Spatiotemporal Variability of Bioaerosols

Katie Amy Elisabeth Miaow

A thesis submitted to Auckland University of Technology in fulfilment of the requirements for the degree of Doctor of Philosophy (PhD) in Biological Sciences.

2021

School of Science

I. Abstract

The aim of this thesis was to understand variation in microbial aerosol (or bioaerosol) communities at differing spatial and temporal scales. Bioaerosols in urban parks were shown to vary as a result of location, sampled air-mass source and, for fungi in particular, time. Modelling was able to explain 38% of the fungal variation and 19% of the bacterial variation. Urban sampling over two years confirmed that bioaerosol communities varied over time in a non-linear fashion, exhibiting marked seasonality, which was especially pronounced for fungi. Non-linear diel variation was detected for Antarctic fungi. Bacteria in common between Antarctica and New Zealand increased markedly when New Zealand air was coming from Antarctica, suggesting intercontinental transport for bacteria at notable rates. Fungi appeared to undergo much less long-range atmospheric transport. This thesis research contributes innovative, validated data collection and processing pipelines for sparse microbial community data to our body of information. Novel patterns in bioaerosol spatiotemporal variation have been revealed that lead to new questions about bioaerosol community structure and ecosystem connectivity via bioaerosols. As understanding of the drivers of bioaerosol variation improves, predictions can be made regarding future ecosystem changes and spread of infectious microorganisms. This will be crucial for managing the impacts of these increasingly likely events in the face of climate change.

Table of Contents

II. List of Figures

III. List of Tables

IV. Attestation of Authorship

"I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person (except where explicitly defined in the acknowledgements), nor material which to a substantial extent has been submitted for the award of any other degree or diploma of a university or other institution of higher learning."

V. Co-authored Works

Chapter 1 K. King-Miaow, K. Lee, T. Maki, D. LaCap-Bugler & S.D.J. Archer. (2019). "Airborne Microorganisms in Antarctica: Transport, Survival and Establishment". *The Ecological Role of Micro-organisms in the Antarctic Environment,* 163-196.	Miaow 80%, Archer 10%, Lee 5%, LaCap-Bugler 3%, Maki 2%
Appendix C K. Miaow, D. LaCap-Bugler & H.L. Buckley. (2021). "Identifying optimal bioinformatics protocols for aerosol microbial community data". *Submitted to PeerJ as at 26 Jan 2021.*	Miaow 80%, Buckley 15%, LaCap-Bugler 5%

Katie Miaow

Associate Professor Hannah Buckley

Dr Kevin Lee

Professor Teruya Maki

Associate Professor Donnabella LaCap-Bugler

Dr Stephen Archer

VI. Dedication

To Murphy, you taught me to always get back on, to never give up. I would not be here without those lessons, and I will always treasure them, no matter how unpleasant they were at the time!

VII. Acknowledgements

Hannah, thank you for taking on this problem child and investing endless energy and enthusiasm into helping me science, I am so so grateful for your supervision.

Donna, thank you for being there through the rough and smooth, and for your wise and timely insights.

To my former research group, thank you for starting me off, and to Tim, Colleen and Kevin for ongoing assistance. To the many others at AUT and Auckland University who have helped me, thank you!

K020 (especially Craig and Charlie) and Antarctica New Zealand, thank you for taking me to Antarctica and keeping me alive, an experience I will treasure forever. Particular thanks to those who helped me collect samples at antisocial times of day!

To my family, thank you for leading by example and (mostly) encouraging me during the PhD process! To Rohan, I wouldn't have started this without you, to Billy, who I wouldn't be finishing without and to my friends who have supported me through my stress and patiently waited for me on the other side, I am so grateful.

To anyone else who I have omitted in time-pressured error, apologies and thanks!

Chapter 1 - General Introduction and Literature Review

This chapter is a modified version of:

K. King-Miaow, K. Lee, T. Maki, D. LaCap-Bugler & S.D.J. Archer. (2019). "Airborne Microorganisms in Antarctica: Transport, Survival and Establishment". *The Ecological Role of Micro-organisms in the Antarctic Environment*, 163-196.

1.1 Abstract

Microorganisms are a globally ubiquitous component of the atmosphere, of vital importance to climate, human health, and environmental processes. Bioaerosols (which include viable fungi, prokaryotes, pollens and viruses as well as biologically derived remnants) are suspected to have a fundamental role in structuring the composition and function of ecosystems globally. Urban aero-microbiomes are poorly understood, yet a majority of the global population is exposed to them on a daily basis. Antarctica presents a tractable opportunity to study the dispersal of airborne microorganisms due to its isolation and its simple, microbially dominated ecosystems. Recent advances in technology have begun to shed light on the poorly understood aerosphere, with most research focusing on bacteria. This chapter summarises the current knowledge regarding the movement and behaviour of bioaerosols in the global atmosphere and drivers of bioaerosol spatiotemporal variation. Urban aero-microbiomes are characterised and the role that the air plays as a vector of microbes to Antarctica is described, with an overview of Antarctic bioaerosol research. Survival mechanisms of microbes in the atmospheric environment are outlined, followed by a discussion of the potential effects that aerial input to vulnerable Antarctic ecosystems may have in the face of climate change. Bioaerosols are highly changeable over space and time, with concentrations and compositions influenced by a myriad of variables, particularly climatic factors such as wind speed and temperature. Although studies of bioaerosols have confirmed the low biomass predicted in the atmosphere, greater biodiversity has been discovered as technology has improved. Multiple lines of evidence indicate that bioaerosols have been globally transported over great distances. While many microbes are believed to survive in the atmosphere as spores, some species may remain metabolically active and could contribute to certain atmospheric processes. The evidence of continual bioaerosol deposition and theorised significance to current ecosystem structuring suggests that as the climate changes, deposited microorganisms could drive rapid community shifts. This chapter identifies numerous knowledge gaps in the field, including the variability, environmental drivers, source (where) and extent (how many) of airborne microorganisms. Given the predicted importance of airborne transportation to global ecosystems, it is essential to substantially increase research effort to gain a more comprehensive view of the extreme aerosphere.

1.2 Introduction

Microorganisms (including fungi, bacteria and viruses) are numerous and ubiquitous in the atmosphere and are an important component of bioaerosols, which include all particles of biological origin (Burrows, Elbert, Lawrence, & Poschl, 2009b; Pearce et al., 2016; Reche, D'Orta, Mladenov, Winget, & Suttle, 2018). Bioaerosols within the natural environment are poorly understood, due to a lack of standardised methodology and little data, resulting in abundant conjectures (Burrows et al., 2009b; Pearce et al., 2016). Recent technological

innovations have enabled DNA recovery and interrogation of airborne microorganisms from a wide range of environments, which allows researchers to address fundamental questions regarding the importance of this biosphere.

The urban aero-microbiome has a disproportionately large effect on the global human population, due to concentration of settlement in urban areas. The research which has been performed to date indicates that various factors can influence urban bioaerosols, but also suggests significant variation over space and time and no consistent urban aero-microbiome. As for bioaerosols in general, the urban aerial biome is affected by seasonal variations, likely to be driven by changes in meteorological variables such as temperature and wind speed. Local bioaerosol emission sources are thought to be more important than distant ones and pathogens are often present. Spatial and temporal coverage remains sparse, and knowledge gaps persist. The Antarctic aerial biome represents one of the most challenging environments on Earth. As a result, the aerosol biomass is one of the lowest and is the least understood in the world (Burrows et al., 2009b). There is strong evidence that bioaerosols use the atmosphere as a long-distance transport vector; for example, South American pollen is frequently found in Antarctica (Smith, 1991; Vincent, 2000; Wynn-Williams, 1991). Continental Antarctica provides an excellent natural laboratory to study global aerial transport processes as it is isolated by the Southern Ocean and prevailing wind and water currents (Pearce et al., 2016). The reduced human and other animal vectoring means the atmosphere becomes virtually the sole transport mechanism available for microbes (Pearce et al., 2016) and understanding transport processes is a tractable problem due to Antarctica's simple, microbially-dominated ecosystems (Bottos, Woo, Zawar-Reza, Pointing, & Cary, 2014). In a warming world, shifts in microbial communities may result in the extinction of unique endemic species. Microbes are first responders to change. Therefore, understanding how Antarctic communities may shift informs what can be expected globally in the future (Bottos et al., 2014). Despite technological developments, the study of bioaerosols is challenged by their stochasticity in distribution, variability and low biomass (Womack, Bohannan, & Green, 2010), which is accentuated in Antarctica (Pearce et al., 2016). Very little is known about bioaerosols, and methodology development remains a barrier to widespread investigation (Pearce et al., 2016).

Most researchers believe that the majority of passively transported bioaerosols are in a dormant form (Pearce et al., 2016; Womack et al., 2010); however, there is growing evidence that some may continue to metabolise whilst in transit. These organisms may represent atmospheric residents (Womack et al., 2010) and could substantially alter the chemical constituents of the atmosphere. Most bacteria in the atmosphere are thought to act as cloud condensation nuclei (Burrows et al., 2009b) and ice nucleation activity could increase cloud formation, precipitation and fundamentally affect global weather patterns (Behzad, Gojobori, & Mineta, 2015; Burrows et al., 2009b; Pearce et al., 2016; Sattler, Puxbaum, & Psenner, 2001). Many important plant and animal diseases are also aerially

transmitted, such as foot and mouth disease and Legionnaires disease (Nguyen, Ilef, Jarraud, Rouil, & Desenclos, 2006), highlighting the need to understand atmospheric microbial transfer. It was once thought that "everything is everywhere and the environment selects" (put by Dutch microbiologist Lourens G. M. Bass Becking (O'Malley, 2008)), however more recent research has revealed bioaerosols are highly variable over space and time (Bowers et al., 2013; Bowers, McLetchie, Knight, & Fierer, 2011a; Fierer, Liu, & Rodríguez-Hernández, 2008; Woo et al., 2013). Many microbes appear to display biogeography (varying species distribution over space), despite their ease of dispersal (Bahl et al., 2011; O'Malley, 2008; Pointing et al., 2015; Womack et al., 2010). A microbe's propensity for aerial transport is thought to be a key influencer of its ability to disperse (Pearce et al., 2016; Sokol, Herbold, Lee, Cary, & Barrett, 2013; Sommaruga & Casamayor, 2009).

This chapter explores what is known about the aero-microbiome and some of the questions that remain. The global atmosphere is defined and described, followed by the process of bioaerosol launching and elucidation of aerosol behaviour once in suspension. Mechanisms of long-range aerial dispersal are discussed, followed by the processes that remove particles from suspension. What is known regarding variation of bioaerosols over space and time is summarised, with a focus on urban areas. Bioaerosol sampling challenges and solutions to difficulties are listed. The importance of the atmosphere as a vector to Antarctica, as well as the status of Antarctic bioaerosol research, is described. The mechanisms that microbes use to survive in the atmosphere are then reviewed and the existence of aerial residents is explored. The chapter finishes with effects of aerial vectoring on Antarctic ecosystem structuring and how microbial communities may respond to continued climate change. The chapter content represents a comprehensive review of the field of bioaerosol study, with a focus on the small body of work that has been performed in urban environments and in Antarctica. Given the hypothesised importance of aerial vectoring to global ecosystems, understanding the aerial contribution is crucial to understanding ecosystem structuring.

1.3 The Aerosphere and Bioaerosol Particles

1.3.1 Particle Movement in the Atmosphere

The atmosphere consists of a layer of gases which surrounds the Earth, containing particles of various sizes and types, such as water droplets, inorganic particles like mineral dust and particles of biological origin. Particles suspended in air are called aerosols, those of biological origin are termed bioaerosols. Most aerosol particle monitoring is focused on human health and as a result uses two health-relevant size classifications, $PM_{2.5}$ and PM_{10}, which refer to the aerodynamic diameter of the particle (< 2.5 µm and < 10 µm). Bioaerosol particles vary widely in size, from pollen which can be up to 1000 µm in diameter, bacteria which range between 0.25 and 8 µm in diameter, to viruses at < 0.3 µm in diameter, and fragments of biological material thereof. Air is in constant motion, driven by differences in

atmospheric pressure, which cause its constituents to be highly variable over time and space. The concentration of aerosol particles in the atmosphere is influenced by a complex set of variables not comprehensively understood, including local biological sources and changes in meteorological conditions (Burrows et al., 2009b).

The vertical structure of the atmosphere is divided into layers which are driven by temperature (**Figure 1-1**), although the altitude at which layers transition varies (Pearce et al., 2009). At greater altitude wind speeds typically increase and conditions become less favourable for survival (Pearce et al., 2009). The characteristics of the atmosphere determine both the transport range (Archer & Caldeira, 2009) and the viability of bioaerosols (Womack et al., 2010). The troposphere stretches from the ground to about 10 km and is where the majority of atmospheric mass, including bioaerosols, is located. The troposphere also contains the atmospheric boundary layer, the region of the atmosphere that transitions from turbulent air flow from the Earth's surface to a calmer, laminar flow layer. The boundary layer mediates exchange of particles between the Earth and the atmosphere and its altitude varies depending on atmospheric conditions and terrestrial topography (Rotach et al., 2015). At between around 10 km and 50 km altitude sits the stratosphere, where airflow is fast, predictable and horizontal. The troposphere and the stratosphere are of most interest for the purposes of bioaerosol research, as most microbial isolates are from these levels (Burrows et al., 2009b), although some studies have found culturable organisms in the mesosphere, at up to 77 km altitude (Imshenetsky, Lysenko, & Kazakov, 1978).

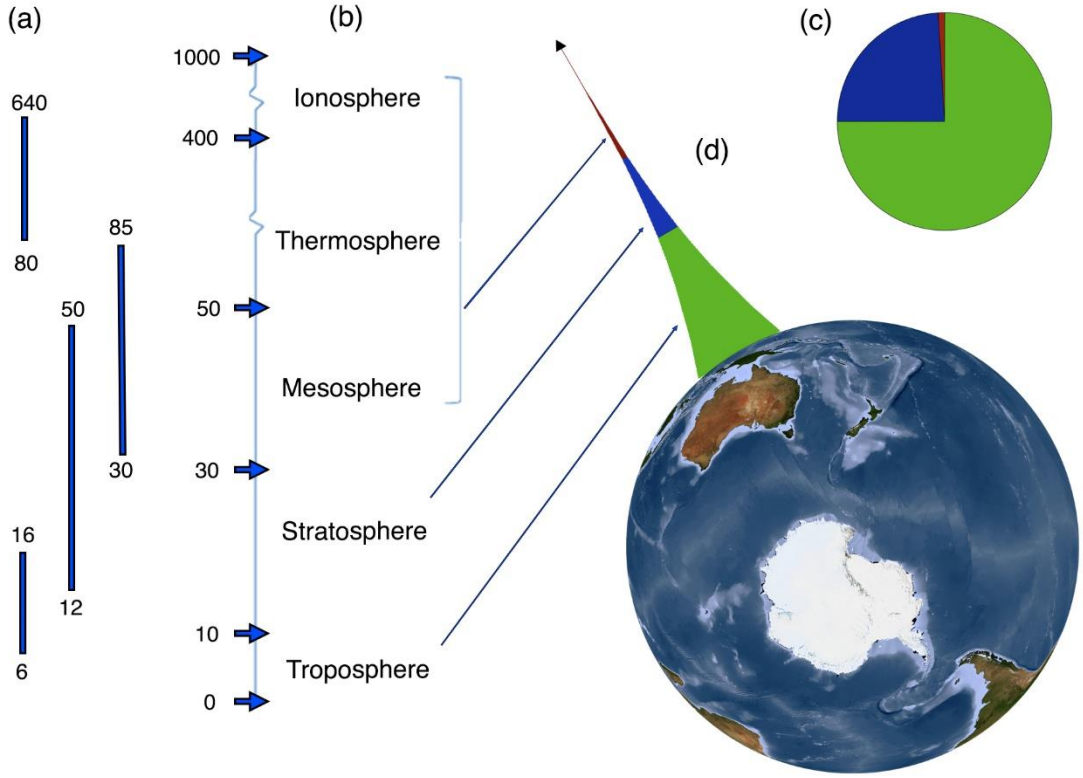


Figure 1-1 – a) Altitude ranges of each atmospheric layer (in km), layer height varies based on temperature and latitude (b) Names of atmospheric layers, (c) Adjusted logarithmic scale by height and density with cumulative area as in (d), (d) Speculative proportions of bioaerosols found in each layer, showing 75% of bioaerosols are expected to remain in the troposphere. Artwork by Chris King, data from Pearce et al. (2009).

The process of aerosolisation of particles is known as launching (Pepper, 2015) and can be either active or passive. Active launching includes forcible ejection of biological material as seen with fungal spores or the process of sneezing. Passive launching results from abiotic processes acting on a reservoir of particles, such as wind blowing over soil or plants, waves breaking on a beach or the bursting of bubbles in water (Burrows et al., 2009b). Rates of passive emission vary based on density of particles on a surface, and local meteorological conditions. Turbulence from the Earth's surface creates chaotic air movement in the atmospheric boundary layer, including vertical movement, which can propel particles to higher atmospheric regions. Turbulence disrupts the flow of particles when air is forced around an object, or subjected to excessive shear. The Reynolds number (velocity x dimension/viscosity) can estimate the amount of turbulence and is based on wind velocity, viscosity of the air and dimensions of the interfering surface (Pepper, 2015). Any result over 2000 is deemed to be turbulent air flow (Pepper, 2015). The higher this number, the more movement of particles will occur in a given time, and the higher the extent of aerosolisation. As events that propel particles to high altitudes are rarer, the majority of particles are thought to remain in the atmospheric boundary layer (**Figure 1-1**) and as a result have short transport ranges. The minority of particles that do escape the

boundary layer, to the tropopause, or stratosphere where the airflow is much faster and more uniform, can be rapidly transported on a global scale. Reche et al. (2018) estimated downwards viral flux above the atmospheric boundary layer of 0.26×10^9 to $>7 \times 10^9$ m^{-2} viral particles per day. They found downwards flux of bacteria of 0.3×10^7 to $>8 \times 10^7$ m^{-2} per day. So, although this represents a minority of bioaerosols, significant numbers still escape the boundary layer. There are few studies of vertical distribution of bioaerosols and most of these are qualitative, only indicating the presence or absence of microbes (Maki et al., 2017; Maki et al., 2008; Wainwright, Wickramasinghe, Narlikar, & P., 2003). Much of the current research supports reductions in concentration with altitude (Fulton, 1966a; Fulton, 1966b; Fulton & Mitchell, 1966), however some studies do not show a clear relationship between concentration and altitude (Andreeva et al., 2002; Matsuki, Iwasaka, & Osada, 2003).

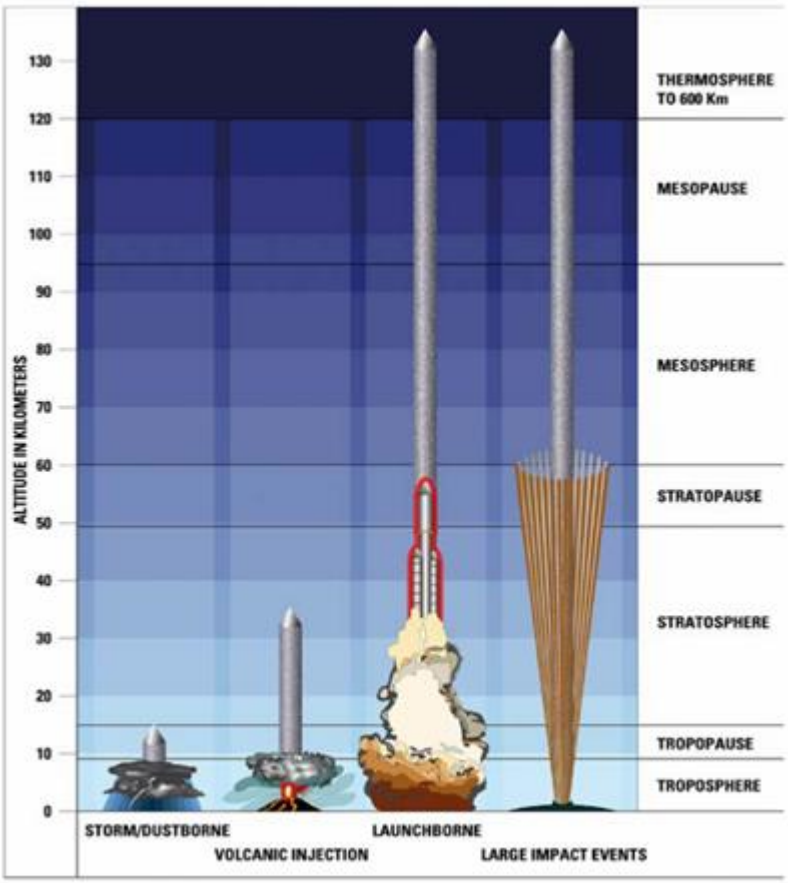


Figure 1-2 - Mechanisms of microbial movement into the atmosphere (Griffin, 2004), artwork by Betsy Boyton. Storms can launch aerosols into the tropopause, volcanoes to the stratosphere, and rockets or meteorites can propel particles as far as space.

Various events can propel particles to different atmospheric layers (**Figure 1-2**). Storms are a good example of frequent processes which transport bioaerosols upwards. DeLeon-Rodriguez et al. (2013) found hurricanes in the Caribbean created large amounts of bioaerosols that were launched into the tropopause. Dust storms in desert areas also launch

large quantities of aerosols, with increases in number and diversity of microbes evident during dust events (Kellogg & Griffin, 2006; Maki et al., 2011). Volcanic eruptions are less frequent in occurrence but have the potential to propel substantial numbers of particles to the stratosphere and some cases even the mesosphere (Diallo et al., 2017). Krakatoa erupted in 1883 and observers at the time estimated that ash was ejected to an altitude of 80 km (O'Connell, 2006). The effective transportation of thermophilic organisms over large distances through volcanic eruption is supported by the detection of similar organisms in widely dispersed geothermal sites (Herbold, Lee, McDonald, & Cary, 2014). Other impact events such as meteorites landing are rarer still but could feasibly transport particles throughout the atmosphere. Survival at these high altitudes is possible, with some experiments showing bacterial survival alongside space craft lift off and re-entry (Rettberg et al., 2002). There is evidence that *Bacillus subtilis*, a spore-forming, Gram-positive bacterium, that is frequently found in bioaerosols, can survive in space for at least six years (Horneck, 1993).

Although large scale movements are driven primarily by air currents, small scale movements of airborne particles depend on Brownian motion (the random movements in a fluid from collisions with fast moving molecules and other suspended particles). Particles diffuse from a source down a concentration gradient, with trajectory and speed of diffusion influenced by air currents and gravity (Pepper, 2015). Particles with high mass require more force to change direction (force = mass x acceleration) meaning that they are more likely to impact on a surface and be removed from the air (Pepper, 2015).

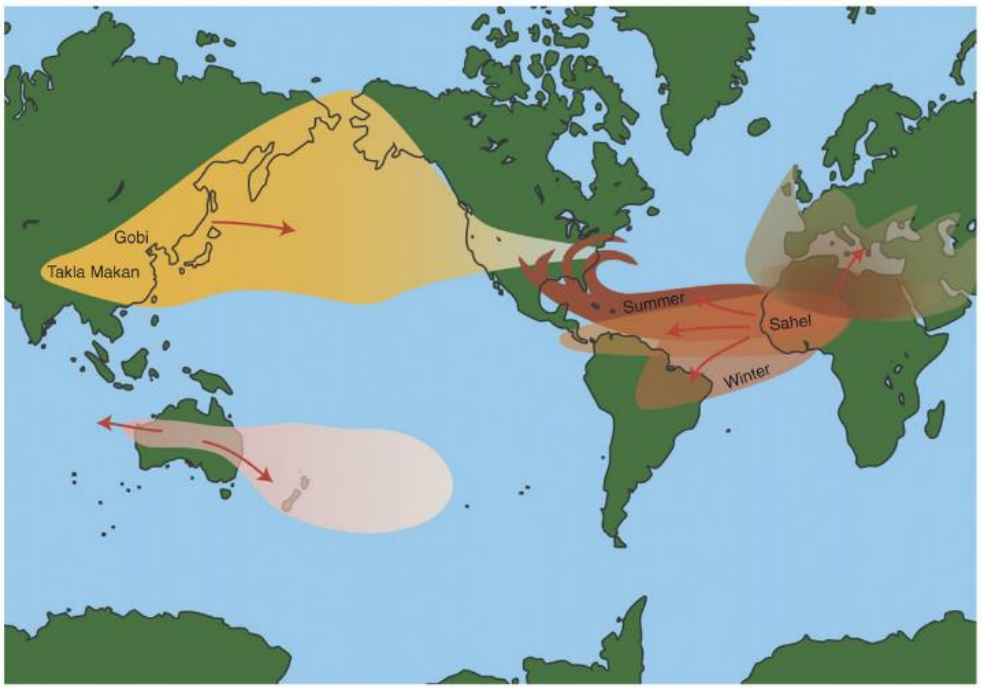


Figure 1-3 - Storms are capable of transporting large numbers of particles (Kellogg & Griffin, 2006). The three main sources of worldwide desert dust are shown (Australia, the Sahel in Africa and China) along with their typical seasonal movements.

Particles from deserts are small in size, have long residence times and are likely to be significant contributors to global transportation of bioaerosols (**Figure 1-3**) (Hara & Zhang, 2012; Kellogg & Griffin, 2006). The transport of dust and associated organisms is supported by similarities of dust in different areas. Caribbean and African isolates show very high ribosomal RNA gene sequence similarity, suggesting a common source of bioaerosols (Kellogg & Griffin, 2006; Kellogg et al., 2004). Reche et al. (2018) found Saharan particles in Spanish air and particles in Florida in summer are often of African origin (Prospero, 1999). Maki et al. (2011) studied Asian kosa dust and found that after dust events in China, there were significantly more culturable bacteria in snows in Japan, which had been contaminated by the dust.

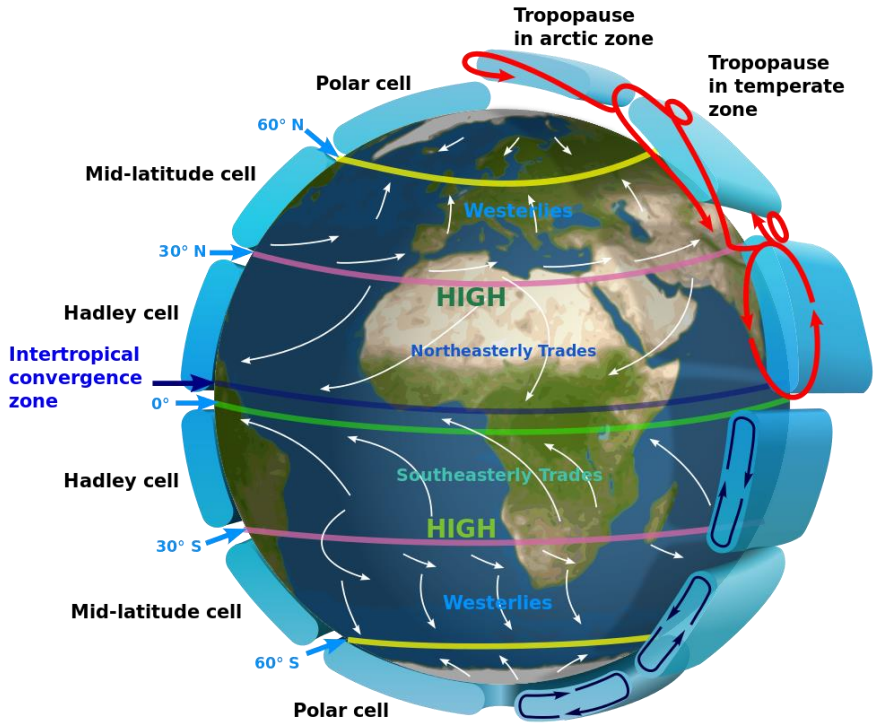


Figure 1-4 - Diagram of the major air cells and wind currents in the Troposphere (Kaidor, 2013). These winds influence trajectory of airborne transport of microbes.

Air cells are created by differential heating of the Earth between the equator and the poles, and the Earth's rotation. Predictable air cells above the boundary layer facilitate movement in the atmosphere horizontally in the prevailing wind direction of the cell (**Figure 1-4**). For long distance transport, bioaerosols need to get above the atmospheric boundary layer, to these air cells, where wind speeds have been estimated to be up to 90 km/h (Miller, Gans, & Kleidon, 2011). Bacterial residence times (the time that a bacterial cell, on average, is expected to remain aloft in the atmosphere) are modelled to be between three to seven days, although this may differ for other bioaerosols (Burrows et al., 2009b). In seven days at 90 km/h, there is sufficient time for a microorganism to be transported over 15,000 km (the distance from the South Pole to London is 15,710 km).

Particles are removed from aerial suspension through particle deposition. Gravitational settling is the gravitational pull of any particle heavier than air downwards and is the most common cause of deposition. Therefore, bioaerosols with lower mass will tend to stay aloft longer, have longer residence times and consequently greater transport range. Gravitational settling can be described by Stokes's law, which considers gravity, particle density, particle diameter and air viscosity.

$$V = [D^2 \times (p_p - p_1) \times g]/18p$$

V = velocity of fall (m/s), g = gravity (9.8m/s^2), D = diameter of particle (m), p_p = density of particle (kg/m^3), p_1 = density of dispersion medium (kg/m^3), p = viscosity of dispersion medium (kg/m*s)

Particles can be deposited on a surface through impaction (when they collide with the surface), also known as dry deposition. The particle loses kinetic energy and can come to rest on the surface, or deflect and return to the air flow with reduced kinetic energy, increasing the chance of settling. Probability of impaction on a surface depends on velocity, particle diameter and the size of the surface. Wet deposition is mediated by rain or other precipitation. As rain or snow falls, it collides and combines with aerosols creating particles of greater mass that can settle faster. Efficiency of wet deposition depends on the spread area of the particle plume, with diffused plumes experiencing stronger impaction. Electrostatic deposition can also occur when particles with opposing charges are attracted to one another, creating particles with greater mass. Bacteria are often negatively charged and therefore have a tendency to become attracted to positively charged particles in the atmosphere or surfaces, which increases deposition (Pepper, 2015).

1.3.2 Bioaerosol Particles: Temporal and Spatial Variation

Bioaerosols constitute a significant and variable portion of aerosol particles, up to 28% of total aerosol volume (Matthias-Maser, Obolkin, Khodzer, & Jaenicke, 2000). Bioaerosols are thought to frequently exist in assemblages with other inorganic or organic particles (Burrows et al., 2009b). A bacterium is around 1 µm in diameter, whereas typical bacterial-associated particles are 2 – 4 µm, suggesting that multiple cells clump together or associate with other particles (Shaffer & Lighthart, 1997). Huffman, Treutlein, and Poschl (2010) observed bioaerosol particle sizes of 1.5 µm, 3 µm, 5 µm and 13 µm. 1.5 µm is likely to represent single bacteria, 3 µm is likely to be multiple bacterial cells or fungal spores, 5 µm is likely to be fungal spores and 13 µm is likely to be pollen (Huffman et al., 2010). Several authors have proposed particle association could improve bioaerosol viability (see section 1.4 Microbial Survival Mechanisms in the Atmosphere onwards).

The broad and rapid dispersal of particles from a source in the air means that experimental determination of microbial residence times remains infeasible, leaving much of our understanding related to atmospheric modelling. Burrows et al. (2009a) modelled the aerosol concentration of bacteria in the global atmosphere for six simulated years, using global meteorological models, estimated emissions from different ecosystems, estimated

residence times and estimated deposition rates (**Figure 1-5**). In the simulation, a mean microbial residence time of around one week was calculated. There was significant variation, with some particles staying aloft for months. Meteorological variables significantly affect residence times, for instance turbulence in the atmosphere will increase residence times, while precipitation in an area reduces residence times. Dry regions have longer expected residence times due to increased vertical movement in the boundary layer, related to greater turbulence from heating and reduced wet deposition. As the number of bioaerosol studies increases across a range of spatial, temporal and environmental variables, our understanding of microbial residence time influences will increase. Physical characteristics of particles, such as size or surface structure, also affect residence time. Particles with a smaller diameter or lower mass tend to have greater residence times, due to reduced deposition. It is thought that particles under 1 μm in diameter (which include viruses and some free-living bacteria) fall within the "scavenger gap" and have the longest residence times (Burrows et al., 2009a). This assumption is one of the bases for Burrows' simulation. The surface structure of pollen can alter its aerodynamic properties and aid its dispersal by wind (Niklas, 1985). Given similar selective pressures apply to other types of bioaerosol, they may use similar mechanisms to enhance dispersal. Some bacteria, such as *Pseudomonas* spp., have cell surface ice nucleation proteins (Pearce et al., 2009) and are believed to act as cloud condensation nuclei (Bauer et al., 2002). It is thought that organisms which act as cloud condensation nuclei have increased wet deposition in the atmosphere and therefore have reduced residence times (Burrows et al., 2009a).

Reliable estimates of bioaerosol concentrations from all ecosystems are not available and variable methodologies make comparison difficult. Estimates based on available information show that there is significant variation in bioaerosol concentration by ecosystem. Generally higher concentrations are observed over more productive ecosystems (grassland and crops), likely due to larger microbial source populations from which microbes can be aerosolized (Harrison et al., 2005; Tong & Lighthart, 2000). Desert areas have low estimated concentrations of bioaerosols by mass due to low source biomass. However, due to the ease of aerosolisation of particles from desert surfaces and long residence times of desert particles, desert dust is still thought to play a significant role in global microbial dispersal (Bowers, Sullivan, & Costello, 2011b; Kellogg & Griffin, 2006; Maki et al., 2017; Shaffer & Lighthart, 1997).

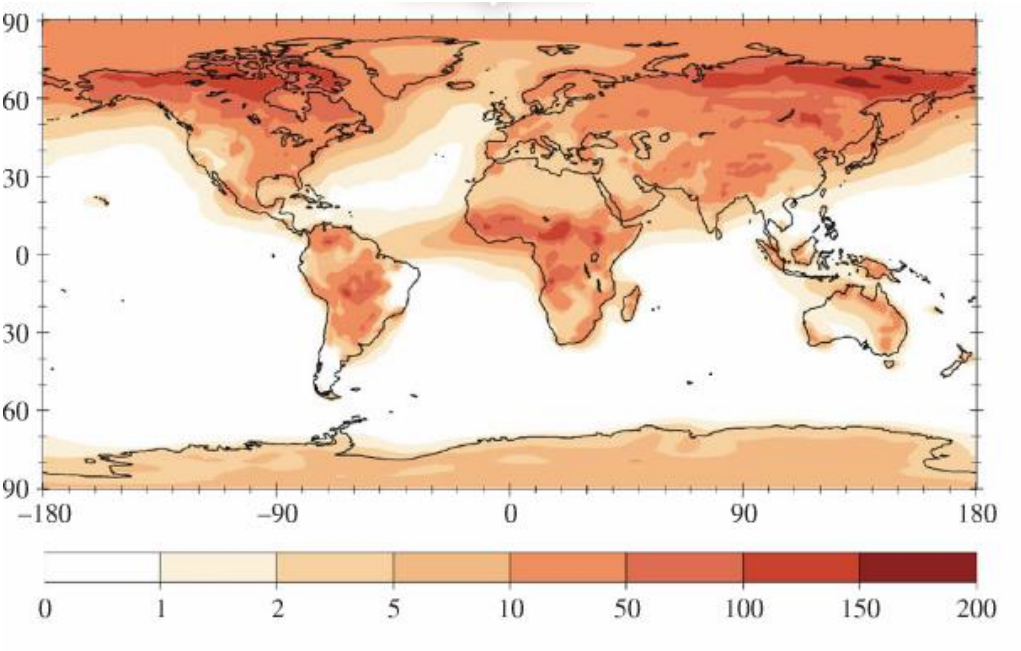


Figure 1-5 - Heat map of the simulated concentration of 1 µm bacteria in the troposphere according to modelling by Burrows et al. (2009a).

Burrows' model shows that after three years, bacteria reach an effective distribution equilibrium and are present in all locations. This suggests that short and long range atmospheric transportation are highly effective at dispersing microorganisms globally (**Figure 1-5**). Later studies of bioaerosols support the high concentration heterogeneity predicted by the model (Barberán, Henley, Fierer, & Casamayor, 2014; Bowers et al., 2013; Bowers et al., 2011b; Grantham et al., 2015). The application of the model to urban areas is discussed in 1.3.3 Spatiotemporal Variation of Urban Bioaerosols and Antarctic bioaerosols, the other focus of the thesis, is discussed in 1.3.4 The Antarctic Aerosphere.

Bioaerosols are highly variable over various time scales with stochasticity in concentration and community structure over a short timeframe frequently more pronounced than seasonal changes (Burrows et al., 2009b; Fierer et al., 2008). Some cultivation-based studies have shown that seasonal variation is very likely to be driven by changes in meteorological conditions, occurring predictably over the course of the year. The highest average concentration of bioaerosols is thought to be at times of maximum productivity, generally in the summer (Lighthart & Stetzenbach, 1994; Tong & Lighthart, 2000). However, this could be confounded by significant seasonal variations in culturability of microbes (Burrows et al., 2009b). A molecular study, (Woo et al., 2013) found the highest microbial loading occurred during summer. This variation extends to observed diurnal patterns linked to solar heating as net upwards flux was highest in the warmest part of the day (Chen et al., 2001; Shaffer & Lighthart, 1997; Tong & Lighthart, 2000). Higher

temperatures increase turbulence, wind speeds, vertical mixing, residence times and division rates, increasing the microbial load and aerosolisation.

1.3.3 Spatiotemporal Variation of Urban Bioaerosols

Understanding characteristics of urban bioaerosols is particularly pertinent, given the majority of the world's population reside in urban areas and this is forecast to increase (UN, 2018). A majority of the bioaerosol studies which investigated differences in land-use and levels of urbanisation detected differences in the urban and rural aero-microbiome. When culturable bacteria were investigated in Oregon using city, field, forest and coastal sites, it was found that bacteria varied by location and time (Shaffer & Lighthart, 1997). The highest concentrations were detected at the urban site. *Bacillus* spp. were frequently present and the majority of bacteria were associated with particles greater than 3 µm aerodynamic diameter (Shaffer & Lighthart, 1997), supporting predicted spatiotemporal variation and suggesting a distinctive urban aero-microbiome. DNA sequencing in Germany detected plants, protists, fungi and bacteria at urban, rural and high alpine locations. Bacterial T-RFLP analysis found Shannon-Weaver diversity was highest at the rural location compared to the urban or the alpine location. Ascomycota and Basidiomycota fungi were detected along with proteobacteria, actinobacteria and firmicutes (Després et al., 2007). This indicated that the urban aero-microbiomes may be less diverse, possibly due to fewer different microbial sources in a more homogenised built environment. Bowers et al. (2011a) performed bioaerosol sampling in Colorado across three land use types: field, suburban and forest. They found no differences in bacterial cell concentration but more biological ice nuclei over agricultural areas. Pyrosequencing revealed that bacterial communities were significantly related to land use type, driven by local sources, again supporting the idea that urban aero-microbiomes differ to others, but not the higher urban concentrations previously observed. Tanaka et al. (2020) reported greater alpha diversity at a suburban site in Japan compared to an urban site and distinctive organisms were detectable at the different locations. The urban site harboured human skin-associated bacteria such *Propionibacterium, Staphylococcus*, and *Corynebacterium*, while soil and plant-associated bacteria were abundant at the suburban site (*Methylobacterium* and *Sphingomonas*). This was consistent with previous findings in terms of the presence of urban microbial signatures and reduced urban diversity. Metabolite fingerprinting performed in the UK in urban parks, industrial areas and farms found seasonal and land use differences in the sampled aerosols and that different compounds were associated with different land uses, signifying the presence of different organisms between locations (Garcia-Alcega et al., 2020). Bowers et al. (2013) found that at two urban and two rural sites, the airborne bacteria varied significantly between sites and displayed seasonality, further supporting spatial variation and in addition predictable temporal variation.

In contrast to these findings, Woo et al. (2013) sampled along an urbanisation gradient in tropical Hong Kong and detected no effect on the bacterial or fungal bioaerosol communities. However, they confirmed distinctive seasonal changes which were attributed

to a different predominant wind source between the summer and winter. In a study of continental wide distribution of bacteria and fungi across the US, external house dust from 1,200 homes was subjected to 16S rRNA (16S) and Internal Transcribed Spacer (ITS) DNA sequencing. Fungi and bacteria were found to be highly spatially variable, driven by climatic and soil variables. A distinctive urban versus rural community was not reported, but they did find that urbanisation tended to lead to homogenisation, with urban communities being less geographically variable than rural ones (Barberán et al., 2015), in some agreement with the notion that sources of bioaerosols could be less diverse in urban areas.

Taken together, it seems that bioaerosols in urban areas can be less diverse than non-urbanised areas. Urban areas do sometimes display higher concentrations of airborne microorganisms. There is evidence for different compositions of bioaerosols in different areas but there does not seem to be a common worldwide urban aerial community. The urban aero-microbiome appears to vary by city, depending on local sources and climatic variables and varies over time. Unfortunately, the areas in which the urban aero-microbiome has been studied are geographically limited, most studies having been performed in the USA, Europe and Asia.

Cultivation studies have been conducted in Beijing. Three sites within the city were compared: a roadside, a human activity enriched area and a park. Bacterial concentrations in the park were found to be significantly lower, postulated to be due to fewer vehicular movements aerosolising microorganisms. *Micrococcus* was the dominant genus detected, along with *Staphylococcus*, *Corynebacterium*, *Bacillus* and *Pseudomonas* (Fang, Ouyang, Zheng, Wang, & Hu, 2007a). The lower concentrations in parks were somewhat consistent with Shaffer and Lighthart (1997) observing higher urban concentrations. Later, a metagenomics study was performed in Beijing, with smog analysed for bacteria, archaea, fungi and double-stranded DNA viruses. The organisms detected were mostly soil related and non-pathogenic, however human allergens and pathogens were found and their abundance seemed to increase with pollution levels (Cao et al., 2014). This indicated pollution is important, and may be a driver of urban bioaerosol variability between cities, but application to less polluted areas may be debatable. In the USA, DNA microarrays were used to characterise the bacterial bioaerosol community in two cities. 1,800 types were found, with a diversity similar to soil. Some of the groups detected had pathogenic members. The authors found temporal and local meteorological influences can be stronger than location in influencing bioaerosol composition (Brodie et al., 2007b). When the aero-microbiome of 96 sites in North American cities were characterised, they were highly seasonal with microbial sources identified including soils, leaves and dog faeces (Bowers et al., 2011b), again supporting importance of local sources and seasonality. Metagenomics techniques applied in the US National Capital Region detected bacteria, plants, fungi, invertebrates, and viruses in the air. Temporal shifts were evident, with bacteria peaking in the summer and fungi in the spring. The bacterial genera *Ralstonia, Cupriavidus*, and *Bacillus* were abundant throughout the year (Be et al., 2015), strengthening observations of

seasonality in previous studies. Cultivatable bacteria and fungi were characterised along the New York waterfront. Microbial counts were found to be higher when the wind was offshore, and bacteria were dominant. When the wind was onshore, fungi were dominant. Microbes were associated with coarse aerosols, which signified importance of local sources and there was a significant correlation between wind speed and microbial culturable aerosol concentrations. Actinobacteria were dominant, specifically the genera *Streptomyces* spp. and *Bacillus* spp. (Montero, Dueker, & O'Mullan, 2016). This again suggests urban areas may have higher bioaerosol concentrations, and dominance of local sources. Urban German fungal communities confirm global consistency of seasonal changes. Plant pathogens were associated with coarse particulates and human pathogens and allergens associated with fine particulates. *Cladosporium* spp., *Alternaria* spp., *Penicillium* spp. and the plant pathogen *Blumeria graminis* (mildew) were found (Fröhlich-Nowoisky, Pickersgill, Després, & Pöschl, 2009). In Milan, pyrosequencing was used to interrogate bacterial bioaerosols. Seasonality was evident, with plant associated bacteria more common in summer and spore formers more prevalent in the winter. Species richness was comparable to soil but species evenness was low (Franzetti, Gandolfi, Gaspari, Ambrosini, & Bestetti, 2010). Low evenness is expected with variable bioaerosol populations, and again seasonality was evident. Diurnal variation was detected in Melbourne bioaerosols, with peaks in abundance at midday of four times the mean night-time values (Jamriska, DuBois, & Skvortsov, 2012), suggesting finer grain temporal variation as thought from modeling. 16S sequencing was performed on bacterial bioaerosols sampled at an elevated site in Tokyo. The air-mass source changed over the sampling period but that did not affect the bioaerosol community, or alpha or beta diversity. There were significant correlations between relative humidity and wind speed and both alpha and beta diversity. Local sea water and soil were identified as constant and predominant sources. Relative humidity appeared to be the most influential variable, due to its correlation with soil moisture and negative correlation with soil emissions (Uetake et al., 2019). This again indicated the importance of local sources versus distant ones, and the impact of meteorological conditions on the aero-mircobiome, suggesting these as a driver of commonly observed seasonal variation. Mhuireach, Wilson, and Johnson (2020) applied 16S metabarcoding sequencing to bioaerosol samples passively deposited in urban parks in Eugene, Oregon. Some sites were forested, and others were grass covered. They defined a core aero-microbiome of plant and soil associated genera. The forested sites were significantly more diverse than the grass covered ones. Seasonal and site-specific effects were detected; vegetation type explained 14% of the differences in communities and site in total accounted for 41% of the variance. The genera *Sphingomonas, Acidiphilium, 1174-90112, Ralstonia, Lactococcus, Methylobacterium, Pantoea, Granulicella, Pseudomonas, Hymenobacter* and *Terriglobus* were common. Increased diversity in forests was consistent with diversity varying by land-use, and seasonality and local effects were in agreement with earlier studies, supporting variation within the urban biome (i.e. between parks and more trafficked areas) previously noted by Fang et al. (2007a).

The urban bioaerosol studies emphasise the spatiotemporal variability of the aero-microbiome. While some genera such as *Bacillus* spp. are very commonly observed, others are inconsistently found, suggesting no consistent global urban aero-microbiome. Climatic variables, seasonality and local sources appeared to be reliably important to the urban aero-microbiome, and land-use within the urban space affected bioaerosols. The multitude of techniques used and varying target organisms make comparison difficult and unifying theories elusive. Bioaerosol communities appear to be driven by many factors, but further research, in different areas and over longer time scales, is needed to understand them in more detail.

1.3.4 The Antarctic Aerosphere

Bacterial concentrations over Antarctica were estimated to be very low from Burrows's simulation and this has since been verified (Bottos et al., 2014; Pearce et al., 2016). Polar areas are thought to have low concentrations of bioaerosols, due to a lack of microbial activity in those areas and the physical stability of frozen surfaces reducing emissions. Pearce et al. (2009) found a large variation of microbial concentrations in different polar and ice ecosystems, although all showed evidence of life. Since there is a lack of data from ice environments Burrows' simulation used average background estimates of bioaerosol concentration as a maximum. The other relevant ecosystem for consideration of Antarctic bioaerosols is the sea, since it surrounds Antarctica. Seas are estimated to have low bacterial concentrations of $1\times10^{4}m^{-3}$ based on cultivation studies (Bauer et al., 2002; Harrison et al., 2005; Kellogg & Griffin, 2006) although this could be an underestimate given marine bacteria are on average less culturable than terrestrial bacteria (Parks et al., 2017). In nutrient rich marine regions, such as those found around Antarctica (Fripiat et al., 2017), concentrations of bacteria are often much higher, likely resulting in higher emissions (Cho & Azam, 1990). Bioaerosol sampling in Antarctica is insufficient to reveal the level of marine aerial input, although a couple of studies show limited marine taxa despite proximity to water (Bottos et al., 2014; Pearce, Hughes, Lachlan-Cope, Harangozo, & Jones, 2010). In Burrows' simulation, Antarctica was decoupled from the rest of the world, due to its isolation by the Southern Ocean. However, residence times of particles in Antarctica were estimated to be high, despite low emissions and most particles that circulated in Antarctica were expected to originate there. Some of these predictions have since been supported by other authors, based on later Antarctic bioaerosol research (Bottos et al., 2014; Crawford et al., 2017; Pearce et al., 2010). Estimates of the bioaerosol fraction of Antarctic aerosols indicate that they constitute only a small proportion of total aerial particles (under around 2%), but also that they can vary significantly (Crawford et al., 2017). This low biological particle fraction in Antarctica is likely driven by low ecosystem productivity and large amounts of dust and other aerosols in the atmosphere.

Westerlies move air from Africa and South America towards Antarctica (**Figure 1-6**). Frequent discoveries of temperate pollen (Wynn-Williams, 1991) in Antarctica indicate that these winds could facilitate intercontinental transport. Within the Antarctic continent there

are also characteristic air movement patterns such as the circumpolar vortex, which is a ring of low-pressure systems that creates circular air movement around the continent and provides a barrier to entry to the continent (**Figure 1-6**). Powerful Antarctic katabatic winds, created as cold high density air flows downwards towards the sea under the force of gravity (Parish & Cassano, 2003), combined with local weather cells, are thought to distribute microorganisms effectively around the continent (Nkem et al., 2006; Pearce et al., 2009). Deuterated methane was released from a plane at 5.5 km altitude in the Maritime Antarctic, and within a week was detected all around Antarctica, indicating rapid dispersal on a continental scale (Mroz et al., 1989). For the purposes of bioaerosol movements, the Antarctic Peninsula can be thought to be somewhat isolated from the rest of the continent, as the Peninsula sits outside of these typical wind patterns and forms a distinct biogeographic zone (Chong, Pearce, & Convey, 2015).

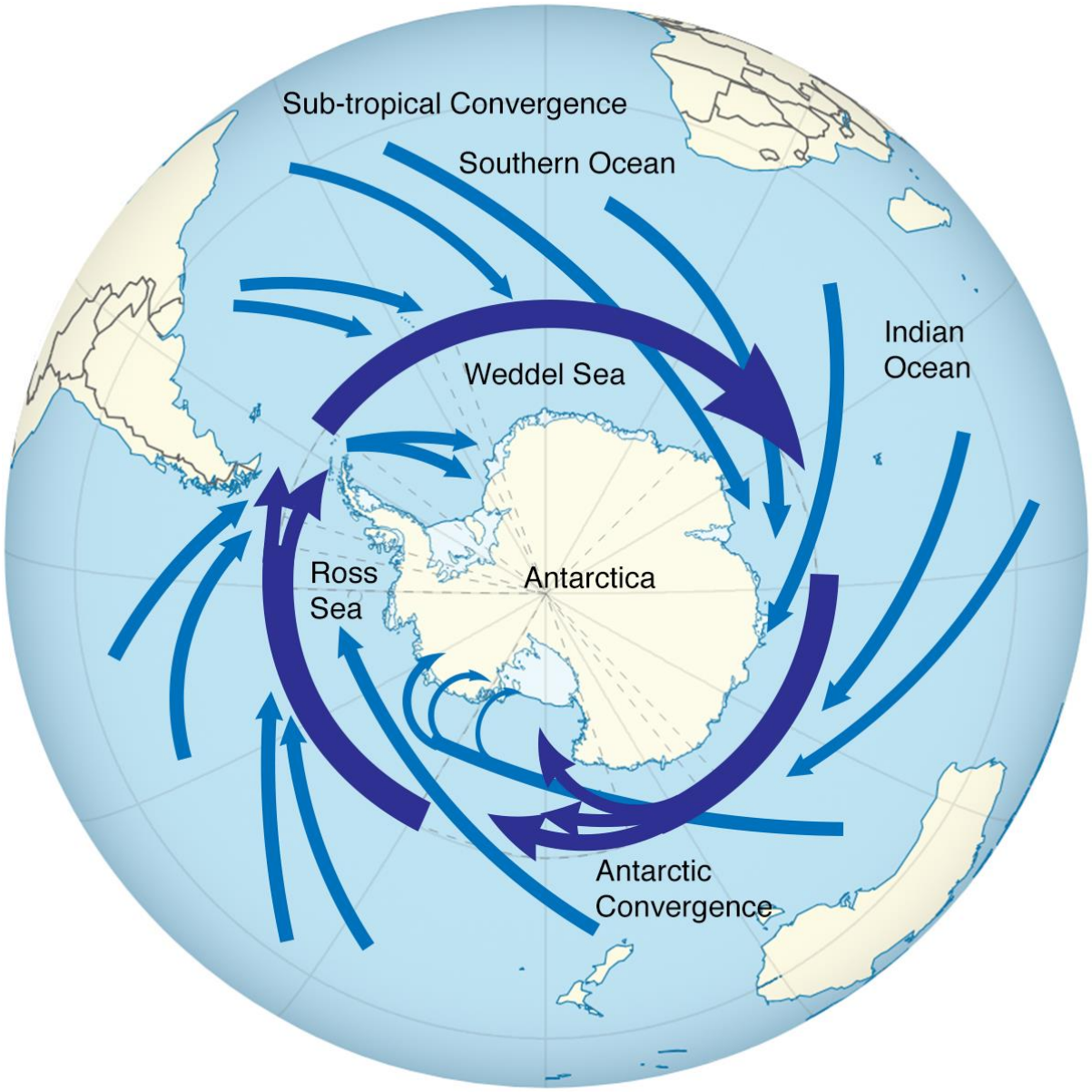


Figure 1-6 - Predominant wind movements over Antarctica during summer at 4 km altitude. Adapted by Chris King from Wynn-Williams (1991). Dark blue arrows are frequent cyclone tracks, lighter blue ones are more occasional tracks. The circum-polar vortex is shown.

1.3.5 Importance of Aerial Transport of Microorganisms to and within Antarctica

Three lines of evidence support the atmosphere as a long-range vector to Antarctica. Firstly, exotic propagules found in Antarctica, secondly isolated thermophiles at remote geothermal sites, and thirdly the presence of globally ubiquitous microorganisms. Exotic propagules have consistently been found in remote locations (Marshall, 1996a; Marshall, 1997; Smith, 1991; Wynn-Williams, 1991). These include lichen spores and pollen granules from Patagonia that have been found at King George Island off the coast of the Antarctic Peninsula (minimum approx. 1,200 kilometres away) (Wynn-Williams, 1991); and exotic

species in Antarctica in ice sheets up to 400,000 years old (Vincent, 2000); as well as plant pollen and fungal spores at Halley Bay (Gregory, 1961), Signy Island and South Georgia Island (Gregory, 1961; Smith, 1991). The lack of alternative transport pathways strongly suggests these propagules arrived via the air and have travelled thousands of kilometres from their source populations in temperate areas of South America (Marshall, 1996a; Vincent, 2000). These propagules could be transported continually from high altitude weather systems or by stochastic powerful low pressure systems capable of translocating large bioaerosol loads from South America, to the Antarctic (Marshall, 1996b), (**Figure 1-7**). These low-pressure systems occur periodically (approximately 1.5 times a year), suggesting that Antarctica has the potential to experience regular microbiological exchange with other land masses.

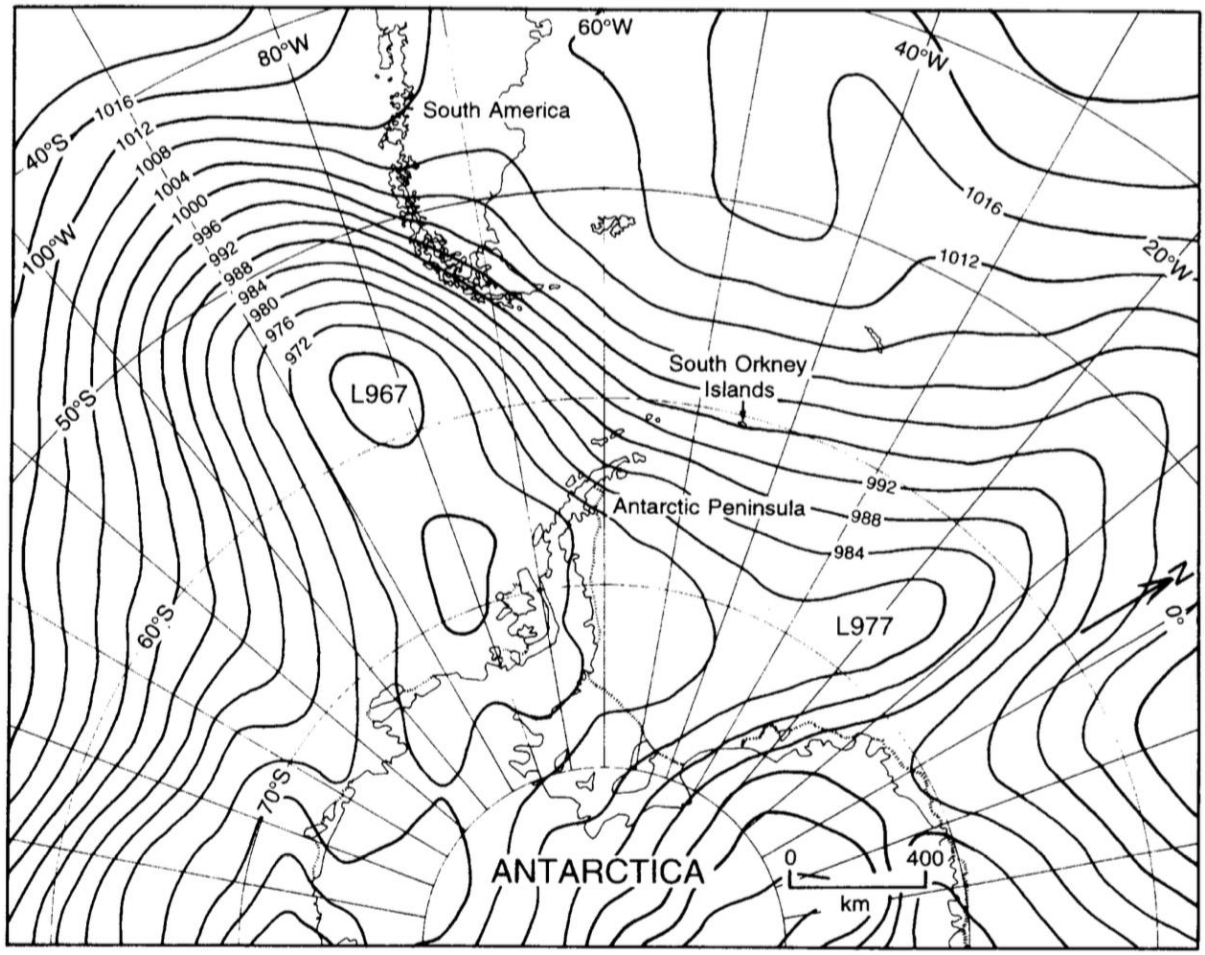


Figure 1-7 - A synoptic chart of 11/11/1993 showing a large low pressure system between South America and the South Orkney Islands (Marshall, 1996b).

As thermophiles are incapable of surviving in the below-freezing temperatures surrounding Antarctic geothermal features, the presence of thermophilic organisms at remote geothermal locations indicate either ancient refugia or recent aerial dispersal. The volatile nature of geothermal sites makes them ideal launchers of material into the upper atmosphere (**Figure 1-2**). A recent molecular study on Mt Erebus microorganisms supports

atmospheric vectoring as a likely contributor to geothermal environments in Antarctica (Herbold et al., 2014). Researchers identified a diverse thermophilic community including *Mastigocladus laminosus*, a thermophile previously isolated from Yellowstone National Park in the USA. This is further supported by thermophiles being identified in a bioaerosol study near Mt Erebus (Bottos et al., 2014). A similar distribution has been observed in multicellular organisms that surround fumaroles (volcanic vents), which are found nowhere else in Antarctica but are common in South America.

The abundance of ubiquitous microorganisms in Antarctic air and terrestrial samples suggests a propensity for aerial transportation and establishment in Antarctica where other means of dispersal are limited. The McMurdo Dry Valleys have been well studied as one of the largest ice-free areas in Antarctica, with comparatively high microbial biomass (Pearce et al., 2016). The valleys are located on the coast, to the west of McMurdo Sound and the ice shelf. Bioaerosol samples taken in the McMurdo Dry Valleys showed that the most prolific bacterial taxa are consistent with bioaerosols from other continents (Bottos et al., 2014). Microorganisms of exotic origin have also been detected in a range of terrestrial and aquatic environments in Antarctica (Archer, McDonald, Herbold, & Cary, 2014; Herbold et al., 2014; Lee, Barbier, Bottos, McDonald, & Cary, 2012; Sokol et al., 2013). Interestingly, many Antarctic studies noted a high sequence similarity between samples within Antarctica, suggesting organisms which reach Antarctica can be effectively distributed around the continent (Bottos et al., 2014; Pearce et al., 2016; Vincent, 2000).

There is evidence that most organisms detected in Antarctica originated there (Bottos et al., 2014; Crawford et al., 2017; Pearce et al., 2010), which can be expected due to Antarctica's extreme isolation. This is consistent with the decoupling of the Antarctic from the rest of the world predicted by modelling (Burrows et al., 2009a). Therefore, it seems long-range aerial dispersal to Antarctica, although present, is rare and acts with more common short-range transport to fundamentally affect Antarctic microbial communities. Local wind is important for local dispersal of cyanobacteria and small eukaryotic organisms (Nkem et al., 2006; Wood, Rueckert, Cowan, & Cary, 2008). Dry Valleys with lakes had more cyanobacteria in nearby soils than valleys without lakes (Wood et al., 2008), (**Figure 1-8**) and wind-borne dispersal of faecal coliforms and avian-associated bacteria has been detected downwind of research stations and bird colonies (Hughes, 2003; Kobayashi et al., 2016; Pearce et al., 2010). Nkem et al. (2006) observed frequent short-range wind dispersal of rotifers and tardigrades (small multicellular invertebrates) in the McMurdo Dry Valleys. Additionally, large volumes of dust are relocated (typically under 30 km but up to 120 km) onto sea ice in the McMurdo Sound, from the Dry Valleys (Atkins & Dunbar, 2009). Given microbes are known to frequently associate with dust and other particles (Burrows et al., 2009b), it is reasonable to assume substantial numbers of microbes would also be transferred.

Figure 1-8 - a) cyanobacterial mat in an Antarctic lake b) These mats can dry out when water levels drop. When dry they break up and are easily transported by the wind.

1.3.6 Bioaerosol Sampling: Challenges and Solutions

The extremely low biomass of the air presents a significant challenge to the collection and analysis of an unbiased representative bioaerosol community. Most air sampling to date either uses passive gravitational settling or active pumping, with impaction or liquid impingement to capture aerosol particles. Although early work on bioaerosols was a critical basis for later study, it drastically underestimated microbial biodiversity (Burrows et al., 2009b; Pearce et al., 2009). Early description of microorganisms in air was based on cultivation or microscopic identification, however a large majority of microorganisms (70 – 99%) have resisted cultivation to date (Burrows et al., 2009b) and microscopic identification cannot differentiate organisms with similar morphologies (Haig, Mackay, Walker, & Williams, 2016). Recent culture independent community analysis based on nucleic acid (DNA or RNA) sequencing has revealed a plethora of diversity not apparent from microscopy or culture. Molecular analysis can provide information on total community identity (amplicon sequencing), whole genomes (metagenomics) or gene transcription (metatranscriptomics) of bioaerosols (Behzad et al., 2015; Yoo et al., 2016).

Aerial sampling challenges are exacerbated in Antarctica where biomass sources in all systems are typically far lower and logistical constraints for studies are greater (Bottos et al., 2014), resulting in a marked lack of data. The current best practice for Antarctic bioaerosol sampling is to pump air through a 0.2 µm polycarbonate filter (Pearce et al., 2016). These filters are easy to run and can be left for long periods of time to collect samples, however they have low flow rates and can take from 24 hours to two months to collect sufficient biomass for analysis (as with Bottos et al. (2014) **Figure 1-9**). The long sample duration on polycarbonate filters disproportionately degrades the DNA of Gram-negative bacteria in under 24 hours due to desiccation stress (Luhung et al., 2015), indicating serious sample bias in existing Antarctic air sampling methodology. Samplers with liquid collection media and higher flow rates have been developed, which aim to reduce sample time and bias but are limited in their use in sub-zero temperatures (Dybwad, Skogan, & Blatny, 2014). Comparative testing reveals different samplers work best for

capturing different types of bioaerosol, so careful choice of sampler for the desired application is crucial (Dybwad et al., 2014; Haig et al., 2016).

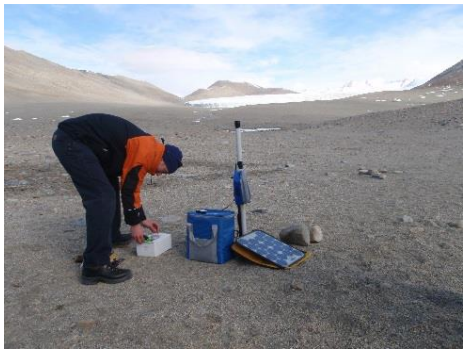


Figure 1-9 - Bottos is pictured setting up the air pump that was used to take the sample leading to the published paper "Airborne Bacterial Populations Above Desert Soils of the McMurdo Dry Valleys, Antarctica" (Bottos et al., 2014).

1.3.7 Bioaerosol Research in Antarctica

Between 1994 and 2014, 12 studies were conducted (both published and unpublished) on bioaerosols in Antarctica (Pearce et al., 2016) (**Table 1-1**, **Figure 1-10**). Most aerobiological studies to date have identified limited biodiversity, however as techniques and resolution have advanced, more recent molecular studies (Archer et al., 2019; Bottos et al., 2014; Hughes, 2003; Pearce et al., 2010) have shown an increasingly diverse and distinctive bioaerosol community, compared to local soils. Unfortunately, the limited studies conducted, samples collected and information gained from the samples have resulted in persistent knowledge gaps. Additionally, the lack of standardised techniques and restricted spatial and temporal coverage mean that it is difficult to make any broad inferences on the significance or extent of bioaerosol transportation to Antarctic ecosystems (Pearce et al., 2016). The studies in **Table 1-1**, and other terrestrial based work in Antarctica strongly suggest that the atmosphere is crucial to both inter and intracontinental Antarctic transport and a distinct microbial community may reside in Antarctic air. Future studies will be able to utilise current knowledge and techniques to conduct bioaerosol sampling in Antarctica that is higher resolution, less biased, inclusive of viruses that may be more abundant in bioaerosols than bacteria (Reche et al., 2018) and determine how these organisms survive and where they originate (see section 1.4 Microbial Survival Mechanisms in the Atmosphere).

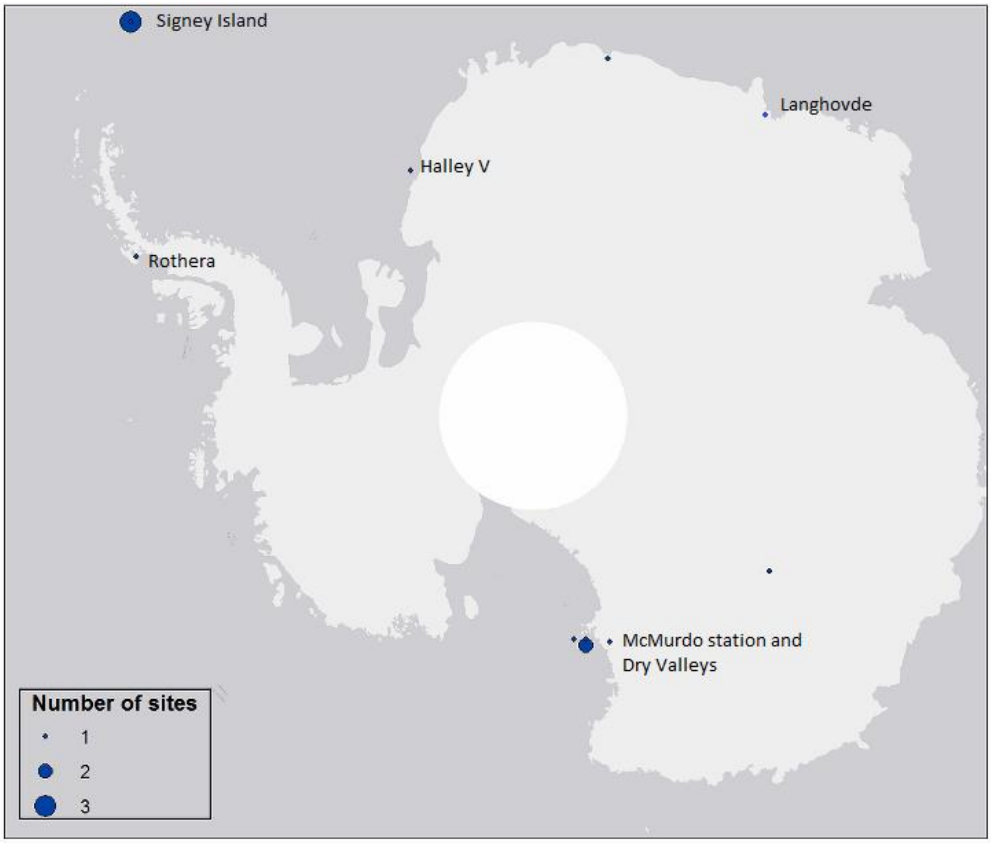


Figure 1-10 - Aerobiological study sites in Antarctica published 1994 – 2014. Adapted from Pearce et al. (2016). Circle width indicates number of sites per study. Most data come either from the Peninsula or the McMurdo Dry Valleys.

Table 1-1 - Summary of aerobiological studies undertaken in Antarctica 1994 – 2021 available on Google Scholar – search terms "Antarctic Aerobiology", "Antarctica Bioaerosol" and "Airborne microorganisms, Antarctica".

