Low-biomass human microbiomes: another piece to the puzzle for non-communicable diseases

Caitlin Alyssa Selway

Department of Molecular and Biomedical Science
School of Biological Sciences
Faculty of Sciences
University of Adelaide

THE UNIVERSITY of ADELAIDE

This thesis is submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

February 2021

“We cannot fathom the marvelous complexity of an organic being; but on the hypothesis here advanced this complexity is much increased. Each living creature must be looked at as a microcosm—a little universe, formed of a host of self-propagating organisms, inconceivably minute and as numerous as the stars in heaven.”

Charles Darwin

Table of Contents

Thesis Abstract

The prevalence of non-communicable diseases (NCDs), such as cancers, cardiovascular, respiratory, and autoimmune diseases, has been increasing since the 1950s. Genetic, environmental and lifestyle factors, including diet, smoking status, and urbanisation, have all been identified as significant contributors to NCDs. More recently, the microbial communities of the human body (microbiota) have also been linked to NCDs. These communities typically exist in a mutually beneficial relationship with their human host, performing critical functions that the human body cannot perform itself. However, an imbalance to these communities may be a causal or perpetuating factor in diseases. While new research has started to unravel the interactions and effects that microbiota have on the human host, the majority of these studies have been focused primarily on the gut, with other body sites remaining neglected.

Understanding microbiota of body sites other than the gut may provide further insight into the cause and effect of NCDs. These 'non-gut' microbiota still play vital roles for the human host, such as defence against pathogens on the skin, or homeostasis of the mouth to prevent or reduce caries and periodontal diseases. However, most microbiota technologies were developed to study the gut, a body site rich in microbial biomass. Hence, application of these technologies to samples from low biomass body sites is difficult due to overwhelming background levels of DNA and contamination. Nevertheless, sufficient information can be obtained from low microbial biomass samples when treated appropriately, and they provide another layer to unravelling the causes of NCDs.

This thesis provides new perspectives on NCDs through the investigation of low microbial biomass body sites. I advocate for the human microbiome (microbiota, their genetic material and surrounding micro-environment) to be used as a new tool in pathology to understand both communicable and NCDs, while also highlighting techniques that can be used to mitigate contamination of low microbial biomass samples. Using a systems biology approach in combination with the microbiome provides a holistic approach to understanding NCDs, which I further explore through a perspectives piece. I then consider NCDs from a 'non-gut'

microbiota perspective across three different studies: I track changes to skin and nasal microbiota after exposure to urban green spaces to improve the understanding of human-environmental interactions and the importance this has for immune-mediated diseases; I investigate the development of oral and lung microbiota in preterm infants and provide insights of a disruption to oral microbiota development in these infants, which can have long-lasting impacts on the immune system; and finally, I trace changes in oral microbiota of children with type 1 diabetes and hyperlipidaemic parents, which shows that changes to fat metabolism in the gut may have repercussions on oral microbiota. Through these three case studies, I provide a deeper understanding on how 'non-gut' microbiota change in response to the environment, which is especially critical in the microbiome development throughout immune training and the prevention of NCDs.

Overall, this thesis provides the groundwork for a holistic approach to understand NCDs. Moving forward, considering the relationships between host genetics, the environment, and microbiota of all body sites will be vital for the treatment, cure, and prevention of NCDs.

Thesis Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I acknowledge that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the
Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program
Scholarship

Caitlin Selway
18th February 2021

Publications

Published articles

Selway C.A., Eisenhofer R. & Weyrich L.S. (2020) Microbiome applications for pathology: challenges of low microbial biomass samples during diagnostic testing, *The Journal of Pathology: Clinical Research,* 6:97-106; https://doi.org/10.1002/cjp2.151 (**Chapter I**)

Selway C.A., Mills J.G., Weinstein P., Skelly C., Yadav S., Lowe A., Breed M.F. & Weyrich L.S. (2020) Transfer of environmental microbes to the skin and respiratory tract of humans after urban green space exposure, *Environment International,* 145:106084; https://doi.org/10.1016/j.envint.2020.106084 (**Chapter III**)

Weyrich L.S., Farrer A.G., Eisenhofer R., Arriola L.A., Young J., Selway C.A., Handsley-Davis M., Adler C.J., Breen J. & Cooper A. (2019) Laboratory contamination over time during low-biomass sample analysis, *Molecular Ecology Resources*, 19:982–996; https://doi.org/10.1111/1755-0998.13011 (**Appendix IV**)

Mills J.G., Bissett A., Gellie N.J.C., Lowe A.J., Selway C.A., Thomas T., Weinstein P., Weyrich L.S. & Breed M.F. (2020) Revegetation of urban green space rewilds soil microbiotas with implications for human health and urban design, *Restoration Ecology*, 28(S4):322-334; https://doi.org/10.1111/rec.13175 (**Appendix V**)

Jensen E.D., Selway C.A., Allen G., Bednarz J., Weyrich L.S., Gue S., Peña A.S. & Couper J. (2020) Early markers of periodontal disease and altered oral microbiota are associated with glycemic control in children with type 1 diabetes. Accepted with minor revisions at *Pediatric Diabetes*, 2020:1-8; https://doi.org/10.1111/pedi.131708 (**Appendix VI)**

Articles in preparation

Selway C.A., Sudarpa J. & Weyrich L.S. Moving beyond the gut microbiome: combining systems biology and multi-site microbiome analyses to combat non-communicable diseases (in prep for The Lancet Microbe; **Chapter II**)

Selway C.A., Collin C.T., Fink N., Makrides M., Gibson R.A., Sullivan T.R., NERO Steering Committee & Weyrich L.S. Initially disrupted preterm infant oral microbiota diversity is restored within three months (in prep for JAMA Pediatrics; **Chapter IV**)

Selway C.A., Jensen E.D., Allen G. & Weyrich L.S. Altered oral microbiota in type 1 diabetic children with hyperlipidemic parents (in prep for Journal of Dental Research; **Chapter V**)

Mills, J.G., Selway, C.A., Weyrich, L.S., Young, J., Thomas, T., Marczylo, E., Skelly C., Weinstein, P., Yadav S., Yadav, V., Lowe, A.J. & Breed, M.F. Rare genera define urban green space soil bacterial communities in three cities across the world. (in prep for Environmental Research)

Bajic, J.E., Howarth G.S., Selway C.A., Weyrich L.S. & Hutchinson M.R. "Little Brain" and "Big Brain" Consequences in 5-Fluorouracil-induced Gut Toxicity and Indomethacin-induced Enteropathy; Microbiota and Neuroimmune implications (in prep)

Acknowledgements

I am truly grateful for participants, their families, and all funding bodies for making this research possible. I sincerely hope this research can improve the quality of lives for all people, big and small, young and old.

A massive thank you to my primary supervisor, Assoc. Prof. Laura Weyrich. You have believed in my abilities and potential from that first meeting in your office. Thank you for helping me to grow as a scientist and a person, and providing me with invaluable opportunities to kick-start my career. I am extremely lucky that I get to carry the skills you have taught me and pass them on to the next generation of scientists. A further thank you to my supervisors both past (Prof. Alan Cooper & Dr. Martin Breed) and present (Dr. Raphael Eisenhofer), for guiding and assisting me through the years. Each of these studies would not have been possible without an extensively network of people from around the globe. I thank each and every one of you, with a special shout-out to Jacob Mills and Phil Weinstein for always checking up on me.

A big shout-out to all my colleagues and friends at ACAD who have supported, trained, and reassured me throughout my PhD—you have made this experience so much easier! I'd also like to thank all of members of the Metagenomics team, both past and present. Thank you for your feedback for analysis and presentations, and for always putting a smile on my face!

To all my friends, you have no idea how much you have helped me on this journey. We have made so many wonderful memories, including sitting by the river drinking wine on a spring afternoon, games nights, going to the Uni Bar, and enjoying themed-party nights. A special thank you to Emily, Nicole, Felicia, and the Plum Chums for all of your support both near and afar, and also for the many G&Ts.

Yasmyn, now who would have thought seven years ago that we would both be completing our PhDs within a couple of weeks of each other! Thank you for the breakfast, lunch and dinner dates, and for always making the time for me whenever I needed it. I can't wait to see what we will achieve over the next seven years.

I owe a special thank you to my motivators and inspirations: Grandma, Nanna and my other angels who watch out for me from above. You remind me how lucky I am every single day. Grandma, you will soon have a doctor in the family!

To Karyn and Mick, thank you for always letting Jaya and I stay over every Wednesday night with a homemade meal. I thank you for all of your support throughout my studies.

Mum, Dad, Dayna and Ethan, thank you for everything. Thank you for being a shoulder to cry on, someone to laugh with, and for just being there. You have no idea how grateful I am for all of your sacrifices. I also now have the time to make you home-made pasta!

Jaya, thank you! I honestly would not have made it through without you. You were always there celebrate my successes, and also to stop me from giving up. Thank you for always reading my papers and listening to my talks (even if you didn't know what they were about). And, thank you for your endless love, support and encouragement.

Thesis Introduction

Non-communicable diseases of the modern world

Non-communicable diseases (NCDs; non-infectious diseases) contribute to more than 70% of deaths worldwide.[1] NCDs affect people of all ages and sex, anywhere around the globe.[2] Over the past 70 years, there has been a subtle decrease in premature deaths from lethal NCDs, such as cardiovascular disease and cancer.[3,4] Despite this minor success, an increased prevalence of less lethal NCDs, including obesity, type 2 diabetes, allergies, asthma, and autoimmune diseases[5–7] has been observed over this same period. The majority of research and interventions (*e.g.*, "25 x 25" Target[8] and Sustainable Development goal 3.4[9]) have focused on reducing premature deaths caused by NCDs. However, there has been less focus on morbidity and disability as a result of NCDs,[10] which can not only hamper quality of life, but also cause a significant economic burden.[11] Research has shown that mortality and morbidity from NCDs can be reduced by adhering to a healthy lifestyle (*e.g.*, balanced diet, regular exercise).[12] In addition to modifiable lifestyle factors, the development and increased prevalence of NCDs is also influenced by environmental factors, genetic predisposition, and the human microbiome.

Since the mid 20th century, there have been rapid changes to the environment and lifestyle in the industrialised and industrialising worlds. Post-World War II, a surge in socioeconomic and earth system trends, termed the Great Acceleration, was witnessed. Increases in pollution, population density, and local and global transportation have not only had profound effects on Earth's climate, they have also contributed to mortality risk for NCDs, especially cardiovascular and respiratory diseases.[13] Urbanisation—a key change through the Great Acceleration—has led to reduced biodiversity and environmental microbial diversity,[14] which has been hypothesised to have consequential impacts on health, including increases in allergy and asthma predisposition.[7,15] Similarly, a dramatic change to human lifestyle has also been recognised through the Great Acceleration.[16,17] Unbalanced diets, high in carbohydrates, fats, and proteins, and an increased sedentary lifestyle have also likely contributed to higher rates of obesity, type 2 diabetes, and cardiovascular diseases.[18,19] Although it is well known that rapid changes to the environment and lifestyle are the main contributing factors for NCD development, there is still limited understanding of other puzzle pieces that can contribute to these diseases.

The number of studies correlating genetics and NCDs has soared following the sequencing of the first human genome in 2003.[20] Genome-wide association

studies (GWAS) have shed a new light on the relationship between genetic variants, such as single nucleotide polymorphisms (SNPs), and traits of the human body (*e.g.*, height and predisposition to diseases).[21] These studies have increased our appreciation of the complexity of diseases, whereby multiple SNPs across multiple genes and chromosomes can contribute to the predisposition of most NCDs. While this research has demonstrated correlations between SNPs and NCDs, genetic predisposition does not account for all of the variation that contributes to diseases. For example, ~8% of the total heritability has been be explained by 212 loci linked to asthma,[22] and coronary artery disease has ~40-60% total genetic heritability.[23] However, most diseases arise from the interaction between genetics and the environment (*e.g.*, obesity can be attributed to both genetic predisposition[24] and lifestyle[25]). Overall, GWAS and other genetic findings only provide one piece to the puzzle for contributors to NCDs.

While there is an increasing appreciation for the contribution of environmental and lifestyle factors and human genetics to NCDs, there remains a lack of understanding of the development, progression, and persistence of NCDs. However, it has been recently identified that the microbial communities of the human body can also play a key role in health and disease.[26] To prevent, treat, and potentially cure NCDs, we need to turn to the trillions of microorganisms that call us home.[27]

The microbial communities of the human body

The role of the human microbiome in health

The human body is a complex ecosystem that houses many tiny organisms—bacteria, fungi, parasites, archaea, and viruses—that work together and form specialised communities (microbiota).[28] Microbiota, their genetic component and the physio-chemical properties of specific micro-environment within the human body, make up the human microbiome.[29] Altogether, the combined microbiomes of the human body have been collectively termed as our 'second genome'.[30] The human microbiome contributes approximately three million genes, which is 150-fold greater than the human genome;[31] as such, the microbiota perform vital functions (*e.g.*, vitamin K production) that the host cannot achieve itself. Essentially, the human microbiome is critical for human health.

Over the last decade, the human microbiome research field has grown significantly, in large part due to the Human Microbiome Project[32] and the American Gut Project.[33] Through these large-scale studies, technologies and methodologies to examine microbiota also greatly improved. Studying the microbes of the human body has moved beyond culture-based methods, which only identify microbes that can grow in specific laboratory environments, to culture-independent methods and high-throughput sequencing (HTS), which provide highly accurate DNA sequences for microbes found in a given sample.[34] Two well-established sequencing approaches are routinely applied in microbiome research: targeted sequencing of specific phylogenetically informative genes (*e.g.*, the 16S ribosomal RNA (rRNA) gene[34] for bacteria or 18S rRNA gene[35,36] for eukaryotic microbes), which tells us 'who's there'; and whole genome shotgun sequencing (WGS), which tells us 'who's there and what they are doing'.[37] While WGS provides more detailed information (*i.e.* obtaining species- and strain-level information with functional data), it is also more costly and less targeted. For initial studies with a lower budget or samples with a large proportion of host DNA, 16S rRNA sequencing is generally a more suitable option to obtain a general baseline for bacterial data.[38,39] Both sequencing approaches can help researchers understand many aspects of the microbiome, including microbiome acquisition and development throughout life, as well as microbial modifications associated to health and disease.

The initial acquisition of the human microbiome for an individual is currently debated. It has been proposed that a microbiome is established in the placenta.[40,41] However, this early result was linked to contamination and subsequent studies failed to provide strong supporting evidence.[42–44] A more established hypothesis is that the first influx of microbes given to an individual occurs through the process of birth.[45] This first microbial encounter is highly dependent on birth mode.[45] For example, an individual born via a vaginal birth will acquire microbes that resemble the mother's vaginal microbiota, whereas an infant born via Caesarean-section will acquire microbes similar to the doctor's, nurses' and mother's skin.[46] This difference in initial microbial communities may have long-term health consequences on the individual.[47] After birth, the composition of the intestinal microbiome fluctuates and eventually stabilises around three years of age.[48] Over time, the human microbiome continues to develop and change depending on many different factors, including host genetics, the environment, and lifestyle factors.[49]

The human microbiota have co-speciated with humans[50] and are adapted to specific surfaces of the human body.[51] Ecological pressures arise from physiological factors of each body site, such as oxygen levels, pH, and nutrient availability;[51] these tightly-regulated physiological factors create an environment for specific microbes at specific sites. A mutually beneficial relationship is formed between the host and the microbes; the host provides the microbes with somewhere to live, and the microbes perform important functions that the host needs to survive. However, it is important to note that the relationship between the host and microbes are not always in harmony, *i.e.* particular microbes can sometimes become pathogenic and cause harm to the host.

The best-understood roles of the microbiota occur in the gut. For instance, gut microbiota can contribute to the digestion of foods to extract and synthesise essential nutrients, such as vitamin K and short chain fatty acids, for the host.[52] Metabolism of medicinal drugs is also influenced by gut microbiota and can explain variation in individuals' responses to specific drugs[53]. Understanding an individual's response to medicines can provide more precise treatment regimens for that individual.[54] Further, gut microbiota also influence the immune system; commensal microbes restrict pathogen establishment through competitive exclusion[55], and they also produce specific microbe-associated molecular patterns (MAMPs), which allow the host's immune cells to distinguish between beneficial or commensal and pathogenic microbes.[56] These are only a few of many examples where the gut microbiota provides vital resources for the human host to survive and maintain good health.

Microbes located in 'non-gut' body sites also play critical roles in human survival and disease prevention. Microbiota of the skin, nose, mouth, vagina, and lungs all carry out essential functions for the human host, although the extent of these functions is less understood. For example, the predominant roles of skin microbiota are to protect the skin against unwanted pathogens and to prime the immune system.[57,58] Mouth microbes play similar protective roles to the skin microbes,[58] but also have additional responsibilities, such as nitrate reduction to maintain low blood pressure and oral homeostasis.[59] Mouth and skin microbiomes are both influenced by location-dependent physiological factors; there are three main skin types across the body (dry, sebaceous, and moist)[60], while the mouth contains even more distinct sites, such as the tongue, gingiva, teeth, cheeks, and hard and soft palates.[61] Each of these different characteristics and areas within the

skin and mouth have different physiology and specific microbiotas to suit that environment.[60–63] However, research on these 'non-gut' body sites is currently limited. Therefore, specific functions from the niches within each of these specialised body sites need further investigation, which will help to develop a strong understanding of and appreciation for NCDs.

The role of the human microbiome in disease

Disease is often associated with an alteration or imbalance of microbiota.[59,64,65] A microbial imbalance—often termed dysbiosis—can arise due to changes in microbial diversity, composition, or even from a loss or introduction of a single species. For example, lower gut microbial diversity has been observed in individuals with inflammatory bowel disease compared to healthy individuals,[66] and shifts in community composition can contribute to periodontal disease.[67] Microbial dysbiosis can occur due to several different factors, including diet, environmental changes, and even medical treatments (*e.g.*, antibiotics).[65] The period at which a disruption takes place in an individual's life can potentially have life-long impacts on the microbiota and general health of the individual.

Disruptions to the microbiome within the first several years of life can have long-lasting impacts on health. Over the first three years of life, the human body and microbiome undergo rapid development, and disruptions to the microbiome through this period may lead to diseases later in life.[68] For example, the neonatal window of opportunity (*i.e.* first three month of life) is a critical period[69,70] whereby the immune system is undergoing intense training. As the microbiome and immune system are tightly intertwined, disruptions to the microbiota or absence of exposure to particular microbes during this window are thought to be partly responsible for the development of certain NCDs, such as asthma[71] and allergies.[68] The development of such diseases due to reduced environmental microbial exposure during immune development has been proposed through multiple hypotheses, including the Hygiene,[72] Old friends,[15] and Biodiversity[7] hypotheses. However, studies investigating the causal relationship between NCDs and biodiversity still remain limited.

A holistic approach: collecting all pieces of the puzzle to understand NCDs

NCDs can arise from an imperfection at any level (*i.e.* molecular, cellular, tissue, organ, system or organismal level) or with any component (*e.g.*, genetic, epigenetic, transcriptomic, proteomic, microbial, and outside forces, such as the environment and lifestyle factors) of the human body. The holistic approach in systems biology considers each component at every level in the system (*i.e.* the human body), which affords an understanding of complex interactions across multiple subsystems.[73] This field of research has rapidly expanded following recent advancements in the 'omics field, and has been incorporated into multiple biological fields such as oncology[74] and immunology.[75] 'Omics approaches allow for multiple molecules, DNA, genes, transcripts, proteins, and even microbial species to be detected simultaneously.[76] Incorporating multiple 'omics fields into systems biology is the best chance to identify emergent properties of NCDs and is the approach required to fully understand these diseases.

While combining multiple 'omics approaches together, as well as environmental and lifestyle factors into systems biology, is the most ideal method, resources, time, and funding still need to be allocated to do so. To further understand the microbiome—an underrecognized component of systems biology—for human health and disease, researchers must consider how microbiomes work together and interact with host genetics. Microbiome-wide association studies, which are analogous to GWAS, aim to link microbial taxa, genes, and/or functions to a particular human phenotype.[77] For example, Turpin et al. (2016), found that one-third of faecal bacteria are inherited, and 58 human SNPs are associated with the relative abundance of 33 taxa across over 1,000 subjects.[78] Other studies have taken an alternative approach whereby co-occurrence and networks of microbes have been determined.[79–81] Whilst these approaches are heading in the right direction, it has been identified that a majority of correlation studies within microbial communities and between host genetics have been focused on the gut microbiome.[82] Expanding the scope to encompass additional microbiomes of the human body (*e.g.*, mouth, skin, and respiratory tract) and considering more than just local effects of microbiota may unlock links between body sites, which could provide a greater understanding of diseases.

Unlike our individual genomes, which are fixed from birth, our microbiome is dynamic and can change over time. This dynamic aspect to microbiome research can be addressed through longitudinal studies,[83,84] in addition to current understanding of human development. For example, disruptions to the microbiome through immune system development may have long-term health implications.[85] By tracking microbiome development and health outcomes for children, the predisposition or initiation of particular diseases may be identified.[86] For this reason, it is paramount that longitudinal microbiome studies are carried out to identify predisposition and the development of NCDs.

Low microbial biomass samples

Most microbiome research has been focussed on the gut, which has resulted in development of technologies more suited to samples with a high microbial load.[87] In comparison to the gut, samples from most other body sites or samples collected with swabs have a much lower microbial load—termed 'low microbial biomass samples'—and are technically more difficult to examine. The key difficulty associated with these samples arises from an abundance of background DNA that originates from the environment, equipment, reagents, and technicians, and can be introduced at any stage from manufacturing of equipment and reagents to sample collection and sequencing (**Chapter I**).[87,88] Crucially, contaminant DNA is preferentially amplified and sequenced in a low microbial biomass context, which lowers the representation of the true microbiota signal from the sample.[87] One of the best practices in microbiome research is to address contamination by collecting controls to identify contaminant sequences, which is discussed further in **Chapter I** and **Appendix IV** as well as the limitations that originate from technologies and processes used in the microbiome space.

In a number of published research articles, microbial contamination has been reported as true, biologically meaningful findings due to lack of diligence and awareness for technical concerns when working with low microbial biomass samples.[89–91] A prime example of this is the reporting of a placental microbiome. In 2014, Aagaard and colleagues set out to determine the relationship between periodontitis and preterm birth.[40] Their hypothesis was that pathogenic microbes from the mouth, such as *Fusobacterium nucleatum*, were being transported to the

placenta via the blood, causing premature birth. Samples were collected under "sterile" conditions, but no control samples were collected to identify potential contamination from the environment, doctors, technicians, or collection equipment. The only form of controls were laboratory controls (extraction blank controls), of which only a small handful were sequenced. The researchers failed to identify and remove contaminant sequences in their data. This oversight resulted in a false positive result and has unfortunately led other researchers down a similar path.[92–94] Since this initial study, other researchers replied to the paper and tried to replicate the study,[42,43,95,96] further validating that the bacterial sequences found in the placenta samples closely resemble those found in controls. As awareness for the challenges of low microbial biomass sample examination has increased, more researchers are now including controls from sampling through to data analysis. Nevertheless, this practice is still not standardised across the board, and researchers urgently need to learn from these past errors to improve microbiome research in the future.

Despite the difficulties of contamination, low microbial biomass samples hold a wealth of knowledge and should not be dismissed. It is now well-recognised that most sites of the human body are not sterile and do house microbiota. For example, it was initially thought that lungs were sterile in health, and pathogens were present in disease. Research has now shown that while the lung microbiota can be stochastic,[97] there are many niche-specific microbes present,[98] and physiological changes to the respiratory tract can cause microbial dysbiosis in diseases, such as cystic fibrosis.[99] Similar non-sterile low microbial biomass sites, such as skin,[100] nasal,[101] vaginal,[102] and eye sites,[103] are now also being investigated in health and disease. If treated carefully, low microbial biomass samples have the power to provide insights into areas of the human body that are not well understood from a microbial perspective and may provide the missing puzzle piece needed to understand the predisposition, prevention, and development of NCDs.

Thesis overview

In this thesis, I expand our knowledge of the development, changes, and relationships between human microbiotas and NCDs. Morbidity and mortality of NCDs are at sky-high levels, and unachievable targets are currently set to reduce premature NCD-related deaths. To decrease premature death rates from NCDs, it is

vital to go beyond genetics—introduce measurable features of the environment, explore the interaction between the two, and incorporate the human microbiome into these assessments. As the human microbiome is modifiable to some extent, it is an important factor to consider for decreasing NCD prevalence and associated death. Incorporating our second genome into the NCD puzzle will provide a greater, more holistic understanding of the initiation, progression, and prevalence of NCDs. Across five manuscripts, this thesis aims to provide new perspectives on NCDs through the investigation of low microbial biomass sites.

Chapter I: Microbiome applications for pathology: challenges of low microbial biomass samples during diagnostic testing

Within this published review, I evaluate current techniques used in pathology and provide insights for the future integration of microbiome analysis for communicable and NCD diagnosis. As we start to move away from Koch's postulates (*i.e.* one pathogen = one disease) to a community-wide approach, it is important to consider changes to technologies, diagnosis, and treatments in the pathology field. As a majority of samples (*e.g.*, skin, blood, urine, etc.) used in pathology are low in microbial biomass, I also highlight pitfalls (contamination and biases) that pathologists may come across when microbiome analyses are incorporated into a pathology setting. To overcome these obstacles, I provide recommendations for minimising and mitigating contamination and biases that can affect the results of low microbial biomass samples. Incorporating microbiome analyses as a new tool for pathology may help to identify and treat NCDs more effectively in the future.

Chapter II: Moving beyond the gut microbiome: combining systems biology and multi-site analyses to combat non-communicable diseases

In this opinion piece, I highlight that human microbiome research is primarily focussed on the gut, and almost all other body sites are forgotten, when considering NCDs. This shortcoming for current microbiome research also misses out on the advantages of considering the human body as system. Appreciation for

the interrelationship between human microbiotas in the body may provide more insights into NCDs.

Chapter III: Transfer of environmental microbes to the skin and respiratory tract after green space exposure

Lifestyle and environmental factors have changed dramatically over the past 70 years, which has correlated with a rise in NCDs. Three key hypotheses (Hygiene, Old friends, and Biodiversity hypotheses) have attempted to explain this phenomenon by linking reduced environmental microbial exposure to the increase in NCDs. To counteract the increase in NCDs, the Microbiome Rewilding hypothesis has proposed that increased exposure to microbially diverse environments can restore health. In this chapter, I track changes to the skin and nasal microbiota before and after urban green space exposure. This study provides a foundation to explore the relationship between human microbiota and the natural environment, and provides new insights into both the past development of NCDs during lifestyle disruptions in the Great Acceleration and the potential prevention of NCDs in the future.

Chapter IV: Initially disrupted preterm infant oral microbiota is restored within three months

Disruptions to the microbiota development over the first 1,000 days of life can have long-lasting health consequences. These disturbances may explain the higher morbidity and mortality rates in preterm infants compared to their full-term counterparts, as well as lifelong health issues. While we know there are alterations to the gut microbiota of preterm infants, there is very little research carried out on other body sites of these fragile preterm neonates. In this chapter, I investigate oral and respiratory microbiota development over three months in preterm neonates who develop respiratory or systemic diseases, compared to healthy preterm and full-term infants and adults. This research detects a disruption to the oral and respiratory microbiota in preterm infants at a critical developmental stage for the immune system. Understanding these changes is vital for reducing morbidity and mortality rates in preterm infants and provides foundational data to understand how these earlier life microbiota alterations go on to influence NCDs later in life.

Chapter V: Altered oral microbiota in type I diabetic children with hyperlipidemic parents

Hyperlipidaemia (elevated blood lipids) is a multifactorial disease that has been proposed as a link between periodontal disease and type 1 diabetes. In this chapter, I explore the oral microbial relationship between periodontal disease of children with type 1 diabetes and the hyperlipidaemia status of their parents. As fats have a direct link to the gut microbiota and periodontal disease is linked to oral microbiota changes in the mouth, most research has investigated local relationships between disease and microbiota. Moving away from this dogma, this research identified a relationship between hyperlipidaemia status of parents and oral microbiota of children, which could suggest that processes in the gut may have an influence on other areas of the body, such as the mouth. In the future, it is vital to collect microbiota samples from across the body (as advocate for in **Chapter II**), to provide a holistic understanding of NCDs.

References

1. Bennett, J. E. *et al.* NCD Countdown 2030: worldwide trends in non-communicable disease mortality and progress towards Sustainable Development Goal target 3.4. *The Lancet* **392**, 1072–1088 (2018).
2. Martinez, R. *et al.* Trends in premature avertable mortality from non-communicable diseases for 195 countries and territories, 1990–2017: a population-based study. *The Lancet Global Health* **8**, e511–e523 (2020).
3. Ritchey, M. D., Wall, H. K., George, M. G. & Wright, J. S. US trends in premature heart disease mortality over the past 50 years: Where do we go from here? *Trends in Cardiovascular Medicine* **30**, 364–374 (2020).
4. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2020. *CA: A Cancer Journal for Clinicians* **70**, 7–30 (2020).
5. Bhupathiraju Shilpa N. & Hu Frank B. Epidemiology of Obesity and Diabetes and Their Cardiovascular Complications. *Circulation Research* **118**, 1723–1735 (2016).
6. Bach, J.-F. The Effect of Infections on Susceptibility to Autoimmune and Allergic Diseases. *New England Journal of Medicine* **347**, 911–920 (2002).

7. von Hertzen, L., Hanski, I. & Haahtela, T. Natural immunity. *EMBO reports* **12**, 1089–1093 (2011).

8. WHO | NCD Global Monitoring Framework. *WHO* http://www.who.int/nmh/global_monitoring_framework/en/.

9. Transforming our world: the 2030 Agenda for Sustainable Development .:. Sustainable Development Knowledge Platform. https://sustainabledevelopment.un.org/post2015/transformingourworld.

10. Richards, N. C. *et al.* Disability, noncommunicable disease and health information. *Bull. World Health Organ.* **94**, 230–232 (2016).

11. Chen, S., Kuhn, M., Prettner, K. & Bloom, D. E. The macroeconomic burden of noncommunicable diseases in the United States: Estimates and projections. *PLoS One* **13**, (2018).

12. Crosland, P., Ananthapavan, J., Davison, J., Lambert, M. & Carter, R. The health burden of preventable disease in Australia: a systematic review. *Australian and New Zealand Journal of Public Health* **43**, 163–170 (2019).

13. Lelieveld, J. Clean air in the Anthropocene. *Faraday Discuss.* **200**, 693–703 (2017).

14. Gillings, M. R. & Paulsen, I. T. Microbiology of the Anthropocene. *Anthropocene* **5**, 1–8 (2014).

15. Rook, G. A. W., Martinelli, R. & Brunet, L. R. Innate immune responses to mycobacteria and the downregulation of atopic responses. *Curr Opin Allergy Clin Immunol* **3**, 337–342 (2003).

16. Gillings, M. R., Paulsen, I. T. & Tetu, S. G. Ecology and Evolution of the Human Microbiota: Fire, Farming and Antibiotics. *Genes (Basel)* **6**, 841–857 (2015).

17. Popkin, B. M. Urbanization, Lifestyle Changes and the Nutrition Transition. *World Development* **27**, 1905–1916 (1999).

18. Popkin, B. M., Adair, L. S. & Ng, S. W. NOW AND THEN: The Global Nutrition Transition: The Pandemic of Obesity in Developing Countries. *Nutr Rev* **70**, 3–21 (2012).

19. Frumkin, H. & Haines, A. Global Environmental Change and Noncommunicable Disease Risks. *Annual Review of Public Health* **40**, 261–282 (2019).

20. Mills, M. C. & Rahal, C. A scientometric review of genome-wide association studies. *Communications Biology* **2**, 1–11 (2019).

21. Visscher, P. M. *et al.* 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am J Hum Genet* **101**, 5–22 (2017).

22. Han, Y. *et al.* Genome-wide analysis highlights contribution of immune system pathways to the genetic architecture of asthma. *Nature Communications* **11**, 1776 (2020).

23. Wang, Y. & Wang, J.-G. Genome-Wide Association Studies of Hypertension and Several Other Cardiovascular Diseases. *PLS* **6**, 169–186 (2018).

24. Cheng, M. *et al.* Computational analyses of obesity associated loci generated by genome-wide association studies. *PLoS One* **13**, (2018).

25. Cha, E. *et al.* Lifestyle habits and obesity progression in overweight and obese American young adults: Lessons for promoting cardiometabolic health. *Nurs Health Sci* **17**, 467–475 (2015).

26. Young, V. B. The role of the microbiome in human health and disease: an introduction for clinicians. *BMJ* **356**, (2017).

27. Sender, R., Fuchs, S. & Milo, R. Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLOS Biology* **14**, e1002533 (2016).

28. Lederberg, J. & McCray, A. T. 'Ome Sweet 'Omics -- A Genealogical Treasury of Words. *The Scientist* **15**, 8 (2001).

29. Whipps, J., Lewis, K. & Cooke, R. Mycoparasitism and plant disease control. in *Fungi in Biological Control Systems* (ed. Burge, M. N.) 161–187 (Manchester University Press, 1988).

30. Grice, E. A. & Segre, J. A. The Human Microbiome: Our Second Genome. *Annu Rev Genomics Hum Genet* **13**, 151–170 (2012).

31. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).

32. NIH HMP Working Group *et al.* The NIH Human Microbiome Project. *Genome Res.* **19**, 2317–2323 (2009).

33. McDonald, D. *et al.* American Gut: an Open Platform for Citizen Science Microbiome Research. *mSystems* **3**, (2018).

34. Caporaso, J. G. *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* **6**, 1621–1624 (2012).

35. Amaral-Zettler, L. A., McCliment, E. A., Ducklow, H. W. & Huse, S. M. A Method for Studying Protistan Diversity Using Massively Parallel Sequencing of V9 Hypervariable Regions of Small-Subunit Ribosomal RNA Genes. *PLOS ONE* **4**, e6372 (2009).

36. Stoeck, T. *et al.* Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Molecular Ecology* **19**, 21–31 (2010).

37. Sharpton, T. J. An introduction to the analysis of shotgun metagenomic data. *Front Plant Sci* **5**, (2014).

38. Hillmann, B. *et al.* Evaluating the Information Content of Shallow Shotgun Metagenomics. *mSystems* **3**, (2018).

39. Jovel, J. *et al.* Characterization of the Gut Microbiome Using 16S or Shotgun Metagenomics. *Front Microbiol* **7**, 459 (2016).

40. Aagaard, K. *et al.* The placenta harbors a unique microbiome. *Sci Transl Med* **6**, 237ra65 (2014).

41. Parnell, L. A. *et al.* Microbial communities in placentas from term normal pregnancy exhibit spatially variable profiles. *Scientific Reports* **7**, 11200 (2017).

42. Leiby, J. S. *et al.* Lack of detection of a human placenta microbiome in samples from preterm and term deliveries. *Microbiome* **6**, 196 (2018).

43. Lim, E. S., Rodriguez, C. & Holtz, L. R. Amniotic fluid from healthy term pregnancies does not harbor a detectable microbial community. *Microbiome* **6**, 87 (2018).

44. Goffau, M. C. de *et al.* Human placenta has no microbiome but can contain potential pathogens. *Nature* **572**, 329–334 (2019).

45. Lif Holgerson, P., Harnevik, L., Hernell, O., Tanner, A. & Johansson, I. Mode of Birth Delivery Affects Oral Microbiota in Infants. *Journal of Dental Research* **90**, 1183–1188 (2011).

46. Dominguez-Bello, M. G. *et al.* Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *PNAS* **107**, 11971–11975 (2010).

47. Arboleya, S. *et al.* C-section and the Neonatal Gut Microbiome Acquisition: Consequences for Future Health. *ANM* **73**, 17–23 (2018).

48. Yatsunenko, T. *et al.* Human gut microbiome viewed across age and geography. *Nature* **486**, 222–227 (2012).

49. Hasan, N. & Yang, H. Factors affecting the composition of the gut microbiota, and its modulation. *PeerJ* **7**, (2019).

50. Groussin, M., Mazel, F. & Alm, E. J. Co-evolution and Co-speciation of Host-Gut Bacteria Systems. *Cell Host & Microbe* **28**, 12–22 (2020).

51. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012).

52. Rowland, I. *et al.* Gut microbiota functions: metabolism of nutrients and other food components. *Eur J Nutr* **57**, 1–24 (2018).

53. Zimmermann, M., Zimmermann-Kogadeeva, M., Wegmann, R. & Goodman, A. L. Mapping human microbiome drug metabolism by gut bacteria and their genes. *Nature* **570**, 462–467 (2019).

54. Kuntz, T. M. & Gilbert, J. A. Introducing the Microbiome into Precision Medicine. *Trends Pharmacol. Sci.* **38**, 81–91 (2017).

55. Elsen, L. W. van den, Poyntz, H. C., Weyrich, L. S., Young, W. & Forbes-Blom, E. E. Embracing the gut microbiota: the new frontier for inflammatory and infectious diseases. *Clinical & Translational Immunology* **6**, e125 (2017).

56. Negi, S., Das, D. K., Pahari, S., Nadeem, S. & Agrewala, J. N. Potential Role of Gut Microbiota in Induction and Regulation of Innate Immune Memory. *Front. Immunol.* **10**, (2019).

57. Sanford, J. A. & Gallo, R. L. Functions of the skin microbiota in health and disease. *Semin Immunol* **25**, 370–377 (2013).

58. Park, Y. J. & Lee, H. K. The Role of Skin and Orogenital Microbiota in Protective Immunity and Chronic Immune-Mediated Inflammatory Disease. *Front. Immunol.* **8**, (2018).

59. Frias-Lopez, J. & Duran-Pinedo, A. E. The Function of the Oral Microbiome in Health and Disease. in *Emerging Therapies in Periodontics* (ed. Sahingur, S. E.) 141–173 (Springer International Publishing, 2020). doi:10.1007/978-3-030-42990-4_10.

60. Byrd, A. L., Belkaid, Y. & Segre, J. A. The human skin microbiome. *Nature Reviews Microbiology* **16**, 143–155 (2018).

61. Mark Welch, J. L., Dewhirst, F. E. & Borisy, G. G. Biogeography of the Oral Microbiome: The Site-Specialist Hypothesis. *Annu. Rev. Microbiol.* **73**, 335–358 (2019).

62. Simón-Soro, A. *et al.* Microbial geography of the oral cavity. *J. Dent. Res.* **92**, 616–621 (2013).

63. Aas, J. A., Paster, B. J., Stokes, L. N., Olsen, I. & Dewhirst, F. E. Defining the normal bacterial flora of the oral cavity. *J. Clin. Microbiol.* **43**, 5721–5732 (2005).

64. Wang, B., Yao, M., Lv, L., Ling, Z. & Li, L. The Human Microbiota in Health and Disease. *Engineering* **3**, 71–82 (2017).

65. Petersen, C. & Round, J. L. Defining dysbiosis and its influence on host immunity and disease. *Cell Microbiol* **16**, 1024–1033 (2014).

66. Sha, S. *et al.* The biodiversity and composition of the dominant fecal microbiota in patients with inflammatory bowel disease. *Diagnostic Microbiology and Infectious Disease* **75**, 245–251 (2013).

67. Chen, C. *et al.* Oral microbiota of periodontal health and disease and their changes after nonsurgical periodontal therapy. *The ISME Journal* **12**, 1210–1224 (2018).

68. Gensollen, T., Iyer, S. S., Kasper, D. L. & Blumberg, R. S. How colonization by microbiota in early life shapes the immune system. *Science* **352**, 539–544 (2016).

69. Torow, N. & Hornef, M. W. The Neonatal Window of Opportunity: Setting the Stage for Life-Long Host-Microbial Interaction and Immune Homeostasis. *The Journal of Immunology* **198**, 557–563 (2017).

70. Renz, H. *et al.* The neonatal window of opportunity—early priming for life. *J Allergy Clin Immunol* **141**, 1212–1214 (2018).

71. Arrieta, M.-C. *et al.* Early infancy microbial and metabolic alterations affect risk of childhood asthma. *Science Translational Medicine* **7**, 307ra152-307ra152 (2015).

72. Strachan, D. P. Hay fever, hygiene, and household size. *BMJ* **299**, 1259–1260 (1989).

73. Alcocer-Cuarón, C., Rivera, A. L. & Castaño, V. M. Hierarchical structure of biological systems. *Bioengineered* **5**, 73–79 (2014).

74. Du, W. & Elemento, O. Cancer systems biology: embracing complexity to develop better anticancer therapeutic strategies. *Oncogene* **34**, 3215–3225 (2015).

75. Meijerink, M. *et al.* The Impact of Immune Interventions: A Systems Biology Strategy for Predicting Adverse and Beneficial Immune Effects. *Front. Immunol.* **10**, (2019).

76. Ma'ayan, A. Complex systems biology. *Journal of The Royal Society Interface* **14**, 20170391 (2017).

77. Gilbert, J. A. *et al.* Microbiome-wide association studies link dynamic microbial consortia to disease. *Nature* **535**, 94–103 (2016).

78. Turpin, W. *et al.* Association of host genome with intestinal microbial composition in a large healthy cohort. *Nature Genetics* **48**, 1413–1417 (2016).

79. Faust, K. *et al.* Microbial Co-occurrence Relationships in the Human Microbiome. *PLoS Comput Biol* **8**, (2012).

80. Einarsson, G. G. *et al.* Community analysis and co-occurrence patterns in airway microbial communities during health and disease. *ERJ Open Res* **5**, (2019).

81. Layeghifard, M. *et al.* Microbiome networks and change-point analysis reveal key community changes associated with cystic fibrosis pulmonary exacerbations. *npj Biofilms and Microbiomes* **5**, 1–12 (2019).

82. Witherden, E. A., Moyes, D. L., Bruce, K. D., Ehrlich, S. D. & Shoaie, S. Using systems biology approaches to elucidate cause and effect in host–microbiome interactions. *Current Opinion in Systems Biology* **3**, 141–146 (2017).

83. Stewart, C. J. *et al.* Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature* **562**, 583–588 (2018).

84. Lif Holgerson, P., Esberg, A., Sjödin, A., West, C. E. & Johansson, I. A longitudinal study of the development of the saliva microbiome in infants 2 days to 5 years compared to the microbiome in adolescents. *Scientific Reports* **10**, 9629 (2020).

85. Hornef, M. W. & Torow, N. 'Layered immunity' and the 'neonatal window of opportunity' – timed succession of non-redundant phases to establish mucosal host–microbial homeostasis after birth. *Immunology* **159**, 15–25 (2020).

86. Dzidic, M. *et al.* Oral microbiota maturation during the first 7 years of life in relation to allergy development. *Allergy* **73**, 2000–2011 (2018).

87. Salter, S. J. *et al.* Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biology* **12**, 87 (2014).

88. Selway, C. A., Eisenhofer, R. & Weyrich, L. S. Microbiome applications for pathology: challenges of low microbial biomass samples during diagnostic testing. *The Journal of Pathology: Clinical Research* **6**, 97–106 (2020).

89. Chan, A. A. *et al.* Characterization of the microbiome of nipple aspirate fluid of breast cancer survivors. *Scientific Reports* **6**, 28061 (2016).

90. Hieken, T. J. *et al.* The Microbiome of Aseptically Collected Human Breast Tissue in Benign and Malignant Disease. *Scientific Reports* **6**, 30751 (2016).

91. Xuan, C. *et al.* Microbial dysbiosis is associated with human breast cancer. *PLoS ONE* **9**, e83744 (2014).

92. Antony, K. M. *et al.* The preterm placental microbiome varies in association with excess maternal gestational weight gain. *American Journal of Obstetrics & Gynecology* **212**, 653.e1-653.e16 (2015).

93. Zheng, J. *et al.* The Placental Microbiome Varies in Association with Low Birth Weight in Full-Term Neonates. *Nutrients* **7**, 6924–6937 (2015).

94. Bassols, J. *et al.* Gestational diabetes is associated with changes in placental microbiota and microbiome. *Pediatric Research* **80**, 777–784 (2016).

95. Lauder, A. P. *et al.* Comparison of placenta samples with contamination controls does not provide evidence for a distinct placenta microbiota. *Microbiome* **4**, 29 (2016).

96. Theis, K. R. *et al.* Does the human placenta delivered at term have a microbiota? Results of cultivation, quantitative real-time PCR, 16S rRNA gene sequencing, and metagenomics. *American Journal of Obstetrics & Gynecology* **220**, 267.e1-267.e39 (2019).

97. Huang, Y. J. *et al.* The Role of the Lung Microbiome in Health and Disease. A National Heart, Lung, and Blood Institute Workshop Report. *Am J Respir Crit Care Med* **187**, 1382–1387 (2013).

98. Man, W. H., de Steenhuijsen Piters, W. A. A. & Bogaert, D. The microbiota of the respiratory tract: gatekeeper to respiratory health. *Nat Rev Micro* **15**, 259–270 (2017).

99. LaSala, P. R., Segal, J., Han, F. S., Tarrand, J. J. & Han, X. Y. First Reported Infections Caused by Three Newly Described Genera in the Family Xanthomonadaceae. *J. Clin. Microbiol.* **45**, 641–644 (2007).

100. Grice, E. A. *et al.* Topographical and temporal diversity of the human skin microbiome. *Science* **324**, 1190–1192 (2009).

101. Bassis, C. M., Tang, A. L., Young, V. B. & Pynnonen, M. A. The nasal cavity microbiota of healthy adults. *Microbiome* **2**, 27 (2014).

102. Ma, B., Forney, L. J. & Ravel, J. The vaginal microbiome: rethinking health and diseases. *Annu Rev Microbiol* **66**, 371–389 (2012).

103. Ozkan, J. & Willcox, M. D. The Ocular Microbiome: Molecular Characterisation of a Unique and Low Microbial Environment. *Curr. Eye Res.* **44**, 685–694 (2019).

Chapter I

Microbiome applications for pathology: challenges of low microbial biomass samples during diagnostic testing

Statement of Authorship

Title of Paper	Microbiome applications for pathology: challenges of low microbial biomass samples during diagnostic testing
Publication Status	☒ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Published in The Journal of Pathology: Clinical Research

Principal Author

Name of Principal Author (Candidate)	Caitlin A. Selway
Contribution to the Paper	Conceptualised the review and wrote the manuscript.
Overall percentage (%)	75%
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature	Date: 26-8-2020

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Raphael Eisenhofer		
Contribution to the Paper	Critically reviewed and edited all versions of the manuscript.		
Signature		Date	12/8/2020

Name of Co-Author	Laura S. Weyrich		
Contribution to the Paper	Conceptualised the review and critically reviewed and edited all versions of the manuscript.		
Signature		Date	8/18/20

The Journal of Pathology: Clinical Research
J Pathol Clin Res; 2020
Published online in Wiley Online Library
(wileyonlinelibrary.com). DOI: 10.1002/cjp2.151

REVIEW

Microbiome applications for pathology: challenges of low microbial biomass samples during diagnostic testing

Caitlin A Selway[1]*, Raphael Eisenhofer[1] and Laura S Weyrich[1,2,3]*

[1]Australian Centre for Ancient DNA, Department of Molecular and Biomedical Science, University of Adelaide, Adelaide, SA, Australia
[2]Department of Anthropology, Pennsylvania State University, University Park, PA, USA
[3]Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, PA, USA

*Correspondence: Caitlin A Selway, Australian Centre for Ancient DNA, University of Adelaide, Adelaide, SA 5005 Australia. E-mail: caitlin.selway@adelaide.edu.au; Laura S Weyrich, Department of Anthropology, Pennsylvania State University, University Park, PA 16801 USA. E-mail: lsw132@psu.edu

Abstract

The human microbiome can play key roles in disease, and diagnostic testing will soon have the ability to examine these roles in the context of clinical applications. Currently, most diagnostic testing in pathology applications focuses on a small number of disease-causing microbes and dismisses the whole microbial community that causes or is modulated by disease. Microbiome modifications have already provided clinically relevant insights in gut and oral diseases, such as irritable bowel disease, but there are currently limitations when clinically examining microbiomes outside of these body sites. This is critical, as the majority of microbial samples used in pathology originate from body sites that contain low concentrations of microbial DNA, including skin, tissue, blood, and urine. These samples, also known as low microbial biomass samples, are difficult to examine without careful consideration and precautions to mitigate contamination and biases. Here, we present the limitations when analysing low microbial biomass samples using current protocols and techniques and highlight the advantages that microbiome testing can offer diagnostics in the future, if the proper precautions are implemented. Specifically, we discuss the sources of contamination and biases that may result in false assessments for these sample types. Finally, we provide recommendations to mitigate contamination and biases from low microbial biomass samples during diagnostic testing, which will be especially important to effectively diagnose and treat patients using microbiome analyses.

Keywords: pathology; microbiome; diagnostic testing; microbiota; personalised medicine; low biomass; contamination; clinical microbiology

Received 5 September 2019; Revised 28 October 2019; Accepted 7 November 2019

No conflicts of interest were declared.

Introduction

Existing pathology techniques currently survey small numbers of disease-causing microbes by applying Koch's postulates. Koch's postulates explain the relationship between a single culturable microbial isolate and a disease, but these postulates only explain a small number of microbial related diseases [1]. The significance of more microbially complex diseases is now well appreciated. In most cases, human genetics and single microbes do not explain the full disease pathology, such as urinary tract infections [2] or periodontitis [3]. Such diseases result from polymicrobial infections or complex interspecies interactions that can only be understood by looking at the microbiome – an ecosystem of complex microbial communities – in conjunction with human genetics and transcriptomics [4,5]. The human microbiome consists of diverse microbial communities (microbiota) that live on external and internal surfaces of the human body [6], as well as the genetic content and environment of these microbes. Disruptions to the microbiota, through factors such as diet, environment, and medical treatment (i.e. antibiotics), can alter the microbiota structure and contribute to disease [7]. It is now known that numerous non-infectious diseases, including inflammatory bowel disease [7], asthma [8], and neurological disorders [9], are associated with alterations in the

microbiome. However, current techniques used in pathology cannot readily detect and characterise ecosystem shifts within these communities, as well as unknown or unculturable pathogens, which can impact the ability to accurately diagnose some diseases.

Most ongoing microbiome research is focused on areas of the body that have high concentrations of microbes (i.e. gut or mouth). However, most samples used for pathological screening have fewer microbes and are known as low microbial biomass samples. Low microbial biomass samples can be easily overwhelmed by contamination from background DNA and are more prone to technical biases, such as overamplification during PCR [10–13]. As investigations into low microbial biomass body sites increase, it is vital that new protocols and techniques are applied to minimise the effect that contamination and biases have on these samples and that the limitations are fully understood when developing diagnostic tools based on the results. Here, we review the current techniques used in diagnostic testing and discuss how microbiome assessments can be incorporated into pathology in the future, as many of these techniques have not yet been successfully developed in a diagnostic setting. Lastly, we discuss issues during analysis of low microbial biomass samples in past studies, while highlighting the sources of contamination and biases, and review techniques that can be applied to minimise these confounding factors as microbiome tools are developed moving forward.

Assessing and treating diseases in pathology: The present and the future

Limitations of current diagnostic testing for single pathogens

Current diagnostic testing is continuously improving with technological advancements, allowing for more accurate detection of diseases and providing information for precise treatment options. Most current practices used to identify a microorganism during an infection involve collecting a sample (e.g. swab, bodily fluid, or tissue) of the infected area. The sample is then prepared for microscopic analysis, culture, or a PCR-based method to identify a single or small range of pathogenic species that were previously characterised to cause those disease symptoms [14]. However, identifying specific disease-causing microbes can be a time-consuming process [15], during a period when patients are potentially left untreated or are administered treatments that are not targeted for a specific condition, such as broad-spectrum antibiotics [16]. In addition, this technology cannot identify unknown pathogens or diverse mixtures of microbes, resulting in delays when investigating rapidly emerging, novel, or unculturable pathogens [17,18]. Lastly, contamination from sample collection and laboratory technicians is not adequately addressed at present. There seems to be no national or international standard that specifically standardises sample collection, with the exception of gloves to be worn to protect the technician. This dismissal of protecting the integrity of the sample could lead to the inadvertent introduction of additional microbes into the sample, which can modify or bias results, and could ultimately result in incorrect diagnoses and treatment of patients [18]. New approaches based on sensitive, high-throughput techniques that investigate unknown microbes, or the microbial community as a whole, are required to mitigate some of these issues and better identify disease sources and complications.

Benefits of microbiome analysis for pathology

Current research efforts assess changes to the microbiome to understand disease pathologies on a case-by-case basis. Approaches used for microbiome analysis can assist in characterising both communicable (infectious) and non-communicable (non-infectious) diseases, with non-communicable conditions being more prevalent in high-income countries [19]. Microbiome analysis provides clinicians with the ability to practice precision medicine, especially in unique or unsolved cases, as it will enable the identification of specific unknown pathogen(s) and the ability to monitor the microbiota and microbiome through time, in relationship to disease status and treatment [4,20,21]. A current set of procedures to explore microbiomes in diagnostic testing could be: sample collection from the affected area; nucleic acid extraction; sequencing library preparation; sequencing using high-throughput sequencing (HTS) approaches; and finally, reconstruction of the microbiota and/or their functions using high-throughput computing resources [22]. Several types of HTS can be employed [23], including amplicon based sequencing (targeting one 'fingerprint' or 'barcode' gene to identify the microbiota present, such as the gene encoding 16S ribosomal RNA) [22]; shotgun metagenomic sequencing (assessing a random sampling of DNA from the biological sample to reconstruct microbial genomes and functions of known, new, or under characterised species/strains in the microbiome) [24]; and metatranscriptomics (examining the actively transcribed genes using RNA based

sequencing approaches) [25]. These approaches can aid in the identification, function, and activity levels of known and novel species/strains that contribute to infectious diseases [4,26,27], which is important for diagnosis and treatment, but also critical for the downstream development of new rapid and cost effective techniques to readily detect these pathogens. Lastly, understanding the functions of the microbiota and how these functions are utilised is also essential to understand the underpinning mechanisms of disease, within both infectious and non-infectious conditions. The human microbiome typically contributes over 3 million genes in every single human, which is approximately 150 times more genes than the human genome [28]. This volume of information is not routinely assessed in current diagnostic testing and could inform more effective treatment strategies or identify unknown infection dynamics. Below, we discuss in detail how microbiome testing can provide additional information when diagnosing communicable and non-communicable diseases in the future.

1. *Identifying infectious disease from new or unknown pathogens*

While the most common approach to diagnose an infectious disease is to test for a single pathogen, this approach does not explore unknown or unidentifiable pathogens. Metagenomics analysis can be used to reconstruct the genomes of novel or unknown pathogens by comparison to a distantly related species or strain, to assemble genomes of unknown species from a sample using *de novo* approaches, to reconstruct genomic information, or to quantify levels of a taxon that is present in one location but absent or lower in another, perhaps providing information about an opportunistic pathogen. For example, in one report, a patient suffered with chronic meningoencephalitis for 3 years with no known disease aetiology using standard pathology tests [4]. An assessment of the microbiome (metagenomics) revealed that the Cache Valley virus (not known to cause meningoencephalitis) was responsible for the disease [4]. In concert with metagenomics analysis, host transcriptomics can also be used to identify if a disease is caused by an infection [4,29] by examining the active transcription of immune genes activated during infection. Certain host genes are transcribed when fighting an infectious disease, and these transcripts, or their absence, can be detected using current transcriptomic approaches, providing additional clues to the type of infection [29]. Quantifying the level of these particular transcripts can also distinguish between an infectious and non-infectious disease [30,31].

2. *Co-infections or diverse poly-microbial infections*

Diagnostic testing is typically limited to one potential pathogen per test; however, several diseases can manifest as co-infections or poly-microbial infections, where multiple microorganisms contribute to the disease. Poly-microbial infections, such as those observed in the urinary and respiratory tracts, are often difficult to treat due to the interactions between different microbes, so understanding the mechanisms that underpin these infections could improve treatment strategies [32,33]. Amplicon or metagenomic analysis can be applied to identify numerous pathogens simultaneously or, potentially, assess levels of opportunistic pathogens, if a healthy sample has been taken previously.

3. *Function(s) of disease-causing pathogen(s)*

Assessing specific pathogen functions is essential to identify the correct treatment option(s), especially in cases where the first line of treatment is ineffective. This is typically done by screening a cultured pathogen against different types of antimicrobials. However, the resistance for these antimicrobials is encoded in the genome of each microbe, which contains specific genes for individualised functions. For example, some pathogens carry antimicrobial resistance genes that can provide broad spectrum or very specific antibiotic resistance [34]. Metagenomic sequencing could identify which (if any) antimicrobial resistance genes a microbe has and provide information to advise which antibiotic would be the most specific and effective [35]. In addition, metatranscriptomics – examination of the RNA in the microbiome – could provide key information on the microbes that are actively playing a role in drug metabolism [36] or antibiotic resistance [37]. This information could then be used for better and more targeted treatment, which is especially critical in the light of rising antibiotic resistance.

4. *Microbiome functions for non-infectious diseases*

Critically, microbiome assessment may prove to be most useful during the diagnosis and treatment of non-infectious diseases. Autoimmune, allergic, and inflammatory disease are on the rise in industrialised countries [38] and, in many cases, their primary causes remain unclear [39–41]. For example, inflammatory bowel disease has recently been linked to disruptions in the gut microbiota, and there is research underway to determine whether faecal microbiota transplants can be an effective treatment [41]. Other microbiome transplants have also been recently suggested, opening the door for different types of microbiome transplantation.

An amplicon or metagenomics-based approach could be utilised to assess the donor's or patient's microbiota or microbiome and identify which specific transplant donor might be best, which microbial functions are missing from the patient's gut, or which probiotic strategies may be the most useful [41]. Overall, characterising and assessing the microbiome can be another tool to help identify non-infectious disease causes or complications and inform more specific and effective treatments.

Implementing microbiome analysis in pathology: Challenges of working with low-biomass samples

While there are many advantages to investigating the microbiome using HTS, the widespread implementation of such technologies in a medical context is still limited by certain factors. Clinicians, pathologists, and bioinformaticians require training to properly collect, process, analyse, and interpret microbiome samples, and minimum standards for laboratory analysis and reporting are needed to ensure robust diagnosis and treatment. In some cases (e.g. non-infectious diseases), the microbial communities and functions linked to disease are still being described; thus, only hypotheses related to causality and function of non-infectious diseases are described [42,43]. The bioinformatics and sequencing technologies needed to completely describe non-infectious diseases are not adequate at present; therefore, more research still needs to be undertaken before the microbiome can be analysed routinely in diagnostics [42,43]. Several issues for implementing reliable microbiome analysis in diagnostic testing are already known and need to be addressed in the diagnostics field as microbiome testing is developed, implemented, and employed. This is largely due to the fact that many of the samples commonly screened in diagnostic testing contain a low microbial biomass.

