Factors that Influence Pygoscelid Microbiomes Across the Scotia Arc



Chloe A. Kaczvinsky
Hertford College
University of Oxford

Thesis submitted for the degree of Doctor of Philosophy

Michaelmas Term 2024-2025
Supervisors: Tom Hart and Adrian Smith

Abstract

Antarctic work is complicated by the factors that make it especially compelling for biological study: extreme seasonality, remote location, and freezing temperatures. Within this environment, penguins represent a major meso-predator and sentinel species. Gentoo penguins (*Pygoscelis papua*) span much of the Scotia arc from the Falkland Islands/Malvinas to continental Antarctica and overlap a wide array of gradients in terms of prey and environment. Examining their relationships with other key species, whether prey, parasite, or key primary producer, is a crucial way of understanding the broader biogeographical interactions within the Scotia Arc. In this thesis, I examine several key aspects of penguin biogeography in the Scotia Arc. First, I investigate the segregation of faecal microbiomes across geography and examine their connections with diet. Then, I use faecal samples to detect potential tapeworm infections in individuals and determine the relationship with bacterial microbiome and diet in those individuals. Next, all of this information is further studied with regards to the algal signatures present in gentoo faeces and the relation of them to other biogeographical factors. Finally, I consider the findings of these chapters within the broader context of extant work on the ecology of the Scotia Arc and for gentoo penguins specifically and lay out further areas for study.

Acknowledgements

I want to thank all the many collaborators and lab mates who helped make this project a reality, there are too many to name, but thank you especially to Meagan Dewar and Samanta Dodino; I really appreciated your insight into the field and science in general. I also want to thank Ignacio Juarez and Steven Fiddaman for being excellent group post docs and the Katzourakis group for being excellent lab and office neighbours. A huge thank you to all my co-authors who have helped make these projects the best they can be.

Thank you to my supervisors, Tom Hart and Adrian Smith, and their lab groups, and to the Maiden, Shephard and Knowles lab groups who made sure I had some other microbiome people to help; your support has been invaluable. Thank you to Andrea Kastner, Heather Green and the rest of the admin team; none of us could do this without you. And finally, thank you to the Oxford Oxitec-Natural Motion Scholarship and the Antarctic Bursary that helped fund this work.

I'd like to thank my parents who read through much of this text to tell me about the approximately 300 misplaced or missing commas (thanks, Mom!), and who supported my academic journey, even though I did not go into the classics despite my dad's earnest entreaties. *Mea culpa*, Dad.

I am especially thankful to my friend, Molly Mitchell, who provided constant support via text despite being many time zones away, and the many housemates who listened to me complain, explain, and, occasionally, rage.

Finally, I'd like to acknowledge the contributions of ethanol, caffeine, and Taylor Swift's "You're on Your Own, Kid." I couldn't have done it without you.

Publication History and Authorship

The following papers are presented in this thesis:

Chapter 2
Kaczvinsky, C., Levy, H., Preston, S., Youngflesh, C., Clucas, G., Lynch, H., Hart, T., Smith, A. (2024). ″The Influence of Biotic and Abiotic Factors on the Bacterial Microbiome of Gentoo Penguins (*Pygoscelis papua*) in Their Natural Environment.″ Scientific Reports **14**(1): 17933.

C.K., H.L., C.Y., H.J.L., A.S., T.H. jointly conceived of the study. T.H., H.L., C.Y., G.C. collected samples. Initial molecular work conducted by H.L. Statistics and coding conducted by C.K., S.P., H.L. C.K. and H.L. drafted the paper, which all authors contributed to. A.S. and T.H. supervised all aspects of the study.

Chapter 3
Kaczvinsky, C., Levy, H., Preston, S., Youngflesh, C., Clucas, G., Lynch, Protasio, Anna V., Christodolou, Maria, Smith, A., Hart, T. "Landscape Level Interactions Between Cestodes and the Microbiome." In Preparation.

C.K., A.S., T.H. jointly conceived of the study. T.H., H.L., C.Y., G.C. collected samples. Initial molecular work conducted by H.L. Statistics and coding conducted by C.K., S.P., M.C. Phylogeny constructed by C.K. and A.P. C.K. drafted the paper, which all authors contributed to. A.S. and T.H. supervised all aspects of the study.

Chapter 4
Kaczvinsky, C., Levy, H., Preston, S., Youngflesh, C., Clucas, G., Lynch, Smith, A., Hart, T. "Connections Between Gentoo Penguin Diet and Algal Signatures." In Preparation.

C.K., A.S., T.H. jointly conceived of the study. T.H., H.L., C.Y., G.C. collected samples. Initial molecular work conducted by H.L. Statistics and coding conducted by C.K., S.P. C.K. drafted the paper, which all authors contributed to. A.S. and T.H. supervised all aspects of the study.

Contents

Abstract i

Acknowledgements ii

Publication History and Authorship iii

Contents iv

Chapter 1: Introduction 1

Chapter 2: The Influence of Biotic and Abiotic Factors on the Bacterial Microbiome of Gentoo Penguins (*Pygoscelis papua*) in Their Natural Environment 17

Chapter 3 Prefix: A Brief Introduction to Helminths in Pygoscelids 29

Chapter 3: Landscape Level Interactions Between Cestodes and the Microbiome 36

Chapter 4: Connections Between Gentoo Penguin Diet and Algal Signatures 64

Chapter 5: Discussion 85

Chapter 6: Conclusion 90

Appendix A 93

Appendix B 122

Appendix C 134

Chapter 1
Introduction

Factors that Influence Pygoscelid Microbiomes Across the Scotia Arc

Understanding how ecological drivers affect organisms usually requires both resolution and replication across some kind of gradient. Most studies of animal microbiomes are restricted to laboratory settings, where experimental manipulations are easier. Despite any difficulties, studying microbiomes in the wild is important because it allows us to test hypotheses about connectivity and disease transmission that would otherwise be impossible. It is especially interesting to test in the Antarctic and Scotia Arc because it has many features that are rare outside of this region. Namely, penguin host species exist across a large latitudinal gradient, and low temperatures and freezing mean that many pathogens may be unable to live outside the host between seasons. This thesis attempts to resolve some of these drivers by studying the microbiome, pathogens and diseases of these same populations across latitudes and gradients of temperature, distance between colonies, and other biogeographic factors.

Antarctic Ecology

Polar biology allows the study of a relatively isolated system that is subject to multiple extreme environments. Testing systems at extremes can provide important insight into their functions by delineating the limits and what informs those limits. In polar ecosystems, organisms are exposed to strong seasonality, including being in darkness for significant periods of the year and the resultant seasonality of photosynthetic biomass. These environments are also subject to freezing temperatures over long time frames.

Perhaps the most obvious example of Antarctica's extreme environment is the freezing temperatures for most of the year and associated annual cycle. The Southern Ocean is dominated by the annual growth and contraction of sea ice that is among the greatest annual physical variations on Earth (Massom and Stammerjohn 2010). Terrestrially, even the sub-Antarctic islands are

subject to temperatures that rarely climb above 5 °C and can be subject to multiple freeze thaw cycles annually (Walton 1982), while the polar oceans are some of the coldest waters to which marine organisms are exposed (Clarke 1991). Organisms living in this environment must contend with the impact of freezing temperatures that can cause dehydration and lethal intracellular ice crystals (Meryman 1956, Clarke 1991). Adaptations range from anti-freeze peptides and proteins in Antarctic ice fish (Bilyk and DeVries 2010) to anti-icing feather structures in penguins (Alizadeh-Birjandi, Tavakoli-Dastjerdi et al. 2020).

Despite the freezing pressures, the Southern Ocean has high productivity in ice pores and layers of sea ice and can have intense algal blooms during summer, leading to high seasonality (Barnes and Clarke 2011). This productivity is, of course, limited by the extreme seasonality of a polar environment, where winter is spent in complete or almost complete darkness that limits the phototrophic species at the bottom of the food web (Marschall 1988, Hunt, Drinkwater et al. 2016). Algal species and other phytoplankton do persist over the dark winter in sea ice, and these reservoirs of phytoplankton feed larval krill and other groups that cannot fast the entire low winter period (Marschall 1988, Meyer, Atkinson et al. 2002, Johnston, Murphy et al. 2022). These phytoplankton and their summer blooms especially feed a variety of zooplankton, including the crucial Antarctic krill, which feed many of the higher order predators in Antarctica, including whales and penguins (Johnston, Murphy et al. 2022).

While subject to the same environmental pressures as the Arctic, Antarctica is also much more ecologically and geographically isolated by the Antarctic Circumpolar Current (ACC). This is a current of west-bound winds that distributes organisms within the same latitude and acts as a boundary to north-south dispersal (Hunt, Drinkwater et al. 2016). The ACC includes the Antarctic Polar Front [hereafter Polar Front] where colder southern waters are subsumed under warmer northern waters (Hunt, Drinkwater et al. 2016, Fraser, Key et al. 2017). The abrupt change from

warmer to cooler water that characterizes the Polar Front provides a strong ecological barrier to species that might otherwise disperse across it (Hunt, Drinkwater et al. 2016, Carpenter-Kling, Handley et al. 2019, Vianna, Fernandes et al. 2020). This temperature barrier has important implications for the potential colonization of Antarctica by invasive and other non-native species, including microbes brought in on other species.

Biodiversity of macroorganisms (plants and animals), tends to peak in the tropics and decreases towards the poles (Yergeau, Newsham et al. 2007). It is likely that the increased stressors of an extreme environment like Antarctica contributes to this pattern. There are several key obstacles with which organisms would have to contend. Equally, a lack of free water (Convey, Chown et al. 2014), near total lack of sunlight at lower latitudes in winter for autotrophic organisms, and the consequent reduced availability of biomass for heterotrophic organisms (Rothschild and Mancinelli 2001, Yergeau, Newsham et al. 2007) create serious limitations for organisms in these environments.

Pygoscelis (Brush-Tailed Penguins)

Penguins provide a potentially useful study system of microbiomes across broad biogeographical gradients. Penguins are pelagic birds that come ashore for breeding and moulting. They have highly variable colony sizes and nest across a variety of environments, including the rocky and ice free areas of both continental Antarctica and islands. In the Scotia Arc, they are subject to the polar environment detailed previously. The cold, remote location will likely limit the degree to which other species are introducing microbes, with the main other species present being predatory birds and seals. Furthermore, these penguin species live in a variety of colony sizes, ranging from half a million (pairs) to dozens, with a wide range of distances from other colonies, and both mixed and single species colonies. These highly diverse conditions allow comparisons between a

wide range of environmental conditions to better tease out different ecological factors affecting associated microbiome.

Pygoscelis is a genus of Antarctic and sub-Antarctic penguins that inhabit Antarctic and sub-Antarctic waters containing three species: gentoo penguins (*Pygoscelis papua*), chinstrap penguins (*Pygoscelis antarctica*), and Adélie penguins (*Pygoscelis adeliae*). They have overlapping ranges covering over 30 degrees latitude (Lynch 2013, Trathan and Grant 2013, Trivelpiece and Trivelpiece 2013), but their diets, habitats, and breeding cycles have different degrees of divergence. Chinstraps and Adélies are classed as krill specialists, generally Antarctic krill (*Euphausia superba*), while gentoo diets include varying degrees of fish (often Antarctic silverfish), krill, squid, and occasionally salps (Clucas, Dunn et al. 2014). Dietary specialization is variable within gentoos between individuals and seems to follow a broader latitudinal gradient, with greater degrees of generalization at the northern part of the range (Herman, Valls et al. 2017). Furthermore, both Adélies and chinstraps forage farther from their colonies, and gentoos often focus on shallower dives than chinstraps (Kokubun, Takahashi et al. 2010, Miller, Kappes et al. 2010, Cimino, Moline et al. 2016, Pickett, Fraser et al. 2018). Gentoos also show greater year-to-year variation in their diving patterns, which is possibly related to the availability and structure of krill populations (Pickett, Fraser et al. 2018). The wide ecological ranges pygoscelids inhabit make pygoscelids useful study subjects to understand biogeographic impacts across these large ecological gradients.

Pygoscelids all breed during the Antarctic summer after an annual moult, but the exact timing and the degree to which it is synchronous across colonies and latitudes varies with species. (Stonehouse 1970, Williams 1995, Levy, Clucas et al. 2016, Clucas, Younger et al. 2018, Cole, Dutoit et al. 2019, Levy 2020). In gentoos, phenology generally varies by a few weeks across the wide number of latitudes that they inhabit, with more northern populations (exposed to higher

temps) beginning settlement and egg laying earlier (Hinke, Polito et al. 2012, Ancel, Beaulieu et al. 2013, Lynch 2013, Black 2016). Gentoos also show the greatest sensitivity to environmental variation, adjusting clutch initiation date twice as much per °C compared with Adélies (Hinke, Polito et al. 2012). Similar latitudinal variation has also been observed for Adélie penguins, while chinstrap breeding is relatively synchronous across the range (Trathan and Grant 2013, Trivelpiece and Trivelpiece 2013). After breeding and moulting, gentoos are more likely to remain at their breeding colony during the winter months than other pygoscelids (Williams 1991, Hinke and Trivelpiece 2011, Black, Rey et al. 2017). This difference in dispersal patterns has been connected with differences in habitats between gentoos and other pygoscelids and could contribute to increased population structuring and reduced migration in gentoos (Clucas, Dunn et al. 2014, Black, Rey et al. 2017, Vianna, Noll et al. 2017).

In addition to being a useful model system to investigate evolution along an ecological gradient, *Pygoscelis* penguins are a family of sister taxa to track the numerous climactic and ecological changes underway in Antarctica and the surrounding islands. Generally, all pygoscelid species are dependent on ice conditions, but gentoos are the least affected by reductions in sea ice (Forcada, Trathan et al. 2006). Gentoos are able to colonize newly uncovered land free of sea ice and are more flexible in their diet, while chinstrap and Adélie populations are more negatively impacted by decreases (Lima and Estay 2013, Trathan and Grant 2013, Herman, Borowicz et al. 2020). For example, Adélie foraging trips are longest in times with the lowest sea ice extent, at least for the breeding period (Fraser and Hofmann 2003, Trathan and Grant 2013). The increasing populations of gentoos in the Western Antarctic Peninsula and the decreasing numbers of Adélies are correlated with the retreating sea ice cover in the area (Lima and Estay 2013) and will likely serve as an important factor in predicting more long-term effects of climate change on Antarctic ecology.

Microbiome

Microbiomes are the bacteria, fungi, and unicellular eukaryotes associated with an environment or large, multicellular host organism. The majority of studies focus on human health and model systems like mice (Clavel, Lagkouvardos et al. 2016, Huwe, Prusty et al. 2019). These studies show complex systems of microbiomes and hosts that interact with the health (Gérard, Béguet et al. 2004), infection (Zaiss and Harris 2016, Hahn, Piecyk et al. 2022), diet (Turnbaugh, Ridaura et al. 2009), season (Bird, Prewer et al. 2019), metabolism (Samuel and Gordon 2006), and immune response (Larsson, Tremaroli et al. 2012) of animals; however, such studies are rarer in wild settings. Thus, we have some understanding of the mechanisms through which microbiomes influence health, but it is unclear if these processes are similar in wild animals.

In host-associated microbiomes, host diet plays a role in microbiome composition (Turnbaugh, Ridaura et al. 2009). Experimental studies in mice have demonstrated the primary importance of diet in determining the microbiome niche available for colonization (DeLong 2014), though to what extent this is true in birds is uncertain. A study of Darwin's finches found that the vampire finch microbiome had several rare taxa associated with carnivorous birds and reptiles, and that these bacteria were rare or absent in other finches that were not sanguivorous or carnivorous (Michel, Ward et al. 2018). Less extreme dietary differences are also reflected in the microbiome. A study on a broader range of finches across habitats found a stronger correlation with observed foraging data and microbiome that differed from the phylogeny of the finches (Loo, Garcia-Loor et al. 2019).

Wild avian microbiomes are dominated by Firmicutes, Proteobacteria, Bacteroides, and Actinobacteria (Grond, Sandercock et al. 2018). In the case of penguin microbiomes, all four are the main phyla present together with Fusobacteria (Banks, Cary et al. 2009, Dewar, Arnould et al. 2013, Barbosa, Balagué et al. 2016, Zeng, Li et al. 2021). Current studies have found higher

alpha diversity in gentoo faecal samples than sympatric Adélies (Zeng, Li et al. 2021), but no significant differences in alpha diversity for sympatric Adélie and chinstrap penguin stomach microbiota (Yew, Pearce et al. 2017). A different study examining sympatric Adélie and gentoo penguins found higher concentrations of Actinobacteria and Cyanobacteria in gentoo penguins, which they attributed to the combination of phylogenetic divergence and differences in diet compared to other penguin species (Zeng, Li et al. 2021). A study of Adélie penguins from the Ross Sea found significant correlation between genetic distance and the similarity of microbiomes, concluding that these bacteria are inherited (Banks, Cary et al. 2009). A different study of chick and adult chinstrap penguins, however, found that age was the primary determining factor of microbiome and that chicks did not group with adults from their nest (Barbosa, Balagué et al. 2016). The divergences between adults and chicks were determined to be due to differences in the prevalence of common taxa, rather than the presence or absence of rare taxa (Barbosa, Balagué et al. 2016). However, given the age-segregation of microbiomes, this thesis will focus on adults. Several known human pathogens have also been found in penguin microbiomes (including *Campylobacter*, *Helicobacter*, and *Streptococcus*) but whether they are pathogenic to penguins is unknown (Dewar, Arnould et al. 2013). Thus, there is evidence for some importance of both genetic and diet differences in determining penguin-associated microbiomes, coupled with the presence of possible pathogens.

We know the microbiome is important for nutrient production, disease resistance, and general evolutionary fitness of the host. Host dynamics are expected to influence the microbiome both within and between populations. Host-associated microbiomes, in turn, are rarely studied in wild populations and tend to focus on humans in biomedical contexts. We use the natural experiment of Antarctica, with the reduced complexity expected in a high latitude ecosystem and taxa that occur over a large range with gradients concerning diet, population size, and temperature to

study how diet and biogeography play a role in the microbiome. In turn, this study can provide important information about temperature impacts and thus conservation risks in a rapidly changing environment.

Disease

Infection and disease are crucial factors in organism fitness. In wild systems, disease in concert with other stressors, such as malnutrition or heat stress, can combine to cause mortality when one of the factors alone would be survivable. Understanding disease and its impacts on organisms allows greater understanding for the survival pressures acting on organisms and the interactions this can have with other ecological factors.

We have several key studies of disease in zoo penguins that provides important information about the potential susceptibility of penguins to various diseases. There are key issues with the health of penguins in captive environments, including common infections with aspergillosis (a fungal infection) and diseases such as avian malaria that are not naturally found in Antarctic and sub-Antarctic species (Itoh, Matsumoto et al. 2023). Recent studies have found aspergillosis in wild, migrating birds, but these were penguins in the temperate regions of South America, rather than the polar regions studied in this work (Melo, Poester et al. 2023).

Generally, wild Antarctic and sub-Antarctic penguins are considered immunologically naïve due to a remote habitat that has historically reduced contact with potential vectors and pathogenic organisms unable to withstand the temperatures (Kerry, Riddle et al. 1998, Clarke and Kerry 2000, Grimaldi, Seddon et al. 2015, Trathan, Garcia-Borboroglu et al. 2015, Levy 2020). Immunological naïveté in penguins has been supported by their relatively high susceptibility to disease in captivity (Grimaldi, Seddon et al. 2015) but there have not been any studies

demonstrating the relative lack of pathogens or increased susceptibility in wild populations. It is, therefore, uncertain if these populations are truly immunologically naïve.

In wild penguins, there is evidence for mass mortality due to diseases like Avian cholera, and necroscopic studies have found lesions and infection of helminths like tapeworms in the digestive tracts of the deceased penguins (Leotta, Chinen et al. 2006, Cooper, Crawford et al. 2009, Diaz, Fusaro et al. 2016, Martin, Ortiz et al. 2016). In human and lab systems, these infections are linked to complex changes in the infected organism, including changes in immunity modulation, host microbiome, and the leaching of nutrients from the host (Dheilly, Poulin et al. 2015, Zaiss and Harris 2016, Hahn, Piecyk et al. 2022). Multiple potentially pathogenic genera have been identified in pygoscelid species by cultures and PCR including *Chlamydia* sp. and *Campylobacter* sp. (Levy 2020). However, many are only potentially pathogenic and have not been actively identified as the causative agent of a disease outbreak. It is possible that many of these are commensal or opportunistically pathogenic, complicating our understanding of level of threats to pygoscelid species.

Dense aggregations of pygoscelids during the breeding season (given the increasing rarity of ice-free ground on continental Antarctica) are important opportunities for the spread of pathogens throughout colonies. The ubiquity of faecal material in these colonies is also a crucial factor in disease transmission, allowing spread through contaminated water and nesting material, in addition to dense population and high faecal concentration (Leotta, Chinen et al. 2006, Griekspoor, Olsen et al. 2009). Finally, penguin colonies are often in contact with other seabird species, allowing disease flow (Grimaldi, Seddon et al. 2015).

There is speculation on the impact of the dense colonial living for the spread of microorganisms, particularly in a disease context (Griekspoor, Olsen et al. 2009, Dewar, Arnould et al. 2013). The degree to which this could be active in mixed colonies, with multiple species of

pygoscelids is uncertain. Studying disease in a wild population across a large area with varying biotic and abiotic factors, like pygoscelid penguins, allows us to examine their impacts in complex systems and some of the factors that may influence them.

Aims

This thesis will examine the connections between biogeography in a polar environment, host-associated microbiomes, diseases, and the presence of closely related sympatric species in pygoscelid penguins. This will be done through three key questions. First, does penguin faecal microbiome correlate with biogeography and diet? Next, does the prescence of tapeworms co-occur with microbiome changes and biogeographic differences? Finally, are there biogeographic links between algal traces in gentoo faeces from across the Scotia Arc, diet, and bacterial microbiomes in gentoos? Harnessing these findings, I will examine how these studies inform our understanding of microbiomes, biogeography, and disease more generally, which can inform conservation strategy.

Works Cited

Alizadeh-Birjandi, E., F. Tavakoli-Dastjerdi, J. S. Leger, K. F. Faull, S. H. Davis, J. P. Rothstein and H. Pirouz Kavehpour (2020). ″Delay of ice formation on penguin feathers.″ The European Physical Journal Special Topics **229**(10): 1881-1896.

Ancel, A., M. Beaulieu and C. Gilbert (2013). ″The different breeding strategies of penguins: A review.″ Comptes Rendus Biologies **336**(1): 1-12.

Banks, J. C., S. C. Cary and I. D. Hogg (2009). ″The phylogeography of Adelie penguin faecal flora.″ Environ Microbiol **11**(3): 577-588.

Barbosa, A., V. Balagué, F. Valera, A. Martínez, J. Benzal, M. Motas, J. I. Diaz, A. Mira and C. Pedrós-Alió (2016). ″Age-Related Differences in the Gastrointestinal Microbiota of Chinstrap Penguins (Pygoscelis antarctica).″ PLOS ONE **11**(4): e0153215.

Barnes, D. K. A. and A. Clarke (2011). ″Antarctic marine biology.″ Current Biology **21**(12): R451-R457.

Bilyk, K. T. and A. L. DeVries (2010). ″Freezing avoidance of the Antarctic icefishes (Channichthyidae) across thermal gradients in the Southern Ocean.″ Polar Biology **33**(2): 203-213.

Bird, S., E. Prewer, S. Kutz, L.-M. Leclerc, S. T. Vilaça and C. J. Kyle (2019). ″Geography, seasonality, and host-associated population structure influence the fecal microbiome of a genetically depauparate Arctic mammal.″ Ecology and Evolution **9**(23): 13202-13217.

Black, C., A. R. Rey and T. Hart (2017). ″Peeking into the bleak midwinter: Investigating nonbreeding strategies of Gentoo Penguins using a camera network.″ The Auk **134**(3): 520-529.

Black, C. E. (2016). ″A comprehensive review of the phenology of Pygoscelis penguins.″ Polar Biology **39**(3): 405-432.

Carpenter-Kling, T., J. M. Handley, M. Connan, R. J. M. Crawford, A. B. Makhado, B. M. Dyer, W. Froneman, T. Lamont, A. C. Wolfaardt, M. Landman, M. Sigqala and P. A. Pistorius (2019). ″Gentoo penguins as sentinels of climate change at the sub-Antarctic Prince Edward Archipelago, Southern Ocean.″ Ecological Indicators **101**: 163-172.

Cimino, M. A., M. A. Moline, W. R. Fraser, D. L. Patterson-Fraser and M. J. Oliver (2016). ″Climate-driven sympatry may not lead to foraging competition between congeneric top-predators.″ Scientific Reports **6**(1): 18820.

Clarke, A. (1991). ″Cold adaptation.″ Journal of Zoology **225**(4): 691-699.

Clarke, J. and K. Kerry (2000). ″Diseases and Parasites of Penguins.″ Penguin Conservation **13**(1): 5-24.

Clavel, T., I. Lagkouvardos, M. Blaut and B. Stecher (2016). ″The mouse gut microbiome revisited: From complex diversity to model ecosystems.″ International Journal of Medical Microbiology **306**(5): 316-327.

Clucas, G. V., M. J. Dunn, G. Dyke, S. D. Emslie, R. Naveen, M. J. Polito, O. G. Pybus, A. D. Rogers and T. Hart (2014). ″A reversal of fortunes: climate change 'winners' and 'losers' in Antarctic Peninsula penguins.″ Sci Rep **4**: 5024.

Clucas, G. V., J. L. Younger, D. Kao, L. Emmerson, C. Southwell, B. Wienecke, A. D. Rogers, C.-A. Bost, G. D. Miller, M. J. Polito, P. Lelliott, J. Handley, S. Crofts, R. A. Phillips, M. J. Dunn, K. J. Miller and T. Hart (2018). ″Comparative Population Genomics Reveals Key Barriers to Dispersal in Southern Ocean Penguins.″ Molecular Ecology **27**: 4680-4697.

Cole, T. L., L. Dutoit, N. Dussex, T. Hart, A. Alexander, J. L. Younger, G. V. Clucas, M. J. Frugone, Y. Cherel, R. J. Cuthbert, U. Ellenberg, S. R. Fiddaman, J. Hiscock, D. Houston, P. Jouventin, T. Mattern, G. D. Miller, C. Miskelly, P. Nolan, B. Wienecke, J. A. Vianna and J. M. Waters (2019). "Receding ice drove parallel expansions in Southern Ocean penguins." PNAS **116**(52): 26690-26696.

Convey, P., S. L. Chown, A. Clarke, D. K. A. Barnes, S. Bokhorst, V. Cummings, H. W. Ducklow, F. Frati, T. G. A. Green, S. Gordon, H. J. Griffiths, C. Howard-Williams, H. L. Huiskes, J. Laybourn-Parry, W. B. Lyons, A. McMinn, S. A. Morley, L. S. Peck, A. Quesada, S. A. Robinson, S. Schiaparelli and D. H. Wall (2014). "The Spatial Structure of Antarctic Biodiversity." Ecological Monographs **84**(2): 203-244.

Cooper, J., R. J. M. Crawford, M. S. De Villiers, B. M. Dyer, G. J. G. Hofmeyr and A. Jonker (2009). "Disease Outbreaks Among Penguins at Sub-Antarctic Marion Island: A Conservation Concern." Marine Ornithology **37**: 193-196.

DeLong, E. F. (2014). "Alien invasions and gut "island biogeography"." Cell **159**(2): 233-235.

Dewar, M. L., J. P. Y. Arnould, P. Dann, P. Trathan, R. Groscolas and S. Smith (2013). "Interspecific variations in the gastrointestinal microbiota in penguins." MicrobiologyOpen **2**(1): 195-204.

Dheilly, N. M., R. Poulin and F. Thomas (2015). "Biological warfare: Microorganisms as drivers of host–parasite interactions." Infection, Genetics and Evolution **34**: 251-259.

Diaz, J. I., B. Fusaro, L. Longarzo, N. R. Coria, V. Vidal, V. D'Amico and A. Barbosa (2016). "Gastrointestinal helminths of Adélie penguins (Pygoscelis adeliae) from Antarctica." Polar Research **35**(1).

Forcada, J., P. N. Trathan, K. Reid, E. J. Murphy and J. P. Croxall (2006). "Contrasting population changes in sympatric penguin species in association with climate warming." Global Change Biology **12**(3): 411-423.

Fraser, C. I., G. M. Key, M. du Plessis and P. G. Ryan (2017). "Breaking Down the Barrier: Dispersal Across the Antarctic Polar Front." Ecography **40**: 235-237.

Fraser, W. R. and E. E. Hofmann (2003). "A predator's perspective on causal links between climate change, physical forcing and ecosystem response." Marine Ecology Progress Series **265**: 1-15.

Gérard, P., F. Béguet, P. Lepercq, L. Rigottier-Gois, V. Rochet, C. Andrieux and C. Juste (2004). "Gnotobiotic rats harboring human intestinal microbiota as a model for studying cholesterol-to-coprostanol conversion." FEMS Microbiol Ecol **47**(3): 337-343.

Griekspoor, P., B. Olsen and J. Waldenström (2009). "Campylobacter jejuni in penguins, Antarctica." Emerg Infect Dis **15**(5): 847-848.

Grimaldi, W. W., P. J. Seddon, P. O. B. Lyver, S. Nakagawa and D. M. Tompkins (2015). "Infectious diseases of Antarctic penguins: current status and future threats." Polar Biology **38**(5): 591-606.

Grond, K., B. K. Sandercock, A. Jumpponen and L. H. Zeglin (2018). "The avian gut microbiota: community, physiology and function in wild birds." Journal of Avian Biology **49**(11).

Hahn, M. A., A. Piecyk, F. Jorge, R. Cerrato, M. Kalbe and N. M. Dheilly (2022). "Host phenotype and microbiome vary with infection status, parasite genotype, and parasite microbiome composition." Molecular Ecology **31**(5): 1577-1594.

Herman, R., A. Borowicz, M. Lynch, P. Trathan, T. Hart and H. Lynch (2020). "Update on the global abundance and distribution of breeding Gentoo Penguins (Pygoscelis papua)." Polar Biology **43**(12): 1947-1956.

Herman, R. W., F. C. L. Valls, T. Hart, M. V. Petry, W. Z. Trivelpiece and M. J. Polito (2017). "Seasonal consistency and individual variation in foraging strategies differ among and

within Pygoscelis penguin species in the Antarctic Peninsula region." Marine Biology **164**(5): 115.

Hinke, J. T., M. J. Polito, C. S. Reiss, S. G. Trivelpiece and W. Z. Trivelpiece (2012). "Flexible reproductive timing can buffer reproductive success of Pygoscelis spp. penguins in the Antarctic Peninsula region." Marine Ecology Progress Series **454**: 91-104.

Hinke, J. T. and W. Z. Trivelpiece (2011). "Daily activity and minimum food requirements during winter for gentoo penguins (Pygoscelis papua) in the South Shetland Islands, Antarctica." Polar Biology **34**(10): 1579-1590.

Hunt, G. L., K. F. Drinkwater, K. Arrigo, J. Berge, K. L. Daly, S. Danielson, M. Daase, H. Hop, E. Isla, N. Karnovsky, K. Laidre, F. J. Mueter, E. J. Murphy, P. E. Renaud, W. O. Smith, P. Trathan, J. Turner and D. Wolf-Gladrow (2016). "Advection in polar and sub-polar environments: Impacts on high latitude marine ecosystems." Progress in Oceanography **149**: 40-81.

Huwe, T., B. K. Prusty, A. Ray, S. Lee, B. Ravindran and E. Michael (2019). "Interactions between Parasitic Infections and the Human Gut Microbiome in Odisha, India." Am J Trop Med Hyg **100**(6): 1486-1489.

Itoh, M., N. Matsumoto, K. Hagino, N. Sawayama, M. Kuwayama, K. Yamada and T. Toyotome (2023). "Nationwide Survey about the Occurrence of Aspergillosis in Captive Penguins in Zoos and Aquariums in Japan." Animals **13**(12): 1913.

Johnston, N. M., E. J. Murphy, A. Atkinson, A. J. Constable, C. Cotté, M. Cox, K. L. Daly, R. Driscoll, H. Flores, S. Halfter, N. Henschke, S. L. Hill, J. Höfer, B. P. V. Hunt, S. Kawaguchi, D. Lindsay, C. Liszka, V. Loeb, C. Manno, B. Meyer, E. A. Pakhomov, M. H. Pinkerton, C. S. Reiss, K. Richerson, W. O. S. Jr., D. K. Steinberg, K. M. Swadling, G. A. Tarling, S. E. Thorpe, D. Veytia, P. Ward, C. K. Weldrick and G. Yang (2022). "Status, Change, and Futures of Zooplankton in the Southern Ocean." Frontiers in Ecology and Evolution **9**.

Kerry, K., M. Riddle and J. Clarke (1998). Diseases of Antarctic Wildlife. Kingston, Australia, Australian Antarctic Division.

Kokubun, N., A. Takahashi, Y. Mori, S. Watanabe and H.-C. Shin (2010). "Comparison of diving behavior and foraging habitat use between chinstrap and gentoo penguins breeding in the South Shetland Islands, Antarctica." Marine Biology **157**(4): 811-825.

Larsson, E., V. Tremaroli, Y. S. Lee, O. Koren, I. Nookaew, A. Fricker, J. Nielsen, R. E. Ley and F. Bäckhed (2012). "Analysis of gut microbial regulation of host gene expression along the length of the gut and regulation of gut microbial ecology through MyD88." Gut **61**(8): 1124-1131.

Leotta, G. A., I. Chinen, G. B. Vigo, M. Pecoraro and M. Rivas (2006). "Outbreaks of avian cholera in Hope Bay, Antarctica." J Wildl Dis **42**(2): 259-270.

Levy, H. (2020). The Molecular Ecology of the Gentoo Penguin (Pygoscelis papua). Doctor of Philosophy, University of Oxford.

Levy, H., G. V. Clucas, A. D. Rogers, A. D. Leache, K. L. Ciborowski, M. J. Polito, H. J. Lynch, M. J. Dunn and T. Hart (2016). "Population structure and phylogeography of the Gentoo Penguin (Pygoscelis papua) across the Scotia Arc." Ecol Evol **6**(6): 1834-1853.

Lima, M. and S. A. Estay (2013). "Warming effects in the western Antarctic Peninsula ecosystem: the role of population dynamic models for explaining and predicting penguin trends." Population Ecology **55**(4): 557-565.

Loo, W. T., J. Garcia-Loor, R. Y. Dudaniec, S. Kleindorfer and C. M. Cavanaugh (2019). "Host phylogeny, diet, and habitat differentiate the gut microbiomes of Darwin's finches on Santa Cruz Island." Sci Rep **9**(1): 18781.

Lynch, H. J. (2013). Gentoo Penguin. Penguins: Natural History and Conservation. P. G. Borboroglu and P. D. Boersma, University of Washington Press.

Marschall, H.-P. (1988). "The overwintering strategy of Antarctic krill under the pack-ice of the Weddell Sea." Polar Biology **9**(2): 129-135.

Martin, M. A., J. M. Ortiz, J. Seva, V. Vidal, F. Valera, J. Benzal, J. J. Cuervo, C. de la Cruz, J. Belliure, A. M. Martinez, J. I. Diaz, M. Motas, S. Jerez, V. L. D'Amico and A. Barbosa (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**(3): 568-575.

Massom, R. A. and S. E. Stammerjohn (2010). "Antarctic sea ice change and variability – Physical and ecological implications." Polar Science **4**(2): 149-186.

Melo, A. M., V. R. Poester, M. R. Trápaga, D. A. Stevens, P. L. Canabarro, A. C. Adornes, A. P. da Silva, S. C. Estima, E. Frere, R. Sabino and M. O. Xavier (2023). "Aspergillosis in free-ranging Magellanic penguins." Medical Mycology **61**(2).

Meryman, H. T. (1956). "Mechanics of Freezing in Living Cells and Tissues." Science **124**(3221): 515-521.

Meyer, B., A. Atkinson, D. Stöbing, B. Oettl, W. Hagen and U. V. Bathmann (2002). "Feeding and energy budgets of Antarctic krill Euphausia superba at the onset of winter–I. Furcilia III larvae." Limnology and Oceanography **47**(4): 943-952.

Michel, A. J., L. M. Ward, S. K. Goffredi, K. S. Dawson, D. T. Baldassarre, A. Brenner, K. M. Gotanda, J. E. McCormack, S. W. Mullin, A. O'Neill, G. S. Tender, J. A. C. Uy, K. Yu, V. J. Orphan and J. A. Chaves (2018). "The gut of the finch: uniqueness of the gut microbiome of the Galapagos vampire finch." Microbiome **6**(1): 167.

Miller, A. K., M. A. Kappes, S. G. Trivelpiece and W. Z. Trivelpiece (2010). "Foraging-Niche Separation of Breeding Gentoo and Chinstrap Penguins, South Shetland Islands, Antarctica." The Condor **112**(4): 683-695.

Pickett, E. P., W. R. Fraser, D. L. Patterson-Fraser, M. A. Cimino, L. G. Torres and A. S. Friedlaender (2018). "Spatial niche partitioning may promote coexistence of Pygoscelis penguins as climate-induced sympatry occurs." Ecol Evol **8**(19): 9764-9778.

Rothschild, L. J. and R. L. Mancinelli (2001). "Life in Extreme Environments." Nature **409**: 1092-1101.

Samuel, B. S. and J. I. Gordon (2006). "A humanized gnotobiotic mouse model of host–archaeal–bacterial mutualism." Proceedings of the National Academy of Sciences **103**(26): 10011-10016.

Stonehouse, B. (1970). "Geographic Variation in Gentoo Penguins Pygoscelis papua." Ibis **112**(1): 52-57.

Trathan, P. N., P. Garcia-Borboroglu, D. Boersma, C. A. Bost, R. J. Crawford, G. T. Crossin, R. J. Cuthbert, P. Dann, L. S. Davis, S. De La Puente, U. Ellenberg, H. J. Lynch, T. Mattern, K. Putz, P. J. Seddon, W. Trivelpiece and B. Wienecke (2015). "Pollution, habitat loss, fishing, and climate change as critical threats to penguins." Conserv Biol **29**(1): 31-41.

Trathan, P. N. and B. Grant (2013). Adélie Penguin. Penguins: Natural History and Conservation. P. G. Borboroglu and P. D. Boersma, University of Washington Press: 37-57.

Trivelpiece, W. and S. Trivelpiece (2013). Chinstrap Penguin. Penguins: Natural History and Conservation. P. G. Borboroglu and P. D. Boersma, University of Washington Press.

Turnbaugh, P. J., V. K. Ridaura, J. J. Faith, F. E. Rey, R. Knight and J. I. Gordon (2009). "The Effect of Diet on the Human Gut Microbiome: A Metagenomic Analysis in Humanized Gnotobiotic Mice." Science Translational Medicine **1**(6): 6ra14-16ra14.

Vianna, J. A., F. A. N. Fernandes, M. J. Frugone, H. V. Figueiro, L. R. Pertierra, D. Noll, K. Bi, C. Y. Wang-Claypool, A. Lowther, P. Parker, C. Le Bohec, F. Bonadonna, B. Wienecke, P. Pistorius, A. Steinfurth, C. P. Burridge, G. P. M. Dantas, E. Poulin, W. B. Simison, J. Henderson, E. Eizirik, M. F. Nery and R. C. K. Bowie (2020). "Genome-wide analyses reveal drivers of penguin diversification." Proc Natl Acad Sci U S A **117**(36): 22303-22310.

Vianna, J. A., D. Noll, G. P. M. Dantas, M. V. Petry, A. Barbosa, D. Gonzalez-Acuna, C. Le Bohec, F. Bonadonna and E. Poulin (2017). "Marked phylogeographic structure of Gentoo penguin reveals an ongoing diversification process along the Southern Ocean." Mol Phylogenet Evol **107**: 486-498.

Walton, D. W. H. (1982). "The Signy Island terrestrial reference sites: XV. Micro-climate monitoring, 1972-74." British Antarctic Survey Bulletin **55**: 11-126.

Williams, T. D. (1991). "Foraging ecology and diet of Gentoo Penguins Pygoscelis papua at South Georgia during winter and an assessment of their winter prey consumption." Ibis (London, England) **133**(1): 3-13.

Williams, T. D. (1995). The penguins : Spheniscidae. Oxford, Oxford University Press.

Yergeau, E., K. K. Newsham, D. A. Pearce and G. A. Kowalchuk (2007). "Patterns of bacterial diversity across a range of Antarctic terrestrial habitats." Environ Microbiol **9**(11): 2670-2682.

Yew, W. C., D. A. Pearce, M. J. Dunn, A. A. Samah and P. Convey (2017). "Bacterial community composition in Adélie (Pygoscelis adeliae) and Chinstrap (Pygoscelis antarctica) Penguin stomach contents from Signy Island, South Orkney Islands." Polar Biology **40**(12): 2517-2530.

Zaiss, M. M. and N. L. Harris (2016). "Interactions between the intestinal microbiome and helminth parasites." Parasite Immunology **38**(1): 5-11.

Zeng, Y.-X., H.-R. Li, W. Han and W. Luo (2021). "Comparison of Gut Microbiota between Gentoo and Adélie Penguins Breeding Sympatrically on Antarctic Ardley Island as Revealed by Fecal DNA Sequencing." Diversity **13**(10).

Chapter 2
The Influence of Biotic and Abiotic Factors on the Bacterial Microbiome of Gentoo Penguins (*Pygoscelis papua*) in Their Natural Environment

scientific reports

Check for updates

OPEN

The influence of biotic and abiotic factors on the bacterial microbiome of gentoo penguins (*Pygoscelis papua*) in their natural environment

Chloe Kaczvinsky[1,9], Hila Levy[2,9], Stephen Preston[1,3], Casey Youngflesh[4], Gemma Clucas[5], Heather J. Lynch[6,7], Tom Hart[8,10]✉ & Adrian L. Smith[1,10]✉

The microbiome is a key factor in the health, well-being, and success of vertebrates, contributing to the adaptive capacity of the host. However, the impact of geographic and biotic factors that may affect the microbiome of wild birds in polar environments is not well defined. To address this, we determined the bacterial 16S rRNA gene sequence profiles in faecal samples from pygoscelid penguin populations in the Scotia Arc, focusing on gentoo penguins. This mesopredatory group breeds in defined colonies across a wide geographic range. Since diet could influence microbiome structure, we extracted dietary profiles from a eukaryotic 18S rRNA gene sequence profile. The bacterial microbiome profiles were considered in the context of a diverse set of environmental and ecological measures. Integrating wide geographic sampling with bacterial 16S and eukaryotic 18S rRNA gene sequencing of over 350 faecal samples identified associations between the microbiome profile and a suite of geographic and ecological factors. Microbiome profiles differed according to host species, colony identity, distance between colonies, and diet. Interestingly there was also a relationship between the proportion of host DNA (in relation to total 18S rRNA gene signal) and the microbiome, which may reflect gut passage time. Colony identity provided the strongest association with differences in microbiome profiles indicating that local factors play a key role in the microbiome structure of these polar seabirds. This may reflect the influence of local transfer of microbes either via faecal-oral routes, during chick feeding or other close contact events. Other factors including diet and host species also associate with variation in microbiome profile, and in at least some locations, the microbiome composition varies considerably between individuals. Given the variation in penguin microbiomes associated with diverse factors there is potential for disruption of microbiome associations at a local scale that could influence host health, productivity, and immunological competence. The microbiome represents a sensitive indicator of changing conditions, and the implications of any changes need to be considered in the wider context of environmental change and other stressors.

