocean Barriers

The major takeaway from the first section (The Host) of this thesis was the pronounced differentiation of Gentoo Penguin populations—at neutral and functional markers—in the Scotia Arc and Antarctic Peninsula. In **Chapter 2**, we found that this region is home to distinct population groupings, separated by the extrinsic barrier of the Antarctic Polar Front, and reinforced by intrinsic factors such as low levels of dispersal and strong philopatric behaviour. Strong population structure revealed that the Falkland Islands/Malvinas, home to the world's largest population of this species, are distinct from populations breeding south of the Polar Front in South Georgia, the South Orkneys, South Shetland Islands, and the Antarctic Peninsula. This publication (Levy *et al.* 2016) was the first to use neutral genetic evidence to call into question the assignment of South Georgia's Gentoo Penguins to a Northern Gentoo subspecies (*P. papua papua*) with the Falkland Islands (Stonehouse 1970), and has played an important role in challenging the taxonomic orthodoxy of a two sub-species model. These findings have been supported and extended by subsequent work that indicates at least four cryptic clades exist worldwide (Clucas *et al.* 2018; Pertierra *et al.* 2020; Vianna *et al.* 2017). The estimated time scale of clade divergence (~0.5-3 million years ago) and level of genetic differentiation present is likely sufficient to propose four sub-species, or even distinct species, as more genomic, morphological, and behavioural evidence of divergence comes to light in the future.

Other evolutionary ecology studies that use island models to test for selection are often set against the backdrop of previous population bottlenecks or founder events that make it difficult to determine whether any genetic shifts might be more attributable to drift (González-Quevedo *et al.* 2015; Spurgin *et al.* 2014). In the case of Gentoo Penguins, there is little evidence of any severe bottlenecks across the South Atlantic range, though there is a history of colonisation events (Cole *et al.* 2019; Pena *et al.* 2014). Within the Western

Antarctic Peninsula, there is rapid Gentoo Penguin population growth and colonisation of southward expansion as a result of climate change and sea ice loss (Lynch *et al.* 2012), yet there is no significantly depleted genetic diversity (microsatellite haplotype diversity in **Chapter 2**, and genome-wide data from Clucas *et al.* (2018)), indicating that localised dynamics still maintain diverse genotypes in the Antarctic Peninsula region.

The South Atlantic and Antarctic Peninsula populations of Gentoo Penguins therefore present an ideal sector of the world to evaluate drivers of evolutionary selection in a wild vertebrate. It contains genetically distinct populations of considerable size (>10,000 breeding pairs) that do not appear to intermix at the regional scale (across large distances or ocean barriers) as well as sustained levels of host diversity demonstrate gradual divergence in neutral loci (Clucas *et al.* 2014). Armed with this knowledge, it was appropriate to consider genes of interest in the immune system, where pathogens and their hosts most closely interact. Do specific innate immune genes, a proxy for understanding certain host-parasite (biotic) interactions over an organism's evolutionary past, also present the same patterns of geographic differentiation found in host neutral and mitochondrial markers?

The innate immune system is the first sentry encountered by an invading pathogen. Thus, the Toll-like receptors (TLRs), which recognise microbe-associated molecular patterns (MAMPs), could provide an indication of whether pathogen pressures existed in the Gentoo Penguin's evolutionary past across its range. **Chapter 3** (Levy *et al.* 2020) considered the genetic diversity of three TLRs across the four clades of Gentoo Penguins, and yielded a number of findings that support the hypothesis that distinct habitats and biotic interactions shaped the innate immune system. We selected TLRs that are well-studied in birds, and examined the full length of their coding genes (which is often not done in these types of studies). The selection of TLR4, 5 and 7 would also allow consideration of both bacterial and viral agonists. TLR4 responds to lipopolysaccharide from Gram-negative bacterial cell walls

(Chow *et al.* 1999; Hopper *et al.* 2012) although some other agonists have been identified (Rallabhandi *et al*. 2012), TLR5 recognises bacterial flagellin (Gewirtz *et al.* 2001), and TLR7 responds to single-stranded RNA, typically U-rich viral RNA (Diebold *et al.* 2006; Lund *et al.* 2004).

This work represented the largest geospatial scale undertaken in any immunogenetic study outside of humans, and our results were informative. The genes encoding bacterial-responding TLRs (*TLR4* and *TLR5*), were highly diverse with patterns of population structure somewhat in line with expectations from the host population structure. There were also clear indications of positive selection at particular amino acid residues in TLR4 and TLR5. Moreover, some of the highly selected sites were in the transmembrane and intracellular Toll/interleukin-1 receptor domains which are usually highly conserved (Mikami *et al.* 2012). This was in contrast to *TLR7*, which was less diverse and appeared highly conserved across the full range of Gentoo penguins. Finally, census population size was not correlated to the detected diversity, with even the smallest population sampled (N_c = 12) maintaining high TLR diversity, while some of the larger populations south of the Polar Front were dominated by single haplotypes.

The contrasting patterns of selection in these three TLRs juxtaposed with their unique haplotype compositions indicate that these differences were not likely to have resulted from geographic isolation or drift. In the opposite case—the absence of selective pressure—we would expect to see similar patterns in each marker, and diversity levels correlating to population size. The functional characterisation of selected haplotypes of TLR5 which dominated in different parts of the Gentoo Penguin range, principally between populations found North and South of the Polar Front revealed key differences in signalling capacity. Functional assessment of genotypes in this context is absolutely key to understanding the likely selective pressures responsible for these patterns, and had never been used for a wild

species' immune genes before. Simply, the southerly dominant haplotype signalled more strongly than the more northerly haplotype which provided a strong foundation for more detailed interpretation. First, the very high frequencies of the derived, southern haplotype indicated a strong selective pressure that resulted in the dominance of this haplotype in some Gentoo Penguin populations. Second, the selection in *TLR4* and *TLR5* but not *TLR7* supports the premise that this selective sweep was due to pressures exerted by a flagellated bacterial pathogen. Thirdly, the higher responsiveness of the southerly dominant haplotype may indicate the need to more sensitively detect a particular flagellated bacterial pathogen rather than being tempered by the need to contain the potential for immune pathology and flagellin-based toxic shock-like syndromes. In this way, **Chapter 3**/Levy et al. (2020) demonstrates the potential for functional analyses to influence interpretations of immunogenetic datasets.

The immune system is a complex network of processes and it is highly likely that other parts of this network will have been under differential selection in the populations of Gentoo Penguins. Hence it would be very surprising that the selective forces that altered *TLR4* and *TLR5* across the Polar Front would not affect other genes associated with immunity. Therefore an extended set of genes should be tested for their association with the pressures detected in *TLR4* and *TLR5*. Indeed, a whole genome approach could be very powerful in this circumstance. Population-level sampling is still critical, given the high diversity in Gentoo Penguins, and it may be useful to include other sympatric species that could have experienced similar changes in pathogen profiles.

Ultimately, the geospatial structure in the host TLRs indicate structure in the past or present pathogen challenge. The big question in an evolutionary context remained: Which pathogen pressures led to this immunogenetic structuring? While difficult to identify the microbe challenges of the past, it may be appropriate to examine the modern pathogen profile in these penguins and the patterns in *TLR4* and *TLR5* indicate the importance of

considering bacterial pathogens in this analysis. Moreover, where strong pathogen pressure has caused such large immunogenetic selection, this could render populations susceptible to changes in pathogen profiles. These might occur by chance, climatic change or other or anthropogenic introductions. Immunogenetics adds another layer of key information on top of the presence/absence of a population of animals when considering their response to change and the types of pressure that might need to be considered within conservation efforts.

Diving Into the Pathogen Milieu

Another main theme of this thesis was the increased diversity of microbes encountered through molecular investigation. Uncovering "unknown unknowns" is largely a result of the limited information available regarding the pathogens that circulate in Gentoo Penguin habitats. Antarctic wildlife disease monitoring presents many of the same fiscal, ethical, and practical challenges as in other wild species, with added layers of logistical and environmental hurdles. This is reflected in the sparse case reports and observations of mortality events in the Sub-Antarctic and Antarctic, discussed in **Chapter 4**, where the pathogen literature is biased toward reporting seabird symptoms or deaths occurring in close proximity to research bases, even among systematic attempts to characterise baseline pathogen presence.

In my review of the literature, I turned my attention to the specific work on pathogens found in the genus *Pygoscelis*. Given that many pathogens are shared across closely related hosts, it was important to consider the sympatric congeneric species of Adélie and Chinstrap Penguins wherever possible, even if in smaller numbers. Increased sampling effort and innovations in field laboratories have advanced modern Antarctic pathogen research, but there are trade-offs in either obtaining many samples from varied sites opportunistically, in short time frames, or studying one breeding area over a longer period with a research base

nearby. Antarctic samples are difficult to obtain and often limited in volume or number. Attempting to test for a series of pathogens one at a time is not typically a suitable strategy. To investigate pathogen and parasite presence in penguins, this thesis pursued the approach of maximising geospatial range by screening large numbers of samples, and molecular techniques held promise for parallel sample processing.

In the data chapters that followed, **Chapters 5-9** used high-throughput sequencing-based approaches to identify novel viruses, profile bacterial microbiomes with a classical 16S rRNA approach, and then identify eukaryotic parasites with an 18S target. These sequences were obtained from penguin gastrointestinal tracts or faeces, so could also derive from penguin prey and other ingesta. These studies underscored the importance of continued sampling in order to increase our knowledge of sub-Antarctic and Antarctic biodiversity.

Among viruses, the work in **Chapters 5** and **6**, showed that Pygoscelid and King Penguins are associated with 99 novel single-stranded DNA viruses or circular DNA molecules. We could not fully attribute all of these as infecting (replicating) in penguins, but their identification builds on our understanding of the Antarctic virome. Aside from Penguin Circovirus (PenCV) in an Adélie Penguin and a Chinstrap Penguin living among Gentoo Penguins on Booth Island, the Antarctic DNA viruses did not fall into any known families, and the proteins encoded by the novel Antarctic circular molecules did not cluster closely with those of classified viruses. More work remains to be done to understand the roles of these molecules in the environment.

Recent research points to a rich Antarctic virome in which penguins (and the seabird tick *Ixodes uriae*) are likely reservoirs of viral diversity (Hurt *et al.* 2016; Hurt *et al.* 2014; Smeele *et al.* 2018; Wille *et al.* 2019; Wille *et al.* 2020). This is important to highlight, particularly when the viral ecology and epidemiology of newly identified viruses remains unclear. Opportunities abound for viral exchange between seabirds capable of long-distance

migrations (e.g. Arctic terns and albatrosses) that intermingle with penguins (or their faeces). Monitoring the circulation or transmission of avian influenza strains or other vertebrate-infecting viruses is critical to a OneHealth approach to protect humans and wildlife from emerging epizootics, while also protecting the Antarctic region from novel introgression.

For similar reasons, to provide a baseline for what bacteria might qualify in the realm of potential pathogens in the Pygoscelid penguin microbiome, **Chapter 8** looked into the bacterial genera that were most likely to contain pathogenic species. Parallel sequencing of 16S amplicons was used as a means of rapidly identifying taxa of interest for future wildlife disease monitoring. In a field that is often resource-constrained, understanding what samples contain targets of interest can save time and money. This technique identified sequence variants attributable to 29 potentially pathogenic genera that might affect penguins, either opportunistically or chronically. Some of these sequences marked first reports in their hosts or in the Antarctic region: *Brucella*, *Chlamydia*, *Haemophilus*, *Klebsiella*, *Neisseria*, *Vibrio*, and *Yersinia*. The southernmost Gentoo Penguins featured comparable levels of present taxa of interest as those in the Falkland Islands or South Georgia.

Other studies have sought to identify a handful of pathogenic bacteria through culture, with negative findings also reported (Bonnedahl *et al.* 2005). *Salmonella* was not identified in this sample set, but frequently gets attention as bacteria (along with *Campylobacter* spp., *E. coli*, and *Yersinia*) that can provide insight for assessment of anthropogenic introduction or natural occurrence (Dimitrov *et al.* 2009; Iveson *et al.* 2009; Vigo *et al.* 2011). Antibiotic resistance is of growing worldwide concern, and the recent reports of resistant bacterial strains in the Antarctic (Bonnedahl *et al.* 2008; Cerdà-Cuéllar *et al.* 2019; González-Acuña *et al.* 2013; Hernández & González-Acuña 2016; Jara *et al.* 2020; Miller *et al.* 2009; Retamal *et al.* 2017; Van Goethem *et al.* 2018; Wang *et al.* 2016) should encourage further screening of varied hosts and sample types to protect against reverse

zoonoses or introductions, with negative findings also presented to aid in baseline comparisons over time.

Finally, in terms of undersampled diversity, **Chapter 9** identified that at least some of the eukaryotic parasites present in the faecal samples from Pygoscelid penguins represent a very diverse set of organisms. The parallel sequencing (targeting 18S) was very sensitive and most likely identified not just parasites that infect penguins but also some that affect organisms in the diet of penguins (e.g. fish). This appears to be a critical observation: the more we look, the more we find when it comes to Antarctic biodiversity, and further molecular and morphological characterisation is needed across taxa in this environment.

Microbial Geospatial and Host-Related Patterning: Here, There, Everywhere?

The unique and understudied nature of the Southern Ocean environment brings us back to the theme of spatial patterning among hosts and their parasites. A great deal of research in evolutionary ecology has been devoted to Theodosius Dobzhansky's observation that "for most groups of organisms, tropical environments support a greater diversity of species than do temperate- or cold-zone environments" (Dobzhansky 1950). Many reviews have been written to document the growing number of hypotheses proposed to explain this gradient in species richness (Pianka 1989; Preisser 2019; Rohde 1992; Willig *et al.* 2003), but in this case, penguins present a notable and recognised exception as a taxon with an extratropical peak in diversity, alongside procellariform seabirds (Chown *et al.* 1998) and pinnipeds (Proches 2001). If penguin hosts deviate from this prevailing macroecological trend, then it is reasonable to hypothesise that at least some groups of their microbes and pathogens may share this "deviant pattern".

Additionally, based on the host structure described in **Chapters 2** and **3,** some differentiation of pathogens north and south of the Polar Front was expected, with Gentoo

Penguins on the Falkland Islands presenting unique microbes from those in South Georgia, the South Shetland Islands, and the Western Antarctic Peninsula. In addition, a latitudinal diversity gradient might have been expected to play a part in differential patterns seen in microbe, pathogen, and parasite presence north and south of the Polar Front.

Unfortunately, our swab sampling scheme did not allow us to compare *viral* diversity across the Polar Front, given we did not have swabs from live penguins on the Falkland Islands. This is one area that would have been beneficial to explore, not just for comparative purposes, but for epizootic monitoring given the close proximity of penguins to human habitation, agricultural activity, and livestock (sheep, cattle, pigs, and chickens, among others) in that archipelago, and the history of avian pox mortality events (Munro 2006; Pistorius 2009). On the other hand, from the molecular surveys of *eukaryotic parasite* and *bacterial* presence, the geospatial picture did not fit neatly into the expectations of host population structure.

Bearing in mind that many eukaryotic parasites require some level of maturation of eggs/cysts/oocysts outside the host before becoming infective (e.g. many helminths and the protozoan Eimerian parasites), these groups could be hypothesised to be affected dramatically by latitudinal differences in climate (especially temperature). For example, areas with milder environmental conditions, such as the Falkland Islands and to some extent, South Georgia, might have different parasite profiles than more southerly (colder) sites. Microscopy results in **Chapter 9** seemed to support this initially, but overall, visual examination of faecal samples underestimated true parasite presence relative to molecular screening, and southerly Antarctic colonies also had high levels of coccidian abundance and diversity. Given that more sequence variants (presumed taxa) were found when larger sample sizes were screened it is clear that more birds/sites need to be sampled to obtain a true picture. However, as alluded

to in the above section on microbial diversity, the Antarctic is not as 'pristine' or free of viruses, bacteria, and macroparasites as the species gradient may have led us to believe.

Clear differentiation across geographic regions was expected among the guano samples that allowed for north/south comparisons in both bacterial (**Chapter 7** and **8**) and eukaryotic parasite (**Chapter 9**) presence and prevalence. However, in **Chapter 7**, there was instead little structuring of the gastrointestinal bacterial microbiota in Gentoo Penguins by region. Within-colony (inter-individual) variation in the microbiome was greater than between-colony variation for this species. Additionally, sampling year had more of an effect (albeit small) than placement within a geographic region. This likely reflects the impact of fluctuations in localised conditions on the host.

Data originated from a population-level sample, and not based on individual-level tracking, which limits what conclusions can be drawn from the inter-annual structuring of the bacterial microbiome. Though conditions could not be controlled as they might in a laboratory-based experiment, future research on wild penguins could investigate what variables have the greatest impact on inter-annual microbiome variations. Potential factors include prey composition, given that gut bacterial composition is linked to diet (Harris *et al.* 2019), environmental variables (precipitation, sea, and air temperature) that could affect host stress levels and energy expenditures (Barbosa *et al.* 2007; Cockrem *et al.* 2006; Colominas-Ciuro *et al.* 2017), sex (Escallón *et al.* 2019) and breeding stage (Dewar *et al.* 2014; Yew *et al.* 2017), all of which we could not precisely account for in our study. Gentoo Penguins are known to experience year-to-year fluctuation and inter-individual differences in phenology (Black 2016), breeding success, and dietary makeup at localised sites based on studies in the Falkland Islands (Baylis *et al.* 2012; Pütz *et al.* 2001), South Georgia (Williams 1990; Xavier *et al.* 2017), the South Orkney Islands (Coria *et al.* 2000), the South Shetland Islands (Juáres *et al.* 2013), and the Antarctic Peninsula (Dunn *et al.* 2019), so a fuller understanding of the

microbial ecology relative to that of the host is worth pursuing in light of rapid environmental change. Environmental stressors and competition for prey with fisheries may alter penguin body condition, particularly in vulnerable periods like the breeding season or moult-fast, which occur during the austral summer and the weeks that follow (Thouzeau *et al.* 2003). In these challenging states, physiological imbalances could arise that lead to negative fitness consequences in the host, especially for threatened species, though these effects are poorly understood (West *et al.* 2019).

Though this thesis could not delve into full assessments of the physiological consequences of potential pathogen detection in penguin hosts, the presence of Penguin Circovirus (PenCV) detected in an Adélie Penguin and a Chinstrap Penguin living among Gentoo Penguins on Booth Island (**Chapter 6**) is valuable to highlight from a wildlife disease epidemiology standpoint. This viral species may have links to a feather disorder noted in Adélie Penguins in the Ross Sea, where the DNA sequences of this circovirus were first detected (Morandini *et al.* 2019). The PenCV sequences detected on the Antarctic Peninsula were closer to each other (across host species) than to the previous findings on Ross Island, thousands of kilometres away (Morandini *et al.* 2019). Though no Gentoo Penguin samples in the data set presented this virus, the presence of a potential pathogen circulating across species in the same genus, breeding in such close physical proximity (illustrated by the photo in **Figure 1**) is noteworthy. All other novel Antarctic viral genotypes (**Chapter 5**) were exclusively found in single host species, with noted cases of adult-chick viral sharing in each Pygoscelid host.

Figure 1. Booth Island, Central Western Antarctica Peninsula is a breeding site for all three *Pygoscelis* species. In this photo, Gentoo Penguins (*P. papua*) are found on the far left and in the foreground, an Adélie Penguin (*P. adeliae*) nest is located in the rear centre, and Chinstrap Penguins (*P. antarcticus*) are seen left and right of centre.

In light of the localised variability of Gentoo Penguin environments, I anticipated finding region-specific eukaryotic parasites for them, but was surprised to see that there was a bimodal distribution in the parasite sequences that were region-unique or broadly distributed across the entire range. The same unexpected pattern emerged across host species, when looking at not just congeneric penguins, but also a small number of sympatric flying seabirds, Blue-Eyed Shags. A number of highly abundant *Eimeria* (coccidian) parasite sequences were host-specific, but there were also several that were present in all hosts and all species. This finding complicates our view of the biology of these protozoans, which have been shown to inhabit very narrow host niches (organ-level or cell-level specificity in some cases) in other

vertebrates (Levine 1973, 1988). An alternative, but less likely, explanation (given the high relative abundances seen in some individuals), might be that host-specific coccidians are present in a common prey item, then seen in seabird excreta. Careful analysis of the co-occurrence of dietary genetic material with these prevalent sequences may help disentangle this.

The underlying biological connectivity that may exist in the Antarctic marine food web remains an understudied part of the puzzle in mapping the geospatial and host-specific patterning of microbes detected in seabirds. For example, we could not determine the extent of the association between penguins and the 97 novel circular molecules and DNA virus genomes identified in **Chapter 5**. Given the paucity of available data, some single-stranded DNA circular molecules were most similar, relatively speaking, to sequences from marine sources and were found in two nearby sites, which might suggest trophic sourcing. However, only one circular molecule genotype was shared across host species, found in both an adult Gentoo Penguin and a Chinstrap Penguin, at Booth Island, Antarctica, muddying the picture of food-derived transmission. Trophic processes may still be relevant for some of the bacteria highlighted in **Chapter 8** (e.g. *Flavobacterium*), found near-ubiquitously in gastrointestinal samples across hosts and regions. Broad host- and region-sharing was also seen among helminths, which may be linked to trophic transmission through vertebrate or invertebrate prey. Findings like these serve to highlight the overwhelming gap in our knowledge of the Antarctic marine and terrestrial ecosystems, where microfaunal diversity requires further study.

Future Directions

The strengths of molecular techniques in uncovering biodiversity must also be balanced by what we cannot tell from these approaches. In ideal circumstances, studies could be undertaken that provide greater resolution by integrating additional genetic markers (e.g.

ITS/5.8S/28S for eukaryotes, multilocus sequence typing for bacteria) or whole-genome sequencing. However, the lack of basic descriptions, whether by culture, microscopic, or molecular characterisation, impede accurate identification of the diverse ecosystem that is an individual host's gastrointestinal tract. Scientific effort must be invested into simultaneous morphological, biochemical and molecular characterization of microbes and parasites to prop up the validity of reference databases.

The restrictions in place on activity in Antarctica (Protocol on Environmental Protection to the Antarctic Treaty, also known as the Madrid Protocol) are intended to protect it from the introduction of harmful chemicals, pollutants, or introduced biota (Hughes & Convey 2010). There are valid biosecurity reasons for some of the research-related restrictions on in-place culture or bacterial isolation, just as importation from Antarctica to analytical laboratories abroad merits caution (Grimaldi *et al.* 2011; Grimaldi *et al.* 2015). Research by humans to detect human-introduced disturbance is "self-policing" in nature, and requires commitment from all parties to the Antarctic Treaty, as well as the tourist and fishing industries. Though policy challenges abound (Kerry & Riddle 2009), there is a critical need for improved methods of characterising microbes and pathogens in the field, or through long-term cold chains or preservative techniques that can withstand the logistics of Antarctic research.

It is clear that Antarctic microbial and environmental biology remains in its infancy when it comes to considering host-pathogen interactions that shape evolutionary processes and underpin conservation management. The broad host-sharing we noted in certain microbial or parasitic taxa hints at underlying connectivity that could be structured by the food web shared among the marine mesopredators we studied, or might be maintained by migratory seabirds whose parasites have never been characterised. For example, this thesis presented the first look at the gastrointestinal contents of Blue-Eyed Shags, just one of many

avian species that receive even less attention in the Antarctic region. Ecological studies that look at the full range of a species are needed, because site-specific studies, even in the long-term, can only tell us a fraction of what a migratory flying or diving bird does when not under human observation. Further assessment of the population genetics, dynamics, migratory patterns, pathogen prevalence, and niche partitioning of a greater number of sympatric species is needed. Only then can we consider the interspecific relationships among hosts and their parasites for a broader view of the network of host-parasite-prey interactions layered onto host immune-pathogen coevolution.

From a host perspective, disease outbreaks and deaths are unfortunately common among captive penguins, which has led to assumptions that penguins are pathogen-naïve in the wild (Grimaldi *et al.* 2015). While we have seen that this is not necessarily the case, and that diverse pathogens circulate in the Southern Ocean, we do not know enough about the role of penguins as reservoirs for viral, bacterial, and eukaryotic pathogens, or their susceptibility to pathogens and parasites outside the scope of this work (i.e. ectoparasites, haemoparasites, and fungi). Research is still needed to understand the penguin immune system, and look beyond the limited markers I examined in this work.

Concluding Remarks

Throughout this thesis, I endeavoured to build upon existing knowledge in the fields of population genetics, immunogenetics, virology, microbiomics, parasitology, and molecular ecology. The work identified Gentoo Penguin population genetic structure strongly influenced by oceanic barriers such as the Polar Front, detected underreported diversity in the microbial and pathogen realms, and elucidated geospatial and cross-host patterns among the detected viruses, bacteria, and gastrointestinal eukaryotic parasites. Additionally, this thesis emphasises the importance of interdisciplinary work, scientific team-building, and the value of basic science research. Not only is there a great deal left to be learned about the Gentoo

Penguin itself, but each of the biotic components of the Southern Ocean merit further study to add to our understanding of the biodiversity of this rapidly changing ecosystem.

In the evolutionary past, I showed evidence that pathogen-mediate pressure shaped the current population-level distribution of certain immune genotypes. Given that history, what will come in the future? Gentoo Penguins face multiple threats from anthropogenic environmental change, habitat loss, fisheries pressure, and pollution (Trathan *et al.* 2015), with the role of infectious disease poorly understood. In the presence of the above risk factors, the introduction of infectious disease into threatened populations can result in temporary or permanent declines in abundance, and in some cases, extinction (Smith et al. 2009; Smith et al. 2006). Wildlife disease management for conservation is increasingly important in this context, particularly as the uniqueness of Antarctic biodiversity comes under greater threat (Barnes *et al.* 2006; Cowan & Tow 2004; Frenot *et al.* 2005; Hughes & Convey 2010; Turner *et al.* 2014).

We need concrete and systematic disease monitoring schemes in wild seabird habitats and Antarctic Specially Protected Areas to ensure humans are not the source of the next wave of pathogen pressure. These monitoring programmes should support spatially-distributed and simultaneous sampling efforts that leverage the strengths of global Antarctic research programmes and incorporates modern technological advances to minimise the impact of the research itself. Identifying zoonotic agents in wild penguin reservoirs, and even safeguarding against reverse zoonoses (from humans to seabirds) is a necessary component of a OneHealth management approach for conservation in our changing world.

References

Barbosa A, Merino S, Benzal J, Martinez J, Garcia-Fraile S (2007) Population variability in heat shock proteins among three Antarctic penguin species. *Polar Biology* **30**, 1239-1244.

Barnes DKA, Hodgson DA, Convey P, Allen CS, Clarke A (2006) Incursion and excursion of Antarctic biota: past, present and future. *Global Ecology and Biogeography* **15**, 121-142.

Baylis AMM, Zuur AF, Brickle P, Pistorius PA (2012) Climate as a driver of population variability in breeding Gentoo Penguins *Pygoscelis papua* at the Falkland Islands. *Ibis* **154**, 30-41.

Black CE (2016) A comprehensive review of the phenology of *Pygoscelis* penguins. *Polar Biology*, 1-28.

Bonnedahl J, Broman T, Waldenstrom J*, et al.* (2005) In search of human-associated bacterial pathogens in Antarctic wildlife: report from six penguin colonies regularly visited by tourists. *Ambio* **34**, 430-432.

Bonnedahl J, Olsen B, Waldenstrom J*, et al.* (2008) Antibiotic susceptibility of faecal bacteria in Antarctic penguins. *Polar Biology* **31**, 759-763.

Cerdà-Cuéllar M, Moré E, Ayats T*, et al.* (2019) Do humans spread zoonotic enteric bacteria in Antarctica? *Science of the Total Environment* **654**, 190-196.

Chow JC, Young DW, Golenbock DT, Christ WJ, Gusovsky F (1999) Toll-like Receptor-4 Mediates Lipopolysaccharide-induced Signal Transduction. *Journal of Biological Chemistry* **274**, 10689-10692.

Chown SL, Gaston KJ, Williams PH (1998) Global patterns in species richness of pelagic seabirds: the Procellariiformes. *Ecography* **21**, 342-350.

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

Clucas GV, Younger JL, Kao D*, et al.* (2018) Comparative population genomics reveals key barriers to dispersal in Southern Ocean penguins. *Mol Ecol* **27**, 4680-4697.

Cockrem JF, Potter MA, Candy EJ (2006) Corticosterone in relation to body mass in Adelie penguins (*Pygoscelis adeliae*) affected by unusual sea ice conditions at Ross Island, Antarctica. *Gen Comp Endocrinol* **149**, 244-252.

Cole TL, Dutoit L, Dussex N*, et al.* (2019) Receding ice drove parallel expansions in Southern Ocean penguins. *Proc Natl Acad Sci U S A*.

Colominas-Ciuro R, Santos M, Coria N, Barbosa A (2017) Reproductive effort affects oxidative status and stress in an Antarctic penguin species: An experimental study. *PLoS One* **12**, e0177124.

Coria N, Libertelli M, Casaux R, Darrieu C (2000) Inter-Annual Variation in the Autumn Diet of the Gentoo Penguin at Laurie Island, Antarctica. *Waterbirds: The International Journal of Waterbird Biology* **23**, 511-517.

Cowan DA, Tow LA (2004) Endangered antarctic environments. *Annu Rev Microbiol* **58**, 649-690.

Dewar ML, Arnould JPY, Krause L*, et al.* (2014) Influence of Fasting during Moult on the Faecal Microbiota of Penguins. *PLoS One* **9**.

Diebold SS, Massacrier C, Akira S*, et al.* (2006) Nucleic acid agonists for Toll-like receptor 7 are defined by the presence of uridine ribonucleotides. *Eur J Immunol* **36**, 3256-3267.

Dimitrov K, Metcheva R, Kenarova A (2009) Salmonella presence - an indicator of direct and indirect human impact on gentoo in Antarctica. *Biotechnology & Biotechnological Equipment* **23**, 246-249.

Dobzhansky T (1950) Evolution in the Tropics. *American Scientist* **38**, 208-221.

Dunn MJ, Forcada J, Jackson JA*, et al.* (2019) A long-term study of gentoo penguin (*Pygoscelis papua*) population trends at a major Antarctic tourist site, Goudier Island, Port Lockroy. *Biodiversity and Conservation* **28**, 37-53.

Escallón C, Belden LK, Moore IT (2019) The Cloacal Microbiome Changes with the Breeding Season in a Wild Bird. *Integrative Organismal Biology* **1**.

Frenot Y, Chown SL, Whinam J*, et al.* (2005) Biological invasions in the Antarctic: extent, impacts and implications. *Biological Reviews* **80**, 45-72.

Gewirtz AT, Navas TA, Lyons S, Godowski PJ, Madara JL (2001) Cutting Edge: Bacterial Flagellin Activates Basolaterally Expressed TLR5 to Induce Epithelial Proinflammatory Gene Expression. *The Journal of Immunology* **167**, 1882-1885.

González-Acuña D, Hernández J, Moreno L*, et al.* (2013) Health evaluation of wild gentoo penguins (*Pygoscelis papua*) in the Antarctic Peninsula. *Polar Biology* **36**, 1749-1760.

González-Quevedo C, Spurgin LG, Illera JC, Richardson DS (2015) Drift, not selection, shapes toll-like receptor variation among oceanic island populations. *Mol Ecol* **24**, 5852-5863.

Grimaldi W, Jabour J, Woehler EJ (2011) Considerations for minimising the spread of infectious disease in Antarctic seabirds and seals. *Polar Record* **47**, 56-66.

Grimaldi W, Seddon P, Lyver PB, Nakagawa S, Tompkins D (2015) Infectious diseases of Antarctic penguins: current status and future threats. *Polar Biology* **38**, 591-606.

Harris EV, de Roode JC, Gerardo NM (2019) Diet–microbiome–disease: Investigating diet's influence on infectious disease resistance through alteration of the gut microbiome. *PLOS Pathogens* **15**, e1007891.

Hernández J, González-Acuña D (2016) Anthropogenic antibiotic resistance genes mobilization to the polar regions. *Infect Ecol Epidemiol* **6**, 32112.

Hopper AT, Campbell BM, Kao H, Pintchovski SA, Staal RGW (2012) Recent Developments in Targeting Neuroinflammation in Disease. In: *Annual Reports in Medicinal Chemistry* (ed. Desai MC), pp. 37-53. Academic Press.

Hughes KA, Convey P (2010) The protection of Antarctic terrestrial ecosystems from inter- and intra-continental transfer of non-indigenous species by human activities: A review of current systems and practices. *Global Environmental Change* **20**, 96-112.

Hurt AC, Su YCF, Aban M*, et al.* (2016) Evidence for the Introduction, Reassortment, and Persistence of Diverse Influenza A Viruses in Antarctica. *J Virol* **90**, 9674-9682.

Hurt AC, Vijaykrishna D, Butler J*, et al.* (2014) Detection of Evolutionarily Distinct Avian Influenza A Viruses in Antarctica. *Mbio* **5**.

Iveson JB, Shellam GR, Bradshaw SD*, et al.* (2009) *Salmonella* infections in Antarctic fauna and island populations of wildlife exposed to human activities in coastal areas of Australia. *Epidemiology and Infection* **137**, 858-870.

Jara D, Bello-Toledo H, Domínguez M*, et al.* (2020) Antibiotic resistance in bacterial isolates from freshwater samples in Fildes Peninsula, King George Island, Antarctica. *Sci Rep* **10**, 3145.

Juáres MA, Santos MM, Negrete J*, et al.* (2013) Better late than never? Interannual and seasonal variability in breeding chronology of gentoo penguins at Stranger Point, Antarctica. *Polar Research* **32**, 18448.

Kerry KR, Riddle MJ (2009) *Health of Antarctic wildlife: a challenge for science and policy* Springer, Berlin ; London ; New York.

Levine ND (1973) Introduction, History, and Taxonomy. In: *The Coccidia: Eimeria, Isospora, Toxoplasma, and related genera* (eds. Hammond DM, Long PL), pp. 1-22. University Park Press, Baltimore, MD.

Levine ND (1988) *The Protozoan Phylum Apicomplexa : Volume 1*, 1st edn. CRC Press,, Boca Raton, FL, USA.

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**, 1834-1853.

Levy H, Fiddaman SR, Vianna JA*, et al.* (2020) Evidence of pathogen-induced immunogenetic selection across the large geographic range of a wild seabird. *Mol Biol Evol* **6**, 1708-1726.

Lund JM, Alexopoulou L, Sato A*, et al.* (2004) Recognition of single-stranded RNA viruses by Toll-like receptor 7. *Proc Natl Acad Sci U S A* **101**, 5598-5603.

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.

Mikami T, Miyashita H, Takatsuka S, Kuroki Y, Matsushima N (2012) Molecular evolution of vertebrate Toll-like receptors: evolutionary rate difference between their leucine-rich repeats and their TIR domains. *Gene* **503**, 235-243.

Miller RV, Gammon K, Day MJ (2009) Antibiotic resistance among bacteria isolated from seawater and penguin fecal samples collected near Palmer Station, Antarctica. *Can J Microbiol* **55**, 37-45.

Morandini V, Dugger KM, Ballard G*, et al.* (2019) Identification of a Novel Adelie Penguin Circovirus at Cape Crozier (Ross Island, Antarctica). *Viruses* **11**.

Munro G (2006) Outbreak of avian pox virus in gentoo penguins in the Falklands. February 2006. Falklands Conservation, Stanley, Falkland Islands.

Pena MF, Poulin E, Dantas GP*, et al.* (2014) Have historical climate changes affected Gentoo penguin (*Pygoscelis papua*) populations in Antarctica? *PLoS One* **9**, e95375.

Pertierra LR, Segovia NI, Noll D*, et al.* (2020) Cryptic speciation in gentoo penguins is driven by geographic isolation and regional marine conditions: Unforeseen vulnerabilities to global change. *Diversity and Distributions* **26**, 958-975.

Pianka ER (1989) Latitudinal gradients in species diversity. *Trends in Ecology & Evolution* **4**, 223.

Pistorius PA (2009) Falkland Island Seabird Monitoring Programme. Annual Report 2008/2009. Falklands Conservation, Stanley, Falkland Islands.

Preisser W (2019) Latitudinal gradients of parasite richness: a review and new insights from helminths of cricetid rodents. *Ecography* **42**, 1315-1330.

Proches Ş (2001) Back to the sea: secondary marine organisms from a biogeographical perspective. *Biological Journal of the Linnean Society* **74**, 197-203.

Pütz K, Ingham RJ, Smith JG, Croxall JP (2001) Population trends, breeding success and diet composition of gentoo *Pygoscelis papua,* magellanic *Spheniscus magellanicus* and rockhopper *Eudyptes chrysocome* penguins in the Falkland Islands. A review. *Polar Biology* **24**, 793-807.

Rallabhandi P, Phillips RL, Boukhvalova MS, *et al*. (2012) Respiratory syncytial virus fusion protein-induced toll-like receptor 4 (TLR4) signaling is inhibited by the TLR4 antagonists *Rhodobacter sphaeroides* lipopolysaccharide and eritoran (E5564) and requires direct interaction with MD-2. *mBio* **3**, e00218-12.

Retamal P, Llanos-Soto S, Salas LM*, et al.* (2017) Isolation of drug-resistant *Salmonella enterica* serovar enteritidis strains in gentoo penguins from Antarctica. *Polar Biology* **40**, 2531-2536.

Rohde K (1992) Latitudinal Gradients in Species Diversity: The Search for the Primary Cause. *Oikos* **65**, 514-527.