There are numerous areas of the body that are now considered to be of low microbial biomass but were originally thought to be sterile. For example, research has shown that microbes colonise and perform critical functions within the lungs, albeit at low concentrations [44]. Other examples of low microbial biomass body sites include skin [45], blood [46], urine [47] and tissue [48] – all of which are typical samples collected for diagnostic tests. These samples typically have small numbers of microbial cells (100–10000 cells/mL) [49,50] and are more difficult to examine than those of high microbial biomass. Even in widespread microbiome research today, common microbiome protocols [51–53] are not optimised for low microbial biomass samples. Recent research has provided improvements in low biomass laboratory and analysis protocols [54,55], but there is still room for further improvement.

The most significant issue in examining low biomass samples is contaminating, or exogenous, DNA (i.e. DNA from sources other than the sample of interest), which is unintentionally introduced during collection and processing of biological samples [10,13,48,56]. Contaminating DNA originates from numerous sources, including cells or small fragments of DNA from the same environment, sampling equipment, laboratory reagents and equipment, technicians, and so on [10,12,13]. Although sterilisation lyses and kills microbial cells, their DNA can be broken up into smaller fragments, which can still be extracted and amplified by PCR. Unsurprisingly, HTS is more sensitive than traditional culture methods in detecting contamination, and if no controls are used to identify contamination, the results may be confounded [57]. Using these HTS approaches, contamination has driven spurious conclusions and drastically altered several reported biological discoveries in the literature [56,58,59]. Especially in low microbial biomass samples, contamination and bias from contaminants has already had severe impacts on the fidelity of reported results [12]. As the application of microbiome research shifts towards clinical use, contamination needs to be monitored and accounted for to prevent patient misdiagnoses.

Lessons from past low biomass microbiome studies

Despite the potential benefits, some researchers working with low microbial biomass samples do not use the necessary precautions to control or limit contamination. Initial studies of the gut microbiome (a source of high microbial biomass) were less prone to contamination, as contaminating DNA was negligible compared to the endogenous microbial DNA (DNA belonging to the sample of interest) present in the sample, and thus downstream analyses were not severely impacted [60]. Researchers then followed similar protocols to examine low microbial biomass samples, including placental tissue [58], nipple aspirate [61], and tumours [62,63], overlooking the possible negative impacts of contamination. This resulted in generating data from low microbial biomass samples that were overwhelmed by the amount of exogenous DNA relative to the endogenous DNA and consequently, contamination was falsely reported as a true result in some studies [56,58,59]. These past errors are lessons for the future, but we need to ensure they are not

repeated, especially when applying microbiome techniques in diagnostic testing.

Lack of controls in microbiome analysis of medical samples has already caused issues within the field. For example, initial investigations into the placenta microbiome sought to answer many questions about infant development and preterm birth during pregnancy [58]. However, an initial study failed to use the necessary precautions to monitor and minimise contamination. Non-template controls were introduced during the extractions, but only a subset were sequenced. Sequences found in the negative controls were not critically compared to those from the biological samples. Additionally, no environmental controls were collected or analysed, preventing the detection of environmental contaminants (e.g. microbes in the air, from the technician, or on sterile dissection equipment). Lastly, the limit of detection was not established, preventing the researchers' ability to determine if a reliable signal could be detected. Contaminant microbial species could therefore not be correctly identified or assessed within the placenta samples to determine if the microbial signature from the placenta was truly endogenous. In 2016, Lauder *et al* [56] replicated the study of the placenta microbiome and found that the microbial placenta profiles resembled those of extraction blank control (EBC) samples and air samples, suggesting that the placenta was likely sterile and did not contain a diverse microbial signal. Since then, many other well-controlled studies have further supported the lack of a placenta microbiome [59,64–67]. These studies highlight the importance of having robust protocols and controls to avoid spurious conclusions.

Sources of contamination and biases in low microbial biomass samples

To better control for contamination and biases, a solid understanding of when and how these factors arise is needed. While contaminating DNA and biases can be introduced at any stage in the sample preparation and analysis process, three predominant sources originate from sampling procedures and the laboratory: (1) doctors, nurses, technicians, and so on; (2) environments; and (3) reagents and equipment (Figure 1). More recently, there has been an increased awareness of contamination and biases introduced into low microbial biomass samples, but widespread inclusion and analysis of controls still needs to be broadly implemented and reported [12,57,68]. Below, we review the current information on sources of contamination and biases and provide recommendations for reducing these confounding factors when using microbiome assessments in diagnostic testing.

Sample collectors and technicians in the clinic and the laboratory

Professionals who collect and process samples may introduce their own microbial DNA into a sample (Figure 1), which is especially problematic when multiple individuals are collecting and processing samples for the same study or test. Most clinical protocols do not address microbial DNA contamination introduced from sample collectors and laboratory technicians, leading to the possibility that signals from individuals may override the signal from the biological sample if precautions and procedures are not put in place. Even the best trained sample collectors and laboratory technicians will introduce contamination into the samples; this is not an error on the part of the individual but now an appreciated signal within microbiome research. For example, sample collectors generally wear gloves and occasionally face masks to protect their health; however, DNA can be shed from unclean gloves, a mask, a lab coat, or the collector's unexposed skin [69] – all sources can now be detected using new HTS methods. To decrease the strength of these signals, all sample collectors and laboratory technicians should wear gloves, face masks, lab coats, and other appropriate clothing to prevent contamination from the technician [12]. This problem is also not mitigated by the use of robotics, rather than human technicians. Automated robots produce more well-to-well contamination (cross-contamination) across samples [70], which leads to specific batch and robot contamination. Additionally, bias can be introduced into samples of low microbial biomass through the technique applied by multiple sample collectors. Simpkins *et al* [71] showed that there was a significant difference in microbial species collected between technicians of varied experience (e.g. how hard a technician, doctor, or nurse presses down on a skin swab to collect a sample). Future research needs to fully characterise the contamination and bias introduced from sample collectors and laboratory technicians and understand how this can influence or alter microbial signatures in low microbial biomass samples.

The sampling and laboratory environments

Low microbial biomass samples are also influenced by the environment where they were collected (Figure 1) [60]. Microbial profiles in different built and outdoor environments are unique, and even different laboratories with similar technicians and purposes have unique profiles [10,13]. Additionally, the same lab can have

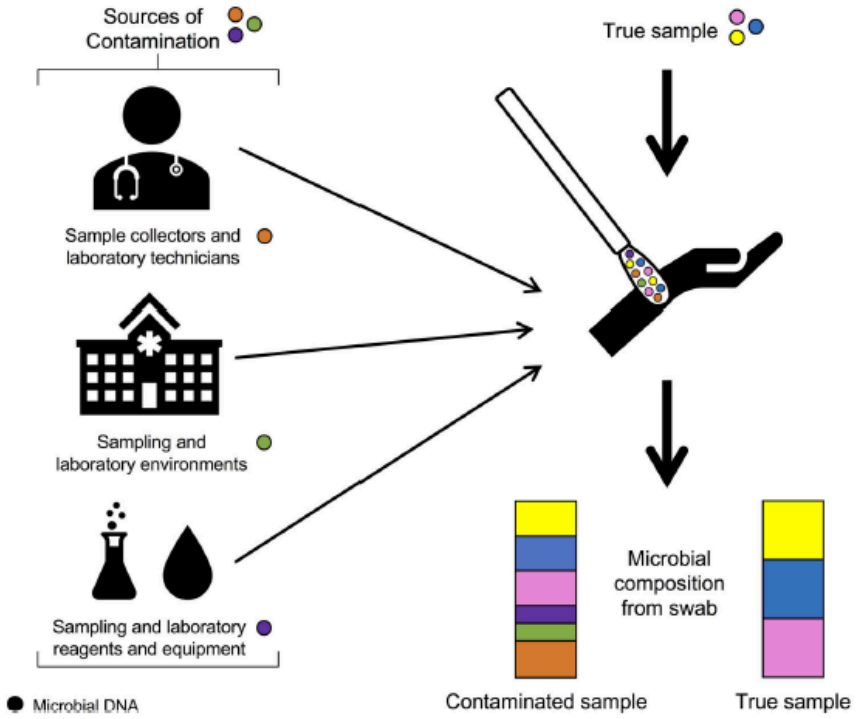

Figure 1. Microbial DNA from sample collectors and laboratory technicians, the environment, and reagents and equipment can contaminate pathological samples, which can distort the microbial profile of the sample.

different contaminant profiles that change according to the season, year, or technician [13]. For example, indoor environments, such as hospitals, resemble a microbial profile similar to that of the inhabiting individual(s) [72,73]. Despite this, the exact mechanisms that result in environmental contamination in low biomass samples are poorly understood. It is likely that the majority of microbes in a sampling environment exist in the air, in bioaerosols or on surfaces.

Reagents and equipment from sampling and laboratory processes

Low microbial biomass samples are highly susceptible to contaminating DNA present in reagents and equipment used for sample collection, DNA extractions, and library preparation (Figure 1). It is well recognised that reagents and equipment, including 'sterile' water, contain microbial DNA [13,48,74,75]. Salter *et al* [10] demonstrated that different extraction kits showed a different amount and composition of microbial DNA, and even the 'cleanest' kits contain reliable DNA signatures. This becomes problematic for low microbial biomass samples as the contaminating microbes can overwhelm the signal from endogenous DNA content [76,77]. During amplification, the exogenous DNA becomes preferentially amplified, lowering the chance of observing the true microbial signal. Varying concentrations of input DNA can also increase the number of artefacts during DNA amplification, which has a significant impact on low microbial biomass samples. Chafee *et al* [11] showed that input DNA concentration biases low microbial biomass samples by increasing duplication levels, favouring AT-rich sequences, and overall biasing population levels. Generally, low microbial biomass samples have low DNA input levels for amplification reactions, which results in over-amplification of the template DNA with high levels of duplication [11,78]. This can significantly bias results, suggesting that a single species or genus, perhaps even a contaminant species, is more dominant in the sample than it truly is. Consequently, these factors cause significant issues when reconstructing microbial communities for diagnostic purposes, especially when attempting to identify and quantify the microbes present.

Recommendations to avoid potential contamination and biases

Below, we discuss the current recommendations to avoid potential contamination and biases when attempting to introduce microbiome analysis into diagnostic testing. For a quick reference guide, the RIDE checklist is also available as minimum standards for low microbial biomass samples [12].

1. *Include controls from the sampling and laboratory environments, equipment, and reagents*

To detect environmental microbes, air samples of the collection room should be gathered. Additional controls should include swabs of the hospital room (e.g. seats, walls, benches) before collection of biological samples. During laboratory processes, EBCs (controls that are run in parallel to the samples during the extraction, but do not include sample DNA) and no-template amplification controls (NTCs; amplification reactions without any extracted DNA) should be included with every extraction or amplification batch, respectively, to monitor DNA incorporated into the sample via laboratory reagents. Additional controls to monitor any tool, substance, or individual that comes in contact with the samples may also be required for specific cases. It should be noted that additional amplification or strategies (e.g. the introduction of carrier DNA) [79] may be required to detect contaminating DNA and should be performed if necessary. Additionally, extraction methods that are known to contain fewer contaminants and are optimised for low microbial biomass samples (e.g. Mo Bio PowerMag with a ClearMag bead [54]) should be used. Contaminating DNA simply cannot be avoided and needs to be monitored to ensure it is not driving the signals present in the collected samples.

2. *Minimise the amount of microbial and human contamination being introduced into samples*

Currently, there are no established protocols to minimise the introduction of microbial and human DNA into low microbial biomass samples in a clinical setting. However, other fields have methods that could be adopted or modified here, including ancient DNA protocols or those utilised in levels of high biosecurity [12,13]. Generally, introducing contaminating microbial DNA could be decreased by wearing clothing that covers exposed skin, such as wearing face masks and gloves. This is similar to the techniques used in ultra-clean labs [12], where technicians are required to wear full disposable body suits, shoe covers, face masks, a plastic visor, and multiple pairs of gloves to minimise the introduction of human and bacterial DNA. Human DNA, which can overwhelm microbial DNA of many low microbial biomass samples, can also be depleted using methods such as Benzonase [33]. Additionally, DNA in reagents and on equipment can be minimised by irradiating reagents with ultraviolet radiation [80]. As microbial DNA is ubiquitous, these strategies can aid in reducing the contaminating DNA profile relative to the biological one, but it is critical to understand that current research supports the idea that it cannot be completely eliminated and should be monitored for the best results.

3. *Consistency and randomisation*

Inconsistency in sample collection (e.g. differences in pressure/duration when swabbing skin) can introduce biases. To minimise these biases, the best practice is to reduce the number of sample collectors. However, this is not possible at some diagnostic testing sites, so the collection and processing of samples should be randomised between sample collectors and laboratory technicians to minimise any possible biases. Alternatively, additional controls may be needed to account for this bias. Furthermore, standardised training and explicit sampling and processing protocols should be implemented to ensure that samples are collected and processed as similarly as possible. In the laboratory, positive controls with non-biological DNA fragments (i.e. mock communities) can be utilised to ensure that technician bias during processing is limited and technician-specific contamination can be more readily detected [12]; however, the user must be careful that these positive controls are handled with consideration to their microbial biomass level. For example, it would be ill advised to add DNA from a positive, high biomass control first before processing low biomass samples in the same batch.

4. *Use quantitative laboratory methods*

In the past, the successful acquisition of DNA from a sample was determined by the presence or absence of bands on a gel. However, this technique is not sensitive to the level of DNA present in many control samples [54]. Quantitative methods, such as fluorescent probes (e.g. PicoGreen) or quantitative PCR, should instead be utilised to determine the total DNA present in both samples and controls [54,56]. This process verifies that the biological sample has more DNA than the controls. It is also recommended that low input samples should be sequenced at a higher depth to capture a sufficient number of unique sequences [11]. Further research is still needed to examine how low copy number biases can be best avoided in low biomass research.

5. *Incorporating bioinformatics approaches to assess or remove contamination*

Recently, several bioinformatic methods have been developed to track and remove contaminating DNA that influences low microbial biomass samples. First, the limit of detection can be applied using positive and negative controls [54], as described above. Additionally, individual contaminating species can be tracked from their source (i.e. environmental controls, reagents, equipment, EBCs, NTCs, *etc*.) using their exact sequence [81] and then can be subsequently removed or identified using publicly available programs, such as SourceTracker [82] and Decontam [64]. While programs can be used to remove or track contaminants, the procedures do not mitigate the requirement to monitor and examine contaminants during sample collection and laboratory processes.

Conclusion

In the near future, the microbiome will become an important asset for diagnostics and treating human diseases. However, most pathology samples contain low numbers of microbial cells, which makes them difficult to extract, amplify, and analyse in a microbiome context. These low microbial biomass samples are more prone to contamination and biases from the sample collectors and laboratory technicians, environment, reagents, and equipment. To avoid misdiagnoses and incorrect treatments, clear procedures and guidelines are needed for pathologists and clinicians to mitigate contamination and biases. Several groups, such as ancient DNA and forensic researchers, have developed similar guidelines and protocols for analyses done in their field [80,83,84]. Proper controls, methodological precautions, and appropriate analytical strategies will allow the benefits of microbiome analysis to be appreciated for diagnostic testing in the future.

Acknowledgements

We thank the thesis writing group from the Australian Centre for Ancient DNA for providing edits to the original manuscript.

Author contributions statement

CAS and LSW conceptualised the review. CAS wrote the original manuscript. LSW and RE critically reviewed and edited all versions of the manuscript.

References

1. Cohen J. 1 – The evolution of Koch's postulates. In: *Infectious Diseases* (4th edn), Cohen J, Powderly WG, Opal SM, (Eds). Elsevier: Amsterdam, 2017; 1–3.e1.
2. de Vos MGJ, Zagorski M, McNally A, *et al*. Interaction networks, ecological stability, and collective antibiotic tolerance in polymicrobial infections. *Proc Natl Acad Sci U S A* 2017; **114**: 10666–10671.
3. Opal SM. 69 – The potential role of infectious agents in diseases of unknown etiology. In: *Infectious Diseases* (4th edn), Cohen J, Powderly WG, Opal SM (Eds). Elsevier: Amsterdam, 2017; 625–630.e1.
4. Langelier C, Kalantar KL, Moazed F, *et al*. Integrating host response and unbiased microbe detection for lower respiratory tract infection diagnosis in critically ill adults. *Proc Natl Acad Sci U S A* 2018; **115**: E12353–E12362.
5. Sandoval-Motta S, Aldana M, Martínez-Romero E, *et al*. The human microbiome and the missing heritability problem. *Front Genet* 2017; **8**: 80.
6. NIH HMP Working Group, Peterson J, Garges S, *et al*. The NIH human microbiome project. *Genome Res* 2009; **19**: 2317–2323.
7. Petersen C, Round JL. Defining dysbiosis and its influence on host immunity and disease. *Cell Microbiol* 2014; **16**: 1024–1033.
8. Stiemsma LT, Turvey SE. Asthma and the microbiome: defining the critical window in early life. *Allergy Asthma Clin Immunol* 2017; **13**: 3.
9. Slingerland AE, Stein-Thoeringer CK. Microbiome and diseases: neurological disorders. In: *The Gut Microbiome in Health and Disease*, Haller D (Ed). Springer International Publishing. 2018; 295–310.
10. Salter SJ, Cox MJ, Turek EM, *et al*. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol* 2014; **12**: 87.
11. Chafee M, Maignien L, Simmons SL. The effects of variable sample biomass on comparative metagenomics. *Environ Microbiol* 2015; **17**: 2239–2253.
12. Eisenhofer R, Minich JJ, Marotz C, *et al*. Contamination in low microbial biomass microbiome studies: issues and recommendations. *Trends Microbiol* 2019; **27**: 105–117.
13. Weyrich LS, Farrer AG, Eisenhofer R, *et al*. Laboratory contamination over time during low-biomass sample analysis. *Mol Ecol Resour* 2019; **19**: 982–996.
14. Washington JA. Principles of diagnosis. In: *Medical Microbiology* (4th edn), Baron S (Ed). University of Texas Medical Branch at Galveston: Galveston, TX, 1996.
15. Bursle E, Robson J. Non-culture methods for detecting infection. *Aust Prescr* 2016; **39**: 171–175.
16. Michael CA, Dominey-Howes D, Labbate M. The antimicrobial resistance crisis: causes, consequences, and management. *Front Public Health* 2014; **2**: 145.
17. Beck ES, Ramachandran PS, Khan LM, *et al*. Clinicopathology conference: 41-year-old woman with chronic relapsing meningitis. *Ann Neurol* 2019; **85**: 161–169.
18. Ni P-X, Ding X, Zhang Y-X, *et al*. Rapid detection and identification of infectious pathogens based on high-throughput sequencing. *Chin Med J (Engl)* 2015; **128**: 877–883.

19. Byndloss MX, Bäumler AJ. The germ-organ theory of non-communicable diseases. *Nat Rev Microbiol* 2018; **16**: 103–110.
20. Kuntz TM, Gilbert JA. Introducing the microbiome into precision medicine. *Trends Pharmacol Sci* 2017; **38**: 81–91.
21. Petrosino JF. The microbiome in precision medicine: the way forward. *Genome Med* 2018; **10**: 12.
22. Kumar R, Eipers P, Little RB, *et al*. Getting started with microbiome analysis: Sample acquisition to bioinformatics. *Curr Protoc Hum Genet* 2014; **82**: 18.8.1–18.8.29.
23. Knight R, Vrbanac A, Taylor BC, *et al*. Best practices for analysing microbiomes. *Nat Rev Microbiol* 2018; **16**: 410–422.
24. Sharpton TJ. An introduction to the analysis of shotgun metagenomic data. *Front Plant Sci* 2014; **5**: 209.
25. Maurice CF, Haiser HJ, Turnbaugh PJ. Xenobiotics shape the physiology and gene expression of the active human gut microbiome. *Cell* 2013; **152**: 39–50.
26. Hasman H, Saputra D, Sicheritz-Ponten T, *et al*. Rapid whole-genome sequencing for detection and characterization of microorganisms directly from clinical samples. *J Clin Microbiol* 2014; **52**: 139–146.
27. Couto N, Schuele L, Raangs EC, *et al*. Critical steps in clinical shotgun metagenomics for the concomitant detection and typing of microbial pathogens. *Sci Rep* 2018; **8**: 13767.
28. Qin J, Li R, Raes J, *et al*. A human gut microbial gene catalog established by metagenomic sequencing. *Nature* 2010; **464**: 59–65.
29. Burel JG, Peters B. Discovering transcriptional signatures of disease for diagnosis versus mechanism. *Nat Rev Immunol* 2018; **18**: 289–290.
30. Holcomb ZE, Tsalik EL, Woods CW, *et al*. Host-based peripheral blood gene expression analysis for diagnosis of infectious diseases. *J Clin Microbiol* 2017; **55**: 360–368.
31. Ramilo O, Mejias A. Host transcriptomics for diagnosis of infectious diseases: one step closer to clinical application. *Eur Respir J* 2017; **49**: 1700993.
32. Birdsell DN, Özsürekci Y, Rawat A, *et al*. Coinfections identified from metagenomic analysis of cervical lymph nodes from tularemia patients. *BMC Infect Dis* 2018; **18**: 319.
33. Nelson MT, Pope CE, Marsh RL, *et al*. Human and extracellular DNA depletion for metagenomic analysis of complex clinical infection samples yields optimized viable microbiome profiles. *Cell Rep* 2019; **26**: 2227–2240.e5.
34. Li B, Webster TJ. Bacteria antibiotic resistance: new challenges and opportunities for implant-associated Orthopaedic infections. *J Orthop Res Off Publ Orthop Res Soc* 2018; **36**: 22–32.
35. Paharik AE, Schreiber HL, Spaulding CN, *et al*. Narrowing the spectrum: the new frontier of precision antimicrobials. *Genome Med* 2017; **9**: 110.
36. Selwyn FP, Cui JY, Klaassen CD. RNA-Seq quantification of hepatic drug processing genes in germ-free mice. *Drug Metab Dispos* 2015; **43**: 1572–1580.
37. Liu Z, Klümper U, Liu Y, *et al*. Metagenomic and metatranscriptomic analyses reveal activity and hosts of antibiotic resistance genes in activated sludge. *Environ Int* 2019; **129**: 208–220.
38. von Hertzen L, Hanski I, Haahtela T. Natural immunity. *EMBO Rep* 2011; **12**: 1089–1093.
39. Edwards MR, Saglani S, Schwarze J, *et al*; Members of the EARIP WP2 Working Group. Addressing unmet needs in understanding asthma mechanisms: from the European Asthma Research and Innovation Partnership (EARIP) Work Package (WP)2 collaborators. *Eur Respir J* 2017; **49**: 1602448
40. Rosenblum MD, Remedios KA, Abbas AK. Mechanisms of human autoimmunity. *J Clin Invest* 2015; **125**: 2228–2233.
41. Wilson BC, Vatanen T, Cutfield WS, *et al*. The super-donor phenomenon in fecal microbiota transplantation. *Front Cell Infect Microbiol* 2019; **9**: 2.
42. Gilbert JA, Blaser MJ, Caporaso JG, *et al*. Current understanding of the human microbiome. *Nat Med* 2018; **24**: 392–400.
43. Staley C, Kaiser T, Khoruts A. Clinician guide to microbiome testing. *Dig Dis Sci* 2018; **63**: 3167–3177.
44. O'Dwyer DN, Dickson RP, Moore BB. The lung microbiome, immunity and the pathogenesis of chronic lung disease. *J Immunol Baltim Md* 2016; **1950**: 4839–4847.
45. Oh J, Byrd AL, Park M, *et al*. Temporal stability of the human skin microbiome. *Cell* 2016; **165**: 854–866.
46. Crowder MF, Conly J, Surette M. A microbiome approach to sepsis: development and case-study application of novel methods for detection and isolation of microbes from whole blood. *Crit Care* 2012; **16**: P111.
47. Gottschick C, Deng Z-L, Vital M, *et al*. The urinary microbiota of men and women and its changes in women during bacterial vaginosis and antibiotic treatment. *Microbiome* 2017; **5**: 99.
48. Glassing A, Dowd SE, Galandiuk S, *et al*. Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of microbiota in low bacterial biomass samples. *Gut Pathog* 2016; **8**: 24.
49. Bacteria in Snow MV, Ice G. In: *Psychrophiles: From Biodiversity to Biotechnology*, Margesin R, Schinner F, Marx J-C, *et al*. (Eds). Springer: Berlin, Heidelberg, 2008; 31–50.
50. Zhong Z-P, Solonenko NE, Gazitúa MC, *et al*. Clean low-biomass procedures and their application to ancient Ice Core microorganisms. *Front Microbiol* 2018; **9**: 1094.
51. Human Microbiome Project Consortium. A framework for human microbiome research. *Nature* 2012; **486**: 215–221.
52. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* 2012; **486**: 207–214.
53. Thompson LR, Sanders JG, McDonald D, *et al*. A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* 2017; **551**: 457–463.
54. Minich JJ, Zhu Q, Janssen S, *et al*. KatharoSeq enables high-throughput microbiome analysis from low-biomass samples. *mSystems* 2018; **3**: e00218–e00217.
55. Zinter MS, Mayday MY, Ryckman KK, *et al*. Towards precision quantification of contamination in metagenomic sequencing experiments. *Microbiome* 2019; **7**: 62.
56. Lauder AP, Roche AM, Sherrill-Mix S, *et al*. Comparison of placenta samples with contamination controls does not provide evidence for a distinct placenta microbiota. *Microbiome* 2016; **4**: 29.
57. Kim D, Hofstaedter CE, Zhao C, *et al*. Optimizing methods and dodging pitfalls in microbiome research. *Microbiome* 2017; **5**: 52.
58. Aagaard K, Ma J, Antony KM, *et al*. The placenta harbors a unique microbiome. *Sci Transl Med* 2014; **6**: 237ra65.

59. Leiby JS, McCormick K, Sherrill-Mix S, *et al*. Lack of detection of a human placenta microbiome in samples from preterm and term deliveries. *Microbiome* 2018; **6**: 196.
60. Weiss S, Amir A, Hyde ER, *et al*. Tracking down the sources of experimental contamination in microbiome studies. *Genome Biol* 2014; **15**: 564.
61. Chan AA, Bashir M, Rivas MN, *et al*. Characterization of the microbiome of nipple aspirate fluid of breast cancer survivors. *Sci Rep* 2016; **6**: 28061.
62. Hieken TJ, Chen J, Hoskin TL, *et al*. The microbiome of aseptically collected human breast tissue in benign and malignant disease. *Sci Rep* 2016; **6**: 30751.
63. Xuan C, Shamonki JM, Chung A, *et al*. Microbial dysbiosis is associated with human breast cancer. *PLoS One* 2014; **9**: e83744.
64. Davis NM, Proctor DM, Holmes SP, *et al*. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome* 2018; **6**: 226.
65. Lim ES, Rodriguez C, Holtz LR. Amniotic fluid from healthy term pregnancies does not harbor a detectable microbial community. *Microbiome* 2018; **6**: 87.
66. Theis KR, Romero R, Winters AD, *et al*. Does the human placenta delivered at term have a microbiota? Results of cultivation, quantitative real-time PCR, 16S rRNA gene sequencing, and metagenomics. *Am J Obstet Gynecol* 2019; **220**: 267.e1–267.e39.
67. de GMC, Lager S, Sovio U, *et al*. Human placenta has no microbiome but can contain potential pathogens. *Nature* 2019; **572**: 329–334.
68. Pollock J, Glendinning L, Wisedchanwet T, *et al*. The madness of microbiome: attempting to find consensus "best practice" for 16S microbiome studies. *Appl Env Microbiol* 2018; **84**: e02627–17.
69. You R, Cui W, Chen C, *et al*. Measuring the short-term emission rates of particles in the 'personal cloud' with different clothes and activity intensities in a sealed chamber. *Aerosol Air Qual Res* 2013; **13**: 911–921.
70. Minich JJ, Sanders JG, Amir A, *et al*. Quantifying and understanding well-to-well contamination in microbiome research. *mSystems* 2019; **4**: e00186–19.
71. Simpkins C, Van Sluys M, Hero J-M. Swabber effect: swabbing technique affects the detectability of *Batrachochytrium dendrobatidis*. *Herpetol Rev* 2014; **45**: 443–445.
72. Fierer N, Lauber CL, Zhou N, *et al*. Forensic identification using skin bacterial communities. *Proc Natl Acad Sci U S A* 2010; **107**: 6477–6481.
73. Kelley ST, Gilbert JA. Studying the microbiology of the indoor environment. *Genome Biol* 2013; **14**: 202.
74. Shen H, Rogelj S, Kieft TL. Sensitive, real-time PCR detects low-levels of contamination by *Legionella pneumophila* in commercial reagents. *Mol Cell Probes* 2006; **20**: 147–153.
75. Kéki Z, Grébner K, Bohus V, *et al*. Application of special oligotrophic media for cultivation of bacterial communities originated from ultrapure water. *Acta Microbiol Immunol Hung* 2013; **60**: 345–357.
76. Tanner MA, Goebel BM, Dojka MA, *et al*. Specific ribosomal DNA sequences from diverse environmental settings correlate with experimental contaminants. *Appl Environ Microbiol* 1998; **64**: 3110–3113.
77. Willerslev E, Hansen AJ, Poinar HN. Isolation of nucleic acids and cultures from fossil ice and permafrost. *Trends Ecol Evol* 2004; **19**: 141–147.
78. Solonenko SA, Ignacio-Espinoza JC, Alberti A, *et al*. Sequencing platform and library preparation choices impact viral metagenomes. *BMC Genomics* 2013; **14**: 320.
79. Xu Z, Zhang F, Xu B, *et al*. Improving the sensitivity of negative controls in ancient DNA extractions. *Electrophoresis* 2009; **30**: 1282–1285.
80. Llamas B, Valverde G, Fehren-Schmitz L, *et al*. From the field to the laboratory: controlling DNA contamination in human ancient DNA research in the high-throughput sequencing era. *STAR: Sci Technol Archaeol Res* 2017; **3**: 1–14.
81. Amir A, McDonald D, Navas-Molina JA, *et al*. Deblur rapidly resolves single-nucleotide community sequence patterns. *mSystems* 2017; **2**: e00191–16.
82. Knights D, Kuczynski J, Charlson ES, *et al*. Bayesian community-wide culture-independent microbial source tracking. *Nat Methods* 2011; **8**: 761–763.
83. Poinar HN. The top 10 list: criteria of authenticity for DNA from ancient and forensic samples. *Int Congr Ser* 2003; **1239**: 575–579.
84. Fitzpatrick R, Raven M. *Guidelines for Conducting Criminal and Environmental Soil Forensic Investigations (Version 10.1)*. Centre for Australian Forensic Soil Science: Adelaide, South Australia, 2017.

Chapter II

Moving beyond the gut microbiome: combining systems biology and multi-site microbiome analyses to combat non-communicable diseases

Statement of Authorship

Title of Paper	The human microbiome is more than just the gut: considering a holistic approach for non-communicable disease prevention, progression, and treatment
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ☒ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Unpublished and unsubmitted work

Principal Author

Name of Principal Author (Candidate)	Caitlin A. Selway		
Contribution to the Paper	Conceptualised and wrote the manuscript.		
Overall percentage (%)	70%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	8/10/20

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);
ii. permission is grantec for the candidate in include the publication in the thesis; and
iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Jaya Sudarpa		
Contribution to the Paper	Conceptualised and critically reviewed and edited the manuscript.		
Signature		Date	7/10/20

Name of Co-Author	Laura S. Weyrich		
Contribution to the Paper	Conceptualised and critically reviewed and edited the manuscript.		
Signature		Date	7/10/2020

Moving beyond the gut microbiome: combining systems biology and multi-site microbiome analyses to combat non-communicable diseases

Caitlin A. Selway[1*], Jaya Sudarpa[2], Laura S. Weyrich[1,3,4*]

Affiliations:

[1] Department of Molecular and Biomedical Science, University of Adelaide, Adelaide, SA 5005 Australia

[2] Nova Systems, Adelaide, SA 5031 Australia

[3] Department of Anthropology, The Pennsylvania State University, University Park, PA 16802 USA

[4]Huck Institutes of Life Sciences, The Pennsylvania State University, University Park, PA 16802 USA

Corresponding authors:

[*] Caitlin A Selway, Department of Molecular and Biomedical Science, University of Adelaide, Adelaide SA 5005, Australia. caitlin.selway@adelaide.edu.au or selway.caitlinalyssa@gmail.com

[*] Laura S Weyrich, Department of Anthropology, Pennsylvania State University, University Park, PA 16802 USA. lsw132@psu.edu

Main Text

For the past 70 years, the prevalence of non-communicable diseases (NCDs) has steadily risen, reaching the point where they are now a leading contributor of death for 70% of the world's population.[1] This key public health crisis costs countries trillions of dollars in treatment, disability, and death,[2] and has prompted the United Nations to adopt a target of a 1/3 reduction in premature NCD-related deaths by

2030 (Sustainable Development Goal 3.4).[3] Despite this, there is still a critical lack of understanding about the initiation and persistence of NCDs, although research has identified a complex web of contributing factors, including genetics, the environment, lifestyle and the human microbiome—the microbial communities of the human body, their genetic content, and their surrounding micro-environment.

Shifts in the human microbiome have now been linked to almost every NCD, including cardiovascular diseases, cancers, and respiratory diseases.[4–6] However, the causal relationships between the human microbiome and disease remain largely unknown. Microbiome research on NCDs has been dominated by studies examining changes in the gut microbiome, as the gut contains the largest reservoir of microbes and has many direct and indirect relationships with other systems of the body, such as the immune, digestive, circulatory, integumentary, neuroendocrine, and central nervous systems.[7] Despite the advances in understanding microbiome relationships with the human body from gut studies, other body sites are also critical to understanding NCDs, and the sheer explosion in gut microbiome research (1168 publications in 2020 alone[a] compared to 680 publications for all other body sides[b]) has dwarfed microbiome studies in interconnected body sites (*e.g.,* mouth, lungs, skin, vagina, and even the brain). Considering the complex factors throughout the body that underpin NCDs, further research on the interrelationships between microbiomes throughout the body is needed to shed new light on the initiation, progression, and potential cures for NCDs.

Microbiome research in non-gut body sites has already provided valuable insights into NCDs. For example, oral disease—one of the most common NCDs to affect an individual throughout their lifetime—is linked both to the oral microbiome and to other systemic NCDs. Overgrowth of particular opportunistic oral pathogens found during periodontal disease, such as *Porphyromonas gingivalis* or *Fusobacterium nucleatum*, can disseminate and colonise elsewhere in the body, initiating other NCDs, including Alzheimer's disease or colorectal cancer,

Scopus search on the 14 Feb 2021

[a] TITLE-ABS-KEY (human) AND TITLE-ABS ("microbiome" OR "microbiota") AND TITLE-ABS-KEY ("gut" OR "fecal" OR "faecal" OR "stool" OR "intestinal" OR "gastric" OR "stomach" OR "enterotype") AND NOT TITLE-ABS-KEY (mice OR larvae OR fish OR "bee" OR insect OR "panda" OR animal OR "ticks" OR "soil" OR "sedimentary" OR "ice" OR "leaf" OR "wastewater" OR "drains" OR water OR "bathroom" OR "fermentation") AND NOT TITLE-ABS ("review" OR "perspectives" OR "model*" OR "in silico" OR "in vitro" OR "github" OR "publication*") AND (LIMIT-TO (SRCTYPE , "j")) AND (LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (PUBYEAR , 2020)) AND (LIMIT-TO (LANGUAGE , "English")) AND (LIMIT-TO (EXACTKEYWORD , "Human"))

[b] TITLE-ABS-KEY (human) AND TITLE-ABS ("microbiome" OR "microbiota") AND NOT TITLE-ABS-KEY ("gut" OR "fecal" OR "faecal" OR "stool" OR "intestinal" OR "gastric" OR "stomach" OR "enterotype") AND NOT TITLE-ABS-KEY (mice OR larvae OR fish OR "bee" OR insect OR "panda" OR animal OR "ticks" OR "soil" OR "sedimentary" OR "ice" OR "leaf" OR "wastewater" OR "drains" OR water OR "bathroom" OR "fermentation") AND NOT TITLE-ABS ("review" OR "perspectives" OR "model*" OR "in silico" OR "in vitro" OR "github" OR "publication*") AND (LIMIT-TO (SRCTYPE , "j")) AND (LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (PUBYEAR , 2020)) AND (LIMIT-TO (LANGUAGE , "English")) AND (LIMIT-TO (EXACTKEYWORD , "Human"))

respectively.[4,8] Further, Streptococci species in the mouth—bacteria shared across nearly all global populations—can also escape the mouth and bind heart tissue, leading to heart disease.[9] Similarly, the skin microbiota also contributes to NCDs in unique ways. As the skin microbiota and the immune system are tightly coupled, low skin microbial diversity and the enrichment of *Staphylococcus spp*. and opportunistic fungi can exacerbate atopic dermatitis symptoms by initiating a more severe immune response.[10] These are several examples of what are likely to be many, yet unknown, associations between the microbial communities of 'non-gut' sites, a systemic response, and a subsequent NCD.

In addition to exploring additional body sites, the microbial connectedness between these sites needs to be further examined. As the human body is a complex system, Systems Theory (defined by the International Council on Systems Engineering, as "…an **arrangement** of parts or elements that **together exhibit behaviour** or meaning that the **individual constituents** do not…") could be applied to understand how these microbiomes, physically isolated and sometimes seemingly unrelated, can be interconnected. For example, the gastrointestinal tract and brain are distinct and physically distant sites of the human body but are interconnected through the hypothalamic-pituitary-adrenal axis and nervous system (altogether, termed the 'gut-brain axis'). Gut metabolites (*e.g.*, short-chain fatty acids), which are a product of gut microbiota, can influence the brain and potentially contribute to mental and neurodegenerative diseases either via the immune system or direct signalling of the nervous system.[11,12] This example highlights the need for more studies investigating microbial connectedness across all body sites.

This systems biology concept has been applied to complex areas of the human body, including immunology and cancer biology, which can help to further understand signalling pathways[13] and identify more targeted treatments and therapies.[14] Systems biology concepts can be applied to the microbiome to: 1) understand the interactions between all components locally (microbiome, genome, transcriptome, metabolome, etc.) for a single body site of interest; and 2) globally assess the interactions of all elements throughout the system that is the human body. To provide the best understanding of disease, the latter approach is the most needed. Practical applications to target a microbiome systems approach could include mining of existing datasets to model whole system interactions; collecting a diverse range of samples from across the body of a single individual and performing

network analyses; collecting both microbiome and genomic information to perform microbiome-wide association studies; or carrying out multi-omics approaches (*e.g.,* transcriptomics, proteomics, metabolomics, *etc.*) on one or more sites of the human body. Each of these examples need to be supplemented with extensive metadata to associate the findings with the environment, lifestyle factors, or NCD outcomes, and we acknowledge that there may be additional requirements for the ethical, legal, and social implications of this work.

This systems biology approach will reveal emergent properties—properties that are realised not by constituent elements, but only by the system once analysed holistically—that can be identified and targeted to combat disease. Systems biology provides a holistic conceptual approach that connects different aspects of the human body to understand the intricacies of health and disease. Therefore, application of systems biology to all sites of the human body should be initiated now so that its benefits can be applied to start combating NCDs. Better informed research can be used to tackle increasingly complex problems, such as understanding microbe-microbe competition, microbe-host interactions, or improving gene ontology for microbes in non-gut sites. Using these approaches, we can start to unravel the underlying factors that contribute to NCDs, and in the future, strive towards reducing morbidity and premature mortality.

References

1. Bennett, J. E. *et al.* NCD Countdown 2030: worldwide trends in non-communicable disease mortality and progress towards Sustainable Development Goal target 3.4. *The Lancet* **392**, 1072–1088 (2018).

2. Chen, S., Kuhn, M., Prettner, K. & Bloom, D. E. The macroeconomic burden of noncommunicable diseases in the United States: Estimates and projections. *PLoS One* **13**, (2018).

3. Transforming our world: the 2030 Agenda for Sustainable Development .:. Sustainable Development Knowledge Platform. https://sustainabledevelopment.un.org/post2015/transformingourworld.

4. Pietiäinen, M., Liljestrand, J. M., Kopra, E. & Pussinen, P. J. Mediators between oral dysbiosis and cardiovascular diseases. *European Journal of Oral Sciences* **126 Suppl 1**, 26–36 (2018).

5. Xavier, J. B. *et al.* The Cancer Microbiome: Distinguishing Direct and Indirect Effects Requires a Systemic View. *Trends in Cancer* **6**, 192–204 (2020).

6. Rivas, M. N., Crother, T. R. & Arditi, M. The Microbiome in Asthma. *Curr Opin Pediatr* **28**, 764–771 (2016).

7. Zheng, D., Liwinski, T. & Elinav, E. Interaction between microbiota and immunity in health and disease. *Cell Research* **30**, 492–506 (2020).

8. Teles, F. R. F., Alawi, F., Castilho, R. M. & Wang, Y. Association or Causation? Exploring the Oral Microbiome and Cancer Links. *J Dent Res* 0022034520945242 (2020) doi:10.1177/0022034520945242.

9. Cahill, T. J. & Prendergast, B. D. Infective endocarditis. *The Lancet* **387**, 882–893 (2016).

10. Oh, J. *et al.* The altered landscape of the human skin microbiome in patients with primary immunodeficiencies. *Genome Res.* **23**, 2103–2114 (2013).

11. Schroeder, B. O. & Bäckhed, F. Signals from the gut microbiota to distant organs in physiology and disease. *Nature Medicine* **22**, 1079–1089 (2016).

12. Carabotti, M., Scirocco, A., Maselli, M. A. & Severi, C. The gut-brain axis: interactions between enteric microbiota, central and enteric nervous systems. *Ann Gastroenterol* **28**, 203–209 (2015).

13. Davis, M. M., Tato, C. M. & Furman, D. Systems immunology: just getting started. *Nature Immunology* **18**, 725–732 (2017).

14. Werner, H. M. J., Mills, G. B. & Ram, P. T. Cancer Systems Biology: a peak into the future of patient care? *Nat Rev Clin Oncol* **11**, 167–176 (2014).

Chapter III

Transfer of environmental microbes to the skin and respiratory tract of humans after urban green space exposure

Statement of Authorship

Title of Paper	Transfer of environmental microbes to the skin and respiratory tract of humans after urban green space exposure
Publication Status	[X] Published [ ] Accepted for Publication [ ] Submitted for Publication [ ] Unpublished and Unsubmitted work written in manuscript style
Publication Details	Published at Environmental International.

Principal Author

Name of Principal Author (Candidate)	Caitlin A. Selway
Contribution to the Paper	Performed laboratory experiments, data curation, bioinformatic analyses, interpretation of data, created all data visualization, and drafted the manuscript.
Overall percentage (%)	60%
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature	Date: 7.10.2020

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's statec contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Jacob G. Mills
Contribution to the Paper	Conceptualisation, sample collection, performed laboratory experiments, and edited and reviewed manuscript.
Signature	Date: 07/08/2020

Name of Co-Author	Philip Weinstein
Contribution to the Paper	Conceptualisation, funding acquisition, and edited and reviewed manuscript.
Signature	Date: 7/8/20

Name of Co-Author	Chris Skelly		
Contribution to the Paper	Conceptualisation, sample collection, and edited and reviewed manuscript.		
Signature		Date	4/7/2020

Name of Co-Author	Sudesh Yadav		
Contribution to the Paper	Conceptualisation, sample collection, and edited and reviewed manuscript.		
Signature		Date	3rd Aug. 2020

Name of Co-Author	Andrew Lowe		
Contribution to the Paper	Conceptualisation, funding acquisition, and edited and reviewed manuscript.		
Signature		Date	4th August 2020

Name of Co-Author	Martin F. Breed		
Contribution to the Paper	Conceptualisation, funding acquisition, sample collection, and edited and reviewed manuscript.		
Signature		Date	3/7/2020

Name of Co-Author	Laura S. Weyrich		
Contribution to the Paper	Conceptualisation, funding acquisition, interpretation of data, and edited and reviewed manuscript.		
Signature		Date	18/8/20

Environment International 145 (2020) 106084

Contents lists available at ScienceDirect

Environment International

journal homepage: www.elsevier.com/locate/envint

Transfer of environmental microbes to the skin and respiratory tract of humans after urban green space exposure

Caitlin A. Selway[a,*], Jacob G. Mills[b], Philip Weinstein[b], Chris Skelly[c], Sudesh Yadav[d], Andrew Lowe[b], Martin F. Breed[e], Laura S. Weyrich[a,f,g,*]

[a] Department of Molecular and Biomedical Science, University of Adelaide, Adelaide, SA 5005, Australia
[b] School of Biological Sciences and the Environment Institute, University of Adelaide, Adelaide, SA 5005, Australia
[c] Public Health Dorset, Dorset County Council, Dorchester DT1 1TP, UK
[d] School of Environmental Sciences, Jawaharlal Nehru University, New Delhi 110067, India
[e] College of Science and Engineering, Flinders University, Bedford Park, SA 5042, Australia
[f] Department of Anthropology, The Pennsylvania State University, University Park, PA 16802, USA
[g] Huck Institutes of Life Sciences, The Pennsylvania State University, University Park, PA 16802, USA

ARTICLE INFO

Handling Editor: Frederic Coulon

Keywords:
Microbiota
Biodiversity
Rewilding
Non-communicable diseases
Allergies
Urban green space

ABSTRACT

Background: In industrialized countries, non-communicable diseases have been increasing in prevalence since the middle of the 20th century. While the causal mechanisms remain poorly understood, increased population density, pollution, sedentary behavior, smoking, changes in diet, and limited outdoor exposure have all been proposed as significant contributors. Several hypotheses (*e.g.* Hygiene, Old Friends, and Biodiversity Hypotheses) also suggest that limited environmental microbial exposures may underpin part of this rise in non-communicable diseases. In response, the Microbiome Rewilding Hypothesis proposes that adequate environmental microbial exposures could be achieved by restoring urban green spaces and could potentially decrease the prevalence of non-communicable diseases. However, the microbial interactions between humans and their surrounding environment and the passaging of microbes between both entities remains poorly understood, especially within an urban context.

Results: Here, we survey human skin (n = 90 swabs) and nasal (n = 90 swabs) microbiota of three subjects that were exposed to air (n = 15), soil (n = 15), and leaves (n = 15) from different urban green space environments in three different cities across different continents (Adelaide, Australia; Bournemouth, United Kingdom; New Delhi, India). Using 16S ribosomal RNA metabarcoding, we examined baseline controls (pre-exposure) of both skin (n = 16) and nasal (n = 16) swabs and tracked microbiota transfer from the environment to the human body after exposure events. Microbial richness and phylogenetic diversity increased after urban green space exposure in skin and nasal samples collected in two of the three locations. The microbial composition of skin samples also became more similar to soil microbiota after exposure, while nasal samples became more similar to air samples. Nasal samples were more variable between sites and individuals than skin samples.

Conclusions: We show that exposure to urban green spaces can increase skin and nasal microbial diversity and alter human microbiota composition. Our study improves our understanding of human-environmental microbial interactions and suggests that increased exposure to diverse outdoor environments may increase the microbial diversity, which could lead to positive health outcomes for non-communicable diseases.

1. Introduction

Over the past 70 years, industrialized countries have experienced an increased prevalence of non-communicable diseases (NCDs), such as cancers, diabetes, asthma, allergies, and cardiovascular, mental and autoimmune diseases (Bach, 2002; Boutayeb and Boutayeb, 2005; Okada et al., 2010). This increase was initially observed in industrialized countries (Wagner and Brath, 2012); however, NCDs are

* Corresponding authors at: Department of Molecular and Biomedical Science, University of Adelaide, Adelaide, SA 5005, Australia (C.A. Selway). Department of Anthropology, Pennsylvania State University, University Park, PA 16802, USA (L.S. Weyrich).

E-mail addresses: caitlin.selway@adelaide.edu.au (C.A. Selway), jacob.mills@adelaide.edu.au (J.G. Mills), philip.weinstein@adelaide.edu.au (P. Weinstein), wchris.skelly@gmail.com (C. Skelly), syadav@mail.jnu.ac.in (S. Yadav), andrew.lowe@adelaide.edu.au (A. Lowe), martin.breed@flinders.edu.au (M.F. Breed), lsw132@psu.edu (L.S. Weyrich).

https://doi.org/10.1016/j.envint.2020.106084
Received 13 June 2020; Received in revised form 19 August 2020; Accepted 19 August 2020

also rapidly increasing in industrializing countries (Okada et al., 2010). Many of these diseases are linked to immune system dysfunction, where the body attacks itself (autoimmune disease) or overreacts to harmless substances (allergy; *e.g.* pollen or dust mites). For example, reduced exposure to these harmless substances during immune development leads to a higher chance of developing asthma (Lynch et al., 2014). Increased NCD rates have been linked to a wide range of factors, including genetics, environmental and lifestyle factors (diet, sedentary behavior and smoking), and more recently documented, microbial communities on and in the human body (microbiota) (Rook, 2012).

Industrialized societies experiencing this higher prevalence of NCDs have undergone an excelled rate of environmental and sociological change in the recent past (*e.g.* The Great Acceleration) (Graham and White, 2016; Steffen et al., 2015). Since the 1950s, urbanization has become one of the biggest drivers of industrialization (Steffen et al., 2015). Urbanization is associated with a decrease in outdoor exposure, fewer green space environments, and reduced environmental microbial diversity (Graham and White, 2016; McMichael, 2013). For instance, there was a 14–55% decline in green cover of single-family homes across 20 cities in Los Angeles County from 2000 to 2009 (Lee et al., 2017). This decrease of urban vegetation, as well as the increased access to medicine (*e.g.* antibiotics), are both hypothesized to reduce microbial diversity within humans (Rook, 2012; Graham and White, 2016).

The Hygiene (Strachan, 1989), Old Friends (Rook et al., 2003), and Biodiversity (von Hertzen et al., 2011) Hypotheses each suggest mechanisms that may explain why an urbanized lifestyle is associated with decreased human microbial diversity. The Hygiene Hypothesis suggests that limiting early childhood exposure to microbes, such as infectious agents, symbiotic microorganisms, and parasites, increases their susceptibility to various allergens (Strachan, 1989). The Old Friends Hypothesis also highlights the impacts of decreased microbial exposure during immune development but suggests that modern exposures contain less microbial diversity than those that occurred throughout human history (Rook et al., 2003). The Biodiversity Hypothesis further expands on the Hygiene and Old Friends Hypotheses and suggests that reduced exposure to microbiota from natural environments may be the cause (von Hertzen et al., 2011). Each of these hypotheses infers causal relationships between reduced microbial exposure and increased incidences of NCDs, despite few studies directly examining the causal links and mechanisms between decreased environmental microbiota exposure and human disease.

The Microbiome Rewilding Hypothesis (Mills et al., 2017) provides a potential solution to reduce NCD incidence. This hypothesis posits that exposure to remnant and revegetated ecosystems may restore the microbial diversity in human habitats and improve current industrialized microbial states. In turn, increased microbial diversity could help reduce NCD prevalence. Natural plant communities (*e.g.* remnant vegetation) tend to be more microbially variable than human-impacted environments (*e.g.* lawns) (Mhuireach et al., 2016, 2019), and therefore, could serve as a source of diverse microbiota to colonize the human body and rewild human microbiomes.

Several studies have suggested that microbiota transfer from the environment to humans. For example, there is more bacterial diversity on the skin of individuals living near agricultural and forest environments compared to urban environments (Hanski et al., 2012). Murine models have further supported this observation; farm dust containing microorganisms can protect the epithelium in the lungs to minimize allergies, perhaps through transfer of protective microorganisms (Schuijs et al., 2015). Further, a study has suggested that spending at least 120 min outdoors over a week does improve health and well-being (White et al., 2019) and may help decrease the NCD epidemic in industrialized countries. If the Microbiome Rewilding Hypothesis is true, exposure to microbially diverse outdoor environments could be a solution to decreasing NCD rates in urban centers. Indeed, it has already been shown in a randomized controlled mouse trial that high-biodiversity soils potentially supply the gut with anxiety-reducing microbes (*e.g.* butyrate-producing bacteria) (Liddicoat et al., 2020).

Despite these hypotheses and observations, specific studies tracking how the human microbiome is affected when visiting urban green spaces have not yet been conducted. In this paper, we monitor changes in human-associated microbial diversity and composition after exposure to urban green spaces on both the skin and nasal passage of volunteers in Adelaide, Australia; Bournemouth, United Kingdom; and New Delhi, India. In Adelaide, we characterized environmental samples from this urban environment to track the direct transfer of specific environmental microbiota to the human subjects. Overall, this study aims to better understand the relationship between humans and their surrounding microbial environment and provide evidence to better understand the Hygiene, Old Friends, and Biodiversity Hypotheses.

2. Methods

2.1. Sampling locations

Samples were collected in urban green spaces of Adelaide, Australia (environmental samples: September 2016; human samples: October-November 2016), Bournemouth, England (human samples: October 2016), and New Delhi, India (human samples: October 2016). Locations were logged with GPS (Table S1). Five different types of urban green space were sampled three times in each city (n = 15 locations per city) to represent different vegetation types and to capture the city's green space heterogeneity (Table S1). The vegetation types varied between cities due to biome differences but were largely analogous. Adelaide's vegetation types were vacant lots (covering in annual weeds), lawns (*e.g.* sports fields), parklands (*i.e.* annual understory with tree canopy), revegetated open woodlands, and remnant open woodlands. Bournemouth's vegetation types consisted of bare ground, lawns, parklands, young regrowth forests (revegetated), and old regrowth forests (remnant). New Delhi's vegetation types were bare ground, lawns, parklands, invaded shrublands (*i.e.* predominantly invasive shrub species), and savannah woodlands (*i.e.* native grassland with sparse tree cover).

2.2. Sample collection

2.2.1. Soils

Soil was collected from the 0–10 cm soil horizon in Adelaide. Within each green space, approximately 100 g of soil was collected from each of nine points using sterilized (flamed) trowels and homogenized in a sterile bag. From the homogenized soil, a 50 g sample was collected in a sterile falcon tube and frozen at −20 °C at the research facility on the evening of collection.

2.2.2. Air

In Adelaide, an Airchek Sampler 224-PCXR7 (SKC Inc.) was run for 1 h to collect air samples. The sampler ran at 2 $L.min^{-1}$ and was calibrated using a rotameter before and after sampling to determine average flow rate. Samples were collected on 1 μm pore-size, 25 mm diameter glass fiber filters (SKC Inc.). The filter cartridge was sterilized before each filter was inserted. After collection, filters were placed into 2 mL sterile microcentrifuge tubes with sterilized forceps. A field control was taken at each site by loading a filter into a sterilized filter cartridge without running the sampler. Air filters were frozen at −20 °C at the research facility until DNA extraction.

2.2.3. Leaves

Leaf samples were collected in Adelaide while wearing sterile gloves from vegetation strata (mixed species) in two sterile Falcon tubes per site and frozen −20 °C at the research facility until washing. Care was taken to collect healthy, undamaged leaves. Samples were standardized for leaf surface area, and surface microbes were washed into 15 mL of

TE buffer from which DNA was extracted.

2.2.4. Skin and nasal swabs

In each city, two of the three subjects directly interacted with elements of urban green space by collecting air, soil, and leaf samples. These interactions, *i.e.* digging in the dirt and brushing against vegetation, were an attempt to mimic behaviors of children interacting with the environment. Subject 1 visited urban green spaces in all three countries, whereas subject 2 visited urban green spaces in Australia, and subject 3 visited UK and India. For subjects that traveled internationally, samples were taken at least three days after landing.

Under ethics approval (H-2016–235; University of Adelaide Human Research Ethics Committee), participants swabbed themselves (skin and nasal) in the morning before leaving their accommodation and then after spending over 1 h in Adelaide green spaces (October-November 2016; n = 4 days) or after ~15 min in Bournemouth and New Delhi green spaces (October 2016; n = 2 days each). Given time constraints, multiple sites were visited on the same day (Table S1). Two drops of sterile saline solution were applied to the inside (volar) left wrist of each participant who then rubbed a swab (Sterile Catch-All™ Sample Collection Swab; Epicentre Biotechnologies) on the area for 30 sec. Participants also swirled a swab in their nasal cavity no more than 1 cm up the nose. Swab tips were snipped into sterile 2 mL microcentrifuge tubes. Swab field controls (sterile swab that followed the same process as sampling, without touching the skin or nasal cavity) to capture background signals from outdoor swab collection were also taken in each green space. At each location, the skin field control involved applying 2 drops of saline to a swab and then enclosing the tip in a 2 mL microcentrifuge tube, and the nasal control was a dry swab placed into separate tube.

2.3. DNA extraction

To allow for a direct comparison, skin and nasal samples were prepared identically to the environmental samples. Bacterial DNA was extracted using the DNeasy Powersoil kit (QIAGEN), following manufacturer's instructions. Extraction blank controls (*i.e.* extractions without any biological sample; EBCs) were performed in parallel to monitor DNA present within the laboratory and reagents.