Reference	Study area	Sampling Method	Analysis Method	Summary of Findings
(Cao et al., 2021)	The researchers circumnavigated Antarctica, taking 25 bioaerosol samples over coastal waters.	A high-volume air sampler and quartz filters. Sample duration was 24 hours.	Molecular 16S rRNA sequencing. Air mass back trajectory.	Proteobacteria dominated bioaerosols, including *Sphingomonas* spp.. They found low alpha diversity, high spatial and temporal variability but consistency in major groups collected over the study. They postulated immigration through long range atmospheric transport. Wind speed, temperature and organic carbon significantly affected the bacterial community. Richness and diversity did not significantly differ due to weather conditions.
(Trout-Haney, Heindel, & Virginia, 2020)	Greenland and the McMurdo Dry Valleys above freshwater and soil.	A pump and a dry filter with a sample duration of five days.	Epifluorescence microscopy on picocyanobacterial aerosols only to count the cells captured.	Picocyanobacterial cells in near surface air were present at a concentration of 2,431 to 28,355 cells m^{-3} air and no significant differences were detected between substrate or regions. Concentrations were lower than temperate areas.
(Archer et al., 2019)	McMurdo Dry Valleys	Soil sampling and air sampling with a Coriolis sampler and a RNAlater collection medium, as designed in Appendix A Method Development	Molecular 16S rRNA sequencing and ITS sequencing. Air mass back trajectory.	Proteobacteria, bacterioids and firmicutes dominated the bacterial bioaerosols, fungal bioaerosols contained yeasts and ascomycetes. These organisms are frequently spore formers and tolerant to extreme conditions. Bacterial bioaerosols showed marine influence. In ecological network analysis, bacterial communities were clustered by habitat, fungi clustered by geographic distance, indicating more dispersal limitations for fungi. No significant distance-decay detected. Nestedness analysis showed fungi were more nested, suggesting dispersal limitations. Intercontinental connectivity to the McMurdo Dry Valleys is limited, in line with previous modelling indicating infrequent microbiological exchange with the rest of the world.
(Weisleitner, Perras, Moissl-Eichinger, Andersen, & Sattler, 2019)	Lake Untersee,	A Coriolis took seven samples over three days. These samples were then pooled.	Molecular 16S rRNA sequencing.	Bioaerosols were dominated by firmicutes, proteobacteria and actinobacteria, while cyanobacteria only constituted 0.9% of the community. They found *Staphylococcus, Bacillus, Corynebacterium, Micrococcus, Streptococcus* and *Neisseria* and noted similar composition to Bottos et al. (2014) and other Antarctic bioaerosol studies.
(Kobayashi et al., 2016)	Hukuro Cove, Langhovde	Sampled for one hour near Adelie penguins on to 0.45 μm filter. One sample upwind and one downwind of the colony.	Molecular 16S rRNA sequencing. Air mass back trajectory.	19.4 times more *Bacillus* downwind of penguins from penguin faeces.
(Bottos et al., 2014)	McMurdo dry Valleys - Miers Valley	Air filtered on to 0.2 μm polycarbonate filters, total sample air volume 75,000 L at each	Molecular 16S rRNA sequencing. Air mass back trajectory.	Aerosols dominated by firmicutes suggesting volcanic activity. Most abundant taxa common to aerosols from other continents, representing a distinct widely dispersed bioaerosol community. Minimal marine input. Air masses originated from Antarctic Plateau. Some taxa in common with Halley

		of two locations. 1 m elevation. 55 day sample time, summer.		station work. Input likely from within Antarctica rather than intercontinental.
(Pearce et al., 2010)	Halley V Research Station	Two-week sampling in summer and winter with Hivol sampler with 0.2 µm polycarbonate filter.	Molecular 16S rRNA sequencing. Air mass back trajectory.	Low diversity, many sequence replicates and sequences from uncultivated organisms. No significant patterns detected between summer and winter. Few marine sequences irrespective of the distance to water. 1/3 sequences similar to those found in human studies indicating possible contamination of local environment from research station.
(Hughes, McCartney, Lachlan-Cope, & Pearce, 2004)	Rothera Point (Antarctic Peninsula).	Hivol sampler with 0.2 µm polycarbonate filter.	Molecular 16S rRNA sequencing. Air mass back trajectory.	Microorganisms, including cyanobacteria, actinomycetes, diatom plastids and other uncultivated bacterial groups were detected. Matches for microorganisms indicative of human contamination were not found. The closest matches were from Antarctic clones or from other cold environments. The majority of the sequences are likely to be of local origin. Back trajectory calculations showed that the sampled air may have travelled over the Antarctic Peninsula immediately prior to reaching the sample site, a proportion of the detected biota may be of non-local origin.
(Hughes, 2003)	Rothera Research Station	Exposed agar plates.	Plate counts.	Faecal coliform bacteria detected 75m downwind of the sewage outfall. Within one hour of deposition UV and desiccation kills most bacteria.
(Marshall, 1997)	Signey Island, South Orkney Islands	Rotorod Samplers. Three sites. 1 m and 0.15 m above ground level. Four rotorods at two separate 24 hour periods each week at all three sites for 14 months.	Microscopy.	Low concentrations of fungal spores in the air compared to the rest of the world. Concentrations increased in summer. *Chlamydospores* and *Vladospoirum* spp. were most and second most abundant spores respectively. Evidence of long-distance transport of spores.
(Marshall, 1996a)	Signey Island, South Orkney Islands	Rotorod Samplers. Three sites. 1 m and 0.15 m elevation. Four rotorods for two separate 24 hour periods each week at all three sites for two years.	Microscopy.	Lichen soredia most abundant bioaerosols. Dominance of soredia over ascospores decreases with more mature fell field sites. No correlation with temperature, humidity or wind speed. 1 m elevation not significantly different to ground level. Soredia peak in numbers after winter snow melt, demonstrating they are produced at sub-zero temperatures.

1.4 Microbial Survival Mechanisms in the Atmosphere

Microorganisms are highly resilient to environmental stressors, as shown by their successful colonisation of almost every niche explored by humans to date (Nuwer, 2014). The global atmosphere and terrestrial Antarctica are both extreme environments which share many survival challenges including high ultra violet (UV) irradiance, low temperature, low humidity and low nutrients (Womack et al., 2010). The extreme selective pressures experienced by microorganisms in the atmosphere should result in microbes that survive transport to Antarctica being well suited to colonise Antarctic ecosystems (Pearce et al., 2009). Dormancy (through spore formation) is thought to be the principle bioaerosol survival mechanism (Bottos et al., 2014; Pearce et al., 2016; Womack et al., 2010).

Spore formation: To survive transportation in the air, microorganisms can expend energy to counteract each stressor individually and remain metabolically active, or they can become dormant through spore formation, becoming resistant to all stressors with a single strategy. Species which become inactive therefore use the atmosphere as a vector only and are not considered to be residents of the aerial habitat. Spore formation is thought to be the most common mechanism for airborne survival, given that the majority of culturable isolates from high altitudes are spore-forming bacteria and fungi (Griffin, 2004; Smith, Griffin, McPeters, Ward, & Schuerger, 2011). Several authors have suggested that the air is dominated by Firmicutes (Bottos et al., 2014; Pearce et al., 2009), a bacterial phylum in which spore formers are common. Bacterial spores are small cells with a highly reduced cytoplasm and a tough outer coating, allowing them to remain dormant for millions of years and rapidly reactivate when conditions are appropriate for growth (Cano & Borucki, 1995). Their DNA is bound with various proteins to protect from UV, heat, cold, desiccation and any other stressors the organism might be likely to encounter (Lennon & Jones, 2011). Suspension of cellular metabolism protects against nutrient starvation and the toughened outer coating may protect from desiccation and UV (Lennon & Jones, 2011). Single-cell eukaryotes and fungi also form spores or cysts which operate on a similar principle. Small multicellular eukaryotes, like rotifers and tardigrades, undergo anhydrobiosis (desiccation of their bodies and significant reduction or suspension of their metabolism). Anhydrobiosis similarly facilitates survival in disadvantageous conditions and aerial dispersal (Nkem et al., 2006). However, multicellular organisms have a much more limited transport range due to their larger body sizes.

UV tolerance: UV is the limiting factor determining survival of aerial microbes (Smith et al., 2011) deactivating foreign microorganisms in Antarctic samples in under an hour (Hughes, 2003). UV intensity rises with altitude in the aerosphere and is elevated in Antarctica, due to its thin atmosphere and the localised depletion of the ozone layer (Pearce et al., 2009). UV radiation is highly damaging to most molecules in cells, but mediates its lethal impacts through DNA damage, which interrupts all cellular functions. UV causes harm to DNA in various ways, the most significant being cyclobutane pyrimidine dimers, where adjacent bases bind together. This erroneous binding causes destruction of normal base

pairing, resulting in mutation and distorts the DNA double helix, preventing genes from being transcribed (Sinha & Häder, 2002). As UV damage has such serious consequences, all organisms have developed defences against it. Pigmentation or carotenoids are commonly employed by microorganisms to absorb UV radiation before it can cause damage within a cell (Sinha & Häder, 2002). Isolates of culturable bacteria and fungi from the stratosphere often show pigmentation, indicating it could be an important factor for their survival (Smith et al., 2011; Womack et al., 2010). Bioaerosols generally exist as aggregates of particles, often including organic and inorganic matter bound together. Several authors have suggested that this might protect against UV radiation, as could persistence within clouds (Burrows et al., 2009b; Pearce et al., 2009; Womack et al., 2010). Microorganisms in challenging environments often grow in biofilms, which are layered communities of microbes that likely provide protection from UV in a similar fashion (Pointing et al., 2015). DNA repair mechanisms, which correct UV induced damage, are ubiquitous in living organisms. They include use of enzymes such as photolyases, which harvest energy from light to repair damaged DNA. Photolyases have been found in UVC-resistant Antarctic microorganisms (Marizcurrena et al., 2017). Excision repair is another very common pathway, which works by cutting out the damaged portion of DNA and resynthesizing it using the complementary strand as a template (Sinha & Häder, 2002). Some microorganisms such as *Deinococcu*s have highly efficient DNA repair mechanisms, which confer extreme resistance to UV radiation (Pepper, 2015).

Cold resistance: Most organisms have growth optima well above temperatures commonly experienced in Antarctica and the aerosphere. With increased altitude temperature drops to below zero towards the tropopause, to around -60 ^{o}C, before recovering to approximately zero in the stratosphere (NASA, 1962). Antarctic winter temperatures become as low as -93.2 °C and summer temperatures reach up to 15 °C, with a summer mean of around zero (Pointing et al., 2015). Below-freezing temperatures cause ice crystal formation on cell surfaces and slowing of metabolic processes, which can either kill cells or severely limit their growth rates (Pepper, 2015). Psychrophilic (cold-loving) microorganisms grow successfully down to -18 °C (Rothschild & Mancinelli, 2001) and various adaptions allow these organisms to remain metabolically active. Ratios of unsaturated to saturated fatty acids in cell membranes can be increased to counteract reductions in membrane fluidity at lower temperatures. Some organisms have developed enzymes with optimal activity at lower temperatures and some employ antifreeze proteins, which help prevent crystallization (Laybourn-Parry, 2002). Cold-tolerant organisms can be isolated from bacterial communities in temperate environments (Wilson & Walker, 2010), indicating psychrotolerant organisms likely originated outside Antarctica, transported from warmer clines. Although there has been no evidence of airborne microorganisms employing these mechanisms to tolerate cold environments, with increasing molecular studies of bioaerosols, similar survival mechanisms are likely to be detected.

Low humidity tolerance: While microorganisms differ in their responses to changes in humidity, very low humidity seems universally intolerable (Pepper, 2015). Both Antarctic and high-atmospheric relative humidity levels are low, with desiccation-tolerant organisms frequently found in both environments (Luhung et al., 2015; Pearce et al., 2009). Low humidity causes damage to the lipid bilayers in the cell membrane and can change the membrane from a crystalline structure to a gel structure. This affects configuration of cell surface proteins, interrupting their function and deactivating the cell (Pepper, 2015). Viral survival is also thought to vary in a way which is dependent on humidity, with encapsulated viruses showing better resilience to lower humidity (Mohr, 2007). As with protection from UV, associations with clouds or other aerosol particles are thought to help protect against low humidity (Pearce et al., 2009; Womack et al., 2010). Biofilms with extracellular matrices adhering the organisms are also thought to reduce water loss (Pointing et al., 2015). Gram-positive cells tend to be more resilient to desiccation (Mohr, 2007), thought to be due to a number of factors such as a thicker peptidoglycan layer, ability to accumulate magnesium and linkages to radiation tolerance (probably due to strong DNA repair capability conferring protection against desiccation as well as UV) (Makarova et al., 2001). Accordingly, the majority of bacterial bioaerosols found to date are Gram-positive (Bottos et al., 2014; Griffin, 2004; Smith et al., 2011). However, this observation could be confounded by sample bias noted against Gram-negative bacteria (Luhung et al., 2015), or the fact that Gram-positive bacteria are also frequently spore formers.

Oligotrophy resistance: Another shared characteristic and challenge to survival, in the aerosphere and Antarctica generally, is low nutrient concentrations. The atmosphere is generally assumed to be lacking in the nutrients required by microorganisms, however key nutrients for microbial survival such as carbon, sulphate and nitrate can be found in cloud water, at similar levels to lake water (Pearce et al., 2009). Bauer et al. (2002) observed carbonaceous material constituted up to 20% of total aerosol mass in atmospheric aerosol samples. In addition to metabolising nutrients in the atmosphere, photosynthetic microbes that can independently fix carbon have been found in air samples, such as cyanobacteria and Chloroflexi (Brodie, DeSantis, & Parker, 2007a). Aerosolised bacteria can multiply and metabolise organic compounds generally present in clouds, some even at super cooled temperatures (Dimmick, Straat, Wolochow, & Levin, 1975; Dimmick & Wolochow, 1979; Sattler et al., 2001) . Vaïtilingom and Deguillaume (2013) and Amato et al. (2007) showed microorganisms are capable of degrading formaldehyde and carboxylic acids, carbon sources that are often present in cloud water. Current bioaerosol work is insufficient to determine the extent of metabolism and division in the atmosphere and what portion of the bioaerosol community typically remains metabolically active. It is assumed that there is unlikely to be significant reproduction of bacteria within clouds (Burrows et al., 2009a), given that most bioaerosols likely only spend a tiny fraction of their time suspended within cloud droplets (Lelieveld & Heintzenberg, 1992). The ability to remain metabolically active is the key differentiator between dormant microbes using the atmosphere as a vector and active microbes using the air as a habitat.

1.5 Impact of Aerial Dispersal on Antarctic Microbial Populations

Long-range aerial transportation of microorganisms has long been suspected to play a significant role in structuring the Antarctic biological community (Horowitz, Cameron, & Hubbard, 1972). We have suggested that the air is likely to be the dominant vector for microbes to Antarctica (see section 1.3 The Aerosphere and Bioaerosol Particles) and that microorganisms transported via the atmosphere are likely to be preselected to colonise Antarctica, with sporulation the likely main survival mechanism (see section 1.4 Microbial Survival Mechanisms in the Atmosphere). Most Antarctic environments have a relatively low turnover rate (Cary, McDonald, Barrett, & Cowan, 2010) so aerially deposited microorganisms are unlikely to be consumed by the resident community. This implies that a large pool of atmospherically vectored temperate organisms is present in Antarctica, which will come out of dormancy when conditions become favourable under continued global warming (Kussell, Kishony, Balaban, & Leibler, 2005). Therefore it is likely that atmospherically vectored microbes have the potential to substantially impact Antarctic ecosystems and understanding the potential impacts from this shifting microbial landscape to the continent is paramount.

1.5.1 The Propagule Bank

The propagule bank describes a reservoir of dormant microorganisms awaiting favourable conditions to reactivate (Wynn-Williams, 1991). It represents the adaptive potential of an ecosystem, allowing microbial communities to rapidly respond to change. Up to 80% of microbial cells in soil are estimated to be in a dormant state (Lennon & Jones, 2011), with no published equivalent figure for the atmosphere. The presence of many dormant organisms represents a significant risk to the existing microbial community, which could change drastically, resulting in biodiversity loss (Lennon & Jones, 2011). In Antarctica, aerially deposited microbes are thought to comprise a large proportion of the propagule bank. An example of this is found at Lake Vostok (**Figure 1-11**), located below the central East Antarctic ice sheet and named for the Russian research station on its surface. Here, viable microorganisms were recovered at different depths. The lake itself is sealed under the ice sheet and remains liquid due to the enormous pressure from the weight of ice above (Vincent, 2000). Ice cores collected from above the lake contained a continuous chronology of microbial deposition since the lake was covered around 400,000 years ago (Sinha & Krishnan, 2013). From the ice cores, viable propagules were found, with culturable yeasts, fungi and bacteria present up to around 3,000 years old. At around 10,000 years old, microbial communities became dominated by spore formers (Vincent, 2000). At around 3.6km depth, ice was derived from the underlying lake water and contains viable bacteria (Karl et al., 1999). In 1998 (around the time **Figure 1-11** was published), drilling was temporarily halted due to concerns about contaminating the pristine ecosystem. In 2012 drilling was completed and sampling revealed unique microbial life, however concerns remained regarding contamination from the drilling process (Bulat, 2016). It is speculated

that as Antarctic ice melts, a multitude of ancient organisms could reactivate from dormancy (Fox-Skelly, 2017), and alter microbial community composition.

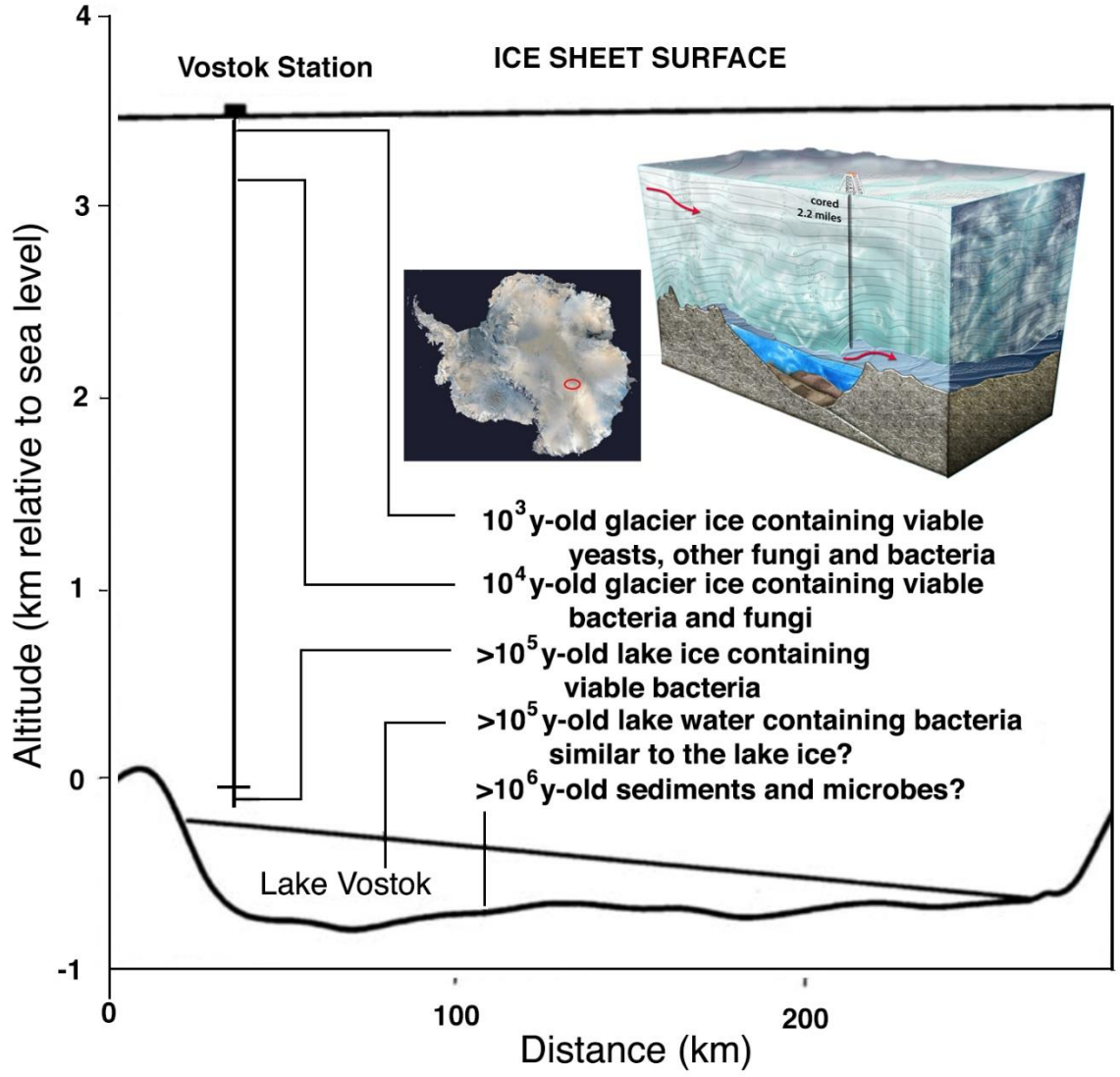


Figure 1-11 - Different types of microorganisms present in cross section of ice layers above Lake Vostok. Adapted from Vincent (2000) by Chris King.

To model changes in ecosystems from the revival of dormant species, several cloche experiments have been performed in Antarctica (Convey & Wynn-Williams, 2002; Kennedy, 1994; Smith, 1991; Smith, 1994; Wynn-Williams, 1996). A cloche is a cover placed on the ground designed to warm the ground beneath and, sometimes, to reduce UV exposure or increase humidity. Although this changes multiple variables simultaneously in an unnatural manner, these experiments provide insight into the potential effects of climate change. Such experiments have shown similar rapid increase in biodiversity and abundance of nematode worms, cyanobacteria, bryophytes and microarthropods (Convey & Wynn-Williams, 2002; Kennedy, 1994; Smith, 1991; Smith, 1994; Wynn-Williams, 1996). The increase in abundance of particular species can be driven by increased reproduction of

existing populations. However, increased diversity can only be driven by activation and reproduction of dormant organisms in the propagule bank. The rapid increase in abundance and diversity in these cloche experiments demonstrated that substantial changes can occur quickly, and furthermore display both the latent potential and the speed at which the propagule bank can respond to even slight increases in average temperature. The outcome of these complex cascading ecological interactions could have rapid and fundamental effects on Antarctic ecosystems.

1.5.2 Future Changes to Antarctica

It is highly likely that climate change will induce substantial changes to Antarctic microbial communities, fuelled by the propagule bank and aerial input (Convey & Wynn-Williams, 2002; Cowan et al., 2011). However, current understanding about the rate and direction of change in microbial communities is limited due to lack of studies in this field (Kennicutt et al., 2014). The warming of Antarctica is providing a more hospitable environment to temperate species. The Antarctic Peninsula has experienced an average temperature increase of 3 °C in the last 50 years (Turner et al., 2005) which has led to an increase in free water availability, an extension of the growing season (Convey, 2006) and precipitation being more frequently observed as rain rather than snow (Pearce et al., 2009). These changing conditions are believed to be responsible for the increased range and abundance of the only two known native vascular plants in Antarctica over the last 25 years (Fowbert & Smith, 1994).

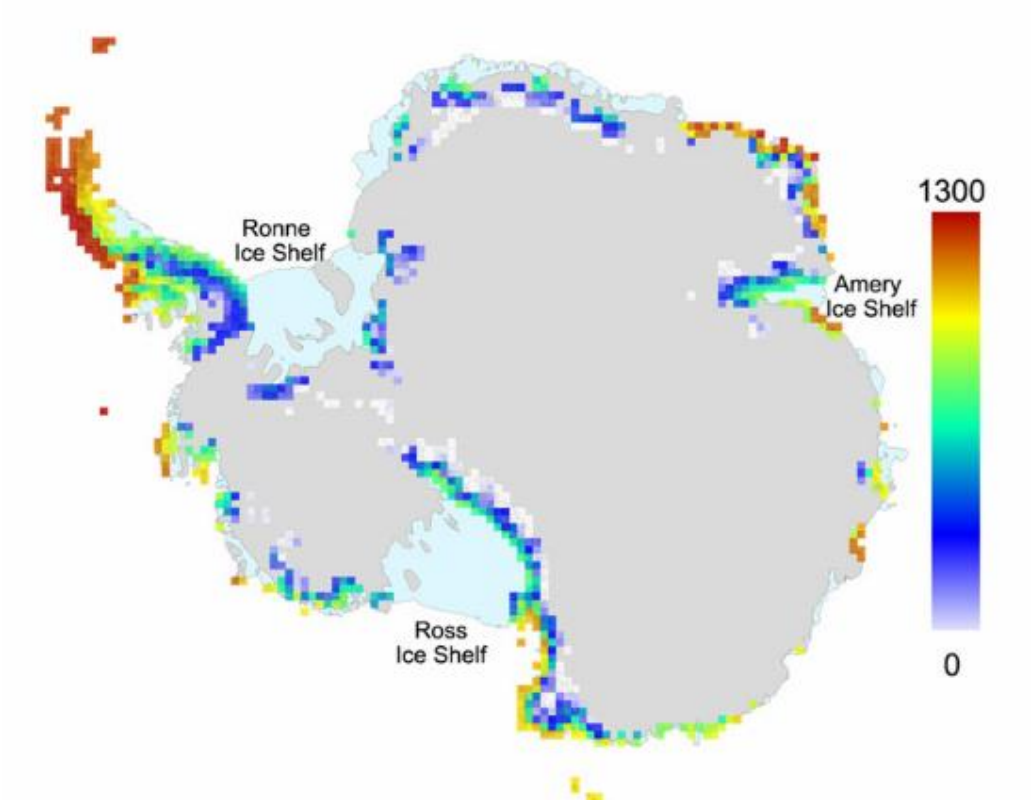


Figure 1-12 - Projected increase in annual cumulative degree days, measuring increase in growing season, between 2007 and 2100 (Chown et al., 2012), indicating risk of alien species establishment.

The Peninsula acts as an early warning indicator to forecast impacts from climate change for the rest of the continent, as it experiences more temperate conditions and is the least isolated part of Antarctica. In the near future, the Antarctic continent is expected to experience widespread cooling, followed by warming as the ozone hole recovers. Generally, coastal or ice free areas are expected to warm first, as they are most exposed to a warming maritime influence, and are less affected by glacial cooling (**Figure 1-12**) (Chown et al., 2012). Observed changes and cloche experiments indicate increase in ground cover and surface greening will occur, as the conditions become more amenable to plants. Lichens and cyanobacteria will likely grow more optimally, given many are psychrotolerant rather than psychrophilic. These changes could create positive feedback effects on atmospheric carbon levels and climate change (Pointing et al., 2015). A green Antarctica reflects less light, and increased carbon fixing from higher photosynthetic rates may be more than offset by increased release of carbon in soils, through greater decomposition (Pointing et al., 2015). Antarctica's future changes continue to make it increasingly hospitable to temperate organisms, therefore dormant bioaerosols originating in nearby regions are likely to activate, as are other temperate microbes previously deposited in soil and water via the atmosphere and other vectors.

1.6 Study of Bioaerosols: Filling the Gaps

This chapter has outlined the evidence for and mechanisms of airborne transportation of microorganisms. The limited studies to date have confirmed speculation that bioaerosols are extremely variable. Although most microorganisms likely originate from local species pools, at least some originate from foreign areas. Given the sensitivity of global biomes to non-local organisms, addressing key knowledge gaps of airborne transportation is essential to understand future responses to climate change. In the future we hope that technological advancements will allow studies to shed more light onto the extent, sources (local or intercontinental), survival mechanisms and significance of airborne microorganisms to global ecosystem structuring. With the technological advancements made in the past few years, we have an unprecedented opportunity to resolve these core biological questions. However, to understand the underlying physical processes discussed in this chapter that drive this transportation, interdisciplinary studies must be conducted. The foci of the remainder of the thesis are urban bioaerosols, as so many people worldwide dwell in cities, and Antarctic bioaerosols, due to the fragility, uniqueness and isolation of Antarctica.

1.7 Thesis Objectives

This thesis had three main objectives:

1.7.1 Quantify Spatiotemporal Bioaerosol Variability in Urban Parks in Auckland, Aotearoa New Zealand

The aim of this research was to understand spatiotemporal variation in bioaerosol communities across city parks, and to determine if it is possible to predict community variation based on environmental variation. Nine parks in Auckland's metropolitan area were repeatedly sampled in autumn and winter to quantify the relative contributions of environmental variables to fungal and bacterial bioaerosol communities and understand bioaerosol exposure in urban parks at a city-wide spatial scale. This is important to understand the microbial exposure of people visiting city parks and to begin to predict and model how this might change over space and time.

1.7.2 Characterise Seasonal Variability of Bioaerosols in Aotearoa New Zealand over a Two-year Period

The aim of this research was to determine if there is detectable seasonal variation in urban bioaerosols and whether this variation was predictable based on measurable environmental variables. Air samples were collected in the centre of Auckland, every week for two years, to understand seasonal variation of bacterial and fungal bioaerosols. This sampling represents the longest continuous urban bioaerosol study to date and is vital for shedding light on the exposure of city dwellers to airborne pathogens and allergens. The relative contributions of different environmental factors to bioaerosol diversity were quantified and insights were gained into temporal dynamics of urban bioaerosols in a sub-tropical climate.

1.7.3 Quantify Diel Variation and Test for Intercontinental Connectivity of Antarctic Bioaerosols

The aims of this research were, firstly, to determine if there is detectable diel variation in Antarctic bioaerosols and if this can be predicted based on environmental variation, and secondly, if connectivity between bioaerosol communities is detectable at very large spatial scales. Sampling of ambient air was performed in the McMurdo Dry Valleys of Antarctica, in two-hourly increments for several days to quantify temporal variation of Antarctic bioaerosols at a fine scale. These samples were compared to those collected at Baring Head Lighthouse, near Wellington in Aotearoa New Zealand, to assess levels of intercontinental bioaerosol exchange between Antarctica and the rest of the world. This research is crucial to understanding bioaerosols in Antarctica's unique and vulnerable environment and predicting ecosystem changes in a warming world.

The general discussion synthesises the main themes and findings across studies and discusses potential future directions for research. Aerobiology is a relatively new sub-field within microbial ecology. Due to this, significant effort was put into developing the protocols followed throughout this thesis. Various sample collection protocols, sample pre-processing, DNA extraction protocols and bioinformatics pipelines were compared for their ability to operate in temperate and cold environments, the quantity of DNA recovered, and the quality of data generated for further analysis. This work provided methods that dramatically reduced sampling times in cold environments and increased assurance over bioaerosol community data generated from bioinformatic analysis. The thesis finishes with supplementary information which pertains to the previous experimental chapters and the reference list.

Chapter 2 - Spatiotemporal Bioaerosol Variability in Urban Parks

2.1 Abstract

The atmosphere provides a conduit to almost all worldwide ecosystems for microorganisms that can exploit it, including invasive species, and organisms that can be pathogenic to humans. Understanding ecosystem connectivity via the atmosphere is crucial to understanding worldwide ecosystem structuring and for predicting changes which may occur in a warming world. Bioaerosols in urban parks have been subjected to almost no molecular interrogation, despite the fact they are poorly understood and are known to harbour pathogens and allergens. Here, the spatiotemporal variation and relative importance of putative drivers of bioaerosol variation were quantified for the first time. Nine parks in Auckland, New Zealand were sampled in multiple blocks between July 2017 and June 2018 with a Coriolis liquid cyclone air sampler, and their fungal and prokaryotic (bacterial and archaeal) communities assessed through internal transcribed spacer (ITS) and 16S rRNA (16S) gene amplicon sequencing. Bioaerosols in urban parks varied as a result of many factors; in particular, location, time and sampled air-mass source. Modelling was able to explain 38% of the fungal variation and 19% of the bacterial variation. Time was more important for fungi than bacteria. Likely due to their ease of dispersal, bioaerosol communities did not show typical microbial distance-dissimilarity patterns, but did display local differentiation, probably due to differences in local sources. This work sheds important light on the scales of variation in bioaerosol communities and the relative contributions of measurable environmental variables to bioaerosol community structure and microbial exposure of city-dwellers.

2.2 Introduction

The aerosphere has been identified as the last unexplored biome on Earth (Pointing, Fierer, Smith, Steinberg, & Wiedmann, 2016) and contains a diverse microbiological component including bacteria, fungi and viruses with putative impacts on atmospheric chemistry, cloud formation, human health and biogeography (Kellogg & Griffin, 2006; Smith et al., 2011; Womack et al., 2010). Very little is known, however, about the dispersal and connectivity of airborne microbial populations. Bioaerosols vary markedly over space and time (Burrows et al., 2009b), but information is lacking on the relative importance of the factors that drive this spatiotemporal variation, and scales over which they operate (Archer et al., 2020; Womack et al., 2010). Addressing this knowledge gap is important as many bacterial and fungal bioaerosols have the potential to be allergens or pathogens (Yoo et al., 2016), or are environmental organisms that affect the ecosystems in which we reside (Burrows et al., 2009b). Understanding movement of aerial microorganisms informs the potential for connectivity among microbial communities in disparate ecosystems. Urban parks represent a tractable opportunity to understand bioaerosols in the urban environment, to which millions of people worldwide are exposed. Recent evidence suggests that exposure to greater microbial diversity has a multitude of health benefits (Mhuireach et al., 2020). Parks are consistently managed ecosystems, spatially discrete, similar in environmental conditions

and numerous, providing natural replication of data points within the same urban area. This ease of replication provides the opportunity to explore the relative effects of factors which are predicted to be important based on previous aerosol research (Burrows et al., 2009b; Pearce et al., 2016) or previous research in other microbial communities such as soil or water (Carini et al., 2020; Lear et al., 2013). These factors include elements in the local environment that emit bioaerosols, meteorological conditions and the effect of wind direction. Further, parks in urban areas represent fine scale land-use changes and thus provide a chance to extend our knowledge on effects of land use patterns on bioaerosols at a coarser scale.

Bioaerosol concentrations and composition vary with different land-use types or biomes (Bowers et al., 2011a; Burrows et al., 2009a). Earlier research focussed on cultivatable numbers of airborne microbes. Bioaerosol concentrations appeared to be highest over more productive areas, such as grassland or crops, and in cities, and lowest over deserts, ice and seas (Burrows et al., 2009b; Harrison et al., 2005; Shaffer & Lighthart, 1997). Very little research has been done on city parks specifically. Fang et al. (2007b) noted lower concentrations of culturable bioaerosols in the single Beijing park sampled compared to surrounding urban areas. This was postulated to be due to fewer vehicle movements aerosolising microorganisms. In the UK, three urban parks were subjected to metabolite fingerprinting (Garcia-Alcega et al., 2020). The results showed seasonal differentiation in microbial communities, and the bacterial genera *Bacillus* and *Pseudomonas* and *Penicillium* fungus, frequently detected in bioaerosols, were identified (Garcia-Alcega et al., 2020). Later, molecular approaches continued to improve understanding of urban bioaerosols. Airborne bacteria in two urban green space areas in Eugene, Oregon were characterised with high throughput metabarcode sequencing (Mhuireach et al., 2020). A core aero-microbiome of plant and soil-associated genera was observed, and the forested site was more diverse than the grass-covered one. Seasonal and other site-specific effects (site explained 41% of the bioaerosol variance) were also detected (Mhuireach et al., 2020). Barberán et al. (2015) showed geographical differentiation of fungal and bacterial species detected in dust gathered from the outside of houses across the US. They reported no distinct urban versus rural community or effect on diversity, but that urban samples were more similar to each other than rural ones. A recent study in Japan, however, reported greater alpha diversity of taxa at suburban compared to urban sites (Tanaka et al., 2020). Some urban molecular bioaerosol studies report diversity comparable to soil (Brodie et al., 2007b). Woo et al. (2013) sampled along an urbanisation gradient in Hong Kong and did not find a significant relationship between bioaerosols and urbanisation. As the number of studies was so limited, and the majority of environmental microbes do not grow in culture (Pearce et al., 2009), there is a need to apply molecular methods to get a more complete picture of bioaerosol diversity in city parks. This study is the first molecular characterisation of fungal bioaerosols (and only the second for bacteria) in urban parks. The number of parks sampled (nine) is well above most other comparable studies, helping to address the knowledge gaps apparent in the literature. No Southern Hemisphere parks appear to have

been studied, so this work is also the first to gather data in this region, and in a sub-tropical climate.

Seasonality has repeatedly been demonstrated to affect bioaerosols. There tends to be higher concentrations of culturable microbes in warmer and wetter periods, but this pattern is not universal (Burrows et al., 2009b; Lighthart & Stetzenbach, 1994; Woo et al., 2013). Long-term observations in the Northern Hemisphere showed higher concentrations of culturable microorganisms in autumn and lower concentrations in winter (Burrows et al., 2009b). Every molecular study of bioaerosol seasonality reviewed detected seasonal variation and, frequently, a relationship with temperature (see section 3.2 Introduction). Higher summer concentrations of bioaerosols are posited to be due to warmer temperatures, increasing microbial reproduction and aerosolisation (see section 1.3.2 Bioaerosol Particles: Temporal and Spatial Variation).

Air-mass source is thought to be another key driver of bioaerosol communities. In January 2020, the New Zealand sky turned sepia, when smoke from the Australian bushfires was carried thousands of kilometres across the Tasman Sea. This observation is supported by several recent molecular studies indicating that air-mass source is a driver of the aero-microbiome (Archer et al., 2020; Maki et al., 2017; Woo et al., 2013). Bottos et al. (2014) detected no local ocean bioaerosol influence in their Antarctic sampling but did observe influence from nearby volcanic activity. Correlation is thought to occur between time of year and air mass origin (Woo et al., 2013). Woo et al. (2013) detected seasonal patterns in alpha diversity and temperature effects on diversity as estimated using the UniFrac distance metric. The driver for the observed seasonal patterns was postulated to be seasonal changes in air mass origin from continental to marine.

Little research has addressed the relevant temporal scales for the influence of wind on bioaerosol concentration and community structure. The source of sampled air can be estimated using back-trajectory analysis. Three-day back-trajectories (revealing the movements of sampled air for the previous three days) are normally used as predictors of bioaerosol variation in the literature (Archer et al., 2020; Archer et al., 2019). It is untested, to date, if three days is the best predictor of variation in bioaerosol community structure. Bacteria have been simulated to remain airborne for a mean of one week, with substantial variation either way (Burrows et al., 2009a), so it is possible another trajectory length may be more informative.

A further factor thought to influence aerial microbial populations is weather. For instance, rain is thought to increase deposition of suspended microorganisms and reduce concentrations of bioaerosols (Burrows et al., 2009b). Species such as *Pseudomonas,* which possess ice nucleation proteins, are thought to be particularly affected by patterns of precipitation. Other meteorological variables such as temperature, humidity and turbulence also impact bioaerosol communities (also see section 1.3.2 Bioaerosol Particles: Temporal and Spatial Variation and section 3.2 Introduction).

The aim of this chapter is to use bioaerosol communities in urban parks within Auckland, in Aotearoa New Zealand to ask two research questions: (1) What is the relationship between temporal scale of wind back-trajectories and variation in bioaerosol community structure? (2) What are the effects of environmental variation among different parks on bioaerosol community structure and population densities? To characterise community structure, bacteria/archaea were identified using 16S rRNA gene sequencing and fungi using sequencing of the ITS gene.

2.2.1 Hypotheses

1. Temporal wind hypothesis: Air-mass trajectories of differing spatiotemporal lengths differ in their ability to predict variation in bioaerosol community structure. Specifically, one-week trajectories provide the greatest explanatory power, as indicated by modelling.
2. Environmental drivers hypothesis:
    a. Spatial variation in bioaerosol community structure (composition and diversity) at Auckland parks can be predicted by time, geographic origin of sampled air, altitude, distance to sea, temperature, relative humidity and park size (small, medium, large).
    b. Relative abundances of common and rarer taxa fluctuate by location, time, and air mass source.
3. Distance-dissimilarity hypothesis: Similarity in community composition of parks declines with their increasing geographic distance.

2.3 Methods

2.3.1 Field Sampling and Environmental Data Collection

Nine Auckland parks were selected for sampling (**Figure 2-1**). The Auckland parks were predominantly grassland, with varying numbers of deciduous and coniferous trees, and subtropical ferns. Some parks had bodies of water in them, such as Western Springs. The parks had differing numbers of pedestrian and vehicular movements. Livestock, such as sheep and cattle, were kept in some parks. The presence of birds and dogs was also variable among parks. The parks differed markedly in area and altitude, with several Auckland parks encompassing volcanic summits.

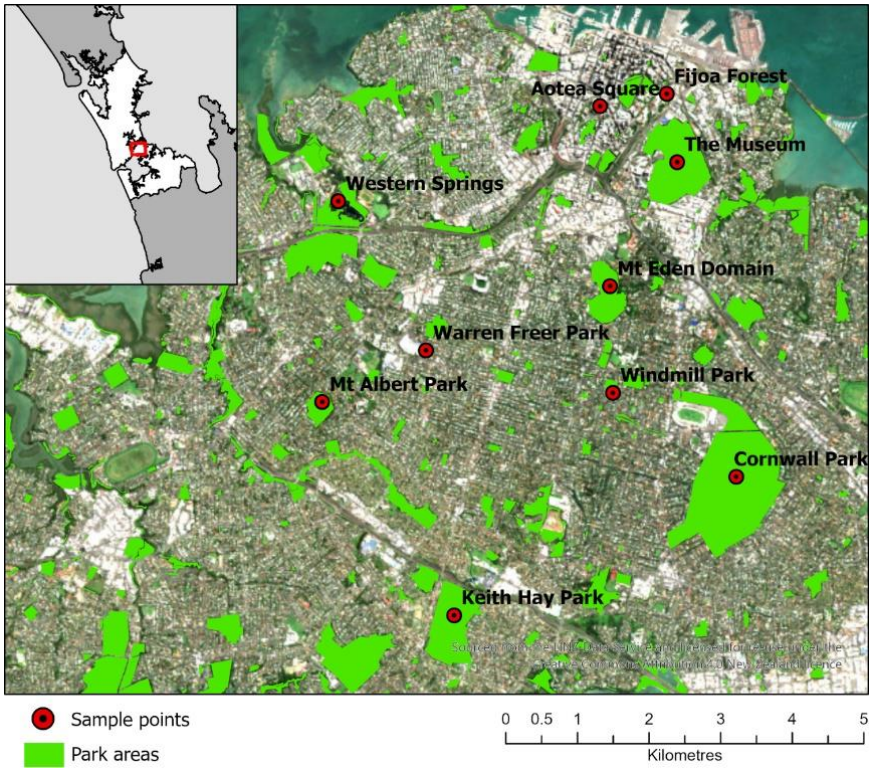


Figure 2-1 – Map of Auckland metropolitan area with the nine urban parks selected for sampling indicated. Aotea Square was a control sampling location.

Each location was visited in a randomised order between 10 am and 4 pm to account for any diurnal effects. The sample was taken in the same location each time, as close to the centre of the park as practicable and at minimal elevation (avoiding effects from change in altitude or being close to an edge confounding results). Each park was visited three times (to provide sample replication) in the first sampling window (12 July 2017 to 14 September 2017) and three times in the second sampling window (20 March 2018 to 1 June 2018). This was to replicate the sampling within two seasons, so that temporal differences could be characterised. The dates of visits were selected based on practical considerations for performing the sampling, such as traffic at particular times and availability of the sample locations on different dates. The timing and order of visits was varied to control for any unforeseen effects related to sampling order or timing. The exact location was saved on Google Maps (Google) and a photo was taken of the Coriolis (see section A.1 Introduction) and location. Sampling was performed 1.8 m above ground level. Gloves were worn and the extender and Coriolis unit cleaned with bleach. The Coriolis neck, head and cone were cleaned with bleach, ethanol and three rinses of milli-Q H_2O (MQH_2O). The cone was filled with 15 mL of phosphate buffered saline (PBS). A negative was taken (PBS put into the cone without running the Coriolis) before running the Coriolis for two – four minutes with MQH_2O to ensure all bleach residue in the head and neck was removed. The MQH_2O was discarded and replaced with 15 mL of PBS and the Coriolis was run at 300 L/m for one hour. The PBS in the cone was topped up to 15 mL after 30 minutes and sampling was completed with 10 mL PBS remaining in the cone. Samples were transferred into a 15 mL falcon tube,

transported in an insulated box with ice blocks and stored at -20 °C within four hours. During sampling, observations were recorded of weather conditions (with a Kestrel 3000) and particle counts were taken (with an AeroTrak particle counter). If rain occurred the Coriolis was sheltered with an umbrella. If rain was heavy the Coriolis was packed up until rain abated.

Air-mass back-trajectories were generated hourly over the sampling periods using the NOAA HYSPLIT model (v5. 0. 0 Ubuntu), (Stein et al., 2016) with GDAS meteorological data. Clustering was performed in HYSPLIT (see Appendix B HYSPLIT Clustering Procedure for details). The number of clusters selected (six) was based on a marked increase in total spatial variance as clusters reduced. The cluster for each sample was manually entered into the metadata, which was imported into R. Back-trajectories were generated for one-day, three-day and one-week durations.

2.3.2 Laboratory Processing

For details on DNA extraction and DNA sequencing methods see section C.1 Laboratory Methods – DNA Extraction, and section C.3 Laboratory Methods – DNA Sequencing.

2.3.3 Bioinformatics

For details on bioinformatic processing see Appendix D Optimising Bioinformatics Protocols for Aerosol Microbial Community Data – a Case Study Using an Urban Parks Dataset. The outputs, the decontaminated amplicon sequence variant (ASV) table (with read counts adjusted to remove contaminant sequences) and combined taxonomy and relevant metadata were further analysed in R with respect to the hypotheses for this study.

2.3.4 Data Analysis

The 16S (the gene for the RNA component of the 30S small subunit of the prokaryotic ribosome) and ITS (spacer DNA situated between the small-subunit RNA and large-subunit ribosomal RNA genes in the fungal genome) ASV tables were prepared for analysis in two ways: (1) with all ASVs (unfiltered ASV tables) and (2) with a filter applied for only ASVs with more than either 100 reads (for 16S) or 500 reads (for ITS) and a coefficient of variation of greater than three (filtered ASV tables). These filters were used to improve any signal which would be masked by invariant taxa and/or low-level stochastic variation. Read counts for ITS were higher, hence the greater stringency of the filter applied.

Next generation sequencing (NGS) data are inherently ‘compositional’ (Gloor, Macklaim, Pawlowsky-Glahn, & Egozcue, 2017), meaning that read counts generated are quantitative descriptions of part of a whole and only reflect relative, not absolute, abundances. Compositionality in NGS data arises because sequencing machines have limited slots and once saturated cannot count further reads. Even with low biomass bioaerosol samples, other steps in the sequencing process including PCR amplification, standardisation and stochastic variation in the read depth for different samples, mean that raw read count is not an accurate measure of microbial abundance. Various problems occur when non-

compositional tools are used on compositional data. Distance matrices are confounded by differences in read depth, spurious correlation upon subsetting occurs and relative abundance measures show a high false-positive rate (Gloor et al., 2017). Compositional tools typically use a log transformation (Aitchison, 1982). Since the log of zero is undefined, zeros in the ASV table present a problem. To address this, Bayesian multiplicative replacement was performed (Gloor et al., 2017; Quinn et al., 2019) and implemented using the function cmultRepl (Aitchison, 1982; Templ, Hron, & Filzmoser, 2011) in the R package, "zCompositions" (v1. 3. 4) (Palarea-Albaladejo & Martín-Fernández, 2015). "RobCompositions" (v2. 2. 1) (Templ et al., 2011) was used to generate the Aitchison distance matrix, which was fed into downstream analysis functions. Hill numbers, D0, D1 and D2, were calculated using the R package "hillR" (v0. 5. 0) (Li, 2018; Oksanen et al., 2012) based on the unfiltered ASV table to quantify alpha diversity. A Bray-Curtis (BC) distance matrix, Hellinger-transformed ASV table and Jaccard presence/absence matrix were also generated using the R package "vegan". Results from non-compositional data analysis were generally consistent with the results from compositional tools, and so are not presented in the main text (see section E.1.3 Results from Non-Compositional Data Analysis for details).

Variance partitioning of the Aitchison distance-based redundancy analysis (db-RDA) of the compositional data using the three different lengths of back trajectories as explanatory variables was used to determine which back trajectory length would be used for the remainder of the analysis. Variance partitioning provided quantification of the relative importance of alternative trajectory lengths to the sampled bioaerosol communities. For these purposes, a higher R^2 indicated better correlation of a trajectory. Variance partitioning was also performed on an RDA of the matrix of Hill numbers related to each sample using the three different back trajectory lengths as explanatory variables. Hill numbers calculated were ${}^{q}D$ where D equals 0 or D0 (raw diversity, i.e. the number of ASVs), D1 (exponential of Shannon's entropy index) and D2 (inverse of Simpson's concentration index) (Chao, Chiu, & Jost, 2016).

Db-RDA with variance partitioning was used to quantify the relative importance of measured variables for predicting bioaerosol variation, including time, wind direction, temperature, relative humidity, distance to sea, altitude, and park size (small, medium or large) and a Euclidean geographical distance matrix to account for any effects of spatial proximity of parks, using the UTM coordinates of each sample site (**Table 2-1**). In addition to quantifying the relative importance of the different variables, this procedure quantified the portion of explained variation that is shared among variables. Firstly, forward selection of variables was performed to reduce the number of explanatory variables in each db-RDA analysis. This was implemented in R using the capscale and ordistep functions in "vegan" (v2. 5. 6) (Oksanen et al., 2012). Variables were included with a *P*-value of 0.01 or lower. Forward selection was run with 999 permutations. Selected variables were included in the db-RDA with variance partitioning (varpart function in "vegan") alongside location and time, which were included to account for spatiotemporal autocorrelation. All numerical variables

were standardised as *Z*-scores prior to analysis so that any effect sizes would be comparable.

Table 2-1 – Summary table of variables showing their inclusion in the db-RDA

Variable	Variable type	Description	Included in 16S model?	Included in ITS model?
1-day back trajectory cluster	Categorical	Path of sampled air-mass in previous 24 hours	Yes	Yes
3-day back trajectory cluster	Categorical	Path of sampled air-mass in previous 72 hours	No	No
1-week back trajectory cluster	Categorical	Path of sampled air-mass in previous 168 hours	No	No
Time	Categorical	Time of year during which sampling took place	Yes	Yes
Location	Numerical	Euclidean distance matrix of UTM coordinates of park sampled	Yes	Yes
Weather	Categorical	Weather during sampling	No	No
Temperature	Numerical	Temperature during sampling	No	No
Humidity	Numerical	Humidity during sampling	No	Yes
Size	Categorical	Park area – small, medium or large	Yes	Yes
Elevation	Numerical	Elevation above sea level of sampling location	No	Yes
Distance to sea	Numerical	Distance from sampling location to sea	No	Yes

Further visualisation of compositional variation among samples was achieved by generating Non-metric Multidimensional Scaling (NMDS) plots using the package "vegan" and the Aitchison distance matrix. The stress for two to six dimensions was compared, with 500 random starts and 999 iterations per run. The lowest dimensional solution with a stress under 0.2 was selected (McCune, Grace, & Urban, 2002). The function stat-ellipse in "ggplot2" (v3.3.2) (Wickham, 2016) was used to create ellipses, assuming a multivariate *t*-distribution. To assess patterns by location, time and trajectory for different taxa, differences in relative read counts were quantified to generate taxon bar plots using "ggplot2" and "phyloseq" (v1.30.0) (McMurdie & Holmes, 2013). The percentage of total read count was calculated on data filtered as follows. For 16S, samples and ASVs with less

than 100 reads were filtered out. Relative abundance by genus for each category included only genera with at least 0.5% of the reads. For ITS, samples with less than 100 reads and ASVs with less than 500 reads were filtered out. Relative abundance by genus for each category included only genera with at least 1% of the reads.

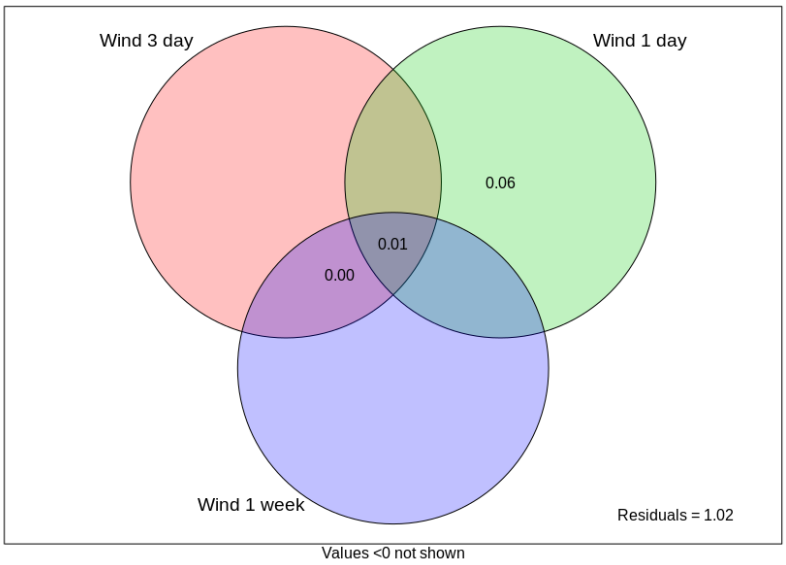
Scatter plots and box plots of Aitchison dissimilarity values were generated for each explanatory variable (Baselga & Orme, 2012). To assess if dissimilarity increased with geographic distance, distance-dissimilarity plots were generated using Aitchison distances with "betapart" (v1. 5. 1) (Baselga & Orme, 2012).

2.4 Results

2.4.1 One-Day Wind Back-Trajectories Correlate Best with the Bioaerosol Community

More compositional variation was explained for ITS (fungi) than for 16S (bacteria) (**Figure 2-2**). For 16S, the wind trajectories overall explained no significant variation in composition, but one-day wind had an R^2 of 6%. For ITS, there was a notable amount of shared variation explained by all three tested trajectory lengths. One-day and one-week wind had a similar R^2, suggesting either trajectory length would be suitable. Results from the Hill number-based RDA indicated one-day wind trajectories consistently had the highest R^2. Based on these results, the one-day back-trajectories were selected for both amplicons for the db-RDAs with variance partitioning.

a)



b)

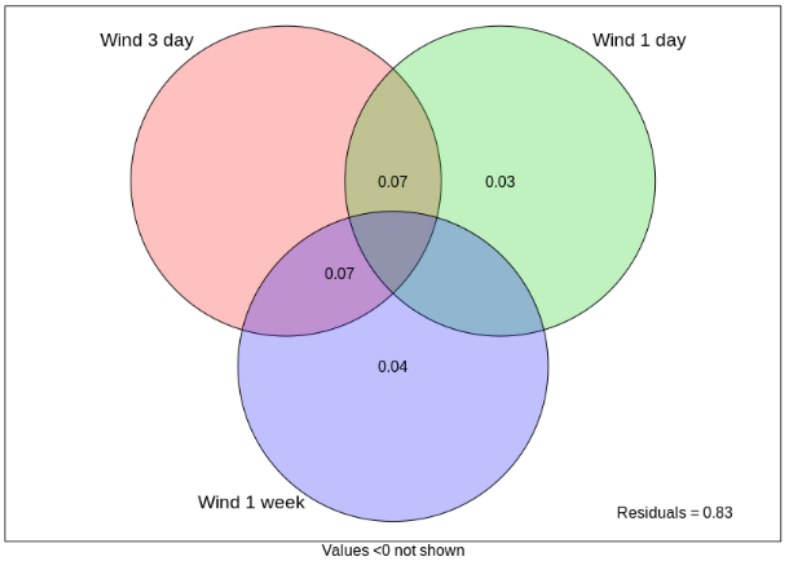


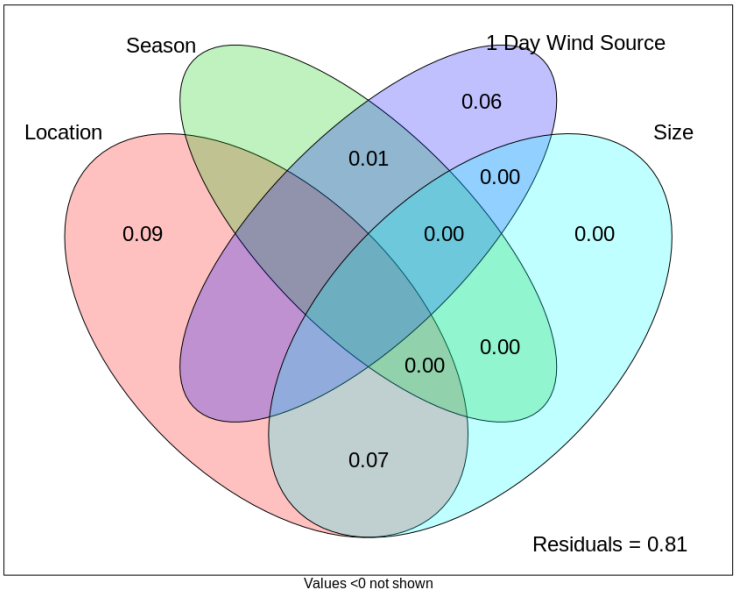
Figure 2-2 – Variance partitioning of the results from a distance-based redundancy analysis on the Aitchison distances among samples in the filtered ASV tables for (a) 16S and (b) ITS (abundance over 100 reads for 16S or 500 reads for ITS, and a coefficient of variability greater than three) showing the variance explained by one-day, three-day, and one-week back trajectories. The 16S variance partitioning shows an R^2 of 6% for one-day

wind. The total R^2 for the model was greater than one due to negative R^2 values associated with several of the other variables. Negative R^2 values occur in RDA analysis where most of the explanatory variables in a model explain very little information and are thus interpreted as zero explained variance (values not shown).

2.4.2 Bioaerosol Community Structure is Noisy, but Partially Explained by Wind, Spatial Location and Time

Wind, location and time were consistently important in explaining compositional variation among samples in bacterial and fungal aerosol communities (**Figure 2-3**). ITS had a much higher R^2 than 16S for the overall model (38% for ITS compared to 19% for 16S). Both amplicons showed shared explained variation between wind and location. There was also consistent shared variation between size and location, elevation, relative humidity and distance to sea where present. Time appeared to be important for ITS but its importance was less evident for 16S. Weather, park area and temperature were consistently not included in the db-RDAs, as they were not identified by the forward selection procedures. Location and one-day wind appeared to be similarly important and the time was also important in some analyses, particularly for ITS. Location was less important for both amplicons when the unfiltered ASV tables (including all ASVs) were used. Strong temporal clustering was confirmed in NMDS analysis of ITS (**Figure 2-5**), whereas 16S showed much weaker temporal differentiation (**Figure 2-4**). Some clustering is present by both location and wind trajectory, and this was more pronounced for ITS than 16S.

a)



b)

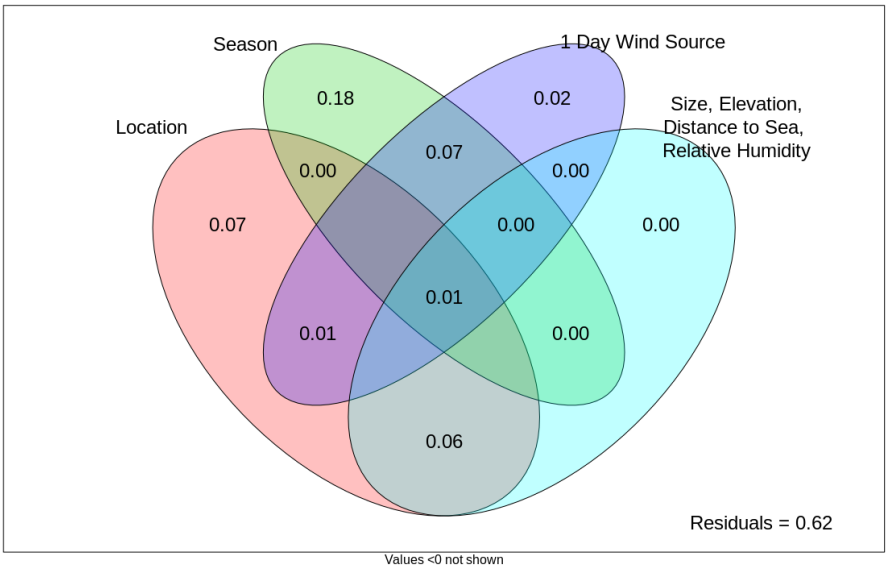


Figure 2-3 - Variance partitioning of the results from a db-RDA on the Aitchison distances among samples in the filtered ASV tables for (a) 16S and (b) ITS (abundance over 100 reads for 16S or 500 reads for ITS, and a coefficient of variability greater than three) showing the variance explained by location, time (referred to as season), one-day back-trajectory and other variables as indicated by forward selection procedures.

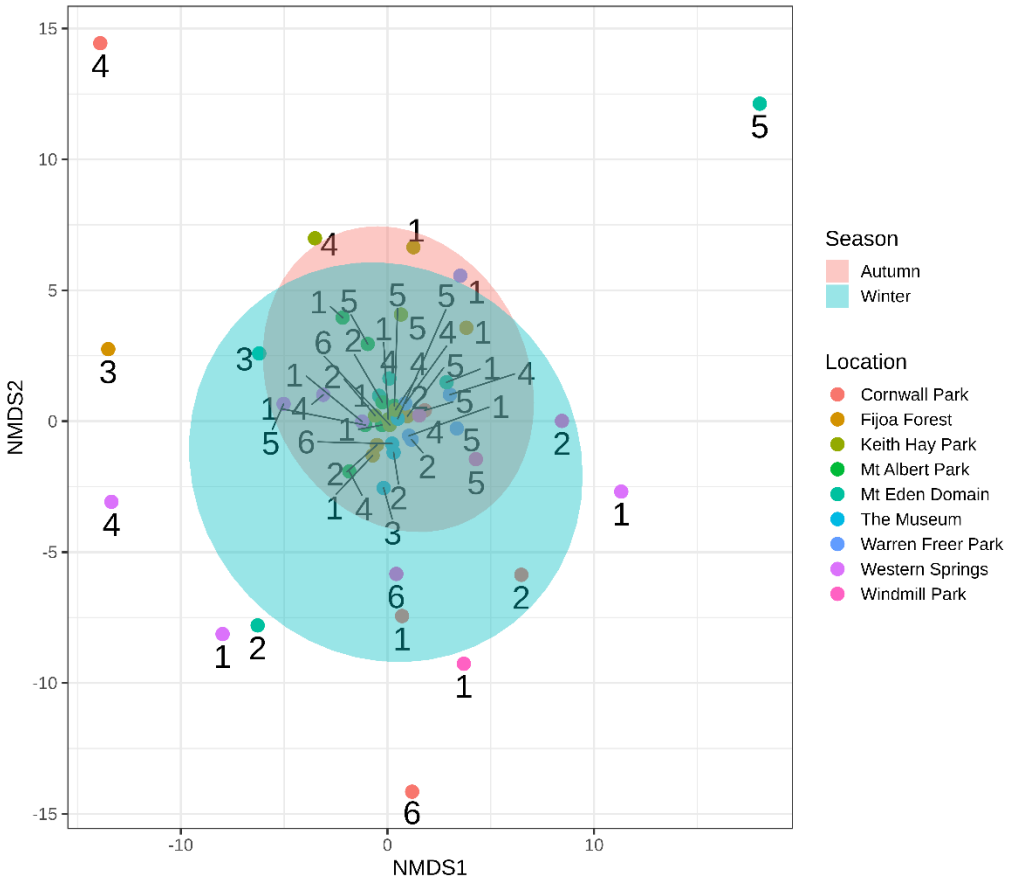


Figure 2-4 - NMDS ordination of the 16S Aitchison distance matrix of the filtered ASV table. Time (season) is indicated by ellipses representing the *t*-distribution of the autumn and winter points. Locations are denoted by colour and each sample is numbered with the trajectory cluster to which it belongs. Stress on 16S NMDS was 0.15 (dimensions one and two are presented above); the three-dimensional solution was selected as the two-dimensional solution had stress greater than 0.2 so could not be relied upon. Dimensions one and three, and two and three are presented in Supplementary Materials (E.1.1).

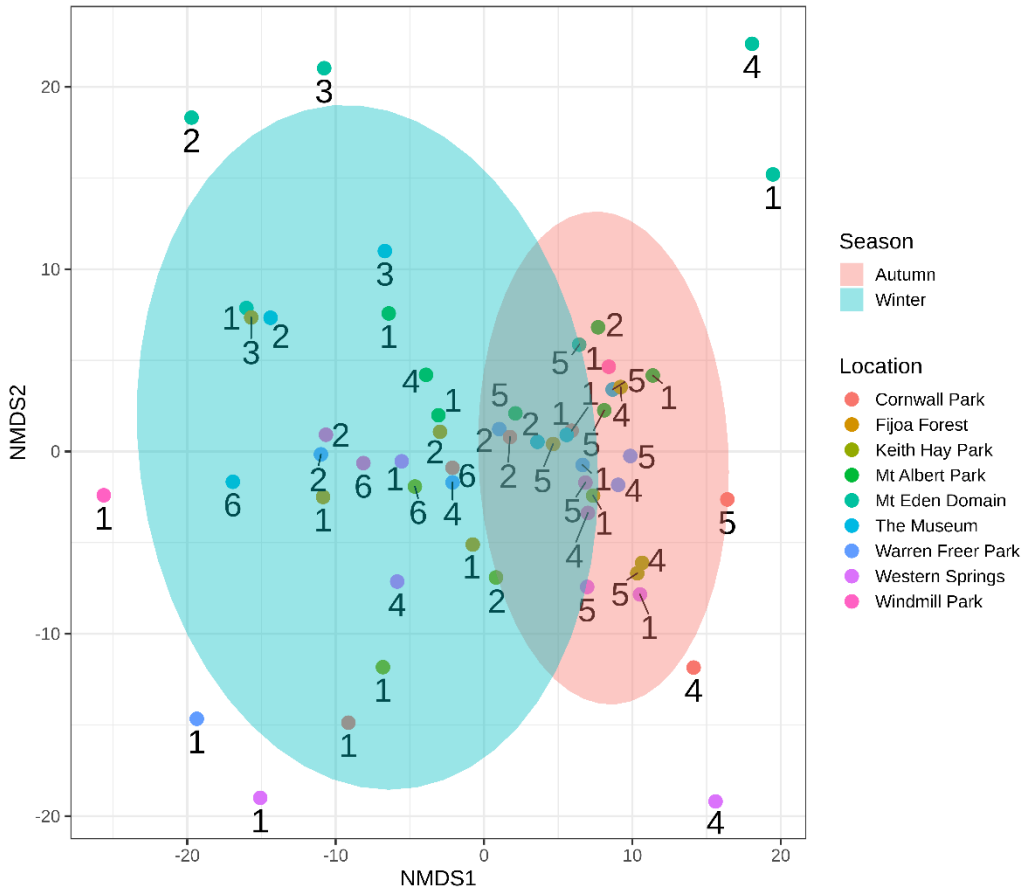


Figure 2-5 - NMDS ordination of the ITS Aitchison distances among samples in the filtered ASV tables. Time (season) is indicated by ellipses representing the *t*-distribution of the autumn and winter points. Locations are denoted by different colours and each sample is numbered with the trajectory cluster to which it belongs. The stress on the ITS NMDS was 0.15 (dimensions one and two are presented above); the three-dimensional solution was selected as the two-dimensional solution had stress greater than 0.2 so could not be relied upon. Dimensions one and three, and two and three are presented in Supplementary Materials (E.1.2). Higher dimensional solutions had lower stress values for both amplicons.

2.4.3 Differences Visible Among Locations, Wind Trajectory Clusters, and Time in Relative Taxon Abundances

Unfiltered bacterial reads numbered 607,740, with 3,828 ASVs inferred. Unfiltered fungal reads numbered 1,796,883, comprised of 5,311 ASVs.

Relative Abundance by Location – 16S

Each sample location had a distinctive aero-microbiome (**Figure 2-6**). While similar genera tended to be abundant in most places, this was not universal. For instance, *Methylobacterium* was abundant at the museum but was largely absent from other areas. An unclassified family of enterobacteria was only present in any significant number at Cornwall Park. *Bacillus* was present at about half of the sites. *Listeria* spp., which can include human pathogens, were widespread. Filtering criteria are described in **Figure 2-6** legend below.

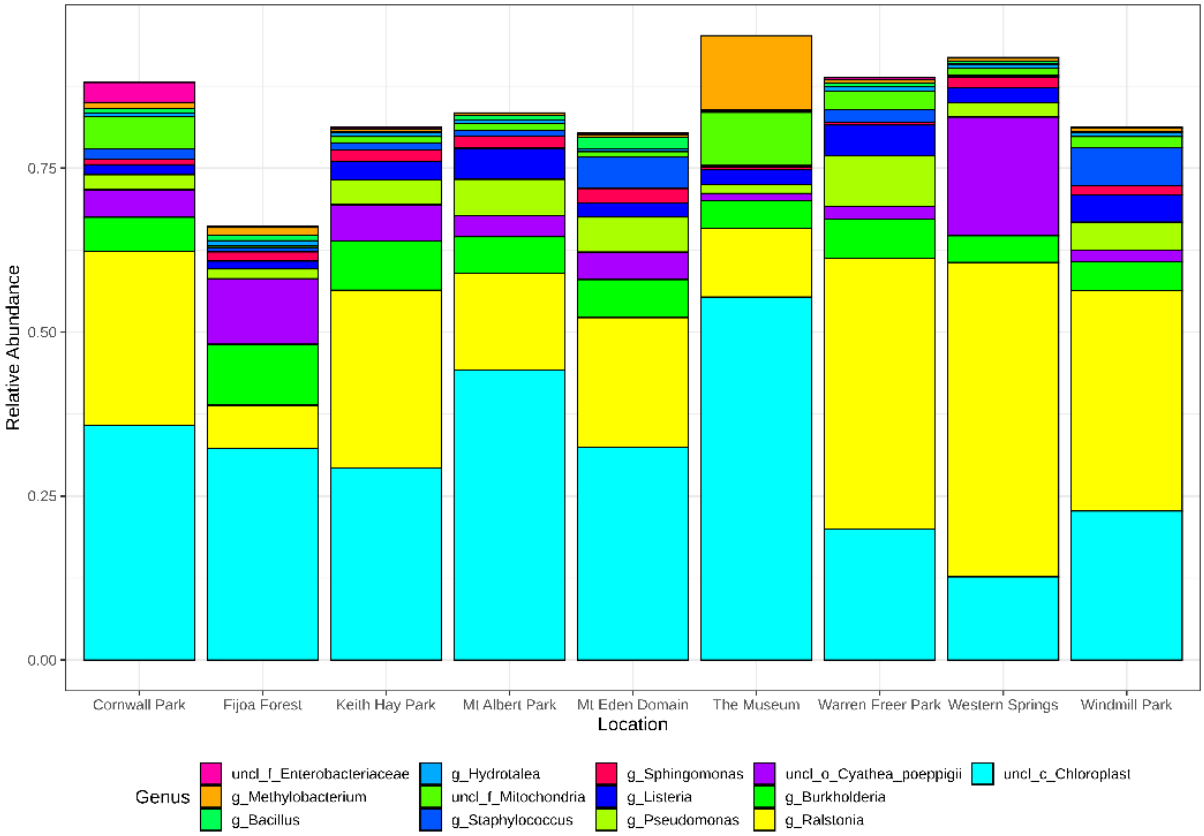


Figure 2-6 - Relative abundance by genus by sample location for 16S. Samples and ASVs with less than 100 reads were filtered out. Only genera with at least 0.5% of the reads per location are included in the bar plot.