Smeele ZE, Ainley DG, Varsani A (2018) Viruses associated with Antarctic wildlife: From serology based detection to identification of genomes using high throughput sequencing. *Virus Res* **243**, 91-105.

Spurgin LG, Illera JC, Jorgensen TH, Dawson DA, Richardson DS (2014) Genetic and phenotypic divergence in an island bird: isolation by distance, by colonization or by adaptation? *Mol Ecol* **23**, 1028-1039.

Stonehouse B (1970) Geographic Variation in Gentoo Penguins *Pygoscelis papua*. *Ibis* **112**, 52-57.

Thouzeau C, Froget G, Monteil H, Le Maho Y, Harf-Monteil C (2003) Evidence of stress in bacteria associated with long-term preservation of food in the stomach of incubating king penguins (Aptenodytes patagonicus). *Polar Biology* **26**, 115-123.

Trathan PN, Garcia-Borboroglu P, Boersma D*, et al.* (2015) Pollution, habitat loss, fishing, and climate change as critical threats to penguins. *Conservation Biology* **29**, 31-41.

Turner J, Barrand NE, Bracegirdle TJ*, et al.* (2014) Antarctic climate change and the environment: an update. *Polar Record* **50**, 237-259.

Van Goethem MW, Pierneef R, Bezuidt OKI*, et al.* (2018) A reservoir of 'historical' antibiotic resistance genes in remote pristine Antarctic soils. *Microbiome* **6**, 40.

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

Vigo GB, Leotta GA, Ines Caffer M*, et al.* (2011) Isolation and characterization of *Salmonella enterica* from Antarctic wildlife. *Polar Biology* **34**, 675-681.

Wang F, Stedtfeld RD, Kim O-S*, et al.* (2016) Influence of Soil Characteristics and Proximity to Antarctic Research Stations on Abundance of Antibiotic Resistance Genes in Soils. *Environmental Science & Technology* **50**, 12621-12629.

West AG, Waite DW, Deines P*, et al.* (2019) The microbiome in threatened species conservation. *Biological Conservation* **229**, 85-98.

Wille M, Aban M, Wang J*, et al.* (2019) Antarctic Penguins as Reservoirs of Diversity for Avian Avulaviruses. *J Virol* **93**.

Wille M, Harvey E, Shi M*, et al.* (2020) Sustained RNA virome diversity in Antarctic penguins and their ticks. *ISME J*.

Williams TD (1990) Annual variation in breeding biology of gentoo penguins, *Pygoscelis papua*, at Bird Island, South Georgia. *Journal of Zoology* **222**, 247-258.

Willig MR, Kaufman DM, Stevens RD (2003) Latitudinal Gradients of Biodiversity: Pattern, Process, Scale, and Synthesis. *Annual Review of Ecology, Evolution, and Systematics* **34**, 273-309.

Xavier JC, Trathan PN, Ceia FR*, et al.* (2017) Sexual and individual foraging segregation in Gentoo penguins *Pygoscelis papua* from the Southern Ocean during an abnormal winter. *PLoS One* **12**, e0174850.

Yew WC, Pearce DA, Dunn MJ, Samah AA, Convey P (2017) Bacterial community composition in Adélie (*Pygoscelis adeliae*) and Chinstrap (*Pygoscelis antarctica*) Penguin stomach contents from Signy Island, South Orkney Islands. *Polar Biology* **40**, 2517-2530.

Appendix A1 - Supplementary Data for Chapter 2

Supplementary Table S1. Genetic diversity of Gentoo Penguins (*Pygoscelis papua*) at fourteen breeding sites across the Scotia Arc. Number of alleles (*A*), observed (H_O) and expected (H_E) heterozygosity for each locus in each population and *P*-values for deviations from Hardy-Weinberg Equilibrium.

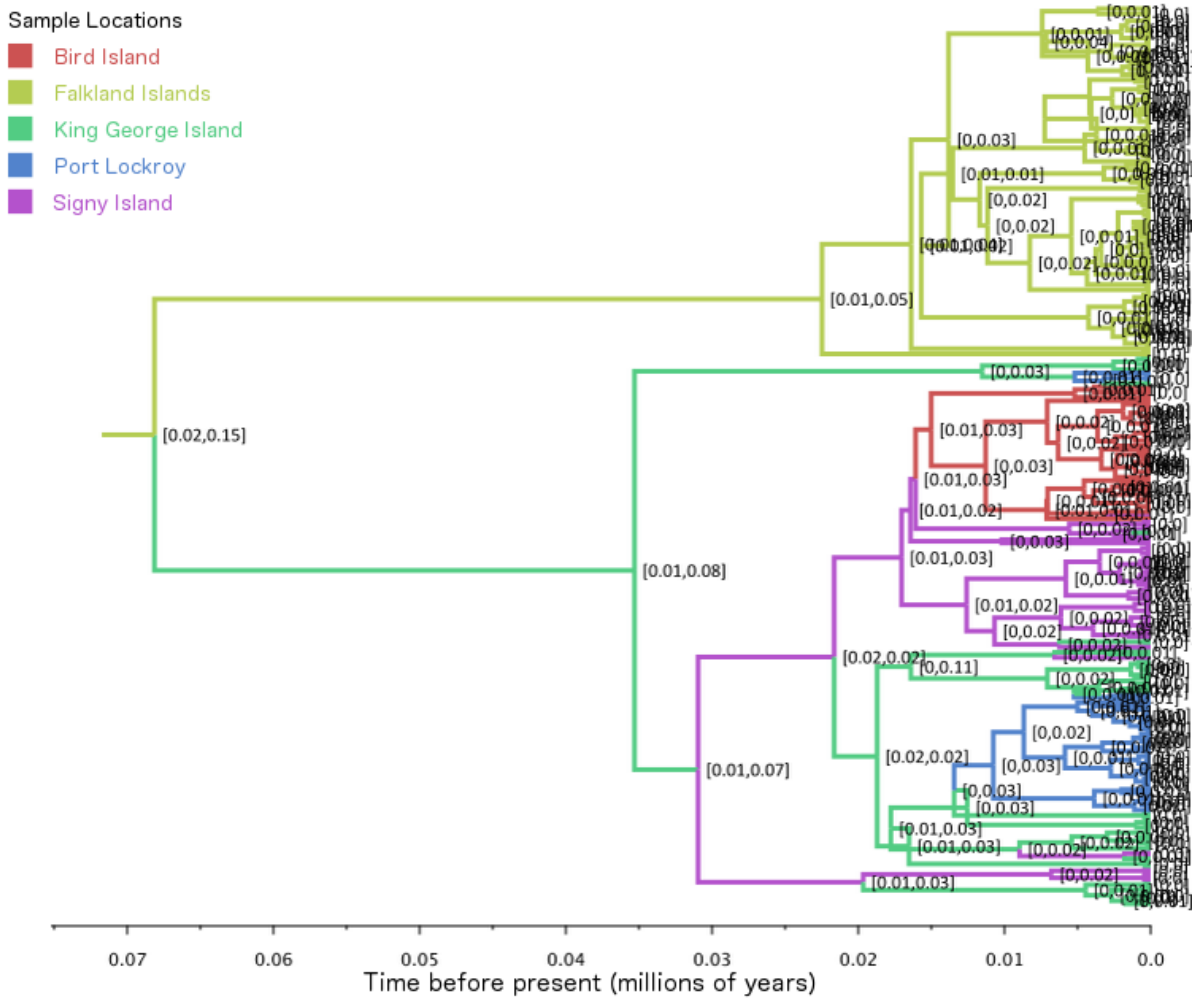
Volunteer Point (Falkland Is.) (n=35)					Kidney Cove (Falkland Is.) (n=46)				Bluff Cove (Falkland Is.) (n=45)			
Locus	*A*	H_O	H_E	*P*	*A*	H_O	H_E	*P*	*A*	H_O	H_E	*P*
Ech030	11	0.62857	0.84720	0.04558	10	0.69565	0.79646	0.02443	10	0.82222	0.83720	0.84380
Ech036	2	0.45714	0.45714	1.00000	2	0.28261	0.37912	0.11401	2	0.53333	0.44944	0.31185
Ech050	2	0.28571	0.47371	0.02879	3	0.21739	0.37291	0.00448	3	0.42222	0.44644	0.62981
Ech065	2	0.02857	0.02857	1.00000	2	0.02174	0.02174	1.00000	2	0.02222	0.06517	0.03343
Ech071	6	0.65714	0.68157	0.97782	5	0.69565	0.70258	0.11752	5	0.66667	0.63446	0.79967
Ech091	6	0.31429	0.55238	**0.00001**	5	0.21739	0.32083	0.00023	5	0.35556	0.46242	0.15642
Emm4	4	0.37143	0.54369	0.02313	4	0.47826	0.55638	0.21176	4	0.71111	0.62996	0.22142
RM3	2	0.40000	0.47371	0.46982	2	0.69565	0.51505	**0.00006**	2	0.35556	0.43346	0.29490
	Mean H_0	0.39286				0.41304				0.48611		
	SD	0.20015				0.26472				0.25405		
Bertha's Beach (Falkland Is.) (n=35)					Ajax Bay (Falkland Is.) (n=34)				New Haven (Falkland Is.) (n=24)			
Locus	*A*	H_O	H_E	*P*	*A*	H_O	H_E	*P*	*A*	H_O	H_E	*P*
Ech030	9	0.65714	0.83395	0.00485	10	0.73529	0.84065	0.09376	8	0.54167	0.84752	0.00190
Ech036	2	0.42857	0.50559	0.49894	2	0.35294	0.50571	0.09422	2	0.50000	0.45390	1.00000
Ech050	2	0.28571	0.35776	0.32953	2	0.26471	0.43327	0.03888	3	0.08333	0.08245	1.00000
Ech065	1	--	--	--	2	0.05882	0.05795	1.00000	1	--	--	--
Ech071	6	0.71429	0.74493	0.59896	4	0.76471	0.69227	0.79393	4	0.75000	0.69592	0.92252
Ech091	4	0.25714	0.50890	**0.00002**	3	0.23529	0.37182	0.05164	3	0.29167	0.46188	0.00835
Emm4	4	0.51429	0.53623	0.94746	4	0.41176	0.48420	0.52436	3	0.45833	0.52039	0.81258
RM3	3	0.45714	0.44596	1.00000	4	0.38235	0.54741	0.00094	2	0.54167	0.48848	0.68327
	Mean H_0	0.47347				0.40073				0.45238		
	SD	0.17211				0.24200				0.21168		

Fox Bay (Falkland Is.) (n=31)					Saunders Penguin Is. (Falkland Is.) (n=25)				Saunders Penarrow Pt. (Falkland Is.) (n=36)			
Locus	*A*	H_O	H_E	*P*	*A*	H_O	H_E	*P*	*A*	H_O	H_E	*P*
Ech030	7	0.64516	0.79376	0.06784	9	0.92000	0.85061	0.11399	9	0.72222	0.79773	0.00655
Ech036	2	0.29032	0.45531	0.05479	2	0.16000	0.21551	0.28696	2	0.50000	0.40689	0.22692
Ech050	2	0.32258	0.27499	0.56760	2	0.08000	0.41143	**0.00009**	2	0.08333	0.17801	0.02091
Ech065	2	0.03226	0.03226	1.00000	1	--	--	--	1	--	--	--
Ech071	5	0.64516	0.71814	0.21469	5	0.52000	0.65469	0.18558	6	0.66667	0.66901	0.06225
Ech091	3	0.22581	0.30196	0.29469	4	0.16000	0.40163	**0.00039**	3	0.19444	0.48083	**0.00000**
Emm4	3	0.61290	0.56108	1.00000	3	0.40000	0.50694	0.30404	3	0.33333	0.49257	0.02637
RM3	3	0.48387	0.44157	0.78190	2	0.28000	0.45796	0.07422	3	0.44444	0.50430	0.70292
	Mean H_0	0.40726				0.36000				0.42063		
	SD	0.22543				0.29028				0.23485		

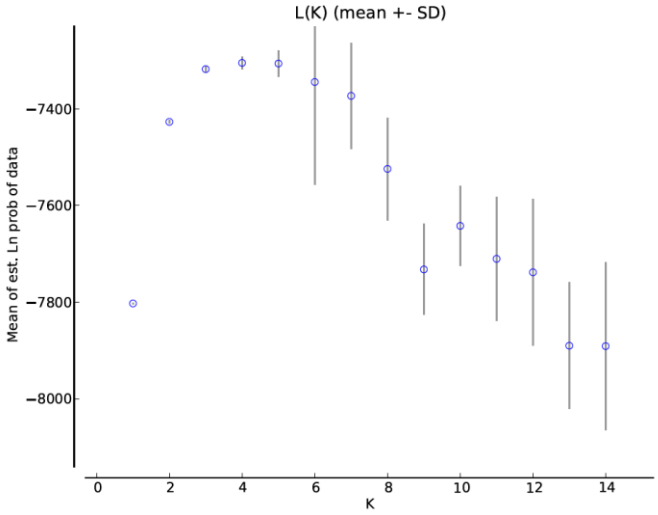
Shallow Harbour (Falkland Is.) (n=45)					Bird Island (S. Georgia) (n=39)				Signy Is. (S. Orkney Is.) (n=37)			
Locus	*A*	H_O	H_E	*P*	*A*	H_O	H_E	*P*	*A*	H_O	H_E	*P*
Ech030	9	0.73333	0.79900	0.25681	9	0.84315	0.88178	0.07640	12	0.83784	0.89448	0.09943
Ech036	2	0.64444	0.45668	0.00704	2	0.28205	0.24542	1.00000	3	0.02703	0.07960	0.01371
Ech050	3	0.24444	0.57353	**0.00000**	2	0.17949	0.16550	1.00000	3	0.32432	0.27953	0.63454
Ech065	1	--	--	--	2	0.02564	0.02564	1.00000	2	0.05405	0.10367	0.08100
Ech071	6	0.75556	0.67516	0.70588	6	0.46154	0.42857	0.72775	5	0.27027	0.31100	0.26213
Ech091	5	0.35556	0.39201	0.17693	5	0.46154	0.47686	0.06045	5	0.32432	0.49056	**0.00062**
Emm4	3	0.22222	0.38926	0.00150	3	0.84615	0.56743	**0.00021**	5	0.48649	0.66679	0.02093
RM3	3	0.31111	0.43121	0.03987	3	0.64103	0.50050	0.06190	4	0.27027	0.28619	0.66324
	Mean H_0	0.46667				0.46757				0.32432		
	SD	0.23518				0.29935				0.25559		

	King George Is. (S. Shetland Is.) (n=40)				Port Lockroy (Western Antarctic Peninsula) (n=38)			
Locus	*A*	H_O	H_E	*P*	*A*	H_O	H_E	*P*
Ech030	9	0.82500	0.84557	0.84932	9	0.79487	0.76190	0.00543
Ech036	2	0.07500	0.07310	1.00000	2	0.02564	0.02564	1.00000
Ech050	2	0.27500	0.33892	0.33383	2	0.15385	0.22644	0.09473
Ech065	1	--	--	--	1	--	--	--
Ech071	6	0.52500	0.47278	0.86475	4	0.15385	0.14685	1.00000
Ech091	5	0.65000	0.55570	0.61187	3	0.51282	0.47519	0.82431
Emm4	4	0.62500	0.56234	0.56857	4	0.28205	0.25408	1.00000
RM3	2	0.25000	0.22152	1.00000	2	0.05128	0.05062	1.00000
	Mean H_0	0.46071				0.28205		
	SD	0.26687				0.27932		

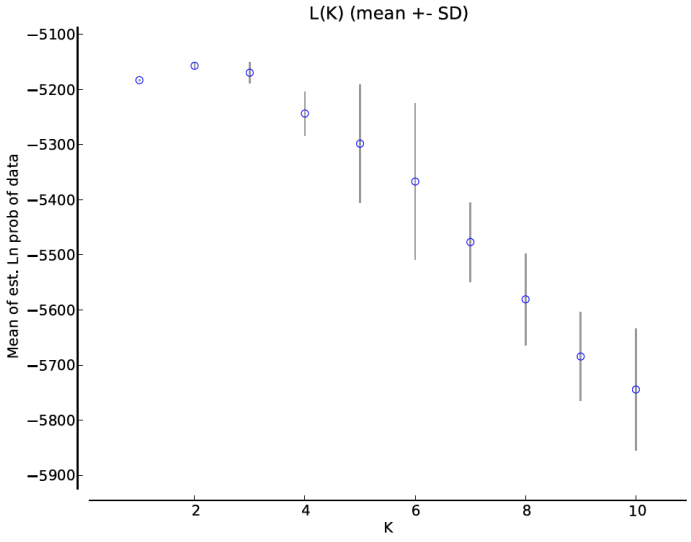
Supplementary Figure S2: Maximum clade credibility tree derived from mtDNA showing the origin and differentiation of *Pygoscelis papua* lineages north (Falkland Islands, light green above) and south of the Polar Front (all other colours and locations). Node colours represent the most likely location of each ancestral node, whilst node labels show the 95% highest posterior densities (HPDs) for node heights (time).



Supplementary Figure 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.



Supplementary Figure 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 A2 - Supplementary Data for Chapter 3

Supplementary Table S1. Polymorphisms and diversity indices for fourteen colonies of Gentoo Penguin at three toll-like receptor loci and the mitochondrial hypervariable control region. (CR = Crozet Island; MAR = Marion Island; COU = Courbet Peninsula, Kerguelen; MO = Pointe du Morne, Kerguelen; CB = Cow Bay, Falkland/Malvinas Islands; BR = Bull Roads, Falkland/Malvinas Islands; BI = Bird Island, South Georgia; MT = Martillo Island, Tierra del Fuego; SIG = Signy Island, South Orkney Islands; COP = Copacabana (Admiralty Bay), King George Island, South Shetland Islands; SP = Stranger Point, King George Island, South Shetland Islands; BO = Bernardo O'Higgins Base, Western Antarctic Peninsula; GGV = Gabriel González Videla Base, Western Antarctic Peninsula; JP = Jougla Point, Western Antarctic Peninsula).

Locus	Colony (*N*)	*S*	*H*	H_d (SD)	π (SD) (x10^{-4})	UPV	AR	*D*	*Fs*
TLR4	CR (7)	5	4	0.6593 (0.1203)	5.86 (1.9)	3	3.802	-0.17122	0.48000
	MAR (8)	5	4	0.7667 (0.0573)	9.80 (1.2)	4	3.858	2.17212*	2.06630
	COU (7)	8	7	0.6923 (0.1366)	8.32 (2.4)	4	5.286	-0.59171	-1.89341
	MO (15)	9	14	0.9034 (0.0362)	11.01 (0.8)	8	7.033	0.73200	-5.86857*
	CB (19)	2	2	0.0526 (0.0492)	0.41 (0.4)	1	1.263	-1.49106*	-0.61060
	BR (7)	0	1	0	0	1	1.000	0	0
	BI (18)	0	1	0	0	1	1.000	0	0
	MT (5)	0	1	0	0	1	1.000	0	0
	SIG (5)	4	3	0.5111 (0.1643)	6.01 (2.3)	1	3.000	0.32418	1.35813
	COP (22)	5	5	0.6342 (0.0511)	7.46 (0.7)	1	3.246	1.63019	1.71487
	SP (5)	4	3	0.5111 (0.1643)	6.45 (2.1)	1	3.000	0.62589	1.51568
	BO (6)	5	3	0.5455 (0.1436)	5.94 (2.3)	2	2.970	-0.31291	1.59874
	GGV (7)	4	2	0.2637 (0.1360)	4.14 (2.1)	1	1.934	-0.53129	2.69666
	JP (24)	4	3	0.0824 (0.0541)	1.12 (0.7)	1	1.417	-1.56204*	-0.73409
	Total (155)	13	21	0.478 (0.035)	6.1 (0.5)	9	3.446	-0.58350	-10.018
TLR5	CR (7)	1	2	0.4396 (0.1120)	1.66 (0.4)	2	1.999	0.84228	0.94377
	MAR (8)	1	2	0.1250 (0.1064)	0.47 (0.4)	2	1.625	-1.16221	-0.70010
	COU (7)	15	8	0.8571 (0.0774)	16.09 (2.9)	7	6.275	-0.40117	-0.88654
	MO (20)	13	17	0.8731 (0.0415)	12.30 (1.3)	14	6.697	0.19907	-6.77889*
	CB (18)	11	17	0.9159 (0.0309)	17.10 (0.7)	12	7.380	2.17976	-4.88768
	BR (7)	9	6	0.7692 (0.0895)	16.09 (1.9)	5	4.856	1.92919	1.04846
	BI (12)	5	5	0.4928 (0.1165)	5.25 (1.6)	4	2.883	0.10549	-0.03412
	MT (5)	9	6	0.8444 (0.1029)	15.72 (2.9)	5	6.000	1.34084	-0.04600
	SIG (5)	5	3	0.5111 (0.1643)	4.37 (2.3)	3	3.000	-1.38818	0.76172
	COP (21)	6	6	0.6527 (0.0485)	4.54 (1.0)	6	3.462	-0.36633	-0.68391
	SP (5)	1	2	0.2000 (0.1541)	0.76 (0.6)	2	2.000	-1.11173	-0.33931
	BO (6)	5	3	0.3182 (0.1637)	3.15 (2.1)	3	2.667	-1.83094*	0.32487
	GGV (7)	2	3	0.3846 (0.1494)	1.54 (0.6)	3	2.648	-0.95919	-0.85452
	JP (24)	1	2	0.1197 (0.0612)	0.45 (0.2)	2	1.512	-0.63670	-0.38233
	Total (152)	20	46	0.781 (0.022)	14.8 (0.6)	32	5.600	0.59898	-23.918

TLR7	CR (7)	0	1	0	0	1	1.000	0	0
	MAR (8)	0	1	0	0	1	1.000	0	0
	COU (7)	3	4	0.4945 (0.1506)	2.32 (0.8)	4	2.978	-0.70770	-1.11799
	MO (18)	2	3	0.3381 (0.0919)	1.13 (0.3)	2	2.275	-0.51702	-0.49065
	CB (17)	1	2	0.0588 (0.0546)	0.19 (0.2)	2	1.235	-1.13783	-1.31506
	BR (7)	1	2	0.3626 (0.1302)	1.16 (0.4)	2	1.945	0.32440	0.64281
	BI (13)	0	1	0	0	1	1.000	0	0
	MT (4)	1	2	0.4286 (0.1687)	1.37 (0.5)	2	2.000	0.33350	0.53626
	SIG (5)	2	3	0.6000 (0.1305)	2.13 (0.6)	2	2.800	-0.18393	-0.27178
	COP (21)	3	4	0.3751 (0.0902)	1.30 (0.3)	3	2.413	-0.90151	-1.43643
	SP (5)	1	2	0.4667 (0.1318)	1.49 (0.4)	1	2.000	0.81980	0.81801
	BO (6)	2	3	0.4394 (0.1581)	1.50 (0.6)	2	2.576	-0.84971	-0.72455
	GGV (7)	2	3	0.3846 (0.1494)	1.30 (0.5)	2	2.407	-0.95919	-0.85452
	JP (23)	4	5	0.3826 (0.0881)	1.34 (0.3)	4	2.510	-1.25133	-2.59143
	Total (148)	9	10	0.296 (0.034)	1.0 (0.1)	8	2.140	-1.67798	-9.379
HVR1	CR (7)	8	4	0.8570 (0.102)	123.5 (33.0)	-	-	0.10944	0.87029
	MAR (8)	5	4	0.7860 (0.113)	68.8 (20.5)	-	-	-0.16751	0.00547
	COU (2)	0	1	0	0	-	-	0	0
	MO (20)	28	16	0.9580 (0.033)	173.2 (10.0)	-	-	-0.93073	-7.01213*
	CB (16)	12	10	0.9000 (0.062)	116.4 (15.4)	-	-	-0.50145	-3.57558
	BR (7)	5	4	0.8570 (0.102)	77.6 (18.5)	-	-	0.13210	-0.04242
	BI (17)	9	8	0.8820 (0.047)	78.2 (13.8)	-	-	-0.75716	-3.70741*
	MT (5)	4	2	0.6000 (0.175)	88.9 (26.0)	-	-	1.64070	3.02249
	SIG (5)	6	5	1.0000 (0.126)	103.7 (18.8)	-	-	-0.19092	-2.37061*
	COP (21)	22	15	0.9670 (0.024)	151.4 (18.9)	-	-	-1.27487	-6.15999*
	SP (6)	16	6	1.0000 (0.096)	227.2 (34.2)	-	-	-0.77134	-1.77289
	JP (24)	18	15	0.9420 (0.031)	127.2 (18.6)	-	-	-1.03686	-7.42519*
	Total (138)	90	79	0.986 (0.003)	744.3 (56.6)	-	-	0.53097	-22.196

S = number of polymorphic sites

H = number of haplotypes

H_d = haplotype diversity (standard deviation)

π = nucleotide diversity

UPV = unique protein variants of translated haplotypes

AR = allelic richness, adjusted for sample size differences

D = Tajima's *D* test statistic, with * indicating p <0.05

Fs = Fu's *Fs* test statistic, with * indicating p <0.02

Supplementary Table S2. Pairwise F_{ST} values (Weir and Cockerham 1984) derived from haplotype frequencies (below diagonal), and pairwise Φ_{ST} values of genetic distance (above the diagonal), calculated using the Tamura and Nei substitution model (Tamura and Nei 1993), with associated p-values, among Gentoo Penguin breeding sites based on toll-like receptor sequence data (tables S2A-C). mtDNA hypervariable region I Φ_{ST} values were calculated using the Kimura 2-Parameter (K80) substitution model (Kimura 1980) with a gamma of 0.27 (table S2D). Comparisons that are significant <0.01 after correction for multiple tests using (SGOF+) are shown in bold (Carvajal-Rodriguez and de Uña-Alvarez 2011).

S2A. *TLR4*

Colony	CR	MAR	COU	MO	CB	BR	BI	MT	SIG	COP	SP	BO	GGV	JP
Crozet Island (CR)	*	0.04120 (0.17969)	0.43041 **(0.00079)**	0.16276 (0.01178)	0.84905 **(0.00000)**	0.78209 **(0.00010)**	0.87027 **(0.00000)**	0.75072 **(0.00000)**	0.54200 **(0.00010)**	0.40124 **(0.00000)**	0.48925 **(0.00109)**	0.55826 **(0.00000)**	0.60966 **(0.00010)**	0.82233 **(0.00000)**
Marion Island (MAR)	0.00110 (0.40541)	*	0.20483 (0.01475)	0.0153 (0.24453)	0.66826 **(0.00000)**	0.54644 **(0.00000)**	0.69000 **(0.00000)**	0.50219 **(0.00158)**	0.31652 (0.01049)	0.25797 **(0.00400)**	0.26514 (0.01416)	0.33837 **(0.00446)**	0.38119 **(0.00228)**	0.65266 **(0.00000)**
Courbet Peninsula (COU)	0.24738 **(0.00158)**	0.13006 (0.14950)	*	0.04851 (0.10870)	0.21641 **(0.00485)**	0.12661 (0.04010)	0.25919 **(0.00030)**	0.08874 (0.10603)	0.00000 (0.41392)	0.06705 (0.08405)	0.00000 (0.50153)	0.00000 (0.39570)	0.00000 (0.35739)	0.18573 (0.00911)
Pointe du Morne (MO)	0.08383 (0.01109)	0.01953 (0.18993)	0.03074 (0.10365)	*	0.39773 **(0.00000)**	0.30278 **(0.00050)**	0.41555 **(0.00000)**	0.27201 **(0.00307)**	0.13365 (0.02901)	0.10606 **(0.00505)**	0.08702 (0.06831)	0.15195 (0.01366)	0.17420 (0.00871)	0.39423 **(0.00000)**
Cow Bay (CB)	0.71152 **(0.00000)**	0.58202 **(0.00000)**	0.28168 **(0.00069**)	0.38276 **(0.00000)**	*	0.00000 (0.99990)	0.00000 (0.99990)	0.00000 (0.99990)	0.30157 (0.01139)	0.31274 **(0.00010)**	0.34009 (0.01257)	0.18803 (0.01119)	0.12601 (0.07079)	0.00000 (0.61390)
Bull Roads (BR)	0.61538 **(0.00000)**	0.46925 **(0.00000)**	0.19231 (0.01554)	0.29735 **(0.00000)**	0.00000 (0.99990)	*	0.00000 (0.99990)	0.00000 (0.99990)	0.19940 (0.06257)	0.24712 **(0.00455)**	0.23125 (0.05425)	0.10916 (0.03168)	0.07692 (0.47787)	0.00000 (0.99990)
Bird Island (BI)	0.75224 **(0.00000)**	0.6231 **(0.00000)**	0.34289 **(0.00010)**	0.40973 **(0.00000)**	0.00000 (0.99990)	0.00000 (0.99990)	*	0.00000 (0.99990)	0.37725 **(0.00743)**	0.33222 **(0.00000)**	0.41492 (0.00812)	0.24677 **(0.00317)**	0.19101 (0.07247)	0.00997 (0.50589)
Martillo Island (MT)	0.57055 **(0.00020)**	0.42329 **(0.00020)**	0.14980 (0.04990)	0.26665 **(0.00010)**	0.00000 (0.99990)	0.00000 (0.99990)	0.00000 (0.99990)	*	0.14818 (0.21107)	0.22308 (0.01901)	0.17778 (0.20889)	0.06971 (0.32304)	0.04328 (0.49352)	0.00000 (0.99990)
Signy Island (SIG)	0.34300 **(0.00050)**	0.21259 (0.00772)	0.00000 (0.42926)	0.10256 (0.01049)	0.27433 (0.01188)	0.19940 (0.05910)	0.37721 **(0.00644)**	0.14815 (0.20602)	*	0.00234 (0.32710)	0.00000 (0.80032)	0.00000 (0.67835)	0.00000 (0.76676)	0.19722 (0.02148)
Copacabana (Admiralty Bay), King George Island (COP)	0.30121 **(0.00000)**	0.20420 **(0.00020)**	0.05519 (0.08088)	0.10420 **(0.00119)**	0.28632 **(0.00000)**	0.22873 **(0.00267)**	0.31195 **(0.00000)**	0.20492 (0.01158)	0.00000 (0.41917)	*	0.00000 (0.48758)	0.05467 (0.14523)	0.07275 (0.10237)	0.28589 **(0.00000)**
Stranger Point, King George Island (SP)	0.34300 **(0.00059)**	0.21259 **(0.00673)**	0.00000 (0.42679)	0.10256 (0.01089)	0.27433 (0.01049)	0.19940 (0.05871)	0.37721 (0.00911)	0.14815 (0.21572)	0.00000 (0.99990)	0.00000 (0.41659)	*	0.00000 (0.60539)	0.00000 (0.61578)	0.24091 (0.02534)
Bernardo O'Higgins Base (BO)	0.33193 **(0.00010)**	0.20624 **(0.00396)**	0.00000 (0.36927)	0.09915 (0.00980)	0.28556 **(0.00228)**	0.20379 (0.03227)	0.36842 **(0.00307)**	0.15691 (0.09583)	0.00000 (0.70231)	0.01547 (0.25443)	0.00000 (0.71211)	*	0.00000 (0.67538)	0.11697 (0.02614)

Gabriel González Videla Base (GGV)	0.47406 **(0.00000)**	0.33467 **(0.00030)**	0.006308 (0.13523)	0.19909 **(0.00010)**	0.10105 (0.06811)	0.07692 (0.45609)	0.19101 (0.07613)	0.04328 (0.49906)	0.00000 (0.61479)	0.08597 (0.06376)	0.00000 (0.61608)	0.00485 (0.38432)	*	0.05175 (0.12306)
Jougla Point (JP)	0.70966 **(0.00000)**	0.58554 **(0.00000)**	0.27327 **(0.00168)**	0.39345 **(0.00000)**	0.00000 (0.99990)	0.00000 (0.99990)	0.00324 (0.50965)	0.00000 (0.99990)	0.23112 (0.02534)	0.28073 **(0.00000)**	0.23112 (0.02237)	0.25542 **(0.00525)**	0.05560 (0.12494)	*

S2B. *TLR5*

Colony	CR	MAR	COU	MO	CB	BR	BI	MT	SIG	COP	SP	BO	GGV	JP
Crozet Island (CR)	*	0.10463 (0.15701)	0.19875 (0.00337)	0.14048 (0.00426)	0.42400 **(0.00000)**	0.61731 **(0.00000)**	0.81884 **(0.00000)**	0.65281 **(0.00000)**	0.86382 **(0.00000)**	0.83274 **(0.00000)**	0.93688 **(0.00000)**	0.88320 **(0.00000)**	0.92114 **(0.00000)**	0.96416 **(0.00000)**
Marion Island (MAR)	0.10463 (0.15602)	*	0.27084 **(0.00030)**	0.18983 **(0.00059)**	0.46395 **(0.00000)**	0.66343 **(0.00000)**	0.84184 **(0.00000)**	0.70960 **(0.00000)**	0.90057 **(0.00000)**	0.84665 **(0.00000)**	0.97033 **(0.00000)**	0.91651 **(0.00000)**	0.95016 **(0.00000)**	0.97642 **(0.00000)**
Courbet Peninsula (COU)	0.12961 (0.02485)	0.27629 **(0.00109)**	*	0.00000 (0.91209)	0.17852 **(0.00050)**	0.33854 **(0.00000)**	0.55967 **(0.00000)**	0.31502 **(0.00000)**	0.50797 **(0.00010)**	0.64872 **(0.00000)**	0.58133 **(0.00000)**	0.54190 **(0.00000)**	0.60325 **(0.00000)**	0.78796 **(0.00000)**
Pointe du Morne (MO)	0.10214 (0.00713)	0.21189 **(0.00000)**	0.00000 (0.56351)	*	0.25438 **(0.00000)**	0.42963 **(0.00000)**	0.56884 **(0.00000)**	0.41306 **(0.00000)**	0.54117 **(0.00000)**	0.63605 **(0.00000)**	0.59046 **(0.00000)**	0.55817 **(0.00000)**	0.59835 **(0.00000)**	0.72147 **(0.00000)**
Cow Bay (CB)	0.28371 **(0.00000)**	0.40690 **(0.00000)**	0.10025 **(0.00000)**	0.09881 **(0.00000)**	*	0.05018 (0.09811)	0.36548 **(0.00000)**	0.05309 (0.11880)	0.33306 **(0.00000)**	0.45208 **(0.00000)**	0.37935 **(0.00000)**	0.34106 **(0.00000)**	0.39910 **(0.00000)**	0.54574 **(0.00000)**
Bull Roads (BR)	0.39560 **(0.00000)**	0.56722 **(0.00000)**	0.18264 **(0.00000)**	0.17234 **(0.00000)**	0.03011 (0.09910)	*	0.49814 **(0.00000)**	0.00000 (0.71864)	0.44207 **(0.00188)**	0.58038 **(0.00000)**	0.50051 **(0.00059)**	0.46193 **(0.00119)**	0.53330 **(0.00020)**	0.73178 **(0.00000)**
Bird Island (BI)	0.52955 **(0.00000)**	0.66409 **(0.00000)**	0.34622 **(0.00000)**	0.30010 **(0.00000)**	0.13386 **(0.00000)**	0.22379 **(0.00059)**	*	0.59937 **(0.00000)**	0.00978 (0.31730)	0.02817 (0.12860)	0.05571 (0.10425)	0.00778 (0.30997)	0.06350 (0.10553)	0.17521 **(0.00000)**
Martillo Island (MT)	0.37560 **(0.00000)**	0.57279 **(0.00000)**	0.14276 **(0.00079)**	0.13905 **(0.00010)**	0.04473 (0.05455)	0.00000 (0.81279)	0.31426 **(0.00010)**	*	0.55172 **(0.00010)**	0.67565 **(0.00000)**	0.62473 **(0.00020)**	0.57910 **(0.00010)**	0.65295 **(0.00000)**	0.83149 **(0.00000)**
Signy Island (SIG)	0.52861 **(0.00000)**	0.71876 **(0.00000)**	0.30275 **(0.00000)**	0.26750 **(0.00000)**	0.10387 (0.00762)	0.19030 (0.01574)	0.00557 (0.31136)	0.27121 **(0.00277)**	*	0.08135 (0.07187)	0.00000 (0.99990)	0.00000 (0.67587)	0.00000 (0.81546)	0.12682 (0.06009)
Copacabana (Admiralty Bay), King George Island (COP)	0.42728 **(0.00000)**	0.54018 **(0.00000)**	0.26264 **(0.00000)**	0.23785 **(0.00000)**	0.11156 **(0.00000)**	0.18623 **(0.00109)**	0.06573 (0.03732)	0.23511 (0.00990)	0.09384 (0.05920)	*	0.09581 (0.03990)	0.08366 (0.04594)	0.11085 (0.01465)	0.20084 **(0.00000)**