2.4. 16S ribosomal RNA amplification and sequencing preparation

Using Illumina primers (515F: 5'-AATGATACGGCGACCACCGAGATCTACACTATGGTAATTGTGTGCCAGCMGCCGCGGTAA-3'; reverse-barcoded 806R: 5'-CAAGCAGAAGACGGCATACGAGATnnnnnnnnnnnnnAGTCAGTCAGCCGGACTACHVGGGTWTCTAAT-3') (Caporaso et al., 2011, 2012), the V4 region of the 16S rRNA gene was targeted and amplified. Each individual polymerase chain reaction (PCR; 25 μL) contained 18.05 μL DNA-free water, 2.5 μL 10× High Fidelity PCR Buffer (Invitrogen), 1.0 μL $MgSO_4$ (50 mM), 0.2 μL dNTPs (25 mM) (Invitrogen), 0.25 μL Platinum® Taq DNA Polymerase High Fidelity (Invitrogen), 1.0 μL of each V4 primer (10 μM), and 1.0 μL of DNA extract. Cycling conditions were as follows: 95 °C for 6 min; 38 cycles of 95 °C for 30 sec, 50 °C for 30 sec, and 72 °C for 90 sec; and a final extension at 60 °C for 10 min. A no-template amplification control (NTC) was included during each PCR group to monitor laboratory and reagent contamination. Additionally, all extracts and NTCs were amplified in triplicate to minimize PCR bias (Goodrich et al., 2014). Once amplified, PCR products were pooled, and the presence of a 16S V4 gene product was verified by gel electrophoresis. Amplifiable products from EBCs and NTCs were prepared for sequencing alongside the samples to characterize laboratory and reagent contamination.

16S rRNA libraries were quantified using a Qubit dsDNA HS assay kit (Invitrogen) and pooled in equimolar concentrations into groups of ~30 samples. Pooled libraries were then cleaned using AxyPrep Mag PCR Clean-up Kit (Axygen Scientific). The pooled, cleaned libraries were then quantified using a broad sensitivity D1000 Screentape on an Agilent 2200 Tapestation. Quantitative PCR (qPCR; KAPA Illumina Primer mix) was carried out to accurately determine the concentration of this final library (LightCycler 96 System, Roche Life Science), before diluting to a 2 nM concentration for sequencing. Libraries were sequenced on the Illumina Miseq using 2 × 150 bp kit with and custom sequencing primers, as previously described (Caporaso et al., 2011, 2012).

2.5. Pre-processing and high-quality sequence selection

Using QIIME2 (v 2019.7) (Bolyen et al., 2019), all sequencing runs were demultiplexed by the unique barcode fused to each individual sample. Using only the forward sequences, sequences were filtered based on the quality scores (minimum 4) and presence of ambiguous base calls (Bokulich et al., 2013). Following this, sequences were denoised using Deblur (Amir et al., 2017), with sequences trimmed to 150 bp. Amplicon sequence variant (ASV) tables and representative sequences from each sequencing run were then merged to allow for comparative analysis. Any ASV that was observed less than 10 times in the dataset and ASVs observed in one sample only were removed. The remaining representative ASVs were assigned taxonomy based on the SILVA database (v 132) (Quast et al., 2013). Next, a SEPP insertion tree was created for phylogenetic diversity analyses (Janssen et al., 2018). Decontam (Davis et al., 2018) was then used to identify contaminants found from EBCs and NTCs at a 0.5 prevalence, and these contaminants were removed. Finally, singletons and any sample with less than 500 sequences was also removed.

2.6. Diversity, statistical and taxonomic analyses

Next, alpha- and beta diversity were calculated at a rarefied depth of 1,000 sequences. For alpha diversity, both the species richness (observed species) and phylogenetic diversity (Faith's PD) (Faith, 1992) were calculated, and a Wilcoxon test (Wilcoxon, 1945) was applied to calculate the significant differences in sample groups. Unweighted UniFrac (Lozupone and Knight, 2005) was used to examine the phylogenetic beta diversity. PERMANOVA (Anderson, 2001) was implemented as a pairwise comparison to compare the UniFrac distances between sample groups. To analyze taxa at the genus level, ASVs were summarized into their taxonomic classification. Presence and absence of taxa between the post-exposure samples and pre-exposure and environmental samples were calculated. To further characterize taxa shared between sample groups, a core microbiome analysis (80% present in all samples) was performed both before and after the removal of pre-exposure taxa from post-exposure samples.

3. Results

3.1. Sequencing analysis and quality control

This study included human skin (n = 16 pre-exposure; n = 90 post-exposure) and nasal (n = 16 pre-exposure; n = 90 post-exposure) swabs, environmental samples (*i.e.* air filters, leaves, and soils; n = 45), environmental background controls (field controls; swab or air filter opened from packaging in the field but no sample was taken; n = 52), and laboratory controls (EBCs and NTCs; n = 51). As these samples range in biomass, we first examined the microbial signal in each sample type for artefacts. First, we removed singletons and low abundance taxa by excising any amplicon sequence variant (ASV) represented by less than 10 sequences and any ASV only observed in one sample over the dataset. Next, we detected 65 contaminant ASVs found in laboratory controls using Decontam (Davis et al., 2018) (Table S2) and removed these ASVs from the analysis. Once laboratory contaminant sequences were removed, EBC and NTC samples were then removed for downstream analysis. We then removed samples with less than 500

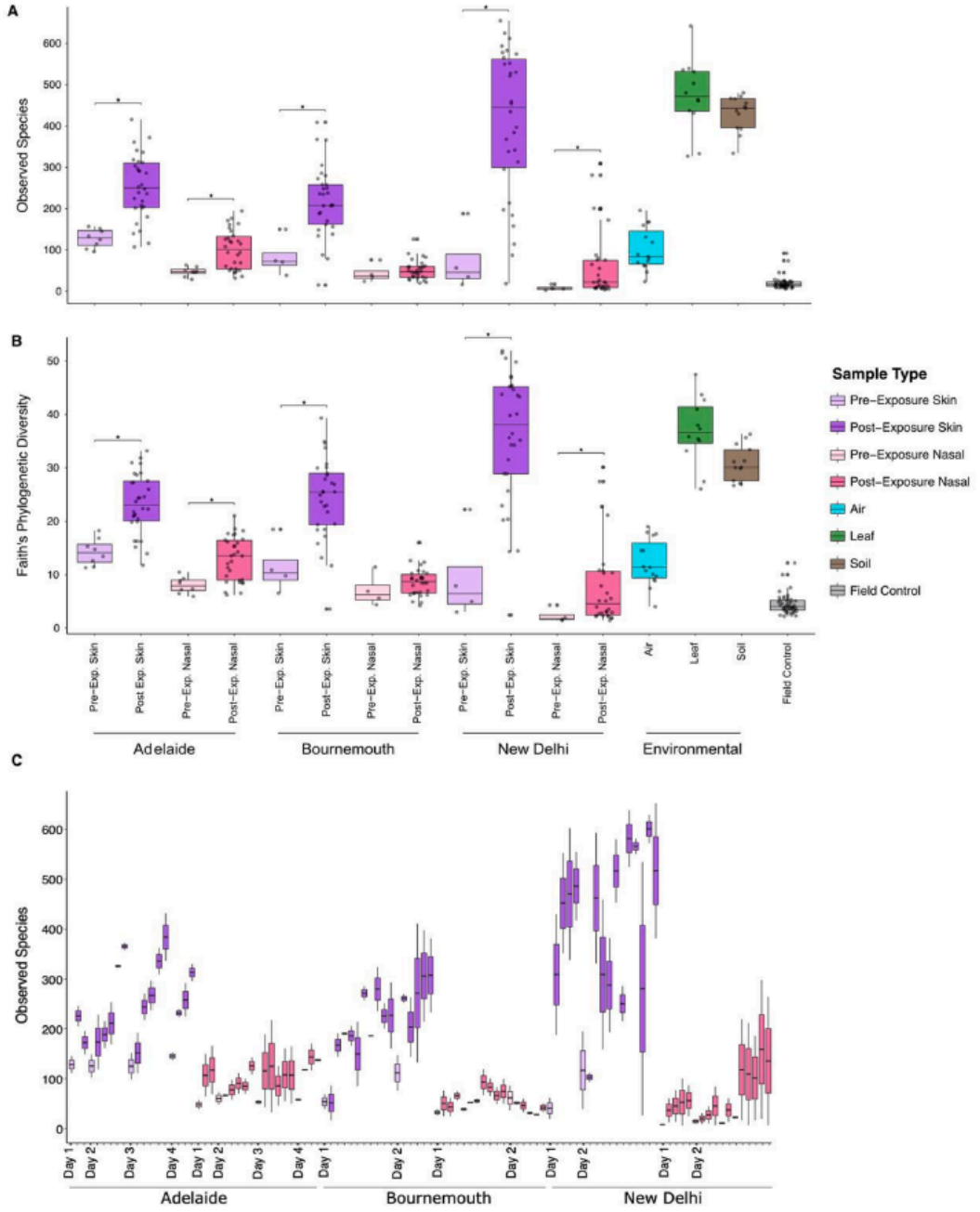

Fig. 1. Skin and nasal diversity increased after exposure to urban green space environments. Alpha diversity (*i.e.* within sample diversity) was calculated at a rarefied depth of 1,000 sequences for each sample, before binning into the respective sample type and location group. Two metrics were measured: A) Observed species and B) Faith's phylogenetic diversity. Statistical significance between pre- and post-exposure samples are denoted with * ($q < 0.05$). C) Alpha diversity was also calculated for each site visited throughout the day for both skin and nasal samples (displayed in order of sites visited). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

sequences (12 samples; Table S3). Post-filtering, 14,300,289 sequences (mean = 48,149 sequences; range = 529–791,282 sequences; see Table S4 for individual sample types) and 23,845 ASVs (see Table S4 for individual sample types) were retained from 297 samples.

3.2. Increased microbial diversity after green space exposure in Adelaide

First, we focused our analysis on a single city—Adelaide, Australia—to explore changes in skin and nasal microbiota after urban

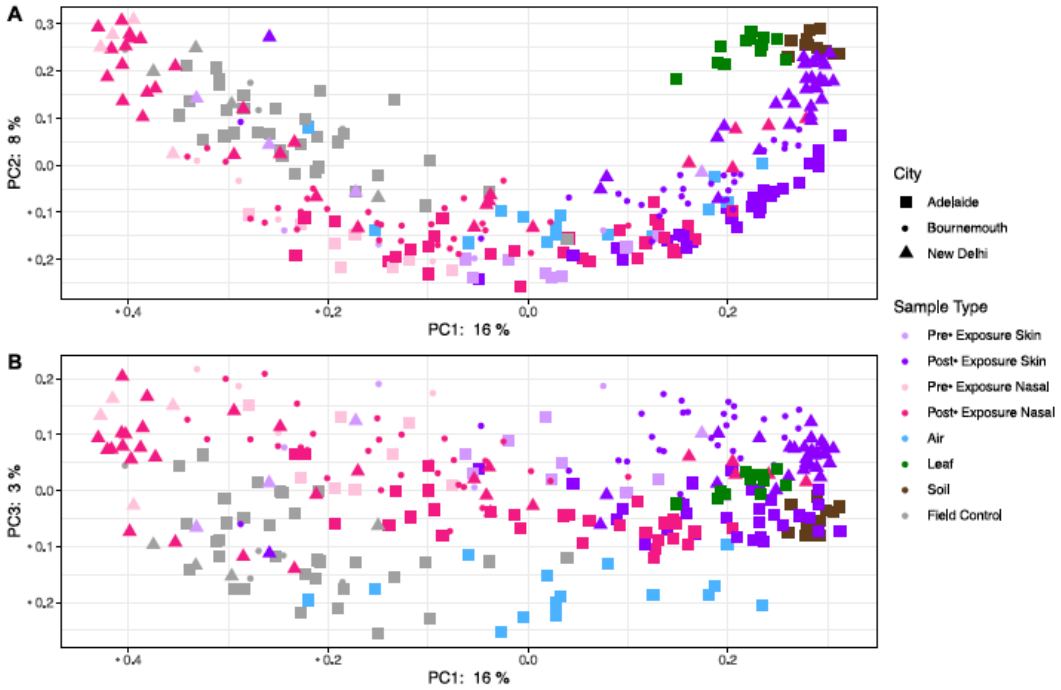

Fig. 2. Skin and nasal samples became more similar to environmental samples after urban green space exposure. PCoA plots were generated to compare compositional differences using the unweighted UniFrac metric. The first three principle component axes are shown as (A) PC1 vs. PC2 and (B) PC1 vs. PC3. Colors represent the sample type, and shapes represent the location. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

green space exposure. We assessed the diversity (microbial richness (observed species; OS) and phylogenetic (Faith's PD)) in skin and nasal samples after ~1.5 h of exposure to 15 different urban green space sites over four days. We observed a significant increase in skin microbiota diversity after urban green space exposure (Fig. 1; pre vs. post; Wilcoxon; OS: q = 0.0002; PD: q = 0.0001). Similarly, we observed an increase in nasal microbiota diversity after urban green space exposure (Fig. 1; OS: q = 0.0035; PD: q = 0.0022). In most cases, we noted an increase in diversity throughout the day, but after showering, the skin and nasal microbiota 'reset' to pre-exposure diversity levels as seen from the previous day (Fig. 1C). Overall, these results suggests that skin and nasal microbiota diversity increases after urban green space exposure.

Next, we evaluated the microbial composition changes (beta diversity) after exposure to urban green space environments in Adelaide. We found that the skin and nasal composition significantly shifted after urban Adelaide green space exposure (Fig. 2; pre vs. post; PERMANOVA; skin: pseudo-F = 3.351, q = 0.001; nasal: pseudo-F = 2.887, q = 0.001), and that microbiota in skin and nasal samples became more similar to the environmental samples after this exposure. More specifically, the skin samples showed closer approximation to soil samples after urban green space exposure (Fig. 2; pre vs. soil: pseudo-F = 10.62, q = 0.001; post vs. soil: pseudo-F = 10.77, q = 0.001). Further, the microbial composition of the nasal samples became more similar to the air samples (Fig. 2; PERMANOVA; pre vs. air: pseudo-F = 4.040, q = 0.001; post vs. air: pseudo-F = 2.687, q = 0.001) compared with the pre-exposure nasal samples. Overall, these results suggest that environmental microbes can colonize the human body during urban green space exposure over the timeframe observed.

We then investigated the origin or source of microbes in skin and nasal swabs after exposure to Adelaide urban green spaces. Before green space exposure, skin samples were consistently dominated (> 3% relative abundance) by *Micrococcus* (11.6% relative abundance), *Staphylococcus* (10.9%), *Terrasphera* (10.3%), *Corynebacterium* (7.5%), *Paracoccus* (6.9%), *Actinetobacter* (6.7%), *Brevundimonas* (3.8%), and *Cutibacterium* (3.5%) (Fig. 3), and these genera decreased in relative abundance post-exposure. Rare taxa (less than 3% relative abundance; combined relative abundance = 45.8%) and *Sphingomonas* (5.2%) increased in abundance post-exposure (Fig. 3). We observed similar trends in the nose. *Staphylococcus* (57.6%), *Corynebacterium* (20.8%), *Lawsonella* (6.6%), and *Anaerococcus* (3.9%) taxa were the dominant genera in the nose before urban green space exposure (Fig. 3). After urban green space exposure, *Staphylococcus* and *Lawsonella* decreased in abundance, while *Corynebacterium* and *Anaerococcus* remained similar (Fig. 3). Similar to the skin, rare ASVs increased by 7.5% post-exposure in the nose (Fig. 3). Altogether, this suggests that increases in diversity and shifts in microbial abundances are linked to an increase in rare taxa and a decrease in common, human-associated taxa after green space exposure at this city.

3.3. Introduced microorganisms originate from the air, plants, and soil

We next identified the relative proportions of ASVs shared between human and environmental samples after green space exposure in Adelaide. In the skin, 18% of ASVs in post-exposure samples were also found in pre-exposure skin samples, while more ASVs were shared with environmental samples: 23% with air, 30% with leaf, and 40% shared with soil samples (Fig. 4). In the nose, only 14% of post-exposure nasal ASVs were shared with pre-exposure nasal samples, while there were more ASVs shared with environmental samples: 28% shared with air, 30% with leaf, and 37% with soil samples (Fig. 4). This highlights the presence of environmental ASVs in both skin and nasal samples post-exposure. Following this, we completed a core microbiome analysis to further explore the taxa consistently found on the human body after

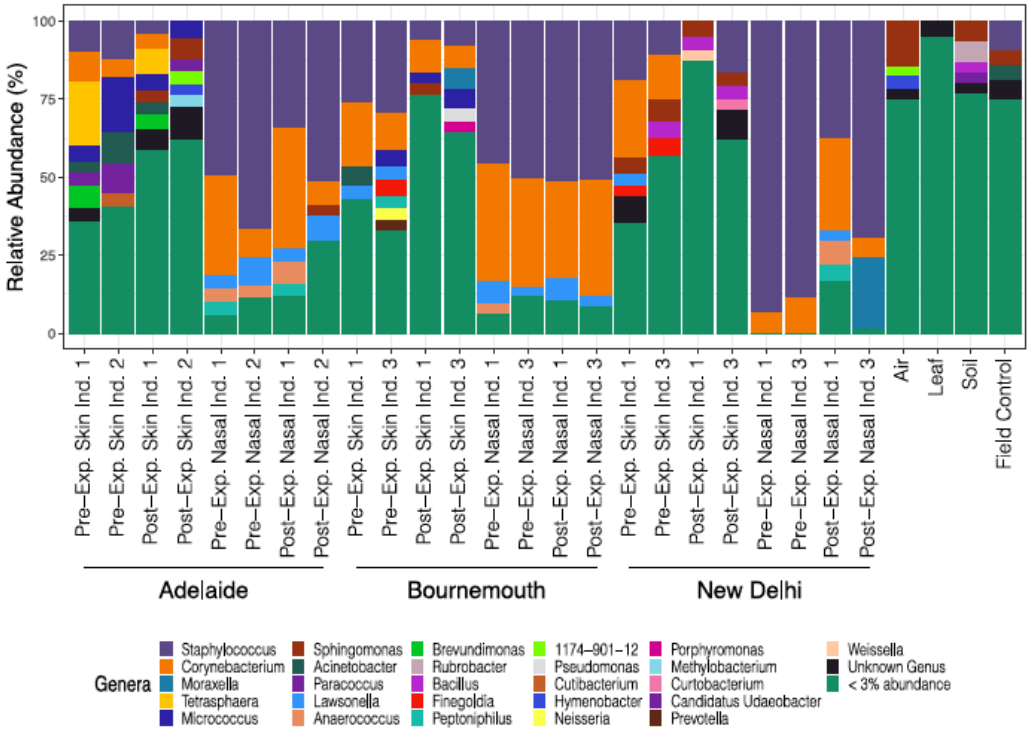

Fig. 3. Urban green space exposure is associated with an increase in low abundance taxa. The proportion of bacterial genera are shown in pre- and post-exposure skin and nasal samples, as well as environmental samples. Samples were grouped according to sample type, location, and subject. Genera with < 3% relative abundance were collapsed. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

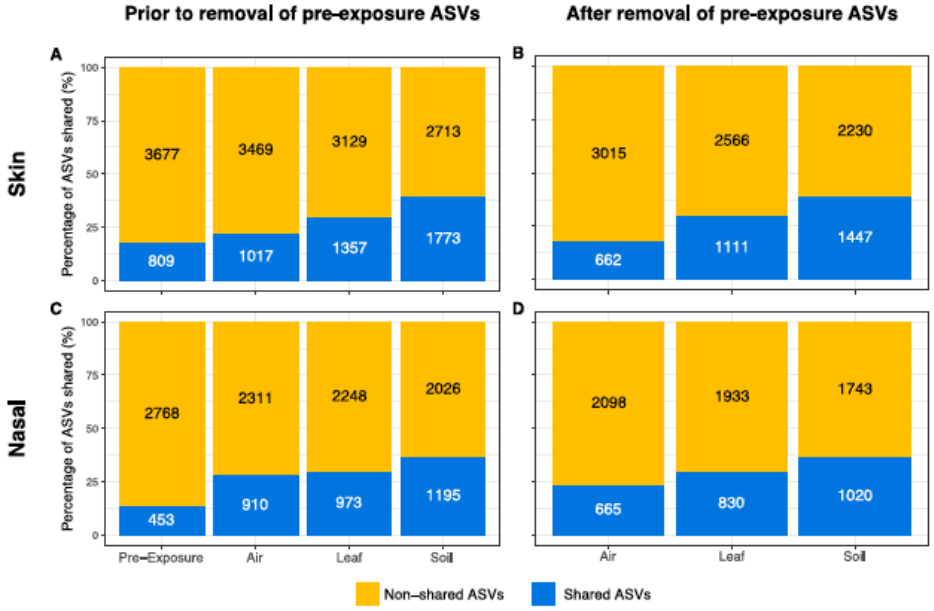

Fig. 4. Environmental ASVs are shared with skin and nasal samples after urban green space exposure. The presence or absence of ASVs from post-exposure samples and pre-exposure and environmental samples was calculated. This observation was calculated both before the removal of pre-exposure ASVs for skin (A) and nasal (C) samples, as well as after the removal of pre-exposure ASVs for skin (B) and nasal (D) samples. The percentage of non-shared ASVs were colored yellow, while the percentage of shared ASVs were blue. The number of shared or non-shared ASVs were noted in the bars for each sample type. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

urban green space exposure (*i.e.* present in 80% of samples). While 40 and 16 core ASVs were shared between the pre- and post-exposure skin and nasal samples, respectively, there were far fewer ASVs consistently shared between the environmental and post-exposure samples. For example, 13 air, 3 leaf, and 3 soil ASVs were shared with post-exposure skin samples, while 9 air, 1 leaf, and 3 soil ASVs were shared with post-exposure nasal samples (Table S5). When examining the core of the biological samples with all environmental samples, only a *Sphingomonas*, *Blastococcus*, and *Solirubrobacter* ASV were shared in 80% of all post-exposure skin and environmental samples, while a *Sphingomonas*, *Cutibacterium*, *Blastococcus*, and *Massilia* ASV were identified in 80% of all post-exposure nasal and environmental samples (Table S5). Overall, this suggests that microbes shared between the environment and the body are variable across environmental sites and the two individuals, but suggests that some environmental organisms can be identified in post-exposure skin and nasal swabs.

As the goal of this paper is to track newly introduced environmental species and not environmental species that may have been present on the skin and in the nose before the study, we excluded any ASV that was present in both pre- and post-exposure samples and then investigated the remaining post-exposure ASVs with environmental ASVs. In post-exposure skin samples, 18% of post-exposure skin ASVs were shared with air, 30% with leaf, and 39% with soil samples (Fig. 4). In post-exposure nasal samples, 24% of ASVs were shared with air, 30% with leaf, and 37% with soil samples (Fig. 4). The proportion of shared ASVs were very similar to before the removal of pre-exposure taxa, which indicates that very few pre-exposure microbes were shared with the environment, and this further confirms that these ASVs were transferred from the environment to the human body.

We next examined the core ASVs between post-exposure and environmental samples to investigate if the same ASVs were introduced onto the skin from the environment. No specific ASVs were present in the core microbiome between post-exposure and environmental samples, suggesting that the transfer of individuals species is highly variable across individuals and sites. While ASVs may be different, it is plausible that certain genera or types of species are preferentially transferred. Therefore, we performed a core microbiome analysis at the genus taxonomic level to provide a general description of the ASVs transferred from the environment. In the post-exposure skin samples, 34 genera were shared with at least one environmental type (Table 1). For example, *Sphingomonas*, *Nocardioides*, and *Solirubrobacter* genera, as well as unknown genera belonging to Sphingomonadaceae, Burkholderiaceae (family), and Solirubrobacterales (order), were shared in all of the environmental and post-exposure skin samples (Table 1). In the post-exposure nasal samples, 22 genera were shared with at least one environmental type (Table 1). *Sphingomonas*, *Nocardioides*, and Burkholderiaceae genera were observed in all three environmental sources and the post-exposure nasal samples (Table 1). As these genera were shared in multiple environmental types and humans, it may indicate their ubiquitous presence across the study sites.

3.4. Skin diversity increased with outdoor exposure in different locations

To investigate whether increased human-associated microbial diversity after urban green space exposure occurs in vastly different outdoor environments, skin and nasal samples from two individuals were collected during 15-minute, urban green space exposures in Bournemouth, England and New Delhi, India. In the skin samples, both microbial richness and phylogenetic diversity significantly increased after exposure to green space environments in Bournemouth (Fig. 1; pre vs. post; Wilcoxon; OS: $q = 0.0075$; PD: $q = 0.013$) and New Delhi (Fig. 1; OS: $q = 0.0016$; PD: $q = 0.0023$). In the nose, microbial richness and phylogenetic diversity significantly increased in nasal samples collected after urban green space exposure in New Delhi (Fig. 1; OS: $q = 0.034$; PD: $q = 0.031$); however, this was not the case in Bournemouth (Fig. 1; OS: $q = 0.27$; PD: $q = 0.15$). Similar to Adelaide, in most cases we saw an increase in microbial richness over the day for both skin and nasal samples in Bournemouth and New Delhi (Fig. 1C). Overall, these findings support an increase in microbial diversity of the skin, but not necessarily the nasal cavity, after green space exposure in different green space locations around the world.

Next, we examined the microbial composition of samples collected before and after green space exposure in Bournemouth and New Delhi. Exposure to urban green spaces resulted in a significant change in microbial composition in skin and nasal samples from Bournemouth (Fig. 2; pre vs. post; PERMANOVA; skin: pseudo-$F = 2.354$, $q = 0.002$; nasal: pseudo-$F = 1.722$, $q = 0.004$). In New Delhi, skin samples collected post-exposure were also significantly different in composition compared to pre-exposure skin samples (Fig. 2; pseudo-$F = 3.449$, $q = 0.001$). However, nasal samples showed no significant compositional changes after green space exposure (Fig. 2; pseudo-$F = 1.747$, $q = 0.063$). These results indicate that skin microbiota composition significantly changes after this type of green space exposure, but nasal microbiota compositional changes were not always observed during green space exposures, possibly due to the shorter exposure time frame.

Lastly, we examined changes in relative abundance of bacterial genera after green space exposure in Bournemouth and New Delhi (Fig. 3). On the skin, exposure to green spaces in both Bournemouth and New Delhi resulted in a decrease in abundance of *Staphylococcus* (20.7% decrease in Bournemouth; 5.4% decrease in New Delhi), *Corynebacterium* (7.7% decrease in Bournemouth; 17.9% decrease in New Delhi), *Finegoldia* (1.5% decrease in Bournemouth; 4.1% decrease in New Delhi), *Lawsonella* (2.8% decrease in Bournemouth), *Acinetobacter* (2.4% decrease in Bournemouth), and *Sphingomonas* (1.6% decrease in New Delhi) (Fig. 3). The genera that increased after green space exposure in Bournemouth were *Micrococcus* (1.8% increase) and *Moraxella* (3.4% increase), while only *Bacillus* (1.4% increase) increased in New Delhi (Fig. 3). Rare taxa also increased in both Bournemouth (26.1% increase) and New Delhi (30.6% increase) (Fig. 3). Similar to Australia, Bournemouth nasal samples were dominated by *Staphylococcus* (47.8% relative abundance), *Corynebacterium* (36.0%), and *Lawsonella* (5.2%) before green space exposure (Fig. 3). However, the relative abundance of these genera did not vary by more than 3% post-exposure, and these genera remained dominant post-exposure in Bournemouth. In contrast, the abundances in post-exposure nasal samples from New Delhi were more variable between individuals and were distinct from pre-exposure nasal samples. Again, *Staphylococcus* (90.7% relative abundance) and *Corynebacterium* (8.9%) were the most abundant taxa pre-exposure; however, *Staphylococcus* (53.1% relative abundance) decreased after green space exposure, and *Corynebacterium* (17.9%) and low abundance taxa (12.3%) increased post-exposure (Fig. 3). Interestingly, both individuals were affected by different taxa post-exposure in New Delhi. For example, *Moraxella* increased to 23.1% relative abundance after green space exposure in one subject, yet *Anaerococcus* (7.9%) and *Peptoniphilus* (5.1%) increased in abundance in the other subject (Fig. 3). While compositional and abundance-based differences were dependent on the city and individual, *Staphylococcus* consistently decreased, and low abundance taxa consistently increased in post-exposure skin and nasal samples in all three cities.

4. Discussion

There is a need to improve our understanding of the interactions between environmental and human microbiota, especially if this interaction is to be used to reduce the prevalence of NCDs in urban human populations. In this study, we monitored the bacterial diversity, composition, and relative abundance of microbiota in the skin and nasal passages from three individuals before and after exposure to several urban green spaces in three cities. The diversity and composition of skin microbiota consistently increased after exposure to urban green space environments. The response of nasal microbiota was unique to specific green spaces and individuals studied. Overall, we show that microbiota

Table 1
Core genera are shared between post-exposure samples and environmental samples after removing of pre-exposure ASVs.

ASV Genus	Skin			Nasal		
	Air	Leaf	Soil	Air	Leaf	Soil
Sphingomonas						
Hymenobacter						
Nocardioides						
Methylobacterium						
Solirubrobacter						
Unknown Genus (Family 67-14)						
Unknown Genus (Family Sphingomonadaceae)						
Unknown Genus (Family Burkholderiaceae)						
Unknown Genus (Family Chitinophagaceae)						
Chthoniobacter						
Unknown Genus (Family Acetobacteraceae)						
Conexibacter						
Candidatus Udaeobacter						
Bryobacter						
Unknown Genus (Family Solirubrobacteraceae)						
Novosphingobium						
Gemmatimonas						
Unknown Genus (Family Beijerinckiaceae)						
Mycobacterium						
Streptomyces						
Pedobacter						
Unknown Genus (Family Caulobacteraceae)						
Jatrophihabitans						
Unknown Genus (Family Isosphaeraceae)						
Rubrobacter						
Microvirga						
Flavisolibacter						
RB41 (Family Pyrinomonadaceae)						
Bacillus						
Segetibacter						
Marmoricola						
Actinoplanes						
Geodermatophilus						
Unknown Genus (Family Longimicrobiaceae)						
Devosia						
Pseudomonas						

A core microbiome analysis (present in at least 80% of samples) was performed for post-exposure samples and environmental samples at the genus taxonomic level. Shaded boxes denote shared genera between post-exposure skin (light grey) and nasal (dark grey) samples and environmental samples.

can transfer from the environment onto the human body, for at least the period observed here. Demonstrating such a colonization pathway lays a foundation to further examine how human microbiota shift during interactions with the natural environment.

4.1. Environmental microbiota transfer onto people

We identified increases in skin and nasal microbial diversity and composition after green space exposure in three different cities. Before exposure to green spaces, we observed that the skin and nasal samples were dominated by typical skin microbiota, including *Staphylococcus* and *Corynebacterium* (Grice et al., 2009; Oh et al., 2014). However, after exposure to urban green spaces, we saw an increase in environmental microbial taxa. For instance, *Sphingomonas*, *Blastococcus* (skin and nose), *Solirubrobacter* (skin), and *Massilia* (nasal)—all common soil microbes—were consistently observed on the individuals in this study after urban green space exposure. *Cutibacterium*, which was shared between the air and post-exposure nasal samples, is normally found in high abundance on the skin (Gannesen et al., 2019), and may possibly be cross-contamination from the individual who collected the air samples. After removing the ASVs found in pre-exposure samples, we also found *Nocardioides* (skin and nose), *Burkholderiaceae* (skin and nose), Solirubrobacterales (skin), and Sphingomonadaceae (skin) taxa transferred from the environment to the human body after green space exposure. Of these aforementioned taxa, both *Sphingomonas* and *Massilia* have also been shown to colonize the human body. For example, some species of *Sphingomonas* have been associated with nosocomial infections (Ryan and Adley, 2010) and others have been observed as a human skin commensal (Cosseau et al., 2016). While most known species of *Massilia* are also found in the air (Weon et al., 2008), freshwater (Gallego et al., 2006), and plant roots (Ofek et al., 2012); *Massilia* has also been detected in blood, cerebrospinal fluid, and bone of immunocompromised patients (Lindquist et al., 2003) and from the eye of a patient suffering with endophthalmitis (Kämpfer et al., 2012). While we found taxa transferred from the environment to both individuals, it is unknown how long these microbes remained present on the human body. Indeed, Bateman (Bateman, 2017) showed that most soil microbes transferred onto an individual were lost after 2 h, but rare taxa still remained present on the skin for longer than 24 h, including after washing (Bateman, 2017). Further research is also needed to explore how exposure to these and similar microbes from different environments (*e.g.* indoor environments) may impact skin and nasal microbiota or how much human use of certain green spaces may also influence microbial transfers. With longer tracking across more individuals and

explorations of different environment types, further research can investigate how these microbes do or do not colonize the human body and whether or not that colonization has health benefits or risks in healthy individuals.

Surprisingly, the change in diversity and composition before and after urban green space exposure was generally similar in the three different cities, despite numerous factors in timing (*e.g.* seasonality), diet, and exposures (*e.g.* temperature, humidity, pollution) differing between the three cities. Several studies have explored human microbiota changes when traveling (Youmans et al., 2015; Vangay et al., 2018; Voorhies et al., 2019), and these studies have implications to our results. For example, microbiota are impacted by circadian rhythms (Deaver et al., 2018), which were likely to be altered in the individuals studied here as they traveled internationally to these locations over a three-week period. We also only sampled our participants in the morning; future studies should examine the maintenance of these microbes throughout the day and collect control samples from these same individuals when they are not exposed to green spaces at different times of day. Further studies have also shown that diseases acquired during traveling, such as travelers' diarrhea, which affects up to 60% of individuals traveling to industrializing countries (Greenwood et al., 2008), can alter gut microflora composition, but not diversity, after the illness (Youmans et al., 2015). Altered diets and immunological responses to microbes in these locations may also have influenced our results. Another study showed that International Space Station astronauts had significant changes to skin and gut microbiota both during and after space flight, suggesting that the exposure to a single indoor environment can influence microbial composition and increase diversity (Voorhies et al., 2019). While space capsules are very different from urban environments, our participants were on long-haul flights during this study traveling from country to country. Lastly, Thai immigrants had altered gut microbial composition, decreased microbial diversity, and an increase in obesity rates after shifting their environment and moving to the United States (Vangay et al., 2018). Nevertheless, these studies support the idea that exposure to environments with different levels of diversity can alter the relative abundance of existing skin taxa, such as *Staphylococcus*, *Streptococcus*, and *Corynebacterium* (Voorhies et al., 2019). However, longitudinal investigations into microbial changes associated with different environmental contexts (*e.g.* indoor, non-green space, and different urban environments), as well as their concomitant functional changes, are needed to understand the role of environmental exposure in shaping the human microbiome, especially in the context of health. Further studies are also needed to examine if environmental exposure influences the microbiota in the mouth, urogenital tract, and other body sites.

4.2. Microbial diversity and health

Associations between decreased environmental or human diversity and disease have been noted in the skin and respiratory tract. For example, lower biodiversity on the skin is associated with atopic dermatitis (Bjerre et al., 2017; Kim and Kim, 2019), and exposure to farm dust—with a higher microbial load compared to house dust—has been shown to protect individuals against respiratory disease, such as asthma (Schuijs et al., 2015). These studies continue to support the idea that either having increased exposure to microbes or being colonized with more diverse microbes may have positive health benefits. In our study, we observed a significant increase in microbial diversity on the skin and in the nasal cavity of individuals that were exposed to diverse urban green spaces, indicating that exposure to these environments may be able to assist in providing a protective effect against immune-mediated diseases. Downstream studies looking at the health outcomes of individuals with regular outdoor exposure are needed to determine if these increases can be applied to successfully prevent or treat human disease.

While a substantial number of studies suggest that increased microbial biodiversity may generally be good for human health, the timing of this exposure to biodiverse microbial communities may also be important. For instance, increased biodiversity exposure is likely critical during immune development, as intense immune training and development occurs during the first three months following birth (Lloyd and Saglani, 2019; Knight et al., 2014) through to the age of three (Dzidic et al., 2018; Gensollen et al., 2016; Yatsunenko et al., 2012). During birth, increased infant microbial diversity has been associated with a vaginal birth and breast-feeding of the infant, compared to caesarean section births and formula feeding (Dominguez-Bello et al., 2019; Dzidic et al., 2018). However, there are many additional factors that contribute to the development of a child's microbiota, including the environment, diet, and medication (Munyaka et al., 2014). Understanding how these environmental factors alter the human microbiota during this critical window is especially important in the context of limiting downstream health issues. There is a growing body of literature suggesting that several diseases and disorders can be reduced by exposing an infant to particular 'allergens' early in immune development (von Hertzen et al., 2011; Burbank et al., 2017). Alternatively, there are negatively associated factors, such as antibiotics and reduced outdoor exposure, that may disturb a naturally developing microbiota and underpin disease development later in life (*e.g.* obesity) (Dominguez-Bello et al., 2019). While our study was conducted on adults, similar studies examining outdoor exposure in children should be conducted to see if an increase in their microbial biodiversity reduces downstream development of NCDs. Health consequences of outdoor exposure in school age children would also be a fruitful avenue to pursue in future research.

In most industrialized countries, individuals now spend an average of 146 h inside each week (~87%) (Spalt et al., 2016). Children also spend less time in outdoor activities than any other point in our history, especially with an increase in the availability of technology; for example, children spent ~30 min less in outdoor play between 1975 and 2015 (Mullan, 2019). In response to this, several studies have identified that limiting indoor exposure—by spending time outside—can improve health. For example, White et al. (2019) revealed that spending at least two hours outdoors each week was associated with self-reported good health and wellbeing (White et al., 2019). While numerous factors may contribute to this improvement in health, such as vitamin D production and exercise, the interaction with a microbially diverse environment could also make a long-term contribution to health. Recently, researchers have identified that some microorganisms in biodiverse soil (*e.g. Mycobacterium vaccae* and *Kineothrix alysoides*) can improve mood and limit anxiety and depression in mouse models (Reber et al., 2016; Liddicoat et al., 2020), suggesting that exposure to biodiverse outdoor microbiota may also contribute to improved human health. Our study indicates that a pathway for colonization of these environmental microbiota can occur through exposure to biodiverse environments (Mills et al., 2019; Aerts et al., 2018), which can happen in addition to eating from home or community-planted vegetable gardens (Tresch et al., 2019; Robinson et al., 2018; Mills et al., 2019), incorporating indoor plants into an office space or at home (Mahnert et al., 2015), planting roof-top gardens on apartment buildings (Robinson et al., 2018), or living with a dog that spends time outdoors.

5. Conclusions

Decreases in environmental microbial diversity and increases in NCDs have been observed over the last 70 years in both industrialized and industrializing countries. Exposure to biodiverse urban green spaces may improve health by increasing microbial diversity and shaping the composition of human microbiota after exposure. This study suggests that short-term (*e.g.* 15–30 min) exposure can introduce environmental microbes to the skin, and in some cases, to the nose. Further research is needed to examine if these microbiota alterations linked to green space exposure have positive health benefits, especially

in children, across different locations, and within other body sites.

Ethics approval and consent to participate

Ethics approval for this project was obtained from the University of Adelaide Human Research Ethics Committee (H-2016–235). Approvals included informed consent from participants and education to safely and reproducibly collect swabs.

Funding

This work was supported by the Australian Research Council (Future Fellowship FT180100407, awarded to LSW) and the University of Adelaide (DVCR Interdisciplinary Research Grant awarded to MFB, PW, AL, and LSW).

Author contributions

LSW, MFB, JGM, PW, CS, SY, and AL conceptualized and designed the study; JGM, MFB, CS, and SY collected samples; CAS and JGM performed experiments; CAS performed bioinformatics analysis; CAS and LSW provided interpretation of the data; CAS wrote the manuscript; all authors contributed to editing the manuscript.

CRediT authorship contribution statement

Caitlin A. Selway: Data curation, Investigation, Formal analysis, Visualization, Validation, Writing - original draft, Writing - review & editing. **Jacob G. Mills:** Conceptualization, Methodology, Data curation, Investigation, Writing - review & editing. **Philip Weinstein:** Conceptualization, Methodology, Writing - review & editing, Funding acquisition. **Chris Skelly:** Conceptualization, Methodology, Writing - review & editing. **Sudesh Yadav:** Conceptualization, Methodology, Writing - review & editing. **Andrew Lowe:** Conceptualization, Methodology, Writing - review & editing, Funding acquisition. **Martin F. Breed:** Conceptualization, Methodology, Resources, Writing - review & editing, Funding acquisition, Supervision. **Laura S. Weyrich:** Conceptualization, Methodology, Resources, Writing - review & editing, Funding acquisition, Supervision.

Data availability statement

The datasets analyzed in this article are available in the QIITA repository with Study ID 13064 (https://qiita.ucsd.edu/study/description/13064). Scripts for QIIME2 and R analysis are available on Figshare (https://adelaide.figshare.com/articles/Microbial_Transfer_To_Humans_Selway2020_txt/12016443).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to thank Soumya Prasad (School of Life Sciences) and Vijay Pal Yadav (deceased; School of Environmental Sciences) from Jawaharlal Nehru University, New Delhi, India, and Ken Daniel and colleagues from City of Playford, Adelaide, Australia for field and collection support. We also thank Jennifer Young from the University of Adelaide and Flinders University for providing laboratory support. Additionally, we thank Paul Gooding from the Australian Genome Research Facility for providing sequencing support. This work was supported by Australian Government Research Training Program Scholarships awarded to CAS and JGM.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envint.2020.106084.

References

Aerts, R., Honnay, O., Van Nieuwenhuyse, A., 2018. Biodiversity and human health: mechanisms and evidence of the positive health effects of diversity in nature and green spaces. Br. Med. Bull. 127, 5–22. https://doi.org/10.1093/bmb/ldy021.
Amir, A., McDonald, D., Navas-Molina, J.A., Kopylova, E., Morton, J.T., Xu, Z.Z., et al., 2017. Deblur rapidly resolves single-nucleotide community sequence patterns. MSystems 2, e00191–e216. https://doi.org/10.1128/mSystems.00191-16.
Anderson, M.J., 2001. A new method for non-parametric multivariate analysis of variance. Austral. Ecol. 26, 32–46. https://doi.org/10.1111/j.1442-9993.2001.01070.pp. x.
Bach, J.-F., 2002. The effect of infections on susceptibility to autoimmune and allergic diseases. N. Engl. J. Med. 347, 911–920. https://doi.org/10.1056/NEJMra020100.
Bateman, A., 2017. The dynamics of microbial transfer and persistence on human skin.
Bjerre, R.D., Bandier, J., Skov, L., Engstrand, L., Johansen, J.D., 2017. The role of the skin microbiome in atopic dermatitis: a systematic review. Br. J. Dermatol. 177, 1272–1278. https://doi.org/10.1111/bjd.15390.
Bokulich, N.A., Subramanian, S., Faith, J.J., Gevers, D., Gordon, J.I., Knight, R., et al., 2013. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nat. Methods 10, 57–59. https://doi.org/10.1038/nmeth.2276.
Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al-Ghalith, G.A., et al., 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat. Biotechnol. 37, 852–857. https://doi.org/10.1038/s41587-019-0209-9.
Boutayeb, A., Boutayeb, S., 2005. The burden of non communicable diseases in developing countries. Int. J. Equity Health 4, 2. https://doi.org/10.1186/1475-9276-4-2.
Burbank, A.J., Sood, A.K., Kesic, M.J., Peden, D.B., Hernandez, M.L., 2017. Environmental determinants of allergy and asthma in early life. J. Allergy Clin. Immunol. 140, 1–12. https://doi.org/10.1016/j.jaci.2017.05.010.
Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Lozupone, C.A., Turnbaugh, P.J., et al., 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc. Natl. Acad. Sci. 108, 4516–4522. https://doi.org/10.1073/pnas.1000080107.
Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., et al., 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. 6, 1621–1624. https://doi.org/10.1038/ismej.2012.8.
Cosseau, C., Romano-Bertrand, S., Duplan, H., Lucas, O., Ingrassia, I., Pigasse, C., et al., 2016. Proteobacteria from the human skin microbiota: species-level diversity and hypotheses. One Health 2, 33–41. https://doi.org/10.1016/j.onehlt.2016.02.002.
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, 226. https://doi.org/10.1186/s40168-018-0605-2.
Deaver, J.A., Eum, S.Y., Toborek, M., 2018. Circadian disruption changes gut microbiome taxa and functional gene composition. Front. Microbiol. 9. https://doi.org/10.3389/fmicb.2018.00737.
Dominguez-Bello, M.G., Godoy-Vitorino, F., Knight, R., Blaser, M.J., 2019. Role of the microbiome in human development. Gut 68, 1108–1114. https://doi.org/10.1136/gutjnl-2018-317503.
Dzidic, M., Boix-Amorós, A., Selma-Royo, M., Mira, A., Collado, M.C., 2018. Gut microbiota and mucosal immunity in the neonate. Med. Sci. 6, 56. https://doi.org/10.3390/medsci6030056.
Faith, D.P., 1992. Conservation evaluation and phylogenetic diversity. Biol. Conserv. 61, 1–10. https://doi.org/10.1016/0006-3207(92)91201-3.
Gallego, V., Sánchez-Porro, C., García, M.T., Ventosa, A., 2006. Massilia aurea sp. nov., isolated from drinking water. Int. J. Syst. Evol. Microbiol. 56, 2449–2453. https://doi.org/10.1099/ijs.0.64389-0.
Gannesen, A.V., Zdorovenko, E.L., Botchkova, E.A., Hardouin, J., Massier, S., Kopitsyn, D.S., et al., 2019. Composition of the biofilm matrix of cutibacterium acnes acneic strain RT5. Front. Microbiol. 10. https://doi.org/10.3389/fmicb.2019.01284.
Gensollen, T., Iyer, S.S., Kasper, D.L., Blumberg, R.S., 2016. How colonization by microbiota in early life shapes the immune system. Science 352, 539–544. https://doi.org/10.1126/science.aad9378.
Goodrich, J.K., Di Rienzi, S.C., Poole, A.C., Koren, O., Walters, W.A., Caporaso, J.G., et al., 2014. Conducting a microbiome study. Cell 158, 250–262. https://doi.org/10.1016/j.cell.2014.06.037.
Graham, H., White, P.C.L., 2016. Social determinants and lifestyles: integrating environmental and public health perspectives. Public Health 141, 270–278. https://doi.org/10.1016/j.puhe.2016.09.019.
Greenwood, Z., Black, J., Weld, L., O'Brien, D., Leder, K., Von Sonnenburg, F., et al., 2008. Gastrointestinal infection among international travelers globally. J. Travel Med. 15, 221–228. https://doi.org/10.1111/j.1708-8305.2008.00203.x.
Grice, E.A., Kong, H.H., Conlan, S., Deming, C.B., Davis, J., Young, A.C., et al., 2009. Topographical and temporal diversity of the human skin microbiome. Science 324, 1190–1192. https://doi.org/10.1126/science.1171700.
Hanski, I., von Hertzen, L., Fyhrquist, N., Koskinen, K., Torppa, K., Laatikainen, T., et al.,

2012. Environmental biodiversity, human microbiota, and allergy are interrelated. Proc. Natl. Acad. Sci. 109, 8334–8339. https://doi.org/10.1073/pnas.1205624109.
Janssen, S., McDonald, D., Gonzalez, A., Navas-Molina, J.A., Jiang, L., Xu, Z.Z., et al., 2018. Phylogenetic placement of exact amplicon sequences improves associations with clinical information. MSystems 3. https://doi.org/10.1128/mSystems.00021-18.
Kämpfer, P., Lodders, N., Martin, K., Falsen, E., 2012. Massilia oculi sp. nov., isolated from a human clinical specimen. Int. J. Syst. Evol. Microbiol. 62, 364–369. https://doi.org/10.1099/ijs.0.032441-0.
Kim, J.E., Kim, H.S., 2019. Microbiome of the skin and gut in atopic dermatitis (AD): understanding the pathophysiology and finding novel management strategies. J. Clin. Med. 8, 444. https://doi.org/10.3390/jcm8040444.
Knight, J.M., Davidson, L.A., Herman, D., Martin, C.R., Goldsby, J.S., Ivanov, I.V., et al., 2014. Non-invasive analysis of intestinal development in preterm and term infants using RNA-Sequencing. Sci. Rep. 4, 5453. https://doi.org/10.1038/srep05453.
Lee, S.J., Longcore, T., Rich, C., Wilson, J.P., 2017. Increased home size and hardscape decreases urban forest cover in Los Angeles County's single-family residential neighborhoods. Urban For Urban Green 24, 222–235. https://doi.org/10.1016/j.ufug.2017.03.004.
Liddicoat, C., Sydnor, H., Cando-Dumancela, C., Dresken, R., Liu, J., Gellie, N.J.C., et al., 2020. Naturally-diverse airborne environmental microbial exposures modulate the gut microbiome and may provide anxiolytic benefits in mice. Sci. Total Environ. 701, 134684. https://doi.org/10.1016/j.scitotenv.2019.134684.
Lindquist, D., Murrill, D., Burran, W.P., Winans, G., Janda, J.M., Probert, W., 2003. Characteristics of massilia timonae and massilia timonae-like isolates from human patients, with an emended description of the species. J. Clin. Microbiol. 41, 192–196. https://doi.org/10.1128/JCM.41.1.192-196.2003.
Lloyd, C.M., Saglani, S., 2019. Opening the window of immune opportunity: treating childhood asthma. Trends Immunol. 40, 786–798. https://doi.org/10.1016/j.it.2019.07.004.
Lozupone, C., Knight, R., 2005. UniFrac: a new phylogenetic method for comparing microbial communities. Appl. Environ. Microbiol. 71, 8228–8235. https://doi.org/10.1128/AEM.71.12.8228-8235.2005.
Lynch, S.V., Wood, R.A., Boushey, H., Bacharier, L.B., Bloomberg, G.R., Kattan, M., et al., 2014. Effects of early-life exposure to allergens and bacteria on recurrent wheeze and atopy in urban children. J. Allergy Clin. Immunol. 134 (593–601), e12. https://doi.org/10.1016/j.jaci.2014.04.018.
Mahnert, A., Moissl-Eichinger, C., Berg, G., 2015. Microbiome interplay: plants alter microbial abundance and diversity within the built environment. Front. Microbiol. 6. https://doi.org/10.3389/fmicb.2015.00887.
McMichael, A.J., 2013. Globalization, climate change, and human health. N. Engl. J. Med. 368, 1335–1343. https://doi.org/10.1056/NEJMra1109341.
Mhuireach, G.Á., Betancourt-Román, C.M., Green, J.L., Johnson, B.R., 2019. Spatiotemporal controls on the urban aerobiome. Front. Ecol. Evol. 7. https://doi.org/10.3389/fevo.2019.00043.
Mhuireach, G., Johnson, B.R., Altrichter, A.E., Ladau, J., Meadow, J.F., Pollard, K.S., et al., 2016. Urban greenness influences airborne bacterial community composition. Sci. Total Environ. 571, 680–687. https://doi.org/10.1016/j.scitotenv.2016.07.037.
Mills, J.G., Weinstein, P., Gellie, N.J.C., Weyrich, L.S., Lowe, A.J., Breed, M.F., 2017. Urban habitat restoration provides a human health benefit through microbiome rewilding: the Microbiome Rewilding Hypothesis. Restor. Ecol. 25, 866–872. https://doi.org/10.1111/rec.12610.
Mills, J.G., Brookes, J.D., Gellie, N.J.C., Liddicoat, C., Lowe, A.J., Sydnor, H.R., et al., 2019. Relating urban biodiversity to human health with the 'holobiont' concept. Front. Microbiol. 10. https://doi.org/10.3389/fmicb.2019.00550.
Mullan, K., 2019. A child's day: trends in time use in the UK from 1975 to 2015. Br. J. Sociol. 70, 997–1024. https://doi.org/10.1111/1468-4446.12369.
Munyaka, P.M., Khafipour, E., Ghia, J.-E., 2014. External influence of early childhood establishment of gut microbiota and subsequent health implications. Front. Pediatr. 2. https://doi.org/10.3389/fped.2014.00109.
Ofek, M., Hadar, Y., Minz, D., 2012. Ecology of root colonizing massilia (oxalobacteraceae). PLoS ONE 7, e40117. https://doi.org/10.1371/journal.pone.0040117.
Oh, J., Byrd, A.L., Deming, C., Conlan, S., NISC Comparative Sequencing Program, Barnabas, B., et al., 2014. Biogeography and individuality shape function in the human skin metagenome. Nature 514, 59–64. Doi: 10.1038/nature13786.
Okada, H., Kuhn, C., Feillet, H., Bach, J.-F., 2010. The 'hygiene hypothesis' for autoimmune and allergic diseases: an update. Clin. Exp. Immunol. 160, 1–9. https://doi.org/10.1111/j.1365-2249.2010.04139.x.
Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al., 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 41, D590–D596. https://doi.org/10.1093/nar/gks1219.
Reber, S.O., Siebler, P.H., Donner, N.C., Morton, J.T., Smith, D.G., Kopelman, J.M., et al., 2016. Immunization with a heat-killed preparation of the environmental bacterium Mycobacterium vaccae promotes stress resilience in mice. Proc. Natl. Acad. Sci. 113, E3130–E3139. https://doi.org/10.1073/pnas.1600324113.
Robinson, J.M., Mills, J.G., Breed, M.F., 2018. Walking ecosystems in microbiome-inspired green infrastructure: an ecological perspective on enhancing personal and planetary health. Challenges 9, 40. https://doi.org/10.3390/challe9020040.
Rook, G.A.W., 2012. Hygiene hypothesis and autoimmune diseases. Clin. Rev. Allergy Immunol. 42, 5–15. https://doi.org/10.1007/s12016-011-8285-8.
Rook, G.A.W., Martinelli, R., Brunet, L.R., 2003. Innate immune responses to mycobacteria and the downregulation of atopic responses. Curr. Opin. Allergy Clin. Immunol. 3, 337–342. https://doi.org/10.1097/01.all.0000092602.76804.ad.
Ryan, M.P., Adley, C.C., 2010. Sphingomonas paucimobilis: a persistent Gram-negative nosocomial infectious organism. J. Hosp. Infect. 75, 153–157. https://doi.org/10.1016/j.jhin.2010.03.007.
Schuijs, M.J., Willart, M.A., Vergote, K., Gras, D., Deswarte, K., Ege, M.J., et al., 2015. Farm dust and endotoxin protect against allergy through A20 induction in lung epithelial cells. Science 349, 1106–1110. https://doi.org/10.1126/science.aac6623.
Spalt, E.W., Curl, C.L., Allen, R.W., Cohen, M., Adar, S.D., Stukovsky, K.H., et al., 2016. Time-location patterns of a diverse population of older adults: the multi-ethnic study of atherosclerosis and air pollution (MESA Air). J. Expo. Sci. Environ. Epidemiol. 26, 349–355. https://doi.org/10.1038/jes.2015.29.
Steffen, W., Broadgate, W., Deutsch, L., Gaffney, O., Ludwig, C., 2015. The trajectory of the anthropocene: the great acceleration. Anthr. Rev. 2, 81–98. https://doi.org/10.1177/2053019614564785.
Strachan, D.P., 1989. Hay fever, hygiene, and household size. BMJ 299, 1259–1260. https://doi.org/10.1136/bmj.299.6710.1259.
Tresch, S., Frey, D., Bayon, R.-C.L., Mäder, P., Stehle, B., Fliessbach, A., et al., 2019. Direct and indirect effects of urban gardening on aboveground and belowground diversity influencing soil multifunctionality. Sci. Rep. 9, 1–13. https://doi.org/10.1038/s41598-019-46024-y.
Vangay, P., Johnson, A.J., Ward, T.L., Al-Ghalith, G.A., Shields-Cutler, R.R., Hillmann, B.M., et al., 2018. U.S. immigration westernizes the human gut microbiome. Cell 175 (962–972), e10. https://doi.org/10.1016/j.cell.2018.10.029.
von Hertzen, L., Hanski, I., Haahtela, T., 2011. Natural immunity. EMBO Rep. 12, 1089–1093. https://doi.org/10.1038/embor.2011.195.
Voorhies, A.A., Mark Ott, C., Mehta, S., Pierson, D.L., Crucian, B.E., Feiveson, A., et al., 2019. Study of the impact of long-duration space missions at the International Space Station on the astronaut microbiome. Sci. Rep. 9. https://doi.org/10.1038/s41598-019-46303-8.
Wagner, K.-H., Brath, H., 2012. A global view on the development of non communicable diseases. Prev. Med. 54, S38–S41. https://doi.org/10.1016/j.ypmed.2011.11.012.
Weon, H.-Y., Kim, B.-Y., Son, J.-A., Jang, H.B., Hong, S.K., Go, S.-J., et al., 2008. Massilia aerilata sp. nov., isolated from an air sample. Int. J. Syst. Evol. Microbiol. 58, 1422–1425. https://doi.org/10.1099/ijs.0.65419-0.
White, M.P., Alcock, I., Grellier, J., Wheeler, B.W., Hartig, T., Warber, S.L., et al., 2019. Spending at least 120 minutes a week in nature is associated with good health and wellbeing. Sci. Rep. 9, 1–11. https://doi.org/10.1038/s41598-019-44097-3.
Wilcoxon, F., 1945. Individual comparisons by ranking methods. Biom. Bull. 1, 80–83. https://doi.org/10.2307/3001968.
Yatsunenko, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M., et al., 2012. Human gut microbiome viewed across age and geography. Nature 486, 222–227. https://doi.org/10.1038/nature11053.
Youmans, B.P., Ajami, N.J., Jiang, Z.-D., Campbell, F., Wadsworth, W.D., Petrosino, J.F., et al., 2015. Characterization of the human gut microbiome during travelers' diarrhea. Gut Microbes 6, 110–119. https://doi.org/10.1080/19490976.2015.1019693.