Relative Abundance by Location – ITS

Fungal genera present were highly variable by location (**Figure 2-7**). There was a greater number of variable genera, and more differentiation between locations compared to 16S data. *Khuskia* spp. were only present at Mount Eden. Many *Trametes* spp. and *Saccharomyces* spp. were present at Fijoa Forest compared to other areas. *Ganoderma* spp. were much more common at Albert Park than elsewhere. Filtering criteria are described in **Figure 2-7** legend below.

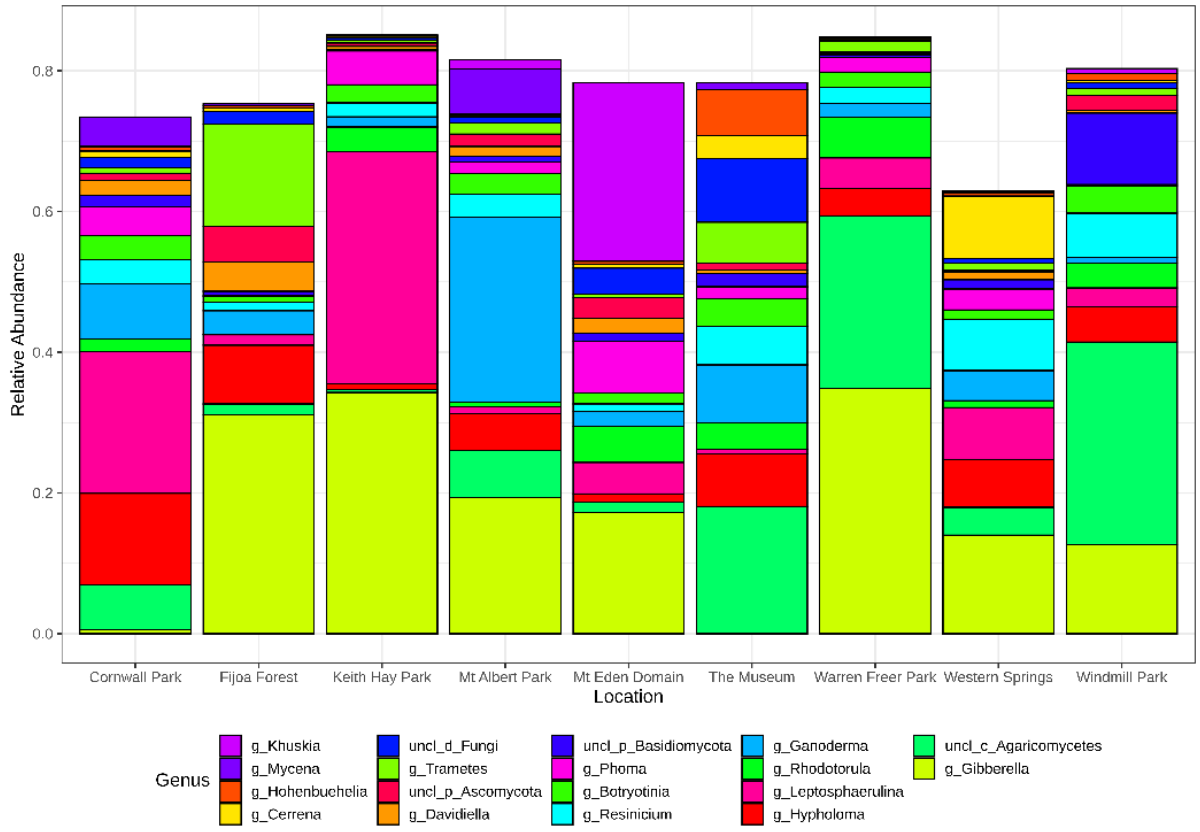


Figure 2-7 - Relative abundance by sample location for ITS. Samples with less than 100 reads and ASVs with less than 500 reads were filtered out. Only genera with at least 1% of the reads per location are included in the bar plot.

Relative Abundance by Time – 16S

The bacterial community composition visibly shifted between autumn and winter (**Figure 2-8**). *Methylobacterium* was present in winter only. Unclassified Chloroplasts were more dominant in winter. Filtering criteria are described in **Figure 2-8** legend below.

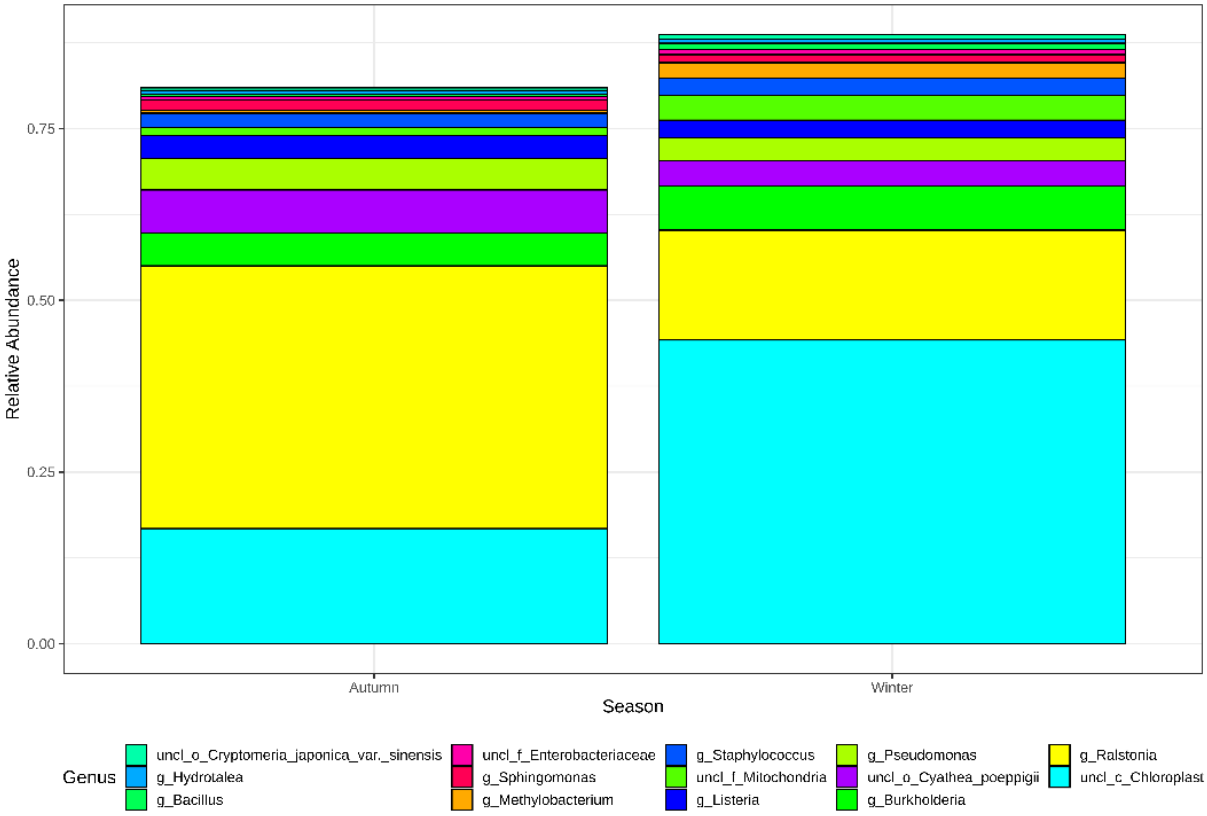


Figure 2-8 – Relative abundance by genus by time (season) for 16S. Samples and ASVs with less than 100 reads were filtered out. Only genera with at least 0.5% of the reads for each time were included in the bar plot.

Relative Abundance by Time – ITS

The fungal community varied markedly between autumn and winter (**Figure 2-9**). *Phoma* spp. were much more dominant in autumn, as were *Leptosphaerulina*, *Rhodotorula* and *Khuskia* spp. In contrast, *Trametes, Botryotinia* and *Ganoderma* spp. were more common in winter. Filtering criteria are described in **Figure 2-9** legend below.

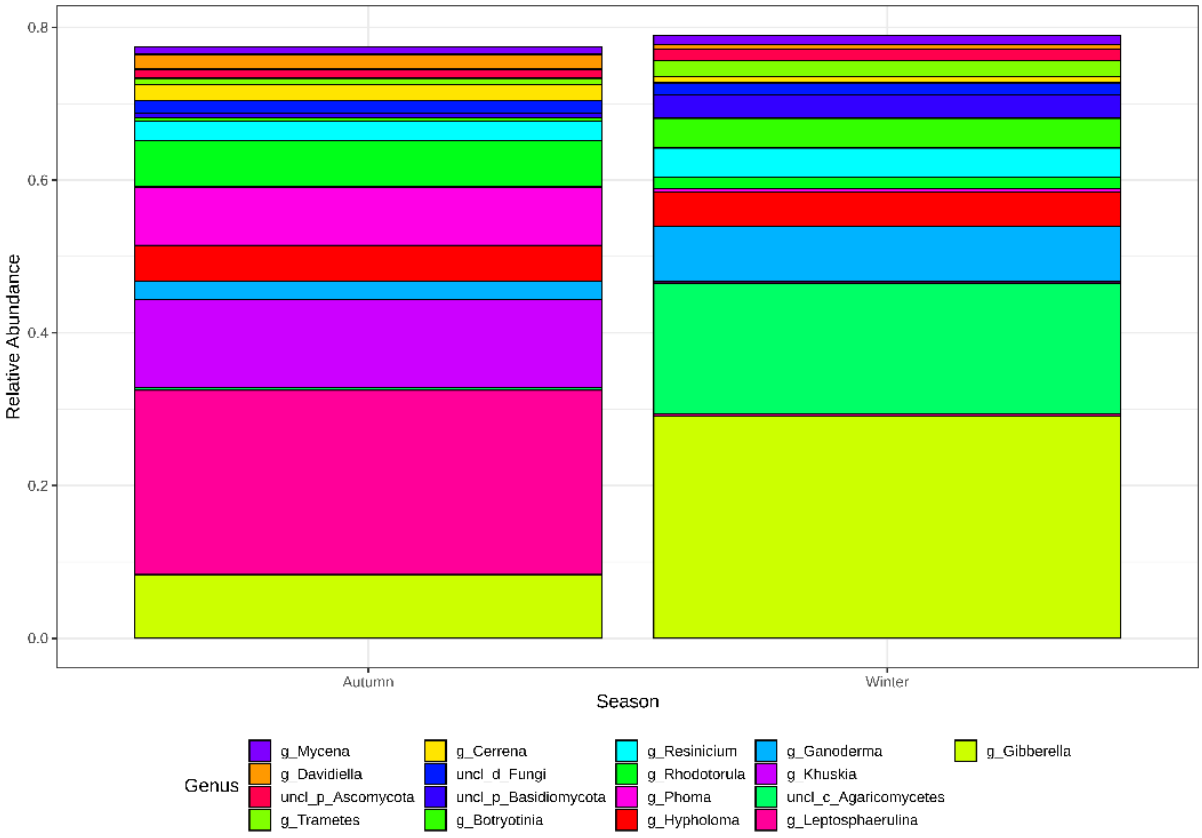
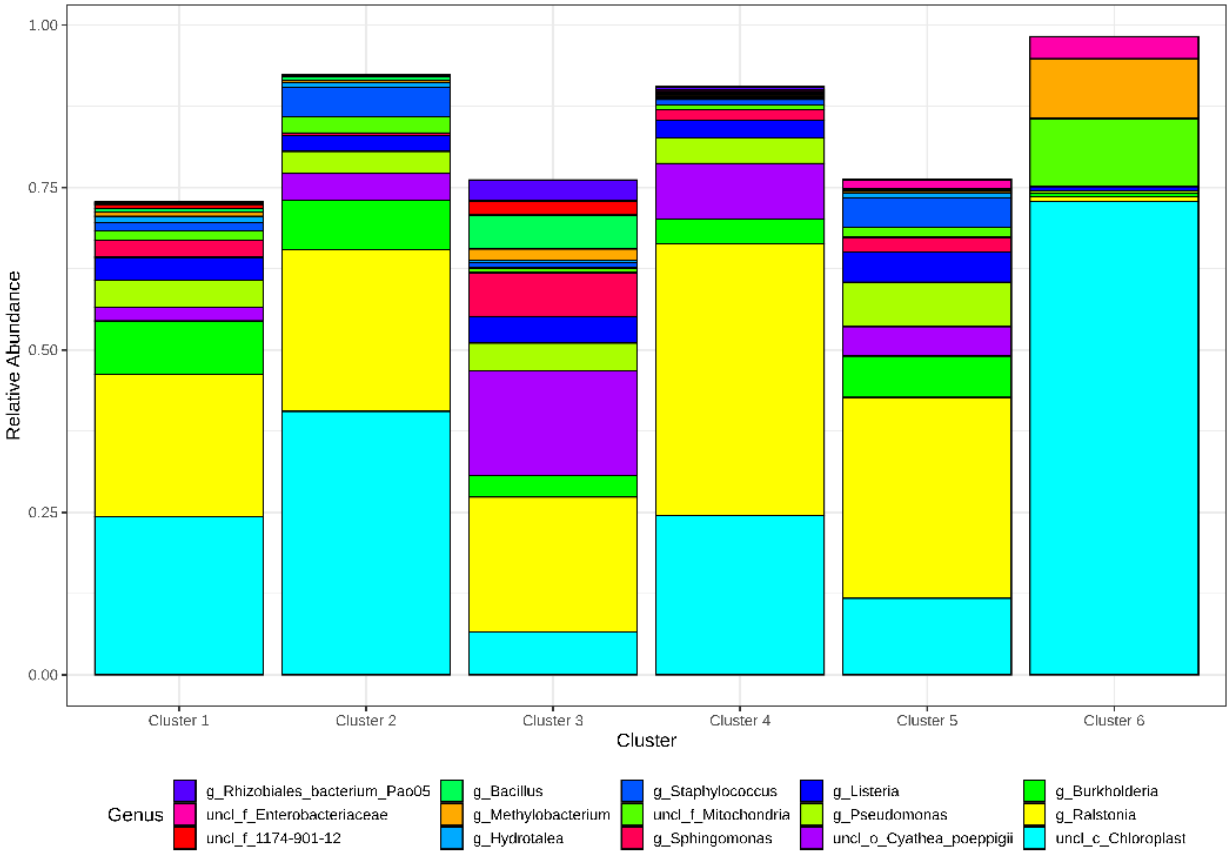


Figure 2-9 - Relative abundance by genus by time (season) for ITS. Samples with less than 100 reads and ASVs with less than 500 reads were filtered out. Genera with at least 1% of the reads for each time were included in the bar plot.

Relative Abundance by Wind Back-Trajectory Cluster – 16S

Bacterial genera were variable depending on the weather system sampled (**Figure 2-10**). Clusters one, two, four and five had reasonably consistent assemblages of microbes. Cluster six had a distinctive assemblage dominated by *Methylobacterium* spp, an unclassified chloroplast and an unclassified mitochondrion. Another unclassified organism (likely a chloroplast from the plant species *Cryptomeria japonica*) was present only in cluster one. Filtering criteria are described in **Figure 2-10** legend below.

a)



b)

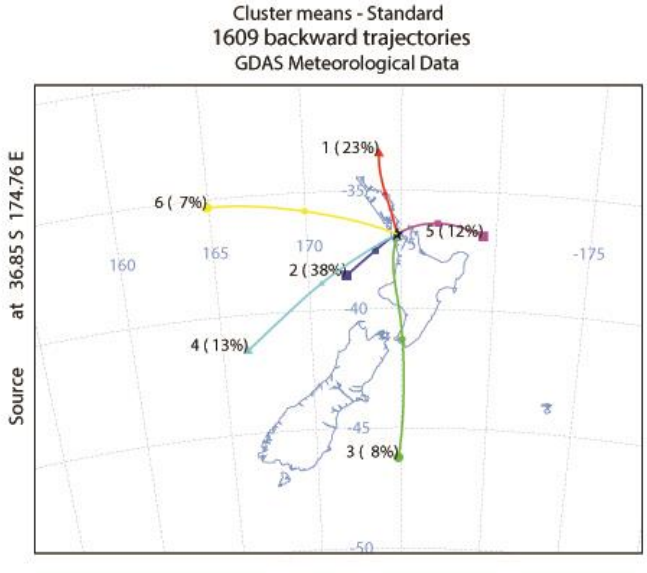


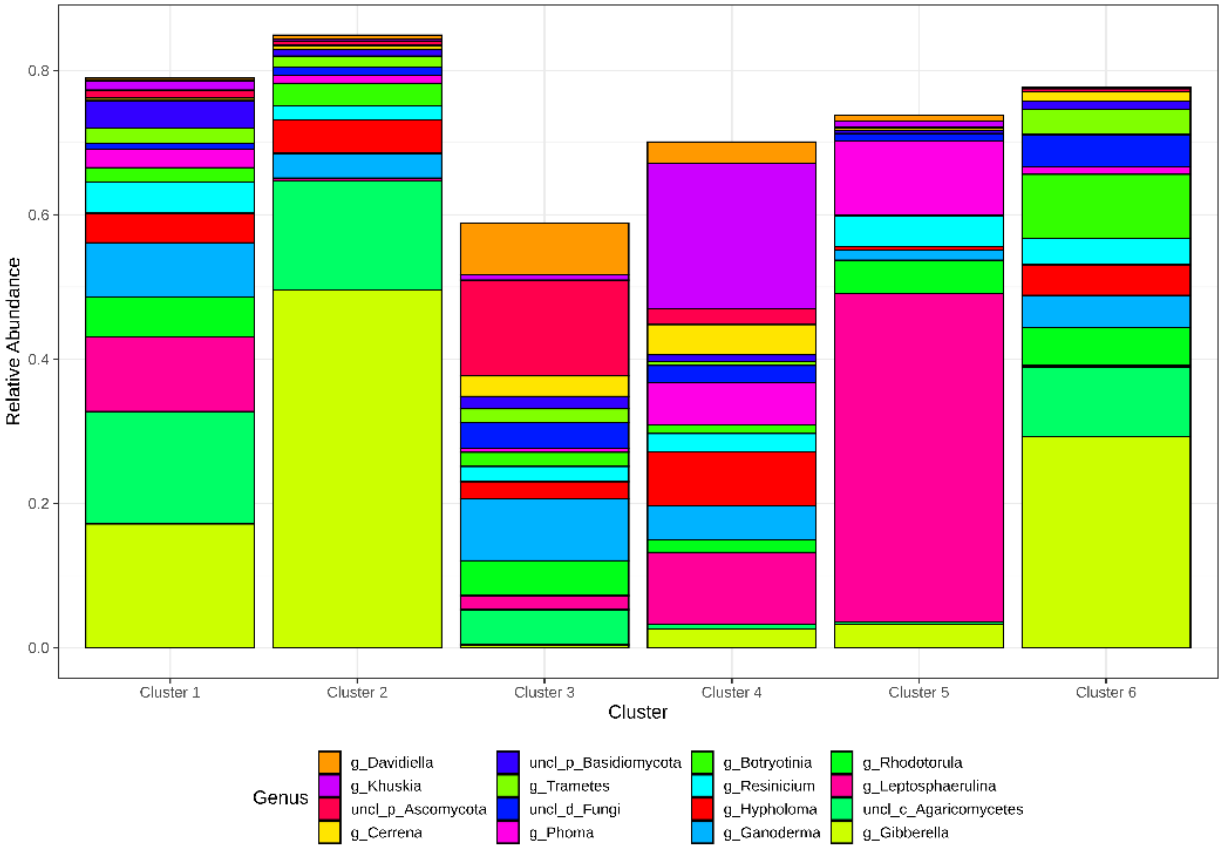
Figure 2-10 – Relative abundance by back trajectory cluster 16S. Samples and ASVs with less than 100 reads were filtered out. a) Relative abundance by genus for each wind trajectory. Only genera with at least 0.5% of the reads for each wind source included in the bar plot. b) Paths of the six clusters identified over the sampling period in the previous

24 hours before sampling, generated by the NOAA HYSPLIT model based on GDAS meteorological data. % next to each cluster indicates proportion of trajectories assigned to that cluster.

Relative Abundance by Wind Back-Trajectory Cluster – ITS

Fungal abundance and diversity varied by back-trajectory of the sampled air mass (**Figure 2-11**). Relative genus abundances in clusters one, two, and six appeared reasonably consistent, while the other clusters appeared quite different. *Khuskia* and *Cerrena* spp. were strongly associated with cluster four. Cluster three was dominated by *Saccharomyces* spp. *Resiniciuem* spp. were associated with cluster six. Filtering criteria are described in **Figure 2-11** legend below.

a)



b)

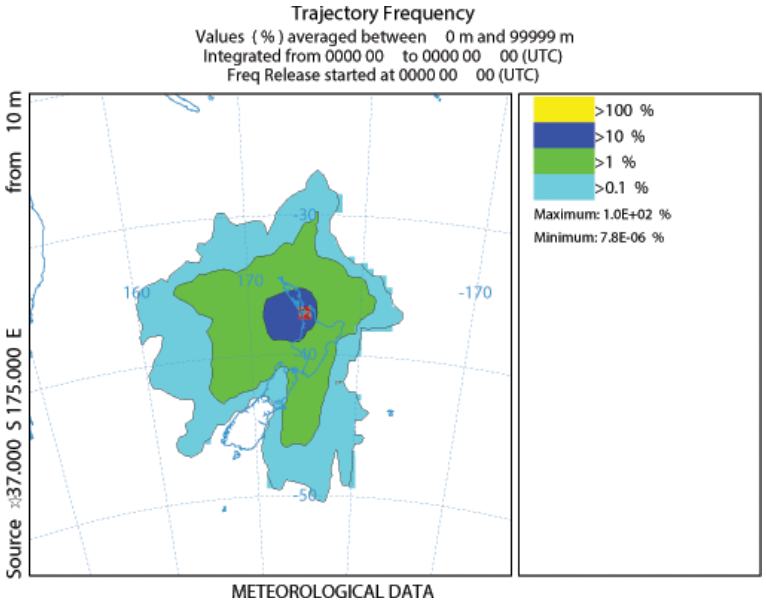


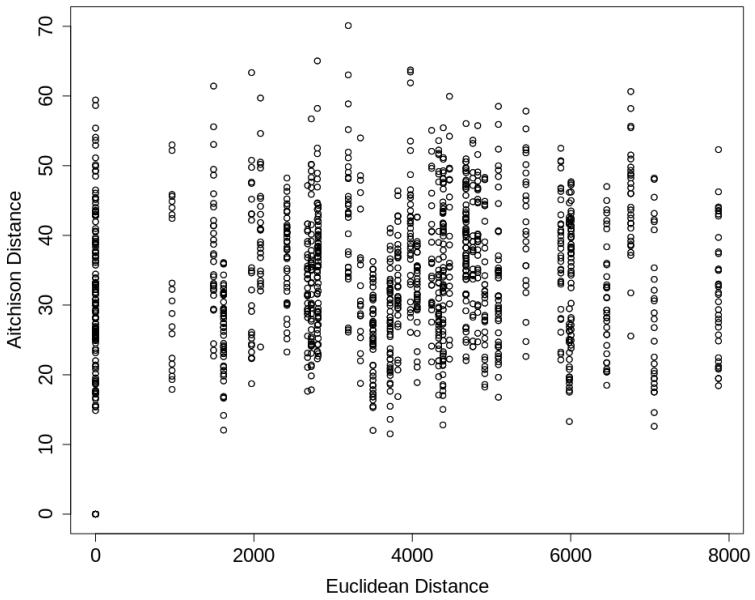
Figure 2-11 - Relative abundance by back trajectory cluster for ITS. Samples with less than 100 reads and ASVs with less than 500 reads were filtered out. a) Relative abundance by genus for each wind trajectory. Only genera with at least 1% of the reads each wind source were included in the bar plot. b) Frequency plot of air source over the sampling period in the previous 24 hours before sampling, generated by the NOAA HYSPLIT model

based on GDAS meteorological data. Colours indicate % of air coming from indicated area on the map.

Distance-Dissimilarity not Evident for Bacteria (16S) or Fungi (ITS)

There was no evidence of distance-dissimilarity across the 8 km study range, as genetic dissimilarity between samples appeared to be unrelated to physical separation of sample locations (**Figure 2-12**) (with dispersal limitations a positive correlation between distance and dissimilarity would be expected). Both amplicons show a consistent lack of distance-dissimilarity when using both filtered and unfiltered ASV tables as a basis for the Aitchison distance matrix.

a)



b)

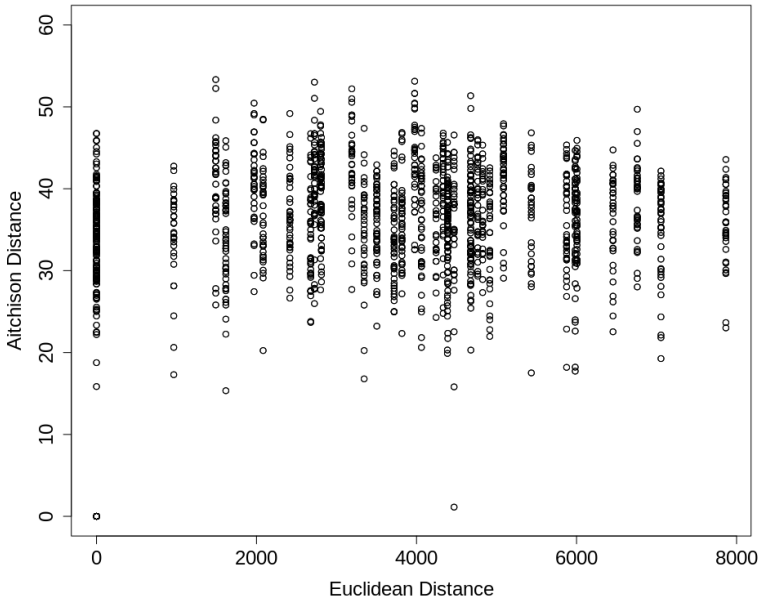


Figure 2-12 - Distance-dissimilarity plots for a) 16S and b) ITS of Aitchison sample distances calculated from the filtered ASV table against Euclidean spatial distances among sample locations.

2.5 Discussion

2.5.1 Summary

The aim of this chapter was to use bioaerosol communities in urban parks within Auckland, Aotearoa New Zealand to answer two research questions: (1) What is the relationship

between temporal scale of wind back-trajectories and variation in bioaerosol community structure? (2) What are the effects of environmental variation among different parks on bioaerosol community structure? Bioaerosols varied depending on spatial location, but wind source and, for fungi, time was of similar or greater importance. Urban parks harboured a diverse range of microorganisms and showed consistency with genera identified in similar bioaerosol studies. Some of the genera detected contained obligate or opportunistic pathogens. Air-mass trajectories of differing spatiotemporal lengths differed in their ability to predict variation in bioaerosol community structure. However, one-day, rather than one-week, trajectories provided the greatest explanatory power for Auckland parks. The temporal wind hypothesis 1 (air-mass trajectories of differing spatiotemporal lengths differ in their ability to predict variation in bioaerosol community structure. Specifically, one-week trajectories provide the greatest explanatory power, as indicated by modelling) therefore partially failed to be rejected. Spatial variation in bioaerosol community structure (composition and diversity) at Auckland parks was correlated with location, time, geographic origin of sampled air and park size for 16S. For ITS, the same factors as for 16S were important, and additionally distance to sea, relative humidity and elevation were correlated with the bioaerosol community. Temperature and weather were not correlated with the bioaerosol community for either amplicon. Environmental drivers hypothesis 2 a) (spatial variation in bioaerosol community structure (composition and diversity) at Auckland parks can be predicted by time, geographic origin of sampled air, altitude, distance to sea, temperature, relative humidity, park size (small, medium, large)), partially failed to be rejected. Relative abundances of common and rarer taxa fluctuated by location, time and air mass source, therefore environmental drivers hypothesis 2 b) (relative abundances of common and rarer taxa fluctuate by location, time, and air mass source) failed to be rejected. Similarity in community composition of parks did not decline with their increasing geographic distance, therefore the distance-dissimilarity hypothesis 3 (similarity in community composition of parks declines with their increasing geographic distance) was rejected. This study is important because urban bioaerosols remain poorly understood yet are thought to have noteworthy impacts on us and the environments we inhabit (Mhuireach et al., 2020). Understanding spatiotemporal variation on a city-wide scale is crucial for modelling microbial exposure in cities.

2.5.2 Wind Back-Trajectory Duration

Back trajectories of differing durations varied in their correlation to the bioaerosol community. One-day trajectories had the highest R^2 values in the wind db-RDAs with variance partitioning (**Figure 2-2**). One or three-day trajectories were generally better than one-week trajectories, suggesting that the standard trajectory length of three days per the literature (Archer et al., 2020) is reasonable. This was supported by frequency back-trajectory analysis, which showed more than 10% of source air over the sample period was from Auckland or close by, and the air coming from outside New Zealand was under 0.1% (**Figure 2-11**). Collinearity between the different trajectory lengths was observed, as air

coming from a certain direction would generally be expected to continue in that direction as it moves towards its destination. In accordance with this, cluster assignments for each sample were reasonably consistent for different trajectory lengths. The data suggest one-day trajectories could have better predictive value than three-day trajectories, and that consideration of alternative trajectory lengths to three days is prudent.

2.5.3 Spatiotemporal Variation in Bioaerosol Community Structure

Wind source and location (geographic proximity) were consistently identified as the most important variables correlated with variation in bioaerosol community structure (

Figure *2-3*), with time being additionally important for ITS. The ITS finding is consistent with patterns in the existing literature. Temporal variation is observed in all bioaerosol community studies to date that have investigated temporal patterns for bacteria and fungi (see section 3.2 Introduction for full details). The lack of temporal variation for the 16S data was unexpected given this context, but perhaps a bigger sample size was needed to detect a smaller effect. Other bioaerosol studies have shown that location or land-use can also be correlated with the bioaerosol community (Balyan, Ghosh, Das, & Banerjee, 2019; Burrows et al., 2009b; Mhuireach et al., 2020; Tanaka et al., 2020). A small number of studies considered the effect of air-mass source; some postulated it to be of key importance (Archer et al., 2020; Woo et al., 2013), while others suggested local sources were the determinants of the bioaerosol community (Bowers et al., 2013). For the first time, the relative importance of wind source, location and time were quantified. Modelling predicted 38% of the variation in fungal communities (ITS) and 19% of the variation in bacterial communities (16S), which is comparatively high for microbial community data (Carini et al., 2020). Some spatial effects were evident in the data; bacteria were mostly correlated with location, followed by one-day wind source. However, differences were not predictable using geographic distance between sample locations in the study area, as evidenced by the lack of a distance-dissimilarity relationship. For fungi, the most important variable driving community variation was time, with one-day wind and location of secondary importance. The greater importance of temporal variation for fungi was not evident in the bioaerosol literature reviewed, but few studies have surveyed both bacteria and fungi. Greater temporal variation for fungi has been demonstrated in studies of soil microbiota (Shigyo, Umeki, & Hirao, 2019). This is possibly due to the fungal genera identified being largely soil and plant associated and wood rotting organisms, so their abundances could be expected to be linked more closely to plant growth cycles, while bacterial genera tended to be more generalist. Shigyo et al.(2019) proposed that greater fungal temporal variation was due to environmental factors like plant traits, somewhat supporting this idea. Weather and temperature were not identified as influential variables in any analysis. Given that precipitation is known to affect the deposition of microorganisms (Burrows et al., 2009b), and temperature has been frequently found to affect bioaerosol communities (Woo et al., 2013), this is a surprising result.

2.5.4 Effect of Location, Time, and Wind Direction on Fungal and Bacterial Genera Identified

All bacteria and fungi observed are known environmental microbes or are associated with plants, humans or other animals. Many bacterial genera identified included pathogens such as *Listeria* spp. Fungal genera present were predominantly plant or soil associated, with many wood rotting species. Some yeasts were present, and some genera contained opportunistic human pathogens and allergens such as *Khuskia* spp. The genera detected, such as bacterial *Pseudomonas, Ralstonia* and *Methylobacterium* spp. and fungal *Penicillium, Alternaria* and *Cladosporium* spp. were consistent with previous bioaerosol studies (Barberán et al., 2015; Be et al., 2015; Garcia-Alcega et al., 2020). While *Pseudomonas, Ralstonia* and *Burkholderia* are commonly reported contaminants, their consistency with other bioaerosol studies, biological niches and presence despite stringent decontamination procedures indicate they also represent a genuine constituent of the bioaerosol community. Chloroplasts were commonly observed, presumably from pollen and other plant fragments, and are abundant in similar bioaerosol studies (Brodie et al., 2007b; Franzetti et al., 2010; Woo et al., 2013).

Fungal genera were observed to be more diverse and variable than bacterial genera (see section 2.4.3 Differences Visible Among Locations, Wind Trajectory Clusters, and Time in Relative Taxon Abundances). Bacterial and fungal genera varied by location. For instance, *Methylobacterium* (an opportunistic human pathogen and soil and water associated bacterium (Lai et al., 2011)), was only present at the museum. *Bacillus* spp., which are frequently present in bioaerosols (Bottos et al., 2014), were interestingly only detected at about half of locations sampled. The fungal genus *Khuskia*, which consists of one species, *K. oryzae*, a known human allergen (Wang, Liu, Crous, & Cai, 2017) was detected only at Mount Eden. The fungal genus*, Phoma*, which is a notable plant pathogen (Bennett, Ponder, & Garcia-Diaz, 2018), however, was more dominant in autumn. This could be due to more rapid plant growth at that time of year. The effect of time at different scales is investigated in later thesis chapters. It was unclear if bacteria were less affected by time, or if the data did not reveal the relationship in this study.

2.5.5 Distance-Dissimilarity of Bioaerosols

No distance-dissimilarity relationships were apparent in either bacterial or fungal bioaerosol communities (**Figure 2-12**). This is not unreasonable since the atmosphere is a fluid medium which could be expected to be well mixed at the scale which sampling was performed (8 km between the most distant Auckland parks). Back-trajectory analysis also placed all sample locations in the same cluster at the same time, supporting consistency in weather patterns across the city. This is in stark contrast to the often very strong distance-dissimilarity patterns generally observed for soil and water microorganisms (Feng et al., 2019; Lear et al., 2013). As noted above, this lack of distance-dissimilarity relationship did not preclude local spatial differentiation. This is likely due to the majority of bioaerosols remaining close to the ground, and having short transport ranges and local differentiation as a result (see section 1.3.1 Particle Movement in the Atmosphere). A minority of bioaerosols are propelled higher

into the atmosphere and easily transported on a regional scale (Kellogg & Griffin, 2006). The idea of there being distinct groups of organisms, some good at regional dispersal and showing little distance-dissimilarity and some being restricted to certain parks, is supported by spatial differentiation only being apparent for the filtered ASV table (with widely dispersed invariant taxa removed). The extent to which mixing of this scale occurs is presently unclear, and different spatial scales are the subject of later chapters of this thesis.

Two common threads emerge through all the preceding analyses. ITS data displayed clearer patterns, and higher R^2 values than 16S data. This is likely due to the presence of many more reads and a greater number of ASVs for ITS. Higher quality information for ITS would have made any relationships easier to detect. Higher numbers of ASVs for ITS are expected since ITS is more variable in length than 16S (Callahan, 2020). Fungi, being eukaryotic, are frequently multicellular and have larger cells in contrast to prokaryotic, smaller, unicellular bacteria. Some of the abundant fungi identified in the study are macroscopic fungi which have relatively large spores. Thus, cells could be expected to be clumped together more often than bacteria, which could explain why more reads were present for ITS. Fungal cells being part of larger assemblages than bacterial cells would result in shorter average atmospheric residence times for fungi (Burrows et al., 2009b), which could also be an explanation for the greater observed variability for ITS. The other notable feature is the high degree of consistency in ecological conclusions reached between the different amplicons tested and between various methods (both compositional and non-compositional) employed for data analysis. This suggests that the relationships detected are ecologically sound.

2.6 Conclusions

Different durations of back trajectories showed different predictive potential for bioaerosol communities. Bioaerosols varied as a result of many factors, but in particular location, time and air mass source. Temporal variation was especially important for fungi. These key findings were consistent with the body of literature on bioaerosols. Due to their ease of dispersal, bioaerosol communities did not show distance-dissimilarity, yet still displayed local differentiation. The type and number of fungal and bacterial bioaerosols detected varied by location, time and air-mass source. The persistence of these patterns at different spatial and temporal scales requires further investigation. However, this study shows that ecological factors can predict community structure of microbial aerosol communities in urban landscapes.

Chapter 3- Two-Year Seasonal Study of Temporal Bioaerosol Variability

3.1 Abstract

Microbes are a fundamental component of Earth's ecosystems, thus knowledge of ecosystem connectivity through microbial dispersal is key to understanding how ecosystems will change in a warming world. However, microbial dispersal through the air remains poorly understood. Relatively few studies have been performed on bioaerosols in urban areas, despite them housing the majority of the world's human population and harbouring important aerial pathogens and allergens. To shed further light on this, weekly sampling was performed in Auckland, Aotearoa New Zealand for two years. Sampled bacterial and fungal communities were characterised through sequencing of 16S rRNA (16S) and internal transcribed spacer (ITS) marker genes. In general, shorter (one to three-day) air-mass back-trajectories correlated more strongly with the bioaerosol community than longer (one-week). However, more importantly, the bioaerosol communities varied over time in a non-linear fashion, exhibiting marked seasonality, which was especially pronounced for fungi. This differential response of fungi and bacteria was possibly due to differences in size and consequent atmospheric residence times.

3.2 Introduction

Microbial communities drive functioning of all terrestrial and aquatic ecosystems worldwide (Pointing et al., 2016). Their composition and function vary dramatically, and this variation has important impacts on ecosystem functions (Fierer et al., 2012). 'Priority effects' in microbial communities can be important in determining composition and functioning over time (Evans, Martiny, & Allison, 2017; Fukami, 2015), therefore understanding how microbes move around ecosystems is crucial (Zhou & Ning, 2017). Understanding microbial dispersal is fundamental within the context of climate change driving ecosystem changes, as selective pressures in both source and destination habitats shift (Chown et al., 2012).

Aerosol microbial communities are an excellent case study system to examine microbial movement because air is easy to sample at a variety of temporal and spatial scales in a highly replicated way. As the aerosphere links all ecosystems, organisms that can utilise it for dispersal appear to have unfettered access to all global habitats (Pearce et al., 2016) but the extent to which "everything is everywhere" holds true for microbes remains unclear (O'Malley, 2008). Aerosol community composition influences composition of novel terrestrial and aquatic habitats, thus bioaerosols provide the source populations for Earth's ecosystems (Kellogg et al., 2004; Šantl-Temkiv et al., 2020; Sokol et al., 2013). The behaviour of microbes in the air follows particle physics in a fluid (Burrows et al., 2009b; Pepper, 2015) (see 1.3.1), therefore predictions can be made about how microbes move in air based on their properties, further supporting bioaerosols as an excellent model system for microbial dispersal.

Insights from urban aerosols specifically are important, since 55% of the world's population currently resides in urban areas, which is projected to increase to 68% by 2050 (UN, 2018), and microbial exposure is emerging as an important human health influence

(Mhuireach et al., 2020). Urban aerosol communities vary across spatiotemporal scales, but fewer studies have been conducted with high replication at scales with fine temporal grain (less than one week between samples and/or sample duration under one day) over a large temporal extent (more than one year between samples) (Mathieu et al., 2020). Shorter duration and coarser grain studies have repeatedly demonstrated temporal variability over daily, seasonal and yearly scales (Bowers, McCubbin, Hallar, & Fierer, 2012; Burrows et al., 2009b; Els et al., 2019). Patterns vary, but generally warming, cooling and precipitation cycles affect bioaerosol communities (Woo et al., 2013). The effect of temperature can be variable, but more studies show increasing concentrations with temperature (Tong & Lighthart, 1999). Spatial variability is ubiquitous, with climate, land-use, population, ecosystem, vegetation and altitude all thought to influence bioaerosols (Burrows et al., 2009a; Burrows et al., 2009b). Very few molecular studies have been performed on seasonal variation in outdoor urban bioaerosols, despite their ability to harbour human pathogens and allergens (Bowers et al., 2011b; Franzetti et al., 2010; Fröhlich-Nowoisky et al., 2009; Priyamvada et al., 2017; Woo et al., 2013). Most work on urban bioaerosols has been limited to Europe, Asia, and the USA, over periods of around a year or less. Every bioaerosol study detected seasonal variability in the micro-organisms surveyed, but how universal these patterns are is unknown.

Many factors have been demonstrated to affect urban bioaerosols, but the relative contribution of each has yet to be quantified. Factors thought to be important drivers of observed seasonal variability are air-mass source, temperature, relative humidity, wind speed, air pressure, particulate load and precipitation. Woo et al.(2013) observed distinctive fungal and bacterial assemblages at different times of year in Hong Kong and postulated the reason for this change was a shift from marine air in summer to continental air in winter. Other recent bioaerosol studies have also suggested air mass source is the key determinant of bioaerosol communities (Archer et al., 2020). In contrast, Bowers et al.(2012), in their non-urban seasonal study, found wind trajectories were not of significant importance for the composition of the sampled bioaerosol communities and postulated that local bacterial emission sources were key. Temperature is the environmental factor most consistently identified as a correlate with the bioaerosol community in studies to date (Bowers et al., 2013; Bowers et al., 2012; Fröhlich-Nowoisky et al., 2009; Woo et al., 2013). This is thought to be due to the indirect effects of warmer temperatures, increasing microbial reproduction and aerosolisation (see section 1.3.2 Bioaerosol Particles: Temporal and Spatial Variation). For the other factors, some studies observed significant effects, and some did not, so their importance is less conclusive (Brodie et al., 2007b; Fröhlich-Nowoisky et al., 2009; Priyamvada et al., 2017). Few studies have compared the seasonal responses of fungi and bacteria, although Woo et al. (2013) hint that while there may be more local influence for fungi, there is currently insufficient data to conclude if systematic differences exist in the responses of bacterial and fungal bioaerosols to seasonality. The results in 2.5.4 Effect of Location, Time, and Wind Direction on Fungal and Bacterial Genera Identified suggest differences are likely, with indications of greater seasonality for fungi. The drivers for this

are presently unclear but could be related to greater size of fungal particles and resulting shorter atmospheric residence times, or fungal life history strategies being more host-specific and less generalist than those of bacteria (Shigyo et al., 2019).

To contribute to filling this knowledge gap, weekly samples were taken in the CBD of Auckland, Aotearoa New Zealand for two years (May 2017 to May 2019), in the longest continuous urban bioaerosol study to date. Samples were subjected to 16S (for bacteria and archaea) and ITS (for fungi) gene amplification and sequencing. Abundance and composition of microorganisms were analysed with respect to air-mass source, temperature, relative humidity, weather conditions and a variety of non-linear functions were compared to quantify the scale and pattern of the temporal variation. Distance-based redundancy analysis (db-RDA) with variance partitioning was used to quantify, for the first time, the relative correlation of each variable with the variation in diversity and composition of the observed bioaerosol community.

3.2.1 Hypotheses

1. Temporal wind hypothesis: Air-mass trajectories of differing spatiotemporal lengths differ in their ability to predict variation in bioaerosol community composition. Specifically, one- or three-day trajectories provide the greatest explanatory power relative to one-week back-trajectories (based on findings in Chapter 2 - Spatiotemporal Bioaerosol Variability in Urban Parks).
2. Time-dissimilarity hypothesis: Similarity in community composition of samples declines with their increasing temporal separation in a non-linear fashion, due to seasonality.
3. Environmental drivers hypothesis:
    a. Temporal variation in bioaerosol community structure (composition and diversity) at the Auckland CBD can be predicted by season, geographic origin of sampled air, temperature, weather and relative humidity.
    b. Relative abundances of common and rarer bacterial and fungal taxa fluctuate by season and air mass source.
4. Taxonomic differences: Temporal patterns of 16S bacterial and ITS fungal microbial aerosol communities are expected to be different. Specifically, fungi display more variability in general and greater seasonality.

3.3 Methods

3.3.1 Field Sampling and Environmental Data Collection

Sampling with the Coriolis into phosphate buffered saline (PBS) was performed for one hour, once a week, for two years on the roof of a 16-floor building in central Auckland. This elevated location was chosen to avoid disruption from stochastic ground level events (e.g. a street sweeper or large vehicle going past) that could mask seasonal trends. The day of the week was randomised to avoid weekends biasing results. These protocols were based on methods in the similar study by Woo et al. (2013). Samples were collected in a designated

corner of the roof at a height of 1.8 m. Gloves were worn and the extender and Coriolis unit cleaned with bleach. The Coriolis neck, head and cone were cleaned with bleach, ethanol and three rinses of milli-Q (MQ) H_2O. The cone was filled with 15 mL of PBS. A negative was taken (PBS put into the cone without running the Coriolis) before running the Coriolis for two – four minutes with MQH_2O to ensure all bleach residue in the head and neck was removed. The MQH_2O was discarded and replaced with 15 mL of PBS and the Coriolis was run at 300 L/m for one hour. The PBS in the cone was manually topped up to 15 mL after 30 minutes and sampling was completed with at least 10 mL of PBS remaining in the cone. Samples were transferred into a 15 mL falcon tube and stored at -20 °C within one hour. During sampling, observations were recorded of weather conditions (with a Kestrel 3000) and particle counts were taken (with an AeroTrak particle counter). If rain occurred, the Coriolis was sheltered with an umbrella. If rain was heavy it was packed up until the rain abated.

Air-mass back-trajectories were generated hourly over the sampling periods using the NOAA HYSPLIT model (v5. 0. 0 Ubuntu), (Stein et al., 2016) with GDAS meteorological data. Clustering was performed in HYSPLIT (see Appendix B HYSPLIT Clustering Procedure for details). The number of clusters selected (four) was based on a marked increase in total spatial variance as clusters reduced. The cluster for each sample was manually entered into the metadata, which was imported into R. Back-trajectories were generated for one day, three day and one week durations.

3.3.2 Laboratory Processing

For details on DNA extraction and DNA sequencing methods see section C.1 Laboratory Methods – DNA Extraction, and section C.3 Laboratory Methods – DNA Sequencing

3.3.3 Bioinformatics

For details on bioinformatic processing see D.3.3 Data Analysis in Appendix D

Optimising Bioinformatics Protocols for Aerosol Microbial Community Data – a Case Study Using an Urban Parks Dataset. The process recommended as a result of that piece of work was followed for the seasonal study dataset presented in this chapter. The decontaminated amplicon sequence variant (ASV) table (with read counts adjusted to remove contaminant sequences) and combined taxonomy and relevant metadata were analysed in R with respect to the hypotheses for this study.

3.3.4 Data Analysis

The 16S (the gene for the RNA component of the 30S small-subunit of the prokaryotic ribosome) and ITS (spacer DNA situated between the small-subunit RNA and large-subunit ribosomal RNA genes in the fungal genome) amplicon sequence variant (ASV) tables were prepared for analysis by retaining only ASVs with more than either 300 reads (for 16S) or 500 reads (for ITS) and a coefficient of variation (CV) of greater than four. This filter was used to improve any signal which could be masked by invariant taxa and/or low-level

stochastic variation. Any problematic samples were completely removed, for instance duplicates or samples contaminated by flying insects. Metadata associated with each sample were adjusted to remove redundant variables (incomplete measurements or variables which were not analysed further), numerical variables were standardised using *Z*-scoring in R (with the function scale in base R). Additional variables to indicate time were included. These were days since the start of the study (DSS), whether a sample related to year one or two of the study, days from the beginning of the study year in circular degrees (count of days into the respective sampling year converted to degrees of a circle, with 180° being six months into either year one or year two) and a categorical variable for season (summer being December to February, autumn being March to May, winter being June to August and spring being September to November).

Bayesian multiplicative replacement was performed on the ASV tables (Gloor et al., 2017; Quinn et al., 2019) using the function cmultRepl (Aitchison, 1982; Templ et al., 2011) in the R package "zCompositions" (v1. 3. 4) (Palarea-Albaladejo & Martín-Fernández, 2015). "RobCompositions" (v2. 2. 1) (Templ et al., 2011) was used to generate an Aitchison distance matrix, which was fed into downstream analysis functions.

Variance partitioning of an Aitchison distance matrix calculated on pairwise comparisons of compositional samples was conducted following a distance-based redundancy analysis (db-RDA) to compare the relative amount of compositional variation explained by the three different lengths of back-trajectories. This result was used to determine which length would be used for the remainder of the analyses; a higher R^2 indicated greater variation explained by a given trajectory.

A pairwise Euclidean distance matrix was calculated using the number of days between samples (days since start; DSS). The Aitchison compositional dissimilarities were plotted against the Euclidean temporal dissimilarities for each pair of samples to visualise the time-dissimilarity relationship. The time-dissimilarity plot using pairwise compositional dissimilarity showed a cubic-non-linear relationship, so a cubic transformation of the DSS variable was added to the db-RDA to model this non-linearity. A non-linear correlation between the Aitchison compositional dissimilarities and the Euclidean temporal dissimilarities was calculated using the R package "nlcor" (v.1.3.2) (Ranjan, 2020). This correlation results in a coefficient that varies between zero and one, and reflects the strength of the non-linear correlation whereby higher values demonstrate a stronger correlation than lower values; a *P*-value is also provided which reports the statistical significance of the correlation.

Forward selection of variables was performed to determine which explanatory variables to include in further db-RDA analysis. This was implemented in R using the capscale and ordi2step functions in "vegan" (v2. 5. 6) (Oksanen et al., 2012). Variables were included with a p in parameter of 0.2 and 999 permutations.

Significant variables resulting from the forward selection procedure were included in the db-RDA with variance partitioning to quantify strength of relationship with bioaerosol community composition (**Table 3-1**). These variables were season and wind-source for 16S. For ITS, variables included were season, DSS to powers one through to three, weather and temperature. In addition to quantifying the relative importance of the different variables, db-RDA quantified the portion of explained variation that was shared among variables.

Table 3-1 - Explanatory variables and whether they were included in modelling compositional variation in bioaerosol communities sampled in urban Auckland.

Variable	Variable type	Description	Included in 16S model?	Included in ITS model?
1-day back trajectory cluster	Categorical	Path of sampled air-mass in previous 24 hours	Yes	No
3-day back trajectory cluster	Categorical	Path of sampled air-mass in previous 72 hours	No	Yes
1-week back trajectory cluster	Categorical	Path of sampled air-mass in previous 168 hours	No	No
Season	Categorical	Season sampling during which took place	Yes	Yes
Days since start (DSS)	Numerical	Number of days since sampling began	No	Yes
Days since start $^{2-3}$ (DSS^{2-3})	Numerical	Number of days since sampling began to the power of 2 and 3	No	Yes
Days in degrees (DD)	Numerical	Number of days into sampling year converted into circular degrees	No	No
Weather	Categorical	Weather during sampling	No	Yes
Temperature	Numerical	Temperature during sampling	No	Yes
Humidity	Numerical	Humidity during sampling	No	No
Month	Categorical	Month sampling took place during	No	No
Year	Categorical	Year of study sampling took place during (year 1 or 2)	No	No

Visualisation of compositional variation among samples was achieved by generating Non-metric Multidimensional Scaling (NMDS) plots on Aitchison distance matrix, implemented in the R package "vegan". The function stat-ellipse in "ggplot2" (v3.3.2) (Wickham, 2016) was used to create ellipses, assuming a multivariate *t*-distribution. The stress for two to six dimensions was compared, with 500 random starts and 999 iterations per run. The lowest dimensional solution with a stress under 0.2 was selected (McCune et al., 2002).

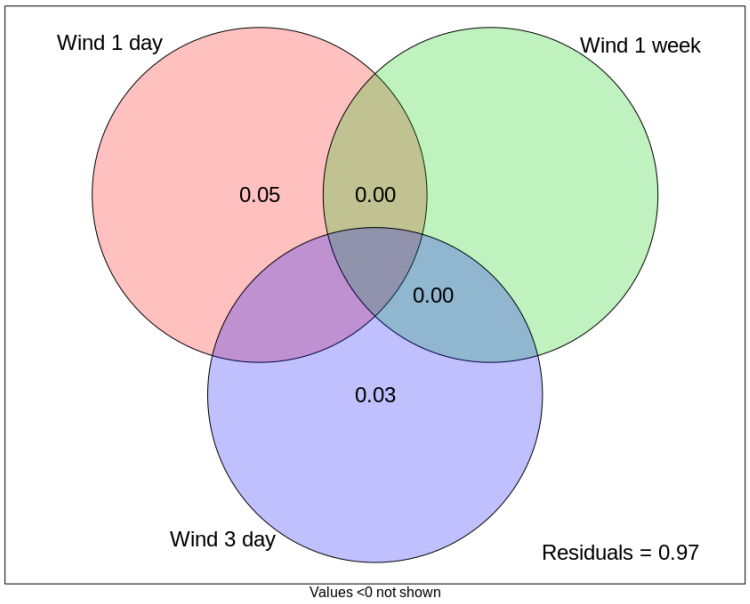
To assess patterns by month and back-trajectory for different genera, the percentage of total read count were calculated on data filtered as follows. For 16S, samples with less than 100 reads and ASVs with less than 300 reads were filtered out. Relative abundance by genus for each category included only genera comprising at least 1.2% of the reads. For ITS, samples with less than 100 reads and ASVs with less than 500 reads were filtered out. Relative abundance by genus for each category included only genera with at least 1% of the reads. Stacked bar plots showing relative abundances of these key taxa in different categories (month, wind) were generated using "ggplot2" (v3.3.2) (Wickham, 2016) and "phyloseq" (v1.30.0) (McMurdie & Holmes, 2013).

3.4 Results

3.4.1 One or Three-Day Back-Trajectories Selected for Air-Mass Source Variable in db-RDA

Wind back trajectories showed similar overall predictive value of community composition for both amplicons (R^2 of 2-3%) (**Figure 3-1**). The different trajectory lengths explained similar variation in community composition. One or three-day trajectories appeared to have the strongest relationships with the bioaerosol communities. One-day trajectories were selected for ongoing 16S analysis and three-day trajectories were selected for the remainder of the ITS analysis.

a)



b)

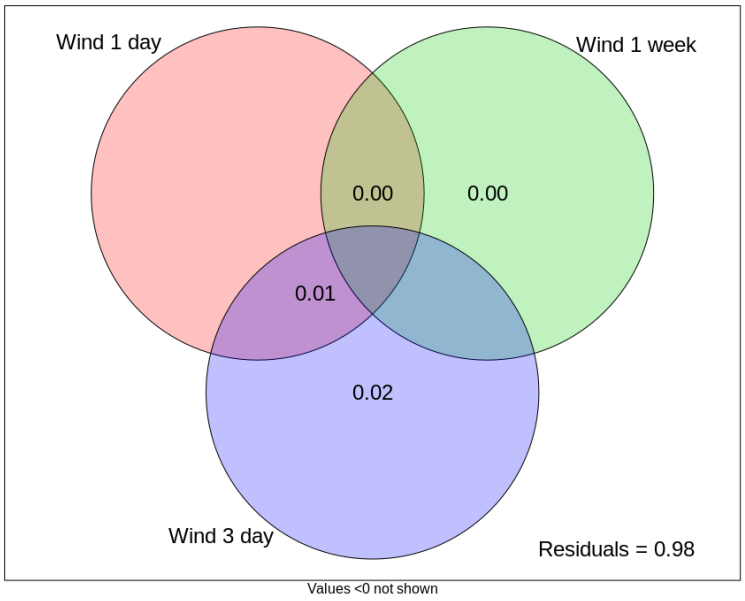
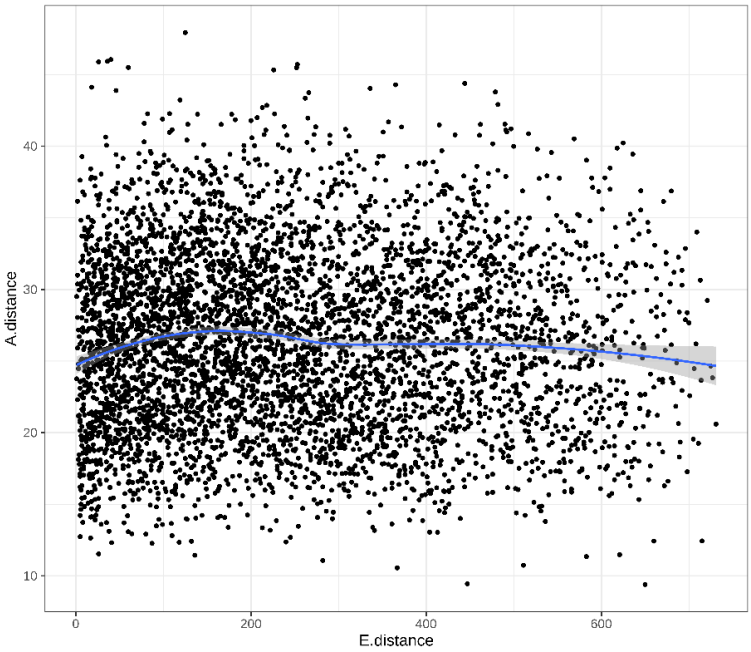


Figure 3-1 - Variance partitioning from a distance-based redundancy analysis on the pairwise Aitchison compositional distances among bioaerosol samples for (a) 16S and (b) ITS showing the variance explained by one-day, three-day, and one-week back trajectories.

3.4.2 Time-Dissimilarity was Non-Linear

Non-linear time-dissimilarity relationships were evident for both amplicons (**Figure 3-2**). The relationships were close to cubic in shape, both showing inflexion points at around 200 days and 400 days, suggesting a cyclical seasonal relationship. The non-linear correlation coefficient for 16S was 0.1 and ITS was 0.244, both were statistically significant (*P*-16S = 0.002, *P*-ITS < 0.0001). The patterns for year one of the study (0 to 365 DSS on the *x*-axis) were similar for both amplicons, but year two showed a flattening in 16S, and a steepening for ITS.

a)



b)

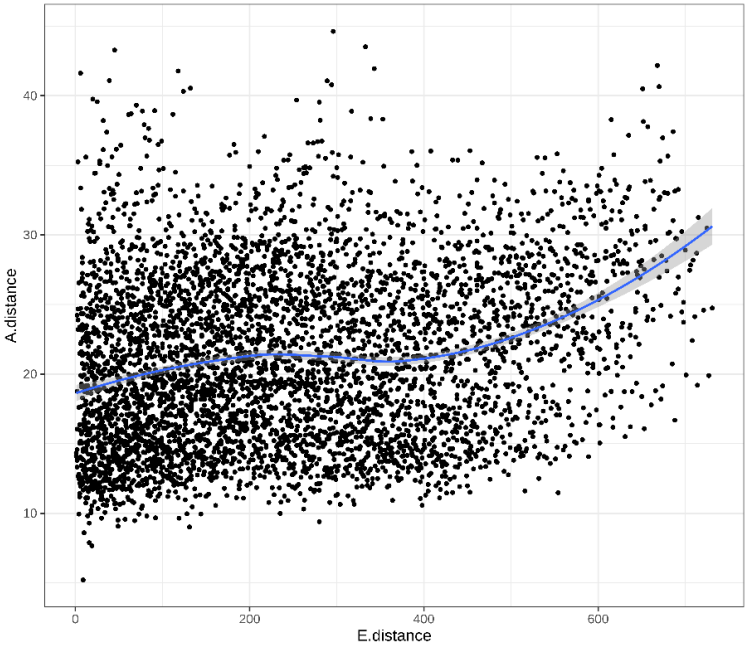


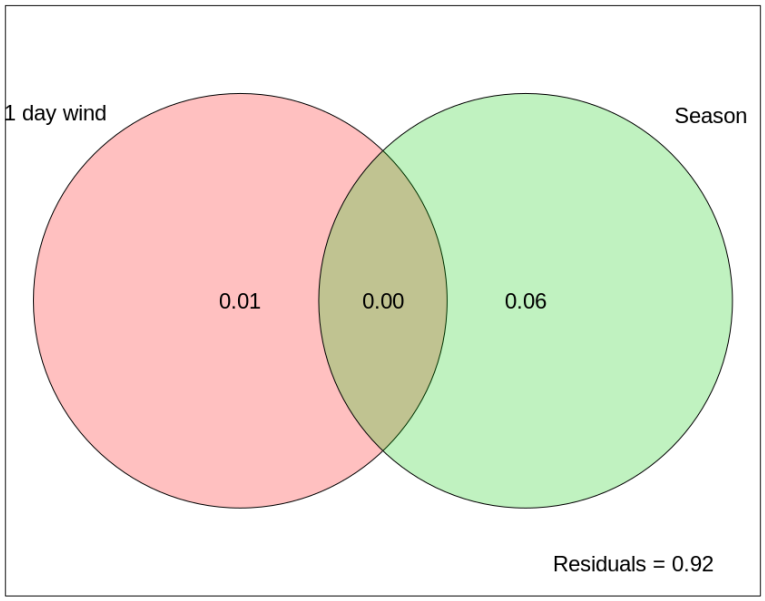
Figure 3-2 - Time-dissimilarity plots for a) 16S and b) ITS of Aitchison compositional sample distances against Euclidean temporal dissimilarities between sample times (number of days). Blue lines show the fit of a loess smoother and the 95% confidence interval (grey envelope).

3.4.3 Strong Temporal Patterns in Bioaerosol Community Composition

Season and DSS[1 - 3] were the most important correlates with the bioaerosol community for both amplicons, explaining the greatest amount of variation in the db-RDAs (6% for 16S, and

23% for ITS) (**Figure 3-3**). Wind (and temperature and weather for ITS) were of secondary importance. The R^2 of the model was substantially higher for ITS than 16S (33% versus 8%). Seasonality and a difference between year one and year two in composition measured for ITS were evident in the NMDS ordination. The 16S ordination showed minimal differentiation between years, consistent with the time-dissimilarity and the db-RDA analysis (**Figure 3-4**). Both NMDS ordinations showed clustering of the day-degree variable over the course of the year, with yellow and dark purple often being adjacent as they represent January and December, respectively.

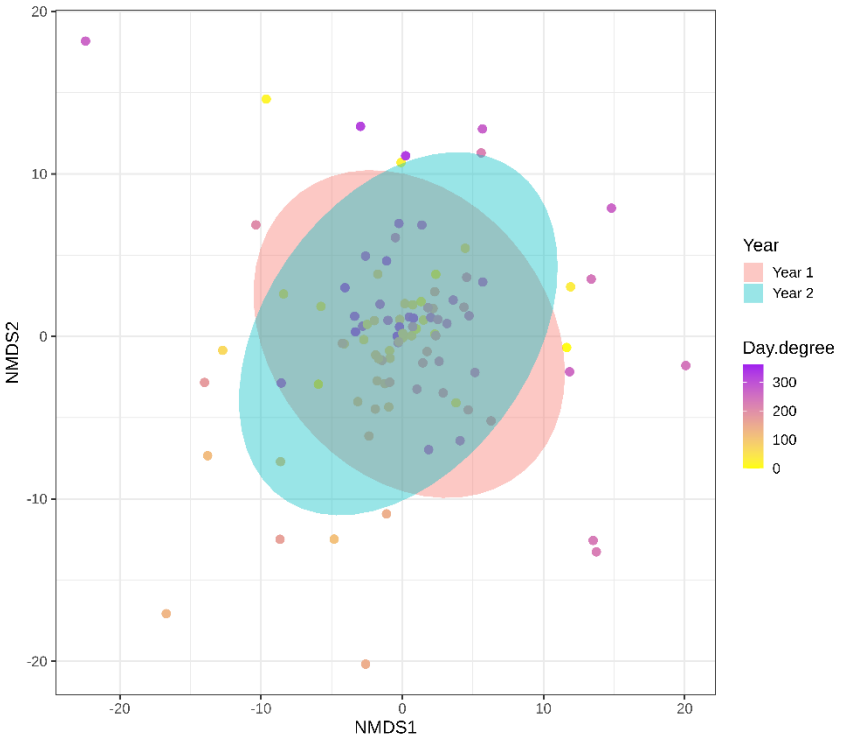
a)



b)

Figure 3-3 - Variance partitioning of the results from a db-RDA on the Aitchison compositional distances among samples for (a) 16S and (b) ITS, showing the variance explained by season, days since start (DSS^{1-3}), one-day wind (16S) or three-day wind (ITS), and temperature and weather for ITS only.

a)



b)

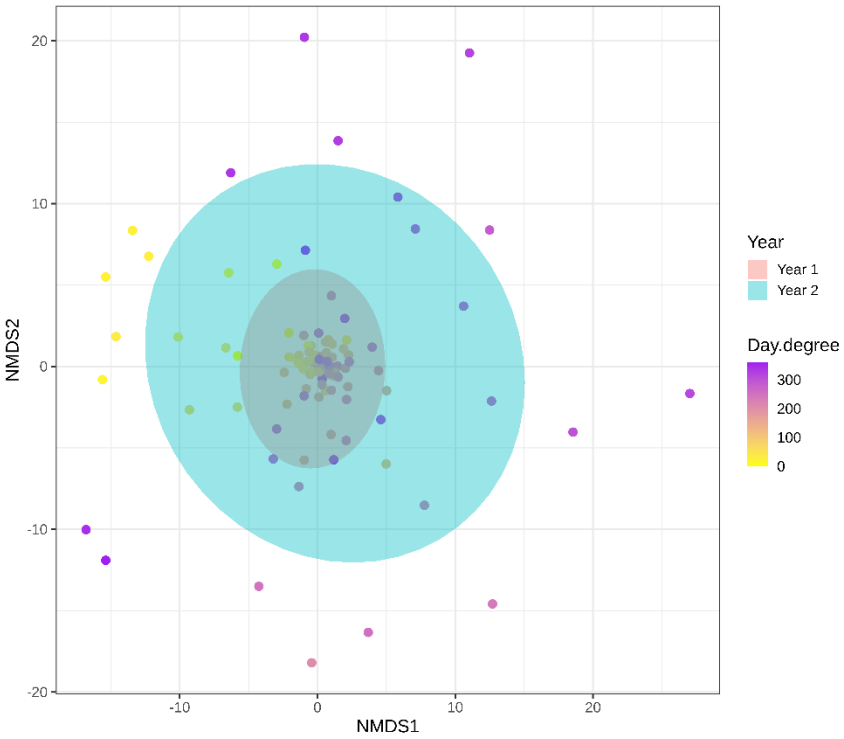


Figure 3-4 - NMDS ordination of the a) 16S or b) ITS Aitchison distances among samples. The study year is indicated by polygons representing the *t*-distribution of the year one and year two points. Days in circle degrees are indicated by the gradient colour. Stress on 16S NMDS was 0.18 (dimensions one and two are presented above), the three-dimensional

solution was selected as the two-dimensional solution had stress greater than 0.2 so was unreliable. Dimensions one and three and two and three are presented in Supplementary Materials (E.2.1). The two-dimensional solution for the ITS NMDS was selected, as stress was 0.18. Higher dimensional solutions had lower stress values for both amplicons.

3.4.4 Differences Visible Among Months and Wind Trajectory Clusters, in Relative Taxon Abundances

Unfiltered bacterial reads numbered 1,171,139, with 5,235 ASVs inferred. Unfiltered fungal reads numbered 2,760,306, comprised of 3,419 ASVs.

Relative Abundance by Month – 16S

Many of the common bacterial genera (defined as per filtering described in **Figure 3-5** legend below) showed clear seasonal trends, for instance *Ralstonia* spp. were common early in the year and much less prevalent between July and October. Conversely, chloroplasts/plastids were much more common in winter months. *Methylobacterium* spp. were much more prevalent in August and September in year one. *Listeria* spp. peak in relative abundance between August and October 2018. Some genera, however, did not show marked seasonal patterns, and were found throughout the two-year sampling period, such as *Pseudomonas* spp. and an unclassified organism in the genus *Cyathea poeppigii* (in red, likely a chloroplast from a tree fern). Sampling year one and two were reasonably consistent.