Keywords Pygoscelid penguins, Gentoo penguins, Microbiome, Diet, Monitoring

The bacterial microbiomes of vertebrates are well established as key components of a healthy host and their composition can be affected by a wide range of factors[1–4]. Experimental approaches with model species have been key to determining the role of microbiome communities on the health[5], metabolism[6], and immune response[7] of animals. While these and many other studies demonstrate the importance of a healthy microbiome in a laboratory

[1]Department of Biology, University of Oxford, 11a Mansfield Road, Oxford OX1 3SZ, UK. [2]Office of Science and Technology Policy, Executive Office of the President, 1650 Pennsylvania Avenue, Washington, DC 20504, USA. [3]UCL School of Pharmacy, University College London, 29-39 Brunswick Square, London WC1N 1AX, UK. [4]Department of Biological Sciences, Clemson University, Clemson, SC 29634, USA. [5]Cornell Lab of Ornithology, 159 Sapsucker Woods Rd., Ithaca, NY 14850, USA. [6]Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY 11794, USA. [7]Institute for Advanced Computational Sciences, Stony Brook University, Stony Brook, NY 11794, USA. [8]Oxford Brookes University, Gypsy Lane, Headington, Oxford OX3 0BP, UK. [9]These authors contributed equally: Chloe Kaczvinsky and Hila Levy. [10]These authors jointly supervised this work: Tom Hart and Adrian L. Smith. ✉email: t.hart@brookes.ac.uk; adrian.smith@biology.ox.ac.uk

environment, the biology of these interactions may be more complicated in wild systems. Studies on humans have associated microbiome changes with a wide range of circumstances, identifying health and nutritional status, disease, host genetics, diet, and geography as key factors that affect microbiome communities[1,3,5,8]. In birds, experimental studies in chickens have demonstrated an important role for diet in determining microbiome composition (e.g.[9,10]), but it is rarely considered in studies of wild vertebrates. Key microbial taxa were detected in sanguivorous finches but were absent in closely related, non-blood-feeding species[11] and the effect of diet was more important than phylogeny on microbiome composition in Darwin's finches[4]. Similarly, the microbiome composition of wild-caught great tits was variable and affected by post capture dietary manipulation (seed versus insect) during an 11-day period in captivity[12], further indicating a role of diet in wild avian microbiomes.

The microbiome of polar species holds particular interest in terms of global biogeographic patterns. Generally, free-living metazoan species distributions follow a pattern of highest diversity in equatorial regions decreasing towards the poles[13], but this pattern is much less studied in microorganisms[14], particularly microbiomes that may be more influenced by the host. Wild avian faecal microbiomes are often dominated by four key phyla—Firmicutes, Proteobacteria, Bacteroides, and Actinobacteria[15]. These same phyla also dominate penguin microbiomes[16–19]. A study of Adélie penguins from two parts of the Ross Sea found a significant correlation between genetic distance and the alpha diversity of microbiomes, concluding that these bacteria are influenced by host genetics[16]. Another study comparing the microbiomes of two sympatric Adélie and gentoo penguins found higher concentrations of Actinobacteria and Cyanobacteria in gentoo penguins compared to Adélie penguins[19]. There is also evidence of several known potential pathogens in penguin microbiomes (including *Campylobacter*, *Helicobacter*, and *Streptococcus*) but it is unclear if they are pathogenic in penguins[18]. Importantly, the aforementioned studies were relatively limited in scope, focusing on one or, at most, a few colonies, over relatively restricted distances. Here we investigate a much greater number of penguin microbiomes in multiple colonies across the broad geographic range of the Scotia Arc.

Pygoscelid penguins are a genus of three sympatric species (gentoo [*Pygoscelis papua*], chinstrap [*P. antarcticus*], and Adélie [*P. adeliae*] penguins) with a range from 46° S to 77° S[20–22]. Pygoscelid breeding colonies vary in size from dozens of pairs to half a million or more with varying distances between colonies. Differentiation also exists among these species in their dietary preferences, with Adélie and chinstrap penguins feeding almost exclusively on krill, whereas diets are more diverse in gentoo penguins, including krill, fish, and squid. These variations offer granularity and replication to study the factors that influence penguin microbiomes.

The work presented in this manuscript tests a series of hypotheses on the associations between environmental and biotic factors and microbiome composition in wild gentoo penguins across a large geographic range. We examined a large dataset of > 350 faecal samples (mostly from gentoo penguins but also including samples from chinstrap and Adélie penguins as well as blue-eyed shags [*Phalacrocorax atriceps*], a sympatric species of flying seabird). This study focussed on gentoo penguins with the other pygoscelid penguins and the blue-eyed shags acting as locally derived comparators. Samples were derived from 25 gentoo colonies across the Scotia Arc, covering a substantial geographic and latitudinal ranges. All samples were subject to molecular assessment of microbiome and diet-derived DNA as well as an indicator of the proportion of host DNA in each sample. By combining these measures, we explored the degree to which a range of biotic and abiotic factors are related to variation in the gentoo penguin microbiome.

Methods

Sample collection

Fresh faecal samples (guano) were collected from within 1–2 m of pygoscelid penguin nests distributed throughout the colony to minimize the chances of duplicating samples. Guano samples were discrete and identifiable as a single event each from a single individual (to avoid double sampling of the same individual samples were taken at least 2 nests apart). Samples were collected between the 2015/16 and 2018/19 breeding seasons (November to February) for brush-tailed penguins along the Scotia Arc (Fig. 1 and Supplemental Table 1). Samples were taken from each colony, placed in RNALater® (MFCD03453003)[23] (at approximately 1:1 volume/volume) in screw-capped Eppendorf® tubes and stored at 4 °C until shipping. These samples were frozen and stored at − 20 °C either in the UK or the Falklands/Malvinas within three weeks. In total, 381 guano samples were collected, with the majority from gentoos (n = 337), and the rest from Adélie or chinstrap penguins (n = 6 and 20 respectively). Eighteen samples were also collected from blue-eyed shags, a sympatric species of flying seabird. Figure 1 summarizes the distribution of sampling across the Scotia Arc.

DNA extraction and amplicon generation

DNA was extracted using the MoBio PowerSoil® kit (Qiagen, 51804, now branded as Qiagen QIAamp® PowerFecal® DNA Kit). The extraction kit protocol was modified such that the initial lysis step was extended to 12–18 h 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 s at 5.5 m/s with a 5-min pause between cycles). All other steps were according to the manufacturer's protocol and extracted DNA was stored at − 20 °C.

The V3-V4 region of the prokaryotic 16S rRNA gene was amplified using primers 16S _338F (ACTCCTACG GGAGGCAGCAGT)[24] and 16S_806R (GGACTACHVGGGTWTCTAAT)[25]. A ~ 170 bp fragment of the V7 region of the eukaryotic nuclear small subunit (SSU) 18S rRNA gene was amplified with the following primers 18S rRNA gene_SSU3_F (GGTCTGTGATGCCCT-TAGATG) and 18S rRNA gene_SSU3_R (GGTGTGTACAAA GGGCAGGG)[26]. In both cases, the forward primer contained a unique 7 nt 'barcode' tag to allow pooling of PCR products prior to library preparation. DNA from each guano sample was amplified in triplicate independent PCR reactions with different barcodes to quality control the output. 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

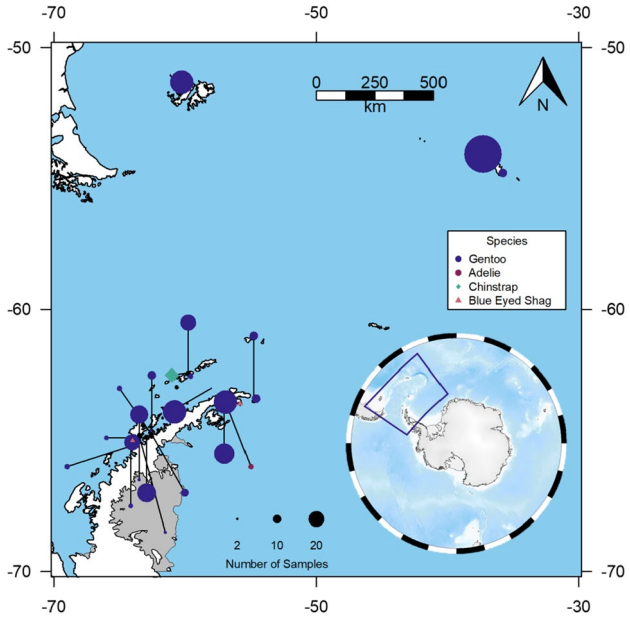


Figure 1. Colony locations and sample size across in the Scotia Arc. Gentoo penguins (n = 337, 23 colonies) are depicted as dark blue circles that vary in diameter according to the number of samples from each location. Adélie penguins (n = 6), chinstrap penguins (n = 20), and blue-eyed shags (n = 18, 3 colonies) are represented by rose square, diamond, and triangle outlines respectively. The insert depicts the location of the Scotia Arc.

manufacturer's recommendations, with 2 µL of 1/10 diluted faecal DNA. Thermal cycling conditions were 98 °C for 1 min; followed by 40 cycles of 98 °C for 15 s, 60 °C for 15 s, 72 °C for 5 s; with a final extension of 72 °C for 30 s. A water control was included for every 48 samples.

PCR products were visualised on a 1% agarose gel stained with SYBR™ Safe DNA Gel Stain. Amplicon pooling was based upon relative band intensity with reference to actual DNA concentration, of a subset of reactions analysed using the Qubit 1X dsDNA High-Sensitivity Assay Kit (Invitrogen, UK, Q33231) using a Qubit Fluorometer. Up to 96 PCR reactions from different samples with individually tagged 16S rRNA gene 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) 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 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 divided 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). Negative controls were included and are available in the data set marked as WC.

Bioinformatic analysis

Sequence data was processed using a custom python script to de-multiplex sequences, remove primer and barcode sequences. These were then processed on *DADA2*, analysing the data as a pool[27] to remove chimeras, align sequences, and assign taxonomy using the *DADA2* formatted NCBI database for 16S rRNA gene data and the SILVA 18S rRNA gene database, version 132.99[28] (script available: https://doi.org/10.6084/m9.figshare.20457378). These assignments were performed[29] using the reference data sets with Amplicon Sequence Variant (ASV) assignment set to 0.99 identity. Reads were trimmed based on error plots of the pooled data sets to minimize degradation at the end of reads. No uncalled positions (Ns) were allowed and *k*-values for the alignment algorithm were set to 3. After demonstrating that technical triplicates replicated the outputs to a high level, these data were pooled for downstream analyses (see Supplemental Table 2 for metrics). Numbers of reads remaining at each stage of processing are given in Table 1.

We used both Simpson's (alpha diversity) and Bray–Curtis (beta diversity) measures in our analyses. For alpha diversity, we first tested Simpson, Shannon, and observed richness for the non-rarefied data which exhibited similar patterns (Supplemental Fig. 1). We elected to avoid phylogenetically informed metrics due to the scarcity of data for robust tree generation, (based on a single gene fragment). Bray–Curtis was used as a measure of beta diversity, in part due to the zero-inflated nature of microbiome data, which undermined the use of methods relying on log-transformations[30]. To correct for the influence of read depth data sets were repeatedly (100 replicates) rarefied to 5098 reads/sample for 16S rRNA gene and to 51 reads/sample for 18S rRNA gene (food, 4 components). Model parameters were estimated for each iteration of rarefied data, and the 5th and 95th quantiles, the

	Raw	Filtered	Merged	Non-Chimera	Filtering
16S	10,394,014	7,862,089	7,214,834	6,295,746	5,925,116
18S	5,563,002	5,510,560	5,323,332	5,114,051	4,560,848

Table 1. Numbers of reads at each stage of bioinformatic processing. The number of reads from the input to those remaining after filtering, merging, removing chimeras, and removing those not identified to order respectively. For 18S data, the total number of reads assigned to one of the four food groups equalled 312,447.

median estimate, and the median significance value were reported. We selected this method as a balance between the problems of rarefying and normalization. The key problem with rarefying is that it removes informative data[31]. By using repeated rarefaction, we could include rare taxa, without biasing the data with differing read depths and gain a more accurate picture of the variation in and between samples. Although the food data rarefication was performed at 51 reads/sample/iteration, this was sufficient as food comprised only 4 groups. The rarefaction depths were selected to balance retaining reads and samples (See Supplemental Fig. 2A,B).

A generalised linear mixed effects model was developed in the *glmmTMB* package[32] in R[33] to test correlation between latitude and alpha diversity for 16S rRNA genes, including colony as a random intercept. This model used a beta distribution to account for the bounds of Simpson's diversity index between 0 and 1. To examine the connection between ecological metrics and diversity, PERMANOVAs were used to examine these factors against beta diversity, measured by Bray Curtis. Mantel tests were used to test whether geographic distance was an indicator of difference in associated microbiomes. This test used geographic distance estimated using Haversine distance in *geosphere*[34] and Bray–Curtis dissimilarity matrices as inputs using the mantel function in the *vegan* package in R[35]. Spearman's distance and 9999 permutations were specified for all tests, except the partial Mantel comparing all three matrices, where the Pearson's distance was used due to restrictions for input into partial Mantel tests[35,36]. The analyses were performed using the adonis2 and mantel functions respectively in the *vegan* package in R[35]. See Supplemental Fig. 3 for analysis diagram. (Full analysis script: https://doi.org/10.6084/m9.figshare.20457378).

Ethics approval

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/Malvinas 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. Animal handling protocols followed Scientific Committee on Antarctic Research Code of Conduct on the use of Animals for Scientific Purposes in Antarctica. All samples were transported in accordance with applicable export permits and United Kingdom Department for Environment, Food and Rural Affairs import permits.

Results

Examination of 6,009,327 16S rRNA gene sequences identified 8648 ASVs, and 6598 specific bacterial phyla that were retained in the analytical pipeline. The majority of these ASVs (4998) could only be assigned at the level of order and the most abundant were the Clostridiales, Bacilliales, and Pseudomonadales. The remaining 2779 ASVs were identified to one of 188 families. The average (mean) 16S rRNA gene profile of gentoo penguins was dominated by Firmicutes (62%), with significant contributions of Fusobacteria (21%), Proteobacteria (13%), and Actinobacteria (3%) (Fig. 2). All other ASVs represented less than 1% of all reads identifiable to phylum. However, it is important to indicate that the composition of the gentoo faecal 16S rRNA gene profiles was variable and influenced by multiple factors (Supplemental Figs. 4–6) across the 347 samples analysed (See Supplemental Table 1 for all samples). This variation was evident both within and between different colonies (Fig. 2) which provided a basis for considering the factors that may affect microbiome composition across different conditions. (See Supplemental Fig. 7 for an NMDS visualization of the 16S data.)

Latitude does not correlate with Simpson's diversity

To test if diversity changed with increased latitude, Simpson's diversity measures were regressed against year (season) using a beta regression in one model and against latitude and species (gentoo, Adélie, or chinstrap penguin, or blue-eyed shag) in a second model restricted to 2018 data. In both models, colony was included as a random intercept. There was no significant association between Simpson's diversity and latitude or species in this model (Tables 2, 3).

One potential determinant in microbiome composition is diet, and since this varies in pygoscelid penguins[20–22], it was relevant to generate a dietary profile relating to individuals assessed for microbiome composition. A eukaryotic 18S rRNA gene sequencing approach was adopted to identify dietary signatures in the faecal samples. From 5,114,051 total eukaryotic 18S rRNA gene reads, 1759 ASVs were identified as eukaryotic taxa including 255 that could be attributed to metazoans. The faecal 18S rRNA gene signature included a wide variety of organisms (that varied considerably across the sample set) including a large amount of algal (12.7% of raw data) and fungal (5% of raw data) sequences (Fig. 3A). A further 183 ASVs were identified as Opisthokonta,

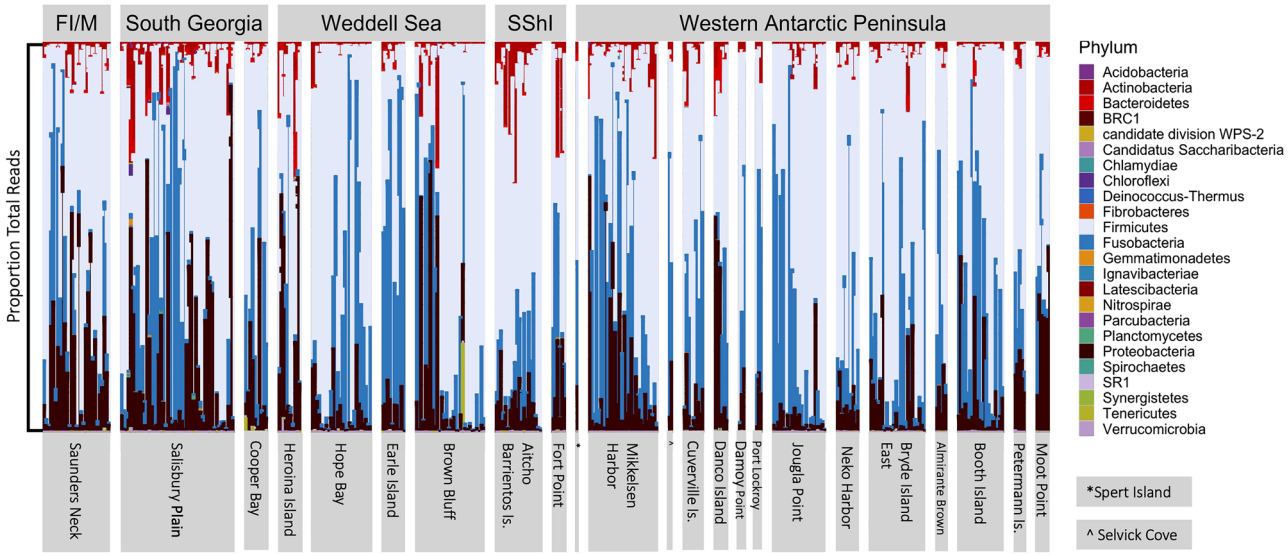


Figure 2. A stacked bar plot depicting the percentage composition of the unrarefied 16S rRNA gene profiles of gentoo penguin samples around the Scotia Sea. Individual microbiomes are reported to Phylum level and samples are arranged into regions (Falkland Islands/Malvinas = FI/M, South Georgia and South Shetland Islands = SShI, Weddell Sea, and Western Antarctic Peninsula). Within these regional groupings, samples are grouped by colony and ordered by latitude, presenting these samples in increasing latitude from the left to the right of the figure. Each bar represents a faecal sample and white space separates colonies.

	Median	5th percentile	95th percentile	Median *p* value
Intercept	0.74	0.73	0.74	1.0e−09
2016 Season	−0.042	−0.051	−0.034	0.85
2018 Season	0.0061	0.0023	0.010	0.96

Table 2. Linear mixed effects model examining season and alpha diversity. Sampling without replacement of the samples to 5098 was repeated 100 times and the model results saved for each run. The median, 5th percentile, and 95th percentile of the estimates are reported as well as the median *p* value. Samples from 2015 were not included in this model as they were limited to a single site. The intercept year was 2015.

	Median	5th percentile	95th percentile	Median *p* value
Intercept	0.21	0.17	0.23	0.90
Latitude	−0.015	−0.016	−0.015	0.56
Blue-Eyed Shag	0.64	0.62	0.66	0.20
Chinstrap	−0.35	−0.37	−0.37	0.43
Gentoo	−0.27	−0.27	−0.26	0.45

Table 3. Linear mixed effects model examining alpha diversity with latitude and species. Sampling without replacement of the samples to 5098 reads was repeated 100 times and the model results saved for each run. The median, 5th percentile, and 95th percentile of the estimates are reported here, as well as the median *p* value. This model is only focused on samples from 2018 as this was the season with the most expansive sampling among different species.

Amoebozoa, or SAR (stramenopiles [heterokonts], alveolates, and Rhizaria), unranked clades below eukaryotes. Many of the non-metazoans were most likely derived from the environment (water) and since diet was the focus of this study, all non-metazoan sequences were excluded from further analysis. Of the metazoans, 111 of 255 taxa were identifiable to Class or superclass. A large but variable proportion of host DNA was present in the metazoan 18S rRNA gene sequences. The proportions of host relative to the metazoan dataset were retained as a potential explanatory variable for other analyses (proportion of host may relate to gut passage time). Finally, the ASVs considered for more detailed analysis were restricted to those identified as food (Eumalacostraca, Decapodiformes, Neopterygii, and Thaliacea), the proportions of which varied across the gentoo faecal 18S rRNA gene dataset and across the Scotia Arc (Figs. 3B, 4). The two most prominent groups in the sample set related to Eumalacostraca (krill) and Neoptergyii (a subclass of ray-finned fish) with a few individuals also containing significant

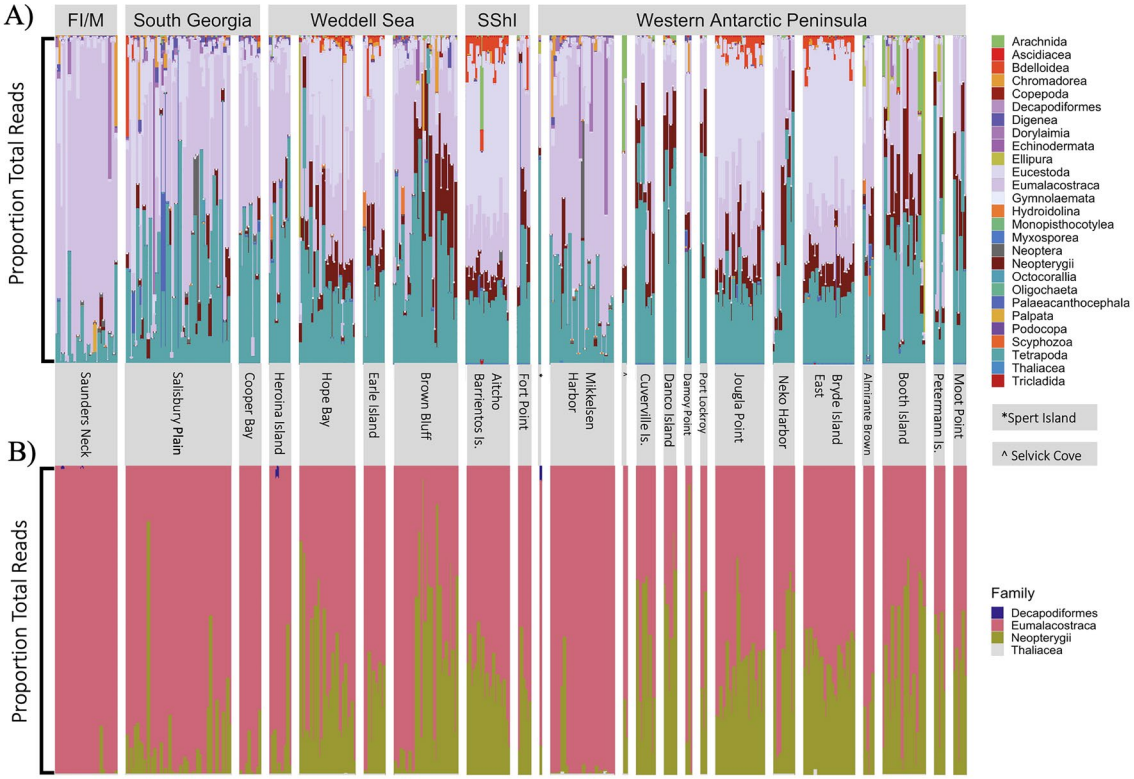


Figure 3. A stacked bar plot depicting the 18S rRNA gene results for gentoo samples, represented as percentages for ease of comparison. Samples are arranged as per this figure by region (top row) and colony (middle row). (**A**) Stacked barplot showing family assignment for all gentoo 18S rRNA gene ASVs in the metazoan data set. (**B**) Stacked barplot showing the proportions of the four determined food groups as ASVs present in faeces of gentoos, where pink represents Eumalacostraca (krill), olive represents Neoptergyii, a subclass of ray-finned fish, dark blue represents Decapodiformes, (squid) and the grey representing Thaliacea (salps). The majority of the estimated diet is attributable to krill, but there is a significant uptake of fish and clear variation between individuals.

proportions of Decapodiformes (squid) (Fig. 3B). Very small amounts of Thaliacea (salps) were detected in a small number of samples but these have been retained in the dietary group as penguins are known to occasionally consume these as a prey species[37]. (See Supplemental Fig. 7 for an NMDS visualization of the 18S data.)

Beta diversity variation according to a range of ecological factors

To test the relationship between ecological factors and diversity in microbiome or dietary signatures, Bray–Curtis dissimilarity was used as the dependent variable in the PERMANOVA models against species, colony, percent crustacean, and the percent host (Table 4). Data were also blocked by species and by colony to account for the multiple samples at each colony. This analysis was run twice, first on the whole dataset and then on the gentoo penguins alone. All analyses were performed on rarefied (sub-sampled) data, which were repeated with random sub-sampling 100 times. Median and 90% confidence intervals for the models are reported (Table 4). The results were robust to variations, providing confidence in the models. For the 16S rRNA gene profiles, colony had the strongest correlation with microbiome composition (~ 18%) with species (~ 3%), percent crustacean (~ 1%) and percent host (1.7%) also making significant contributions. Although there were differences between the host species according to this model, we are reluctant to make strong inferences due to the unbalanced dataset (between gentoo and other species). Broadly similar results were obtained when considering gentoos in isolation, with colony being the dominant explanatory factor (18.3%) and percent host (1.9%) remaining significant (Table 4). The percent of crustacean (Eumalacostraca) was not significantly correlated with microbiome when colony was included as a random effect. All other variables were statistically significant, with Pr(> F) values of less than 0.05 (Table 4).

The 18S rRNA gene models were similar, ~ 2.8% of the groupings were explained by percent host DNA when included as a factor with colony in the "all species" food model, versus ~ 5% when just examined in gentoos. Colony had the most explanatory power, explaining ~ 30% of the variation in the "all species" food model and ~ 35% in the gentoo model (Table 4). Bird species accounted for ~ 7% of the variation when examined in the "all species" food model. All variables were statistically significant, with Pr(> F) values of less than 0.05 (Table 4). PERMANOVA includes an assessment of intra-site versus inter-site variability. For a visualization with NMDS of both 16S and 18S, please see Supplemental Figs. 7 and 8.

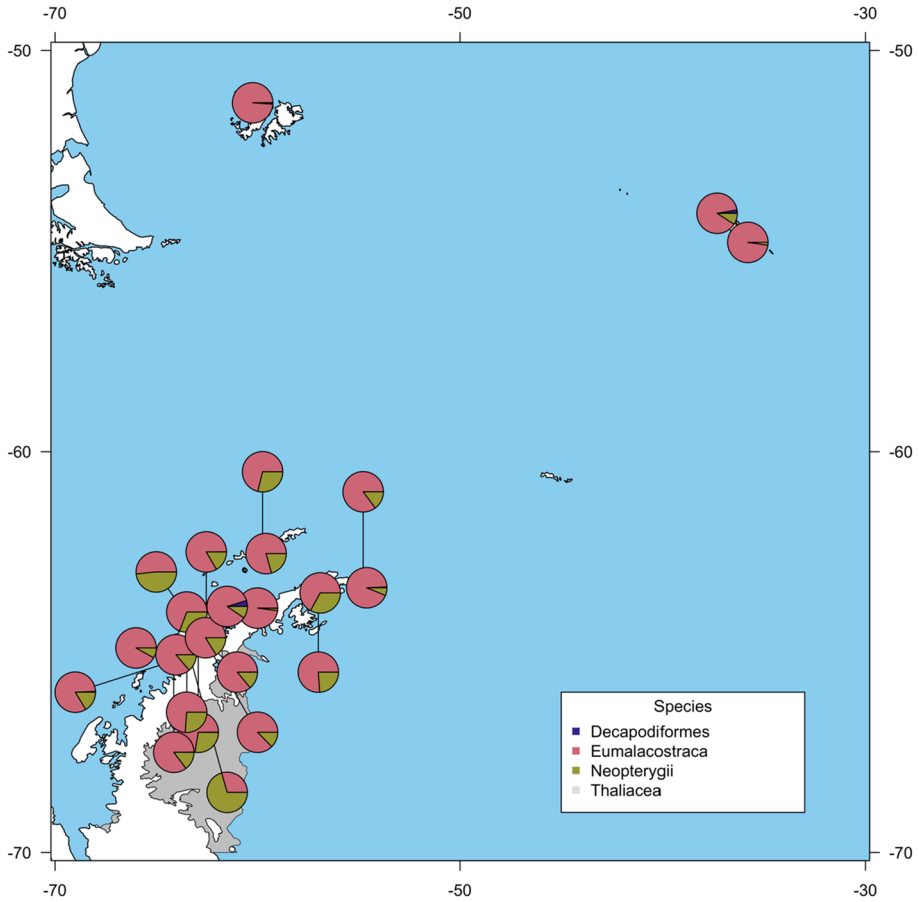


Figure 4. The relative composition of food from gentoo samples across the Scotia Arc as inferred from the 18S rRNA gene data set. Pie charts are constructed as mean proportions of food items in each colony. Pink represents krill (Eumalacostraca) and comprises most of the samples, while olive represents fish (Neopterygii). Where multiple colonies are close together, black lines lead from the expanded colony composition to its location on the map.

Beta diversity is correlated with distance

To test for a potential relationship between geographic distance and 16S rRNA gene profiles or dietary profiles (derived from18S rRNA gene sequence profiles), Bray Curtis dissimilarity matrices and distance were compared using Mantel tests. The 16S rRNA gene analysis reported a median Mantel statistic of $r = 0.200$ and p value < 0.0001. For the 18S rRNA gene food dataset, the median Mantel statistic was $r = 0.060$ and p value < 0.0036. Both tests were statistically significant, indicating that samples closer in geographic space were more similar in beta diversity (Table 5).

While the models comparing composition of 16S rRNA gene sequence profile with geographic distance and the dietary profile with geographic distance indicate that both diet and microbiome are more similar for samples closer in geographic space, the models do not support a direct correlation between diet and 16S rRNA gene profiles. Similarly, there was no significant correlation between 16S rRNA gene diversity, diet diversity, and geographic distance (Table 5).

Discussion

The biology of wild Antarctic seabirds provides an opportunity to test the relative influences of biotic and abiotic effects on the microbiome. We show that faecal microbiome profiles were variable among Antarctic seabirds but broadly similar to other wild birds in a review study of over 25 species, including penguins, ostriches, gulls, and geese[38]. In our study, host species explain about 3% of the observed variation in microbiome structure. While other studies have reported on differences in microbiome composition of penguin species in small numbers of individuals at one or a few study sites[17–19,39,40], our study encompasses over 350 samples and 25 colonies.

Colony identity had the largest effect on both 16S rRNA gene and dietary 18S rRNA gene signatures, though distance was also strongly associated with differences in these profiles. The Adélie penguin faecal microbiome composition has previously been shown to vary between colonies, though there was little evidence for an association with distance between the colonies in the Ross Sea[16]. Our study included a greater diversity of sites across the entire Scotia Arc and focussed on gentoo penguins which exhibit a greater level of philopatry than Adélie penguins[41]. Interestingly, latitude did not have a detectable association with microbiome structure

Type	Subgroup	Variable	Strata	R^2 (median)	R^2 (5th percentile)	R^2 (95th percentile)	Pr(>F)
16S rRNA Gene Profile	All species	Species		0.036	0.036	0.036	0.001
	All species	Percent Crustacean	Species	0.013	0.013	0.013	0.001
	All species	Colony	Species	0.180	0.180	0.181	0.001
		Percent Crustacean	Species	0.004	0.004	0.004	0.059
	All species	Percent Crustacean	Species	0.014	0.013	0.014	0.001
		Colony	Species	0.172	0.172	0.172	0.001
	All species	Percent Host	Species	0.017	0.017	0.017	0.001
		Colony	Species	0.173	0.172	0.173	0.001
	Gentoos only	Colony		0.183	0.183	0.184	0.001
	Gentoos only	Percent Crustacean	Colony	0.019	0.018	0.019	0.03
	Gentoos only	Percent Host	Colony	0.019	0.019	0.019	0.001
Dietary Profile	All species	Species		0.070	0.064	0.076	0.001
	All species	Colony	Species	0.340	0.326	0.354	0.001
	All species	Percent Host	Species	0.0278	0.024	0.033	0.001
		Colony	Species	0.314	0.301	0.327	0.001
	Gentoos only	Colony		0.396	0.382	0.415	0.001
	Gentoos only	Percent Host		0.053	0.047	0.059	0.001
		Colony		0.346	0.333	0.365	0.001
	Gentoos only	Percent Host	Colony	0.053	0.047	0.059	0.001

Table 4. Reporting the PERMANOVA results of bray Curtis Beta diversity. The R^2 was greatest in all models for colony. Dietary profile models only included the four groups identified as part of gentoo diets. PERMANOVA model terms are not interchangeable, so the model examining percent crustacean (Eumalacostra*ca*) and colony have been run twice with the order of terms switched. A dotted border indicates this variable is a continuation from the previous line as part of a multi-variable formula.

	16S rRNA genes versus Geography	Sig.*	18S rRNA genes versus Geography	Sig.*	16S rRNA genes versus 18S rRNA genes	Sig.*	Partial	Sig.*
median	0.20	1e−04	0.060	0.0036	−0.032	0.92	−0.024	0.84
5th percentile	0.20	1e−04	0.049	0.00069	−0.038	0.85	−0.032	0.75
95th percentile	0.20	1e−04	0.072	0.012	−0.023	0.96	−0.017	0.91

Table 5. Reporting the Mantel results using Bray Curtis Beta diversity. Both 16S rRNA gene and 18S rRNA gene data sets were correlated with geography, but tests examining correlation between 16S rRNA genes and 18S rRNA genes were non-significant. A partial Mantel test looking for correlation between 16S rRNA genes, 18S rRNA genes, and geographic distance was similarly non-significant. (*Sig. here stands for significance.)

in gentoo penguins. The effect of latitudinal gradients on diversity has been the topic of numerous studies but these have focused on free=living organisms including environmental microbes[42,43]. The internal microbiome of endothermic vertebrates could be considered more likely to be independent of climate (or latitude), but these microbes could still be exposed to latitude-associated environmental pressures during transfer between individuals. The lack of a detectable latitudinal association indicates that the pygoscelid microbiome structure may be more dependent on the host environment than exposure to the external environment. We hypothesise that microbiome transfer between penguins is common (as suggested by the dominant effect of colony identity on microbiome profiles), which might be homogenised by faecal-oral transfer (for example during stone stealing). It is also formally possible that the latitudinal range examined was too small to detect an association.

Our results for the gentoo penguins provided general support for an "isolation by distance" model in both microbiome and diet, and the importance of colony identity on both. Compared with many other factors, colony was one of the largest factors associated with microbiome diversity, accounting for around 18% of the observed variation. While there is support for dietary patterns at the colony or smaller geographic scale, there are no significant biogeographic patterns across the entire range, as indicated by the combined Mantel tests. Individually, both diet and microbiome were more similar among nearby individuals. However, there was no support for co-variation of diet and microbiome. Because the percent crustacean in gentoos was correlated with microbiome within colony, it seems that there is some influence of diet on microbiome, simply not one that appears at the broader geographical scale examined in the Mantel tests. While there is the potential confounding factor of increased relatedness within colonies due to the philopatric lifecycle of gentoo penguins, there is evidence for significant variation in both microbiome profiles and diet within colonies in this and other studies[19,44]. Despite the probability of increased genetic relatedness confounding the results, there is clear variation in diet and microbiome within and between colonies.

In some colonies there was considerable variation in the proportions of krill or bony fish in the diets, yet the colony effect (around 18% explanatory power) remained consistently stronger than the effect of percent crustacean (Eumalacostraca) (statistically significant but only explaining 2% of variance in gentoos) on the microbiome profiles. It could be that the microbiome of prey is part of what determines penguin microbiomes, thus making the geographic origin of diet a factor in penguin microbiomes. In contrast, distance and segregation (but not host genetics or environmental factors) were shown to be key drivers of microbiome structure in Antarctic krill[45]. Interestingly, the krill microbiome composition also diverged when populations were housed in different aquaria under identical physical conditions[45] showing the importance of isolation between populations in the development of divergent microbiomes. While prey species also contain microbiomes, the penguin faecal signature most closely mapped to other avian faecal microbiome compositions (with very different diets) suggesting any direct effect of prey microbiome-derived DNA was minimal. Given the philopatric nature of gentoo penguins, isolation may be a key driver of the colony-based divergence in microbiome composition[46]. It may be difficult to disentangle the independent effects of factors that covary between colonies and across geographic distances. More extensive sampling of key locations might support greater discrimination of factors in the future. Even with these potential complicating factors, the key observation that distance and colony identity represent strong associations with microbiome composition is an important finding.

Diet has been shown to be a factor in shaping the microbiome in a variety of systems[4,8,12,47–49]. Gentoo penguins are a good example where wild individuals vary in their dietary habits with some being considered individual specialists or individual generalists in terms of the primary prey species being targeted[50]. Moreover, the availability of different diet items is variable across the range of gentoo penguins and may also be subject to seasonal and annual variation which may then modify the microbiome and health of gentoo penguin populations. In this study, most of the dietary 18S rRNA gene signature comprised variable proportions of crustaceans and fish with a few individual samples showing high proportions of squid. This pattern is consistent with the known dietary preference of gentoo penguins, with squid being an occasional opportunistic food source in the region[51]. The interrelationship of diet and microbiome is further demonstrated by the correlation of percent crustacean to beta diversity of the microbiome. It is important to note that the dietary signature calculated from faecal samples most likely represents the most recent meal rather than a longer-term diet of the individual, but it is interesting to note that at some sites most individuals contained similar profiles whereas at other sites the dietary signature was more variable. A more detailed analysis of dietary effect might be possible with greater sampling of sites where diet was identified as a variable within a single colony.

Within the 18S rRNA gene signature, we were not surprised to find that a significant proportion of the sequences were derived from the host although, unexpectedly, the proportion was highly variable. When we considered proportion of host in relation to the observed microbiome structure, this was found to have a relatively small (around 4%) but significant association. We propose that this effect may be related to gut passage time. With longer gut passage times (and a longer time since feeding), we might expect a greater proportion of host to other metazoan (mostly diet) DNA, since we would expect digestive processes to reduce the amount of diet-associated DNA. Gut passage time in humans has been shown to affect the microbiome structure[52]. It could be useful to test the correlation of host DNA percentage with the colour of the faeces in the sample, as green faeces have been attributed to higher amounts of bile and longer fasts[40,53]. Interestingly, the gut microbiomes of a range of penguins were shown to alter during the moult-fast period[39,40], although with king penguins the changes were more dramatic than other species; a finding proposed to be related to the length of the moult-fast period. It is also important to note that differing passage times might affect detectable dietary 18S rRNA gene signatures with soft-bodied prey being more prone to degradation than those where DNA is protected by hard structures (e.g., bone). There is also evidence of lipid-based bias in diet determination from DNA[54] that should be taken into account. These factors demonstrate the importance of considering the 18S rRNA gene signatures including host DNA content and how host physiology might affect both diet and microbiome signatures. Given the well-documented differences between gentoo penguins as more generalist predators compared to the more specialist chinstrap and Adélie penguins, a more balanced study of all three species may be useful to segregate the diet/host/colony effects.

Conclusion

The integration of 16S rRNA gene and dietary 18S rRNA gene sequence data provided a powerful platform to determine the relationship between microbiome and diet and the factors that drive variation in these processes over space and time. In the most studied systems (humans and laboratory animals), the microbiome is well known to be a product of complex interactions including host genetics, location, diet, general health, stress and infection (reviewed in[1,3,48]). This study shows that penguin microbiomes are associated with a similarly complex framework of ecological, physiological, and dietary factors. The most important factors included colony identity and geographic distance between colonies, but there were others (host, species, proportion of host, and dietary composition) that were also associated with microbiome composition. An integrated approach is relevant to our understanding of variation in microbiomes in natural environments. Moreover, these data provide a baseline for continued monitoring of the microbiome in polar birds and may provide a tool to monitor the health of these species as they experience the ongoing effects of environmental change into the future.

Data availability

The datasets generated and/or analysed during the current study are available in the Sequence Read Archive (SRA) repository under project PRJNA956456, https://www.ncbi.nlm.nih.gov/bioproject/956456. All scripts used are available at figshare, https://doi.org/10.6084/m9.figshare.20457378.