Chapter IV

Initially disrupted preterm infant oral microbiota diversity is restored within three months

Statement of Authorship

Title of Paper	Initial disrupted preterm infant oral microbiota diversity is restored within three months
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ☒ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Unpublished work prepared for submission to JAMA pediatrics

Principal Author

Name of Principal Author (Candidate)	Caitlin A. Selway
Contribution to the Paper	Performed laboratory experiments, data curation, bioinformatic analyses, interpretation of data, created all data visualization, and drafted the manuscript.
Overall percentage (%)	60%
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature	Date 8/10/20

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Carmel T. Collins
Contribution to the Paper	Conceptualisation, data interpretation, and critically reviewed manuscript.
Signature	Date 24/08/2020

Name of Co-Author	Naomi Fink
Contribution to the Paper	Sample collection, data interpretation, and critically reviewed manuscript.
Signature	Date 26/08/2020

Name of Co-Author	Maria Makrides		
Contribution to the Paper	Conceptualisation, funding acquisition, data interpretation, and critically reviewed manuscript.		
Signature		Date	21/08/20

Name of Co-Author	Robert A. Gibson		
Contribution to the Paper	Conceptualisation, funding acquisition, data interpretation, and critically reviewed manuscript.		
Signature		Date	26/08/20

Name of Co-Author	Thomas R. Sullivan		
Contribution to the Paper	Data interpretation and critically reviewed manuscript.		
Signature		Date	28/08/2020

Name of Co-Author	Laura S. Weyrich		
Contribution to the Paper	Conceptualisation, funding acquisition, interpretation of data, and critically reviewed manuscript		
Signature		Date	7/10/20

Initially disrupted preterm infant oral microbiota diversity is restored within three months

Caitlin A. Selway[1], Carmel T Collins[2,3], Naomi Fink[2,4], Maria Makrides[2,3], Robert A Gibson[2,5], Thomas R Sullivan[2,6], NERO Steering Committee, and Laura S. Weyrich[1,7,8]

Affiliations:

[1] School of Biological Sciences, The University of Adelaide, South Australia, Australia

[2] SAHMRI Women and Kids, South Australian Health and Medical Research Institute, Adelaide, South Australia, Australia

[3] Adelaide Medical School, Discipline of Paediatrics, The University of Adelaide, Adelaide, South Australia, Australia

[4] Department of Pediatrics, McMaster University Hamilton, Ontario, Canada

[5] School of Agriculture, Food and Wine, The University of Adelaide, South Australia, Australia

[6] School of Public Health, The University of Adelaide, South Australia, Australia

[7] Department of Anthropology, Pennsylvania State University, University Park, PA 16802 USA

[8] Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, PA, 16802, USA

Corresponding authors:

[*] Caitlin A Selway, Department of Molecular and Biomedical Science, University of Adelaide, Adelaide SA 5005, Australia. caitlin.selway@adelaide.edu.au

[*] Laura S Weyrich, Department of Anthropology, Pennsylvania State University, University Park, PA 16802 USA. lsw132@psu.edu

Keywords: preterm infants, paediatrics, microbiota, microbiome, neonatal immunity

4.1. Key points

Question: How does the preterm oral microbiota develop through the neonatal window of opportunity, and is this different to full-term infants?

Findings: Using longitudinal data from 50 preterm infants, 14 full-term infants, and 16 adults, the oral microbiota of preterm infants within a week of birth is significantly distinct from all other preterm and full-term infant and adult oral microbiota samples. Potential opportunistic pathogens were observed in the mouth of preterm infants within a week of birth.

Meaning: Disruptions to oral microbiota during the neonatal window of opportunity may have profound implications for short- and long-term health in preterm infants.

4.2. Abstract

Importance: Preterm infants suffer higher morbidity and mortality rates compared to full-term infants, with infections serving as a leading cause of death. However, little is known about the origin and development of these infections over time, especially within oral and respiratory tract microbiota.

Objective: To investigate oral and respiratory microbiota development over the first two-to-three months in preterm infants who develop respiratory or systemic diseases, compared to healthy preterm and full-term infants and adults.

Design: A subset of preterm neonates from the N3RO cohort were included in an observational, longitudinal study, conducted between March 2014 to October 2015 in Australia, to study oral and respiratory microbiota development across the neonatal window of opportunity.

Setting: Preterm infants were administered to the Women's and Children's Hospital in Adelaide, Australia.

Participants: This study included 50 preterm (P) infants born <29 gestational weeks. Twenty-nine individuals were diagnosed with bronchopulmonary dysplasia (BPD) or sepsis (BPD, n=17; sepsis, n=9; BPD and sepsis, n=3). Fourteen preterm infants required intubation. Comparative data originated from 14 healthy full-term (F) infants and 16 adults.

Main outcomes and Measures: Oral and lung microbiota diversity, composition, and species abundances were calculated using amplicon sequence variations obtained from sequencing the 16S ribosomal RNA gene for two time points (T1: within a week post-birth; T2: approximately one-to-two months post-birth).

Results: PT1 sample diversity was more variable than all other groups, and PT1 composition was significantly different from PT2, full-term infant, and adult samples. However, PT2 samples were not significantly different to FT1 samples. BPD and sepsis influenced the preterm oral microbiota diversity and composition in infants at different ages. Opportunistic pathogens associated with BPD and sepsis were identified more frequently at T1 than T2 for both healthy preterm neonates and those with BPD and/or sepsis.

Conclusion and Relevance: After an initial delay, preterm oral microbiota appears to shift over time to resemble full-term infants. The presence of opportunistic pathogens in the oral cavity of infected neonates suggests that oral biomarkers should be developed for early warning signs of infection. The short- and long-term consequences of oral microbiota development need further examination in preterm infant infections and later development.

4.3. Introduction

Compared to full-term neonates, infants born prior to 37 weeks gestion (preterm) have a higher predisposition to chronic and infectious diseases, as many essential organs are underdeveloped at birth.[1] For example, underdeveloped lungs often require mechanical ventilation, which can damage the lungs and cause bronchopulmonary dysplasia (BPD).[1] Pathogens, such as *Ureaplasma,*[2] *Streptococcus anginosus,*[3] and *Streptococcus agalactiae* (Group B *Streptococcus*)[4], have been identified in the mouth and lungs of individuals with BPD,[5] and they can cause infections that lead to sepsis or pneumonia[1], although causes that underpin these diseases remain an active area of research.

Current research suggests that preterm health complications may be associated with the infant's postnatal environment.[6] Preterm infants live in a comparatively 'unnatural' environment within the neonatal intensive care unit (NICU), which is in stark contrast to most full-term infants. This altered environment, as well as delivery mode,[7,8] feeding method,[9,10] skin-to-skin contact (kangaroo care),[11] and antibiotic use,[12] can disturb or delay the commensal bacteria that colonize infants (microbiota) and are critical for health.[13] Preterm infant microbiota research focused on non-gut sites is limited, although several studies are now investigating oral[11,14–16] and skin[15,17] microbiota in these children. Even fewer studies have examined preterm infant microbiota development over time, especially outside of the gut[18]. In particular, oral microbiota development needs further attention, as oral microbiota can underpin both local and systemic diseases,[19,20] and may provide key insights into short- and long-term health impacts for preterm infants.

Early alterations to preterm infant microbiota could have downstream effects on development and health outcomes. Many alterations to preterm microbiota have been identified during the neonatal immunity window of opportunity—a non-redundant priming phase of the immune system.[21,22] As preterm infants are highly susceptible to infections, it is possible that critical immune system processes, such as innate immune tolerization and reprogramming, are also altered in these infants.[23] Indeed, Olin *et al.* (2018) demonstrated that preterm and full-term infant immune systems initially respond differently after birth, but profiles eventually converge after three months.[24] Existing studies have identified a mirrored delay in gut microbiota development in preterm infants, suggesting that a linked delay in microbiota and immune system development during this critical window may

contribute to both short- and long-term health consequences.[24,25] Despite this, we know very little about how similar microbiota alterations outside of the gut in preterm infants may also play roles in immediate and life-long health.

In this study, microbiota assessment was performed on a sub-set of infants born <29 weeks' gestation who were participating in the n-3 Fatty Acids for Improvement in Respiratory Outcomes (N3RO) randomised controlled trial.[26] Specifically, we characterised the development of oral and respiratory tract microbiota in healthy infants compared to those that develop disease (*i.e.* BPD and sepsis). We used oral swabs collected within one week of birth and at 36 weeks' postmenstrual age to compare oral microbiota diversity and compositional development over time with a published full-term infant and adult dataset[27]. Overall, this study aims to investigate oral microbiota development in preterm infants during the neonatal immune window of opportunity and identify unique signatures in the oral microbiota that may be linked to disease development.

4.4. Materials and Methods

4.4.1. Study population and sample collection

All microbiota samples were collected under the Human Research Ethics Committee approval obtained for both preterm infants (Women and Children's Hospital; HREC 2434/12/16; 27/11/2013) and adults (University of Adelaide; H2012-108). A subset of preterm infants primarily enrolled in a large clinical trial (N3RO)[26], which investigated the effects of docosahexaenoic acid (DHA), were selected for this study. For preterm (P) infants, patient metadata and population information was collected by medical record examination (Table 1). Preterm infant buccal swabs were collected at two time points—2-12 days post-birth (PT1) and 36 weeks' postmenstrual age (PT2)—from 50 preterm infants delivered at the Women's and Children's Hospital in Adelaide, Australia (Figure 1). Twenty-nine of these preterm infants developed BPD and/or sepsis (BPD, n=17; sepsis, n=9; BPD and sepsis, n=3). Tracheal aspirate samples were collected from fourteen preterm neonates who required intubation. All samples were frozen (-20°C) immediately in empty tubes after collection to preserve the bacterial composition. For a healthy, mature oral microbiota comparison, three adult samples were collected and processed following the same protocols as the preterm infant samples. To further

explore infant oral maturation, published oral microbiota data were obtained for 14 full-term (F) infants collected at two similar time points—0-7 days post-birth (FT1) and 4th-5th week post-birth (FT2)—and for 13 mothers (Study ID 2010, QIITA data repository; Figure 1).

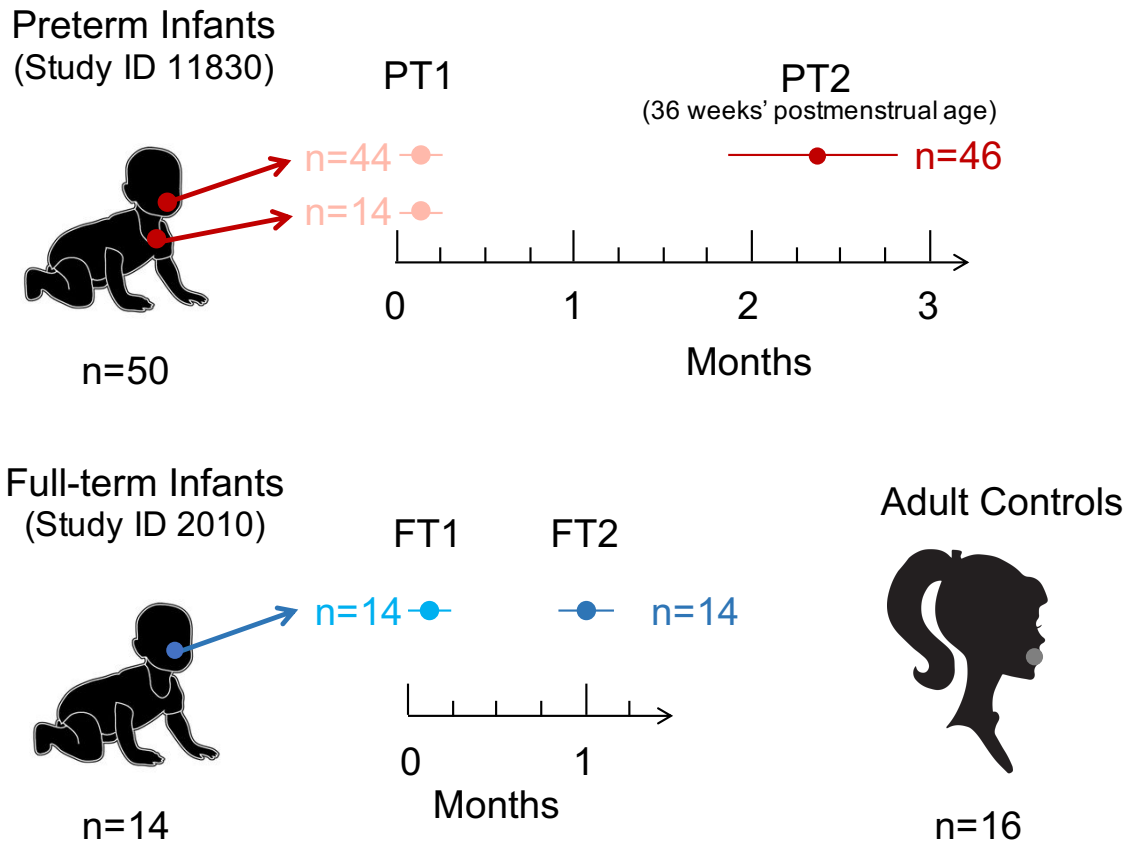

Figure 1: Sample were collected from preterm infants, full-term neonates, and adults over time. Oral samples were collected at analogous time points from preterm and full-term infants (approximately during the first week post-birth and one month later). Adult oral samples were to understand microbiota maturation of the infants. For preterm infants that were intubated, tracheal aspirates were also collected to identify shared taxa between the mouth and lungs of these infants.

Table 1: Study cohort characteristics

	Overall (n=50)	**With BPD* (n=20)**	**Without BPD* (n=28)**
Sex (male, %)	27 (54.0)	11 (55.0)	15 (53.6)
Gestational age at birth (weeks, mean, S.D.)	26.43 (±1.70)	25.86 (±1.78)	27.79 (±1.40)
Age at collection (days; mean, S.D.) T1 T2	 4.24 (±1.55) 65.74 (±14.59)	 4.26 (±1.33) 72.32 (±13.48)	 4.28 (±1.72) 60.50 (±13.63)
Birth type (n, %) Vaginal C.S. in labour C.S. no labour	 18 (36.0) 11 (22.0) 21 (42.0)	 8 (40.0) 4 (20.0) 8 (40.0)	 8 (28.6) 7 (25.0) 13 (46.4)
Enteral feeds on discharge (n, %) Breastmilk Formula Breastmilk & formula	 26 (52.0) 18 (36.0) 6 (12.0)	 10 (50.0) 7 (35.0) 3 (15.0)	 15 (53.6) 11 (39.3) 2 (7.1)
Maternal steroids (n, %) Steroids not given First dose <24hr before birth Given >7 days before birth Complete	 4 (8.0) 11 (22.0) 1 (2.0) 34 (68.0)	 0 (0.0) 3 (15.0) 0 (0.0) 1 17 (85.0)	 4 (14.3) 7 (25.0) 1 (3.6) 16 (57.1)
Postnatal steroids (n, %)	17 (34.0)	12 (60.0)	3 (10.7)
Sepsis (n, %)	13 (26.0)	7 (35.0)	6 (21.4)
Surgery (n, %)	15 (30.0)	9 (45.0)	6 (21.4)
Collected samples for both time points (n, %)	41 (82.0)	18 (90.0)	23 (82.1)
Individuals with tracheal aspirate collected (n, %)	14 (28.0)	9 (45.0)	5 (17.6)
	Overall (n=40)†	**With BPD (n=17)**	**Without BPD (n=23)**
Probiotics (n, %)	39 (97.5)	17 (100.0)	22 (95.7)
Antibiotics (n, %)	39 (97.5)	17 (100.0)	22 (95.7)
Antifungals (n, %)	39 (97.5)	17 (100.0)	22 (95.7)

BPD = bronchopulmonary dysplasia; C.S. = Caesarean-Section; S.D. = standard deviation. *Two subjects did not have information recorded for BPD. †Antibiotic information was only available for 40 subjects.

4.4.2. Sample DNA extraction, 16S Ribosomal RNA Library Preparation, and DNA Sequencing

All samples were prepared in a still-air room designed for low-biomass microbiota analysis using strict measures to reduce cross-contamination and the introduction of background DNA.[28] DNA was extracted from buccal swabs and tracheal aspirates using a previously published, in-house silica DNA extraction method designed to enhance DNA recovery;[29] the method was modified to include mechanical lysis (see **Appendix II** Supplementary Methods). DNA present within the laboratory and reagents was monitored using extraction blank controls (EBCs).

Bacterial DNA from each sample, including EBCs, was amplified in triplicate using primers that target the V4 region of the 16S ribosomal RNA (rRNA) gene[30], using previously described amplification conditions.[31] No-template controls (NTCs) were also included in each amplification batch. 16S rRNA libraries were prepared for sequencing (see **Appendix II** Supplementary Methods). All libraries were sequenced on an Illumina MiSeq (2x150 bp) at the Australian Genomics Research Facility.

4.4.3. Pre-processing, ASV selection, and contaminant removal

Preterm infant sequences (PT1, PT2, and tracheal aspirates) and adult (n=3) sequences obtained over two sequencing runs were uploaded to QIITA data repository (Study ID 11832; https://qiita.ucsd.edu/study/description/11832). Demultiplexed sequences were trimmed to 150 bp and amplicon sequence variants (ASVs) were generated via Deblur.[32] ASVs were merged with a full-term infant and maternal dataset (Study ID 2010; FT1, FT2 and adult) to create a SEPP insertion tree[33] and the study dataset. Contaminant sequences from EBCs and NTCs were identified via Decontam,[34] and were subsequently removed from all biological samples (**Appendix II** Table S1) in QIIME 2 (v2019.7)[35]. Following this, ASVs with less than 10 reads assigned were removed. In total, 156 samples were retained and represented 8,738,663 sequences and 1,553 ASVs.

4.4.4. Diversity analyses, taxonomic classification, and statistical comparisons

In QIIME2 (v2019.7)[35], ASVs were summarized into their taxonomic classification using the SILVA database (v132; 16S 515-806).[36] Using a rarefied depth of 2,000 sequences, (alpha) diversity was measured using observed species (OS) and Faith's phylogenetic diversity (PD)[37] metrics, and beta diversity (composition) was calculated using the unweighted UniFrac metric.[38] Significant associations between diversity and sample metadata were detected using pairwise Kruskal-Wallis tests[39] and the pairwise Fligner-Killeen test,[40,41] while significant links between composition and metadata were examined using adonis[42,43] and PERMANOVA.[42] Lastly, ANCOM[44] was used to identify significant changes in an ASV abundance across samples.

4.5. Results

4.5.1 Preterm oral microbiota at two months are similar to that in full-term infants

We first compared all preterm and full-term infant oral microbiota by investigating diversity at two analogous time points (PT1 and PT2 vs. FT1 and FT2). There was high variability and diversity in PT1 samples (Figure 2A&B), suggesting that each preterm infant responds uniquely to their environment. However, diversity significantly appeared to stabilize at 36 weeks' postmenstrual age, as a significant reduction in diversity variability was observed between PT1 and PT2 (PT1 vs. PT2; Fligner-Killeen; OS: chi-squared=33.149, p=8.537e-09; PD: chi-squared=29.913, p=4.519e-08; Figure 2A&B). Interestingly, PT2 and FT1 samples both contained similar, yet low, microbial diversity and were not significantly different from one another (Kruskal-Wallis; OS: H=0.041, q=0.840; PD: H=0.257, q=0.612). Adult oral microbiota diversity were significantly different from both preterm and full-term infant samples (Figure 2A&B; **Appendix II** Table S2; q>0.05), suggesting that further maturation of infant microbiota may continue over time, as previously reported.[45] Overall, this suggests that oral microbiota diversity in preterm infants decreases and stabilizes within two months to resemble full-term neonates.

Next, we analysed the composition of preterm and full-term infant oral samples using the unweighted UniFrac metric. PC1 separated preterm infants according to timepoint (PT1 vs. PT2; PERMANOVA; pseudo-F=8.982, q=0.001). We also observed a progression of microbial maturation across the first two principal coordinates (PC1 and PC2; Figure 2C&D), as PT1 samples moved towards the PT2 and FT1 samples (PT1 vs. FT1; pseudo-F=4.672, q=0.001; Figure 2C&D). Similarly, the PT2 and FT1 samples moved towards the FT2 samples (PT2 vs. FT2; pseudo-F=6.093, q=0.001; FT1 vs. FT2; pseudo-F=2.372, q=0.010; Figure 2C&D), suggesting that preterm oral microbiota may become more similar to full-term infants over time in some individuals. Nevertheless, preterm and full-term infant microbiota were significantly different from one another (Figure 2C&D; **Appendix II** Table S2; $q<0.05$). Lastly, the FT2 samples clustered near adult samples, although they remained discrete (pseudo-F=8.506, q=0.001). Overall, these results indicate that the neonatal oral microbiota may mature over time.

To better understand the compositional diversity differences, we investigated the microbial taxa abundance at the genus taxonomic level. While PT1 samples were highly variable between individuals (**Appendix II** Figure S1), a clear distinction was observed between PT1 and all other samples. Specifically, *Staphylococcus* (average relative abundance; 61.8%) dominated PT1 samples (Figure 3) and was significantly more abundant compared to all other groups (**Appendix II** Table S3). Conversely, PT2 samples were dominated (relative abundance >1%) by *Streptococcus* (71.6%), *Gemella* (6.1%), *Rothia* (5.7%), *Veillonella* (4.1%), *Staphylococcus* (3.6%), *Prevotella* (2.5%), and *Haemophilus* (1.2%), which were shared with FT1 samples (Figure 3), with the exception of *Prevotella*. The only taxa that significantly differed in abundance between PT2 and FT1 samples were *Rothia*, *Paenibacillus*, and *Erythrobacteraceae bacterium K-2-3* ASVs, all of which were more abundant in PT2 samples (**Appendix II** Table S3) and also observed in PT1 samples. Lastly, we identified several genera present in all infant and adult mouths, including *Streptococcus*, *Gemella*, *Rothia*, *Veillonella*, *Staphylococcus*, *Prevotella*, and *Haemophilus* ASVs, suggesting that these microbes are introduced from a young age and maintained into adulthood.

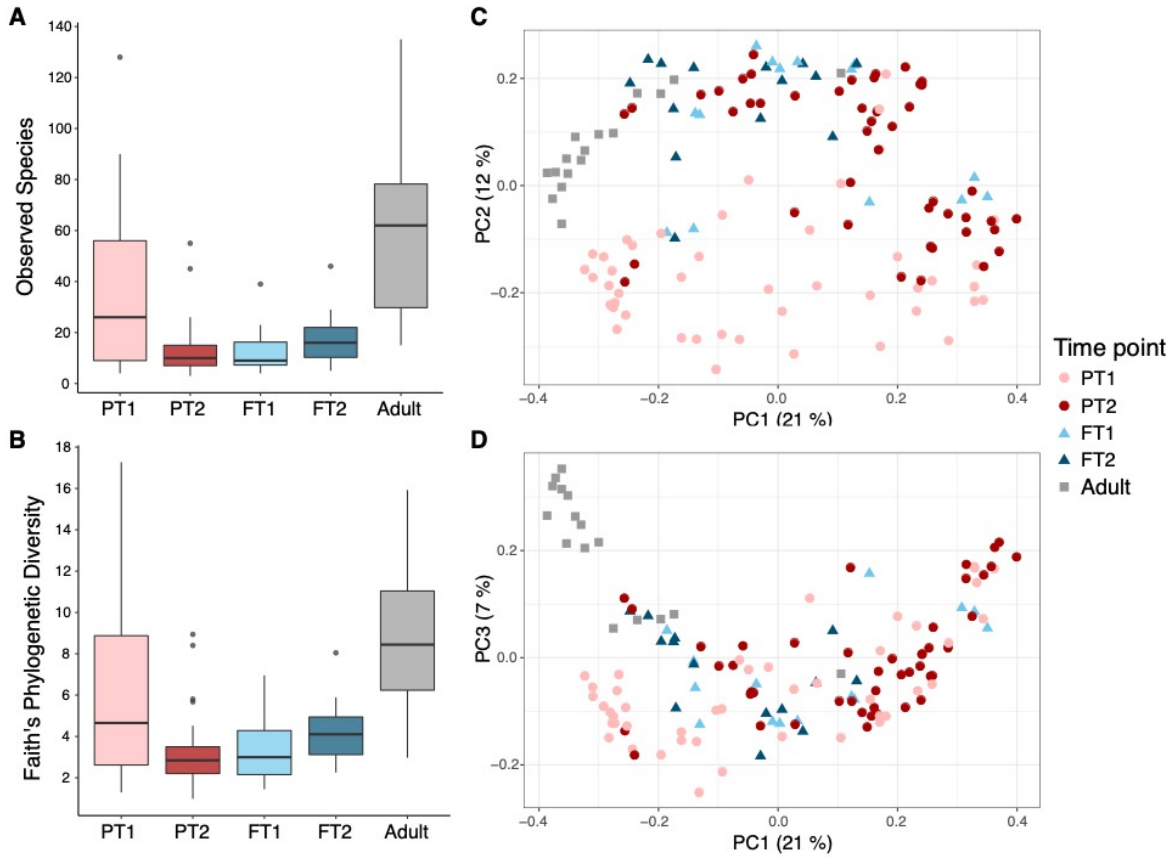

Figure 2: Within three months, preterm infant oral microbiota diversity and composition resembles full-term infants. Using 2,000 sequences per sample, alpha diversity variation decreases over time in preterm infants using both (A) observed species and (B) Faith's phylogenetic diversity metrics. Composition of preterm infants was initially distinct from PT2 and full-term infant samples, although this normalized over time using both unweighed UniFrac metric (C) over Axis 1 and 2. Adults were significantly different from infant samples over Axis 2 and 3 (D). Samples were coloured according to the time of collection and if they were a preterm or full-term infant, or adult.

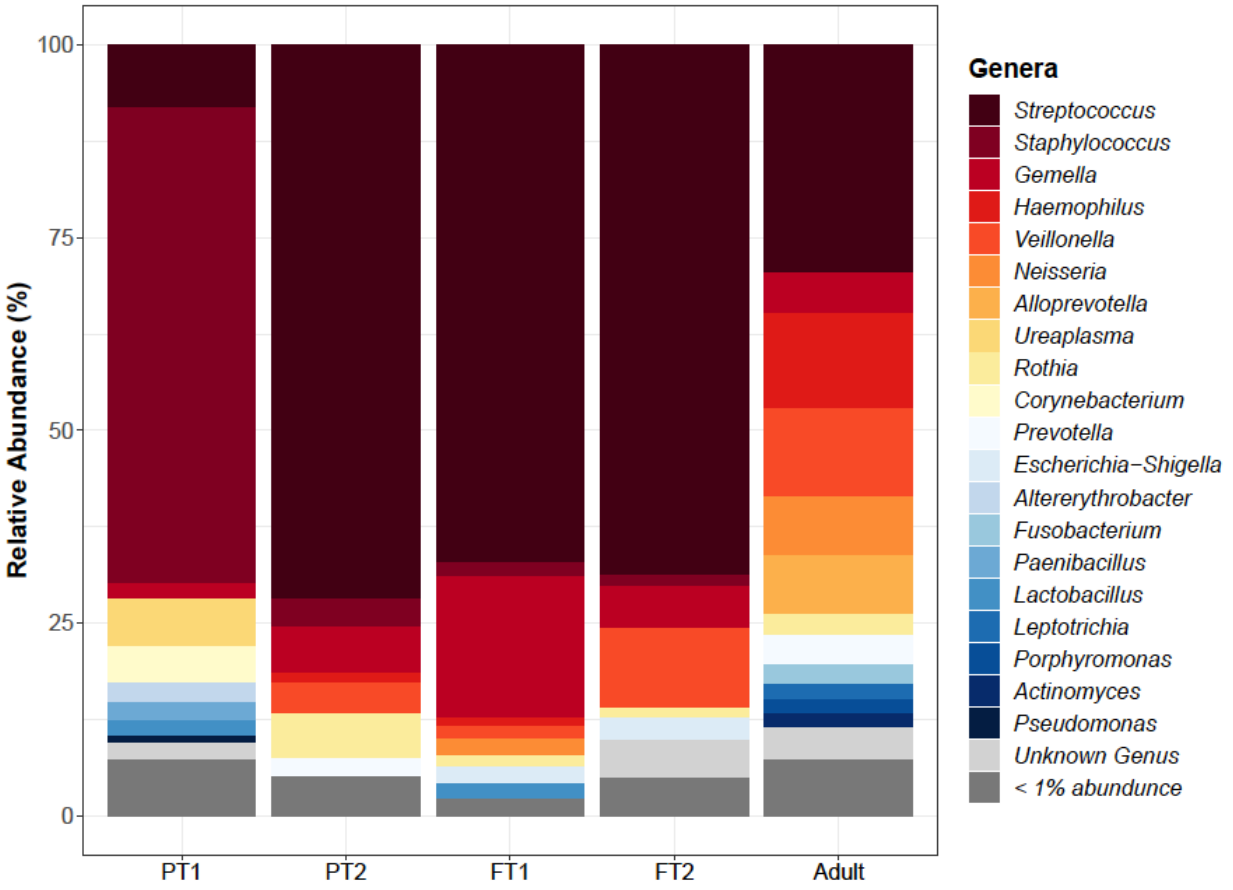

Figure 3: *Staphylococcus* dominates the PT1 samples, but taxonomic composition is restored over time for preterm infants. The relative abundance of oral bacterial genera are shown for preterm and full-term infants at two time points, as well as adult samples. Genera with <1% relative abundance were collapsed.

4.5.2. Disease influences the preterm oral microbiota

As BPD and sepsis commonly affect preterm infants, we wanted to determine if lethal BPD and sepsis are associated with the oral microbiota during the first few months post-birth. The presence of lethal BPD in preterm infants approached significance for oral microbiota diversity (OS: H=2.935, q=0.087; PD: H=3.593, q=0.058) but not composition (adonis; R^2=0.032, p=0.112) in infants at ~4 days old. At this age, infants that presented lethal BPD during this study had a lower microbial diversity at T1 (median: 13 ASVs) compared to infants who did not present BPD (median: 35 ASVs). Interestingly, *Corynebacterium* was found at 11.3% relative abundance in infants that developed BPD, but only at 0.9% relative

abundance in infants that did not develop BPD. Using ANCOM, no specific species were driving differences in composition. At ~65 days of age, however, infants that were diagnosed with sepsis had a lower microbiota diversity (median of 7 ASVs for children with sepsis vs median of 10 ASVs for children without sepsis; OS: H=3.949, q=0.04; PD: H=2.569, q=0.109) and composition was significantly influenced (adonis: R^2=0.046, p=0.031), but these differences were not detected at an earlier age. In all individuals at this age, *Streptococcus* dominated the mouth, which reduced the diversity observed in both infants with and without sepsis diagnosis. While no species or genera were significantly more abundant in infants that presented sepsis, we observed *Staphylococcus* at a higher relative abundance (13.6%) in infants with sepsis diagnosis, compared to infants that were not diagnosed with sepsis (0.6%). Sex of the infant at this age was also an important contributor to the oral microbiota (OS: H=5.189, q=0.02; PD: H=7.782, q=0.005; adonis: R^2=608, p=0.009). Interestingly, nine of thirteen infants diagnosed with sepsis were males. As these results are difficult to disentangle, more samples are needed to determine if sepsis is contributing to differences in the oral microbiota.

4.5.3. Oral microbiota are shared with lung microbiota

In preterm infants that were intubated, we identified ASVs shared between PT1 oral samples and tracheal aspirate samples in the same individual (n=12). Overall, 32 ASVs were shared between a PT1 oral sample and tracheal aspirate in at least one individual. Although lung samples were low-biomass and stochastic, the most dominant shared ASVs (>1,000 reads across individuals) included *Paenibacillus*, *Staphylococcus*, *Erythrobacteraceae bacterium K-2-3*, *Streptococcus anginosus*, *Ureaplasma*, Burkholderiaceae, and *Streptococcus agalactiae* sequences (**Appendix II** Table S4). *Ureaplasma* increases the risk for developing BPD,[3] and was also found as a dominant taxa in PT1 samples of children who did not receive intubation (FigureS1; >1,000 sequences; healthy=8, BPD=2). Additionally *Streptococcus anginosus* and *Streptococcus agalactiae* have previously been associated with BPD and sepsis, respectively. In this study, *Streptococcus anginosus* was found as a dominant species in one individual with BPD, while *Streptococcus agalactiae* was found as dominant species in one individual with sepsis. *Ureaplasma*, *Streptococcus anginosus*, and *Streptococcus agalactiae* all reduced in abundance by the T2 collection point. Together, this

suggests that early oral colonizers in preterm infants can also spread to the lungs and be linked to disease complications. However, these results also suggest that stabilization and maturation of the oral microbiota is linked to a reduction of opportunistic pathogens in these preterm children.

4.6. Discussion

By comparing oral and lung microbiota of preterm and full-term infants over the first few months of life, we observed marked levels of interindividual variation and compositional differences in newborn preterm infants. However, these observations were ameliorated in two-month-old preterm infants, as their microbes were more similar to full-term infants. While BPD is associated with early oral microbial diversity in preterm newborns, sepsis was significantly associated with diversity and compositional differences in preterm infants two months post-birth. Lastly, potential pathogens identified in both the mouth and lungs of preterm infants with BPD or sepsis also provide a unique window into microbial colonisation of the lungs linked to poor health outcomes. Overall, monitoring the oral microbiota through the neonatal immune window of opportunity may provide new insights into understanding higher morbidity and mortality rates in preterm infants compared to their full-term counterparts.

Higher morbidity and mortality rates in preterm infants typically arise from respiratory and infectious diseases, including BPD and sepsis. A single previous study surveyed the cross-over between the oral and lung microbiota of preterm infants in relation to BPD and found dominant genera that were also observed in our study.[5] Using exact sequences (ASVs), we identified these taxa at the species level and they were assigned as key pathogens associated with BPD and sepsis (*i.e. Ureaplasma,*[2] *Streptococcus anginosus,*[3] and *Streptococcus agalactiae* (Group B *Streptococcus*)[4]). In this study, these pathogens were highly abundant in oral samples at the first time point but they reduced in abundance by the second time point. Interestingly, we observed *Corynebacterium* in high abundance in the mouth of infants with BPD, which has also been observed in at high abundance in the lungs of infants with severe BPD.[46] We also detected *Staphylococcus* at a higher abundance in preterm infants with sepsis diagnosis. Indeed, *Staphylococcus aureus* has been shown to cause early onset sepsis, as detected by cultures of blood or cerebrospinal fluid.[47] Biomarkers for these opportunistic pathogens could be used

early in life or to explore children at risk of lethal diseases, due to the high proportion of these taxa found in the mouth. Further, microbiota diversity and composition were distinct in preterm infants with sepsis, compared to those without, but only one other study has examined the link between sepsis and the oral microbiota, although no significant differences in the mouths of infants diagnosed with sepsis were detected, likely due to a small sample size (total n=7).[18] Interestingly, more male infants in this study were affected by sepsis and overall we observed lower microbial diversity in male infants at ~65 days of age. In general, male infants more frequently suffer from diseases, such as early onset sepsis and BPD.[48,49] Overall, these results suggest that monitoring oral microbiota diversity in preterm infants may provide insights into respiratory health outcomes, although further work needs to account for high levels of microbial diversity variation in newly born preterm infants.

If preterm infants avoid or survive infectious or respiratory disease early in life, they still maintain a high predisposition to immune-mediated disease later in life, such as asthma.[50] Several studies postulate a disruption in the neonatal immunity window of immune training, which occurs during the first three years of life in layered phases (peri/postnatal, weaning, and post-weaning)[23], could be responsible for diseases later in life. The neonatal window notably includes a critical period for interactions between the immune system and gut microbiota during the first 100 days,[25] so examining how microbiota develop over this period throughout the entire body is vital to understanding long-term immune system development. Critically examining how the oral and respiratory tract microbiota develops over this time period is needed to understand the development of chronic respiratory disease development later in life. Here, we observed that the postnatal day was significantly associated with oral microbial composition in preterm infants; this could be impacted by the continual changes to the immune system observed during the first week of birth.[24] Further, we see a convergence of microbial diversity and composition over time that resembles full-term infants; this effect has also been reported in the immune system.[24] Despite this similar trend, we lack an understanding of the early-life, non-redundant immune processes altered during the peri/postnatal phase of immune development that cannot be modified later in the life. For example, Hornef *et al.* (2020) suggests that immune tolerization and reprogramming may be reduced due to prematurity.[23] In addition, we have even less of an understanding how oral and respiratory microbes may contribute to these

immune development processes, as most existing research is conducted on gut microbiota. Future studies should examine these intertwined processes at a much finer time scale, *i.e.* days/weeks during the peri/postnatal phase of immune system development, as well as how the outcomes of these processes are also related to non-immunological factors, such as the environment.

Our conclusions are limited by our population size, which included a sub-group of preterm infants from a single hospital in Australia, and a small number of full-term infants from a previously published study. The PT1 samples showed a significant amount of variation within the first week of birth. If samples were collected on the same day post-birth, *e.g.*, day 3, then this may reduce some of the variation associated this collection point. We also recognise that these results are limited to the mouth; collecting samples across the whole body would be required to understand how microbiomes of an individual work together. Additionally, pairing the whole-body approach with immune function, better environmental information (*i.e.* where the microbes originate from), social surveys, and more metadata would provide a holistic approach to the preterm infant microbiota development, could provide more information on the impacts on the neonatal window of opportunity and long-term health consequences.

4.7. Conclusions

Over the first few years of life, preterm infants are exposed to unnatural environments, which influences microbiota colonisation and immune system development. Our study demonstrated that the oral microbiota of preterm infants is disrupted up to the two months after birth and may start to resemble the oral microbiota of full-term infants over time. This delayed effect of the oral microbiota may have downstream implications on the health of preterm infants.

4.8 Declarations

Sources of funding and support

Supported by Australian Government Research Training Program Scholarship (CAS); Women's and Children's Hospital Foundation MS McLeod Research Fund Postdoctoral Fellowship (CTC); National Health and Medical

Research Council (NHMRC) Fellowships 1132596 (CTC), 1061704 and 1154912 (MM) and 1046207 (RAG). The N3RO trial was supported by a grant (1022112) from the NHMRC.

Authors Contributions

C.T.C., I.P., N.F., L.S.W, M.M, R.A.G Conceptualization; N.F. Sample collection; C.A.S. laboratory and bioinformatic analyses; C.A.S. drafted the initial manuscript; all authors interpreted the data and critically reviewed and revised manuscript

Acknowledgements

We would firstly like to thank all participants and their families for providing samples. In addition, we would like to thank Jennifer Young for lab troubleshooting, Pool Gooding at the Australian Genome Research Facility for providing sequencing, and Salva Herrando-Perez for statistical support.

Conflicts of interest

The authors declare that they have no conflicts of interest

4.9. References

1. Behrman RE, Butler AS, Outcomes I of M (US) C on UPB and AH. *Mortality and Acute Complications in Preterm Infants*. National Academies Press (US); 2007. Accessed April 22, 2020. https://www.ncbi.nlm.nih.gov/books/NBK11385/
2. Pammi M, Lal CV, Wagner BD, et al. Airway Microbiome and Development of Bronchopulmonary Dysplasia in Preterm Infants: A Systematic Review. *J Pediatr*. 2019;204:126-133.e2. doi:10.1016/j.jpeds.2018.08.042
3. Glaser K, Gradzka-Luczewska A, Szymankiewicz-Breborowicz M, et al. Perinatal Ureaplasma Exposure Is Associated With Increased Risk of Late Onset Sepsis and Imbalanced Inflammation in Preterm Infants and May Add to Lung Injury. *Front Cell Infect Microbiol*. 2019;9. doi:10.3389/fcimb.2019.00068

4. Doran KS, Nizet V. Molecular pathogenesis of neonatal group B streptococcal infection: no longer in its infancy. *Molecular Microbiology*. 2004;54(1):23-31. doi:10.1111/j.1365-2958.2004.04266.x

5. Brewer MR, Maffei D, Cerise J, et al. Determinants of the lung microbiome in intubated premature infants at risk for bronchopulmonary dysplasia. *The Journal of Maternal-Fetal & Neonatal Medicine*. 2019;0(0):1-7. doi:10.1080/14767058.2019.1681961

6. Santos J, Pearce SE, Stroustrup A. Impact of Hospital-Based Environmental Exposures on Neurodevelopmental Outcomes of Preterm Infants. *Curr Opin Pediatr*. 2015;27(2):254-260. doi:10.1097/MOP.0000000000000190

7. Dominguez-Bello MG, Costello EK, Contreras M, et al. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *PNAS*. 2010;107(26):11971-11975. doi:10.1073/pnas.1002601107

8. Chu DM, Ma J, Prince AL, Antony KM, Seferovic MD, Aagaard KM. Maturation of the infant microbiome community structure and function across multiple body sites and in relation to mode of delivery. *Nature Medicine*. 2017;23(3):314-326. doi:10.1038/nm.4272

9. Lif Holgerson P, Vestman NR, Claesson R, et al. Oral microbial profile discriminates breast-fed from formula-fed infants. *J Pediatr Gastroenterol Nutr*. 2013;56(2):127-136. doi:10.1097/MPG.0b013e31826f2bc6

10. Ho NT, Li F, Lee-Sarwar KA, et al. Meta-analysis of effects of exclusive breastfeeding on infant gut microbiota across populations. *Nat Commun*. 2018;9. doi:10.1038/s41467-018-06473-x

11. Hendricks-Muñoz KD, Xu J, Parikh HI, et al. Skin-to-Skin Care and the Development of the Preterm Infant Oral Microbiome. *Am J Perinatol*. 2015;32(13):1205-1216. doi:10.1055/s-0035-1552941

12. Gasparrini AJ, Wang B, Sun X, et al. Persistent metagenomic signatures of early-life hospitalization and antibiotic treatment in the infant gut microbiota and resistome. *Nature Microbiology*. 2019;4(12):2285-2297. doi:10.1038/s41564-019-0550-2

13. Tanaka M, Nakayama J. Development of the gut microbiota in infancy and its impact on health in later life. *Allergology International*. 2017;66(4):515-522. doi:10.1016/j.alit.2017.07.010

14. Sohn K, Kalanetra KM, Mills DA, Underwood MA. Buccal administration of human colostrum: impact on the oral microbiota of premature infants. *Journal of Perinatology*. 2016;36(2):106-111. doi:10.1038/jp.2015.157

15. Younge NE, Araújo-Pérez F, Brandon D, Seed PC. Early-life skin microbiota in hospitalized preterm and full-term infants. *Microbiome*. 2018;6(1):98. doi:10.1186/s40168-018-0486-4

16. Li H, Zhang Y, Xiao B, Xiao S, Wu J, Huang W. Impacts of delivery mode on very low birth weight infants' oral microbiome. *Pediatrics & Neonatology*. 2020;61(2):201-209. doi:10.1016/j.pedneo.2019.10.004

17. Pammi M, O'Brien JL, Ajami NJ, Wong MC, Versalovic J, Petrosino JF. Development of the cutaneous microbiome in the preterm infant: A prospective longitudinal study. *PLoS ONE*. 2017;12(4):e0176669. doi:10.1371/journal.pone.0176669

18. Young GR, van der Gast CJ, Smith DL, Berrington JE, Embleton ND, Lanyon C. Acquisition and Development of the Extremely Preterm Infant Microbiota Across Multiple Anatomical Sites. *Journal of Pediatric Gastroenterology and Nutrition*. 2020;70(1):12–19. doi:10.1097/MPG.0000000000002549

19. Lu M, Xuan S, Wang Z. Oral microbiota: A new view of body health. *Food Science and Human Wellness*. 2019;8(1):8-15. doi:10.1016/j.fshw.2018.12.001

20. Graves DT, Corrêa JD, Silva TA. The Oral Microbiota Is Modified by Systemic Diseases. *J Dent Res*. 2019;98(2):148-156. doi:10.1177/0022034518805739

21. Torow N, Hornef MW. The Neonatal Window of Opportunity: Setting the Stage for Life-Long Host-Microbial Interaction and Immune Homeostasis. *The Journal of Immunology*. 2017;198(2):557-563. doi:10.4049/jimmunol.1601253

22. Renz H, Adkins BD, Bartfeld S, et al. The neonatal window of opportunity—early priming for life. *J Allergy Clin Immunol*. 2018;141(4):1212-1214. doi:10.1016/j.jaci.2017.11.019

23. Hornef MW, Torow N. 'Layered immunity' and the 'neonatal window of opportunity' – timed succession of non-redundant phases to establish mucosal host–microbial homeostasis after birth. *Immunology*. 2020;159(1):15-25. doi:10.1111/imm.13149

24. Olin A, Henckel E, Chen Y, et al. Stereotypic Immune System Development in Newborn Children. *Cell*. 2018;174(5):1277-1292.e14. doi:10.1016/j.cell.2018.06.045

25. Arrieta M-C, Stiemsma LT, Dimitriu PA, et al. Early infancy microbial and metabolic alterations affect risk of childhood asthma. *Science Translational Medicine*. 2015;7(307):307ra152-307ra152. doi:10.1126/scitranslmed.aab2271

26. Collins CT, Makrides M, McPhee AJ, et al. Docosahexaenoic Acid and Bronchopulmonary Dysplasia in Preterm Infants. *New England Journal of Medicine*. 2017;376(13):1245-1255. doi:10.1056/NEJMoa1611942

27. Dominguez-Bello MG, De Jesus-Laboy KM, Shen N, et al. Partial restoration of the microbiota of cesarean-born infants via vaginal microbial transfer. *Nat Med*. 2016;22(3):250-253. http://dx.doi.org/10.1038/nm.4039

28. Weyrich LS, Farrer AG, Eisenhofer R, et al. Laboratory contamination over time during low-biomass sample analysis. *Molecular Ecology Resources*. 2019;19(4):982-996. doi:10.1111/1755-0998.13011

29. Rohland N, Hofreiter M. Comparison and optimization of ancient DNA extraction. *BioTechniques*. 2007;42(3):343-352. doi:10.2144/000112383

30. Caporaso JG, Lauber CL, Walters WA, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J*. 2012;6(8):1621-1624. doi:10.1038/ismej.2012.8

31. Selway CA, Mills JG, Weinstein P, et al. Transfer of environmental microbes to the skin and respiratory tract of humans after urban green space exposure. *Environment International*. 2020;145:106084. doi:10.1016/j.envint.2020.106084

32. Amir A, McDonald D, Navas-Molina JA, et al. Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. *mSystems*. 2017;2(2):e00191-16. doi:10.1128/mSystems.00191-16

33. Janssen S, McDonald D, Gonzalez A, et al. Phylogenetic Placement of Exact Amplicon Sequences Improves Associations with Clinical Information. *mSystems*. 2018;3(3). doi:10.1128/mSystems.00021-18

34. Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. Simple statistical identification and removal of contaminant sequences in marker-gene

and metagenomics data. *Microbiome*. 2018;6(1):226. doi:10.1186/s40168-018-0605-2

35. Bolyen E, Rideout JR, Dillon MR, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol*. 2019;37(8):852-857. doi:10.1038/s41587-019-0209-9

36. Quast C, Pruesse E, Yilmaz P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res*. 2013;41(D1):D590-D596. doi:10.1093/nar/gks1219

37. Faith DP. Conservation evaluation and phylogenetic diversity. *Biological Conservation*. 1992;61(1):1-10. doi:10.1016/0006-3207(92)91201-3

38. Lozupone C, Knight R. UniFrac: a New Phylogenetic Method for Comparing Microbial Communities. *Appl Environ Microbiol*. 2005;71(12):8228-8235. doi:10.1128/AEM.71.12.8228-8235.2005

39. Kruskal WH, Wallis WA. Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association*. 1952;47(260):583-621. doi:10.2307/2280779

40. Fligner MA, Killeen TJ. Distribution-Free Two-Sample Tests for Scale. *Journal of the American Statistical Association*. 1976;71(353):210-213. doi:10.1080/01621459.1976.10481517

41. Conover WJ, Johnson ME, Johnson MM. A Comparative Study of Tests for Homogeneity of Variances, with Applications to the Outer Continental Shelf Bidding Data. *Technometrics*. 1981;23(4):351-361. doi:10.2307/1268225

42. Anderson MJ. A new method for non-parametric multivariate analysis of variance. *Austral Ecology*. 2001;26(1):32-46. doi:10.1111/j.1442-9993.2001.01070.pp.x

43. Oksanen J, Blanchet FG, Friendly M, et al. *Vegan: Community Ecology Package*.; 2018. https://CRAN.R-project.org/package=vegan

44. Mandal S, Treuren WV, White RA, Eggesbø M, Knight R, Peddada SD. Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microbial Ecology in Health and Disease*. 2015;26(1):27663. doi:10.3402/mehd.v26.27663

45. Lif Holgerson P, Esberg A, Sjödin A, West CE, Johansson I. A longitudinal study of the development of the saliva microbiome in infants 2 days to 5 years compared to the microbiome in adolescents. *Scientific Reports*. 2020;10(1):9629. doi:10.1038/s41598-020-66658-7

46. Imamura T, Sato M, Go H, et al. The Microbiome of the Lower Respiratory Tract in Premature Infants with and without Severe Bronchopulmonary Dysplasia. *Am J Perinatol*. 2017;34(1):80-87. doi:10.1055/s-0036-1584301

47. Isaacs D, Fraser S, Hogg G, Li HY. Staphylococcus aureus infections in Australasian neonatal nurseries. *Archives of Disease in Childhood - Fetal and Neonatal Edition*. 2004;89(4):F331-F335. doi:10.1136/adc.2002.009480

48. Neubauer V, Griesmaier E, Ralser E, Kiechl-Kohlendorfer U. The effect of sex on outcome of preterm infants - a population-based survey. *Acta Paediatr*. 2012;101(9):906-911. doi:10.1111/j.1651-2227.2012.02709.x

49. Trembath A, Laughon M. Predictors of Bronchopulmonary Dysplasia. *Clin Perinatol*. 2012;39(3):585-601. doi:10.1016/j.clp.2012.06.014

50. Goedicke-Fritz S, Härtel C, Krasteva-Christ G, Kopp MV, Meyer S, Zemlin M. Preterm Birth Affects the Risk of Developing Immune-Mediated Diseases. *Front Immunol*. 2017;8. doi:10.3389/fimmu.2017.01266

Chapter V

Altered oral microbiota in type 1 diabetic children with hyperlipidemic parents

Statement of Authorship

Title of Paper	Altered oral microbiota in type I diabetic children with hyperlipidemic parents
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ☒ Unpublished and Unsubmitted work written in manuscript style
Publication Details	In preparation for Journal of Dental Research

Principal Author

Name of Principal Author (Candidate)	Caitlin A Selway		
Contribution to the Paper	Conceptualised and designed the research study, analysed and interpreted data, and wrote the manuscript.		
Overall percentage (%)			
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	26·8·2020

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Emilija D. Jensen		
Contribution to the Paper	Conceptualised and designed the research study and critically revised the manuscript.		
Signature		Date	04/08/2020

Name of Co-Author	Gabrielle Allen		
Contribution to the Paper	Conceptualised and designed the research study and critically revised the manuscript.		
Signature		Date	19/08/20

Name of Co-Author	Laura S. Weyrich		
Contribution to the Paper	Conceptualised and designed the research study, interpreted the data, and critically revised the manuscript.		
Signature		Date	18/8/20

Altered oral microbiota in type 1 diabetic children with hyperlipidemic parents

Caitlin A. Selway[1*], Emilija D. Jensen[2,3], Gabrielle Allen[2,3], Laura S. Weyrich[1,4,5*]

Affiliations:

[1] School of Biological Sciences, University of Adelaide, South Australia, Australia

[2] Discipline of Paediatric Dentistry, Adelaide Dental School, University of Adelaide, South Australia, Australia

[3] Department of Paediatric Dentistry, Women's and Children's Hospital, South Australia, Australia

[4] Department of Anthropology, The Pennsylvania State University, University Park, PA USA

[5] Huck Institutes of Life Sciences, The Pennsylvania State University, University Park, PA USA

Corresponding authors:

[*] Caitlin A Selway, School of Biological Sciences, University of Adelaide, Adelaide SA 5005, Australia. caitlin.selway@adelaide.edu.au

[*] Laura S Weyrich, Department of Anthropology, Pennsylvania State University, University Park, PA 16802 USA. lsw132@psu.edu

5.1. Abstract

Hyperlipidemia may play a significant role in the interrelationship between type 1 diabetes and periodontal disease. However, the effect that hyperlipidemia has on the oral microbiota of children with type 1 diabetes has not yet been explored. We examined the bacterial composition of gingival swab samples from 72 South Australian children with type 1 diabetes in relation to periodontal risk factors and hyperlipidemia status of first-degree relatives. In periodontally healthy (no periodontal pockets with depth >3 mm) children with type 1 diabetes, bacterial diversity (richness and phylogenetic diversity) was significantly reduced in children with a family history of hyperlipidemia compared to those without. Further, bacterial compositional diversity was significantly altered in these children with a family history of hyperlipidemia. In contrast, no significant differences in diversity

or compositional diversity between those with or without a family history of hyperlipidemia were observed in children with high-risk periodontal markers (at least one periodontal pocket with depth >3mm) and type 1 diabetes. These results were supported by co-occurrence networks, which showed fewer networks in periodontally healthy children with a family history of hyperlipidemia, and similar networks across all children with high-risk periodontal markers. When investigating additional factors that may contribute to these observations, we found that hyperlipidemia and periodontal pocket depth dependently impacted the oral microbiota, while glycated hemoglobin (HbA1c) independently affected the oral microbiota of these children. Our findings support an interrelationship between the familial history of hyperlipidemia, periodontal risk factors, and type 1 diabetes, and show that a family history of hyperlipidemia is linked to a significant effect on the oral microbiota of periodontally healthy children with type 1 diabetes. Further investigations are required to understand if these differences in oral microbiota composition are reflective of lipid profiles from the children themselves.

5.2. Introduction

Dyslipidemia is a condition where there is an abnormal concentration of blood lipids, such as fats, triglycerides, and cholesterol. In industrialized countries, most dyslipidemia cases are classified as hyperlipidemia (elevated blood lipids) and are a consequence of many factors, including poor diet, smoking, sedentary lifestyle, diabetes, and familial hyperlipidemia (Expert Panel on Integrated Guidelines for Cardiovascular Health and Risk Reduction in Children and Adolescents 2011). Many hyperlipidemia cases go undetected during the early stages of the condition, putting individuals at higher risk for developing cardiovascular diseases, such as atherosclerosis (Elkins et al. 2019). However, individuals who have a genetic predisposition to particular hyperlipidemias, especially hypercholesterolemia (high abundance of cholesterol), can be given statin-based treatments to control blood lipids and slow hyperlipidemia progression (Luirink et al. 2019).

Hyperlipidemia is a multifactorial condition that has been linked to both genetic and environmental factors (Costanza et al. 2005). From genome wide association studies, over 150 human genomic loci were associated with abnormal lipid levels in the bloodstream (Matey-Hernandez et al. 2017). However, more than half of the variation associated with circulating lipid levels in the blood has been linked to non-genetic factors, including the microbial communities that colonize the body (microbiota) (Matey-Hernandez et al. 2017). In mouse models, gut microbiota have been associated with regulation of blood lipids using particular lipoprotein lipase inhibitors, such as fast-inducing adipose factor (FIAF; Bäckhed et al. 2004). Further, gut microbiota co-produce secondary acids, such as short-chain fatty acids, bile acids, and conjugated linoleic acids, which are used in metabolic pathways, such as regulating lipids (Allayee and Hazen 2015) and cholesterols (Le Roy et al. 2019). Although the relationship between hyperlipidemia and gut microbiota is emerging, the specific, systemic mechanisms that underpin these interactions needs further investigation.

Interactions between hyperlipidemia and microbiota elsewhere in the body also needs further exploration, as relationships between hyperlipidemia and other microbiota-associated diseases, such as type 1 diabetes (T1D) and periodontitis, have been proposed (Zhou et al. 2015). In fact, 29-66% of children with T1D have also been shown to have hyperlipidemia (Zabeen et al. 2018), and hyperlipidemia is a risk factor for periodontal disease (Lee et al. 2018). Although T1D has a strong genetic inheritance (Steck and Rewers 2011), recent mouse and human studies

identified additional correlations between gut microbiota and T1D (Zhou et al. 2020). For instance, decreased abundances of *Bifidobacterium* and butyrate-producing bacteria in the gut, as well as lower overall gut microbial diversity, have been observed in human T1D compared to healthy individuals (Zhou et al. 2020). We hypothesize that microbiota may play a role in the connection between both diseases, as well as their interactions with other health outcomes. For example, uncontrolled T1D and hyperlipidemia can negatively impact wound healing (Abraham et al. 2019) and increases the risk for developing periodontal disease (Zhou et al. 2015; Abraham et al. 2019). Indeed, our previous study showed that glycemic control and periodontal markers in children with T1D can influence the oral microbiota (**Appendix III** Jensen et al. 2020), although the additional interactions with hyperlipidemia have not yet been explored.

In this post-hoc study, we explored the effect of hyperlipidemia (parent status) on the oral microbiota of children with T1D (n=72), who were enrolled in a study to investigate the effects of glycemic control (HbA1c) and periodontal risk markers, including bleeding on probing, plaque index, gingival index, brushing frequency, and periodontal pocket depth (PD), on the oral microbiota (**Appendix III** Jensen et al. 2020). We previously confirmed links between periodontal disease characteristics and changes in oral microbiota (**Appendix III** Jensen et al. 2020), but wanted to further explore the links between lipid levels and the oral microbiota of these children, while accounting for periodontal risk markers. Our goal was to explore the oral microbiota relationship between T1D, periodontal PD, and the family history (FHx) of hyperlipidemia in these children.

5.3. Materials and methods

5.3.1. Study cohort and sample collection

Seventy-six children with T1D were recruited at the Pediatric Diabetes Clinic at the Women's and Children's Hospital, Adelaide, Australia between February 2018 to February 2019, under ethics approval (Women's and Children's Health Network Human Research Ethics Committee; HREC/17/WCHN/165). Children were between 8-18 years and had T1D that was diagnosed by detectable islet cell autoantibodies. It was advised that children who had a fever or infection, diabetic ketosis, or were currently taking a course of antibiotics, reschedule their appointment. Informed written consent was acquired from all parents or guardians

of children under 16 of age, or from the child themselves if they were 16 years or older. As part of the study, participants, or their parents/guardians were verbally interviewed and diabetes data were obtained from participant's medical records, and periodontal risk markers were collected by a practitioner at the time of dental examination (see Table 1 for further details).

A gingival swab was collected from the buccal-gingival margin of the lower left first permanent molar of each child by a dentist undertaking pediatric specialist training. Swabs were stored in sterile tubes and immediately placed in a -80°C freezer. Frozen samples were later transported in a cool box with freezer packs to a dedicated low-biomass microbiome laboratory at the University of Adelaide. Samples were kept frozen until DNA extraction.

Table 1. Summary of study participant characteristics.