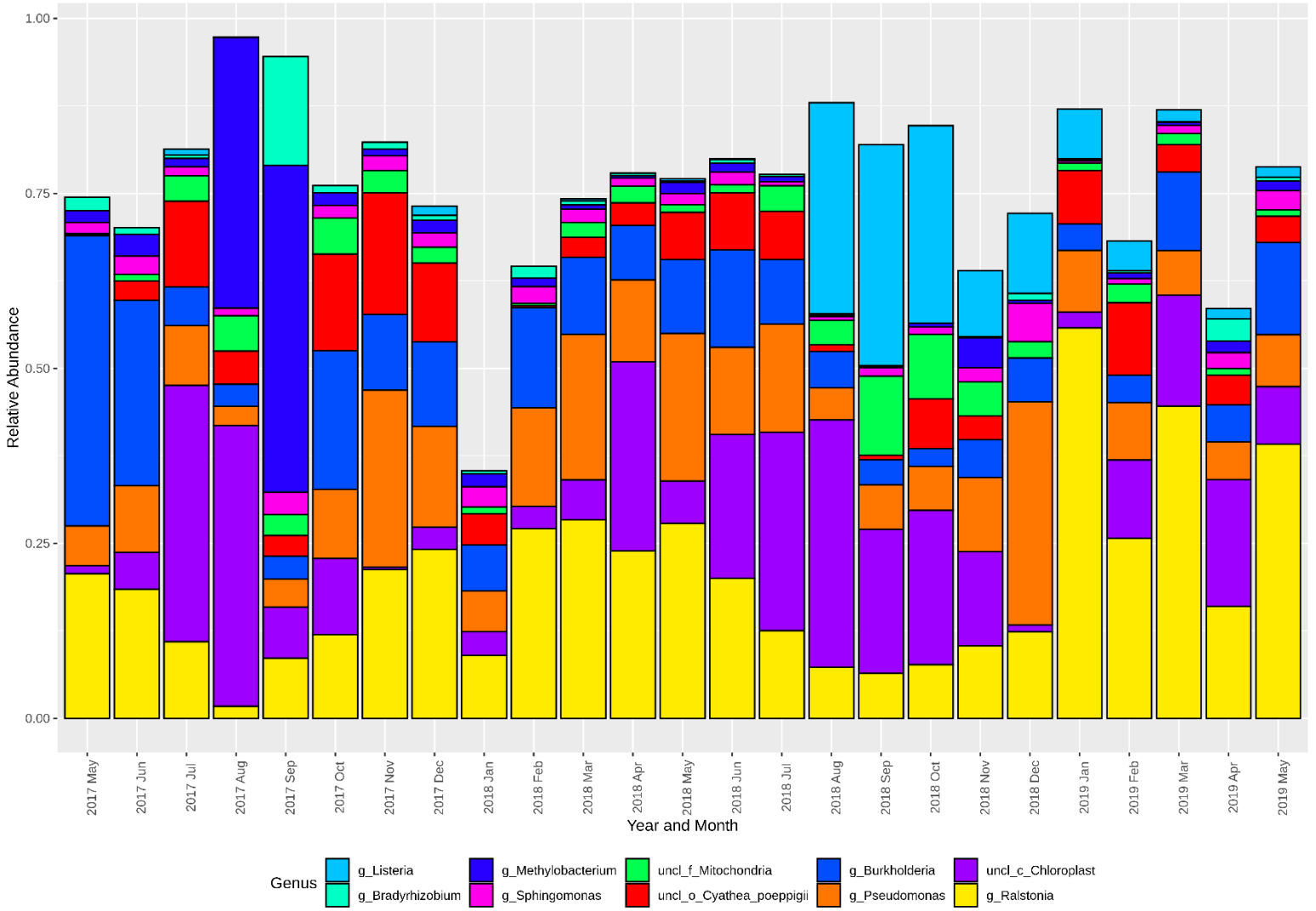


Figure 3-5 - Relative abundance by month for 16S. ASVs with less than 300 reads and samples with less than 100 reads were filtered out. Genera with greater than 1.2% of reads for each season were included in the bar plots.

Relative Abundance by Month – ITS

Marked seasonal variation was observed amongst the fungal genera (**Figure 3-6**). The dominant genera totally shifted several times over the course of the two years (filtering criteria described in **Figure 3-6** legend below). *Phoma* spp. were very common from December 2018 until May 2019 and in June 2017. *Gibberella* spp. dominated the bioaerosol samples in August 2017, December 2017, and most months between February 2018 and September 2018. *Meyerozyma* spp. dominated in January 2018, while *Botryotinia* spp. were dominant around October in both years. *Leptosphaerulina* spp. were common between February and April 2019. Fungi displayed greater variability between year one and year two than bacteria.

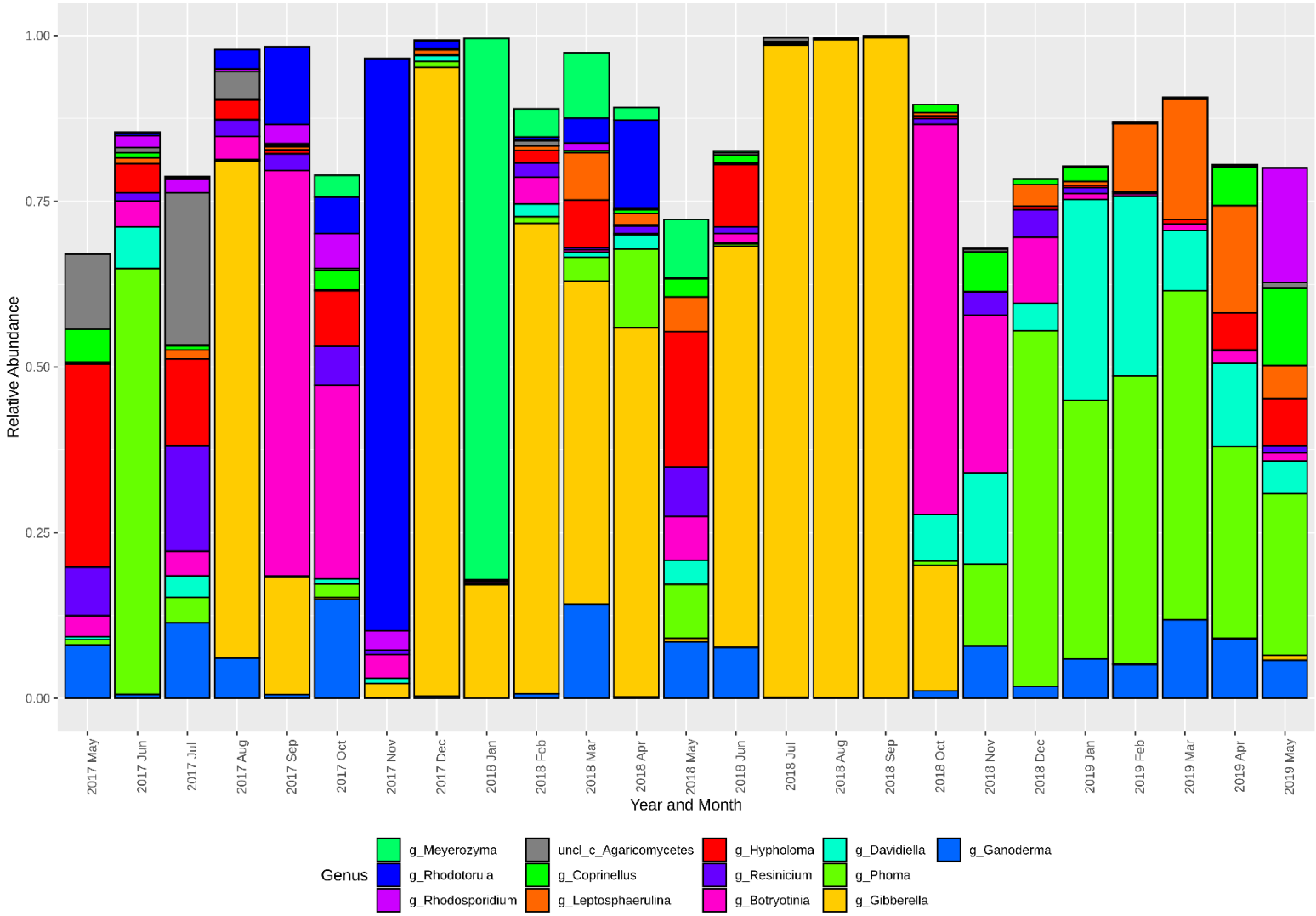
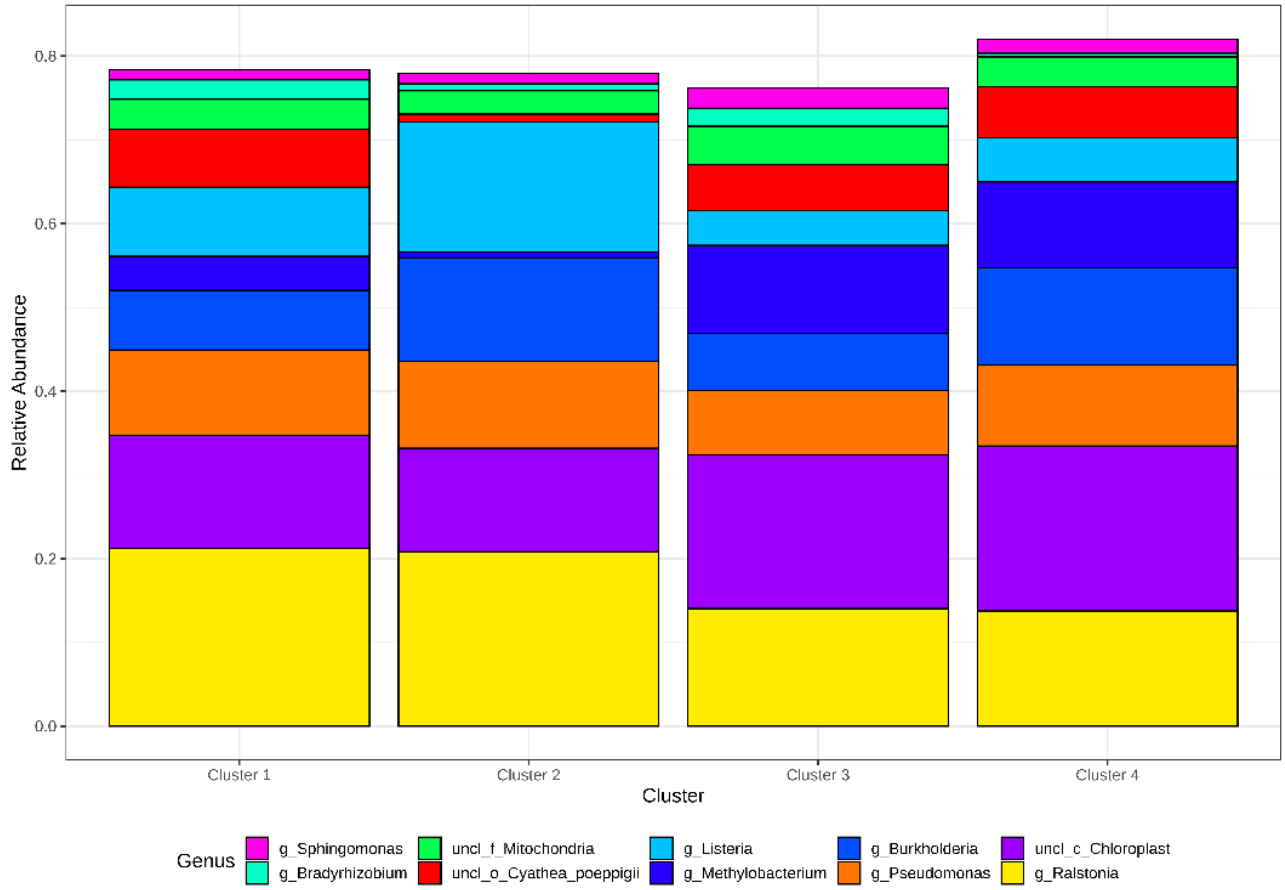


Figure 3-6 - Relative abundance by genus by month for ITS. ASVs with less than 500 reads and samples with less than 100 reads were filtered out. Only genera with at least 1% of the reads were included in the bar plot.

Relative Abundance by Wind Back-Trajectory Cluster – 16S

The four back-trajectory clusters appeared to have similar bacterial compositions (**Figure 3-7**). Cluster two had a greater proportion of *Burkholderia* spp., fewer *Methylobacterium* spp. and more *Listeria* spp. Filtering criteria described in **Figure 3-7** legend below.

a)


Relative Abundance
0.0
0.2
0.4
0.6
0.8
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster
Genus
g_Sphingomonas
g_Bradyrhizobium
uncl_f_Mitochondria
uncl_o_Cyathea_poeppigii
g_Listeria
g_Methylobacterium
g_Burkholderia
g_Pseudomonas
uncl_c_Chloroplast
g_Ralstonia

b)

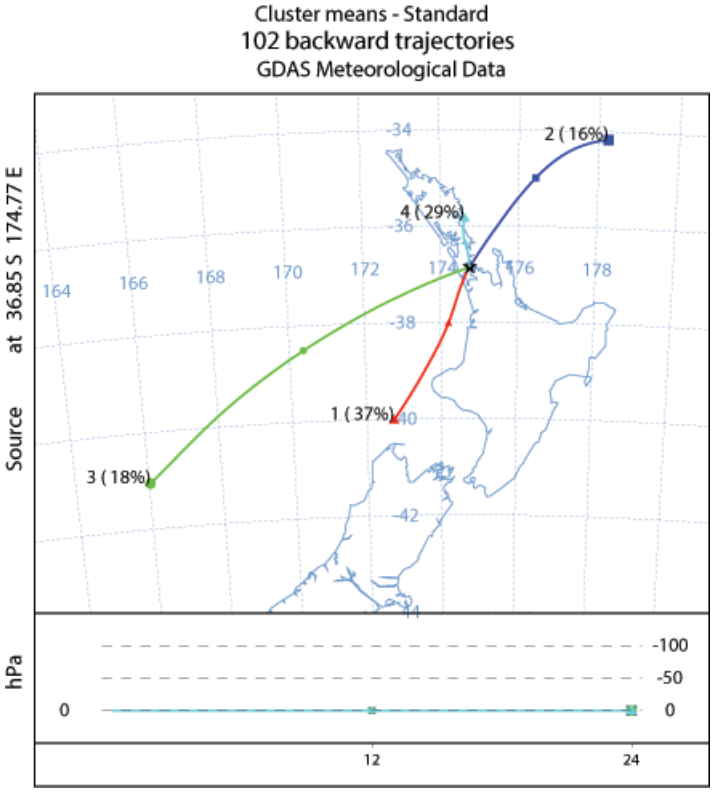
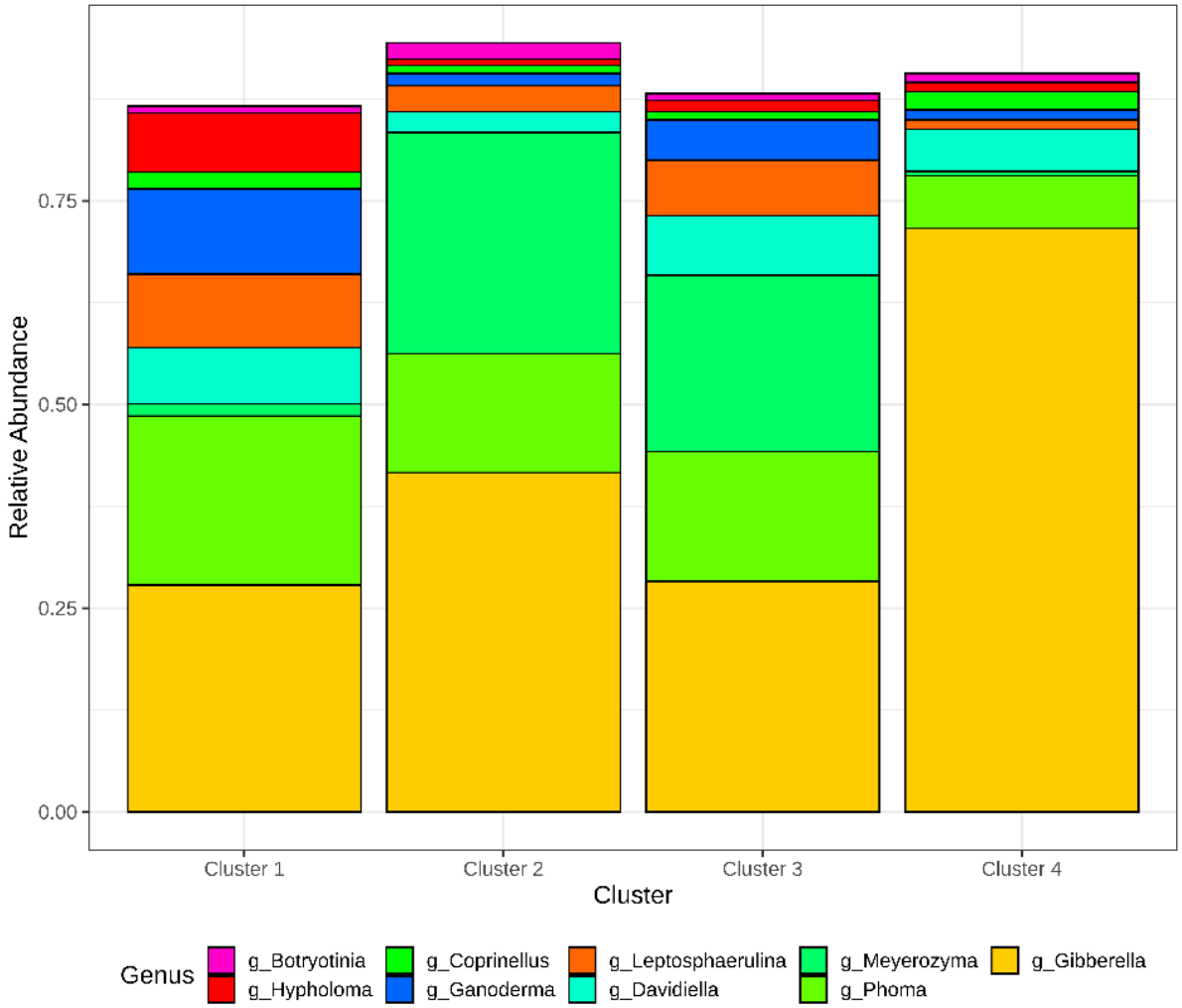

Cluster means - Standard
102 backward trajectories
GDAS Meteorological Data
Source at 36.85 S 174.77 E
2 (16%)
4 (29%)
1 (37%)
3 (18%)
-34
-36
-38
-40
-42
164
166
168
170
172
174
176
178
hPa
-100
-50
0
0
12
24

Figure 3-7 - Relative abundance by back-trajectory cluster for 16S. ASVs with under 300 reads and samples with less than 100 reads were filtered out. a) abundance by genus for each wind trajectory. Only genera with at least 1.2% of the reads were included in the bar plot. b) Paths of the four clusters identified over the sampling period in the previous 24 hours before sampling, generated by the NOAA HYSPLIT model based on GDAS meteorological data. % next to each cluster indicates proportion of trajectories assigned to that cluster.

Relative Abundance by Wind Back-Trajectory Cluster – ITS

Fungal assemblages showed some variability for different trajectory clusters (**Figure 3-8**). *Hypholoma* and *Ganoderma* spp. were more prevalent in cluster one, while *Meyerozyma* spp. were more prevalent in clusters two and three. Filtering criteria described in **Figure 3-8** legend below.

a)


Relative Abundance
0.75
0.50
0.25
0.00
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster
Genus
g_Botryotinia
g_Hypholoma
g_Coprinellus
g_Ganoderma
g_Leptosphaerulina
g_Davidiella
g_Meyerozyma
g_Phoma
g_Gibberella

b)

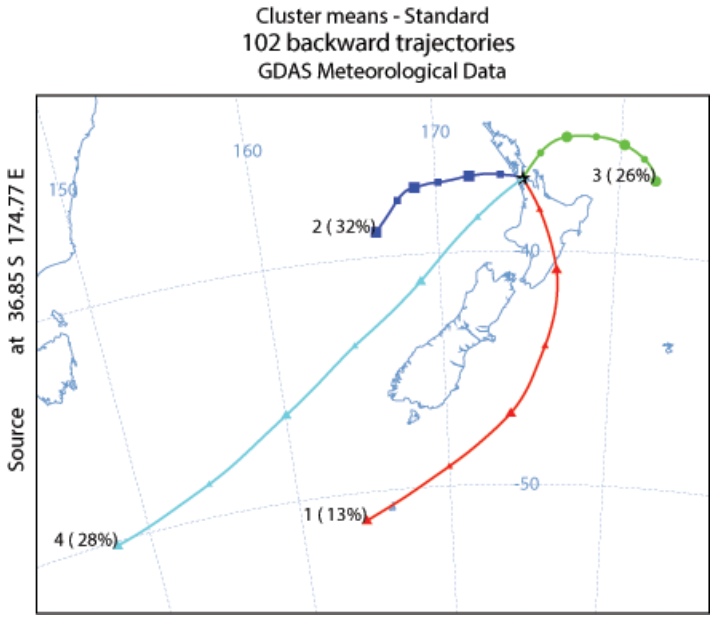

Cluster means - Standard
102 backward trajectories
GDAS Meteorological Data
Source at 36.85 S 174.77 E
150
160
170
-40
-50
3 (26%)
2 (32%)
1 (13%)
4 (28%)

Figure 3-8 - Relative abundance by back-trajectory cluster for ITS. ASVs with under 500 reads and samples with less than 100 reads were filtered out. a) abundance by genus for each wind trajectory. Only genera with at least 1% of the reads were included in the bar plot. b) Paths of the four clusters identified over the sampling period in the previous 72 hours before sampling, generated by the NOAA HYSPLIT model based on GDAS meteorological data. % next to each cluster indicates proportion of trajectories assigned to that cluster.

3.5 Discussion

Urban aerosol microbial community composition varied strongly and, to some extent, predictably in time, showing clear seasonal variation. Air-mass trajectories of differing spatiotemporal lengths differed in their ability to predict variation in bioaerosol community structure and one or three-day trajectories provided the greatest explanatory power in this dataset. Therefore, hypothesis 1 - Temporal wind hypothesis (air-mass trajectories of differing spatiotemporal lengths differ in their ability to predict variation in bioaerosol community composition. Specifically, one- or three-day trajectories provide the greatest explanatory power relative to one-week back-trajectories) failed to be rejected. Similarity in community composition of samples declined with their increasing temporal separation in a non-linear fashion, therefore hypothesis 2 - Time-dissimilarity hypothesis (similarity in community composition of samples declines with their increasing temporal separation in a non-linear fashion, due to seasonality) failed to be rejected. Temporal variation in bioaerosol community structure (composition and diversity) at the Auckland CBD was correlated with season, days since the start of the study cubed, geographic origin of sampled air and temperature. Relative humidity and other measured variables were not correlated with bioaerosol communities recovered in this study. Therefore, hypothesis 3. a) environmental drivers (temporal variation in bioaerosol community structure (composition and diversity) at the Auckland CBD can be predicted by season, geographic origin of sampled air, temperature, weather and relative humidity) failed to be rejected with respect to seasonality, temperature and air mass source only. The relative abundances of common and rarer taxa fluctuated by season and air mass source. Therefore, the alternative hypothesis 3. b) (relative abundances of common and rarer bacterial and fungal taxa fluctuate by season and air mass source) failed to be rejected. Fungi displayed greater seasonality, therefore, hypothesis 4 (temporal patterns of 16S bacterial and ITS fungal microbial aerosol communities are expected to be different. Specifically, fungi display more variability in general and greater seasonality) failed to be rejected. This study represents the longest continuous time series data currently available on urban bioaerosols. Their variability is substantial and correlated with measurable factors (from 8% for 16S, to 33% for ITS), so it is now possible to make predictions regarding future changes. This means that estimates of microbial exposure of pathogens and allergens of city dwellers can be improved and that we can begin to understand ecosystem connectivity via the atmosphere.

3.5.1 One and Three-Day Wind Back-Trajectories were not Strongly Correlated to the Bioaerosol Community

The overall correlation of air mass source only with bioaerosol communities sampled was low for both amplicons, with an R^2 of 2-3%, which contrasts with some of the body of literature suggesting that air mass source is the key determinant of bioaerosol variation (Archer et al., 2020; Woo et al., 2013), but is consistent with other studies which suggest local sources are more important than distant ones (Bowers et al., 2013; Bowers et al., 2012). One and three-day trajectories had higher R^2 values than one-week trajectories, (see section 2.5.2 Wind Back-Trajectory Duration), consistent with the standard practice in the literature of using a three-day trajectory (Archer et al., 2020). These results suggest that one-day trajectories may have better predictive value for fine-scale temporal compositional variation than three-day trajectories, and that certainly comparison of alternative trajectory lengths would be prudent for all studies.

3.5.2 Time-Dissimilarity was Non-Linear

ITS and 16S showed similar cubic trends in the time-dissimilarity analysis, suggesting seasonal variability, as expected. Sampling began in May 2017, and the inflexion point at 200 days (samples more dissimilar than expected if a linear relationship with time existed) is the following November or December. A larger difference between summer and winter supports seasonality. A dip in the line of best fit occurs at 370 days – May again, where samples are more similar than expected with a linear relationship. Given this is the same season, this also suggests seasonal effects. From 400 days, or June 2018 onwards, samples become increasingly dissimilar, again this is moving from the same to different seasons being compared; increasing differences suggest seasonality. Clear seasonality was consistent with the body of literature (Bowers et al., 2013; Bowers et al., 2012; Els et al., 2019; Franzetti et al., 2010; Mhuireach et al., 2020; Woo et al., 2013). Although no other urban bioaerosol studies sampled across multiple years, seasonal variations were always reported within the sampling period. The possibly more pronounced temporal variation for fungi was also suggested in literature that looked at both bacteria and fungi (Woo et al., 2013).

3.5.3 Clear Temporal Patterns in Bioaerosol Variability

Season and DSS^{1-3} had an R^2 of 6% for 16S, and 23% for ITS in the db-RDA, and were the most important correlates of variation in microbial composition, as hypothesised. DSS^3 modelled cyclical changes in the data, as well as the year-on-year change. The season categorical variable appeared to model more complex relationships in the data, as it explained further variation on top of the DSS^3 variable. NMDS ordination was consistent with the db-RDA result for both amplicons. Separation of communities based on the day degree variable supporting seasonal variation over the year was present and more pronounced with ITS. Differentiation between the first and second year of study was evident with ITS but not 16S. Both ordinations, particularly 16S, are clumped towards a central distribution. Wind (and weather and temperature for ITS) were of secondary importance. Both the seasonality, and lesser influences from temperature, are consistent with previous

research (Franzetti et al., 2010; Woo et al., 2013), but weather was not noted as a correlate in the literature reviewed. Franzetti et al. (2010) pooled samples for three-month periods representing summer and winter only, whereas Woo et al. (2013) sampled weekly for one year, for 24 hours at a time. These differing approaches converging on the same result lend it strong support. The secondary importance of wind origin is consistent with some literature (Bowers et al., 2013; Bowers et al., 2012), but contrasts with other studies which postulated wind being the key driver behind bioaerosol variation (Archer et al., 2020; Woo et al., 2013). The studies suggesting wind is of greater importance tend to be over a much shorter duration; however, Woo et al. (2013) had a very similar protocol to the one adopted here. While it is well known that some bioaerosols can be transported over great distances in the atmosphere (Burrows et al., 2009a), it is more intuitive that local sources of bioaerosols have greater impact than distant ones. Other variables, which were not observed to correlate with the bioaerosol community, were of inconsistent importance in the literature (Fröhlich-Nowoisky et al., 2009; Woo et al., 2013).

3.5.4 Bacterial and Fungal Genera Show Different Relative Abundances over Time and by Wind Back-Trajectory Cluster

The bacterial and fungal genera detected were consistent with previous bioaerosol studies and known to be environmental organisms or associated with plants, humans or other animals. Genera frequently identified included pathogens such as *Listeria* spp. and allergens such as *Alternaria* spp. (Priyamvada et al., 2017). While *Pseudomonas, Ralstonia* and *Burkholderia* are commonly reported contaminants, their consistency with other bioaerosol studies, biological niches and presence despite stringent decontamination procedures indicate they also represent a genuine constituent of the bioaerosol community.

Several bacterial genera showed monthly trends, for instance *Ralstonia* spp. (a widely distributed soil and water bacterium) were common early in the year while chloroplasts/plastids were much more common in winter months. Chloroplasts, presumably from pollen and other plant fragments, are abundant in similar bioaerosol studies, and have been observed to be seasonally variable (Brodie et al., 2007b; Franzetti et al., 2010; Woo et al., 2013). Generally, they are most common in spring or summer (Woo et al., 2013), which is inconsistent with the data from this study. It is likely that Auckland's mild sub-tropical climate results in flowering less tied to a season. Other genera had variable distribution through the year. *Bradyrhizobium* spp., a plant-associated nitrogen fixing bacterium (Ormeño-Orrillo & Martínez-Romero, 2019) peaked in September 2017, at a time of rapid plant growth. *Methylobacterium* spp. (an opportunistic human pathogen and soil and water associated bacterium (Green, 2006)) was much more prevalent in August and September 2017. *Listeria* spp. peaked in relative abundance between August and October 2018. Many of the above peaks were around late winter or early spring, when the Auckland climate is mild and damp, encouraging bacterial growth (Dannemiller, Weschler, & Peccia, 2017; De Silvestri, Ferrari, Gozzi, Marchi, & Foschino, 2018). Thus, these results support the conclusion that bioaerosol communities display strong seasonality and that this relates

directly to the responses of microbes to environmental conditions such as temperature and humidity.

Fungal genera showed marked monthly variation, with the dominant genera changing several times, in line with previous analyses suggesting greater seasonal variation in fungi (Bowers et al., 2013; Shigyo et al., 2019). *Meyerozyma* spp. yeasts dominated in January 2018, *Phoma* spp. plant pathogens from December 2018 until May 2019 and in June 2017, *Gibberella* spp. (a plant pathogenic mould) dominated the bioaerosol samples in August 2017, December 2017, and most months between February 2018 and September 2018, and another plant pathogen, *Botryotinia* spp. was dominant around October in both years. *Leptosphaerulina* spp., a further plant pathogen, were common between February and April 2019. Many fungi sporulate at particular times of the year (Lagomarsino Oneto, Golan, Mazzino, Pringle, & Seminara, 2020) and it is supposed that each different genus sporulating at different times causes shifts in the dominant genera. The case of the between year variation is unclear.

Bacterial genera showed slight differentiation by back-trajectory cluster. A Tasman Sea-sourced trajectory (cluster one) was associated with *Ralstonia* and *Pseudomonas* spp., while *Bradyrhizobium* spp. was associated with clusters approaching from the Pacific Ocean and the east coast of New Zealand (clusters two and four). No associations between certain wind paths and genera were apparent. Fungal assemblages showed some variability for different trajectory clusters. *Hypholoma* and *Ganoderma* spp. were more prevalent in the Tasman Sea cluster (cluster one), while *Meyerozyma* spp. were more prevalent in the Pacific and Tasman clusters (two and three). Cluster one transited over much of the North Island whereas the other clusters were marine, suggesting possible associations between a terrestrial path and *Hypholoma* and *Ganoderma* spp.

3.5.5 Bacterial and Fungal Genera Show Differential Responses to Time and Other Factors

The non-linear correlation coefficient was greater for ITS than 16S. The first year of the study looked consistent for both amplicons, but interestingly, the second year showed greater variation for ITS, whereas there was not much visible year-on-year variation for 16S, which was consistent with the species bar plots. This could be due to the apparent greater variability overall with ITS, emphasising trends which are more subtle for 16S. Since the ecological niches of fungi can be quite different to those of bacteria, it is also possible that fungi responded differently to environmental changes between year one and year two to bacteria. Bacteria are smaller than fungi, so likely have longer residence times in the atmosphere (Archer et al., 2019; Burrows et al., 2009a), which could have reduced their observed temporal variability. In agreement with a greater non-linear correlation coefficient for ITS, the R^2 of the db-RDA was substantially higher for ITS than 16S (33% versus 8%). This could be driven by differences in read counts between bacteria and fungi. While more bacterial ASVs were identified, the fungal read counts were more than double the bacterial

ones. This was caused by several fungal samples with very high read counts, while the bacterial samples had a narrow range of read counts. As many fungi are multicellular, versus unicellular bacteria, fragments with many cells could sometimes be sampled for fungi, creating this pattern. Fungi may also be more seasonally dependant, which could explain some of the differences in observed R^2 values. Increased seasonal dependence for fungi has been previously observed (Shigyo et al., 2019), and this is supported by the large shifts in dominant genera, visible in the genus bar plots for fungi.

3.6 Conclusions

Different durations of back-trajectories can show different predictive potential for bioaerosol communities, with shorter trajectories showing higher R^2 values in this case. Different back-trajectory lengths should be considered based on the characteristics of the study they are used for. From this work, it appears that back-trajectory cluster is of secondary importance to the bioaerosol communities compared to local variables (particularly temperature), in contrast to some recent literature. Fungi in particular, exhibit marked seasonality, probably due to timing of sporulation events. Fungi also seem to have higher read counts and correlation coefficients from modelling than bacteria. Biological features appear to influence patterns of bioaerosol variation, with fungi and bacteria responding differently to the variables measured, possibly due to differences in size and presumed atmospheric residence times.

Chapter 4 - Diel Variation and Intercontinental Connectivity of Antarctic Bioaerosols

4.1 Abstract

The isolated continent of Antarctica provides an ideal model system for understanding long-range bioaerosol transport, of crucial importance in a rapidly warming world undergoing unprecedented ecosystem change. Antarctic air is little studied and poorly understood, due to technical and logistical constraints, but harbours a surprising diversity of microorganisms. Here, the first Antarctic high-temporal resolution bioaerosol dataset is presented. The aim of this work was to understand the nature and drivers of diurnal variation in Antarctic bioaerosols and the level of atmospherically mediated microbial transport from Antarctica to Aotearoa New Zealand. A Coriolis high-volume air-sampler was specially modified to operate in sub-zero temperatures, dramatically reducing required sampling durations. Non-linear temporal variation was detected for fungi; bacterial data, however, were inconclusive, with no obvious correlations detected with any measured variable. Bacteria and fungi showed little variance in composition over time. Fungal temporal variation was unlinked to measured variables thought to affect bioaerosols (i.e. wind speed, UV, temperature), suggesting a need for further understanding of influences of Antarctic bioaerosols. Bacteria consistant between Antarctica and New Zealand (NZ) increased noticeably when NZ air was coming from Antarctica, suggesting intercontinental transport for bacteria at notable rates. Fungi did not show this pattern, postulated to be due to their larger size and suspected shorter atmospheric residence times resulting in reduced rates of long-range transport.

4.2 Introduction

Antarctica's pristine and unique landscape is highly vulnerable to climate change and is rapidly warming. Antarctica is extremely isolated, due to minimal human and animal movements, physical distance and air and water currents, which further reduce movement from the open air and sea onto the continent (Pearce et al., 2016). This means that the atmosphere is the key transit route for colonising microbes to the Antarctic (Pearce et al., 2009). Urgent understanding of the current ecosystems and processes is needed, before they change forever (Smith, 1994). Understanding of propagule pressure from the atmosphere is essential for projecting ecosystem change in the future under a warming climate. As temperatures increase, less selective environmental filters will allow more temperate organisms to gain a foothold (Kennedy, 1994). The dominance of the atmosphere for microbial transport to Antarctica, coupled with the simple microbial ecosystems in-situ also provide a fantastic model system to understand global bioaerosol circulation. Since bioaerosols include pathogens and invasive species and ecosystem change is accelerating, this understanding is crucial (Pearce et al., 2009). The ongoing COVID-19 pandemic is a salient case in point. It demonstrates how an aerially transmitted pathogen can wreak havoc in our day-to-day lives and how understanding of transmission via the atmosphere can be crucial to limiting the spread of microorganisms across spatiotemporal scales.

Despite its importance, the Antarctic aerosphere and intercontinental connectivity remain poorly understood, due to technical and logistical constraints (see section 1.3.6 Bioaerosol Sampling: Challenges and Solutions). Only seven DNA-based studies have been

performed on Antarctic bioaerosols to date. Just one of these studies included fungi, with the rest limited to bacteria. Antarctica represents a very challenging environ to inhabit. Temperature, UV exposure, wind and relative humidity are extreme (Pearce et al., 2009). As a result, the continent harbours very few vascular plants or land animals. This limited biological activity and stability of frozen surfaces reducing emissions, results in low bioaerosol concentrations and diversity (Burrows et al., 2009a). There is evidence for atmospheric biological transfer from elsewhere, particularly South America to the Antarctic peninsula (Smith, 1991), and the bioaerosol studies that have been done to date consistently find globally distributed taxa (Bottos et al., 2014). However, extra-continental exchange appears to be limited for bacteria and fungi (Archer et al., 2019; Bottos et al., 2014), which is in line with previous estimates from modelling (Burrows et al., 2009a). Further, data available suggest that bacteria are more able to undergo long-distance transport than fungi, because they have greater residence times in the atmosphere (Archer et al., 2019). There is minimal evidence of distance-decay with airborne bacteria or fungi (Archer et al., 2019). Marine influence on terrestrial aerosols appears to be variable, with one study detecting marine taxa, while others did not (Archer et al., 2019; Bottos et al., 2014). Proteobacteria and firmicutes (such as *Bacillus* spp.) and basidiomycetes yeasts dominate Antarctic bioaerosols sampled (Archer et al., 2019; Bottos et al., 2014). These organisms are often spore formers, which could facilitate their survival in the atmosphere (Bottos et al., 2014). Thermophiles were present, thought to be launched by steam from vents at local volcanic sites (Bottos et al., 2014). This limited literature shows more studies at a greater range of spatiotemporal scales are required to develop a basic understanding of how the Antarctic aero-microbiome varies and what the drivers are.

It is understood that wind source affects bioaerosols (Archer et al., 2020; Woo et al., 2013), but optimal trajectory lengths for correlation with Antarctic bioaerosols remain unclear. Studies which considered air-mass source have used either three-day and two-week back-trajectories or one-week back-trajectories (Archer et al., 2019; Bottos et al., 2014), but no comparison of the different trajectory lengths is available. It has been suggested that Antarctic bioaerosols may have longer residence times than average (Archer et al., 2019; Bottos et al., 2014; Burrows et al., 2009b), so possibly longer trajectory lengths may be more informative. Sampling performed by Bottos et al. (2014) used dry filters and took 55 days per sample. Archer et al. (2019) improved on this by sampling for four hours on different days in several locations, but no datasets exist at higher temporal resolutions. Therefore, we do not know how Antarctic bioaerosols change at a fine temporal scale – hour-by-hour or even day-by-day. Diel variation is expected, as although the sun does not set in summer, nor rise in winter, predictable changes in temperature, wind direction, UV, etc. occur over a 24-hour period. These factors are known to influence bioaerosols in other parts of the world (Burrows et al., 2009b). No quantification of the relative impacts of the different environmental variables which are thought to affect Antarctic bioaerosols has occurred, such as wind speed, temperature, and humidity. Understanding of global connectivity of Antarctic aerial microbes requires enhancement, particularly for fungi, which

often outnumber bacterial reads in aerosol samples but are far less studied. Sampling locations in Antarctica are also extremely limited (**Figure 1-10**).

To start to address some of these gaps, bioaerosol sampling was performed in January 2018 at Taylor Valley in the McMurdo Dry Valleys, one of the largest-ice free and most biologically diverse areas in Antarctica (Archer et al., 2019). A modified Coriolis high-volume liquid impinger (Carvalho et al., 2008) was run for several days with samples removed every hour. Modifications were required to prevent freezing of the collection liquid during sampling (see section A.2.1 Methods for Operation in Sub-Zero Environments). Samples were then pooled in two-hour blocks. Sampling occurred close to camp at Spaulding Pond in the Taylor Valley and about a kilometre from camp in the centre of Taylor Valley (referred to as the Remote Location or the Deposition Zone). In order to assess extra-continental exchange, sampling was also performed in Aotearoa New Zealand. Baring Head lighthouse, near Wellington, was selected, as air masses from the Southern Ocean frequently arrive there. Therefore, it was thought that this would be the best place in New Zealand to detect microorganisms of Antarctic origin. The number of taxa (amplicon sequence variant or ASVs and genera) that were the same between Antarctica and Baring Head air of Antarctic origin, and between Antarctica and Baring Head air of non-Antarctic origin, was calculated. This was used to assess the degree of microbial exchange between the two locations, and if shared ASVs increased when air was coming from Antarctica. ASVs in common were expected as globally dispersed bioaerosols are consistently detected in bioaerosol studies (Bottos et al., 2014), so does not necessarily provide evidence for substantial biological exchange between the regions. However, an increase in shared ASVs when wind was from Antarctica would strongly suggest wind-mediated transport was occurring between the two locations.

4.2.1 Hypotheses

The aims of this research were to understand the nature and drivers of fine-temporal scale variation of Antarctic bioaerosols and to quantify the level of microbial transport mediated by the wind between Antarctica and Aotearoa New Zealand. To address these aims, the following hypotheses were tested:

1. Temporal wind hypothesis: Biological (i.e. particle size) and environmental characteristics (affecting atmospheric residence times) affect optimal spatiotemporal lengths of air-mass trajectories in relation to their ability to predict fine scale temporal variation in related bioaerosol community structure. Specifically, Antarctic bioaerosols were expected to have longer optimal trajectory lengths than elsewhere (one week), and longer trajectory lengths were anticipated for bacteria.
2. Time-dissimilarity hypothesis: Changes in environmental variables that are known to affect bioaerosols (such as UV and temperature) were expected to drive temporal variation in the bioaerosol community, within a 24-hour period, and between different days studied. The relationship between pairwise sample dissimilarity and time was expected to be non-linear due to cyclical day-night periods.

3. Environmental drivers hypotheses:
    a. Geographic origin of sampled air, sample location, temperature, relative humidity, wind speed and amount of ultraviolet light (UVA, UVB and UVC) drive fine-scale (hourly and daily) temporal variation in Antarctic bioaerosol community structure (composition and diversity), with differing levels of relative importance.
    b. Relative abundances of taxa fluctuate by day, time of day and air mass source.
4. Intercontinental connectivity hypotheses:
    a. A minority of globally distributed ASVs (identical sequences) were expected to be observed in both Antarctica and New Zealand. Bacteria were expected to show a higher propensity for long-range aerial transport due to their longer atmospheric residence times and smaller size than fungi, therefore higher numbers of bacterial globally distributed ASVs and genera were expected.
    b. Aerial transport was expected to occur from Antarctica to New Zealand, at greatest rates for bacteria, resulting in more ASVs/genera in common (greater overlap in taxa) between Antarctica and New Zealand when Baring Head air came from Antarctica.
    c. Bacterial composition was expected to be more affected by air mass source than fungal composition, and fungi were expected to be more dependent on local environmental variables (temperature, relative humidity and weather) than bacteria.
    d. Relative abundance of bacteria and fungi were expected to vary depending on air mass source, as different trajectories passed over different regions with variable microbial emissions.

4.3 Methods

4.3.1 Field Sampling and Environmental Data Collection

Sampling was performed in the McMurdo Dry Valleys, Antarctica and at Baring Head Lighthouse, Wellington, NZ. Baring Head was sampled continuously in one-hour increments using the Coriolis into phosphate buffered saline (PBS) for a period of three days in July 2017. In the Dry Valleys, the Coriolis was used to sample continually in one-hour increments in PBS for one day. This was repeated three times at the Spaulding Pond location, and for one whole day and two partial days at the remote sample location (Deposition Zone) in Taylor Valley in January 2018 (**Figure 4-1**). Partial days occurred due to snowfall halting sampling. Two sites were used in Antarctica to control for site specific effects, in particular

proximity to the camp at Spaulding pond.

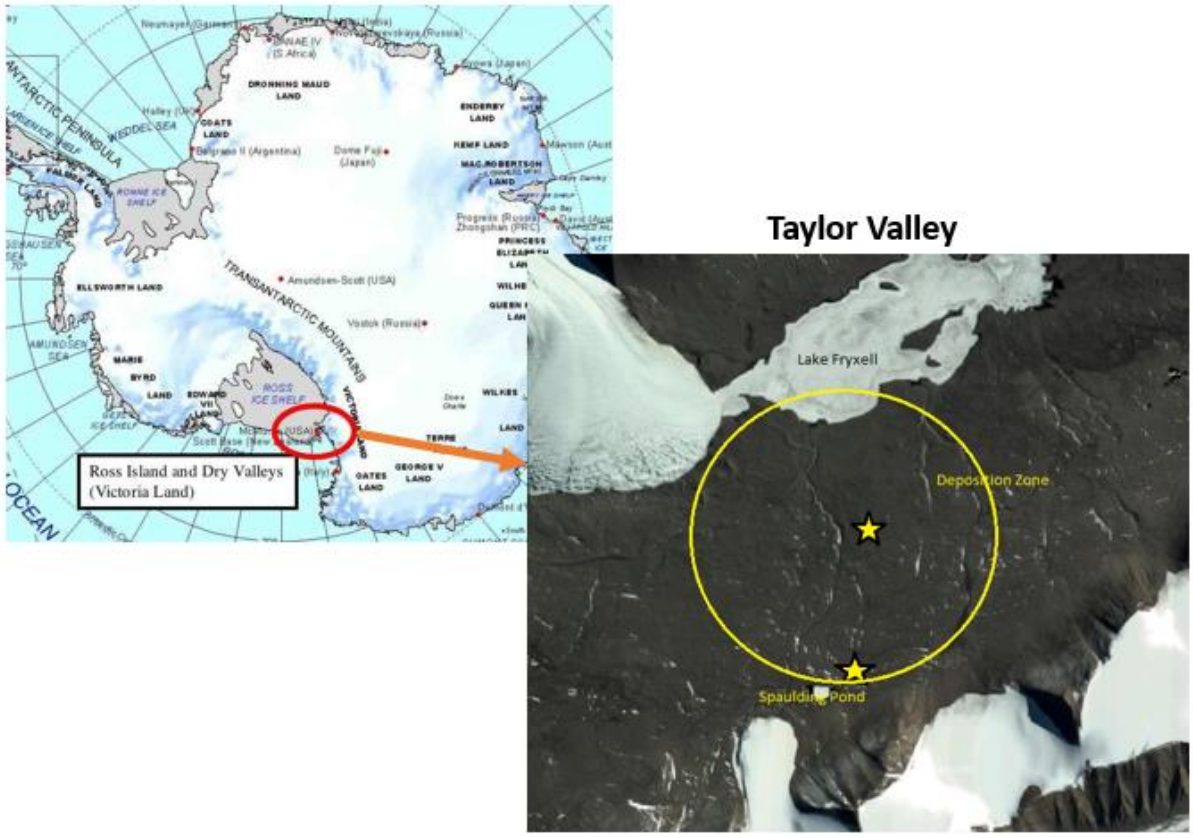


Figure 4-1 - Map of Antarctic sampling locations

The sample sites were pre-determined, and the exact location was saved in a GPS locator. Baring Head was sampled at table height (0.75 m) and the Coriolis extender and tripod at 1.8 m were used in the Dry Valleys. A photo was taken of the machine and location. Gloves were worn and the extender and Coriolis unit cleaned with bleach. The Coriolis neck, head and cone were cleaned with bleach, ethanol and three rinses of milli-Q water (MQH_2O). The cone was filled with 15 mL of PBS. A negative was taken (PBS put into the cone without running the Coriolis) before running the Coriolis for two – four minutes with MQH_2O to ensure all bleach residue in the head and neck was removed. The MQH_2O was discarded and replaced with 15 mL of PBS and the Coriolis was run at 300 L/m for one hour. In the Dry Valleys, the extender was used at the maximum top up rate. The heating rig was set to 20 °C, but windchill meant it was always heating at maximum rate and maintained a temperature around 5 °C depending on the weather conditions. No heating was required at Baring Head. After one hour the sample was transferred to a 15 mL falcon tube and stored in an insulated bin with ice. 15 mL of PBS was loaded into the Coriolis cone and the machine restarted. Antarctic samples were filtered in two-hour pools in the field and stored in CTAB buffer in the insulated bin until they were transported to Scott Base. At Scott Base they were stored at -20 °C before being transported to Auckland in an insulated box containing dry ice. Once they were received in Auckland, they were stored at -20 °C. Baring Head samples were transported by road back to Auckland in the insulated box with

ice then stored at -20 °C. Samples were pooled in two – four-hour increments. Observations of temperature, wind chill, relative humidity, max wind speed, mean wind speed and weather conditions were recorded at each sample changeover (with a Kestrel 3000). Particle counts were taken (with an AeroTrak particle counter) in New Zealand (NZ). It was too cold for the particle counter to function in Antarctica. In Antarctica UVA, B and C were measured with a UVC radiometer (UVP Inc, Upland, CA). If light rain occurred during sampling, the Coriolis was sheltered with an umbrella. If rain was heavy, the unit was packed up until the rain abated.

Air-mass back-trajectories were generated hourly over the sampling periods using the NOAA HYSPLIT model (v5. 0. 0 Ubuntu), (Stein et al., 2016) with GDAS meteorological data. Clustering was performed in HYSPLIT (see Appendix B HYSPLIT Clustering Procedure for details). The number of clusters selected (two - three) was based on a marked increase in total spatial variance as clusters reduced. The cluster for each sample was manually entered into the metadata, which was imported into R. Back-trajectories were generated for one-day, three-day and one-week durations.

4.3.2 Laboratory Processing

For details on DNA extraction and DNA sequencing methods see section C.1 Laboratory Methods – DNA Extraction, and section C.3 Laboratory Methods – DNA Sequencing.

4.3.3 Bioinformatics

For details on bioinformatic processing see Appendix D Optimising Bioinformatics Protocols for Aerosol Microbial Community Data – a Case Study Using an Urban Parks Dataset. FastQ files pertaining to Baring Head and Antarctica were analysed as per the recommended protocols. The decontaminated amplicon sequence variant (ASV) table (with read counts adjusted to remove contaminant sequences) and combined taxonomy and relevant metadata were analysed in R with respect to the hypotheses for this study.

4.3.4 Data Analysis

The 16S (the gene for the RNA component of the 30S small-subunit of the prokaryotic ribosome) and internal transcribed spacer (or ITS, which is spacer DNA situated between the small-subunit RNA and large-subunit ribosomal RNA genes in the fungal genome) ASV tables were prepared for analysis by removing duplicate or reattempted samples (some samples did not sequence well, so sequencing was repeated). Samples with the highest read counts were retained. The ASV table was filtered with variable stringency, and the results submitted to redundancy analysis (RDA) and compared to determine the optimal level of filtering. An unfiltered ASV table was created. These filters were used to improve any signal which could be masked by invariant taxa and/or low-level stochastic variation. For Antarctica 16S, an ASV table with only ASVs with over 300 reads and a coefficient of variation (CV) over three, and an ASV table with only ASVs with over 400 reads and a CV over four was created. For Antarctica ITS, an ASV table with only ASVs with over 500 reads and a CV over three, and an ASV table with only ASVs with over 500 reads and a CV over

four was created. Filtering was often performed at a higher read count threshold for ITS as ITS read counts were generally higher than 16S, and CV also tended to be higher. The thresholds were set based on the characteristics of each dataset. For Baring Head 16S, an ASV table with only ASVs with over 100 reads and a CV over three, and an ASV table with only ASVs with over 200 reads and a CV over four was created. For Baring Head ITS, an ASV table with only ASVs with over 100 reads and a CV over three, and an ASV table with only ASVs with over 200 reads and a CV over four was created. Days five and six in the Antarctic dataset were filtered out of all downstream analysis, as they were not full days, so represented very few sampling points and were not comparable with the remainder of the dataset so risked distorting the result. Metadata associated with each sample were adjusted to remove redundant variables (incomplete measurements or variables which were not analysed further). Numerical variables were standardised as *Z*-scores in R (with the function scale in base R). An additional numerical variable was included to indicate time of day using "lubridate" (v. 1.7.9, (Grolemund & Wickham, 2011)) - hours since midnight. As time on a 24-hour clock is a circular variable, this was subjected to both sine and cosine transformations, and a Euclidean distance matrix was calculated using the transformed variables to model temporal cyclicity (London, 2017). Additional categorical variables representing different temporal scales of variation were added – day/night (day 8am – 8pm, night 8pm – 8am), time window (1am – 3am, 3am – 5am, 5am – 7am, 7am – 9am, 9am – 11am, 11am – 1pm, 1pm – 3pm, 3pm – 5pm, 5pm - 7pm, 7pm – 9pm, 9pm – 11pm, 11pm – 1am), time of day (morning 8am – 12pm, afternoon 12pm – 4pm, evening 4pm – 8pm, night 8pm – 12am, late night 12am – 4am, early morning 4am – 8am).

To adjust zeros in the ASV table, which prevent log transformations (as used in the Aitchison distance matrix), Bayesian multiplicative replacement was performed on the ASV tables (Gloor et al., 2017; Quinn et al., 2019) using the function cmultRepl (Aitchison, 1982; Templ et al., 2011) in the R package "zCompositions" (v1. 3. 4) (Palarea-Albaladejo & Martín-Fernández, 2015). "RobCompositions" (v2. 2. 1) (Templ et al., 2011) was used to generate an Aitchison distance matrix, which was fed into downstream analysis functions.

Variance partitioning of the Aitchison distance matrix was conducted using distance-based redundancy analysis (db-RDA) to compare the three different lengths of back-trajectories to determine which length would be used for the remainder of the analysis. Variance partitioning provided quantification of the relative correlation of alternative trajectory lengths with the sampled bioaerosol communities; a higher R^2 indicated greater variation explained by a given trajectory. For the Antarctic and Baring Head data, the 16S unfiltered ASV table with a one-week trajectory had the highest R^2. For ITS, an unfiltered ASV table with a one-day back-trajectory had the highest R^2, therefore these parameters were chosen for the remainder of the analysis.

The Aitchison compositional dissimilarities were plotted against the Euclidean cos-sin transformed temporal dissimilarities for each pair of samples to visualise the time-dissimilarity relationship. The time-dissimilarity plot showed a linear relationship once

transformed. Non-linear correlation was calculated using the R package "nlcor" (v.1.3.2) (Ranjan, 2020).

Forward selection of variables was performed to determine which explanatory variables to include in further db-RDA analysis. This was implemented in R using the capscale and ordi2step functions in "vegan" (v2. 5. 6) (Oksanen et al., 2012). Variables (other than wind and time, as these were pre-selected) were included if indicated by forward selection. Forward selection was run with 999 permutations with a *P*-value parameter of one or lower for all amplicons. Variables were further excluded if all their variation was shared with other variables, so the simplest RDA with the highest R^2 value was accepted as the solution.

Significant variables resulting from the forward selection procedure were included in the db-RDA with variance partitioning to quantify strength of relationship with bioaerosol community composition. These variables were one-week wind, day and the Euclidean time matrix (time matrix) for Antarctica 16S, one-day wind, day and the time matrix for Antarctica ITS (**Table 4-1**). For Baring Head 16S, one-week wind and temperature were included. For Baring Head ITS, temperature, humidity and weather were included (**Table 4-2**). In addition to quantifying the relative importance of the different variables, db-RDA quantified the portion of explained variation that was shared among variables.

Table 4-1 - Explanatory variables and whether they were included in modelling compositional variation in bioaerosol communities sampled in Antarctica

Variable	Variable type	Description	Included in 16S model?	Included in ITS model?
1-day back trajectory cluster	Categorical	Path of sampled air-mass in previous 24 hours	No	Yes
3-day back trajectory cluster	Categorical	Path of sampled air-mass in previous 72 hours	No	No
1-week back trajectory cluster	Categorical	Path of sampled air-mass in previous 168 hours	Yes	No
Location	Categorical	Location where sampling took place	No	No
Day/night	Categorical	Time of sampling	No	No
Time of day	Categorical	Time of sampling	No	No
Time window	Numerical	Time of sampling	No	No
Max. wind speed	Numerical	Maximum wind speed in observation window	No	No

Mean wind speed	Numerical	Mean wind speed in observation window	No	No
Wind chill	Numerical	Wind chill during sampling	No	No
UVA/B/C	Numerical	UV measurements during sampling	No	No
Weather	Categorical	Weather during sampling	No	No
Temperature	Numerical	Temperature during sampling	No	No
Humidity	Numerical	Humidity during sampling	No	No
Day	Categorical	Day sampling took place during	Yes	Yes
Time matrix	Categorical	Euclidean distance matrix on cos and sin transformed hour since midnight variable	Yes	Yes

Table 4-2 - Explanatory variables and whether they were included in modelling compositional variation in bioaerosol communities sampled at Baring Head

Variable	Variable type	Description	Included in 16S model?	Included in ITS model?
1-day back trajectory cluster	Categorical	Path of sampled air-mass in previous 24 hours	No	No
3-day back trajectory cluster	Categorical	Path of sampled air-mass in previous 72 hours	No	No
1-week back trajectory cluster	Categorical	Path of sampled air-mass in previous 168 hours	Yes	No
Weather	Categorical	Weather during sampling	No	Yes
Temperature	Numerical	Temperature during sampling	Yes	Yes
Humidity	Numerical	Humidity during sampling	No	Yes

Visualisation of compositional variation among samples was achieved by generating Non-metric Multidimensional Scaling (NMDS) plots on the Aitchison distance matrices, implemented in the R package “vegan” (Oksanen et al., 2012) for the Antarctic data. The stress for two to six dimensions was compared, with 500 random starts and 999 iterations per run. The lowest dimensional solution with a stress under 0.2 was selected. No NMDS was presented for Baring Head, as R was unable to compute an NMDS for 16S due to insufficient data.

To assess patterns over time and by back-trajectory for different genera, for Antarctic data, the percentage of total read count were calculated for the relatively more common genera. Bar plots were generated using "ggplot2" (v3.3.2) (Wickham, 2016) and "phyloseq" (v1.30.0) (McMurdie & Holmes, 2013) on data filtered as follows. For 16S, samples with less than 100 reads were filtered out. Relative abundance by genus for each category included only genera with at least 2% of the reads. For ITS, samples with less than 100 reads were filtered out. Relative abundance by genus for each category included only genera with at least 2% of the reads. To assess patterns by back-trajectory for different genera, for Baring Head data, the percentage of total read count were calculated for the relatively more common genera. Bar plots were generated using "ggplot2" (v3.3.2) (Wickham, 2016) and "phyloseq" (v1.30.0) (McMurdie & Holmes, 2013) on data filtered as follows. For 16S, samples with less than 100 reads were filtered out. Relative abundance by genus for each category included only genera with at least 2% of the reads. For ITS, samples with less than 100 reads were filtered out. Relative abundance by genus for each category included only genera with at least 2% of the reads.

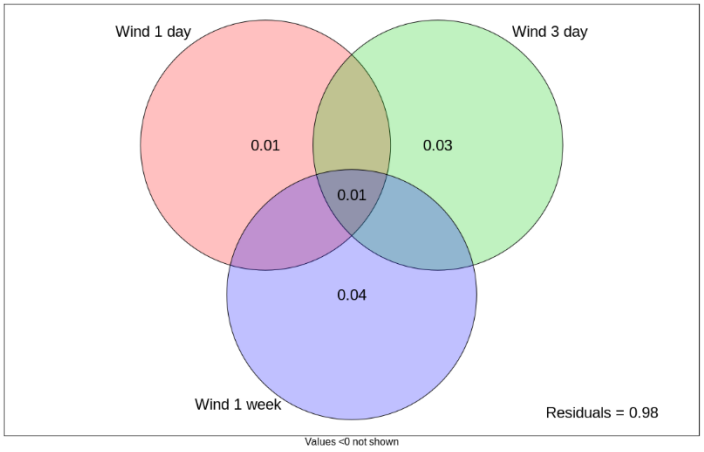
The number of ASVs (exact sequence match) and genera shared between Antarctica and Baring Head was calculated using the R package "zetadiv" (v. 1.2.0 (Latombe, McGeoch, Nipperess, & Hui, 2018)). A presence/absence matrix was created for Antarctica, Baring Head air of Antarctic origin and Baring Head air of non-Antarctic origin including any ASV/genera identified at that site in any sample. The number of overlapping taxa between Antarctic air and Baring Head air of Antarctic origin was calculated and compared to the number for the comparison of Antarctic air and Baring Head air of non-Antarctic origin. A comparison of air sampled in Antarctica of NZ and non-NZ origin with NZ air could not be performed as no air sampled in January 2018 appeared to be of New Zealand origin, based on one-week back-trajectories run.

4.4 Results

4.4.1 One-Week and One-Day Back-Trajectories Selected for Air-Mass Source Variable in Antarctic db-RDA

Wind back-trajectories correlated with similarly low levels of variation in community composition for 16S (R^2 = 2%), and ITS (R^2 = 1%) (**Figure 4-2**). The different trajectory lengths showed consistent shared variation. For 16S, one-week trajectories had the highest R^2 value. For ITS, one-day trajectories had the highest R^2 values. One-week trajectories were selected for the remainder of the analysis for 16S, and one-day trajectories were selected for ITS.

a)



b)

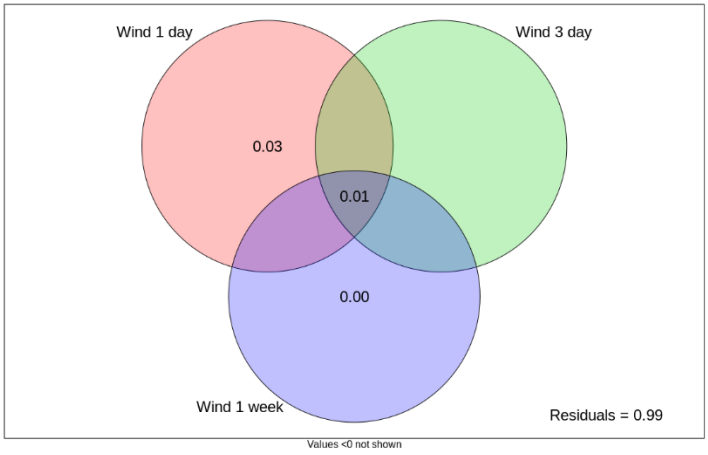


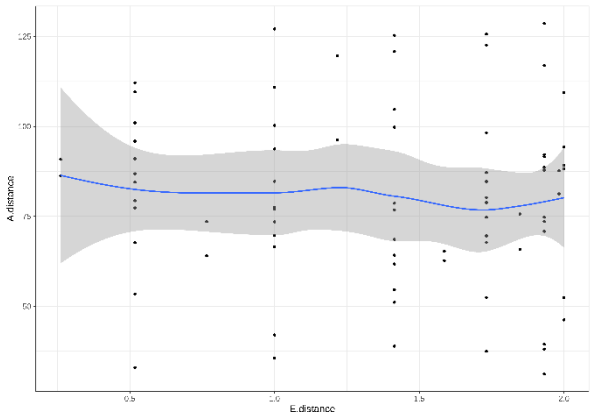
Figure 4-2 - Variance partitioning of the results from a distance-based redundancy analysis on the Aitchison compositional distances among samples for (a) 16S and (b) ITS showing the variance explained by one-day, three-day, and one-week back-trajectories. Only R^2 above 1% are displayed, and negative values offset the positive values resulting in the low overall R^2 values for the model.

4.4.2 Time was Transformed to Reflect Cyclicity

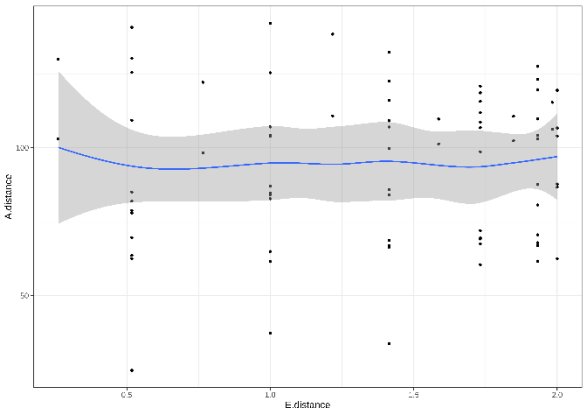
A Euclidean distance matrix calculated from time in hours since midnight and the Aitchison dissimilarity matrix were non-linearly related for both amplicons, with significant non-linear correlation coefficients overall for days one to four (16S: 0.19, ITS: 0.22). The plots for individual days show noticeable variation between days and non-linearity persisting over the 24-hour period despite the transformation. However, these variations appeared to cancel out overall, as the Euclidean distance matrix of the sine and cosine transformed fine

scale temporal variables for days one to four in aggregate appeared to have little obvious relationship with Aitchison dissimilarity in the time-dissimilarity plots (**Figure 4-3**).

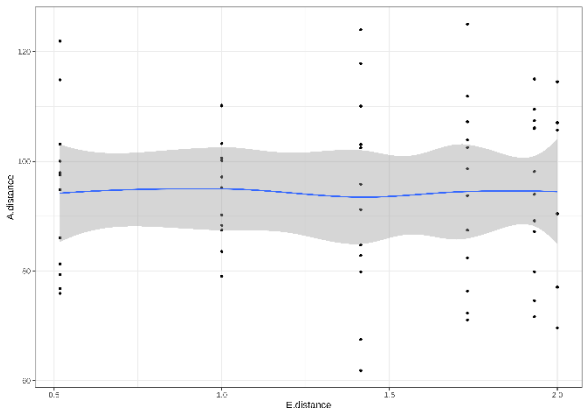
a)


A.distance
E.distance

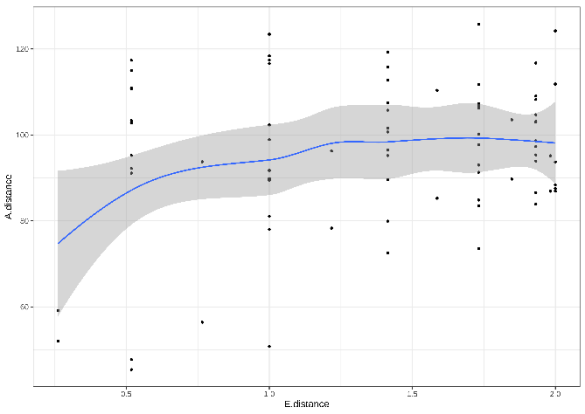
b)


A.distance
E.distance

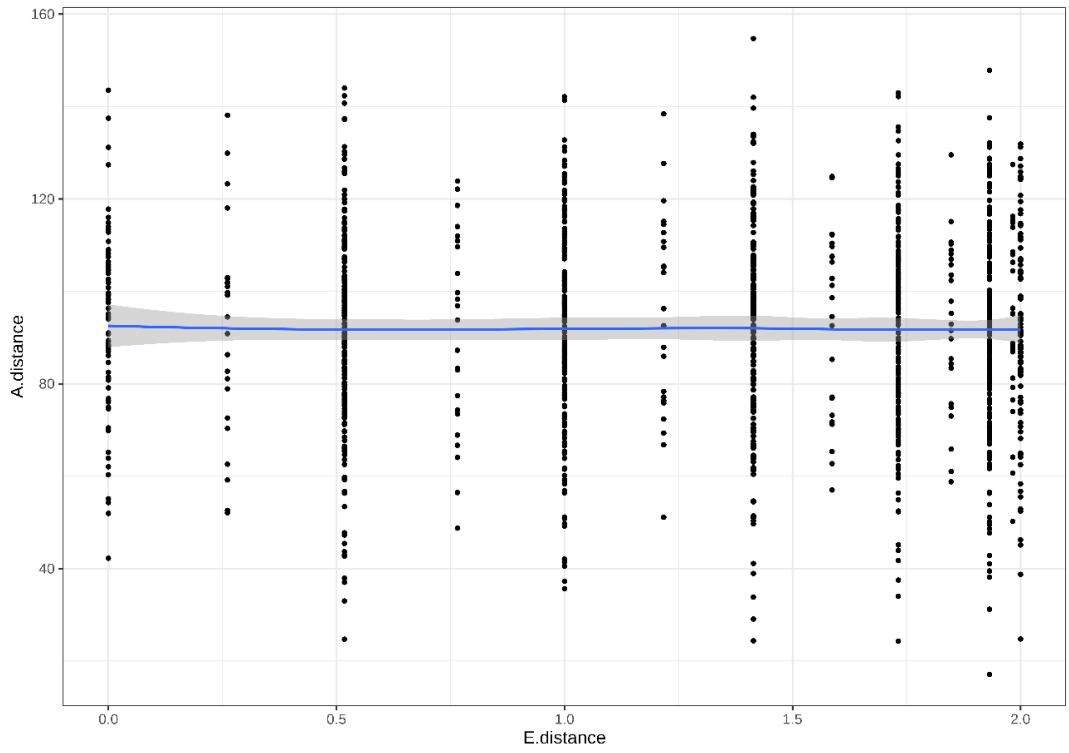
c)


A.distance
E.distance

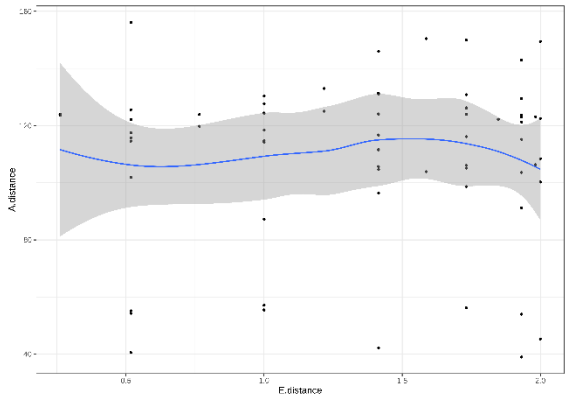
d)


A.distance
E.distance

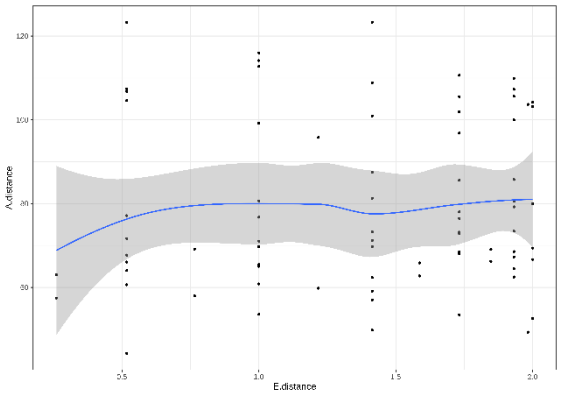
e)


160
120
80
40
A.distance
0.0
0.5
1.0
1.5
2.0
E.distance

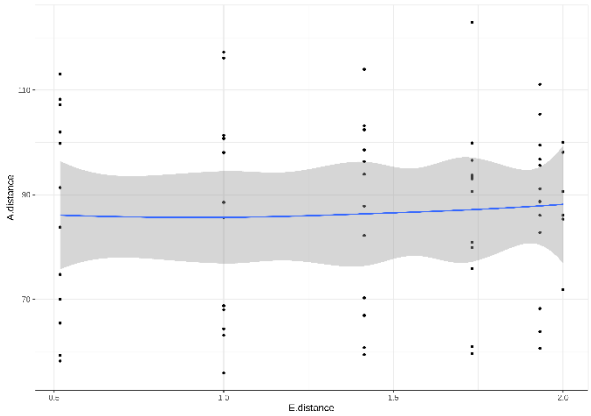
f)


A.distance
E.distance

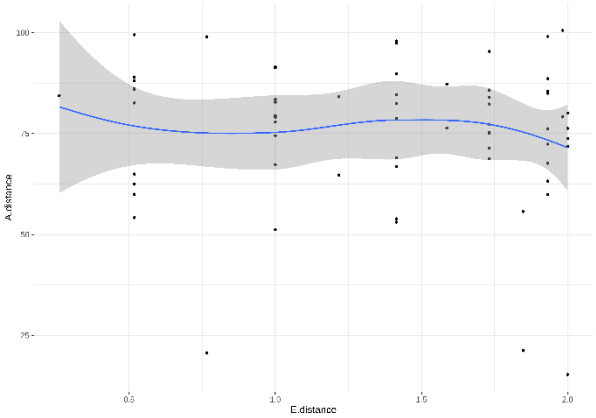
g)


A.distance
E.distance

h)


A.distance
E.distance

i)


A.distance
E.distance

j)

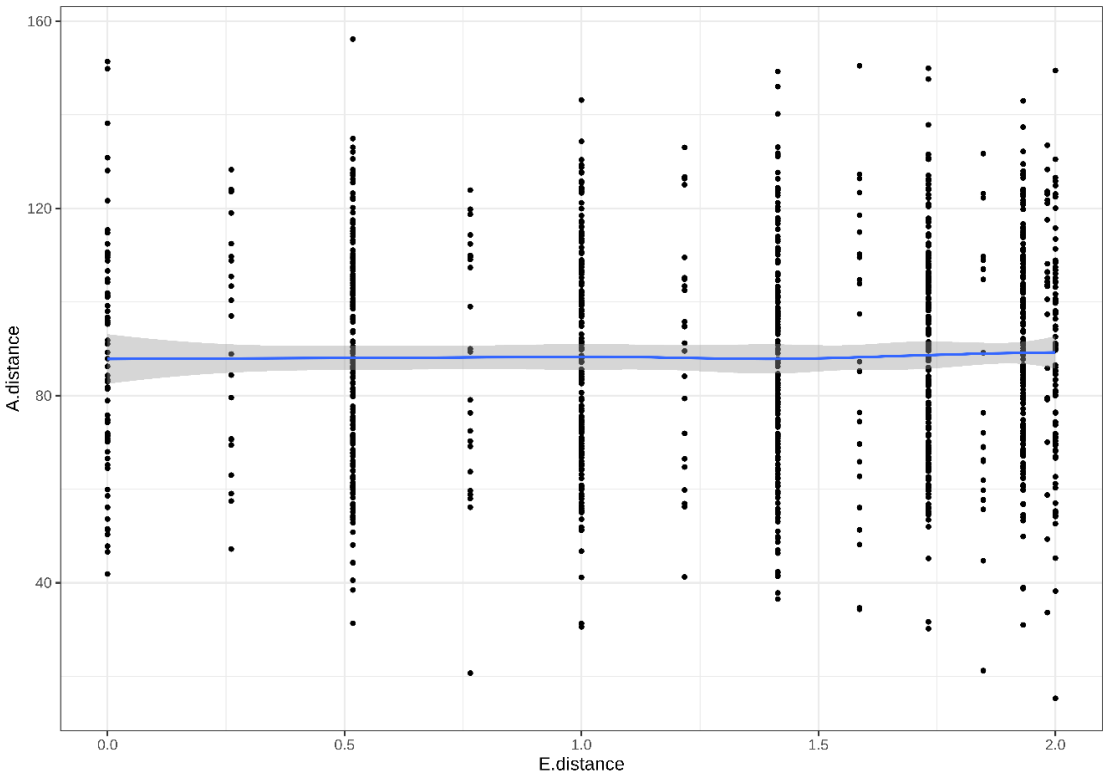
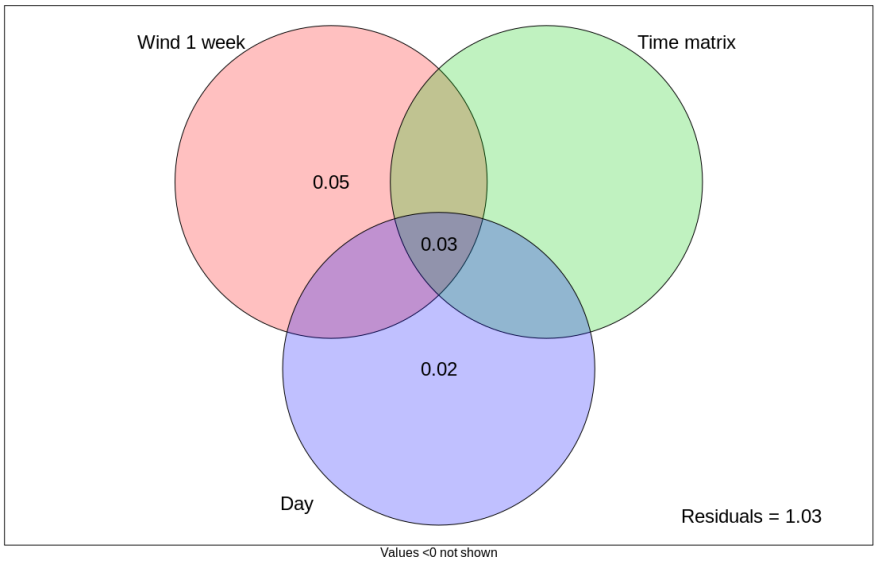

160
120
80
40
A.distance
0.0
0.5
1.0
1.5
2.0
E.distance

Figure 4-3 - Distance time-lag plots for a) 16S day one b) 16S day two c) 16S day three d) 16S day four e) 16S days one to four and f) ITS day one g) ITS day two h) ITS day three i) ITS day four j) ITS days one to four of Aitchison compositional sample distances against Euclidean time temporal dissimilarities between sine and cosine transformed sample times (number of hours since midnight).