Received: 5 July 2023; Accepted: 1 July 2024
Published online: 02 August 2024

References

1. Blaser, M., Bork, P., Fraser, C., Knight, R. & Wang, J. The microbiome explored: Recent insights and future challenges. *Nat. Rev. Microbiol.* **11**, 213–217 (2013).
2. Bodawatta, K. H., Hird, S. M., Grond, K., Poulsen, M. & Jønsson, K. A. Avian gut microbiomes taking flight. *Trends Microbiol.* **30**, 268–280 (2022).
3. Knight, R. *et al.* The microbiome and human biology. *Annu. Rev. Genom. Hum. Genet.* **18**, 65–86 (2017).
4. Loo, W. T., Garcia-Loor, J., Dudaniec, R. Y., Kleindorfer, S. & Cavanaugh, C. M. Host phylogeny, diet, and habitat differentiate the gut microbiomes of Darwin's finches on Santa Cruz Island. *Sci. Rep.* **9**, 18781 (2019).
5. Gérard, P. *et al.* Gnotobiotic rats harboring human intestinal microbiota as a model for studying cholesterol-to-coprostanol conversion. *FEMS Microbiol. Ecol.* **47**, 337–343 (2004).
6. Samuel, B. S. & Gordon, J. I. A humanized gnotobiotic mouse model of host–archaeal–bacterial mutualism. *Proc. Natl. Acad. Sci.* **103**, 10011–10016 (2006).
7. Larsson, E. *et al.* Analysis of gut microbial regulation of host gene expression along the length of the gut and regulation of gut microbial ecology through MyD88. *Gut* **61**, 1124–1131 (2012).
8. Turnbaugh, P. J. *et al.* The effect of diet on the human gut microbiome: A metagenomic analysis in humanized gnotobiotic mice. *Sci. Transl. Med.* **1**, 6ra14-16ra14 (2009).
9. De Cesare, A. *et al.* Effect of a low protein diet on chicken ceca microbiome and productive performances. *Poult. Sci.* **98**, 3963–3976 (2019).
10. Thibodeau, A. *et al.* Chicken caecal microbiome modifications induced by campylobacter jejuni colonization and by a non-antibiotic feed additive. *PLoS ONE* **10**, e0131978 (2015).
11. Michel, A. J. *et al.* The gut of the finch: Uniqueness of the gut microbiome of the Galapagos vampire finch. *Microbiome* **6**, 167 (2018).
12. Davidson, G. L. *et al.* Diet induces parallel changes to the gut microbiota and problem solving performance in a wild bird. *Sci. Rep.* **10**, 20783 (2020).
13. Brown, J. H. *Analytical Biogeography* 57–89 (Springer, 1988).
14. Martiny, J. B. *et al.* Microbial biogeography: Putting microorganisms on the map. *Nat. Rev. Microbiol.* **4**, 102–112 (2006).
15. Grond, K., Sandercock, B. K., Jumpponen, A. & Zeglin, L. H. The avian gut microbiota: Community, physiology and function in wild birds. *J. Avian Biol.* **49**, e01788 (2018).
16. Banks, J. C., Cary, S. C. & Hogg, I. D. The phylogeography of Adelie penguin faecal flora. *Environ. Microbiol.* **11**, 577–588 (2009).
17. Barbosa, A. *et al.* Age-related differences in the gastrointestinal microbiota of chinstrap penguins (*Pygoscelis antarctica*). *PLoS ONE* **11**, e0153215 (2016).
18. Dewar, M. L. *et al.* Interspecific variations in the gastrointestinal microbiota in penguins. *MicrobiologyOpen* **2**, 195–204 (2013).
19. Zeng, Y.-X., Li, H.-R., Han, W. & Luo, W. Comparison of gut microbiota between gentoo and Adélie penguins breeding sympatrically on Antarctic Ardley Island as revealed by fecal DNA sequencing. *Diversity* **13**, 500 (2021).
20. Lynch, H. J. Gentoo penguin. In *Penguins: Natural History and Conservation* (eds Borboroglu, P. G. & Boersma, P. D.) (University of Washington Press, 2013).
21. Trathan, P. N. & Grant, B. Adélie penguin. In *Penguins: Natural History and Conservation* (ed. Borboroglu, P. G.) 37–57 (University of Washington Press, 2013).
22. Trivelpiece, W. & Trivelpiece, S. Chinstrap penguin. In *Penguins: Natural History and Conservation* (eds Borboroglu, P. G. & Boersma, P. D.) (University of Washington Press, 2013).
23. Edwards, J., Hoffbeck, C., West, A. G., Pas, A. & Taylor, M. W. 16S rRNA gene-based microbiota profiles from diverse avian faeces are largely independent of DNA preservation and extraction method. *Front. Microbiol.* **14**, 1239167 (2023).
24. Lane, D. J. 16S/23S rRNA sequencing. In *Nucleic Acid Techniques in Bacterial Systematic* (eds Stackebrandt, E. & Goodfellow, M.) 115–175 (Wiley, 1991).
25. Caporaso, J. G. *et al.* Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci. USA* **108**(Suppl 1), 4516–4522 (2011).
26. McInnes, J. C., Emmerson, L., Southwell, C., Faux, C. & Jarman, S. N. Simultaneous DNA-based diet analysis of breeding, non-breeding and chick Adélie penguins. *R. Soc. Open Sci.* **3**, 150443 (2016).
27. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
28. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590-596 (2013).
29. McMurdie, P. J. & Holmes, S. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* **8**, e61217 (2013).
30. Weiss, S. *et al.* Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* **5**, 27 (2017).
31. McMurdie, P. J. & Holmes, S. Waste not, want not: Why rarefying microbiome data is inadmissible. *PLoS Comput. Biol.* **10**, e1003531 (2014).
32. Brooks, M. E. *et al.* glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *R J.* **9**, 378–400 (2017).
33. R Development Core Team. *R: A Language and Environment for Statistical Computing* (2021).
34. Hijmans, R. J. *geosphere: Spherical Trigonometry* (2021).
35. Oksanen, J., Simpson, G. L., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Solymos, P., Stevens, M. H. H., Szoecs, E., *et al. vegan: Community Ecology Package* (2022).
36. Legendre, P. & Legendre, L. *Numerical Ecology* (Elsevier, Oxford, 2012).
37. Thiebot, J.-B. *et al.* Jellyfish and other gelata as food for four penguin species—Insights from predator-borne videos. *Front. Ecol. Environ.* **15**, 437–441 (2017).
38. Grond, K., Sandercock, B. K., Jumpponen, A. & Zeglin, L. H. The avian gut microbiota: Community, physiology and function in wild birds. *J. Avian Biol.* **49**, e01788 (2018).
39. Dewar, M. L. *et al.* Influence of fasting during moult on the faecal microbiota of penguins. *PLoS ONE* **9**, e99996 (2014).
40. Lee, W. Y. *et al.* Faecal microbiota changes associated with the moult fast in Chinstrap and Gentoo penguins. *PLoS ONE* **14**, e0216565 (2019).
41. Clucas, G. V. *et al.* Comparative population genomics reveals key barriers to dispersal in southern ocean penguins. *Mol. Ecol.* **27**, 4680–4697 (2018).
42. Yergeau, E., Newsham, K. K., Pearce, D. A. & Kowalchuk, G. A. Patterns of bacterial diversity across a range of Antarctic terrestrial habitats. *Environ. Microbiol.* **9**, 2670–2682 (2007).
43. Zhou, J. *et al.* Temperature mediates continental-scale diversity of microbes in forest soils. *Nat. Commun.* **7**, 12083 (2016).

44. Polito, M. J. *et al.* Contrasting specialist and generalist patterns facilitate foraging nichepartitioning in sympatric populations of Pygoscelis penguins. *Mar. Ecol. Prog. Seri.* **519**,221–237 (2015).
45. Clarke, L. J. *et al.* Bacterial epibiont communities of panmictic Antarctic krill are spatially structured. *Mol. Ecol.* **30**, 1042–1052 (2021).
46. Pertierra, L. R. *et al.* Cryptic speciation in Gentoo penguins is driven by geographic isolation and regional marine conditions: Unforeseen vulnerabilities to global change. *Divers. Distrib.* **26**, 958–975 (2020).
47. Loo, W. T., Dudaniec, R. Y., Kleindorfer, S. & Cavanaugh, C. M. An inter-island comparison of Darwin's finches reveals the impact of habitat, host phylogeny, and island on the gut microbiome. *PLoS ONE* **14**, e0226432 (2019).
48. Clavel, T., Lagkouvardos, I., Blaut, M. & Stecher, B. The mouse gut microbiome revisited: From complex diversity to model ecosystems. *Int. J. Med. Microbiol.* **306**, 316–327 (2016).
49. Davies, C. S. *et al.* Immunogenetic variation shapes the gut microbiome in a natural vertebrate population. *Microbiome* **10**, 41 (2022).
50. Herman, R. W. *et al.* Seasonal consistency and individual variation in foraging strategies differ among and within Pygoscelis penguin species in the Antarctic Peninsula region. *Mar. Biol.* **164**, 115 (2017).
51. Clausen, A. P. & Pütz, K. Recent trends in diet composition and productivity of Gentoo, Magellanic and Rockhopper penguins in the Falkland Islands. *Aquat. Conserv. Mar. Freshw. Ecosyst.* **12**, 51–61 (2002).
52. Asnicar, F. *et al.* Blue poo: Impact of gut transit time on the gut microbiome using a novel marker. *Gut* **70**, 1665–1674 (2021).
53. Lee, W. Y. Avian gut microbiota and behavioral studies. 한국조류학회 **22**, 1–11 (2015).
54. Thomas, A. C., Jarman, S. N., Haman, K. H., Trites, A. W. & Deagle, B. E. Improving accuracy of DNA diet estimates using food tissue control materials and an evaluation of proxies for digestion bias. *Mol. Ecol.* **23**, 3706–3718 (2014).

Acknowledgements

We thank Quark Expeditions Ltd, Cheeseman Ecology Safaris Ltd, Pelagic Expeditions Ltd and Oceanwide Expeditions Ltd for logistical support in the field. We also gratefully acknowledge fieldworkers, including some of those who contributed over multiple seasons, like Caitlin Black, Fiona Jones and others from Penguin Watch, Oceanites and Stony Brook University. We thank all of those associated with the Comparative Infection and Immunology Group and the Polar Ecology and Conservation Research Group, especially Laure Cugniere for support and comments on the study and manuscript. The work conducted by H Levy was carried out in a non-governmental capacity and does not represent the views of the White House or the Office of Science and Technology Policy.

Author contributions

C.K., H.L., C.Y., H.J.L., A.S., T.H. jointly conceived of the study. T.H., H.L., C.Y., G.C. collected samples. Initial molecular work conducted by H.L. Statistics and coding conducted by C.K., S.P., H.L. C.K. and H.L. drafted the paper, which all authors contributed to. A.S. and T.H. supervised all aspects of the study.

Funding

This work was funded by an Oxford-Oxitec DPhil Scholarship and an Antarctic Science Bursary (CK), by the John Ellerman Foundation, the Save our Seas Foundation and by numerous donations on board Quark Expeditions and Cheesemans' Ecology Surfaris (TH), and University of Oxford Clarendon Fund Scholarship and the U.S. Department of Veteran's Affairs G.I. Bill (HL).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-66460-9.

Correspondence and requests for materials should be addressed to T.H. or A.L.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2024

Chapter 3 Prefix

A Brief Introduction to Helminths in Pygoscelids

A Brief Introduction to Helminths in Pygoscelids

There is a general scarcity of information on intestinal parasites of pygoscelids. Generally, intestinal parasite diversity has been found to be lower in pelagic birds than other seabirds that feed in shallower waters and closer to shore (Hoberg 2005, Diaz, Fusaro et al. 2017). However, even in comparison with similar taxa, pygoscelids have a relatively sparse set of internal helminth parasites, possibly due to a restricted diet that would also limit exposure to these parasites in intermediate hosts (Diaz, Fusaro et al. 2017). Despite this, there have been 6 recorded helminth species in Antarctic and sub-Antarctic pygoscelids, though some are identified only to genus (Brandão, Moreira et al. 2014, Diaz, Fusaro et al. 2017). These seem to span a range of orders within tapeworms (Cestoda), including Tetrabothriidea, Cyclophyllidea, and Diphyllobothriidea (Brandão, Moreira et al. 2014, Diaz, Fusaro et al. 2017).

Studies of internal parasites in penguins have generally been reliant on necroscopies (Martin, Ortiz et al. 2016, González-Acuña, Moreno et al. 2021), which limits the understanding of parasite loads throughout the population as these could bias findings towards sick and non-representative birds. There have been some studies looking at fecal samples from birds to detect the presence of parasites that focus primarily on the detection and identification of tapeworm eggs in faeces (Kleinertz, Christmann et al. 2014). Further, an experimental study in chinstrap chicks found a six percent lesser body mass for control chicks versus those treated with anti-helminth drugs (Palacios, Valera et al. 2012), providing experimental evidence for the impact of helminth infection on average specimens.

There are two primary potential routes for tapeworm infection in pygoscelids, through an intermediate fish host and through an intermediate krill host (Diaz, Fusaro et al. 2017). Given the dietary differences between these species, a fish route would likely lead to much greater infections

in gentoo (*Pygoscelis papua*), as postulated for *Tetrabothrius* species (Diaz, Fusaro et al. 2017). However, for many of the species found in pygoscelids, the intermediate hosts are unknown.

Two key parasite species from different orders seem to predominate in pygoscelids, *Parochites zederi and Tetrabothrius pauliani* (Martin, Ortiz et al. 2016, Diaz, Fusaro et al. 2017, González-Acuña, Moreno et al. 2021). In a necroscopy focused study, Martin et al. found almost half of the pygoscelid small intestines contained *Parochites zederi* and over 60% of those with worms had lesions associated with intestinal parasites (Martin, Ortiz et al. 2016). The presence of these helminths was associated with lesions, hemorrhage, and edemas, which probably affected the heath of the birds; (Martin, Ortiz et al. 2016) however, the reliance on necroscopy could predispose the work to those birds that were already sick or weak.

Tetrabothriidea

The life cycle of *Tetrabothrius* species is currently unknown, though crustaceans and fish have been suggested as first and second intermediate hosts respectively (Diaz, Fusaro et al. 2017, Rocka 2017). Further understanding of the lifecycle of these helminths is unresolved, though the bothridia (the adhesive mouthpiece) is unlikely to develop until ingested by the definitive host, as is the case in some *Tetrabothrius* (Mariaux 2017). Ultimately, *Tetrabothrius* species are known to parasitize pygoscelids, likely through fish and or crustaceans, but further information on their life cycle is uncertain.

Cyclophillidea

One of the primary helminths found in pygoscelids is *Parochites zederi*, a member of the Cyclophillidea and the only known example of such in pelagic birds (Diaz, Fusaro et al. 2017). *Parochites zederi* (Dilepididae) is one of the main tapeworms in Antarctic penguins, possibly due to a wide distribution of *Euphausia superba*, their posited intermediate host (Hoberg 2005, Diaz, Fusaro et al. 2016, Diaz, Fusaro et al. 2017). Lesions in the intestines associated with multiple

attached worms have been documented for this species (Martin, Ortiz et al. 2016), providing some evidence for detrimental impacts for affected birds. Distributions for species affected by *Parochites zederi* appear different between the species: prevalence is higher in the more northern South Shetland Islands for gentoos, but lower in northern locations in Adélie penguins (Martin, Ortiz et al. 2016).

Diphyllobothriidea

A *Diphyllobothrium sp.* has been reported in a gentoo carcass at Ardley Island (González-Acuña, Hernández et al. 2013), as well as other pygoscelids adults and on pygoscelid nests (Diaz, Fusaro et al. 2017). This order is common in Antarctica, but not in birds (Diaz, Fusaro et al. 2017), However, when present in birds, Diphyllobothiidea species are generally non-specific in their final host (Waeschenbach, Brabec et al. 2017). Further, this order is a common fish parasite, part of the penguin diet, and strongly associated with marine and freshwater habitats (Waeschenbach, Brabec et al. 2017, Scholz, Kuchta et al. 2019). Whether or to what degree it parasitizes penguins, however, has not been definitively established.

Infection Impacts

While it is difficult to assess the impact of helminth infection on pygoscelid mortality, there have been some studies demonstrating probable health impacts from infections. In fact, the presence of these helminths was associated with lesions, hemorrhage, and edemas, which likely affected the health of the birds (Martin, Ortiz et al. 2016). Helminth infections are not believed to be sufficient to cause mortality in infected populations, but they likely contribute to mortality in combination with other stressors (Clarke and Kerry 2000). For example, the study from (Palacios, Valera et al. 2012) described earlier concerning helminths and body mass should be considered in concert with the fact that body mass at fledging is a major factor in survival for juveniles (Ainley, Dugger et al. 2018). A separate study found an increased immune response for chicks that were

treated with anti-helminth drugs (Bertellotti, D'Amico et al. 2015). These indicate complex but potentially important impacts of helminth infection on multiple aspects of chick health and that infection could indeed contribute to lowered fitness.

Works Cited

Ainley, D. G., K. M. Dugger, M. La Mesa, G. Ballard, K. J. Barton, S. Jennings, B. J. Karl, A. Lescroël, P. O. Lyver, A. Schmidt and P. Wilson (2018). "Post-fledging survival of Adélie penguins at multiple colonies: chicks raised on fish do well." Marine Ecology Progress Series **601**: 239-251.

Bertellotti, M., V. L. D'Amico, M. G. Palacios, A. Barbosa and N. Coria (2015). "Effects of antihelminthic treatment on cell-mediated immunity in Gentoo penguin chicks." Polar Biology **39**(7): 1207-1212.

Brandão, M. L., J. Moreira and J. L. Luque (2014). "Checklist of Platyhelminthes, Acanthocephala, Nematoda and Arthropoda parasitizing penguins of the world." Check List **10**(3).

Clarke, J. and K. Kerry (2000). "Diseases and Parasites of Penguins." Penguin Conservation **13**(1): 5-24.

Diaz, J. I., B. Fusaro, L. Longarzo, N. R. Coria, V. Vidal, V. D'Amico and A. Barbosa (2016). "Gastrointestinal helminths of Adélie penguins (Pygoscelis adeliae) from Antarctica." Polar Research **35**(1).

Diaz, J. I., B. Fusaro, V. Vidal, D. González-Acuña, E. S. Costa, M. Dewar, R. Gray, M. Power, G. Miller, M. Blyton, R. Vanstreels and A. Barbosa (2017). Macroparasites in Antarctic Penguins. Biodiversity and Evolution of Parasitic Life in the Southern Ocean**:** 183-204.

González-Acuña, D., J. Hernández, L. Moreno, B. Herrmann, R. Palma, A. Latorre, G. Medina-Vogel, M. J. Kinsella, N. Martín, K. Araya, I. Torres, N. Fernandez and B. Olsen (2013). "Health evaluation of wild gentoo penguins (Pygoscelis papua) in the Antarctic Peninsula." Polar Biology **36**(12): 1749-1760.

González-Acuña, D. A., L. Moreno, M. Wille, B. Herrmann, M. J. Kinsella and R. L. Palma (2021). "Parasites of chinstrap penguins (Pygoscelis antarctica) from three localities in the Antarctic Peninsula and a review of their parasitic fauna." Polar Biology **44**(11): 2099-2105.

Hoberg, E. P. (2005). Marine Birds and Their Helminth Parasites. Marine Parasitology. K. Rohde. Collingwood, Vic, SIRO Publishing.

Kleinertz, S., S. Christmann, L. M. Silva, J. Hirzmann, C. Hermosilla and A. Taubert (2014). "Gastrointestinal parasite fauna of Emperor Penguins (Aptenodytes forsteri) at the Atka Bay, Antarctica." Parasitol Res **113**(11): 4133-4139.

Mariaux, J. K. R. H. E. P. (2017). "Tetrabothriidea Baer, 1954." In Planetary Biodiversity Inventory (2008-2017): Tapeworms from Vertebrate Bowels of the Earth.

Martin, M. A., J. M. Ortiz, J. Seva, V. Vidal, F. Valera, J. Benzal, J. J. Cuervo, C. de la Cruz, J. Belliure, A. M. Martinez, J. I. Diaz, M. Motas, S. Jerez, V. L. D'Amico and A. Barbosa (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**(3): 568-575.

Palacios, M. J., F. Valera and A. Barbosa (2012). "Experimental assessment of the effects of gastrointestinal parasites on offspring quality in chinstrap penguins (Pygoscelis antarctica)." Parasitology **139**(6): 819-824.

Rocka, A. (2017). Cestodes and Nematodes of Antarctic Fishes and Birds. Biodiversity and Evolution of Parasitic Life in the Southern Ocean. S. Klimpel, T. Kuhn and H. Mehlhorn. Cham, Springer International Publishing**:** 77-107.

Scholz, T., R. Kuchta and J. Brabec (2019). "Broad tapeworms (Diphyllobothriidae), parasites of wildlife and humans: Recent progress and future challenges." Int J Parasitol Parasites Wildl **9**: 359-369.
Waeschenbach, A., J. Brabec, T. Scholz, D. T. J. Littlewood and R. Kuchta (2017). "The catholic taste of broad tapeworms – multiple routes to human infection." International Journal for Parasitology **47**(13): 831-843.

Chapter 3
Landscape Level Interactions Between Cestodes and the Microbiome

Landscape Level Interactions Between Cestodes and the Microbiome

Authors: Chloe Kaczvinsky[1], Hila Levy[2], Stephen Preston[1,3], Casey Youngflesh,[4] Gemma Clucas[5], Heather J. Lynch[6,7], Anna V. Protasio[8], Maria Christodolou[9], Adrian L. Smith[1]†, Tom Hart[10]†

†co-senior authors and co-corresponding adrian.smith@biology.ox.ac.uk and t.hart@brookes.ac.uk

Affiliations:

1. Department of Biology, University of Oxford, 11a Mansfield Road, Oxford, OX1 3SZ. https://orcid.org/0000-0003-2802-2976

2. Office of Science and Technology Policy, Executive Office of the President, 1650 Pennsylvania Avenue, Washington, DC, 20504

3. UCL School of Pharmacy, University College London, 29-39 Brunswick Square London WC1N 1AX

4. Department of Biological Sciences, Clemson University, Clemson, SC 29634 ORCID: 0000-0001-6343-3311

5. Cornell Lab of Ornithology, 159 Sapsucker Woods Rd., Ithaca, NY 14850, USA

6. Department of Ecology & Evolution, Stony Brook University, Stony Brook, NY 11794

7. Institute for Advanced Computational Sciences, Stony Brook University, Stony Brook, NY 11794

8. Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge, CB2 1QP.

9. Department of Statistics, University of Oxford, 24-29 St Giles′, Oxford OX1 3LB.

10. Oxford Brookes University, Gypsy Lane, Headington, Oxford. OX3 0BP. https://orcid.org/0000-0002-4527-5046

Abstract

Parasitic infections are an important stress factor for wildlife worldwide and one that may become increasingly important in a multiple stressor scenario, particularly since the impacts on the health of free-living animals can be nonlinear. In Antarctic birds, particularly gentoo penguins, studies on helminth infections have been limited, often focusing on necropsies of a few individuals, with scarce epidemiological data. We employed a molecular approach to investigate the epidemiology of tapeworms (Cestoda) across more than 325 faecal samples collected from gentoo penguins at more than 20 locations within the Scotia Arc. We screened for tapeworm 18S DNA sequences and found a wide variation in tapeworm DNA abundance, ranging from undetectable levels to over 80% of the sequences, which is indicative of an over-dispersed distribution typical of tapeworm infections. The proportion of tapeworm DNA was significantly correlated with alterations in the microbiome composition, suggesting that these infections may influence gut microbiota, with potential consequences for penguin health. This study highlights the need for a comprehensive understanding of both direct and indirect effects of these pathogens on individual and population health.

Introduction

Parasite life cycles are usually studied in laboratory or manipulated studies with a relatively small sample size of host individuals. Scaling such studies to large wild populations has been difficult historically due to the level of sampling required and the lack of experimental manipulation in many cases, but molecular microbiome approaches make large-scale surveys of wild populations tractable. In Antarctic penguins, the difficulty of these parasite surveys is compounded by the remote location and freezing conditions. Such landscape level surveys are, however, a priority in the polar regions, since the extreme seasonality and large variations in climate across latitudes might influence parasite persistence and diversity. Features of polar ecosystems (such as freezing of colony sites outside the breeding season) likely make polar parasitology differ from better documented temperate and tropical models. Freezing temperatures may kill parasites outside the host and impact some parasites with direct transmission. This impact also likely differs between higher and lower latitudes.

While there is limited research on parasites in Antarctic fauna, helminths, i.e. parasitic worms, generally have lower diversity compared to their temperate and terrestrial counterparts when studied. Intestinal parasite diversity has been found to be lower in pelagic species than those that spend the majority of time on land (Diaz, Fusaro et al. 2017). The majority of helminths in these pelagic species are generally tapeworms (Cestoda), and this remains the case in penguins (Hoberg 2005, Diaz, Fusaro et al. 2017). There have been 13 recorded helminth species in Antarctic and sub-Antarctic penguins (Diaz, Fusaro et al. 2017). Studies of internal parasites in penguins have generally been reliant on necroscopies (Martin, Ortiz et al. 2016, González-Acuña, Moreno et al. 2021), which limit the understanding of parasite loads throughout the population as these could bias findings towards sick and non-representative birds. There have been some studies looking at faecal samples from birds to detect the presence of parasites, but these focus

primarily on the detection and identification of tapeworm eggs in faeces (Kleinertz, Christmann et al. 2014), rather than molecular methods that can be more easily scaled.

In pygoscelid penguins, which comprise Adélie, chinstrap, and gentoo penguins in the Antarctic and sub-Antarctic islands, two key species of parasites seem to predominate, *Parochites zederi* and *Tetrabothrius pauliani* (Cielecka, Wojciechowska et al. 1992, Martin, Ortiz et al. 2016, Diaz, Fusaro et al. 2017, González-Acuña, Moreno et al. 2021). In a necroscopy focused study, Martin et al. 2016 found that 7 of 8 adult pygoscelid small intestines contained *Parochites zederi* (Martin, Ortiz et al. 2016) and were associated with lesions, haemorrhage, and oedemas, which probably affected the heath of the birds (Martin, Ortiz et al. 2016). The experimental study in chinstrap chicks that connected helminths to lesser mass increases in chicks detailed earlier (Palacios, Valera et al. 2012), also gives evidence for health impacts on birds, at least in early development stages.

Little is known about the lifecycles of the tapeworm species that have been associated with gentoo penguins and other pygoscelids. Generally, intermediate hosts are thought to be the krill and fish that dominate pygoscelid diets, though the greater reliance of chinstrap and Adélie penguins on krill versus the more flexible gentoos (Ratcliffe and Trathan 2012, Clucas, Dunn et al. 2014), does provide a potential difference in parasitic fauna and their introduction. As *Parochites zederi* belongs to the Dilepididae, which usually involve crustaceans in their lifecycle, there is an unproven supposition of krill as an intermediate host (Martin, Ortiz et al. 2016, Ahyong, Boyko et al. 2024). In one study, 7 of 9 adult pygoscelids had multiple *P. zederi* present per specimen, almost all of which were small or immature (Cielecka, Wojciechowska et al. 1992). *Tetrabothrius* species are a common parasite of pelagic birds, and thus unsurprising in pygoscelids. A lifecycle encompassing a first and second intermediate host of crustacean and fish

or cephalopod for those *Tetreaborthius* that infect pygoscelids has been proposed (Cielecka, Wojciechowska et al. 1992).

Generally, while helminth infection is not believed to be sufficient to cause mortality in infected populations, it likely contributes to mortality when in combination with other factors such as starvation or viral infection (Clarke and Kerry 2000). There have been studies highlighting the mixed and complicated relationship between host microbiomes and parasites, which can interact to protect against or promote helminth infection (Zaiss and Harris 2016). If freezing temperatures influence the life cycle or the diet of penguins, or the life cycle of helminths, then it seems likely that the rapidly changing climate in the Western Antarctic Peninsula will also lead to changes in the distribution of parasites in these populations. Therefore, a systematic examination of current helminth infection and potential ecological correlations is important to understand potential impacts of climate change on the transmission and distribution of these species.

Materials and Methods

We reanalysed the data set in Kaczvinsky et al. 2024 (Kaczvinsky, Levy et al. 2024). Broadly, these samples were collected from penguin colonies at more than 20 sites around the Scotia Arc. Within each colony, scats were taken from nests at least 2 nests apart in penguin colonies to minimize resampling of an individual. Samples were then preserved in a roughly even volume of RNALater and were frozen upon arrival to the UK. Samples were then extracted using MoBio PowerSoil® kit (Qiagen, 51804, now branded as Qiagen QIAamp® PowerFecal® DNA Kit), with the lysis step modified to a 12-18 hour bead beating on a shaker block at 65°C. Samples were then frozen at -20°C and amplified using barcoded PCR primers. Prokaryotic barcodes were amplified using the V3-V4 region of the 16S rRNA using primers 16S _338F (ACTCCTACGGGAGGCAGCAGT) (Lane 1991) and 16S_806R (GGACTACHVGGGTWTCTAAT) (Caporaso, Lauber et al. 2011). Eukaryotic DNA was

amplified using a ~170 bp fragment of the V7 region of the nuclear small subunit (SSU) 18S rRNA gene using primers 18S rRNA gene_SSU3_F (GGTCTGTGATGCCCT-TAGATG) and 18S rRNA gene_SSU3_R (GGTGTGTACAAAGGGCAGGG) (McInnes, Emmerson et al. 2016). PCR used standard reactions with the high fidelity Phusion Hot Start Flex DNA polymerase enzyme (New England Biolabs, UK, M0535), with 2 µL of a 1:10 dilution of extracted sample. Negative controls were included at the PCR step and are available in the data set as WC.

PCR products were visualised on agarose gels and normalized with a sub-group of samples quantified with the Qubit 1X dsDNA High-Sensitivity Assay Kit (Invitrogen, UK, Q33231) using a Qubit Fluorometer. Up to 96 samples were pooled and cleaned with the MinElute PCR purification kit (Qiagen, UK, 28004). Libraries were then prepared with the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® (New England Biolabs, UK, E7103S) 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 to allow multiple libraries in a lane. MiSeq TapeStation (Agilent Genomics) and qPCR (NEBNext® Library Quant Kit for Illumina®, E7630S, New England Biolabs, UK) were used to check for successful preparation of libraries. The libraries and 5% phiX were added to the MiSeq at the University of Oxford, Department of Zoology. Sequencing was divided 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) for the 16S data. 18S samples used 150 bp paired end reads on the 300-cycle MiSeq Reagent Kit v2 (Illumina, UK, MS-102-2002).

Bioinformatic Analysis

Sequence data was processed using a custom python (Foundation) script to de-multiplex sequences, remove primer and barcode sequences with demultiplex (Laros. 2023). These were

then processed on *DADA2*, analysing the data as a pool (Callahan, McMurdie et al. 2016) to remove chimeras, align sequences, and assign taxonomy using the *DADA2* formatted NCBI database for 16S rRNA gene data and the SILVA 18S rRNA gene database, version 132.99 (Quast, Pruesse et al. 2013) (script available: 10.6084/m9.figshare.20457378). These assignments were performed (McMurdie and Holmes 2013) using the reference data sets with Amplicon Sequence Variant (ASV) assignment set to .99 identity. Reads were trimmed based on error plots of the pooled data sets to minimize degradation at the end of reads. No uncalled positions (Ns) were allowed and *k*-values for the alignment algorithm were set to 3. After demonstrating that technical triplicates replicated the outputs to a high level, these data were pooled for downstream analyses as in (Kaczvinsky, Levy et al. 2024).

16S data was then filtered to exclude cyanobacteria and chloroplast DNA, and 18S filtered to remove chloroplasts and fungi. Reads that could not be identified to phyla were excluded from both 16S and 18S data sets to improve data quality. Percent host/bird was the percent of reads attributed to Aves divided by the total reads of non-chloroplasts and fungi. All avian reads were then removed from the 18S data set for downstream analysis as per Kaczvinsky et al. 2024 (Kaczvinsky, Levy et al. 2024). All non-metazoans and non-identified families were also removed at this point. Next, a food subset was created (Eumalacostraca, Decapodiformes, Neopterygii, and Thaliacea), and the percent crustacean was recorded relative to the total food reads. This was selected as an informative metric as most diet for these species was either krill or fish, with tiny amounts of the other food sources, making percent crustacean a meaningful descriptor of the most recent meal (Kaczvinsky et al 2024). For 18S, a tapeworm data subset was created that included all ASVs identified as belonging to tapeworm (recorded as Cestoda). Then, ASVs with fewer than 100 reads were excluded from the dataset, which left 10 ASVs present in the analysis, with between 50,860 reads and 142 reads in the unrarefied data set. The percentage of reads identified

as tapeworm was then identified as a percentage of the remaining metazoan reads (minus host and unidentified families) and added to the sample data for the filtered 16S gene profiles.

While the Silva 18S (Quast, Pruesse et al. 2013) database used with phyloseq was able to identify the 10 ASVs as tapeworms, further identification was not possible with that database. Therefore, the exact sequences that were collapsed into ASVs by DADA2 (from 1 to 6 sequence variations) were run using megaBLAST (Morgulis, Coulouris et al. 2008) with the nucleotide collection to refine identification. The percentage of reads for the top four tapeworm ASVs (10, 25, 48, and 50) relative to the filtered 18S dataset were also added as a variable to the 16S data.

We generated phylogenies using knowns from the BLAST database that were generally model systems or common zoonotic parasites. The phylogenies were all generated using the NJS algorithm in ape (Paradis and Schliep 2018) in R (R Development Core Team 2021), but different models (TN94, K80, JC69, and GC95) were used to better ascertain likely relationships.

Beta diversity of the associated microbial profile was calculated using Bray-Curtis diversity on the 16S gene profiles from the same samples (see Supplemental Figure 1 for PCoA). To account for the read depth bias of this metric, each analysis was run using a randomized rarefication down to 50 reads (for samples subset down to food), and samples with fewer than this excluded (removing 40 of 381). The models were then run 100 times using this rarefication process and the median, 5%, and 95% co-efficient estimate values were reported along with the median p-value. Each model and hypothesis is explored below. This was repeated for 16S data, where all samples with fewer than 5000 reads were pruned (leaving 333 samples).

Tapeworm and Microbiome Variations

We then ran a PERMANOVA using the 100 replications of rarefaction detailed above comparing the beta diversity metrics with the variation in percent tapeworm. Colony was included as an additional term in the model and as a strata term as a sensitivity test of the model. We also

generated a phylogeny with bootstraps to examine the groupings used in this study. While support was relatively low for bootstrap values, this is likely due to the limitations of 170 bps. The same tree was supported by multiple algorithms (Figure 1, See Supplemental Figures 2-4 for all trees). While none of the branches had very high support, *Parochites zederi* samples and ASV247 and ASV48 grouped consistently together with decent support for a 170 bp segment. We also built a heatmap to examine the co-occurrence of individual tapeworm ASVs (Figure 2).

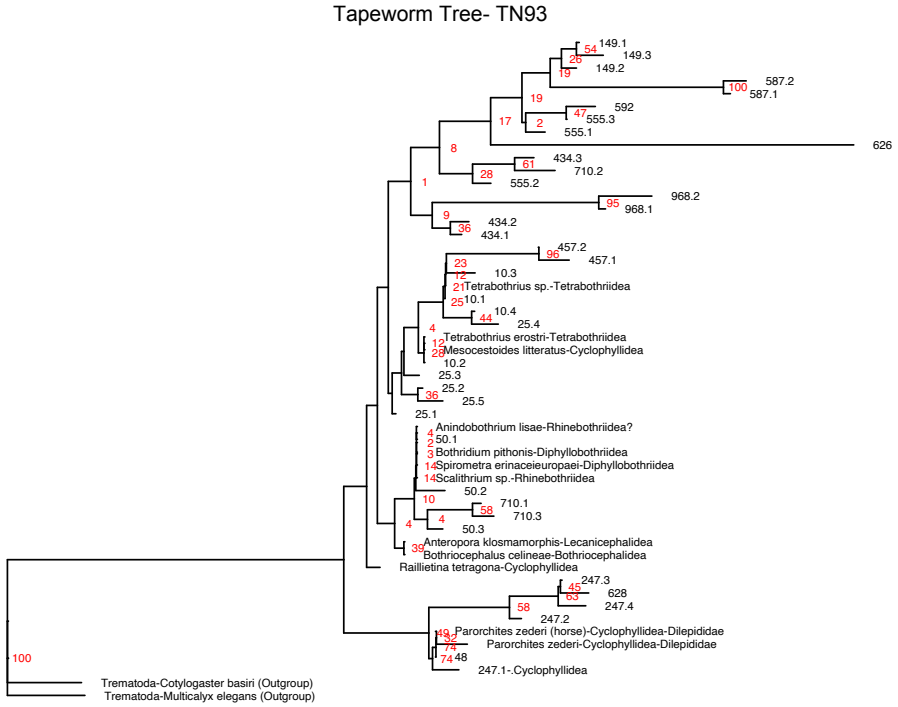


Figure 1: A phylogeny of the detected Tapeworm ASVs (numbered tips) with knowns from the NCBI BLAST database (named tips). ASVs include all of the sequences grouped under that ASV and use decimal and integers to indicate which the number of the subsequence (e.g. 149.1 is the first subsequence of ASV 149). The tree was generated using TN93 nucleotide model and the NJS neighbour joining algorithm. Bootstrap values are shown in red.

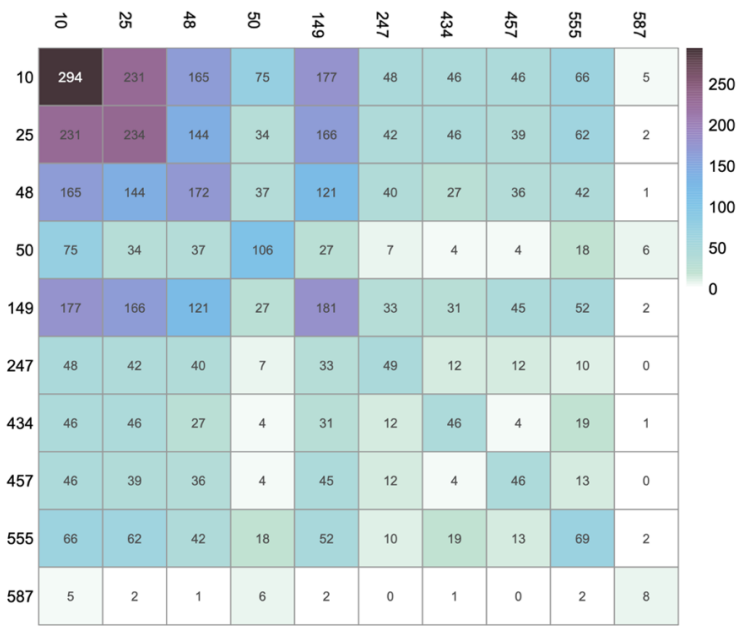


Figure 2: A heatmap showing the co-occurrence of each ASV identified as tapeworms. The center line from the top left to bottom right reports the total number of times the ASV was present in at least one read for the 337 total samples.

Species Comparison

Finally, we also had a data set of 20 samples each from chinstraps and gentoos on Aitcho Islands, both collected in 2018. A set of simple linear models examining the differences in percent tapeworm and the percent of the top three ASVs between the species were run. We also ran a differential abundance test of the microbiomes of the two species using DESeq2 (Love, Huber et al. 2014).

Results

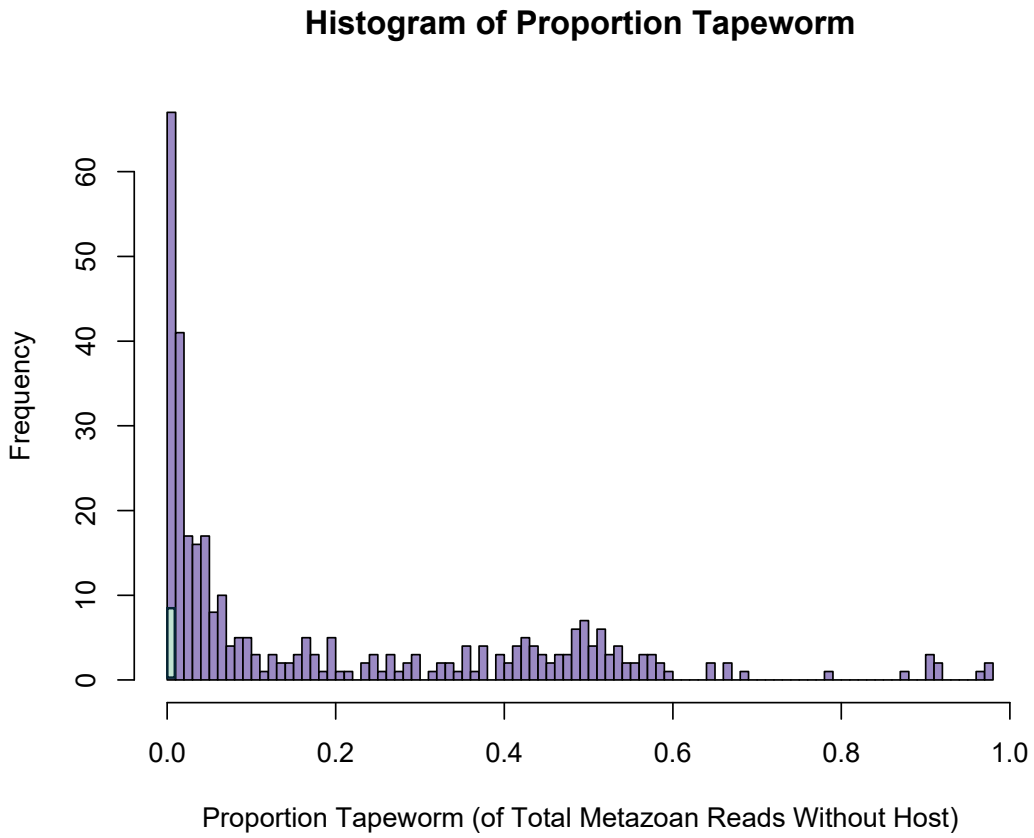


Figure 3: Histogram looking at the proportion of tapeworm (of all metazoan reads identifiable to phyla minus those identified as host) in 18S profiles. Frequency is the total number of samples in each proportion with 100 breaks (0.01 per bar). Of the 337 gentoo samples examined, 9 (2.7% of the total samples) had 0 reads identified as tapeworm, represented by the green bar on the far left of the histogram.

The percent tapeworm in the metazoan data set (minus host) ranged from 0% to 97.8% and showed a long right tail consistent with tapeworm infections across populations (Wilson, ørnstad et al. 2002) (Figure 3). Of the 337 samples in the dataset, 9 (2.7% had no reads identified as tapeworms). Infections also appeared variable by colony, with some colonies having well above the average amount of tapeworm in samples and others well below the average (mean: 0.198, std: 0.236). The mean and standard deviation were calculated from the proportion tapeworm for all 337 gentoo samples relative to the total metazoan reads identifiable to phyla minus those recorded as "Aves." All samples were weighted equally and numbers were calculated before samples were removed for rarefaction. Colonies did not show clear geographic differentiation (Figure 4).