Characteristic	Participants (n = 76)	Participants after filtering samples (n=72)
Age (years)	13.3 ± 2.6	12.9 ± 2.5
Female gender (n, %)	39 (51.3)	38 (52.7)
Duration of T1D (years)	5.50 ± 3.82	5.56 ± 3.90
HbA1c (%, median, range)	8.1 (5.8–13.3)	7.95 (5.8–13.3)
BMI z-score	0.80 ± 0.76	0.79 ± 0.77
Plaque index	0.93 ± 0.52	0.92 ± 0.52
Gingival index	0.67 ± 0.43	0.66 ± 0.43
Bleeding on probing (n, %)	21.52 ± 19.48	21.18 ± 19.26
Pocket depth >3 mm (count, %)	36 (47.4)	34 (47.2)
Family history of disease (n, %)		
Hyperlipidemia	30 (39.5)	30 (41.7)
Cardiovascular disease	20 (27.6)	20 (27.8)
Obesity	35 (46.1)	34 (47.2)
T1D	26 (34.2)	24 (33.3)
T2D	47 (61.8)	45 (62.5)

Mean ± SD unless specified; T1D = type 1 diabetes; BMI = body mass index.

5.3.2. Bioinformatic, microbial diversity, and statistical analyses

As the data for this study was publicly available, demultiplexed DNA sequences were downloaded from the QIITA repository (Study ID 13235) and were joined using vsearch (Rognes et al. 2016) in QIIME2 (Bolyen et al. 2019; v2019.7). Next, sequences were quality assessed and denoised at 250 bp using Deblur (Amir et al. 2017). A SEPP insertion tree (Mirarab et al. 2012) was created, and sequences were assigned using the SILVA database (v132; 16S rRNA gene 515-806) using default parameters. Using decontam (Davis et al. 2018), contaminant amplicon sequence variants (ASVs) were identified from EBCs and NTCs (prevalence threshold set to 0.6) and were removed from gingival samples. Samples with low sequencing depth (<5,000 sequences) or incomplete data and ASVs with <11 sequences per ASV were also removed.

Microbial diversity and statistical analyses were performed in QIIME2. At a rarefied depth of 5,000 sequences, diversity (alpha diversity) was calculated using observed ASVs (richness) and Faith's phylogenetic diversity (Faith 1992) metrics. Significant associations between diversity and categorical metadata were tested using a Kruskal-Wallis (Kruskal and Wallis 1952) pairwise statistical test, whereby FDR adjusted p-values (q-values) were reported. Similarly, differences in bacterial composition diversity were calculated using unweighted UniFrac (Lozupone and Knight 2005) and Bray-Curtis metrics at a rarefied depth of 5,000 sequences. Both PERMANOVA (Anderson 2017) and adonis (Anderson 2001) tests were used to determine statistical differences between groups and the variation that a metadata category contributes to the composition, respectively. We also explored confounding factors between HbA1c, periodontal PD ,and FHx hyperlipidemia using linear mixed models in R (version 3.6.1).

Genera co-correlations were examined by exporting biom tables from QIIME2 and importing them into MEGAN6 (Huson et al. 2016). Taxonomy for each ASV was checked and adjusted manually due to differences between the SILVA and BLAST taxonomy strings. Equal numbers of samples were compared in each instance (n=11 per group). In MEGAN6, ASV data was collapsed to the genus taxonomic level and visualized as co-occurrence plots using the Pearson correlation test with the following parameters: observed in 80% to 100% of samples and had an edge threshold of 80%.

5.4. Results

5.4.1. A robust oral microbiota signal was achieved from gingival swabs

To ensure a robust oral microbiota signal was achieved, contaminant ASVs and poorly sequenced samples were investigated. Using decontam, 54 ASVs were identified as contaminants from EBCs and NTCs (**Appendix III** Table S1) and removed from the biological samples. From 76 samples, three individuals had incomplete metadata, while one sample had a low sequencing depth; these four samples were discarded from all downstream analyses. In the remaining 72 samples (Table1), 905 total ASVs were identified from 2,567,677 sequences (median sequences per sample: 33,920).

We next characterized the gingival microbiota of all children with T1D. A total of 12 phyla were detected from 72 gingival samples. The most abundant phyla (>2% mean relative abundance) were Proteobacteria (43.3%), Firmicutes (30.7%), Bacteroidetes (9.7%), Actinobacteria (8.3%), and Fusobacteria (6.3%) (**Appendix III** Figure S1). From 131 total identified genera, 12 genera were highly abundant (>2% relative abundance): *Haemophilus* (19.2%), *Streptococcus* (16.3%), *Neisseria* (10.3%), *Veillonella* (9.9%), *Aggregatibacter* (5.3%), *Fusobacterium* (3.8%), *Actinobacillus* (3.2%), *Actinomyces* (2.8%), *Leptotrichia* (2.5%), *Corynebacterium* (2.4%), *Prevotella* (2.0%), and *Rothia* (2.0%).

Previous studies identified that glycated hemoglobin (HbA1c) and blood lipid counts are correlated (Giuffrida et al. 2012) and can increase the risk and severity of periodontal disease (Zhou et al. 2015). Therefore, we wanted to explore whether or not HbA1c, periodontal PD, and FHx hyperlipidemia were confounded. Using linear mixed models, we determined all three factors were not interacting with one another (**Appendix III** Table S2). Once we removed this three-way insignificant interaction and any two-way insignificant interactions, we determined that glycated hemoglobin influenced the oral microbiota independently of periodontal PD and FHx hyperlipidemia (**Appendix III** Table S3), and that periodontal PD was confounded by a FHx hyperlipidemia. Therefore, we accounted for periodontal PD when investigating FHx hyperlipidemia by grouping samples as low periodontal risk (periodontally healthy; no periodontal pockets with depth >3 mm; individuals=38) or high periodontal risk (at least one periodontal pocket with depth >3 mm; individuals=34) for the following analyses.

5.4.2. Periodontally healthy children with a family history of hyperlipidemia have decreased oral microbiota diversity

As hyperlipidemia is related to both gut microbiota changes and periodontal disease (Matey-Hernandez et al. 2017; Lee et al. 2018), we wanted to examine links between oral microbiota health and FHx hyperlipidemia. To determine if a FHx hyperlipidemia altered oral diversity of children with T1D, we measured diversity using two metrics: microbial richness and Faith's phylogenetic diversity. In periodontally healthy children, children with a FHx hyperlipidemia had significantly lower oral microbial diversity compared to those without a FHx hyperlipidemia (Kruskal-Wallis; phylogenetic: H=11.174, q=0.001; richness: H=8.616, q=0.003; Figure 1A-B). However, this was not in the case in children with high periodontal risk, as there were no significant differences between children with and without a FHx hyperlipidemia (Faith's PD: H=0.009, q=0.927; richness: H=0.049, q=0.825; Figure 1C-D). These results suggest that gingival microbial diversity is lower in periodontally healthy children with a FHx hyperlipidemia.

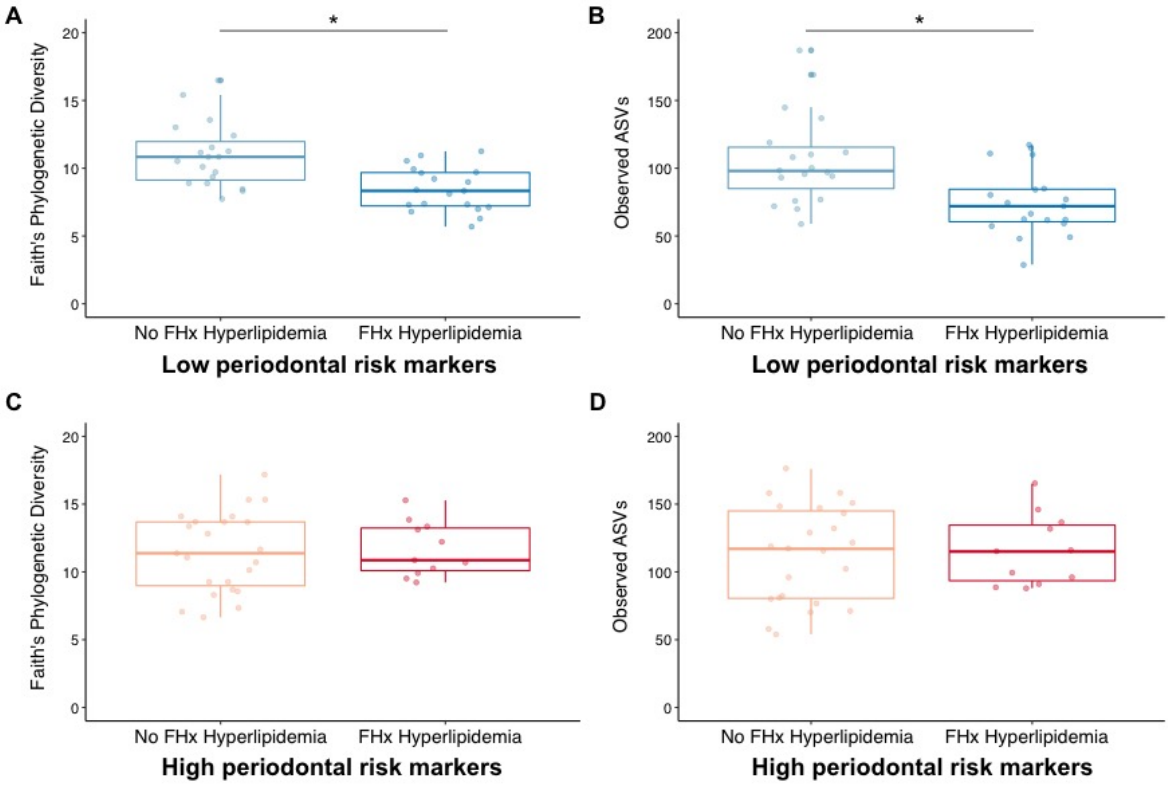

Figure 1: Periodontally healthy children with a family history of hyperlipidemia have lower oral microbial diversity. Samples were rarefied to 5,000 sequences and alpha diversity was measured using faith's phylogenetic diversity (A and B) and microbial richness (observed ASVs; C and D) for children with low (blue) and high (red) periodontal risk markers. Within each analysis, diversity was compared between children without (light) and with (dark) a FHx hyperlipidemia. * $p<0.05$

5.4.3. Periodontally healthy children with a family history of hyperlipidemia have altered oral microbiota composition

We wanted to see if the gingival microbiota composition was different between children with and without a FHx hyperlipidemia, given the altered gingival microbial diversity in these children. We measured compositional variation using unweighted UniFrac and found a significant difference in phylogenetic composition between periodontally healthy children with and without a FHx of hyperlipidemia (unweighted UniFrac; PERMANOVA: pseudo-F=4.281, q=0.001; adonis: R2=0.117, q=0.02; Figure 2A). However, there were no significant differences for weighted non-phylogenetic composition between these groups (Bray-Curtis: pseudo-F=1.164, q=0.265; Figure 2B). In children with high-risk periodontal markers, we saw no

compositional differences between children with and without a FHx hyperlipidemia (PERMANOVA; unweighted UniFrac: pseudo-F=0.743, q=0.725; Bray-Curtis: pseudo-F=0.948, q=0.511; Figure 2C&D). Overall, we see a phylogenetic compositional shift in periodontally healthy children with FHx of hyperlipidemia.

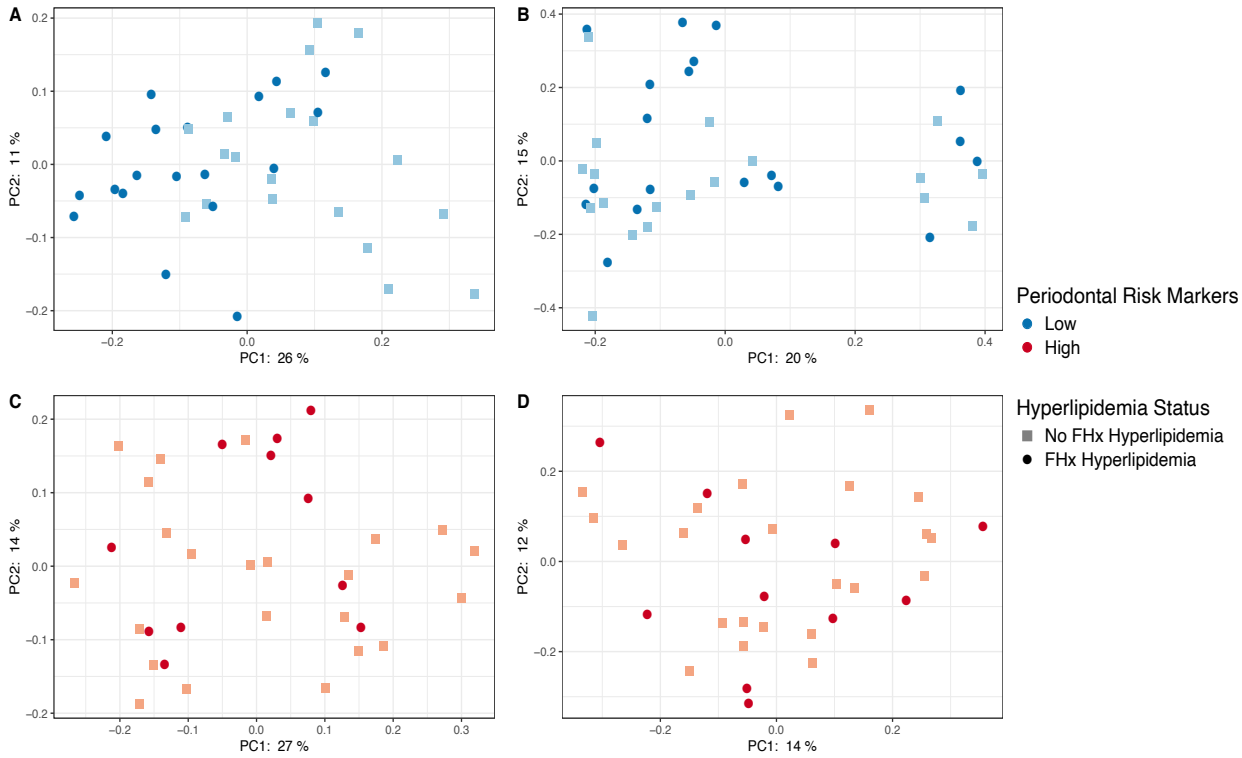

Figure 2: Periodontally healthy children with a family history of hyperlipidemia have a distinctly different phylogenetic microbial composition to those that do not have a family history of hyperlipidemia. PCoA plots were generated to compare compositional differences using unweighted UniFrac and Bray-Curtis metrics. Children were separated based on having low-risk (blue; A-B) or high-risk (red; C-D) periodontal markers before comparing children with and without with a family history of hyperlipidemia.

5.4.4. Periodontally healthy children with a family history of hyperlipidemia have less unique taxa

To investigate the species that underpin diversity and composition differences in periodontally healthy children with FHx hyperlipidemia, we explored

specific taxa that could contribute to these differences. Using ANCOM, a *Prevotella* ASV was found to be significantly more abundant in periodontally healthy children with a FHx hyperlipidemia compared to children without (W=202). As only one ASV was significantly more abundant, we looked for the presence of any unique ASVs in periodontally healthy FHx children compared to those with no family history of hyperlipidemia. There were more unique ASVs observed in children without a FHx hyperlipidemia (412 unique ASVs; n=19 samples) in comparison to children with a FHx hyperlipidemia (84 unique ASVs; n=18 samples). Overall, periodontally healthy children with a FHx hyperlipidemia have increased levels of a specific *Prevotella* ASV and fewer unique species.

5.4.5 Periodontally healthy children with a family history of hyperlipidemia have unique microbial networks compared to children with high-risk periodontal markers

As we observed differences in periodontally healthy children with and without FHx dyslipidemia, we wanted to explore how these shifts may alter biofilm related co-occurrence networks in the mouth. Using the Pearson metric, we identified distinct networks of co-occurring genera between periodontally healthy children with a FHx hyperlipidemia and children without FHx. Specifically, four networks were observed in these children with no FHx hyperlipidemia (**Appendix III** Figure S2A), which included networks of 1.) *Campylobacter* and *Fusobacterium*; 2.) *Rothia* and *Corynebacterium*; 3.) *Alloprevotella*, *Porphyromonas,* and *Capnocytophaga*; and 4.) *Streptococcus*, *Gemella*, and *Bergeyella*. In periodontally healthy children with a FHx hyperlipidemia, there were only two co-occurrence networks: 1.) *Corynebacterium* and *Capnocytophaga*; and 2.) *Fusobacterium* and *Leptotrichia* (**Appendix III** Figure S2B). This indicates that FHx hyperlipidemia may reduce the microbial networks in periodontally healthy children and shift the microbial community to a more mature plaque structure.

We next examined co-occurrence networks in children with periodontal disease who did or did not have a FHx hyperlipidemia, as these networks can be altered in individual's with periodontal disease (Chen et al. 2018; Lamont et al. 2018). In children with high-risk periodontal markers and no FHx hyperlipidemia, a single network of six species commonly associated with periodontal disease (*Prevotella*, *Porphyromonas*, *Fusobacterium*, *Campylobacter*, *Capnocytophaga*,

and *Leptotrichia*) was detected (**Appendix III** Figure S2C). However, children with high periodontal risk markers and a FHx hyperlipidemia maintained four distinct networks (**Appendix III** Figure S2D) and notably did not include a network with *Porphyromonas*. One network did link *Prevotella*, *Fusobacterium*, *Leptotrichia*, and *Corynebacterium* species together, again to the exclusion of *Porphyromonas*. As the prevalence in *Porphyromonas* is typically considered critical in the development of periodontal disease, this observation may suggest that the networks of microbes that underpin periodontal disease development in children with a FHx hyperlipidemia may be unique compared to children without this familial history.

5.5. Discussion

Hyperlipidemia is a multifactorial disease that may provide a link between T1D and periodontal disease. In this study, we explored the relationship between hyperlipidemia and periodontal pocket depth (PD) in children with T1D using the parent's hyperlipidemia status (FHx hyperlipidemia) as a proxy for abnormal lipid counts in their child, given genetic and environmental factors similarities between parents and children (Ang et al. 2013; Robledo and Siccardi 2016; Filgueiras et al. 2019). Lower microbial diversity and changes to microbial composition were correlated with a FHx hyperlipidemia in periodontally healthy children (no periodontal pockets with depth >3 mm), but this was not observed for children with high-risk periodontal markers (at least one periodontal pocket with depth >3 mm). Further, increased abundance of a *Prevotella* ASV and fewer unique ASVs were also observed in this group. In periodontally healthy children with FHx hyperlipidemia, we also observed networks likely associated with more mature plaque structures. Altogether, these microbial differences may indicate that either genetic and/or environmental factors related to FHx hyperlipidemia may be altering the oral microbiota in children with T1D.

Although we did not have direct lipid information of children in this study, we used the status of hyperlipidemia from first-degree relatives (*i.e.* parent/s) as an indication of potentially abnormal lipid levels in children. This assumption is derived from the likelihood that children are exposed to similar lifestyle habits, *e.g.*, diet and exercise patterns, and they share the same genetic make-up as their parents (Ang et al. 2013; Robledo and Siccardi 2016; Filgueiras et al. 2019). Previous work

also suggests that between 29-66% of children with T1D also had hyperlipidemia (Mona et al. 2015; Zabeen et al. 2018), aligning well with our assumptions about FHx hyperlipidemia in this cohort. Further, Guy et al. (2009) showed that children with T1D had elevated apolipoprotein B and LDL levels (*i.e.* early stages of hyperlipidemia), but this did not affect HbA1c levels, again similar to what we observed in this cohort. Nevertheless, it is possible that children within this cohort would not go on to develop hyperlipidemia, and future studies should directly examine blood lipid levels and lifestyle factors in children to correlate changes in microbiota more accurately.

A bi-directional relationship between hyperlipidemia and periodontal disease has been described, but how this affects the oral microbiota has not been previously explored. Both mouse and human studies have shown that high fat diets and high blood lipids are associated with periodontitis (Zhou et al. 2015). Increases in blood lipids can elevate proinflammatory cytokines, which can then reside in the gingival crevicular fluid and promote inflammation of the gingiva. This process of gingival inflammation and the infiltration of proinflammatory cytokines is indicative of periodontal disease (Zhou 2017). Most periodontal studies have also suggested that gingival inflammation is also likely linked to the outgrowth of *Porphyromonas gingivalis* species during the development of periodontal disease. In periodontal disease, inflammation is also typically preceded by the development of large dental plaque structures. Our co-occurrence analysis of children with high periodontal risk markers revealed that *Porphyromonas* was not involved in the main networks of microbes in children with FHx hyperlipidemia, although species were present for mature plaque formation. However, *Prevotella* ASVs, which was also the only ASV with a significant association with FHx hyperlipidemia children, were identified in the networks of children with high-risk periodontal markers and FHx hyperlipidemia. It is possible that non-*Porphyromonas* species, such as *Prevotella* species, can significantly contribute to the development of periodontal disease in these children, or that the microbial ecosystem in these children is disrupted in unique ways with putative pathogens that can increase in abundance during periodontal disease and stimulate an immune response (Curtis et al. 2020). Monitoring the microbiota in these children over time and investigating the presence of a tipping point or control of oral microbiota in relationship to periodontal disease is needed in the future. This analysis should involve examining

both taxonomic and functional differences of oral microbiota in children with and without a FHx hyperlipidemia.

Currently, hyperlipidemia goes undetected in the early stage of disease progression (Elkins et al. 2019) and is only recommended to be monitored every five years in children with diabetes (American Diabetes Association 2003). Further, the salivary lipid profiles reflect serum lipid levels (Al-Rawi 2011), which suggests that oral microbiota may be responding to changes in salivary lipid prevalence. Downstream studies should investigate if the oral microbiota could be used as an early prediction of hyperlipidemia in children with diabetes. For example, gingival microbiota swab samples could be collected during a dental visit to determine if changes to the oral microbiota, *i.e.* a decrease in microbial diversity has occurred and warrants more invasive testing. Collection of gingival samples is less invasive than blood samples and can be collected easily at routine dental visits, by individuals with less medical training, and from patients with less co-operative ability. Similarly, oral microbiota samples could also be taken to monitor changes in periodontal health in these children. Increasing our understanding of periodontal and metabolic diseases from a holistic approach can be more beneficial to not only detect multifactorial diseases, but also to prevent and treat them.

As the relationship between diabetes, hyperlipidemia, and periodontal disease ranges from the gut to the mouth and includes the circulatory system, it is important to think of the body as a whole system, rather than focusing on one body site (Faust et al. 2012). For example, diabetes studies conventionally focus on gut microbiota, while periodontal studies predominantly focus on the oral microbiota. While a few emerging studies have now begun to examine links between oral and gut microbiota in periodontal disease (Gatej et al. 2020), future studies should focus on the mechanistic interactions between increased circulating glycated hemoglobin, lipids, and proinflammatory molecules. As we start to know more about T1D and potential treatments, it is important to consider the entire ecosystem of both the mouth and gut, which includes microbiota and the effects of circulating metabolites and cytokines.

5.6. Declarations

Ethics approval and consent to participate

Ethics approval for this project was obtained from the Women's and Children's Health Network Human Research Ethics Committee (HREC/17/WCHN/165). Informed written consent was acquired from parents, caretakers, or from the child themselves (if they were at least 16 years of age). Participants, or their parents/caretakers filled out information sheets, and medical records were obtained by practitioners

Availability of data and material

The dataset analyzed in this article was obtained from the QIITA repository with Study ID 13235 (https://qiita.ucsd.edu/study/description/13235).

Funding

This study was funded by LSW Australian Research Council Future Fellowship (FT180100407).

Author contributions

C.A.S, L.S.W., E.J., and G.A., designed the research study; E.J. recruited participants and obtained consent; C.A.S analyzed the microbiota data, C.A.S. and L.S.W interpreted the data; C.A.S. wrote the initial draft of the manuscript; L.S.W., E.J., and G.A. critically revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This research was supported by Australian Research Council Future Fellowship (FT180100407). CAS was supported by an Australian Government Research Training Program Scholarship. We would like to thank Jennifer Couper and Alexia Pena for designing the initial project and for contributing their expertise to this project. We would like to thank Ray Tobler for assisting with mixed model analyses. We would also like to thank the thesis writing group at the Australian Centre for Ancient DNA for editing an early draft. We are grateful to the children who participated in this study and their families.

Competing Interests

The authors have no conflict of interests to disclose.

5.7. References

Al-Rawi NH. 2011. Oxidative stress, antioxidant status and lipid profile in the saliva of type 2 diabetics. Diabetes and Vascular Disease Research. 8(1):22–28. doi:10.1177/1479164110390243. [accessed 2020 Jul 29]. https://doi.org/10.1177/1479164110390243.

American Diabetes Association. 2003. Management of Dyslipidemia in Children and Adolescents With Diabetes. Diabetes Care. 26(7):2194–2197. doi:10.2337/diacare.26.7.2194. [accessed 2020 Aug 3]. https://care.diabetesjournals.org/content/26/7/2194.

Amir A, McDonald D, Navas-Molina JA, Kopylova E, Morton JT, Xu ZZ, Kightley EP, Thompson LR, Hyde ER, Gonzalez A, et al. 2017. Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. mSystems. 2(2):e00191-16. doi:10.1128/mSystems.00191-16.

Anderson MJ. 2001. A new method for non-parametric multivariate analysis of variance. Austral ecology. 26(1):32–46. doi:https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x.

Anderson MJ. 2017. Permutational Multivariate Analysis of Variance (PERMANOVA). In: Wiley StatsRef: Statistics Reference Online. American Cancer Society. p. 1–15. https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118445112.stat07841.

Ang YN, Wee BS, Poh BK, Ismail MN. 2013. Multifactorial Influences of Childhood Obesity. Curr Obes Rep. 2(1):10–22. doi:10.1007/s13679-012-0042-7. [accessed 2020 Jul 29]. https://doi.org/10.1007/s13679-012-0042-7.

Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, et al. 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nature Biotechnology. doi:10.1038/s41587-019-0209-9. https://doi.org/10.1038/s41587-019-0209-9.

Chen C, Hemme C, Beleno J, Shi ZJ, Ning D, Qin Y, Tu Q, Jorgensen M, He Z, Wu L, et al. 2018. Oral microbiota of periodontal health and disease and their changes after nonsurgical periodontal therapy. The ISME Journal. 12(5):1210–1224. doi:10.1038/s41396-017-0037-1. [accessed 2020 Jul 29]. https://www.nature.com/articles/s41396-017-0037-1.

Curtis MA, Diaz PI, Dyke TEV. 2020. The role of the microbiota in periodontal disease. Periodontology 2000. 83(1):14–25. doi:10.1111/prd.12296. [accessed 2020 Aug 3]. https://onlinelibrary.wiley.com/doi/abs/10.1111/prd.12296.

Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. 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. [accessed 2019 Jan 30]. https://doi.org/10.1186/s40168-018-0605-2.

Elkins C, Fruh S, Jones L, Bydalek K. 2019. Clinical Practice Recommendations for Pediatric Dyslipidemia. Journal of Pediatric Health Care. 33(4):494–504. doi:10.1016/j.pedhc.2019.02.009. [accessed 2020 Jul 6]. https://www.jpedhc.org/article/S0891-5245(18)30408-5/abstract.

Faith DP. 1992. Conservation evaluation and phylogenetic diversity. Biological Conservation. 61(1):1–10. doi:10.1016/0006-3207(92)91201-3. [accessed 2020 Jan 8]. http://www.sciencedirect.com/science/article/pii/0006320792912013.

Faust K, Sathirapongsasuti JF, Izard J, Segata N, Gevers D, Raes J, Huttenhower C. 2012. Microbial Co-occurrence Relationships in the Human Microbiome. PLoS Comput Biol. 8(7). doi:10.1371/journal.pcbi.1002606. [accessed 2020 Aug 3]. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3395616/.

Filgueiras M de S, Vieira SA, Ribeiro AQ, Novaes JF de, Filgueiras M de S, Vieira SA, Ribeiro AQ, Novaes JF de. 2019. FAMILY HISTORY IS ASSOCIATED WITH THE PRESENCE OF DYSLIPIDEMIA IN PRE-SCHOOL CHILDREN. Revista Paulista de Pediatria. 37(1):41–48. doi:10.1590/1984-0462/;2019;37;1;00005. [accessed 2020 Jul 29]. http://www.scielo.br/scielo.php?script=sci_abstract&pid=S0103-05822019000100041&lng=en&nrm=iso&tlng=en.

Gatej SM, Bright R, Weyrich LS, Marino V, Christophersen CT, Gibson RJ, Gully N, Zilm P, Bartold PM. 2020. Probiotic Lactobacillus Rhamnosus GG Protects Against P. Gingivalis And F. Nucleatum Gut Dysbiosis. J Int Acad Periodontol. 22(2):18–27.

Giuffrida FM, Guedes AD, Rocco ER, Mory DB, Dualib P, Matos OS, Chaves-Fonseca RM, Cobas RA, Negrato CA, Gomes MB, et al. 2012. Heterogeneous behavior of lipids according to HbA1c levels undermines the plausibility of metabolic syndrome in type 1 diabetes: data from a nationwide multicenter survey. Cardiovascular Diabetology. 11(1):156. doi:10.1186/1475-2840-11-156. [accessed 2020 Aug 3]. https://doi.org/10.1186/1475-2840-11-156.

Guy J, Ogden L, Wadwa RP, Hamman RF, Mayer-Davis EJ, Liese AD, D'Agostino R, Marcovina S, Dabelea D. 2009. Lipid and lipoprotein profiles in youth with and without type 1 diabetes: the SEARCH for Diabetes in Youth case-control study. Diabetes Care. 32(3):416–420. doi:10.2337/dc08-1775.

Huson DH, Beier S, Flade I, Górska A, El-Hadidi M, Mitra S, Ruscheweyh H-J, Tappu R. 2016. MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. PLOS Computational Biology. 12(6):e1004957. doi:10.1371/journal.pcbi.1004957. [accessed 2020 Jul 14]. https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004957.

Jensen ED, Selway CA, Allen G, Bednarz J, Weyrich LS, Gue S, Peña AS, Couper J. 2020. Early markers of periodontal disease and altered oral microbiota are associated with glycemic control in children with type 1 diabetes. Pediatric Diabetes. n/a(n/a). doi:https://doi.org/10.1111/pedi.13170. [accessed 2021 Feb 2]. https://onlinelibrary.wiley.com/doi/abs/10.1111/pedi.13170.

Kruskal WH, Wallis WA. 1952. Use of Ranks in One-Criterion Variance Analysis. Journal of the American Statistical Association. 47(260):583–621. doi:10.2307/2280779. [accessed 2020 Jan 8]. https://www.jstor.org/stable/2280779.

Lamont RJ, Koo H, Hajishengallis G. 2018. The oral microbiota: dynamic communities and host interactions. Nat Rev Microbiol. 16(12):745–759. doi:10.1038/s41579-018-0089-x. [accessed 2020 Jul 29]. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6278837/.

Lee S, Im A, Burm E, Ha M. 2018. Association between periodontitis and blood lipid levels in a Korean population. Journal of Periodontology. 89(1):28–35. doi:10.1902/jop.2017.170111. [accessed 2020 Jul 24]. https://aap.onlinelibrary.wiley.com/doi/abs/10.1902/jop.2017.170111.

Lozupone C, Knight R. 2005. UniFrac: a new phylogenetic method for comparing microbial communities. Applied and environmental microbiology. 71(12):8228–8235. doi:10.1128/AEM.71.12.8228-8235.2005.

Matey-Hernandez ML, Williams FMK, Potter T, Valdes AM, Spector TD, Menni C. 2017. Genetic and microbiome influence on lipid metabolism and dyslipidemia. Physiological Genomics. 50(2):117–126. doi:10.1152/physiolgenomics.00053.2017. [accessed 2020 Jul 24]. https://journals.physiology.org/doi/full/10.1152/physiolgenomics.00053.2017.

Mirarab S, Nguyen N, Warnow T. 2012. SEPP: SATé-enabled phylogenetic placement. Pac Symp Biocomput.:247–258. doi:10.1142/9789814366496_0024.

Mona HM, Sahar SA, Hend SM, Nanees A-WA. 2015. Dyslipidemia in type 1 diabetes mellitus: Relation to diabetes duration, glycemic control, body habitus, dietary intake and other epidemiological risk factors. Egyptian Pediatric Association Gazette. 63(2):63–68. doi:10.1016/j.epag.2015.03.001. [accessed 2020 Jul 29]. http://www.sciencedirect.com/science/article/pii/S1110663815000075.

Robledo JA, Siccardi LJ. 2016. Relationship between genetic and environmental factors and hypercholesterolemia in children. Arch Argent Pediatr. 114(5):419–425. doi:10.5546/aap.2016.419.

Rognes T, Flouri T, Nichols B, Quince C, Mahé F. 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ. 4:e2584. doi:10.7717/peerj.2584. [accessed 2020 Jul 13]. https://peerj.com/articles/2584.

Zabeen B, Balsa AM, Islam N, Parveen M, Nahar J, Azad K. 2018. Lipid Profile in Relation to Glycemic Control in Type 1 Diabetes Children and Adolescents in Bangladesh. Indian J Endocrinol Metab. 22(1):89–92. doi:10.4103/ijem.IJEM_217_17. [accessed 2020 Jul 24]. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5838919/.

Zhou X, Zhang Wenyi, Liu X, Zhang Wu, Li Y. 2015. Interrelationship between diabetes and periodontitis: role of hyperlipidemia. Arch Oral Biol. 60(4):667–674. doi:10.1016/j.archoralbio.2014.11.008.

Thesis Discussion

Thesis summary and significance

Non-communicable diseases (NCDs) are the world's biggest killer.[1] Since the 1950's, the prevalence, morbidity, and mortality of NCDs have increased steadily to the point where NCDs are now a public health crisis. The sudden increase in NCDs is not reflected by a rapid genetic change in humans, suggesting that environmental and behavioural factors and the human microbiome are important contributors. Unnatural changes to the environment (*e.g.*, climate change, pollution, urbanisation, *etc.*) and certain lifestyle factors (*e.g.*, diet, smoking status, activity levels, *etc.*) have been identified as detrimental to human health and are the targeted factors in the Sustainable Development Goals to reduced NCD-related premature death.[2] However, microbiome research is in its infancy, and less is known about how the microbiota impacts the initiation, progression, and prevalence of NCDs. As such, this thesis aimed to provide more insights into development of and disruption to the human microbiome associated with NCDs.

The overarching theme of this thesis is unravelling how the environment and disease impact the microbiome in areas of the human body with low microbial biomass and expands our knowledge of interactions and relationships between body sites and systems of the human body. With low microbial biomass samples comes unique challenges (see **Chapter I**) that must be overcome to appreciate how the environment impacts the skin, nasal, oral and lung microbiota (see **Chapters III** and **IV**) and how disease alters these microbiota (see **Chapters IV** and **V**). It is important to highlight that while low microbial biomass sites provide information for local sites, there are hints of systemic interactions on other microbiotas and systems of the human ecosystem (see **Chapters II** and **V**). Using a holistic approach may provide key insights into the development and progression of NCDs. Using low microbial biomass samples, this thesis provides more information on the microbiome contribution to NCDs in non-gut sites.

In this discussion chapter, I summarise, interpret, and highlight the significance of my research and suggest future applications of 'non-gut' microbiotas for local and systemic research in relation to NCDs. This discussion explores my results through three main themes:

1. Increase awareness of the challenges, benefits, and technical approaches used to examine low microbial biomass samples.
2. Understanding disease from a holistic perspective—the systems and microbiotas of our body are interconnected.

3. Provide insights into the predisposition, initiation, progression, and potential treatments of NCDs.

The challenges and benefits of low biomass samples

Challenges of low biomass samples and considerations for future research

DNA within low microbial biomass samples can be scarce, which presents technical difficulties for obtaining, processing, and bioinformatically analysing these samples. In this thesis, I observed and assessed three dominant issues: 1) contamination of background DNA; 2) the effects of differing biomass within a study; and 3) the stochasticity of samples. Below, I discuss each of these issues in greater detail and provide strategies to mitigate these potential issues in future studies.

Contamination

Background levels of DNA—commonly referred to as contamination—can be unintentionally incorporated into a sample at any stage, from sample collection to sequencing. It is now recognised that contamination is unavoidable, and every microbial sample will be affected by contamination to some extent.[3] For high microbial biomass samples, such as those from the gut, the effect of contamination can be negligible.[4,5] With this unconscious bias in mind, contamination in earlier microbiome studies focused on the gut failed to consider this, and therefore, was not addressed correctly when researchers started analysing low microbial biomass samples.[6–8] Microbial DNA contamination is proportional to the biological DNA in most cases. For example, Karstens *et al.* (2019) showed that a mock community diluted to a level equivalent to a low microbial biomass samples (10^8 dilution) contained ~80% of contaminant sequences, while the undiluted mock community contained <0.05% of contaminant sequences.[5] For this reason, it is essential that contamination is considered, mitigated, and monitored for low microbial biomass samples.

An appreciation for identifying and analysing contaminating DNA through both positive and negative controls has become more widespread in the microbiome field. As the popularity and recognised importance of microbiome research

increases, researchers from a wide range of fields are incorporating microbiome analyses into their studies, but many are not adhering to key practices, such as contamination mitigation and monitoring. Other fields, such as forensics and ancient DNA, faced with similar issues with contamination in their field's infancy, which has now seen strict protocols put into place. For example, Cooper and Poinar (2002) presented a letter in *Science* that emphasised key practices, such as using a dedicated ancient lab and including negative controls, that should be applied for any study working with ancient DNA.[9] Similar publications have been presented in the microbiome field, which target researchers,[3,10–13] clinicians,[14,15] and doctors in their respected area of expertise.[16–20] In addition, **Chapter I** provides a summary of common pitfalls and paired mitigation strategies for contamination and biases that pathologists may come across.[21] It is imperative that pathologists and technicians abide by these mitigation strategies, as negligence may result in incorrect diagnosis and treatment that could have disastrous outcomes for patients.

As technology has improved with an increased interest in microbiome research, new strategies have been implemented to mitigate and identify contamination. Below, I present strategies that were identified in **Chapter I** and newer strategies that have since been presented that can help to minimise contamination in low microbial biomass samples. These strategies should be considered and applied in all future microbiome studies.

1) *Working in a sterile environment and incorporating methods that minimise human, environmental and reagent contamination*

Due to the strict conditions utilised in ancient DNA studies, similar methods could be adopted for microbiome research. Technicians should minimise the amount of their own human and microbial DNA that can be potentially incorporated into a sample. Clean and protective clothing, including freshly washed clothes, disposable body suits or clean laboratory coats, face masks, shoe coverings, and multiple layers of gloves, could be worn to minimise transfer of the technician's DNA into samples.[13] The working environment, which should be a still-air dedicated clean room,[9,13] must also be extensively cleaned before samples are exposed to the laboratory environment. For example, laboratory surfaces, inside and outside of the hood, and all contents within the hood should be cleaned with bleach[13] or degrading detergents (*e.g.*, Decon®90) and ethanol before and after completing any laboratory processes. In addition, ultraviolet light exposure (30 mins or more) can

also be used to cross-link DNA and create thymine dimers, which prevents polymerase from reading those DNA templates.[22] It is also highly recommended that reagents should be aliquoted prior to sample exposure. This technique reduces the risk of contaminating samples from different DNA extraction batches, and samples within an extraction batch. These contamination mitigation strategies were used extensively in **Chapters III**, **IV** and **V** to minimise contamination in these studies. In addition to the aforementioned strategies, reagents can also be decontaminated via exposure to UV radiation[23] or through the inclusion of dsDNase to PCR master mixes,[24] but some research have found these methods to be compromise PCR efforts.[25,26]

2) *Collecting both positive and negative controls for all stages of laboratory processing*

For low microbial biomass samples, collecting and monitoring positive and negative controls are critical to distinguish between background DNA from the sampling and laboratory environment and biological samples. Negative controls, such as extraction blank controls (EBC; an extraction without biological sample added) and no-template amplification controls (NTC; an amplification reaction without an DNA extract added), are the most commonly employed controls currently used in microbiome research.[27] As noted by Hornung *et al.* (2019), negative controls should be collected from reagents and equipment and at every stage of processing between sample collection and sequencing[27] *(i.e.* all reagents and equipment used in sample collection, environment of the sample collection location, EBCs and NTCs for each extraction/amplification batch, and at least one sequencing negative control). In **Chapters III** and **V**, negative controls were collected during sampling and within each extraction amplification batch. However, only EBCs and NTCs were collected in **Chapter IV**. It is worth noting that negative sequencing controls were not included in these chapters but should be considered in future studies.

In addition to negative controls, positive controls should also be collected. Positive controls are just as important as negative controls in microbiome research.[17] These controls—mainly in the form of mock communities or a single culture species—are included to ensure that the extractions, amplifications, and sequencing has performed as expected and to identify the limit of detection and cross-contamination.[13,17] A positive control should be included in the extraction

process (and subsequently amplified before sequencing), and at least one positive control should be included in each sequencing run.[27] However, one area of concern for positive controls is that they can be magnitudes higher in biomass compared to low biomass samples, which can lead to cross-contamination. The effects from positive controls can be minimised by semi-isolation of positive controls (*e.g.*, physically locating the positive control further from the samples in a 96-well plate format).

It has also been proposed that journals should only accept articles that have included both negative and positive controls.[27] While negative controls should be incorporated in every low microbial biomass study, there is currently no best-practice for including positive controls for low microbial biomass samples without the low microbial biomass samples being affected by the high biomass positive controls. Eisenhofer *et al.* (2019) instead provides a description for contamination assessment with the 'RIDE' checklist, which may be more suitable for low microbial biomass samples in the current climate.[13] Importantly, this checklist should be adhered to and mediated by researchers, reviewers, and journal editors. Studies that do not include positive controls should not be dismissed from journals until new strategies for minimising cross-contamination (see '**Differences in microbial biomass across samples**' section for more information) between low microbial biomass samples and positive controls are implemented. Assessments between study findings and common contaminant taxa lists[4,13,28–31] may be a better alternative to a simple rejection for studies that do not include the correct controls.

3) Assess, identify, and remove contamination

Programs and bioinformatic analyses can be used to identify and remove contamination. The main techniques used to identify and subsequently remove contamination include the comparison of microbial communities through principal components plots[18] and the use of decontam[32] or SourceTracker.[33] Principal components plots can be used to identify similarities (or dissimilarities) between samples and controls.[18] However, this method for identifying differences between controls and biological samples is more practical for high microbial biomass samples and is not ideal for low biomass samples, as samples can be very similar to controls. Principal components plots can also be used to identify single biological samples that show high similarity to control samples. In **Chapter V**, for example, a principal components plot was used to identify one sample that had a similar

microbial composition to the controls, which was subsequently removed. In **Chapters III**, **IV**, and **V**, I used decontam, which is a prevalence-based method for identifying contaminants based on proportion of contaminant sequences in negative controls to biological samples.[32] Decontam (set at 0.5 frequency threshold) has been shown to remove ~70-90% of contaminant sequences from a mock community, without identifying species from the mock community as contaminants.[5] Another well-known and trusted program for identification of contaminants is SourceTracker.[33] SourceTracker uses a Bayesian approach to identify contaminants by using controls as the 'source' and biological samples as the 'sink'. SourceTracker was also tested by Karstens *et al.* (2019), and was also shown to be very effective at identifying contaminant sequences.[5] Both decontam and Sourcetracker are the most reliable methods to identify and quantify contaminants based on statistical methods, especially for low microbial biomass samples. For each study though, parameters within the selected program should be tested (*e.g.*, frequency threshold for decontam or the leave-one-out method for SourceTracker).

While a majority of these above strategies were applied in the data chapters of this thesis (**Chapters III**, **IV**, **and V**), some strategies were not considered, not available, or were not appropriate for these datasets. For example, positive controls were not as widespread when samples from this thesis were processed; thus were not applied to **Chapters III**, **IV**, and **V**. Although there is concern for high biomass mock communities (commonly used as positive controls) being extracted with low microbial biomass samples, positive controls would have been especially useful for **Chapter IV**. In **Chapter IV**, an in-solution silica extraction method—commonly employed in ancient DNA studies—was used, but it may not be as efficient as kit extractions. Serial dilutions of a mock community or single species (*e.g.*, similar to the Katharoseq pipeline[11]) may have aided in determining if the in-solution silica extraction method was suitable for oral and lung microbiota samples from **Chapter IV**. Unfortunately, negative controls were not collected during sampling of preterm infants (**Chapter IV**). Blank swabs, sterile tracheal aspirate tubes, and samples taken from the sampling room (*i.e.* air and surface swabs) would have provided more insights into potential contamination, as well as bacteria that may be transferred to preterm infants. Lastly, healthy full-term and adult samples that were available on QIITA data repository and incorporated into **Chapter IV** did not contain their own negative controls. To filter potential contaminants, I removed amplicon sequence variants (ASVs; 100% operational taxonomic units) that were

found in negative controls within the preterm extraction and amplification batches from the full-term and adult samples. This is not ideal and most likely resulted in removal of sequences that were not considered contaminants in the original study. For every study that is uploaded to an online data repository, I recommend uploading negative and positive controls with biological samples to account for contaminants found in separate laboratories. In future microbiome studies, it is crucial to include all controls possible and to monitor and remove controls, regardless of the sample's microbial biomass.

Differences in microbial biomass across samples

Variation in microbial biomass of samples within a study can be problematic for laboratory processing and bioinformatic analyses. It has been noted in the field that performance of extraction methods can vary depending on sample biomass, and that efficiency of extraction techniques should be considered for low microbial biomass samples. In 2019, Davis *et al.* showed that the QIAamp BiOstic Bacteremia kit with a modification—adaptable with a 96-well silica membrane for high-throughout—performed better (*i.e.* high DNA recovery and equal fastest hands on time) than other commonly used kits (FastDNA-96 Soil Microbe DNA kit, DNEASY POWERSOIL HTP kit, QIAamp BiOstic Bacteremia DNA Kit without modifications, and MagAttract PowerSoil DNA KF kit).[34] Another kit and workflow that can be used for low microbial biomass samples is KatharoSeq.[11] This workflow allows for microbial profile recovery of only 50-cell input, uses positive and negative controls to improve the identification of contaminants, and determines the limit of detection (lowest detectable threshold).[11] As part of this workflow, the Qiagen Ultraclean Pathogen kit was utilised.[11] Another method that can be used to improve amplification efficiency of low microbial biomass samples is to spike in DNA or RNA to increase the overall DNA content of a sample. As the carrier DNA is known, sequences can be subsequently removed in the bioinformatics stage.[13,35] Most importantly, it is vital to use the same extraction method throughout a study to reduce biases between methods, which was implemented in **Chapters III**, **IV**, and **V**. Overall, the extraction method should be well-considered for the study prior to commencement (*e.g.*, if there are differing biomasses in the one study, all low microbial biomass, and high-throughput), as changes to laboratory protocols through in the middle of a study can have implications in downstream analyses and the final results.

Cross-contamination—the exchange of DNA from one sample to another within an extraction or amplification batch—can be observed when samples with a wide range of biomasses are extracted in the same batch.[13,36] Traditionally, microbiome extractions were performed as single-tube protocols; however, with the expected large volume of samples that may be used in real-world applications for microbiome analysis, researchers are now turning to high-throughput methods, such as 96-well plate formats and automated robots.[11,34,36] Although single tube extractions take longer to process the same amount of samples, they are generally considered 'cleaner' and present a lower risk for cross-contamination across samples (providing that individual tubes are only open while the sample/solution is added). As mentioned above, it is preferred to extract and amplify samples in higher batches (*e.g.*, 96-well format) for the sake of efficiency, but this comes with a higher risk of cross-contamination.[36] Indeed, Minich *et al.* (2019) showed that cross-contamination was more prevalent in plate extractions compared to single-tube extractions, and that DNA could be exchanged up to 10 wells away.[36] To reduce cross-contamination in **Chapters III**, **IV**, and **V**, all samples were extracted in single-tubes, and samples with a similar biomass were extracted on the same day. Recently, new methods and techniques have been developed to identify and examine cross-contamination (*e.g.*, such as spike-in of synthetic DNA during extractions)[37]. While cross-contamination cannot be completely eradicated, it is important to consider these potential issues and use strategies to mitigate the effects.

Differing input in biomass across samples in a given study can also have downstream effects on bioinformatic analyses. Even if samples are added to a sequencing pool at equimolar concentrations, duplications, complexity and quality of the sequences can impact the number of sequences obtained post-sequencing.[38] To minimise any biases that may arise due to differing sequence depth, bioinformaticians generally rarefy (select an equal number of sequences for each sample) samples to the same threshold before analysing alpha and beta diversity metrics. For instance, rarefaction curves can first be used to identify diversity saturation (*i.e.* when no more new sequences are identified at a particular depth).[39] This saturation depth can then be used to subsample sequences and directly compare them to one another. For beta diversity, Weiss *et al.* (2017) also showed that rarefying sequence data consistently outperforms other subsampling methods, for common microbiome diversity metrics, such as Bray-Curtis, Jaccard, unweighted UniFrac, and weighted UniFrac.[39] While rarefaction is ideal for samples

with similar biomass, there are implications when rarefying samples in a study with varying biomasses. In **Chapter III**, biomass and correlated sequencing depth were observed (*i.e.* low microbial biomass and sequencing depth in pre-exposure nasal and skin samples compared to high microbial biomass and sequencing depth in environmental samples). As a result, a lower rarefied depth (*i.e.* 1,000 sequences) was chosen to retain as many samples as possible, but this compromised sequence diversity. To confirm that a similar pattern was still observed at a higher rarefied depth, diversity statistical tests were also measured with a higher rarefied depth of 5,000 sequences. Further research is still required to investigate and mitigate the effects for a range of biomasses in a given study, especially for the assessment of multiple body sites.

Stochasticity

Stochasticity can arise in different scenarios, including introduction of microbes into sterile environments, contamination influencing samples, or through laboratory and bioinformatic processes. Ecologically, initial colonization of microbes is random until selection acts to remove unwanted microbes, and the microbes may appear stochastic. Depending on the environment, stochasticity may be observed for longer periods of time and can potentially have an impact on health. In the most natural setting, a full-term infant born will first be exposed to the mother's vagina, the mother's skin (kangaroo care), and breast milk. All of these maternal and environmental factors contain specific microbes that are passed on from the mother and aid in selection of microbes for short-term and long-term health of the infant.[40] However, changes to this process can result in alternative microbes fighting for niches in the infant, and it may take longer for the mutually beneficial microbes to fully colonise the infant. This phenomenon was observed in the mouth of preterm infants (**Chapter IV**), whereby samples collected across the first week following birth had high variability between individuals. However, over time, ecological pressures (*e.g.*, competition between species) acted on the mouth and selected mutually beneficial microbes. As such, microbial diversity was reduced, and the community became more stabilised. Stochasticity from environmental exposures was also observed in **Chapter III**. Pre-exposure samples (*i.e.* skin and nasal samples collected before exposure to urban green spaces) were low in diversity, but post-exposure samples comprised a higher microbial diversity after green space exposure. Additionally, after exposure to urban green spaces in

Adelaide, it was evident that species were randomly transferred to the skin and nose, as no ASVs were transferred to all post-exposure samples after visiting each location. Indeed, Bateman (2017) demonstrated that rare taxa can remain persistent on the skin, but the majority of transferred microbes disappear (as they are out-competed by commensal taxa) after 48 hours.[41] Overall, this could suggest that exposure to maternal and environmental microbes of less diverse human niches is a stochastic process, and that these processes can be captured in low microbial biomass samples.

Apart from stochasticity occurring due to ecological selection, stochasticity can also arise as part of sample collection and processing and in bioinformatic analyses. During sampling, variation in sample collection may exist due to the volume of a sample or from technicians collecting samples in an inconsistent manner.[42] For example, there were differences in the amount of aspirate collected from preterm infants that were intubated (**Chapter IV**), which led to samples being collected over multiple tubes. Due to limited information on the method used to collect tracheal aspirate samples, each tube was treated as a separate sample, even if there were multiple samples from a single individual. As a result, variation between samples from the same individual (*e.g.*, microbial content may have been higher in one sample compared to another) may have resulted, although only presence/absence comparisons were used when performing analyses on these samples to reduce any bias. This potential bias may have been avoided if liquid was collected in a single tube, then centrifuged to create a pellet of microbial cells. The pellet could have then been resuspended in a small amount of solution, which would be equal across samples, and then added to the extraction. However, this technique would need to be verified. In addition to volumes, the sampling area and pressure applied to the sampling area for the collection of skin samples can potentially lead to variation in the microbiota profile.[43] To reduce stochasticity for skin swabs collected in **Chapter III**, it was ensured that subjects sampled in the same location, for the same duration, and applied the same pressure. Importantly, subjects may have a slightly different technique, which could explain the difference in microbial diversity between individuals.

Random processes also occur through laboratory processes (extractions and amplification). Technicians/automated robots, batches, seasonality, and volumes can create stochasticity in low microbial biomass samples during extractions and amplification preparations. In some microbiome studies, technicians/automated

robots and extraction batch has been shown to have an effect on the microbial composition.[44] To avoid batch effects and technician bias in this thesis, samples processed in **Chapters IV** and **V** were processed by a single technician (*i.e.* myself). Although these precautions were taken, extraction date was still significantly contributing to the oral microbial composition in preterm infants (**Chapter IV**), which may be explained by a higher proportion of T2 (samples collected at the second timepoint) in an extraction batch. While I tried to reduce technician biases, this could not be avoided in **Chapter III**, as soils could not enter Australia unless they were already extracted. Unfortunately, samples processed in India may have been processed under less strict conditions and were significantly contaminated by Streptophyta (plant) sequences, which may have been a result of processing under conditions less strict to samples processed by myself. Therefore, these samples could not be included in the study. Especially for low microbial biomass samples, batch effects and contamination from kits and seasonality can have a significant impact on samples. For example, longitudinal studies can be impacted by variations across kits and reagents batches, if kits are purchased as required.[3] It is therefore recommend to purchase all kits at the same time to reduce variation and minimise batch effects across samples.[3] Furthermore, it has been shown that seasonality can have an influence on samples (**Appendix IV**).[31] To minimise these effects in this thesis, extraction kits were purchased within the same timeframe to minimise variance between kit or reagent batches, and in most cases, samples were also processed within the same timeframe and season to reduce random microbes being incorporated in the sample.

Stochasticity may also arise from the volume of sample taken at each laboratory step. Volumes can affect the samples in two potential manners: 1) a fraction of sample volume is acquired for each step, and 2) standard DNA elution volumes (*e.g.*, 100uL) can dilute low microbial biomass samples severely. In the first case, some of the solution collected at each stage in extraction kits is left behind, due to the max volume used in the following step. In addition, only a small fraction of DNA eluant (3uL of 100uL) is transferred into the amplification stage, and a small proportion of the amplified library is included in the sequencing pool (1-20uL of 75uL). To ensure that these small volumes of DNA from the sample is representative of the sample, the solution is pipette-mixed before being added to the next stage. For the second case, DNA for each sample is normally eluted in 100uL of elution buffer, which can severely dilute DNA from low microbial biomass samples,

potentially resulting in a loss of rare taxa. To reduce this effect, low biomass samples could be eluted in lower amounts of elution buffer (TLE or water) to concentrate the sample and completely represent microbial taxa.

In the bioinformatic process, sequences are subsampled to compare samples at the same sequence depth. Low microbial biomass samples normally have a low sequencing depth and a high number of duplicates[38] (*i.e.* even if the sample was sequenced deeper, no to very little new unique sequences will be identified). In these type of samples, contaminant sequences comprise a large proportion of sequences in the samples (approximately 70-80% of sequences).[5] Post-contaminant filtering, few sequences are generally retained in low microbial biomass samples, which can make it difficult to compare samples through diversity metrics. Most often, samples are usually rarefied for diversity comparisons in microbiome studies (see '**Differences in microbial biomass across samples**'). Some rarefaction methods randomly subsample sequences for a given sample (usually through a resampling/permutation process), which may have an effect on low biomass samples. To avoid this potential issue, researchers should run a high number of permutations for statistical tests to avoid any biases (*i.e.* 999 permutations or more). Once challenges with low microbial biomass samples are appreciated, mitigated, and controlled for, high quality information can be extracted from low microbial samples.

Low microbial biomass samples provide a wealth of knowledge for disease

Although working with low microbial biomass samples presents difficulties and challenges, the potential resulting wealth of knowledge for health and disease makes the effort well worthwhile. In this thesis, hundreds of samples across numerous body sites were analysed in the context of health, including nasal and skin samples analysed to provide insights on green space exposure and immune-mediated diseases (**Chapter III**); oral and lung microbiota used to track microbiota development, and indicators of bronchopulmonary dysplasia (BPD) and sepsis (**Chapter IV**); and, gingival swabs used to investigate the effects of periodontal diseases and a family history of hyperlipidaemia on children with type 1 diabetes (T1D; **Chapter V**). Other researchers have also investigated other low microbial

biomass sites of the human body, revealing additional information regarding NCDs that I discuss below.

Skin is the biggest organ of the human body, is the first point of contact with the environment, and has a range of different physiological sites for different microbiota to reside. The skin's main function is to create a barrier for the body, and skin microbiota aid in this process. Skin microbiota are also tightly linked to the immune system to respond appropriately to the external environment.[45] In **Chapter III**, skin and nasal microbial diversity increased after environmental exposure, whereby higher human microbial diversity is thought to reduce immune-mediated diseases. However, disruptions to the skin microbiota have been linked to NCDs. For example, lower diversity[46] and an enrichment of *Staphylococcus spp*. and opportunistic fungi have been associated with atopic dermatitis severity and flare-ups.[47] It has also been proposed that psoriasis is associated with disruption to the skin microbiota.[48] Although low in biomass, the skin microbiota still provides a wealth of knowledge for NCDs, and microbiota associations to other skin diseases such as gout, warts, or chicken pox (Varicella) could be investigated in the future.