Stranger Point, King George Island (SP)	0.66455 **(0.00000)**	0.84565 **(0.00000)**	0.43966 **(0.00000)**	0.36950 **(0.00000)**	0.19855 **(0.00020)**	0.31769 **(0.00267)**	0.01752 (0.28591)	0.42613 **(0.00040)**	0.00000 (0.57499)	0.17399 (0.01495)	*	0.00000 (0.99990)	0.00000 (0.99990)	0.00000 (0.99990)
Bernardo O'Higgins Base (BO)	0.61772 **(0.00000)**	0.79060 **(0.00000)**	0.40132 **(0.00000)**	0.34057 **(0.00000)**	0.16943 **(0.00050)**	0.27133 **(0.00238)**	0.00117 (0.37323)	0.36809 **(0.00040)**	0.00000 (0.56757)	0.14866 (0.01505)	0.00000 (0.99990)	*	0.00000 (0.79764)	0.05537 (0.21948)
Gabriel González Videla Base (GGV)	0.58791 **(0.00000)**	0.75247 **(0.00000)**	0.37912 **(0.00000)**	0.32380 **(0.00000)**	0.15489 **(0.00010)**	0.25607 **(0.00248)**	0.00682 (0.32373)	0.35394 **(0.00020)**	0.00000 (0.86714)	0.13331 (0.01802)	0.00000 (0.81378)	0.00000 (0.99990)	*	0.01709 (0.18543)
Jougla Point (JP)	0.79404 **(0.00000)**	0.87888 **(0.00000)**	0.65499 **(0.00000)**	0.52270 **(0.00000)**	0.35294 **(0.00000)**	0.53893 **(0.00000)**	0.12200 **(0.00257)**	0.65751 **(0.00000)**	0.14763 (0.05821)	0.30787 **(0.00000)**	0.00000 (0.99990)	0.00582 (0.57668)	0.04903 (0.11989)	*

S2C. *TLR7*

Colony	CR	MAR	COU	MO	CB	BR	BI	MT	SIG	COP	SP	BO	GGV	JP
Crozet Island (CR)	*	0.00000 (0.99990)	0.15385 (0.10049)	0.03575 (0.25166)	0.00000 (0.99990)	0.15385 (0.2206)	0.24153 (0.1199)	0.00000 (0.99990)	0.21932 (0.02099)	0.00575 (0.41451)	0.27835 (0.05821)	0.07738 (0.08643)	0.05127 (0.48193)	0.00000 (0.62479)
Marion Island (MAR)	0.00000 (0.99990)	*	0.17066 (0.03406)	0.04320 (0.12395)	0.00000 (0.99990)	0.17065 (0.09524)	0.26718 (0.0988)	0.00000 (0.99990)	0.24170 (0.01584)	0.01208 (0.25423)	0.30218 (0.04445)	0.09237 (0.06316)	0.06381 (0.09049)	0.00013 (0.26186)
Courbet Peninsula (COU)	0.13462 (0.09613)	0.15073 (0.03673)	*	0.02638 (0.13038)	0.24508 **(0.00515)**	0.00000 (0.91486)	0.00000 (0.99990)	0.23966 (0.01079)	0.09619 (0.11098)	0.08448 (0.02208)	0.17212 (0.02287)	0.00000 (0.52817)	0.06001 (0.09534)	0.10587 (0.00871)
Pointe du Morne (MO)	0.06757 (0.16939)	0.07580 (0.09940)	0.00000 (0.58420)	*	0.07551 (0.02198)	0.00000 (0.59489)	0.00000 (0.60311)	0.06997 (0.03643)	0.13982 (0.02564)	0.02337 (0.11662)	0.18981 (0.00812)	0.00000 (0.50470)	0.02590 (0.20364)	0.03442 (0.04425)
Cow Bay (CB)	0.00000 (0.99990)	0.00000 (0.99990)	0.17804 (0.01307)	0.08759 (0.01653)	*	0.21440 (0.01950)	0.30347 (0.03059)	0.00000 (0.99990)	0.31917 **(0.00099)**	0.04066 (0.03465)	0.35766 (0.00941)	0.14082 (0.02594)	0.10513 (0.02277)	0.02675 (0.04277)
Bull Roads (BR)	0.15385 (0.21988)	0.17065 (0.08841)	0.00000 (0.99990)	0.00000 (0.86308)	0.18657 (0.0202)	*	0.00000 (0.99990)	0.23965 (0.03346)	0.11188 (0.0694)	0.03171 (0.1389)	0.19988 (0.01792)	0.00000 (0.80804)	0.03355 (0.23849)	0.05073 (0.10286)
Bird Island (BI)	0.24153 (0.11771)	0.26718 (0.10019)	0.00000 (0.99990)	0.00000 (0.61113)	0.27337 (0.03158)	0.00000 (0.99990)	*	0.36906 (0.05029)	0.08050 (0.21364)	0.03041 (0.25869)	0.18463 (0.07247)	0.00000 (0.99990)	0.02727 (0.34026)	0.05262 (0.19166)
Martillo Island (MT)	0.00000 (0.99990)	0.00000 (0.99990)	0.21679 (0.01238)	0.10671 (0.03554)	0.00000 (0.99990)	0.23965 (0.03772)	0.36906 (0.04960)	*	0.33218 **(0.00307)**	0.03301 (0.08039)	0.39735 (0.01673)	0.15325 (0.02525)	0.11382 (0.03812)	0.01883 (0.13583)

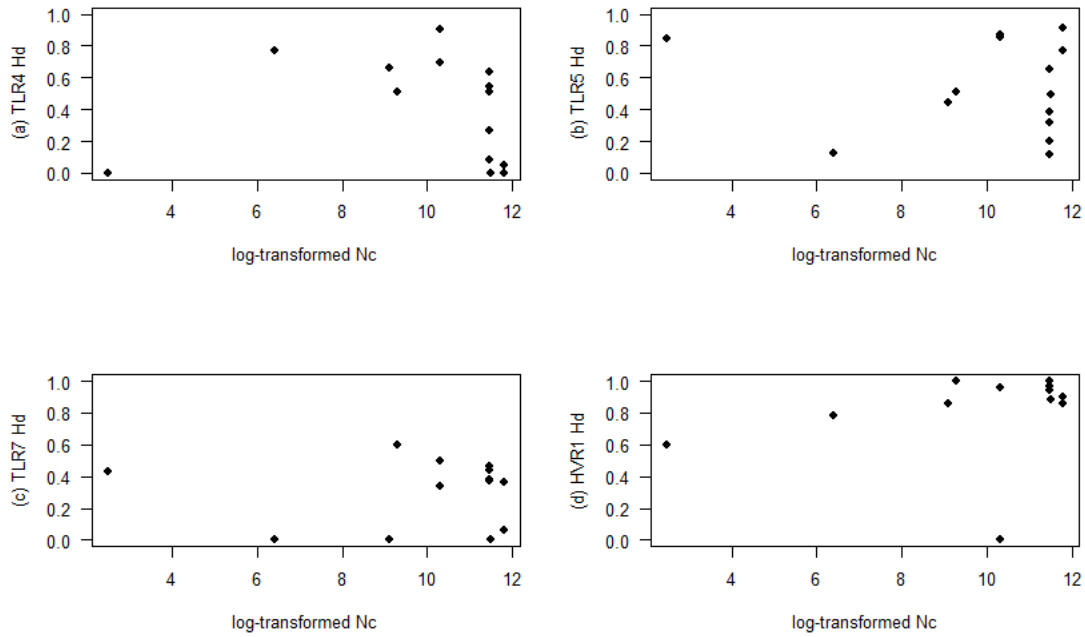
Signy Island (SIG)	0.30744 (0.02069)	0.33179 (0.01445)	0.02088 (0.38392)	0.09976 (0.05287)	0.38145 (0.00376)	0.06005 (0.21335)	0.01582 (0.49480)	0.42834 **(0.00267)**	*	0.03625 (0.20493)	0.00000 (0.99990)	0.00039 (0.44204)	0.00000 (0.68350)	0.05056 (0.16315)
Copacabana (Admiralty Bay), King George Island (COP)	0.05824 (0.12672)	0.06574 (0.10247)	0.00000 (0.51431)	0.00125 (0.34294)	0.07694 (0.00802)	0.00000 (0.47381)	0.00000 (0.51579)	0.09310 (0.02079)	0.03264 (0.16563)	*	0.05349 (0.14266)	0.00000 (0.83318)	0.00000 (0.86249)	0.00000 (0.57965)
Stranger Point, King George Island (SP)	0.27835 (0.05693)	0.30218 (0.04792)	0.03784 (0.22651)	0.09551 (0.04534)	0.33172 (0.00931)	0.08347 (0.14523)	0.05636 (0.26324)	0.39735 (0.01832)	0.00000 (0.99990)	0.01031 (0.37036)	*	0.06359 (0.19364)	0.00000 (0.60915)	0.05583 (0.14642)
Bernardo O'Higgins Base (BO)	0.14089 (0.08128)	0.15866 (0.07019)	0.00000 (0.99990)	0.00000 (0.76052)	0.17845 (0.02544)	0.00000 (0.99990)	0.00000 (0.99990)	0.23146 (0.02584)	0.00000 (0.57440)	0.00000 (0.82398)	0.00000 (0.43817)	*	0.00000 (0.83586)	0.00000 (0.48965)
Gabriel González Videla Base (GGV)	0.10256 (0.21968)	0.11742 (0.08841)	0.00000 (0.68023)	0.00000 (0.45857)	0.13061 (0.03425)	0.00000 (0.7821)	0.00000 (0.58202)	0.17796 (0.03861)	0.00000 (0.58103)	0.00000 (0.99990)	0.00000 (0.61053)	0.00000 (0.99990)	*	0.00000 (0.85031)
Jougla Point (JP)	0.05241 (0.11712)	0.05952 (0.09356)	0.00000 (0.39155)	0.00830 (0.22443)	0.06981 (0.01099)	0.00696 (0.32888)	0.00000 (0.38491)	0.08498 (0.01634)	0.03980 (0.14929)	0.00000 (0.76933)	0.01196 (0.36442)	0.00000 (0.70765)	0.00000 (0.93100)	*

S2D. HVR1

Colony	CR	MAR	COU	MO	CB	BR	BI	MT	SIG	COP	SP	JP
Crozet Island (CR)	*	0.00000	0.98071	0.97191	0.96572	0.97018	0.97094	0.96879	0.96497	0.9557	0.94835	0.96015
		(0.56638)	(0.03178)	**(0.00000)**	**(0.00000)**	**(0.00040)**	**(0.00000)**	**(0.00149)**	**(0.00149)**	**(0.00000)**	**(0.00010)**	**(0.00000)**
Marion Island (MAR)	0.00000	*	0.98950	0.97523	0.97019	0.97819	0.97591	0.97818	0.97557	0.96038	0.95916	0.96415
	(0.66974)		(0.02376)	**(0.00000)**	**(0.00000)**	**(0.00020)**	**(0.00000)**	**(0.00079)**	**(0.00069)**	**(0.00000)**	**(0.00040)**	**(0.00000)**
Courbet Peninsula (COU)	0.35052	0.39100	*	0.29453	0.94336	0.96564	0.96057	0.96480	0.95757	0.92969	0.90586	0.93788
	(0.11405)	(0.02455)		(0.01287)	(0.00683)	(0.02673)	**(0.00594)**	(0.04821)	(0.04831)	**(0.00475)**	(0.03534)	**(0.00297)**
Pointe du Morne (MO)	0.08579	0.11761	0.21195	*	0.92836	0.92743	0.92721	0.92737	0.91436	0.91865	0.90576	0.92253
	(0.00238)	**(0.00030)**	(0.02079)		**(0.00000)**	**(0.00000)**	**(0.00000)**	**(0.00000)**	**(0.00010)**	**(0.00000)**	**(0.00000)**	**(0.00000)**
Cow Bay (CB)	0.11875	0.15084	0.28772	0.07047	*	0.10968	0.7874	0.28031	0.74375	0.71836	0.70848	0.73538
	(0.00257)	**(0.00099)**	(0.02138)	**(0.00010)**		(0.03356)	**(0.00000)**	**(0.00208)**	**(0.00000)**	**(0.00000)**	**(0.00000)**	**(0.00000)**
Bull Roads (BR)	0.14286	0.17963	0.35052	0.08579	0.08636	*	0.84043	0.28029	0.82431	0.75894	0.74524	0.77946
	(0.01188)	**(0.00564)**	(0.10781)	**(0.00248)**	(0.03257)		**(0.00000)**	(0.03346)	**(0.00158)**	**(0.00000)**	**(0.00069**)	**(0.00000)**
Bird Island (BI)	0.12855	0.16014	0.29812	0.07930	0.10888	0.12855	*	0.84374	0.32908	0.24287	0.25257	0.22048
	(0.00653)	**(0.00059)**	(0.01228)	**(0.00000)**	**(0.00010)**	(0.00624)		**(0.00000)**	**(0.00030)**	**(0.00000)**	**(0.00020)**	**(0.00000)**
Martillo Island (MT)	0.25878	0.29433	0.55224	0.17732	0.21433	0.10862	0.22363	*	0.80733	0.74466	0.70428	0.76889
	(0.00465)	**(0.00218)**	(0.10375)	**(0.00040)**	**(0.00238)**	(0.12751)	**(0.00168)**		(0.00762)	**(0.00000)**	**(0.00287)**	**(0.00000)**
Signy Island (SIG)	0.07667	0.11833	0.28571	0.02460	0.05851	0.07667	0.06939	0.20000	*	0.01749	0.03152	0.01385
	(0.10969)	(0.05465)	(0.04663)	(0.19840)	(0.11118)	(0.10920)	(0.07970)	(0.07801)		(0.29324)	(0.26304)	(0.29621)
Copacabana (Admiralty Bay), King George Island (COP)	0.08074	0.11247	0.23804	0.03770	0.06588	0.08074	0.07469	0.17159	0.01948	*	0.00000	0.01858
	(0.00208)	**(0.00010)**	(0.01861)	**(0.00000)**	**(0.00020)**	**(0.00257)**	**(0.00000)**	**(0.00030)**	(0.22285)		(0.53480)	(0.15929)
Stranger Point, King George Island (SP)	0.07363	0.11346	0.26829	0.02384	0.05657	0.07363	0.06709	0.18978	0.00000	0.01097	*	0.06616
	(0.10019)	(0.04574)	(0.03198)	(0.16553)	(0.07574)	(0.10217)	(0.05653)	**(0.00119)**	(0.99990)	(0.31829)		(0.08742)
Jougla Point (JP)	0.09405	0.12503	0.25198	0.05014	0.07823	0.09405	0.06869	0.18393	0.01765	0.02247	0.03323	*
	(0.00475)	**(0.00069)**	(0.01376)	**(0.00030)**	**(0.00010)**	**(0.00465)**	**(0.00050)**	**(0.00149)**	(0.26225)	(0.04208)	(0.14088)	

Supplementary Table S3. Census population size (number of breeding pairs) applied to analyses related to each colony, based on best available census data for each island group or region, with census year and reference information.

Colony	Census Region	Census Population (breeding pairs)	Year	Reference / Notes
CR	Crozet Archipelago	9,000	1994	Jouventin, 1994.
MAR	Marion Island	600	2012/13	Crawford, et al. 2014.
COU	Kerguelen Islands	30,000	1988	Weimerskirch, et al. 1988. (*Lescroel, and Bost, 2006. show significant decrease in 2002 of Kerguelen population, but was localized to the sampled portion of the archipelago, leading us to use 30,000 as the lower range number for Weimerskirch, et al. 1998's estimate of 30-40,000)
MO				
CB	Falkland/ Malvinas Islands	132,321	2010	Baylis, et al. 2013.
BR				
BI	South Georgia	98,867	1996	Trathan, et al. 1996.
MT	Martillo Island	12	2006/07	Ghys, et al. 2008.
SIG	South Orkney Islands	10,760	2013	Lynch 2013.
COP	South Shetland Islands and Western Antarctic Peninsula	94,751	2013	Lynch 2013.
SP				
BO				
GGV				
JP				



Supplementary Figure S1. Haplotype diversity (H_d) for (a) *TLR4*, (b), *TLR5*, (c) *TLR7*, and (d) mtDNA HVR1 in relation to census population size across Gentoo Penguin colonies. Census population sizes and their source publications used for analyses are found in Supplementary Table S3.

Supplementary Table S4. Spearman's rank correlations and *p* values calculated for each haplotype diversity Hd vs. census population size comparison depicted in Supplementary Figure S1.

Locus	Spearman's rank correlation	*p*
TLR4	-0.5029245	0.06679
TLR5	0.09462364	0.7476
TLR7	-0.1380377	0.6379
HVR1	N/A	0.3263

Supplementary Table S5A. Mantel Test results for Isolation by Distance for each locus, using F_{ST} (Weir and Cockerham 1984) values against a matrix of geographic distance (km), with significance evaluated using 10,000 permutations in Arlequin.

Mantel Test Locus	Correlation Coefficient (R)	*P* value
TLR4	0.514998	0.0011
TLR5	0.592887	0.0010
TLR7	0.026662	0.3677
HVR1	0.311577	0.0153

Supplementary Table S5B. Pairwise geographic distances (most direct distance by sea during summer sea ice extent, in km) used for Mantel's test, derived from Google Earth v7.3.2.5776.

Colony	CR	MAR	COU	MO	CB	BR	BI	MT	SIG	COP	SP	BO	GGV	JP
Crozet Island (CR)	0													
Marion Island (MAR)	960	0												
Courbet Peninsula (COU)	1577	2461	0											
Pointe du Morne (MO)	1565	2449	14	0										
Cow Bay (CB)	7190	6481	7850	7837	0									
Bull Roads (BR)	7185	6504	7833	7819	168	0								
Bird Island (BI)	5958	5204	6786	6773	1359	1432	0							
Martillo Island (MT)	7354	6743	7915	7903	792	639	1892	0						
Signy Island (SIG)	5955	5315	6565	6555	1282	1258	901	1437	0					
Copacabana (Admiralty Bay), King George Island (COP)	6419	5848	6978	6970	1275	1173	1501	1010	705	0				
Stranger Point, King George Island (SP)	6421	5849	6979	6971	1262	1161	1506	995	711	13	0			
Bernardo O'Higgins Base (BO)	6402	5804	6936	6928	1351	1248	1536	1085	708	130	125	0		
Gabriel González Videla Base (GGV)	6705	6107	7237	7229	1522	1414	1837	1147	1009	372	360	302	0	
Jougla Point (JP)	6719	6138	7264	7256	1543	1435	1861	1170	1035	400	391	328	45	0

Supplementary Table S6A. Log-likelihood comparisons for PAML (Yang 1997; Yang 2007) analysis of *TLR4*.

Model	lnL	$2(\ln L_1 - \ln L_0)$	df	*p-value*
M1a	-3605.474038	8.33452799999986	2	0.015494595
M2a	-3597.13951			
M7	-3605.717024	8.33604300000025	2	0.015482863
M8	-3597.380981			
M8a	-3605.474029	8.09304800000018	2	0.01748304
M8	-3597.380981			

Supplementary Table S6B. Bayes Empirical Bayes (Yang, et al. 2005) analysis of *P. papua TLR4* sites under positive selection with posterior probability >0.90. Sites obtained from *codeml* M2a.

Amino acid pos.	Amino acid	Pr(ω>1)	Mean ω ± SE
12	V	0.932	7.669 ± 2.761
236	I	0.932	7.669 ± 2.761
659	A	0.933	7.671 ± 2.759

Supplementary Table S6C. Bayes Empirical Bayes analysis of *P. papua TLR4* sites under positive selection with posterior probability >0.90. Sites obtained from *codeml* M8.

Amino acid pos.	Amino acid	Pr(ω>1)	Mean ω ± SE
12	V	0.947	1.461 ± 0.163
82	N	0.936	1.485 ± 0.000
236	I	0.947	1.461 ± 0.163
316	C	0.936	1.485 ± 0.000
445	Q	0.936	1.485 ± 0.000
659	A	0.947	1.461 ± 0.163

Supplementary Table S6D. Log-likelihood comparisons for PAML analysis of *TLR5*.

Model	lnL	$2(\ln L_1 - \ln L_0)$	df	*p-value*
M1a	-3929.201728	102.078912	2	6.821e-23
M2a	-3878.162272			
M7	-3929.323036	102.311604	2	6.072e-23
M8	-3878.167234			
M8a	-3928.465238	100.596008	2	7.16e-23
M8	-3878.167234			

Supplementary Table S6E. Bayes Empirical Bayes analysis of *P. papua TLR5* sites under positive selection with posterior probability >0.90. Sites obtained from *codeml* M2a.

Amino acid pos.	Amino acid	Pr(ω>1)	Mean ω ± SE
10	Y	1.000	9.645 ± 0.504
285	M	0.980	9.472 ± 1.320
442	A	0.998	9.631 ± 0.609
535	S	0.981	9.482 ± 1.286
667	S	1.000	9.643 ± 0.514
698	M	0.980	9.469 ± 1.329
747	L	0.985	9.518 ± 1.163
788	C	0.982	9.492 ± 1.255
845	V	1.000	9.645 ± 0.503

Supplementary Table S6F. Bayes Empirical Bayes analysis of *P. papua TLR5* sites under positive selection with posterior probability >0.90. Sites obtained from *codeml* M8a.

Amino acid pos.	Amino acid	Pr(ω>1)	Mean ω ± SE
10	Y	1.000	8.947 ± 1.200
285	M	0.989	8.855 ± 1.487
442	A	0.999	8.941 ± 1.219
535	S	0.989	8.860 ± 1.472
667	S	1.000	8.946 ± 1.202
698	M	0.988	8.853 ± 1.490
747	L	0.991	8.878 ± 1.421
788	C	0.990	8.865 ± 1.459
845	V	1.000	8.947 ± 1.200

Supplementary Table S7. Physiochemical distances between amino acid variants at positively selected sites in TLR4 and TLR5, as given by several authors, detailed in the lower table. Red colors indicate a greater physiochemical distance, while green colors indicate a lower physiochemical distance. *Epstein's distance has an element of directionality, so both directions are given: Epstein_1 = variant_1 → variant_2; Epstein_2 = variant_2 → variant_1.

TLR	nt_site	aa_site	variant_1	variant_2	Grantham [1]	Sneath [2]	Epstein_1 [3]	Epstein_2 [3]	Miyata [4]	Urbina [5]
TLR4	35	12	V	A	64	12	0.4	0.41	1.85	3.125
TLR4	706	236	V	I	29	7	0.03	0.05	0.85	0.873
TLR4	1977	659	A	T	58	20	0.41	0.4	0.9	1.795
TLR5	29	10	C	Y	194	34	0.25	0.13	2.38	3.591
TLR5	854	285	T	M	81	25	0.8	0.8	1.86	2.404
TLR5	1325	442	A	V	64	12	0.41	0.4	1.85	3.125
TLR5	1604	535	S	T	58	12	0.03	0.03	0.89	1.697
TLR5	2000	667	F	S	155	25	0.81	0.81	3.45	4.544
TLR5	2094	698	I	M	10	22	0.05	0.03	0.29	1.565
TLR5	2239	747	V	L	32	9	0.05	0.03	0.91	0.831
TLR5	2362	788	S	C	112	13	0.6	0.6	2.84	2.758
TLR5	2533	845	V	I	29	7	0.05	0.03	0.85	0.873

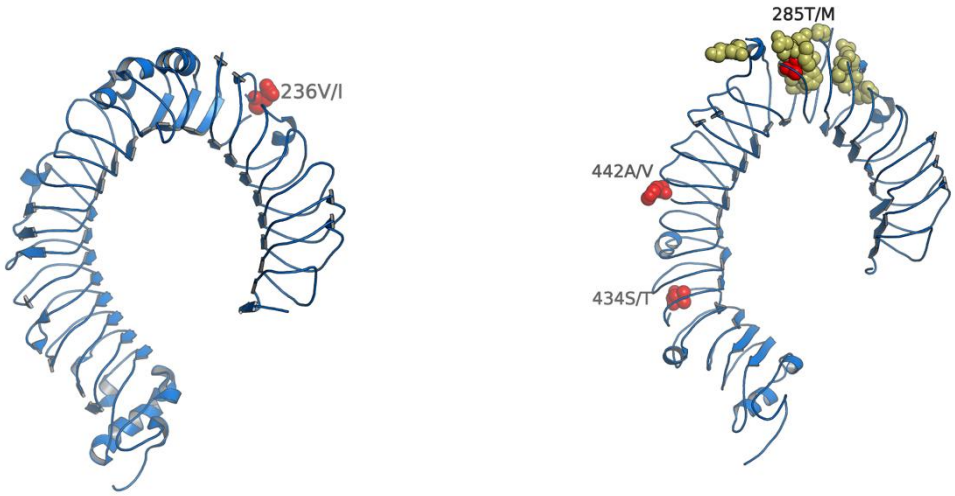
Number	Criteria	Range	Reference
[1]	composition, polarity, molecular volume	5(low)-215(high)	Grantham (1974)
[2]	134 categories of activity and structure	5(low)-45(high)	Sneath (1966)
[3]*	polarity, size	0(low)-1.02(high)	Epstein (1967)
[4]	volume, polarity	0.06(low)-5.13(high)	Miyata (1979)
[5]	volume, bulkiness, polarity, isoelectric point, hydrophobicity, surface area accessible to water, fraction of accessible area lost when protein folds	1.031(low)-7.08(high)	Urbina, et al. (2006)

Supplementary Table S8. Positions of Gentoo Penguin positively selected sites with respect to sites of known function in other vertebrates. The position and name of the corresponding chicken amino acid is given for each Gentoo TLR position, as chicken TLRs are typically the best-studied and best-annotated TLRs. "other_literature" denotes a study in which the described sites have also been found to be positively selected in a particular taxon (references are given below). "function" denotes the potential for change in function, due to proximity to sites of known function. "N/A" indicates that the Gentoo TLR5 open reading frame begins upstream of other known TLRs, so no functional information is known.

TLR	nt_site	aa_site	variant_1	variant_2	chicken_posn	chicken_aa	other_literature	function
TLR4	35	12	V	A	6	A	Velová, et al. 2018	No
TLR4	706	236	V	I	230	I	No	No
TLR4	1977	659	A	T	653	A	No	No
TLR5	29	10	C	Y	N/A	N/A	N/A	N/A
TLR5	854	285	T	M	267	M	No	Adjacent to two residues in flagellin-binding interface B (266,268) Yoon, et al. 2012; Song, et al. 2017
TLR5	1325	442	A	V	424	A	Velová, et al. 2018; Grueber, et al. 2014	No
TLR5	1604	535	S	T	517	S	No	No
TLR5	2000	667	F	S	649	S	No	No
TLR5	2094	698	I	M	680	L	Velová, et al. 2018	No
TLR5	2239	747	V	L	729	L	No	No
TLR5	2362	788	S	C	770	C	No	No
TLR5	2533	845	V	I	827	I	No	No

TLR4

TLR5



Supplementary Figure S2. Positions of positively selected sites in Gentoo Penguin TLR4 and TLR5 ectodomains. Positively selected sites from PAML analysis (highlighted in red) are superimposed on homologous sites from the crystal structure of mouse TLR4 ectodomain (PDB: 2Z64) and a homology model of human TLR5 ectodomain (PDB: 3J0A). Further positively selected sites in the signal peptide of TLR4 (12V/A) and TLR5 (10C/Y), as well as TMD/TIR domain sites (TLR4: 659A/T; TLR5: 667F/S, 698I/M, 747V/L, 788S/C, 845V/I) were omitted because the residues are not included in the available model structure, or were of low confidence. Relative positions of these residues are depicted schematically in main text Figure 6. Site 285T/M is located within interface-B of the flagellin binding site (Yoon, et al. 2012; Song, et al. 2017), which is highlighted in yellow. No other positively selected sites were found to be proximal to sites of known function in other vertebrates.

Supplementary Table S9. *In silico* prediction of change of function of positively selected sites using the homology-based methods SIFT (Ng and Henikoff 2003) and PolyPhen-2 (Adzhubei, et al. 2013). Both SIFT and PolyPhen-2 are directional, so the scores for each variant 1 → variant 2 and variant 2 → variant 1 are given. Red colours indicate the change is likely to be of high consequence, while green colours indicate milder consequences.

TLR	nt_site	aa_site	variant_1	variant_2	SIFT				Polyphen2					
					v1→v2		v2→v1		v1→v2			v2→v1		
					score	MSC	score	MSC	score	sensitivity	specificity	score	sensitivity	specificity
TLR4	35	12	V	A	0.15	3.34	0.58	3.38	N/A	N/A	N/A	0.997	0.410	0.980
TLR4	706	236	V	I	1.00	3.20	0.18	3.21	0.001	0.990	0.150	0.049	0.940	0.830
TLR4	1977	659	A	T	0.20	3.04	0.25	3.05	0.810	0.840	0.930	0.000	1.000	0.000
TLR5	29	10	C	Y	0.00	4.32	0.00	4.32	0.004	0.970	0.590	N/A	N/A	N/A
TLR5	854	285	T	M	0.43	3.36	0.10	3.36	0.000	1.000	0.000	0.990	0.720	0.970
TLR5	1325	442	A	V	0.51	3.55	0.72	3.36	0.165	0.920	0.870	0.000	1.000	0.000
TLR5	1604	535	S	T	0.38	3.36	0.82	3.36	0.253	0.910	0.880	0.000	1.000	0.000
TLR5	2000	667	F	S	0.34	3.36	0.01	3.36	0.000	1.000	0.000	0.998	0.270	0.990
TLR5	2094	698	I	M	0.06	3.36	0.45	3.36	0.009	0.960	0.770	0.005	0.970	0.740
TLR5	2239	747	V	L	1.00	3.36	0.00	3.36	0.000	1.000	0.000	0.000	1.000	0.000
TLR5	2362	788	S	C	1.00	3.36	0.00	3.36	0.000	1.000	0.000	0.992	0.700	0.970
TLR5	2533	845	V	I	0.45	3.36	1.00	3.36	0.024	0.950	0.810	0.000	1.000	0.000

Supplementary Figure S3. Alignment of the transmembrane region and surrounding sequence of available avian TLR5 amino acid sequences. The conserved serine residue is highlighted in pink and any variant amino acids are highlighted in red.

Supplementary Table S10. Diversity statistics of Gentoo Penguin (*Pygoscelis papua*) vs. avian Toll-like receptors from published studies. Chicken diversity data was drawn directly from Świderská, et al. (2018), while grey partridge study sequences were obtained from GenBank along with supplementary frequency data for analysis in DnaSP 6.12.10 as in main text methods to obtain comparable diversity measures.

Gene	Species	Length (bp)	*k*	H_d ± SD	π ± SD	UPV	A
TLR4	Gentoo Penguin	2550	1.550	0.478 ±0.035	0.00061 ± 0.00005	9	21
	Domestic chicken	2532	6.540	0.947±0.010	0.00258±0.00005	24	70
	Grey partridge	2529	3.063	0.863 ±0.063	0.00121 ± 0.00018	5	10
TLR5	Gentoo Penguin	2643	3.919	0.781 ±0.022	0.00148 ± 0.00006	32	46
	Domestic chicken	2586	2.866	0.854±0.013	0.00111±0.00004	10	22
	Grey partridge	2589-2592	1.053	0.747 ± 0.076	0.00041 ± 0.00008	4	6
TLR7	Gentoo Penguin	3126	0.327	0.296 ±0.034	0.00010 ± 0.00001	8	10
	Domestic chicken	3180	3.363	0.849±0.017	0.00106±0.00005	13	26
	Grey partridge	3141	1.753	0.532 ± 0.100	0.00056 ± 0.00013	3	3

k = average number of nucleotide differences between alleles; Hd = haplotype diversity; SD: standard deviation; π: nucleotide diversity (average number of nucleotide differences per site between two sequences); UPV: number of unique protein variants; A: number of alleles.

Notes: To compare diversity levels in our study population with those of other species, we sought datasets that sequenced *TLR4*, *TLR5*, and *TLR7* in birds. Because diversity measures can be biased by gene length and whether sequenced fragments are under selection, we restricted our comparison only to studies that sequenced the full length of these genes: a study in 110 domestic chickens (*Gallus gallus domesticus*, Świderská, et al. 2018) and a study of 10 European grey partridges (*Perdix perdix*, Vinkler, et al. 2015). The difference in gene length among these species is minimal enough to allow comparison (*TLR4*: <0.01%, *TLR5*: 2.15%, *TLR7*: 1.69%). However, sampling strategy differs significantly between studies, and also affects diversity measures (i.e. chicken study sampled few individuals from 25 distinct farmed breeds, partridge study sampled a small geographic subset of the species, while Gentoo Penguin study had varying numbers of samples from multiple wild populations). Given the differences in sampling strategy, only relative statements comparing TLRs within a species/study are appropriate, and care should be given to drawing conclusions from interspecific comparisons.

Supplementary Table S11. Sampling sites, location details, and number of samples obtained from Gentoo Penguins for immunogenetic analysis.

Sampling Site Name	Region	Latitude	Longitude	Sample size			
				mtDNA	*TLR4*	*TLR5*	*TLR7*
Crozet Island (CR)	Crozet Is.	46.4298 S	50.4046 E	7	7	7	7
Marion Island (MAR)	Prince Edward Is.	46.8872 S	37.8671 E	8	8	8	8
Courbet Peninsula (COU)	Kerguelen Is.	49.2811 S	70.5410 E	2	7	7	7
Pointe du Morne (MO)	Kerguelen Is.	49.3805 S	70.4423 E	20	15	20	18
Cow Bay (CB)	Falkland/Malvinas Is.	51.4343 S	57.8506 W	16	19	18	17
Bull Roads (BR)	Falkland/Malvinas Is.	52.3380 S	59.3633 W	7	7	7	7
Bird Island (BI)	South Georgia	54.0094 S	38.0631 W	17	18	12	13
Martillo Island (MT)	Tierra del Fuego	54.9083 S	67.3833 W	5	5	5	4
Signy Island (SIG)	South Orkney Is.	60.6747 S	45.6278 W	5	5	5	5
Copacabana (Admiralty Bay), King George Island (COP)	South Shetland Is.	62.1792 S	58.4472 W	21	22	21	21
Stranger Point, King George Island (SP)	South Shetland Is.	62.2375 S	58.5939 W	6	5	5	5
Bernardo O'Higgins Base (BO)	Western Antarctic Peninsula	63.3207 S	57.8987 W	0	6	6	6
Gabriel González Videla Base (GGV)	Western Antarctic Peninsula	64.8238 S	62.8581 W	0	7	7	7
Jougla Point (JP)	Western Antarctic Peninsula	64.8270 S	63.4932 W	24	24	24	23
			Total	138	155	152	148

Supplementary Table S12. List of primers used for PCR amplification and sequencing of Gentoo Penguin TLR genes, along with reaction conditions.

Gene	Description[1] and (Primer Names)	Primer Sequences	Ta (°C)	Product Size	Polymerase / reaction volume / PCR program
TLR4	Exon 1 (PengTLR4_Ex1_F2 + PengTLR4_Ex1_R2_long)	F: CTCACTCCAGGTCCTTGAGTG R: ACTGCAGTACCCTCAGAGGT	55	419 bp	Qiagen *Taq* PCR Master Mix Kit (12 µL) 94° C / 3:00 40 cycles: 94° C / 0:45 55° C / 0:45 72° C / 1:00 72° C / 10:00
TLR4	Exon 2 (PengTLR4_Ex2_F2 + (PengTLR4_Ex2_R1)	F: TGCTGCTTTCCTGACAGTGT R: AGCCAAGACCAACAGACACC	55	358 bp	Qiagen *Taq* PCR Master Mix Kit (12 µL) 94° C / 3:00 40 cycles: 94° C / 0:45 55° C / 0:45 72° C / 1:00 72° C / 10:00
TLR4	Exon 3 (Start) (PengTLR4_Ex3_Start_F + PengTLR4_Ex3_Start_R2)	F: AAGGGACAGTGTTGCATGCA R: CACTCAAATCTACAACCTCCAGG	55	945 bp	Qiagen *Taq* PCR Master Mix Kit (12 µL) 94° C / 3:00 40 cycles: 94° C / 0:45 55° C / 1:00 72° C / 1:00 72° C / 10:00
TLR4	Exon 3 (Middle) (PengTLR4_Ex3_Mid_F1 + PengTLR4_Ex3_MidR2)	F: GGCAACGTCTCTAGTATTCGGT R: TGCAGGTGTGTGGCATATCA	55	932 bp	Qiagen *Taq* PCR Master Mix Kit (12 µL) 94° C / 3:00 40 cycles: 94° C / 0:45 55° C / 1:00 72° C / 1:00 72° C / 10:00
TLR4	Exon 3 (End)[2] (PengTLR4_Ex3_Mid_F1 + PengTLR4_Ex3_EndR2 or PengTLR4_Ex3_EndR)	F: GGCAACGTCTCTAGTATTCGGT R: TGCCTTCTAGCAGGACTCCT or R: ACTTCTGAGCACCAGCCATC	50	1563 bp or 1671 bp	Qiagen *Taq* PCR Master Mix Kit (12 µL) 94° C / 3:00 40 cycles:

					94° C / 0:45 50° C / 1:00 72° C / 2:00 72° C / 10:00
TLR5	TLR5 (A) (TLR5_full_new_F + TLR5_1000_R)	F: AGTGTGACATTAGTCTTTTGTTCTACA R: AGCCAAATCCCGAACCCATT	64	1021 bp	NEB Phusion Hot Start Flex DNA Polymerase (25 µL) 98° C / 3:00 40 cycles: 98° C / 0:30 64° C / 0:30 72° C / 1:00 72° C / 10:00
TLR5	TLR5 (B) (Gentoo_TLR5_sequencing_2_F + Gentoo_TLR5_full_R)	F: ACCTTAGTGACAACGGCTGG R: GTGTTCATTCCTGCCATGGC	68	1948 bp	NEB Phusion Hot Start Flex DNA Polymerase (25 µL) 98° C / 3:00 40 cycles: 98° C / 0:30 68° C / 0:30 72° C / 2:00 72° C / 10:00
TLR5	TLR5 (C) (Gentoo_TLR5_sequencing_3_F + Gentoo_TLR5_full_R)	F: TCCCCTGAGCCTGAAGTCTT R: GTGTTCATTCCTGCCATGGC	69	962 bp	NEB Phusion Hot Start Flex DNA Polymerase (25 µL) 98° C / 3:00 40 cycles: 98° C / 0:30 69° C / 0:30 72° C / 1:00 72° C / 10:00
TLR7	TLR7 Coding Exon (A) (TLR7_StartF_212 + TLR7_AR_882)	F: AGCCAGTGTGAGAGACAGGA R: ACGGTAGTAACAGTTCTGTC	62	670 bp	NEB Phusion Hot Start Flex DNA Polymerase (25 µL) 98° C / 3:00 40 cycles: 98° C / 0:30 62° C / 0:30 72° C / 0:30 72° C / 10:00
TLR7	TLR7 Coding Exon (B) (TLR7_BF_832 + TLR7_BR_1733)	F: GAGCTAGGAAACATAGAAAT R: TCATCATACCTGAAATAATG	55	901 bp	NEB Phusion Hot Start Flex DNA Polymerase (25 µL) 98° C / 3:00 40 cycles: 98° C / 0:30 55° C / 0:30

TLR7[3]	TLR7 Coding Exon (C) (TLR7_midF_1512 + TLR7_midR_2741)	F: AAATCTAACAGTGTTGGATC R: GTCACTTGAGTCTGATTGAT	50	1229 bp	NEB Phusion Hot Start Flex DNA Polymerase (25 µL) 98° C / 3:00 40 cycles: 98° C / 0:30 50° C / 0:30 72° C / 0:30 72° C / 10:00
TLR7	TLR7 Coding Exon (D) (TLR7_DF_2613 + TLR7_End_DR_3597)	F: AATTAAGAAATCTAGCTTCC R: TTTAGGGTACAATATCACAA	54	984 bp	NEB Phusion Hot Start Flex DNA Polymerase (25 µL) 98° C / 3:00 40 cycles: 98° C / 0:30 54° C / 0:45 72° C / 0:30 72° C / 10:00

[1] Longer exons are broken up into fragments due to the practical limitations of Sanger sequencing. Fragments are listed sequentially in the table, with sufficient overlaps to allow for sequence assembly. The descriptors Start/Middle/End or A/B/C/D also reflect this ordering of fragments of the gene.