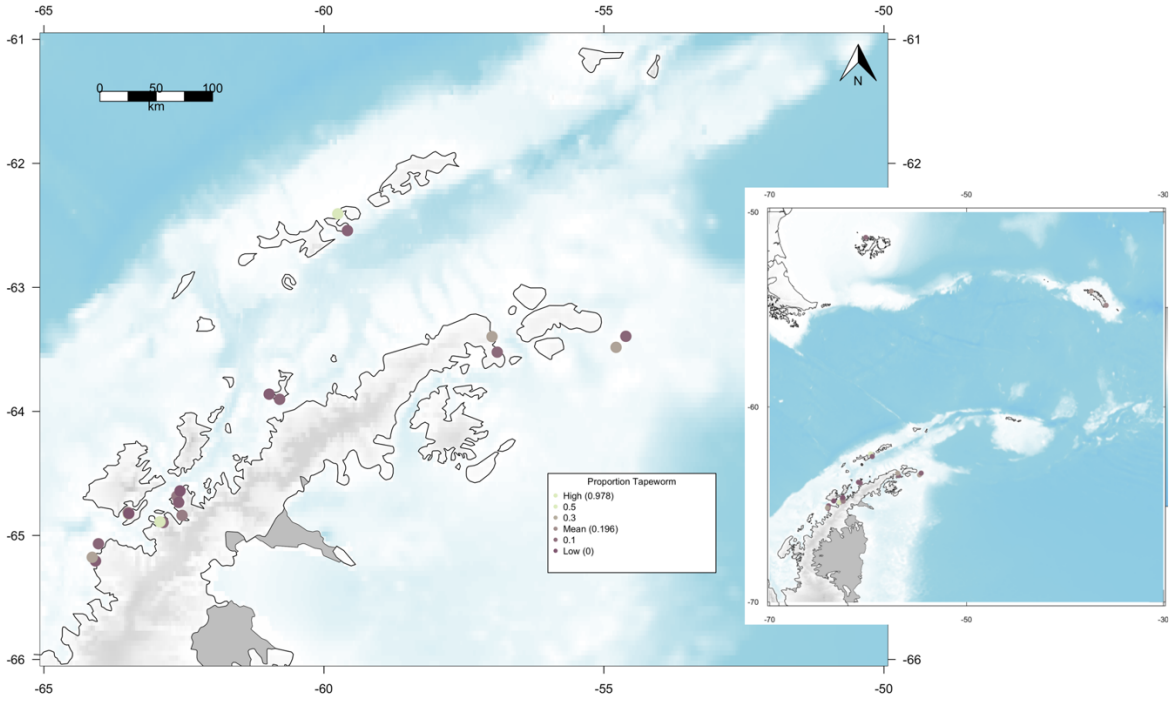


Figure 4: A map of all sample sites with coloured by proportion tapeworm (from metazoan reads excluding host), centred around the mean value across all samples of 0.196. Each colony reports the mean proportion of tapeworm for that colony of proportion tapeworm (from metazoan reads excluding host). Color scale is graduated, with examples of rough proportions shown in the legend, ranging from high proportions in green to low proportions in purple.

Identification

ASV10 is likely a *Tetrabothrius* sp. (100%- 99.42% identity though it had the same to slightly lower identity with *Mesocestoides sp.*), which is a known pathogen in pyogoscelids. The encompassing orders Tetrabothriidea and Mesocestoididea, respectively, are monophyletic in recent phylogeny with the Cyclophellidea and Nippotaeniidea and sometimes collapsed within Cyclophellidea (Waeschenbach and Littlewood 2017, Jesudoss Chelladurai and Brewer 2021). The close relationship between the orders, the absence of records of *Mesocestoides* sp. in Antarctica, and its status as a carnivore parasite (though birds are intermediate hosts) leans towards a *Tetrobothrius* sp. identification (Jesudoss Chelladurai and Brewer 2021). ASV48 had 100% identity with *Parochites zederi*, the most commonly reported pygoscelid tapeworm.

The first sequence of ASV50 had a majority (64%) of 100% identity matches in BLAST from Diphyllobothridae. This was consistent with the other two sequences grouped to this ASV,

which each had 100 matches at 94.2% identity, also with 64 matches to Diphyllobothridae. Given the current estimations of tapeworm phylogeny, this family is relatively basal (Waeschenbach and Littlewood 2017), meaning that a match to Diphyllobothridae is not improbable. However, given the uncertainty of this data, we are unwilling to further speculate on the identity of this. (See Supplemental Table 1 for example BLASTN results for ASV10). All four ASVs showed a similar right-tail skew distribution as was present in the histogram of percent tapeworm DNA distribution (see Supplemental Figure 5).

The vast majority of ASVs seem to co-occur with each other. Of the 337 samples, only nine (2.7%) had none of the top ten tapeworm ASVs, and 27 (8%) had only one of the top ten tapeworm ASVs. 89.3% of the samples had at least two tapeworm ASVs. However, ASV50 seems to be occurring primarily separately from other infections, possibly indicating a different infection mechanism, such as different intermediate host or some kind of protection against infection by a second parasite. Of the top three, all seem to have significant overlap in their occurrence with ASV10, but it could be possible that they are indicating co-infection rather than genetic variation within a species or a sequencing error. Based on the putative phylogeny of the samples and the heatmap showing high co-occurrence, we merged ASV247 with ASV48 and ASV457 with ASV10.

Diet and Infection

Many tapeworms enter the gut through diet, which we can examine through 18S signatures. Four generalized linear mixed models run with glmm (Brooks, Kristensen et al. 2017) using the ordered beta linear model looked at percent crustacean as a function of the percent ASV with colony as a random effect. None of the models showed a correlation between the tapeworm ASV and the percent crustacean in diet (p-values: 0.602, 0.266, 0.0531, 0.0531).

Tapeworm and Microbiome Variation

Table 1: PERMANOVA comparing the total percent of tapeworm reads (of all metazoan reads identifiable to phyla minus those identified as host) and beta diversity of the microbiome. Solid lines divide separate models; dashed lines indicate factors included in the same model. Colony is included as both a factor in the first model and a strata term in the second model. Models were run 100 times on separately subset (without replacement) data to 5,000 reads, and the median, 5%, and 95% percent estimates are reported as well as the median p-value.

		Strata	Median	5%	95%	P-value (median)
Gentoo	Colony		0.183	0.183	0.184	0.001*
	Percent Tapeworm		0.00932	0.00921	0.00944	0.001*
Gentoo	Percent Tapeworm	Colony	0.0114	0.0113	0.0115	0.043*

Table 2: PERMANOVA comparing the total percent of the of top tapewormASV (ASV 10, putative Iidentification as *Tetrabothrius sp.*) (of all metazoan reads identifiable to phyla minus those identified as host) and beta diversity of the microbiome. Solid lines divide separate models; dashed lines indicate factors included in the same model. Colony is included as both a factor in the first model and a strata term in the second model. Models were run 100 times on separately subset (without replacement) data to 5,000 reads, and the median, 5%, and 95% percent estimates are reported as well as the median p-value.

		Strata	Median	5%	95%	P-value (median)
Gentoo	% ASV 10		0.0117	0.0116	0.0118	0.001*
	Colony		0.179	0.178	0.179	0.001*
Gentoo	% ASV 10	Colony	0.0117	0.0116	0.0118	0.055

Table 3: PERMANOVA comparing the total percent of the of top 2nd tapeworm ASV (ASV 25) (of all metazoan reads identifiable to phyla minus those identified as host) and beta diversity of the microbiome. Solid lines divide separate models; dashed lines indicate factors included in the same model. Colony is included as both a factor in the first model and a strata term in the second model. Models were run 100 times on separately subset (without replacement) data to 5,000 reads, and the median, 5%, and 95% percent estimates are reported as well as the median p-value.

		Strata	Median	5%	95%	P-value (median)
Gentoo	% ASV 25		0.00368	0.00362	0.00374	0.153
	Colony		0.183	0.183	0.184	0.001*
Gentoo	% ASV 25	Colony	0.00368	0.00362	0.00374	0.616

Table 4: Percentage of top 3rd tapewormASV (ASV 48, Putative Identification as *Parochites zederi*) (of all metazoan reads identifiable to phyla minus those identified as host) and beta diversity of the microbiome. Solid lines divide separate models; dashed lines indicate factors included in the same model. Colony is included as both a factor in the first model and a strata term in the second model. Models were run 100 times on separately subset (without replacement) data to 5,000 reads, and the median, 5%, and 95% percent estimates are reported as well as the median p-value.

		Strata	Median	5%	95%	P-value (median)
Gentoo	% ASV 48		0.00455	0.00447	0.00463	0.057
	Colony		0.183	0.182	0.183	0.001*
Gentoo	% ASV 48	Colony	0.00455	0.00447	0.00463	0.172

Table 5: Percentage of top 3rd tapeworm ASV (ASV 50) (of all metazoan reads identifiable to phyla minus those identified as host) and beta diversity of the microbiome. Solid lines divide separate models; dashed lines indicate factors included in the same model. Colony is included as both a factor in the first model and a strata term in the second model. Models were run 100 times on separately subset (without replacement) data to 5,000 reads, and the median, 5%, and 95% percent estimates are reported as well as the median p-value.

		Strata	Median	5%	95%	P-value (median)
Gentoo	% ASV 50		0.00478	0.00471	0.00487	0.048*
	Colony		0.183	0.183	0.184	0.001*
Gentoo	% ASV 50	Colony	0.00478	0.00471	0.00487	0.026*

The same PERMANOVA models for the total amount of tapeworm present in samples were repeated only for the percent tapeworm attributed to the top four ASVs. While the presence of ASV25 was significant in all of the tests, neither ASV48 nor ASV10 were significant when the variation was examined only within colony (Tables 1-5). Generally, the presence of tapeworms is associated with significant differences in the beta diversity of the associated host microbiome, i.e. samples with greater amounts of tapeworm DNA have different host microbiomes than those with less or no tapeworm DNA present. The estimate was much lower when colony was a stratum term, i.e. the differences in variables were considered only within colony. This likely indicates that there are significant differences in microbiome beta diversity that are related to the different infection rates between colonies as demonstrated in Figure 4. That is, structure in microbiome and infection rates are associated with specific colonies.

Infection Cut-Offs

Table 6: Infection cut-off models examining different cut-offs for percentage of tapeworm as a factor in the differences in beta diversity of microbiome across samples. Each model was run as an individual PERMANOVA, with infection being a binary state set by the reported cut off for percent tapeworm present in the sample (of all metazoan reads identifiable to phyla minus those identified as host). Each model was run 100 times on a data set individually rarefied to 5,000 reads and the 5%, 95% and mean values of the model are reported here as well as the p-value.

Cut-Off	Median	5%	95%	P-value (median)
0.5%	0.0069	0.0068	0.0070	0.063
1%	0.0051	0.0050	0.0051	0.24
2%	0.010	0.010	0.010	0.010*
3%	0.0083	0.0082	0.0084	0.059
4%	0.010	0.010	0.010	0.031*
5%	0.011	0.011	0.011	0.013*
10%	0.012	0.012	0.012	0.043*
50%	0.0061	0.0060	0.0062	0.17

Based on the distribution of percent tapeworm, different PERMANOVAs were run using different cut-offs of reads identified as a percentage of the remaining metazoan reads (minus host and unidentified families), 0.5%, 1%, 2%, 3%, 4%, 5%, 10%, and 50%,. These cur-offs created a categorical model where all samples with a percent tapeworm above the cut-off would be considered infected and below which it would not. These cut-offs served as ways to try to capture the fact that the percentage representations likely contain a mix of true infections and incidental tapeworm (for example tapeworms of prey that were not infectious to penguins). Models linking host infection status and beta diversity of microbiomes showed colony and infection cut-offs varied in significance depending on the percent of tapeworm DNA selected as the cut-off. Based on (Kaczvinsky et al 2024), we know that colony has a significant impact on the microbiome of gentoos, so we restricted permutations of the data to those within colonies.

One thing to note about these results is the change from 1 to 2 %, where both the median and significance changes. While 3% infection is not significant, it has an estimate consistent with others at the middle values identified from the histogram. Above 10%, the model is not significant and the estimate changes, but there are only a few samples that would meet this threshold, reducing n, and these samples are likely ones that had proglottids contained in the extracted portion. Given this, we would speculate that for these results, around 2% tapeworm DNA is the point at which key changes can be seen in microbiome and might represent true tapeworm infections in the sample.

Differential Abundance

We found 27 bacterial ASVs with log2fold changes in abundance with a differential abundance test comparing the percent tapeworm present in each sample (compared with all metazoan reads identifiable to phyla that are not avian). These were all in one of the phyla: Fusobacteria (8), Firmicutes (12), or Proteobacteria (7) (Figure 5). 14 of the bacterial ASVs

increased with an increasing proportion of tapeworm (7 of which were classed as Fusobacteria). The other 13 showed reduced amounts with increased amounts of tapeworm in the sample. Both Firmicutes and Proteobacteria are generally considered part of the core avian microbiome, which is also the case in penguins (Banks, Cary et al. 2009, Dewar, Arnould et al. 2013, Barbosa, Balagué et al. 2016, Grond, Sandercock et al. 2018, Zeng, Li et al. 2021). The identified taxa comprise a wide variety of bacteria, including methyltropic (Ardley and Green) and spore forming bacteria (Gerritsen, Fuentes et al. 2014). When differential abundance tests were conducted on the tapeworm ASV groups isolated, they had different lists of differentially abundant bacteria, that nevertheless overlapped with each other and the broader tapeworm numbers (Supplemental Figures 6-9; Supplemental Tables 2-6).

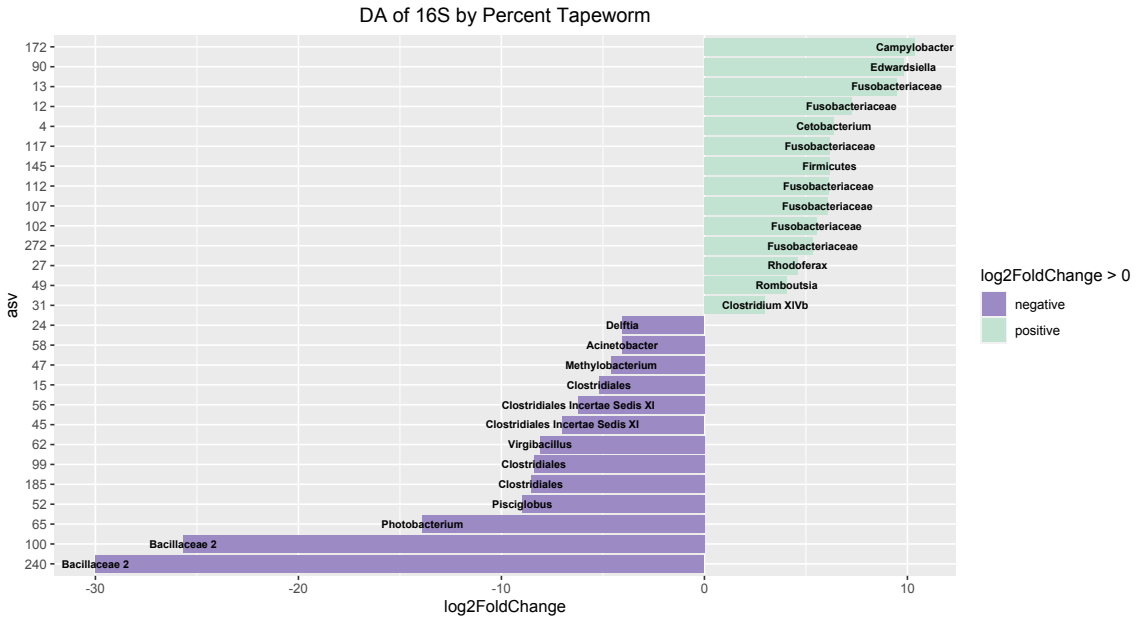


Figure 5: A visualization of the results of the differential abundance test in DESeq2 for the percent tapeworm in sample controlling for colony, showing all ASVs with log2fold changes to their lowest determined identity. Those in purple showed negative log fold changes and those in green showed positive log fold changes. The ASV number for each ASV is on the y-axis and the closest identified taxa is recorded on the bar. The length of the bar shows the magnitude of the log fold change over the model.

ASV10, ASV25, and ASV48 all showed significant negative log-fold changes in the bacterial ASV15 associated with increasing amounts of the respective tapeworm ASV. This barcode was identified as a member of the Clostridiales. Increasing amounts of tapeworm ASV10

and ASV25 were associated with significant log-fold reductions in the presence of Carnobacterium (bacterial ASV23). Increases in the presence of both tapeworm ASV48 and ASV50 were associated with log-fold changes in bacterial ASV115, though in opposite directions. This bacterial signature was identified as a *Dermaccocus* sp.

Direct Species Comparison

We also had a data set of 20 samples each from chinstraps and gentoos on Aitcho Islands, both collected in 2018. A simple linear model examining differences between the species in how much tapeworm was recorded had an estimate of 0.40 intercept for gentoos compared to 0.078 for chinstraps with a p-value of 2.1e-14 and 0.0023, respectively. The gentoos had lower percentages of ASV10 (estimates of 0.36530 for chinstrap intercept and -.21, with p value of 0.0018 for the gentoos) and ASV25 (estimates of 0.29 for chinstrap intercept and -0.25 for gentoos, a p-value of 3.4e-05 for the gentoos). However, there were no significant differences between the amount of ASV48 between the two groups, which had estimates of 0.063 for chinstraps and a 0.0066 higher estimate for gentoos. This could be because of relatively fewer hits for ASV48 than ASV10 and ASV25, but it could also indicate genuine differences between the parasite load of these species for the first two ASVs that is not the case for ASV48.

Table 7: PERMANOVAs examining differences in the beta diversity of the microbiome between gentoos and chinstraps at Aitcho Island. When included in the model, percent of tapeworm is the percent of reads identified as tapeworm out of all metazoan reads identifiable to phyla minus those identified as host. Solid lines divide separate models; dashed lines indicate factors included in the same model. Species was included both as a factor in the first two models and a strata term in the third model. Solid horizontal lines indicate separate models, while a dashed line indicates different factors in the same model. Models were run 100 times on separately subset (without replacement) data to 5,000 reads, and the median, 5%, and 95% percent estimates are reported as well as the median p-value.

	Strata	Median	5%	95%	P-value (median)
Species		0.21	0.20	0.21	0.001*
Species		0.21	0.20	0.21	0.001*
Percent Tapeworm		0.026	0.025	0.026	0.27
Percent Tapeworm	Species	0.17	0.17	0.18	0.27

Table 8: PERMANOVAs comparing the total percent of the of top tapeworm ASV (ASV 10, putative identification as *Tetrabothrius sp.*) (of all metazoan reads identifiable to phyla minus those identified as host) and beta diversity of the microbiome between chinstraps and gentoos at Aitcho Island. Solid lines divide separate models; dashed lines indicate factors included in the same model. Species is included as both a factor in the first model and a strata term in the second model. Models were run 100 times on separately subset (without replacement) data to 5,000 reads, and the median, 5%, and 95% percent estimates are reported as well as the median p-value.

	Strata	Median	5%	95%	P-value (median)
% ASV 10		0.17	0.17	0.18	0.001*
Species		0.05	0.05	0.05	0.010*
% ASV 10	Species	0.17	0.078	0.080	0.49

Table 9: PERMANOVAs comparing the total percent of the of top 2nd tapeworm ASV (ASV 25) (of all metazoan reads identifiable to phyla minus those identified as host) and beta diversity of the microbiome between chinstraps and gentoos at Aitcho Island. Solid lines divide separate models; dashed lines indicate factors included in the same model. Species is included as both a factor in the first model and a strata term in the second model. Models were run 100 times on separately subset (without replacement) data to 5,000 reads, and the median, 5%, and 95% percent estimates are reported as well as the median p-value.

	Strata	Median	5%	95%	P-value (median)
% ASV 25		0.13	0.13	0.08	0.001 *
Species		0.09	0.09	0.09	0.001*
% ASV 25	Species	0.013	0.080	0.082	0.56

Table 10: PERMANOVAs comparing the total percent of the of top 3nd tapeworm ASV (ASV 48, putative Identification as *Parochites zederi*) (of all metazoan reads identifiable to phyla minus those identified as host) and beta diversity of the microbiome between chinstraps and gentoos at Aitcho Island. Solid lines divide separate models; dashed lines indicate factors included in the same model. Species is included as both a factor in the first model and a strata term in the second model. Models were run 100 times on separately subset (without replacement) data to 5,000 reads, and the median, 5%, and 95% percent estimates are reported as well as the median p-value.

	Strata	Median	5%	95%	P-value (median)
% ASV 48		0.03	0.03	0.03	0.20
Species		0.20	0.20	0.21	0.001*
% ASV 48	Species	0.03	0.053	0.055	0.24

Table 11: PERMANOVAs comparing the total percent of the of top 4th tapeworm ASV (ASV 50) (of all metazoan reads identifiable to phyla minus those identified as host) and beta diversity of the microbiome between chinstraps and gentoos at Aitcho Island. Solid lines divide separate models; dashed lines indicate factors included in the same model. Species is included as both a factor in the first model and a strata term in the second model. Models were run 100 times on separately subset (without replacement) data to 5,000 reads, and the median, 5%, and 95% percent estimates are reported as well as the median p-value.

	Strata	Median	5%	95%	P-value (median)
% ASV 50		0.034	0.033	0.034	0.12
Species		0.19	0.19	0.19	0.001*
% ASV 50	Species	0.03	0.03	0.03	0.357

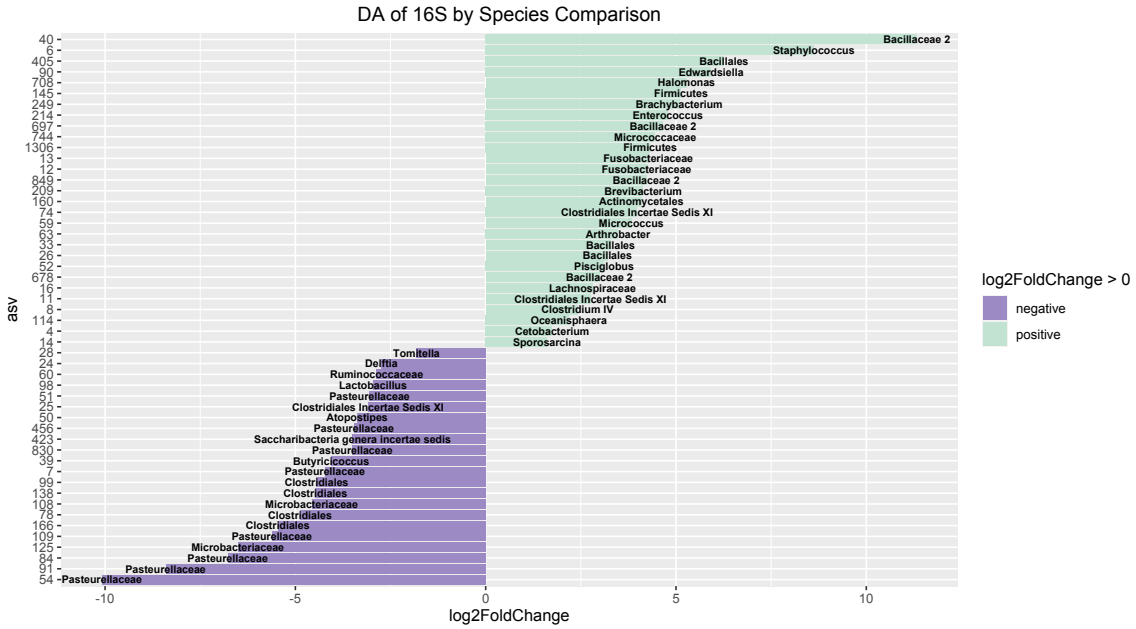


Figure 6: A visualization of the results of the differential abundance test in DESeq2 for 16S microbiome ASVs examining the difference between chinstrap and gentoo penguins at Aitcho Island, showing all microbial ASVs with log2fold changes to their lowest determined identity. Those in purple showed negative log fold changes and those in green showed positive log fold changes. The ASV number for each ASV is on the y-axis and the closest identified taxa is recorded on the bar. The length of the bar shows the magnitude of the log fold change over the model. The chinstrap is the reference species in the graph, so logfold changes are for gentoos relative to chinstraps.

The simple PERMANOVAs comparing the 16S data for the two species and including percent tapeworm as a factor indicate that while there are significant differences between the microbiomes of the two species, these microbiomes do not appear to differ significantly due to percent tapeworm (Tables 7-11). However, there are species differences not correlated to percent tapeworm. This is consistent with the PERMANOVAs looking at specific ASVs and microbiome differences with one exception. ASV48 (potentially *Parochites zederi*) is correlated with different microbiomes, even when considered only within a single species (strata set to species). The species did show significant microbiome differential abundances, with 22 ASVs showing log-fold reductions in gentoos compared to chinstraps, and 29 showing log-fold increases (Figure 6).

Discussion

This study shows there are correlations between microbiome beta diversity and tapeworm DNA that are variable by both colony and the specific ASV of the tapeworm DNA. Patterns of

tapeworm distribution as measured by the presence of tapeworm DNA in gentoo samples show strong differences between colonies, but do not show clear biogeographic patterns. Note especially the different results with permutations limited to within colonies versus permutations throughout the dataset (strata term). This indicates the differences associated with increasing amounts of tapeworm DNA seems to be operating more at a colony than individual level, though there are correlations between the amount of tapeworm DNA and microbiome. This is roughly consistent with the geographic and colony infection patterns we found.

Diet is the main source for tapeworm infection in penguin populations. However, the sparsity of research on these species in Antarctic systems means their lifecycle and intermediate hosts are unknown. For *Parochites zederi*, the primary speculation has been a krill intermediate host, though fish has been suggested (Martin, Ortiz et al. 2016). *Tetrobothrius* is more uncertain. While *Tetrabothrius* species have been reported in pygoscelids (Hoberg 2005, González-Acuña, Moreno et al. 2021), potential intermediate hosts have not been identified beyond that krill, fish, and cephalopods are suitable hosts in the pelagic environment(Hoberg 2005). While our study did not find any correlations between diet and tapeworm infection, it should be noted that this study encompasses only the most recent meal these individuals consumed. It is entirely possible that a study examining longer term diet using methods like isotope analysis can find these connections.

In the species comparison, there are strong differences between the species that are independent of the amount of tapeworm DNA recorded. However, the percent of tapeworm DNA present was different between the species for everything but ASV 48, while this was the only ASV where the amount present was connected to differences in microbiome even when considered only within the species. This ASV has been identified as having 100% identity with *Parochites zederi*, a known parasite in this species, which provides evidence that infections with the species may have

broad impacts on the microbiomes of pygoscelids in general. The differential abundance test also gives some of the largest differences and provides a foundation for future studies examining the interactions between parasites and specific microbes in infected birds.

Given the potential impacts of climate change on Antarctica, especially affecting krill and food webs, understanding the complex interactions between tapeworms, host-associated microbiomes, and penguins is of growing importance. This is consistent with existing evidence of the impact of parasite infection on chicks (Bertellotti, D'Amico et al. 2015). Climate change is likely to disrupt components of the food web, for example krill, which are highly reliant on sea ice.

Given their reliance on sea ice, krill and other components of the food web are likely to be disrupted by climate change, as well as potentially increase stressors (such as thermal stress and more extreme weather systems) (Rogers, Frinault et al. 2020, Gutt, Isla et al. 2021, Lewden, Halna du Fretay et al. 2024). Together, these additional stressors have the potential to interact with fitness costs from parasitic infections, further impacting vulnerable organisms and populations.

Regardless, the detection of tapeworm DNA in gentoo faeces that corresponds to known parasites using common 18S markers opens an important potential tool for further ecological studies in Antarctic systems, work which has mostly been limited to necroscopies and the attendant biases of reliance on such methods. In addition, this study joins others that demonstrate complex links between helminth infection and microbiomes in a wide variety of wild and experimental systems (Zaiss and Harris 2016, Huwe, Prusty et al. 2019).

Acknowledgements

We thank Quark Expeditions Ltd, Lindblad Expeditions, Cheeseman Ecology Safaris Ltd, Pelagic Expeditions Ltd, Golden Fleece Expeditions and Oceanwide Expeditions Ltd for logistical support in the field. We also gratefully acknowledge fieldworkers, including some of those who contributed over multiple seasons, like Caitlin Black, Fiona Jones and others from Penguin Watch, Oceanites and Stony Brook University. We thank all of those associated with the Comparative Infection and Immunology Group and the Polar Ecology and Conservation Research Group, especially Laure Cugniere for support and comments on the study and manuscript. The work conducted by H. Levy was carried out in a non-governmental capacity and does not represent the views of the White House or the Office of Science and Technology Policy.

Funding

Funding for this study is from the Oxford-Oxitec DPhil Scholarship and an Antarctic Science Bursary (CK), by the John Ellerman Foundation, the Save our Seas Foundation and by numerous donations on board Quark Expeditions and Cheesemans' Ecology Surfaris (TH), and University of Oxford Clarendon Fund Scholarship and the U.S. Department of Veteran's Affairs G.I. Bill (HL).

Author Contributions

C.K., A.S., T.H. jointly conceived of the study. T.H., H.L., C.Y., G.C. collected samples. Initial molecular work conducted by H.L. Statistics and coding conducted by C.K., S.P., M.C. Phylogeny constructed by C.K. and A.P. C.K. drafted the paper, which all authors contributed to. A.S. and T.H. supervised all aspects of the study.

Ethics Approval

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/Malvinas 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. Animal handling protocols followed Scientific Committee on Antarctic Research Code of Conduct on the use of Animals for Scientific Purposes in Antarctica. All samples were transported in accordance with applicable export permits and United Kingdom Department for Environment, Food and Rural Affairs import permits.

Works Cited

Ahyong, S., C. B. Boyko, N. Bailly, et al. (2024). World Register of Marine Species (WoRMS), WoRMS Editorial Board.

Ardley, J. K. and P. N. Green Methylobacterium. Bergey's Manual of Systematics of Archaea and Bacteria: 1-23.

Banks, J. C., S. C. Cary and I. D. Hogg (2009). "The phylogeography of Adelie penguin faecal flora." Environ Microbiol **11**(3): 577-588.

Barbosa, A., V. Balagué, F. Valera, A. Martínez, J. Benzal, M. Motas, J. I. Diaz, A. Mira and C. Pedrós-Alió (2016). "Age-Related Differences in the Gastrointestinal Microbiota of Chinstrap Penguins (Pygoscelis antarctica)." PLOS ONE **11**(4): e0153215.

Bertellotti, M., V. L. D'Amico, M. G. Palacios, A. Barbosa and N. Coria (2015). "Effects of antihelminthic treatment on cell-mediated immunity in Gentoo penguin chicks." Polar Biology **39**(7): 1207-1212.

Brooks, M. E., K. Kristensen, K. J. van Benthem, A. Magnusson, C. W. Berg, A. Nielsen, H. J. Skaug, M. Mächler and B. M. Bolker (2017). "glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling." The R Journal **9**(2): 378-400.

Callahan, B. J., P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. Johnson and S. P. Holmes (2016). "DADA2: High-resolution sample inference from Illumina amplicon data." Nat Methods **13**(7): 581-583.

Caporaso, J. G., C. L. Lauber, W. A. Walters, D. Berg-Lyons, C. A. Lozupone, P. J. Turnbaugh, N. Fierer and R. Knight (2011). "Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample." Proc Natl Acad Sci U S A **108 Suppl 1**: 4516-4522.

Cielecka, D., A. Wojciechowska and K. Zdzitowiecki (1992). "Cestodes from penguins on King George Island [South Shetlands, Antarctic]." Acta Parasitologica **2**(37).

Clarke, J. and K. Kerry (2000). "Diseases and Parasites of Penguins." Penguin Conservation **13**(1): 5-24.

Clucas, G. V., M. J. Dunn, G. Dyke, S. D. Emslie, R. Naveen, M. J. Polito, O. G. Pybus, A. D. Rogers and T. Hart (2014). "A reversal of fortunes: climate change 'winners' and 'losers' in Antarctic Peninsula penguins." Sci Rep **4**: 5024.

Dewar, M. L., J. P. Y. Arnould, P. Dann, P. Trathan, R. Groscolas and S. Smith (2013). "Interspecific variations in the gastrointestinal microbiota in penguins." MicrobiologyOpen **2**(1): 195-204.

Diaz, J. I., B. Fusaro, V. Vidal, D. González-Acuña, E. S. Costa, M. Dewar, R. Gray, M. Power, G. Miller, M. Blyton, R. Vanstreels and A. Barbosa (2017). Macroparasites in Antarctic Penguins. Biodiversity and Evolution of Parasitic Life in the Southern Ocean: 183-204.

Foundation, P. S. Python Language Reference, version 2.7. .

Gerritsen, J., S. Fuentes, W. Grievink, L. van Niftrik, B. J. Tindall, H. M. Timmerman, G. T. Rijkers and H. Smidt (2014). "Characterization of Romboutsia ilealis gen. nov., sp. nov., isolated from the gastro-intestinal tract of a rat, and proposal for the reclassification of five closely related members of the genus Clostridium into the genera Romboutsia gen. nov., Intestinibacter gen. nov., Terrisporobacter gen. nov. and Asaccharospora gen. nov." Int J Syst Evol Microbiol **64**(Pt 5): 1600-1616.

González-Acuña, D. A., L. Moreno, M. Wille, B. Herrmann, M. J. Kinsella and R. L. Palma (2021). "Parasites of chinstrap penguins (Pygoscelis antarctica) from three localities in the Antarctic Peninsula and a review of their parasitic fauna." Polar Biology **44**(11): 2099-2105.

Grond, K., B. K. Sandercock, A. Jumpponen and L. H. Zeglin (2018). "The avian gut microbiota: community, physiology and function in wild birds." Journal of Avian Biology **49**(11): e01788.

Gutt, J., E. Isla, J. C. Xavier, B. J. Adams, I. Y. Ahn, C. C. Cheng, C. Colesie, V. J. Cummings, G. di Prisco, H. Griffiths, I. Hawes, I. Hogg, T. McIntyre, K. M. Meiners, D. A. Pearce, L. Peck, D. Piepenburg, R. R. Reisinger, G. K. Saba, I. R. Schloss, C. N. Signori, C. R. Smith, M. Vacchi, C. Verde and D. H. Wall (2021). "Antarctic ecosystems in transition - life between stresses and opportunities." Biol Rev Camb Philos Soc **96**(3): 798-821.

Hoberg, E. P. (2005). Marine Birds and Their Helminth Parasites. Marine Parasitology. K. Rohde. Collingwood, Vic, SIRO Publishing.

Huwe, T., B. K. Prusty, A. Ray, S. Lee, B. Ravindran and E. Michael (2019). "Interactions between Parasitic Infections and the Human Gut Microbiome in Odisha, India." Am J Trop Med Hyg **100**(6): 1486-1489.

Jesudoss Chelladurai, J. R. J. and M. T. Brewer (2021). "Global prevalence of Mesocestoides infections in animals - A systematic review and meta-analysis." Vet Parasitol **298**: 109537.

Kaczvinsky, C., H. Levy, S. Preston, C. Youngflesh, G. Clucas, H. J. Lynch, T. Hart and A. L. Smith (2024). "The influence of biotic and abiotic factors on the bacterial microbiome of gentoo penguins (Pygoscelis papua) in their natural environment." Scientific Reports **14**(1): 17933.

Kleinertz, S., S. Christmann, L. M. Silva, J. Hirzmann, C. Hermosilla and A. Taubert (2014). "Gastrointestinal parasite fauna of Emperor Penguins (Aptenodytes forsteri) at the Atka Bay, Antarctica." Parasitol Res **113**(11): 4133-4139.

Lane, D. (1991). "16S/23S rRNA sequencing." Nucleic acid techniques in bacterial systematics: 115-175.

Laros., J. F. J. (2023). Demultiplex: FASTA/FASTQ demultiplexer (1.2.1).

Lewden, A., T. Halna du Fretay and A. Stier (2024). "Changes in body surface temperature reveal the thermal challenge associated with catastrophic moult in captive gentoo penguins." Journal of Experimental Biology **227**(11).

Love, M. I., W. Huber and S. Anders (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." Genome Biology **15**(12): 550.

Martin, M. A., J. M. Ortiz, J. Seva, V. Vidal, F. Valera, J. Benzal, J. J. Cuervo, C. de la Cruz, J. Belliure, A. M. Martinez, J. I. Diaz, M. Motas, S. Jerez, V. L. D'Amico and A. Barbosa (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**(3): 568-575.

McInnes, J. C., L. Emmerson, C. Southwell, C. Faux and S. N. Jarman (2016). "Simultaneous DNA-based diet analysis of breeding, non-breeding and chick Adélie penguins." Royal Society Open Science **3**(1): 150443.

McMurdie, P. J. and S. Holmes (2013). "phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data." PLOS ONE **8**(4): e61217.

Morgulis, A., G. Coulouris, Y. Raytselis, T. L. Madden, R. Agarwala and A. A. Schäffer (2008). "Database indexing for production MegaBLAST searches." Bioinformatics **24**(16): 1757-1764.

Palacios, M. J., F. Valera and A. Barbosa (2012). "Experimental assessment of the effects of gastrointestinal parasites on offspring quality in chinstrap penguins (Pygoscelis antarctica)." Parasitology **139**(6): 819-824.

Paradis, E. and K. Schliep (2018). "ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R." Bioinformatics **35**(3): 526-528.

Quast, C., E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies and F. O. Glockner (2013). "The SILVA ribosomal RNA gene database project: improved data processing and web-based tools." Nucleic Acids Res **41**(Database issue): D590-596.

R Development Core Team (2021). "R: A Language and Environment for Statistical Computing."

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

Rogers, A. D., B. A. V. Frinault, D. K. A. Barnes, N. L. Bindoff, R. Downie, H. W. Ducklow, A. S. Friedlaender, T. Hart, S. L. Hill, E. E. Hofmann, K. Linse, C. R. McMahon, E. J. Murphy, E. A. Pakhomov, G. Reygondeau, I. J. Staniland, D. A. Wolf-Gladrow and R. M. Wright (2020). "Antarctic Futures: An Assessment of Climate-Driven Changes in Ecosystem Structure, Function, and Service Provisioning in the Southern Ocean." Annual Review of Marine Science **12**(1): 87-120.

Waeschenbach, A. and T. Littlewood (2017). A molecular framework for the Cestoda, The KU Natural History Museum.

Wilson, K., O. N. B. ørnstad, A. P. Dobson, S. Merler, G. Poglayen, S. E. Randolph, A. F. Read and A. Skorping (2002). Heterogeneities in macroparasite infections: patterns and processes. The Ecology of Wildlife Diseases. P. J. Hudson, A. Rizzoli, B. T. Grenfell, H. Heesterbeek and A. P. Dobson, Oxford University Press**:** 0.

Zaiss, M. M. and N. L. Harris (2016). "Interactions between the intestinal microbiome and helminth parasites." Parasite Immunology **38**(1): 5-11.

Zeng, Y.-X., H.-R. Li, W. Han and W. Luo (2021). "Comparison of Gut Microbiota between Gentoo and Adélie Penguins Breeding Sympatrically on Antarctic Ardley Island as Revealed by Fecal DNA Sequencing." Diversity **13**(10).

Chapter 4
Connections Between Gentoo Penguin Diet and Algal Signatures

Connections Between Gentoo Penguin Diet and Algal Signatures

Authors: Chloe Kaczvinsky[1], Hila Levy[2], Stephen Preston[1,3], Casey Youngflesh,[4] Gemma Clucas[5], Heather J. Lynch[6,7], Adrian L. Smith[1]†, Tom Hart[8]†

†co-senior authors and co-corresponding adrian.smith@biology.ox.ac.uk and t.hart@brookes.ac.uk

Affiliations:

1. Department of Biology, University of Oxford, 11a Mansfield Road, Oxford, OX1 3SZ. https://orcid.org/0000-0003-2802-2976
2. Office of Science and Technology Policy, Executive Office of the President, 1650 Pennsylvania Avenue, Washington, DC, 20504
3. UCL School of Pharmacy, University College London, 29-39 Brunswick Square London WC1N 1AX
4. Ecology, Evolution, and Behavior Program, Michigan State University, East Lansing, MI 48824, USA
5. Cornell Lab of Ornithology, 159 Sapsucker Woods Rd., Ithaca, NY 14850, USA
6. Department of Ecology & Evolution, Stony Brook University, Stony Brook, NY 11794
7. Institute for Advanced Computational Sciences, Stony Brook University, Stony Brook, NY 11794
8. Oxford Brookes University, Gypsy Lane, Headington, Oxford. OX3 0BP. https://orcid.org/0000-0002-4527-5046

Abstract

Photosynthetic microorganisms like algae are key productivity drivers in the world's oceans, including the Southern Ocean. In the Scotia Arc, these algae are a key food source for the krill that support much of the larger organisms throughout the region, such as penguins, seals, and whales. Algal sampling at sea is difficult and often requires specialised equipment and large amounts of sea water for studies. Here, we used penguin samples to determine algal presence and diversity in these waters. To do this, we examined over 300 samples of penguin faeces from across the Scotia Arc to better understand the connections between algal diversity and these marine meso-predators. The algal signatures present in these faeces showed connections with the diet of the penguins, the location where the sample was collected, and the broader bacterial microbiome detected in the same samples. Considering the links between geographic distance, diet, and algal signature, the algal signatures in penguin faeces contain preliminary indications of the feeding grounds the penguins are exploiting, whether directly through the algae present in the water or through the stomach contents of the krill that are in turn eaten by the penguins. This study highlights the possibility of indirect, but informative, sampling of oceanic conditions through meso-predator faeces that could demonstrate larger links in the Antarctic food web.

Introduction

Like most marine ecosystems, Antarctic productivity is primarily driven by algae and similar aquatic photosynthetic organisms. These species movements and patterns are dictated by the currents and ecological barriers and gradients in the Southern Ocean. Southern Ocean movements are driven by the interactions between warmer, northern waters and cooler, southern waters and the uninterrupted clockwise winds that move around the continent (Hunt, Drinkwater et al. 2016). The Antarctic Circumpolar Current (ACC) is made up of four fronts that encompass

the Antarctic continent (see Figure 1 for a map of currents from (Hunt, Drinkwater et al. 2016)). These are the sub-Antarctic Front (SAF), the Antarctic Polar Front, hereafter, the Polar Front (PF), the Southern Antarctic Current Front, and the Southern Boundary of the Antarctic Circumpolar Current (Orsi, Whitworth et al. 1995, Hunt, Drinkwater et al. 2016). The Polar Front is where Antarctic waters are subsumed below the more northern waters and is generally a strong barrier between populations and species due to the sudden temperature changes (~1.5 °C over less than 45 km) in the water (Patarnello, Bargelloni et al. 1996, Moore, Abbott et al. 1999, Griffiths 2010, Hunt, Drinkwater et al. 2016). All of these current fronts are functional barriers between different environments in Antarctic waters and play a role in species distributions and patterns (Hunt, Drinkwater et al. 2016).