The oral microbiota is arguably the second most studied site of the human body following the gut and plays key roles in oral homeostasis. Disruptions to the oral microbiota can result in NCDs. In this thesis, I investigated the oral microbiota of two different cohorts: preterm infants (**Chapter IV**) and children with T1D (**Chapter V**). In preterm infants, microbiota diversity and composition was distinct from full-term infants and may be linked to immune-mediated diseases later in life (**Chapter IV**). Further, alterations to microbial diversity were linked to BPD development within a week of birth, while microbial composition was linked to the development of sepsis approximately two months after birth (**Chapter IV**). In **Chapter V**, differences in oral microbiota composition and diversity was observed in children that had a first-degree relative with hyperlipidaemia. However, these associations were not identified if the child had periodontal disease. In addition to these studies, the oral microbiome has been extensively studied for periodontal disease and dental caries (tooth decay). Periodontal disease can range from mild inflammation of the gums (gingivitis) to gum recession and tooth loss (periodontitis). In individuals with periodontal disease, *Porphyromonas gingivalis* and other Gram-negative bacteria increase in abundance and create large plaque structures that contribute to the lesions of the gums.[49,50] However, more information on the relationship between periodontal disease and other systemic

diseases is still required (see '**Interconnected systems and microbiotas of the human body reveal more information for NCDs**'). Over-abundance of particular bacteria, such as *Streptococcus mutans*, have been associated with the development of caries.[51] Most research conducted on dental caries have focussed on *S. mutans*, but the underlying cause for caries is clearly polymicrobial and multifactorial, *i.e.* diet, fluoride, *etc.* contribute to caries development.[51] In addition, some research has begun on the relationship between oral microbiota and oral cancer; this area is still in its infancy and requires more investigation.[52]

While initially thought to be sterile, it is now appreciated that the lungs are colonised with a low biomass of microbes that are important for health. In **Chapter IV**, I investigated the first colonisers of the lungs in preterm infants. Interestingly, some microbes that were observed in the lungs were also found in the mouth, alluding to a potential route for colonisation. Additionally, it was also noted that potential pathogens were observed in the lungs of infants that developed BPD and/or sepsis. Further investigation is required with a larger sample size to determine if these findings are consistent with the general infant population. The lung microbiota has also been examined for other NCDs, including chronic obstructive pulmonary disease (COPD) and cystic fibrosis. Increased abundance of *Moraxella* and *Haemophilus* are thought to play a role in prevalence of COPD, as well as COPD exacerbations. However, the contribution of each of these taxa for COPD are not well understood.[53] The contribution of *Pseudomonas aeruginosa* has been comprehensively studied in cystic fibrosis (CF) patients through culture-based methods, but the disruption to the lung microbiota through non-culture methods has only been investigated in a limited number of studies. For example, Cuthbertson *et al.* (2020) showed that microbial diversity decreased with disease severity and common CF pathogens increased in dominance (*i.e. P. aeruginosa*, *Staphylococcus* aureus, *Burkholderia cepacia*, and *Stenotrophomonas maltophilia*) with severity.[54] Although less research has been performed on the lung microbiome compared to its gut and oral counterparts, it will be important to further understand the lung's contribution to other local and systemic diseases, such as rheumatoid arthritis.[55]

Other low microbial biomass sites have also been investigated in microbiome analysis. However, microbiota studies on sites including the urinary tract, vagina, and eye have been focussed on infectious diseases, with little to no research being conducted on NCDs. For instance, the microbiota of the urogenital system have been examined in light of bacterial vaginosis, sexually transmitted diseases,[56] and urinary

tract infections.[57] However, it is just as important to look at other complications of the urogenital system that may be affected by the microbiome, including infertility, preterm birth,[58] and polycystic ovary syndrome.[59] Similarly, the eye microbiota has only been examined for infectious diseases, such as conjunctivitis.[60] Nevertheless, it is possible that the eye microbiota may play a role in eye NCDs with an unknown cause, such as glaucoma and pseudo-exfoliation syndrome. Overall, it is important to consider alterations to the microbiome for both infectious and non-infectious diseases (NCDs), as well as the potential influence of an infectious disease leading to an NCD, or vice versa.

While it is very fascinating to investigate each of these sites, it is still crucial to apply strategies to reduce contamination and other biases, as mentioned in **Chapter I** and in the '**Challenges of low biomass samples and considerations for future research**' section. Each of these examples above demonstrate that low microbial biomass samples hold a wealth of knowledge, and that a greater understanding of NCDs can be obtained using these samples.

Interconnected systems and microbiotas of the human body can reveal more information on NCDs

Many factors contribute to health and disease, including the environment, lifestyle factors, genetics, and the microbiome. However, these factors are often considered independently when studying health and disease. In addition to this, local effects of the microbiome are mostly considered. In recent years, some studies, including **Chapters III**, **IV**, and **V**, consider the interaction between distinct, seemingly independent body sites. These interactions can occur through many different systems of the human body including the immune, nervous, integumentary (skin), respiratory, and circulatory systems. In most microbiome studies, the focus on interactions across the human body have been centred on the gut, but in this thesis, I looked at the interaction across the skin, nasal, oral, and lung microbiota, and considered potential impacts of the gut.

The connection between the gastrointestinal tract and brain (commonly referred to as the gut-brain axis) has been one of the most fascinating links in the human body. The gut-brain axis is a bidirectional communication network comprised of nervous systems (*i.e.* central, autonomic, and enteric nervous systems)

and the hypothalamic-pituitary-adrenal axis, which has been linked to mental health and degenerative diseases.[61] While not fully elucidated, short chain fatty acids (SCFAs) such as acetate, butyrate, and propionate have been shown to play a role in these types of diseases.[62] For example, anti-depressant like effects from a mixture of SCFAs can reduce chronic psychosocial stress in mice,[63] and SCFAs have been associated to underlying neuropathological processes in Alzheimer's disease[64]. In addition, gut microbes transported from patients with Schizophrenia or Parkinson's Disease into mouse models resulted in mice which displayed differences in behavioural phenotypes[65] and altered motor function[66], respectively. Overall, this demonstrates that alterations to gut microbiota can have profound implications on the brain.

Alterations of gut microbiota have also been implicated in skin diseases, including atopic dermatitis and acne. Atopic dermatitis is thought to be impacted by the gut across three different pathways: neuroendocrine, immunologic, and metabolite pathways.[67] Tryptophan, which is a neurotransmitter precursor and is produced in the gut, is thought to play a role in itchiness in the skin—a common symptom of atopic dermatitis. However, products formed by particular *Lactobacillus* and *Bifidobacterium* species are known to reduce itchiness of the skin.[67] Additionally, low microbial diversity and disruption of the gut microbiota during immune development in infancy has also been observed to contribute to atopic dermatitis.[68,69] Lastly, diet and metabolites may also impact on the skin through inflammation. As the typical Western-diet has a low fibre content, less SCFAs—metabolites with anti-inflammatory properties—are produced in the gut, leading to increased local and systemic inflammation.[70] This change in diet may explain the increases in atopic dermatitis and allergies. Acne is also a common skin condition that is thought to related to the carbohydrate-rich Western diet.[71] Increases in carbohydrates through the diet can increase glycaemic concentrations. This can then activate the mTORC1 pathway and result in the increased production of fats in the sebaceous glands, leading to alterations of the skin microbiota, causing acne.[71,72] Interestingly, symptoms of both atopic dermatitis[73] and acne[74] have been improved by the probiotic supplements. Again, this suggests that the gut is likely to play a role in skin diseases and has systemic effects on other body sites.

The nasal and skin microbiota may respond similarly to environmental exposures (**Chapter III**), but nasal microbiota may have a pathogenic effect on other locations of the body, such as the foot. *Staphylococcus aureus* is a commensal

of the nasal passage in ~32% of individuals.[75] However, *S. aureus* from the nasal passage has also been implicated as a risk factor for nosocomial and surgery infections. For example, colonisation of *S. aureus* is significantly associated with diabetic foot infections, but only if the patient has nasal colonisation of *S. aureus* (<20% of patients).[76] However, it is still not well-understood how other infections arise in patients with diabetic foot infections or ulcers. Future studies could investigate if small colony variants of *S. aureus* can be detected through microbiome analyses, as their morphology (*i.e.* smaller and thicker cell wall compared to wildtype)[77] may go undetected with current microbiome techniques.

As part of the respiratory tract, the oral and lung microbiome are physically close but have unique physiologies and microbiomes. Nonetheless, microbes are shared between the mouth and the lungs,[78] which is even evident in preterm infants (**Chapter IV**). Microbes of the mouth under certain conditions have been associated with lung infections and other chronic lung diseases. For example, lung pathogens have been detected in plaque of individuals that are in intensive care for pneumonia.[79] However, it is unknown if the lung pathogens colonised the mouth or if these pathogens originated in the mouth and then contributed to the lung infection. While initial studies have been conducted on connections between oral microbiota and lungs diseases for COPD[80,81] and cystic fibrosis,[82] more studies are needed to determine if the oral microbiota has negative health outcomes on the lung microbiota, or if oral commensals are just shared with lung microbiota.

Poor diet and disruptions to the gut may influence the oral microbiota through the circulatory and immune systems. High fat diets and subsequent high blood lipids have connected to periodontal disease.[83] In this thesis, **Chapter V** demonstrated that familial history of hyperlipidaemia—which is thought to also affect the descendants through genetics and the environment—reduced oral microbial diversity and had a distinct microbial composition in children with T1D. In addition, periodontal disease and familial history of hyperlipidaemia were dependently contributing to the oral microbiota. This systematic association between the periodontal disease, hyperlipidaemia, and T1D may be exacerbating disease outcomes, as both circulating lipids and a higher immune profile from hyperlipidaemia and T1D are attributable to periodontal disease. Indeed, links between the gut microbiota and periodontitis have been demonstrated in mouse models.[84] Lastly, *Fusobacterium nucleatum*—a common oral species—is a likely contributor to carcinogenesis not only in the mouth but in other body sites. For

example, *F. nucleatum* have been associated with colorectal cancer, although the mechanism is unknown.[85] Altogether, these associations demonstrate that human microbiomes are connected and that more research should incorporate and consider systemic influences on human microbiomes.

Future approaches to understanding connections of the microbiomes

While understanding the interaction between two or more microbial sites are important to improve our understanding of NCDs, a 'whole-system' and functional approach is essential to work towards a complete picture of NCD predisposition, development, and persistence. In **Chapter II**, I suggest many avenues to understand systemic relationships between microbiomes, which may provide more insights into NCDs. In this section, I will expand on two of those ideas: microbiome-wide association studies and multi-omics approaches.

Microbiome-wide association studies

Microbiome-wide association studies (MWAS) are analogues to genome-wide association studies and investigate the relationship between microbial species and functions, host genomic single-nucleotide polymorphisms (SNPs), and host traits.[86] So far, MWAS has been primarily utilised to correlate microbial species that are inherited.[87–90] Recently though, a newly formed consortium, MiBioGen, have obtained 16S rRNA sequencing data with human genomic SNP information for over 19,000 individuals across 18 cohorts.[91] Their goals are to explore the associations between human diseases, human SNPs, and the human gut microbiome, as well as explore human gene-environment interaction in light of gut microbiome composition.[91] While this project contributes a significant effort to understanding key questions surrounding the relationship between the host and microbiome, other studies are still required to understand these relationships in sites other than the gut.

Multi-omics approaches

A more informative and powerful approach to understanding the initiation and progression of NCDs is to perform multiple 'omics analyses. Numerous 'omics-based approaches, such as genomics, epigenomics, transcriptomics, proteomics,

metabolomics, and metagenomics have advanced our way of thinking about the human body and disease. Incorporating the microbiome with other multi-omics approaches is a relatively new field, and very little research has been conducted to understand the relationship between these components of the body and disease. Nevertheless, metagenomics and transcriptomics have been used together to identity the relationship between bacterial species and colorectal cancer.[92] In addition, the combination of metagenomics and metabolomics have revealed associations of disturbances to both circulating metabolites and airway microbiota for children with asthma.[93] As this area of research is new, there are currently many limitations including cost to generates multi-omics data, computing resources to analyse large datasets, and statistical methods that are able to compare multiple datatypes at once. Moving forward, multi-omics approaches will be a valuable tool to understand the interaction of multiple components of the human body and NCDs.

The human microbiota provides more insights into NCDs

The microbiome is not the only component that contributes to NCDs, but it is one of the most understudied components to NCDs. As such, it is still unknown to what extent disruptions to the immune system or microbiota, either early or later in life, lead to NCDs. As we understand more about the impacts of the microbiome on NCDs, we can scrutinise microbiota predisposition and development of NCDs and determine subsequent diagnosis and treatments, which will be discussed below.

Development of the microbiota and immune systems early in life

Disruptions to microbiota at birth, during the neonatal window of opportunity, and through immune system development can have implications for long-term health outcomes. Unnatural procedures and exposures, including Caesarean-section births,[94] formula milk feeding,[94] and antibiotics,[95] have each been shown to impact the microbiota and may also contribute to NCDs, such as asthma and allergies, that are observed later in life. For example, infants born via C-section have an increased abundance of *Enterococcus* and *Klebsiella* (potential pathogens) and decreased abundance of *Bifidobacterium* (beneficial bacteria),[94] as well as altered immune profiles. C-section disruptions to the microbiota and

immune system are associated with respiratory infections.[94] Exclusive formula feeding at approximately 3-months of age can alter gut microbiota and increase the risk for overweightness in infants at 12-months old.[96] However, long-term studies are needed to determine if formula feeding can have long-term implications on obesity. Additionally, antibiotic administration at an early age has been shown to delay colonisation of Bacteroidetes.[95] Although gut microbiota disruptions can recover from antibiotics, metabolic changes can persist and result in a higher risk for developing obesity.[97] Indeed, most of this research has been biased toward gut microbiota, but, as mentioned in the previous section, it is best to consider the body as a whole entity. To fully understand the extent to which alterations to the microbiota have on health, more extensive studies investigating other body sites are required.

In **Chapter IV**, I investigated oral microbiota in preterm and full-term infants at analogous time points after birth. Through this study, I identified high variation in preterm infant oral microbial diversity of samples collected within the first week of birth. This variability due to the disruption of the oral microbiota was restored approximately two months later. Interestingly, Olin *et al.* (2018) demonstrated that the immune system of preterm and full-term infants initially respond differently soon after birth but became similar after three months.[98] In both of these cases, long-term health data was unavailable, and it is unknown if the alterations to oral microbiota and immune profiles will have an impact on the development of NCDs later in life. Indeed, one longitudinal study has shown that gut microbiota alterations early in life can contribute to islet autoimmunity and type 1 diabetes.[99] At present, these studies are correlative. To truly understand whether microbial alterations early in life have long-term health implications, it is important to consider the interaction between the human microbiome and immune system, as well as potential irreversible changes, to either of these components. Future studies should collect both microbiota and immune profile information and medical records to elucidate if the microbiome and immune system in early life contribute to long-term health outcomes.

Lifestyle contributes to NCDs and the microbiota

Common lifestyle factors, including the environment, diet, smoking, medications, and stress, contribute to NCDs and also affect the human microbiome.

It is important to understand the mechanisms by which lifestyle factors contribute to NCDs. In this section, I will discuss the impact of lifestyle factors on the microbiome, and how they individually or collectively contribute to NCD prevalence.

Environment

Reduced or minimal exposure to biodiverse environments has been linked to allergies and asthma. Numerous hypotheses (Hygiene,[100] Old Friends,[101] and Biodiversity[102] hypotheses) have suggested a link between increases in inflammatory-mediated diseases (allergies and asthma) and environmental microbial exposure; however, very few studies have experimentally investigated these links. Initial studies have shown that microbial endotoxin found in the environment, such as indoor house dust, can mediate gut microbiota and reduce incidence of asthma and allergic sensitisation.[103,104] Further, the diversity of the soil has an impact on the faecal microbiota of mouse models, *i.e.* compositional difference are observed in faecal microbiota of mice exposed to low vs. high biodiverse soils.[105] Additionally, high biodiverse soils were found to have butyrate-producing bacteria, which may reduce anxiety in mice.[105] **Chapter III** also demonstrated changes to the microbiota after exposure to the environment. Increased microbial diversity and altered microbial composition was observed on the skin and in the nasal cavity after urban green space exposure. While increases in microbial diversity were observed in this study, it was also noted that diversity was reduced by the following morning (*i.e.* post-shower). As such, more research is needed to determine how much time individuals need to spend outside and whether individuals need to spend time in biodiverse environments every day to see positive health outcomes. Importantly, it is essential to determine if exposure to microbially-rich environments will have a greater reduction in developing immune-mediated diseases if exposure occurs during immune development (*i.e.* up to three years of age), or if positive health outcomes can also be seen in adults after this type of exposure.

Diet

A poor diet (*e.g.*, excessive consumption of fats, salt, and sugar) is a well-known contributor to NCDs, such as obesity, type 2 diabetes, cardiovascular disease, and hyperlipidaemia.[106] The types of macronutrients consumed can also influence

the microbial communities, especially in the gut. For example, both fibre and sugar are carbohydrates, but the gut microbiota responds differently to these carbohydrates. Indeed a plant-rich and high-fibre diet has been associated with bacteria that are SCFA fermenters,[44] which can have anti-inflammatory effects on the body.[107] In contrast, excess sugars (*e.g.*, glucose or fructose) can negatively alter the gut microbiota, leading to a higher abundance of Proteobacteria,[108] which are generally associated with inflammation.[109] Similarly, excess fats can also lead to a greater proportion of Proteobacteria and are also thought to contribute to a pro-inflammatory state. Disruption to the gut microbiota through high sugar and fat consumption can cause metabolic dysregulation, potentially inducing diseases such as hyperlipidaemia, type 2 diabetes, and cardiovascular diseases. However, both high sugar and fat also have negative effects on other body sites, such as the mouth. Foods and drinks high in sugar can have devastating outcomes for oral health. For example, there is a strong relationship between total added sugar and number of tooth surface affected by dental caries in children.[110] The microbial digestion of sugar in the mouth is known to lower the pH, creating a more acidic environment. This type of environment causes tooth demineralisation and leads to tooth decay.[111] As stated earlier, excessive fats may also contribute to oral disease, such as periodontal disease, through metabolic dysfunction and circulating inflammatory cytokines. Such a state was possibly observed in children with type I diabetes and a family history of hyperlipidaemia (**Chapter V**). However, further studies are required to determine if alterations in lipid profile of the children themselves could lead a pro-inflammatory state and eventually contribute to periodontal disease. Overall, diet can influence the human microbiota and contribute to NCDs, but further research is needed to indicate whether dependent associations between microbiota with the immune system are the link between diet and NCDs or if microbiota and the immune system are independently contributing to NCDs.

Smoking

Although it is well-known that tobacco smoking leads to numerous NCDs, including cancers, cardiovascular disease, respiratory diseases, *etc.,* approximately 20% of individuals across the globe smoke tobacco.[112] Tobacco smoke can contain addictive and dangerous chemicals, such as nicotine, tar, carbon monoxide, metals, and many more. In addition, tobacco smoke can alter community composition of microbiotas across the human body, including the mouth, lungs and gut. For

instance, tobacco smoking can alter the oral microbiota, shifting the community to a pathogen-rich state and may contribute to the high proportion of smokers that develop periodontal disease.[113] Traditional tobacco smoking is not the only type of smoking that can affect the mouth. Electronic-cigarettes—a nicotine-based liquid that is vaporised and inhaled—are a new form of smoking and the long-term health consequences have not been explored. One study, however, found that the oral microbiota of electronic-cigarettes smokers was distinct from tobacco cigarette smokers and non-smokers, likely due to the glycerol content in the electronic-cigarette nicotine solution.[114] Tobacco smoking can also cause physical injury and microbial disruption to the lungs. While physical lung injury has been more widely studied, the association of altered microbial communities have not been studied to the same extent. Initial studies have identified potential association between disruption to the microbial community and COPD[115] and acute respiratory stress syndrome,[116] but associations with other lung diseases, including pneumonia, emphysema, lung cancer, and bronchitis, are still yet to be explored. Lastly, Crohn's disease, ulcerative colitis, and colorectal cancer have each been linked to changes in the gut microbiota, as a result of smoking.[117] Even though smoking is one of the biggest risk factors for developing many different NCDs, the connections of the microbiota with smoking is severely understudied. With the introductions of many harmful chemicals introduced to the body from smoking, it would be unsurprising if alterations to the microbiota increase systemic inflammation and contribute to NCDs.

Medical treatments

Medicines are substances that are taken to relieve symptoms or cure a disease. However, some medications can have side-effects that contribute to NCDs.[118] For example, oral corticosteroids can contribute to high blood pressure and cataracts.[119] Some medical treatments can also influence the gut microbiome. Disruptions to gut microbiota from administration of antibiotics have been extensively studies. For instance, antibiotics administered during infancy can disrupt the gut microbiota, which can increase the risk for developing NCDs, such as asthma, allergies, and eczema.[120] In most cases, adult gut microbiota return to baseline ~1.5 months after antibiotic administration, but some taxa can remain missing even after this period of time.[98] Consequences of medications on the human

microbiome have been understudied. More research is required to understand if the microbiome contributes to medication side-effects.

Stress

Although stressors can be different in nature (physical, psychological, environmental, *etc.*), they elicit the same stress response.[121] Stress has been linked to NCDs, including cardiovascular diseases, metabolic diseases, cancers, gastrointestinal diseases, and neuropsychiatric diseases.[122] Recently, it has been appreciated that stress can also alter gut microbiota—through the gut-brain axis—which then contributes to NCD development.[121] For instance, psychological stress can disrupt gut microbiota and contribute to inflammatory bowel disease (IBD) flare-ups. Furthermore, it has also been suspected that psychological stress can contribute to the exacerbation of skin diseases, through gut disturbances.[70] While the association between stress and disruptions to the human microbiome have not been extensively studied, there is evidence to suggest a relationship between the two.

Using characteristics of the human microbiome to track NCD risk and progression

Environmental and behavioural factors are important to identify and track NCD risk and progression, which could be traced through regular clinical sampling and interviews. However, before this can be achieved, longitudinal studies must be used to identify consistent changes to the microbiome that are associated with NCD risk or initiation.

Within this thesis, two studies (**Chapter III** and **IV**) tracked changes of the microbiome over time. In **Chapter III**, skin and nasal microbiota were tracked over a three-week period. In this time frame, subjects were exposed to urban green spaces across three countries. While the major finding showed increased diversity after urban green space exposure, there were still differences in microbial diversity and composition in skin and nasal samples across the different countries. Longer tracking of individuals travelling abroad would provide more information if these microbial differences were due to exposure to different urban green spaces or if these changes are a consequence of travelling abroad. Interestingly, exposure to

urban green space did show an increase in nasal and skin microbiota in most countries, suggesting that this kind of exposure can increase microbial diversity, and have potential positive outcomes for NCDs.

Chapter IV tracked the oral microbiota of preterm and full-term infants over a two-month period. Over this period, the preterm infant oral microbiota went through dramatic changes. In comparison to a steady increase in microbial diversity over time for healthy full-term infants, an initial disruption was observed in preterm infants, but was later restored to resemble the oral microbiota of full-term infants. In addition, a potentially higher-pathogen load was identified in preterm infants within the first week of birth, with *Ureaplasma*, *Streptococcus anginosus*, and *Streptococcus agalactiae* ASVs found in both the mouth and lungs of children with BPD or sepsis. Although, tracheal aspirate samples were not collected at the second time point (~2 months later as intubation was not required at that time), a lower abundance of these potential pathogens were observed in the oral microbiota at this second time point. Identification of when the initial microbiota disruption stabilised and when the potential pathogen-load decreased was limited by the study design. More frequent (*e.g.*, daily, weekly, or monthly) sampling would be required to identify when these changes occurred.

Although there has only been a small number of longitudinal microbiome studies, some researchers have studied humans over long periods of time to reach a more informative view of microbial development and how lifestyle changes may contribute to NCDs. For example, the oral microbiota was tracked from birth to five-years of age and found that diversity increased with age, but no tracking of disease was pursued in this study.[123] The Canadian Healthy Infant Longitudinal Development (CHILD) cohort has be extensively researched to identify key factors that can alter the gut microbiota and result in NCDs. Birth mode,[124] feeding method,[96] vitamin D supplementation,[125] and timing of introduction to allergenic foods[126] were shown to increase risk for particular NCDs, including obesity, asthma, and allergies. Indeed, there are fewer studies that investigate the microbiome in relation to lifestyle factors and NCDs. In most cases, "healthy" baselines are currently being identified through longitudinal studies for the gut,[127] mouth,[127] vaginal,[128] and skin.[129] Nevertheless, some longitudinal studies have identified: differences in gut microbiota for IBD patients that are undergoing disease activity, as well as[130,131] US immigrants having reduced gut microbial diversity and presenting a higher risk for obesity, which is compounded over generations.[132]

Overall, longitudinal studies have the power to detect dynamic changes in relation to health and disease, rather than a snapshot of the microbiota for an individual on a particular day.

Longitudinal sampling can be used to detect microbiome changes and biological markers (biomarkers) in a clinical setting to predict disease risk and progression. Biomarkers are indicators that can aid in identification of distinct alterations to microbial communities—diversity and compositional changes or an outgrowth of a particular species—and may indicate NCD risk and progression. For example, low diversity in the gut can be an indicator for IBD. Distinct increases in certain bacteria (*P. gingivalis*, *Tannerella forsythia*, and *Treponema denticola*) can be an indication of periodontal disease.[133] Additionally, in **Chapter V** children with hyperlipidaemia had increased abundance of a *Prevotella* ASV, which may also be an indication of periodontal disease progression. Lastly, as potential pathogens were identified in both the mouth and lungs of infants with BPD or sepsis, the oral microbiota could be used as a proxy for pathogen detection in young children (**Chapter IV**). As these studies show distinct changes to the microbiome composition, they could possibly be used as indicators for disease risk and progression in the future. However, it is important to note that these biomarkers must be detected in other studies and undergo clinical examinations before they can be used as a diagnostic test.

Potential novel avenues for diagnosis and treatment of NCDs

With the use of longitudinal collection of samples and biomarkers, I propose methods whereby the microbiome can be incorporated into pathology testing for more accurate diagnoses and the administration of targeted treatments (**Chapter I**). Microbiome techniques can provide more information for both infectious diseases with unknown aetiologies, as well as for NCDs. Detection of microbiota alterations and associated functional dysfunction is one of many benefits for creating diagnostic tests for NCDs within a pathology setting (**Chapter I**). For example, routine longitudinal sampling—say, as part of a doctor's visit—could be used to identify sudden disruptions to diversity or compositional changes to the gut. These samples could then indicate that specific treatments, such as a faecal transplant or probiotics, could be used to remedy the gut microbiota and minimise the chance of developing a serious and potentially chronic disease. Additionally,

biomarkers could be used in a similar fashion. For instance, oral swabs could be collected to identify shifts in microbial communities (*e.g.*, oral microbiota showing signs of periodontitis before inflammation is present) or an increased abundance of opportunistic pathogens. In this case, oral swabs could be collected as part of routine dental check-ups. Additionally, information from multiple 'omics technologies (see **Chapter II**) could be collected simultaneously for patients with an unknown aetiology. Using this strategy, more detailed and specific information can be extracted and could improve diagnosis, and subsequently treat patients quicker.

Certain NCDs may also be remedied through microbiome-based treatments. Probiotics and faecal gut transplants can be employed to correct microbiome imbalance and treat disease. Probiotic treatments have been shown to be successful in infectious diseases, such as necrotizing enterocolitis, diarrhoea, and respiratory tract infections. However, recent studies have shown promising results for alleviating symptoms of certain NCDs with probiotics. Allergic diseases including asthma, allergic rhinitis, atopic dermatitis, and food allergies[73,134] have each shown to benefit from probiotics. In individuals with peanut allergy, for example, *Lactobacillus rhamnosus* in conjunction with oral immunotherapy resulted in 82% of individuals becoming unresponsive to the peanut challenge (*i.e.* no longer reacted to oral injection of peanuts).[135]

Faecal microbiome transplants (FMTs) are another microbiome-based treatment that be used to restore gut dysbiosis. FMT is the process of collecting stool from a healthy donor and transplanting it to the colon of another individual. This type of treatment has been relatively successful in patients with reoccurring *Clostridium difficile* infections.[136] However, FMT success rate is much lower in IBD patients.[137] In addition, the delivery of microbiome transplants (*e.g.*, oral pill, colonoscopic or nasogastric delivery) should also be highly considered, as effectiveness varies across individual.[138] Microbiome transplants for oral diseases (transfer of saliva from a healthy individual to another individual with an oral disease) are currently in the conceptual stage but may be another potential microbiome-based therapy.[139] Importantly, an improved understanding of the complexity and mechanisms of NCDs is required to ensure microbiome transplants are safe and effective before being routinely applied in patients suffering from NCDs.

Future directions for a better understanding of NCDs with the addition of low microbial biomass samples

The importance and contribution of the human microbiome in NCDs is becoming more recognised. As microbiome research is in its infancy, there remains numerous considerations for obtaining robust, comparable, and informative datasets. By utilising larger studies to examine local and systematic microbial links between all body sites, a better understanding of the microbiome contribution to NCDs will be obtained. In this section, I will broaden the discussion and suggest four main considerations for future microbiome research in relation to NCDs: 1) assess multiple microbiotas of the human body; 2) obtain functional information of the human microbiome; 3) collect extensive and complete metadata; and 4) crowdsource low microbial biomass samples.

1) Assess multiple microbiotas of the human body

As this thesis suggests, it is important to consider the connection between multiple microbiotas of the human body, and their indirect effects on each other. By assessing multiple sites, key questions can be answered, such as, is there cross-talk between microbiota of other sites? Is there a system that is connecting two or more microbiotas that can have an influence on disease? Can commensals of one body site become a pathogen of another, and how does it get there? Should these questions be answered, they may shed light on NCD initiation and prevalence.

2) Obtain functional information of the human microbiome

In this thesis, community composition was characterised; however, it is more powerful to understand the functions of the microbiota. Functional information can provide insights into what function a microbe within a community is performing, as well as the overall function of the whole community itself.

Both 16S rRNA gene sequencing and shotgun metagenomic sequencing can provide functional information, but the latter is more informative. In most cases for 16S rRNA sequencing, only one hypervariable region of a gene is sequenced and matched to a database. Using a functionally informative database, such as KEGG orthologs, functions of a microbiota can be predicted based on the main function performed by taxa in that sample. However, these results mirror taxonomic

information; so, the identification of microbiota through 16S rRNA sequencing, is predominantly useful if there are abundance differences for particular taxa. Alternatively, shotgun metagenomic sequencing is more specific as functional information is directly obtained from the fragments of DNA within a sample. With more funding, obtaining functional information for **Chapters III**, **IV**, and **V** could provide a deeper understanding for functional consequences in changes of microbial communities.

3) Collect extensive and complete metadata

Metadata—information or characteristics collected to inform other data—is one of the most important aspects to microbiome research. Differences in microbiota based on body sites, diseases, treatments, ethnicity, socio-economic status *etc.* are only detected if the information is recorded. In this thesis, extensive metadata was provided for each of the studies, which especially led to the detection of hyperlipidaemia status of parents playing a role in the oral microbiota composition of children (**Chapter V**). While the metadata was extensive, issues of missing metadata remained. Missing metadata ultimately results in a reduction of sample size for statistical testing. For example, three samples were removed from **Chapter V** due to missing data. If samples with missing data are still included in statistical tests, such as Kruskal-Wallis pairwise tests for alpha diversity, it can lead to altered conclusions. For instance, the oral microbiota of preterm infants with BPD was compared to infants without—if the unknown status of BPD was included in the pairwise test—although not directly considered—there was no significance between infants with compared to without BPD; however, there was a significant difference between infants with and without BPD if the samples with unknown status were removed before running the test. If metadata is extensive and complete, it can detect new characteristics that contribute to the microbiota and can aid in detection of co-variates. Like in **Chapter V**, testing all characteristics collected, post hoc, can shine a light on characteristics that were not thought to be driving changes in the microbiome. Additionally, using multivariate testing (*e.g.*, adonis), the interaction between variates in the metadata can be identified. This type of analysis is very powerful, as new associations between characteristics of individuals can provide a deeper knowledge of the microbiome with respect to NCDs.

4) Crowdsource low microbial biomass samples

Crowdsourcing is one of the most cost-effective ways to collect a large number of samples and has been particularly successful for ancestry and gut microbiome research. For example, over 10,000 samples have been collected,[44] and almost two million dollars has been raised[140] for the American Gut Project. However, with the issues of contamination and self-reported metadata, there are some further considerations for crowdsourcing low microbial biomass samples. Collection of low microbial biomass samples is currently difficult, even for the best trained personnel, so encouraging the public to take due care with low microbial biomass samples may be difficult. Further, extensive metadata is beneficial, but lengthy questionnaires and self-reporting of sensitive information may discourage contributors or lead to false reporting. Perhaps, dedicated mobile vehicles could be used as collection centres, especially as most low microbial biomass sites, such as skin, nasal cavity and mouth, are less invasive, compared to faecal samples. Dedicated personnel could take samples in the cleanest possible manner and fill out questionnaires. In doing so, the many benefits to crowdsourcing low microbial biomass samples may be realised, especially for understanding NCDs,including the majority of the costs associated with processing microbiome samples are covered by the consumer; large numbers of individuals can be recruited, which improves the statistical power of a study; correlations and networks of microbiota can be identified across multiple sites of the human body; and, individuals may have samples taken multiple times, which can provide longitudinal data. Once larger studies are carried out, the microbiota's contribution to NCDs may be elucidated.

Conclusion

Overall, this thesis provides new perspectives on the development and prevalence of non-communicable diseases using information from low microbial biomass sites of the human body. Each chapter highlights important contributions from low microbial biomass samples in light of non-communicable diseases, which include best practices for incorporating low microbial biomass samples as a diagnostic tool; highlighting the role and connections across low biomass microbiotas; exposure to the environment can alter low microbial biomass body sites and potentially reduce immune-mediated diseases; development of preterm infant oral microbiota is disrupted at an early age, which may have long-term health consequences; and, metabolic and oral diseases could be linked in children with type

1 diabetes. Through these studies, an appreciation for system-wide interactions between microbiotas and the complexity of non-communicable diseases has come to light. My thesis provides foundational knowledge to non-communicable diseases that can be drawn upon in future research and it provides another piece to the non-communicable disease puzzle.

References

1. Bennett, J. E. *et al.* NCD Countdown 2030: worldwide trends in non-communicable disease mortality and progress towards Sustainable Development Goal target 3.4. *The Lancet* **392**, 1072–1088 (2018).
2. Transforming our world: the 2030 Agenda for Sustainable Development .:. Sustainable Development Knowledge Platform. https://sustainabledevelopment.un.org/post2015/transformingourworld.
3. Kim, D. *et al.* Optimizing methods and dodging pitfalls in microbiome research. *Microbiome* **5**, 52 (2017).
4. Salter, S. J. *et al.* Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biology* **12**, 87 (2014).
5. Karstens, L. *et al.* Controlling for Contaminants in Low-Biomass 16S rRNA Gene Sequencing Experiments. *mSystems* **4**, (2019).
6. Aagaard, K. *et al.* The placenta harbors a unique microbiome. *Sci Transl Med* **6**, 237ra65 (2014).
7. Chan, A. A. *et al.* Characterization of the microbiome of nipple aspirate fluid of breast cancer survivors. *Scientific Reports* **6**, 28061 (2016).
8. Hieken, T. J. *et al.* The Microbiome of Aseptically Collected Human Breast Tissue in Benign and Malignant Disease. *Scientific Reports* **6**, 30751 (2016).
9. Cooper, A. & Poinar, H. N. Ancient DNA: Do It Right or Not at All. *Science* **289**, 1139–1139 (2000).
10. Goodrich, J. K. *et al.* Conducting a Microbiome Study. *Cell* **158**, 250–262 (2014).
11. Minich, J. J. *et al.* KatharoSeq Enables High-Throughput Microbiome Analysis from Low-Biomass Samples. *mSystems* **3**, e00218-17 (2018).

12. Knight, R. *et al.* Best practices for analysing microbiomes. *Nat Rev Microbiol* **16**, 410–422 (2018).

13. Eisenhofer, R. *et al.* Contamination in Low Microbial Biomass Microbiome Studies: Issues and Recommendations. *Trends Microbiol.* **27**, 105–117 (2019).

14. Allaband, C. *et al.* Microbiome 101: Studying, Analyzing, and Interpreting Gut Microbiome Data for Clinicians. *Clin Gastroenterol Hepatol* **17**, 218–230 (2019).

15. Martin, T. C., Visconti, A., Spector, T. D. & Falchi, M. Conducting metagenomic studies in microbiology and clinical research. *Appl Microbiol Biotechnol* **102**, 8629–8646 (2018).

16. Leiby, J. S. *et al.* Lack of detection of a human placenta microbiome in samples from preterm and term deliveries. *Microbiome* **6**, 196 (2018).

17. Karstens, L. *et al.* Community profiling of the urinary microbiota: considerations for low-biomass samples. *Nat Rev Urol* **15**, 735–749 (2018).

18. Schierwagen, R. *et al.* Trust is good, control is better: technical considerations in blood microbiome analysis. *Gut* **69**, 1362–1363 (2020).

19. Zegans, M. E. & Gelder, R. N. V. Considerations in Understanding the Ocular Surface Microbiome. *American Journal of Ophthalmology* **158**, 420–422 (2014).

20. Casals-Pascual, C. *et al.* Microbial Diversity in Clinical Microbiome Studies: Sample Size and Statistical Power Considerations. *Gastroenterology* **158**, 1524–1528 (2020).

21. Selway, C. A., Eisenhofer, R. & Weyrich, L. S. Microbiome applications for pathology: challenges of low microbial biomass samples during diagnostic testing. *The Journal of Pathology: Clinical Research* **6**, 97–106 (2020).

22. Llamas, B. *et al.* From the field to the laboratory: Controlling DNA contamination in human ancient DNA research in the high-throughput sequencing era. *STAR: Science & Technology of Archaeological Research* **3**, 1–14 (2017).

23. Tamariz, J., Voynarovska, K., Prinz, M. & Caragine, T. The application of ultraviolet irradiation to exogenous sources of DNA in plasticware and water for

the amplification of low copy number DNA. *Journal of Forensic Sciences* **51**, 790–794 (2006).

24. Stinson, L. F., Keelan, J. A. & Payne, M. S. Identification and removal of contaminating microbial DNA from PCR reagents: impact on low-biomass microbiome analyses. *Letters in Applied Microbiology* **68**, 2–8 (2019).

25. Corless, C. E. *et al.* Contamination and sensitivity issues with a real-time universal 16S rRNA PCR. *Journal of Clinical Microbiology* **38**, 1747–1752 (2000).

26. Klaschik, S., Lehmann, L. E., Raadts, A., Hoeft, A. & Stuber, F. Comparison of different decontamination methods for reagents to detect low concentrations of bacterial 16S DNA by real-time-PCR. *Mol Biotechnol* **22**, 231–242 (2002).

27. Hornung, B. V. H., Zwittink, R. D. & Kuijper, E. J. Issues and current standards of controls in microbiome research. *FEMS Microbiol Ecol* **95**, (2019).

28. Weyrich, L. S. *et al.* Neanderthal behaviour, diet, and disease inferred from ancient DNA in dental calculus. *Nature* **544**, 357–361 (2017).

29. Lauder, A. P. *et al.* Comparison of placenta samples with contamination controls does not provide evidence for a distinct placenta microbiota. *Microbiome* **4**, 29 (2016).

30. Glassing, A., Dowd, S. E., Galandiuk, S., Davis, B. & Chiodini, R. J. Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of microbiota in low bacterial biomass samples. *Gut Pathogens* **8**, 24 (2016).

31. Weyrich, L. S. *et al.* Laboratory contamination over time during low-biomass sample analysis. *Molecular Ecology Resources* **19**, 982–996 (2019).

32. Davis, N. M., Proctor, D. M., Holmes, S. P., Relman, D. A. & Callahan, B. J. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome* **6**, 226 (2018).

33. Knights, D. *et al.* Bayesian community-wide culture-independent microbial source tracking. *Nature Methods* **8**, 761–763 (2011).

34. Davis, A. *et al.* Improved yield and accuracy for DNA extraction in microbiome studies with variation in microbial biomass. *BioTechniques* **66**, 285–289 (2019).

35. Xu, Z. *et al.* Improving the sensitivity of negative controls in ancient DNA extractions. *ELECTROPHORESIS* **30**, 1282–1285 (2009).

36. Minich, J. J. *et al.* Quantifying and Understanding Well-to-Well Contamination in Microbiome Research. *mSystems* **4**, (2019).

37. Tourlousse, D. M., Ohashi, A. & Sekiguchi, Y. Sample tracking in microbiome community profiling assays using synthetic 16S rRNA gene spike-in controls. *Scientific Reports* **8**, 9095 (2018).

38. Chafee, M., Maignien, L. & Simmons, S. L. The effects of variable sample biomass on comparative metagenomics. *Environ Microbiol* **17**, 2239–2253 (2015).

39. Weiss, S. *et al.* Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* **5**, 27 (2017).

40. Wang, S. *et al.* Maternal Vertical Transmission Affecting Early-life Microbiota Development. *Trends in Microbiology* **28**, 28–45 (2020).

41. Bateman, A. The Dynamics of Microbial Transfer and Persistence on Human Skin. (2017).

42. Poussin, C. *et al.* Interrogating the microbiome: experimental and computational considerations in support of study reproducibility. *Drug Discovery Today* **23**, 1644–1657 (2018).

43. Kong, H. H. *et al.* Performing skin microbiome research: A method to the madness. *J Invest Dermatol* **137**, 561–568 (2017).

44. McDonald, D. *et al.* American Gut: an Open Platform for Citizen Science Microbiome Research. *mSystems* **3**, (2018).

45. Sanford, J. A. & Gallo, R. L. Functions of the skin microbiota in health and disease. *Semin Immunol* **25**, 370–377 (2013).

46. Kong, H. H. *et al.* Temporal shifts in the skin microbiome associated with disease flares and treatment in children with atopic dermatitis. *Genome Res.* **22**, 850–859 (2012).

47. Oh, J. *et al.* The altered landscape of the human skin microbiome in patients with primary immunodeficiencies. *Genome Res.* **23**, 2103–2114 (2013).

48. Nørreslet, L. B., Agner, T. & Clausen, M.-L. The Skin Microbiome in Inflammatory Skin Diseases. *Curr Derm Rep* **9**, 141–151 (2020).

49. How, K. Y., Song, K. P. & Chan, K. G. Porphyromonas gingivalis: An Overview of Periodontopathic Pathogen below the Gum Line. *Front Microbiol* **7**, (2016).

50. Curtis, M. A., Diaz, P. I. & Dyke, T. E. V. The role of the microbiota in periodontal disease. *Periodontology 2000* **83**, 14–25 (2020).

51. Baker, J. L. & Edlund, A. Exploiting the Oral Microbiome to Prevent Tooth Decay: Has Evolution Already Provided the Best Tools? *Front Microbiol* **9**, (2019).

52. Zhang, L., Liu, Y., Zheng, H. J. & Zhang, C. P. The Oral Microbiota May Have Influence on Oral Cancer. *Front Cell Infect Microbiol* **9**, (2020).

53. Wang, Z. *et al.* Airway host-microbiome interactions in chronic obstructive pulmonary disease. *Respiratory Research* **20**, 113 (2019).

54. Cuthbertson, L. *et al.* Lung function and microbiota diversity in cystic fibrosis. *Microbiome* **8**, 45 (2020).

55. Scher, J. U. *et al.* The lung microbiota in early rheumatoid arthritis and autoimmunity. *Microbiome* **4**, (2016).

56. Greenbaum, S., Greenbaum, G., Moran-Gilad, J. & Weintraub, A. Y. Ecological dynamics of the vaginal microbiome in relation to health and disease. *American Journal of Obstetrics and Gynecology* **220**, 324–335 (2019).

57. Moustafa, A. *et al.* Microbial metagenome of urinary tract infection. *Scientific Reports* **8**, 4333 (2018).

58. Fettweis, J. M. *et al.* The vaginal microbiome and preterm birth. *Nature Medicine* **25**, 1012–1021 (2019).

59. Hong, X. *et al.* Association between polycystic ovary syndrome and the vaginal microbiome: A case-control study. *Clinical Endocrinology* **93**, 52–60 (2020).

60. Ozkan, J., Coroneo, M., Willcox, M., Wemheuer, B. & Thomas, T. Identification and Visualization of a Distinct Microbiome in Ocular Surface Conjunctival Tissue. *Invest. Ophthalmol. Vis. Sci.* **59**, 4268–4276 (2018).

61. Carabotti, M., Scirocco, A., Maselli, M. A. & Severi, C. The gut-brain axis: interactions between enteric microbiota, central and enteric nervous systems. *Ann Gastroenterol* **28**, 203–209 (2015).

62. Schroeder, B. O. & Bäckhed, F. Signals from the gut microbiota to distant organs in physiology and disease. *Nature Medicine* **22**, 1079–1089 (2016).

63. van de Wouw, M. *et al.* Short-chain fatty acids: microbial metabolites that alleviate stress-induced brain-gut axis alterations. *J. Physiol. (Lond.)* **596**, 4923–4944 (2018).

64. Ho, L. *et al.* Protective roles of intestinal microbiota derived short chain fatty acids in Alzheimer's disease-type beta-amyloid neuropathological mechanisms. *Expert Review of Neurotherapeutics* **18**, 83–90 (2018).

65. Zheng, P. *et al.* The gut microbiome from patients with schizophrenia modulates the glutamate-glutamine-GABA cycle and schizophrenia-relevant behaviors in mice. *Science Advances* **5**, eaau8317 (2019).

66. Sampson, T. R. *et al.* Gut Microbiota Regulate Motor Deficits and Neuroinflammation in a Model of Parkinson's Disease. *Cell* **167**, 1469-1480.e12 (2016).

67. Lee, S.-Y., Lee, E., Park, Y. M. & Hong, S.-J. Microbiome in the Gut-Skin Axis in Atopic Dermatitis. *Allergy Asthma Immunol Res* **10**, 354–362 (2018).

68. Abrahamsson, T. R. *et al.* Low diversity of the gut microbiota in infants with atopic eczema. *Journal of Allergy and Clinical Immunology* **129**, 434-440.e2 (2012).

69. Galazzo, G. *et al.* Development of the Microbiota and Associations With Birth Mode, Diet, and Atopic Disorders in a Longitudinal Analysis of Stool Samples, Collected From Infancy Through Early Childhood. *Gastroenterology* **158**, 1584–1596 (2020).

70. Vaughn, A. R., Notay, M., Clark, A. K. & Sivamani, R. K. Skin-gut axis: The relationship between intestinal bacteria and skin health. *World Journal of Dermatology* **6**, 52–58 (2017).

71. O'Neill, C. A., Monteleone, G., McLaughlin, J. T. & Paus, R. The gut-skin axis in health and disease: A paradigm with therapeutic implications. *BioEssays* **38**, 1167–1176 (2016).

72. Agamia, N. F., Abdallah, D. M., Sorour, O., Mourad, B. & Younan, D. N. Skin expression of mammalian target of rapamycin and forkhead box transcription factor O1, and serum insulin-like growth factor-1 in patients with acne vulgaris and their relationship with diet. *British Journal of Dermatology* **174**, 1299–1307 (2016).

73. Blanchet-Réthoré, S. *et al.* Effect of a lotion containing the heat-treated probiotic strain Lactobacillus johnsonii NCC 533 on Staphylococcus aureus colonization in atopic dermatitis. *Clinical, Cosmetic and Investigational Dermatology* **10**, 249–257 (2017).

74. Fabbrocini, G. *et al.* Supplementation with Lactobacillus rhamnosus SP1 normalises skin expression of genes implicated in insulin signalling and improves adult acne. *Beneficial Microbes* **7**, 625–630 (2016).

75. Mainous, A. G., Hueston, W. J., Everett, C. J. & Diaz, V. A. Nasal Carriage of Staphylococcus aureus and Methicillin-Resistant S aureus in the United States, 2001–2002. *Ann Fam Med* **4**, 132–137 (2006).

76. Lin, S.-Y., Lin, N.-Y., Huang, Y.-Y., Hsieh, C.-C. & Huang, Y.-C. Methicillin-resistant Staphylococcus aureus nasal carriage and infection among patients with diabetic foot ulcer. *Journal of Microbiology, Immunology and Infection* **53**, 292–299 (2020).

77. Onyango, L. A., Hugh Dunstan, R., Roberts, T. K., Macdonald, M. M. & Gottfries, J. Phenotypic Variants of Staphylococci and Their Underlying Population Distributions Following Exposure to Stress. *PLoS One* **8**, (2013).

78. Charlson, E. S. *et al.* Topographical continuity of bacterial populations in the healthy human respiratory tract. *Am. J. Respir. Crit. Care Med.* **184**, 957–963 (2011).

79. Scannapieco, F. A., Stewart, E. M. & Mylotte, J. M. Colonization of dental plaque by respiratory pathogens in medical intensive care patients. *Critical Care Medicine* **20**, 740–745 (1992).

80. Pragman, A. A. *et al.* The lung tissue microbiota of mild and moderate chronic obstructive pulmonary disease. *Microbiome* **6**, (2018).

81. Lin, M. *et al.* Saliva Microbiome Changes in Patients With Periodontitis With and Without Chronic Obstructive Pulmonary Disease. *Front. Cell. Infect. Microbiol.* **10**, (2020).

82. Narang, A., Maguire, A., Nunn, J. H. & Bush, A. Oral health and related factors in cystic fibrosis and other chronic respiratory disorders. *Archives of Disease in Childhood* **88**, 702–707 (2003).

83. Zhou, X., Zhang, W., Liu, X., Zhang, W. & Li, Y. Interrelationship between diabetes and periodontitis: role of hyperlipidemia. *Arch. Oral Biol.* **60**, 667–674 (2015).

84. Gatej, S. M. *et al.* Probiotic Lactobacillus Rhamnosus GG Protects Against P. Gingivalis And F. Nucleatum Gut Dysbiosis. *J Int Acad Periodontol* **22**, 18–27 (2020).

85. Teles, F. R. F., Alawi, F., Castilho, R. M. & Wang, Y. Association or Causation? Exploring the Oral Microbiome and Cancer Links. *J Dent Res* 0022034520945242 (2020) doi:10.1177/0022034520945242.

86. Gilbert, J. A. *et al.* Microbiome-wide association studies link dynamic microbial consortia to disease. *Nature* **535**, 94–103 (2016).

87. Turpin, W. *et al.* Association of host genome with intestinal microbial composition in a large healthy cohort. *Nature Genetics* **48**, 1413–1417 (2016).

88. Goodrich, J. K. *et al.* Genetic Determinants of the Gut Microbiome in UK Twins. *Cell Host & Microbe* **19**, 731–743 (2016).

89. Blekhman, R. *et al.* Host genetic variation impacts microbiome composition across human body sites. *Genome Biology* **16**, 191 (2015).

90. Wang, J. *et al.* Genome-wide association analysis identifies variation in vitamin D receptor and other host factors influencing the gut microbiota. *Nature Genetics* **48**, 1396–1406 (2016).

91. Wang, J. *et al.* Meta-analysis of human genome-microbiome association studies: the MiBioGen consortium initiative. *Microbiome* **6**, 101 (2018).

92. Purcell, R. V., Visnovska, M., Biggs, P. J., Schmeier, S. & Frizelle, F. A. Distinct gut microbiome patterns associate with consensus molecular subtypes of colorectal cancer. *Scientific Reports* **7**, 11590 (2017).

93. Chiu, C.-Y. *et al.* Integration of metagenomics-metabolomics reveals specific signatures and functions of airway microbiota in mite-sensitized childhood asthma. *Allergy* **n/a**,.

94. Reyman, M. *et al.* Impact of delivery mode-associated gut microbiota dynamics on health in the first year of life. *Nature Communications* **10**, 4997 (2019).

95. Eck, A. *et al.* Neonatal microbiota development and the effect of early life antibiotics are determined by two distinct settler types. *PLOS ONE* **15**, e0228133 (2020).

96. Forbes, J. D. *et al.* Association of Exposure to Formula in the Hospital and Subsequent Infant Feeding Practices With Gut Microbiota and Risk of Overweight in the First Year of Life. *JAMA Pediatr* **172**, e181161–e181161 (2018).

97. Cox, L. M. *et al.* Altering the Intestinal Microbiota during a Critical Developmental Window Has Lasting Metabolic Consequences. *Cell* **158**, 705–721 (2014).

98. Olin, A. *et al.* Stereotypic Immune System Development in Newborn Children. *Cell* **174**, 1277-1292.e14 (2018).

99. Stewart, C. J. *et al.* Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature* **562**, 583–588 (2018).

100. Strachan, D. P. Hay fever, hygiene, and household size. *BMJ* **299**, 1259–1260 (1989).

101. Rook, G. A. W., Martinelli, R. & Brunet, L. R. Innate immune responses to mycobacteria and the downregulation of atopic responses. *Curr Opin Allergy Clin Immunol* **3**, 337–342 (2003).

102. von Hertzen, L., Hanski, I. & Haahtela, T. Natural immunity. *EMBO reports* **12**, 1089–1093 (2011).

103. Fujimura, K. E. *et al.* House dust exposure mediates gut microbiome Lactobacillus enrichment and airway immune defense against allergens and virus infection. *PNAS* **111**, 805–810 (2014).

104. Stein, M. M. *et al.* Innate Immunity and Asthma Risk in Amish and Hutterite Farm Children. *New England Journal of Medicine* **375**, 411–421 (2016).

105. Liddicoat, C. *et al.* Naturally-diverse airborne environmental microbial exposures modulate the gut microbiome and may provide anxiolytic benefits in mice. *Science of The Total Environment* **701**, 134684 (2020).

106. Di Daniele, N. The Role of Preventive Nutrition in Chronic Non-Communicable Diseases. *Nutrients* **11**, (2019).

107. Matt, S. M. *et al.* Butyrate and Dietary Soluble Fiber Improve Neuroinflammation Associated With Aging in Mice. *Front. Immunol.* **9**, (2018).

108. Do, M. H., Lee, E., Oh, M.-J., Kim, Y. & Park, H.-Y. High-Glucose or -Fructose Diet Cause Changes of the Gut Microbiota and Metabolic Disorders in Mice without Body Weight Change. *Nutrients* **10**, (2018).

109. Satokari, R. High Intake of Sugar and the Balance between Pro- and Anti-Inflammatory Gut Bacteria. *Nutrients* **12**, (2020).

110. Chi, D. L. & Scott, J. M. Added Sugar and Dental Caries in Children: A Scientific Update and Future Steps. *Dent Clin North Am* **63**, 17–33 (2019).

111. Touger-Decker, R. & van Loveren, C. Sugars and dental caries. *Am J Clin Nutr* **78**, 881S-892S (2003).

112. Ritchie, H. & Roser, M. Smoking. *Our World in Data* (2013).

113. Mason, M. R. *et al.* The subgingival microbiome of clinically healthy current and never smokers. *ISME J* **9**, 268–272 (2015).

114. Ganesan, S. M. *et al.* Adverse effects of electronic cigarettes on the disease-naive oral microbiome. *Science Advances* **6**, eaaz0108 (2020).

115. Erb-Downward, J. R. *et al.* Analysis of the Lung Microbiome in the "Healthy" Smoker and in COPD. *PLOS ONE* **6**, e16384 (2011).

116. Segal, L. N. Distressing microbial consequences of smoking. *Science Translational Medicine* **10**, (2018).

117. Huang, C. & Shi, G. Smoking and microbiome in oral, airway, gut and some systemic diseases. *Journal of Translational Medicine* **17**, 225 (2019).

118. Dubin, K. *et al.* Intestinal microbiome analyses identify melanoma patients at risk for checkpoint-blockade-induced colitis. *Nature Communications* **7**, 10391 (2016).

119. Mundell, L., Lindemann, R. & Douglas, J. Monitoring long-term oral corticosteroids. *BMJ Open Qual* **6**, e000209 (2017).

120. Ni, J. *et al.* Early antibiotic exposure and development of asthma and allergic rhinitis in childhood. *BMC Pediatrics* **19**, 225 (2019).

121. Karl, J. P. *et al.* Effects of Psychological, Environmental and Physical Stressors on the Gut Microbiota. *Front Microbiol* **9**, (2018).

122. Fricchione, G. L. The Challenge of Stress-Related Non-Communicable Diseases. *Med Sci Monit Basic Res* **24**, 93–95 (2018).

123. Lif Holgerson, P., Esberg, A., Sjödin, A., West, C. E. & Johansson, I. A longitudinal study of the development of the saliva microbiome in infants 2 days to 5 years compared to the microbiome in adolescents. *Scientific Reports* **10**, 9629 (2020).

124. Vu, K. *et al.* From birth to overweight and atopic disease: multiple and common pathways of the infant gut microbiome. *Gastroenterology* **0**, (2020).

125. Drall, K. M. *et al.* Vitamin D supplementation in pregnancy and early infancy in relation to gut microbiota composition and C. difficile colonization: implications for viral respiratory infections. *Gut Microbes* **12**, 1799734 (2020).

126. Simons, E. *et al.* Timing of Introduction, Sensitization, and Allergy to Highly Allergenic Foods at Age 3 Years in a General-Population Canadian Cohort. *The Journal of Allergy and Clinical Immunology: In Practice* **8**, 166-175.e10 (2020).

127. David, L. A. *et al.* Host lifestyle affects human microbiota on daily timescales. *Genome Biology* **15**, R89 (2014).

128. Gajer, P. *et al.* Temporal Dynamics of the Human Vaginal Microbiota. *Science Translational Medicine* **4**, 132ra52-132ra52 (2012).

129. Oh, J., Byrd, A. L., Park, M., Kong, H. H. & Segre, J. A. Temporal Stability of the Human Skin Microbiome. *Cell* **165**, 854–866 (2016).

130. Lloyd-Price, J. *et al.* Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* **569**, 655–662 (2019).

131. Clooney, A. G. *et al.* Ranking microbiome variance in inflammatory bowel disease: a large longitudinal intercontinental study. *Gut* (2020) doi:10.1136/gutjnl-2020-321106.

132. Vangay, P. *et al.* U.S. immigration westernizes the human gut microbiome. *Cell* **175**, 962-972.e10 (2018).

133. Socransky, S. S., Haffajee, A. D., Cugini, M. A., Smith, C. & Kent, R. L. Microbial complexes in subgingival plaque. *J. Clin. Periodontol.* **25**, 134–144 (1998).

134. Liu, Y., Tran, D. Q. & Rhoads, J. M. Probiotics in Disease Prevention and Treatment. *J Clin Pharmacol* **58**, S164–S179 (2018).

135. Tang, M. L. K. *et al.* Administration of a probiotic with peanut oral immunotherapy: A randomized trial. *Journal of Allergy and Clinical Immunology* **135**, 737-744.e8 (2015).

136. Cammarota, G. *et al.* Randomised clinical trial: faecal microbiota transplantation by colonoscopy vs. vancomycin for the treatment of recurrent Clostridium difficile infection. *Alimentary Pharmacology & Therapeutics* **41**, 835–843 (2015).