[2] Certain samples yielded cleaner PCR amplicons and sequencing results with either of these two Reverse primers, so both are listed. PCRs were conducted under the same conditions for both combinations.

[3] This fragment was GC-rich and required Phusion GC buffer (NEB) rather than HF buffer for successful PCR amplification. The presence of a secondary structure here required repeat bi-directional sequencing for most samples.

Appendix A3 – Supplementary Data for Chapter 5

Supplementary Table S8. Sampling site details and coordinates for four penguin species, along with cloacal swabs obtained in terms of total numbers of individuals, and breakdown by adults and chicks.

Colony and Antarctic Site Inventory Subarea	**Coordinates**	**Penguin Species (Individuals Sampled)**	**Adults Sampled**	**Chicks Sampled**
St. Andrew's Bay (STA), South Georgia	54°26'S, 36°11'W	King (20)	10	10
Yankee Harbor (YANK), Greenwich Island, South Shetland Islands	62˚32'S, 59˚47'W	Gentoo (9)	6	3
Half Moon Island (HALF), South Shetland Islands	62˚36'S, 59˚55'W	Chinstrap (10)	10	-
Baily Head (BAIL), Deception Island, South Shetland Islands	62°58'S, 60°30'W	Chinstrap (10)	5	5
Kinnes Cove (KINN, Madder Cliff), Joinville Island, Northeast Antarctic Peninsula	63°18'S, 56°29'W	Adélie (14)	7	7
Georges Point (GEOR), Ronge Island, Central Western Antarctic Peninsula	64°40'S, 62°40'W	Chinstrap (6) Gentoo (6)	3 3	3 3
Port Charcot (BOOT), Booth Island, Central Western Antarctic Peninsula	65°05'S, 64°00'W	Adélie (3) Chinstrap (6) Gentoo (6)	2 3 3	1 3 3
Moot Point (MOOT), Central Western Antarctic Peninsula	65°12'S, 64°06'W	Gentoo (5)	4	1
Total by Species		**Adélie (17)** **Chinstrap (32)** **Gentoo (26)** **King (20)**	**9** **21** **16** **10**	**8** **11** **10** **10**

Supplementary Table S2: Summary of AntV and AntCM sequences identified in this study together with the sampling date, location, host and the primer pairs used to recover the sequences.

Accession	Virus / circular molecule	Genotype	Isolate	Collection date	Location	Host	
MT196247	AntV1	I	CPGEORsw001Ad	6-Jan-16	Georges Point, West Antarctic peninsula	*Pygoscelis antarcticus*	F: 5'-CATAAATCATTTTGTGTTACGATCAACAACTACAC-3' R: 5'-ACTCATCTTGAAAGAAAAGTGGTGGTCTATATTAC-3'
MT196248	AntV1	I	CPGEORsw002Ad	6-Jan-16	Georges Point, West Antarctic peninsula	*Pygoscelis antarcticus*	F: 5'-CATAAATCATTTTGTGTTACGATCAACAACTACAC-3'; R: 5'-ACTCATCTTGAAAGAAAAGTGGTGGTCTATATTAC-3'
MT196249	AntV1	II	CPGEORsw003Ad	6-Jan-16	Georges Point, West Antarctic peninsula	*Pygoscelis antarcticus*	F: 5'-CATAAATCATTTTGTGTTACGATCAACAACTACAC-3'; R: 5'-ACTCATCTTGAAAGAAAAGTGGTGGTCTATATTAC-3'
MT196250	AntV1	II	CPGEORsw003Ch	6-Jan-16	Georges Point, West Antarctic peninsula	*Pygoscelis antarcticus*	F: 5'-CATAAATCATTTTGTGTTACGATCAACAACTACAC-3' R: 5'-ACTCATCTTGAAAGAAAAGTGGTGGTCTATATTAC-3'
MT196289	AntV2	I	GPYANKsw001Ad	17-Jan-16	Booth Island, West Antarctic peninsula	*Pygoscelis papua*	F: 5'-GCTAATAATTCAAATAATGGTTTACCGTCG-3' R: 5'-AGTTATATTTCACAAAACCCCCTCTTATAT-3'
MT196290	AntV2	I	GPYANKsw001Ch	17-Jan-16	Booth Island, West Antarctic peninsula	*Pygoscelis papua*	F: 5'-GCTAATAATTCAAATAATGGTTTACCGTCG-3' R: 5'-AGTTATATTTCACAAAACCCCCTCTTATAT-3'
MT196291	AntV2	I	GPYANKsw002Ad	17-Jan-16	Booth Island, West Antarctic peninsula	*Pygoscelis papua*	F: 5'-GCTAATAATTCAAATAATGGTTTACCGTCG-3' \R: 5'-AGTTATATTTCACAAAACCCCCTCTTATAT-3'
MT196292	AntV2	I	GPYANKsw003Ad	17-Jan-16	Booth Island, West Antarctic peninsula	*Pygoscelis papua*	F: 5'-GCTAATAATTCAAATAATGGTTTACCGTCG-3' R: 5'-AGTTATATTTCACAAAACCCCCTCTTATAT-3'
MT196293	AntV2	I	GPYANKsw004Ad	17-Jan-16	Booth Island, West Antarctic peninsula	*Pygoscelis papua*	F: 5'-GCTAATAATTCAAATAATGGTTTACCGTCG-3' R: 5'-AGTTATATTTCACAAAACCCCCTCTTATAT-3'
MT196252	AntV3	I	CPHALFsw004Ad	3-Jan-16	Half Moon Island, SouthShetland	*Pygoscelis antarcticus*	F: 5'-AGATCATACTTAAGAGTCTTAACAATCCTATACGC-3' R: 5'-TAGTCCCAATGAGATTATTCAAAAGAAGTTTTTCT-3'
MT196261	AntV3	II	CPHALFsw005Ad	3-Jan-16	Half Moon Island, SouthShetland	*Pygoscelis antarcticus*	F: 5'-TTCATCACCCTTAGTCTCTTCATCTAAGTTGTTTA-3' R: 5'-TAAATTACATTTTGACTTAAATAAGTGTTGGACGT-3'
MT196253	AntV3	II	CPHALFsw005Ad	3-Jan-16	Half Moon Island, SouthShetland	*Pygoscelis antarcticus*	F: 5'-AGATCATACTTAAGAGTCTTAACAATCCTATACGC-3' R: 5'-TAGTCCCAATGAGATTATTCAAAAGAAGTTTTTCT-3'
MT196294	AntV3	III	KPSTAsw010Ad	22-Dec-15	St. Andrews, South Georgia	*Aptenodytes patagonicus*	F: 5'-CTTTAGTCTCTTCATCTAAGTTGTTTATCACATCG-3' R: 5'-GCGATGAATAAACTACATGTTTTGCATTTAAATAA-3'
MT196295	AntV3	III	KPSTAsw014Ch	22-Dec-15	St. Andrews, South Georgia	*Aptenodytes patagonicus*	F: 5'-CTTTAGTCTCTTCATCTAAGTTGTTTATCACATCG-3' R: 5'-GCGATGAATAAACTACATGTTTTGCATTTAAATAA-3'
MT196297	AntV3	III	KPSTAsw016Ch	22-Dec-15	St. Andrews, South Georgia	*Aptenodytes patagonicus*	F: 5'-CTTTAGTCTCTTCATCTAAGTTGTTTATCACATCG-3' R: 5'-GCGATGAATAAACTACATGTTTTGCATTTAAATAA-3'
MT196298	AntV3	III	KPSTAsw018Ch	22-Dec-15	St. Andrews, South Georgia	*Aptenodytes patagonicus*	F: 5'-CTTTAGTCTCTTCATCTAAGTTGTTTATCACATCG-3' R: 5'-GCGATGAATAAACTACATGTTTTGCATTTAAATAA-3'
MT196299	AntV3	III	KPSTAsw019Ch	22-Dec-15	St. Andrews, South Georgia	*Aptenodytes patagonicus*	F: 5'-CTTTAGTCTCTTCATCTAAGTTGTTTATCACATCG-3' R: 5'-GCGATGAATAAACTACATGTTTTGCATTTAAATAA-3'

MT196296	AntV3	IV	KPSTAsw015Ch	22-Dec-15	St. Andrews, South Georgia	*Aptenodytes patagonicus*	F: 5'-CTTTAGTCTCTTCATCTAAGTTGTTTATCACATCG-3' R: 5'-GCGATGAATAAACTACATGTTTTGCATTTAAATAA-3'
MT196279	AntV3	V	CPHALFsw003Ad	3-Jan-16	Half Moon Island, SouthShetland	*Pygoscelis antarcticus*	F: 5'-AACTACATGTTTTACATTTAAATAAGTGTTGGACG-3' R: 5'-TATTCATCTCCTTTAGTTTCCTCGTCTAAATTATG-3'
MT196300	AntV4	I	KPSTAsw001Ad	22-Dec-15	St. Andrews, South Georgia	*Aptenodytes patagonicus*	F: 5'-CACATCGACAAGTATATTGTTGACTCCCGTATTCA-3' R: 5'-GAGGTTGTAGTTTGTATCATTGGTAAGATCAAACA-3'
MT196308	AntV4	I	KPSTAsw002Ad	22-Dec-15	St. Andrews, South Georgia	*Aptenodytes patagonicus*	F: 5'-CAACACAATGCCAACAACAATTCATTTTTCTAATA-3' R: 5'-GGGAATGTATTAAATATGGGCGAGTGGGCATTAAA-3'
MT196309	AntV4	I	KPSTAsw008Ad	22-Dec-15	St. Andrews, South Georgia	*Aptenodytes patagonicus*	F: 5'-CAACACAATGCCAACAACAATTCATTTTTCTAATA-3' R: 5'-GGGAATGTATTAAATATGGGCGAGTGGGCATTAAA-3'
MT196310	AntV4	I	KPSTAsw009Ad	22-Dec-15	St. Andrews, South Georgia	*Aptenodytes patagonicus*	F: 5'-CAACACAATGCCAACAACAATTCATTTTTCTAATA-3' R: 5'-GGGAATGTATTAAATATGGGCGAGTGGGCATTAAA-3'
MT196311	AntV4	I	KPSTAsw010Ad	22-Dec-15	St. Andrews, South Georgia	*Aptenodytes patagonicus*	F: 5'-CAACACAATGCCAACAACAATTCATTTTTCTAATA-3' F: 5'-GGGAATGTATTAAATATGGGCGAGTGGGCATTAAA-3'
MT196301	AntV4	II	KPSTAsw004Ad	22-Dec-15	St. Andrews, South Georgia	*Aptenodytes patagonicus*	F: 5'-CACATCGACAAGTATATTGTTGACTCCCGTATTCA-3' R: 5'-GAGGTTGTAGTTTGTATCATTGGTAAGATCAAACA-3'
MT196302	AntV4	II	KPSTAsw014Ch	22-Dec-15	St. Andrews, South Georgia	*Aptenodytes patagonicus*	F: 5'-CACATCGACAAGTATATTGTTGACTCCCGTATTCA-3' R: 5'-GAGGTTGTAGTTTGTATCATTGGTAAGATCAAACA-3'
MT196303	AntV4	II	KPSTAsw015Ch	22-Dec-15	St. Andrews, South Georgia	*Aptenodytes patagonicus*	F: 5'-CACATCGACAAGTATATTGTTGACTCCCGTATTCA-3' R: 5'-GAGGTTGTAGTTTGTATCATTGGTAAGATCAAACA-3'
MT196304	AntV4	II	KPSTAsw016Ch	22-Dec-15	St. Andrews, South Georgia	*Aptenodytes patagonicus*	F: 5'-CACATCGACAAGTATATTGTTGACTCCCGTATTCA-3' R: 5'-GAGGTTGTAGTTTGTATCATTGGTAAGATCAAACA-3'
MT196305	AntV4	II	KPSTAsw017Ch	22-Dec-15	St. Andrews, South Georgia	*Aptenodytes patagonicus*	F: 5'-CACATCGACAAGTATATTGTTGACTCCCGTATTCA-3' R: 5'-GAGGTTGTAGTTTGTATCATTGGTAAGATCAAACA-3'
MT196306	AntV4	II	KPSTAsw018Ch	22-Dec-15	St. Andrews, South Georgia	*Aptenodytes patagonicus*	F: 5'-CACATCGACAAGTATATTGTTGACTCCCGTATTCA-3' R: 5'-GAGGTTGTAGTTTGTATCATTGGTAAGATCAAACA-3'
MT196307	AntV4	II	KPSTAsw019Ch	22-Dec-15	St. Andrews, South Georgia	*Aptenodytes patagonicus*	F: 5'-CACATCGACAAGTATATTGTTGACTCCCGTATTCA-3' R: 5'-GAGGTTGTAGTTTGTATCATTGGTAAGATCAAACA-3'
MT196312	AntV5	I	KPSTAsw001Ad	22-Dec-15	St. Andrews, South Georgia	*Aptenodytes patagonicus*	F: 5'-CCCTAACCAGATCCTTCCTATGAAATTTTATCTC-3' R: 5'-GAAAGATCAAGCTTGTGAGTTTTTATGATCTTAA-3'
MT196313	AntV5	I	KPSTAsw002Ad	22-Dec-15	St. Andrews, South Georgia	*Aptenodytes patagonicus*	F: 5'-CCCTAACCAGATCCTTCCTATGAAATTTTATCTC-3' R: 5'-GAAAGATCAAGCTTGTGAGTTTTTATGATCTTAA-3'
MT196314	AntV5	I	KPSTAsw004Ad	22-Dec-15	St. Andrews, South Georgia	*Aptenodytes patagonicus*	F: 5'-CCCTAACCAGATCCTTCCTATGAAATTTTATCTC-3' R: 5'-GAAAGATCAAGCTTGTGAGTTTTTATGATCTTAA-3'
MT196315	AntV5	I	KPSTAsw008Ad	22-Dec-15	St. Andrews, South Georgia	*Aptenodytes patagonicus*	F: 5'-CCCTAACCAGATCCTTCCTATGAAATTTTATCTC-3' R: 5'-GAAAGATCAAGCTTGTGAGTTTTTATGATCTTAA-3'
MT196316	AntV5	I	KPSTAsw009Ad	22-Dec-15	St. Andrews, South Georgia	*Aptenodytes patagonicus*	F: 5'-CCCTAACCAGATCCTTCCTATGAAATTTTATCTC-3' R: 5'-GAAAGATCAAGCTTGTGAGTTTTTATGATCTTAA-3'
MT196317	AntV5	I	KPSTAsw018Ch	22-Dec-15	St. Andrews, South Georgia	*Aptenodytes patagonicus*	F: 5'-CCCTAACCAGATCCTTCCTATGAAATTTTATCTC-3' R: 5'-GAAAGATCAAGCTTGTGAGTTTTTATGATCTTAA-3'
MT196318	AntV5	I	KPSTAsw020Ch	22-Dec-15	St. Andrews, South Georgia	*Aptenodytes patagonicus*	F: 5'-CCCTAACCAGATCCTTCCTATGAAATTTTATCTC-3' R: 5'-GAAAGATCAAGCTTGTGAGTTTTTATGATCTTAA-3'
MT196222	AntV6	I	APBOOTsw001Ad	7-Jan-16	Booth Island, West Antarctic peninsula	*Pygoscelis adeliae*	F: 5'-AGGACAAAATATAATATTATGAGAAGGCAA-3' R: 5'-AATACTTTTTCTATAACTACAAAACGTCGC-3'

MT196223	AntV6	I	APBOOTsw001Ch	7-Jan-16	Booth Island, West Antarctic peninsula	*Pygoscelis adeliae*	F: 5'-AGGACAAAATATAATATTATGAGAAGGCAA-3' R: 5'-AATACTTTTTCTATAACTACAAAACGTCGC-3'
MT196226	AntCM1	I	CPBAILsw002Ad	29-Dec-15	Baily Head, Deception Island	*Pygoscelis antarcticus*	F: 5'-CCGAAACATTATAAAAACCAGAAATATCCTCAATC-3' R: 5'-AATTTAAGGAAGAGAGATTGGAGGTTGATGAAGAT-3'
MT196251	AntCM2	I	CPHALFsw001Ad	3-Jan-16	Half Moon Island, SouthShetland	*Pygoscelis antarcticus*	F: 5'-GAGAGAGTATGGAAGATAAGAAGAATAGAGATACG-3' R: 5'-ATCCCTCCAAACTTTGATACAAGATACATAAGTTG-3'
MT196224	AntCM3	I	CPBAILsw002Ad	29-Dec-15	Baily Head, Deception Island	*Pygoscelis antarcticus*	F: 5'-ATAGCGCAGAATAATATCATAACACCTCTCTATAA-3' R: 5'-TGTTCCTGTCATATCCATTAGCTTAGTATATTTCT-3'
MT196272	AntCM3	I	CPHALFsw001Ad	3-Jan-16	Half Moon Island, SouthShetland	*Pygoscelis antarcticus*	F: 5'-TGTTCCTGTCATATCCATTAGCTTAGTATATTTCT-3' R: 5'-ATAGCGCAGAATAATATCATAACACCTCTCTATAA-3'
MT196273	AntCM3	I	CPHALFsw003Ad	3-Jan-16	Half Moon Island, SouthShetland	*Pygoscelis antarcticus*	F: 5'-TGTTCCTGTCATATCCATTAGCTTAGTATATTTCT-3' R: 5'-ATAGCGCAGAATAATATCATAACACCTCTCTATAA-3'
MT196274	AntCM3	I	CPHALFsw004Ad	3-Jan-16	Half Moon Island, SouthShetland	*Pygoscelis antarcticus*	F: 5'-TGTTCCTGTCATATCCATTAGCTTAGTATATTTCT-3' R: 5'-ATAGCGCAGAATAATATCATAACACCTCTCTATAA-3'
MT196276	AntCM3	I	CPHALFsw008Ad	3-Jan-16	Half Moon Island, SouthShetland	*Pygoscelis antarcticus*	F: 5'-TGTTCCTGTCATATCCATTAGCTTAGTATATTTCT-3' R: 5'-ATAGCGCAGAATAATATCATAACACCTCTCTATAA-3'
MT196225	AntCM3	II	CPBAILsw005Ad	29-Dec-15	Baily Head, Deception Island	*Pygoscelis antarcticus*	F: 5'-ATAGCGCAGAATAATATCATAACACCTCTCTATAA-3' R: 5'-TGTTCCTGTCATATCCATTAGCTTAGTATATTTCT-3'
MT196275	AntCM3	III	CPHALFsw005Ad	3-Jan-16	Half Moon Island, SouthShetland	*Pygoscelis antarcticus*	F: 5'-TGTTCCTGTCATATCCATTAGCTTAGTATATTTCT-3' R: 5'-ATAGCGCAGAATAATATCATAACACCTCTCTATAA-3'
MT196277	AntCM3	III	CPHALFsw009Ad	3-Jan-16	Half Moon Island, SouthShetland	*Pygoscelis antarcticus*	F: 5'-TGTTCCTGTCATATCCATTAGCTTAGTATATTTCT-3' R: 5'-ATAGCGCAGAATAATATCATAACACCTCTCTATAA-3'
MT196278	AntCM3	III	CPHALFsw010Ad	3-Jan-16	Half Moon Island, SouthShetland	*Pygoscelis antarcticus*	F: 5'-TGTTCCTGTCATATCCATTAGCTTAGTATATTTCT-3' R: 5'-ATAGCGCAGAATAATATCATAACACCTCTCTATAA-3'
MT196227	AntCM4	I	CPBAILsw001Ad	29-Dec-15	Baily Head, Deception Island	*Pygoscelis antarcticus*	F: 5'-CAGCAAAACTCATAGAATACGTAGATGATAACATA-3' R: 5'-TTATATCATCTGTTTTCTTTTCGTATGCTAACTCT-3'
MT196228	AntCM4	I	CPBAILsw002Ad	29-Dec-15	Baily Head, Deception Island	*Pygoscelis antarcticus*	F: 5'-CAGCAAAACTCATAGAATACGTAGATGATAACATA-3' R: 5'-TTATATCATCTGTTTTCTTTTCGTATGCTAACTCT-3'
MT196229	AntCM4	I	CPBAILsw004Ad	29-Dec-15	Baily Head, Deception Island	*Pygoscelis antarcticus*	F: 5'-CAGCAAAACTCATAGAATACGTAGATGATAACATA-3' R: 5'-TTATATCATCTGTTTTCTTTTCGTATGCTAACTCT-3'
MT196262	AntCM4	I	CPHALFsw003Ad	3-Jan-16	Half Moon Island, SouthShetland	*Pygoscelis antarcticus*	F: 5'-TTATATCATCTGTTTTCTTTTCGTATGCTAACTCT-3' R: 5'-CAGCAAAACTCATAGAATACGTAGATGATAACATA-3'
MT196263	AntCM4	I	CPHALFsw004Ad	3-Jan-16	Half Moon Island, SouthShetland	*Pygoscelis antarcticus*	F: 5'-TTATATCATCTGTTTTCTTTTCGTATGCTAACTCT-3' R: 5'-CAGCAAAACTCATAGAATACGTAGATGATAACATA-3'
MT196264	AntCM4	I	CPHALFsw008Ad	3-Jan-16	Half Moon Island, SouthShetland	*Pygoscelis antarcticus*	F: 5'-TTATATCATCTGTTTTCTTTTCGTATGCTAACTCT-3' R: 5'-CAGCAAAACTCATAGAATACGTAGATGATAACATA-3'
MT196265	AntCM4	I	CPHALFsw009Ad	3-Jan-16	Half Moon Island, SouthShetland	*Pygoscelis antarcticus*	F: 5'-TTATATCATCTGTTTTCTTTTCGTATGCTAACTCT-3' R: 5'-CAGCAAAACTCATAGAATACGTAGATGATAACATA-3'
MT196266	AntCM4	I	CPHALFsw010Ad	3-Jan-16	Half Moon Island, SouthShetland	*Pygoscelis antarcticus*	F: 5'-TTATATCATCTGTTTTCTTTTCGTATGCTAACTCT-3' R: 5'-CAGCAAAACTCATAGAATACGTAGATGATAACATA-3'
MT196230	AntCM5	I	CPBAILsw001Ad	29-Dec-15	Baily Head, Deception Island	*Pygoscelis antarcticus*	F: 5'-GGGATTTACAGTTACGAATATATGGCTGTATTTTA-3' R: 5'-CCTCCCTCATTGCTACTAGATGATTTTATAAAAAT-3'
MT196231	AntCM5	I	CPBAILsw002Ad	29-Dec-15	Baily Head, Deception Island	*Pygoscelis antarcticus*	F: 5'-GGGATTTACAGTTACGAATATATGGCTGTATTTTA-3' R: 5'-CCTCCCTCATTGCTACTAGATGATTTTATAAAAAT-3'

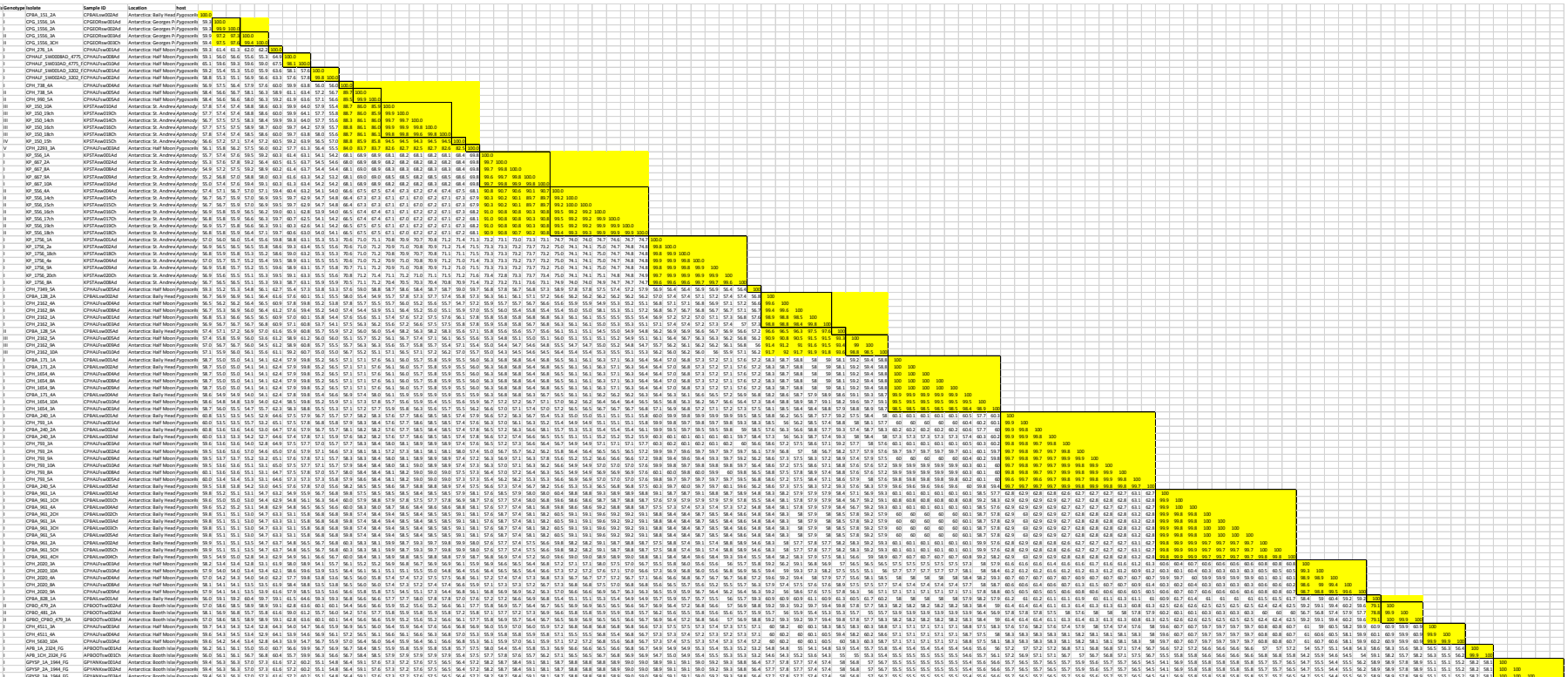
MT196232	AntCM5	I	CPBAILsw003Ad	29-Dec-15	Baily Head, Deception Island	*Pygoscelis antarcticus*	F: 5'-GGGATTTACAGTTACGAATATATGGCTGTATTTTA-3' R: 5'-CCTCCCTCATTGCTACTAGATGATTTTATAAAAAT-3'
MT196233	AntCM5	I	CPBAILsw005Ad	29-Dec-15	Baily Head, Deception Island	*Pygoscelis antarcticus*	F: 5'-GGGATTTACAGTTACGAATATATGGCTGTATTTTA-3' R: 5'-CCTCCCTCATTGCTACTAGATGATTTTATAAAAAT-3'
MT196254	AntCM5	I	CPHALFsw001Ad	3-Jan-16	Half Moon Island, SouthShetland	*Pygoscelis antarcticus*	F: 5'-GGGATTTACAGTTACGAATATATGGGTGTATTTTA-3' R: 5'-CCTCCCTCATTGCTACTAGATGATTTTATAAAAAT-3'
MT196255	AntCM5	I	CPHALFsw002Ad	3-Jan-16	Half Moon Island, SouthShetland	*Pygoscelis antarcticus*	F: 5'-GGGATTTACAGTTACGAATATATGGGTGTATTTTA-3' R: 5'-CCTCCCTCATTGCTACTAGATGATTTTATAAAAAT-3'
MT196256	AntCM5	I	CPHALFsw003Ad	3-Jan-16	Half Moon Island, SouthShetland	*Pygoscelis antarcticus*	F: 5'-GGGATTTACAGTTACGAATATATGGGTGTATTTTA-3' R: 5'-CCTCCCTCATTGCTACTAGATGATTTTATAAAAAT-3'
MT196257	AntCM5	I	CPHALFsw005Ad	3-Jan-16	Half Moon Island, SouthShetland	*Pygoscelis antarcticus*	F: 5'-GGGATTTACAGTTACGAATATATGGGTGTATTTTA-3' R: 5'-CCTCCCTCATTGCTACTAGATGATTTTATAAAAAT-3'
MT196258	AntCM5	I	CPHALFsw008Ad	3-Jan-16	Half Moon Island, SouthShetland	*Pygoscelis antarcticus*	F: 5'-GGGATTTACAGTTACGAATATATGGGTGTATTTTA-3' R: 5'-CCTCCCTCATTGCTACTAGATGATTTTATAAAAAT-3'
MT196259	AntCM5	I	CPHALFsw009Ad	3-Jan-16	Half Moon Island, SouthShetland	*Pygoscelis antarcticus*	F: 5'-GGGATTTACAGTTACGAATATATGGGTGTATTTTA-3' R: 5'-CCTCCCTCATTGCTACTAGATGATTTTATAAAAAT-3'
MT196260	AntCM5	I	CPHALFsw010Ad	3-Jan-16	Half Moon Island, SouthShetland	*Pygoscelis antarcticus*	F: 5'-GGGATTTACAGTTACGAATATATGGGTGTATTTTA-3' R: 5'-CCTCCCTCATTGCTACTAGATGATTTTATAAAAAT-3'
MT196235	AntCM6	I	CPBAILsw001Ad	29-Dec-15	Baily Head, Deception Island	*Pygoscelis antarcticus*	F: 5'-ATGTATTGTAGGTGTCATAGAAGGAAAGTAGTATC-3' R: 5'-CATCAACTACTGATATTGTTTGTACAAAAGTAACA-3'
MT196236	AntCM6	I	CPBAILsw001Ch	29-Dec-15	Baily Head, Deception Island	*Pygoscelis antarcticus*	F: 5'-ATGTATTGTAGGTGTCATAGAAGGAAAGTAGTATC-3' R: 5'-CATCAACTACTGATATTGTTTGTACAAAAGTAACA-3'
MT196237	AntCM6	I	CPBAILsw002Ad	29-Dec-15	Baily Head, Deception Island	*Pygoscelis antarcticus*	F: 5'-ATGTATTGTAGGTGTCATAGAAGGAAAGTAGTATC-3' R: 5'-CATCAACTACTGATATTGTTTGTACAAAAGTAACA-3'
MT196238	AntCM6	I	CPBAILsw002Ch	29-Dec-15	Baily Head, Deception Island	*Pygoscelis antarcticus*	F: 5'-ATGTATTGTAGGTGTCATAGAAGGAAAGTAGTATC-3' R: 5'-CATCAACTACTGATATTGTTTGTACAAAAGTAACA-3'
MT196239	AntCM6	I	CPBAILsw003Ad	29-Dec-15	Baily Head, Deception Island	*Pygoscelis antarcticus*	F: 5'-ATGTATTGTAGGTGTCATAGAAGGAAAGTAGTATC-3' R: 5'-CATCAACTACTGATATTGTTTGTACAAAAGTAACA-3'
MT196240	AntCM6	I	CPBAILsw003Ch	29-Dec-15	Baily Head, Deception Island	*Pygoscelis antarcticus*	F: 5'-ATGTATTGTAGGTGTCATAGAAGGAAAGTAGTATC-3' R: 5'-CATCAACTACTGATATTGTTTGTACAAAAGTAACA-3'
MT196241	AntCM6	I	CPBAILsw004Ad	29-Dec-15	Baily Head, Deception Island	*Pygoscelis antarcticus*	F: 5'-ATGTATTGTAGGTGTCATAGAAGGAAAGTAGTATC-3' R: 5'-CATCAACTACTGATATTGTTTGTACAAAAGTAACA-3'
MT196242	AntCM6	I	CPBAILsw004Ch	29-Dec-15	Baily Head, Deception Island	*Pygoscelis antarcticus*	F: 5'-ATGTATTGTAGGTGTCATAGAAGGAAAGTAGTATC-3' R: 5'-CATCAACTACTGATATTGTTTGTACAAAAGTAACA-3'
MT196243	AntCM6	I	CPBAILsw005Ad	29-Dec-15	Baily Head, Deception Island	*Pygoscelis antarcticus*	F: 5'-ATGTATTGTAGGTGTCATAGAAGGAAAGTAGTATC-3' R: 5'-CATCAACTACTGATATTGTTTGTACAAAAGTAACA-3'
MT196244	AntCM6	I	CPBAILsw005Ch	29-Dec-15	Baily Head, Deception Island	*Pygoscelis antarcticus*	F: 5'-ATGTATTGTAGGTGTCATAGAAGGAAAGTAGTATC-3' R: 5'-CATCAACTACTGATATTGTTTGTACAAAAGTAACA-3'
MT196280	AntCM7	I	CPHALFsw003Ad	3-Jan-16	Half Moon Island, SouthShetland	*Pygoscelis antarcticus*	F: 5'-CAAAACAGAAAATAATATAACAGATCTAGACGCAG-3' R: 5'-TAAACTTCAATAATATCTACATGCATTGTTGTTGC-3'
MT196281	AntCM7	I	CPHALFsw004Ad	3-Jan-16	Half Moon Island, SouthShetland	*Pygoscelis antarcticus*	F: 5'-CAAAACAGAAAATAATATAACAGATCTAGACGCAG-3' R: 5'-TAAACTTCAATAATATCTACATGCATTGTTGTTGC-3'
MT196282	AntCM7	I	CPHALFsw010Ad	3-Jan-16	Half Moon Island, SouthShetland	*Pygoscelis antarcticus*	F: 5'-GAAGAATTTTACTTGACTTGAGGCATTTTCATATT-3' R: 5'-GATTACCATCTAATTTTATACGCTTACAGCAATTA-3'

MT196234	AntCM8	I	CPBAILsw001Ad	29-Dec-15	Baily Head, Deception Island	*Pygoscelis antarcticus*	F: 5'-GTATTAATCCGACCTACAAAACGATTGTATATGAT-3' R: 5'-CATCATAAGGATGTTTATAATCATCGACTAGCCAC-3'
MT196246	AntCM8	II	CPBOOTsw002Ad	7-Jan-16	Booth Island, West Antarctic peninsula	*Pygoscelis antarcticus*	F: 5'-TATAGCAGAAAGGCAAGAATAATGAATAATACAGC-3' R: 5'-GCCTGGTATAACGCATTATATCATCAATAATCATA-3'
MT196245	AntCM8	II	CPBOOTsw002Ad	7-Jan-16	Booth Island, West Antarctic peninsula	*Pygoscelis antarcticus*	F: 5'-AGGAAGATAAAAATCCTACTGTAAAAGGTGAATTT-3' R: 5'-TTTGAATATATTTGAGCACGTCATGTATGTTTCTA-3'
MT196288	AntCM8	II	GPBOOTsw003Ad	17-Jan-16	Booth Island, West Antarctic peninsula	*Pygoscelis papua*	F: 5'-TTTGAATATATTTGAGCACGTCATGTATGTTTCTA-3' R: 5'-AGGAAGATAAAAATCCTACTGTAAAAGGTGAATTT-3'
MT196284	AntCM9	I	CPHALFsw001Ad	3-Jan-16	Half Moon Island, SouthShetland	*Pygoscelis antarcticus*	F: 5'-GTATTAGCAAATTAGCAGAATTAGCATACG-3' R: 5'-GCTCTTGACTCATATTGGTTCTTATTTTTC-3'
MT196285	AntCM9	I	CPHALFsw002Ad	3-Jan-16	Half Moon Island, SouthShetland	*Pygoscelis antarcticus*	F: 5'-GTATTAGCAAATTAGCAGAATTAGCATACG-3' R: 5'-GCTCTTGACTCATATTGGTTCTTATTTTTC-3'
MT196286	AntCM10	I	CPHALFsw008Ad	3-Jan-16	Half Moon Island, SouthShetland	*Pygoscelis antarcticus*	F: 5'-CGATAATAACACTGCGGTTCTAGATATACT-3' R: 5'-CTAAATATTTTTAGCGATACACTTTTGGCG-3'
MT196287	AntCM10	I	CPHALFsw010Ad	3-Jan-16	Half Moon Island, SouthShetland	*Pygoscelis antarcticus*	F: 5'-CGATAATAACACTGCGGTTCTAGATATACT-3' R: 5'-CTAAATATTTTTAGCGATACACTTTTGGCG-3'
MT196283	AntCM11	I	CPHALFsw005Ad	3-Jan-16	Half Moon Island, SouthShetland	*Pygoscelis antarcticus*	F: 5'-GGAGATTTGTTTGTAGTTGGGATGAGGTATTAAAG-3' R: 5'-GGTAAAATCATTAATTGCATGTAAATCTTAGGTAATCTTACCC-3'
MT196267	AntCM12	I	CPHALFsw003Ad	3-Jan-16	Half Moon Island, SouthShetland	*Pygoscelis antarcticus*	F: 5'-CTTTATTAGCATACTGGTCTTAATACAACACTTCT-3' R: 5'-CCATAAGTGAAAAATTAAGTAAGACACCATAACAA-3'
MT196268	AntCM12	I	CPHALFsw004Ad	3-Jan-16	Half Moon Island, SouthShetland	*Pygoscelis antarcticus*	F: 5'-CTTTATTAGCATACTGGTCTTAATACAACACTTCT-3' R: 5'-CCATAAGTGAAAAATTAAGTAAGACACCATAACAA-3'
MT196269	AntCM12	I	CPHALFsw008Ad	3-Jan-16	Half Moon Island, SouthShetland	*Pygoscelis antarcticus*	F: 5'-CTTTATTAGCATACTGGTCTTAATACAACACTTCT-3' R: 5'-CCATAAGTGAAAAATTAAGTAAGACACCATAACAA-3'
MT196270	AntCM12	I	CPHALFsw009Ad	3-Jan-16	Half Moon Island, SouthShetland	*Pygoscelis antarcticus*	F: 5'-CTTTATTAGCATACTGGTCTTAATACAACACTTCT-3' R: 5'-CCATAAGTGAAAAATTAAGTAAGACACCATAACAA-3'
MT196271	AntCM12	I	CPHALFsw010Ad	3-Jan-16	Half Moon Island, SouthShetland	*Pygoscelis antarcticus*	F: 5'-CTTTATTAGCATACTGGTCTTAATACAACACTTCT-3' R: 5'-CCATAAGTGAAAAATTAAGTAAGACACCATAACAA-3'

Supplementary Data 1: Details of cloacal swab samples of the four penguin species. [Electronic attachment with preview below]

| | | | | | | Antarctic viral-like circular molecule (AntCM) | | | | | | | | | | | | Rep-encoding AntCM | | | Antarctic (CRESS DNA) viruses (AntV) | | | | | | | | | | | |

Supplementary Data 2: Pairwise identity matrix for determination of AntV and AntCM species and genotypes. [Electronic attachment with preview below]



Appendix A4 – Supplementary Data for Chapter 6

Supplementary Data 1: Spreadsheet containing pairwise identities of the avian circovirus genomes, and the Rep and CP amino acid sequences.