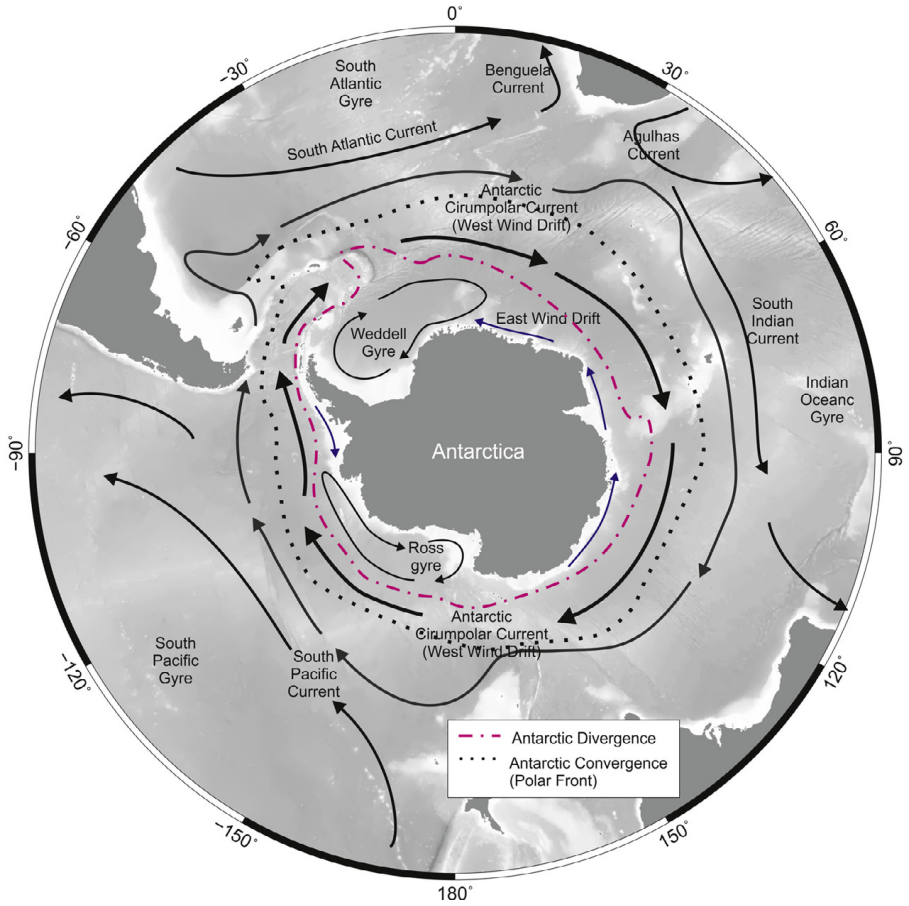


Figure 1: Map from (Hunt, Drinkwater et al. 2016) showing the main currents and jet streams affecting Antarctic waters.

In the Scotia Arc, there are important differences in algal diversity above and below the ecological boundary of the Antarctic Convergence, showing a high degree of endemism (Clayton 1994). Where these species are not exclusively endemic, they are most often overlapping with the sub-Antarctic Islands and southern tip of South America that is the northern-most portion of the Scotia Arc, with Antarctic species below the Polar Front (Smith and Schnack-Schiel 1990, Clayton 1994). However, one of the strongest distinguishers between the various algal microbiomes seems to be the distinction between the cold and warm waters. A survey of dinoflagellates and other phytoplankton showed the sub-Antarctic Front (SAF) to be the primary divider between marine phytoplankton, with polar-associated species assemblages being found south of the front (Olguín Salinas, Brandini et al. 2015). Across latitudinal gradients, beta diversity of algal microbiomes was correlated with the temperature gradients, most especially with temperature breakpoints associated with the polar oceans (Martin, Schmidt et al. 2021).

As is the case for most organisms in Antarctica, algal growth is strongly impacted by seasonality (Smith and Schnack-Schiel 1990). Thick winter ice and snow accumulation can further limit the potential habitat for photosynthetic organisms that must withstand the meagre sunlight of polar winters (Clayton 1994). Generally, ice extent the previous year is tied to phytoplankton biomass and general nutrient availability in the ecosystem (Gallagher, Dinniman et al. 2023). The light limitations of these communities are further compounded by the limitation of some key nutrients in these environments, including iron (Griffiths 2010, Nunes, Latasa et al. 2019). These limitations have strong impacts on species found here, which must be adapted to periods of high productivity and biomass followed by scarcity and freezing temperatures.

The geographic distribution of algae and other micro-photosynthetic organisms are inextricably connected to the keystone of Southern Ocean food-web and their primary consumer, euphasiid krill, especially *Euphasia superba*. Generally, *E. superba* appears to be a single

population, with groups mixing throughout the Southern Ocean, likely driven by the currents (Patarnello, Bargelloni et al. 1996). *E. superba* is sister to the other main Antarctic krill species, *E. crystallorophias* (Patarnello, Bargelloni et al. 1996). Both are closely related to, but distinct from, the sub-Antarctic species *E. vallentini*, reinforcing the importance of the Antarctic Polar Front as a barrier to dispersal (Patarnello, Bargelloni et al. 1996). However, in these currents, the biggest distinction is that between euphasiid larvae and adults, which are generally spatially segregated (Smith and Schnack-Schiel 1990). While immature krill are usually concentrated in shelf areas near the continent, adults are primarily found in deeper waters and where different water bodies converge (Smith and Schnack-Schiel 1990). In the Southern Ocean itself, geographic distance is correlated with moult microbiomes (Clarke, Suter et al. 2021). Clarke et al. combines this study with an experimental system that showed krill microbiomes from the same swarm rapidly diverges when kept in separate tanks, indicating geographic divergence is likely due to drift (Clarke, Suter et al. 2021).

Many studies in the region have focused on broad scale productivity, examining things like satellite images for indicators of biomass accumulation or chlorophyll-a concentration (Ward, Whitehouse et al. 2007). Studies looking at specific species distributions have involved specialized vessels using continuous plankton recorders (CPRs) (Deagle, Clarke et al. 2018) and Niskin bottles with later filtering, either from near a number of stations (Stirnimann, Bornman et al. 2021) or ocean transects (Ward, Whitehouse et al. 2007) to record species assemblages. Some studies did mix these methods, using satellites for initial indications of phytoplankton blooms and then engaging transecting methods (Ward, Whitehouse et al. 2007). However, those studies that examine species assemblages rely on specialized methods and vessels, as well as costly and time-dependent transect paths for the designated study area.

In this study, we will test the use of penguin faeces to trace algal signatures from their feeding grounds. Similar to other seabird species, penguins are often considered a sentinel species in the Antarctic (Boersma 2008). For pygoscelids (brush-tailed penguins) in particular, their ranges are broadly distributed throughout the Antarctic and sub-Antarctic islands, and they are at sea for the vast majority of the year, coming ashore only for moulting and breeding (Lynch 2013, Trathan and Grant 2013, Trivelpiece and Trivelpiece 2013). This broad integration into the marine ecosystem and their reflection of issues like pollution (Metcheva, Yurukova et al. 2011, Souza, Pacyna-Kuchta et al. 2021), krill availability (Fraser and Hofmann 2003), and environmental change (Clucas, Dunn et al. 2014) demonstrate their strong interrelationship with the Southern Ocean ecosystem.

Because penguins come ashore for moulting and breeding, their faeces provide a relatively simple form of biological sampling in Antarctica. Here, we use this penguin faecal sampling as a source of important information on the Southern Ocean food web at large. The potential for this type of data usage exists already in other ecosystems. For example, site-specific algae patterns in faeces from a fish-eating gull species have been found (Yu, Lee et al. 2024), demonstrating the possibility of detecting colony segregation through faecal traces of non-prey species. Given the connections between penguins, krill, and Antarctic algae, algal signatures in penguin faeces have the potential to inform us of connections between these populations and Antarctic ecology more broadly, providing vital information about Antarctic food webs.

Materials and Methods

The data comes from the dataset published in Kaczvinsky et al 2024 (Kaczvinsky, Levy et al. 2024). Samples were collected from the Scotia Arc from 22 gentoo penguin sites (Figure 2). Sample numbers retained in the data set varied from 1 to 30 replicates, with a mean of 11.6 and a

standard deviation of 9.0. Each sample was a discrete faecal event and separated by at least 2 nests to minimize the chances of resampling, was preserved in a 1:1 ratio of RNALater to sample, and frozen after reaching the UK. DNA was extracted with the using MoBio PowerSoil® kit (Qiagen, 51804, now branded as Qiagen QIAamp® PowerFecal® DNA Kit) with a modified protocol for overnight incubation at 65° C on a heated shaker block rather than the shorter bead beating. Samples were then run through PCR using primers 18S rRNA gene_SSU3_F (GGTCTGTGATGCCCT-TAGATG) and 18S rRNA gene_SSU3_R (GGTGTGTACAAAGGG CAGGG) (McInnes, Emmerson et al. 2016) with individual barcodes for later pooling. PCRs were done in triplicate with 1:10 dilution of the sample using Phusion Hot Start Flex DNA polymerase enzyme (New England Biolabs, UK, M0535) and viewed on gels with a selected subset quantified with Qubit 1X dsDNA High-Sensitivity Assay Kit (Invitrogen, UK, Q33231) to standardize amounts for pooling. Libraries were prepared with NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® (New England Biolabs, UK, E7103S) without size selection using NEBNext Multiplex Oligos for Illumina (New England Biolabs, UK, E7335S and E7500S). DNA concentrations were checked on TapeStation (Agilent Genomics) and qPCR (NEBNext® Library Quant Kit for Illumina®, E7630S, New England Biolabs, UK) and loaded onto the MiSeq with 5% phiX at the University of Oxford, Department of Zoology. 150 bp paired-end reads with the 300-cycle MiSeq Reagent Kit v2 (Illumina, UK, MS-102-2002) was used to cover the entire portion of the selected 18S region.

The full 18S data was processed using DADA2 (Callahan, McMurdie et al. 2016) and phyloseq (McMurdie and Holmes 2013) in R (R Development Core Team 2021), with the Silva 18S database used for taxonomic assignment (Morien and Parfrey 2018). The 18S data was subset to algal signatures including all taxa identified as Charophyta, Chlorophyta, Ochrophyta, Dinoflagellata, or Phaeocystis. Since the ultimate aim in this chapter is looking at the variations in

general algal signatures found in penguin faeces (as opposed to changes related to specific species), the data is not further broken down into respective ASVs and the taxonomic distribution. More important in this study is the broad scale community and how similar they are to those found in other samples. Therefore, this analysis focuses on the beta diversity of those groups classed as algae.

PERMANOVA and Mantel tests using the vegan (Oksanen, Simpson et al. 2022) and ape (Paradis and Schliep 2018) packages in R (R Development Core Team 2021) were used to examine the connections between algal signatures, colony, and diet, and algal signatures, microbiome, and geographic distance respectively. The data was also put into a food subset of crustaceans, squid, fish, and salps. To best assess diet, a single value was calculated for each sample indicating the percentage of DNA attributed to crustacean relative to all of the ASVs associated with diet. Given that the majority of penguin diet is generally krill and fish (Ratcliffe and Trathan 2012), this was a functional single explainer term for diet to include in models for this study. While the PCoAs show some level of differentiation by colony and diet, a formal test is a better assessment of the impact of these factors on algal signature differentiation. Therefore, to better quantify the impact of colony and diet on algal signatures, percent crustacean and colony were included as factors in PERMANOVAs. Factors included in the mantel tests were algal signature, geographic distance, and bacterial microbiome. Each variable was limited to samples with 500 reads (for algae) or 5,000 reads (for microbiome) and then rarefied without replacement for 100 iterations before being used in the models. Bray-Curtis values for both algal signature and bacterial microbiome were calculated. Geographic distance for the mantel tests was estimated using the geosphere (Hijmans 2021) package from the site coordinates. The results were then averaged and are reported with mean, 5%, and 95% values for estimates and the mean for significance.

The mantel test indicated a few groupings of colonies that had more similar algal signatures, which were grouped and then mapped. For grouping, the main groupings were completely reciprocal, where all colonies grouped as more similar to each other. A few colonies grouped with some but not all colonies of a grouping, These were distinguised on the mapping and the full color coded table with Bray-Curtis values is available as Supplementary Table 1. The base cut-off for grouping was < 0.5. Because the Bray-Curtis values for colonies were averages of all the samples from the respective colony, the matrix was not perfectly identical between pairs. However, the numbers are broadly close, and the groupings can be examined in Supplementary Table 1.

Results

First, the beta diversity of algal signatures was calculated using Bray-Curtis. Initial PCoAs of rarefied data demonstrate roughly 60% of variation was explained by the first two axes. However, these axes did not clearly map on to any one variable considered. These visualizations show some degree of segregation by colony and latitude and a greater segregation by the percent crustacean in diet signatures (Figure 3). These tests demonstrate that while both colony and percent crustacean appear to be connected with the algal signatures detected in penguin faeces, the relationship is complex and multi-factored.

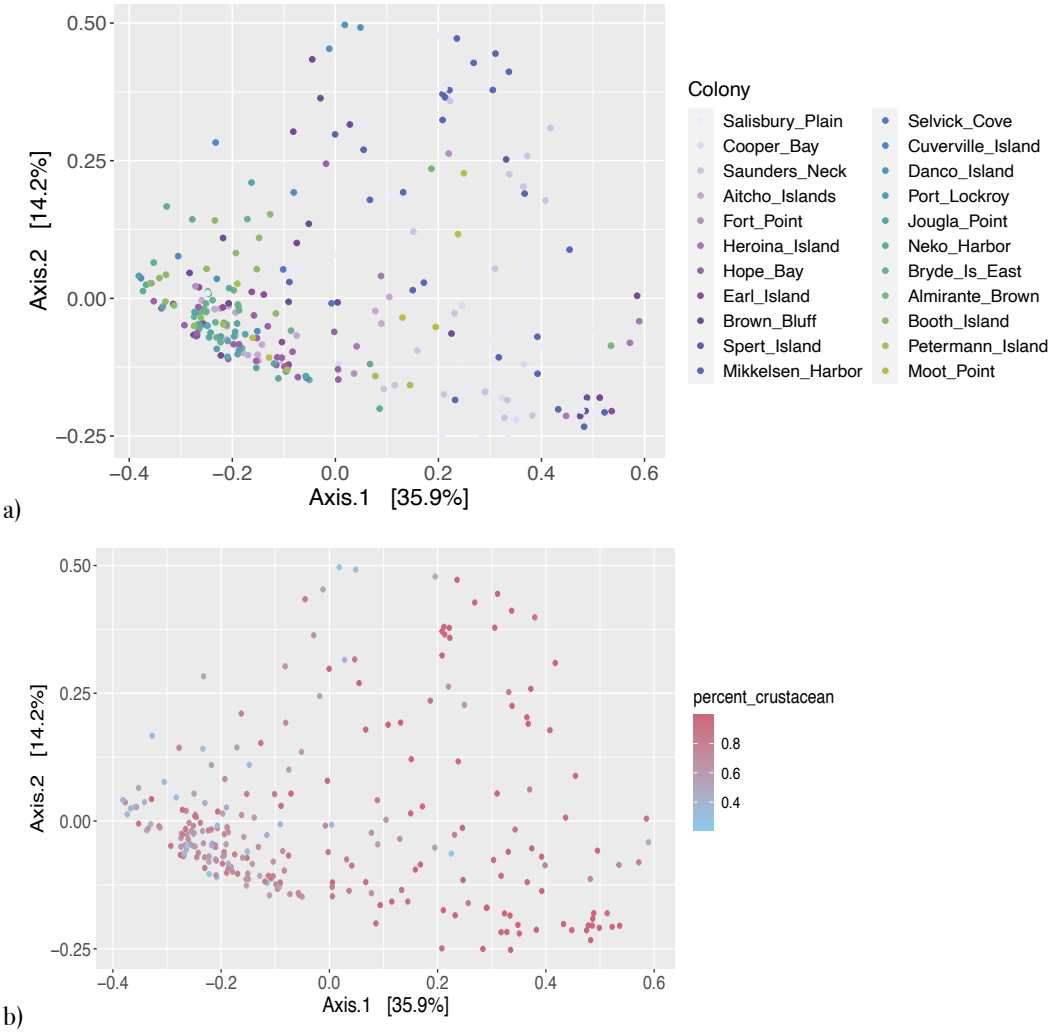


Figure 2: PCoA visualisations of algal signatures, coloured by a) the colony of the sample and b) percent crustacean. Colonies were arranged by latitude in this visualisation to account for the possible influence of latitude on algal signatures.

In these models, colony accounted for roughly 41% of variation in algal signature diversity, and the percent crustacean accounted for roughly 15.5% percent of variation, even when permutations were limited to within the same colony (strata term) (Table 1). Both of these terms were significant (p-value =0.001).

		Strata	Median	5%	95%	P-value (median)
Algal Signature	Colony		0.410	0.407	0.412	0.001
Algal Signature	Percent Crustacean		0.155	0.153	0.157	0.001
Algal Signature	Percent Crustacean	Colony	0.155	0.153	0.157	0.001

Table 1: Percentage of Tapeworm and Microbiome. The algal signature is significantly correlated with colony and percent crustacean.

We then more directly examined the role of geography in algal signatures. To do this, we ran mantel tests examining the correlations between algal signatures and geographic distance, algal signatures and bacterial microbiome, and finally, all three factors in a partial mantel test. The mantel test checks for co-correlation between matrices, such as Bray-Curtis beta diversity tables and matrices of the distance between each pair of colonies. In this test, all three models were significant (Table 2), indicating co-variation between the algal signatures, bacterial microbiome, and geographic distance across the Scotia Arc. When visualised in a by colony average (Figure 3), these comparisons also show some degree of grouping though it does not appear to be a simple set of distinct colonies.

	Algal Signature vs. Geography	Sig.*	Algal Signature vs. Bacterial Microbiome	Sig.*	Partial	Sig.*
median	0.208	1e-04	0.131	2e-04	0.057	0.028
5th percentile	0.206	1e-04	0.127	1e-04	0.054	0.021
95th percentile	0.212	1e-04	0.134	4e-04	0.060	0.035

Table 2: Reporting the Mantel Results Using Bray Curtis Beta Diversity. The algal signature is significantly correlated with geography and the bacterial microbiome. When all three factors are considered in a partial Mantel test, they are also significantly correlated. See Figure 4 for a visualization of this comparison

a)

b)

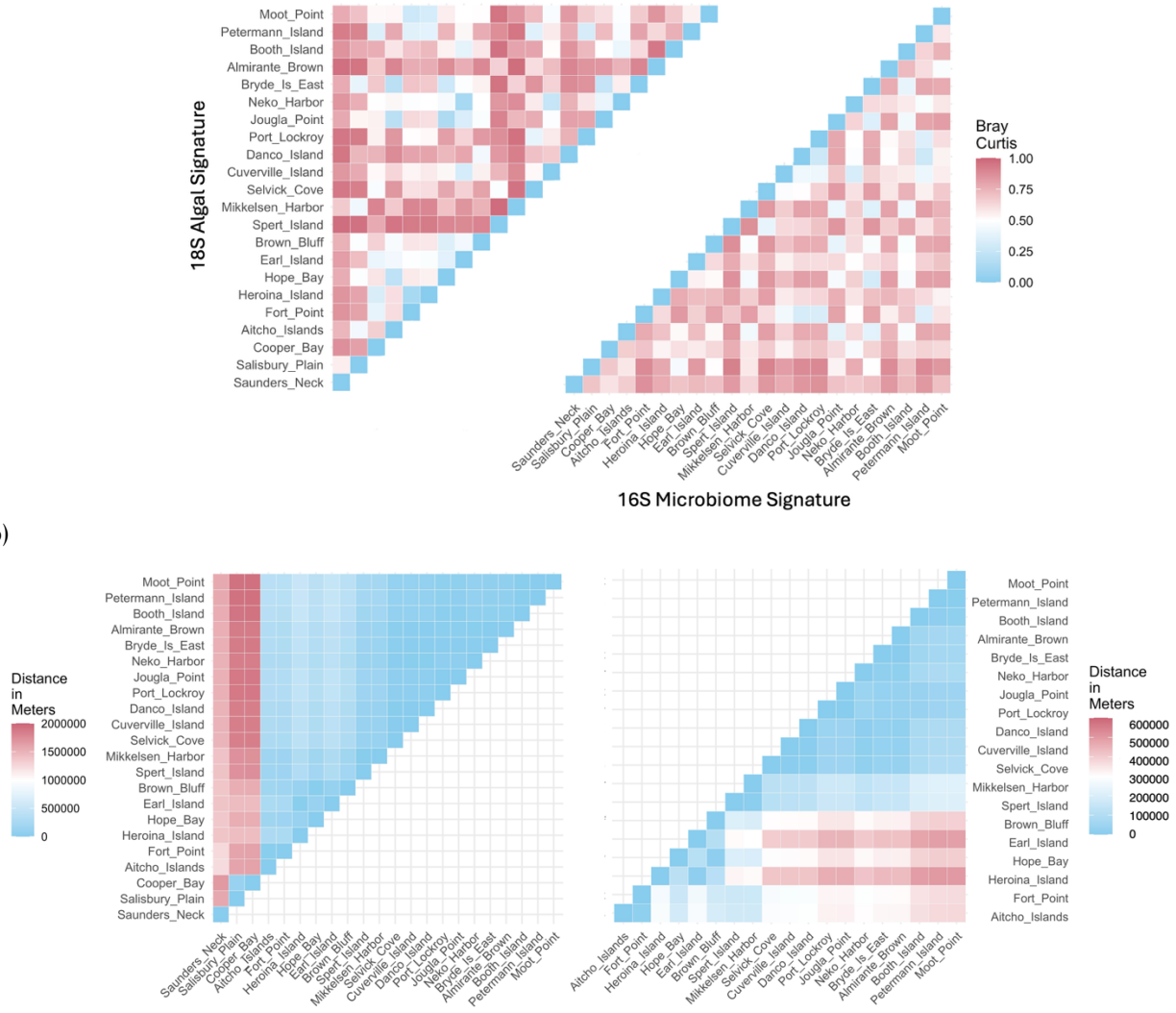


Figure 3: Visualization of Mantel Tests (Table 2) showing the compared matrices for the Bray-Curtis value of the 18S algal signatures, Bray-Curtis Value for the microbiome signatures, and geographic distance. All matrices show the colony average value for the measures involved and are a simplified visualization of the co-correlation tests on the matrices performed by the Mantel test.

a) Matrices for the 18S algal signature on the left and the microbiome signature on the right. Both are set on the same visual scale of Bray Curtis values.

b) Matrices for the distance between colonies. The top right includes all colonies in the comparison, while the bottom right excludes the farthest three colonies (Saunders Neck, Salisbury Plain, and Cooper Bay) for a better visualization of the differences between the remaining colonies. The scales are different for these reflecting the different maximum distances involved.

The Mantel test indicated a few groupings of colonies that had more similar algal signature, which were grouped and then mapped (Figure 4). The groupings came in two types. The main two were fully reciprocal, where each colony was more similar in beta diversity (<0.5) to all of the others in the group. These groups also had a few colonies that were more similar to some but not all the colonies in the group proper, which were distinguished as partial groups. These groups show no obvious latitudinal or longitudinal segreation, as they sometimes occur directly

next to a colony that groups differently. However, the more distant colonies in South Georgia did group as the partial groups (i.e. more similar to some but not all colonies in the group). This could indicate some degree of segreation due to factors other than strict geography that bears further investigating.

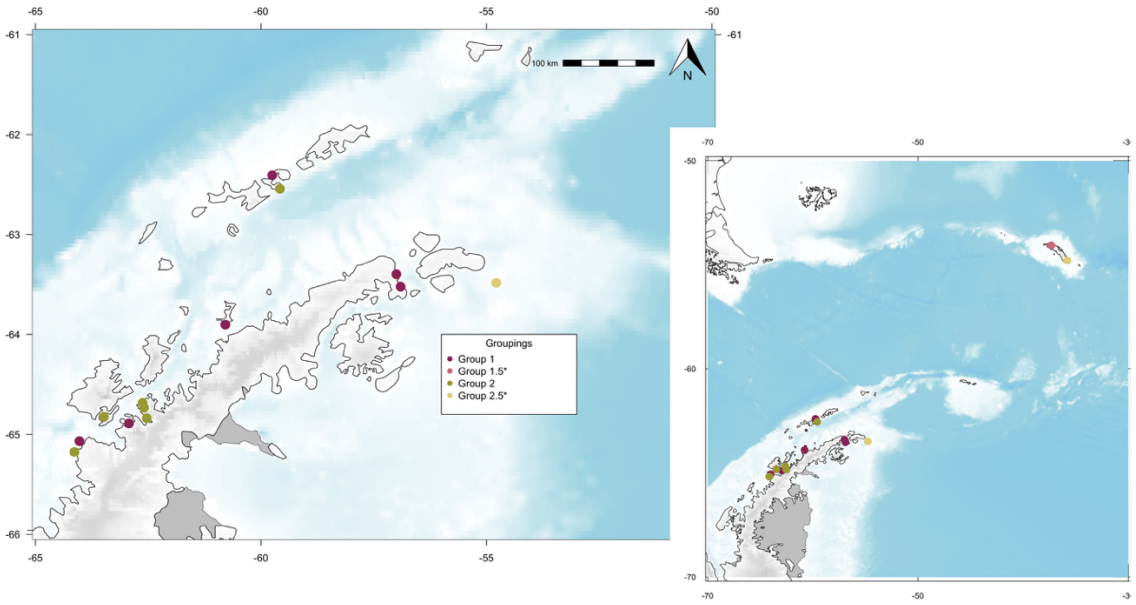


Figure 4: A map showing the putative groupings of algal signatures using averaged Bray-Curtis values. Groups 1 and 2 were fully reciprocal, where each colony was more similar (<0.5) to all of the others in the group. For groups 1.5 and 2.5 the colonies were more similar to some but not all the colonies in the group proper, hence their being distinguished. There were also a few pairs of similar colonies and some that did not group with any others. For a full map including all colonies evaluated see Supplemental Figure 1. For the spreadsheet of Bray Curtis average values, see Supplemental Table 1.

Discussion

Though the connections between algal traces in penguin faeces, diet, and geography are not clear, they are nonetheless informative in terms of this system and the ecology of the Southern Ocean at large. Colony accounted for roughly 40% of algal signature variation, while diet accounted for 15%. Together, this indicates that the location of the penguin (and its probable feeding grounds) and its diet composition have a major influence on the algal signatures detected in its faeces.

While we can't distinguish between those algae and dinoflagellates that were in the guts of prey from those that were in the water that was swallowed by the penguins, it does not matter for our ultimate understanding of connections between algae signatures and diet. Regardless of the way they were introduced to the penguin gut, the algae signatures show strong differences between different gentoo colonies and the diet of the individual.

Further, the results of this study demonstrate that researchers can use penguins as a sentinel species through which they can examine the status of the algae that is the foundation of the Southern Ocean food web. We have validated that not only is this possible in seabirds in general (Yu, Lee et al. 2024), but that it is feasible in the Antarctic ecosystem. Given the difficulty and expense of field work in this environment, penguin faecal samples are a potential source for long-term and continuous time series, especially for research stations near colonies. These samples can complement the more intensive, direct sampling methods like Niskin bottle transects. Future studies pairing these methods could provide vital information about discrepancies or degradation due to digestion or similar factors and provide information for future studies utilizing faecal samples for monitoring.

The large presence of algal genetic material in penguin faeces opens up the possibility of examining broader elements in the Southern Ocean through non-invasive faecal samples. One potential avenue for further exploration is a better integration of algal signatures with current and hunting information for these populations. This could include camera and GPS trackers on individuals from these colonies to gain further understanding of the feeding grounds frequented by penguins at these colonies. Additionally, this data could be integrated with information on currents, oceanic boundaries and krill movements, similar to those seen in (Gallagher, Dinniman et al. 2023), to better understand the interconnections between algae, krill, and higher level predators like penguins.

Regardless, this study demonstrates a strong relationship between penguins, diet, and algal signatures that further highlights the complex relationships within ecosystems. There is no doubt that there is additional work that can be done to clarify how these different factors interact, but being able to monitor a general ecosystem using more accessible faecal samples is an important method for tracking changes in an ecosystem that is highly vulnerable to climate change. Finally, examining the complex interactions between penguins, biogeography, diet, and algal organisms allows for a better understanding of how various biotic and abiotic factors are interacting and impacting microbiomes and penguins in general.

Acknowledgements

We thank Quark Expeditions Ltd, Lindblad Expeditions, Cheeseman Ecology Safaris Ltd, Pelagic Expeditions Ltd, Golden Fleece Expeditions and Oceanwide Expeditions Ltd for logistical support in the field. We also gratefully acknowledge fieldworkers, including some of those who contributed over multiple seasons, like Caitlin Black, Fiona Jones and others from Penguin Watch, Oceanites and Stony Brook University. We thank all of those associated with the Comparative Infection and Immunology Group and the Polar Ecology and Conservation Research Group, especially Ignacio Juarez for support and comments on the study and manuscript. The work conducted by H. Levy was carried out in a non-governmental capacity and does not represent the views of the White House or the Office of Science and Technology Policy.

Author Contributions

C.K., A.S., T.H. jointly conceived of the study. T.H., H.L., C.Y., G.C. collected samples. Initial molecular work conducted by H.L. Statistics and coding conducted by C.K., S.P. C.K. drafted the paper, which all authors contributed to. A.S. and T.H. supervised all aspects of the study.

Funding

Funding for this study is from the Oxford-Oxitec DPhil Scholarship and an Antarctic Science Bursary (CK), by the John Ellerman Foundation, the Save our Seas Foundation and by numerous donations on board Quark Expeditions and Cheesemans' Ecology Surfaris (TH), and University of Oxford Clarendon Fund Scholarship and the U.S. Department of Veteran's Affairs G.I. Bill (HL).

Ethics Approval

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/Malvinas 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. Animal handling protocols followed Scientific Committee on Antarctic Research Code of Conduct on the use of Animals for Scientific Purposes in Antarctica. All samples were transported in accordance with applicable export permits and United Kingdom Department for Environment, Food and Rural Affairs import permits.

Works Cited

Boersma, P. D. (2008). "Penguins as Marine Sentinels." BioScience **58**(7): 597-607.

Callahan, B. J., P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. Johnson and S. P. Holmes (2016). "DADA2: High-resolution sample inference from Illumina amplicon data." Nat Methods **13**(7): 581-583.

Clarke, L. J., L. Suter, R. King, A. Bissett, S. Bestley and B. E. Deagle (2021). "Bacterial epibiont communities of panmictic Antarctic krill are spatially structured." Molecular Ecology **30**(4): 1042-1052.

Clayton, M. N. (1994). "- Evolution of the Antarctic Marine Benthic Algal Flora." Journal of Phycology **30**: 897-904.

Clucas, G. V., M. J. Dunn, G. Dyke, S. D. Emslie, R. Naveen, M. J. Polito, O. G. Pybus, A. D. Rogers and T. Hart (2014). "A reversal of fortunes: climate change 'winners' and 'losers' in Antarctic Peninsula penguins." Sci Rep **4**: 5024.

Deagle, B. E., L. J. Clarke, J. A. Kitchener, A. M. Polanowski and A. T. Davidson (2018). "Genetic monitoring of open ocean biodiversity: An evaluation of DNA metabarcoding for processing continuous plankton recorder samples." Molecular Ecology Resources **18**(3): 391-406.

Fraser, W. R. and E. E. Hofmann (2003). "A predator's perspective on causal links between climate change, physical forcing and ecosystem response." Marine Ecology Progress Series **265**: 1-15.

Gallagher, K. L., M. S. Dinniman and H. J. Lynch (2023). "Quantifying Antarctic krill connectivity across the West Antarctic Peninsula and its role in large-scale Pygoscelis penguin population dynamics." Scientific Reports **13**(1): 12072.

Griffiths, H. J. (2010). "Antarctic Marine Biodiversity – What Do We Know About the Distribution of Life in the Southern Ocean?" PLOS ONE **5**(8): e11683.

Hijmans, R. J. (2021). "geosphere: Spherical Trigonometry."

Hunt, G. L., K. F. Drinkwater, K. Arrigo, J. Berge, K. L. Daly, S. Danielson, M. Daase, H. Hop, E. Isla, N. Karnovsky, K. Laidre, F. J. Mueter, E. J. Murphy, P. E. Renaud, W. O. Smith, P. Trathan, J. Turner and D. Wolf-Gladrow (2016). "Advection in polar and sub-polar environments: Impacts on high latitude marine ecosystems." Progress in Oceanography **149**: 40-81.

Kaczvinsky, C., H. Levy, S. Preston, C. Youngflesh, G. Clucas, H. J. Lynch, T. Hart and A. L. Smith (2024). "The influence of biotic and abiotic factors on the bacterial microbiome of gentoo penguins (Pygoscelis papua) in their natural environment." Scientific Reports **14**(1): 17933.

Lynch, H. J. (2013). Gentoo Penguin. Penguins: Natural History and Conservation. P. G. Borboroglu and P. D. Boersma, University of Washington Press.

Martin, K., K. Schmidt, A. Toseland, C. A. Boulton, K. Barry, B. Beszteri, C. P. D. Brussaard, A. Clum, C. G. Daum, E. Eloe-Fadrosh, A. Fong, B. Foster, B. Foster, M. Ginzburg, M. Huntemann, N. N. Ivanova, N. C. Kyrpides, E. Lindquist, S. Mukherjee, K. Palaniappan, T. B. K. Reddy, M. R. Rizkallah, S. Roux, K. Timmermans, S. G. Tringe, W. H. van de Poll, N. Varghese, K. U. Valentin, T. M. Lenton, I. V. Grigoriev, R. M. Leggett, V. Moulton and T. Mock (2021). "The biogeographic differentiation of algal microbiomes in the upper ocean from pole to pole." Nature Communications **12**(1): 5483.

McInnes, J. C., L. Emmerson, C. Southwell, C. Faux and S. N. Jarman (2016). "Simultaneous DNA-based diet analysis of breeding, non-breeding and chick Adélie penguins." Royal Society Open Science **3**(1): 150443.

McMurdie, P. J. and S. Holmes (2013). "phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data." PLOS ONE **8**(4): e61217.

Metcheva, R., L. Yurukova and S. E. Teodorova (2011). "Biogenic and toxic elements in feathers, eggs, and excreta of Gentoo penguin (Pygoscelis papua ellsworthii) in the Antarctic." Environ Monit Assess **182**(1-4): 571-585.

Moore, J. K., M. R. Abbott and J. G. Richman (1999). "Location and dynamics of the Antarctic Polar Front from satellite sea surface temperature data." Journal of Geophysical Research: Oceans **104**(C2): 3059-3073.

Morien, E. and L. W. Parfrey (2018). SILVA v128 and v132 dada2 formatted 18s 'train sets' (1.0) [Data set]. Zenodo.

Nunes, S., M. Latasa, M. Delgado, M. Emelianov, R. Simó and M. Estrada (2019). "Phytoplankton community structure in contrasting ecosystems of the Southern Ocean: South Georgia, South Orkneys and western Antarctic Peninsula." Deep Sea Research Part I: Oceanographic Research Papers **151**: 103059.

Oksanen, J., G. L. Simpson, F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O'Hara, P. Solymos, M. H. H. Stevens, E. Szoecs, H. Wagner, M. Barbour, M. Bedward, B. Bolker, D. Borcard, G. Carvalho, M. Chirico, M. De Caceres, S. Durand, H. B. A. Evangelista, R. FitzJohn, M. Friendly, B. Furneaux, G. Hannigan, M. O. Hill, L. Lahti, D. McGlinn, M.-H. Ouellette, E. Ribeiro Cunha, T. Smith, A. Stier, C. J. F. Ter Braak and J. Weedon (2022). "vegan: Community Ecology Package."

Olguín Salinas, H. F., F. Brandini and D. Boltovskoy (2015). "Latitudinal patterns and interannual variations of spring phytoplankton in relation to hydrographic conditions of the southwestern Atlantic Ocean (34°–62° S)." Helgoland Marine Research **69**(2): 177-192.

Orsi, A. H., T. Whitworth and W. D. Nowlin (1995). "On the meridional extent and fronts of the Antarctic Circumpolar Current." Deep Sea Research Part I: Oceanographic Research Papers **42**(5): 641-673.

Paradis, E. and K. Schliep (2018). "ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R." Bioinformatics **35**(3): 526-528.

Patarnello, T., L. Bargelloni, V. Varotto and B. Battaglia (1996). "Krill evolution and the Antarctic ocean currents: evidence of vicariant speciation as inferred by molecular data." Marine Biology **126**(4): 603-608.

R Development Core Team (2021). "R: A Language and Environment for Statistical Computing."

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

Smith, S. L. and S. B. Schnack-Schiel (1990). Polar Zooplankton. San Diego ; London :, Academic Press.

Souza, J. S., A. D. Pacyna-Kuchta, L. S. Teixeira da Cunha, E. S. Costa, P. Niedzielski and J. P. Machado Torres (2021). "Interspecific and intraspecific variation in organochlorine pesticides and polychlorinated biphenyls using non-destructive samples from Pygoscelis penguins." Environ Pollut **275**: 116590.

Stirnimann, L., T. G. Bornman, H. M. Verheye, M.-L. Bachèlery, J. van der Poel and S. E. Fawcett (2021). "Plankton community composition and productivity near the Subantarctic Prince

Edward Islands archipelago in autumn." Limnology and Oceanography **66**(12): 4140-4158.

Trathan, P. N. and B. Grant (2013). Adélie Penguin. Penguins: Natural History and Conservation. P. G. Borboroglu and P. D. Boersma, University of Washington Press: 37-57.

Trivelpiece, W. and S. Trivelpiece (2013). Chinstrap Penguin. Penguins: Natural History and Conservation. P. G. Borboroglu and P. D. Boersma, University of Washington Press.

Ward, P., M. Whitehouse, R. Shreeve, S. Thorpe, A. Atkinson, R. Korb, D. Pond and E. Young (2007). "Plankton community structure south and west of South Georgia (Southern Ocean): Links with production and physical forcing." Deep Sea Research Part I: Oceanographic Research Papers **54**(11): 1871-1889.

Yu, T.-S., S. Y. Lee, J. Lee, T. Y. Suman and I.-S. Kwak (2024). "Insights into trophic interactions: DNA metabarcoding reveals species composition in black-tailed gull fecal samples in coastal environments of South Korea." Environmental Monitoring and Assessment **196**(8): 761.

Chapter 5
Discussion

Discussion

A broad-scale study examining multiple factors concerned with pygoscelids in the Scotia Arc and their connections with the greater ecology of the region provides a deeper understanding of the factors impacting species and the region as a whole. The connections between colony, parasites, algae, and microbiome are complex, but better illuminated through the presented biogeographic landscape scale studies of pygoscelids in the Scotia Arc. It is clear that the breeding colony of gentoos is one of the major factors associated with differences in microbiome diversity across the Scotia Arc. That diet and parasite load are also factors in microbiome is apparent, though the interaction between algal species in the water and penguin microbiomes, while present, is less clear-cut.

Colony is crucial to microbiomes and accounts for over 20% of variation in the sites studied (Chapter 2). However, it is uncertain what degree is due to relatedness. In general, gentoos are philopatric and tend to stay in their natal colonies, leading to greater sub-structuring than in other pygoscelid species (Levy, Clucas et al. 2016, Clucas, Younger et al. 2018, Tyler, Bonfitto et al. 2020). To what degree this is the driver of microbiome differentiation between colonies is unclear, as gentoos also demonstrate more individual dietary habits that likely play into this effect (Polito, Trivelpiece et al. 2015). While we included diet in these studies and examined the smaller but significant correlation it had with microbiome and geography, the direction of causality and the difference between penguin microbiome and the microbiome of prey are not sufficiently disentangled in this study. Further work incorporating the genetic relatedness of samples and microbiome and prey of krill to broader penguin microbiome studies is a crucial next step in disentangling the relationship of diet, geography, and microbiome in this system.

It is clear, however, that penguin microbiomes vary with distance and diet, though the effect is much less than that for colony. While there is a broad geographic pattern, there were no

clear break points or latitudinal relationship within this data set (Chapter 2). However, the scarcity of data above the Polar Front does leave the possibility that gentoos above and below this point show microbiome differences similar with the previously detected immunological differences (Levy, Fiddaman et al. 2020). The incorporation of a greater number of samples from above the polar front would be a logical next step in evaluating the geography of penguin microbiomes in the Scotia Arc.

This thesis also examined the role of helminths, specifically tapeworms, in the microbiomes of penguins and the general use of less invasive faecal sampling to detect them. Considering the multiple interactions documented between helminths and microbiomes in a variety of systems (Granroth-Wilding, Burthe et al. 2014, Granroth-Wilding, Burthe et al. 2015, Zaiss and Harris 2016, Huwe, Prusty et al. 2019), it is clear that understanding the role of helminths in Antarctica is crucial to understanding the ecology of this region. Within this study, the amount of tapeworm DNA present in sequenced samples was also connected with different microbiomes, and this connection mostly operates on a colony-based level in gentoos (Chapter 3). However, this study relied on Illumina sequencing rather than qPCR, introducing potential biases in amplification. A qPCR or similar method that could more directly assess the percentage of tapeworm DNA would be an important next step in confirming the utility of molecular detection methods for helminths. Microscopy based confirmation of infections would also help validate the methods for future work. Despite these limitations, however, Chapter 3 demonstrates the interactions between parasitic infections, microbiomes, and geography in penguins, providing important information of factors that influence health and microbiomes of these species.

Algal signatures can be detected in penguin faeces and are strongly correlated both with colony and diet, as well as showing connections with the penguin bacterial microbiome and the geographic distance between where the samples were collected. Much like penguin bacterial

microbiome (Chapter 2), algal signatures showed correlations with geographic distance, but distance was not the sole factor affecting them (Chapter 4). Interestingly, the connection between colony and algal signature (~40%, Chapter 4) was even stronger than that for bacterial microbiome and colony (~20%, Chapter 2). It could be that algal signature became more concentrated as it moved up the food chain from water to krill to penguin. Potential next steps for research in this area could focus on disentangling factors that were examined together in this study. These further explorations could include things like GPS or similar trackers to determine the feeding grounds of penguins paired with faecal studies. The incorporation of water currents and krill sources and population structure could also clarify the interactions between diet and microbiome by indicating which portions of algal signatures and microbiome are from the locations of penguin feeding and which are carried over from prey.

Works Cited

Clucas, G. V., J. L. Younger, D. Kao, L. Emmerson, C. Southwell, B. Wienecke, A. D. Rogers, C.-A. Bost, G. D. Miller, M. J. Polito, P. Lelliott, J. Handley, S. Crofts, R. A. Phillips, M. J. Dunn, K. J. Miller and T. Hart (2018). "Comparative Population Genomics Reveals Key Barriers to Dispersal in Southern Ocean Penguins." Molecular Ecology **27**: 4680-4697.