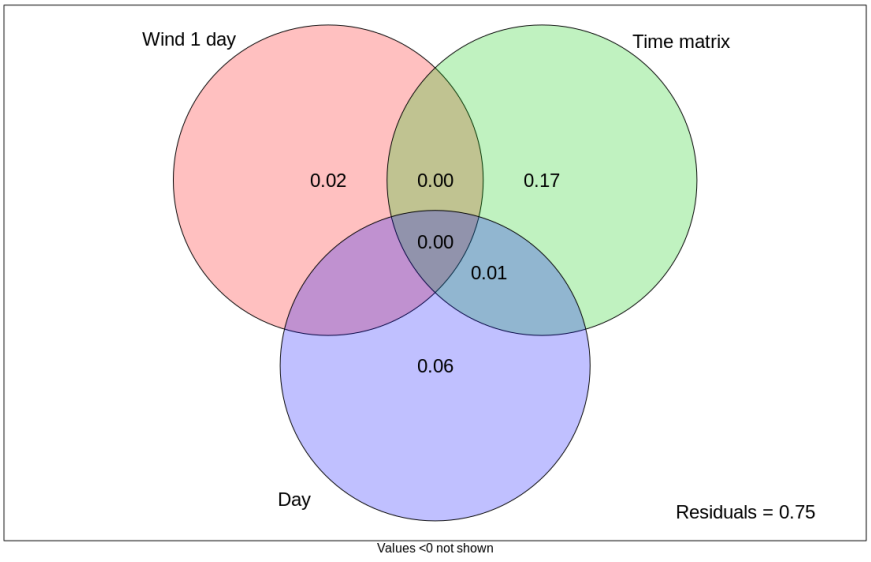
4.4.3 Time of Day, Air-Mass Source and Location Affect ITS Bioaerosol Community Composition

The 16S data appeared to be too sparse, with a negative R^2 from modelling, and no convincing relationships with time and air-mass source detected (**Figure 4-4**). However, the ITS model had an R^2 of 25%, and indicated that time of day, air-mass source and sampling day were correlated with variability in the Antarctic bioaerosol community. Variation over the course of a day appeared to be much greater than between days for ITS. Temperature, relative humidity, wind speed and UVA, UVB and UVC were not correlated with the bioaerosol community per forward selection procedures. Location was autocorrelated with day (as different locations were sampled on different days), so only day was included. Both NMDS ordinations showed clustering of samples with different air-mass sources, the hours since midnight variable, with blue and red often being adjacent as they represent midnight and 1am respectively, and sampling day. Clustering appeared to be more pronounced for ITS.

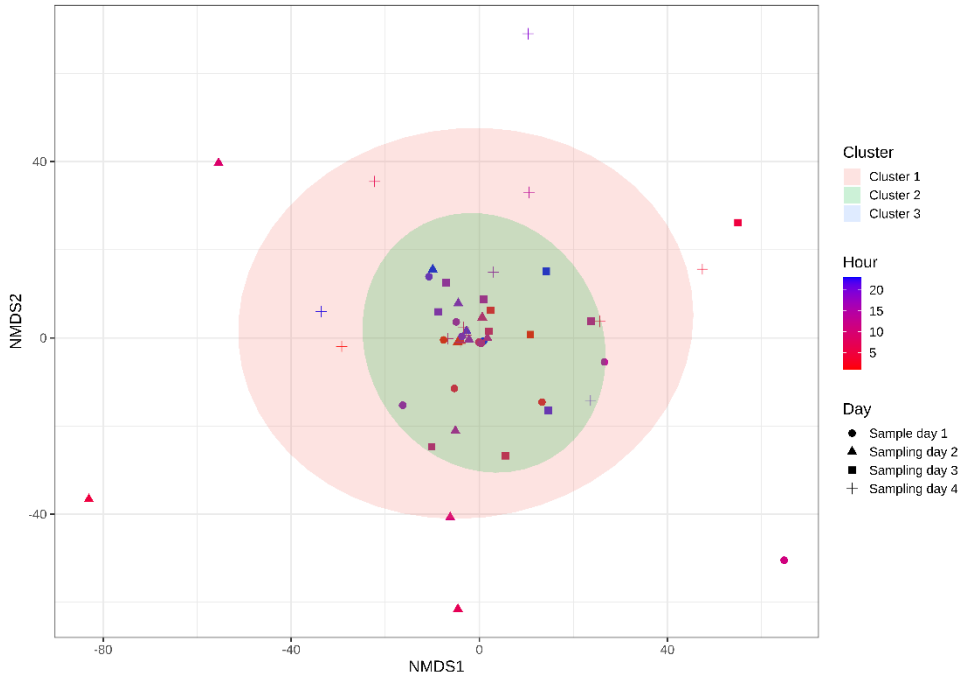
a)


Wind 1 week
Time matrix
0.05
0.03
0.02
Day
Residuals = 1.03
Values <0 not shown

b)


Wind 1 day
Time matrix
0.02
0.00
0.17
0.00
0.01
0.06
Day
Residuals = 0.75
Values <0 not shown

c)



d)

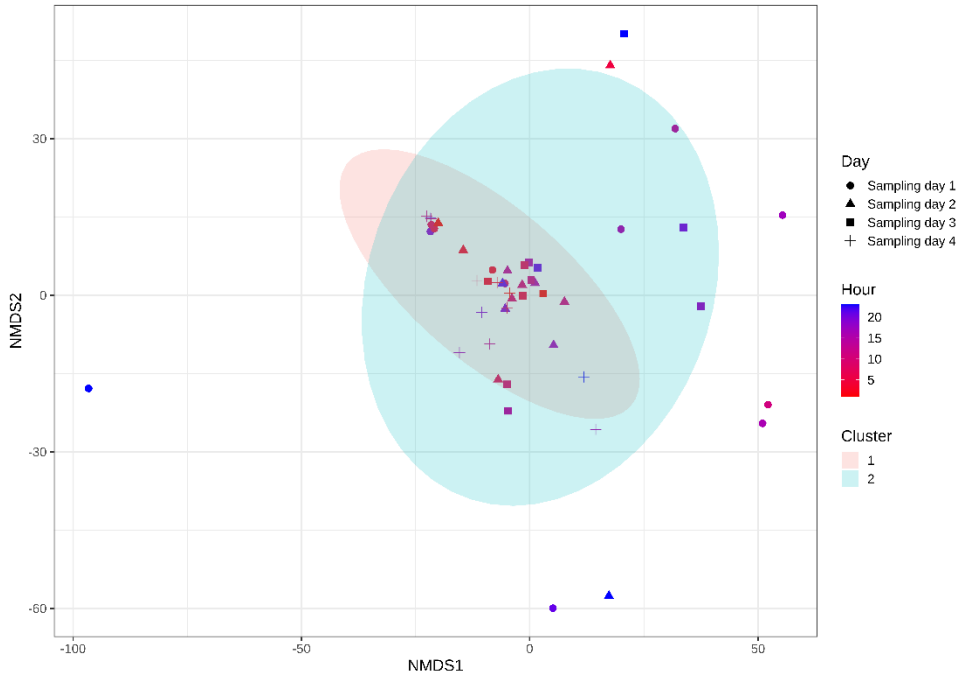


Figure 4-4 - Variance partitioning of the results from a db-RDA on the Aitchison compositional distances among samples for (a) 16S and (b) ITS, showing the variance explained by time matrix, wind and day. NMDS ordination of the c) 16S and d) ITS Aitchison distances among samples. The back-trajectory cluster is indicated by ellipses representing the *t* distribution of the points relating to each cluster. Hours are indicated by the gradient colour. Sampling day is indicated by shape. Stress on 16S NMDS was 0.17 (dimensions one and two are presented above), the three-dimensional solution was

selected as the two-dimensional solution had stress greater than 0.2 so could not be relied upon. Dimensions one and three, and two and three are presented in Supplementary Materials (Figure E-10). Stress on ITS NMDS was 0.15 (dimensions one and two are presented above), the three-dimensional solution was selected as the two-dimensional solution had stress greater than 0.2. Dimensions one and three, and two and three are presented in Supplementary Materials (Figure E-11). Higher dimensional solutions had lower stress values for both amplicons.

Relative Abundance by Time Window – 16S

Unfiltered bacterial reads numbered 897,244, with 2,276 ASVs inferred. Unfiltered fungal reads numbered 1,579,022, comprised of 1,569 ASVs.

Stenatrophomonas spp., *Acetinobacter* spp. and *Pseudomonas* spp. were common in Antarctic bioaerosols (**Figure 4-5**). Genera were consistently relatively abundant over a 24-hour period. Filtering criteria described in **Figure 4-5** legend below.

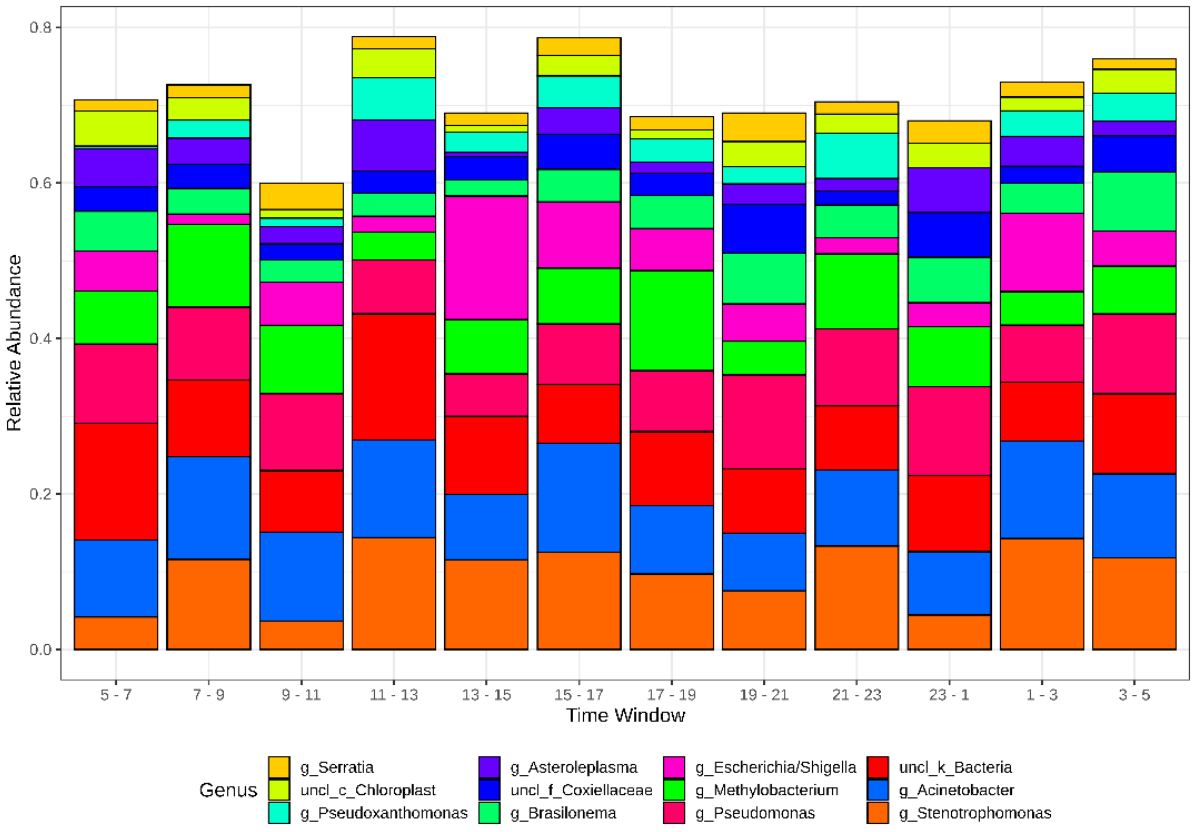


Figure 4-5 - Relative abundance by genus by time window for 16S. Samples with less than 100 reads were removed. Only genera with at least 2% of the reads were included in the bar plot.

Relative Abundance by Time Window – ITS

Peniophora spp., an unclassified organism in the family *Polyporaceae* and *Rigidoporus* spp. were common in Antarctic bioaerosols (**Figure 4-6**). Community composition appeared to be reasonably consistent over a 24-hour period. Filtering criteria is described in **Figure 4-6** legend below.

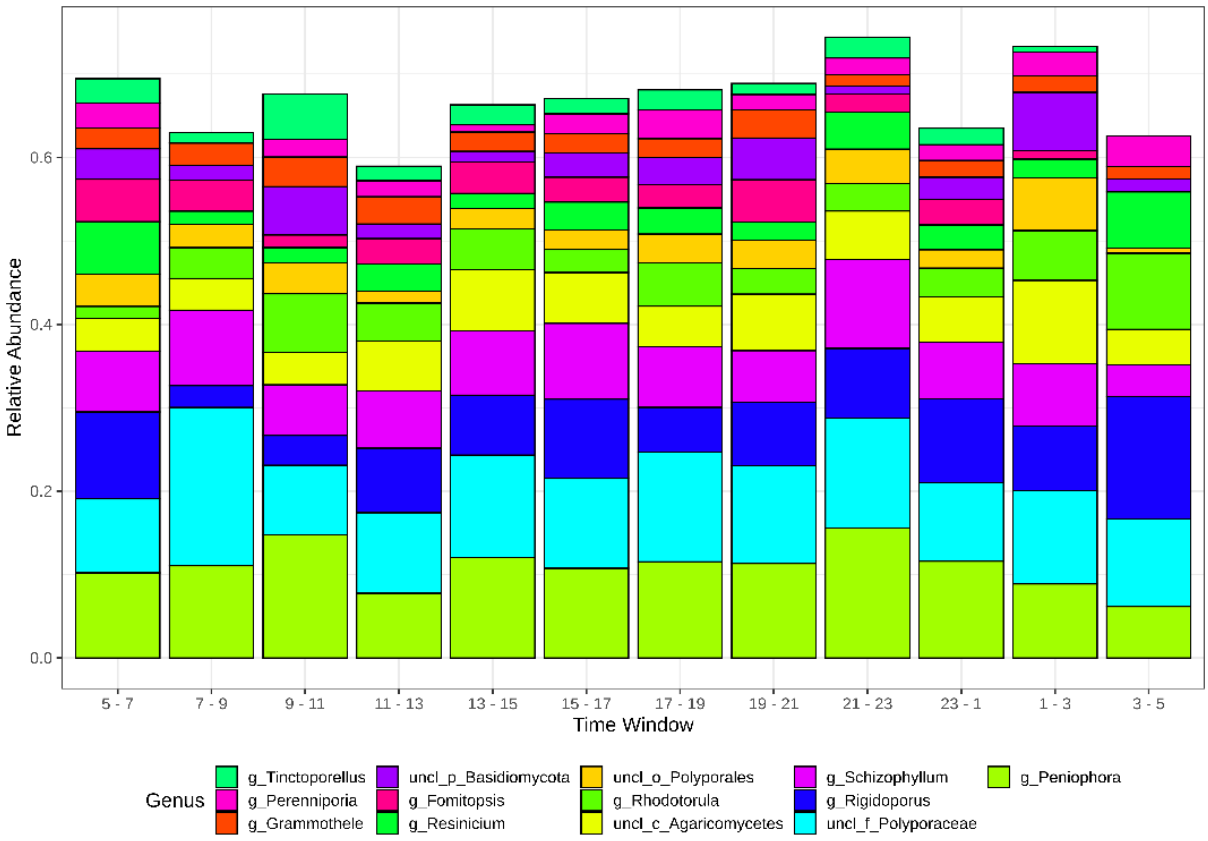


Figure 4-6 - Relative abundance by genus by time window for ITS. Samples with less than 100 reads were removed. Only genera with at least 2% of the reads were included in the bar plot.

Relative Abundance by Sampling Day – 16S

Genera were reasonably consistently abundant on different days (**Figure 4-7**). Filtering criteria is described in **Figure 4-7** legend below.

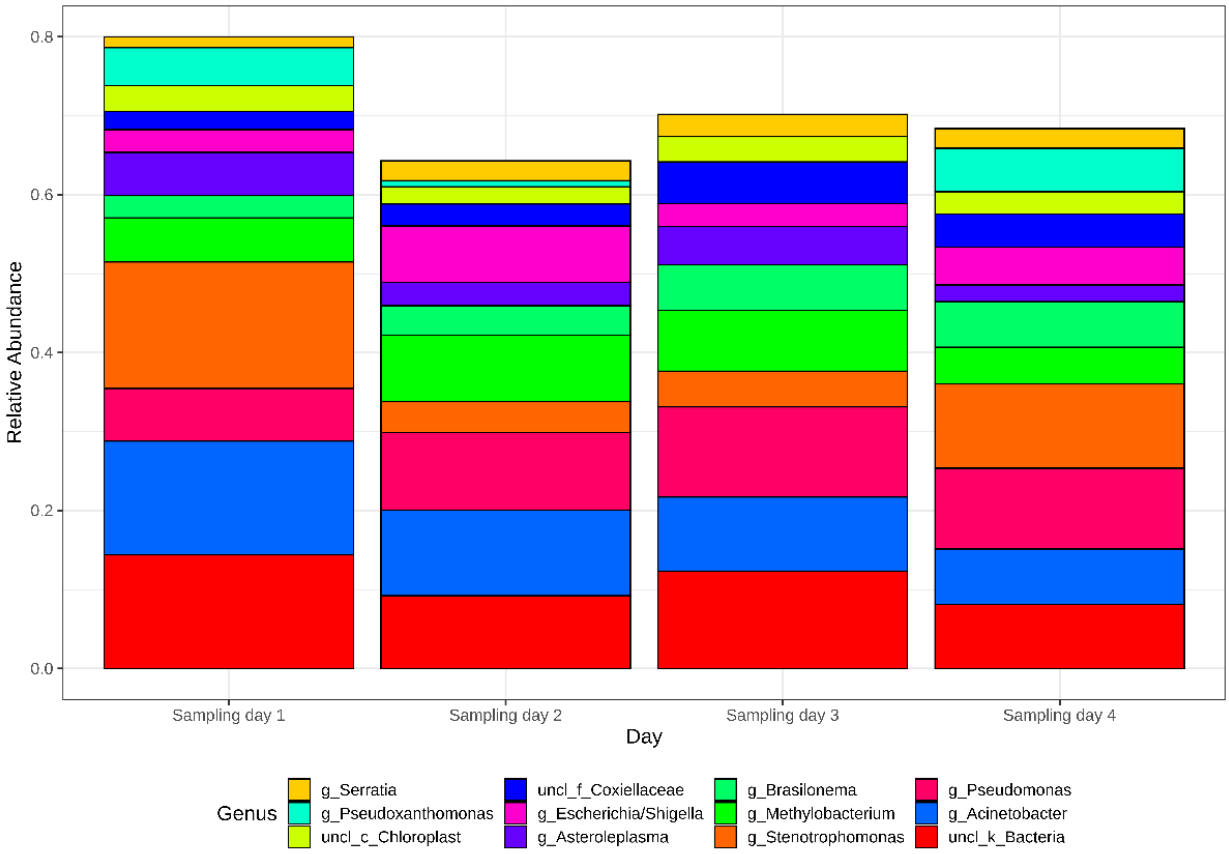


Figure 4-7 - Relative abundance by genus by sampling day for 16S. Samples with less than 100 reads were removed. Only genera with at least 2% of the reads were included in the bar plot.

Relative Abundance by Sampling Day – ITS

Genera were consistently relatively abundant on different days (**Figure 4-8**). Filtering criteria is described in **Figure 4-8** legend below.

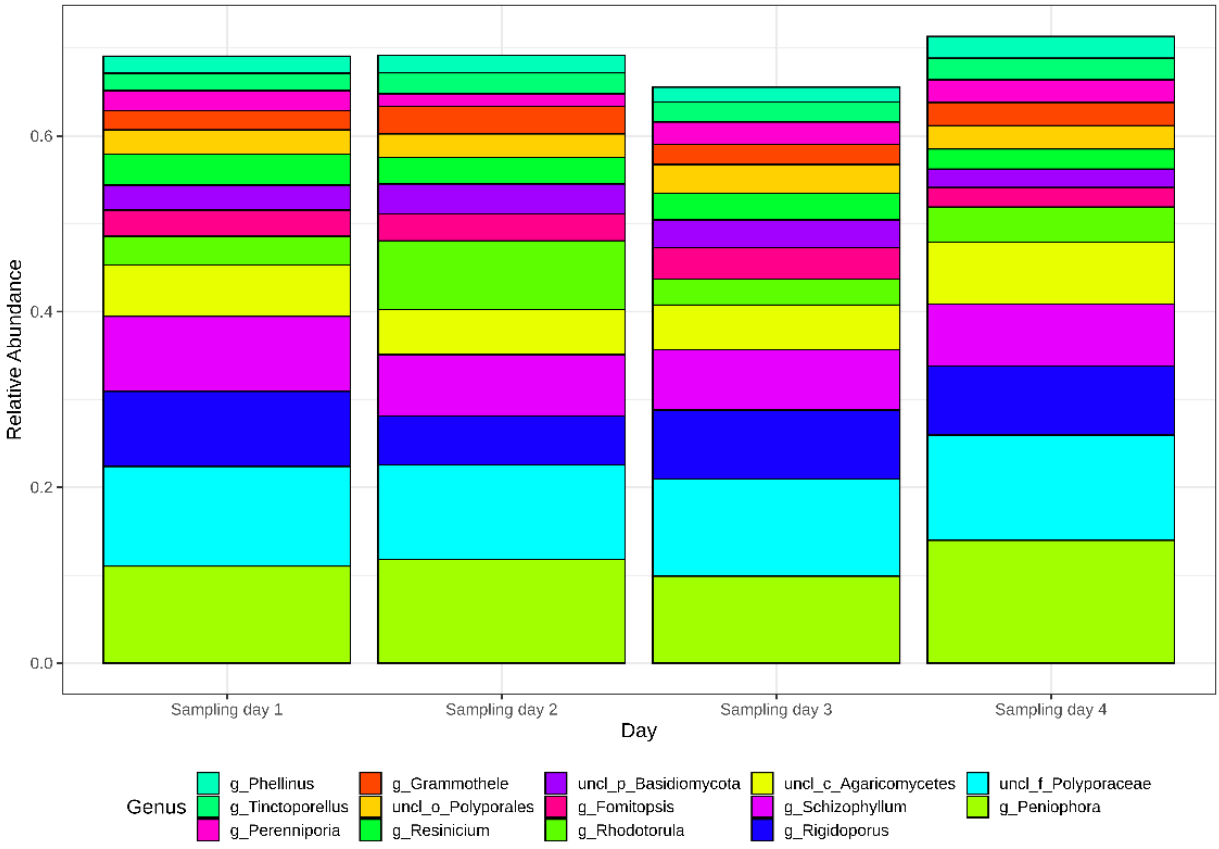
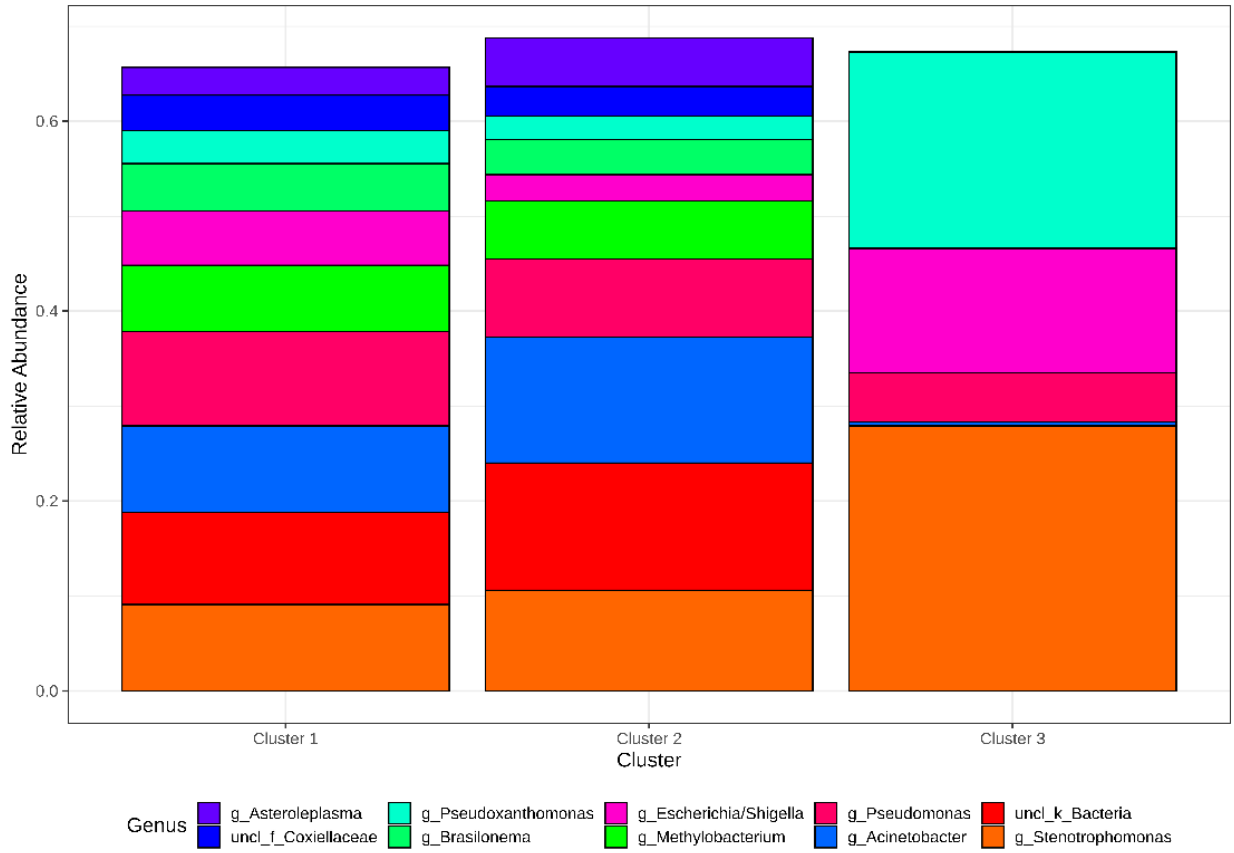


Figure 4-8 - Relative abundance by genus by sampling day for ITS. Samples with less than 100 reads were removed. Only genera with at least 2% of the reads were included in the bar plot.

Relative Abundance by Back-Trajectory Cluster – 16S

Clusters one and two harboured consistent assemblages of bacteria (**Figure 4-9**). Cluster three appeared to be distinct, but that was likely driven by low total reads from few samples mapping to cluster three, which affected apparent community composition. Filtering criteria is described in **Figure 4-9** legend below.

a)



b)

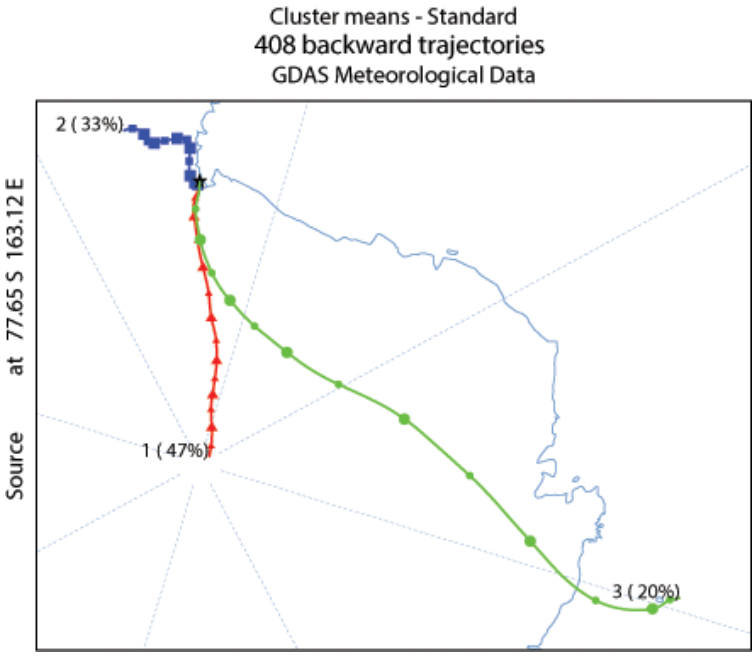
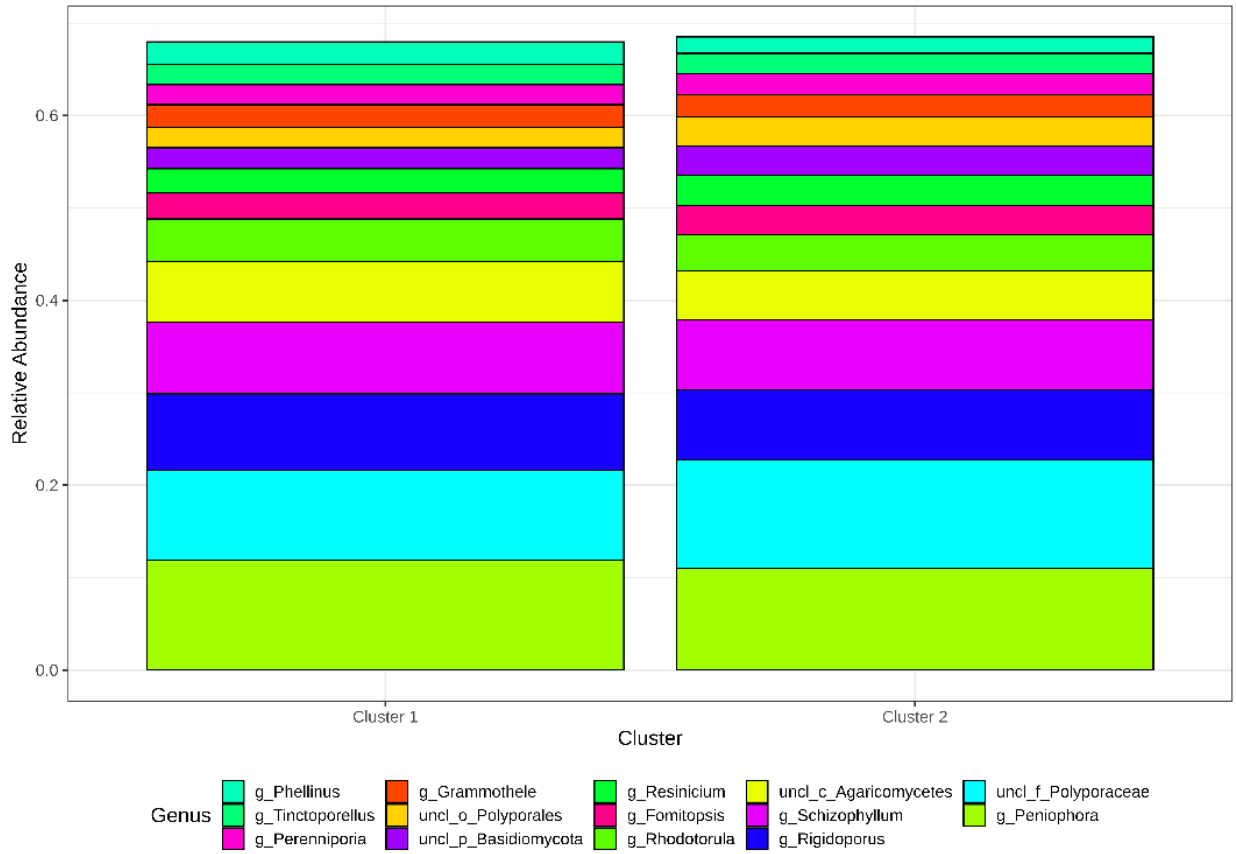


Figure 4-9 – a) Relative abundance by genus by back-trajectory cluster for 16S. Samples with less than 100 reads were removed. Only genera with at least 2% of the reads were included in the bar plot. b) Map of routes taken by one-week back-trajectory clusters. % next to each cluster indicates proportion of trajectories assigned to that cluster.

Relative Abundance by Back-Trajectory Cluster – ITS

Clusters one and two harboured consistent assemblages of fungi (**Figure 4-10**). Cluster three had such low numbers of fungal reads that no genera passed the filters used for genus bar plots. Filtering criteria is described in **Figure 4-10** legend below.

a)



b)

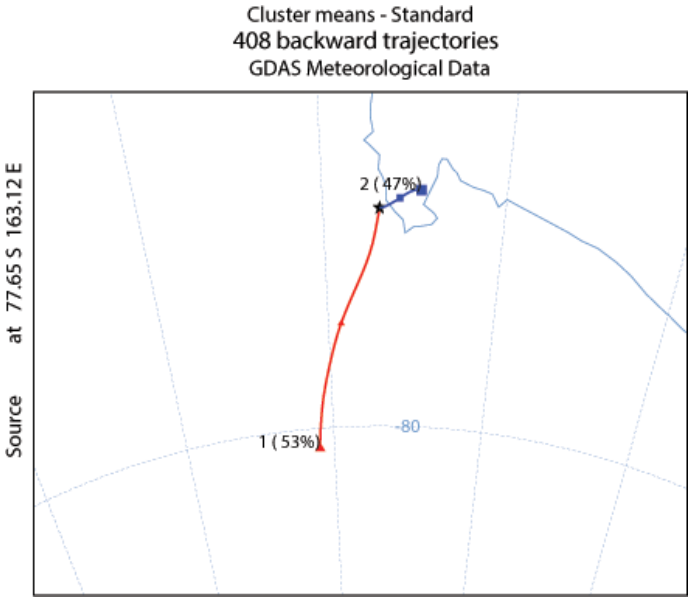


Figure 4-10 - a) Relative abundance by back-trajectory cluster for ITS. Samples with less than 100 reads were removed. Abundance by genus for each cluster, only genera with at least 2% of the reads were included in the bar plot. b) Map of routes taken by one-day back-trajectory clusters. % next to each cluster indicates proportion of trajectories assigned to that cluster.

4.4.4 Taxonomic Overlap Between Antarctica and Baring Head Increased when Baring Head Air was of Antarctic Origin

ASVs with identical sequences for both bacteria and fungi were found in Antarctica and New Zealand (**Table 4-3**). For bacteria, 150 - 212 ASVs were shared between the two locations, and 144 – 153 fungal ASVs were shared, depending on wind direction. There were more ASVs in Baring Head air of non-Antarctic origin, than in Baring Head air of Antarctic origin. The pattern was similar for genera, with 174 – 247 bacterial genera in common, and 205 – 208 for fungi. There were greater numbers of genera in Baring Head air of non-Antarctic origin, than in Baring Head air of Antarctic origin. As a proportion of ASVs present, bacterial shared ASVs increased from 11.8% to 18.9% when the sampled air mass originated in Antarctica. Fungal shared ASVs increased from 3% to 3.4% when sampled air was from Antarctica. Unsurprisingly, the number of genera shared between locations was much greater than ASVs. As a proportion of genera present, bacterial shared genera increased from 61.4% to 72.2% when the sampled air mass originated in Antarctica. Fungal shared genera increased from 41.3% to 41.8% when sampled air was from Antarctica.

Table 4-3 – Table showing shared numbers of ASVs or genera between Antarctica and New Zealand, and the change in shared diversity depending on origin of air sampled in New Zealand.

ASV Amplicon	**Antarctica v Baring Head Antarctic air (BHA)** *Shared ASVs/total ASVs BHA*	**Antarctica v Baring Head Antarctic air** *%*	**Antarctica v Baring Head non-Antarctic air (BHnA)** *zeta diversity/total ASVs BHnA*	**Antarctica v Baring Head non-Antarctic air** *%*
16S	150/793	18.9	212/1,793	11.8
ITS	153/4,555	3.4	144/4,860	3.0
Genus Amplicon	**Antarctica v Baring Head Antarctic air** *Shared ASVs /total genera BHA*	**Antarctica v Baring Head Antarctic air** *%*	**Antarctica v Baring Head non-Antarctic air** *zeta diversity/total genera BHnA*	**Antarctica v Baring Head non-Antarctic air** *%*
16S	174/241	72.2	247/402	61.4
ITS	205/491	41.8	208/504	41.3

4.4.5 One-Week (16S) or One-Day (ITS) Back-Trajectories Selected for Air-Mass Source Variable in Baring Head db-RDA

Wind back-trajectories correlated with more of the variation in community composition for ITS (R^2 11%), than for 16S (R^2 4%) overall (**Figure 4-11**). The different trajectory lengths showed consistent shared variation. For 16S, one-week trajectories had the highest R^2 value (15%). For ITS, one-day trajectories had slightly higher R^2 values (7%). One-week trajectories were selected for the db-RDA for bacteria, and one-day trajectories were selected for the db-RDA for fungi.

a)

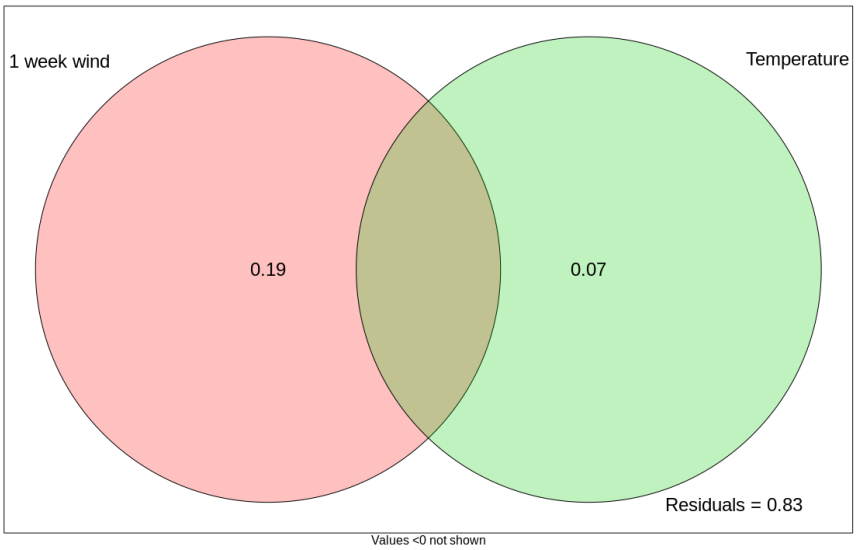
b)

Figure 4-11 - Variance partitioning of the results from a distance-based redundancy analysis on the Aitchison compositional distances among samples for (a) 16S and (b) ITS showing the variance explained by one-day, three-day, and one-week back trajectories.

4.4.6 Air-Mass Source, Temperature, Humidity and Weather Affected Baring Head Bioaerosol Community Composition

The 16S model had an R^2 of 17%, suggesting that air-mass source and temperature were the principal correlates with the Baring Head bioaerosol community for bacteria (**Figure 4-12**) and that weather and relative humidity had little impact. The ITS model had a lower R^2 of 13%, and indicated that temperature, relative humidity and weather were correlated with variability in the Baring Head bioaerosol community. One-day wind was all shared variation for ITS, so was not included as it did not independently explain any of the variation.

a)



b)

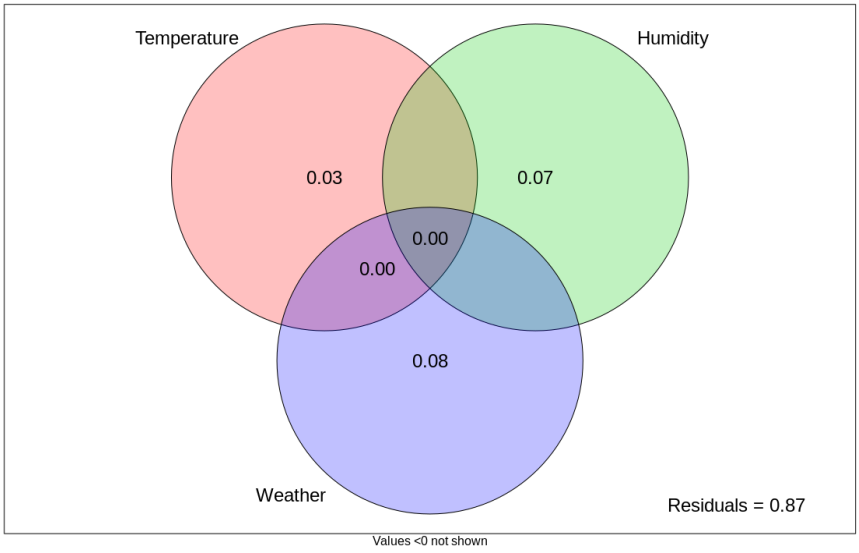


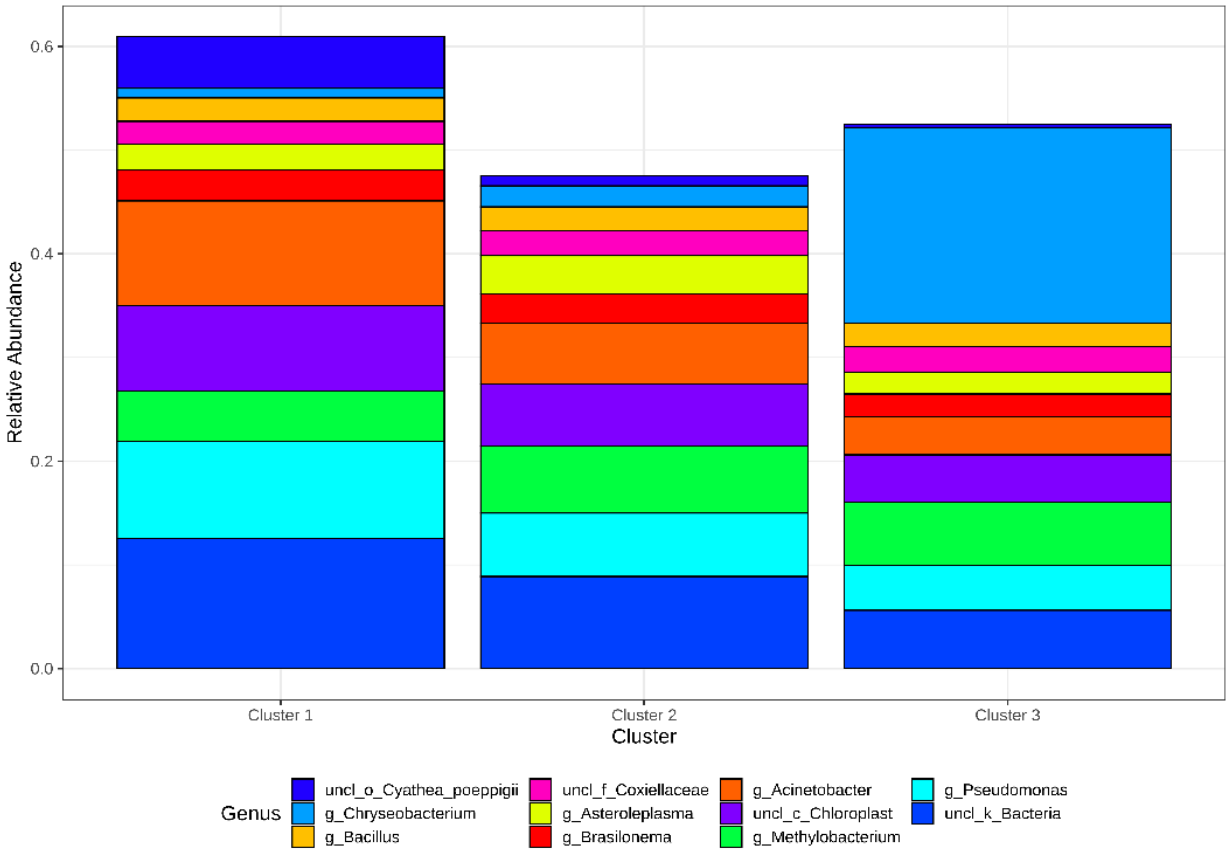
Figure 4-12 - Variance partitioning of the results from a distance-based redundancy analysis on the Aitchison compositional distances among samples for (a) 16S showing variance explained by one-week wind and temperature and (b) ITS showing the variance explained by temperature, relative humidity and weather.

Relative Abundance by Back-Trajectory Cluster for Baring Head – 16S

Unfiltered bacterial reads numbered 310,354, with 2,009 ASVs inferred. Unfiltered fungal reads numbered 2,470,453, comprised of 5,469 ASVs.

Methylobacterium spp., an unclassified bacterium, an unclassified chloroplast, *Actinobacteria* spp., *Chrysobacterium* spp. and *Pseudomonas* spp. were common in Baring Head bioaerosols (**Figure 4-13**). Community composition was similar in the different back-trajectory clusters. *Chrysobacterium* spp. were more prevalent in cluster three. Filtering criteria are described in **Figure 4-13** legend below.

a)



b)

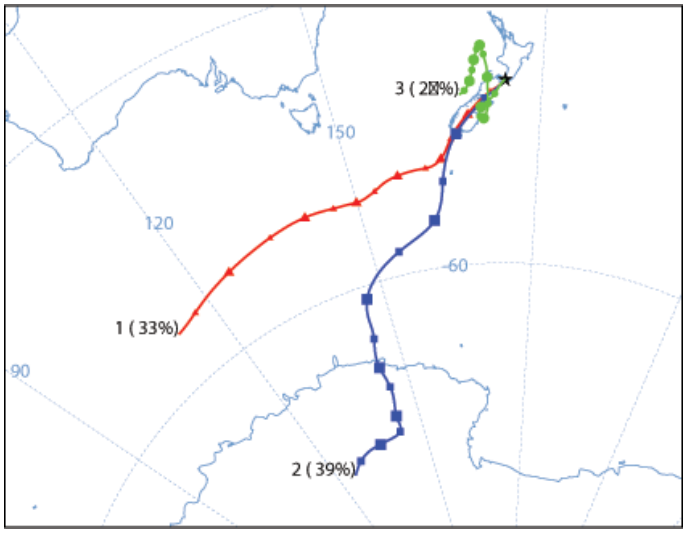
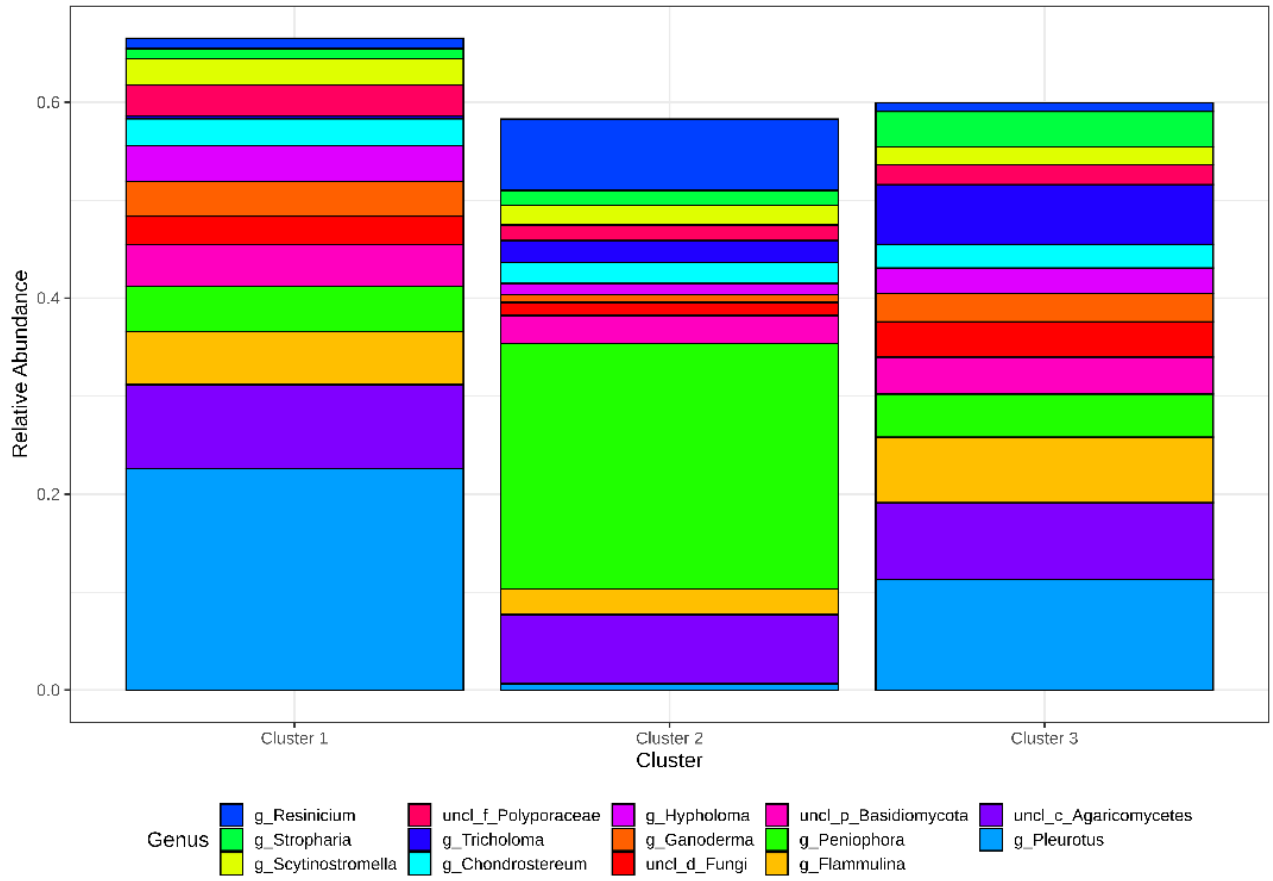


Figure 4-13 – a) Relative abundance by back-trajectory cluster for 16S. Samples with less than 100 reads were removed. Only genera with at least 2% of the reads were included in the bar plot. b) Map of routes taken by back-trajectory clusters. % next to each cluster indicates proportion of trajectories assigned to that cluster.

Relative Abundance by Back-Trajectory Cluster for Baring Head – ITS

Pleurotus spp., an unclassified organism in the class Agaricomycetes and *Flammulina* spp. were common in Baring Head bioaerosols (**Figure 4-14**). Cluster two was distinctive, with higher relative abundances of *Peniophora* spp. and fewer *Pleurotus* spp. Filtering criteria are described in **Figure 4-14** legend below.

a)



b)

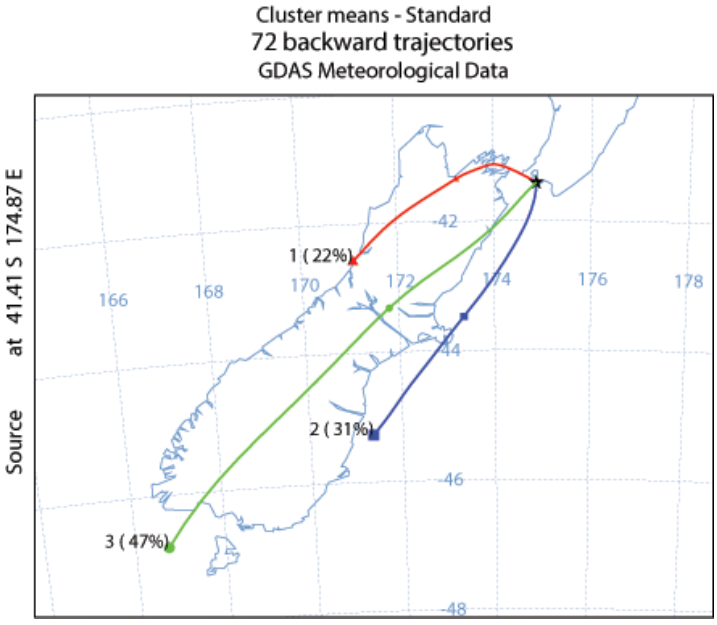


Figure 4-14 - a) Relative abundance by genus by back-trajectory cluster for ITS. Samples with less than 100 reads were removed. Only genera with at least 2% of the reads were included in the bar plot. b) Map of routes taken by back-trajectory clusters. % next to each cluster indicates proportion of trajectories assigned to that cluster.

4.5 Discussion

The purpose of this study was to understand the nature and drivers of diel variation in Antarctic bioaerosols, and the frequency of intercontinental exchange with Aotearoa New Zealand. Antarctic fungal bioaerosols were correlated with one-day wind and exhibited diel variation that was independent of measured variables such as temperature, which is known to vary within a 24 hour period and affect bioaerosols (Burrows et al., 2009b). Relationships between environmental variables and the 16S amplicon data for bacteria were less conclusive. Antarctica and New Zealand have distinct bioaerosol communities but share a portion of bacterial ASVs that varies, depending on air source. Taxonomic overlap between Antarctica and New Zealand increased when New Zealand air was coming from Antarctica, suggesting frequent intercontinental transport for bacteria and much less frequent transport for fungi. This research suggests that diel variation is present in Antarctic fungi, but the drivers remain unclear. Bacterial diel variation was less apparent. Bacteria appear to have longer atmospheric residence times and transport ranges than fungi and, as a consequence, undergo wind-mediated intercontinental transport to Aotearoa New Zealand more often.

4.5.1 Optimal Back-Trajectory Lengths were Longer for Bacteria than Fungi

Air back-trajectories of different lengths varied in their correlation with bioaerosol community structure. Therefore, hypothesis 1 (Biological (i.e. particle size) and environmental characteristics (affecting atmospheric residence times) affect optimal spatiotemporal lengths of air-mass trajectories in relation to their ability to predict fine scale temporal variation in related bioaerosol community structure) failed to be rejected. Specifically, one-week trajectories appeared best (highest R^2) for Antarctic bacteria. This trajectory length was longer than is typical in the worldwide literature (three days is often used (Archer et al., 2020)). For Antarctica specifically, Bottos et al. (2014) used seven days and Archer et al. (2019) used three days and two weeks. Antarctic fungi had an optimal trajectory length of one-day, consistent with other areas globally. Shared variation indicated other trajectory length choices are acceptable. Longer trajectory lengths for bacteria are supported by suggestions in the literature of greater bioaerosol residence times in Antarctica (Archer et al., 2019; Burrows et al., 2009b) and greater propensity for aerial dispersal of bacteria opposed to fungi (Archer et al., 2019). Greater residence times could be due to high average wind speeds, keeping particles in suspension for longer (Parish & J., 2003), although why that would affect bacteria more than fungi is unclear. Fungal spores are generally larger than bacteria (Bowers et al., 2013). Larger particles would be expected to move out of suspension more easily due to increased gravitational settling and increased chance of wet or dry deposition (see section 1.3.1 Particle Movement in the Atmosphere), so shorter predicative trajectory lengths and presumed residence times make intuitive sense. However, interpretation of these findings must be tempered, since both bacterial and fungal models had low overall R^2 values (16S – 2%, ITS - 1%). Further, bioaerosols are

thought to often persist in an aggregate form (Huffman et al., 2010), so the size of the whole particle may not be larger for fungi.

4.5.2 Time was Non-Linearly Correlated with Bioaerosol Communities due to Daily Cyclicity and Bioaerosol Communities Varied on Different Days

Dissimilarity in Antarctic bioaerosol communities increased with time over a 24-hour period in a non-linear fashion due to daily cyclicality. Differences in time-dissimilarity plots were evident between different sampling days, with variation in the pattern between days. Therefore, hypothesis 2 (time-dissimilarity hypothesis: Changes in environmental variables that are known to affect bioaerosols (such as UV and temperature) were expected to drive temporal variation in the bioaerosol community, within a 24-hour period, and between different days studied. The relationship between pairwise sample dissimilarity and time was expected to be non-linear due to cyclical day-night periods) failed to be rejected. The transformed time variable was convincingly correlated with ITS variation (R^2 = 17%) and both amplicons had significant non-linear correlation coefficients with transformed time of day. Transformation of the time variable improved R^2 in ITS db-RDA, supporting transformational validity. Other time variables displayed only shared variation with the transformed variable so were excluded. While the db-RDA for 16S was inconclusive, some clustering by time in the NMDS was visible. Clear diel variation for ITS is consistent with bioaerosol literature, with several studies detecting variation over the course of a day in culturable bioaerosol concentrations. Multiple studies observed higher concentrations in the morning and evening (Burrows et al., 2009b; Tong & Lighthart, 1999), while studies of continental bacterial flux detected highest upwards movement in the warmest part of the day (Chen et al., 2001; Shaffer & Lighthart, 1997; Tong & Lighthart, 2000), thought to be driven by higher temperatures increasing turbulence, wind speeds, vertical mixing, residence times and division rates, increasing the microbial load and aerosolisation. Given this background, minimal evidence of diel variation for bacteria was unexpected. While Antarctic read counts were lower for bacteria than fungi, relationships should be detectable with the available data (based on the R^2 values for Baring Head data discussed below, which were much more conclusive with a lower bacterial read count and ASV number than Antarctica). For ITS, variation within a day appeared to be greater than between days (R^2 – 6% between days, versus 17% within days), while the 16S relationship was less clear. Bioaerosol literature indicates patterns in bioaerosol variability can be obscured by stochasticity over short time scales (Burrows et al., 2009b), sampling days one through four also had similar weather conditions, with less variation than over each day so this finding was consistent with field observations. Due to the inherent variability of bioaerosols and the lack of comparable molecular data over short timescales, further fine-grain sampling is needed to crystallise this understanding.

4.5.3 Day, Time of Day and Air-Mass Source Affected ITS Bioaerosol Community Composition

Wind source, time and day were correlated with variation in Antarctic ITS communities. Hypothesis 3 a. (Environmental drivers hypotheses: diel variation in Antarctic bioaerosol community structure (composition and diversity) can be predicted by time of day, geographic origin of sampled air, sample location, temperature, relative humidity, wind speed and UVA, UVB and UVC) therefore failed to be rejected for ITS. It could be rejected for 16S, as the R^2 values for the 16S data were very low. As noted above, 16S reads were lower than ITS meaning bacterial relationships, if they exist, would be harder to detect. Whether due to minimal relationship or insufficient read counts to detect them, the data do not support correlation of Antarctic bacterial bioaerosol communities with any of the measured variables. There was no correlation of temperature, humidity, UV or wind speed with Antarctic bioaerosols for either amplicon. This is in contrast to temperature and humidity being observed to be frequently correlated with bioaerosols elsewhere (Els et al., 2019; Woo et al., 2013), thus is somewhat surprising, and indicates that further Antarctic studies across more days and in more locations are required to confirm this observation. Hypothesis 3 a. is rejected for these variables. Diel variation was expected, as it is known that temperature, humidity, and UV exposure vary predictably over the course of the day and these variables affect bioaerosol concentrations (Burrows et al., 2009b; Els et al., 2019; Hughes, 2003; Woo et al., 2013). Here, diel variation has been observed but does not appear to be linked to any of the variables measured which are thought to drive it. This suggests that some other variables could be affecting Antarctic bioaerosols on a daily cyclical basis, but their nature is unclear from available literature (Burrows et al., 2009b). Identification of further relevant variables which could be measured in a more intensive sampling regime would be prudent. Interrogation of sampling protocols to remove any sources of contamination would be beneficial, as although the samples were processed to correct for contamination, it is possible that this or some other stochastic event could be causing the unexplained temporal variation.

4.5.4 Relative Bioaerosol Abundance did not Vary by Day, Time of Day and Wind Source

Common bacterial genera, such as *Stenotrophomonas* spp. and *Pseudomonas* spp., are found in other bioaerosol studies. They are widely dispersed environmental bacteria. Some more surprising genera are abundant, for instance, *Astreloplasma* spp., which is obligately anaerobic and associated with the bovine rumen (Weisburg et al., 1989). *Brasilomema* spp. is a cyanobacterium, which inhabits water, and the leaves of bromeliads (Fiore et al., 2007). It could have been aerosolised and transported from Spaulding Pond or the nearby sea. Finding marine associated bacteria is consistent with Archer et al. (2019) but not Bottos et al. (2014). Chloroplasts were much less prevalent than in bioaerosol studies in more temperate areas, which is consistent with the lack of higher plant life in Antarctica. Common fungal genera included widespread fungi (such as *Peniophora* spp.) and, interestingly, plant

associated fungi and yeasts (*Rhodosporidium* spp.). Yeasts were identified as the most common fungi in Antarctica by Archer et al. (2019).

Hypothesis 3 b. (relative abundances of common and rarer taxa fluctuate by day, time of day and air-mass source) is rejected for relative abundance, as there was not much evidence of substantial variations in relative abundances. Genera seemed to stay consistent between samples. Changes in microbial source populations which could affect the mix of genera present would be expected to take longer (for instance, bacterial population turnovers tend to be measured in days at a minimum rather than hours (Kevorkian, Bird, Shumaker, & Lloyd, 2018)). Differences in air-mass source could rapidly affect relative abundances, but this did not seem to be the case here. Cluster three appeared distinctive, but this was likely due to few samples mapping to it, skewing the relative abundance data.

4.5.5 Taxonomic Overlap Between Antarctica and Baring Head Increased when Baring Head Air was of Antarctic Origin

Since ASVs with identical sequences were found in Antarctica and New Zealand, at greater rates for bacteria, hypothesis 4 a. (intercontinental connectivity hypotheses: a minority of globally distributed ASVs (identical sequences) were expected to be observed in Antarctica and New Zealand. Bacteria were expected to show a higher propensity for long-range aerial transport due to their longer atmospheric residence times and smaller size than fungi, therefore globally distributed ASVs and genera were expected to be higher for bacteria) failed to be rejected. Bacterial ASVs of 11.8% and 3% of fungal ASVs were shared between all New Zealand and Antarctica samples. This suggested a minority of bioaerosols are very widely dispersed, presumably due to being well adapted to aerial dispersal; this is consistent with observations by Bottos et al. (2014) and Archer et al. (2019) in Antarctica, as well as in other locations (Be et al., 2015; Maki et al., 2017).

Taxonomic overlap in ASVs and genera increased when air sampled in New Zealand transited from Antarctica. Hypothesis 4 b. (aerial transport was expected to occur from Antarctica to New Zealand, at greatest rates for bacteria, resulting in more ASVs/Genera in common (greater overlap in taxa) between Antarctica and New Zealand when Baring Head air came from Antarctica) failed to be rejected. Approximately 7% more ASVs and 11% more genera were identical for 16S when air was coming from Antarctica. There were greater numbers of shared genera than shared ASVs for both amplicons, as an exact sequence match was not required for different ASVs to be placed in the same genus. ITS did not share the same degree of change, only 0.4% more ASVs and 0.5% more genera matching when air was coming from Antarctica. Given samples were taken during one continuous block of time in each location and other variables changing was reasonably controlled for through repeats, this suggests that there could be aerial transport via the wind between Antarctica and New Zealand, in addition to the widely dispersed ASVs always present. Bacteria had many more ASVs in common between the locations, and a bigger increase when the wind was coming from Antarctica, suggesting that they are possibly more adept at long distance

transport, which is in line with the observations in Archer et al. (2019). They found in an NRI (net relatedness index) analysis that fungi were more nested and had less non-local input than bacteria. They inferred greater dispersal limitation for fungi and, therefore, less long-range transport. Few other studies compared bacteria and fungi in this manner, although as previously noted smaller particle size would make long-range transport more likely (Burrows et al., 2009b) and this is consistent with longer back-trajectories correlating with bacterial variation already discussed.

There was a temporal delay between the Antarctic and the Baring Head sampling of approximately six months. The result above however is still relevant for discussion here for various reasons. Modeling and bioaerosol sampling in Antarctica consistently suggest that bioaerosol transport from elsewhere to Antarctica is uncommon and that most bioaerosols sampled in Antarctica originated there (Archer et al., 2019; Bottos et al., 2014; Burrows et al., 2009b). It follows that the ASVs present in Antarctica would be consistent over time, particularly relatively short time scales of a few months. Since it is reasonable to expect consistency over these time periods in the ASVs present in Antarctica it remains reasonable to perform a comparison with New Zealand with a temporal difference between sampling dates. The mismatch in timing however does represent a limitation in the current work. It would be beneficial to perform sampling during an overlapping time window at both locations to strengthen the case for intercontinental transport between them. Observation of consistent ASVs, even with concurrent sampling, would only suggest that transport could be occurring, and could not prove transport. In order to do that, a marked tracer would need to be released and re-captured at the distant location, which would be another potentially fruitful area to focus future effort.

Bacteria were more affected by air-mass source than the fungi, which were more affected by local variables. Therefore, hypothesis 4 c. (bacterial composition was expected to be more affected by air-mass source than fungal composition, and fungi were expected to be more dependent on local environmental variables (temperature, relative humidity and weather) than bacteria) failed to be rejected. Db-RDA with variance partitioning for the three different wind back trajectory lengths indicated that one-week wind was the best correlate with the bioaerosol community for bacteria, and one-day wind was the best correlate for fungi. The complete db-RDA with variance partitioning for Baring Head showed an R^2 of 11% for one-week trajectories for bacteria, and also indicated temperature was important. For ITS, one-day wind was all shared variation with local variables so did not feature in the final model. Temperature, humidity and weather were indicated as the best correlates with the bioaerosol community. This suggested that local sources were more dominant for fungi and distal sources were more important for bacteria, which is consistent with Archer et al. (2019) and the higher observed bacterial taxonomic overlap in this study.