137. Costello, S. P. *et al.* Effect of Fecal Microbiota Transplantation on 8-Week Remission in Patients With Ulcerative Colitis: A Randomized Clinical Trial. *JAMA* **321**, 156–164 (2019).

138. Li, N. *et al.* [Efficacy analysis of fecal microbiota transplantation in the treatment of 406 cases with gastrointestinal disorders]. *Zhonghua Wei Chang Wai Ke Za Zhi = Chinese Journal of Gastrointestinal Surgery* **20**, 40–46 (2017).

139. Nascimento, M. M. Oral microbiota transplant: a potential new therapy for oral diseases. *J Calif Dent Assoc* **45**, 565–568 (2017).

140. American Gut. *FundRazr* https://fundrazr.com/americangut.

Appendix I

Chapter III Supplementary Materials

Supplementary tables can be found as electronic files

The following files are provided separately to this document:

AppendixI_TableS1.xlsx - Table S1. Metadata for sampling locations.

AppendixI_TableS2.xlsx - Table S2. Contaminant ASVs identified from laboratory controls.

AppendixI_TableS3.xlsx - Table S3. Number of sequences assigned to each sample.

AppendixI_TableS4.xlsx - Table S4. Sequencing information for each sample type.

AppendixI_TableS5.xlsx - Table S5. Core ASVs are shared between post-exposure and pre-exposure and environmental samples.

Appendix II

Chapter IV Supplementary Materials

Supplementary Methods

Study population and sample collection

All microbiota samples were collected under the Human Research Ethics Committee approval obtained for both preterm infants (Women and Children's Hospital; HREC 2434/12/16; 27/11/2013) and adults (University of Adelaide; H2012-108). A subset of preterm infants primarily enrolled in a large clinical trial (N3RO)[1], which investigated the effects of docosahexaenoic acid (DHA), were selected for this study. For preterm infants (P), patient metadata and population information was collected by medical record examination (Table 1). Preterm infant buccal swabs were collected at two time points: preterm timepoint one (PT1, 2-12 days post-birth) and preterm timepoint two (PT2, 36 weeks' postmenstrual age) from 50 preterm infants delivered at the Women's and Children's Hospital in Adelaide, Australia (Figure 1). Twenty-nine of these preterm infants developed BPD or sepsis (BPD, n=17; sepsis, n=9; BPD and sepsis, n=3). Fourteen preterm neonates required intubation, whereby tracheal aspirate samples were collected. All samples were frozen immediately in empty tubes after collection at -20°C to preserve the bacterial composition. For a healthy, mature oral microbiota comparison, three adult samples were collected and processed following the same protocols as the preterm infant samples. To further explore infant oral maturation, published oral microbiota data were obtained for 14 full-term infants (F) collected at two similar time points (FT1, 0-7 days post-birth; FT2; 4th-5th week post-birth) and 13 mothers (Study ID 2010, QIITA data repository; Figure 1).

Sample DNA extraction, 16S Ribosomal RNA Library Preparation, and DNA Sequencing

All samples were prepared in a still-air room designed for low-biomass microbiota analysis using strict measures to reduce cross-contamination and the introduction of background DNA.[2] DNA was extracted from buccal swabs and tracheal aspirates using a previously published, in-house silica DNA extraction method designed to enhance DNA recovery;[3] the method was modified to include mechanical lysis. The DNA extraction was performed with the following modifications: initial bead-beating step with 500µL lysis buffer (470µL EDTA and 30µL SDS) and 0.1mm glass beads in a 2mL screw cap tube (BeadBug™, Sigma); 1-hour incubation with 20µL

of 20mg/mL proteinase K; and 1.5mL modified guanidine DNA-binding buffer. DNA present within the laboratory and reagents was monitored using extraction blank controls (EBCs).

Bacterial DNA from each sample, including EBCs, was amplified in triplicate using primers that target the V4 region of the 16S ribosomal RNA (rRNA) gene[4], using previously described amplification conditions.[5] No-template controls (NTCs) were also included in each amplification batch. 16S rRNA libraries were prepared for sequencing. Pooled 16S rRNA library triplicates were quantified and pooled in equimolar concentrations into groups of ~30 samples. Pooled libraries were then purified (SeraPure; Homemade AMPure[6]) and quantified using a TapeStation (Agilent 2200) before a final pooling and quantification using a KAPA kit (LightCycler 96 System, Roche Life Science). All libraries were sequenced on an Illumina MiSeq (2x150 bp) at the Australian Genomics Research Facility.

Pre-processing, ASV selection, and contaminant removal

Preterm infant sequences (PT1, PT2, and tracheal aspirates) and adult (n=3) sequences obtained over two sequencing runs were uploaded to QIITA data repository (Study ID 11832; https://qiita.ucsd.edu/study/description/11832). Demultiplexed sequences were trimmed to 150 bp and amplicon sequence variants (ASVs) were generated via Deblur.[7] ASVs from the 'reference hit' biom file were merged with a full-term infant and maternal dataset (Study ID 2010; FT1 and FT2) to create a SEPP insertion tree[8] and the study dataset. Contaminant sequences from EBCs and NTCs within the preterm data were identified via Decontam,[9] as there were no laboratory controls from the full-term dataset, and were subsequently removed from all biological samples (Table S1) in QIIME 2 (v2019.7)[10]. Following this, ASVs with less than 10 reads assigned were removed. In total, 156 samples were retained and represented 8,738,663 sequences and 1,553 ASVs.

Diversity analyses, taxonomic classification, and statistical comparisons

In QIIME2 (v2019.7)[10], ASVs were summarized into their taxonomic classification using the SILVA database (v132.9; 16S 515-806).[11] Using a rarefied depth of 2,000 sequences, (alpha) diversity was measured using observed species (OS) and

Faith's phylogenetic diversity (PD)[12] metrics, and beta diversity (composition) was calculated using the unweighted UniFrac metric.[13] Significant associations between diversity and sample metadata were detected using pairwise Kruskal-Wallis tests[14] and the pairwise Fligner-Killeen test,[15,16] while significant links between composition and metadata were examined using adonis[17,18] and PERMANOVA.[17] In all tests, "Extraction Date" significantly contributed to the diversity in the dataset (*e.g.* ~18% of the total variation in the data; adonis); therefore, the impact of 'Extraction Date' was treated as a confounding variable in all downstream analyses. Lastly, ANCOM[19] was used to identify significant changes in an ASV abundance across samples.

References

1. Collins CT, Makrides M, McPhee AJ, et al. Docosahexaenoic Acid and Bronchopulmonary Dysplasia in Preterm Infants. *New England Journal of Medicine*. 2017;376(13):1245-1255. doi:10.1056/NEJMoa1611942
2. Weyrich LS, Farrer AG, Eisenhofer R, et al. Laboratory contamination over time during low-biomass sample analysis. *Molecular Ecology Resources*. 2019;19(4):982-996. doi:10.1111/1755-0998.13011
3. Rohland N, Hofreiter M. Comparison and optimization of ancient DNA extraction. *BioTechniques*. 2007;42(3):343-352. doi:10.2144/000112383
4. Caporaso JG, Lauber CL, Walters WA, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J*. 2012;6(8):1621-1624. doi:10.1038/ismej.2012.8
5. Selway CA, Mills JG, Weinstein P, et al. Transfer of environmental microbes to the skin and respiratory tract of humans after urban green space exposure. *Environment International*. 2020;145:106084. doi:10.1016/j.envint.2020.106084
6. Rohland N, Reich D. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Res*. 2012;22(5):939-946. doi:10.1101/gr.128124.111

7. Amir A, McDonald D, Navas-Molina JA, et al. Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. *mSystems*. 2017;2(2):e00191-16. doi:10.1128/mSystems.00191-16

8. Janssen S, McDonald D, Gonzalez A, et al. Phylogenetic Placement of Exact Amplicon Sequences Improves Associations with Clinical Information. *mSystems*. 2018;3(3). doi:10.1128/mSystems.00021-18

9. Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome*. 2018;6(1):226. doi:10.1186/s40168-018-0605-2

10. Bolyen E, Rideout JR, Dillon MR, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol*. 2019;37(8):852-857. doi:10.1038/s41587-019-0209-9

11. Quast C, Pruesse E, Yilmaz P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res*. 2013;41(D1):D590-D596. doi:10.1093/nar/gks1219

12. Faith DP. Conservation evaluation and phylogenetic diversity. *Biological Conservation*. 1992;61(1):1-10. doi:10.1016/0006-3207(92)91201-3

13. Lozupone C, Knight R. UniFrac: a New Phylogenetic Method for Comparing Microbial Communities. *Appl Environ Microbiol*. 2005;71(12):8228-8235. doi:10.1128/AEM.71.12.8228-8235.2005

14. Kruskal WH, Wallis WA. Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association*. 1952;47(260):583-621. doi:10.2307/2280779

15. Fligner MA, Killeen TJ. Distribution-Free Two-Sample Tests for Scale. *Journal of the American Statistical Association*. 1976;71(353):210-213. doi:10.1080/01621459.1976.10481517

16. Conover WJ, Johnson ME, Johnson MM. A Comparative Study of Tests for Homogeneity of Variances, with Applications to the Outer Continental Shelf Bidding Data. *Technometrics*. 1981;23(4):351-361. doi:10.2307/1268225

17. Anderson MJ. A new method for non-parametric multivariate analysis of variance. *Austral Ecology*. 2001;26(1):32-46. doi:10.1111/j.1442-9993.2001.01070.pp.x

18. Oksanen J, Blanchet FG, Friendly M, et al. *Vegan: Community Ecology Package*.; 2018. https://CRAN.R-project.org/package=vegan

19. Mandal S, Treuren WV, White RA, Eggesbø M, Knight R, Peddada SD. Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microbial Ecology in Health and Disease*. 2015;26(1):27663. doi:10.3402/mehd.v26.27663

Supplementary tables can be found as electronic files

The following files are provided separately to this document:

AppendixII_Tables_ChapterIV.xlsx - All tables for Chapter IV

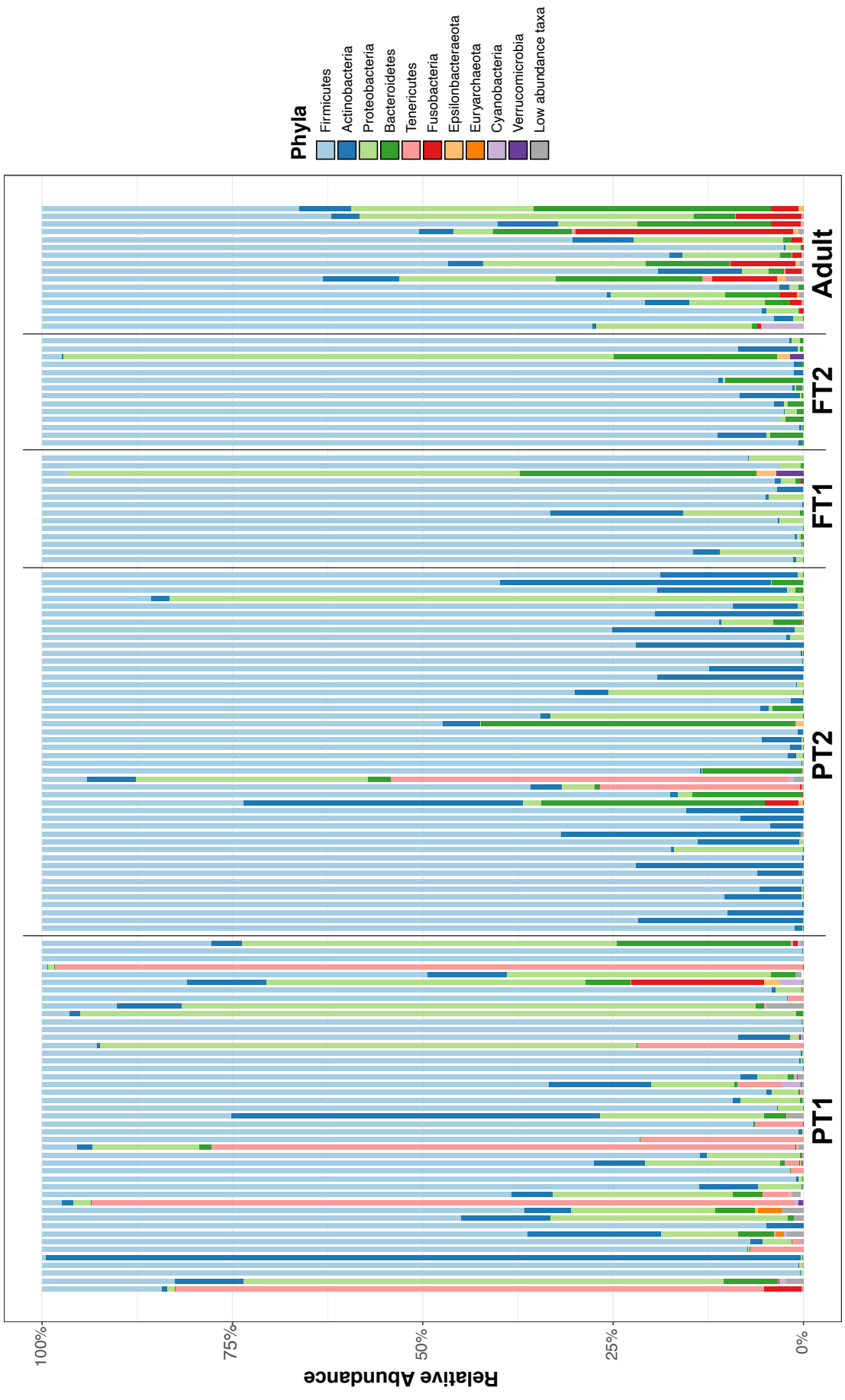

Figure S1: PT1 samples are stochastic. The relative abundance of oral bacterial genera are shown for preterm and full-term infants at two time points, as well as adult samples. Genera with <1% relative abundance were collapsed.

Appendix III

Chapter V Supplementary Materials

Supplementary tables can be found as electronic files

The following files are provided separately to this document:

AppendixIII_TablesS1.xlsx - Table S1. Contaminant ASVs identified from laboratory controls

Table S2: No interaction between periodontal pocket depth, glycated haemoglobin and a family history of hyperlipidemia.

	Sum of Squares	F value	Pr (>F)
Periodontal PD	6.5486	1.2083	0.27578
HbA1c	25.6292	4.7290	0.03336*
FHx Hyperlipidemia	12.6063	2.3261	0.13215
Periodontal PD:HbA1c	6.7710	1.2494	0.26785
Periodontal PD:FHx Hyperlipidemia	3.3294	0.6143	0.43606
HbA1c:FHx Hyperlipidemia	21.7358	4.0106	0.04946*
Periodontal PD:HbA1c:FHx Hyperlipidemia	6.9440	1.2813	0.26189

PD = pocket depth; FHx = family history of hyperlipidemia; HbA1c = glycated hemoglobin; * 0.05 > p > 0.01.

Table S3: Periodontal pocket depth and a family history of hyperlipidemia dependently affect the oral microbiota.

	Sum of Squares	F value	Pr (>F)
Periodontal PD	0.121	0.0220	0.882476
HbA1c	25.386	4.6078	0.035452 *
FHx Hyperlipidemia	56.926	10.3324	0.002012 **
Periodontal PD:FHx Hyperlipidemia	39.573	7.1828	0.009255 **

PD = pocket depth; FHx = family history of hyperlipidemia; HbA1c = glycated hemoglobin; * 0.05 > p > 0.01, ** p < 0.01.

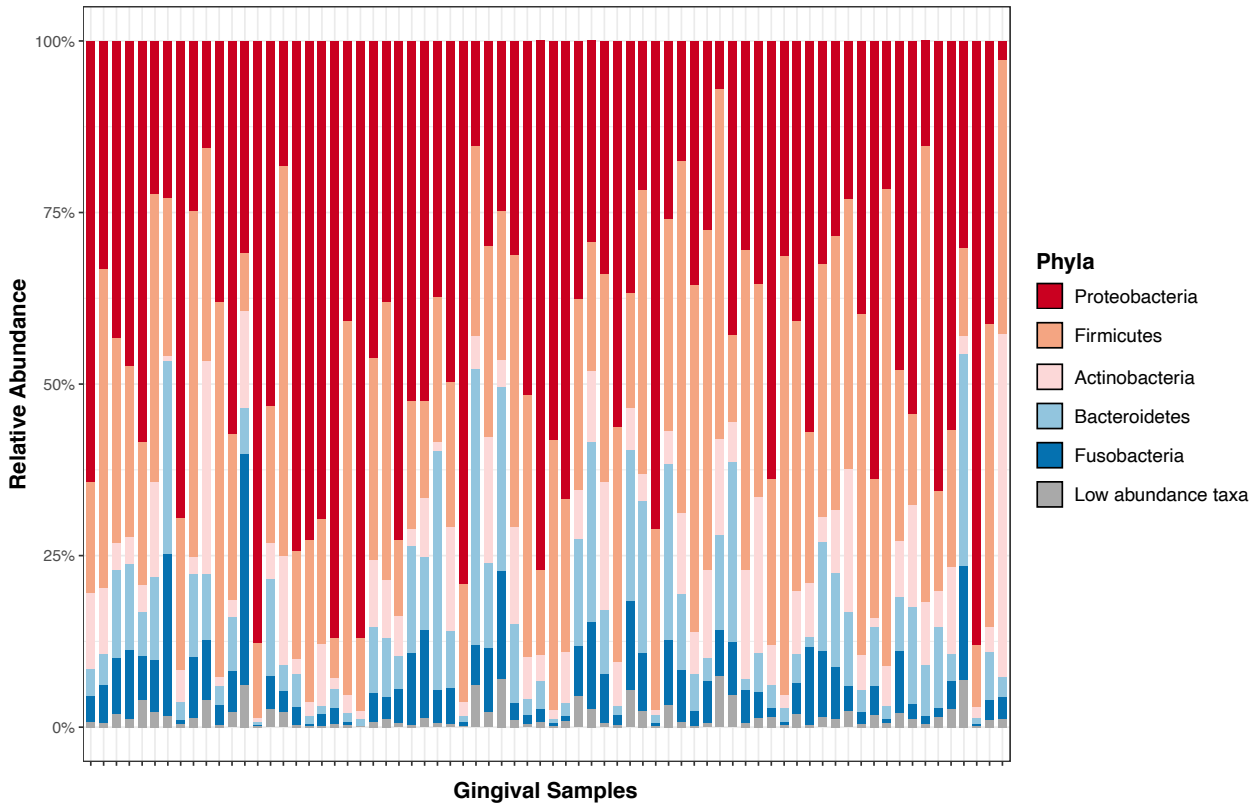

Figure S1: The relative abundance of gingival samples at the phyla taxonomic level. Five dominant phyla (>2% relative abundance) were observed across all gingival samples from children with type 1 diabetes.

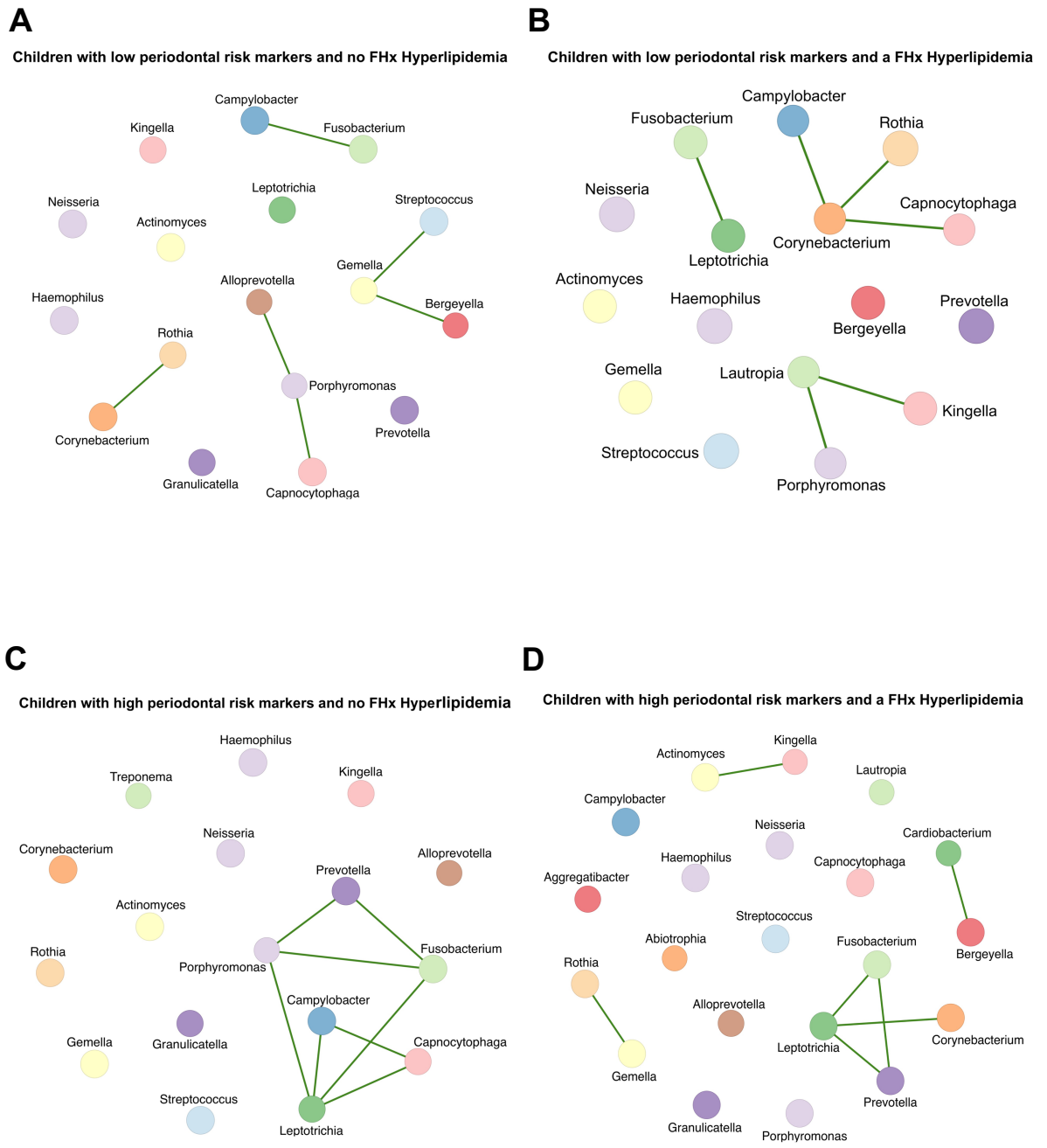

Figure S2: Fewer co-occurrence networks were observed in periodontally healthy children with a family history of hyperlipidemia. Co-occurrence plots were generated using the Pearson co-efficient, to compare differences in hyperlipidemia and periodontal status. Networks between periodontally healthy children without (A) and with (B) a family history of hyperlipidemia were distinct, with fewer networks observed in periodontally healthy children with a family history of hyperlipidemia. In children with high-risk periodontal markers, frequently observed periodontal pathogens were shown to co-occur in both children without (C) and (D) with a family history of hyperlipidemia.

Appendix IV

Laboratory contamination over time during low-biomass sample analysis (Weyrich *et al.* 2018)

Received: 15 January 2018 | Accepted: 8 March 2019
DOI: 10.1111/1755-0998.13011

RESOURCE ARTICLE

MOLECULAR ECOLOGY RESOURCES WILEY

Laboratory contamination over time during low-biomass sample analysis

Laura S. Weyrich[1,2] | Andrew G. Farrer[1] | Raphael Eisenhofer[1,2] | Luis A. Arriola[1] | Jennifer Young[1] | Caitlin A. Selway[1] | Matilda Handsley-Davis[1,2] | Christina J. Adler[3] | James Breen[1] | Alan Cooper[1,2]

[1]Australian Centre for Ancient DNA, University of Adelaide, Adelaide, South Australia, Australia

[2]ARC Centre of Excellence for Australian Biodiversity and Heritage, University of Adelaide, Adelaide, South Australia, Australia

[3]School of Dentistry, Faculty of Medicine and Health, University of Sydney, Sydney, New South Wales, Australia

Correspondence
Laura S. Weyrich, Australian Centre for Ancient DNA, University of Adelaide, Adelaide, SA, Australia.
Email: laura.weyrich@adelaide.edu.au

Present Address
Jennifer Young, Advanced DNA, Identification, and Forensic Facility, University of Adelaide, Adelaide, South Australia, Australia

James Breen, South Australian Health and Medical Research Institute (SAHMRI), Adelaide, South Australia, Australia

Funding information
Australian Research Council, Grant/Award Number: DECRA 150101574

Abstract

Bacteria are not only ubiquitous on earth but can also be incredibly diverse within clean laboratories and reagents. The presence of both living and dead bacteria in laboratory environments and reagents is especially problematic when examining samples with low endogenous content (e.g., skin swabs, tissue biopsies, ice, water, degraded forensic samples or ancient material), where contaminants can outnumber endogenous microorganisms within samples. The contribution of contaminants within high-throughput studies remains poorly understood because of the relatively low number of contaminant surveys. Here, we examined 144 negative control samples (extraction blank and no-template amplification controls) collected in both typical molecular laboratories and an ultraclean ancient DNA laboratory over 5 years to characterize long-term contaminant diversity. We additionally compared the contaminant content within a home-made silica-based extraction method, commonly used to analyse low endogenous content samples, with a widely used commercial DNA extraction kit. The contaminant taxonomic profile of the ultraclean ancient DNA laboratory was unique compared to modern molecular biology laboratories, and changed over time according to researcher, month and season. The commercial kit also contained higher microbial diversity and several human-associated taxa in comparison to the home-made silica extraction protocol. We recommend a minimum of two strategies to reduce the impacts of laboratory contaminants within low-biomass metagenomic studies: (a) extraction blank controls should be included and sequenced with every batch of extractions and (b) the contributions of laboratory contamination should be assessed and reported in each high-throughput metagenomic study.

KEYWORDS
ancient DNA, contaminant, contamination, metagenomics, microbiome, microbiota

1 | INTRODUCTION

In the new era of culture-independent microbiome research, targeted amplicon or "metabarcoding" approaches are now routinely used to amplify DNA from microbial species across the tree of life. However, these methods lack the ability to select for specific species or exclude contaminants (Caporaso et al., 2012). Although these techniques have provided invaluable insight into otherwise cryptic microbial communities, the increased sensitivity and lack of target specificity leave microbiota studies particularly susceptible to the effects of contamination. Such effects are widespread, as several recent studies have indicated that contaminant microbial DNA can be routinely isolated from laboratory reagents and surfaces (Laurence, Hatzis, & Brash, 2014; Salter et al., 2014; Tanner, Goebel, & Dojka, 1998) and that this signal has significantly impacted the past interpretation and characterization of microbiota in high-throughput sequencing studies. For example, Salter et al. (2014) demonstrated that bacterial DNA present in laboratory reagents is present in both quality-filtered 16S rRNA gene and shotgun metagenomic data sets and significantly impacts the interpretation of results. Multiple microbial contaminants have also been identified within the published 1,000 Genomes data set and other medical genomic studies (Kearney et al., 2012; Laurence et al., 2014). Despite these findings, the routine assessment of microbial background contamination is still not required, or fully reported, in microbiota studies.

While the presence of contaminant DNA is widespread, the effects are particularly problematic in low-biomass samples that contain very little endogenous DNA (Weiss et al., 2014; e.g., preterm infant swabs, tissue samples, such as placenta, tumour biopsies or breast tissue, and some environmental samples, such as ice or calcite). In low-biomass samples, a small contaminant signal from laboratory reagents can easily overpower the intrinsic signal from the sample. This is similarly an issue in current palaeomicrobiology studies that examine ancient, degraded microbiota, such as mummified human tissue, preserved faeces (coprolites) or calcified dental plaque (calculus; Warinner, Speller, & Collins, 2015; Weiss et al., 2014; Weyrich, Dobney, & Cooper, 2015). In ancient samples, the amount of endogenous DNA attributed to the original source can be extremely low (e.g., <0.05% of the total DNA in the sample) and is damaged, fragmented and intermixed with longer, higher-quality modern DNA fragments from contaminant species (Cooper & Poinar, 2000). Therefore, monitoring and understanding the contributions of contaminant DNA, especially in low-biomass or ancient samples, is critical to ensure that reported results are only based on the endogenous DNA.

Microbial contaminant DNA (i.e., background or exogenous DNA) is a mixture of DNA from both environmental and laboratory sources, with the former including factors such as soil from a burial site, air within the sampling facility and microorganisms from people touching the sample, while the latter involves reagents, glassware, labware and surfaces (Weyrich et al., 2015). Environmental contamination in low-biomass samples may be difficult to control or monitor, but the laboratory contaminants can be monitored by including extraction blank (EBC) and no-template amplification (NTC) controls and assessed using bioinformatics tools (e.g., SOURCETRACKER Knights et al., 2011). An EBC is an empty tube introduced during the extraction steps to collect DNA from the laboratory environment and the reagents throughout processing (Adler et al., 2013). Similarly, an NTC is an amplification reaction that lacks the addition of DNA from biological samples. These controls should be amplified and sequenced along with other samples and are critical to identify and exclude contaminant taxa from downstream analyses, reducing noise and ensuring any results are based solely on endogenous DNA (Weyrich et al., 2017). Despite this, there are surprisingly few published resources describing contaminant taxa found in EBCs or NTCs (Glassing, Dowd, Galandiuk, Davis, & Chiodini, 2016; Lauder et al., 2016; Salter et al., 2014).

In this study, we used 16S rRNA metabarcoding to characterize the contaminant diversity in 144 EBCs and NTCs using laboratory techniques specifically designed for low-biomass material. We also explored differences in microbial contamination within two different types of laboratory facilities: a state-of-the-art, purpose-built ancient DNA clean laboratory over the course of 5 years, and three typical modern molecular biology laboratories over 1 year. Lastly, we investigated differences between a common commercial DNA extraction kit and a home-made DNA extraction method typically applied in the ancient DNA field. Overall, this study is designed to assess contaminant profiles over time and identify more potential contaminant sequences in both high- and low-biomass research.

2 | MATERIALS AND METHODS

2.1 | Sample collection

Four different types of sample were used: ancient dental calculus (calcified dental plaque), modern dental calculus, EBCs and NTCs. Dental calculus samples were obtained from ancient and modern humans as described by Adler et al. (2013). A single EBC was included in each batch of extractions by treating an empty tube as if it was a biological sample throughout the DNA extraction and library preparation process. Similarly, NTC samples were created during the 16S rRNA library amplification stage by processing additional reactions without adding any known template DNA. Both EBCs and NTCs were subsequently included through to DNA sequencing and were included at a ratio of one control sample for every 10 biological samples.

2.2 | Description of laboratory facilities

DNA extraction occurred in two different types of laboratory facilities: a purpose-built, ultraclean ancient DNA laboratory (ancient lab) and three typical modern molecular biology laboratories (modern labs). The ancient lab is physically remote from the university campus in a building with no other molecular biology

laboratories and contains a HEPA (high-efficiency particulate air)-filtered, positive pressure air system to remove DNA and bacteria from external sources. The HEPA filter function is checked annually and changed every 10 years. The surface and floors within the laboratory are cleaned weekly with a 5% bleach (NaClO) solution and are illuminated with ceiling-mounted UV lights for 30 min each night. UV light bulbs are changed annually. Users entering the ancient lab are required to have showered, wear freshly laundered clothing, avoid the university campus prior to entry, and cannot bring personal equipment (e.g., phones, writing equipment and bags) into the facility. Standard personal laboratory wear includes disposable full-body suits, surgical facemasks, plastic see-through visors, and three layers of gloves to allow frequent changing without skin exposure (including one pair of inner elbow-length surgical gloves). All liquid reagents within the ancient lab are certified DNA-free, and the outer surface of all plasticware and reagent bottles are decontaminated prior to entering the laboratory (i.e., cleaned with 5% bleach and treated with UV (2×, 40-W, 254-nm UV tubes at a distance of 10 cm for 10 min) within a UV oven (Ultra Violet Products). All DNA extractions and amplification preparations are performed in a room separate to sample preparation and are completed in still-air cabinets that are cleaned with 5% bleach and UV treated for 30 min (3×, 15-W, 253.7-nm tube lamps; AURA PCR) prior to beginning any work. In addition, ancient samples from different sources (e.g., soil, plants and other animals) are processed in separate, dedicated rooms to minimize cross-contamination. In contrast, the modern laboratories are located over 2 km away from the ancient lab at the University of Adelaide (n = 2) and at the University of Sydney (n = 1). All three modern labs are typical of most molecular biology laboratories, are not routinely decontaminated and contain users who routinely use latex gloves but are not required to wear body suits or masks. DNA extracted within the modern labs comes from a wide range of sources (e.g., humans, mammals and environmental samples), although microbiome extractions were only performed on days when no other material was being extracted. In all facilities, DNA was extracted and prepared for amplification in still-air cabinets that were cleaned before and after each use with 5% bleach.

2.3 | DNA extractions

Several specialized DNA extraction protocols have been developed within ancient DNA studies to remove environmental contamination and enhance the recovery of the endogenous DNA. The extraction method selected for this study has previously been described for work on ancient dental calculus (Weyrich et al., 2017). Each ancient sample was first decontaminated using a published protocol (Adler et al., 2013), while modern samples were not decontaminated. The decontamination procedure included exposure to UV radiation for 15 min on each side of the sample, submersion of the sample in 5% bleach for 5 min, followed by submersion in 90% ethanol for 3 min to remove any residual bleach, and 5 min of drying. Decontaminated ancient calculus was then wrapped in aluminium foil and pulverized into power with a steel hammer and placed into a sterile 2-ml tube. The EBCs were empty tubes exposed to air for 30 s in the same room during sample decontamination and were included in the extraction process as if they contained a biological sample.

Following decontamination, DNA was extracted using the QG-based method previously described for the extraction of ancient microbiome material (Weyrich et al., 2017; referred to as "QG"). All reagents for the QG extraction method were prepared in a "sample-free" room in the ancient DNA facility, and all reagents were aliquoted immediately upon opening and frozen until further use to avoid cross contamination. Where possible, certified "DNA-free" reagents and labware were purchased (e.g., water and plastic tubes). All other reagents were opened solely within a sterilized hood within the ancient DNA facility. All chemicals were prepared for the extraction with previously unopened DNA- and RNA-free certified water (Ultrapure water; Invitrogen). Briefly, 1.8 ml of 0.5 ethylenediaminetetraacetic acid (EDTA; Life Tech), 100 µl of 10% sodium dodecyl sulphate (SDS; Life Tech) and 20 µl of 20 mg/ml proteinase K (proK; Life Tech) were added to each sample, and the mixture was rotated at 55°C overnight to decalcify the sample. Released DNA was then purified by adding silica (silicon dioxide; Sigma Aldrich) and 3 ml of binding buffer (e.g., QG buffer; Qiagen; modified to contain 5.0 m GuSCN; 18.1 mm Tris-HCl; 25 mm NaCl; 1.3% Triton X-100; Rohland & Hofreiter, 2007). The silica was pelleted, washed twice in 80% ethanol, dried and resuspended in 100 µl of TLE buffer (10 mm Tris, 1 mm EDTA, pH 8) twice to elute the DNA, which was then stored at −20°C until amplification. All chemicals were prepared for the extraction with previously unopened DNA- and RNA-free certified water (Ultrapure water; Invitrogen). For QG extractions performed in the modern laboratories, unopened aliquots of DNA extraction reagents were transported to the modern laboratory, and the modern samples were extracted following the ancient DNA approach described above.

In contrast to ancient DNA extractions, many modern microbiome studies decrease cost and time by using commercial DNA extraction kits to isolate DNA. To compare the nature and extent of contaminant DNA in the ancient method to a typical commercial microbiome DNA extraction kit, we analysed an additional set of EBCs created during extractions using a PowerBiofilm DNA Isolation Kit (MOBIO) from concurrent oral microbiome research conducted in the same modern labs (referred to as "kit" EBCs).

2.4 | Library preparation

To minimize additional variables, a simple 16S rRNA amplicon sequencing approach was used in this study to compare the different sample types. Briefly, the V4 region of the bacterial 16S rRNA encoding gene was targeted for amplification using degenerate Illumina fusion primers, as previously described (Caporaso et al., 2012): forward primer 515F (AATGATACGGCGACCACCGAGATCTACACTATG GTAATTGTGTGCCA GCMGCCGCGGTAA) and barcoded reverse primer 806R (CAAGCAGAAGACGGCATACGAGATnnnnnnnnnnnn AGTCAGTCAGCCGGACTACHVGGGTW TCTAAT) (Caporaso et al.,

2012). The string of n's in the reverse primer refers to the unique 12-bp barcode used for each sample. Primers were resuspended in TLE buffer within the ancient facility and distributed to the modern laboratory. In both facilities, all PCR amplification reactions were prepared using ultraclean reagents with strict ancient DNA protocols (Cooper & Poinar, 2000). Each PCR contained 17.25 µl DNA-free water (Ultrapure water; Invitrogen), 2.5 µl 10× reaction buffer (20 mm Tris-HCl, 10 mm $(NH_4)_2SO_4$, 10 mm KCl, 2 mm $MgSO_4$, 0.1% Triton X-100, pH 8.8 at 25°C; ThermoPol Buffer; New England Biolabs), 0.25 µl Taq polymerase (Platinum Taq DNA Polymerase High Fidelity; Thermo Fisher Scientific), 1.0 µl $MgCl_2$ (Thermo Fisher Scientific), 1.0 µl of each primer at 10 µm (IDT) and 2.0 µl of genomic DNA; each reaction was performed in triplicate. 16S rRNA amplification occurred under the following conditions: 95°C for 5 min; 37 cycles of 95°C for 0.5 min, 55°C for 0.5 min and 75°C for 1 min; and 75°C for 10 min. NTC reactions were also included in triplicate. PCR products were quantified (QuBit; Thermo Fisher Scientific) and pooled in batches of 30 samples at equal nanomolar concentrations prior to purification (Ampure; New England Biolabs). Each pool of purified PCR products was quantified (TapeStation; Agilent) before being combined into a single library. All amplicons were sequenced using the Illumina MiSeq 2x150-bp (300 cycle) kit.

2.5 | Bioinformatics analysis

After sequencing, fastq files for the forward and reverse reads were created using the Illumina CASAVA pipeline (version 1.8.2). Overlapping forward and reverse reads were joined based on a maximum of 5% nucleotide difference over a minimum 5-bp overlap using BBMERGE (sourceforge.net/projects/bbmap/). Only successfully merged sequences were used in downstream analyses. The resulting fastq file was then imported into QIIME1 (MACQIIME version 1.8.0), a bioinformatics pipeline-based software for the analysis of metagenomic data (Caporaso et al., 2010). All further analysis of the amplicon data sets was conducted within the QIIME1 package. Libraries were demultiplexed using a Phred base quality threshold of ≤20, with no errors allowed in the barcodes. Operational taxonomic units (OTUs) were determined by clustering sequences at 97% similarity using UCLUST (Edgar, 2010), and representative sequences (i.e., cluster seed) were selected for each cluster. By default, clusters with fewer than five sequences were eliminated from the analysis to reduce noise and spurious findings. Lastly, 16S rRNA gene sequences were given taxonomic assignments using the GREENGENES 13_8 database if the sequence was at least 80% similar (DeSantis et al., 2006; Wang, Garrity, Tiedje, & Cole, 2007). Taxonomic diversity measurements (alpha- and beta-diversity) and statistical analyses were performed and visualized in QIIME1. Samples were rarefied to a minimum of 150 sequences (see Figure 2) and a maximum of 1,000 sequences for diversity analyses, as many controls contained low sequence counts. Statistical differences between groups were identified using a PERMANOVA test for beta diversity (adonis), nonparametric *t*-test for alpha diversity (Monte Carlo), or Kruskal–Wallis and G-tests for detection of specific taxa associated with different treatments.

3 | RESULTS

3.1 | Low bacterial diversity is routinely obtained from laboratory extraction controls

The EBCs and NTCs were sequenced alongside the ancient and modern biological samples; all sample types were pooled together at equimolar concentrations. Despite the equimolar pooling, we routinely obtained fewer reads from control samples (EBCs and NTCs) compared to the dental calculus samples, probably due to poor amplification of control samples, the quantification of poor DNA libraries and the clean-up strategy employed. Compared to the ancient and modern calculus samples, 6.4-fold fewer reads on average were obtained from EBCs, and 7.6-fold fewer were obtained from NTCs (Figure 1a). As well as containing fewer reads overall, the control samples contained fewer taxa that could be identified than the biological samples. In the ancient laboratory, 719 total OTUs were observed in ancient biological samples

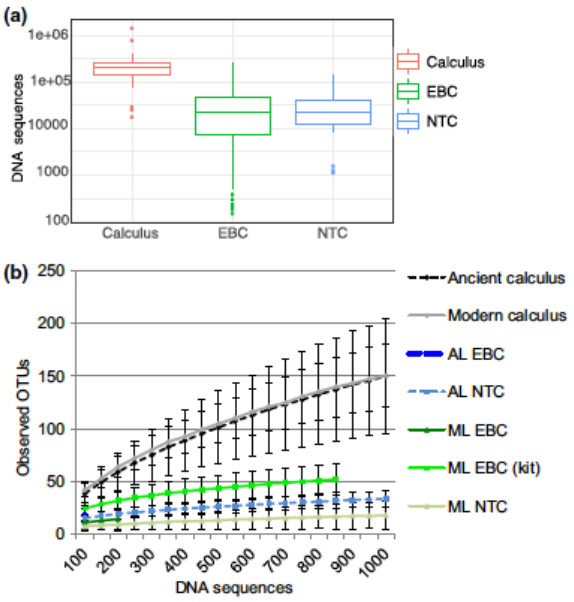

FIGURE 1 Lower diversity is observed in EBCs and NTCs compared to biological samples. (a) The number of sequenced reads from samples that were all pooled at equimolar concentrations is displayed on a box and whisker plot. (b) The alpha diversity of each type sample (i.e., the within sample diversity) was calculated using observed species metrics in QIIME1 for rarefied 16S rRNA data. Each sample was rarefied up to 1,000 sequences in 100 sequence intervals; the standard error at each subsampling event is displayed using error bars. Calculus samples are shown in blue, while control samples (extraction blank controls [EBCs] and no-template controls [NTCs]) from the ancient laboratory (AL) and the modern laboratory (ML) are shown in red and green, respectively [Colour figure can be viewed at wileyonlinelibrary.com]

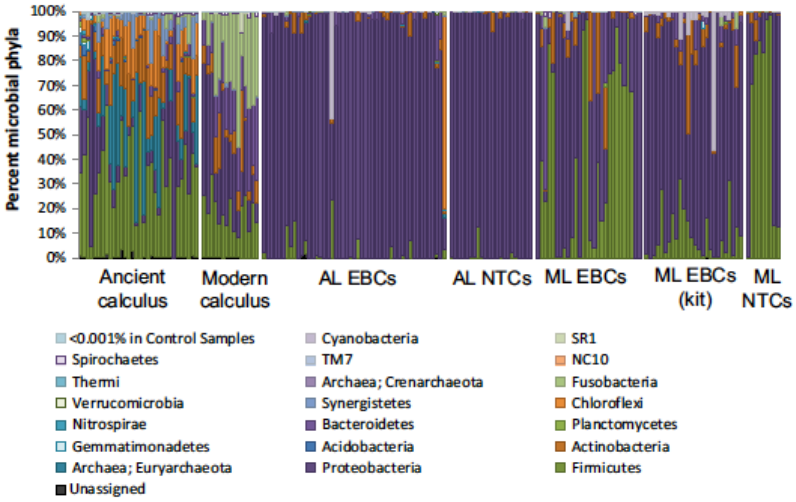

FIGURE 2 Microbial phyla within controls are distinct from biological samples. The proportion of different microbial phyla are shown for a wide array of modern and ancient calculus samples and control samples (EBCs and NTCs) from both laboratory facilities (modern lab [ML] and ancient lab [AL]) and two different extraction methods: the method employed in ancient DNA research and a commercially available DNA extraction kit (kit). Rare phyla were collapsed if they represented <0.001% of the total phyla observed [Colour figure can be viewed at wileyonlinelibrary.com]

(calculus), while only 415 were identified in the EBCs and 228 in NTCs (Figure 1b). In the modern laboratories, 286 total OTUs were described in the modern calculus samples, versus 208 in the EBCs and 102 in the NTCs. The OTU diversity within the EBCs and NTCs is reminiscent of the diversity observed in modern and ancient biological specimens, potentially reflecting minor cross contamination during DNA extraction. Across different extraction methods, the EBCs for the commercial extraction kit contained 261 OTUs, around 25% more than the QG method conducted in the modern laboratory. Overall, the laboratory controls were largely dominated by a single phylum, Proteobacteria (Figure 2), and alpha diversity was significantly lower than in the biological samples extracted within the same laboratory (Monte Carlo; $p \leq 0.0001$ and $T = >11.0$ in all comparisons between any group of controls and all biological samples). While the diversity within laboratory controls was considerably lower than the biological samples, these results demonstrate microbial DNA contamination within an ultraclean laboratory with "DNA-free" reagents, and clearly highlight the need to routinely monitor and report background contamination within all research facilities.

3.2 | EBCs detect >50% more contaminant taxa than NTCs

Many studies, including some in palaeomicrobiological research, reported failed EBC and NTC amplification reactions (often via simple visual comparison on an agarose gel) as a means to determine that their samples are free from contamination (Aagaard et al., 2014; Santiago-Rodriguez et al., 2015). This approach is clearly inadequate, and importantly, also fails to appreciate the extent of contamination introduced during the extraction process, even though this issue is well described in the literature (Kliman, 2014; Lauder et al., 2016; Weyrich, Llamas, & Cooper, 2014). In our comparisons, EBCs were taxonomically far more diverse than NTCs (Figure 1b) and contained more microbial genera (415 vs. 228 genera in the ancient lab, and 208 vs. 102 genera in the modern labs). This pattern suggests that if just NTCs were used to monitor the presence of laboratory contamination, at least 53% of the total laboratory contamination may go undetected. These results highlight the need for the standard reporting of both EBCs and NTCs in both modern and ancient metagenomics research.

We also examined the impact of overall laboratory contamination on ancient samples by bioinformatically filtering (removing) all contaminant OTUs from ancient dental calculus samples. For the ancient samples prepared with the specialized facility, an average 92.5% of the sequence reads were contaminants, but importantly, accounted for only 28% of the genera identified within these samples. This indicates that endogenous signal can be identified even in samples of low endogenous content once contaminant taxa are removed.

3.3 | EBCs and NTCs reflect laboratory environment

Previous studies have detected differences in the contaminants present in different laboratory facilities (Salter et al., 2014). In our study, the laboratory environments explained more of the taxonomic diversity observed in the EBCs and NTCs than the extraction or amplification methods used to generate them (Figure 3). For example, Proteobacteria dominated the EBCs and NTCs from the ancient laboratory, while Firmicutes were more dominant in EBCs and NTCs from the modern laboratories. In fact, different types of controls (i.e., EBC or NTC) from the same laboratory clustered with others of the same sample type in a principle coordinates analysis (PCoA) of unweighted UniFrac values ($p \leq 0.001$, $R^2 = 0.083$; Figure 3a), despite large variation and significant differences in each lab (Figure 1b). Despite the sample type (e.g., EBC or NTC) driving the majority of the signal, taxa distinguishing each laboratory could also be detected, as a *Paenibacillus* taxa was only found in the modern laboratories, while the ancient laboratory contained both bacterial (*Comamonas*, *Pseudomonas*, *Acinetobacter*, *Enterobacter*)

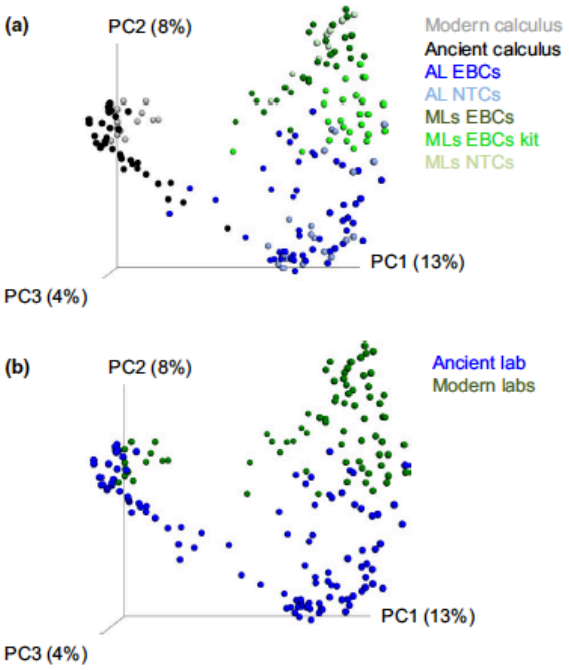

FIGURE 3 PCoA plots of control samples highlight differences in method and laboratory. PCoA plots of unweighted UniFrac values were plotted in QIIME1 to compare beta diversity differences (between samples differences) in all samples (a) or in different laboratories (b). The different laboratory facilities are represented by ML (modern lab) and AL (ancient lab), and the two control types are represented by EBC (extraction blank control) or no-template control (NTC) [Colour figure can be viewed at wileyonlinelibrary.com]

and archaeal (*Methanobrevibacter*) taxa that were not observed in the modern labs. In addition, several bacterial taxa were identified in both lab types, but were significantly increased in one location. The ancient laboratory contained significantly higher levels of certain *Acinetobacter*, *Comamonas* and *Pseudomonas* taxa compared to the modern laboratories (Kruskal–Wallis; Bonferroni-corrected $p \leq 0.05$), while Erythrobacteraceae and *Staphylococcus* taxa were increased in abundance in the modern laboratories. With the exception of the *Staphylococcus* taxa, each of these taxa had been previously identified in laboratory reagents (Salter et al., 2014). This suggests that some contaminant taxa are relatively universal across laboratories and are therefore either introduced in the manufacturing of laboratory reagents and labware or have a fundamental niche in low-nutrient, laboratory environments.

We next examined the genera that were likely to be in the reagents themselves, rather than the laboratories, by looking for shared taxa within the EBCs generated during extractions in both the ancient lab and the modern labs. Of the 69 dominant genera (i.e., observed at >0.1%), 17 were present in the reagents used in the in-house QG DNA extraction process used in both types of facility. These included *Cloacibacterium*, *Flavobacterium*, *Paenibacillus*, *Novosphingobium*, *Sphingomonas*, *Limnohabitans*, *Tepidomonas*, *Cupriavidus*, *Ralstonia*, *Acinetobacter*, *Enhydrobacter*, *Pseudomonas* and *Stenotrophomonas* taxa and four unidentified genera within Comamonadaceae, Erythrobacteraceae, Enterobacteriaceae and Pseudomonadaceae (Table 1). Within the ancient laboratory EBCs, the 26 most dominant genera included *Acinetobacter* (39%), followed by three genera within the family Comamonadaceae (totalling 11.3%), *Pseudomonas* (8%), *Novosphingobium* (1.5%), *Ralstonia* (1%), *Cloacibacterium* (1%) and others (Table 1). In the EBCs from the modern laboratories, *Paenibacillus* was the most prevalent of the 43 dominant genera (46%), while two Erythrobacteraceae (16.5%), Comamonadaceae (6.1%), *Cloacibacterium* (3.9%), *Corynebacterium* (2.5%), *Enterococccus* (2.5%), *Staphylococcus* (2.2%), *Enhydrobacter* (1.8%), Microbacteriaceae (1.7%), a Pseudomonadaceae (1.4%), *Ralstonia* (1.3%) and *N09* (1.2%) taxa were the next most prevalent within the reagents (Table 1). Although the same extraction method and reagents were used, only three of the most dominant taxa (i.e., identified at >1% prevalence) were the same within both laboratories (Comamonadaceae, *Cloacibacterium*, Pseudomonadaceae), highlighting the heterogeneity of taxa identified with EBCs. While many of these taxa have been previously identified as laboratory contaminants, the diversity within the modern laboratories also includes some human-associated taxa that have been cultured from the oral cavity, gut and skin (e.g., *Corynebacterium*, *Enterococcus* and *Staphylococcus*, respectively). This suggests that the additional precautionary measures used within the ancient laboratory may help to reduce the introduction of human-associated microorganisms in metagenomic data sets.

3.4 | DNA extraction kits contain microbiota indicative of the human mouth

We compared the diversity of taxa present within EBCs from the widely used ancient DNA extraction method and the commercial PowerBiofilm DNA Isolation Kit, used in the same modern laboratory. While the latter kit has been shown to have the lowest bacterial background contamination of standard microbiome kits (Salter et al., 2014), microbial diversity within the kit EBCs was significantly higher than the QG method (Figure 1b), suggesting that kit-based DNA extractions are more prone to background contamination. On a PCoA plot constructed using unweighted UniFrac distances, the kit EBCs clustered away from the QG EBCs and NTCs, including those processed in the same laboratory (adonis; $p \leq 0.001$, $R^2 = 0.04$; Figure 4a), demonstrating that a unique microbial community profile originates from the kit. This profile was not solely dominated by Firmicutes, similar to the other control samples from the modern lab, but contained taxa from several unique phyla (Acidobacteria, Gemmatimonadetes and Verrucomicrobia). These unique phyla included 15 distinct taxa that were also not observed in the extractions using the ancient DNA extraction method, including *Alicyclobacillus* ($n = 9$), *Halomonas*, *Pseudonocardia*, *Vogesella*, *Allobaculum* ($n = 2$) and *Akkermansia* taxa (Kruskal–Wallis; $p \leq 0.05$; Table 2). Several of these taxa are known to be resistant to sterilization treatments, including pasteurization (Chang & Kang, 2004). In addition,

TABLE 1 Dominant contaminant genera are largely unique within each laboratory

Genera taxonomy	AL EBC	ML EBC	ML EBC (kit)	AL NTC	ML NTC	Identified previously (reference)
Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Actinomycetaceae;g__Actinomyces	0.0002426	0.0011594	0.0028975	9.89E−05	1.33E−05	
Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Actinomycetaceae;g__N09	0	0.012119	0	0	0	
Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Corynebacteriaceae;g__Corynebacterium	0.0002939	0.0254723	0.0104781	0.001034	0.0003335	3
Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Dermacoccaceae;g__Dermacoccus	0	0.0059866	0	0	0.0039892	
Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Microbacteriaceae;g__	0.0006442	0.0172251	0.0002926	5.38E−05	0	
Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Micrococcaceae;g__	6.30E−05	0.0016361	0.0024467	2.50E−06	0.0001067	
Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Micrococcaceae;g__Micrococcus	0.0002292	0.0021648	0.0028065	2.50E−06	4.00E−05	3
Bacteria;p__Bacteroidetes;c__[Saprospirae];o__[Saprospirales];f__Chitinophagaceae;g__Sediminibacterium	0.0017288	0	2.17E−06	0	0	
Bacteria;p__Bacteroidetes;c__Cytophagia;o__Cytophagales;f__Cytophagaceae;g__	0.0017301	0	0.0008842	1.25E−06	0	
Bacteria;p__Bacteroidetes;c__Flavobacteriia;o__Flavobacteriales;f__[Weeksellaceae];g__	6.68E−06	0.0014925	0	5.01E−06	6.67E−06	
Bacteria;p__Bacteroidetes;c__Flavobacteriia;o__Flavobacteriales;f__[Weeksellaceae];g__Chryseobacterium	0.0005711	0.0049479	0.0001279	0	0	3
Bacteria;p__Bacteroidetes;c__Flavobacteriia;o__Flavobacteriales;f__[Weeksellaceae];g__Cloacibacterium	0.0101374	0.0399393	0.0060247	0.0007686	0.0020747	
Bacteria;p__Bacteroidetes;c__Flavobacteriia;o__Flavobacteriales;f__[Weeksellaceae];g__Wautersiella	0.0024707	8.33E−05	0.0001127	0.0014559	0	
Bacteria;p__Bacteroidetes;c__Flavobacteriia;o__Flavobacteriales;f__Cryomorphaceae;g__Fluviicola	0.0010454	0	0	0	0	
Bacteria;p__Bacteroidetes;c__Flavobacteriia;o__Flavobacteriales;f__Flavobacteriaceae;g__Flavobacterium	0.0026519	0.0040362	0.0003251	0	2.00E−05	3, 4
Bacteria;p__Bacteroidetes;c__Sphingobacteriia;o__Sphingobacteriales;f__;g__	0.0017217	0	6.28E−05	0	0	
Bacteria;p__Bacteroidetes;c__Sphingobacteriia;o__Sphingobacteriales;f__Sphingobacteriaceae;g__	0	0.0025812	0	0	0	
Bacteria;p__Bacteroidetes;c__Sphingobacteriia;o__Sphingobacteriales;f__Sphingobacteriaceae;g__Pedobacter	0.001991	2.08E−06	0	0	6.67E−06	3
Bacteria;p__Cyanobacteria;c__Chloroplast;o__Streptophyta;f__;g__	0.009702	0.0007785	0.0511164	2.50E−06	6.67E−06	
Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__[Thermicanaceae];g__Thermicanus	0	0.0028455	0	0	0	
Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae;g__Bacillus	0.0001536	0.009569	0.0017424	2.75E−05	6.67E−06	3
Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae;g__Geobacillus	0	0.0027123	0.0001647	0	0	
Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Paenibacillaceae;g__Paenibacillus	0.0012621	0.4657476	0.0001214	2.13E−05	0.8236393	3
Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus	0.0008838	0.0223562	0.0026417	0.0018664	0.008759	

(Continues)

TABLE 1 (Continued)

Genera taxonomy	AL EBC	ML EBC	ML EBC (kit)	AL NTC	ML NTC	Identified previously (reference)
Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Aerococcaceae;g__	0.0003678	0.0063967	0.0010077	1.25E−06	0	
Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Enterococcaceae;g__Enterococcus	0	0.0249665	2.82E−05	0	8.01E−05	
Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Leuconostocaceae;g__Leuconostoc	0.0010562	0.0006328	0	0	0	
Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Lactococcus	0.0038977	0	0.0039615	0	0	
Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus	0.0009439	0.004338	0.0226336	4.01E−05	0.0038825	3
Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae;g__Clostridium	0.0001019	0.0010491	0.0029256	0.0008788	0.0010207	
Bacteria;p__Planctomycetes;c__Planctomycetia;o__Pirellulales;f__Pirellulaceae;g__	0.000177	0.002577	0	0	6.67E−06	
Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__;g__	0.0011802	0.0005183	0.0009384	0.0011642	6.67E−06	
Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Bradyrhizobiaceae;g__	0.0018115	4.16E−06	0.016817	0.0011679	0	
Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Bradyrhizobiaceae;g__Bradyrhizobium	0.0062786	0.0003997	0.0039659	0.0136547	0.0016944	3, 4
Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Hyphomicrobiaceae;g__Devosia	0.0010312	0	0.0062825	0	0	3
Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Methylobacteriaceae;g__Methylobacterium	0.0217428	0.0001624	0.0011984	0.0465498	0	3, 28
Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Phyllobacteriaceae;g__Mesorhizobium	0.0056686	0.0005516	0.0005981	0	0	3
Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodospirillales;f__Acetobacteraceae;g__	4.17E−07	0.0013967	0.0015538	0	0.0070245	
Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rickettsiales;f__mitochondria;Other	0.0015063	0	0.0012071	0.0004757	0	
Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__Erythrobacteraceae;g__	0.0007168	0.0851929	0.0092212	0	0.0590982	
Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__Erythrobacteraceae;Other	0	0.0014925	8.67E−06	0	4.00E−05	
Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__Sphingomonadaceae;g__	0.0019877	0.0003268	0.001296	9.51E−05	6.67E−06	
Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__Sphingomonadaceae;g__Novosphingobium	0.0159422	0.0028788	0.0068785	0.0449963	0.0016477	3
Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__Sphingomonadaceae;g__Sphingobium	0.0024018	0	0.0099147	1.88E−05	0	3
Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__Sphingomonadaceae;g__Sphingomonas	0.0078676	0.0041132	0.0078797	0.0175166	0.0079785	3, 28, 4
Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__;f__;g__	0.000291	0.0011157	0	1.25E−06	0	
Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Alcaligenaceae;g__Achromobacter	0.0010253	2.08E−06	0	5.13E−05	0	

(Continues)

TABLE 1 (Continued)

Genera taxonomy	AL EBC	ML EBC	ML EBC (kit)	AL NTC	ML NTC	Identified previously (reference)
Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Burkholderia	0.0298854	0	0.0002817	0.0010202	0	3, 4
Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__	0.0272699	0.0616356	0.0149186	0.0043613	0.0167909	
Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Comamonas	0.1663902	0.0001811	0.0552752	0.1315213	0.0001001	3, 28
Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Limnohabitans	0.0070234	0.0013405	0.0029668	0.0106216	0	
Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Tepidimonas	0.0324847	0.0035241	0.0004833	0.0026751	0	
Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;Other	0.0548562	0.0040008	0.0253556	0.0721054	0.0019546	
Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Oxalobacteraceae;g__Cupriavidus	0.0020211	0.0014176	0.0005418	0.0003405	0	3, 28
Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Oxalobacteraceae;g__Ralstonia	0.0109828	0.0137093	0.0158006	0.0459339	0.0092727	3, 28, 4, 29
Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Oxalobacteraceae;Other	0.0025362	0.0001103	0.0016362	0.0180573	0	
Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Methylophilales;f__Methylophilaceae;g__	0.0020728	0	0	0	0	
Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Rhodocyclales;f__Rhodocyclaceae;g__	0.0002296	0.0019713	0.0029213	0.0010478	0	
Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Aeromonadales;f__Aeromonadaceae;g__	7.81E−05	0.001607	0.0040504	1.25E−06	0	
Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__	0.0192299	0.0828282	0.0399578	0.0364377	0.0001668	
Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;Other	0.0147874	1.25E−05	0.0053984	0.030091	0.0043762	
Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Moraxellaceae;g__Acinetobacter	0.3930931	0.0016528	0.2645106	0.4111166	0.0002335	3, 2, 28
Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Moraxellaceae;g__Enhydrobacter	0.001156	0.0181764	0.0005028	0.0026776	0.0194392	3
Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__	0.0001904	0.0052768	0.0060052	7.51E−05	6.67E−05	
Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas	0.0813901	0.0048709	0.0253902	0.0820035	0.0087723	3, 29, 4
Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;Other	0.0052716	0.0140465	0.0001777	5.01E−06	0.0017144	
Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales;f__Xanthomonadaceae;g__	0.0054783	0.0003768	0.0018919	0.0016449	0	
Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales;f__Xanthomonadaceae;g__Lysobacter	0.0018641	2.08E−06	0	0	0	
Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales;f__Xanthomonadaceae;g__Stenotrophomonas	0.0018086	0.0014051	0.0007607	0.0009689	0.0003869	3, 2, 28, 29, 4

Notes. The 69 genera that dominated EBC control samples are displayed for all sample types and include the proportion identified in each sample type. Genera were identified as dominant if they were found to be above 0.01% of the total genera identified within each laboratory. Taxa highlighted in green represent genera that dominated EBCs in the ancient laboratory, while unhighlighted are those from the modern EBC samples. If the genera were identified in previous studies that examined contamination, the reference number is shown in the right-hand column.