Granroth-Wilding, H. M., S. J. Burthe, S. Lewis, K. A. Herborn, E. A. Takahashi, F. Daunt and E. J. Cunningham (2015). "Indirect effects of parasitism: costs of infection to other individuals can be greater than direct costs borne by the host." Proc Biol Sci **282**(1811).

Granroth-Wilding, H. M., S. J. Burthe, S. Lewis, T. E. Reed, K. A. Herborn, M. A. Newell, E. A. Takahashi, F. Daunt and E. J. Cunningham (2014). "Parasitism in early life: environmental conditions shape within-brood variation in responses to infection." Ecol Evol **4**(17): 3408-3419.

Huwe, T., B. K. Prusty, A. Ray, S. Lee, B. Ravindran and E. Michael (2019). "Interactions between Parasitic Infections and the Human Gut Microbiome in Odisha, India." Am J Trop Med Hyg **100**(6): 1486-1489.

Levy, H., G. V. Clucas, A. D. Rogers, A. D. Leache, K. L. Ciborowski, M. J. Polito, H. J. Lynch, M. J. Dunn and T. Hart (2016). "Population structure and phylogeography of the Gentoo Penguin (Pygoscelis papua) across the Scotia Arc." Ecol Evol **6**(6): 1834-1853.

Levy, H., S. R. Fiddaman, J. A. Vianna, D. Noll, G. V. Clucas, J. K. H. Sidhu, M. J. Polito, C. A. Bost, R. A. Phillips, S. Crofts, G. D. Miller, P. Pistorius, F. Bonnadonna, C. Le Bohec, A. Barbosa, P. Trathan, A. Raya Rey, L. A. F. Frantz, T. Hart and A. L. Smith (2020). "Evidence of Pathogen-Induced Immunogenetic Selection across the Large Geographic Range of a Wild Seabird." Mol Biol Evol **37**(6): 1708-1726.

Polito, M. J., W. Z. Trivelpiece, W. P. Patterson, N. J. Karnovsky, C. S. Reiss and S. D. Emslie (2015). "Contrasting specialist and generalist patterns facilitate foraging niche partitioning in sympatric populations of Pygoscelis penguins." Marine Ecology Progress Series **519**: 221-237.

Tyler, J., M. T. Bonfitto, G. V. Clucas, S. Reddy and J. L. Younger (2020). "Morphometric and genetic evidence for four species of gentoo penguin." Ecol Evol **10**(24): 13836-13846.

Zaiss, M. M. and N. L. Harris (2016). "Interactions between the intestinal microbiome and helminth parasites." Parasite Immunology **38**(1): 5-11.

Chapter 6
Conclusion

Conclusion

Together, these chapters demonstrate the multi-factored and complex interactions in Scotia Arc ecology that can be explored through relatively low invasive faecal sampling of a major marine meso-predator. We have also shown how many complex interactions are present even in the relatively isolated Antarctic ecosystem. Considering the relatively higher population structuring of gentoo penguins relative to the other pygoscelids, testing the degree to which the conclusions presented here expand to closely related species with more specialized and panmictic ecologies would be a logical next step for the applicability of our conclusions in other species and systems (Levy, Clucas et al. 2016, Vianna, Noll et al. 2017). Further, when considered in the context of a changing climate, gentoo penguins can serve as a crucial sentinel in the Antarctic. The methods demonstrated can be used for large-scale and relatively low-invasive sampling throughout time to monitor shifts in the phytoplankton and other organisms that make up the base of the Southern Ocean food web as well as changes in parasite and penguin populations. Given the different responses of these closely related species to a changing climate (Clucas, Dunn et al. 2014), long-term faecal sampling will also provide vital data on the differing responses that can be expected from climate change, even when the species are both reliant on the phytoplankton and algal basis of the food web.

Works Cited

Clucas, G. V., M. J. Dunn, G. Dyke, S. D. Emslie, R. Naveen, M. J. Polito, O. G. Pybus, A. D. Rogers and T. Hart (2014). ″A reversal of fortunes: climate change ′winners′ and ′losers′ in Antarctic Peninsula penguins.″ Sci Rep **4**: 5024.

Levy, H., G. V. Clucas, A. D. Rogers, A. D. Leache, K. L. Ciborowski, M. J. Polito, H. J. Lynch, M. J. Dunn and T. Hart (2016). ″Population structure and phylogeography of the Gentoo Penguin (Pygoscelis papua) across the Scotia Arc.″ Ecol Evol **6**(6): 1834-1853.

Vianna, J. A., D. Noll, G. P. M. Dantas, M. V. Petry, A. Barbosa, D. Gonzalez-Acuna, C. Le Bohec, F. Bonadonna and E. Poulin (2017). ″Marked phylogeographic structure of Gentoo penguin reveals an ongoing diversification process along the Southern Ocean.″ Mol Phylogenet Evol **107**: 486-498.

Appendix A
Supplemental Material for Chapter 2

Supplemental Table 1. All samples used in this study.

Number of Samples	Species	Colony	Site ID[a]	Latitude	Longitude	Region[b]	Season
20	Chinstrap	Aitcho_Islands	AITC	-62.407	-59.752	SSI	2018
20	Gentoo	Aitcho_Islands	AITC	-62.407	-59.752	SSI	2018
5	Gentoo	Almirante_Brown	ALMI	-64.896	-62.870	WAP	2018
20	Gentoo	Booth_Island	BOOT	-65.067	-64.026	WAP	2018
10	Gentoo	Brown_Bluff	BROW	-63.522	-56.905	NAP	2015
20	Gentoo	Brown_Bluff	BROW	-63.522	-56.905	NAP	2018
24	Gentoo	Bryde_Is_East	BRYE	-64.890	-62.927	WAP	2018
10	Gentoo	Cooper_Bay	CBNO	-54.79	-35.81	SG	2015
9	Gentoo	Cuverville_Island	CUVE	-64.684	-62.626	WAP	2018
3	Gentoo	Damoy_Point	DAMO	-64.8149	-63.4968	WAP	2018
6	Gentoo	Danco_Island	DANC	-64.734	-62.594	WAP	2018
10	Gentoo	Earl_Island	EARL	-63.483	-54.783	NAP	2016
6	Gentoo	Fort_Point	FORT	-62.542	-59.580	SSI	2018
10	Gentoo	Heroina_Island	HERO	-63.394	-54.608	NAP	2016
6	Adelie	Hope_Bay	HOPE	-63.397	-56.998	NAP	2018
26	Gentoo	Hope_Bay	HOPE	-63.397	-56.998	NAP	2018
5	Blue_Eyed_Shag	Jougla_Point	JOUG	-64.827	-63.493	WAP	2018
23	Gentoo	Jougla_Point	JOUG	-64.827	-63.493	WAP	2018
3	Gentoo	Port_Lockroy	LOCK	-64.823	-63.484	WAP	2018
10	Gentoo	Mikkelsen_Harbor	MIKK	-63.902	-60.791	WAP	2015
20	Gentoo	Mikkelsen_Harbor	MIKK	-63.902	-60.791	WAP	2018
6	Gentoo	Moot_Point	MOOT	-65.206	-64.078	WAP	2018
10	Gentoo	Neko_Harbor	NEKO	-64.838	-62.533	WAP	2018
6	Blue_Eyed_Shag	Paulet_Island	PAUL	-63.580	-55.788	NAP	2018
5	Gentoo	Petermann_Island	PETE	-65.176	-64.137	WAP	2018
13	Gentoo	Salisbury_Plain	SALI	-54.063	-37.329	SG	2015
15	Gentoo	Salisbury_Plain	SALI	-54.063	-37.329	SG	2017
21	Gentoo	Salisbury_Plain	SALI	-54.063	-37.329	SG	2018
2	Gentoo	Selvick_Cove	SELV	-64.641	-62.569	WAP	2018
10	Gentoo	Saunders_Neck	SNEC	-60.241	-51.307	FI	2015
19	Gentoo	Saunders_Neck	SNEC	-60.241	-51.307	FI	2018
1	Gentoo	Spert_Island	SPER	-63.861	-60.978	WAP	2018
7	Blue_Eyed_Shag	Tetrad_Islands	TETR	-63.918	-60.734	WAP	2018

[a] Site ID is the standardized four-letter code.

[b] Region indicates the general area of the site as follows: SSI (South Sandwich Islands), FI (Falkland Islands/Malvinas), SG (South Georgia), NAP (Northern Antarctic Peninsula/Weddell Sea), WAP (Western Antarctic Peninsula).

Supplemental Table 2. Metrics for Triplicate Analysis.

16S Samples

Sample[a]	Mean	Std Dev.
APHOPE18s085R	0.48138268	0.02110283
APHOPE18s086R	0.41574457	0.09970095
APHOPE18s087R	0.50155675	0.29167085
APHOPE18s088R	0.86674656	0.07085066
APHOPE18s089R	0.91060658	0.0304523
APHOPE18s095R	0.74207708	0.25566228
BESHJOUG18s575R	0.8564419	0.02729098
BESHJOUG18s576R	0.82089736	0.12949757
BESHJOUG18s577R	0.74347315	0.01930452
BESHJOUG18s578R	0.78901691	0.05418444
BESHJOUG18s579R	0.79940504	0.05753791
BESHPAUL18s096R	0.77765772	0.12383316
BESHPAUL18s097R	0.90354861	0.07295099
BESHPAUL18s098R	0.93449799	0.00071252
BESHPAUL18s099R	0.93835847	0.01288352
BESHPAUL18s100R	0.93077747	0.01569247
BESHPAUL18s101R	0.89487727	0.01254377
CPAITC18s428R	0.36563712	0.15512698
CPAITC18s429R	0.33136582	0.38525393
CPAITC18s430R	0.61156879	0.18057657
CPAITC18s431R	0.59787759	0.24251259
CPAITC18s432R	0.90743522	0.03912449
CPAITC18s433R	0.88351154	0.11063915
CPAITC18s434R	0.65661268	0.15167344
CPAITC18s435R	0.70645936	0.02308536
CPAITC18s436R	0.69275272	0.06044632
CPAITC18s437R	0.48559308	0.34424917
CPAITC18s438R	0.3469437	0.52391504
CPAITC18s439R	0.9052929	0.09019334
CPAITC18s440R	0.77755423	0.10724207
CPAITC18s441R	0.64201467	0.02899399
CPAITC18s442R	0.74543411	0.16710329
CPAITC18s443R	0.93274454	0.02195853
CPAITC18s444R	0.84030367	0.1281665
CPAITC18s445R	0.56464228	0.10012819
CPAITC18s446R	0.59132824	0.30698013
CPAITC18s447R	0.91105618	0.07619754

GPAITC18s448R	0.79569967	0.07812322
GPAITC18s449R	0.86397335	0.06452285
GPAITC18s450R	0.75714308	0.14104569
GPAITC18s451R	0.54341245	0.03017972
GPAITC18s452R	0.90120729	0.02861124
GPAITC18s453R	0.58922483	0.06274595
GPAITC18s454R	0.79208997	0.01192095
GPAITC18s455R	0.69439785	0.25086172
GPAITC18s456R	0.93001553	0.02917117
GPAITC18s457R	0.77382604	0.12829834
GPAITC18s458R	0.80019834	0.05017106
GPAITC18s459R	0.69246265	0.05441826
GPAITC18s460R	0.88391643	0.02305881
GPAITC18s461R	0.56747123	0.14343539
GPAITC18s462R	0.84240014	0.07783859
GPAITC18s463R	0.74052346	0.04945331
GPAITC18s464R	0.76411415	0.12333366
GPAITC18s465R	0.73533323	0.06113158
GPAITC18s466R	0.41643913	0.12705056
GPAITC18s467R	0.53520361	0.01492713
GPALMI18s046R	0.86040641	0.04110618
GPALMI18s047R	0.83368846	0.00199373
GPALMI18s052R	0.73839644	0.02302204
GPALMI18s053R	0.85291191	0.02810455
GPALMI18s054R	0.88433727	0.0037888
GPBR15f001	0.78680027	0.02018631
GPBR15f002	0.5003052	0.01234854
GPBR15f003	0.88786508	0.05467389
GPBR15f004	0.39016867	0.03469783
GPBR15f005	0.74986746	0.04837295
GPBR15f006	0.66792225	0.01354724
GPBR15f007	0.84698807	0.03044368
GPBR15f008	0.76368288	0.0073357
GPBR15f009	0.63797017	0.01770153
GPBR15f010	0.86122382	0.02210729
GPBROW18s508R	0.06788269	0.00723815
GPBROW18s509R	0.70623544	0.03428846
GPBROW18s510R	0.47615781	0.03133459
GPBROW18s511R	0.72205824	0.01995452
GPBROW18s512R	0.23854026	0.05165886
GPBROW18s513R	0.35466275	0.16753019
GPBROW18s514R	0.60082296	0.02953072

GPBROW18s515R	0.46138963	0.02748902
GPBROW18s516R	0.16097068	0.03243716
GPBROW18s517R	0.74610938	0.00801026
GPBROW18s518R	0.7167194	0.0050425
GPBROW18s519R	0.2810319	0.01252235
GPBROW18s520R	0.35423751	0.03397911
GPBROW18s521R	0.41620607	0.02062777
GPBROW18s522R	0.03452494	0.01389198
GPBROW18s523R	0.08976463	0.01971711
GPBROW18s524R	0.63618367	0.02556051
GPBROW18s525R	0.31665619	0.00921864
GPBROW18s526R	0.63774323	0.02155899
GPBROW18s527R	0.61237262	0.01627355
GPBRYE18s048R	0.90498089	0.0046183
GPBRYE18s049R	0.80731219	0.01525982
GPBRYE18s055R	0.78200828	0.02067487
GPBRYE18s056R	0.90861358	0.0101553
GPBRYE18s057R	0.9250326	0.00108761
GPBRYE18s389R	0.65857201	0.04800479
GPBRYE18s390R	0.48260134	0.02488924
GPBRYE18s391R	0.44402066	0.14832865
GPBRYE18s392R	0.36562377	0.05076476
GPBRYE18s393R	0.51517806	0.00515954
GPBRYE18s394R	0.55990816	0.00713159
GPBRYE18s395R	0.744151	0.09650346
GPBRYE18s396R	0.50054748	0.03274357
GPBRYE18s397R	0.71735067	0.06019582
GPBRYE18s398R	0.77374967	0.00959904
GPBRYE18s399R	0.65924495	0.03180428
GPBRYE18s400R	0.94127859	0.00230368
GPBRYE18s401R	0.72302201	0.01471076
GPBRYE18s402R	0.88130324	0.0090079
GPBRYE18s403R	0.58348972	0.0239039
GPBRYE18s404R	0.6632736	0.01823278
GPBRYE18s405R	0.67127031	0.19969673
GPBRYE18s406R	0.65187454	0.0962776
GPBRYE18s407R	0.67590688	0.03435108
GPCOOP15f001	0.62912082	0.05436841
GPCOOP15f002	0.8097549	0.01408295
GPCOOP15f003	0.80393305	0.01843821
GPCOOP15f004	0.72185096	0.08098824
GPCOOP15f005	0.71089513	0.06916416

GPCOOP15f006	0.60994064	0.07829512
GPCOOP15f007	0.83143998	0.01932449
GPCOOP15f008	0.45851424	0.0522181
GPCOOP15f009	0.67764473	0.00549372
GPCOOP15f010	0.57198886	0.07766614
GPCUVE18s037R	0.85979172	0.01671723
GPCUVE18s038R	0.9210019	0.00515167
GPCUVE18s039R	0.69375593	0.04410737
GPCUVE18s040R	0.6626311	0.02617245
GPCUVE18s041R	0.92640676	0.01255893
GPCUVE18s042R	0.89733586	0.00376708
GPCUVE18s043R	0.70739208	0.0196903
GPCUVE18s044R	0.47617539	0.04870432
GPCUVE18s045R	0.82060868	0.00720328
GPDAMO18s058R	0.82930576	0.01063361
GPDAMO18s059R	0.86426388	0.09544888
GPDAMO18s060R	0.73525826	0.07344669
GPDANC18s001R	0.92048227	0.00195845
GPDANC18s002R	0.93919016	0.00519947
GPDANC18s003R	0.90320458	0.00352829
GPDANC18s004R	0.74575599	0.0252434
GPDANC18s005R	0.83936947	0.01387301
GPDANC18s006R	0.62064996	0.03132324
GPEARL16f001R	0.80318354	0.02020422
GPEARL16f002R	0.67923174	0.01238194
GPEARL16f003R	0.82922507	0.01245429
GPEARL16f004R	0.77252607	0.00634271
GPEARL16f005R	0.45393855	0.09966901
GPEARL16f006R	0.88483069	0.0052538
GPEARL16f007R	0.67457241	0.04968445
GPEARL16f008R	0.59269611	0.01090935
GPEARL16f009R	0.43271106	0.04543431
GPEARL16f010R	0.45174192	0.11776156
GPFORT18s028R	0.50336373	0.11650023
GPFORT18s029R	0.83516346	0.01356536
GPFORT18s030R	0.94487439	0.00540113
GPFORT18s034R	0.62369691	0.03251615
GPFORT18s035R	0.95007319	0.00490597
GPFORT18s036R	0.90619523	0.01293106
GPHER16f001R	0.90738218	0.00502442
GPHER16f002R	0.91861933	0.00455387
GPHER16f003R	0.93091006	NA

GPHER16f004R	0.77213252	0.05219262
GPHER16f005R	0.46715479	0.01963267
GPHER16f006R	0.5246841	0.0172406
GPHER16f007R	0.13841382	0.03286673
GPHER16f008R	0.90469935	0.01034978
GPHER16f009R	0.76839309	0.04976959
GPHER16f010R	0.44754118	0.02777886
GPHOPE18s066R	0.8963717	0.02662478
GPHOPE18s090R	0.89299454	0.01464956
GPHOPE18s091R	0.79256424	0.01524383
GPHOPE18s092R	0.60525578	0.08394587
GPHOPE18s093R	0.22160736	0.05353173
GPHOPE18s094R	0.39948159	0.14147966
GPHOPE18s468R	0.57727874	0.01238172
GPHOPE18s469R	0.45878615	0.01763002
GPHOPE18s470R	0.71998078	0.03068569
GPHOPE18s471R	0.8136244	0.00137973
GPHOPE18s472R	0.74989257	0.00781382
GPHOPE18s473R	0.82272532	0.04529459
GPHOPE18s474R	0.86411168	0.00699854
GPHOPE18s475R	0.85520184	0.01320499
GPHOPE18s476R	0.68002843	0.00730217
GPHOPE18s477R	0.0729713	0.00990648
GPHOPE18s478R	0.46232561	0.02330732
GPHOPE18s479R	0.52646838	0.0407695
GPHOPE18s480R	0.41093598	0.0891824
GPHOPE18s481R	0.75320663	0.06166757
GPHOPE18s482R	0.75454249	0.021976
GPHOPE18s483R	0.78539763	0.00809668
GPHOPE18s484R	0.69478082	0.05947429
GPHOPE18s485R	0.52010765	0.03560739
GPHOPE18s486R	0.09081625	0.02408647
GPHOPE18s487R	0.68438071	0.02605017
GPJOUG18s007R	0.72225956	0.01322423
GPJOUG18s008R	0.57732393	0.03408041
GPJOUG18s009R	0.55379015	0.02243489
GPJOUG18s555R	0.69560502	0.06787588
GPJOUG18s556R	0.80585956	0.00805521
GPJOUG18s557R	0.67055957	0.01010372
GPJOUG18s558R	0.21339963	0.09928869
GPJOUG18s559R	0.5400424	0.03134813
GPJOUG18s560R	0.82063358	0.0395065

GPJOUG18s561R	0.77408796	0.01998779
GPJOUG18s562R	0.65203168	0.06675467
GPJOUG18s563R	0.70872779	0.03414498
GPJOUG18s564R	0.57580702	0.00658421
GPJOUG18s565R	0.70312971	0.05602825
GPJOUG18s566R	0.70087782	0.03446403
GPJOUG18s567R	0.33763045	0.04829229
GPJOUG18s568R	0.65482316	0.03198427
GPJOUG18s569R	0.79907544	0.0119327
GPJOUG18s570R	0.86171063	0.01196561
GPJOUG18s571R	0.69954408	0.02669665
GPJOUG18s572R	0.5865637	0.01835517
GPJOUG18s573R	0.77713792	0.00336394
GPJOUG18s574R	0.50077359	0.17033643
GPLOCK18s010R	0.50455466	0.04438493
GPLOCK18s011R	0.84818657	0.01101493
GPLOCK18s012R	0.87155973	0.00821306
GPMH15f001	0.89996411	0.02383095
GPMH15f002	0.87106927	0.00481653
GPMH15f003	0.50885537	0.0349147
GPMH15f004	0.74550712	0.00507541
GPMH15f005	0.50449567	0.02190535
GPMH15f006	0.81876993	0.00480597
GPMH15f007	0.76882182	0.01146866
GPMH15f008	0.66525025	0.09583339
GPMH15f009	0.76195846	0.00775936
GPMH15f010	0.84445222	0.0069504
GPMIKK18s369R	0.77531658	0.03216553
GPMIKK18s370R	0.86487735	0.03884358
GPMIKK18s371R	0.47961773	0.0436394
GPMIKK18s372R	0.84159377	0.01957078
GPMIKK18s373R	0.60058869	0.46447036
GPMIKK18s374R	0.91994081	0.02184017
GPMIKK18s375R	0.67569907	0.06519554
GPMIKK18s376R	0.76056095	0.02931767
GPMIKK18s377R	0.88037446	0.00945153
GPMIKK18s378R	0.7071129	0.17474991
GPMIKK18s379R	0.88435004	0.02133428
GPMIKK18s380R	0.73832377	0.02625951
GPMIKK18s381R	0.59353133	0.05398561
GPMIKK18s382R	0.58883049	0.11511962
GPMIKK18s383R	0.89359427	0.00750427

GPMIKK18s384R	0.46657983	0.08600754
GPMIKK18s385R	0.92876762	0.00288443
GPMIKK18s386R	0.91464211	0.02094934
GPMIKK18s387R	0.95195248	0.00797388
GPMIKK18s388R	0.244856	0.12707749
GPMOOT18s050R	0.84108375	0.00442412
GPMOOT18s051R	0.87541043	0.0111506
GPMOOT18s067R	0.90500691	0.00683278
GPMOOT18s068R	0.91891553	0.00793375
GPMOOT18s069R	0.92867939	0.00527228
GPMOOT18s070R	0.79691267	0.01509217
GPNEKO18s061R	0.68529415	0.02996422
GPNEKO18s062R	0.91445525	0.01420272
GPNEKO18s063R	0.77306785	0.08848669
GPNEKO18s064R	0.77935396	0.00350715
GPNEKO18s065R	0.87064273	0.01506281
GPNEKO18s080R	0.50035829	0.02164065
GPNEKO18s081R	0.69400564	0.03905668
GPNEKO18s082R	0.89200234	0.00656304
GPNEKO18s083R	0.89636697	0.00352825
GPNEKO18s084R	0.8829144	0.01569603
GPPETE18s071R	0.93021051	0.0001326
GPPETE18s072R	0.92194341	0.01234893
GPPETE18s073R	0.90860349	0.02607394
GPPETE18s074R	0.90419705	0.01625162
GPPETE18s075R	0.49841716	0.01403345
GPPTCH18s488R	0.58460855	0.01739911
GPPTCH18s489R	0.86850409	0.00325053
GPPTCH18s490R	0.77962922	0.00749144
GPPTCH18s491R	0.85996332	0.00536195
GPPTCH18s492R	0.5491658	0.02665882
GPPTCH18s493R	0.64678822	0.0211475
GPPTCH18s494R	0.60147262	0.01445302
GPPTCH18s495R	0.94507185	0.00747974
GPPTCH18s496R	0.79768684	0.00237429
GPPTCH18s497R	0.80968571	0.01502255
GPPTCH18s498R	0.23086623	0.04442063
GPPTCH18s499R	0.83302004	0.01916026
GPPTCH18s500R	0.73679021	0.0143159
GPPTCH18s501R	0.56749028	0.01152262
GPPTCH18s502R	0.35770189	0.03212128
GPPTCH18s503R	0.45125487	0.05550428

GPPTCH18s504R	0.64759927	0.03333048
GPPTCH18s505R	0.75330223	0.00642371
GPPTCH18s506R	0.77667658	0.03578116
GPPTCH18s507R	0.73284534	0.01159267
GPSALI15f001	0.59488286	0.06251097
GPSALI15f002	0.41863432	0.00182417
GPSALI15f003	0.4778043	0.02102995
GPSALI15f004	0.9058331	0.01667296
GPSALI15f005	0.97824	0.00234685
GPSALI15f006	0.92124162	0.01739854
GPSALI15f007	0.76760067	0.01321109
GPSALI15f008	0.72713389	0.00587703
GPSALI15f009	0.60960477	0.17209125
GPSALI15f010	0.6982262	0.03940885
GPSALI15f011	0.42646599	0.10761076
GPSALI15f012	0.87449315	0.02041883
GPSALI15f013	0.80687829	0.0536351
GPSALI17f001	0.63029803	0.02794331
GPSALI17f002	0.37379004	0.00612145
GPSALI17f003	0.74449755	0.02022386
GPSALI17f004	0.53719756	0.02805102
GPSALI17f005	0.83811679	0.03941844
GPSALI17f006	0.51444406	0.1076852
GPSALI17f007	0.09520009	0.04093985
GPSALI17f008	0.91950366	0.02206736
GPSALI17f009	0.5459838	0.03525071
GPSALI17f010	0.17615977	0.08624928
GPSALI17f011	0.19506836	0.05541315
GPSALI17f012	0.18060645	0.0283507
GPSALI17f013	0.58099538	0.04066706
GPSALI17f014	0.60345609	0.20153362
GPSALI17f015	0.49115986	0.00229352
GPSALI18s601R	0.88928517	0.04008047
GPSALI18s602R	0.81578101	0.0593572
GPSALI18s603R	0.72287789	0.088231
GPSALI18s604R	0.54810767	0.07138455
GPSALI18s605R	0.83269526	0.06240305
GPSALI18s606R	0.76338319	0.02648491
GPSALI18s607R	0.45370417	0.05529167
GPSALI18s608R	0.85836208	0.00623714
GPSALI18s609R	0.76509598	0.02597403
GPSALI18s610R	0.75230927	0.03098417

GPSALI18s611R	0.88853322	0.00809487
GPSALI18s612R	0.74796623	0.01518004
GPSALI18s613R	0.4422284	0.02355602
GPSALI18s614R	0.78685713	0.04470148
GPSALI18s615R	0.12932175	0.01985723
GPSALI18s616R	0.82689552	0.01072474
GPSALI18s617R	0.60430472	0.05985242
GPSALI18s618R	0.11442185	0.03303654
GPSALI18s619R	0.82693822	0.03044803
GPSALI18s620R	0.81825008	0.00745413
GPSALI18s621R	0.63873536	0.04184636
GPSAUN15f001	0.53672114	0.04307591
GPSAUN15f002	0.17548197	0.04532249
GPSAUN15f003	0.6879001	0.02857929
GPSAUN15f004	0.68481101	0.00658952
GPSAUN15f005	0.55929947	0.0276034
GPSAUN15f006	0.26060752	0.10043548
GPSAUN15f007	0.64670261	0.03024825
GPSAUN15f008	0.76516433	0.01732224
GPSAUN15f009	0.55615486	0.01358464
GPSAUN15f010	0.68028399	0.02861385
GPSAUN18f001R	0.75971884	0.03787906
GPSAUN18f002R	0.66478857	0.10558953
GPSAUN18f003R	0.81200292	0.04674802
GPSAUN18f004R	0.88211903	0.02371801
GPSAUN18f005R	0.87790659	0.03829418
GPSAUN18f006R	0.78604241	0.10384003
GPSAUN18f007R	0.82945001	0.05003966
GPSAUN18f008R	0.84661652	0.02298996
GPSAUN18f009R	0.69347091	0.01584909
GPSAUN18f010R	0.84257232	0.01167618
GPSAUN18f011R	0.74141692	0.06405973
GPSAUN18f012R	0.85170713	0.03101047
GPSAUN18f013R	0.79257737	0.02989728
GPSAUN18f014R	0.25608188	0.13538801
GPSAUN18f015R	0.81467116	0.05427893
GPSAUN18f016R	0.79317391	0.02881891
GPSAUN18f017R	0.88241504	0.01045703
GPSAUN18f018R	0.8738636	0.01976297
GPSAUN18f019R	0.31294145	0.20019533
GPSAUN18sw020	0.87982047	0.0290321
GPSELV18s020R	0.55713174	0.03838825

GPSELV18s021R	0.87789755	0.00465537
GPSPRT18s024R	0.71370616	0.10586386
WC	0.79752704	0.11143905

[a] Samples names are ordered by the two or three letter species code, the four letter site ID, and the 2 number year identification, and the individual information.

18S Samples

Samples[a]	**Mean**	**Std. Dev.**
APHOPE18s085R	0.60047161	0.16764692
APHOPE18s086R	0.80052452	0.00160497
APHOPE18s087R	0.49958372	0.02559225
APHOPE18s088R	0.69326557	0.133397
APHOPE18s089R	0.73617583	0.04998504
APHOPE18s095R	0.66926702	0.27213396
BESHJOUG18s575R	0.12328086	0.05935364
BESHJOUG18s576R	0.85354078	0.00929709
BESHJOUG18s577R	0.64873993	0.05497437
BESHJOUG18s578R	0.52687278	0.01759195
BESHJOUG18s579R	0.69034639	0.03009588
BESHPAUL18s096R	0.08102897	0.04644051
BESHPAUL18s097R	0.77458508	0.06086778
BESHPAUL18s098R	0.4754623	0.01214094
BESHPAUL18s099R	0.6680364	0.10028201
BESHPAUL18s100R	0.77865536	0.0431553
BESHPAUL18s101R	0.71673875	0.05590957
BESHTETR18s548R	0.51087684	0.40410408
BESHTETR18s549R	0.67253715	0.12839789
BESHTETR18s550R	0.78762428	0.07637568
BESHTETR18s551R	0.36988375	0.02116532
BESHTETR18s552R	0.72792173	0.18590717
BESHTETR18s553R	0.84392136	0.00946442
BESHTETR18s554R	0.85626496	0.02243947
CPAITC18s428R	0.86716126	0.03148217
CPAITC18s429R	0.77646166	0.02056694
CPAITC18s430R	0.72467638	0.20589901
CPAITC18s431R	0.61432002	0.21127755
CPAITC18s432R	0.72173968	0.05127314
CPAITC18s433R	0.79821806	0.06651062
CPAITC18s434R	0.75933544	0.02933772
CPAITC18s435R	0.84115206	0.08769046
CPAITC18s436R	0.79269354	0.171511
CPAITC18s437R	0.80008481	0.03972776

CPAITC18s438R	0.74941383	0.10375841
CPAITC18s439R	0.62978719	0.2091236
CPAITC18s440R	0.58072892	0.35032439
CPAITC18s441R	0.73321634	0.05791123
CPAITC18s442R	0.73564555	0.05226761
CPAITC18s443R	0.80288363	0.03826108
CPAITC18s444R	0.81564012	0.06857793
CPAITC18s445R	0.7129972	0.19488009
CPAITC18s446R	0.82232295	0.05146667
CPAITC18s447R	0.82598452	0.02213781
GPAITC18s448R	0.71541134	0.0619193
GPAITC18s449R	0.6729985	0.0287099
GPAITC18s450R	0.67212682	0.03154179
GPAITC18s451R	0.73442874	0.08469356
GPAITC18s452R	0.71567509	0.08887729
GPAITC18s453R	0.61591008	0.05511413
GPAITC18s454R	0.69015219	0.01911619
GPAITC18s455R	0.71848678	0.03033026
GPAITC18s456R	0.70834459	0.02718677
GPAITC18s457R	0.69131572	0.02915796
GPAITC18s458R	0.6931619	0.03487129
GPAITC18s459R	0.68163478	0.02722495
GPAITC18s460R	0.70595854	0.05393134
GPAITC18s461R	0.68596786	0.04228298
GPAITC18s462R	0.67870642	0.04093175
GPAITC18s463R	0.69428859	0.04477209
GPAITC18s464R	0.6713176	0.035594
GPAITC18s465R	0.66801042	0.04307432
GPAITC18s466R	0.67422223	0.02372942
GPAITC18s467R	0.71688083	0.05156419
GPALMI18s046R	0.50848845	0.03288807
GPALMI18s047R	0.64270401	0.04920682
GPALMI18s052R	0.6579803	0.00524376
GPALMI18s053R	0.56273978	0.03775527
GPALMI18s054R	0.59716406	0.00891585
GPBOOT16f001dc	0.69659105	0.01951099
GPBOOT16f002dc	0.77984056	0.02728002
GPBOOT16f003dc	0.77580314	0.01843975
GPBOOT16f004dc	0.76424323	0.02918077
GPBOOT16f005dc	0.75447038	0.03423223
GPBOOT16f006dc	0.75067024	0.08283698
GPBOOT16f007dc	0.79196907	0.00551636

GPBOOT16f008dc	0.86003818	0.01162463
GPBOOT16f009dc	0.81081658	0.029272
GPBOOT16f010dc	0.8478063	0.01695748
GPBR15f001	0.37337748	0.2197614
GPBR15f002	0.23606932	0.06970355
GPBR15f003	0.80910584	0.06209368
GPBR15f004	0.47565004	0.18084637
GPBR15f005	0.46731204	0.05669645
GPBR15f006	0.44145157	0.11178004
GPBR15f007	0.73101025	0.12347821
GPBR15f008	0.46778916	0.26186093
GPBR15f009	0.73702077	0.08042044
GPBR15f010	0.82622074	0.03243314
GPBROW18s508R	0.47995352	0.12353886
GPBROW18s509R	0.51012318	0.12021252
GPBROW18s510R	0.61805602	0.02456496
GPBROW18s511R	0.46726375	0.12483679
GPBROW18s512R	0.28837206	0.3129311
GPBROW18s513R	0.43619119	0.22876553
GPBROW18s514R	0.64410033	0.08756395
GPBROW18s515R	0.54347983	0.09641731
GPBROW18s516R	0.46811846	0.17190538
GPBROW18s517R	0.6019646	0.16965764
GPBROW18s518R	0.61497212	0.02731591
GPBROW18s519R	0.55528098	0.16222509
GPBROW18s520R	0.51494875	0.13649122
GPBROW18s521R	0.47963643	0.12084679
GPBROW18s522R	0.54079516	0.09567033
GPBROW18s523R	0.46622297	0.11583688
GPBROW18s524R	0.30873556	0.25482837
GPBROW18s525R	0.74032239	0.09053366
GPBROW18s526R	0.68302231	0.03333789
GPBROW18s527R	0.44384588	0.13012794
GPBRYE16f001dc	0.88805924	0.02378788
GPBRYE16f002dc	0.76576071	0.02032177
GPBRYE16f003dc	0.81277869	0.02904637
GPBRYE16f004dc	0.57256804	0.01671752
GPBRYE16f005dc	0.5779681	0.02583939
GPBRYE16f006dc	0.84627208	0.03358715
GPBRYE16f007dc	0.67228542	0.02227229
GPBRYE16f008dc	0.7536485	0.02492898
GPBRYE16f009dc	0.81828937	0.01626284

GPBRYE16f010dc	0.7937078	0.03207149
GPBRYE18s048R	0.68343178	0.05419438
GPBRYE18s049R	0.71822371	0.05790171
GPBRYE18s055R	0.68911395	0.02001395
GPBRYE18s056R	0.65969806	0.00578865
GPBRYE18s057R	0.69976892	0.04706256
GPBRYE18s389R	0.73900393	0.04852792
GPBRYE18s390R	0.68941393	0.04467276
GPBRYE18s391R	0.66587335	0.03794697
GPBRYE18s392R	0.63640672	0.03240965
GPBRYE18s393R	0.69832318	0.01703583
GPBRYE18s394R	0.70300893	0.04786788
GPBRYE18s395R	0.66908283	0.00930713
GPBRYE18s396R	0.71366124	0.0272049
GPBRYE18s397R	0.69973212	0.05889459
GPBRYE18s398R	0.67737897	0.03284398
GPBRYE18s399R	0.66643252	0.02997447
GPBRYE18s400R	0.70247601	0.05010815
GPBRYE18s401R	0.68024784	0.00992455
GPBRYE18s402R	0.68431615	0.02370352
GPBRYE18s403R	0.66382898	0.03464454
GPBRYE18s404R	0.67882959	0.04488669
GPBRYE18s405R	0.70161197	0.02768637
GPBRYE18s406R	0.69624941	0.01860395
GPBRYE18s407R	0.76894511	0.0116183
GPCOOP15f001	0.42999644	0.19901182
GPCOOP15f002	0.7236941	0.07936105
GPCOOP15f003	0.88257152	0.01088059
GPCOOP15f004	0.42187927	0.13206908
GPCOOP15f005	0.52092252	0.20391429
GPCOOP15f006	0.71543927	0.1403668
GPCOOP15f007	0.75267308	0.05478122
GPCOOP15f008	0.58521933	0.0904917
GPCOOP15f009	0.6413244	0.08499252
GPCOOP15f010	0.43710522	0.01030117
GPCUVE16f001dc	0.81909212	0.01900715
GPCUVE16f002dc	0.82895266	0.05049819
GPCUVE16f003dc	0.83707847	0.02180125
GPCUVE16f004dc	0.89042059	0.00537021
GPCUVE16f005dc	0.79479829	0.02604022
GPCUVE16f006dc	0.77400473	0.01542185
GPCUVE16f007dc	0.86395672	0.0164007

GPCUVE16f008dc	0.79803017	0.0331714
GPCUVE16f009dc	0.61138676	0.04013971
GPCUVE16f010dc	0.78313605	0.01277329
GPCUVE18s037R	0.63489429	0.01808832
GPCUVE18s038R	0.69151564	0.07948267
GPCUVE18s039R	0.61076757	0.0162165
GPCUVE18s040R	0.46836928	0.13595236
GPCUVE18s041R	0.61569445	0.20376905
GPCUVE18s042R	0.5867741	0.24537551
GPCUVE18s043R	0.7177933	0.06549583
GPCUVE18s044R	0.51301379	0.11975803
GPCUVE18s045R	0.54662119	0.14236312
GPDAMO18s058R	0.18659983	0.01920006
GPDAMO18s059R	0.41513491	0.01957806
GPDAMO18s060R	0.28541584	0.06959956
GPDANC18s001R	0.71306994	0.04611017
GPDANC18s002R	0.58614256	0.05699764
GPDANC18s003R	0.78122341	0.07190369
GPDANC18s004R	0.60369465	0.04782108
GPDANC18s005R	0.63628621	0.0454112
GPDANC18s006R	0.49818223	0.1269357
GPEARL16f001dc	0.64496681	0.05216957
GPEARL16f001R	0.51334843	0.06497375
GPEARL16f002dc	0.77655496	0.04428945
GPEARL16f002R	0.66847856	0.09498662
GPEARL16f003dc	0.67801716	0.13890944
GPEARL16f003R	0.75299206	0.10448056
GPEARL16f004dc	0.63449162	0.02048453
GPEARL16f004R	0.58228224	0.22291143
GPEARL16f005dc	0.6411561	0.07265923
GPEARL16f005R	0.6497038	0.17709567
GPEARL16f006dc	0.80513226	0.05589031
GPEARL16f006R	0.69472163	0.04211201
GPEARL16f007dc	0.67991539	0.21403255
GPEARL16f007R	0.5357355	0.10178355
GPEARL16f008dc	0.44298294	0.22276011
GPEARL16f008R	0.623297	0.03295214
GPEARL16f009dc	0.445814	0.14460436
GPEARL16f009R	0.67103078	0.01210053
GPEARL16f010dc	0.73792182	0.03892897
GPEARL16f010R	0.45688916	0.13013423
GPFORT18s028R	0.3731311	0.22088114

GPFORT18s029R	0.41392123	0.23286498
GPFORT18s030R	0.71241146	0.07192241
GPFORT18s034R	0.5705119	0.09965927
GPFORT18s035R	0.628867	0.03975176
GPFORT18s036R	0.37324002	0.19212983
GPGEOR16f001dc	0.73795271	0.02630322
GPGEOR16f003dc	0.81005909	0.02807745
GPGEOR16f004dc	0.80251561	0.04085648
GPGEOR16f005dc	0.81763793	0.027382
GPGEOR16f006dc	0.85959904	0.01762607
GPGEOR16f007dc	0.8046883	0.01148959
GPGEOR16f008dc	0.78905467	0.04659328
GPGEOR16f009dc	0.872447	0.0358412
GPGEOR16f010dc	0.51541565	0.01171957
GPHER16f001dc	0.48172764	0.05388012
GPHER16f001R	0.78756034	0.06148311
GPHER16f002dc	0.80170591	0.03097826
GPHER16f002R	0.75870093	0.02384975
GPHER16f003dc	0.79465806	0.00672277
GPHER16f003R	0.80657036	0.05398514
GPHER16f004dc	0.85547668	0.03521438
GPHER16f004R	0.36532303	0.0728277
GPHER16f005dc	0.50025546	0.14287651
GPHER16f005R	0.42776943	0.20215994
GPHER16f006dc	0.59669094	0.16348464
GPHER16f006R	0.31766906	0.09759627
GPHER16f007dc	0.66644671	0.06518628
GPHER16f007R	0.69729997	0.08110864
GPHER16f008dc	0.69783297	0.02939564
GPHER16f008R	0.81479011	0.05920639
GPHER16f009dc	0.63232153	0.00595627
GPHER16f009R	0.43454332	0.12434456
GPHER16f010dc	0.57056115	0.04798148
GPHER16f010R	0.25527255	0.063833
GPHOPE18s066R	0.5525976	0.10576635
GPHOPE18s090R	0.65287878	0.1024479
GPHOPE18s091R	0.50743069	0.1411251
GPHOPE18s092R	0.40768717	0.23781732
GPHOPE18s093R	0.32000005	0.28990437
GPHOPE18s094R	0.29931374	0.30637018
GPHOPE18s468R	0.4056983	0.17326789
GPHOPE18s469R	0.61101113	0.08102157