Relative abundances of bacteria and fungi varied depending on back trajectory cluster. Therefore, hypothesis 4 d. (relative abundance of bacteria and fungi were expected to vary depending on air mass source, as different trajectories passed over different regions

with variable microbial emissions) failed to be rejected. Common bacterial bioaerosols reflecting widespread environmental organisms featured in Baring Head samples. Chloroplasts were again common, expected since there are many more higher plants in New Zealand than Antarctica. Fungal species also reflected common bioaerosol associated organisms. Cluster two had a distinctive species composition and transited over the sea, which may have caused this variance (Amend et al., 2019).

4.6 Conclusions

The first fine-temporal-grain bioaerosol dataset collected in Antarctica revealed notable diel variation for fungi but not bacteria, and comparison to samples taken in Aotearoa New Zealand hinted at surprisingly frequent long-distance intercontinental bacterial transport. Several key questions have emerged from this: (1) Do Antarctic bacteria vary within a 24-hour period at all? (2) What factors drive the observed temporal variation in the fungi? (3) How does the biology and size of an organism relate to its ability to undergo long-range aerial transport? (4) How microbially connected is Antarctica to the rest of the world? (5) How can sampling protocols be improved to get the data needed to answer those questions quickly and easily? Further hourly sampling in Antarctica and surrounding areas is recommended, with measurement of additional variables. Concurrent sampling with New Zealand is also suggested to shed more light on the potential intercontinental transport that could be occurring. This research makes a tantalising start that highlights pertinent future research directions.

Chapter 5 - General Discussion

Earth's atmosphere provides a habitat and conduit linking all global ecosystems for microorganisms which can tolerate its inhospitable conditions (Womack et al., 2010). Bioaerosols can be pathogens, or allergens, invasive species, or affect atmospheric chemistry, and yet are little studied (Burrows et al., 2009b; Pearce et al., 2009). Quantifying variation and connectivity of the aero-microbiome and understanding its drivers is crucial for predicting future ecosystem changes and managing outbreaks of dangerous microorganisms, such as the SARS-CoV-2 coronavirus that is the causative agent of COVID-19 (Pearce et al., 2016). In this thesis, I explored the variation in bioaerosol communities over varying spatial and temporal scales to begin to understand what variation in community structure there is, and how and why it changes over space and time. Sampling methods, DNA extraction methods and bioinformatics pipelines were optimised and used to describe the aero-microbiome of urban parks in Auckland, Aotearoa New Zealand. This work provided insights into microbial exposure of city-dwellers and bioaerosol variation over smaller spatial scales, revealing spatial differentiation despite local population connectivity. In the longest seasonal study of urban bioaerosols to date, Auckland bioaerosols were sampled continuously for two years, which provided information on how bioaerosols vary over longer timescales. Fungi showed correlation with temperature and pronounced seasonality, while bacterial bioaerosols were less temporally variable. Finally, the aero-microbiome of pristine Antarctica was described, and its fine-grain temporal variation and intercontinental connectivity were quantified for the first time. Antarctic fungi exhibited marked diel variation and bacteria showed evidence of regular intercontinental transport to New Zealand. In this final chapter, the overarching themes emerging from the body of work presented are teased out and put into wider context, before potential avenues for future research are described.

5.1 Optimisation of Research Methods

Appendix D Optimising Bioinformatics Protocols for Aerosol Microbial Community Data – a Case Study Using an Urban Parks Dataset and Appendix A Method Development detail the method development work underpinning the sampling, laboratory processing and bioinformatic inference procedures for the experimental chapters of the thesis. Method improvement was important since a large portion of the bioaerosol studies reviewed in previous chapters of this thesis relied upon dated non-molecular methods, generally either cultivation or microscopy. However, it is known that 70 to 90% (Burrows et al., 2009b) of environmental microbes do not grow in culture, and it can be difficult to distinguish morphologically similar organisms through microscopy. As a result, it is very likely that diversity is underestimated when these methods are employed, and use of next generation sequencing (NGS) approaches is preferred to capture more of the diversity present (see section 1.3.6 Bioaerosol Sampling: Challenges and Solutions). NGS approaches allow characterisation of large numbers of samples at a reasonable cost and produce millions of reads for analysis. Unfortunately, bioaerosol work presents particular challenges for NGS, as air typically has very low biomass, which tends to result in very long sampling durations (24

hours to two months) in order to collect enough material (Bottos et al., 2014). As a result, no fine-scale NGS temporal data existed in the bioaerosol literature reviewed, as sampling bioaerosols on an hour-by-hour basis was not feasible. Dry filters are often used to capture microorganisms, but these are biased against Gram-negative bacteria due to the desiccation stress over the long sample durations (Luhung et al., 2015). Using a liquid impinger circumvents these issues, with much higher flow rates and reduced desiccation stress on the sampled microorganisms.

The Coriolis µ (Haig et al., 2016) liquid cyclone impinger was investigated for generating the data for this thesis. It has high flow rates (300 L/min), allows sample durations from 10 minutes to six hours, and it has been utilised in a small number of very recent studies (Archer et al., 2020; Archer et al., 2019; Els et al., 2019). However, water based liquid collection media are problematic in cold environments, like Antarctica, due to freezing. Following recommendations based on work performed in Appendix A Method Development, Archer et al. (2019) used the Coriolis with an RNAlater collection medium (which has a depressed freezing point) in Antarctica, collecting samples in two – four hours. I thought it beneficial to invest time in further method development for this thesis, as preliminary testing indicated DNA recovered from RNAlater was approximately half that recovered from PBS. It was also beneficial to have a consistent approach for the whole PhD to allow comparisons to be made across datasets and for the convenience of one sample processing pipeline. A modification to the Coriolis including heating elements was developed, which kept the liquid collection medium from freezing in cold conditions (see section A.2.1 Methods for Operation in Sub-Zero Environments). A sampling duration of one hour was found best to reliably return quantifiable DNA in New Zealand, which is a significant improvement on previous studies. For some studies, the samples were pooled later on into two-hour windows to improve biomass (Antarctic and Baring Head studies). DNA extraction methods were optimised to work with the new sampling protocols (see section A.3.1 Collection Liquid and Sample Processing and Appendix C Laboratory Methods). These approaches were used consistently for all samples collected for this PhD thesis.

Optimisation was also performed in relation to bioinformatic pipelines used for inferring ASVs present and taxonomy from the NGS data produced (see Appendix D Optimising Bioinformatics Protocols for Aerosol Microbial Community Data – a Case Study Using an Urban Parks Dataset). Tools used for bioinformatics and taxonomy databases can markedly affect the outputs produced (Edgar, 2018). USEARCH was chosen as it was observed to produce fewer spurious ASVs. ITS data, in particular, often has a high percentage of ASVs which cannot be assigned taxonomy based on one database (Archer et al., 2020). It is much harder to make inferences about an ASV without taxonomic information. Therefore, a combined taxonomy approach was developed for this PhD thesis, which utilised information in multiple taxonomy databases simultaneously and improved the rate of taxonomic assignments by 12 – 15%.

Removal of contamination is very important for bioaerosol studies, as members of the aero-microbiome are often contaminants as well (presumedly due to their ease of atmospheric dispersal). Similar studies used whole ASV deletion (Archer et al., 2020), but an approach calculating a background contamination profile and deducting that from each sample was assessed as more appropriate for this work.

Consideration of compositionality is important for all NGS based studies, as data generated are inherently compositional (Gloor et al., 2017). Methods that account for compositionality were not seen in any of the literature reviewed for this PhD. Compositionality means that read counts represent part of a whole and only contain relative not absolute abundance information (Aitchison, 1982). This occurs due to standardisation steps in the sequencing process and random variation in the read depth during sequencing. Using non-compositional methods on compositional data causes various problems, such as confounding distance matrices, like Bray-Curtis (BC), which are very commonly used (Gloor et al., 2017). In this PhD, data from the Urban Parks study were analysed with a range of compositional and non-compositional methods and the outputs compared. The compositional approach utilised an Aitchison distance matrix (a centred log ratio transformation) and the non-compositional approach used BC distance matrices, a Jaccard presence-absence matrix and Hill numbers for diversity. The Jaccard presence-absence matrix circumvents the non-compositional/compositional issue, as read counts are converted to presence or absence data so problems with inaccurate raw read counts disappear, providing a good reference point. The results from all tools were broadly ecologically consistent and the main themes did not change, suggesting choice of method at this stage was not of critical importance. Compositional tools were used for the remainder of the PhD. Consideration of approach and comparison of different tools made the results reported more robust.

Method enhancements have broad applications outside of this PhD thesis, as bioaerosol methods in even relatively recent literature are often dated, inconsistent and slow (Bottos et al., 2014; Pearce et al., 2016). The methods used in this thesis are consistent, fast, flexible in different environments and accurate in inferring ASVs, removing contamination and assigning taxonomy. They allow improvement in the data generated and conclusions drawn and future comparisons of data on a global scale.

5.2 Bacterial and Fungal Genera Identified

Fungal and bacterial genera identified included common environmental organisms (plant, soil, water and animal associated). Some surprising genera were identified in Antarctica, such as *Astreloplasma* spp., which is associated with the cow rumen. This could be due to contamination or incomplete understanding of this organism's biological niche. Genera containing pathogens (such as *Listeria*) and allergens (such as *Khuskia* spp.) were identified in Auckland. Fungi were largely wood or plant associated or yeast species. Common genera identified consistent with previous bioaerosol studies were *Pseudomonas*, *Ralstonia*, *Penicillium* and *Alternaria* spp. A large proportion of chloroplasts were identified in New

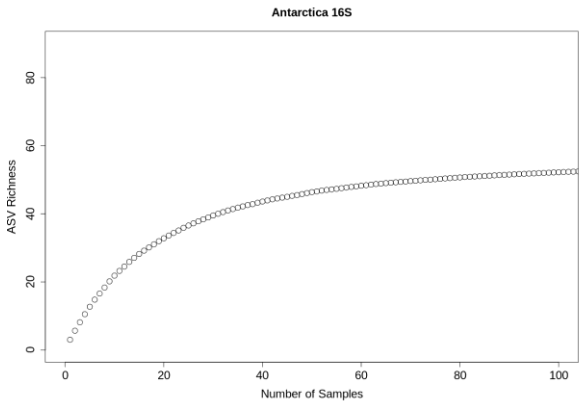
Zealand, which is consistent with previous bioaerosol studies in temperate locations, likely from plant pollen. The lack of higher plants in Antarctica explains the absence of chloroplasts there.

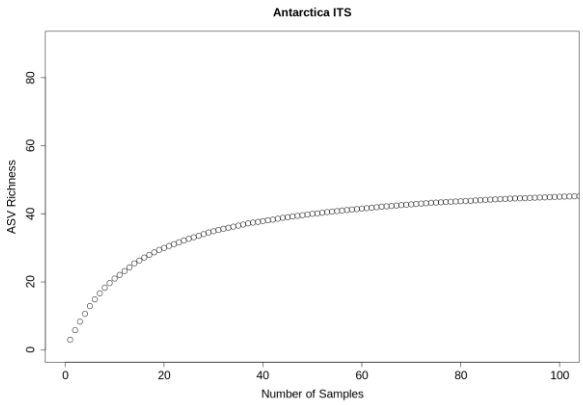
5.3 Spatial Variation in Bioaerosol Communities

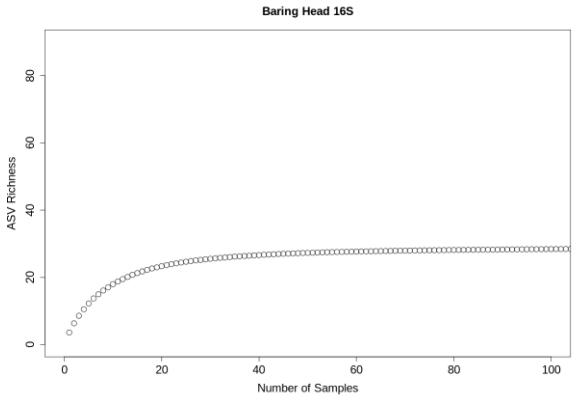
Over small distances, bacteria and fungi appeared to disperse effectively via the atmosphere, whereas long-distance dispersal appeared to be much more frequent for bacteria than fungi. Local spatial differentiation was apparent despite this ease of dispersal. The Urban Parks study showed local differentiation within one city, likely due to differences in local sources in each sample location (variation in bacterial composition explained by location was 9% plus 7% was shared with park size; for fungal composition, 7% was explained by location plus 8% shared variation with other variables). As predicted, there was minimal evidence of distance decay in compositional similarity over the 8km study range. This was likely because some organisms were not adept at aerial dispersal and remained local, differentiating the parks, while others were good at aerial disposal and easily dispersed at that scale. Filtering out invariant (presumably well dispersed) ASVs prior to analysis was required to detect local differentiation, supporting the idea of different members of the community having different propensities for aerial dispersal. Lack of distance decay has been observed in other small scale bioaerosol studies (Archer et al., 2019). The Antarctic study revealed indications of surprisingly common bacterial exchange with New Zealand; 12 to 19% of ASVs were shared between the two locations dependent on wind direction, although few fungi appeared to disperse effectively at these scales. This is consistent with a very small amount of literature which has compared bacterial and fungal bioaerosols, which suggested that bacterial bioaerosols showed less evidence of local nesting and hypothesised that bacteria are more adept at long range aerial dispersal (Archer et al., 2019). Diversity and species assemblages varied between Antarctica and New Zealand as expected. Surprising diversity was found in Antarctica, which was comparable to New Zealand in terms of numbers of ASVs/reads and in species accumulation curves (**Figure 5-1** and **Table 5-1**). More Antarctic sampling would be needed to draw more detailed inferences about differences between locations.

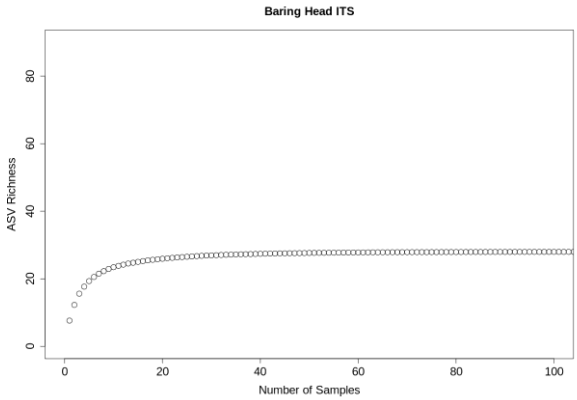
Table 5-1 – Read count and ASV numbers (raw alpha diversity/Hill D0) and sample number comparison for each study

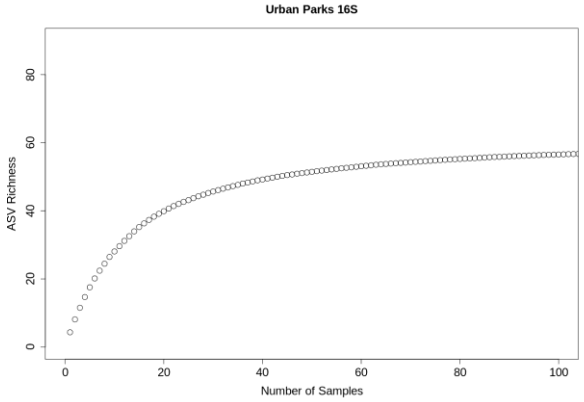
	Reads	ASVs	Samples	Reads/Sample	ASVs/Sample
Antarctica 16S	897,244	2,276	58	15,470	39
Antarctica ITS	1,579,022	1,569	55	28,709	29
Baring Head 16S	310,354	2,009	29	10,702	69
Baring Head ITS	2,470,453	5,469	28	88,230	195
Parks 16S	607,740	3,828	68	8,937	56
Parks ITS	1,796,883	5,311	66	27,226	80
Seasonal study 16S	1,171,139	5,235	100	11,711	52
Seasonal study ITS	2,760,306	3,419	100	27,603	34

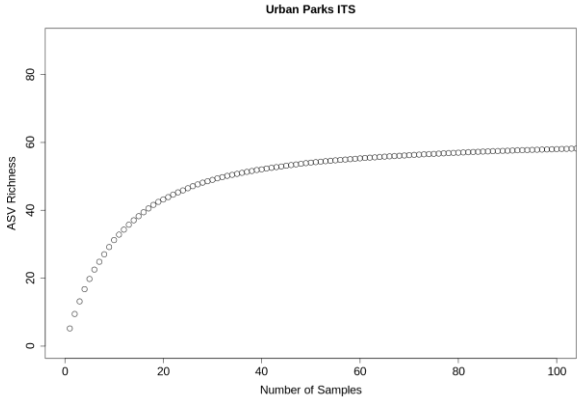

Antarctica 16S
ASV Richness
Number of Samples
0
20
40
60
80
100

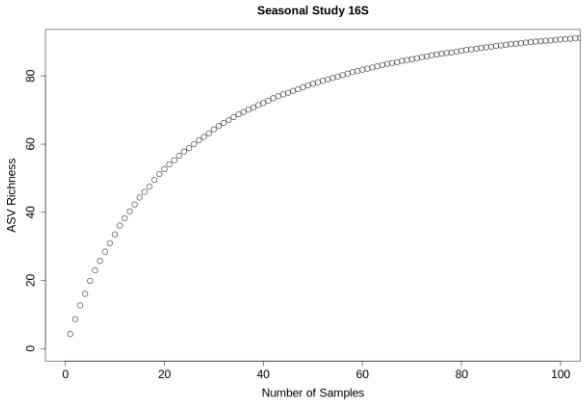

Antarctica ITS
ASV Richness
Number of Samples
0
20
40
60
80
100


Baring Head 16S
ASV Richness
Number of Samples
0
20
40
60
80
100


Baring Head ITS
ASV Richness
Number of Samples
0
20
40
60
80
100


Urban Parks 16S
ASV Richness
Number of Samples
0
20
40
60
80
100


Urban Parks ITS
ASV Richness
Number of Samples
0
20
40
60
80
100


Seasonal Study 16S
ASV Richness
Number of Samples
0
20
40
60
80
100

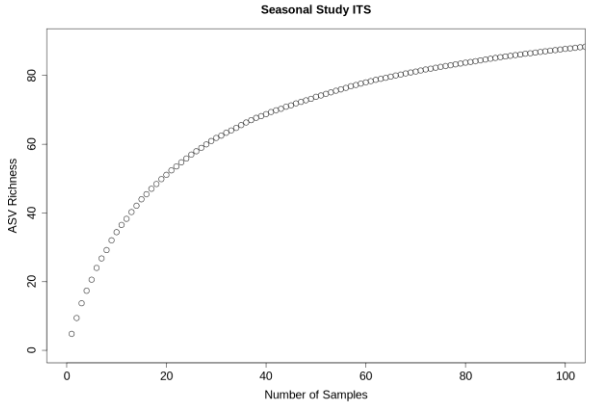

Seasonal Study ITS
ASV Richness
Number of Samples
0
20
40
60
80
100

Figure 5-1 – ASV accumulation curves for each study. ASV accumulation curves were calculated on unfiltered data with the function specaccum in the R package "vegan" (Oksanen et al., 2012), using random sample order and 999 permutations.

5.4 Temporal Variation in Bioaerosol Communities

Fungal bioaerosols were highly variable over time across all timescales measured. Bacterial bioaerosols appeared to show much less temporal variability overall. The results in the urban parks study suggested much more temporal variation for fungi than bacteria (variation in bacterial composition explained by time was 0%; for fungal composition, 18% was explained by time). The seasonal study revealed a non-linear temporal response for fungi and bacteria, showing seasonal and annual variation. Again, fungi responded more than bacteria (variation in bacterial composition explained by time was 6%; for fungal composition, 23% was explained by time). The Antarctic study revealed diel variation for fungi only, and had somewhat inconclusive results for bacteria (variation in bacterial composition explained by time was 0%; for fungal composition, 18% was explained by time of day and 6% was explained by day). The diel variation did not appear to dramatically affect composition. Over longer timescales, variation in composition was observed. This was especially true for fungi, where the dominant genera changed multiple times over the course of the year. Changes in microbial source populations, which are needed to alter composition, would be expected to occur more slowly over time, based on bacterial population turnover times (Kevorkian et al., 2018). Temporal variation was expected to be observed as factors which are thought to affect bioaerosols (such as temperature) vary over time, and seasonality is very frequently detected in bioaerosol literature (Burrows et al., 2009b; Mhuireach et al., 2020; Woo et al., 2013). The lack of seasonality in bacterial bioaerosols was surprising given this context. The importance of variables appeared to differ over different timescales, suggesting investigation at multiple timescales is pertinent. Wind appeared to be more important at shorter timescales, while season was the main driver of temporal variation for longer timescales. There also appeared to be an interaction between the timescale of the study and optimal filtering applied to ASVs. The shorter timescale studies appeared to require no filter to best extract signal from the bioaerosol data, while for longer studies filtering was required to reduce noise in the data. Perhaps over longer timescales there is more time for stochastic events to occur, disrupting observed patterns, making filtering more valuable. This consideration will be useful for future studies.

5.5 Relative Importance of Environmental Variables Measured to the Sampled Bioaerosol Communities

Variability was observed in the effect of environmental variables thought to be important to bioaerosols on the aero-microbiomes recovered. In the Urban Parks study, all variables that were indicated by forward selection as correlated with the bioaerosol community were related to each location, wind trajectory or time. Temperature, weather and relative humidity were not correlated with bioaerosols. However, in the seasonal study, temperature and weather were correlated with the bioaerosol community, but relative

humidity was not. The Antarctic data showed no correlation with relative humidity, temperature, weather, wind speed or UV with bioaerosols. Baring Head sampling suggested that 16S was correlated with temperature but not relative humidity, while ITS was correlated with temperature, relative humidity and weather. From these data, it seems that temperature is likely to impact bioaerosols, which is consistent with the literature (Priyamvada et al., 2017; Woo et al., 2013). The other variables remain somewhat inconclusive, relative humidity and weather may be impactful to fungi as per this thesis but they seem unlikely to be key drivers of bioaerosol communities. This was surprising but is also consistent with the body of literature reviewed (Mhuireach et al., 2020; Priyamvada et al., 2017). All models had large unexplained residual values (which is not unusual for these types of studies (Mhuireach et al., 2020)) so it is likely variables yet to be identified are important. Variation occurred over time that was not explained by the measured variables (such as temperature) which are thought to drive changes. Some of this variability could be the effects of stochasticity, but the data suggest uncharacterised variables are at play.

5.6 Relative Importance of Back-Trajectory Cluster to the Aero-Microbiome

The bioaerosol literature is divided on the importance of wind back-trajectory to aerial microbial communities. Several recent papers postulate it is the "key determinant" or similar of the bioaerosol community (Archer et al., 2020; Woo et al., 2013). **Table 5-2** below does not suggest that wind origin is the key driver of sampled bioaerosols. Wind source has consistent low measurable R^2 values across most studies. The 16S amplicon does seem to have more dependence on wind source than ITS, and wind source appears to be more important in the shorter duration studies (Antarctica and Baring Head). Wind source is not likely to be the key determinant in the longer-term studies. This could be understood as it is possible to rapidly alter microbial input over short time frames with a wind source change (potentially thousands of kilometres shift in air-mass origin in a matter of hours). Over a longer timescale, local influences are more impactful as microbial populations change and the majority of bioaerosols are thought to be locally dispersed (Bowers et al., 2013). Wind should be considered among the variables which affect bioaerosols, but especially for longer term studies, does not appear to be the key driver of the bioaerosol community.

Table 5-2 – Comparison of the R^2 in db-RDA with variance partitioning models in each study pertaining to wind versus other variables to indicate relative importance of wind as a driver of the bioaerosol community.

Study Amplicon	R^2 wind back-trajectories (%)	Highest non-wind R^2 in db-RDA (%)
Baring Head 16S	19	7
Parks 16S	6	9
Antarctica 16S	5	2
Parks ITS	2	18
Antarctica ITS	2	17
Seasonal study ITS	2	13
Seasonal study 16S	1	7

Baring Head ITS	0	8

5.7 Differences Between Bacterial Bioaerosols (16S) and Fungal Bioaerosols (ITS)

Multiple differences in the behaviour of bacterial and fungal bioaerosols were observed. This was not unexpected given they are very different types of organisms and few studies in the literature have sequenced both, so this is one of the first opportunities to make comparisons (Archer et al., 2020; Archer et al., 2019; Bowers et al., 2013; Woo et al., 2013). Fungal read counts were consistently higher (by two times or more), but the number of ASVs were not always greater. R^2 values for fungal modelling were reliably higher than for bacterial models (other than from Baring Head) (**Table 5-4**). Other coefficients, for example the non-linear correlation coefficient, were also higher for ITS. It is unclear if fungi were more variable and more responsive to changes in conditions or if the R^2 values were higher due to the higher read counts. It is likely that both factors played a part, as fungi are known to be highly seasonal and have more host linked niches and supposed variability (Shigyo et al., 2019). Higher read counts could have occurred due to fungal spores being larger than bacteria, and as they are multicellular more incidents of clumping of multiple cells together are likely to occur (Lagomarsino Oneto et al., 2020). Wind was observed to be more important for 16S, season was more important for ITS. These observations are likely linked, as it is thought that the assumed larger size of the fungal particles mean lower atmospheric residence times, and therefore more abrupt changes in fungi over time and increased local-source importance. This is supported by the observation that bacteria generally are linked to longer wind back-trajectories than fungi (**Table 5-3**). Less long-range fungal dispersal has been suggested in the literature. To investigate, with the data available, whether fungal bioaerosols sampled are larger than bacterial bioaerosols, the read counts for bacteria and fungi for the seasonal study were correlated against two size fractions of aerosol particles measured at the time of sampling. 0.3 to 3 µm in diameter were assumed to be likely bacterial and 5 to 10 µm in diameter were assumed likely to be fungal, based on available literature (Tanaka et al., 2020). Neither size fraction correlated with either bacterial or fungal read counts, with very low R^2 values and insignificant P values, and this failed to shed any light on the relative sizes of bacterial and fungal bioaerosols.

Table 5-3 – Optimal back-trajectory (highest R^2 from db-RDA) chosen for further modelling for each study

Study Amplicon	Trajectory chosen for analysis
Antarctica 16S	One-week
Antarctica ITS	One-day
Baring Head 16S	One-week
Baring Head ITS	One-day
Parks 16S	One-day
Parks ITS	One-day

Seasonal study 16S	One-day
Seasonal study ITS	Three-days

Table 5-4 – Comparison of R^2 values of the full db-RDA model for each study

Study Amplicon	Overall R^2 in db-RDA
Parks ITS	38
Seasonal study ITS	33
Antarctica ITS	25
Parks 16S	19
Baring Head 16S	17
Baring Head ITS	13
Seasonal study 16S	8
Antarctica 16S	0

5.8 Future Directions

This thesis presents foundational insights into how bioaerosols vary over different spatial and temporal scales, bringing threads together from existing research and augmenting them with new information, to create an overview of current knowledge and emerging patterns. The research reveals intriguing observations, which raise further questions, as well as inconclusive information or surprising lack of relationships which require confirmation or suggest missing information. These avenues for further research are:

1. Study of bioaerosol variability at further spatial and temporal scales. The studies in this thesis show that the nature of bioaerosol sampling performed, in terms of time scale or spatial location, affects not only the results, but the drivers of the community (i.e. wind is important over short time scales, such as days, but a much less important driver over longer scales like months or years). In order to understand drivers of bioaerosol communities, it holds that further investigation of spatial and temporal scale is needed. In particular, further years in the Seasonal study would reveal if patterns persisted over three or more years, and if bacteria and fungi aligned more in response over longer time scales. The Antarctica study accumulated four full 24-hour periods of samples, each day varied and it would be very beneficial to have several more days to understand this variation better, and to have more conclusive information for the bacteria. It would be valuable to perform sampling in the same time window in Antarctica and New Zealand to strengthen the case for intercontinental transport between them. Observation of consistent ASVs, even with concurrent sampling, only suggests that transport could be occurring, and does not prove transport. A marked tracer could be released in Antarctica and re-captured in New Zealand to provide this proof. The field will also benefit from much more data in general, which covers more areas globally, to confirm the validly of patterns identified in differing sets of circumstances. Further molecular studies are needed, as still much of the literature informing this thesis was based on outdated methods for which we now have much better alternatives (Burrows et al., 2009b). In particular, more fine temporal scale molecular work is needed, as this area is particularly

depauperate due to the technical difficulties of sampling sufficient biomass in a short time (Pearce et al., 2016).

2. Identification of novel variables that impact bioaerosols and measuring them alongside future sampling to confirm their degree of impact. Every chapter of this thesis showed large amounts of unexplained variation in the models produced, and even when variation was associated with time of day, or location or day or season, it was not shared with measured variables which were predicted to drive this variability (Womack et al., 2010). These observations confirm that bioaerosol communities can be highly stochastic (Burrows et al., 2009b), but they also suggest that our current picture is incomplete and further work is needed to understand the drivers of variation in the aero-microbiome. New technology may be needed to identify and measure these variables.
3. The studies in this thesis revealed consistent and dramatic differences in the bacterial and fungal results. As few studies compared these types of organisms side by side, little insight exists in the current literature as to possible causes of these differences (Archer et al., 2020; Archer et al., 2019; Woo et al., 2013). There were some suggestions of reduced fungal population connectivity, residence times and atmospheric transport range, due to their larger size, but this has not been empirically tested (Archer et al., 2019). Larger size of fungal particles could explain most of the differences, in terms of greater read counts and also more local influence and less evidence for long-distance aerial transport. Few attempts have been made to understand the size range of bioaerosol particles, for example Tanaka et al. (2020) sampled bacterial bioaerosols using a size resolved sampler from 11 µm to 0.43 µm and found different species in different size fractions. As a starting point for understanding the fungal and bacterial differences, further work is necessary to determine if environmental fungal bioaerosols are larger than bacterial ones. Similar methods with staged samplers could be employed with fungal and bacterial amplification performed from each size fraction and differences understood. Further follow up work could then investigate if bacteria do have longer atmospheric residence times than fungi in practice, perhaps by using labelled particles in a controlled environment. Based on the observed patterns fungal particles are likely to be larger and persist in the atmosphere for shorter periods of time.
4. Further consideration of different types of bioaerosols are needed. The existing literature focusses very heavily on bacteria, with fungal work in the minority and very few studies assessing both types of organisms (Burrows et al., 2009b). Viral bioaerosols are thought to be extremely numerous and potentially important but have been studied even less (Reche et al., 2018; Whon, Kim, Roh, Shin, & Lee, 2012). Other types of bioaerosols, such as viruses, pollen (using molecular methods, some work with microscopy has been performed), cell fragments, toxins and protists represent a largely ignored black box, for which technology now allows investigation (per below), and which need to be understood to place the existing bioaerosol data in context.
5. Usage of more advanced molecular techniques to resolve further questions about bioaerosol communities other than presence of particular 16S and ITS sequences.

While NGS has revolutionised studies of environmental microorganisms, single amplicon studies, as used in this thesis and commonly in the literature have various drawbacks (Archer et al., 2020). Since bacteria, in particular, are adept at horizontal gene transfer and can have multiple copies of their 16S gene, the 16S taxonomy may not represent the taxonomy of the organism as a whole (Behzad et al., 2015). Quantitative PCR provides information on absolute abundances and would be very useful to understand shifts in absolute, as well as relative abundance. Metagenomics allows retrieval of the whole genome of an organism, and a much better understanding of its phylogeny and metabolic capability. This has been attempted a handful of times for bioaerosols, but low biomass remains a challenge (Jiang et al., 2015). Metagenomics detects all DNA, so encompasses a much wider range of organisms, helping to address the knowledge gaps outside bacteria discussed above. Transcriptomics (retrieval of mRNA) could also be extremely useful to resolve further questions about bioaerosols. It would allow detection of RNA viruses and understanding of which genes are actively being expressed in the air. To my knowledge, this has rarely been applied to environmental bioaerosols (Amato et al., 2019), but would be invaluable to shed light on uncertainties surrounding their existence either in a dormant state, purely transiting as spores, or if they are metabolically active in the atmosphere such that it represents a further microbial habitat (Womack et al., 2010). Using new methods for removal of relic DNA (DNA from dead cells or extracellular DNA) to clarify its impact on observed temporal and spatial patterns would also be a valuable avenue for further investigation of bioaerosol communities (Carini et al., 2020).

5.9 Concluding Remarks

The research in this thesis built upon existing bioaerosol literature and explored bioaerosol communities at different spatial and temporal scales in an attempt to understand broad patterns of variation, key community drivers, and dispersal abilities of microorganisms through the atmosphere. Methods from sampling to bioinformatics were improved and optimised for a wider range of conditions, ease, comparability and accuracy. The enhancement of methods provides further options for future study. I demonstrated that bioaerosols vary on a local scale yet show minimal city-wide distance-dissimilarity and substantial seasonal variation, especially for fungi. I also showed that wind back-trajectory was an influence on the aero-microbiome, but unlikely to be the key determinant. I demonstrated diurnal variation in Antarctic fungal bioaerosols and indications of regular transport of bacteria from Antarctica to New Zealand. This thesis contributes to our understanding of the aero-microbiome, and highlights promising future research directions. As understanding moves from what is there, to how and why it changes, we can begin to make predictions about future events, which will be of crucial importance in mitigating potential future ecosystem changes and disease spread in an increasingly uncertain world.

Appendix A. Method Development

A.1 Introduction

Air has extremely low biomass, so high-volume sampling is critical to ensure adequate DNA is recovered to understand the microbial community present. Volumes used in typical next-generation sequencing surveys vary widely, but it can take as much as 75,000 L of air (Bottos et al., 2014) to gain sufficient biomass for molecular analysis. While dry filters and long exposure times (24 hours +) have been recommended for use in these studies (Pearce et al., 2009), a recent survey has shown this method of sampling may cause significant loss (up to 98%) of selected microbes, particularly more vulnerable Gram-negative bacteria (Luhung et al., 2015). Therefore, method improvement to address these known issues was pertinent.

Investigation of cold regions is critical to complete the picture of world-wide microbial distribution and transport, given such regions are typically more isolated, warming rapidly and have wide-ranging impacts on the rest of the world (for example through ice cap melt)(Pearce et al., 2016).

The use of high-volume cyclone samplers has been investigated to address the issues of both the long sample duration and DNA loss of certain types of microbes. The Coriolis μ (Haig et al., 2016) high-volume liquid-cyclone air sampler was selected. The Coriolis processes air at a rate of 300 L/min and can theoretically capture sufficient biomass for molecular analysis in just over four hours. Comparative trials were performed to determine optimal sampling durations. The Coriolis' liquid collection medium reduces shearing forces and desiccation experienced by microbes during the collection process, which should improve DNA recovery for more fragile microbes. However, most liquids commonly used have a freezing point of 0 °C so are not suitable for very cold environments. Various collection media and heating of the sampling apparatus were investigated with the aim of expanding usage of the Coriolis in sub-zero locations by preventing freezing of the collection liquid.

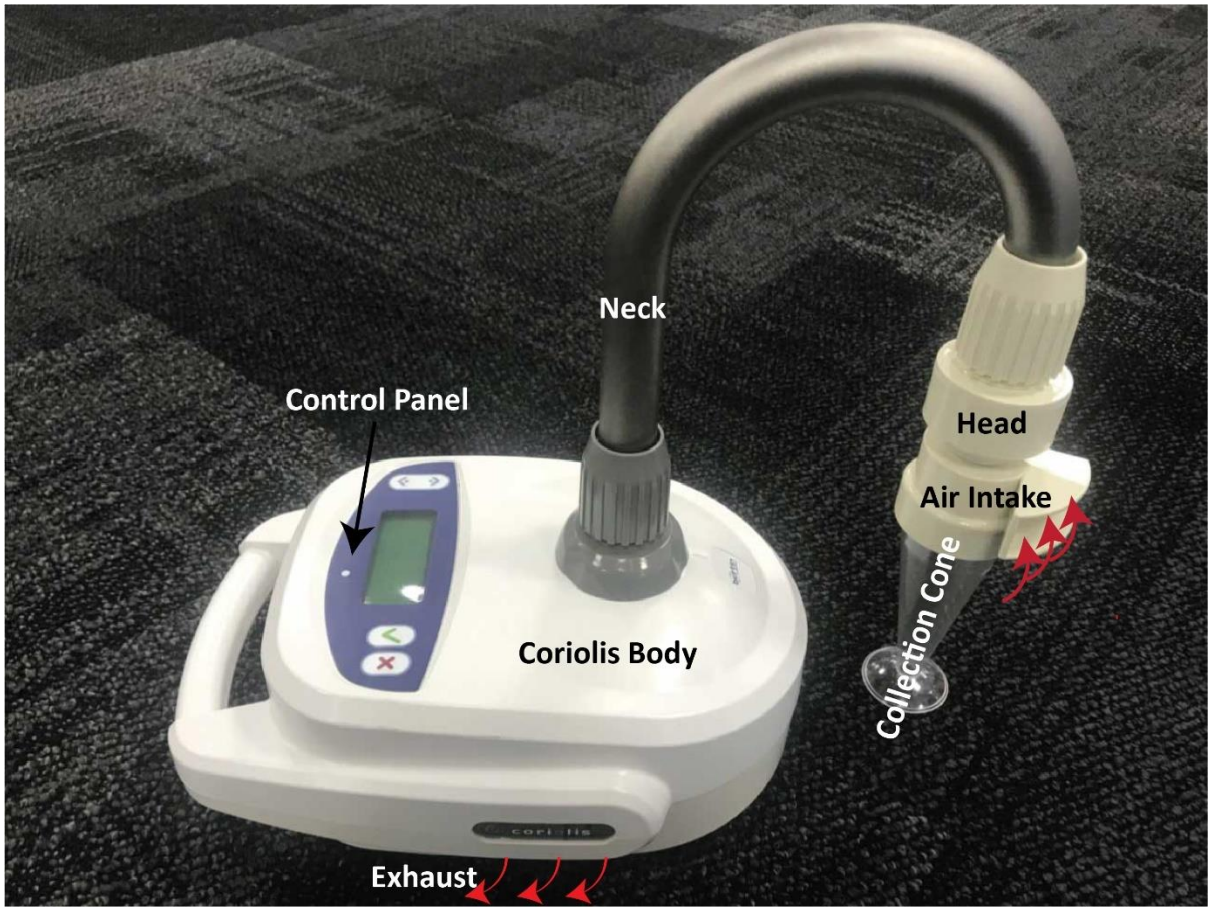


Figure A-1 - Labelled image of a Coriolis high-volume liquid cyclone air sampler. Red arrows indicate air flow.

A.2 Methods

A.2.1 Methods for Operation in Sub-Zero Environments

Trials to determine the most appropriate collection liquid were carried out by adding 100 µl of *Escherichia coli* broth to a 15 mL sample (this is the volume in the collection cone of the Coriolis) of the various collection liquids tested. The collection mediums included ethanol, glycerol, ethylene glycol and RNAlater. They all have a freezing point below 0 °C and were expected to remain liquid in polar and high-altitude conditions. The sample was processed via several methods and the DNA concentration returned by each quantified via Qubit (Invitrogen, 2010). Testing was performed in triplicate. A more easily damaged Gram-negative bacterium was chosen, to confirm the method did not discriminate against these types of bacteria. The Coriolis sampler was tested in a shipping container refrigerated to -20 °C (under the standard operating instructions) for nearly an hour, to confirm that the battery and sampler were able to withstand those conditions.

The optimal method as described below (collection in RNAlater, filtration of sample fluid and CTAB DNA extraction) was fully field tested in Antarctica in January 2017 and the resulting data and analysis have been published (Archer et al., 2019). Various methods were tested to separate the microbial cells for DNA extraction from the collection liquid. The

superior method used a vacuum flask that had been treated with bleach and rinsed with 70% ethanol and milli-Q H_2O (MQH_20) to remove any traces of DNA. The *E. coli* spiked liquid was run through a 0.2 µm polycarbonate filter and the filter was washed with phosphate buffered saline (PBS) to remove any residual liquid that may interfere with the downstream DNA extraction process. The filters were cut up into small pieces and half was put through a manual CTAB extraction protocol (see details in C.1 Laboratory Methods – DNA Extraction) and half through a Power Soil kit using the standard protocol as per the manufacturer's instructions (Qiagen, 2017). On-filter DNA extractions were performed in line with methods previously described (Archer et al., 2014). The DNA concentration in each sample was assessed using Qubit as per the manufacturer's instructions, with 2 µl of the sample added to the assay tubes (Invitrogen, 2010).

A heated head for the Coriolis was developed that allowed the use of any collection liquid in the Coriolis cone, greatly increasing the flexibility and ease of operation of the machine. This was field tested in Antarctica in January 2018 with the methods and results of this work described in Chapter 4 - Diel Variation and Intercontinental Connectivity of Antarctic Bioaerosols.

A.2.2 Optimisation of Sample Duration

The Coriolis was run for one, two, three and four hours in temperate NZ conditions (operating the unit as per field methods in section 2.3 Methods) and DNA extracted via filtration, CTAB and quantification via Qubit as described above.

A.3 Results and Discussion

A.3.1 Collection Liquid and Sample Processing

DNA concentrations were highest for ethanol, using filtration for pre-processing with either a Power Soil kit or CTAB extraction (**Figure A-2**). Ethanol was difficult to use as its low viscosity and volatility caused it to be thrown out of the sample cone and evaporate rapidly. Therefore, RNAlater was chosen as the sample collection liquid. Filtration was chosen for pre-processing the samples as it had the highest DNA concentrations recovered. A CTAB DNA extraction protocol was determined to be optimal since it returned highest DNA concentrations, on average. While for the RNAlater the Power Soil kit performed slightly better, the small sample size of the test extractions meant that the difference seen for RNAlater between Power Soil and CTAB was unlikely to be statistically significant. The test extractions were spiked, and were therefore an imperfect replication of air samples, as their biomass was likely to be higher. Previous experience with DNA extractions indicated the CTAB extraction protocol performed better than a kit on average when biomass is very low, which is usual for air samples.

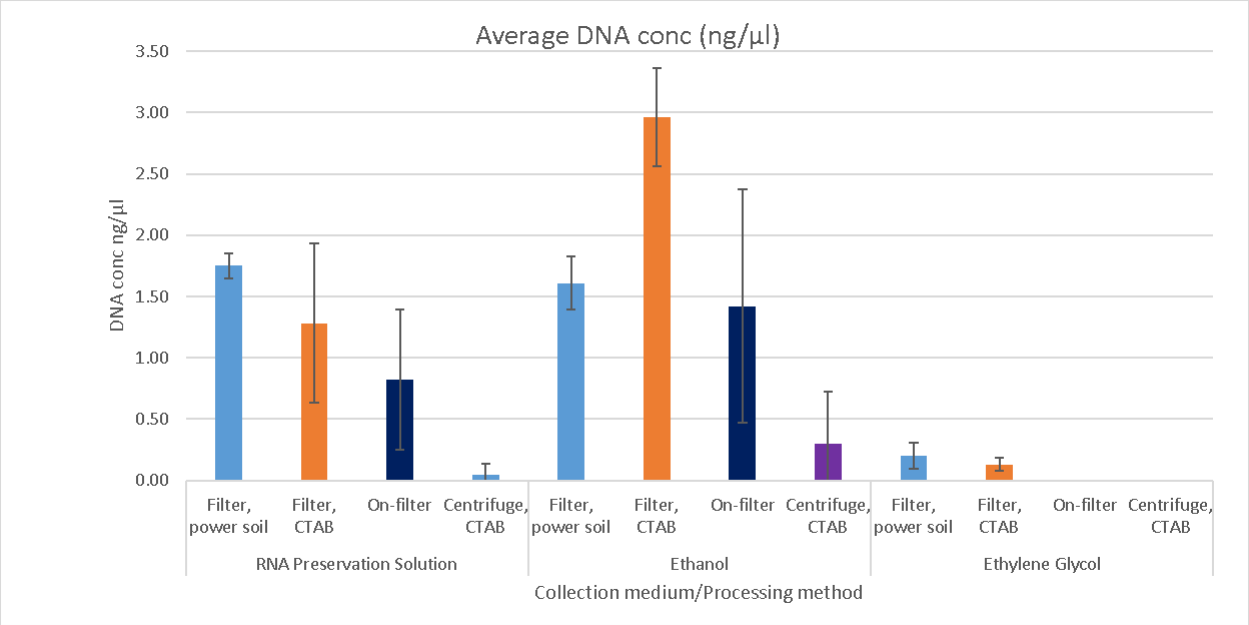


Figure A-2 - Comparison of the DNA concentration recovered from three of the collection mediums tested (RNAlater, Ethanol, Ethylene Glycol), with different combinations of sample pre-processing (filtration or centrifugation) and DNA extraction protocols (Power Soil kit, CTAB or on-filter extraction using CTAB). Glycerol is not included as it was found to be too viscous to function properly in the Coriolis so trials stopped at that point.

A.3.2 Sample Duration

Sampling for an hour reliably yielded quantifiable DNA and therefore this sample duration was selected (**Figure A-3**). This allows for fine temporal resolution while having a reasonable chance of returning acceptable microbial community data. A shorter sample duration increases flexibility, as samples can be combined later if needed to improve DNA concentrations, but samples cannot be separated if longer durations are used. These data also allowed for direct comparison between various sampling mediums, and PBS was found to yield approximately twice the DNA concentration of RNAlater. Dry collection yielded similar amounts of DNA to RNAlater, so that was selected as a back-up option for cold environments, given a liquid medium was preferred if possible to preserve Gram-negative bacteria and maintain normal operation of the cyclone and resulting size range of captured particles. Glycerol performed well in terms of DNA concentrations recovered but was not chosen as it became distributed around the Coriolis cone, head and neck and was difficult to remove effectively. For high temporal resolution sampling this means it would be difficult to separate one sample from the next and risk of contamination increases. PBS has a freezing point of 0 °C, so is well suited to sampling in NZ but not in very cold environments. The clear superiority of PBS as a collection medium led to interest in application of heat to the sample liquid, to allow PBS or a wide range of other liquids to be used in cold environments.

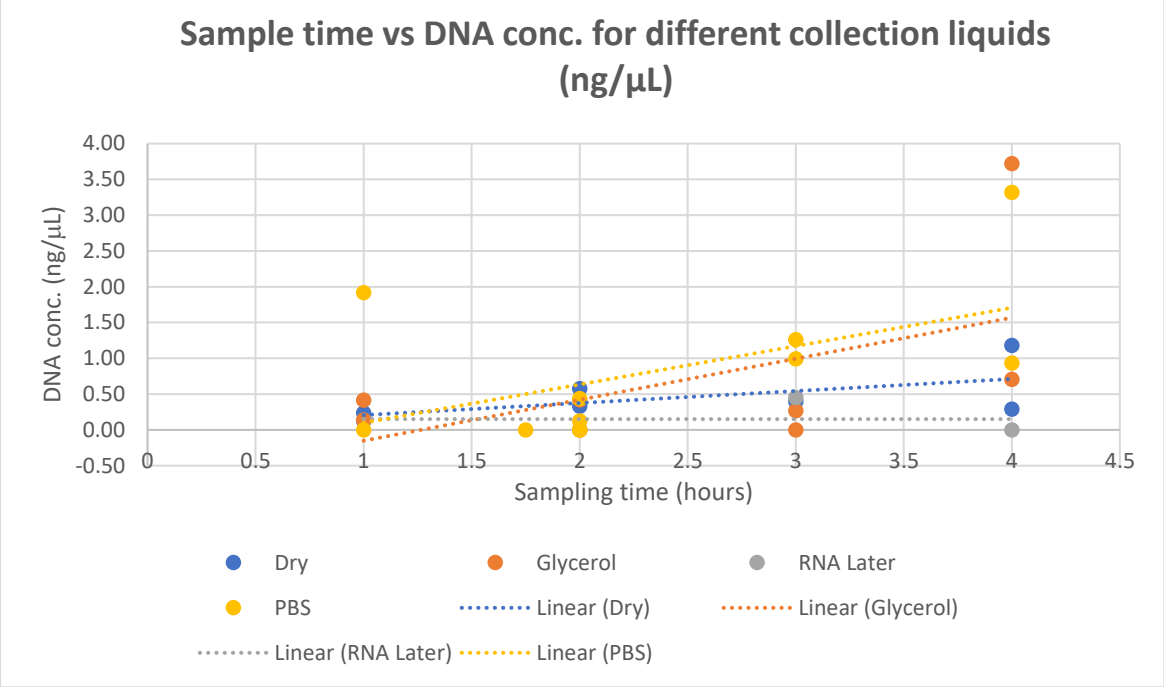


Figure A-3 - Results of trials of different sampling times averaged across various collection mediums (RNAlater, dry, glycerol and PBS) run in NZ temperate conditions. The relationship is linear as expected but sampling for twice as long does not appear to yield twice the average DNA concentration for all collection liquids. Sampling for one hour appears to reliably yield quantifiable DNA.

A.3.3 Development of Heated Coriolis Head and Cone Clip-Ons

Between the Antarctic field seasons of January 2017 and 2018 a heating solution for the Coriolis was developed. This consisted of replicating the sampling head exactly in aluminium, and machining aluminium clip-ons to wrap around the Coriolis cone (**Figure A-4**). The aluminium blocks had heating elements and temperature sensors embedded in them, attached to a control box with simple firmware to heat the blocks to the desired temperature and switch the heaters on or off as needed to maintain temperature.

Figure A-4 - The heater – showing the aluminium head (top left) and one of the two cone clip-ons (top right) with the collection cone inside. This is used with two identical clip-ons which surround the cone and heat the liquid inside.

The Coriolis unit and heater were field tested in Taylor Valley, Antarctica in January 2018 and allowed PBS to be used successfully to take samples. Wind chill was found to be a significant cooling factor which meant the heated head was only able to reach a temperature of 3 – 8 °C despite the heating elements being at maximum output. To sample at lower ambient temperatures further heating elements would need to be added to the head. When sampling was carried out overnight some freezing in the top up reservoir and piping occurred. This was only an issue for prolonged sampling, and could be addressed by extending the heater to apply heat to these additional areas, see section 4.3 Methods for full details of sampling protocols and results of Antarctic field sampling.

A.4 Conclusion

The method development work supports the methods used for this PhD. This was needed since aerobiology is a relatively new area of study without an extensive literature base to draw from. The toolkit now exists to sample in sub-zero conditions using both RNAlater and the heater with any liquid in the Coriolis. Both methods have been field proven. Air samples in the lowest biomass and most challenging environments can be taken in an hour or two versus days to weeks with reduced bias. This work represents a significant method advancement for aerobiology studies.

Appendix B. HYSPLIT Clustering Procedure

The back-trajectories generated by Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) were simplified through cluster analysis. Clustering created a mean trajectory which represented each cluster. A simple categorical variable for each mean trajectory was included in the db-RDA. Trajectories are assigned to clusters by minimising the differences between trajectories within a cluster, while maximising the differences between clusters (Stein et al., 2016). The clustering equations are as follows:

The spatial variance (SV) is computed between each endpoint (k) along trajectory (j) within its cluster (i):

$SV_{i,j} = \Sigma_k (P_{j,k} - M_{i,k})^2$

The sum is taken for all endpoints along the trajectory. P and M are position vectors for the individual trajectory and its cluster mean trajectory. The cluster spatial variance (CSV) is the sum of the spatial variance of all trajectories within the cluster:

$CSV_i = \Sigma_j SV_{i,j}$

The total spatial variance (TSV) is the sum of the CSV over all clusters:

$TSV = \Sigma_i CSV_{j,k}$ (NOAA, 2021)

HYSPLIT starts by assigning each trajectory its own cluster and then combines the two together which are most similar (computes TSV for all combinations of trajectories and chooses the lowest), so that the total clusters are now number of trajectories minus one. The clusters are then merged in this fashion until there is only one, or the specified cluster number is reached. The TSV rises initially, then levels through intermediate iterations, before rising at the end when combining dissimilar clusters (NOAA, 2021). There is subjectivity in deciding on the cluster number and HYSPLIT produces a plot of TSV against cluster number to aid in this judgment. The optimal number of clusters (lowest TSV) is indicated by the position of the final peak in TSV. In the example TSV plot below (**Figure B-1**) either six or four clusters appeared appropriate, as TSV increased at five and three. If there was any lack of clarity on review of the TSV plot, clustering would be performed with each cluster number and visually inspected. If two cluster means appeared very similar on the map, then a lower cluster number was selected. In **Figure B-1** clusters one and five looked very similar, as did clusters two and four. Therefore, four clusters were selected.

a) b)

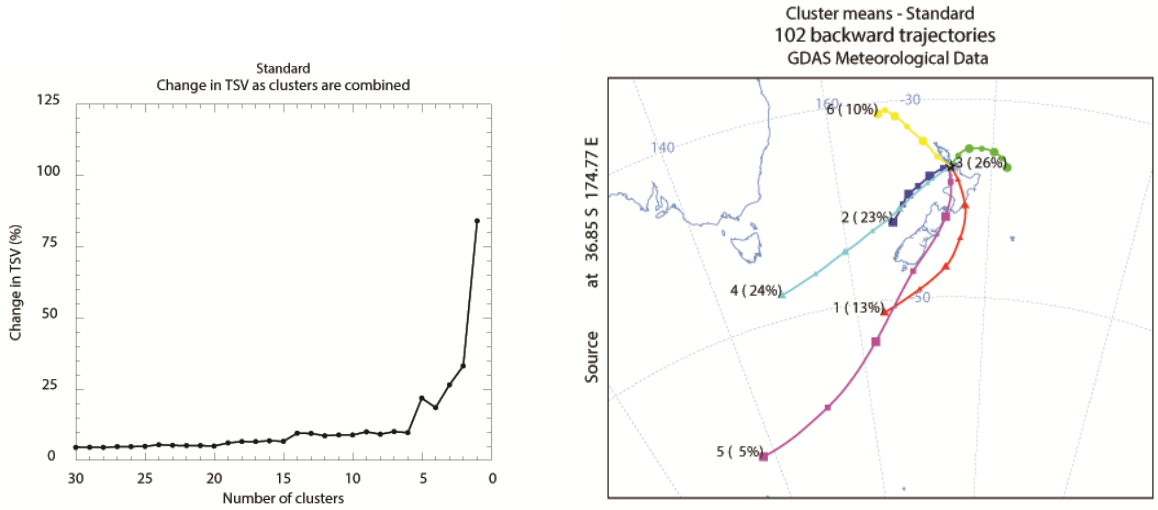


Figure B-1 - HYSPLIT outputs for 72-hour clusters in the Urban Parks study. a) TSV plot showing increasing % change in TSV as cluster number reduced b) Plot of cluster means when six clusters were selected. Percentages next to each cluster mean indicate the percentage of trajectories assigned to each cluster.

Appendix C. Laboratory Methods

C.1 Laboratory Methods – DNA Extraction

The aim of this PhD was to understand and compare bioaerosol community diversity and composition between samples taken across various temporal and spatial scales. DNA analysis of the samples was chosen as it has been shown to reveal a more complete picture of the bioaerosol community than culture or microscopy based studies (Yoo et al., 2016). The bacterial 16S rRNA gene (16S) was sequenced as it is a well conserved marker across the bacteria and archaea and the most commonly used approach in the literature (Pearce et al., 2016). The internal transcribed spacer region (ITS) was also analysed, so fungal sequences could be detected, per the literature (Archer et al., 2019). Samples were stored at -20 °C and defrosted overnight at 4 °C. The liquid was passed through a 0.2 μm GTTP polycarbonate filter using a disposable sterile syringe. The filter holder was cleaned in bleach, ethanol and milli-Q H_2O (MQH_2O) between each sample. The liquid was discarded and the filter processed using a CTAB based DNA extraction protocol. A positive extraction control (spiked with 100 μL *Escherichia coli* suspension in phosphate buffered saline or PBS) and a negative extraction control (no filter added) was used for each DNA extraction run. The CTAB extraction protocol entailed placing the filter in a 2 mL nucleospin bead tube, filled with 0.2 – 0.4 mL of Qiagen 1.4 mm ceramic beads. 270 μL of phosphate buffer (PBS) (100 mM NaH_2PO_4) and 270 μL of SDS lysis buffer (100 mM NaCl, 500 mM Tris pH 8.0, 10% SDS) was added to the tube. Tubes were vortexed at top speed for 15 seconds each, then shaken horizontally on a Vortex Genie 2 for 10 minutes. Samples were centrifuged at 13,200 rpm (19,627 rcf) for 3 minutes. 8 μL of BME was added per 1 mL of CTAB buffer and vortexed. 180 μl of the CTAB buffer and BME solution was added to the samples and incubated at 300 rpm at 60 °C for 30 minutes in a S1-300R Lab Companion shaker-incubator. The samples were centrifuged at 13,200 rpm for 1 minute to reduce bubbles. 350 μL chloroform: isoamyl alcohol (24:1) was added, the samples were vortexed for 15 seconds then centrifuged at 13,200 rpm for 5 minutes. The upper aqueous layer was transferred into a new 1.5 mL sterile Eppendorf tube. 500 μL chloroform: isoamyl alcohol (24:1) was added and the samples were vortexed for 10 seconds, then left on a rocking bed (Life Technologies HulaMixer) for 20 minutes at room temperature. The samples were centrifuged at 13,200 rpm for 5 minutes and the upper aqueous layers moved into a new 1.5 mL sterile Eppendorf tube. Ten M ammonium acetate to a final conc. of 2.5 M (an amount equal to 1/3 of tube volume) was added. The samples were mixed gently by repeated inversion (25 times) and centrifuged at 13,200 rpm for 5 minutes. The upper layer was transferred to a new sterile Eppendorf tube and 0.5 times the tube volume of isopropyl alcohol was added. The samples were mixed by repeated inversion (20 times) then incubated at -80 °C for 48 hours. The samples were centrifuged at 14,000 rpm (20,817 rcf) for 20 minutes at 4 °C and the supernatant discarded, leaving a pellet of DNA. The pellets were washed with 1 mL 70% ethanol (at -20 °C) and centrifuged at 14,000 rpm at 4 °C for 5 minutes. 70% ethanol was gently pipetted off (firstly with a 1000 μL pipette, then recentrifuged at 14,000 rpm at 4 °C for 5 minutes, then a 20 μL pipette used to remove remaining liquid). The pellets were dried in an Eppendorf concentrator plus for 8 - 12

minutes. DNA was re-suspended in 20 µL ultra-pure H_2O (Invitrogen Ultra-Pure Distilled Water – DNAase, RNAase, Free) by pipetting up and down 25 times and scraping the pipette tip on the side of the tubes. The samples were heated at 55 °C for 10 minutes, vortexed for 10 seconds and stored at -20 °C. The DNA was quantified using Qubit (as per the manufacturer's instructions, with 2 µl of the sample added to the assay tubes (Invitrogen, 2010) and test PCRs were performed for two samples (using the same protocol as described for the 16S PCR in section C.3 Laboratory Methods – DNA Sequencing below) in each batch to determine if further processing was needed before sequencing. All samples in batches that failed to amplify were purified with AMPure XP beads using the standard protocol (in section C.2 AMPure XP Beads Protocol below) and PCR reattempted to ensure they amplified. Sample filtration and DNA extraction were performed in a Biosafety cabinet (Gelman Sciences BioHazard-Protection class II) where possible.

C.2 AMPure XP Beads Protocol

Beads were removed from the fridge 30 minutes before use to allow them to heat up to room temperature. The beads were vortexed for one minute until homogenous. For clean-up of DNA samples before PCR a ratio of 1.8 mL of beads to 1 mL of sample was used (a high ratio to capture all DNA). For clean-up of DNA samples after PCR a ratio of 0.8 mL of beads to 1 mL of sample was used (a lower ratio to capture longer pieces of DNA only). The beads were added to the DNA samples and mixed by pipetting up and down 10 times. The mixture was incubated at room temperature for five minutes. The 96 well tray was placed on a magnetic stand and left for three – five minutes until the solution had cleared. With the tray still on the magnetic stand the liquid was carefully pipetted off and discarded. The tray remained in the stand and the bead pellet was washed twice with 80% ethanol (enough volume was added to cover the beads). Each time the ethanol was left for one minute then pipetted off. The tray was then left for a few minutes for the remaining ethanol to evaporate. When the bead pellets were dry the 96 well tray was moved off the magnetic rack and 20 (pre-PCR) or 30 µL (post-PCR) of nuclease free water (Ultra-Pure H_2O) was added. The water and beads were mixed by pipetting up and down 10 times and left to incubate for three minutes. The tray was moved back on to the magnetic rack and left for three – five minutes for the solution to clear. 25 µL of the liquid was taken off post-PCR and as much as possible (with a 0.1 µL tip) was removed from the pre-PCR samples, up to around 20 µL.

C.3 Laboratory Methods – DNA Sequencing

The bacterial and fungal community structure was determined using MiSeq DNA sequencing, the following method was adapted from Maki et al. (2017). Fragments of 16S rDNA (approximately 460 bp) were amplified from the extracted gDNA by PCR using the universal 16S rDNA bacterial primers 314F and 785R (IDT Forward 5'- *TCG-TCG-GCA-GCG-TCA-GAT-GTG-TAT-AAG-AGA-CAG*-CCT-ACG-GGN-GGC-WGC-AG IDT Reverse 16S 5'- *G-TCT-CGT-GGG-CTC-GGA-GAT-GTG-TAT-AAG-AGA-CAG*-GAC-TAC-HVG-GGT-ATC-TAA-TCC) and the ITS fungal primers (ITS 1 Forward, 5'-*TCG-TCG-GCA-GCG-TCA-GAT-GTG-TAT-AAG-AGA-CAG*-

CTT-GGT-CAT-TTA-GAG-GAA-GTA-A ITS2, Reverse 5'-*G-TCT-CGT-GGG-CTC-GGA-GAT-GTG-TAT-AAG-AGA-CAG*-GCT-GCG-TTC-TTC-ATC-GAT-GC). The section of the primer in italics is the Nextera adapter region, for binding of the primers for the indexing PCR. The remainder of the primer is for binding to the target area in the genome for amplification. The PCR amplicon sequences covered the variable regions V3 and V4 of the 16S rRNA gene and the fungal ITS1 region between the 18S and 5.8S rRNA genes. Thermal cycling on an Eppendorf vapo protect was performed under the following conditions for both 16S and ITS: initial denaturation at 95 °C for three minutes then 35 cycles of 95 °C for 30 seconds, annealing at 55 °C for 30 seconds, and extension at 72 °C for 30 seconds. Finally, the sample was held at 72 °C for five minutes then the temperature was reduced to 4 °C. KAPA Hi-Fi Hot Start ReadyMix (KAPA) of 12.5 µL was used, with 5 µL of each primer (at 1 mM concentration) and 5 µl of DNA in a 27.5 µL reaction. A total amount of 20 ng DNA was targeted in the PCR reaction. DNA was diluted if needed, therefore input concentration was variable and was often so low as to be unquantifiable, due to the minimal biomass in aerosol samples. Small DNA fragments were removed with AMPure XP beads (per protocol in section C.2 AMPure XP Beads Protocol) then PCR products were quantified using Qubit , with 2 µl of the sample added to the assay tubes (Invitrogen, 2010). Dilution was performed to standardise DNA concentrations of each sample (16S at 5 ng/µL and ITS at 1 ng/µL) before samples were indexed to allow identification after sequencing. 16S and ITS rDNA in PCR products were amplified again using the indexing PCR primer pair, sequences binding to the regions in italics above with the addition of a unique 8 nucleotide barcode. The samples for this PhD were included on two sequencing runs, each position in the 96 well plates were coded by a unique pair of Nextera indexing primers that were consistent for 16S and ITS. The Antarctic and Baring Head samples were sequenced on the first run, and Urban Parks and the Seasonal study on the second run. Thermo-cycling on the GeneAmp PCR System 9700 was performed under the following conditions: initial denaturation at 95 °C for three minutes, then (8 cycles for 16S and 12 cycles for ITS) of 95 °C for 30 seconds, annealing at 55 °C for 30 seconds, and extension at 72 °C for 30 seconds. Finally, the samples were held at 72 °C for five minutes then at 4 °C. For the indexing PCR 12.5 µL KAPA was used, 2.5 µL for each indexing primer, for ITS 2.5 µL of ultra-pure H_2O was added and 5 µL of sample and for 16S, 5 µL of ultra-pure H_2O and 2.5 µL of sample was used in a reaction totalling 25 µL. PCR amplicons from each sample were pooled, mixed, sub-sampled, purified with AMPure XP beads (as per section C.2 AMPure XP Beads Protocol), quantified (2 µl of the sample added to the assay tubes (Invitrogen, 2010)), tested with a Bioanalyzer (to check for presence/absence of primers and correct library size) and included at approximately equal amounts into a single sequencing run on a MiSeq Genome Sequencer (Illumina, MiSeq CA, USA) machine. The sequences obtained for each sample were demultiplexed based on the 8-nucleotide barcode in the indexing primers. Negative controls were sequenced so contamination could be identified and corrected for during data analysis.

Appendix D. Optimising Bioinformatics Protocols for Aerosol Microbial Community Data – a Case Study Using an Urban Parks Dataset

A modified form of this appendix has been accepted for publication with revisions by PeerJ as at 5 July 2021.

D.1 Abstract

Microbes are fundamental to Earth's ecosystems, thus understanding ecosystem connectivity through microbial dispersal is key to predicting future ecosystem changes in a warming world. However, aerial microbial dispersal remains poorly understood. Few studies have been performed on bioaerosols (microorganisms and biological fragments suspended in the atmosphere), despite them harbouring pathogens and allergens. Most environmental microbes grow poorly in culture, therefore molecular approaches are required to characterize aerial diversity. Bioinformatic tools are needed for processing the next generation sequencing (NGS) data generated from these molecular approaches; however, there are numerous options and choices in the process. These choices can markedly affect key aspects of the data output including relative abundances, diversity, and taxonomy. Bioaerosol samples have relatively little DNA, and often contain novel and proportionally high levels of contaminant organisms, that are difficult to identify. Therefore, bioinformatics choices are of crucial importance. A bioaerosol dataset for bacteria and fungi based on the 16S rRNA gene (16S) and internal transcribed spacer (ITS) DNA sequencing from parks in the metropolitan area of Auckland, Aotearoa New Zealand was used to develop a process for determining the bioinformatics pipeline that would maximize the data amount and quality generated. Two popular tools (Dada2 and USEARCH) were compared for amplicon sequence variant (ASV) inference and generation of an ASV table. A scorecard was created and used to assess multiple outputs and make systematic choices about the most suitable option. The read number and ASVs were assessed, alpha diversity was calculated (Hill numbers), beta diversity (Bray-Curtis distances), differential abundance by site and consistency of ASVs were considered. USEARCH was selected, due to higher consistency in ASVs identified and greater read counts. Taxonomic assignment is highly dependent on the taxonomic database used. Two popular taxonomy databases were compared in terms of number and confidence of assignments, and a combined approach developed that uses information in both databases to maximize the number and confidence of taxonomic assignments. This approach increased the assignment rate by 12 – 15%, depending on amplicon and the overall assignment was 77% for bacteria and 47% for fungi. Assessment of decontamination using "decontam" and "microDecon" was performed, based on review of ASVs identified as contaminants by each and consideration of the probability of them being legitimate members of the bioaerosol community. For this example, "microDecon's" subtraction approach for removing background contamination was selected. This study demonstrates a systematic approach to determining the optimal bioinformatics pipeline using a multi-criteria scorecard for microbial bioaerosol data.

D.2 Introduction

Next Generation Sequencing (NGS) of microbial marker genes has revolutionized microbiome studies (Pearce et al., 2016). Modern techniques are extremely sensitive, and can detect even one copy of a target gene (McKnight et al., 2019). Most environmental microorganisms grow poorly in culture and can be present in very small numbers (Burrows et al., 2009b). NGS circumvents these issues, as different genetic variants present can be inferred, and their taxonomy predicted. The microbiome of the aerosphere is challenging to decode, even with these new approaches. It is a very low biomass environment, and many

common contaminants (microbial taxa that are not true constituents of the environmental population) are naturally present in air. Indeed, these microbes are probably contaminants themselves due to their ease of atmospheric dispersal. Study of the aero-microbiome is relatively recent, and unknown taxa are frequently detected (Bottos et al., 2014). The low biomass and many ubiquitous taxa present make identification and removal of contamination difficult. While biomass can be improved with greater sample durations (Pearce et al., 2009), DNA degradation and practical considerations can preclude prolonged sampling campaigns (Luhung et al., 2015). Furthermore, it is challenging to tell which taxa represent true variants and to assign taxonomy to them with existing databases. Due to these difficulties, single amplicon sequencing (commonly 16S, 18S ribosomal RNA or ITS) has been used for the majority of NGS bioaerosol studies to date (Archer et al., 2020; Barberán et al., 2015; Bottos et al., 2014; Smith et al., 2018; Tanaka et al., 2020; Woo et al., 2013). However, bioaerosol metagenomic and transcriptomic techniques have been developed (Jiang et al., 2015) and successfully applied in a handful of very recent studies. (Amato et al., 2019; Amato et al., 2017; Jaing et al., 2020). Jaing et al. (2020) performed metagenomic sequencing on bioaerosols above the Sierra Nevada mountains in the US and obtained 5,000,000 reads, although only 640,000 were successfully assigned at the genus level. Transcriptomics on microbial communities in clouds has been performed for rRNA genes (Amato et al., 2017) and untargeted amplification has been applied to full cloud metagenomes and transcriptomes to compensate for low biomass (Amato et al., 2019). These omics approaches represent the cutting edge of NGS based bioaerosol research and this study focused on a process for method optimization for the frequently used 16S and ITS amplicons.