A. Genome-wide pairwise identities:

	MN877414_PenCV	MN877415_PenCV	MN164706_PenCV	MN164704_PenCV	MN164703_PenCV	MN164705_PenCV	DQ845074_GuCV	AF071878_BFDV	DQ146997_RaCV	KX987146_TiACV2	AJ301633_CaPV	KP793918_ZfiCV	DQ845075_FiCV	AF252610_PiCV	MH188038_MosACV	DQ172906_StCV	KU230452_TiACV1	GQ404851_ChimpACV	DQ100076_DuCV	EU056309_SwCV	AJ304456_GoCV
MN877414_PenCV	100.0																				
MN877415_PenCV	99.2	100.0																			
MN164706_PenCV	87.2	87.1	100.0																		
MN164704_PenCV	87.8	87.4	99.1	100.0																	
MN164703_PenCV	87.7	87.6	99.3	99.5	100.0																
MN164705_PenCV	87.8	87.7	98.9	99.1	99.5	100.0															
DQ845074_GuCV	66.9	67.1	66.3	66.4	66.6	66.9	100.0														
AF071878_BFDV	62.3	62.3	63.4	63.4	63.2	63.7	61.3	100.0													
DQ146997_RaCV	63.1	63.3	63.3	63.3	63.7	63.7	62.5	64.1	100.0												
KX987146_TiACV2	63.1	63.3	65.7	65.5	65.6	64.5	64.4	64.3	77.1	100.0											
AJ301633_CaPV	62.5	62.9	62.0	62.1	62.5	62.4	63.1	62.9	70.5	70.5	100.0										
KP793918_ZfiCV	62.3	62.2	62.9	63.0	63.1	63.2	62.6	62.1	69.4	70.4	67.8	100.0									
DQ845075_FiCV	62.5	62.8	61.7	61.3	61.5	61.7	65.5	62.7	68.8	69.7	68.9	79.4	100.0								
AF252610_PiCV	61.4	61.6	62.4	62.4	62.7	62.3	65.3	65.2	68.3	67.5	64.4	64.9	65.3	100.0							
MH188038_MosACV	64.6	64.4	64.6	65.1	64.8	64.7	61.7	62.9	63.5	64.2	64.0	65.2	62.5	69.6	100.0						
DQ172906_StCV	64.0	64.1	62.1	62.9	62.8	63.4	62.9	62.2	66.3	68.9	69.1	66.4	67.3	68.0	66.6	100.0					
KU230452_TiACV1	61.7	62.9	62.2	61.8	62.0	61.9	62.9	60.3	68.5	67.9	66.8	68.3	69.0	69.3	66.9	71.3	100.0				
GQ404851_ChimpACV	63.4	63.9	61.6	62.1	62.1	62.4	63.7	63.3	67.1	67.6	66.0	67.4	68.0	68.7	66.8	71.3	73.1	100.0			
DQ100076_DuCV	59.0	59.4	58.8	58.4	58.4	58.3	57.7	56.8	58.7	58.1	57.7	58.6	56.9	58.2	60.8	57.5	58.7	59.1	100.0		
EU056309_SwCV	57.4	57.3	57.9	58.0	58.2	57.6	59.2	60.3	61.8	61.3	57.1	59.2	58.8	58.1	59.4	60.1	59.1	57.5	69.3	100.0	
AJ304456_GoCV	59.6	60.1	58.4	58.8	58.5	57.9	60.4	60.9	58.9	58.4	58.4	61.6	57.2	60.3	58.5	57.1	59.9	59.0	71.0	75.6	100.0
Supplementary Data 1: Spreadsheet containing pairwise identities of the avian circovirus genomes, and the Rep and CP amino acid sequences.																					
This tab = Genome identity																					

B. Replication-associated protein pairwise identities:

	MN877414_PenCV	MN877415_PenCV	MN164706_PenCV	MN164704_PenCV	MN164703_PenCV	MN164705_PenCV	DQ845074_GuCV	DQ146997_RaCV	KX987146_TiACV2	AJ301633_CaPV	KP793918_ZfiCV	DQ845075_FiCV	AF252610_PiCV	MH188038_MosACV	DQ172906_StCV	KU230452_TiACV1	GQ404851_ChimpACV	AF071878_BFDV	DQ100076_DuCV	EU056309_SwCV	AJ304456_GoCV
MN877414_PenCV	100.00																				
MN877415_PenCV	99.30	100.00																			
MN164706_PenCV	95.80	95.80	100.00																		
MN164704_PenCV	95.80	95.80	99.70	100.00																	
MN164703_PenCV	95.80	95.80	100.00	99.70	100.00																
MN164705_PenCV	95.80	95.80	100.00	99.70	100.00	100.00															
DQ845074_GuCV	67.10	66.40	67.40	67.40	67.10	67.10	100.00														
DQ146997_RaCV	65.40	66.10	67.40	67.40	67.50	67.50	61.20	100.00													
KX987146_TiACV2	65.70	66.80	66.30	66.30	66.40	66.40	60.20	85.90	100.00												
AJ301633_CaPV	66.80	66.80	67.70	67.70	67.80	67.80	66.10	76.90	75.50	100.00											
KP793918_ZfiCV	65.10	65.40	66.00	66.00	66.10	66.10	61.60	75.60	76.30	74.10	100.00										
DQ845075_FiCV	67.10	67.50	67.70	67.70	67.80	67.80	65.10	77.70	77.00	75.50	87.60	100.00									
AF252610_PiCV	65.70	65.40	66.00	66.00	65.70	65.70	61.60	72.90	72.50	69.30	70.40	72.50	100.00								
MH188038_MosACV	63.70	63.70	65.00	65.00	64.80	64.80	58.20	67.70	68.40	65.20	66.30	67.40	72.20	100.00							
DQ172906_StCV	67.70	67.40	67.90	67.90	67.70	67.70	63.50	78.90	81.30	76.50	80.30	83.40	73.40	69.80	100.00						
KU230452_TiACV1	67.50	68.20	67.40	67.40	67.10	67.10	60.90	80.70	82.10	77.20	79.70	78.60	75.90	67.70	78.90	100.00					
GQ404851_ChimpACV	65.40	65.40	66.70	66.70	66.80	66.80	60.60	80.10	78.30	75.50	80.40	79.40	73.90	69.80	79.20	82.10	100.00				
AF071878_BFDV	62.30	62.30	61.50	61.50	61.20	61.20	55.00	58.10	62.80	58.80	60.60	62.30	61.20	58.50	60.80	58.80	61.90	100.00			
DQ100076_DuCV	48.30	47.90	48.40	48.40	48.60	48.60	47.80	48.60	48.60	47.60	51.70	51.00	51.70	47.00	50.20	50.00	50.30	46.70	100.00		
EU056309_SwCV	47.90	47.60	48.80	48.80	48.60	48.60	46.70	47.40	49.30	46.70	51.20	51.60	48.30	45.90	51.20	48.80	48.60	45.50	78.40	100.00	
AJ304456_GoCV	48.30	47.90	48.10	48.80	47.90	47.90	46.70	47.90	50.00	47.60	52.10	51.40	49.00	46.60	51.20	49.30	49.70	45.50	82.90	85.30	100.00

Supplementary Data 1: Spreadsheet containing pairwise identities of the avian circovirus genomes, and the Rep and CP amino acid sequences.

This tab = Rep identity

C. Capsid protein pairwise identities:

	MN877414_PenCV	MN877415_PenCV	MN164706_PenCV	MN164704_PenCV	MN164705_PenCV	MN164703_PenCV	DQ845074_GuCV	AF071878_BFDV	DQ146997_RaCV	KX987146_TiACV2	AJ301633_CaPV	KP793918_ZfiCV	DQ845075_FiCV	AF252610_PiCV	MH188038_MosACV	DQ172906_StCV	KU230452_TiACV1	GQ404851_ChimpACV	DQ100076_DuCV	EU056309_SwCV	AJ304456_GoCV
MN877414_PenCV	100.00																				
MN877415_PenCV	99.60	100.00																			
MN164706_PenCV	90.90	90.90	100.00																		
MN164704_PenCV	91.70	91.70	99.20	100.00																	
MN164705_PenCV	91.70	91.70	99.20	100.00	100.00																
MN164703_PenCV	91.70	91.70	99.20	100.00	100.00	100.00															
DQ845074_GuCV	57.60	57.60	56.80	57.60	57.60	57.60	100.00														
AF071878_BFDV	51.70	51.20	52.60	53.10	53.10	53.10	50.20	100.00													
DQ146997_RaCV	48.80	48.80	46.40	47.40	47.40	47.40	48.10	43.20	100.00												
KX987146_TiACV2	48.10	48.10	47.60	48.60	48.60	48.60	47.20	47.80	87.00	100.00											
AJ301633_CaPV	45.10	45.10	45.10	46.00	46.00	46.00	45.00	42.40	75.30	76.60	100.00										
KP793918_ZfiCV	51.00	51.00	49.50	50.50	50.50	50.50	51.10	49.10	61.00	62.00	63.10	100.00									
DQ845075_FiCV	48.80	48.80	48.30	49.30	49.30	49.30	46.90	48.20	61.20	62.20	67.70	75.20	100.00								
AF252610_PiCV	40.70	40.70	39.00	40.30	40.30	40.30	48.00	46.80	46.20	44.80	47.50	49.40	48.70	100.00							
MH188038_MosACV	41.50	41.50	41.40	41.40	41.40	41.40	42.00	45.60	46.50	46.50	46.20	48.50	46.10	65.80	100.00						
DQ172906_StCV	43.60	44.10	43.20	44.10	44.10	44.10	45.60	47.70	50.20	50.40	55.20	50.20	51.70	57.00	60.20	100.00					
KU230452_TiACV1	44.80	46.10	44.30	45.20	45.20	45.20	45.90	46.40	46.50	46.50	51.90	51.20	52.40	60.70	63.70	71.40	100.00				
GQ404851_ChimpACV	44.40	44.40	45.30	45.70	45.70	45.70	45.70	47.00	44.60	44.60	49.30	48.50	50.80	60.90	62.20	75.00	79.80	100.00			
DQ100076_DuCV	27.20	27.20	25.40	25.40	25.40	25.40	28.80	30.60	25.70	26.80	26.00	22.50	25.30	29.20	28.10	28.30	30.80	26.10	100.00		
EU056309_SwCV	29.00	29.00	30.00	30.50	30.50	30.50	32.40	30.70	28.50	27.70	27.50	28.60	29.90	34.10	31.40	32.00	30.50	29.00	55.10	100.00	
AJ304456_GoCV	25.50	25.50	26.50	26.50	26.50	26.50	26.70	28.40	27.90	27.50	26.60	26.50	29.00	34.10	28.70	30.40	28.90	25.50	49.40	65.20	100.00

Supplementary Data 1: Spreadsheet containing pairwise identities of the avian circovirus genomes, and the Rep and CP amino acid sequences.

This tab = CP identity

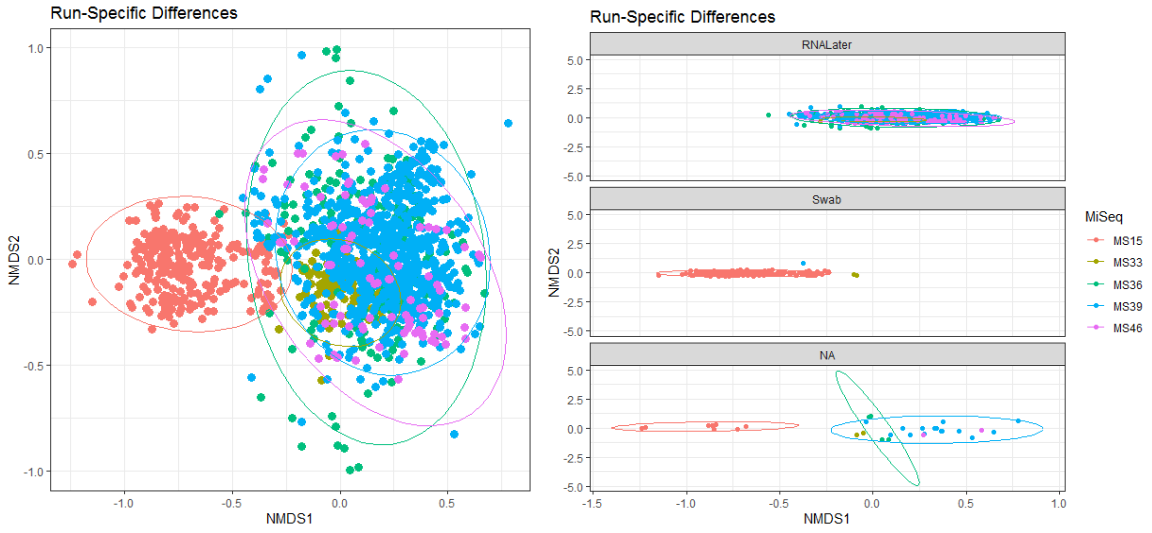
Supplementary Data 9. Bayes Empirical Bayes inference of the probability of positively selected sites in PenCV protein-coding genes obtained from various codeml models (Yang et al., 2005).

Gene	**Model**	**Log-likelihood**	**2 x (ΔlnL)**	**Chi^2**	**Selected sites (Pr)**
cp	M1	-1206.358425	3.2881	0.193	177 (0.640); 201 (0.934)
cp	M2a	-1204.714375			
cp	M7	-1207.004397	4.573534	0.102	177 (0.733); 201 (**0.973**)
cp	M8	-1204.717630			
rep	M1	-1609.231633	0.284004	0.868	33 (0.755)
rep	M2a	-1609.089631			
rep	M7	-1609.865526	1.545972	0.462	33 (0.887)
rep	M8	-1609.092540			

Appendix A5 - Supplementary Data for Chapter 7

Effect of Methodological Differences on Microbiome Results

The communities recovered from different sample types (swab vs. guano – each stored in either Amies liquid transport medium or RNALater, respectively) differed substantially in bacterial diversity and composition. There could be several underlying mechanisms for these evident differences: (1) fundamental differences in the bacterial composition of faeces compared to the cloacal surface/lumen, (2) differential preservation of bacterial DNA in the two respective transport media, (3) differential recovery of DNA during extraction, given differing starting material and potential chemical interactions or lysis steps, (4) Illumina MiSeq run-specific differences (quality/error rates) as shown in **Supplementary Figure S1**. Though differences could have also occurred in PCR amplification, this is less likely given the similar end results from triplicate reactions.



Supplementary Figure S1. MiSeq run-specific differences as visualised using NMDS plots with a Bray-Curtis distance matrix. On the right, RNALater reflects guano samples, Swab indicates cloacal swab sample types, and N/A represents water controls performed alongside all sample amplification and sequencing steps.

In terms of storage media, Amies transport medium is intended for processing of specimens containing aerobes, anaerobes and fastidious bacteria, though it was designed for rapid processing (<24-48 hours) in culture (Copan eSwab™ package insert). RNAlater® on the other hand preserves DNA, RNA, and tissue for genetic profiling (Qiagen package insert). A separate study performed on these same swab samples after transport and freezing at -20°C for 6 months still succeeded at extracting novel DNA viruses. Genomic DNA for this study was extracted from swabs while in the field (within 7 days of collection), so although the swabs would not have maintained viability for culture, it is less likely that significant bacterial DNA degradation would have occurred.

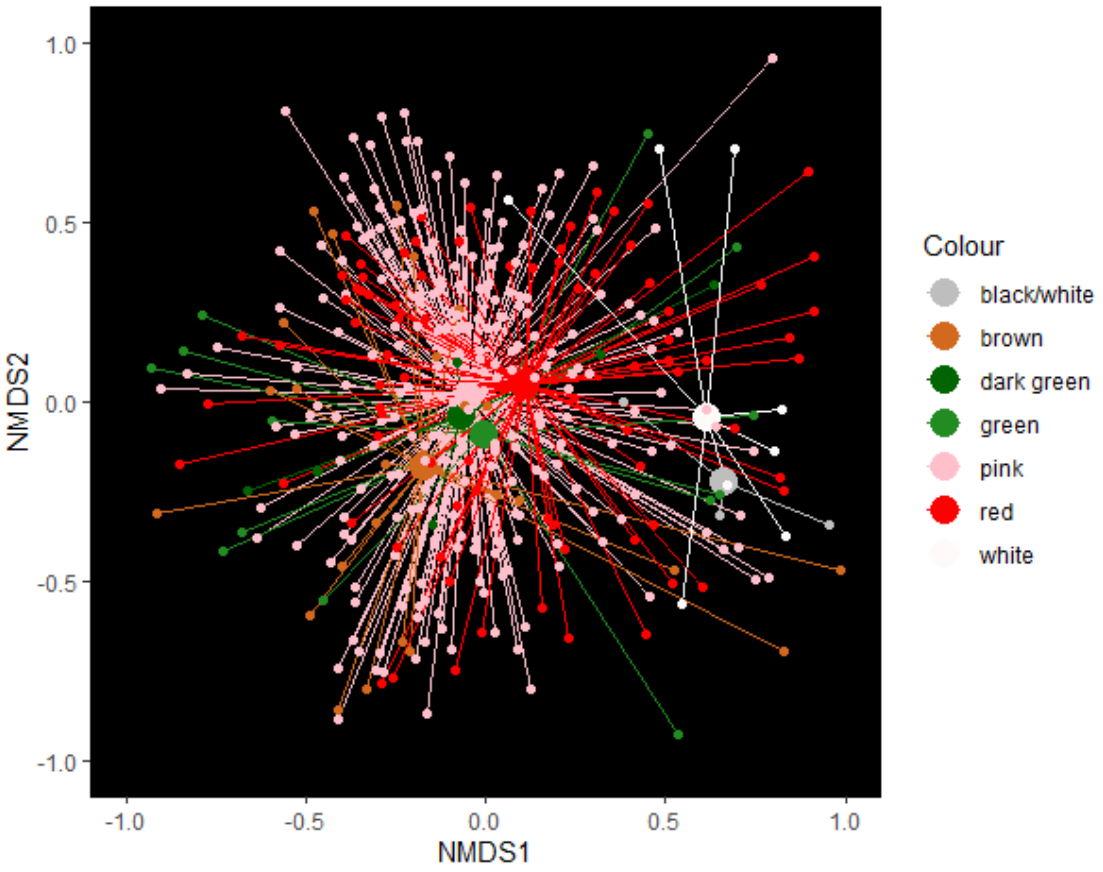
This study's design cannot fully account for whether any of the aforementioned mechanisms are in play, nor to what extent they may each influence the communities described. However, aside from the evident run-specific differences, it is also likely that the inherent composition of the microbiota obtained from the two methods will differ. Studies of differences along the avian gastrointestinal tract have shown that caution is warranted in generalising diversity from particular sample types to the full gut. Studies comparing the effect of sampling differences in seabird microbiomes or wild species are lacking. Recent literature indicates that in chickens, there are quantitative differences in structure and diversity between the microbiome obtained from cloacal sampling relative to caecal sampling (which requires euthanizing and dissecting the bird), though over 99% of sequences were shared across sample types (Stanley *et al.* 2015). A separate study in ostriches that directly compared faecal samples to cloacal swabs as well as three parts of the gut (ileum, caecum, and colon), found that faeces was actually more representative of the colon than cloacal swabs (Videvall *et al.* 2018).

Videvall et al. also contend that the cloaca cannot be truly representative of the gastrointestinal lumen for several theoretical reasons: (1) the cloaca is not a wholly anaerobic setting, given that it is the singular opening for the digestive, urinary, and reproductive tract; (2)

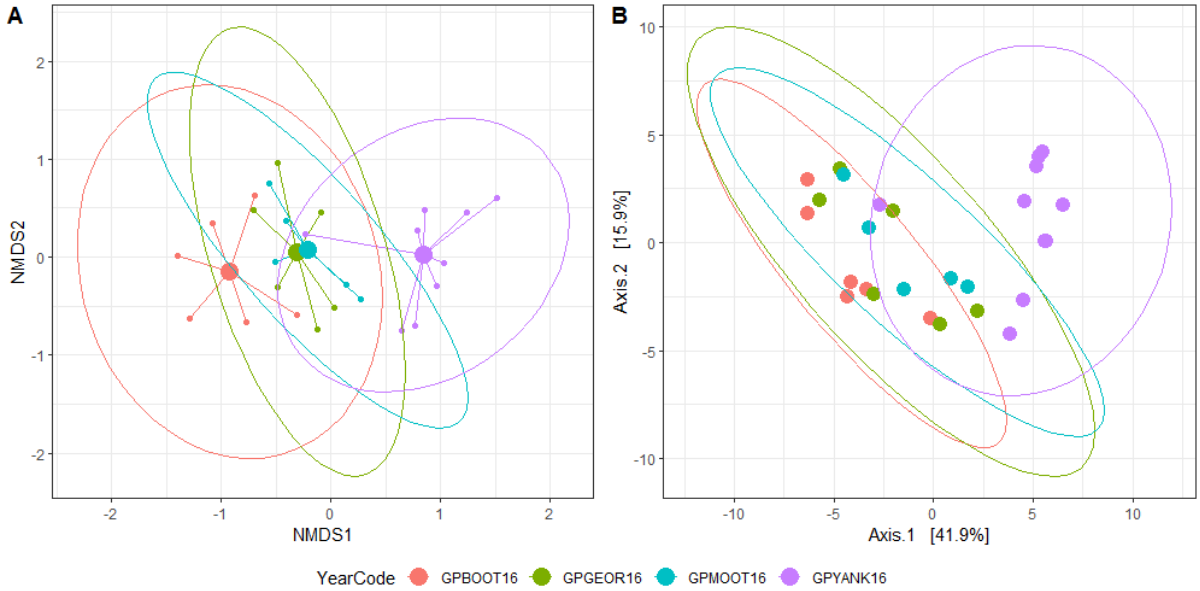
avian cloacae host the Bursa of Fabricius, an immune organ where B lymphocytes and antibodies are produced, such that cloacal microbiota are in contact with the lymphoid system; (3) reproductive transmission of microbes between mates can occur during copulation. While previous studies were performed in species of commercial interest in captive settings, these differences may still be relevant in wild penguins and would suggest that faecal samples will have greater abundance of anaerobic bacteria than cloacal samples. Though we cannot justifiably compare the differences statistically across sample types within a host species, there are trends in relative abundances we can qualitatively describe that would support this hypothesis (see "Bacterial Composition" in **Chapter 7** by species with both sample types). Rather than focus on whether our samples are representative of the gastrointestinal tract, we are more interested in whether we uncovered the microbial diversity and potential pathogens present in seabirds across their colonies.

Among faecal samples, sample freshness and substrate have been known to influence DNA amplification success in an albatross dietary study (McInnes *et al.* 2017), though less is known about how these variables would influence microbiome results. Though sampling bias was avoided as much as possible by collecting scats with a variety of colours and from multiple portions of the colony, there is inherent variation in guano composition highly influenced by diet. Antarctic seabird guano is composed of three fractions that vary in colour and chemical makeup: a white phase with nitrogen metabolism byproducts (urates), oxalates, and phosphates; a red (carotenoid-tinted) phase of undigested chitinous krill shells; a green phase with proteins, lipids, and cholic acid as well as algal contents likely transferred from krill stomach contents (Grzesiak *et al.* 2020; Myrcha & Tatur 1991). Scats do not always contain all of these in equal proportions and can vary in colour and composition. Studies of the microbial differences associated with differing guano appearance have also not been conducted previously. In this case, colour descriptions were documented for each guano sample as metadata, allowing us to visualise community structuring. Though there was significant overlap, samples that were white or mostly white (n =

12) clustered somewhat (but note high stress of 0.279), though we found no other consistent structuring of bacterial communities by dominant guano sample colour (see **Supplementary Figure S2**).

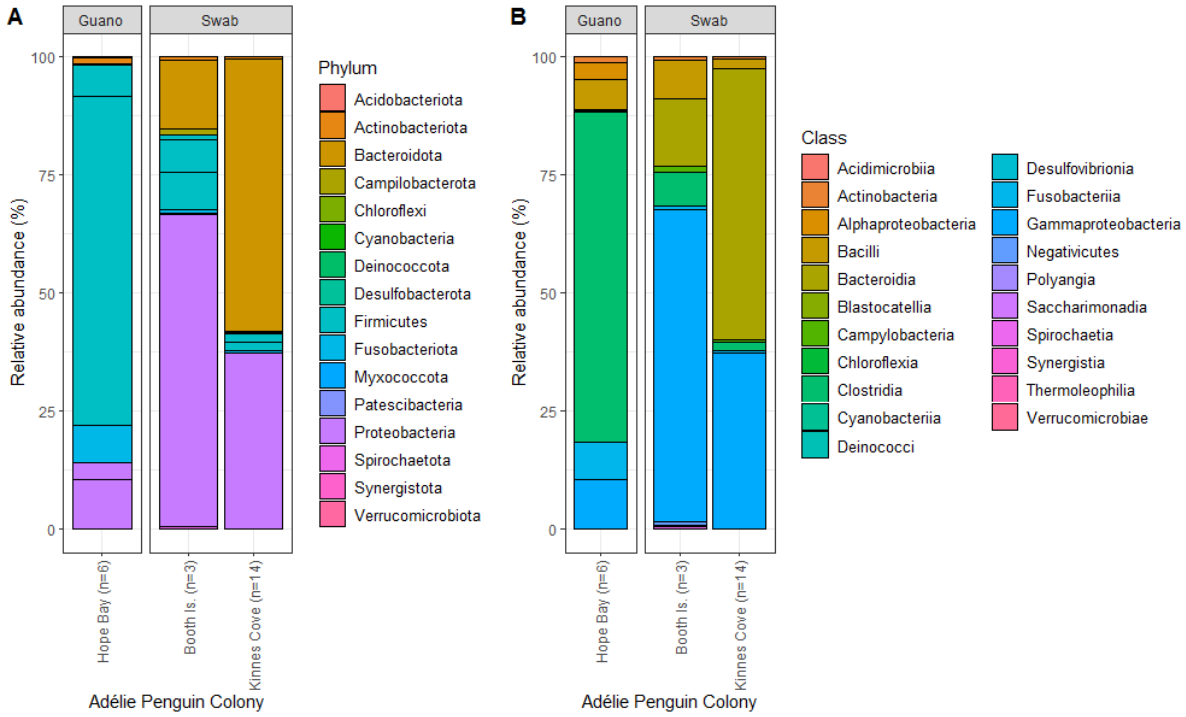


Supplementary Figure S2. NMDS plot (Bray-Curtis distance) of all guano samples by dominant sample colour. Larger circles indicate centroids of each colour group.



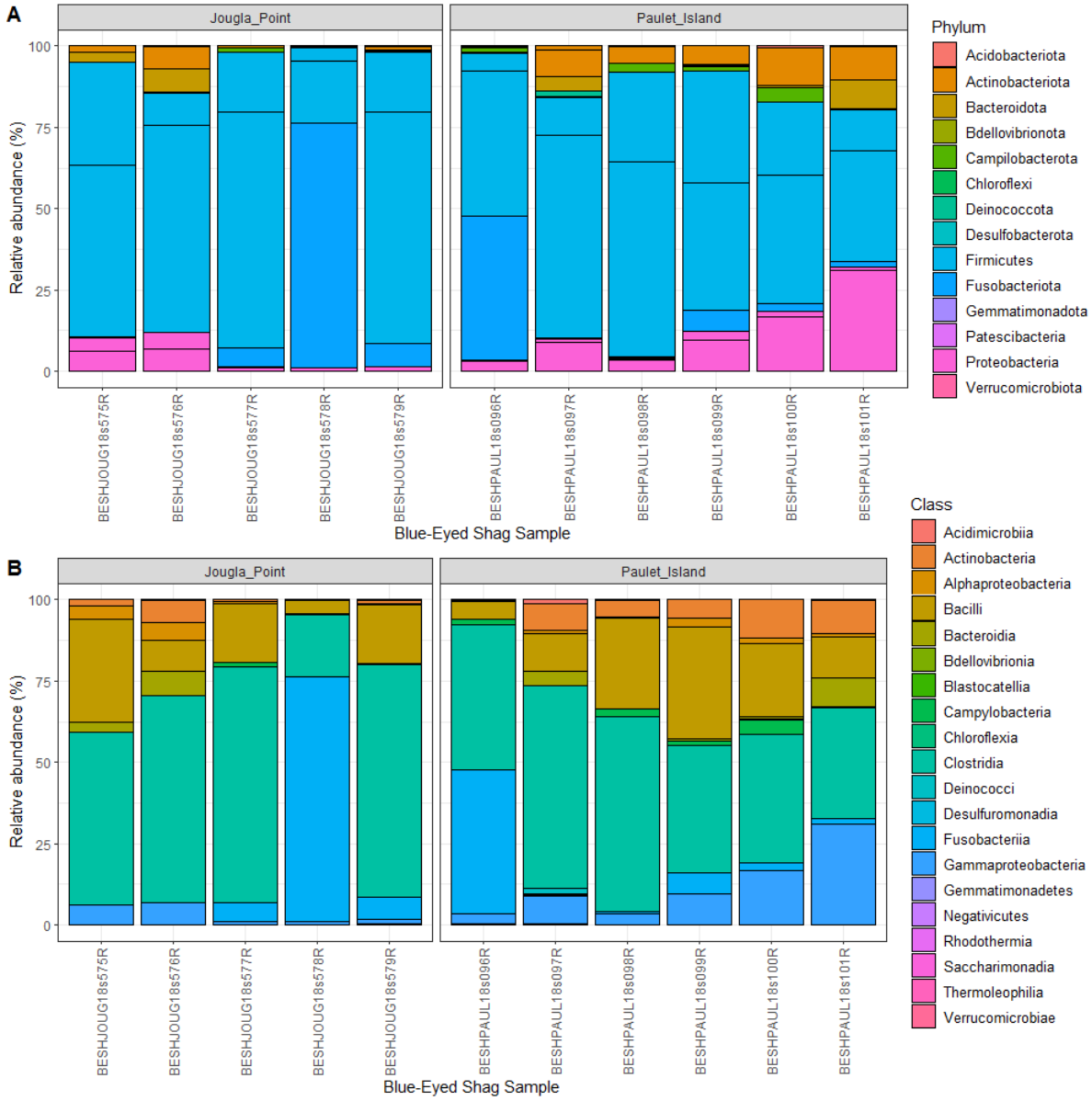
Supplementary Figure S3. Ordination plots of the total microbiome of Gentoo Penguin cloacal swab samples using (A) NMDS by Bray-Curtis dissimilarity and (B) PCoA with PhILR transformation, coloured by colony.

Adélie Penguin Microbiome

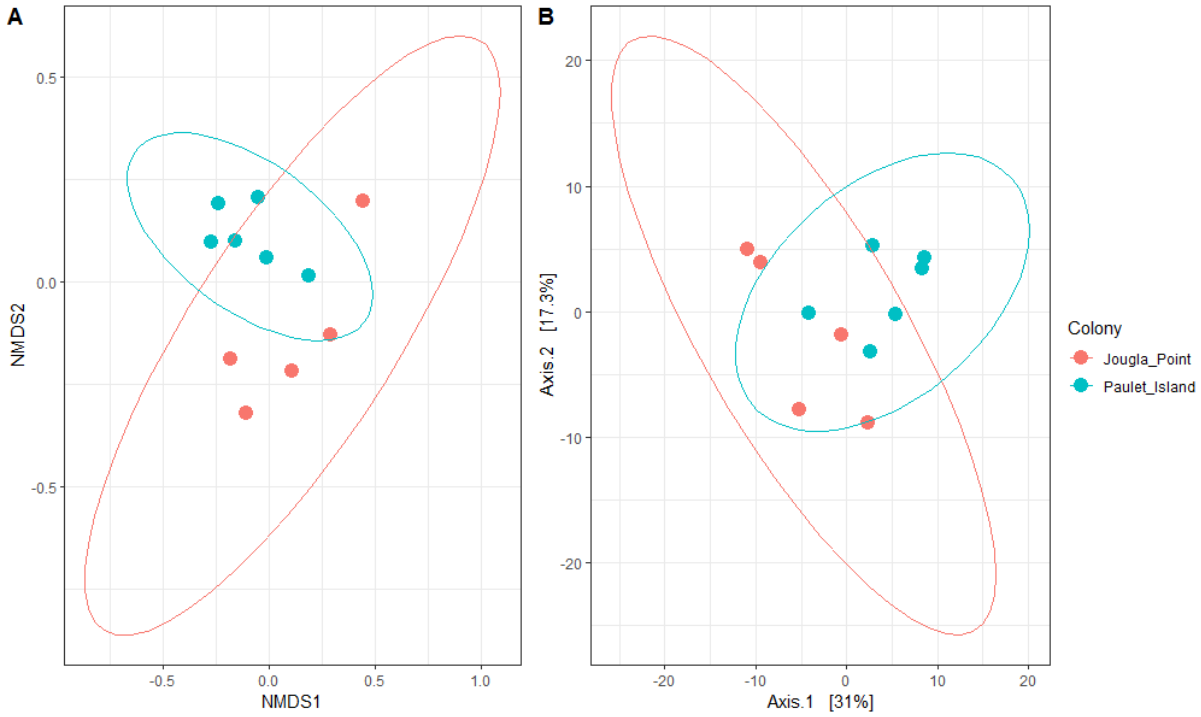


Supplementary Figure S4. Relative abundance of bacterial phyla (A) and classes (B) in Adélie Penguin samples, shown collectively by colony and differentiated by sample type (guano or swab).