GPHOPE18s470R	0.33161075	0.21123518
GPHOPE18s471R	0.33186258	0.24730019
GPHOPE18s472R	0.52318047	0.12597193
GPHOPE18s473R	0.27067279	0.13326946
GPHOPE18s474R	0.69214697	0.11997338
GPHOPE18s475R	0.61732611	0.1441457
GPHOPE18s476R	0.66407807	0.10947441
GPHOPE18s477R	0.63879472	0.03156814
GPHOPE18s478R	0.68285299	0.06902008
GPHOPE18s479R	0.72744477	0.05885746
GPHOPE18s480R	0.73044806	0.07098066
GPHOPE18s481R	0.70481739	0.08302106
GPHOPE18s482R	0.71380013	0.08778635
GPHOPE18s483R	0.64330597	0.09826017
GPHOPE18s484R	0.51226715	0.17700264
GPHOPE18s485R	0.74823232	0.0927991
GPHOPE18s486R	0.66936518	0.10818687
GPHOPE18s487R	0.42953719	0.17978405
GPJOUG18s007R	0.62721462	0.03668769
GPJOUG18s008R	0.60118372	0.03476684
GPJOUG18s009R	0.60807149	0.03506836
GPJOUG18s555R	0.68253498	0.03557219
GPJOUG18s556R	0.55541102	0.06809086
GPJOUG18s557R	0.36161444	0.17213216
GPJOUG18s558R	0.60266292	0.05984985
GPJOUG18s559R	0.69853653	0.04807629
GPJOUG18s560R	0.6303542	0.02836938
GPJOUG18s561R	0.66974936	0.11291095
GPJOUG18s562R	0.6029004	0.01445034
GPJOUG18s563R	0.44210533	0.15635475
GPJOUG18s564R	0.67247621	0.04345366
GPJOUG18s565R	0.69656235	0.01047508
GPJOUG18s566R	0.58267851	0.06785152
GPJOUG18s567R	0.65781084	0.03515246
GPJOUG18s568R	0.67779086	0.04007032
GPJOUG18s569R	0.54190189	0.10567264
GPJOUG18s570R	0.6373603	0.05323078
GPJOUG18s571R	0.59857962	0.09060407
GPJOUG18s572R	0.42362576	0.17095582
GPJOUG18s573R	0.66304174	0.0261787
GPJOUG18s574R	0.34074632	0.23255485
GPLOCK18s010R	0.72713819	0.12506454

GPLOCK18s011R	0.57992277	0.07295208
GPLOCK18s012R	0.54721157	0.17973478
GPMH15f001	0.81162082	0.06687951
GPMH15f002	0.64198419	0.14551916
GPMH15f003	0.47559014	0.19416939
GPMH15f004	0.51647202	0.11829856
GPMH15f005	0.61915181	0.11312625
GPMH15f006	0.30680498	0.12275952
GPMH15f007	0.2199524	0.097158
GPMH15f008	0.74810787	0.04969584
GPMH15f009	0.72299757	0.03584556
GPMH15f010	0.8149492	0.0339378
GPMH16f001dc	0.85326966	0.01225284
GPMH16f002dc	0.85585572	0.00377252
GPMH16f003dc	0.86017509	0.00906086
GPMH16f004dc	0.62869427	0.03609568
GPMH16f005dc	0.86013195	0.02598698
GPMH16f006dc	0.76415354	0.03871408
GPMH16f007dc	0.7414345	0.01478122
GPMH16f008dc	0.52873193	0.09792434
GPMH16f009dc	0.66149029	0.01415366
GPMH16f010dc	0.64522853	0.06610295
GPMIKK18s369R	0.79895919	0.00291857
GPMIKK18s370R	0.80563849	0.06131464
GPMIKK18s371R	0.75809137	0.07281109
GPMIKK18s372R	0.64000549	0.01923349
GPMIKK18s373R	0.81418851	0.02167797
GPMIKK18s374R	0.80381737	0.0447908
GPMIKK18s375R	0.81408418	0.05966309
GPMIKK18s376R	0.79814086	0.04845961
GPMIKK18s377R	0.80302435	0.02140709
GPMIKK18s378R	0.86079678	0.02187231
GPMIKK18s379R	0.80551382	0.02917491
GPMIKK18s380R	0.81125159	0.00322236
GPMIKK18s381R	0.74950825	0.06265927
GPMIKK18s382R	0.79124089	0.02538042
GPMIKK18s383R	0.78192098	0.01688794
GPMIKK18s384R	0.7074762	0.04755154
GPMIKK18s385R	0.79924687	0.03126533
GPMIKK18s386R	0.77887108	0.06639264
GPMIKK18s387R	0.79375167	0.01483906
GPMIKK18s388R	0.80266919	0.02959544

GPMOOT16f001dc	0.70878758	0.02378262
GPMOOT16f002dc	0.72334423	0.02605752
GPMOOT16f003dc	0.76001317	0.01342755
GPMOOT16f004dc	0.76102821	0.02842808
GPMOOT16f006dc	0.81031513	0.02696123
GPMOOT16f007dc	0.73833788	0.04753731
GPMOOT16f008dc	0.76951467	0.04887831
GPMOOT16f009dc	0.59298532	0.02237925
GPMOOT16f010dc	0.73335637	0.02017722
GPMOOT18s050R	0.64610504	0.081959
GPMOOT18s051R	0.7592227	0.12105395
GPMOOT18s067R	0.74465075	0.11231753
GPMOOT18s068R	0.65589879	0.04895728
GPMOOT18s069R	0.67797913	0.1203341
GPMOOT18s070R	0.80329506	0.13410317
GPNEKO16f001dc	0.77378125	0.08881424
GPNEKO16f002dc	0.53146532	0.05814267
GPNEKO16f003dc	0.81034764	0.02208866
GPNEKO16f004dc	0.81869319	0.01564473
GPNEKO16f005dc	0.79154531	0.02024082
GPNEKO16f006dc	0.8018225	0.04585793
GPNEKO16f007dc	0.65566139	0.04442302
GPNEKO16f008dc	0.8353471	0.02349613
GPNEKO16f009dc	0.82443042	0.04180539
GPNEKO16f010dc	0.67009959	0.01436114
GPNEKO18s061R	0.65310871	0.02765017
GPNEKO18s062R	0.6544181	0.0226092
GPNEKO18s063R	0.69628834	0.0932666
GPNEKO18s064R	0.70000624	0.07827793
GPNEKO18s065R	0.57551075	0.13919317
GPNEKO18s080R	0.59058547	0.03540924
GPNEKO18s081R	0.72194096	0.08681431
GPNEKO18s082R	0.7417642	0.07631518
GPNEKO18s083R	0.57199466	0.06796386
GPNEKO18s084R	0.43576007	0.20880427
GPPETE16f001dc	0.57782395	0.15327734
GPPETE16f002dc	0.68520707	0.01102204
GPPETE16f003dc	0.86662062	0.0160203
GPPETE16f004dc	0.85524938	0.02791542
GPPETE16f005dc	0.7742328	0.04881666
GPPETE16f006dc	0.77531381	0.01680798
GPPETE16f007dc	0.78802609	0.01958733

GPPETE16f008dc	0.81918791	0.03502908
GPPETE16f009dc	0.67202901	0.02062262
GPPETE16f010dc	0.80214524	0.00722832
GPPETE18s071R	0.49609749	0.40301029
GPPETE18s072R	0.66920076	0.06832436
GPPETE18s073R	0.72818831	0.11310846
GPPETE18s074R	0.59861692	0.10943131
GPPETE18s075R	0.533702	0.08526994
GPPTCH18s488R	0.81610591	0.14346204
GPPTCH18s489R	0.81090002	0.12440246
GPPTCH18s490R	0.49317648	0.15838425
GPPTCH18s491R	0.35804081	0.2413506
GPPTCH18s492R	0.33896676	0.25701617
GPPTCH18s493R	0.35916195	0.23127053
GPPTCH18s494R	0.43848545	0.16486667
GPPTCH18s495R	0.41287372	0.2196092
GPPTCH18s496R	0.70075724	0.02985413
GPPTCH18s497R	0.51874145	0.12299514
GPPTCH18s498R	0.47436382	0.41445825
GPPTCH18s499R	0.69179105	0.05743681
GPPTCH18s500R	0.30583699	0.29378923
GPPTCH18s501R	0.57418729	0.05941035
GPPTCH18s502R	0.55627734	0.06939831
GPPTCH18s503R	0.69733287	0.05157878
GPPTCH18s504R	0.80749149	0.13617407
GPPTCH18s505R	0.45922966	0.16860364
GPPTCH18s506R	0.77530969	0.09836263
GPPTCH18s507R	0.65256982	0.0989184
GPSALI15f001	0.72027815	0.07766861
GPSALI15f002	0.54181174	0.17983597
GPSALI15f003	0.67354992	0.044462
GPSALI15f004	0.68353474	0.03403068
GPSALI15f005	0.48765839	0.02499153
GPSALI15f006	0.81272527	0.06112022
GPSALI15f007	0.50342528	0.06906937
GPSALI15f008	0.44907481	0.11181607
GPSALI15f009	0.66503116	0.05022732
GPSALI15f010	0.69264234	0.0409721
GPSALI15f011	0.66346352	0.10689086
GPSALI15f012	0.7570753	0.1075137
GPSALI15f013	0.49940351	0.03867
GPSALI16f001dc	0.92013291	0.00646444

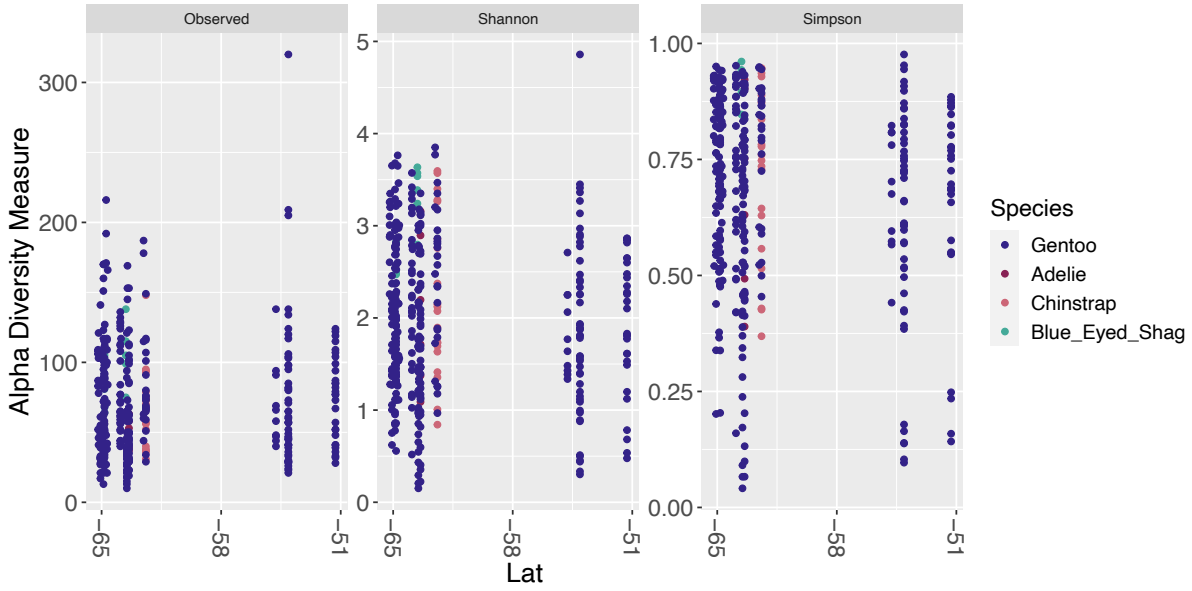
GPSALI16f002dc	0.89187966	0.00511677
GPSALI16f003dc	0.79872244	0.00551673
GPSALI16f004dc	0.87059577	0.00867562
GPSALI16f005dc	0.89300653	0.00360722
GPSALI16f006dc	0.9199684	0.00499026
GPSALI16f007dc	0.84225515	0.09736561
GPSALI16f008dc	0.90119194	0.00320134
GPSALI16f009dc	0.91467877	0.00349113
GPSALI16f010dc	0.78314456	0.0242349
GPSALI17f001	0.61458094	0.01687397
GPSALI17f002	0.43874952	0.15715716
GPSALI17f003	0.23395624	0.05511928
GPSALI17f004	0.78890997	0.0221317
GPSALI17f005	0.79953857	0.08047425
GPSALI17f006	0.84625042	0.03712352
GPSALI17f007	0.83552621	0.03921296
GPSALI17f008	0.80965805	0.05611695
GPSALI17f009	0.71623713	0.03384764
GPSALI17f010	0.77267883	0.06180453
GPSALI17f011	0.76097916	0.07910522
GPSALI17f012	0.73121713	0.12259048
GPSALI17f013	0.72625769	0.06923153
GPSALI17f014	0.27653679	0.03884632
GPSALI17f015	0.77013055	0.05675524
GPSALI18s601R	0.55592021	0.04644545
GPSALI18s602R	0.82119196	0.03914569
GPSALI18s603R	0.83030654	0.02183254
GPSALI18s604R	0.79583653	0.05743455
GPSALI18s605R	0.75349391	0.06833198
GPSALI18s606R	0.78479862	0.07034949
GPSALI18s607R	0.67650927	0.23125092
GPSALI18s608R	0.84445841	0.02935723
GPSALI18s609R	0.82443825	0.01355083
GPSALI18s610R	0.79009022	0.07778937
GPSALI18s611R	0.80193082	0.03945358
GPSALI18s612R	0.70899661	0.06831146
GPSALI18s613R	0.86803945	0.04408301
GPSALI18s614R	0.8448041	0.01572991
GPSALI18s615R	0.86869194	0.04658675
GPSALI18s616R	0.63554717	0.12100534
GPSALI18s617R	0.81904043	0.02891951
GPSALI18s618R	0.84532795	0.0318249

GPSALI18s619R	0.70223458	0.07029802
GPSALI18s620R	0.84471122	0.01973343
GPSALI18s621R	0.79695444	0.04917948
GPSAUN15f001	0.66826657	0.07872371
GPSAUN15f002	0.23475663	0.17145947
GPSAUN15f003	0.72657701	0.02449834
GPSAUN15f004	0.44259407	0.07359449
GPSAUN15f005	0.6055993	0.01456393
GPSAUN15f006	0.21607582	0.0614146
GPSAUN15f007	0.66179171	0.14611144
GPSAUN15f008	0.61685523	0.06949894
GPSAUN15f009	0.49848381	0.00705228
GPSAUN15f010	0.32723361	0.01958521
GPSAUN16f001dc	0.68510471	0.1039342
GPSAUN16f002dc	0.7775724	0.10026616
GPSAUN16f003dc	0.73369348	0.00788685
GPSAUN16f004dc	0.8007488	0.01408176
GPSAUN16f005dc	0.88142857	0.00354873
GPSAUN16f006dc	0.77603614	0.00273383
GPSAUN16f007dc	0.65113434	0.01446681
GPSAUN16f008dc	0.71162925	0.01746324
GPSAUN16f009dc	0.75842529	0.02089638
GPSAUN16f010dc	0.77744384	0.01469732
GPSAUN18f001R	0.48261948	0.18391685
GPSAUN18f002R	0.77695647	0.08562079
GPSAUN18f003R	0.69431963	0.06529495
GPSAUN18f004R	0.73188695	0.03612927
GPSAUN18f005R	0.6206838	0.00454187
GPSAUN18f006R	0.75875376	0.0193333
GPSAUN18f007R	0.90722281	0.00076609
GPSAUN18f008R	0.8760268	0.01515296
GPSAUN18f009R	0.81554291	0.04495688
GPSAUN18f010R	0.46687199	0.05426706
GPSAUN18f011R	0.36830189	0.05448481
GPSAUN18f012R	0.75327443	0.07161063
GPSAUN18f013R	0.45202271	0.02832715
GPSAUN18f014R	0.5740097	0.06157112
GPSAUN18f015R	0.58629995	0.02773191
GPSAUN18f016R	0.70525511	0.09693851
GPSAUN18f017R	0.35379962	0.07432091
GPSAUN18f018R	0.85746636	0.01406517
GPSAUN18f019R	0.27495825	0.04039998

GPSAUN18sw020	0.85192928	0.01247754
GPSELV18s020R	0.54914456	0.07650624
GPSELV18s021R	0.58972391	0.03976701
GPSPRT18s024R	0.37190951	0.04424242
WC	0.73557536	0.09423724

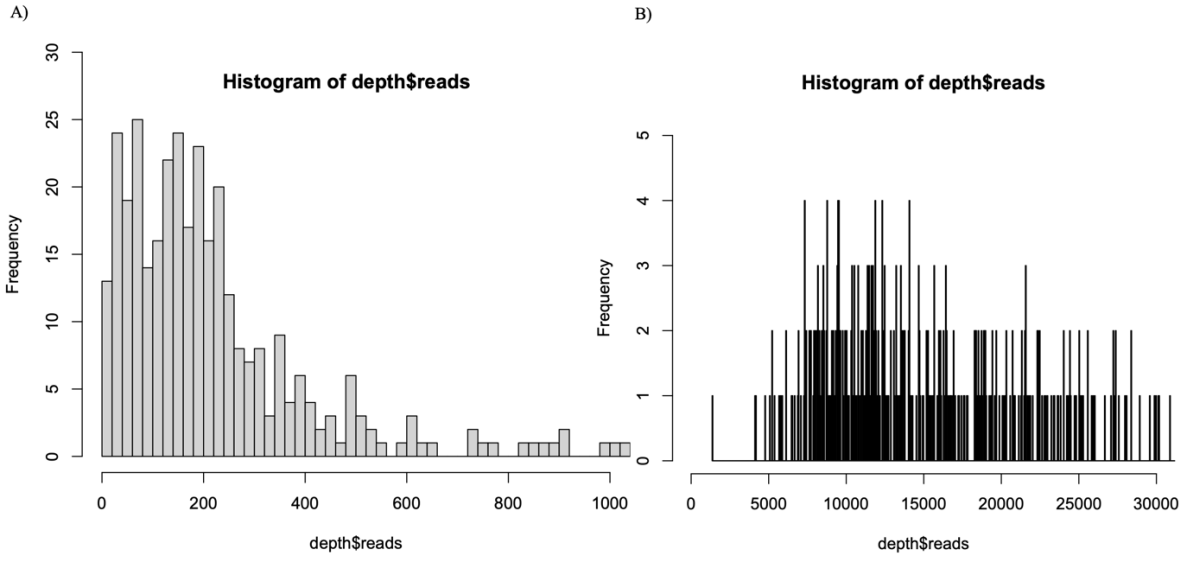
[a] Samples names are ordered by the two or three letter species code, the four letter site ID, and the 2 number year identification, and the individual information.

Supplemental Figure 1. Comparison of Alpha Diversity Measures



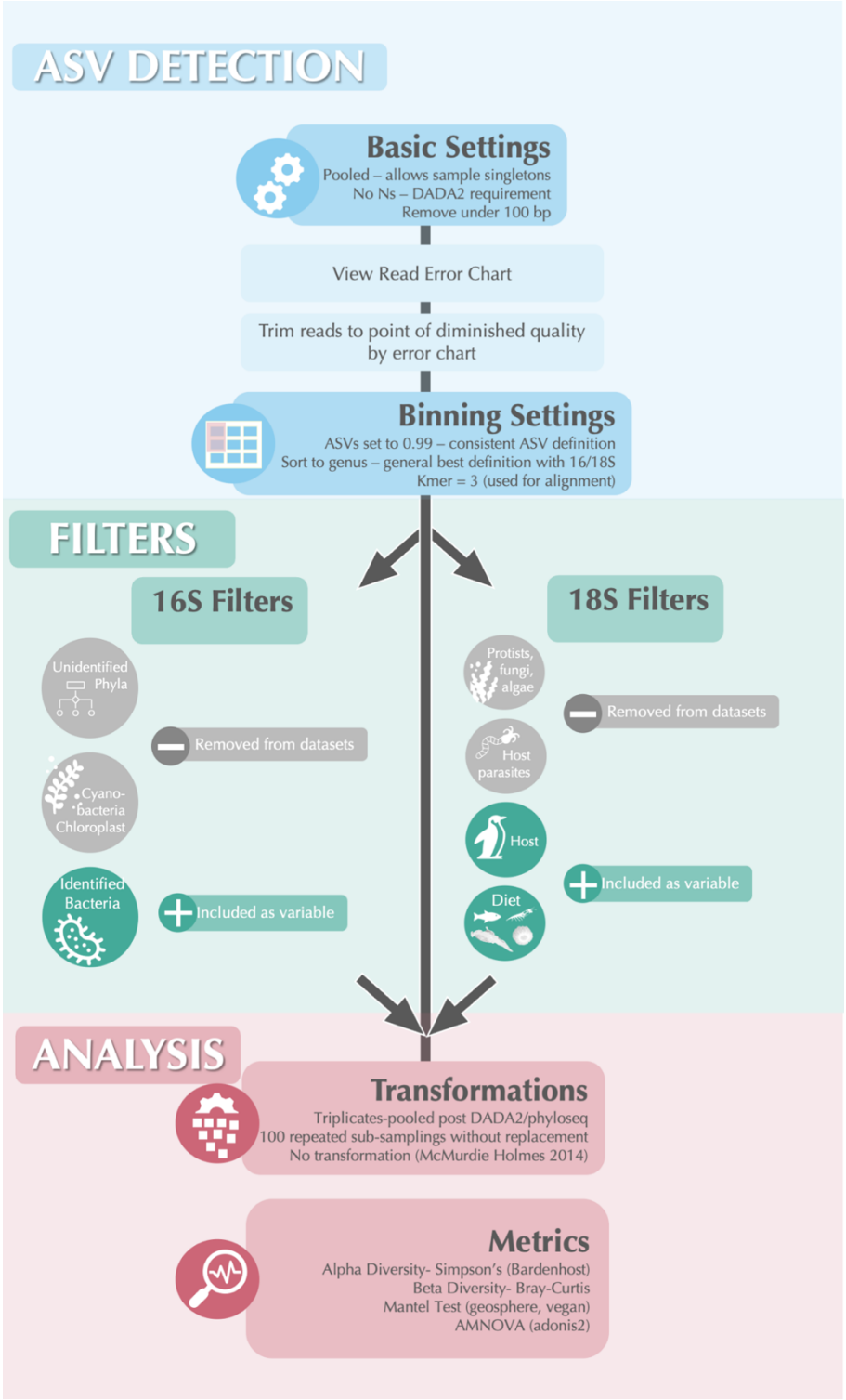
Supplemental Figure 1. Plots comparing Observed, Shannon, and Simpson metrics of Alpha diversity as a function of latitude.

Supplemental Figure 2. Histograms of Read Depth Distribution



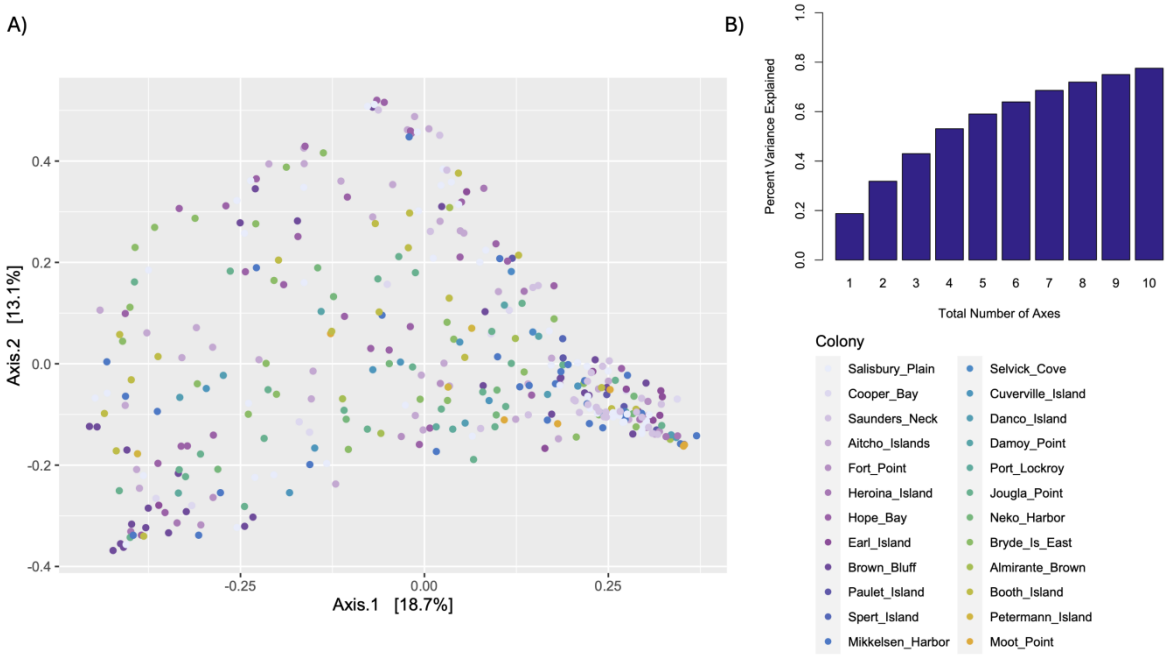
Supplemental Figure 2. Histograms of read depth distribution for A) dietary 18S rRNA gene profiles and B) 16S rRNA gene profiles. For both A and B, the right tail has been cut off.

Supplemental Figure 3. Data Analysis Pipeline



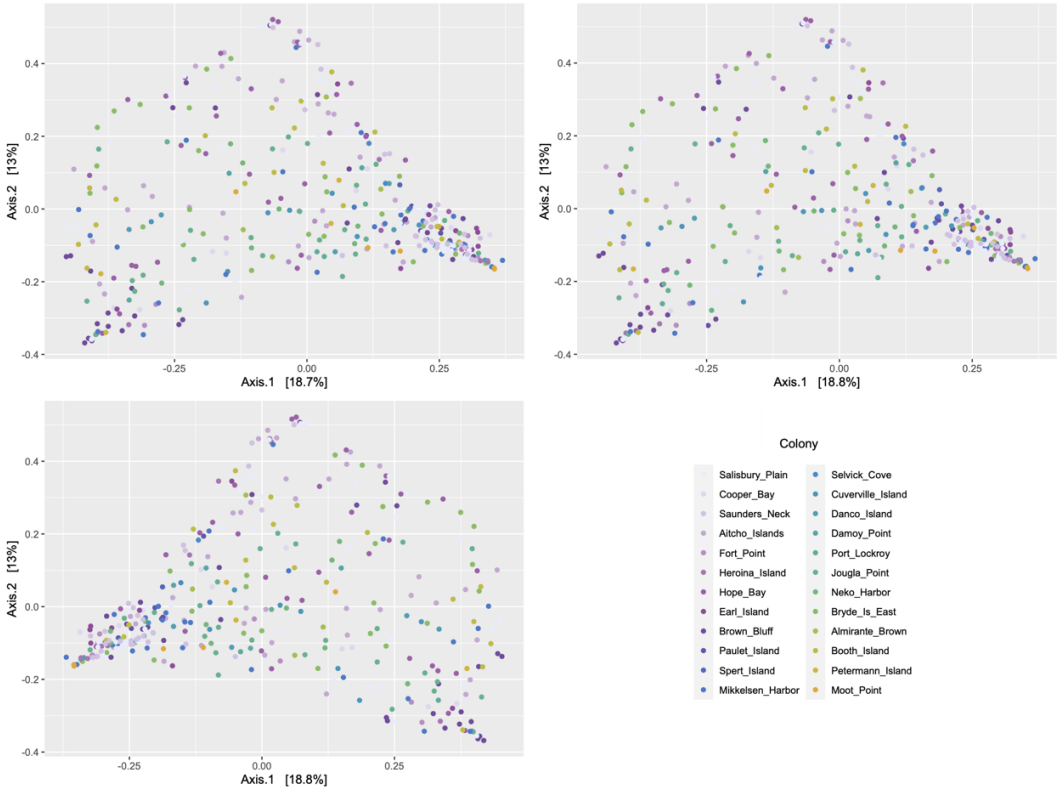
Supplemental Figure 3. A flowchart showing the steps in data analysis, including the various settings used in the processing. For filters, green data (bacteria identified to phylum and the percentage host and all diet information) was kept, while grey data indicates data that was filtered out from the dataset. Metrics also includes the diversity measures and tests used in our analyses.

Supplemental Figure 4. PCoA of 16S rRNA Gene Data



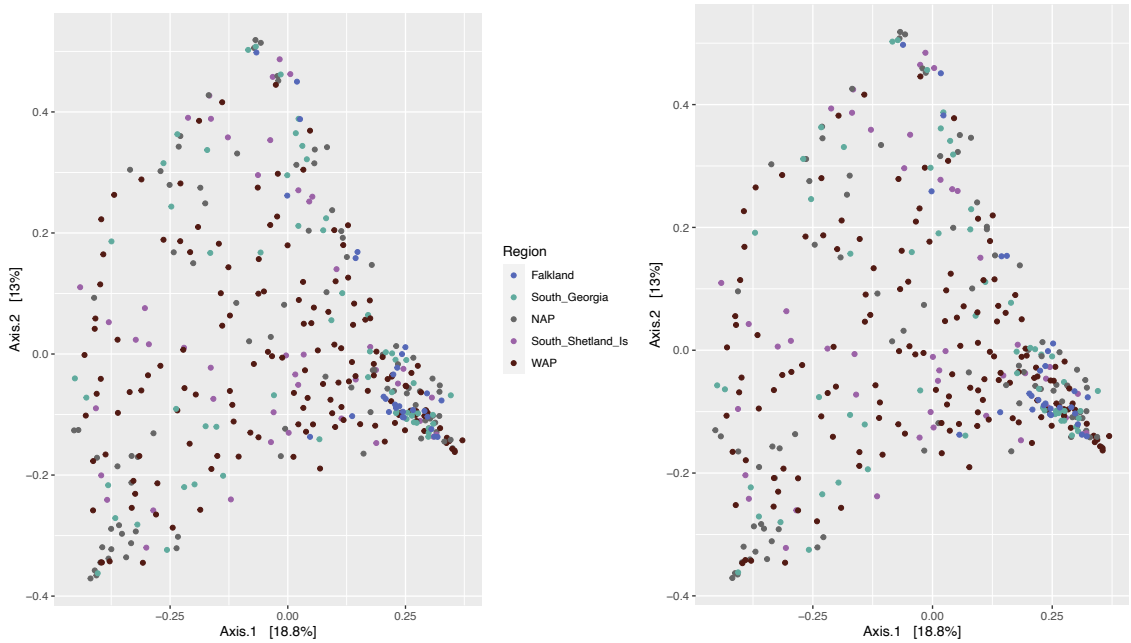
Supplemental Figure 4: PCoA of rarefied 16S rRNA gene data (to 5098 read depth without replacement) using Bray-Curtis, coloured by colony (A). The barplot insert (B) shows the cumulative explained variance for the first 10 axes. The colonies in the reference are arranged in latitudinal order. For other rarefication examples and regional colourings, see Supplemental Figures 5 and 6.

Supplemental Figure 5. Additional PCoAs of 16S rRNA Gene Data



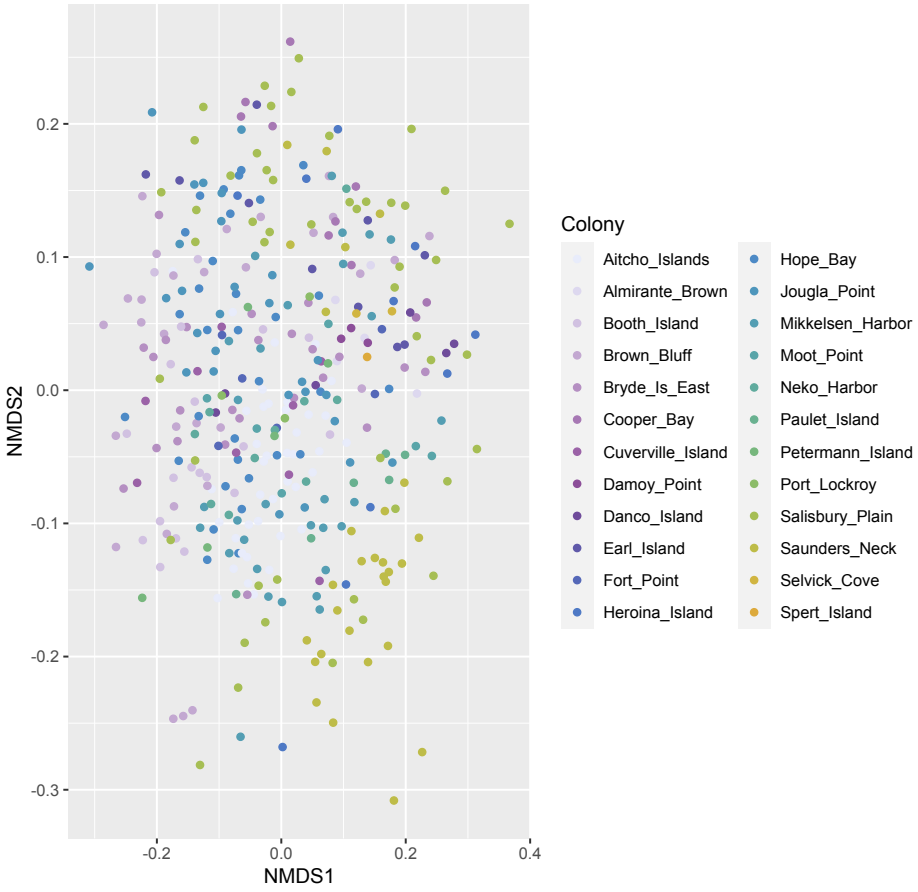
Supplemental Figure 5: 3 additional PCoAs of pseudo-randomly rarefied 16S rRNA gene data (to 5098 read depth without replacement) using Bray-Curtis, coloured by colony. The colonies in the reference are arranged in latitudinal order.

Supplemental Figure 6. Regional PCoAs of 16S rRNA Gene Data



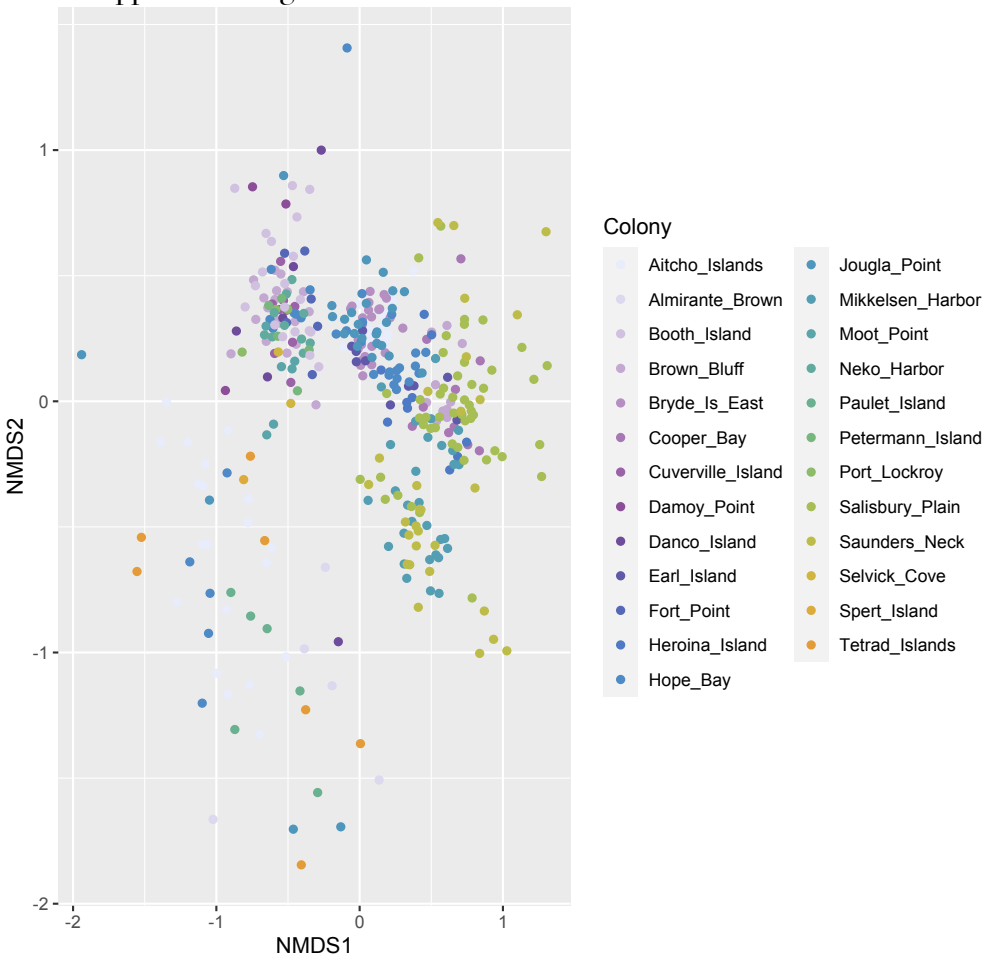
Supplemental Figure 6: PCoAs of pseudo-randomly rarefied 16S rRNA gene data (to 5098 read depth without replacement) using Bray-Curtis, coloured by region.

Supplemental Figure 7. NMDS Visualization of 16S rRNA Gene Data



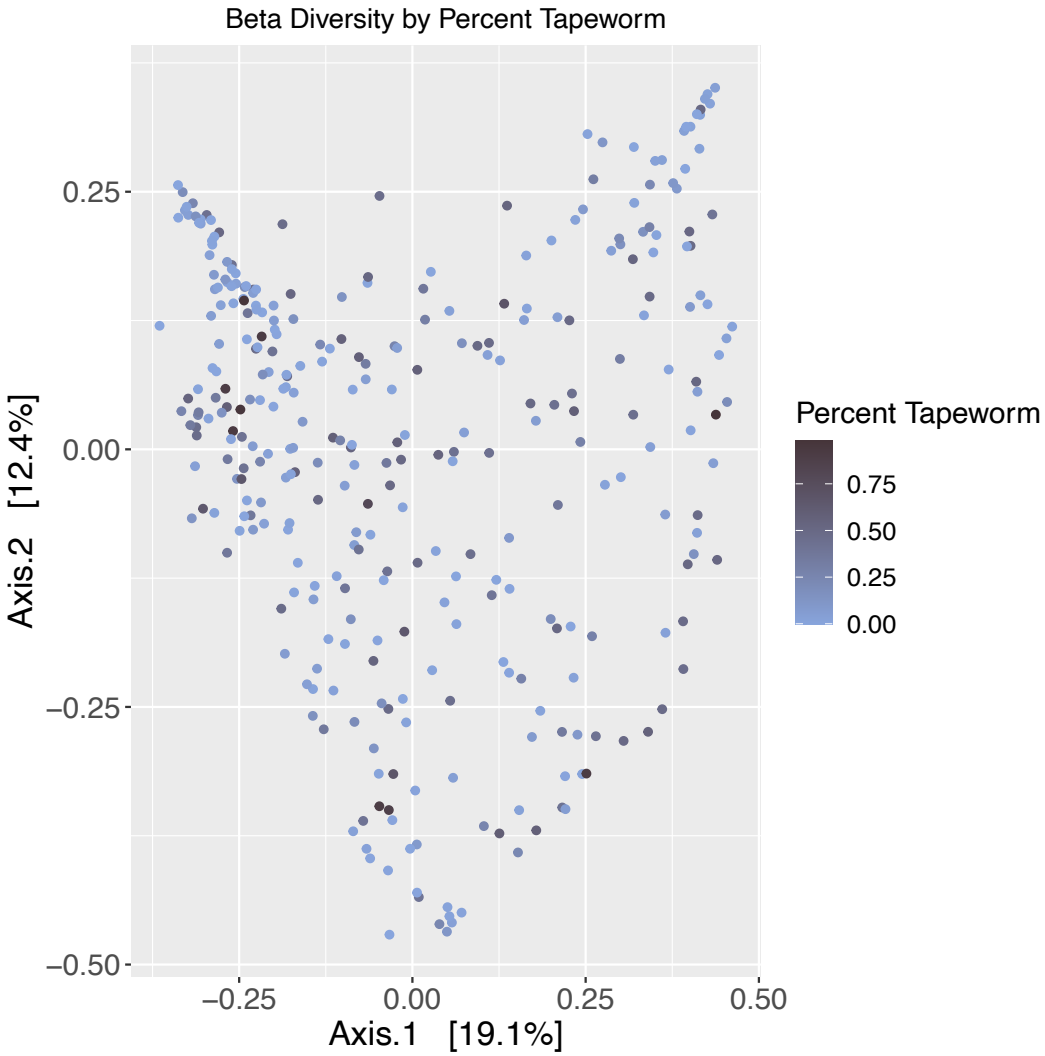
Supplemental Figure 7: NMDS of 16S rRNA gene data using Bray-Curtis, coloured by colony.

Supplemental Figure 8. NMDS Visualization of 18S rRNA Gene Data



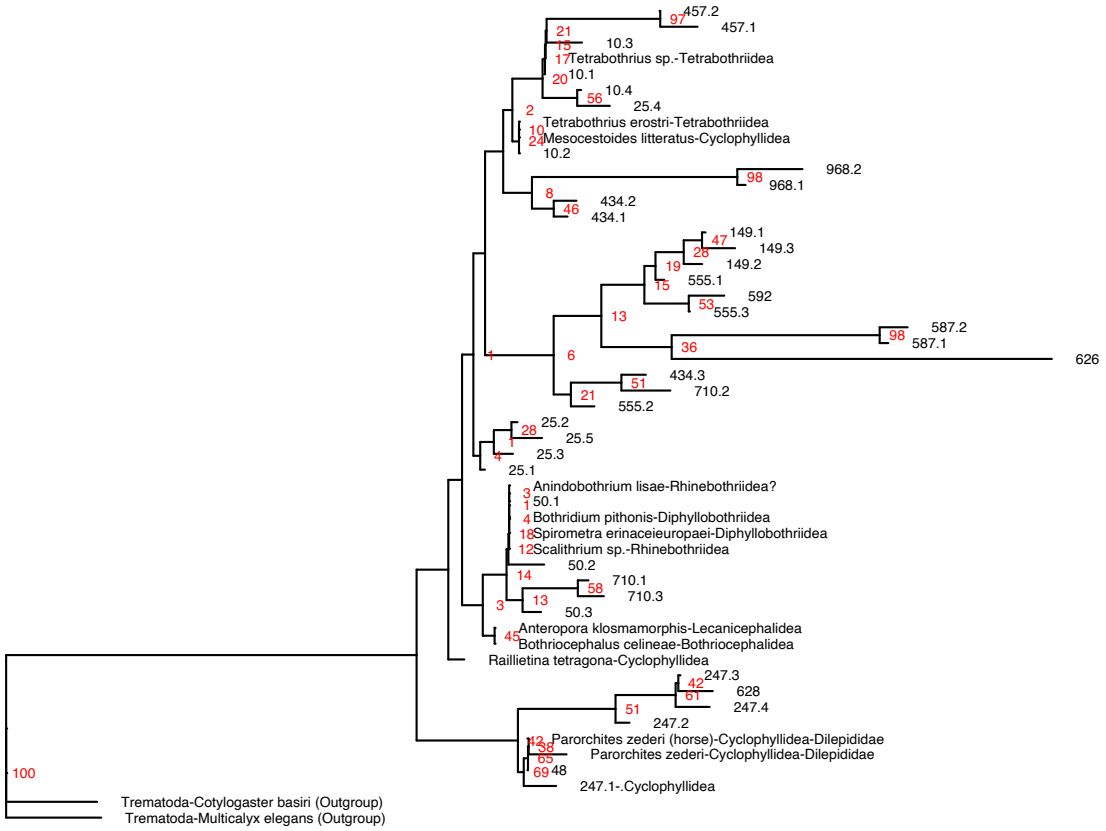
Supplemental Figure 8: NMDS of 18S rRNA gene data using Bray-Curtis, coloured by colony.