The information produced by single amplicon NGS and conclusions drawn are sensitive to choices made when performing bioinformatic processing (Edgar, 2013, 2017). A constantly evolving multitude of environments and tools exist to process 16S or ITS NGS data (Callahan, 2020; Edgar, 2010, 2013). A typical dataset contains millions of reads, and thousands of unique sequences (Amplicon Sequence Variants or ASVs). Sequencing errors and artefacts, even if they occur at low rates, can become significant over large datasets (Edgar, 2013). There is total reliance on automated processing and algorithms which make assumptions and introduce bias (Callahan et al., 2016). It is very difficult to objectively verify the results and determine the “true” aero-microbiome. If the environment is well understood, mock community sequencing can be performed alongside experimental samples (Hermans, Buckley, & Lear, 2018). For less well understood environments like the aerosphere, more surety can be gained where results from multiple bioinformatics approaches converge. Choosing amongst the plethora of data processing options at each stage can become challenging. Especially in somewhat specialized fields, with unusual data characteristics, such as the study of bioaerosols, standard pipelines or assumptions about data can be hazardous. For instance, most decontamination pipelines assume relativity low levels of contaminating DNA compared to target DNA, which is not necessarily the case for the aero-microbiome (Pearce et al., 2016). Further, when tools are constantly being updated and “best practice” in an area can be outdated or undefined, method selection is not necessarily straightforward. An approach to select an optimal pipeline for specialized

datasets, using defined selection criteria, is critically useful in the challenge to extract signal out of noise in NGS data.

D.2.1 Bioinformatic data processing

Processing of 16S and ITS NGS data is a complex, multi-stage procedure and is further complicated by the necessary choices among many alternative tools available (Edgar, 2017) (**Figure D-1**). With so many modifications being made to the data, it is crucial to ensure each step is improving quality rather than introducing bias or errors. The key stages of the process are summarized in Figure 1 and begin with sampling, DNA extraction, amplification of target genes and sequencing of PCR products. The sequence data is processed by separating amplicons using their respective primer sequences, trimming to remove primers, merging forward and reverse reads, quality filtering and denoising (ASV inference). Chimeras and low abundance sequences are removed. An ASV table is constructed and taxonomy is predicted with reference to sequence databases (Callahan et al., 2016; Edgar, 2017). Finally, contaminants should be filtered out of the dataset (Davis, Proctor, Holmes, Relman, & Callahan, 2018; McKnight et al., 2019).

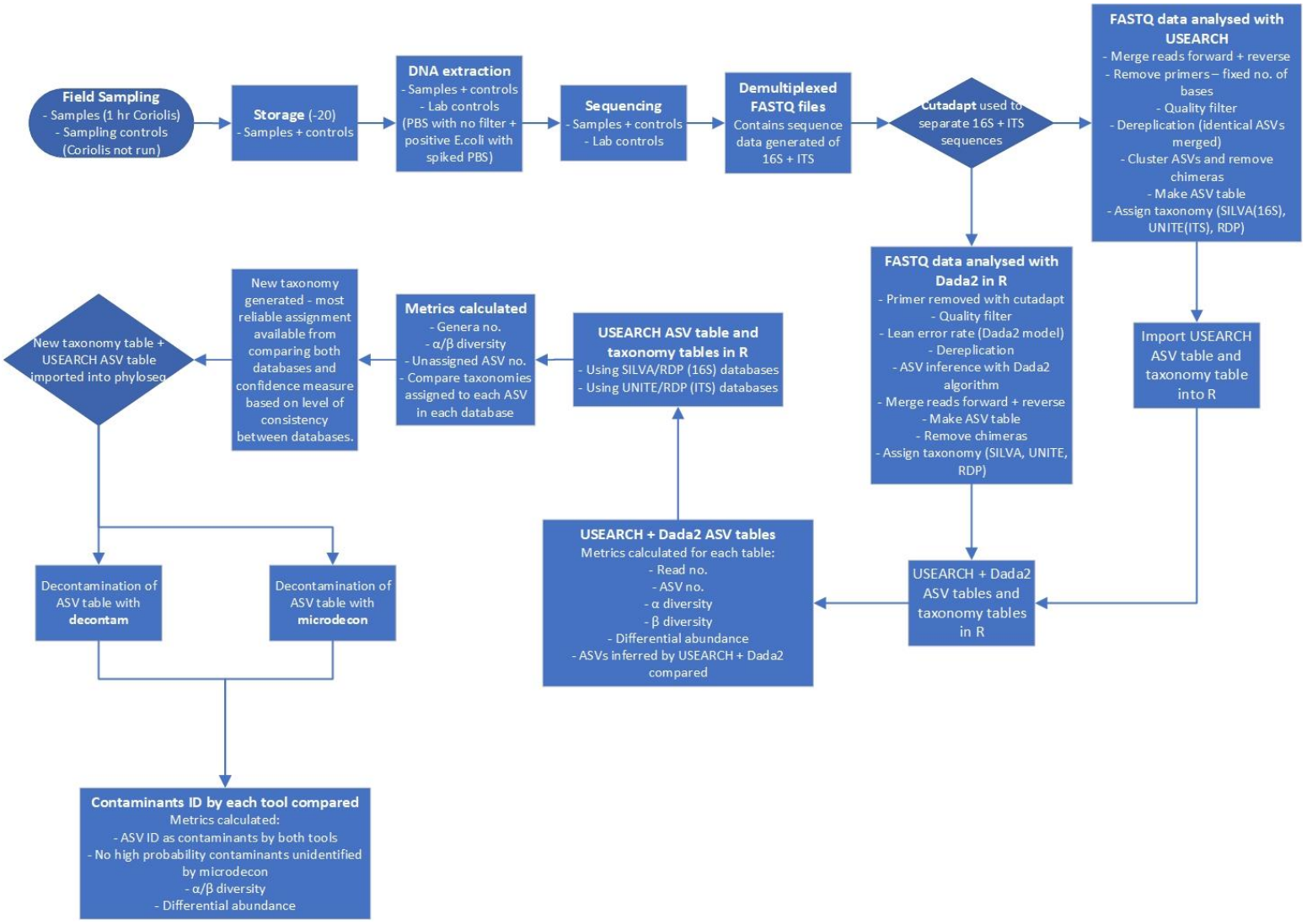


Figure D-1 - Process flow for sampling and data analysis of microbial aerosol communities at urban parks. Hr is hour. PBS is phosphate buffered saline. E. coli is *Escherichia coli*. No. is number. ID is identified.

Inferring the presence of the true sequence variants requires care. Approximately 0.5% (Mardis, 2013) to 0.24% (Pfeiffer et al., 2018) of base calls during Illumina sequencing are incorrect. This means that for the V3 to V4 region of the prokaryotic 16S gene, which is around 450 nucleotides long, each sequence could be expected to have two wrong nucleotides. As there is only one correct sequence but many incorrect versions (as base call errors are random (Edgar, 2013)) the correct variant is present at high abundance in sequence data, with many incorrect versions of the same sequence at low abundance. Most processing pipelines attempt to filter out low abundance sequences to try to remove these spurious ASVs, but filtering risks removing true low abundance organisms as well. Other sequencing artefacts, like chimeras and cross-talk (incorrect sample bar codes causing sequences to be assigned to the wrong sample) can also introduce error (Callahan et al., 2016). Comparative studies show that some tools produce many more spurious variants than others (Edgar, 2017). Two of the most popular tools currently used in bioaerosol research (Archer et al., 2020; Archer et al., 2019) are "Dada2" in R (Callahan et al., 2016) and USEARCH (Edgar, 2010). The use of ASVs has been recommended for bioaerosol studies, as this helps increase the resolution of the data produced (Archer et al., 2019). Operational taxonomic units (or OTUs, normally matched at 97% sequence identity) reduce the information quality compared to ASVs but lower the chance of spurious variants being detected, as low abundance incorrect variants are merged into higher abundance OTUs (Callahan et al., 2016).

The taxonomy database selected can significantly impact the number, nature and confidence of taxonomic assignments of ASVs that can be achieved. There is a trade-off between the size and the accuracy of the databases available. Larger ones, such as SILVA (Quast et al., 2012) and Greengenes (Balvočiūtė & Huson, 2017) for 16S rRNA gene, contain many uncultivated environmental sequences with taxonomy algorithmically predicted. There are thousands of conflicts in assignment of identical sequences between SILVA and Greengenes (Edgar, 2018). It is unclear which one is right, but at least one of them must be wrong. As a result, 17% of the taxonomy annotations in these databases are estimated to be incorrect (Edgar, 2018). The smaller Ribosomal Database Project training set database (RDP) is considered to be more reliable as most taxonomy is assigned based on authoritative examination of isolate strains (Edgar, 2018). However, using a smaller database, especially with novel taxa, is more likely to result in many unassigned ASVs. In addition, this does not improve the situation for unculturable organisms.

Removal of contaminant microbial taxa is critical. In addition to the challenges presented by the aero-microbiome, limited decontamination tools are available. Contaminants (especially bacteria) are ubiquitous in many reagents (McKnight et al., 2019). Clean sampling, laboratory protocols and thorough use of negative controls help. However, reads still exist in negative controls which need to be addressed. There are two principal approaches to deal with the contamination bioinformatically. Identification and removal of entire ASVs that appear to be contaminants, for example, with the R package, "decontam" (Davis et al., 2018). Contaminating ASVs are identified in "decontam" by their higher abundance in negative controls compared to samples (prevalence method) and by their concentrations inversely correlating with sample DNA concentration (frequency method)

(Davis et al., 2018). Alternatively, negative controls can be used to calculate a background contamination profile, which can then be deducted from each sample, in this case with the R package "microDecon" (McKnight et al., 2019).

The aim of this protocol development study was to define a process for optimizing the bioinformatics approach for future bioaerosol studies. To achieve this, a dataset from microbial aerosol communities in urban parks was used to systematically investigate the effects of different bioinformatics choices on the amplicon ASV table and taxonomic assignments of organisms present. An optimal approach based on this information was developed. R code for the bioinformatics pipelines investigated is presented.

D.3 Methods

D.3.1 Field Sampling

Ten parks in urban Auckland, and a reference rural location north of Auckland, New Zealand were selected for sampling (**Figure 2-1**). The vegetation at these parks was primarily cultivated lawns, with variable cover of deciduous and coniferous trees, shrubs and ferns. Some parks had bodies of water in them, such as Western Springs. The parks differed in occurrences of pedestrians, rates of vehicular movements and the abundance of birds and dogs. Livestock, such as sheep and cattle, were present at some. The parks differed markedly in area and altitude, as several encompassed volcanic cones. Each park was visited three times in the first sampling window (12 July 2017 to 10 August 2017) and three times in the second sampling window (20 of March 2018 to 1 June 2018). On different days during the sampling window, a single sampling location within each park was visited in a randomized order between 10 am and 4 pm and was sampled for 1 hour with the Coriolis µ liquid cyclone impinger (Haig et al., 2016) at 300L/m into phosphate buffered saline (PBS) at 1.8 m. The total volume of air in each sample was approximately 18m^3, which sits comfortably in the range of volumes (2.7 – 144m^3) used in similar bioaerosol studies (Amato et al., 2017). The sample was taken in the same location each time, as close to the center of the park as practicable and at minimal elevation (avoiding effects from change in altitude at volcanic peaks or being close to an edge confounding results). The exact location was saved on Google Maps (Google; Stein et al., 2016) and a photo was taken of the Coriolis aerosol sampling unit and location. During sampling, gloves were worn and the extender kit and Coriolis neck, head and cone (**Figure A-1**) were cleaned with bleach, ethanol and three rinses of MilliQ water (MQH_2O). The cone was filled with 15 mL of PBS. A negative control was taken (PBS put into the cone without running the Coriolis) referred to as a "sampling control". The Coriolis was run for 2 – 4 minutes with MQH_2O. The MQH_2O was discarded and replaced with 15 mL of PBS and the Coriolis was run at 300 L/m for 1 hour. The PBS in the cone was topped up to 15 mL after 30 minutes and sampling was completed with 10 mL of PBS remaining in the cone. Samples were transferred into a 15 mL falcon tube, transported in an insulated box with ice blocks and stored at -20 °C within 4 hours. During sampling, observations of weather conditions were recorded (with a Kestrel 3000) and particle counts were taken (with an AeroTrak particle counter). If light rain occurred during sampling, the Coriolis was sheltered with an umbrella. If rain was heavy, the unit was

packed up until the rain abated. While precipitation can affect bioaerosol communities (Reche et al., 2018), only 12 of the 66 samples in the example dataset were affected by precipitation, and therefore the data was considered representative for the purposes of defining a process for optimizing the bioinformatics approach for bioaerosol NGS data.

D.3.2 Laboratory Processing

For details on DNA extraction and DNA sequencing methods see section C.1 Laboratory Methods – DNA Extraction, and section C.3 Laboratory Methods – DNA Sequencing.

D.3.3 Data Analysis

Data analysis was performed in R 3.6.3 x86_64 (R Development Core Team, 2010) and USEARCH (11.0.667_i86linux32 and 9.0.2132_i86linux64) (Edgar, 2010). Details of all code and packages used are in the supplementary materials. Demultiplexed FastQ files were generated. Cutadapt 2.6 (Martin, 2011) with python 3.6.9 (Van Rossum, 2009) was used to separate the two amplicons based on their primer sequences. 16S and ITS were analyzed with "Dada2 1.14.1" (Callahan et al., 2016) in R and USEARCH (Edgar, 2010) with recommended workflows (Callahan; Edgar). USEARCH outputs were imported into R using "RDPutils 1.4.1" (Quensen, 2018). Metrics were calculated for comparison between pipelines. These included numbers of reads, numbers of ASVs, alpha diversity, specifically Hill numbers, (Chao et al., 2016) in "phyloseq 1.30.0" (McMurdie & Holmes, 2013), beta diversity (Bray-Curtis distance or BC) in "vegan 2.5.6" (Dixon, 2003) and differential abundance using "DESeq2 1.26.0" (Love, Huber, & Anders, 2014).

ASVs inferred from USEARCH and Dada2 were compared. ASVs identified by USEARCH were used to compare different taxonomy databases. Consistent and conflicting taxonomic assignments were considered to identify the most appropriate strategy for the data. For 16S, RDP 16 (Balvočiūtė & Huson, 2017) and SILVA 132 (Quast et al., 2012) were used. For ITS, RDP 2 (Balvočiūtė & Huson, 2017) and UNITE UTAX02.02.2019 (Abarenkov et al., 2010) were used. Metrics were calculated for comparison of databases. The number of genera identified, alpha and beta diversity as above, and number of unassigned ASVs were determined. Statistical significance of differences in these metrics between pipelines was tested in base R using a non-parametric Mann-Whitney U test. The taxonomic assignment of each ASV with both databases was compared. A taxonomy file was created, which used classification from both databases and included a confidence measure. "High confidence" (assignments where both databases agreed at the genus level). "Medium confidence" (different assignments from each database). "Low confidence" (assignments only in one database). The RDP database annotation was used in preference for both ITS and 16S. If no RDP classification at the genus level was available, the alternative database was used.

The taxonomy table created was imported into "phyloseq" and subsequently processed with "decontam 1.6.0" (Davis et al., 2018). For comparison, the same dataset was also run with "microDecon 1.0.2" (McKnight et al., 2019) on default settings. The three main methods to identify contaminants available in "decontam" are: frequency, prevalence and combination (Davis et al., 2018). The combination method in "decontam" was chosen for comparison to "microDecon" as it considers the most data (both the frequency and

prevalence methods combined) (Davis et al., 2018). The ASV table, after removal of contaminants, was compared to the contaminant ASVs identified by “decontam” to check consistency and identify if any ASVs highly likely to be contaminants remained. Metrics were calculated from both datasets: ASVs identified as contaminants by both tools, the number of high probability contaminants missed by “microDecon”, alpha and beta diversity and differential abundance, as above.

D.4 Results

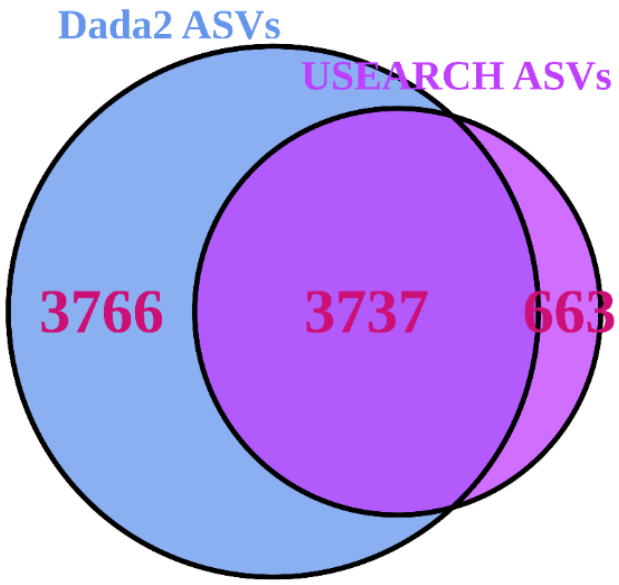
D.4.1 ASV Inference

USEARCH detected 25% to 33% (significantly for 16S) more reads than “Dada2”(**Table D-1**) . USEARCH was expected to return higher read counts since it matches the pre-filtered reads to ASVs if the read quality is sufficient (Edgar, 2013) while “Dada2” matches reads post-filtering (Callahan et al., 2016). More reads also meant that more samples passed through all stages in USEARCH, while in the “Dada2” pipeline several samples failed to complete the pipeline. “Dada2” consistently identified significantly many more ASVs than USEARCH, but USEARCH detected more ASVs per sample, and returned generally higher alpha and beta diversity metrics as a result. Observed alpha diversity was significantly higher for USEARCH and exponential Shannon and inverse Simpson were generally greater for USEARCH also. USEARCH data, once processed with “DESeq2”, showed slightly more ASVs were significantly differentially abundant between sample locations. Significantly differentially abundant ASVs by location were significantly higher for USEARCH. On consideration of consistency between approaches, for 16S, 50% of the “Dada2” ASVs were also present in the USEARCH data (**Figure D-2**). Of the ASVs inferred by USEARCH, 85% were supported by “Dada2”. A similar pattern was apparent with ITS, with 56% of “Dada2” ASVs supported by USEARCH and 76% of USEARCH ASVs supported by “Dada2”.

Table D-1 - Summary of metrics computed for datasets derived from the USEARCH and "Dada2" pipelines for ITS and 16S amplicons for microbial aerosol communities at urban parks. SD is the standard deviation. Observed (D0) is the raw diversity or number of unique ASVs detected. ExpShannon (D1) refers to the exponential of the Shannon diversity index and invSimpson (D2) refers to the inverse of the Simpson diversity index. No. is number, sig. is significantly. Ranges are not appropriate for statistical testing so an N/A appears instead of a *P* value for them. *P* values that are significant at the 0.05 threshold appear in bold.

		16S			ITS		
Metrics		16S Dada2	16S USEARCH	*P* value 16S Dada2 v USEARCH	ITS Dada2	ITS USEARCH	*P* value ITS Dada2 v USEARCH
Raw abundance	*Total reads*	507,586	754,752	**0.03**	1,687,012	2,222,530	0.10
	Mean (SD) reads per sample	9,399 (7,808)	10,938 (9,381)	0.37 (0.3)	25,179 (26,742)	33,172 (35,256)	0.09 (0.07)
Raw diversity	*Total ASVs*	7,503	4,400	**0.01**	7,507	5,519	**0.02**
Samples with no reads completing pipeline		3	0	0.08	1	0	0.36
Alpha diversity (Hill numbers)	*Observed (D0)*	151	315	**0.00**	255	441	**0.01**
	Exp Shannon (D1)	16.46	23.8	0.19	1 8. 88	29.48	0.10
	Inv Simpson (D2)	17.25	16.66	0.90	14.8	17.15	0.61
Beta diversity (Bray-Curtis)	*Mean*	0.84	0.87	0.37	0.9	0.9	0.70
	Median	0.89	0.91	0.95	0.94	0.93	0.57
	Range	0.12-1	0.13-1	N/A	0.17-1	0.22-1	N/A
	Standard deviation	0.16	0.15	0.40	0.12	0.11	0.90
No. ASVs sig. differentially abundant (by location)		174	180	**0.01**	41	51	0.25

a)



b)

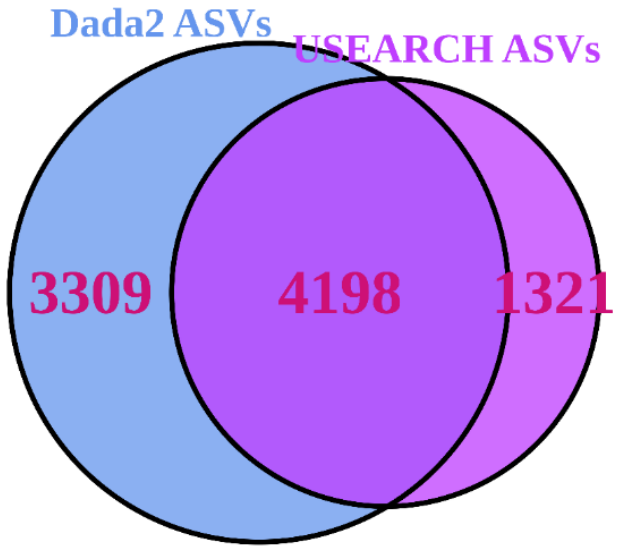
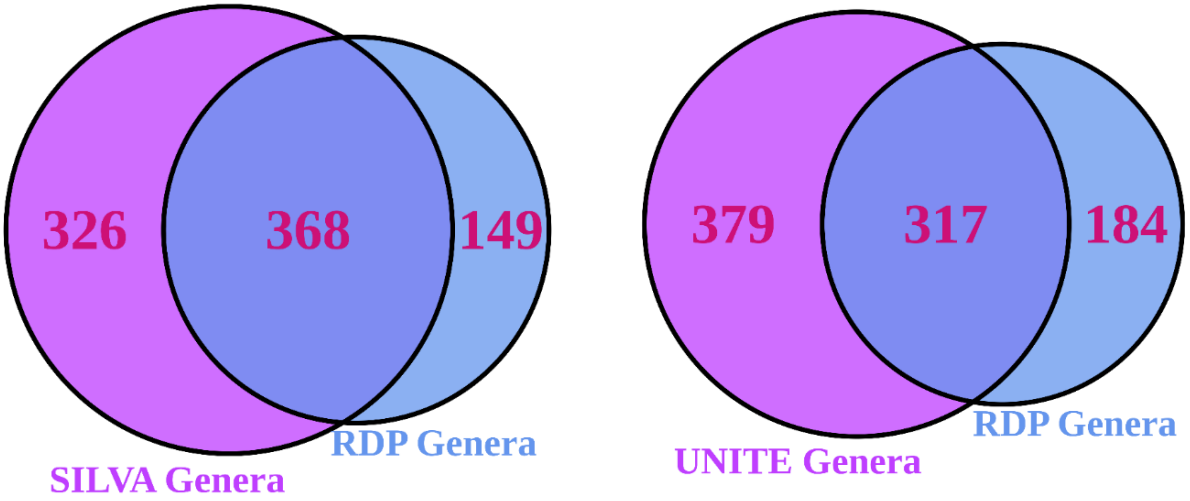


Figure D-2 - Venn diagrams showing the number of ASVs inferred from USEARCH and Dada2 for (a) 16S and (b) ITS. ASVs identified by both pipelines are very likely to be true variants. ASVs identified by both pipelines are very likely to be true variants.

More genera were identified with larger databases (UNITE for ITS and significantly more with SILVA for 16S), as expected (**Table D-2**) (Edgar, 2018). 16S showed greater alpha and beta diversities with a larger database. ITS showed an inconsistent pattern of alpha and beta diversities with database size (most differences were statistically insignificant other than exponential Shannon and inverse Simpson for ITS, where UNITE was less diverse). Larger databases showed significantly fewer unassigned ASVs; RDP 16S had 56% ASVs unassigned, RDP ITS had 71% ASVs unassigned compared to SILVA with 38% unassigned ASVs and UNITE with 64% unassigned ASVs. ITS suffered more overall from unassigned ASVs than 16S. For both amplicons, roughly half of the genera identified in the dataset matched (**Figure D-3**). When assignments for individual ASVs were reviewed, only 5% of 16S taxonomic assignments were consistent between RDP and SILVA at the genus level. However, for the top 100 ASVs by read count, 58% of the assignments were consistent at the individual ASV level and therefore, high confidence. For ITS, 31% of assignments were consistent between RDP and UNITE, and high confidence. When both databases were employed together to provide taxonomic information, unassigned ASVs were substantially lower for both amplicons than the largest individual database (23% for 16S and 53% for ITS ASVs remined unassigned).

Table D-2 - Comparison of taxonomy assignments of 16S and ITS USEARCH ASVs for urban parks bioaerosol communities with the RDP and SILVA (16S)/UNITE(ITS) taxonomy databases, and with a combined database approach. Observed (D0) is the raw diversity or number of unique ASVs detected. ExpShannon (D1) refers to the exponential of the Shannon diversity index and invSimpson (D2) refers to the inverse of the Simpson diversity index. Ranges are not appropriate for statistical testing so an N/A appears instead of a *P* value for them. *P* values that are significant at the 0.05 threshold appear in bold.

Taxonomy database used		RDP training set (16S)	SILVA (16S)	*P*-value between 16S databases	RDP training set (ITS)	UNITE (ITS)	*P*-value Between ITS databases
Metrics							
Total genera identified by pipeline		517	694	**0.02**	501	696	0.06
Alpha diversity of genera (Hill numbers)	*Observed (D0)*	112	127	0.30	121	123	0.75
	expShannon (D1)	11.31	12.56	0.52	14	8.56	**0.02**
	invSimpson (D2)	7.99	8.98	0.56	8.71	5.3	**0.00**
Beta diversity of genera (Bray-Curtis)	*Mean*	0.78	0.8	0.56	0.84	0.82	0.61
	Median	0.8	0.83	0.52	0.86	0.85	0.75
	Range	0 .08– 1	0.09 – 1	NA	0.17 – 1	0.10 – 1	NA
	Standard deviation	0.19	0.18	0.44	0.14	0.18	0.17
Unassigned ASVs at genus level	*Number*	2,479	1,681	**0.02**	3,899	3,559	0.28
	% total ASVs	56%	38%	**0.00**	71%	64%	**0.01**
Unassigned ASVs at genus level with both databases	*Number*		1,011			2,871	
	% total ASVs		23%			53%	



a) b)

Figure D-3 - Venn diagrams showing the number of genera identified in the RDP database only compared to a) SILVA for 16S and b) UNITE for ITS. This is a comparison of the total list of genera and does not address how individual ASVs are mapped, which was addressed separately below. Venn diagrams showing the number of genera identified in the RDP database only compared to a) SILVA for 16S and b) UNITE for ITS.

D.4.2 Decontamination

ASVs identified as contaminants were reasonably consistent between "microDecon" and "decontam"(**Table D-3)**. The large numbers of negative controls used in this study (around 1/3 of total samples) appeared useful in robust identification of contamination. 16S had 350 ASVs adjusted in "microDecon" and 344 flagged as contaminants by "decontam". For ITS, 303 ASVs were adjusted by "microDecon" compared to 251 identified by "decontam". Alpha and beta diversity metrics dropped slightly after processing with "microDecon", as would be expected with removal of reads. For 16S, the number of significantly differentially abundant ASVs by location declined, while for ITS it increased slightly post decontamination.

Table D-3 - Contamination correction method tested on post USEARCH data for microbial aerosol communities at urban parks measured using 16S and ITS DNA sequencing with a combined taxonomy table. For both ITS and 16S "microDecon" was used on default settings. The decontaminated ASV table produced was checked for consistency against contaminant ASVs identified in "decontam" on a range of its options (frequency, prevalence and combination). Observed (D0) is the raw diversity or number of unique ASVs detected. ExpShannon (D1) refers to the exponential of the Shannon diversity index and invSimpson (D2) refers to the inverse of the Simpson diversity index.

Dataset		16S	ITS
Metrics			
ASVs flagged as contaminates		"microDecon" adjusted 350 ASVs for contamination. "decontam" flagged frequency – 117 combination – 340, prevalence 13 – 217 ASVs depending on threshold.	"microDecon" adjusted 303 for contamination, "decontam" flagged frequency – 104 combination – 251 prevalence 1 – 77 ASVs depending on threshold.
Alpha diversity (Hills numbers)	*Observed (D0)*	236	388
	expShannon (D1)	15.92	28.2
	invSimpson (D2)	10.45	18.03
Beta diversity (Bray-Curtis)	*Mean*	0.86	0.91
	Median	0.91	0.94
	Range	0.21-1	0.25-1
	Standard deviation	0.14	0.1
Number of ASVs significantly differentially abundant (by location)		144	54

D.5 Discussion

Bacteria and fungi observed were broadly consistent between pipelines and known to be environmental microbes or are associated with plants, humans or other animals. Fungal genera present were predominantly plant or soil associated, with many wood rotting species. Bacterial *Pseudomonas, Ralstonia* and *Methylobacterium* spp. and fungal *Penicillium, Alternaria* and *Cladosporiudum* spp. were consistent with previous bioaerosol studies (Barberán et al., 2015; Be et al., 2015; Garcia-Alcega et al., 2020). Chloroplasts were commonly observed, presumably from pollen and other plant fragments, and are abundant in similar bioaerosol studies (Brodie et al., 2007b; Franzetti et al., 2010; Woo et al., 2013). Abundance and type of bacterial and fungal genera varied by location. For instance, *Bacillus* spp., which are frequently present in bioaerosols (Bottos et al., 2014), were interestingly only detected at about half of locations sampled. The results from both pipelines appeared consistent with previous bioaerosol studies.

D.5.1 ASV Inference

Low biomass is a particular problem for bioaerosol studies (Amato et al., 2019; Pearce et al., 2016). Therefore, preserving as many reads as possible is invaluable. The higher read counts consistently detected by USEARCH due to matching pre-filtering were preferred, since less information was lost, and raw read counts should be more accurate. Greater numbers of samples completing the pipeline for USEARCH was beneficial, since this provides more data to work with. Fewer ASVs were identified in total for USEARCH, but more per sample, such that alpha and beta diversities were generally greater than for "Dada2" (**Table D-1**). Fewer ASVs may indicate that USEARCH is missing variants, but it may also mean that "Dada2" is detecting spurious ASVs. Greater alpha and beta diversities are likely to be desirable as more information should be available to answer the research questions posed. The number of ASVs identified as significantly differentially abundant by location was significantly higher for USEARCH, suggesting it is likely to be more informative, with the objective in mind for this dataset of detecting differences in the aero-microbiome among locations. Consistency between pipelines can offer support for the ASVs inferred, or not inferred, and indicate whether an ASV might be spurious. USEARCH ASVs appeared to be much more likely to be supported by "Dada2" than the inverse. This provided further indications that "Dada2" ASVs may be spurious. To shed light on this, read counts identified as ASVs exclusive to "Dada2" were compared to those exclusive to USEARCH. Very low abundance ASVs are likely to represent sequencing errors or artefacts, although with low biomass data genuine singletons are possible (Edgar, 2010). The more common an ASV is, the more likely it is to be a true variant. "Dada2" identified 3,766 unique 16S ASVs and just 25 had more than 100 reads in the entire dataset. A large number of very low abundance sequences made up this total. USEARCH identified 663 unique 16S ASVs and 22 of them had more than 100 reads in the entire dataset. There were also far fewer very low abundance sequences. ITS showed a similar pattern, 29 "Dada2" ASVs had over 500 reads, whereas 40 USEARCH ASVs had over 500 reads in total. Higher abundance ASVs were identified consistently by both pipelines. USEARCH appeared to be a superior choice, as it lost fewer reads, appeared to identify fewer extraneous ASVs and a much higher percentage of USEARCH ASVs were supported by "Dada2" than the inverse. On some levels the differences are minor, as most later filtering would remove all the low abundance sequences from further analysis (or they would be unlikely to impact it significantly), but the differences in diversity metrics and differentially abundant ASVs are potentially more impactful. This result demonstrates the value of assessing different pipelines for the users' specific NGS datasets, as the recent literature typically uses "Dada2" and has moved away from USEARCH on the assumption that newer equals better (Archer et al., 2020; Archer et al., 2019; Mhuireach et al., 2020). While "Dada2" performs well in test community sequencing trials (Callahan et al., 2016), and is easy to implement as part of the R environment (Callahan, 2020), it doesn't appear to be universally superior for all kinds of data. Especially when a biome diverges from the typical test community, testing a tool for the data on hand, with the specific research questions in mind, is advisable.

D.5.2 Taxonomic Assignments

Fewer unassigned ASVs are desirable as more inferences can be made about identified ASVs. As with ASV inference, consistency of classification between databases can lend support to the taxonomy predicted for an ASV. The trade-off between database size and accuracy is problematic (Edgar, 2018). While unassigned ASVs are disadvantageous, incorrect assignments are potentially disastrous (Edgar, 2018). Unassigned ASVs is a particular problem for fungi (Archer et al., 2020). In this dataset, ITS had up to 71% unassigned ASVs, which makes inference about the fungi difficult, but more, potentially inaccurate, assignments could lead to the wrong conclusions being drawn (**Table D-2**). The results showed that when more genera are assigned to ASVs, alpha and beta diversity using genera generally improve, and any biological signal should be more apparent. Combining assignments from both databases was investigated to circumvent the trade-off. When information from both databases was used only 23% of 16S and 53% of ITS ASVs remained unassigned (**Table D-2**). The process was fully automated in R to reduce error and it would be possible to add information from additional databases if more assignments and certainty were desired. With two databases, 12% to 15% more ASVs were assigned depending on amplicon, at the same or greater confidence, than in the larger database alone. A combined database approach was able to circumvent the trade-off. It performed better than classification with either database alone, and provided additional confidence information about assignments. All recent bioaerosol studies have only used a single database (Archer et al., 2020; Archer et al., 2019; Maki et al., 2017; Mhuireach et al., 2020; Woo et al., 2013), while suffering from a lower rate of assignments, particularly for fungi, even while using a larger and less accurate database (Archer et al., 2020). A further benefit of multi-database use arises as different databases cover different domains of life, for instance SILVA covers all three domains of life (Bowers et al., 2013). If a target region amplified is one that is shared across multiple domains, it could be beneficial to use multiple specialized databases to assign as many sequences accurately as possible, and not have to default to using the database that covers all the relevant domains. Even without using multiple databases, having an understanding of the differences in the taxonomic output when different databases are used is valuable to inform the subsequent discussion of results, and confidence placed in taxonomic conclusions, which is not the approach taken in the current literature (Archer et al., 2020; Archer et al., 2019; Maki et al., 2017; Mhuireach et al., 2020; Woo et al., 2013). Use, or at least consideration of, information in different taxonomic databases is suggested, as it is clear from this result that database choice can substantially alter taxonomic inference.

D.5.3 Decontamination

ITS data were less contaminated overall, which is consistent with other studies (McKnight et al., 2019). ASVs identified by each approach were compared to understand the degree of differentiation between whole ASV deletion (“decontam”) or background contamination deduction (“microDecon”) approaches. At a higher level, the number of ASVs identified as potential contaminants with each tool (“microDecon” and the combination method in “decontam”) was similar (**Table D-3**). Common ASVs, such as *Ralstonia* spp., were partially removed by “microDecon” and flagged as potential contaminants by “decontam”. These

ASVs appeared to be both genuine contaminants and genuine constituents of the bioaerosol population (Waugh, Granger, & Gaggar, 2010). This supports the subtraction approach in "microDecon", as opposed to the whole ASV deletion with "decontam" for use with this dataset. There was no evidence that "microDecon" had missed contaminant ASVs. All contaminant ASVs per "decontam" with over 400 reads in the whole dataset were adjusted by "microDecon", suggesting that no ASV deletion was required after processing with "microDecon".

This conclusion was further tested by considering the two different types of controls that were used. Sampling controls went through the entire process and should therefore identify environmental and laboratory contamination. Laboratory controls (unused filters which went through the described protocols from DNA extraction onwards) should identify contaminating ASVs which were not present in the sampled environment. These ASVs were likely to entirely be contaminants and should therefore be deleted from the data, rather than partially subtracted. Very few ASVs in each dataset were definitive laboratory contaminants according to "decontam". Of the few identified, they were either totally removed by "microDecon" or had under 100 total reads remaining in the whole dataset after correction. Decontamination with "microDecon" alone is likely to be the superior approach for these data, due to the cross-over between contaminating ASVs and legitimate members of the bioaerosol community and no evidence of "microDecon" omitting important contaminants. The large numbers of controls used in this study, of the two types (laboratory and sampling) proved useful in both packages for robustly identifying potential contaminating sequences, and this procedure is recommended (Davis et al., 2018; McKnight et al., 2019), particularly for the aero-microbiome or similar, where the ratio of legitimate to contaminating sequences is likely to be higher than other biomes. While ASVs were consistently identified by each package, the output ASV tables could be very different if a high abundance sequence was flagged as a contaminant and totally deleted ("decontam") verses being partially a contaminant so that ASV was only partly adjusted ("microDecon"). Instances of this occurrence were seen in the test data used here. Therefore, these results indicate that consideration of the decontamination removal in NGS data is important and can have a large impact on the reported results and conclusions drawn. This difference is particularly marked in situations where ASVs could be both contaminants and community members, and where contaminants are prevalent. Background subtraction does not appear to be considered in the existing literature for bioaerosols, with solely the use of "decontam" appearing to be the dominant approach (Archer et al., 2020; Archer et al., 2019; Maki et al., 2017; Mhuireach et al., 2020; Woo et al., 2013). This work highlights the value of considering contamination correction with reference to the nature of the data set and research questions posed of it.

D.6 Conclusions

While metagenomics and transcriptomics approaches have recently been applied to bioaerosol communities, single amplicon NGS remains the most common method used to interrogate these communities. Therefore, consideration of the process chosen to analyse single amplicon NGS data is pertinent. This study has demonstrated a methodical approach

to selecting an optimized bioinformatics pipeline from a plethora of available options. Further, it has confirmed that bioinformatic data processing choices can make substantial differences in the ASV table and the taxonomy produced from 16S and ITS NGS data. Key differences were in the number of reads, ASVs inferred and ASVs remaining after decontamination. However, at a high level, datasets processed with varying combinations of approaches showed similar high abundance ASVs and taxonomic assignments. Microbiome datasets are very hard to verify, especially for poorly characterised environments like air. The utility of comparison of different approaches to sense check aero-microbiome sequence data was demonstrated. Explicit consideration of optimal approaches and use of multiple taxonomic databases, for example, were not noted in bioaerosol literature reviewed. Here, a process is laid out that researchers can step through and use to develop their own protocol, which is justified by the use of the statistics and comparisons.

D.6.1 Recommendations

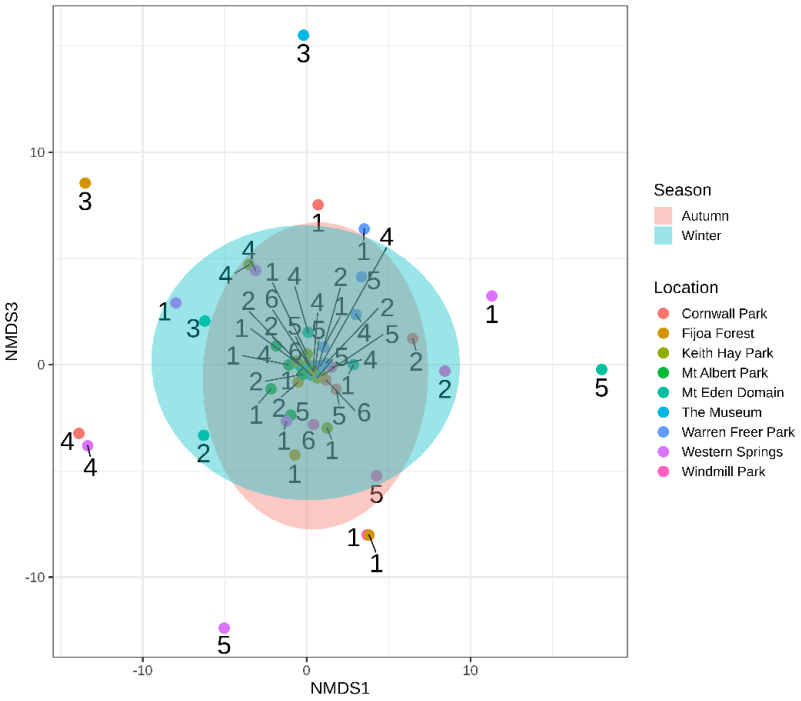
This study shows that understanding the characteristics of the dataset to be analysed and choosing a bioinformatics approach using a systematic procedure is crucial for generating a high-quality dataset. For bioaerosol data, in particular, it is important to have large numbers of negative controls, with different types of controls at different times during sample collection and processing. Stringent laboratory practices are advised to reduce contamination and allow easier identification of contaminating reads in data. Comparing a wide variety of diversity metrics and outputs from multiple bioinformatics pipelines for the relevant dataset in a systematic way is recommended.

Appendix E. Supplementary Information

E.1 Chapter 2 Supplementary Information

E.1.1 NMDS Ordination for 16S

a)



b)

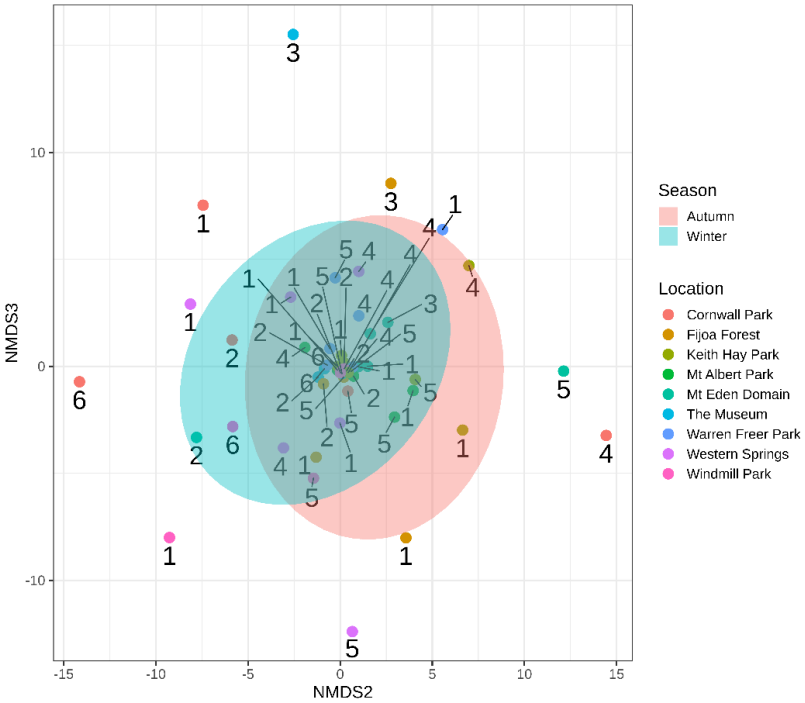
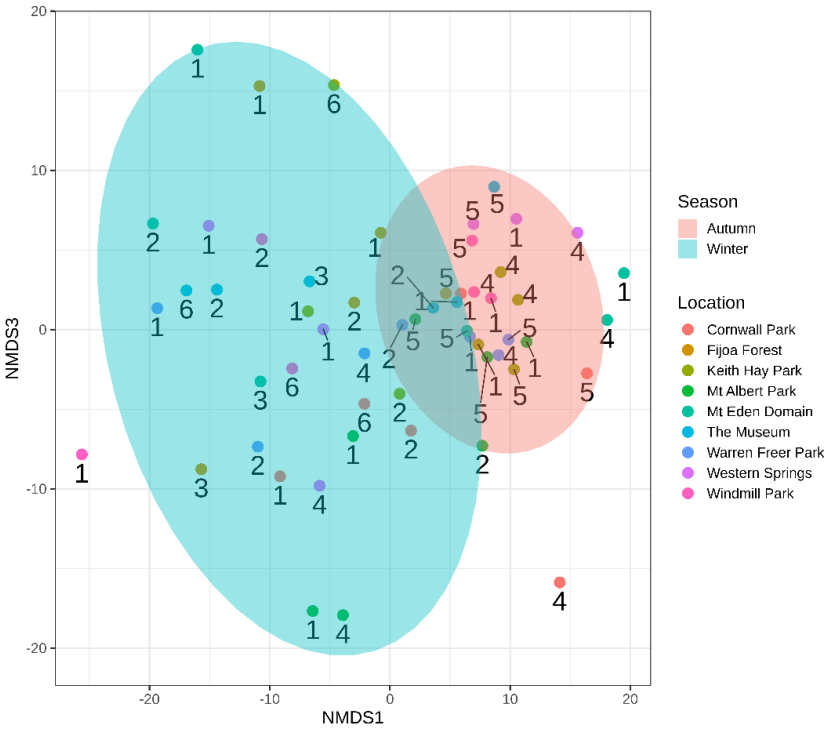


Figure E-1 - NMDS ordination of the a) 16S dimensions one and three and b) 16S dimensions two and three of the Aitchison distances among samples. Season is indicated by ellipses representing the *t* distribution of the autumn and winter points. Locations are

denoted by colour and each sample is numbered with the trajectory cluster to which it belongs.

E.1.2 NMDS Ordination for ITS

a)



b)

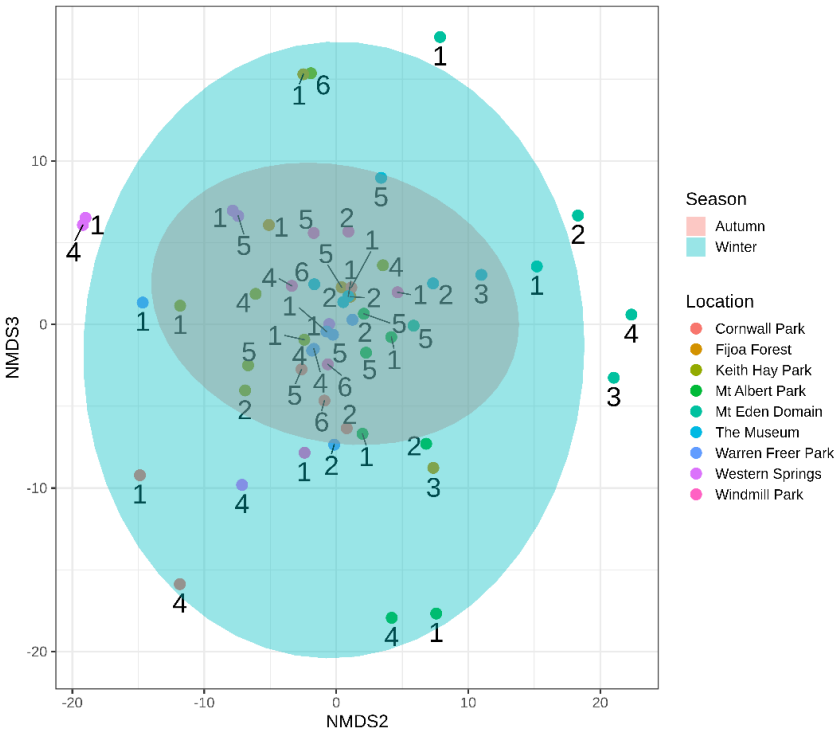


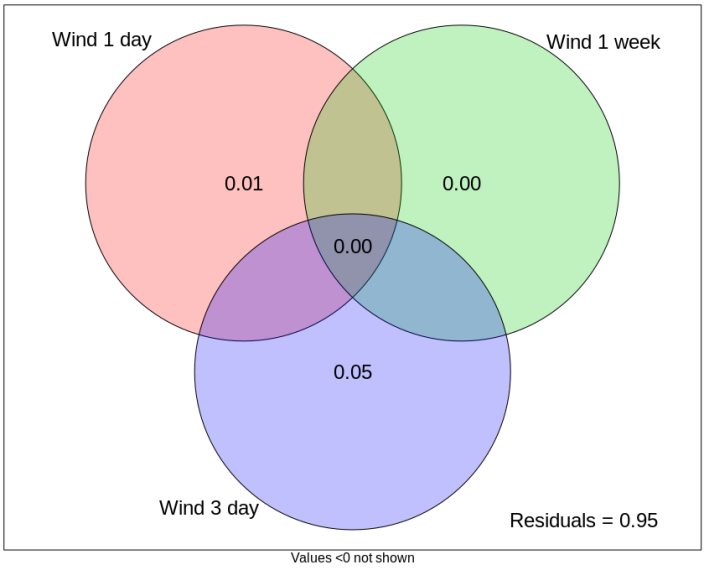
Figure E-2 - NMDS ordination of the a) ITS dimensions 1 and 3 and b) ITS dimensions 2 and 3 of the Aitchison distances among samples. Season is indicated by ellipses representing

the *t* distribution of the autumn and winter points. Locations are denoted by colour and each sample is numbered with the trajectory cluster to which it belongs.

E.1.3 Results from Non-Compositional Data Analysis

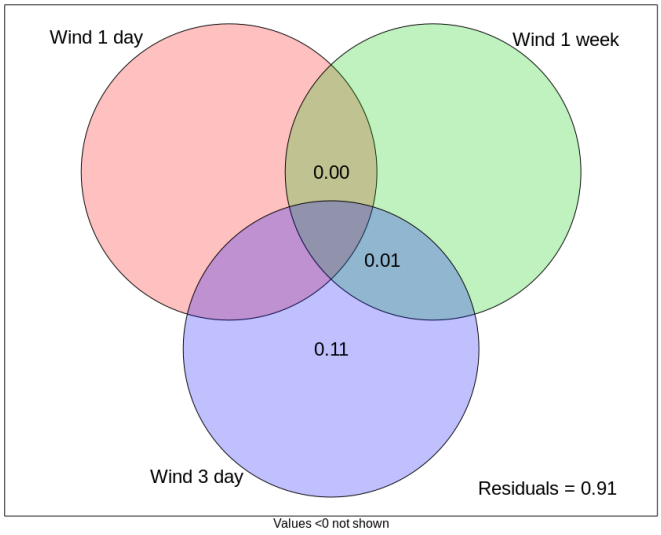
Comparison of Wind Back-Trajectory Effect to a Non-Compositional Dataset

a)



b)

c)



d)

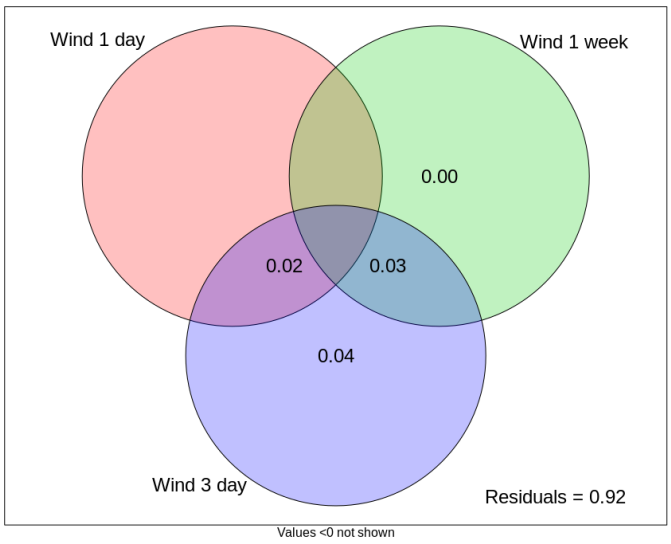
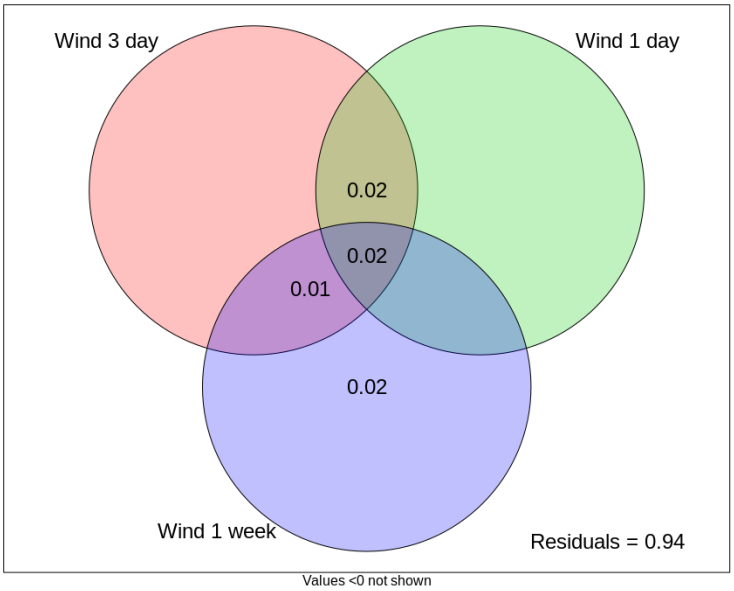
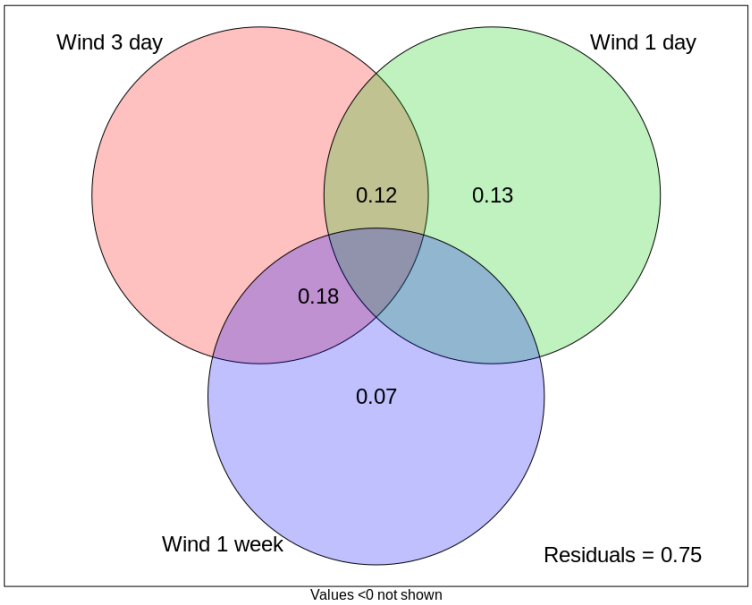


Figure E-3 - db-RDA with variance partitioning for one day, three day and one-week back trajectories for 16S. a) uses Bray-Curtis dissimilarities, b) uses richness data, measured by Hill D0 or raw diversity, c) uses Hill number D1 and d) uses Hill number D2. The Bray Curtis metrics were calculated based on the filtered ASV table. The Hill numbers were based on the unfiltered data, given the need to account for rare species which would be disrupted by the filtering applied. These non-compositional analyses indicate that three-day wind has the greater predictive value of the bioaerosol community for 16S. The frequency of shared variation, however, also indicates that multiple trajectory lengths have a similar predictive value.

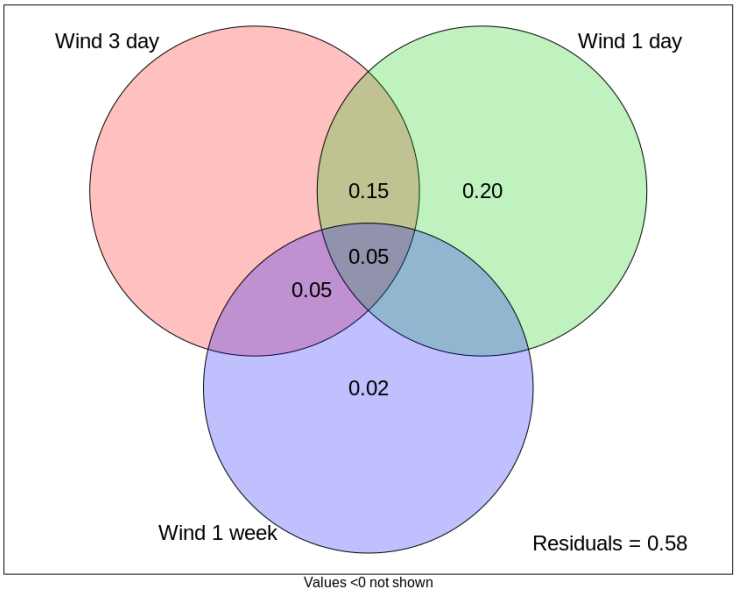
a)


Wind 3 day
Wind 1 day
0.02
0.02
0.01
0.02
Wind 1 week
Residuals = 0.94
Values <0 not shown

b)


Wind 3 day
Wind 1 day
0.12
0.13
0.18
0.07
Wind 1 week
Residuals = 0.75
Values <0 not shown

c)



d)

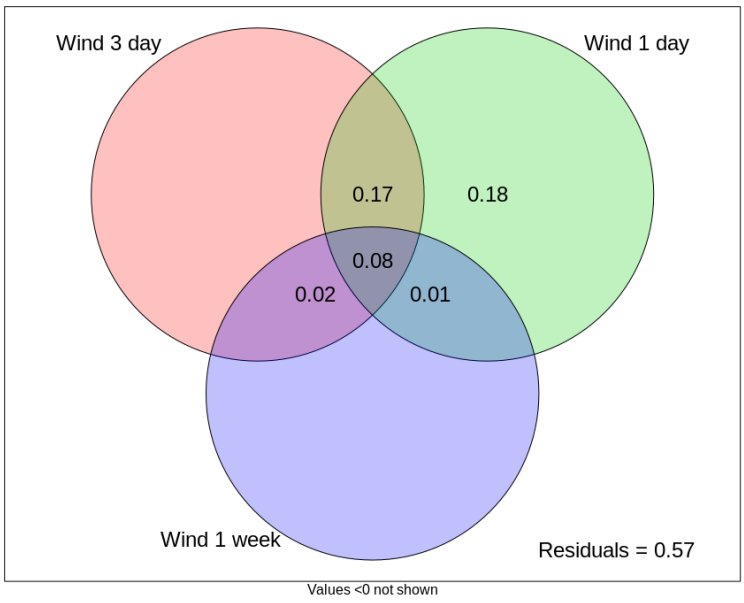
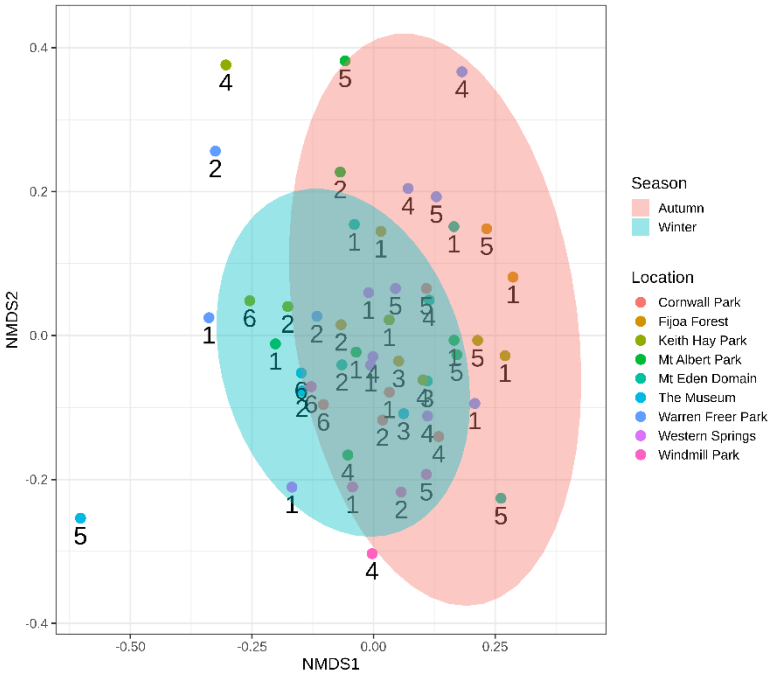


Figure E-4 - db-RDA with variance partitioning for one day, three day and one-week back trajectories for ITS. a) uses Bray-Curtis dissimilarities, b) uses richness data, measured by Hill D0 or raw diversity, c) uses Hill number D1 and d) uses Hill number D2. The Bray Curtis metrics were calculated based on the filtered ASV table. The Hill numbers were based on the unfiltered data, given the need to account for rare species which would be disrupted

by the filtering applied. For ITS one-day wind has the highest R^2 value in comparison to the other trajectory lengths in almost all the analyses. As with 16S there was frequent shared variation, demonstrating the predictive potential of multiple trajectory lengths. Both amplicon's results from non-compositional and compositional tools are broadly similar. For ITS an optimal trajectory length of one day is supported. For 16S non-compositional tools suggested that three-day trajectory could be better for predicting bioaerosol communities, whereas compositional tools suggested that one-day trajectories were superior. These results generally suggest that either one or three-day length trajectories would be reasonable predictors for bioaerosol communities.

Non-Compositional NMDS Ordination

a)



b)

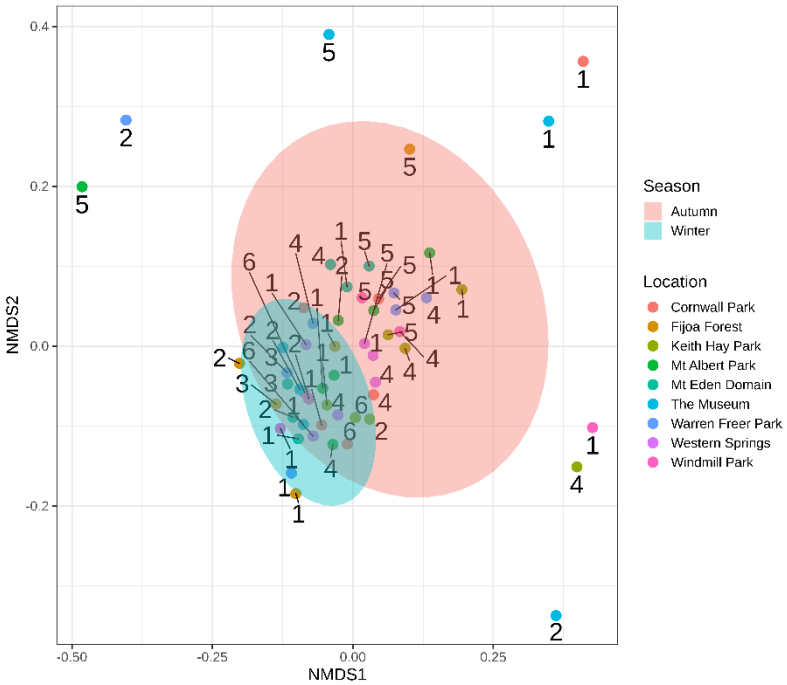
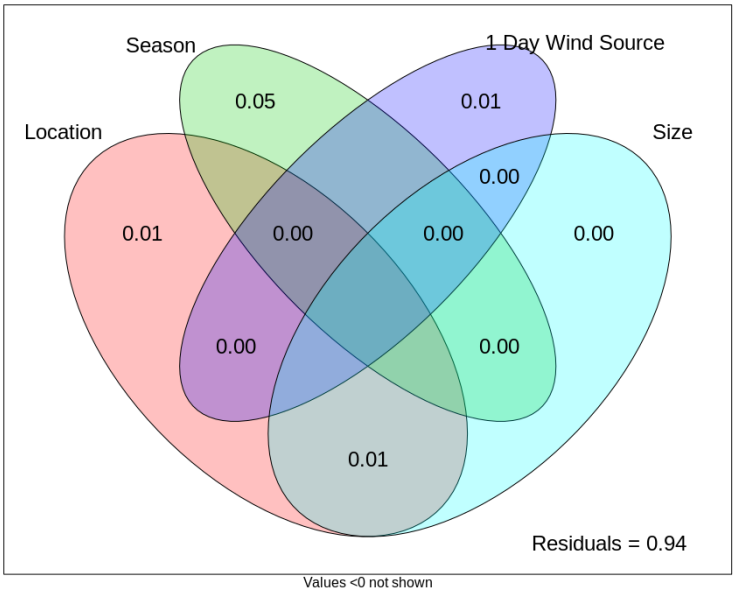


Figure E-5 - NMDS ordination of BC dissimilarities on a filtered Hellinger transformed ASV table for a) 16S and b) ITS. Seasons were indicated by an ellipse based on the *t* distribution of the summer and winter points. Colour of the points denotes the location and the number next to each point indicates which trajectory cluster it belongs to. The non-compositional analyses showed generally reduced levels of clustering in comparison to

compositional tools, with quite substantial variation in the location of individual points. Seasonal patterns remained evident. ITS appeared to be more clumped than its compositional equivalent, while 16S was less clumped. It was necessary to apply a Hellinger transformation before generating BC distances in order to resolve a large degree of clumping which was present with BC distances alone. No additional transformations were required with the Aitchison distance matrix. Both non-compositional analyses are generally consistent with the results of the compositional tools, supporting the importance of season, location and trajectory cluster.

Non-Compositional db-RDA with Variance Partitioning

a)



b)

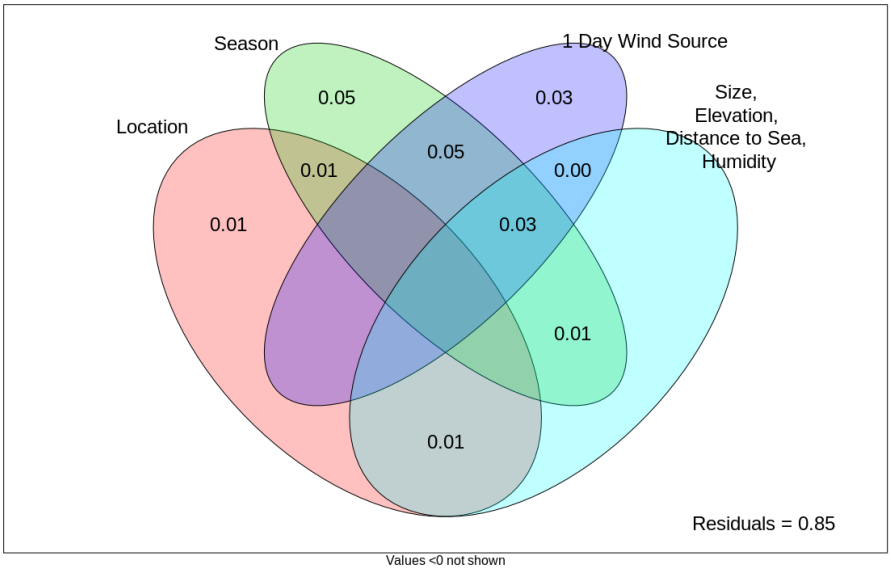
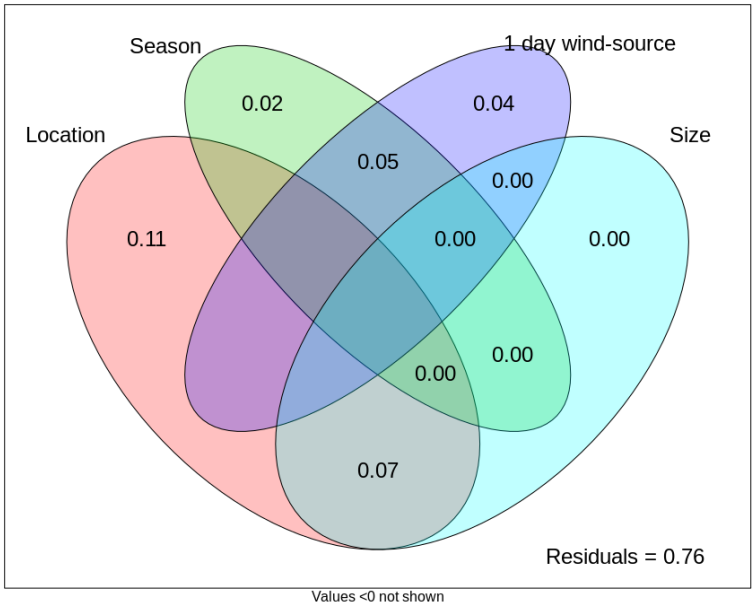


Figure E-6 - BC distance-based redundancy analysis with variance partitioning of the filtered ASV table for a) 16S and b) ITS. Season, location, and one-day wind source are consistently estimated to have comparatively high R^2 values for both amplicons, in broad agreement with compositional methods. As with the compositional methods the ITS

model has greater predictive power, and lower overall residuals (unexplained variation). While the importance of seasonality was less apparent with compositional methods for 16S, it has a comparatively high R^2 in this analysis.

a)



b)

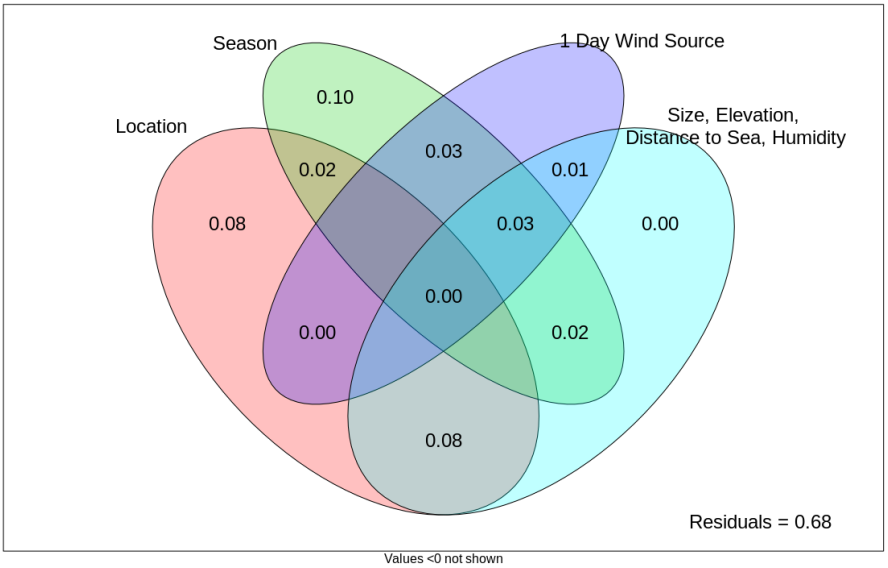
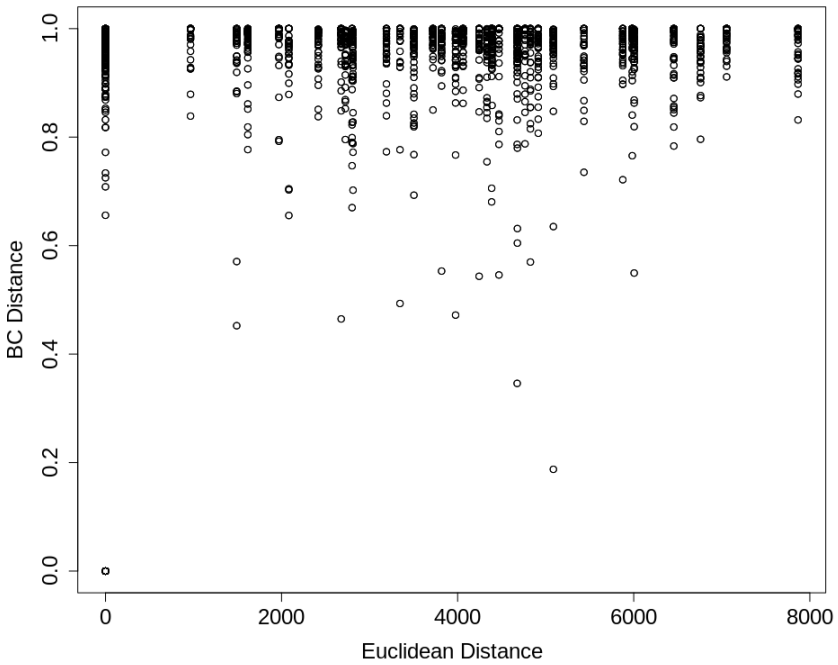


Figure E-7 - Jaccard distance-based redundancy analysis with variance partitioning of the filtered ASV table for a) 16S and b) ITS. Season, location, and one-day wind source consistently have comparatively high R^2 values for both amplicons, in broad agreement with the previous non-compositional and compositional methods presented. As with the compositional methods the ITS model has greater predictive power, and lower overall residuals (unexplained variation). While the importance of seasonality was less apparent

with compositional methods for 16S, it had a comparatively high R^2 in the Jaccard based analysis.