Blue-Eyed Shag Microbiome

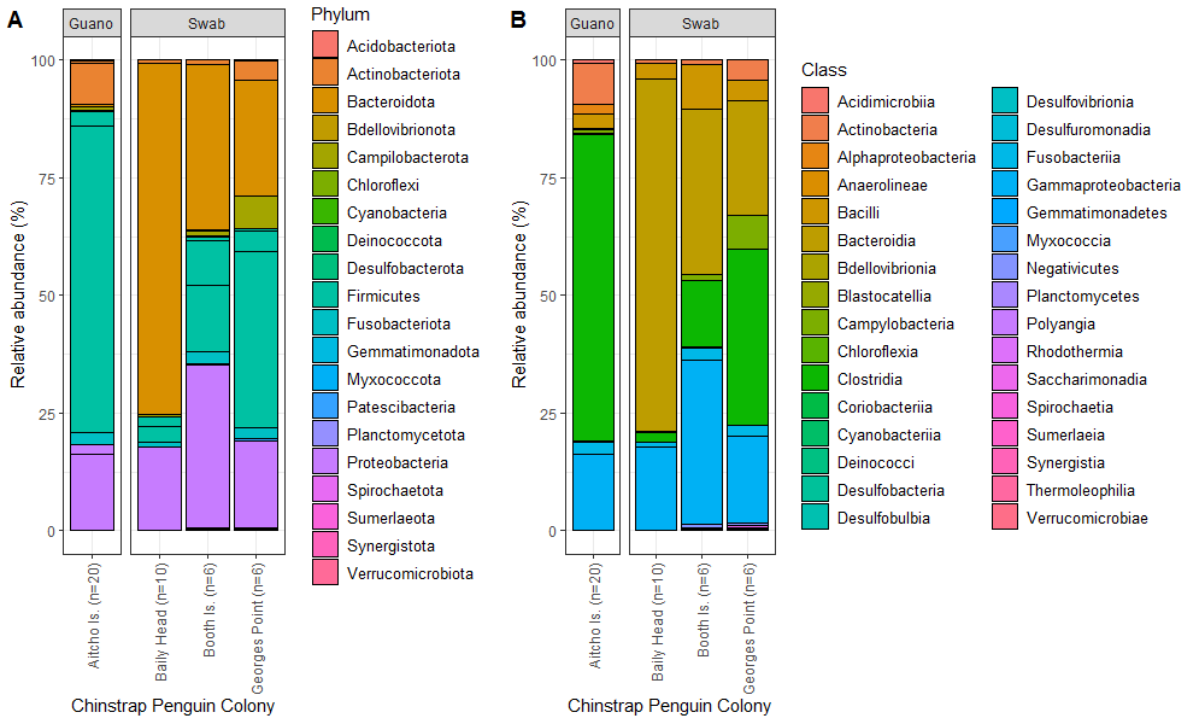


Supplementary Figure S5. Relative abundance of bacterial phyla (A) and classes (B) in Blue-Eyed Shag guano samples, shown individually and differentiated by colony.

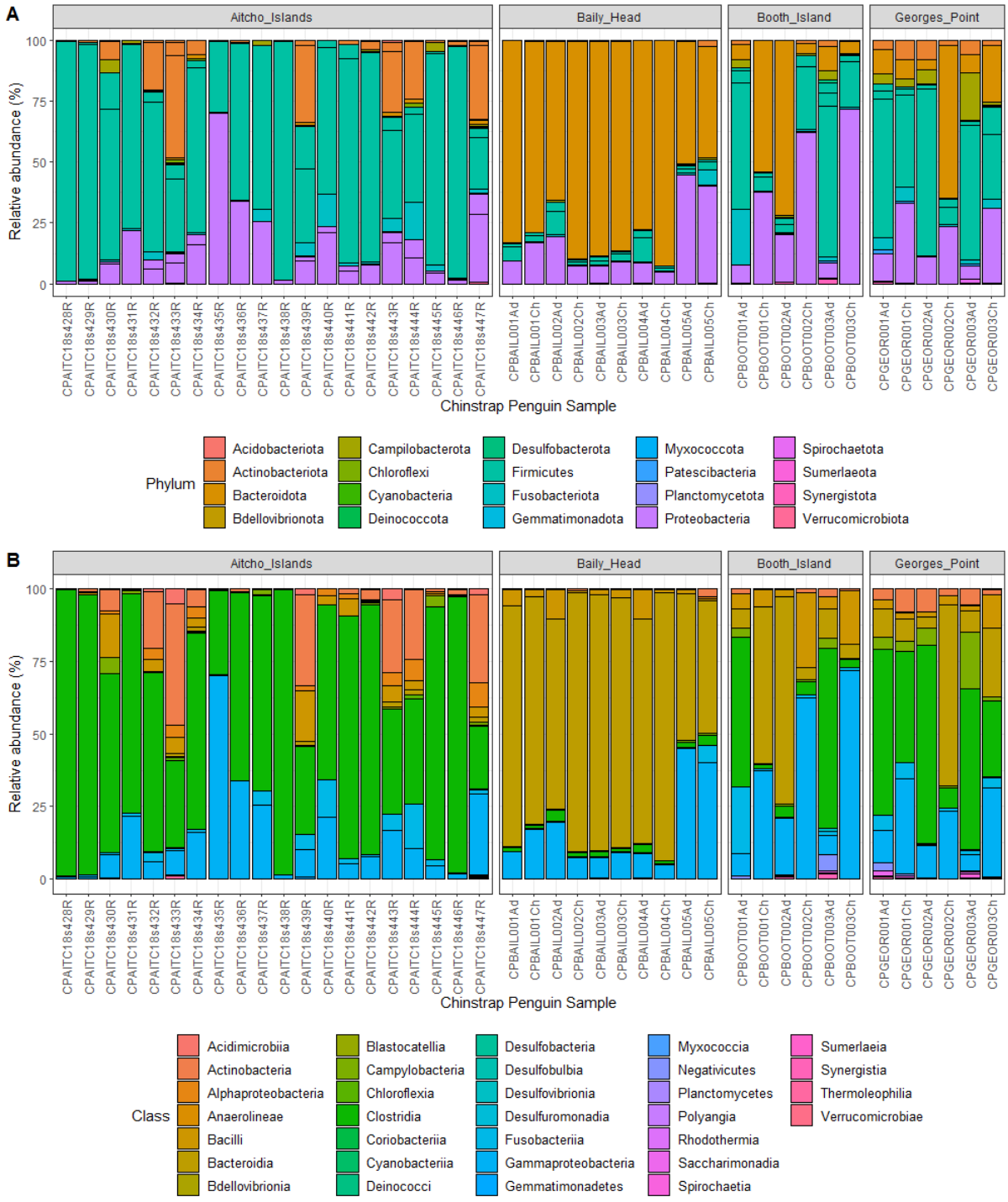


Supplementary Figure S6. Ordination plots of the total microbiome of Blue-Eyed Shag guano samples using (A) NMDS by Bray-Curtis dissimilarity and (B) PCoA with PhILR transformation, coloured by colony.

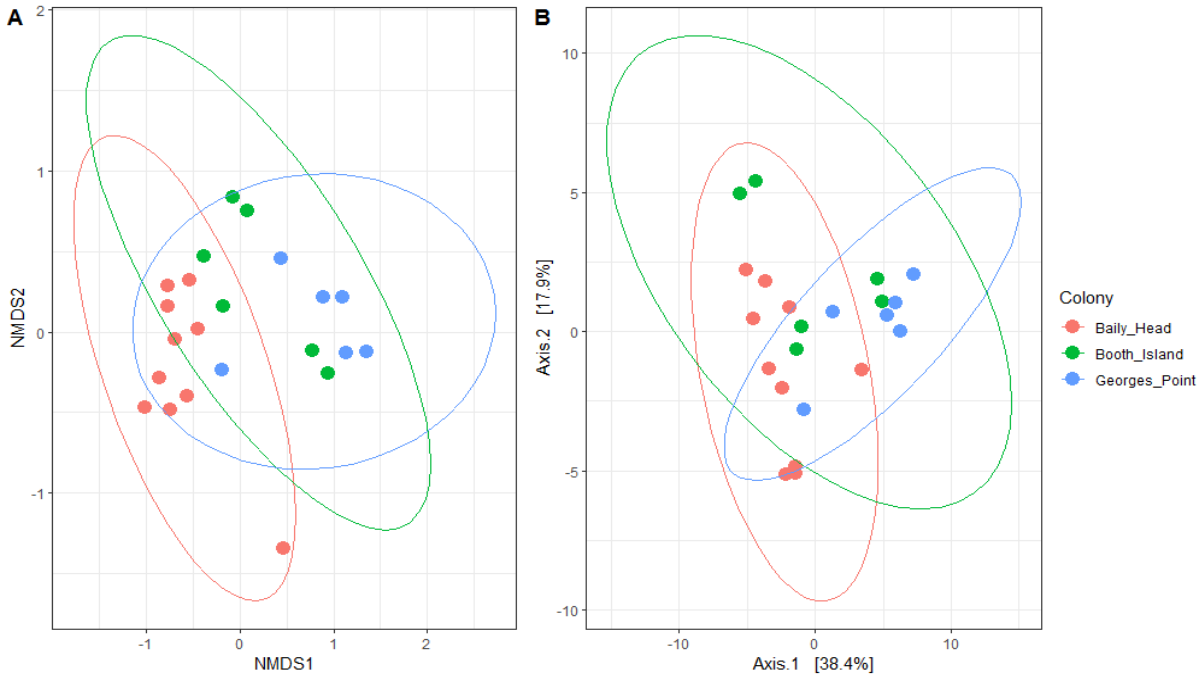
Chinstrap Penguin Microbiome



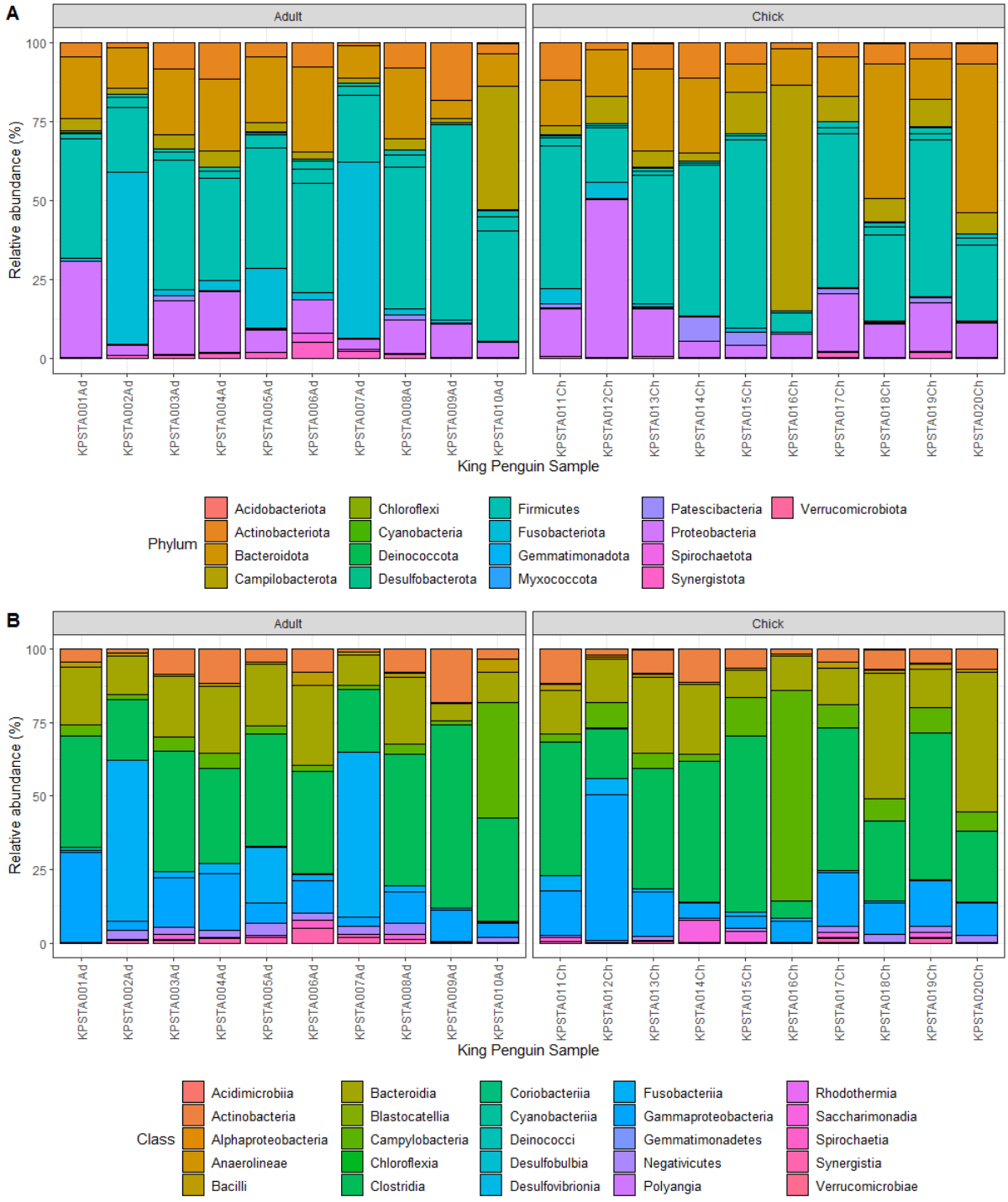
Supplementary Figure S7. Relative abundance of bacterial phyla (A) and classes (B) in Chinstrap Penguin samples, shown collectively by colony and differentiated by sample type (guano or swab).



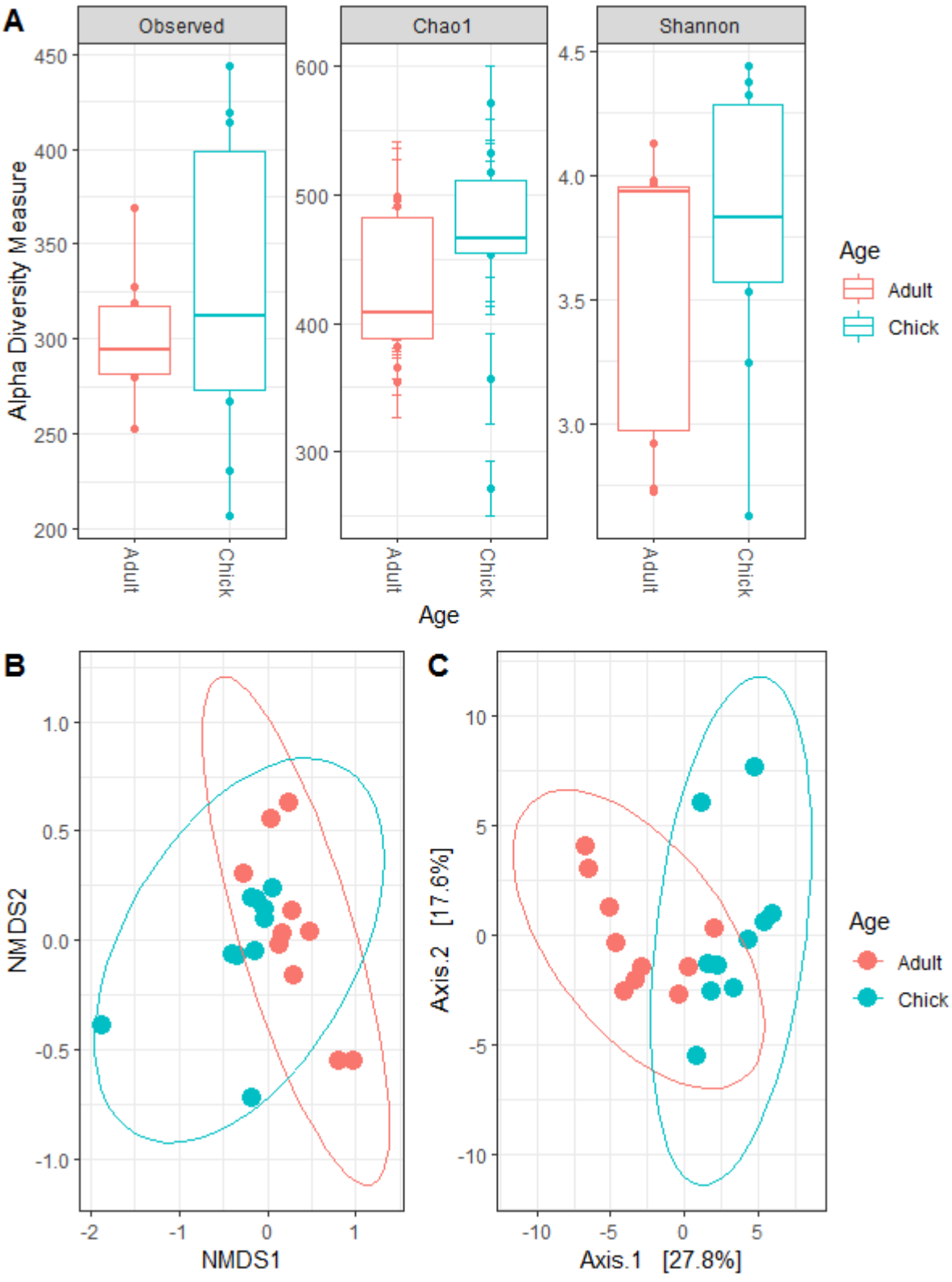
Supplementary Figure S8. Relative abundance of bacterial phyla (A) and classes (B) in Chinstrap Penguin samples. Aitcho Island samples (left) are guano, whereas all others depicted are cloacal swab samples.



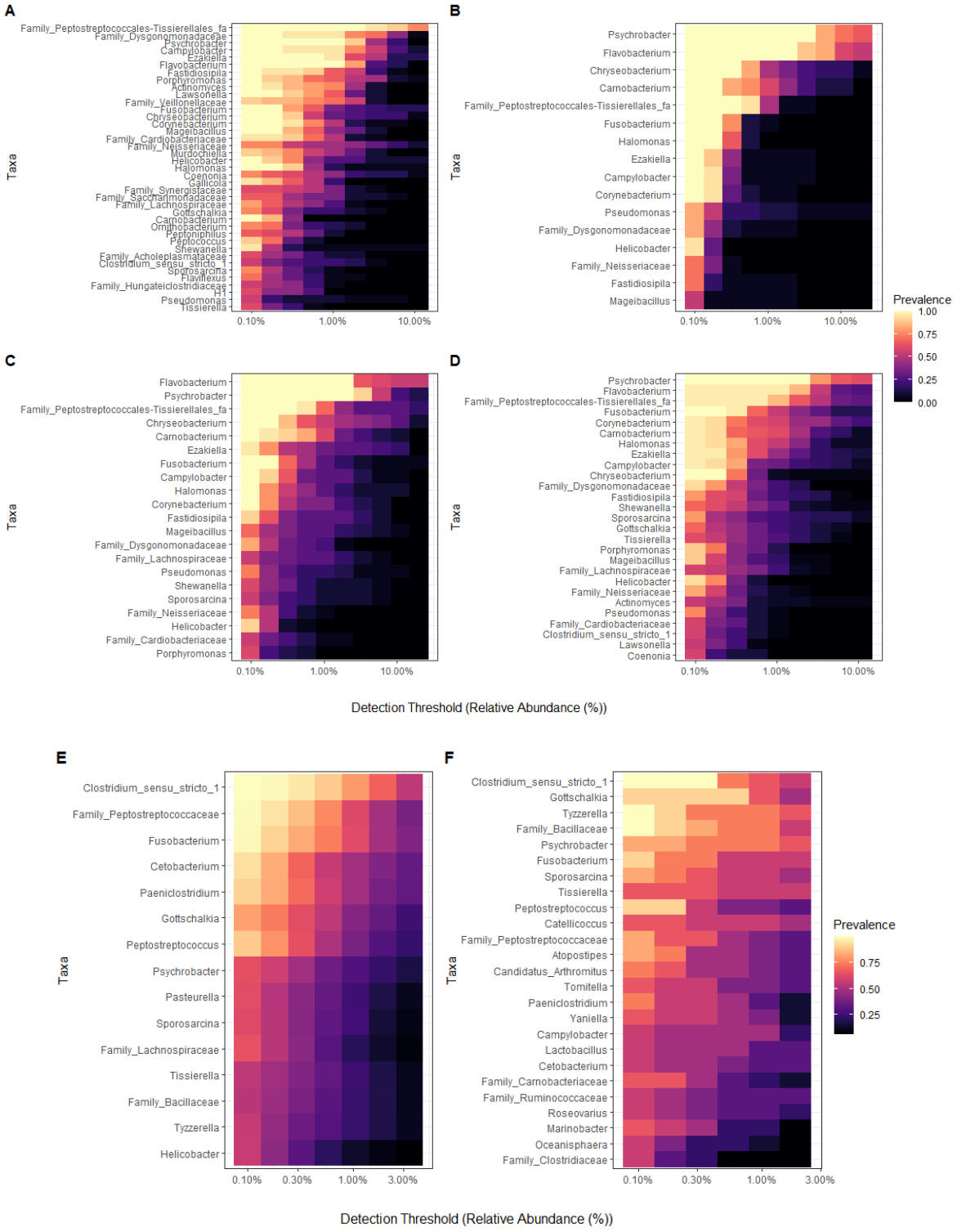
Supplementary Figure S9. Ordination plots of the total microbiome of Chinstrap Penguin cloacal swab samples using (A) NMDS by Bray-Curtis dissimilarity and (B) PCoA with PhILR transformation, coloured by colony.



Supplementary Figure S10. Relative abundance of bacterial phyla (A) and classes (B) in King Penguin cloacal swab samples collected at St Andrew's Bay, South Georgia. Adults and chicks were sampled independently.



Supplementary Figure S11.Alpha diversity measures (A) and ordination plots of the total microbiome of King Penguin cloacal swab samples using (B) NMDS by Bray-Curtis dissimilarity and (C) PCoA with PhILR transformation, coloured by age group (adult vs chick).



Supplementary Figure S12. Core Microbiome of swab samples taken from (A) King Penguins, (B) Adélie Penguins, (C) Chinstrap Penguins, and (D) Gentoo Penguins, and guano samples from (E) Gentoo Penguins and (F) Blue-Eyed Shags, depicting all taxa (genus-level) present at detection levels of ≥ 0.001% of relative abundance in at least 50% of those samples. Prevalence is depicted by the colour scale.

References for Appendix A5

Grzesiak J, Kaczynska A, Gawor J*, et al.* (2020) A smelly business: Microbiology of Adelie penguin guano (Point Thomas rookery, Antarctica). *Sci Total Environ* **714**, 136714.

McInnes JC, Alderman R, Deagle BE*, et al.* (2017) Optimised scat collection protocols for dietary DNA metabarcoding in vertebrates. *Methods in Ecology and Evolution* **8**, 192-202.

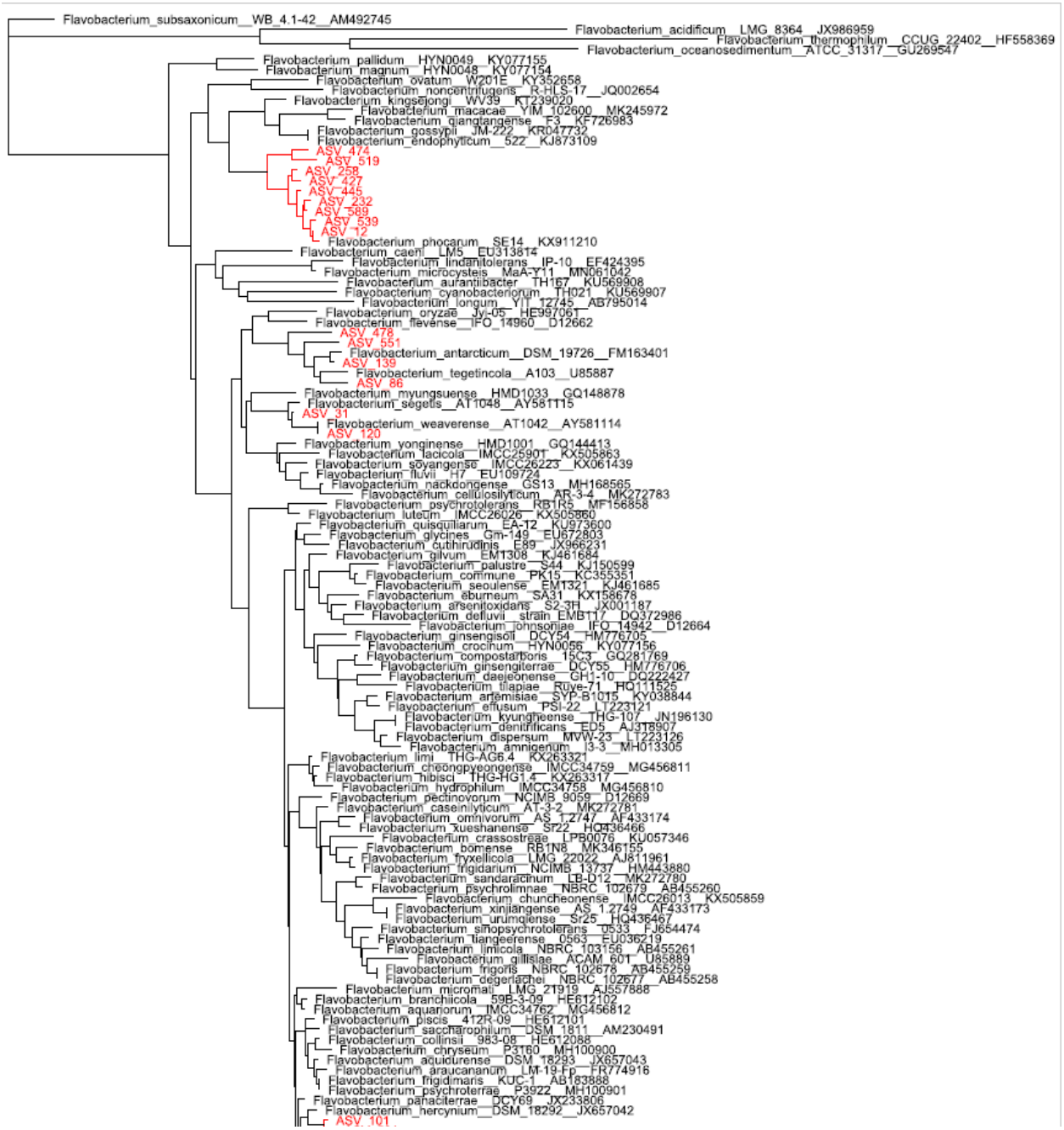
Myrcha A, Tatur A (1991) Ecological Role of the Current and Abandoned Penguin Rookeries in the Land Environment of the Maritime Antarctic. *Polish Polar Research* **12**, 3-24.

Stanley D, Geier MS, Chen H, Hughes RJ, Moore RJ (2015) Comparison of fecal and cecal microbiotas reveals qualitative similarities but quantitative differences. *BMC Microbiology* **15**, 51.

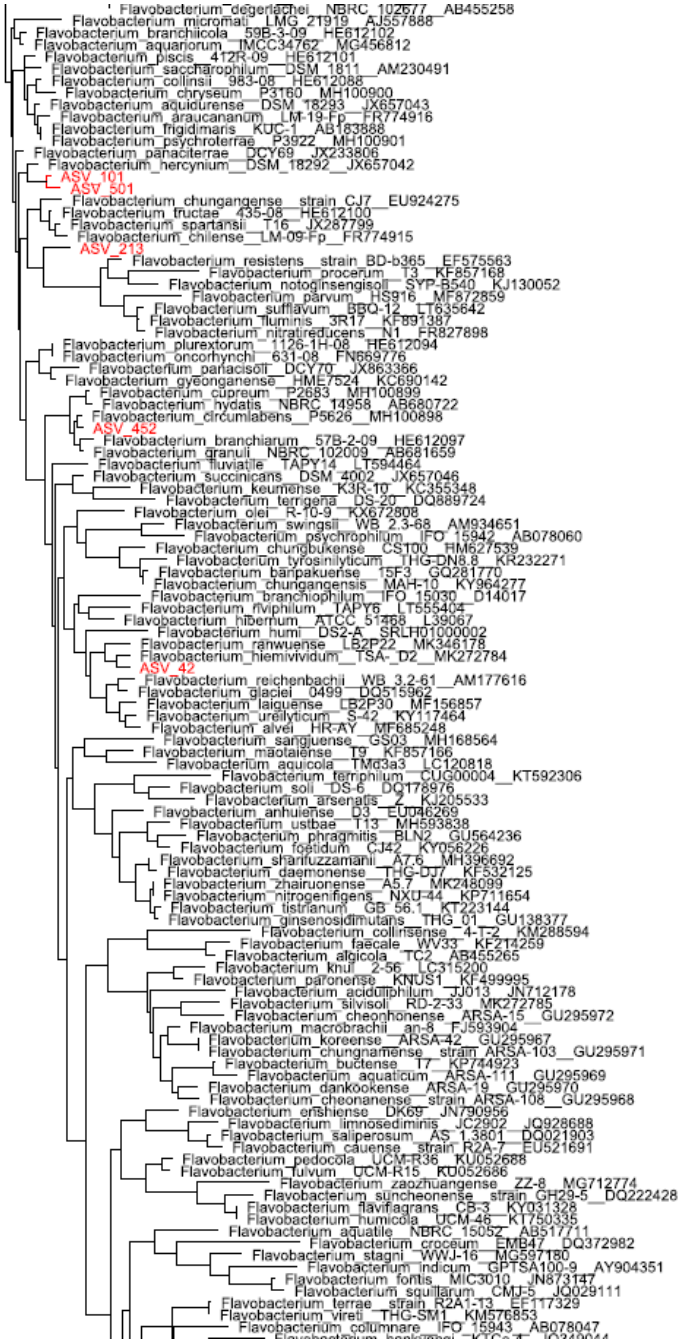
Videvall E, Strandh M, Engelbrecht A, Cloete S, Cornwallis CK (2018) Measuring the gut microbiome in birds: Comparison of faecal and cloacal sampling. *Mol Ecol Resour* **18**, 424-434.

Appendix A6 - Supplementary Information for Chapter 8

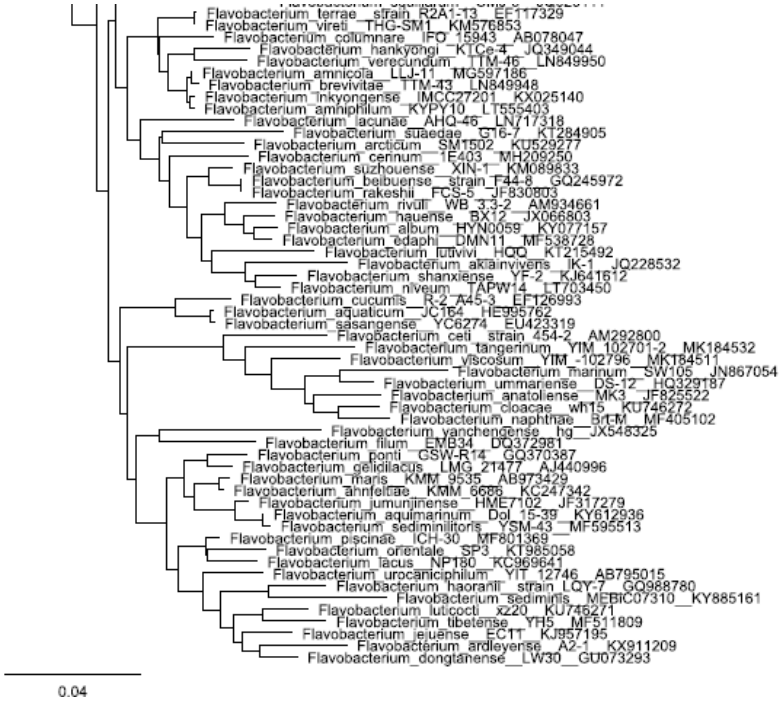
Supplementary Figure S1. Rooted neighbour-joining tree (HKY distance model) built from 16S reference sequences for *Mycobacterium* and emended genera (using a *Corynebacterium* outgroup), with sequence variants from this study shown in red.



S2.1, Continued…

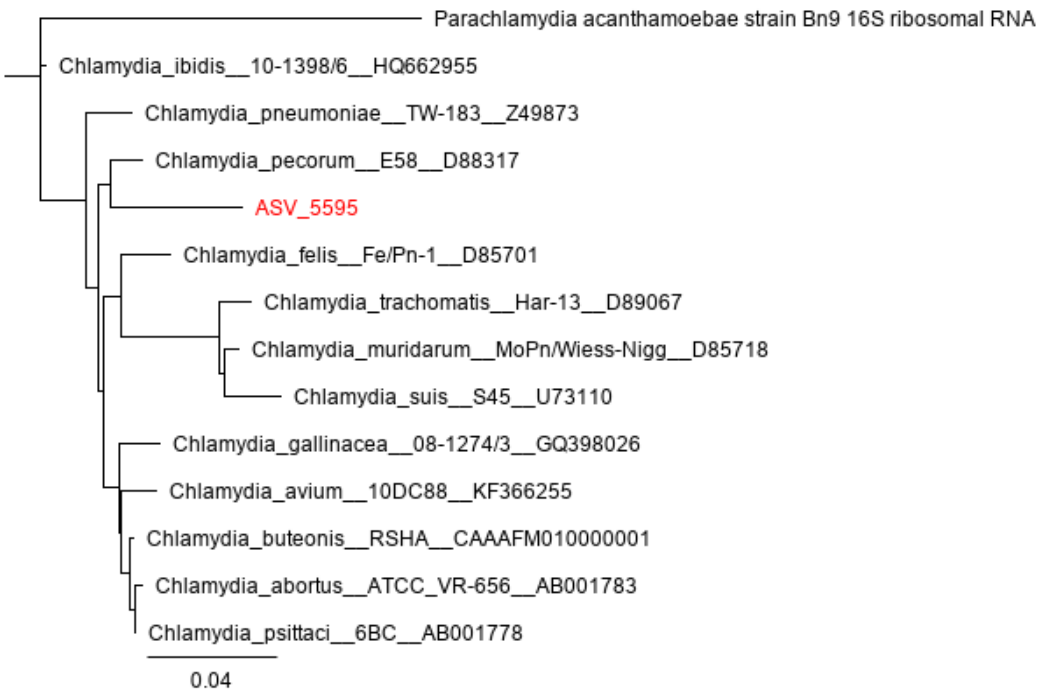


S2.2, Continued…

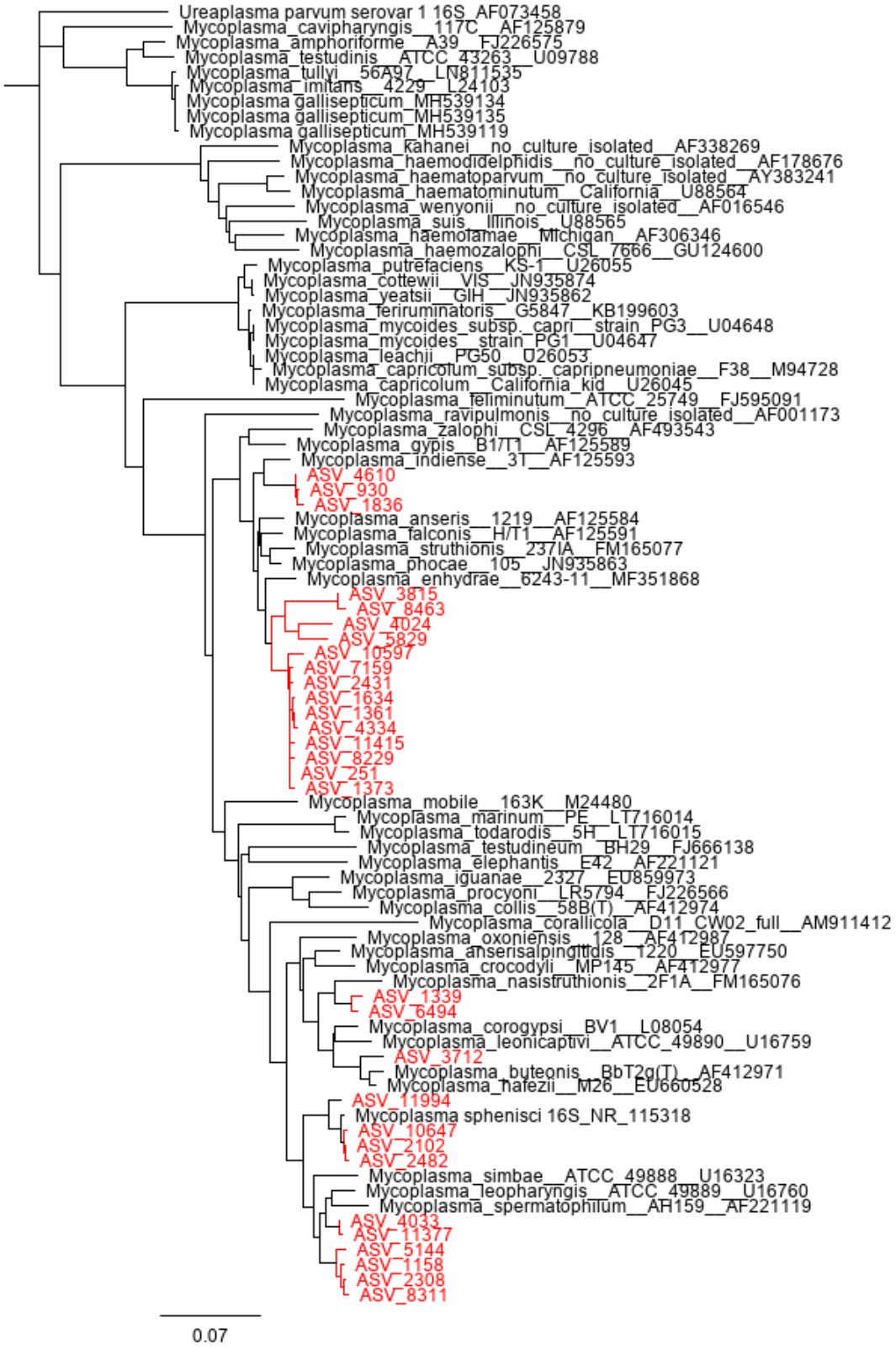


S2.3.