Appendix B
Supplemental Material for Chapter 3

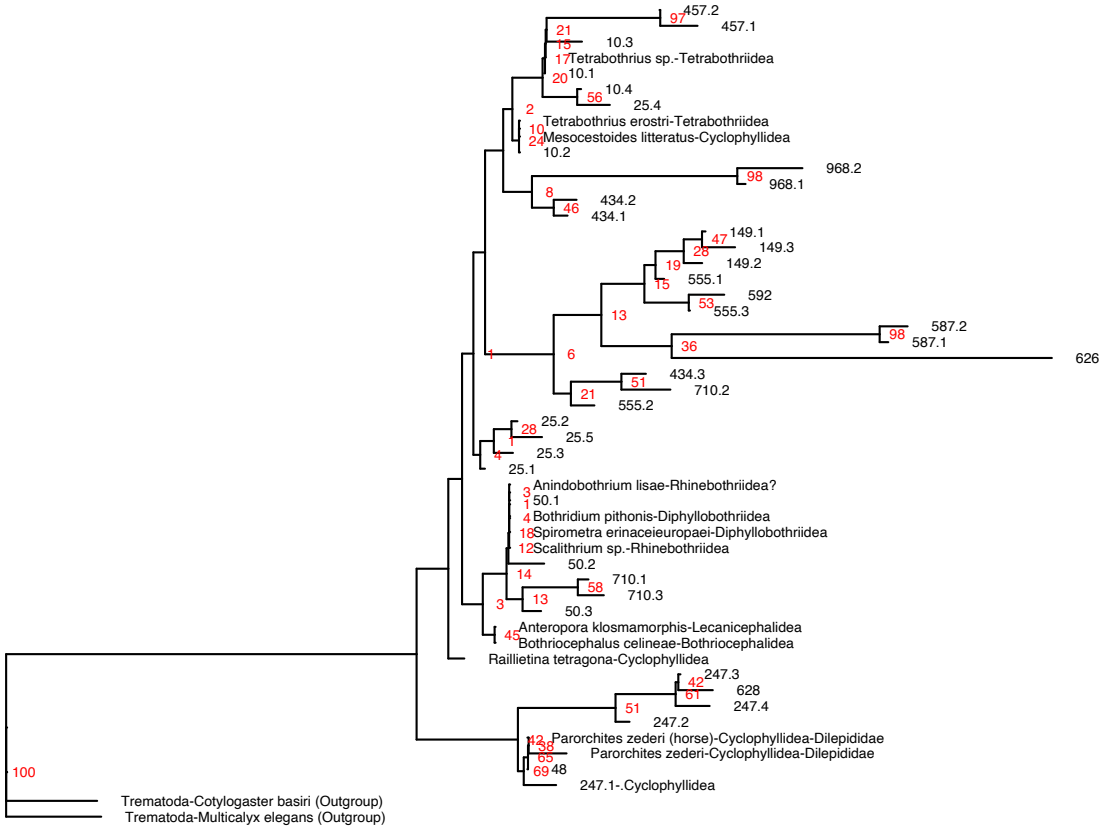


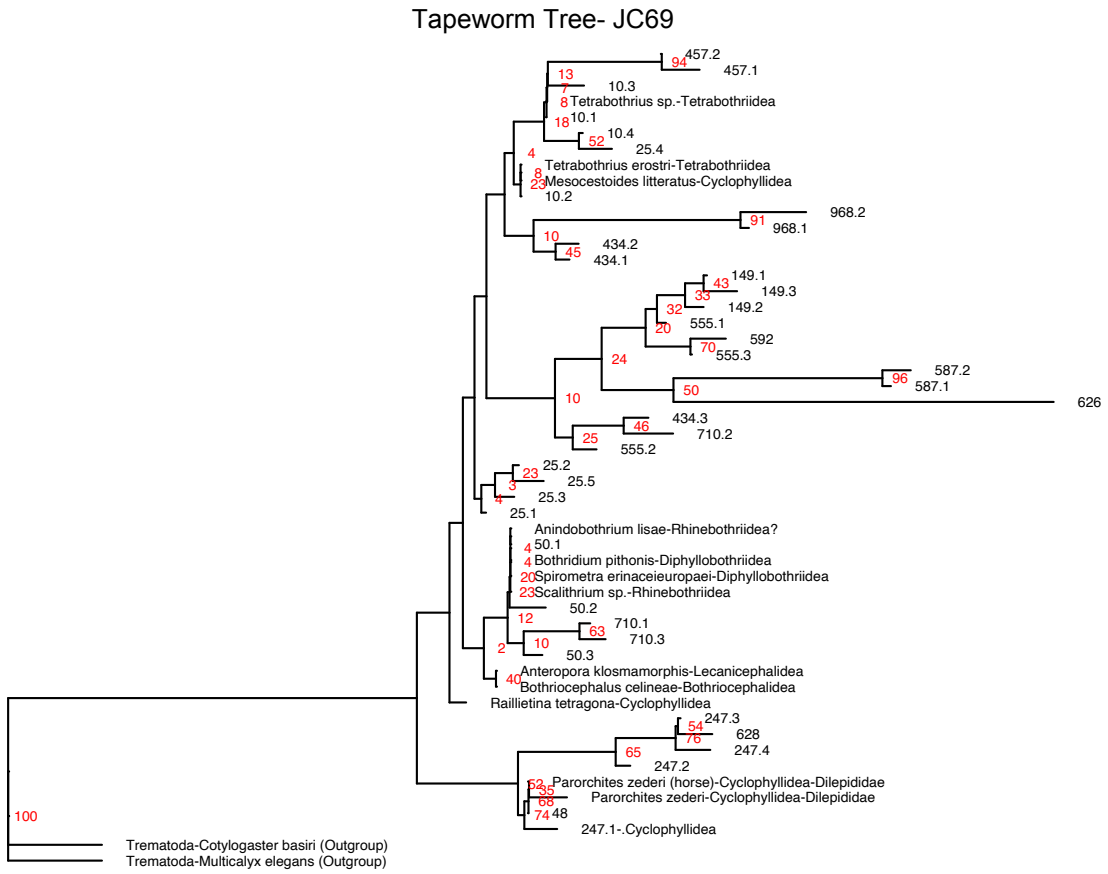
Supplemental Figure 1: PCoA of beta-diversity of 16S bacterial profile, rarefied to even depth and coloured according to the percent tapeworm (of total metazoan reads minus those identified as host). The first two axes of the PCoA explain 31.5% of the variance of the beta diversity of the 16S bacterial profile.

Tapeworm Tree- K80

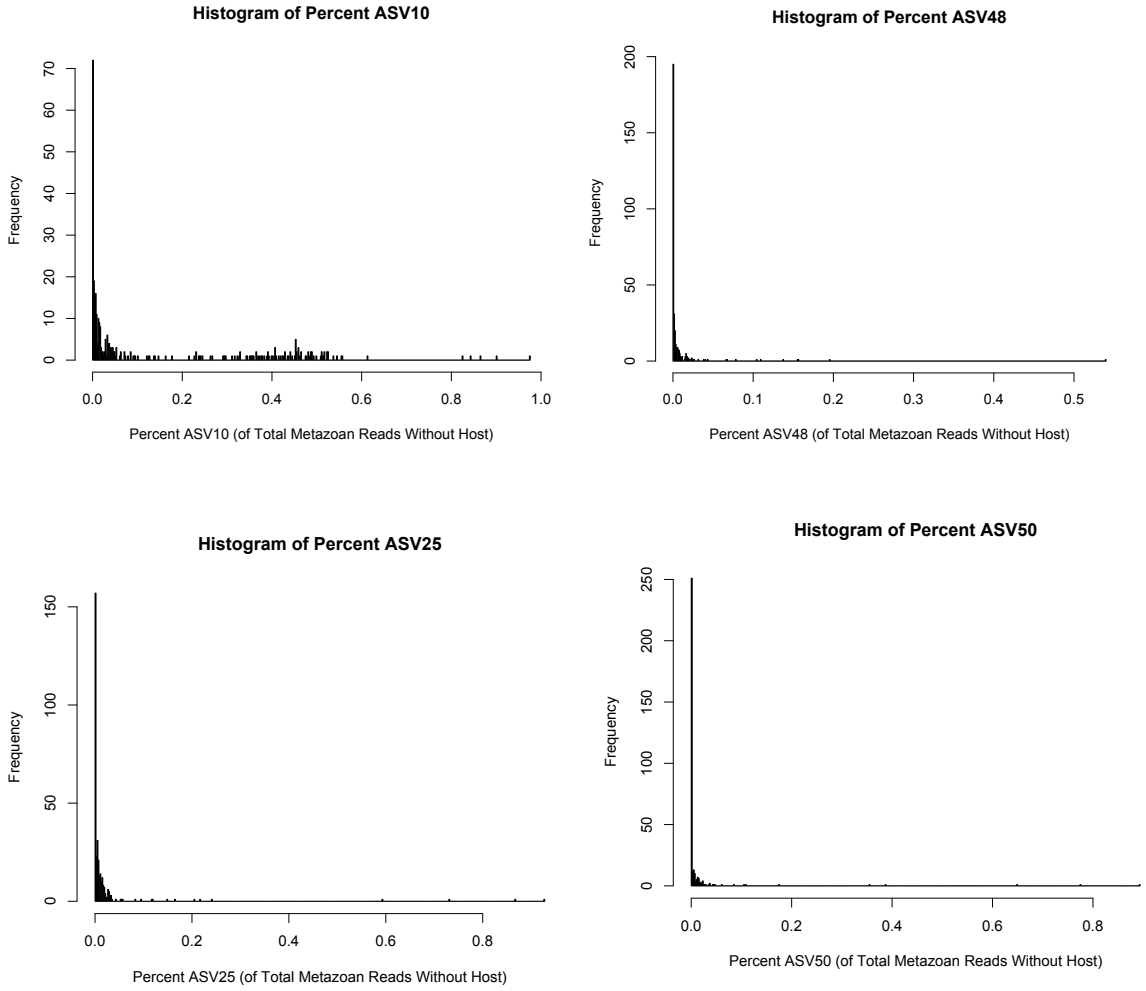


Tapeworm Tree- GG95

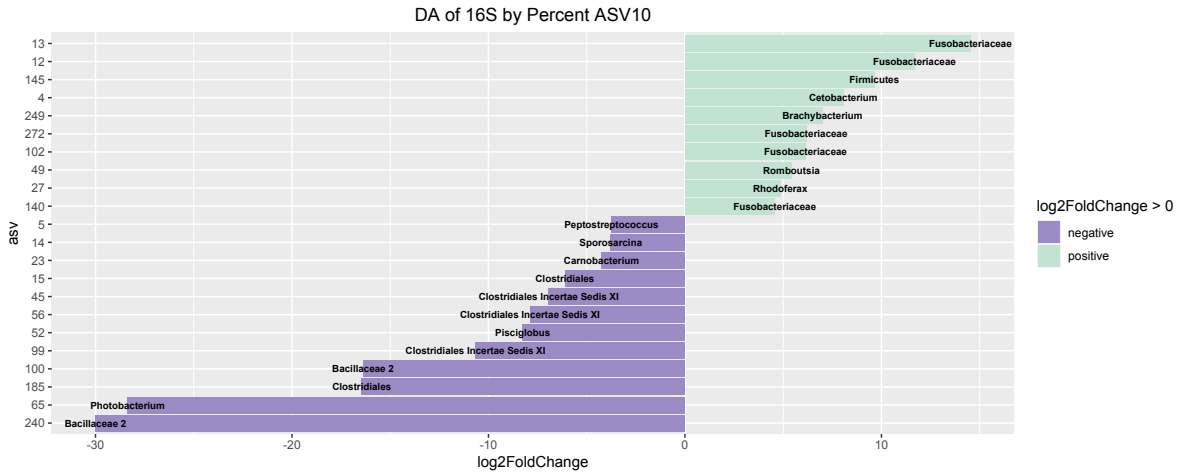




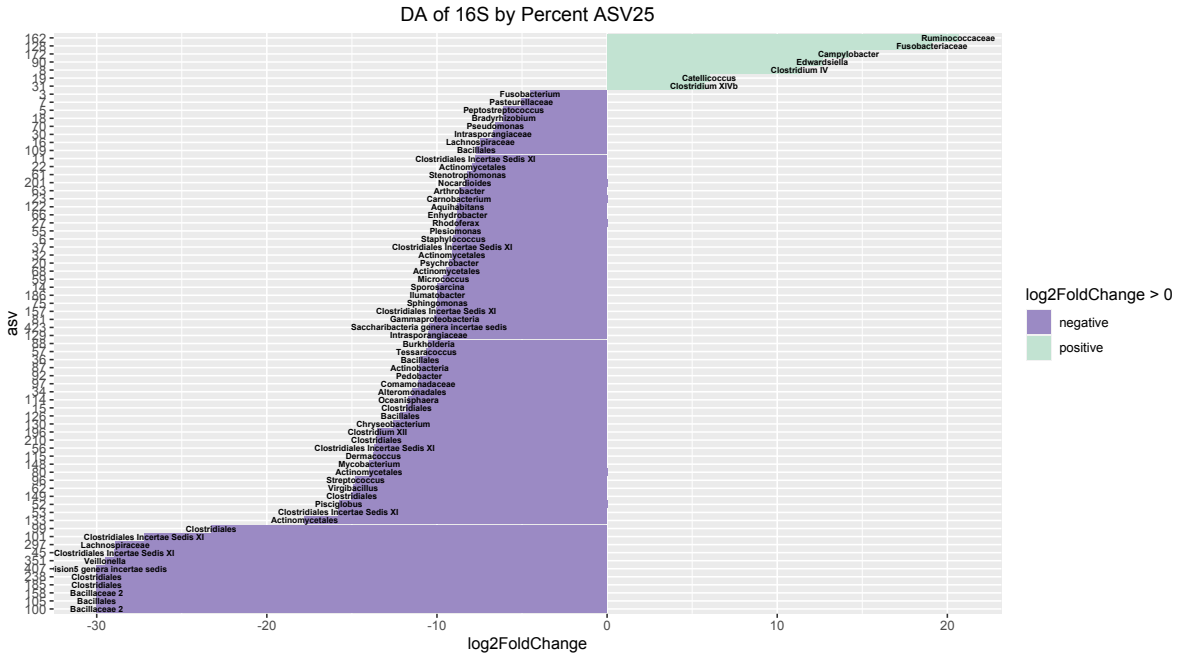
Supplemental Figures 2-4: Phylogenies of the detected tapeworm ASVs (numbered tips) with knowns from the NCBI BLAST database (named tips), using alternative algorithms (K80, GG95, and JC69 respectively). ASVs include all of the sequences grouped under that ASV and use decimal and integers to indicate which the number of the subsequence (e.g. 149.1 is the first subsequence of ASV 149). The tree was generated using TN93 nucleotide model and the NJS neighbour joining algorithm. Bootstrap values are shown in red.



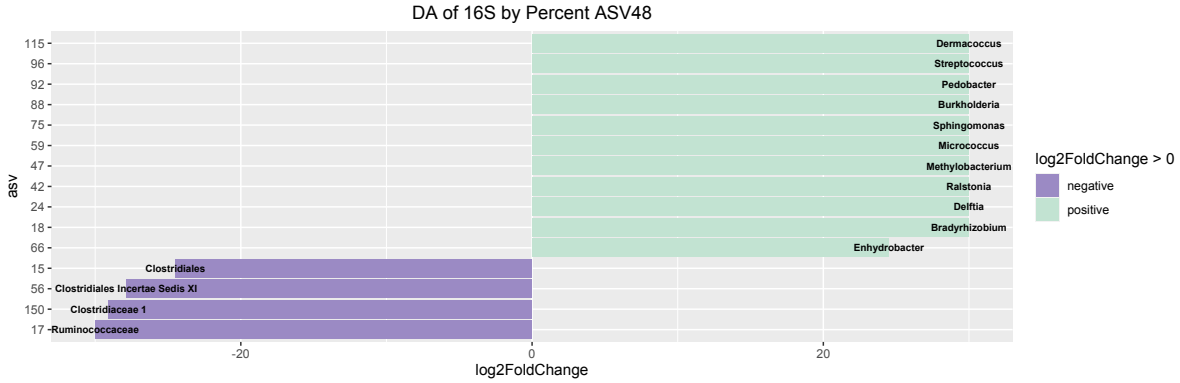
Supplemental Figure 5: Histograms looking at the proportion of each of the top four tapeworm ASVs in samples (of all metazoan reads identifiable to phyla minus those identified as host). Frequency is the total number of samples in each proportion with 100 breaks (0.01 per bar).



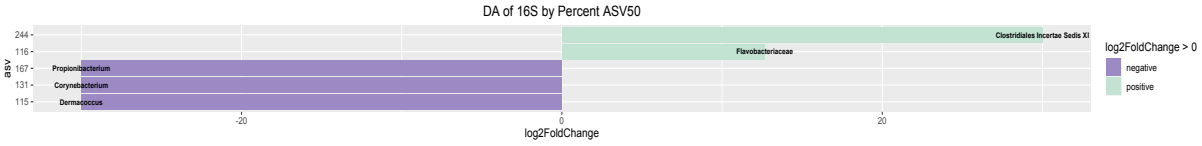
Supplemental Figure 6. The results of the differential abundance test in DESeq2, for the percent of tapeworm ASV10 in samples controlling for colony, showing all bacterial ASVs with log2fold changes to their lowest determined identity. Those in purple showed negative log fold changes and those in green showed positive log fold changes. The ASV number for each ASV is on the y-axis and the closest identified taxa is recorded on the bar. The length of the bar shows the magnitude of the log fold change over the model.



Supplemental Figure 7. The results of the differential abundance test in DESeq2, for the percent of tapeworm ASV25 in samples controlling for colony, showing all bacterial ASVs with log2fold changes to their lowest determined identity. Those in purple showed negative log fold changes and those in green showed positive log fold changes. The ASV number for each ASV is on the y-axis and the closest identified taxa is recorded on the bar. The length of the bar shows the magnitude of the log fold change over the model.



Supplemental Figure 8. The results of the differential abundance test in DESeq2, for the percent of tapeworm ASV48 in samples controlling for colony, showing all bacterial ASVs with log2fold changes to their lowest determined identity. Those in purple showed negative log fold changes and those in green showed positive log fold changes. The ASV number for each ASV is on the y-axis and the closest identified taxa is recorded on the bar. The length of the bar shows the magnitude of the log fold change over the model.



Supplemental Figure 9. The results of the differential abundance test in DESeq2, for the percent of tapeworm ASV50 in samples controlling for colony, showing all bacterial ASVs with log2fold changes to their lowest determined identity. Those in purple showed negative log fold changes and those in green showed positive log fold changes. The ASV number for each ASV is on the y-axis and the closest identified taxa is recorded on the bar. The length of the bar shows the magnitude of the log fold change over the model.

Supplemental Table 1: ASV10 BLASTN Results. The first 5 BLASTN query matches for each sequence contained in tapeworm ASV10.

ASV	Description	TaxID	Total Score	E-Value	Percent Identity
10.1	Tetrabothrius sp. DTJL-2000 18S rRNA gene	117258	318	3e-82	100.00
10.1	Mesocestoides sp. PH-2019 small subunit ribosomal RNA gene	2563108	313	1e-80	99.42
10.1	Mesocestoides sp. JH-2015 clone K21 18S ribosomal RNA gene	1708138	313	1e-80	99.42
10.1	Mesocestoides litteratus isolate Mlit3 18S ribosomal RNA gene	363290	313	1e-80	99.42
10.1	Mesocestoides litteratus isolate Mesme 18S ribosomal RNA gene	363290	313	1e-80	99.42
10.2	Mesocestoides sp. PH-2019 small subunit ribosomal RNA gene	2563108	318	3e-82	100.00
10.2	Mesocestoides sp. JH-2015 clone K21 18S ribosomal RNA gene	1708138	318	3e-82	100.00
10.2	Mesocestoides litteratus isolate Mlit3 18S ribosomal RNA gene	363290	318	3e-82	100.00
10.2	Mesocestoides litteratus isolate Mesme 18S ribosomal RNA gene	363290	318	3e-82	100.00
10.2	Mesocestoides litteratus isolate Mlit1 18S ribosomal RNA gene	363290	318	3e-82	100.00
10.3	Tetrabothrius sp. DTJL-2000 18S rRNA gene	17258	313	1e-80	99.42
10.3	Mesocestoides sp. PH-2019 small subunit ribosomal RNA gene	2563108	307	6e-79	98.84
10.3	Mesocestoides sp. JH-2015 clone K21 18S ribosomal RNA gene	1708138	307	6e-79	98.84
10.3	Mesocestoides litteratus isolate Mlit3 18S ribosomal RNA gene	363290	307	6e-79	98.84
10.3	Mesocestoides litteratus isolate Mesme 18S ribosomal RNA gene	363290	307	6e-79	98.84
10.4	Tetrabothrius sp. DTJL-2000 18S rRNA gene	117258	313	1e-80	99.42
10.4	Mesocestoides sp. PH-2019 small subunit ribosomal RNA gene	2563108	307	6e-79	98.84
10.4	Mesocestoides sp. JH-2015 clone K21 18S ribosomal RNA gene	1708138	307	6e-79	98.84
10.4	Mesocestoides litteratus isolate Mlit3 18S ribosomal RNA gene	363290	307	6e-79	98.84
10.4	Mesocestoides litteratus isolate Mesme 18S ribosomal RNA gene	363290	307	6e-79	98.84

Supplemental Table 2. The results of the differential abundance test in DESeq2, showing all ASVs with log2fold changes to their lowest determined identity.

ASV	baseMean	log2FoldChange	padj	Lowest Identified Taxa
4	2043.34223	6.37992521	2.9595425754027e-15	Cetobacterium
12	247.199032	7.24797352	3.99881019155885e-15	Fusobacteriaceae

13	192.222257	9.47149059	4.140855995460 65e-21	Fusobacteriaceae
15	275.350909	-5.1987042	0.00027777	Clostridiales
24	401.364013	-4.0327413	0.00616906	Delftia
27	276.13604	4.5951879	9.377586561183 2e-06	Rhodoferax
31	370.521395	2.99763847	0.00088451	Clostridium XlVb
45	39.8617623	-6.9932316	2.138113990469 32e-05	Clostridiales Incertae Sedis XI
47	153.162287	-4.5886089	0.00051097	Methylobacterium
49	338.259203	4.06091431	0.00693614	Romboutsia
52	76.9108965	-8.968248	2.182217935940 59e-07	Pisciglobus
56	41.0699818	-6.2118195	0.00057673	Clostridiales Incertae Sedis XI
58	76.0424832	-4.0541226	0.00128183	Acinetobacter
62	165.110925	-8.0948289	0.00051097	Virgibacillus
65	78.6774174	-13.89225	1.005081856593 6e-06	Photobacterium
90	55.9302904	9.80451001	2.354614578994 82e-18	Edwardsiella
99	14.5862593	-8.4051281	0.00118373	Clostridiales
100	62.4843355	-25.678106	2.157040343838 67e-18	Bacillaceae 2
102	8.78233421	5.52666398	0.00027688	Fusobacteriaceae
107	20.1442957	6.06200448	0.00358979	Fusobacteriaceae
112	3.0024217	6.10986326	0.00051142	Fusobacteriaceae
117	2.28369942	6.18688645	0.00116524	Fusobacteriaceae
145	9.08134764	6.16611315	4.082765424759 03e-06	Firmicutes
172	7.37753293	10.3734303	4.686618691225 56e-10	Campylobacter
185	13.3275347	-8.553663	0.00064014	Clostridiales
240	32.916479	-30	1.476786286920 25e-08	Bacillaceae 2
272	3.4170832	5.33272305	0.00026279	Fusobacteriaceae

Supplemental Table 3. The results of the differential abundance test in DESeq2, showing for ASV10 with log2fold changes to their lowest determined identity.

ASV	baseMean	log2FoldChange	padj	Lowest Identified Taxa
4	2043.34223	8.06	3.79	Cetobacterium
5	718.678455	-3.74	0.00047596	Peptostreptococcus
12	247.199032	11.68	1.14e-23	Fusobacteriaceae
13	192.222257	14.53	9.36e-31	Fusobacteriaceae
14	872.528883	-3.7	0.0079	Sporosarcina
15	275.350909	-6.11	0.0031	Clostridiales
23	249.341219	-4.25	0.0063	Carnobacterium
27	276.13604	4.85	0.0017	Rhodoferax

45	39.8617623	-6.93	0.0049	Clostridiales Incertae Sedis XI
49	338.259203	5.42	0.0058	Romboutsia
52	76.9108965	-8.25	0.0019	Pisciglobus
56	41.0699818	-7.89	0.0020	Clostridiales Incertae Sedis XI
65	78.6774174	-28.36	4.15e-16	Photobacterium
99	14.5862593	-10.66	0.00324759	Clostridiales Incertae Sedis XI
100	62.4843355	-16.38	0.00037986	Bacillaceae 2
102	8.78233421	6.14	0.0052421	Fusobacteriaceae
140	4.47493257	4.57	0.00484106	Fusobacteriaceae
145	9.08134764	9.64	3.092e-09	Firmicutes
185	13.3275347	-16.44	2.057e-07	Clostridiales
240	32.916479	-30	9.67e-05	Bacillaceae 2
249	10.5037844	7.00	0.00246201	Brachybacterium
272	3.4170832	6.16475156	0.00221509	Fusobacteriaceae

Supplemental Table 4. The results of the differential abundance test in DESeq2, showing for ASV25 with log2fold changes to their lowest determined identity.

ASV	baseMean	log2FoldChange	padj	Lowest Identified Taxa
3	1540.439	-4.5231388	0.00389086	Fusobacterium
5	718.678455	-6.0695027	0.00013005	Peptostreptococcus
6	2539.802	-9.0339236	3.39102584450932e-05	Staphylococcus
7	218.064711	-5.0345522	0.00477439	Pasteurellaceae
8	364.539256	11.3106589	2.2740646355244e-07	Clostridium IV
11	428.210025	-7.7352644	0.0003424	Clostridiales Incertae Sedis XI
14	872.528883	-9.9583478	2.2740646355244e-07	Sporosarcina
15	275.350909	-11.779335	0.00014682	Clostridiales
16	150.198784	-7.4433113	9.98215159753592e-05	Lachnospiraceae
18	277.745965	-6.0714255	0.00525945	Bradyrhizobium
19	72.453687	5.99966075	0.00409151	Catellicoccus
20	1512.15959	-9.2697126	4.12540851599712e-06	Psychrobacter
22	72.2951522	-7.8943344	0.00037784	Actinomycetales
23	249.341219	-8.7065186	5.7891156955503e-05	Carnobacterium
27	276.13604	-8.8838585	0.00013805	Rhodoferax
30	61.2993584	-6.6958924	0.0083948	Intrasporangiaceae
31	370.521395	5.7030863	0.0018727	Clostridium XlVb
32	41.401694	-9.1092661	0.00075575	Actinomycetales
34	187.394292	-11.443587	4.35729047508976e-05	Alteromonadales
36	148.304387	-10.985592	0.00024012	Bacillales
37	158.601709	-9.1021939	0.00706404	Clostridiales Incertae Sedis XI

45	39.6050288	-28.971546	4.19311681339884e-19	Clostridiales Incertae Sedis XI
52	76.9108965	-15.761064	3.39102584450932e-05	Pisciglobus
53	46.2875345	-15.795997	1.27776729996992e-08	Clostridiales Incertae Sedis XI
55	21.452019	-8.9000262	0.00327525	Plesiomonas
56	41.0699818	-13.65695	0.00021358	Clostridiales Incertae Sedis XI
57	33.9265468	-10.611171	2.30439071034858e-05	Tessaracoccus
59	110.365078	-9.6042151	0.00255061	Micrococcus
61	87.7829651	-8.1978579	0.00335968	Stenotrophomonas
62	165.110925	-14.92268	0.00189767	Virgibacillus
63	117.617244	-8.6814518	7.71928317477809e-05	Arthrobacter
66	55.617233	-8.7811441	0.00249553	Enhydrobacter
68	19.6106747	-9.4347371	0.00116792	Actinomycetales
70	263.906598	-6.561848	0.00233162	Pseudomonas
75	94.6125543	-9.9738749	0.00370906	Sphingomonas
80	61.4777852	-13.998833	0.00024435	Actinomycetales
81	42.2482981	-10.127363	0.0063864	Gammaproteobacteria
87	18.4114813	-11.020251	0.00010729	Actinobacteria
88	40.9994086	-10.47726	0.00378972	Burkholderia
90	55.9302904	12.652619	2.2740646355244e-07	Edwardsiella
92	107.42164	-11.041464	0.00102211	Pedobacter
96	79.1723323	-14.779263	0.00013005	Streptococcus
97	48.4643389	-11.103584	0.00011415	Comamonadaceae
99	14.5862593	-23.271344	4.3454805037273e-06	Clostridiales
100	62.4843355	-30	8.77840657621428e-06	Bacillaceae 2
101	18.4356167	-27.214088	1.59319207892437e-08	Clostridiales Incertae Sedis XI
105	64.1663474	-30	2.69681081959901e-15	Bacillales
109	7.78113109	-7.6787986	0.00378972	Bacillales
114	232.907193	-11.66866	3.29926032319055e-07	Oceanisphaera
115	21.7977113	-13.734476	0.00093851	Dermacoccus
122	46.7827505	-8.7792015	0.00109519	Aquihabitans
126	13.0804003	-12.150602	0.00134899	Bacillales
128	6.27906609	19.0943483	4.1964680946058e-09	Fusobacteriaceae
129	9.73991196	-10.476387	0.00015879	Intrasporangiaceae
130	33.1983097	-12.555128	2.71279691047842e-06	Chryseobacterium
133	8.19147159	-17.778182	0.00024435	Actinomycetales
148	20.5858078	-13.978961	0.00142058	Mycobacterium
149	9.88038843	-15.007831	0.00049691	Clostridiales

157	17.2858698	-10.047698	0.00409151	Clostridiales Incertae Sedis XI
158	14.243774	-30	2.93667728109616e-16	Bacillaceae 2
162	12.5343529	20.6451768	4.01381404047771e-07	Ruminococcaceae
172	7.37753293	14.1952434	5.77210008448697e-05	Campylobacter
185	13.3275347	-30	5.19264335459962e-10	Clostridiales
186	14.4994973	-9.9648813	0.00039541	Ilumatobacter
196	13.2322206	-13.498255	0.00143256	Clostridium XII
201	7.75712308	-8.3550095	0.0083948	Nocardioides
210	19.268984	-13.548806	0.00356266	Clostridiales
238	23.6667576	-30	2.01395773206702e-07	Clostridiales
297	6.14134878	-28.906908	0.00061385	Lachnospiraceae
351	4.77321269	-29.507047	0.00356266	Veillonella
407	16.9920094	-30	1.79307982720399e-06	Subdivision5 genera incertae sedis
423	14.8456137	-10.458549	0.00080556	Saccharibacteria genera incertae sedis

Supplemental Table 5. The results of the differential abundance test in DESeq2, showing for ASV48 with log2fold changes to their lowest determined identity.

ASV	baseMean	log2FoldChange	padj	Genus
15	272.188114	-24.485986	0.0033188	Clostridiales
17	133.343213	-30	0.00049596	Ruminococcaceae
18	277.745965	30	3.89785296605815e-09	Bradyrhizobium
24	401.364013	30	6.82639885191452e-06	Delftia
42	113.827444	30	7.51501640784e-07	Ralstonia
47	122.387242	30	1.60372933358e-07	Methylobacterium
56	41.0699818	-27.87368	0.00952107	Clostridiales Incertae Sedis XI
59	97.0269583	30	0.00015735	Micrococcus
66	55.617233	24.5073259	0.00100039	Enhydrobacter
75	94.6125543	30	0.00047635	Sphingomonas
88	40.9994086	30	0.00065897	Burkholderia
92	107.42164	30	0.00047635	Pedobacter
96	66.2055905	30	0.00212415	Streptococcus
115	14.6963444	30	0.00229862	Dermacoccus
150	10.2461224	-29.072213	0.00100039	Clostridiaceae 1

Supplemental Table 6. The results of the differential abundance test in DESeq2, showing for ASV50 with log2fold changes to their lowest determined identity.

ASV	baseMean	log2FoldChange	padj	Genus
115	21.7977113	-30	8.89212594259412e-10	Dermacoccus
116	39.1360753	12.6593307	0.00205952	Flavobacteriaceae
131	45.801994	-30	5.12355318542311e-21	Corynebacterium
167	29.6304817	-30	3.22127454426192e-09	Propionibacterium
244	2.20497752	30	0.0005454	Clostridiales Incertae Sedis XI

Appendix C
Supplemental Material for Chapter 4

Supplemental Table 1: Average Bray-Curtis beta diversity values comparing all pairs of colonies with coloured grouping for sets of colonies that have a Bray Curtis value of <0.5. The partial groups (where some but not all colonies in a grouping had a Bray-Curtis value of <0.5 for a respective colony) are indicated in paler colour than the main group.

	Saunders_Neck	Salisbury_Plain	Cooper_Bay	Aitcho_Islands	Fort_Point	Heroina_Island	Hope_Bay	Earl_Island
Saunders_Neck	0	0.69690113	0.59245976	0.66222067	0.86468621	0.77280565	0.69640511	0.72866667
Salisbury_Plain	0.69690113	0	0.67858673	0.5787516	0.88482501	0.78511995	0.45572613	0.83465589
Cooper_Bay	0.59245976	0.67858673	0	0.66325112	0.69413566	0.53895256	0.60269719	0.57080623
Aitcho_Islands	0.66222067	0.5787516	0.66325112	0	0.77436014	0.70146396	0.52019448	0.71662698
Fort_Point	0.86468621	0.88482501	0.69413566	0.77436014	0	0.61574252	0.83211432	0.65344042
Heroina_Island	0.77280565	0.78511995	0.53895256	0.70146396	0.61574252	0	0.7707335	0.71478648
Hope_Bay	0.69640511	0.45572613	0.60269719	0.52019448	0.83211432	0.7707335	0	0.5444061
Earl_Island	0.72866667	0.83465589	0.57080623	0.71662698	0.65344042	0.71478648	0.5444061	0
Brown_Bluff	0.70640584	0.50558504	0.62755081	0.51917911	0.80079637	0.72344334	0.51656668	0.61896363
Spert_Island	0.82891949	0.91459245	0.76475553	0.83434622	0.68022031	0.80648328	0.86029974	0.77394064
Mikkelsen_Harbor	0.63850769	0.7509778	0.57576513	0.44215888	0.81675514	0.65990215	0.44822229	0.58624487
Selvick_Cove	0.84125568	0.90419747	0.66957995	0.81596319	0.60929646	0.81627012	0.84382029	0.73468747
Cuverville_Island	0.77126419	0.81719378	0.53971963	0.64381892	0.44304005	0.56725527	0.71531373	0.54794343
Danco_Island	0.84315187	0.88364397	0.67406605	0.77955882	0.31834458	0.61479179	0.81515337	0.61695599
Port_Lockroy	0.86224983	0.88812558	0.69050895	0.76455971	0.32869981	0.63732761	0.80316527	0.61187906
Jougla_Point	0.70236841	0.54047506	0.61675733	0.37731845	0.80673199	0.76245599	0.45427162	0.62217807
Neko_Harbor	0.67158408	0.77212401	0.46159053	0.62391614	0.50418837	0.5878728	0.65443482	0.4932172
Bryde_Is_East	0.70691945	0.49459567	0.6460059	0.37477092	0.80234816	0.70920818	0.30605523	0.60825403
Almirante_Brown	0.83826625	0.87942485	0.76485038	0.7395454	0.55594804	0.73863209	0.81776893	0.65518954
Booth_Island	0.72918078	0.70902445	0.60414924	0.47280127	0.68769102	0.67108849	0.45999479	0.50128661
Petermann_Island	0.83668283	0.88679645	0.6455866	0.78483672	0.34826428	0.62136437	0.81547224	0.65129177
Moot_Point	0.70920167	0.83991574	0.62431304	0.76485722	0.58745888	0.54085933	0.81152136	0.69743021

	Brown_Bluff	Spert_Island	Mikkelsen_Harbor	Selvick_Cove	Cuverville_Island	Danco_Island	Port_Lockroy	Jougla_Point
Saunders_Neck	0.70640584	0.82891949	0.63850769	0.84125568	0.77126419	0.84315187	0.86224983	0.70236841
Salisbury_Plain	0.50558504	0.91459245	0.57509778	0.90419747	0.81719378	0.88364397	0.88812558	0.54047506
Cooper_Bay	0.62755081	0.76475553	0.57576513	0.66957995	0.53971963	0.67406605	0.69050895	0.61675733
Aitcho_Islands	0.51917911	0.83434622	0.44215888	0.81596319	0.64381892	0.77955882	0.76455971	0.37731845
Fort_Point	0.80079637	0.68022031	0.81675514	0.60929646	0.44304005	0.31834458	0.32869981	0.80673199
Heroina_Island	0.72344334	0.80648328	0.65990215	0.81627012	0.56725527	0.61479179	0.63732761	0.76245599
Hope_Bay	0.51656668	0.86029974	0.44822229	0.84382029	0.71531373	0.81515337	0.80316527	0.45427162
Earl_Island	0.61896363	0.77394064	0.58624487	0.73468747	0.54794343	0.61695599	0.61187906	0.62217807
Brown_Bluff	0	0.87248492	0.47596486	0.8464582	0.70093192	0.77128079	0.76399941	0.39014847
Spert_Island	0.87248492	0	0.86826865	0.39509968	0.62674592	0.63414115	0.69879332	0.84424579
Mikkelsen_Harbor	0.47596486	0.86826865	0	0.83358596	0.66211603	0.8043658	0.7909239	0.41029297
Selvick_Cove	0.8464582	0.39509968	0.83358596	0	0.47563163	0.50526176	0.64925295	0.8252297
Cuverville_Island	0.70093192	0.62674592	0.66211603	0.47563163	0	0.42645241	0.4190324	0.68642184
Danco_Island	0.77128079	0.63414115	0.8043658	0.50526176	0.42645241	0	0.2756847	0.79447289
Port_Lockroy	0.76399941	0.69879332	0.7909239	0.64925295	0.4190324	0.2756847	0	0.78560546
Jougla_Point	0.39014847	0.84424579	0.41029297	0.8252297	0.68642184	0.79447289	0.78560546	0
Neko_Harbor	0.61765135	0.62700583	0.66186466	0.59605187	0.32161405	0.4818561	0.49811315	0.64325775
Bryde_Is_East	0.47294995	0.8450803	0.36010269	0.81907301	0.65288315	0.7866002	0.77511975	0.34777127
Almirante_Brown	0.78977396	0.65883675	0.80300554	0.58220655	0.58680846	0.49405041	0.51592115	0.80266334
Booth_Island	0.46757779	0.78085356	0.43019889	0.74755783	0.56488634	0.62873993	0.60285195	0.43987891
Petermann_Island	0.78421903	0.59451536	0.81317403	0.5522956	0.41024299	0.27771939	0.34812114	0.7991994
Moot_Point	0.80103646	0.71462092	0.74783956	0.67178308	0.55907432	0.54005283	0.60992078	0.81890324

	Neko_Harbor	Bryde_Is_East	Almirante_Brown	Booth_Island	Petermann_Island	Moot_Point
Saunders_Neck	0.67158408	0.70691945	0.83826625	0.72918078	0.83668283	0.70920167
Salisbury_Plain	0.77212401	0.49459567	0.87942485	0.70902445	0.88679645	0.83991574
Cooper_Bay	0.46159053	0.6460059	0.76485038	0.60414924	0.6455866	0.62431304
Aitcho_Islands	0.62391614	0.37477092	0.7395454	0.47280127	0.78483672	0.76485722
Fort_Point	0.50418837	0.80234816	0.55594804	0.68769102	0.34826428	0.58745888
Heroina_Island	0.5878728	0.70920818	0.73863209	0.67108849	0.62136437	0.54085933
Hope_Bay	0.65443482	0.30605523	0.81776893	0.45999479	0.81547224	0.81152136
Earl_Island	0.4932172	0.60825403	0.65518954	0.50128661	0.65129177	0.69743021
Brown_Bluff	0.61765135	0.47294995	0.78977396	0.46757779	0.78421903	0.80103646
Spert_Island	0.62700583	0.8450803	0.65883675	0.78085356	0.59451536	0.71462092
Mikkelsen_Harbor	0.66186466	0.36010269	0.80300554	0.43019889	0.81317403	0.74783956
Selvick_Cove	0.59605187	0.81907301	0.58220655	0.74755783	0.5522956	0.67178308
Cuverville_Island	0.32161405	0.65288315	0.58680846	0.56488634	0.41024299	0.55907432
Danco_Island	0.4818561	0.7866002	0.49405041	0.62873993	0.27771939	0.54005283
Port_Lockroy	0.49811315	0.77511975	0.51592115	0.60285195	0.34812114	0.60992078
Jougla_Point	0.64325775	0.34777127	0.80266334	0.43987891	0.7991994	0.81890324
Neko_Harbor	0	0.6210695	0.61134984	0.56837151	0.46740508	0.58420826
Bryde_Is_East	0.6210695	0	0.77259677	0.39567338	0.78797513	0.77492424
Almirante_Brown	0.61134984	0.77259677	0	0.72459835	0.60875185	0.47945103
Booth_Island	0.56837151	0.39567338	0.72459835	0	0.64428308	0.75248455
Petermann_Island	0.46740508	0.78797513	0.60875185	0.64428308	0	0.57618863
Moot_Point	0.58420826	0.77492424	0.47945103	0.75248455	0.57618863	0