No Evidence of Distance-Dissimilarity from Non-Compositional Analysis

a)



b)

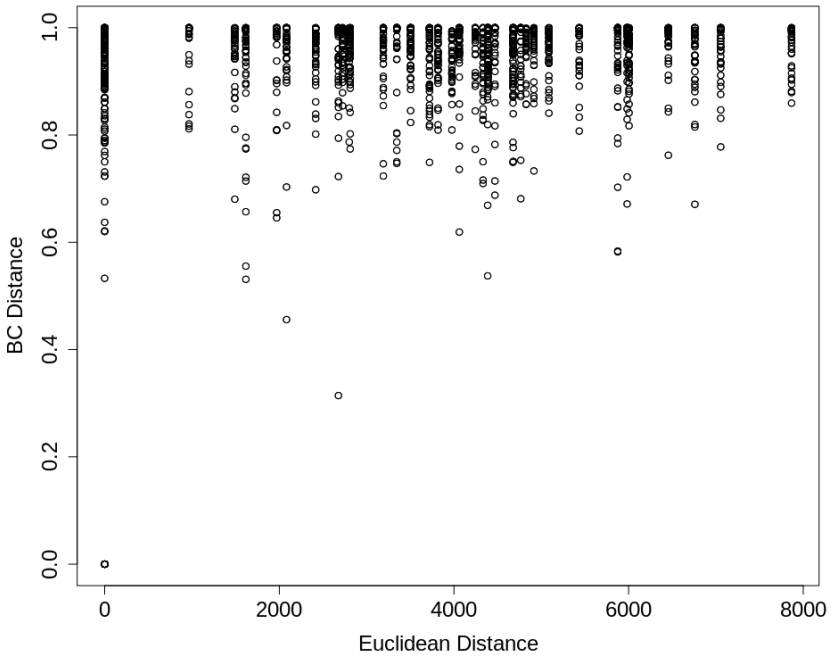


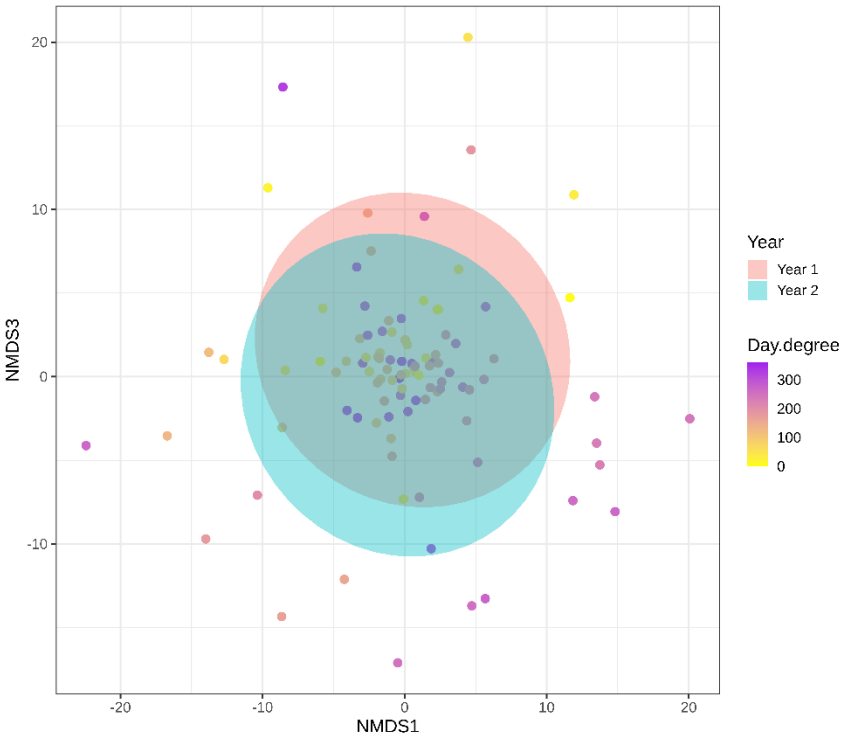
Figure E-8 - Distance-dissimilarity plots for a) 16S and b) ITS of BC distances calculated from the filtered ASV table against Euclidean spatial distances. In agreement with

compositional methods, non-compositional based distance-dissimilarity is not apparent for either bacteria or fungi across the distances sampled.

E.2 Chapter 3 Supplementary Information

E.2.1 NMDS Ordination for 16S

a)



b)

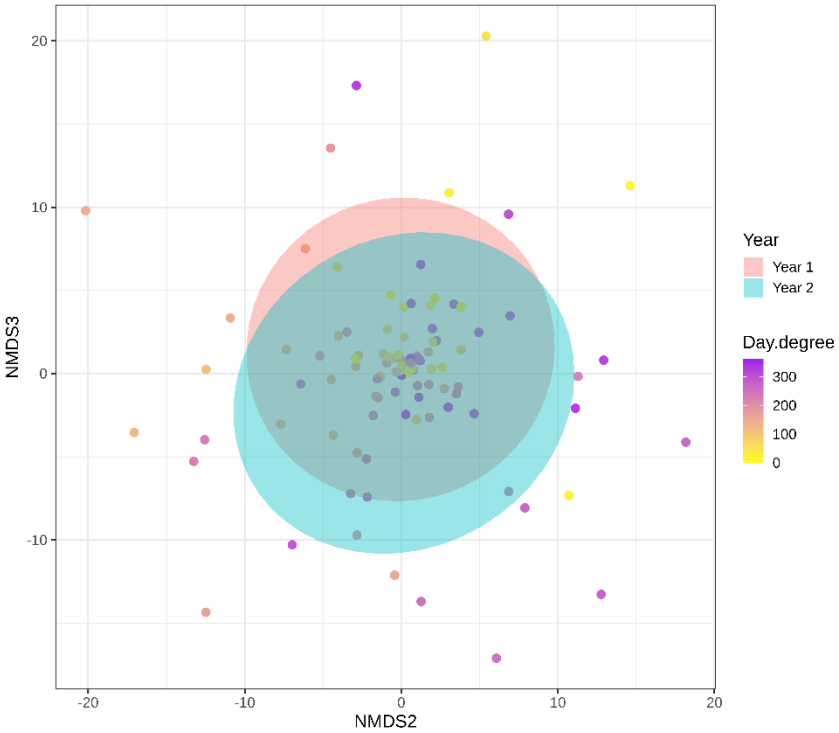
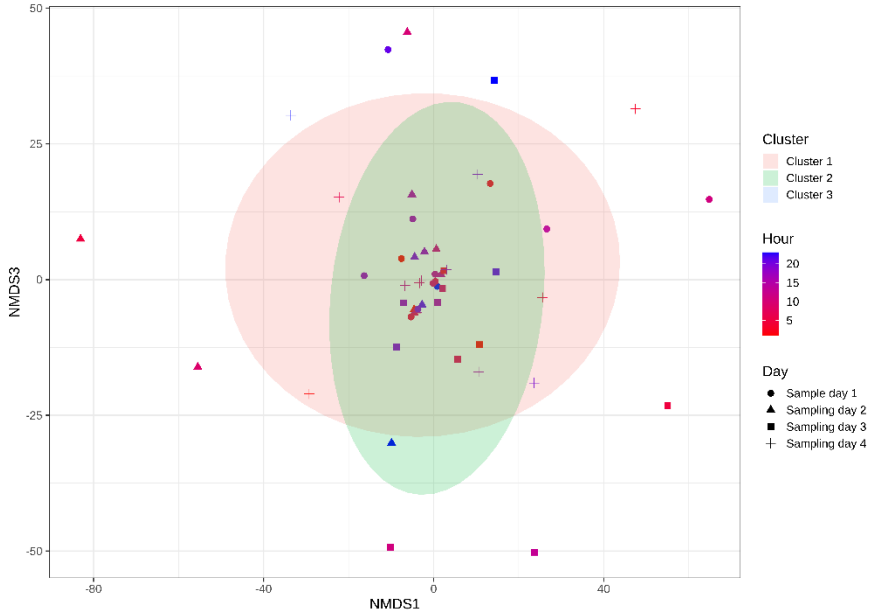


Figure E-9 - NMDS ordination of the a) 16S dimensions one and three and b) 16S dimensions two and three of the Aitchison distances among samples. The study year is indicated by ellipses representing the *t* distribution of the year one and year two points. Days in circle degrees are indicated by the gradient colour.

E.3 Chapter 4 Supplementary Information

E.3.1 NMDS Ordination for Antarctica 16S

a)



b)

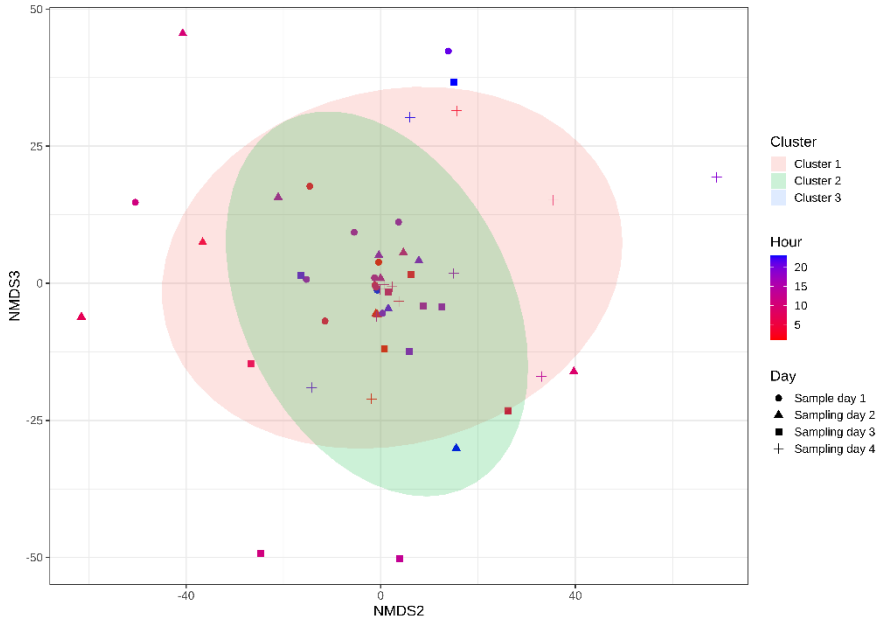
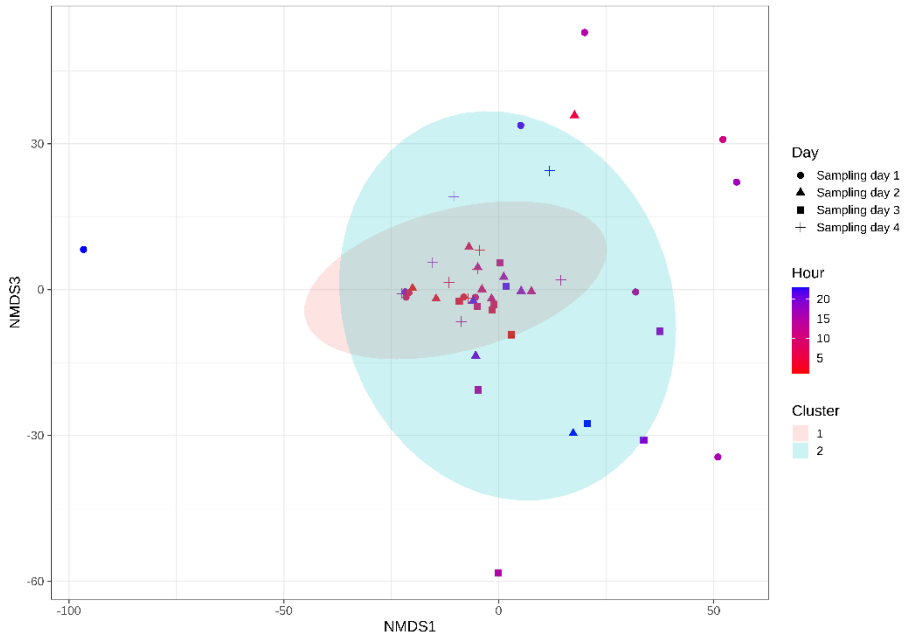


Figure E-10 - NMDS ordination of the a) 16S dimensions one and three and b) 16S dimensions two and three of the Aitchison distances among samples. The back-trajectory cluster is indicated by ellipses representing the *t* distribution of the points relating to each cluster. Hours are indicated by the gradient colour, sampling day is indicated by shape.

E.3.2 NMDS Ordination for Antarctica ITS

a)



b)

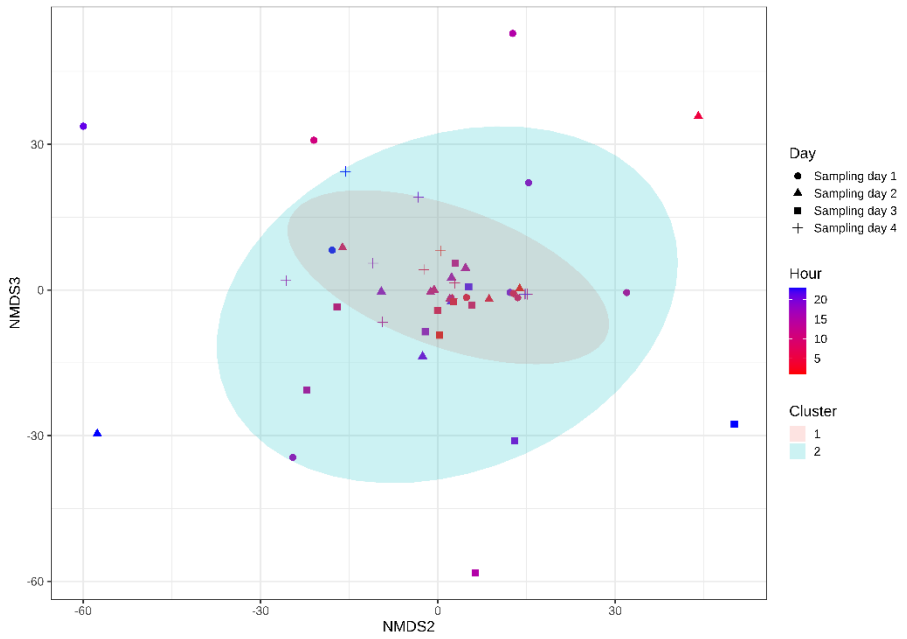


Figure E-11 - NMDS ordination of the a) ITS dimensions one and three or b) ITS dimensions two and three of the Aitchison distances among samples. The back-trajectory cluster is indicated by ellipses representing the *t* distribution of the points relating to each cluster. Hours are indicated by the gradient colour, sampling day is indicated by shape.

Bibliography

Abarenkov, K., Henrik Nilsson, R., Larsson, K.-H., Alexander, I. J., Eberhardt, U., Erland, S., . . . Kõljalg, U. (2010). The UNITE database for molecular identification of fungi – recent updates and future perspectives. *New Phytologist, 186*(2), 281-285. doi:10.1111/j.1469-8137.2009.03160.x

Aitchison, J. (1982). The Statistical Analysis of Compositional Data. *Journal of the Royal Statistical Society. Series B (Methodological), 44*(2), 139-177.

Amato, P., Besaury, L., Joly, M., Penaud, B., Deguillaume, L., & Delort, A.-M. (2019). Metatranscriptomic exploration of microbial functioning in clouds. *Scientific Reports, 9*(1), 4383. doi:10.1038/s41598-019-41032-4

Amato, P., Demeer, F., Melaouhi, A., Fontanella, S., Martin-Biesse, A. S., Sancelme, M., . . . Delort, A. M. (2007). A fate for organic acids, formaldehyde and methanol in cloud water: their biotransformation by micro-organisms. *Atmospheric Chemistry and Physics Discussions, 7*(2), 5253-5276. doi:10.5194/acpd-7-5253-2007

Amato, P., Joly, M., Besaury, L., Oudart, A., Taib, N., Moné, A. I., . . . Debroas, D. (2017). Active microorganisms thrive among extremely diverse communities in cloud water. *PLOS ONE, 12*(8), e0182869. doi:10.1371/journal.pone.0182869

Amend, A., Burgaud, G., Cunliffe, M., Edgcomb, V. P., Ettinger, C. L., Gutiérrez, M. H., . . . Gladfelter, A. S. (2019). Fungi in the Marine Environment: Open Questions and Unsolved Problems. *mBio, 10*(2), e01189-01118. doi:10.1128/mBio.01189-18

Andreeva, I. S., Borodulin, A. I., Buryak, G. A., Zhukov, V. A., Zykov, S. V., Marchenko, Y. V., . . . Churkina, T. V. (2002). Biogenic Component of Atmospheric Aerosol in the South of West Siberia. *Chemistry for Sustainable Development, 10*, 523 - 537. doi:10.1134/S1024856010010148

Archer, C. L., & Caldeira, K. (2009). Global assessment of high-altitude wind power. *Energies, 2*, 307 - 319. doi:10.3390/en20200307

Archer, S. D., McDonald, I. R., Herbold, C. W., & Cary, S. C. (2014). Characterisation of bacterioplankton communities in the meltwater ponds of Bratina Island, Victoria Land, Antarctica. *FEMS Microbiol Ecol, 89*(2), 451-464. doi:10.1111/1574-6941.12358

Archer, S. D. J., Lee, K. C., Caruso, T., King-Miaow, K., Harvey, M., Huang, D., . . . Pointing, S. B. (2020). Air mass source determines airborne microbial diversity at the ocean–atmosphere interface of the Great Barrier Reef marine ecosystem. *The ISME Journal, 14*(3), 871-876. doi:10.1038/s41396-019-0555-0

Archer, S. D. J., Lee, K. C., Caruso, T., Maki, T., Lee, C. K., Cary, S. C., . . . Pointing, S. B. (2019). Airborne microbial transport limitation to isolated Antarctic soil habitats. *Nature Microbiology, 4*(6), 925-932. doi:10.1038/s41564-019-0370-4

Atkins, C. B., & Dunbar, G. B. (2009). Aeolian sediment flux from sea ice into Southern McMurdo Sound, Antarctica. *Global and Planetary Change, 69*(3), 133-141. doi:10.1016/j.gloplacha.2009.04.006

Bahl, J., Lau, M. C. Y., Smith, G. J. D., Vijaykrishna, D., Cary, C. S., Lacap, D. C., . . . Pointing, S. B. (2011). Ancient origins determine global biogeography of hot and cold desert cyanobacteria. *Nature Communications, 2*, 163. doi:10.1038/ncomms1167

Balvočiūtė, M., & Huson, D. H. (2017). SILVA, RDP, Greengenes, NCBI and OTT — how do these taxonomies compare? *BMC Genomics, 18*(2), 114. doi:10.1186/s12864-017-3501-4

Balyan, P., Ghosh, C., Das, S., & Banerjee, B. (2019). Spatio-temporal characterisation of bioaerosols at diverse outdoor land-use sites in an urban environment. *Aerobiologia*, 1-5.

Barberán, A., Henley, J., Fierer, N., & Casamayor, E. O. (2014). Structure, inter-annual recurrence, and global-scale connectivity of airborne microbial communities. *Science of the Total Environment, 487*, 187-195.

Barberán, A., Ladau, J., Leff, J. W., Pollard, K. S., Menninger, H. L., Dunn, R. R., & Fierer, N. (2015). Continental-scale distributions of dust-associated bacteria and fungi. *Proceedings of the National Academy of Sciences, 112*(18), 5756-5761. doi:10.1073/pnas.1420815112

Baselga, A., & Orme, C. D. L. (2012). betapart: an R package for the study of beta diversity. *Methods in Ecology and Evolution, 3*(5), 808-812. doi:10.1111/j.2041-210X.2012.00224.x

Bauer, H., Kasper-Giebl, A., Löflund, M., Giebl, H., Hitzenberger, R., Zibuschka, F., & Puxbaum, H. (2002). The contribution of bacteria and fungal spores to the organic carbon content of cloud water, precipitation and aerosols. *Atmospheric Research, 64*(1), 109-119. doi:10.1016/S0169-8095(02)00084-4

Be, N. A., Thissen, J. B., Fofanov, V. Y., Allen, J. E., Rojas, M., Golovko, G., . . . Jaing, C. J. (2015). Metagenomic analysis of the airborne environment in urban spaces. *Microb Ecol, 69*(2), 346-355. doi:10.1007/s00248-014-0517-z

Behzad, H., Gojobori, T., & Mineta, K. (2015). Challenges and Opportunities of Airborne Metagenomics. *Genome Biology and Evolution, 7*(5), 1216-1226. doi:10.1093/gbe/evv064

Bennett, A., Ponder, M. M., & Garcia-Diaz, J. (2018). Phoma Infections: Classification, Potential Food Sources, and Its Clinical Impact. *Microorganisms, 6*(3), 58. doi:10.3390/microorganisms6030058

Bottos, E. M., Woo, A. C., Zawar-Reza, P., Pointing, S. B., & Cary, S. C. (2014). Airborne Bacterial Populations Above Desert Soils of the McMurdo Dry Valleys, Antarctica. *Microbial Ecology, 67*(1), 120-128. doi:10.1007/s00248-013-0296-y

Bowers, R. M., Clements, N., Emerson, J. B., Wiedinmyer, C., Hannigan, M. P., & Fierer, N. (2013). Seasonal Variability in Bacterial and Fungal Diversity of the Near-Surface Atmosphere. *Environmental Science & Technology, 47*(21), 12097-12106. doi:10.1021/es402970s

Bowers, R. M., McCubbin, I. B., Hallar, A. G., & Fierer, N. (2012). Seasonal variability in airborne bacterial communities at a high-elevation site. *Atmospheric Environment, 50*, 41-49.

Bowers, R. M., McLetchie, S., Knight, R., & Fierer, N. (2011a). Spatial variability in airborne bacterial communities across land-use types and their relationship to the bacterial communities of potential source environments. *The ISME Journal, 5*, 601 - 612. doi:10.1038/ismej.2010.167

Bowers, R. M., Sullivan, A. P., & Costello, E. K. (2011b). Sources of bacteria in outdoor air across cities in the midwestern United States. *Applied and Environmental Microbiology, 77*(18), 6350 - 6356. doi:10.1128/AEM.05498-11

Brodie, E. L., DeSantis, T. Z., & Parker, J. P. M. (2007a). Urban aerosols harbor diverse and dynamic bacterial populations. *Proceedings of the National Academy of Sciences, 104*(1), 299 - 304. doi:10.1073pnas.0608255104

Brodie, E. L., DeSantis, T. Z., Parker, J. P. M., Zubietta, I. X., Piceno, Y. M., & Andersen, G. L. (2007b). Urban aerosols harbor diverse and dynamic bacterial populations. *Proceedings of the National Academy of Sciences, 104*(1), 299-304.

Bulat, S. A. R. (2016). Microbiology of the subglacial Lake Vostok: first results of borehole-frozen lake water analysis and prospects for searching for lake inhabitants. *Phil. Trans. R. Soc. A, 374*(20140292). doi:10.1098/rsta.2014.0292

Burrows, S. M., Butler, T., Jöckel, P., Tost, H., Kerkweg, A., Poschl, U., & Lawrence, M. G. (2009a). Bacteria in the global atmosphere–Part 2: Modeling of emissions and transport between different ecosystems. *Atmospheric Chemistry and Physics, 9*, 9281 - 9297. doi:10.5194/acp-9-9281-2009

Burrows, S. M., Elbert, W., Lawrence, M. G., & Poschl, U. (2009b). Bacteria in the global atmosphere – Part 1: Review and synthesis of literature data for different ecosystems. *Atmospheric Chemistry and Physics, 9*, 9263 - 9280. doi:10.5194/acp-9-9263-2009

Callahan, B. J. DADA2 Pipeline Tutorial (1.16). Retrieved from https://benjjneb.github.io/dada2/tutorial.html

Callahan, B. J. (2020). DADA2 ITS Pipeline Workflow (1.8). Retrieved from https://benjjneb.github.io/dada2/ITS_workflow.html

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods, 13*(7), 581-583. doi:10.1038/nmeth.3869

Cano, R. J., & Borucki, M. K. (1995). Revival and identification of bacterial spores in 25- to 40-million-year-old Dominican amber. *Science, 268*(5213), 1060-1064. doi:10.1126/science.7538699

Cao, C., Jiang, W., Wang, B., Fang, J., Lang, J., Tian, G., . . . Zhu, T. F. (2014). Inhalable Microorganisms in Beijing's PM2.5 and PM10 Pollutants during a Severe Smog Event. *Environmental Science & Technology, 48*(3), 1499-1507. doi:10.1021/es4048472

Cao, Y., Yu, X., Ju, F., Zhan, H., Jiang, B., Kang, H., & Xie, Z. (2021). Airborne bacterial community diversity, source and function along the Antarctic Coast. *Science of the Total Environment, 765*, 142700. doi:https://doi.org/10.1016/j.scitotenv.2020.142700

Carini, P., Delgado-Baquerizo, M., Hinckley, E., Holland-Moritz, H., Brewer, T., Rue, G., . . . Fierer, N. (2020). Effects of spatial variability and relic DNA removal on the detection of temporal dynamics in soil microbial communities. mBio 11: e02776-19. In.

Carvalho, E., Sindt, C., Verdier, A., Galan, C., O'Donoghue, L., Parks, S., & Thibaudon, M. (2008). Performance of the Coriolis air sampler, a high-volume aerosol-collection system for quantification of airborne spores and pollen grains. *Aerobiologia, 24*(4), 191-201. doi:10.1007/s10453-008-9098-y

Cary, C. S., McDonald, I. R., Barrett, J. E., & Cowan, D. A. (2010). On the rocks: the microbiology of Antarctic Dry Valley soils. *Nature Reviews Microbiology, 8*(2). doi:10.1038/nrmicro2281

Chao, A., Chiu, C.-H., & Jost, L. (2016). Phylogenetic Diversity Measures and Their Decomposition: A Framework Based on Hill Numbers. In R. Pellens & P. Grandcolas (Eds.), *Biodiversity Conservation and Phylogenetic Systematics: Preserving our evolutionary heritage in an extinction crisis* (pp. 141-172). Cham: Springer International Publishing.

Chen, M., Jin, L., Sun, Z., Lu, J., Wang, Q., & Hu, Q. (2001). Concentration and flux of bioaerosol and environmental factors. *Progress in Natural Science, 11*(9), 686 - 687.

Cho, B. C., & Azam, F. (1990). Biogeochemical significance of bacterial biomass in the ocean's euphotic zone. *Marine Ecology Progress Series, 63*(2/3), 253 - 259. doi:10.3354/meps063253

Chong, C. W., Pearce, D. A., & Convey, P. (2015). Emerging spatial patterns in Antarctic prokaryotes. *Frontiers in Microbiology, 6*, 1058. doi:10.3389/fmicb.2015.01058

Chown, S. L., Huiskes, A. H. L., Gremmen, N. J. M., Lee, J. E., Terauds, A., Crosbie, K., . . . Bergstrom, D. (2012). Continent-wide risk assessment for the establishment of nonindigenous species in Antarctica. *Proceedings of the National Academy of Sciences, 109*(13), 4938-4943. doi:10.1073/pnas.1119787109

Convey, P. (2006). *Antarctic climate change and its influence on terrestrial ecosystems* (D. Bergstrom, P. Convey, & H. A.H.L. Eds.): Springer.

Convey, P., & Wynn-Williams, D. D. (2002). Antarctic soil nematode response to artificial climate amelioration. *European Journal of Soil Biology, 38*, 255 - 259. doi:10.1016/S1164-5563(02)01155-X

Cowan, D. A., Chown, S. L., Convey, P., Tuffin, M., Hughes, K., Pointing, S., & Vincent, W. F. (2011). Non-indigenous microorganisms in the Antarctic: assessing the risks. *Trends in Microbiology, 19*(11), 540-548. doi:10.1016/j.tim.2011.07.008

Crawford, I., Gallagher, M. W., Bower, K. N., Choularton, T. W., Flynn, M. J., Ruske, S., . . . Stanley, W. R. (2017). Real-time detection of airborne fluorescent bioparticles in Antarctica. *Atmospheric Chemistry and Physics, 17*(23), 14291-14307. doi:10.5194/acp-17-14291-2017

Dannemiller, K. C., Weschler, C. J., & Peccia, J. (2017). Fungal and bacterial growth in floor dust at elevated relative humidity levels. *Indoor Air, 27*(2), 354-363. doi:https://doi.org/10.1111/ina.12313

Davis, N. M., Proctor, D. M., Holmes, S. P., Relman, D. A., & Callahan, B. J. (2018). Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome, 6*(1), 226. doi:10.1186/s40168-018-0605-2

De Silvestri, A., Ferrari, E., Gozzi, S., Marchi, F., & Foschino, R. (2018). Determination of Temperature Dependent Growth Parameters in Psychrotrophic Pathogen Bacteria and Tentative Use of Mean Kinetic Temperature for the Microbiological Control of Food. *Frontiers in Microbiology, 9*(3023). doi:10.3389/fmicb.2018.03023

DeLeon-Rodriguez, N., Lathem, T. L., Rodriguez-R, L. M., Barazesh, J. M., Anderson, B. E., Beyersdorf, A. J., . . . Konstantinidis, K. T. (2013). Microbiome of the upper troposphere: Species composition and prevalence, effects of tropical storms, and atmospheric implications. *Proceedings of the National Academy of Sciences, 110*(7), 2575-2580. doi:10.1073/pnas.1212089110

Després, V., Nowoisky, J., Klose, M., Conrad, R., Andreae, M., & Pöschl, U. (2007). Characterization of primary biogenic aerosol particles in urban, rural, and high-alpine air by DNA sequence and restriction fragment analysis of ribosomal RNA genes. *Biogeosciences, 4*(6), 1127-1141.

Diallo, M., Ploeger, F., Konopka, P., Birner, T., Müller, R., Riese, M., . . . Jegou, F. (2017). Significant Contributions of Volcanic Aerosols to Decadal Changes in the Stratospheric Circulation. *Geophysical Research Letters, 44*(20). doi:10.1002/2017GL074662

Dimmick, R. L., Straat, P. A., Wolochow, H., & Levin, G. V. (1975). Evidence for metabolic activity of airborne bacteria. *Journal of aerosol science, 6*, 387 - 393.

Dimmick, R. L., & Wolochow, H. (1979). Evidence for more than one division of bacteria within airborne particles. *Applied and Environmental Microbiology, 38*(4), 642 - 643.

Dixon, P. (2003). VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science, 14*(6), 927-930. doi:10.1111/j.1654-1103.2003.tb02228.x

Dybwad, M., Skogan, G., & Blatny, J. (2014). Comparative Testing and Evaluation of Nine Different Air Samplers: End-to-End Sampling Efficiencies as Specific Performance Measurements for Bioaerosol Applications. *Aerosol Science and Technology, 48*(3), 282-295. doi:10.1080/02786826.2013.871501

Edgar, R. C. Tutorials, exercises and example scripts for OTU analysis. Retrieved from https://drive5.com/usearch/manual/tutorials.html

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics, 26*(19), 2460-2461. doi:10.1093/bioinformatics/btq461

Edgar, R. C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods, 10*(10), 996-998. doi:10.1038/nmeth.2604

Edgar, R. C. (2017). Accuracy of microbial community diversity estimated by closed- and open-reference OTUs. *PeerJ, 5*, e3889. doi:10.7717/peerj.3889

Edgar, R. C. (2018). Taxonomy annotation and guide tree errors in 16S rRNA databases. *PeerJ, 6*, e5030. doi:10.7717/peerj.5030

Els, N., Larose, C., Baumann-Stanzer, K., Tignat-Perrier, R., Keuschnig, C., Vogel, T. M., & Sattler, B. (2019). Microbial composition in seasonal time series of free tropospheric air and precipitation reveals community separation. *Aerobiologia, 35*(4), 671-701. doi:10.1007/s10453-019-09606-x

Evans, S., Martiny, J. B., & Allison, S. D. (2017). Effects of dispersal and selection on stochastic assembly in microbial communities. *The ISME Journal, 11*(1), 176-185.

Fang, Z., Ouyang, Z., Zheng, H., Wang, X., & Hu, L. (2007a). Culturable Airborne Bacteria in Outdoor Environments in Beijing, China. *Microbial Ecology, 54*(3), 487-496. doi:10.1007/s00248-007-9216-3

Fang, Z., Ouyang, Z., Zheng, H., Wang, X., & Hu, L. (2007b). Culturable airborne bacteria in outdoor environments in Beijing,China. *Microbial Ecology, 54*(3), 487-496. doi:10.1007/s00248-007-9216-3

Feng, M., Tripathi, B. M., Shi, Y., Adams, J. M., Zhu, Y.-G., & Chu, H. (2019). Interpreting distance-decay pattern of soil bacteria via quantifying the assembly processes at multiple spatial scales. *MicrobiologyOpen, 8*(9), e00851-e00851. doi:10.1002/mbo3.851

Fierer, N., Leff, J. W., Adams, B. J., Nielsen, U. N., Bates, S., Lauber, C. L., . . . Caporaso, G. J. (2012). Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proceedings of the National Academy of Sciences, 109*(52), 21390-21395. doi:10.1073/pnas.1215210110

Fierer, N., Liu, Z., & Rodríguez-Hernández, M. (2008). Short-term temporal variability in airborne bacterial and fungal populations. *Applied and Environmental Microbiology, 74*(1), 200 - 207. doi:10.1128/AEM.01467-07

Fiore, M., C.L, S. A., Azevedo, M., Komárek, J., Kaštovský, J., Sulek, J., & Lorenzi, A. (2007). The cyanobacterial genus Brasilonema, gen. nov., a molecular and phenotypic evaluation. *Journal of Phycology, 43*, 789-798. doi:10.1111/j.1529-8817.2007.00376.x

Fowbert, J. A., & Smith, R. I. L. (1994). Rapid Population Increases in Native Vascular Plants in the Argentine Islands, Antarctic Peninsula. *Arctic and Alpine Research, 26*(3), 290-296. doi:10.1080/00040851.1994.12003068

Fox-Skelly, J. (2017). There are diseases in the ice and they are waking up. Retrieved from http://www.bbc.com/earth/story/20170504-there-are-diseases-hidden-in-ice-and-they-are-waking-up

Franzetti, A., Gandolfi, I., Gaspari, E., Ambrosini, R., & Bestetti, G. (2010). Seasonal variability of bacteria in fine and coarse urban air particulate matter. *Applied microbiology and biotechnology, 90*, 745-753. doi:10.1007/s00253-010-3048-7

Fripiat, F., Meiners, K. M., Vancoppenolle, M., Papadimitriou, S., Thomas, D. N., Ackley, S. F., . . . Delille, B. (2017). Macro-nutrient concentrations in Antarctic pack ice: Overall patterns and overlooked processes. *Elem Sci Anth, 5*(13). doi:10.1525/elementa.217

Fröhlich-Nowoisky, J., Pickersgill, D. A., Després, V. R., & Pöschl, U. (2009). High diversity of fungi in air particulate matter. *Proceedings of the National Academy of Sciences, 106*(31), 12814-12819.

Fukami, T. (2015). Historical contingency in community assembly: integrating niches, species pools, and priority effects. *Annual Review of Ecology, Evolution, and Systematics, 46*, 1-23.

Fulton, J. D. (1966a). Microorganisms of the upper atmosphere. V. Relationship between frontal activity and the micropopulation at altitude. *Appl Microbiol, 14*(2), 245-250.

Fulton, J. D. (1966b). Microorganisms of the Upper Atmosphere: III. Relationship between Altitude and Micropopulation. *Applied Microbiology, 14*(2), 237-240.

Fulton, J. D., & Mitchell, R. B. (1966). Microorganisms of the upper atmosphere. II. Microorganisms in two types of air masses at 690 meters over a city. *Appl Microbiol, 14*(2), 232-236.

Garcia-Alcega, S., Nasir, Z. A., Cipullo, S., Ferguson, R., Yan, C., Whitby, C., . . . Tyrrel, S. (2020). Fingerprinting ambient air to understand bioaerosol profiles in three different environments in the south east of England. *Science of the Total Environment*, 137542.

Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V., & Egozcue, J. J. (2017). Microbiome Datasets Are Compositional: And This Is Not Optional. *Frontiers in Microbiology, 8*(2224). doi:10.3389/fmicb.2017.02224

Google. New Zealand. Retrieved from https://www.google.co.nz/maps/@-36.8816781,174.7696692,13z

Grantham, N. S., Reich, B. J., Pacifici, K., Laber, E. B., Menninger, H. L., Henley, J. B., . . . Dunn, R. R. (2015). Fungi Identify the Geographic Origin of Dust Samples. *PLOS ONE, 10*(4). doi:10.1371/journal.pone.0122605

Green, P. N. (2006). Methylobacterium. In M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer, & E. Stackebrandt (Eds.), *The Prokaryotes: Volume 5: Proteobacteria: Alpha and Beta Subclasses* (pp. 257-265). New York, NY: Springer New York.

Gregory, P. H. (1961). *The microbiology of the atmosphere*.

Griffin, D. W. (2004). Terrestrial microorganisms at an altitude of 20,000 m in Earth's atmosphere. *Aerobiologia, 20*, 135 - 140. doi:10.1023/B:AERO.0000032948.84077.12.

Grolemund, G., & Wickham, H. (2011). Dates and times made easy with lubridate. *Journal of statistical software, 40*(3), 1-25.
Haig, C. W., Mackay, W. G., Walker, J. T., & Williams, C. (2016). Bioaerosol sampling: sampling mechanisms, bioefficiency and field studies. *Journal of Hospital Infection, 93*(3), 242-255. doi:10.1016/j.jhin.2016.03.017
Hara, K., & Zhang, D. (2012). Bacterial abundance and viability in long-range transported dust. *Atmospheric Environment, 47*, 20-25. doi:10.1016/j.atmosenv.2011.11.050
Harrison, R. M., Jones, A. M., Biggins, P. D. E., Pomeroy, N., Cox, C. S., Kidd, S. P., . . . Beswick, A. (2005). Climate factors influencing bacterial count in background air samples. *International journal of Biometeorology, 49*(3), 167 - 178. doi:10.1007/s00484-004-0225-3
Herbold, C. W., Lee, C. K., McDonald, I. R., & Cary, C. S. (2014). Evidence of global-scale aeolian dispersal and endemism in isolated geothermal microbial communities of Antarctica. *Nature Communications, 5*. doi:10.1038/ncomms4875
Hermans, S. M., Buckley, H. L., & Lear, G. (2018). Optimal extraction methods for the simultaneous analysis of DNA from diverse organisms and sample types. *Molecular Ecology Resources, 18*(3), 557-569. doi:https://doi.org/10.1111/1755-0998.12762
Horneck, G. (1993). Responses of Bacillus subtilis spores to space environment: results from experiments in space. *Origins of life and evolution of the biosphere : the journal of the International Society for the Study of the Origin of Life, 23*(1), 37-52. doi:10.1007/BF01581989
Horowitz, N., Cameron, R. E., & Hubbard, J. S. (1972). Microbiology of the dry valleys of Antarctica. *Science, 176*(4032), 242-245.
Huffman, J. A., Treutlein, B., & Poschl, U. (2010). Fluorescent biological aerosol particle concentrations and size distributions measured with an Ultraviolet Aerodynamic Particle Sizer (UV-APS) in Central Europe. *Atmospheric Chemistry and Physics, 10*, 3215 - 3233. doi:10.5194/acp-10-3215-2010
Hughes, K. A. (2003). Aerial dispersal and survival of sewage-derived faecal coliforms in Antarctica. *Atmospheric Environment, 37*(22), 3147-3155. doi:10.1016/S1352-2310(03)00207-3
Hughes, K. A., McCartney, H. A., Lachlan-Cope, T. A., & Pearce, D. A. (2004). A preliminary study of airborne microbial biodiversity over Peninsular Antarctica. *Cell Mol Biol (Noisy-le-grand), 50*(5), 537-542.
Imshenetsky, A. A., Lysenko, S. V., & Kazakov, G. A. (1978). Upper boundary of the biosphere. *Appl Environ Microbiol, 35*(1), 1-5.
Invitrogen. (2010). Qubit Assays - quick reference card. Retrieved from https://www.ieg.uu.se/digitalAssets/176/c_176882-l_3-k_qubitquickrefcard.pdf
Jackson, B. (2016). Clean-up using AMPure XP beads. Retrieved from https://www.protocols.io/view/clean-up-using-ampure-xp-beads-f3ebqje
Jaing, C., Thissen, J., Morrison, M., Dillon, M. B., Waters, S. M., Graham, G. T., . . . Smith, D. J. (2020). Sierra Nevada sweep: metagenomic measurements of bioaerosols vertically distributed across the troposphere. *Scientific Reports, 10*(1), 12399. doi:10.1038/s41598-020-69188-4
Jamriska, M., DuBois, T. C., & Skvortsov, A. (2012). Statistical characterisation of bio-aerosol background in an urban environment. *Atmospheric Environment, 54*, 439-448.
Jiang, W., Liang, P., Wang, B., Fang, J., Lang, J., Tian, G., . . . Zhu, T. F. (2015). Optimized DNA extraction and metagenomic sequencing of airborne microbial communities. *Nature Protocols, 10*(5), 768-779. doi:10.1038/nprot.2015.046
Kaidor. (2013). Earth Global Circulation. In https://en.wikipedia.org/wiki/Atmospheric_circulation#/media/File:Earth_Global_Circulation_-_en.svg (Ed.), *Inkscape*. Wikipedia.
Karl, D. M., Bird, D. F., Björkman, K., Houlihan, T., Shackelford, R., & Tupas, L. (1999). Microorganisms in the accreted ice of Lake Vostok, Antarctica. *Science, 286*(5447), 2144 - 2147. doi:10.1126/science.286.5447.2144

Kellogg, C. A., & Griffin, D. W. (2006). Aerobiology and the global transport of desert dust. *Trends in Ecology & Evolution, 21*(11), 638-644. doi:10.1016/j.tree.2006.07.004

Kellogg, C. A., Griffin, D. W., Garrison, V. H., Peak, K. K., Royall, N., Smith, R. R., & Shinn, E. A. (2004). Characterization of aerosolized bacteria and fungi from desert dust events in Mali, West Africa. *Aerobiologia, 20*(2), 99-110. doi:10.1023/B:AERO.0000032947.88335.bb

Kennedy, A. D. (1994). Simulated climate change: a field manipulation study of polar microarthropod community response to global warming. *Ecography, 17*(2), 131-140. doi:doi:10.1111/j.1600-0587.1994.tb00085.x

Kennicutt, M. C., Chown, S. L., Cassano, J. J., Liggett, D., Massom, R., Peck, L. S., . . . Sutherland, W. J. (2014). Polar research: Six priorities for Antarctic science. *Nature News, 512*(7512), 23. doi:10.1038/512023a

Kevorkian, R., Bird, J. T., Shumaker, A., & Lloyd, K. G. (2018). Estimating Population Turnover Rates by Relative Quantification Methods Reveals Microbial Dynamics in Marine Sediment. *Appl Environ Microbiol, 84*(1). doi:10.1128/aem.01443-17

Kobayashi, F., Maki, T., Kakikawa, M., Noda, T., Mitamura, H., Takahashi, A., . . . Iwasaka, Y. (2016). Atmospheric bioaerosols originating from Adélie penguins (Pygoscelis adeliae): Ecological observations of airborne bacteria at Hukuro Cove, Langhovde, Antarctica. *Polar Science, 10*. doi:10.1016/j.polar.2015.12.002

Kussell, E., Kishony, R., Balaban, N. Q., & Leibler, S. (2005). Bacterial persistence: a model of survival in changing environments. *Genetics, 169*, 1807 - 1814. doi:10.1534/genetics.104.035352

Lagomarsino Oneto, D., Golan, J., Mazzino, A., Pringle, A., & Seminara, A. (2020). Timing of fungal spore release dictates survival during atmospheric transport. *Proceedings of the National Academy of Sciences, 117*(10), 5134-5143. doi:10.1073/pnas.1913752117

Lai, C., Cheng, A., Liu, W., Tan, C., Huang, Y., Chung, K., . . . Hsueh, P. (2011). Infections caused by unusual Methylobacterium species. *Journal of clinical microbiology, 49*(9), 3329-3331. doi:10.1128/JCM.01241-11

Latombe, G., McGeoch, M. A., Nipperess, D. A., & Hui, C. (2018). zetadiv: an R package for computing compositional change across multiple sites, assemblages or cases. *bioRxiv*, 324897.

Laybourn-Parry, J. (2002). Survival mechanisms in Antarctic lakes. *Philosophical Transactions of the Royal Society B: Biological Sciences, 357*, 863 - 869. doi:10.1098/rstb.2002.1075

Lear, G., Washington, V., Neale, M., Case, B., Buckley, H., & Lewis, G. (2013). The biogeography of stream bacteria. *Global Ecology and Biogeography*(22), 544-554. doi:10.1111/geb.12046

Lee, C. K., Barbier, B. A. A., Bottos, E. M., McDonald, I. R., & Cary, S. C. (2012). The Inter-Valley Soil Comparative Survey: the ecology of Dry Valley edaphic microbial communities. *The ISME Journal, 6*(5), 1046-1057. doi:10.1038/ismej.2011.170

Lelieveld, J., & Heintzenberg, J. (1992). Sulfate cooling effect on climate through in-cloud oxidation of anthropogenic SO2. *Science, 258*(5079), 117 - 120. doi:10.1126/science.258.5079.117

Lennon, J. T., & Jones, S. E. (2011). Microbial seed banks: the ecological and evolutionary implications of dormancy. *Nature reviews. Microbiology, 9*(2), 119-130. doi:10.1038/nrmicro2504

Li. (2018). hillR: taxonomic, functional, and phylogenetic diversity and similarity through Hill Numbers. *Journal of Open Source Software, 3*(31). doi:https://doi.org/10.21105/joss.01041

Lighthart, B., & Stetzenbach, L. D. (1994). Distribution of Microbial Bioaerosol. In B. Lighthart & A. J. Mohr (Eds.), *Atmospheric Microbial Aerosols: Theory and Applications* (pp. 68-98). Boston, MA: Springer US.

London, I. (2017). Encoding cyclical continuous features - 24-hour time. Retrieved from https://ianlondon.github.io/blog/encoding-cyclical-features-24hour-time/

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology, 15*(12), 550. doi:10.1186/s13059-014-0550-8

Luhung, I., Wu, Y., Ng, C., Miller, D., Cao, B., & Chang, V. (2015). Protocol Improvements for Low Concentration DNA-Based Bioaerosol Sampling and Analysis. *PLOS ONE, 10*(11). doi:10.1371/journal.pone.0141158

Makarova, K. S., Aravind, L., Wolf, Y. I., Tatusov, R. L., Minton, K. W., Koonin, E. V., & Daly, M. J. (2001). Genome of the extremely radiation-resistant bacterium Deinococcus radiodurans viewed from the perspective of comparative genomics. *Microbiology and molecular biology reviews, 65*(1), 44-79. doi:10.1128/MMBR.65.1.44-79.2001

Maki, T., Aoki, K., Kobayashi, F., Kakikawa, M., Tobo, Y., Matsuki, A., . . . Iwasaka, Y. (2011). Characterization of halotolerant and oligotrophic bacterial communities in Asian desert dust (KOSA) bioaerosol accumulated in layers of snow on Mount Tateyama Central Japan. *Aerobiologia, 27*, 277 - 290. doi:10.1007/s10453-011-9196-0

Maki, T., Hara, K., Iwata, A., Lee, K. C., Kawai, K., Kai, K., . . . Iwasaka, Y. (2017). Variations in airborne bacterial communities at high altitudes over the Noto Peninsula (Japan) in response to Asian dust events. *Atmospheric Chemistry and Physics, 17*(19), 11877-11897. doi:10.5194/acp-17-11877-2017

Maki, T., Susuki, S., Kobayashi, F., Kakikawa, M., Yamada, M., Higashi, T., . . . Tobo, Y. (2008). Phylogenetic diversity and vertical distribution of a halobacterial community in the atmosphere of an Asian dust (KOSA) source region, Dunhuang City. *Air Quality, Atmosphere & Health, 1*(2), 81-89. doi:10.1007/s11869-008-0016-9

Mardis, E. R. (2013). Next-Generation Sequencing Platforms. *Annual Review of Analytical Chemistry, 6*(1), 287-303. doi:10.1146/annurev-anchem-062012-092628

Marizcurrena, J. J., Morel, M. A., Braña, V., Morales, D., Martinez-López, W., & Castro-Sowinski, S. (2017). Searching for novel photolyases in UVC-resistant Antarctic bacteria. *Extremophiles, 21*(2), 409-418. doi:10.1007/s00792-016-0914-y

Marshall, W. A. (1996a). Aerial dispersal of lichen soredia in the maritime Antarctic. *New Phytologist, 134*(3), 523-530. doi:10.1111/j.1469-8137.1996.tb04370.x

Marshall, W. A. (1996b). Biological particles over Antarctica. *Nature, 383*(6602), 680-680. doi:10.1038/383680a0

Marshall, W. A. (1997). Seasonality in antarctic airborne fungal spores. *Applied and Environmental Microbiology, 63*(6), 2240 - 2245.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011, 17*(1), 3. doi:10.14806/ej.17.1.200

Mathieu, C., Hermans, S. M., Lear, G., Buckley, T. R., Lee, K. C., & Buckley, H. L. (2020). A Systematic Review of Sources of Variability and Uncertainty in eDNA Data for Environmental Monitoring. *Frontiers in Ecology and Evolution, 8*(135). doi:10.3389/fevo.2020.00135

Matsuki, A., Iwasaka, Y., & Osada, K. (2003). Seasonal dependence of the long-range transport and vertical distribution of free tropospheric aerosols over east Asia: On the basis of aircraft and lidar measurements and isentropic trajectory analysis. *Journal of Geophysical Research, 108*(D23 8663). doi:10.1029/2002JD003266

Matthias-Maser, S., Obolkin, V., Khodzer, T., & Jaenicke, R. (2000). Seasonal variation of primary biological aerosol particles in the remote continental region of Lake Baikal/Siberia. *Atmospheric Environment, 34*(22), 3805-3811. doi:10.1016/S1352-2310(00)00139-4

McCune, B., Grace, J. B., & Urban, D. L. (2002). *Analysis of ecological communities* (Vol. 28): MjM software design Gleneden Beach, OR.

McKnight, D. T., Huerlimann, R., Bower, D. S., Schwarzkopf, L., Alford, R. A., & Zenger, K. R. (2019). microDecon: A highly accurate read-subtraction tool for the post-sequencing removal of contamination in metabarcoding studies. *Environmental DNA, 1*(1), 14-25. doi:10.1002/edn3.11

McMurdie, P. J., & Holmes, S. (2013). phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLOS ONE, 8*(4), e61217. doi:10.1371/journal.pone.0061217

Mhuireach, G. Á., Wilson, H., & Johnson, B. R. (2020). Urban Aerobiomes are Influenced by Season, Vegetation, and Individual Site Characteristics. *EcoHealth*. doi:10.1007/s10393-020-01493-w

Miller, L. M., Gans, F., & Kleidon, A. (2011). Jet stream wind power as a renewable energy resource: little power, big impacts. *Earth Syst. Dynam., 2*(2), 201-212. doi:10.5194/esd-2-201-2011

Mohr. (2007). *Manual of Environmental Microbiology* (Third Edition ed.).

Montero, A., Dueker, M. E., & O'Mullan, G. D. (2016). Culturable bioaerosols along an urban waterfront are primarily associated with coarse particles. *PeerJ, 4*, e2827.

Mroz, E., Mohammed, A., Cappis, J. H., Guthals, P. R., Mason, A. S., & Rokop, D. J. (1989). Antarctic atmospheric tracer experiments. *Journal of Geophysical Research: Atmospheres, 94*(D6), 8577-8583. doi:doi:10.1029/JD094iD06p08577

NASA. (1962). *US Standard Atmosphere 1962*. Retrieved from https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19630003300.pdf

Nguyen, T. M. N., Ilef, D., Jarraud, S., Rouil, L., & Desenclos, J.-C. (2006). A community-wide outbreak of legionnaires disease linked to industrial cooling towers—how far can contaminated aerosols spread? *Journal of Infectious Diseases, 193*, 102 - 111. doi:10.1093/oxfordjournals.aje.a112917

Niklas, K. J. (1985). The aerodynamics of wind pollination. *The Botanical Review, 51*(3), 328. doi:10.1007/bf02861079

Nkem, J. N., Wall, D. H., Virginia, R. A., Barrett, J. E., Broos, E. J., Porazinska, D. L., & Adams, B. J. (2006). Wind dispersal of soil invertebrates in the McMurdo Dry Valleys, Antarctica. *Polar Biology, 29*(4), 346-352. doi:10.1007/s00300-005-0061-x

NOAA. (2021). HYSPLIT tutorial 6.3 Clustering Equations. Retrieved from https://www.ready.noaa.gov/documents/Tutorial/html/traj_cluseqn.html

Nuwer, R. (2014). The last place on Earth without life. Retrieved from http://www.bbc.com/future/story/20140303-last-place-on-earth-without-life

O'Connell, S. (2006). How Krakatoa made the biggest bang. Retrieved from https://www.independent.co.uk/news/science/how-krakatoa-made-the-biggest-bang-5336165.html

O'Malley, M. A. (2008). 'Everything is everywhere: but the environment selects': ubiquitous distribution and ecological determinism in microbial biogeography. *Stud Hist Philos Biol Biomed Sci, 39*(3), 314-325. doi:10.1016/j.shpsc.2008.06.005

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P., O'Hara, R. B., . . . Wagner, H. (2012). Vegan: Community Ecology Package. R package version 2.0-2.

Ormeño-Orrillo, E., & Martínez-Romero, E. (2019). A Genomotaxonomy View of the Bradyrhizobium Genus. *Frontiers in Microbiology, 10*, 1334-1334. doi:10.3389/fmicb.2019.01334

Palarea-Albaladejo, J., & Martín-Fernández, J. A. (2015). zCompositions — R package for multivariate imputation of left-censored data under a compositional approach. *Chemometrics and Intelligent Laboratory Systems, 143*, 85-96. doi:https://doi.org/10.1016/j.chemolab.2015.02.019

Parish, T. R., & Cassano, J. J. (2003). The Role of Katabatic Winds on the Antarctic Surface Wind Regime. *Monthly Weather Review, 131*(2), 317-333. doi:10.1175/1520-0493(2003)131<0317:Trokwo>2.0.Co;2

Parish, T. R., & J., C. J. (2003). The role of katabatic winds on the Antarctic surface wind regime. *The role of katabatic winds on the Antarctic surface wind regime*.

Parks, D. H., Rinke, C., Chuvochina, M., Chaumeil, P.-A., Woodcroft, B. J., Evans, P. N., . . . Tyson, G. W. (2017). Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nature Microbiology, 2*(11), 1533-1542. doi:10.1038/s41564-017-0012-7

Pearce, D. A., Alekhina, I. A., Terauds, A., Wilmotte, A., Quesada, A., Edwards, A., . . . Vincent, W. F. (2016). Aerobiology Over Antarctica – A New Initiative for Atmospheric Ecology. *Frontiers in Microbiology, 7*, 16. doi:10.3389/fmicb.2016.00016

Pearce, D. A., Bridge, P. D., Hughes, K. A., Sattler, B., Psenner, R., & Russell, N. J. (2009). Microorganisms in the atmosphere over Antarctica. *FEMS Microbiology Ecology, 69*(2), 143-157. doi:10.1111/j.1574-6941.2009.00706.x

Pearce, D. A., Hughes, K. A., Lachlan-Cope, T., Harangozo, S. A., & Jones, A. E. (2010). Biodiversity of air-borne microorganisms at Halley Station, Antarctica. *Extremophiles : life under extreme conditions, 14*(2), 145-159. doi:10.1007/s00792-009-0293-8

Pepper, I., Gerber, C., Gentry, T. (Ed.) (2015). *Environmental Microbiology* ((third edition) ed.).

Pfeiffer, F., Gröber, C., Blank, M., Händler, K., Beyer, M., Schultze, J. L., & Mayer, G. (2018). Systematic evaluation of error rates and causes in short samples in next-generation sequencing. *Scientific Reports, 8*(1), 10950. doi:10.1038/s41598-018-29325-6

Pointing, S. B., Büdel, B., Convey, P., Gillman, L. N., Körner, C., Leuzinger, S., & Vincent, W. F. (2015). Biogeography of photoautotrophs in the high polar biome. *Frontiers in Plant Science, 6*, 692. doi:10.3389/fpls.2015.00692

Pointing, S. B., Fierer, N., Smith, G. J. D., Steinberg, P. D., & Wiedmann, M. (2016). Quantifying human impact on Earth's microbiome. *Nature Microbiology, 1*(9), 16145. doi:10.1038/nmicrobiol.2016.145

Priyamvada, H., Singh, R. K., Akila, M., Ravikrishna, R., Verma, R. S., & Gunthe, S. S. (2017). Seasonal variation of the dominant allergenic fungal aerosols - One year study from southern Indian region. *Sci Rep, 7*(1), 11171. doi:10.1038/s41598-017-11727-7

Prospero, J. M. (1999). Assessing the impact of advected African dust on air quality and health in the eastern United States. *Human and Ecological Risk Assessment: An International Journal, 5*(3), 471-479. doi:10.1080/10807039.1999.10518872

Qiagen. (2017). DNeasy PowerSoil Kit Handbook. Retrieved from https://www.qiagen.com/us/resources/resourcedetail?id=5a0517a7-711d-4085-8a28-2bb25fab828a&lang=en

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., . . . Glöckner, F. O. (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research, 41*(D1), D590-D596. doi:10.1093/nar/gks1219

Quensen, J. (2018). "RDPutils: R Utilities for processing RDPTool output," in R package version 1.4.1 ed. Retrieved from https://github.com/jfq3/RDPutils.

Quinn, T. P., Erb, I., Gloor, G., Notredame, C., Richardson, M. F., & Crowley, T. M. (2019). A field guide for the compositional analysis of any-omics data. *GigaScience, 8*(9). doi:10.1093/gigascience/giz107

R Development Core Team. (2010). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistcal Computing. Retrieved from https://www.r-project.org/

Ranjan, C. (2020). *Package 'nlcor': Compute Nonlinear Correlations*.

Reche, I., D'Orta, G., Mladenov, N., Winget, D. M., & Suttle, C. A. (2018). Deposition rates of viruses and bacteria above the atmospheric boundary layer. *The ISME Journal, 12*(4), 1154-1162. doi:10.1038/s41396-017-0042-4

Rettberg, P., Eschweiler, U., Strauch, K., Reitz, G., Horneck, G., Wänke, H., . . . Barbier, B. (2002). Survival of microorganisms in space protected by meteorite material: results of the experiment EXOBIOLOGIE of the PERSEUS mission. *Advances in Space Research, 30*(6), 1539-1545. doi:10.1016/S0273-1177(02)00369-1

Rotach, M. W., Gohm, A., Lang, M. N., Leukauf, D., Stiperski, I., & Wagner, J. S. (2015). On the Vertical Exchange of Heat, Mass, and Momentum Over Complex, Mountainous Terrain. *Frontiers in Earth Science, 3*, 76. doi:10.3389/feart.2015.00076

Rothschild, L. J., & Mancinelli, R. L. (2001). Life in extreme environments. *Nature, 409*, 1092 - 1101. doi:10.1038/35059215

Šantl-Temkiv, T., Sikoparija, B., Maki, T., Carotenuto, F., Amato, P., Yao, M., . . . Huffman, J. A. (2020). Bioaerosol field measurements: Challenges and perspectives in outdoor studies. *Aerosol Science and Technology, 54*(5), 520-546. doi:10.1080/02786826.2019.1676395

Sattler, B., Puxbaum, H., & Psenner, R. (2001). Bacterial growth in supercooled cloud droplets. *Geophysical Research Letters, 28*(2), 239 - 242. doi:10.1029/2000GL011684

Shaffer, B. T., & Lighthart, B. (1997). Survey of culturable airborne bacteria at four diverse locations in Oregon: urban, rural, forest, and coastal. *Microbial Ecology, 34*, 167 - 177. doi:10.1007/s002489900046

Shigyo, N., Umeki, K., & Hirao, T. (2019). Seasonal Dynamics of Soil Fungal and Bacterial Communities in Cool-Temperate Montane Forests. *Frontiers in Microbiology, 10*(1944). doi:10.3389/fmicb.2019.01944

Sinha, R. K., & Krishnan, K. P. (2013). Novel opportunity for understanding origin and evolution of life: perspectives on the exploration of subglacial environment of Lake Vostok, Antarctica. *Annals of microbiology, 63*(2), 409 - 415. doi:10.1007/s13213-012-0525-5

Sinha, R. P., & Häder, D.-P. (2002). UV-induced DNA damage and repair: a review. *Photochemical & Photobiological Sciences, 1*(4), 225-236. doi:10.1039/B201230H

Smith, D. J., Griffin, D. W., McPeters, R. D., Ward, P. D., & Schuerger, A. C. (2011). Microbial survival in the stratosphere and implications for global dispersal. *Aerobiologia, 27*(4), 319-332. doi:10.1007/s10453-011-9203-5

Smith, D. J., Ravichandar, J. D., Jain, S., Griffin, D. W., Yu, H., Tan, Q., . . . McGrath, J. (2018). Airborne Bacteria in Earth's Lower Stratosphere Resemble Taxa Detected in the Troposphere: Results From a New NASA Aircraft Bioaerosol Collector (ABC). *Frontiers in Microbiology, 9*(1752). doi:10.3389/fmicb.2018.01752

Smith, L. R. I. (1991). Exotic sporomorpha as indicators of potential immigrant colonists in Antarctica. *Grana, 30*(2), 313-324. doi:10.1080/00173139109431986

Smith, R. I. L. (1994). Vascular plants as bioindicators of regional warming in Antarctica. *Oecologia, 99*(3), 322-328. doi:10.1007/bf00627745

Sokol, E. R., Herbold, C. W., Lee, C. K., Cary, S. C., & Barrett, J. E. (2013). Local and regional influences over soil microbial metacommunities in the Transantarctic Mountains. *Ecosphere, 4*(11), 136. doi:10.1890/ES13-00136.1

Sommaruga, R., & Casamayor, E. O. (2009). Bacterial 'cosmopolitanism'and importance of local environmental factors for community composition in remote high-altitude lakes. *Freshwater Biology, 54*, 944 - 1005. doi:10.1111/j.1365-2427.2008.02146.x

Stein, A. F., Draxler, R. R., Rolph, G. D., Stunder, B. J. B., Cohen, M. D., & Ngan, F. (2016). NOAA's HYSPLIT Atmospheric Transport and Dispersion Modeling System. *Bulletin of the American Meteorological Society, 96*(12), 2059-2077. doi:10.1175/bams-d-14-00110.1

Tanaka, D., Fujiyoshi, S., Maruyama, F., Goto, M., Koyama, S., Kanatani, J.-i., . . . Kagaya, S. (2020). Size resolved characteristics of urban and suburban bacterial bioaerosols in Japan as assessed by 16S rRNA amplicon sequencing. *Scientific Reports, 10*(1), 1-10.

Templ, M., Hron, K., & Filzmoser, P. (2011). robCompositions: An R-package for Robust Statistical Analysis of Compositional Data. In (pp. 341-355).

Tong, & Lighthart. (1999). Diurnal distribution of total and culturable atmospheric bacteria at a rural site. *Aerosol Science and Technology*. doi:10.1080/027868299304822

Tong, Y., & Lighthart, B. (2000). The annual bacterial particle concentration and size distribution in the ambient atmosphere in a rural area of the Willamette Valley, Oregon. *Aerosol Science & Technology, 32*(5), 393 - 403. doi:10.1080/027868200303533

Trout-Haney, J. V., Heindel, R. C., & Virginia, R. A. (2020). Picocyanobacterial cells in near-surface air above terrestrial and freshwater substrates in Greenland and Antarctica. *Environmental Microbiology Reports, 12*(3), 296-305. doi:https://doi.org/10.1111/1758-2229.12832

Turner, J., Colwell, S. R., Marshall, G. J., Lachlan-Cope, T. A., Carleton, A. M., Jones, P. D., . . . Iagovkina, S. (2005). Antarctic climate change during the last 50 years. *International Journal of Climatology, 25*(3), 279-294. doi:doi:10.1002/joc.1130

Uetake, J., Tobo, Y., Uji, Y., Hill, T. C. J., DeMott, P. J., Kreidenweis, S. M., & Misumi, R. (2019). Seasonal Changes of Airborne Bacterial Communities Over Tokyo and Influence of Local Meteorology. *Frontiers in Microbiology, 10*(1572). doi:10.3389/fmicb.2019.01572

UN. (2018). 68% of the world population projected to live in urban areas by 2050, says UN. Retrieved from https://www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanization-prospects.html

Vaïtilingom, M., & Deguillaume, L. (2013). Potential impact of microbial activity on the oxidant capacity and organic carbon budget in clouds. *Proceedings of the National Academy of Sciences, 110*(2), 59 - 564. doi:10.1073/pnas.1205743110

Van Rossum, G. a. D., Fred L. (2009). *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace.

Vincent, W. F. (2000). Evolutionary origins of Antarctic microbiota: invasion, selection and endemism. *Antarctic Science, 12*(3), 374 - 385. doi:10.1017/S0954102000000420

Wainwright, M., Wickramasinghe, N. C., Narlikar, J. V., & P., R. (2003). Microorganisms cultured from stratospheric air samples obtained at 41 km. *FEMS Microbiology Letters, 218*, 161 - 165. doi:10.1111/j.1574-6968.2003.tb11513.x

Wang, M., Liu, F., Crous, P. W., & Cai, L. (2017). Phylogenetic reassessment of Nigrospora: Ubiquitous endophytes, plant and human pathogens. *Persoonia, 39*, 118-142. doi:10.3767/persoonia.2017.39.06

Waugh, J. B., Granger, W. M., & Gaggar, A. (2010). Incidence, relevance and response for Ralsfonia respiratory infections. *Clinical laboratory science : journal of the American Society for Medical Technology, 23*(2), 99-106.

Weisburg, W., Tully, J., Rose, D., Petzel, J., Oyaizu, H., Yang, D., . . . Van Etten, J. (1989). A phylogenetic analysis of the mycoplasmas: basis for their classification. *Journal of bacteriology, 171*(12), 6455-6467.

Weisleitner, K., Perras, A., Moissl-Eichinger, C., Andersen, D. T., & Sattler, B. (2019). Source Environments of the Microbiome in Perennially Ice-Covered Lake Untersee, Antarctica. *Frontiers in Microbiology, 10*(1019). doi:10.3389/fmicb.2019.01019

Whon, T. W., Kim, M. S., Roh, S. W., Shin, N. R., & Lee, H. W. (2012). Metagenomic characterization of airborne viral DNA diversity in the near-surface atmosphere. *Journal of …*. doi:10.1128/JVI.00293-12

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*: Springer-Verlag New York.

Wilson, S. L., & Walker, V. K. (2010). Selection of low temperature resistance in bacteria and potential applications. *Environmental technology, 31*(8-9), 943-956. doi:10.1080/09593331003782417

Womack, A. M., Bohannan, B. J. M., & Green, J. L. (2010). Biodiversity and biogeography of the atmosphere. *Philosophical Transactions of the Royal Society B: Biological Sciences, 365*(1558), 3645-3653. doi:10.1098/rstb.2010.0283

Woo, A. C., Brar, M. S., Chan, Y., Lau, M., Leung, F., Scott, J. A., . . . Pointing, S. B. (2013). Temporal variation in airborne microbial populations and microbially-derived allergens in a tropical urban landscape. *Atmospheric Environment, 74*, 291-300. doi:10.1016/j.atmosenv.2013.03.047

Wood, S. A., Rueckert, A., Cowan, D. A., & Cary, C. S. (2008). Sources of edaphic cyanobacterial diversity in the Dry Valleys of Eastern Antarctica. *The ISME Journal, 2*(3), 308-320. doi:10.1038/ismej.2007.104

Wynn-Williams, D. D. (1991). Aerobiology and colonization in Antarctica — the BIOTAS Programme. *Grana, 30*(2), 380-393. doi:10.1080/00173139109431994

Wynn-Williams, D. D. (1996). Antarctic microbial diversity: the basis of polar ecosystem processes. *Biodiversity and Conservation, 5*(11), 1271 - 1293. doi:10.1007/BF00051979

Yoo, K., Lee, T., Choi, E., Yang, J., Shukla, S., Hwang, S.-i., & Park, J. (2016). Approach of molecular methods for the detection and monitoring of microbial communities in bioaerosols: A review. *Journal of Environmental Sciences*. doi:10.1016/j.jes.2016.07.002

Zhou, J., & Ning, D. (2017). Stochastic Community Assembly: Does It Matter in Microbial Ecology? *Microbiology and molecular biology reviews, 81*(4), e00002-00017. doi:10.1128/mmbr.00002-17