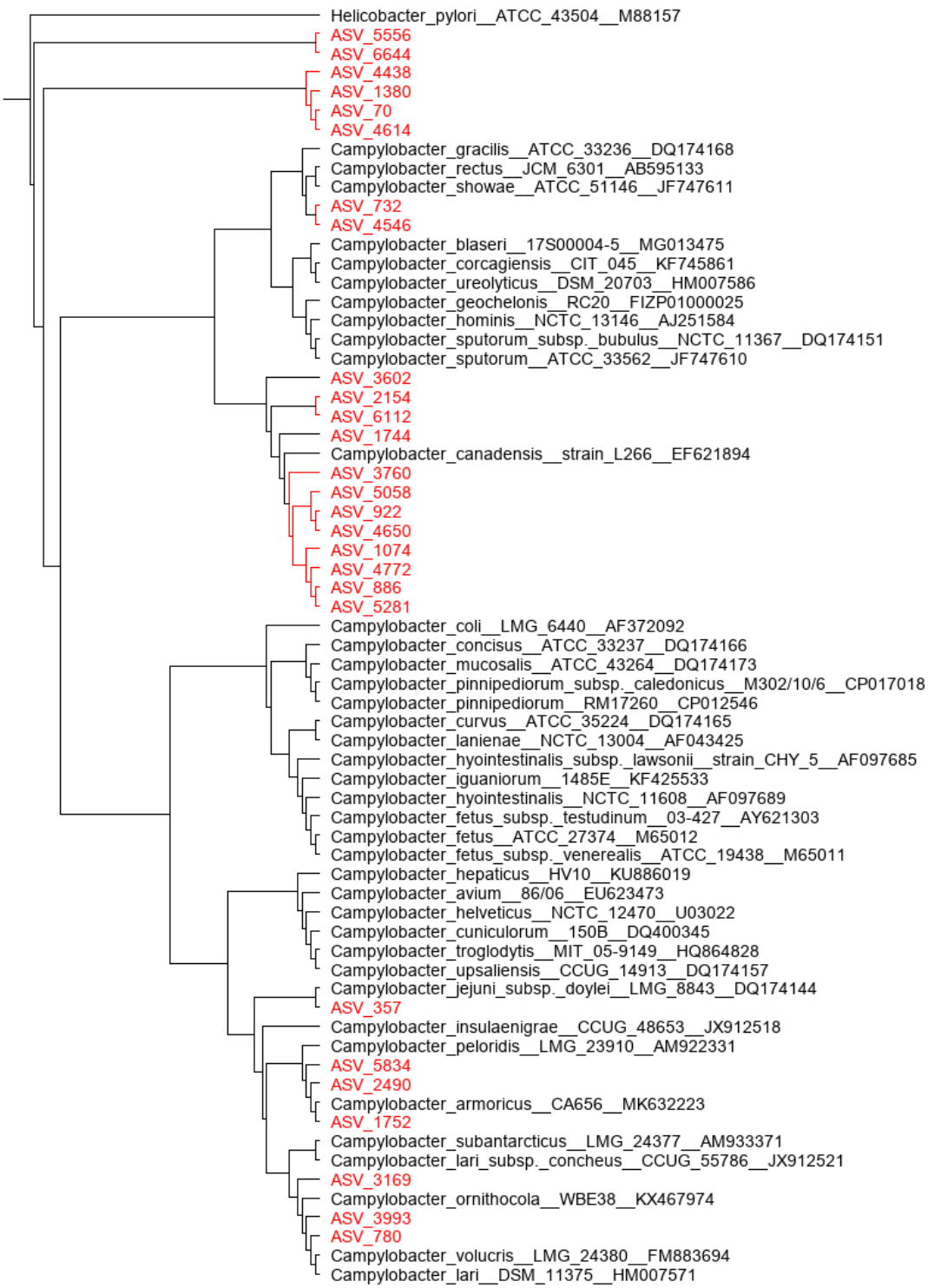
Supplementary Figure S2. Unrooted neighbour-joining tree (HKY distance model) built from 16S reference sequences for the genus *Flavobacterium*, with sequence variants from this study shown in red, continued across 3 pages.



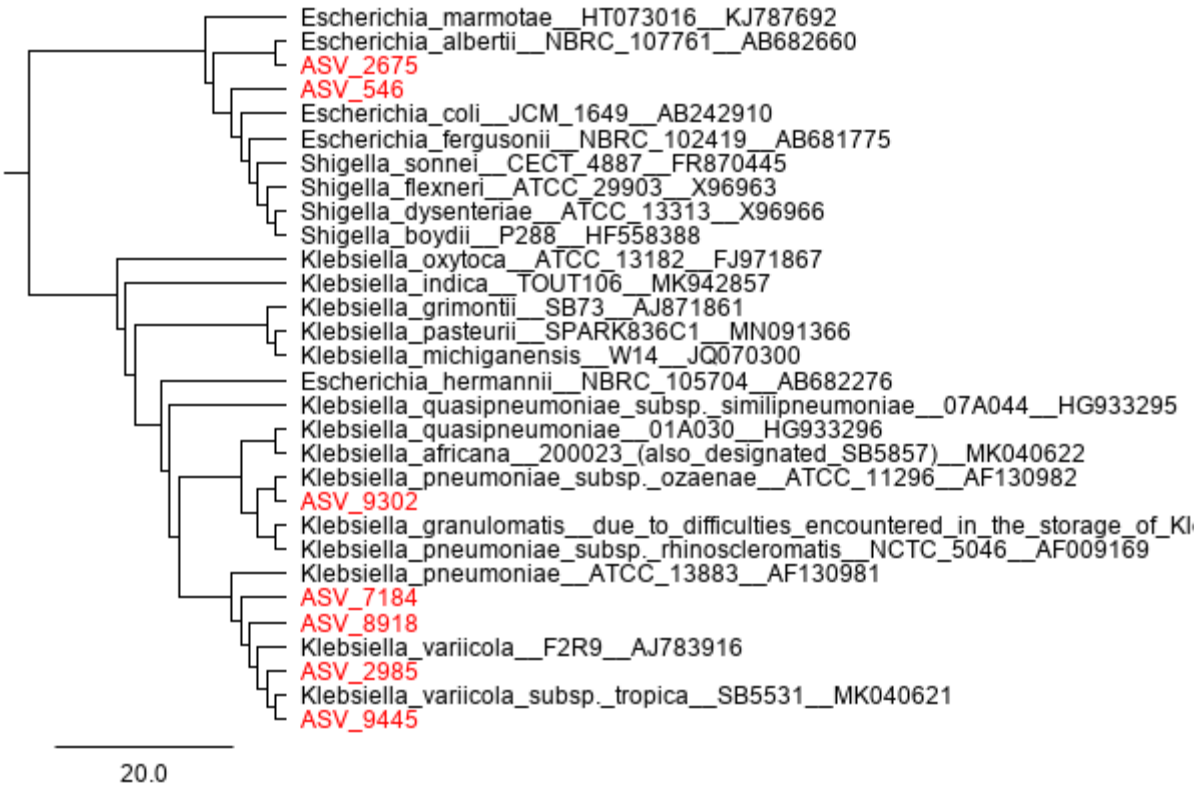
Supplementary Figure S3. Rooted neighbour-joining tree (HKY distance model) built from 16S reference sequences for the genus *Chlamydia* (using a *Parachlamydia* outgroup), with the single *Chlamydia* sequence variant from this study, found in a Gentoo Penguin from the Falkland Islands, shown in red.



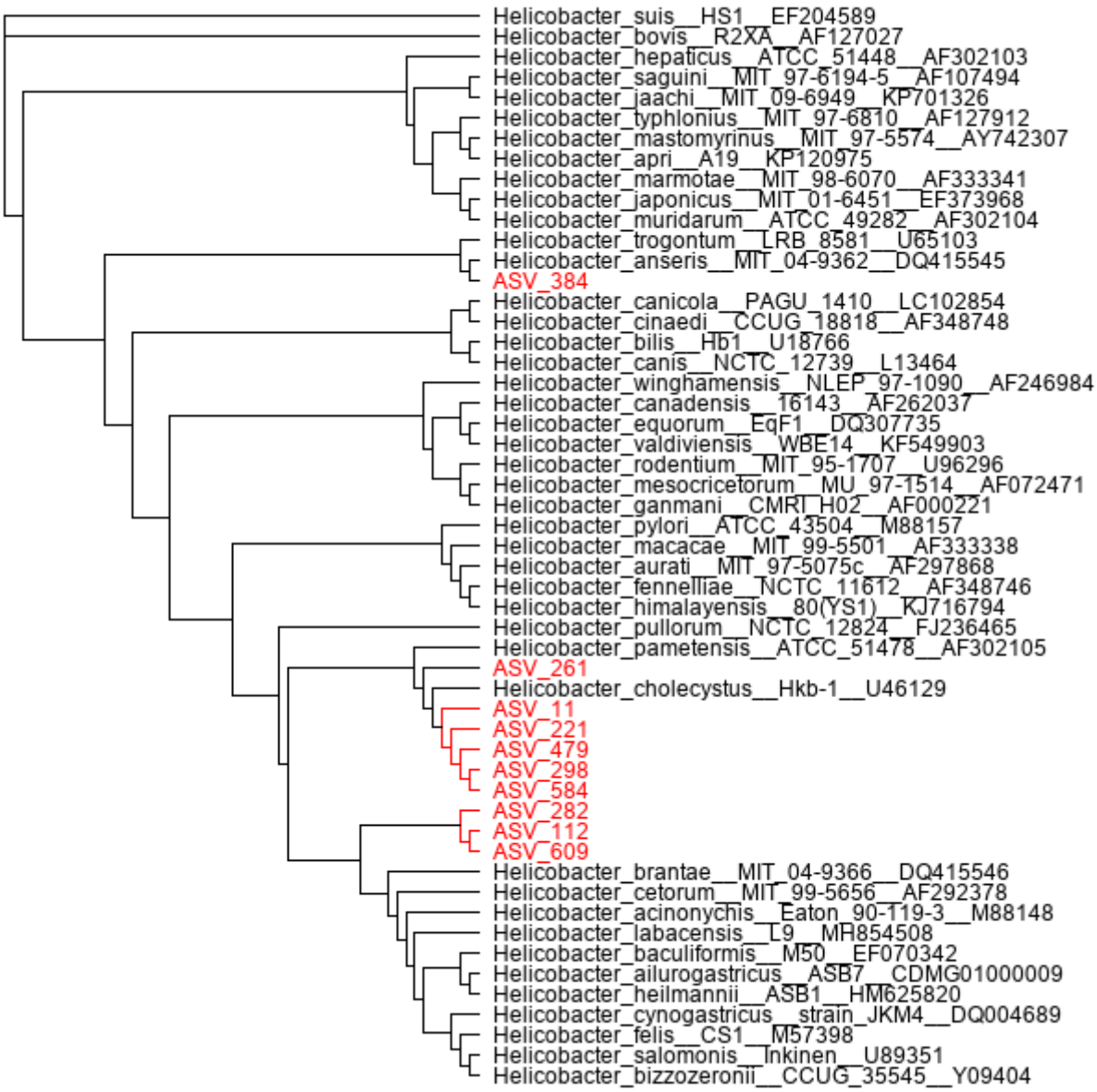
Supplementary Figure S4. Rooted neighbour-joining tree (HKY distance model) built from 16S reference sequences for the genus *Mycoplasma* (using a *Ureaplasma* outgroup), with sequence variants from this study shown in red.



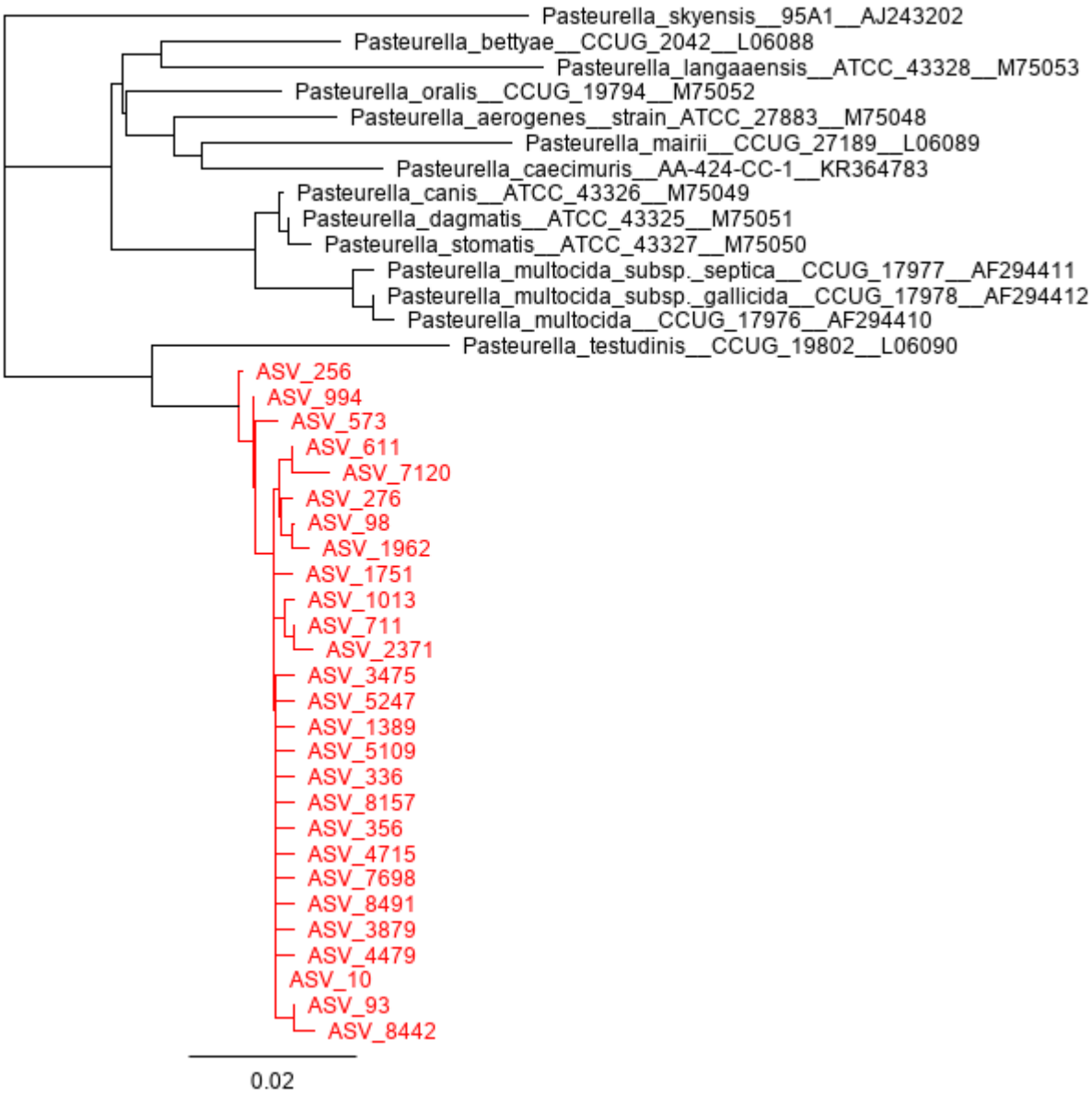
Supplementary Figure S5. Rooted neighbour-joining tree (HKY distance model) built from 16S reference sequences for the genus Campylobacter (using a Helicobacter pylori outgroup), with sequence variants from this study shown in red.



Supplementary Figure S6. Unrooted neighbour-joining tree (HKY distance model) built from 16S reference sequences for the genera *Escherichia, Shigella* and *Klebsiella*, with sequence variants from this study shown in red, nested within their assigned genera.

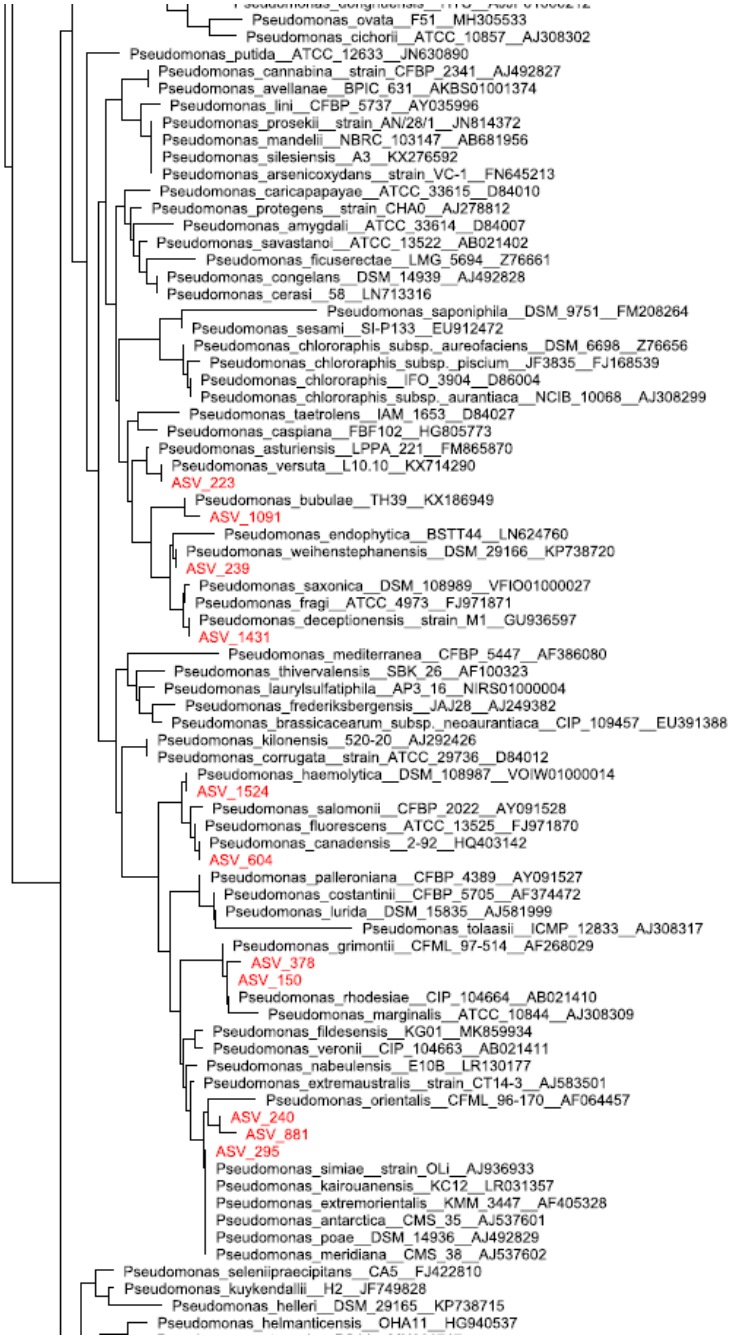


Supplementary Figure S7. Unrooted neighbour-joining tree (HKY distance model) built from 16S reference sequences for the genus *Helicobacter*, with the top ten most abundant sequence variants of this genus found in this study shown in red.

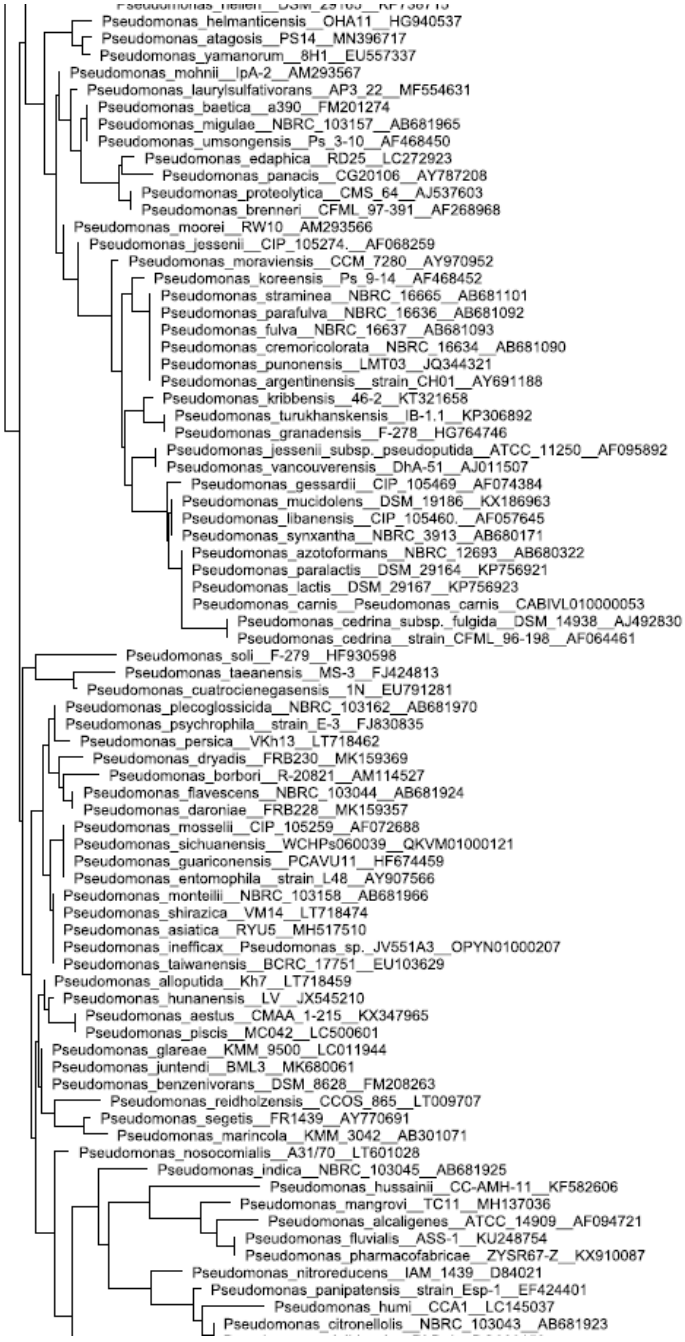


Supplementary Figure S8. Unrooted neighbour-joining tree (HKY distance model) built from 16S reference sequences for the genus *Pasteurella*, with the sequence variants of this genus found in this study shown in red.

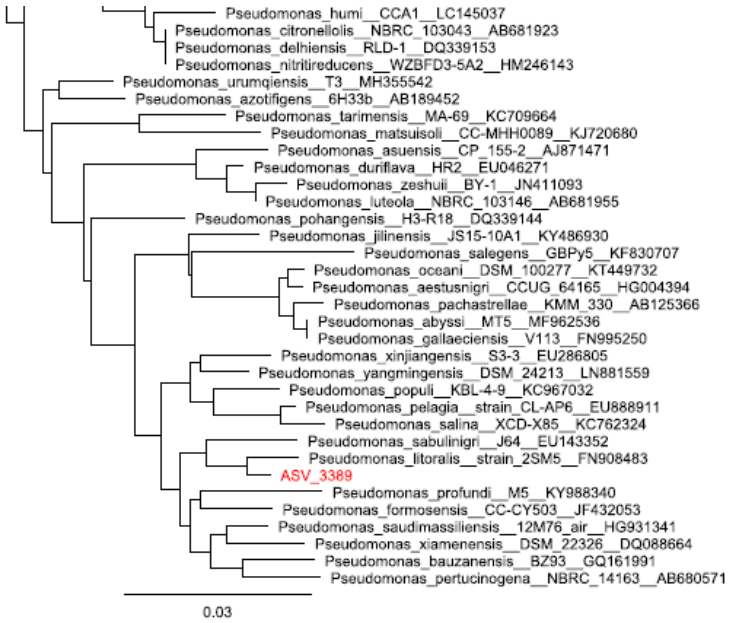
S9.1, Continued…



S9.2, Continued…

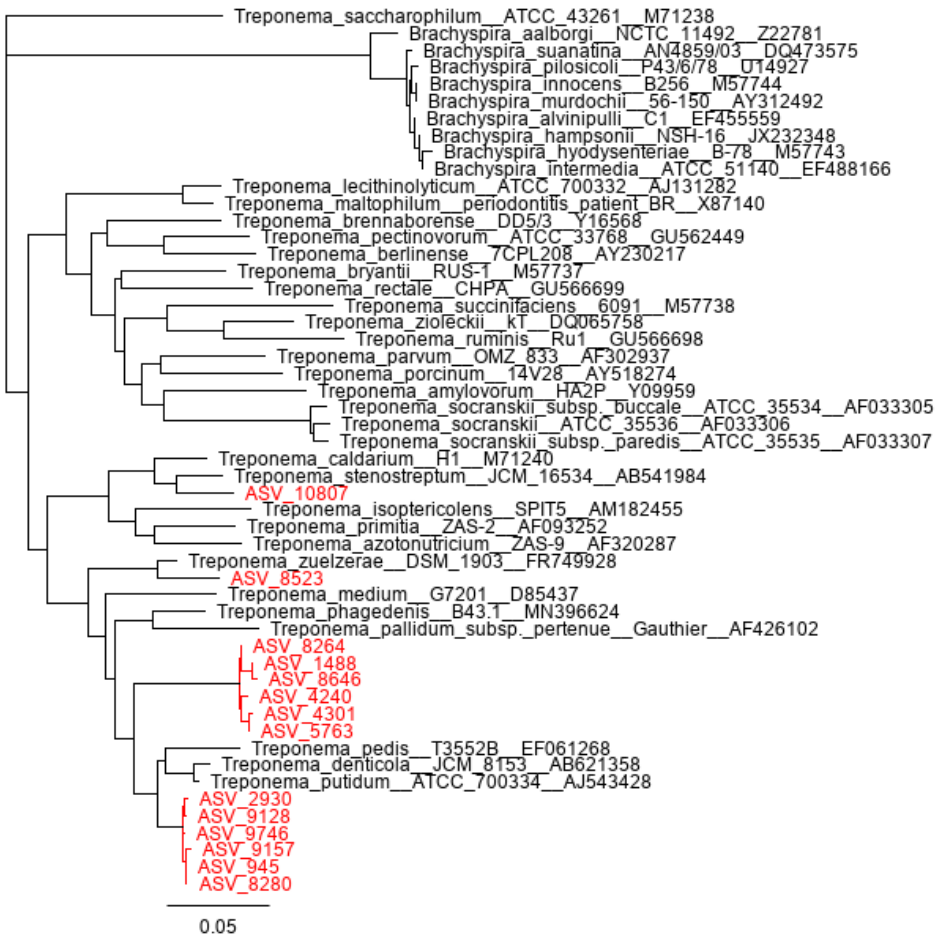


S9.3, Continued…



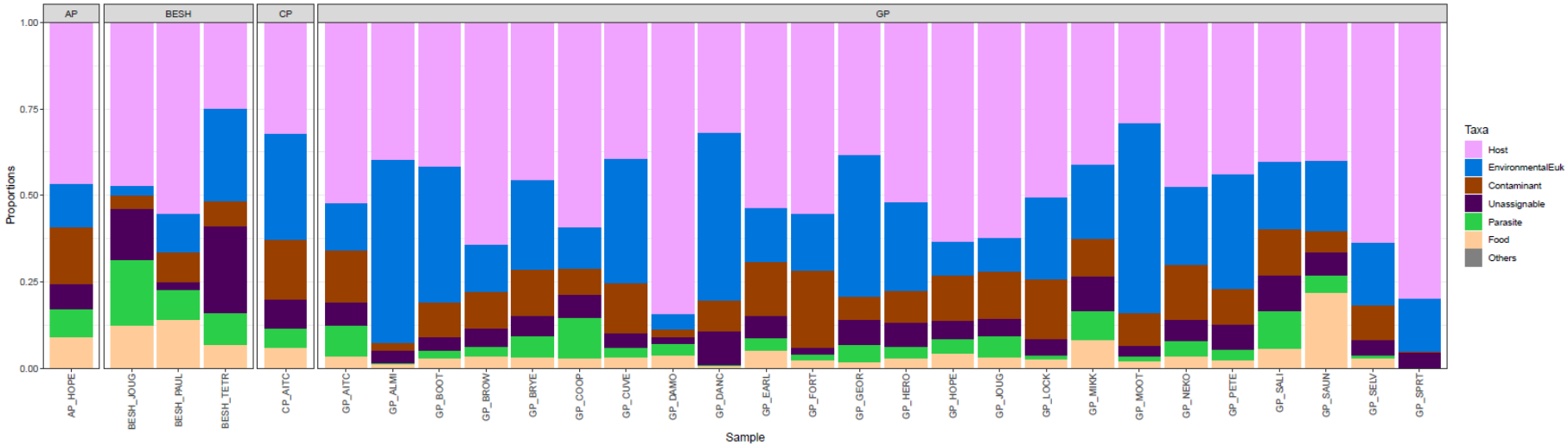
S9.4

Supplementary Figure S9. Unrooted neighbour-joining tree (HKY distance model) built from 16S reference sequences for the genus *Pseudomonas*, with sequence variants from this study shown in red (tree is continued across 4 pages).

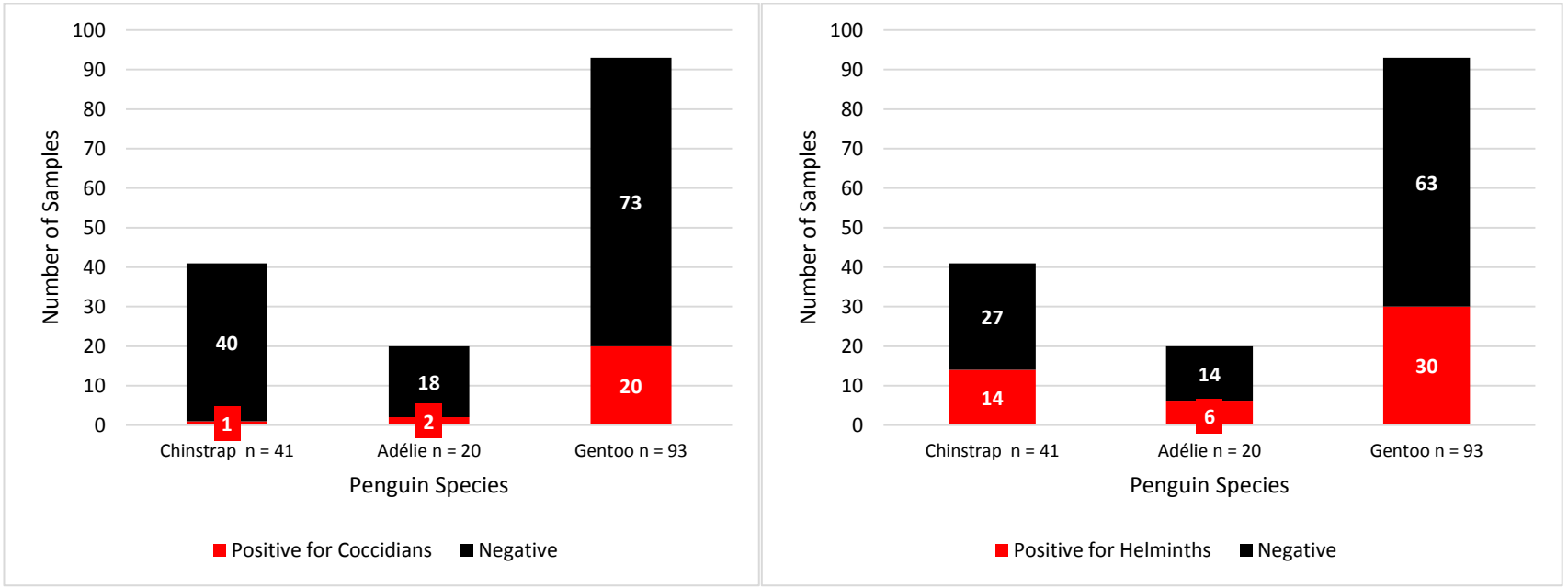


Supplementary Figure S10. Unrooted neighbour-joining tree (HKY distance model) built from 16S reference sequences for the genera *Treponema* and *Brachyspira*, with sequence variants from this study shown in red.

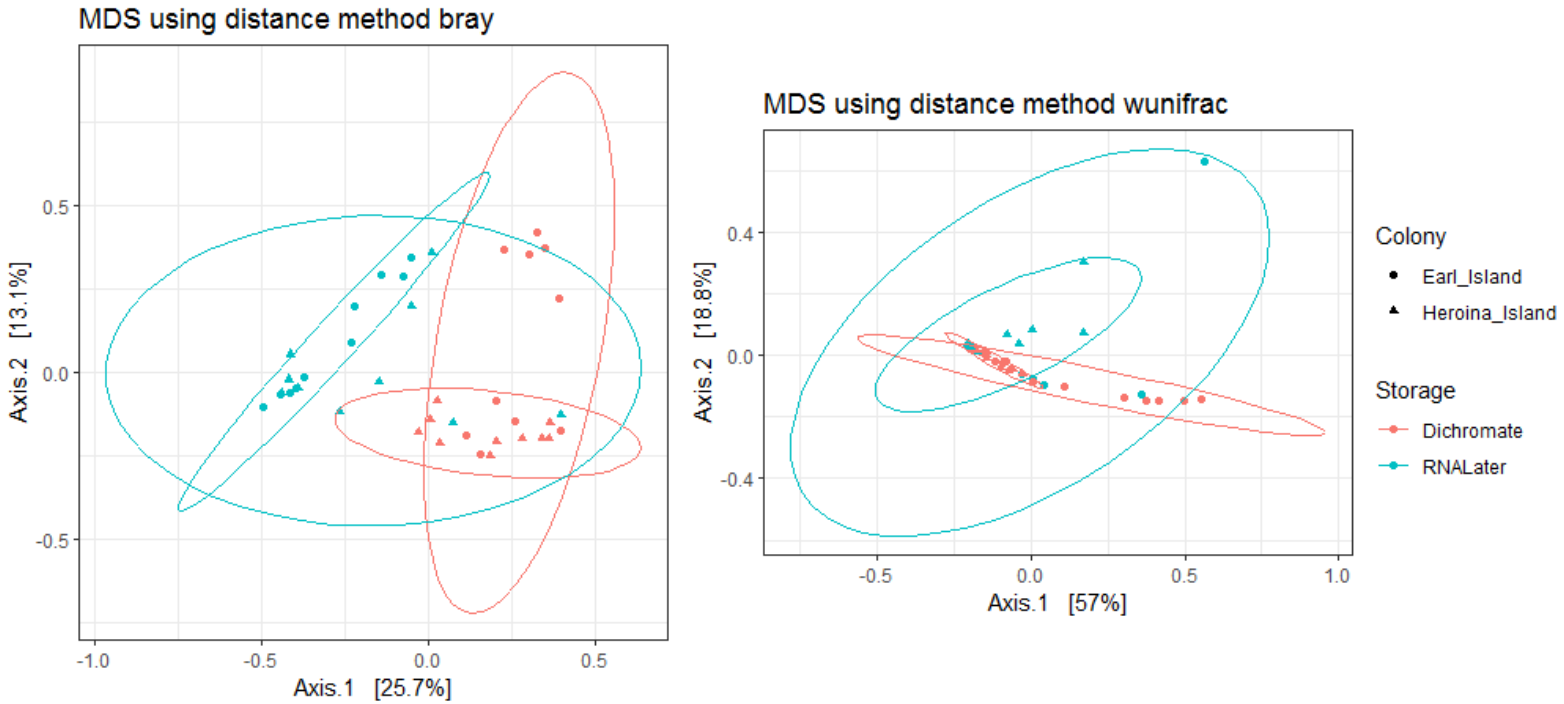
Appendix A7 – Supplementary Information for Chapter 9



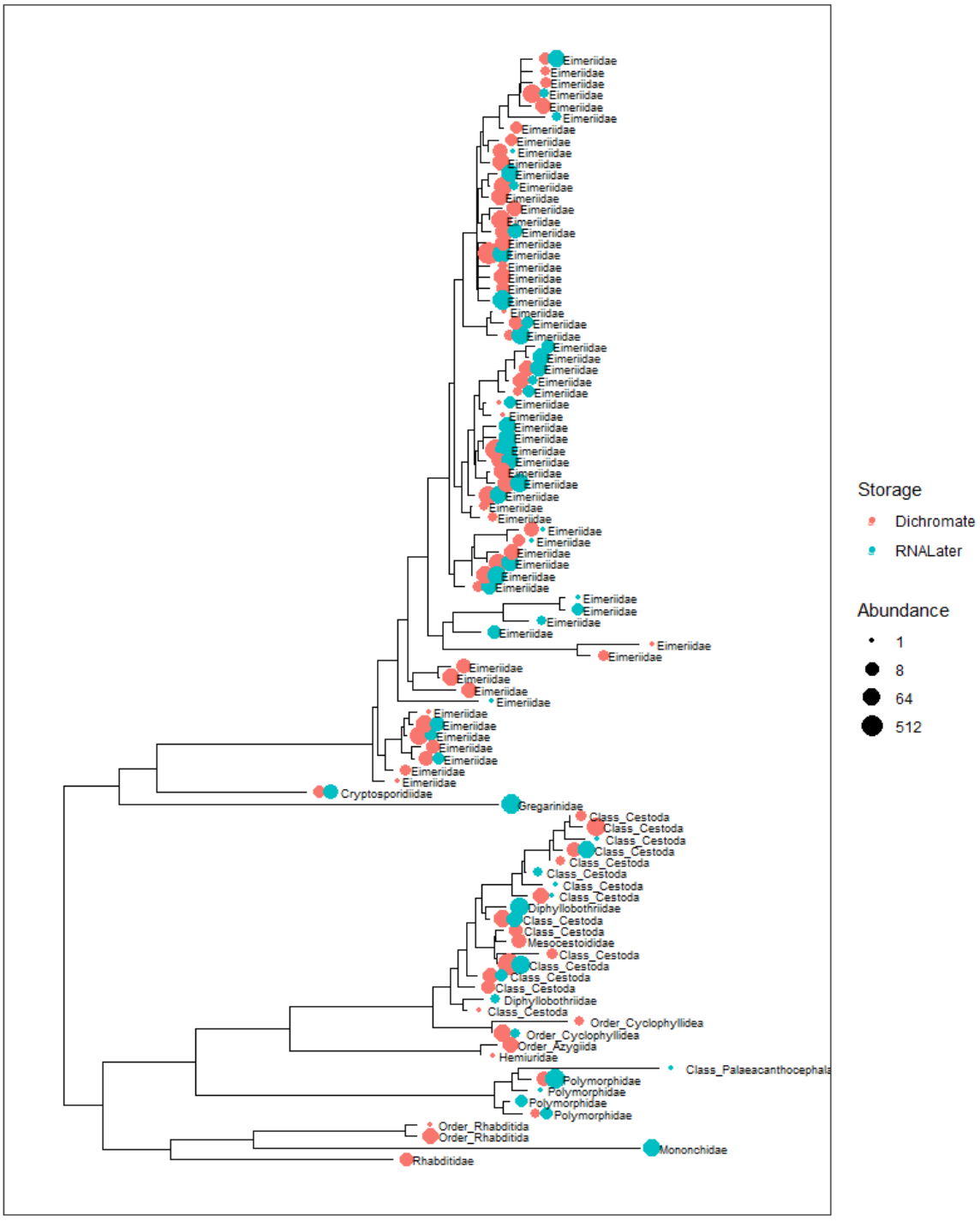
Supplementary Figure S1. Relative abundances of sequence categories in a given host species and colony, with pink representing avian (host) sequences, blue representing environmental eukaryotes (phyotoplankton, algae), brown representing contaminants (to include mammals and fungi), dark purple representing unassignable sequences (not classified at the phylum level), green representing parasitic taxa (as discussed in this paper), and sand colour representing food/prey sequences (to include crustaceans, fish, cephalopods, and other marine metazoans).



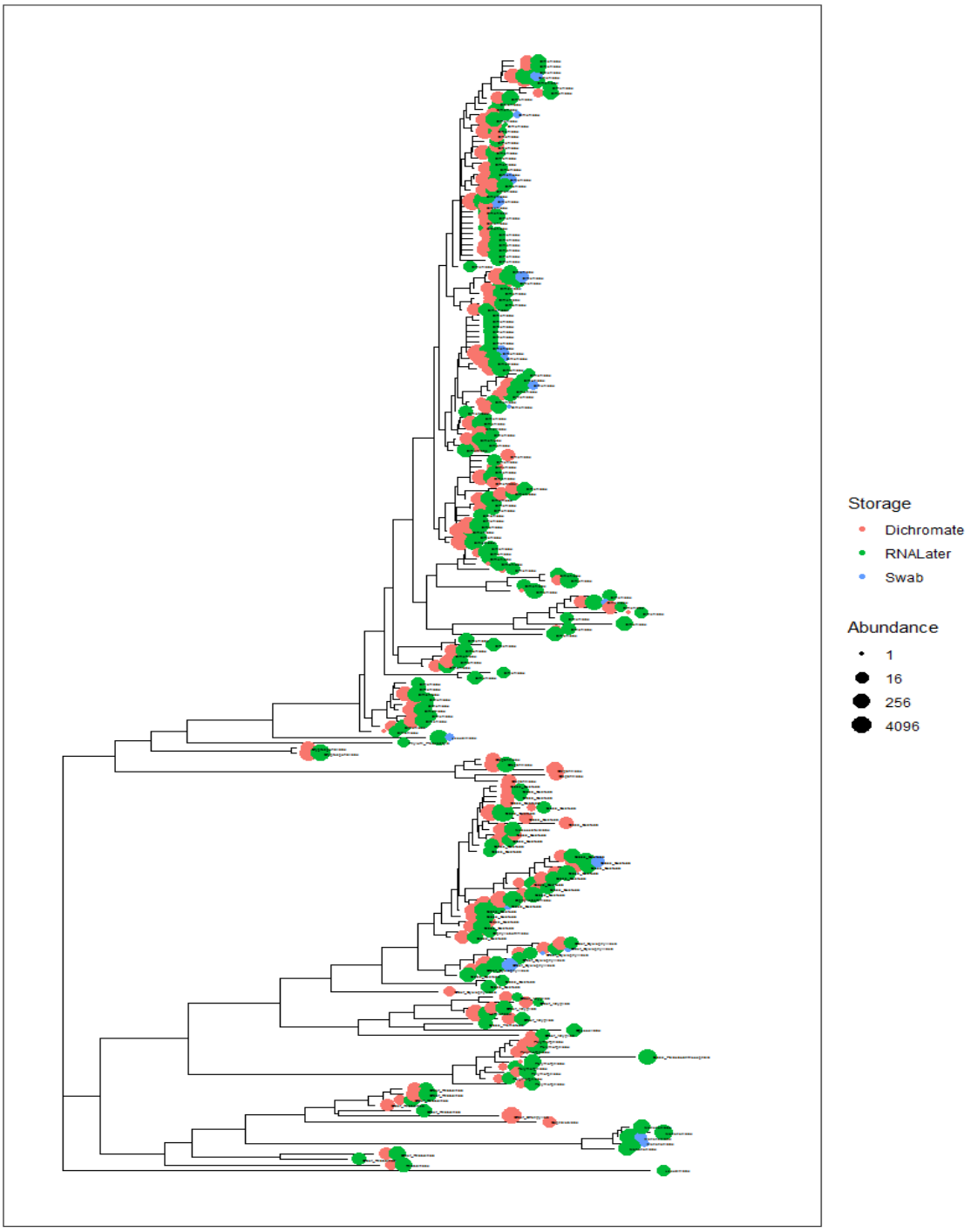
Supplementary Figure S2. Stacked bar plots depicting samples positive for coccidian oocysts (left) and helminths (right) by Pygoscelid host species.



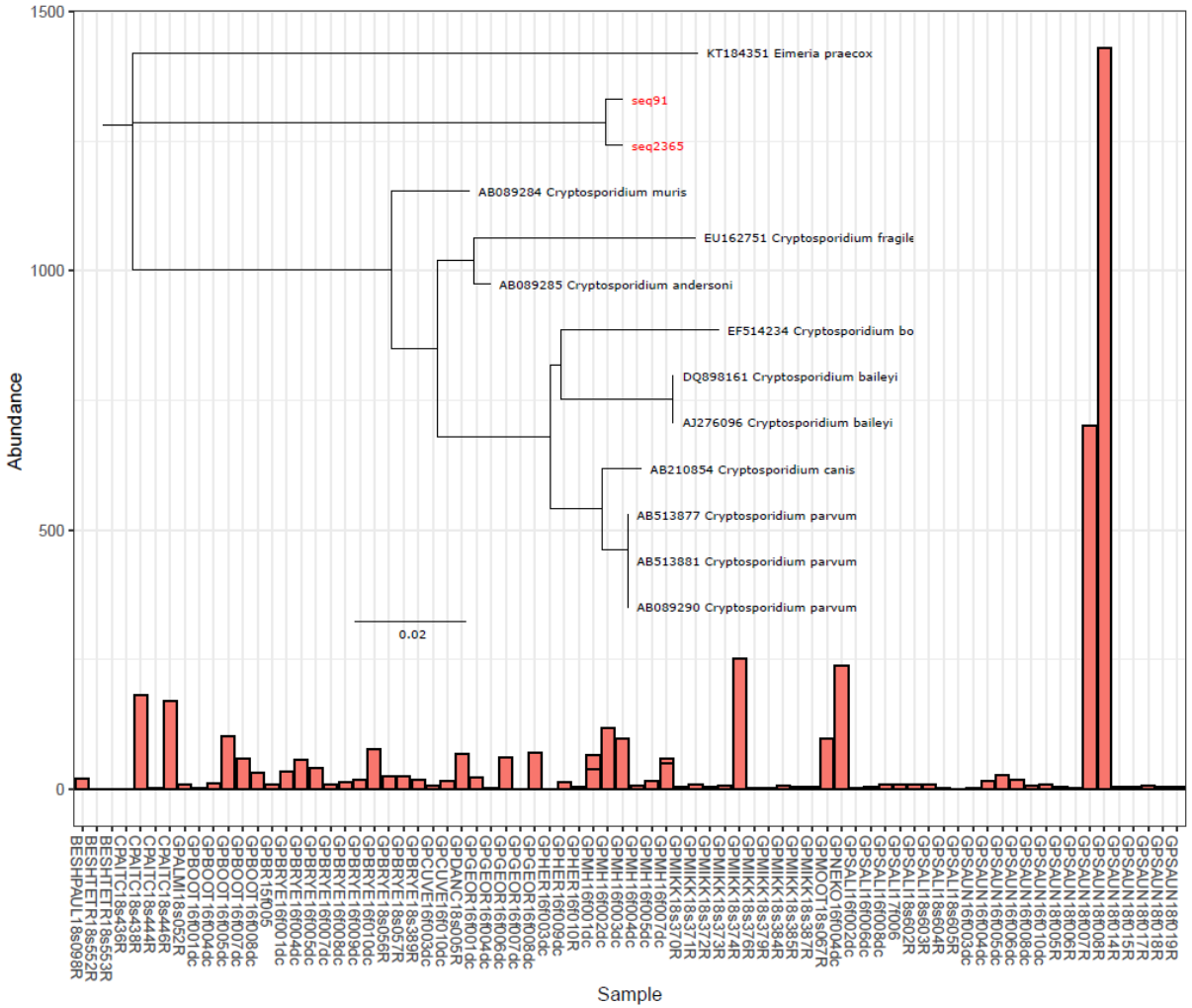
Supplementary Figure S3. NMDS ordinations of parasite community composition from two nearby Gentoo Penguin colonies in the Northeast Antarctic Peninsula, Earle Island and Heroina Island, carried out using Bray-Curtis distance (left) and Weighted UniFrac (right), with colour representing distinct storage media into which the same scat sample was subdivided for storage.



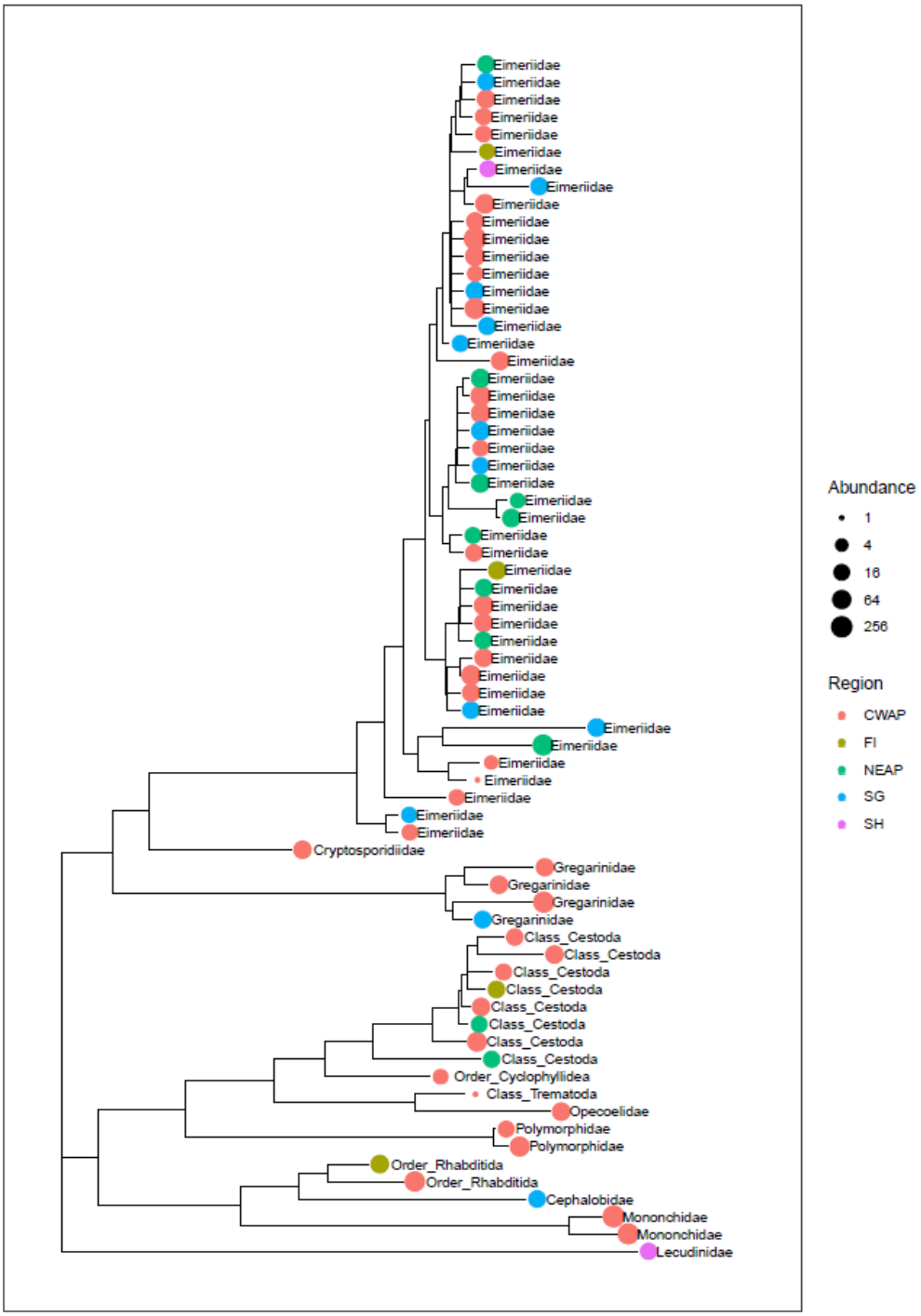
Supplementary Figure S4. Phylogenetic tree (Neighbour-joining, HKY distance) of endoparasite sequences retrieved from two nearby Gentoo Penguin colonies in the Northeast Antarctic Peninsula, Earle Island and Heroina Island, with dot colour representing distinct storage media into which the same scat sample was subdivided for storage, and dot size scaled by sequence abundance. Tip labels depict the Family level, unless the sequence could not be classified at that resolution, in which case the Phylum, Class, or Order is shown.



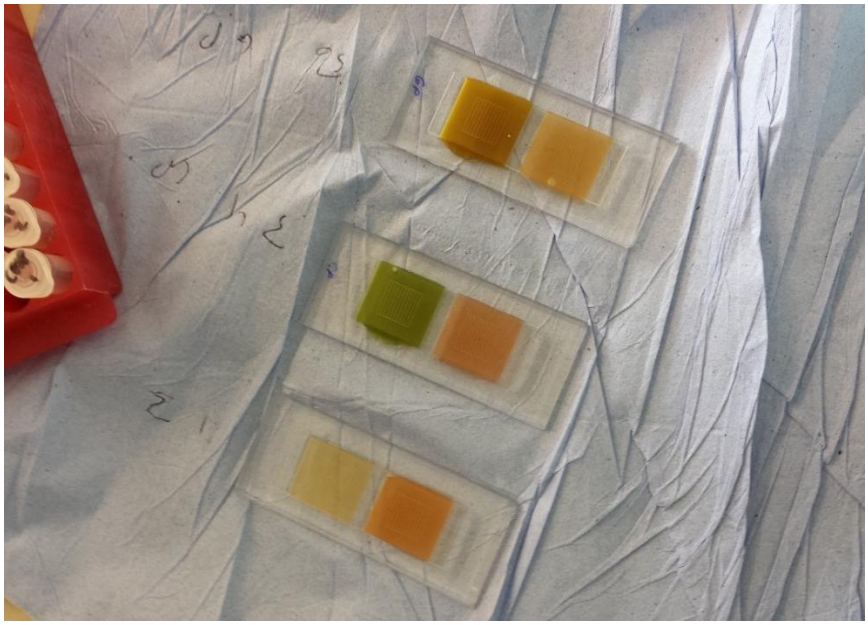
Supplementary Figure S5. Phylogenetic tree (Neighbour-joining, HKY distance) of endoparasite sequences retrieved from the full data set, with dot colour representing distinct storage media, and dot size scaled by sequence abundance. Tip labels depict the Family level, unless the sequence could not be classified at that resolution, in which case the Phylum, Class, or Order is shown.



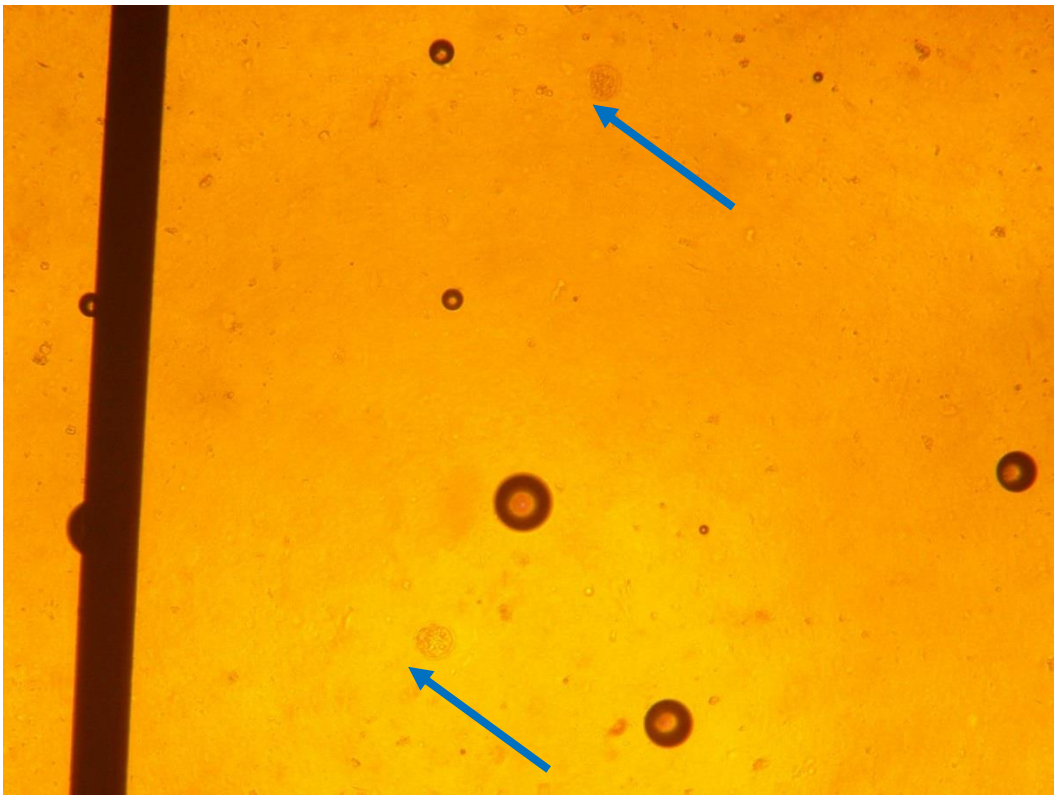
Supplementary Figure S6. Sample-wise read abundance of sequences in the family Cryptosporidiidae retrieved from the full data set, along with an overlaid phylogenetic tree (Neighbour-joining, HKY distance) of the sequences aligned with reference 18S sequences obtained from other vertebrate *Cryptosporidium* sequences. BESH = Blue-Eyed Shag, CP = Chinstrap Penguin, GP = Gentoo Penguin. Four-letter site names follow, as shown in Table 1, as well as the year (two digits), sample number, and storage designation (R = RNAlater, dc = Potassium dichromate).



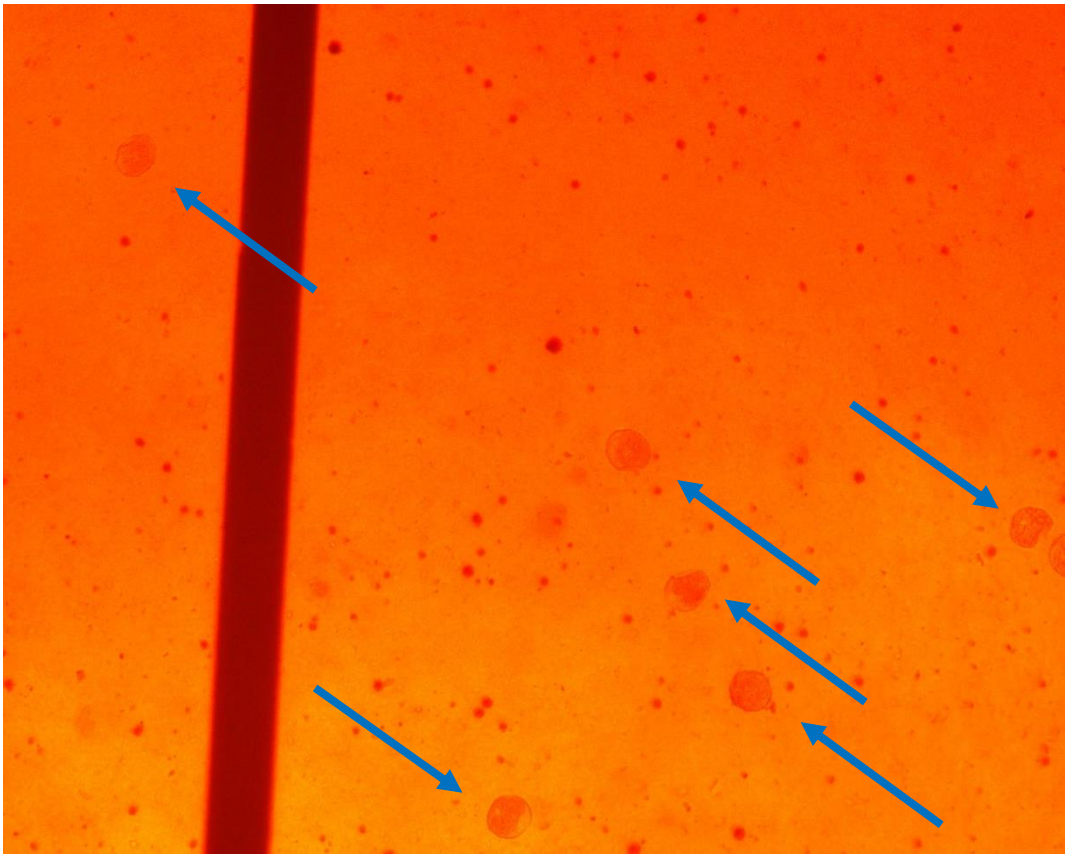
Supplementary Figure S7. Phylogenetic tree (Neighbour-joining, HKY distance) of endoparasite sequences retrieved from only one geographic region in Gentoo Penguins, with dot colour representing these distinct geographic regions (as defined by the Antarctic Site Inventory). CWAP = Central Western Antarctic Peninsula, FI = Falkland Islands/Malvinas, NEAP = Northeast Antarctic Peninsula, SG = South Georgia, and SH = South Shetland Islands. Tip labels depict the Family level, unless the sequence could not be classified at that resolution, in which case the Phylum, Class, or Order is shown.



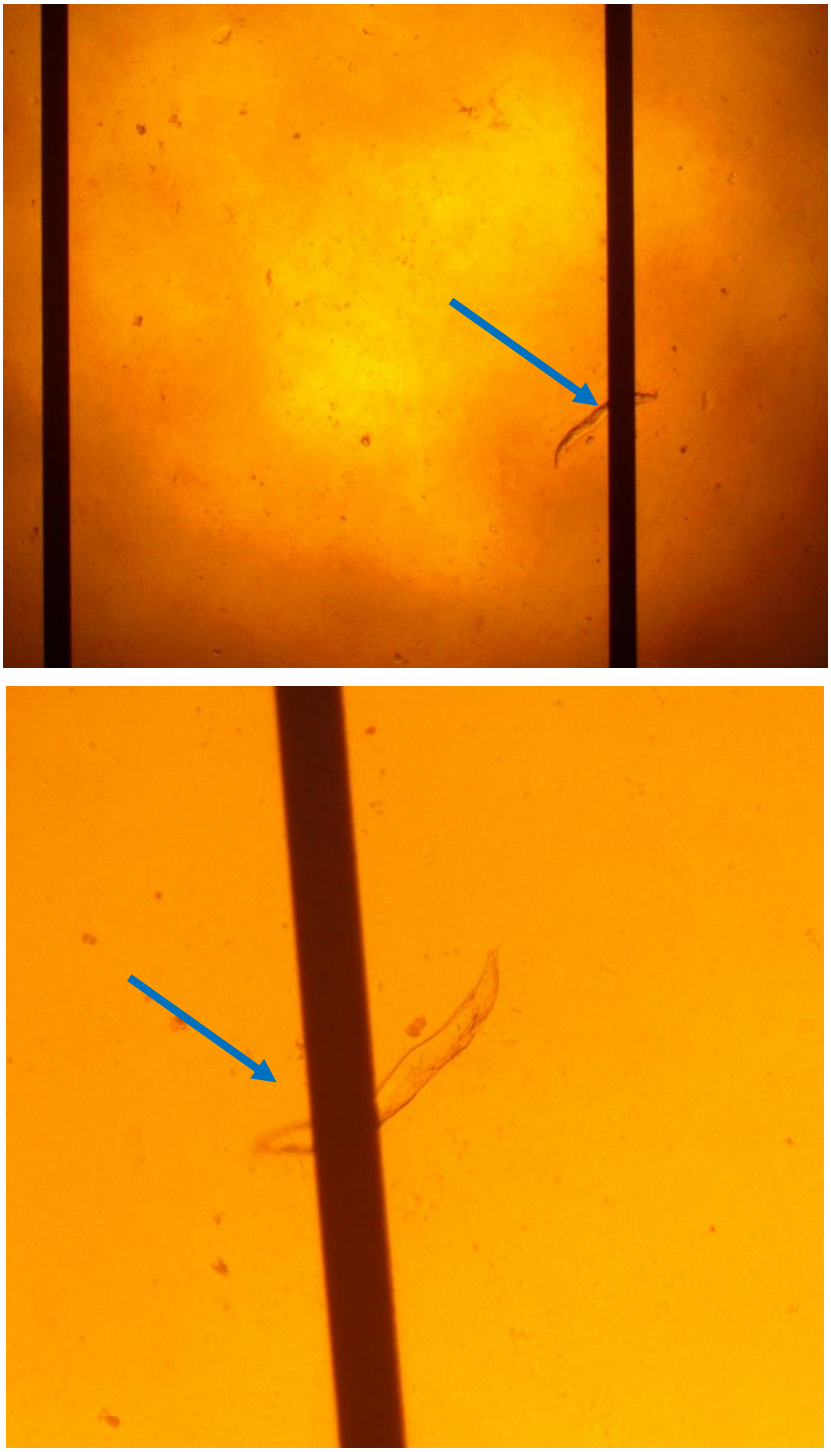
Supplementary Photograph 1. A series of three McMaster counting chambers filled with guano suspensions stand during the salt flotation step prior to microscopic counting of oocysts, eggs, and helminths. This photograph also illustrates the diversity of Pygoscelid penguin guano colouration.



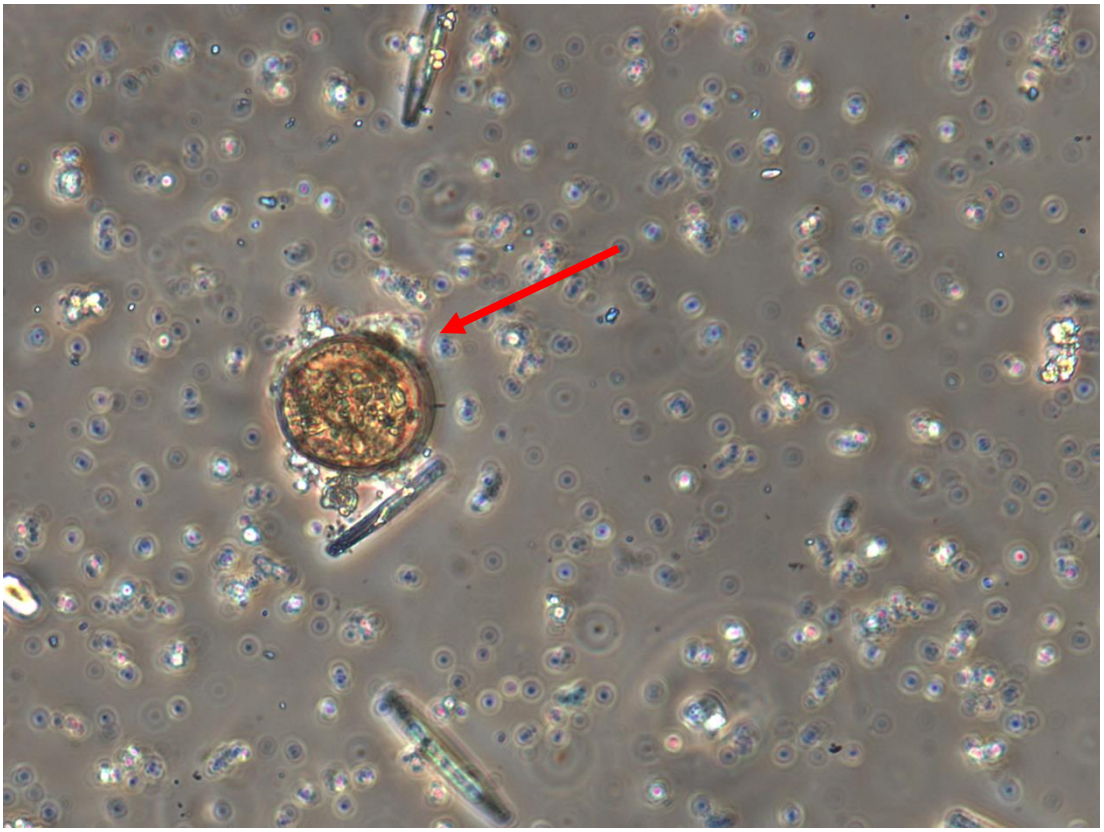
Supplementary Photograph 2. Coccidian oocysts (suspected *Eimeria*) indicated by blue arrows, imaged at 20x power in a McMaster counting chamber (black bar visible from counting cell). Faecal sample GPSALI15f009 – Gentoo Penguin, Salisbury Plain, South Georgia (2014).



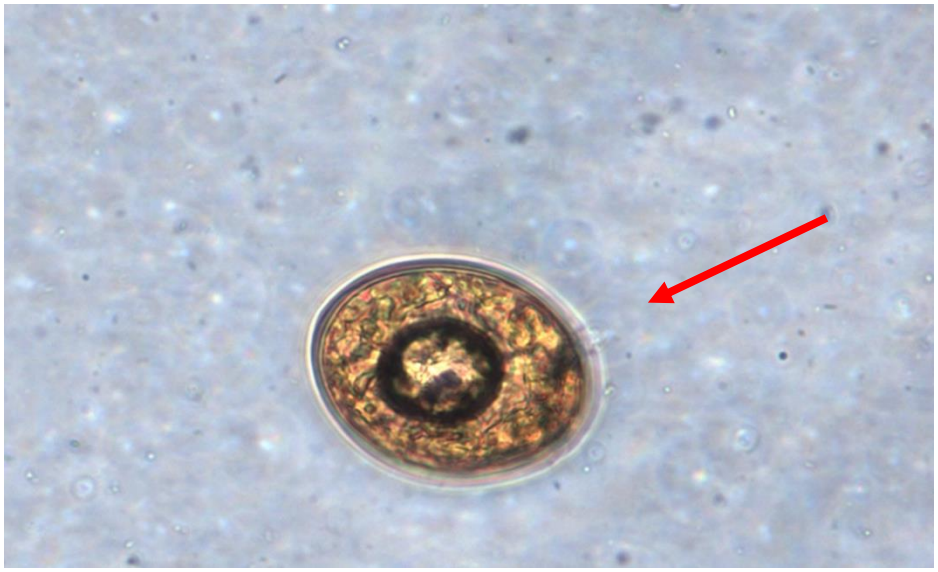
Supplementary Photograph 3. Coccidian oocysts (suspected *Eimeria*) indicated by blue arrows, imaged at 20x power in a McMaster counting chamber (black bar visible from counting cell). Faecal sample GPSAUN15f003 – Gentoo Penguin, Saunders Neck, Falkland Islands (2014).



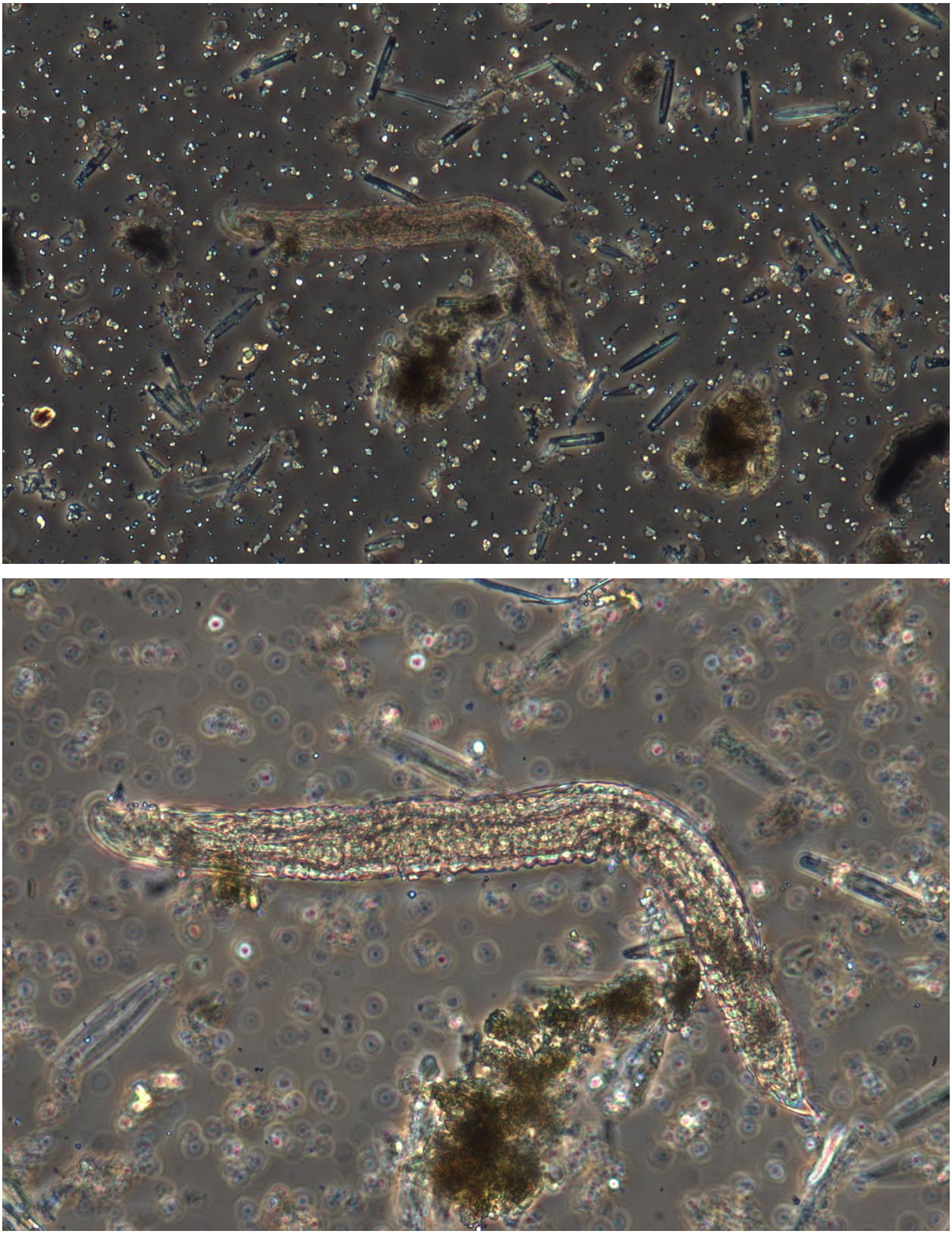
Supplementary Photograph 4. An unidentified helminth indicated by blue arrows, imaged at 10x power (top) and 20x power (bottom) in a McMaster counting chamber (black bar visible from counting cell, distance between bars = 1 mm). Faecal sample CPAITC15f007 – Chisntrap Penguin, Barrientos Island, Aitcho Islands, South Shetland Islands (2014).



Supplementary Photograph 5. Unidentified structure suspected to be a helminth egg indicated with red arrow, imaged at 20x power on a slide. Other debris and diatoms are visible in background. Faecal sample BESHTETR18s554 – Blue-Eyed Shag, Tetrad Islands, Antarctic Peninsula (2018).



Supplementary Photograph 6. Unidentified structure suspected to be a helminth egg indicated with red arrow, imaged at 20x power on a slide. Faecal sample GPALMI15f008 – Gentoo Penguin, Almirante Brown Base, Antarctic Peninsula (2014).



Supplementary Photograph 7. Unidentified helminth, imaged at 20x power (top) and 40x power (bottom) on a slide. Other debris (krill appendages), algae, and diatoms are visible, highlighting the difficulty of detecting parasites in seabird faeces. Faecal sample BESHTETR18s554 – Blue-Eyed Shag, Tetrad Islands, Antarctic Peninsula (2018).

*Alex the Lion: "The **wild**? Are you nuts? That is the worst idea I have ever heard!"*

*Melman the Giraffe: "It's **unsanitary**!"*

Marty the Zebra: "The penguins are going, so why can't I?"

Alex the Lion: "The penguins are psychotic!"

(Madagascar, Pixar Inc., 2005)

Moot Point, Antarctica (Photo: Hila Levy)