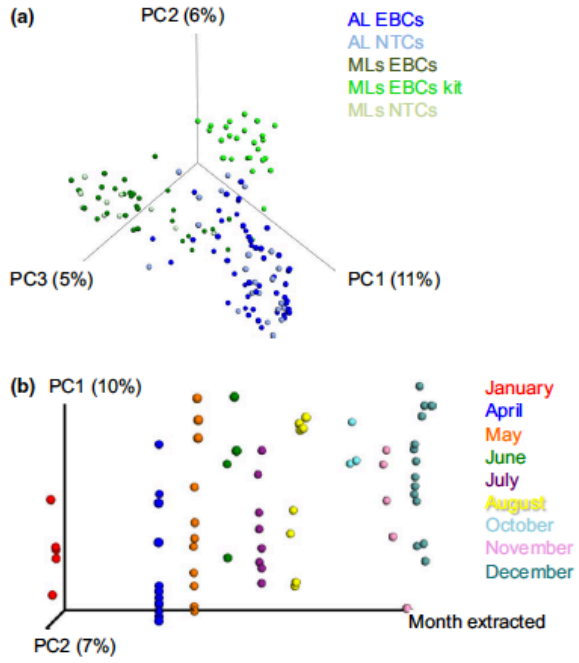

FIGURE 4 PCoA of the extraction method and seasonal variation in contaminant communities. The modern and ancient calculus samples were removed from the analysis presented in Figure 3, and a PCoA plot was constructed of only control samples to identify differences between the extraction method and laboratory in control samples (a). (b) UniFrac values from control samples (EBCs and NTCs) from the ancient laboratory over a 5-year period (2012–2016) are coloured on a PCoA plot according to month [Colour figure can be viewed at wileyonlinelibrary.com]

several OTUs were more likely to be found in higher abundances in the kit EBCs than any other control samples (G-test; $p \leq 0.05$) and include specific Bradyrhizobiaceae, *Neisseria*, *Corynebacterium*, *Fusobacterium*, *Streptococcus*, *Micrococcus* and *Halomonas* taxa. While *Bradyrhizobium* and *Micrococcus* have previously been identified as laboratory contaminants (Laurence et al., 2014; Salter et al., 2014), the remaining taxa are commonly found in the human mouth. Concerningly, many of these human oral taxa have been previously reported from low-biomass samples, such as placenta and tumour tissue, which were examined without EBCs (Aagaard et al., 2014; Hieken et al., 2016). This suggests that DNA extraction kits used in modern molecular biology laboratories may be contributing unique microbial signals in addition to those generated within the laboratory environment.

3.5 | Contaminant taxa change over time

Much of the variation identified in this study is laboratory-specific. In order to test how seasonal changes, different researchers, or time might alter the microbial diversity observed in controls, we assessed the EBC and NTC records from the ancient lab facility over 5 years (2012–2016). Bacterial community structure in the ancient lab was linked to the researcher (adonis; $p = 0.001$, $R^2 = 0.073$), the extraction year (adonis; $p \leq 0.01$, $R^2 = 0.022$), the extraction month (adonis; $p \leq 0.001$, $R^2 = 0.044$; Figure 4b), and wet/dry seasons (adonis; $p = 0.001$, $R^2 = 0.081$). However, each of these signals was less significant and drove less variation within the data set when compared to the differences observed between laboratory facilities or between extraction methods. Very few specific taxa were significantly associated with temporal variation, although linked changes in overall diversity were observed. In total, 32 OTUs were associated with the month in which the extraction was performed and were largely present during dry months (October–January; dominated by Comamonadaceae (2), Bradyrhizobiaceae (11), and Gemmatimonadetes (2) taxa; Kruskal–Wallis; Bonferroni corrected $p \leq 0.05$), while only two OTUs (*Thermobispora* and Actinomycetales taxa) were linked to wet seasons. Interestingly, five OTUs (*Leptotrichia*, Comamonadaceae (3) and *Burkholderia*) were also associated with the lab researcher (Kruskal–Wallis; Bonferroni-corrected $p \leq 0.05$). While we cannot rule out the confounding nature of these variables (e.g., links between different researchers being more active in the laboratories at different times), these observations suggest that contaminant taxa change over time and need to be continually monitored, even in the cleanest molecular facilities.

4 | DISCUSSION

4.1 | Overview

While several studies have now reported on contaminant DNA within laboratory reagents, the systematic inclusion of EBCs has not yet been widely embraced in metagenomic research. Several studies on human microbiota have been criticised for their lack of careful controls (Eisenhofer, Cooper, & Weyrich, 2017; Kliman, 2014; Lauder et al., 2016), as the unfounded results of such studies have potentially serious repercussions and have hindered scientific progress. A similar phenomenon occurred with the new field of ancient DNA in the early 1990s, when research teams, reviewers and editors failed to adequately test for contamination (Austin, Smith, Fortey, & Thomas, ; Beckenbach, 1995; Priest, 1995), leading to many spurious results. This seriously undermined the credibility of ancient DNA research (Weyrich et al., 2014) and resulted in the formation of a robust set of guidelines (Cooper & Poinar, 2000). Here, we surveyed the largest collection of extraction blank and no-template amplification negative control samples to date ($n = 144$) with the goal of better describing contaminant DNA in microbiome studies to avoid pitfalls similar to those observed in the ancient DNA field.

4.2 | Contaminant diversity remains underestimated

We identified 861 contaminant taxa over 5 years within a single ultraclean laboratory facility. Before this publication, the largest collection of contaminant taxa was published by Salter et al. (2014) and included 93 contaminant genera. Within our study, we found 71 of the taxa identified by Salter *et al.* across all labs and

TABLE 2 Extraction methods contain unique taxa

OTU Taxonomy	Mean sequences/sample	
	Kit	QG
Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales	6.5357143	0.0086957
Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Alicyclobacillaceae;g__Alicyclobacillus;s__	2.8214286	0
Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Alicyclobacillaceae;g__Alicyclobacillus;s__	7.2142857	0
Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Alicyclobacillaceae;g__Alicyclobacillus;s__	1.9285714	0
Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Alicyclobacillaceae;g__Alicyclobacillus;s__	2.0714286	0
Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Alicyclobacillaceae;g__Alicyclobacillus;s__	1.7142857	0
Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Oceanospirillales;f__Halomonadaceae;g__Halomonas;s__	0.0357143	0
Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Pseudonocardiaceae;g__Pseudonocardia;s__	0.0357143	0
Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Moraxellaceae;g__Acinetobacter;s__	385.28571	0.026087
Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Bradyrhizobiaceae;g__;s__	93.428571	0
Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Bradyrhizobiaceae;g__;s__	91.857143	11.721739
Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Alicyclobacillaceae;g__Alicyclobacillus;s__	0.8571429	0
Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Alicyclobacillaceae;g__Alicyclobacillus;s__	0.8571429	0
Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae;g__Vogesella;s__	184.46429	0
Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Bradyrhizobiaceae	2.3571429	0.0086957
Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Moraxellaceae;g__Acinetobacter;s__	395.39286	0.626087
Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__Sphingomonadaceae;g__Sphingobium;s__	160.75	33.243478
Bacteria;p__Firmicutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae;g__Allobaculum;s__	0.6071429	0
Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Alicyclobacillaceae;g__Alicyclobacillus;s__	0.4285714	0
Bacteria;p__Verrucomicrobia;c__Verrucomicrobiae;o__Verrucomicrobiales;f__Verrucomicrobiaceae;g__Akkermansia;s__muciniphila	0.5	0
Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Moraxellaceae;g__Acinetobacter;s__	82.142857	0.7391304
Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Bradyrhizobiaceae	1.9642857	1.7217391
Bacteria;p__Firmicutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae;g__Allobaculum;s__	0.4642857	0
Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Alicyclobacillaceae;g__Alicyclobacillus;s__	0.3928571	0
Bacteria;p__Cyanobacteria;c__Chloroplast;o__Streptophyta;f__;g__;s__	2	0.0086957
Bacteria;p__Cyanobacteria;c__Chloroplast;o__Streptophyta;f__;g__;s__	26.107143	1.5478261
Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales	1.9642857	1.9913043
Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__;s__	5.3928571	23.46087
Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas;s__	0	0.0086957
Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__;s__	0	0.0086957
Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas;s__	0	0.0086957
Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__;s__	0	0.0086957
Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__;s__	5.5714286	44.6
Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Oceanospirillales;f__Alcanivoracaceae;g__Alcanivorax;s__	0	0.0173913
Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__;s__	4.7857143	20.730435
Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__;s__	4.9285714	23.373913

(Continues)

TABLE 2 (Continued)

OTU Taxonomy	Mean sequences/sample	
	Kit	QG
Bacteria;p__Proteobacteria;c__Epsilonproteobacteria;o__Campylobacterales;f__Helicobacteraceae;g__;s__	0	0.0173913
Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__;s__	0	0.0173913
Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Oceanospirillales;f__Halomonadaceae;g__ Halomonas;s__	0	0.026087
Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__;s__	1.8928571	4.3391304
Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Oxalobacteraceae;g__Cupriavidus;s__	0	0.0347826
Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__;s__	1.1071429	5.9217391
Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__;s__	0	0.026087
Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Alteromonadales;f__Alteromonadaceae;g__ Marinobacter;s__	0	0.0347826

Notes. Operational taxonomic units identified as significantly (Kruskal–Wallis Bonferroni-corrected p-value <0.05) associated with one of the two extraction methods in the modern laboratory are listed. OTUs highlighted in green were significant within the QG method, while the nonhighlighted OTUs were significant in the kit extraction method.

methodologies. However, only 29.5% of the Salter *et al*. taxa (21 of their 71 taxa) were identified as dominant taxa within our study across all methods and labs. This indicates that laboratory microbial contamination is not yet well described and is likely to be unique across different laboratories, protocols, seasons and researchers. Of the 21 taxa shared across studies, four genera (*Ralstonia*, *Acinetobacter*, *Pseudomonas* and *Stenotrophomonas*) have now been routinely identified in at least four of the six publications that examine laboratory contamination (Barton, Taylor, Lubbers, & Pemberton, 2006; Glassing et al., 2016; Grahn, Olofsson, Ellnebo-Svedlund, Monstein, & Jonasson, 2003; Laurence et al., 2014; Salter et al., 2014; Tanner et al., 1998). All of these taxa are classified as Proteobacteria, as are 55% of the dominant contaminant taxa (38/69) identified within our study and 63% (34/92) within the Salter et al. study. Proteobacteria encompass several bacterial families that are known to be UV- and oxidation-resistant, suggesting that they may have a fundamental niche within laboratory settings. While contamination is highly diverse, this finding indicates that Proteobacteria appear to be the most widespread source of laboratory contamination.

4.3 | Analysing contaminants is critical for the successful interpretation of low-biomass samples

We identified several human oral microbiota taxa present in the commercial extraction kit, including *Fusobacterium*, *Streptococcus* and *Corynebacterium* (Chen et al., 2010), while previous studies have previously identified additional human oral taxa contaminants, including *Haemophilus* and *Peptostreptococcus* (Barton et al., 2006). Worryingly, one of these taxa in particular, *Fusobacterium*, has recently been identified both as a component of the "placental microbiome" and of breast cancer tissue, in low-biomass studies that did not consider background contamination from laboratory reagents or environments (Aagaard et al., 2014; Hieken et al., 2016; Kostic et al., 2012). It remains unclear whether this taxon is a laboratory contaminant, or whether it can escape the oral cavity and contribute to inflammatory processes elsewhere in the body. Other nonoral taxa identified within this study as contaminants have also previously been reported as important taxa within studies that failed to use controls (Mayneris-Perxachs et al., 2016). There is clearly a need for more detailed metagenomic studies, or the use of improved "oligotyping" 16S rRNA gene analysis methods of contaminant taxa, to better identify specific strain differences and determine whether such taxa are contaminants or are actually present in the body and can cause systemic disease. The lack of contaminant assessment has already negatively impacted the metagenomics field (Lauder et al., 2016), and it is critical that editors and reviewers are aware of this issue.

4.4 | Bacterial DNA is still obtained from ultraclean reagents in ultraclean facilities – no facility is contaminant free

Contaminant taxa were identified in EBCs and NTCs within five different laboratory facilities, including a state-of-the-art, ultraclean ancient DNA facility. In the latter, the specialized conditions and procedures did not prevent low levels of bacterial diversity, and a wide range of contaminant taxa was still observed – with the dominant taxa all known to resist disinfectant measures, including treatment with aromatic or oxidative compounds (i.e., bleach, *Acinetobacter*; Ridgway & Olson, 1982, *Comamonas*; Liu et al., 2015;2015 or other disinfectant compounds (*Pseudomonas*; Sagripanti & Bonifacino, 2000). These mechanisms of disinfection resistance have contributed to nosocomial infections in hospitals (i.e., *Acinetobacter*; Dent, Marshall, Pratap, & Hulette, 2010) and to contamination of cell culture reagents (e.g., *Achromobacter*; Gray, Birmingham, & Fenton, 2010). Of note, *Deinococcus*, a taxon that can notoriously survive UV irradiation (Krisko & Radman, 2013),

Alicyclobacillus, known to survive pasteurization (Chang & Kang, 2004), and other species known to degrade oxidative compounds (e.g., *Pasteurella*; Wackett, Logan, Blocki, & Bao-li, 1992) were not observed in the specialized ancient DNA facility, but were identified within the modern laboratory. While measures to reduce contamination have prevented the introduction of human-associated microorganisms into the ancient lab EBCs, these numerous strategies did not eliminate or completely prevent the introduction of bacterial contaminant DNA. This suggests that each research facility will probably contain unique microorganisms able to resist decontamination measures, although it is plausible that contaminant DNA could be routinely introduced into the facility from other sources and represents living species found elsewhere, rather than in the actual facilities utilized in this study. Regardless, this finding reiterates that every laboratory is susceptible to bacterial DNA contamination and that researchers should consistently monitor the contamination present within their own facility as a best practice.

4.5 | Nonkit and approaches provide unique contaminant signals

In this study, we identified several taxa in a commonly used DNA extraction kit that were absent in the home-made ancient DNA extraction method (QG). The home-made ancient DNA method was developed to obtain more DNA from samples with low endogenous DNA, and this and other similar extraction methods are now routinely applied in ancient DNA studies to examine ancient microbiota and metagenomes (Gilbert et al., 2008; Weyrich et al., 2017; Willerslev et al., 2003). In this study, the ancient DNA method produced extraction blanks that had lower microbial diversity and were less likely to contain human oral taxa than extraction blanks generated using a commercial kit. This suggests that commercially available kits may contain more DNA contamination than home-made methods that source clean materials. It is likely that the assembly of kit-based reagents in a separate facility provides an additional opportunity to contaminate reagents with laboratory DNA. Lastly, this suggests that ancient DNA extraction methods and strategies could be applied in modern low-biomass studies to potentially reduce contaminants that originate from humans.

In the future, studies of low biomass or low endogenous count routinely use shotgun sequencing to better identify contaminant taxa, as strain-level identifications increase specificity in tracking contaminants. In many cases, the ancient DNA field has now shifted to utilizing shotgun DNA sequencing as the gold-standard method (Weyrich et al., 2017). Shotgun sequencing also produces many other important molecular signals (e.g., signatures of ancient DNA damage), functional analysis and strain markers to delineate which species are endogenous and which are contaminants. For example, distinct strains within a single genus could be identified as either a contaminant or an endogenous species, which would be critical for examining oral species in low-biomass tissues. In addition, damage profiles of DNA contamination could be used to distinguish fragmented, extracellular DNA within reagents versus species living within the laboratory. Current approaches aimed at eliminating contamination in shotgun sequenced metagenomes have had varied levels of success (reviewed by Salter et al., 2014), and new bioinformatic tools and models will undoubtedly improve our ability to identify and account for contaminant signals within metagenomic data sets (Lu & Salzberg, 2018). However, the need to routinely monitor background contamination will always be necessary when examining low-biomass samples, even when other methodologies, such as shotgun metagenomic sequencing, are applied.

4.6 | Contamination assessment needs to be routinely reported as a publication requirement

Contaminant sequences introduced during sample processing and library construction significantly contribute to signals from biological samples, especially those that are low-endogenous or low-biomass in nature. This study confirms that contaminant taxa unique to the extraction method and facility are related to the material being extracted, and change over time within a single facility, although these levels of contamination can be somewhat mitigated by routine decontamination measures of the facility and potentially the reagents themselves (Borst, Box, & Fluit, 2004). Therefore, the presence of contaminants needs to be considered in all future studies of both human and environmental microbiota. We recommend that all researchers routinely record potential sources of contamination DNA (reagent batches or lot numbers; dates of extractions and amplifications; researchers performing such duties; etc.) and critically propose that researchers routinely include extraction blank controls during the extraction process to monitor the bacterial DNA introduced into their samples, as recently recommended by Eisenhofer et al. (2018). Minimally, one control should be included in at least every batch of extractions and amplifications performed. Adding carrier DNA into control samples may also improve contaminant DNA detection (Xu et al., 2009). If controls were not included in existing data sets, an assessment of previously identified contaminant taxa should also be minimally included in the published analysis. For example, researchers could report how many known contaminant taxa are present within a data set or provide evidence to demonstrate that the removal of known contaminants does not impact the sample signal or conclusions of the paper. To facilitate this process, we have included a text file that includes a list of all the contaminant taxa observed here, as well as a separate file of only the dominant taxa. The inclusion of negative extraction blank controls should be regarded as minimal requirements for any metagenomics research and should become standard requirements of reviewers and journal editors.

ACKNOWLEDGEMENTS

We thank Paul Gooding at the Australian Genomic Research Facility for technical help during DNA sequencing. This research was funded by the Australian Research Council (L.S.W; A.C.).

AUTHOR CONTRIBUTIONS

L.S.W. and A.G.F. conceived of the study. L.S.W., A.G.F., R.E., J.Y., C.S., M.H.D. and C.A. contributed samples and completed lab work. L.S.W., L.A. and J.B. completed bioinformatic analysis of the data. L.S.W. wrote the paper, and all authors edited and contributed to the final manuscript.

DATA ACCESSIBILITY

QIIME demultiplexed sequences (16SContam_seqs_forpub.fna), a phylogenetic tree of representative sequences (rep_set.tre), a biom table (otu_table_clean.biom), and sample metadata (SampleInformation_20180820.txt) can be accessed from https://figshare.com/account/articles/7283816 (https://doi.org/10.25909/5bdaa4431a941).

ORCID

Laura S. Weyrich https://orcid.org/0000-0001-5243-4634

REFERENCES

Aagaard, K., Ma, J., Antony, K. M., Ganu, R., Petrosino, J., & Versalovic, J. (2014). The placenta harbors a unique microbiome. *Science Translational Medicine*, *6*, 237ra65-237ra65. https://doi.org/10.1126/scitranslmed.3008599

Adler, C. J., Dobney, K., Weyrich, L. S., Kaidonis, J., Walker, A. W., Haak, W., … Cooper, A. (2013). Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the Neolithic and Industrial revolutions. *Nature Genetics*, *45*, 450–455. https://doi.org/10.1038/ng.2536

Austin, J. J., Smith, A. B., Fortey, R. A., Thomas, R. H. (1998). Ancient DNA from amber inclusions: A review of the evidence. *Ancient Biomolecules*, *2*(2), 167–176.

Barton, H. A., Taylor, N. M., Lubbers, B. R., & Pemberton, A. C. (2006). DNA extraction from low-biomass carbonate rock: An improved method with reduced contamination and the low-biomass contaminant database. *Journal of Microbiological Methods*, *66*, 21–31. https://doi.org/10.1016/j.mimet.2005.10.005

Beckenbach, A. T. (1995). Age of bacteria from amber. *Science*, *270*, 2015–2016; author reply 2016–2017.

Borst, A., Box, A., & Fluit, A. (2004). False-positive results and contamination in nucleic acid amplification assays: Suggestions for a prevent and destroy strategy. *European Journal of Clinical Microbiology and Infectious Diseases*, *23*(4), 289–299. https://doi.org/10.1007/s10096-004-1100-1

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, … G. A. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, *7*, 335–336.

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., … Knight, R. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME Journal*, *6*, 1621–1624. https://doi.org/10.1038/ismej.2012.8

Chang, S.-S., & Kang, D.-H. (2004). *Alicyclobacillus* spp. in the fruit juice industry: History, characteristics, and current isolation/detection procedures. *Critical Reviews in Microbiology*, *30*, 55–74.

Chen, T., Yu, W.-H., Izard, J., Baranova, O. V., Lakshmanan, A., & Dewhirst, F. E. (2010). The Human Oral Microbiome Database: A web accessible resource for investigating oral microbe taxonomic and genomic information. *Database (Oxford)*, *2010*, baq013.

Cooper, A., & Poinar, H. N. (2000). Ancient DNA: Do it right or not at all. *Science*, *289*, 1139–1139. https://doi.org/10.1126/science.289.5482.1139b

Dent, L. L., Marshall, D. R., Pratap, S., & Hulette, R. B. (2010). Multidrug resistant *Acinetobacter baumannii*: A descriptive study in a city hospital. *BMC Infectious Diseases*, *10*, 196. https://doi.org/10.1186/1471-2334-10-196

DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., … Andersen, G. L. (2006). Greengenes, a Chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and Environment Microbiology*, *72*, 5069–5072. https://doi.org/10.1128/AEM.03006-05

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

Eisenhofer, R., & Cooper, A., Weyrich, L. S. (2017). Reply to Santiago-Rodriguez et al.: Proper authentication of ancient DNA is essential. *FEMS Microbiology Ecology*, *93*(5). https://doi.org/10.1093/femsec/fix042

Eisenhofer, R., Minich, J., Marotz, L., Cooper, A., Knight, R., & Weyrich, L. S. (2018). Contamination in low-biomass microbiome studies: Issues and recommendations. *Trends in Microbiology*, *27*(2), 105–117. https://doi.org/10.1016/j.tim.2018.11.003

Gilbert, M. T. P., Jenkins, D. L., Go therstrom, A., Naveran, N., Sanchez, J. J., Hofreiter, M., … Willerslev, E. (2008). DNA from Pre-Clovis Human Coprolites in Oregon. *North America. Science*, *320*, 786–789. https://doi.org/10.1126/science.1154116

Glassing, A., Dowd, S. E., Galandiuk, S., Davis, B., & Chiodini, R. J. (2016). Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of microbiota in low bacterial biomass samples. *Gut Pathogens*, *8*, 24. https://doi.org/10.1186/s13099-016-0103-7

Grahn, N., Olofsson, M., Ellnebo-Svedlund, K., Monstein, H.-J Ã¼rG., & Jonasson, J. (2003). Identification of mixed bacterial DNA contamination in broad-range PCR amplification of 16S rDNA V1 and V3 variable regions by pyrosequencing of cloned amplicons. *FEMS Microbiology Letters*, *219*, 87–91. https://doi.org/10.1016/S0378-1097(02)01190-4

Gray, J. S., Birmingham, J. M., & Fenton, J. I. (2010). Got black swimming dots in your cell culture? Identification of *Achromobacter* as a novel cell culture contaminant. *Biologicals*, *38*, 273–277. https://doi.org/10.1016/j.biologicals.2009.09.006

Hieken, T. J., Chen, J., Hoskin, T. L., Walther-Antonio, M., Johnson, S., Ramaker, S., … Degnim, A. C. (2016). The microbiome of aseptically collected human breast tissue in benign and malignant disease. *Scientific Reports*, *6*, 30751. https://doi.org/10.1038/srep30751

Kearney, M. F., Spindler, J., Wiegand, A., Shao, W., Anderson, E. M., Maldarelli, F., … Coffin, J. M. (2012). Multiple Sources of Contamination in Samples from Patients Reported to Have XMRV Infection. *PLoS ONE*, *7*(2), e30889. https://doi.org/10.1371/journal.pone.0030889

Kliman, H. J. (2014). Comment on "The placenta harbors a unique microbiome". *Science Translational Medicine*, *6*, 254le4. https://doi.org/10.1126/scitranslmed.3009864

Knights, D., Kuczynski, J., Charlson, E. S., Zaneveld, J., Mozer, M. C., Collman, R. G., … Kelley, S. T. (2011). Bayesian community-wide culture-independent microbial source tracking. *Nat Meth*, *8*, 761–763. https://doi.org/10.1038/nmeth.1650

Kostic, A. D., Gevers, D., Pedamallu, C. S., Michaud, M., Duke, F., Earl, A. M., … Meyerson, M. (2012). Genomic analysis identifies association of *Fusobacterium* with colorectal carcinoma. *Genome Research*, *22*, 292–298. https://doi.org/10.1101/gr.126573.111

Krisko, A., & Radman, M. (2013). Biology of extreme radiation resistance: The way of *Deinococcus radiodurans*. *Cold Spring Harbor Perspectives in Biology*, *5*, a012765.

Lauder, A. P., Roche, A. M., Sherrill-Mix, S., Bailey, A., Laughlin, A. L., Bittinger, K., … Bushman, F. D. (2016). Comparison of placenta samples with contamination controls does not provide evidence for a distinct placenta microbiota. *Microbiome*, *4*, 29. https://doi.org/10.1186/s40168-016-0172-3

Laurence, M., Hatzis, C., & Brash, D. E. (2014). Common contaminants in next-generation sequencing that hinder discovery of low-abundance microbes. *PLoS ONE*, *9*(5), e97876. https://doi.org/10.1371/journal.pone.0097876

Liu, L., Zhu, W., Cao, Z., Xu, B., Wang, G., & Luo, M. (2015). High correlation between genotypes and phenotypes of environmental bacteria *Comamonas testosteroni* strains. *BMC Genomics*, *16*(1), 110. https://doi.org/10.1186/s12864-015-1314-x

Lu, J., & Salzberg, S. L. (2018). Removing contaminants from metagenomic databases. *Biorxiv*. bioRxiv 261859; https://doi.org/10.1101/261859

Mayneris-Perxachs, J., Bolick, D. T., Leng, J., Medlock, G. L., Kolling, G. L., Papin, J. A., … Guerrant, R. L. (2016). Protein- and zinc-deficient diets modulate the murine microbiome and metabolic phenotype. *American Journal of Clinical Nutrition*, *104*(5), 1253-1262.

Priest, F. G. (1995). Age of bacteria from amber. *Science*, *270*, 2015; author reply 2016–2017.

Ridgway, H. F., & Olson, B. H. (1982). Chlorine resistance patterns of bacteria from two drinking water distribution systems. *Applied and Environment Microbiology*, *44*, 972–987.

Rohland, N., & Hofreiter, M. (2007). Ancient DNA extraction from bones and teeth. *Nat Protocols*, *2*, 1756–1762. https://doi.org/10.1038/nprot.2007.247

Sagripanti, J. L., & Bonifacino, A. (2000). Resistance of *Pseudomonas aeruginosa* to liquid disinfectants on contaminated surfaces before formation of biofilms. *Journal of AOAC International*, *83*, 1415–1422.

Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., … Walker, A. W. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biology*, *12*, 87. https://doi.org/10.1186/s12915-014-0087-z

Santiago-Rodriguez, T. M., Fornaciari, G., Luciani, S., Dowd, S. E., Toranzos, G. A., Marota, I., & Cano, R. J. (2015). Gut microbiome of an 11th century A.D. pre-Columbian Andean mummy. *PLoS ONE*, *10*, e0138135. https://doi.org/10.1371/journal.pone.0138135

Tanner, M. A., Goebel, B. M., Dojka, M. A., Pace, N. R. (1998). Specific ribosomal DNA sequences from diverse environmental settings correlate with experimental contaminants. *Applied and Environment Microbiology*, *64*, 3110–3113.

Wackett, L. P., Logan, M. S. P., Blocki, F. A., & Bao-li, C. (1992). A mechanistic perspective on bacterial metabolism of chlorinated methanes. *Biodegradation*, *3*, 19–36. https://doi.org/10.1007/BF00189633

Wang, Q., Garrity, G. M., Tiedje, J. M., Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environment Microbiology*, *73*, 5261–5267.

Warinner, C., Speller, C., & Collins, M. J. (2015). A new era in palaeomicrobiology: Prospects for ancient dental calculus as a long-term record of the human oral microbiome. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *370*, 20130376. https://doi.org/10.1098/rstb.2013.0376

Weiss, S., Amir, A., Hyde, E. R., Metcalf, J. L., Song, S. J., & Knight, R. (2014). Tracking down the sources of experimental contamination in microbiome studies. *Genome Biology*, *15*, 564. https://doi.org/10.1186/s13059-014-0564-2

Weyrich, L. S., Dobney, K., & Cooper, A. (2015). Ancient DNA analysis of dental calculus. *Journal of Human Evolution*, *79*, 119–124. https://doi.org/10.1016/j.jhevol.2014.06.018

Weyrich, L. S., Duchene, S., Soubrier, J., Arriola, L., Llamas, B., Breen, J., … Cooper, A. (2017). Neanderthal behaviour, diet, and disease inferred from ancient DNA in dental calculus. *Nature*, *544*, 357–361. https://doi.org/10.1038/nature21674

Weyrich, L. S., Llamas, B., & Cooper, A. (2014). Reply to Santiago-Rodriguez et al.: Was luxS really isolated from 25- to 40-million-year-old bacteria?. *FEMS Microbiol Letters*, *353*, 85–86.

Willerslev, E., Hansen, A. J., Binladen, J., Brand, T. B., Gilbert, M. T. P., Shapiro, B., … Cooper, A. (2003). Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science*, *300*, 791–795. https://doi.org/10.1126/science.1084114

Xu, Z., Zhang, F., Xu, B., Tan, J., Li, S., & Jin, L. (2009). Improving the sensitivity of negative controls in ancient DNA extractions. *Electrophoresis*, *30*, 1282–1285. https://doi.org/10.1002/elps.200800473

How to cite this article: Weyrich LS, Farrer AG, Eisenhofer R, et al. Laboratory contamination over time during low-biomass sample analysis. *Mol Ecol Resour*. 2019;19:982–996. https://doi.org/10.1111/1755-0998.13011

Appendix V

Revegetation of urban green space rewilds soil microbiotas with implications for human health and urban design (Mills *et al*. 2020)

RESTORATION ECOLOGY
The Journal of the Society for Ecological Restoration

SER SOCIETY FOR ECOLOGICAL RESTORATION

RESEARCH ARTICLE

Revegetation of urban green space rewilds soil microbiotas with implications for human health and urban design

Jacob G. Mills[1,2], Andrew Bissett[3], Nicholas J. C. Gellie[1], Andrew J. Lowe[1], Caitlin A. Selway[4], Torsten Thomas[5], Philip Weinstein[1], Laura S. Weyrich[4,6], Martin F. Breed[1,7]

Many noncommunicable diseases are linked to degraded diversity in the human and environmental microbiota and are rising globally in epidemic proportions in industrialized urban populations. Reducing this disease burden may be aided by the ecological restoration of microbiota and their habitat in urban green spaces—a process termed microbiome rewilding. Microbiome rewilding could serve as a mechanism to increase urban exposure to biodiversity; biodiversity could introduce microbiota species or functional diversity to improve immune training and regulation in urban populations. As a first step in examining this hypothesis, we explored the microbial diversity and composition of a variety of urban green space vegetation types relative to urban revegetated woodlands of varying levels of vegetation diversity, including lawns, vacant lots, parklands, and remnant woodlands. We generated amplicon sequence variant community profiles from bacterial and archaeal 16S rRNA, fungal ITS1 region, and eukaryotic 18S rRNA marker genes. We also made trophic-mode predictions of the fungal amplicon sequence variants. Across sites, soil microbiotas in revegetated urban green spaces were similar to remnant woodland microbiotas and differed greatly from lawns and vacant lots. There were several differentially abundant genera likely driving these differences that had strong correlations to plant species richness, soil pH, and conductivity. We provide the first evidence, as far as we know, that revegetation can improve urban soil microbiota diversity toward a more natural, biodiverse state by creating more wild habitat conditions. This evidence supports initiating further studies within the growing field of microbiome rewilding.

Key words: ecological restoration, exposure, immune protection, microbiome, non-communicable disease, public health

Implications for Practice

- Urban design principles that incorporate biologically diverse communities (above and below ground) within green spaces can improve microbiota diversity, which has been linked to human health benefits.
- This study provides support for global and local initiatives at the nexus of restoration ecology, urban design, and public health that seek to establish urban green space interventions to improve human health outcomes.
- This study encourages further empirical studies to explore the mechanistic interactions between ecological restoration, urban design, human microbial exposure, and public health.

Introduction

Microbial "old friends" are a more-than-human part of the body that, through coevolution, has become involved in training and regulating our immune systems, among many other functions (Rook et al. 2003). Many of these "old friends" come from our environment and colonize us, particularly during childhood. A lack of exposure to these "old friends," perhaps due to poor quality environment or diet, is implicated in noncommunicable diseases (Gilbert et al. 2018).

Given that most of human evolution has occurred in biodiverse and wild environments, it is not surprising that

Author contributions: JM, NG, AL, PW, MB conceived and designed the study; JM did the fieldwork; AB did the bioinformatics; JM, AB, CS, LW, TT analyzed and interpreted the data; JM wrote the manuscript; all authors contributed to editing and guiding the manuscript; JM did the revisions.

[1]School of Biological Sciences and Environment Institute, The University of Adelaide, Adelaide, South Australia, 5005, Australia
[2]Address correspondence to J. G. Mills, email millsj515@gmail.com and jacob.mills@adelaide.edu.au
[3]Oceans and Atmosphere, CSIRO, Hobart, Tasmania, 7004, Australia
[4]Australian Centre for Ancient DNA, School of Biological Sciences, The University of Adelaide, Adelaide, South Australia, 5005, Australia
[5]Centre for Marine Science and Innovation, School of Biological, Environmental and Earth Sciences, University of New South Wales, Sydney, New South Wales, 2052, Australia
[6]Department of Anthropology, The Pennsylvania State University, University Park, Pennsylvania, 16802, USA
[7]College of Science and Engineering, Flinders University, Adelaide, South Australia, 5042, Australia

doi: 10.1111/rec.13175
Supporting information at:
http://onlinelibrary.wiley.com/doi/10.1111/rec.13175/suppinfo

noncommunicable disease rates are disproportionately rising in industrialized urban populations more than in rural populations (von Hertzen et al. 2011). For example, good respiratory health positively correlates with biodiversity exposure in rural residents (Stein et al. 2016; Liddicoat et al. 2018).

The global burden of urban-associated diseases, such as many auto-immune and inflammatory diseases (reviewed in Flies et al. 2019), have partly been attributed to a loss of human microbial diversity. This loss is evident in the decrease of diversity in urban versus rural populations (von Hertzen et al. 2011; Lehtimäki et al. 2017). Further, microbiota—a community of microorganisms in a defined environment (Marchesi & Ravel 2015)—of indoor and outdoor built environments have been linked in many ways to human health status and the structuring of the human microbiota in early life (reviewed in Sbihi et al. 2019). Indeed, like urban people, urban built environments have been shown to be relatively low in microbial diversity compared to rural environments (Kembel et al. 2012; Kirjavainen et al. 2019). Moreover, built environments, along with differences in diet and sedentary lifestyles, were found to be the main factors attributed to an allergy epidemic in the Finnish Karelia (Haahtela et al. 2015).

Recent studies have directly shown the influence of nature on the human skin microbiota. For example, children that attend nature-based childcare facilities have been shown to have a more diverse skin microbiota (Lehtimäki et al. 12018*a*), as have those exposed to sandboxes inoculated with microbially rich organic soil (Hui et al. 2019). However, these approaches are unlikely to be socio-economically transcendent nor sustainable where access to a freely available and ever-present microbial source is needed. Therefore, the ecological restoration of environmental microbiomes—microbiomes being the microbiota, their genomes, and the surrounding environmental conditions (Marchesi & Ravel 2015)—into urban habitats may provide the long-term microbial exposure that is required for positive health outcomes for all people living in a built environment (Mills et al. 2017; Robinson et al. 2018; Lehtimäki et al. 12018*b*).

Previous studies suggest that ecological restoration within urban environments has the potential to "rewild" microbiotas. For example, wild-land restoration interventions that increase vegetation diversity and rewild soil microbiota (Gellie et al. 2017; Yan et al. 2018) may support more niche-adapted, rather than opportunistic pathogenic bacteria in degraded ecosystems (Liddicoat et al. 12019*a*). Moreover, a randomized controlled trial found that exposure to soils from high vegetation diversity areas can supply the mouse gut with anxiety-reducing bacteria (Liddicoat et al. 12019*b*). Further, natural soil microbiota are found in rural house dusts, which are causatively linked to immune protection in experiments with laboratory mice and comparative studies of rural and urban people (Haahtela et al. 2015; Stein et al. 2016; Lehtimäki et al. 12018*b*).

As such, microbiome rewilding of urban outdoor spaces could provide "nature-like" microbiota to urban homes. Therefore, it is important to understand the influence of revegetation on urban soil microbiota if this solution is to be offered as a cost-effective and permanent source of diverse environmental microorganisms for early-life exposure in the urban environment (Mills et al. 2019).

In this study, we compared the soil microbial diversity and composition of revegetated urban green spaces to other common urban vegetation types with the expectation that they would be more similar to local remnant patches. We examined the structure, diversity, and differential abundance of bacterial, fungal, archaeal, and eukaryotic soil microbiota in the context of the vegetation and soil characteristics in different vegetation types of urban green space. Specifically, we wanted to know whether revegetated urban green spaces have soil microbiome characteristics more similar to local remnant patches or to more traditional urban green spaces with lower vegetation diversity.

Methods

Study Sites

We focused our study on five vegetation types of urban green space in the City of Playford, Kaurna Country, South Australia, Australia. The five vegetation types were lawns, vacant lots, parklands, revegetated open woodlands, and remnant open woodlands in urban patches. Each vegetation type was replicated three times in a spatially independent design (Fig. 1). Two randomly located quadrats within each replicate were sampled. Each quadrat was a 25 × 25 m plot in a NSEW orientation, with geo-references and photos taken at the SW corner. The vegetation types represented different land-uses across a gradient of vegetation structure and diversity.

Vegetation Surveys

Vegetation surveys were undertaken between 31 October and 4 November, 2016. All revegetated woodlands had been restored with locally native species more than 15 years before sampling to resemble the pre-European open grassy woodlands that existed in the area. The vegetation of each replicate was described by a point-intercept survey of five 25-m transects running EW (modified from White et al. 2012) and an assessment based on the National Vegetation Information System (ESCAVI 2003) to determine functional groups and structural characteristics of the vegetation in the green spaces. We collected specimens of all plant species within the quadrat. The point-intercept survey provided count data of each plant species within the quadrat and these data were used to calculate Shannon's diversity and effective species number (taken as the exponent of Shannon's diversity).

Soil Sampling

Soils were sampled for microbiota and physicochemical analysis according to the Biomes of Australian Soil Environments (BASE) project protocol (Bissett et al. 2016) between 1 and 16 September, 2016. In brief, 100–200 g

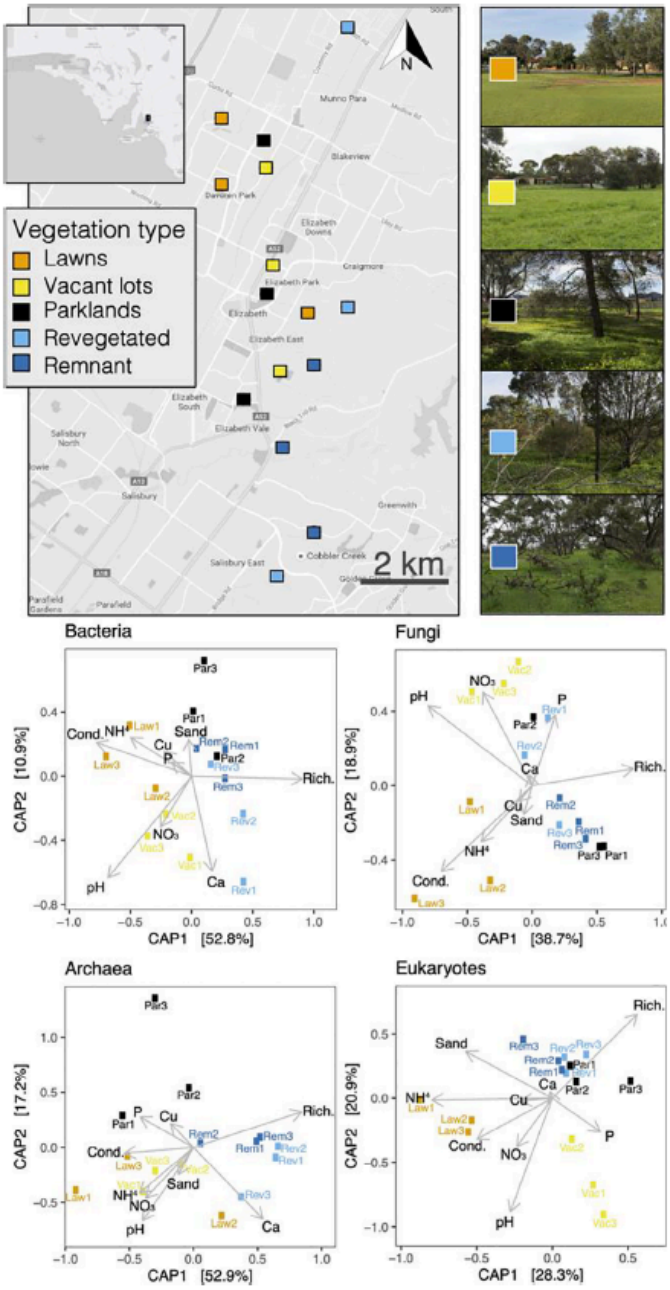

Figure 1. Sampling locations and photos of urban green space vegetation types in the City of Playford, South Australia. Constrained analysis of principal coordinates (CAP) of soil bacteria, fungi, archaea, and eukaryotes constrained by uncorrelated soil physicochemical variables (pH; sand; ammonium, NH^4; nitrate, NO_3; conductivity, Cond.; calcium, Ca; copper, cu; phosphorus, P) and plant species richness (Rich.). Wald test for the difference between revegetated woodland (Rev) soil microbiotas to those of lawns (Law), vacant lots (Vac), parklands (Par), and remnant woodlands (Rem); negative binomial multivariate GLM; bacteria, fungi, archaea, eukaryotes, $W = 285.4, 177.9, 235.4, 156.6$, $p = <0.001, <0.001, <0.001, <0.001$.

of soil was randomly sampled from nine points within the quadrat and mixed. Of this pooled sample, 50 g was taken for DNA analysis and 250 g for physicochemical analysis. Soil physicochemical characteristics were analyzed by CSBP Laboratories (Perth, Western Australia) (see Table S1 for the variables analyzed).

DNA Extraction, Library Preparation, and Sequencing

Soil DNA extraction, library preparation of bacterial and archaeal 16S rRNA, fungal ITS1 region, and eukaryotic 18S rRNA marker genes, and next generation sequencing were done according to Bissett et al. (2016) by the Australian Genome Research Facility and Bioplatforms Australia under the BASE protocol. Briefly, DNA was extracted in triplicate (3×0.25 g) using the MOBIO Power Soil kit, as per the manufacturer's recommendations. The fungal ITS region primers ITS1F (Gardes & Bruns 1993) and ITS4 (White et al. 1990), partial bacterial 16S rRNA genes (primers 27f [Lane 1991]) and 519r (Lane et al. 1985)) and partial 18S rRNA genes (primers 1391f (Amaral-Zettler et al. 2009), and EukBr (Caporaso et al. 2012)) were amplified and sequenced using the Illumina MiSEQ platform. The 16S and ITS amplicons were sequenced using 300 bp paired end sequencing, while 18S amplicon reads were generated using 150 bp paired end sequencing (Bissett et al. 2016).

Sequenced 16S and 18S rRNA gene amplicons were merged with FLASH (Magoč & Salzberg 2011), and merged reads containing Ns or homopolymer runs >8 bp were removed using MOTHUR v1.34.1 (Schloss et al. 2009). For ITS sequences, the ITS1 region was extracted from Illumina R1 reads using ITSx (Bengtsson-Palme et al. 2013), and sequences containing ambiguous bases removed. The remaining analysis steps were the same for all amplicons and used to produce sequence abundance tables per amplicon. Sequences were denoised using the UNOISE3 algorithm (Edgar & Flyvbjerg 2015) with USEARCH v11 (Edgar 2010) (–minsize 4) and abundance profiles built by mapping all data to identified amplicon sequence variants (ASVs) (USEARCH–otutab–maxaccepts 0). ASVs were classified using the RDP classifier (Wang et al. 2007) in MOTHUR at a 60% probability cut-off, against SILVA132 (Quast et al. 2012) for 16S and 18S amplicons and UNITE SH v7 (Nilsson et al. 2018) for fungal ITS.

Data Access

Sequence data are available from the Australian Microbiome (BASE) data portal (https://data.bioplatforms.com/organization/about/australian-microbiome) under sample IDs 39101-39159 (odd numbers only) or from the NCBI Sequence read archive under Bioproject accession no. PRJNA317932 and Biosamples ID's SAMN07449116-SAMN07449174 (even numbers only). Marker gene sequences, ASV and taxonomy tables, and soil and vegetation metadata is available on figshare.com with the DOI 10.25909/5db8cd2782ebd.

Statistics

Data Preparation. Prior to analysis, ASVs assigned to mitochondria, chloroplast, "unknown" or incorrect domain or kingdom (e.g. archaea in the bacterial database) were deleted. The eukaryote dataset was further filtered at the phylum level to focus on micro-eukaryotes. This filtering excluded fungi (captured in the fungal dataset), land plants, brown and red algae, vertebrates, and macro-invertebrates. The read abundances of each ASV between the two quadrats of each replicate were summed because we only had vegetation data for one quadrat within each replicate of vegetation type. Soil physicochemical variables of the two quadrats of each replicate were averaged as, for example, we cannot sum particle percentages or pH.

Environmental Predictor Analysis. All statistics were done in R version 3.5.3 (R Core Team 2019). Generalized linear models (GLMs) with negative-binomial link functions were used to test for significant differences of the soil physicochemical variables between the revegetated woodlands and the other vegetation types with the *MASS* package version 7.3.51.4 (Ripley et al. 2013). GLMs with a Poisson link function (in base R) were used to test for differences of the vegetation community characteristics between the revegetated woodlands and the other vegetation types (i.e. number of growth forms and height classes, species richness, Shannon's diversity, effective species number [Shannon's]). All GLMs were tested and confirmed for goodness-of-fit with chi-squared tests, with $p > 0.05$ for a good fit.

Microbiome Analysis—Diversity. The *phyloseq* package version 1.26.1 was used for microbiota analysis of alpha- and beta-diversity and differential abundance testing (McMurdie & Holmes 2013). All ASVs that had fewer than five sequence reads across all samples were removed to account for possible sequencing errors. Phylogenetic trees were constructed using the *msa* package version 1.16.0 (Bodenhofer et al. 2015) for multiple sequence alignment and the *phangorn* package version 2.5.5 (Schliep 2010) for phylogenetic tree building.

Samples were rarefied to the read count of the sample with fewest reads (bacteria, 30,074 reads; fungi, 86,026; archaea, 123,363; eukaryotes, 12,261) for alpha-diversity analysis of ASVs (observed ASV richness, Faith's phylogenetic diversity, Shannon's diversity index). ASV richness and Shannon's diversity were calculated with the "estimate_richness" function in *phyloseq* and Faith's phylogenetic diversity was calculated with the "pd" function of the *picante* package version 1.8 (Kembel et al. 2010). GLMs with negative-binomial link functions were used to test for differences in the alpha-diversity of soil microbiotas between revegetated woodlands and the other vegetation types with the *MASS* package.

Microbiome Analysis—Composition. Before ordination, the environmental predictor variables were mean centered using the "scale" function (in base R). Environmental predictor variables (soil physicochemical and vegetation metrics) were initially checked for multicollinearity with the "pairs.panels" function of the *psych* package version 1.8.12 (Revelle 2015) with Spearman's rho >0.7 or <–0.7 as cut-offs and the most relevant variables to microbiotas, such as pH and ammonium, were retained for constrained ordination. We then ran variance inflation factor analysis on the remaining predictor variables to further reduce multicollinearity with the "stepVIF" function of the *pedometrics* package version 0.7.0 (Samuel-Rosa 2020). The predictor variables remaining after collinearity reduction

were plant species richness, soil calcium, copper, ammonium, nitrate, phosphorus, pH, and percentage of sand (Table S1).

Visualization of beta-diversity was based on unrarefied data with weighted UniFrac distance and ordinated using the "ordinate" function in *phyloseq* with unconstrained principal coordinates analysis (PCoA) and then with the predictor variables in constrained analysis of principal coordinates (CAP). We ran four CAP models as follows, distance ~ vegetation type, distance ~ vegetation type + predictor variables, distance ~ predictor variables, distance ~ predictor variables + vegetation type. Both models including the predictor variables and vegetation type overdetermined the variation (constrained proportion explained >0.94 for bacteria, fungi, archaea, and eukaryotes). However, the model with only predictor variables explained more variation in microbiota structure for all four taxonomic groups (constrained proportion explained = 0.81 [bacteria]; 0.77 [fungi]; 0.84 [archaea]; 0.77 [eukaryotes]) than the model with only vegetation type (constrained proportion explained = 0.71 [bacteria]; 0.59 [fungi]; 0.60 [archaea]; and 0.59 [eukaryotes]). As such, we visualized the constrained model with only the predictor variables as they provided a better explanation of the variance between green spaces than the vegetation types themselves.

We ran permuted analysis of variance (ANOVA; $n = 999$) on the constrained axes of each ordination to identify statistically important axes with the "anova" function of the *base* R package. Differences in beta-diversity of the microbial community structure of the vegetation types were analyzed using multivariate negative-binomial GLMs and Wald tests implemented in the "manyglm" function of the *mvabund* package version 4.0.1 (Wang et al. 2012). Predictor variables used in the ordination models were correlated to the two main axes (CAP1 and CAP2) of the CAP analyses with two-tailed Spearman's correlation tests with the "cor.test" function of the *stats* package version 3.6.1 (R Core Team 2019).

Microbiome Analysis—Differential Abundance Testing. We tested for differentially abundant microbial taxa across the vegetation types, using revegetated woodlands as the reference group. ASVs were agglomerated to genus level for differential abundance testing of genera using the "tax_glom" function of *phyloseq*. Log-2 fold-change measurement of genera within bacteria, fungi, archaea, and eukaryote datasets was done using the "DESeq" function in the *DESeq2* package version 1.22.2 (Love et al. 2014). *DESeq2* has been shown to work well for designs with unrarefied data, high sequencing depth, and low sample numbers per treatment (Weiss et al. 2017), such as in our study. Differentially abundant genera (alpha level of 0.01) were plotted into heatmaps using the "pheatmap" function of the *pheatmap* package version 1.0.12 (Kolde & Kolde 2015). The differential abundance heatmap (Fig. 2) rows and columns were clustered based on Manhattan distance. The differential abundance heatmap (Fig. 2) scale represents the mean abundance of each genus across samples as 0 with ±3 standard deviations. Unidentified genera were removed for plotting. Archaea only had 1 identified differentially abundant genus; therefore, a heatmap was not produced for them. Differentially abundant genera were correlated with environmental variables using Spearman's correlation with the "cor" function of the *base* R package to produce a correlation matrix. This matrix was pruned to have genera as rows and environmental variables as columns, and then these correlations were plotted into heatmaps using the "pheatmap" function of the *pheatmap* package version 1.0.12 (Kolde & Kolde 2015). The correlations heatmap (Fig. 3) rows and columns were clustered based on Manhattan distance.

Microbiome Analysis—Functional Assignment. The online version of *FUNguild* (http://www.stbates.org/guilds/app.php) was used to make trophic predictions for the fungal ASV dataset (Nguyen et al. 2016). We used GLMs with a binomial link function to test proportions (derived from unrarefied sequence reads) of fungal trophic modes for significant differences between revegetated woodlands and the other vegetation types. Please see Supplement S1 for R script.

Results

Environmental Characterization of Urban Green Spaces

We collected 122 plant specimens across the 15 urban vegetation-type replicates. Fifty-three Australian native and 45 introduced plant species were identified, with 24 specimens unidentified. The revegetated and remnant woodlands comprised more native plant species than the other vegetation types, although they still harbored exotic species. Revegetated woodlands had a significantly higher number of growth forms (i.e. proxy for functional diversity) of plants compared to lawns, vacant lots, and parklands (Table 1 and Fig. S1). Revegetated woodlands also had a significantly greater number of plant height classes (i.e. proxy for structural diversity) compared to lawns and vacant lots (Table 1 and Fig. S1); and significantly higher plant species richness and effective species number compared to lawns and vacant lots (Table 1 and Fig. S1). The revegetated woodlands were not significantly different from the remnant woodlands in any of the vegetation variables (Table 1 and Fig. S1).

Soil physicochemical variables that significantly differed between the revegetated and other vegetation types were silt, clay, ammonium, potassium, sulfur, iron, manganese, zinc, and calcium (Table S1). Ammonium was higher in lawns than revegetated green spaces. Organic carbon was also higher in lawns, revegetated, and remnant green spaces; however, it was also highly variable and therefore not significantly different across these urban green space types. Overall, soil and plant community variables are important to these vegetation types and these are explored further in the context of the soil microbiome below.

Soil Microbial Diversity

Soil in revegetated woodlands had a significantly higher bacterial and fungal ASV richness when compared to lawns; however, the opposite was true for eukaryotes (Table 2 and Fig. S2). Richness of soil bacteria, fungi, archaea, and eukaryote

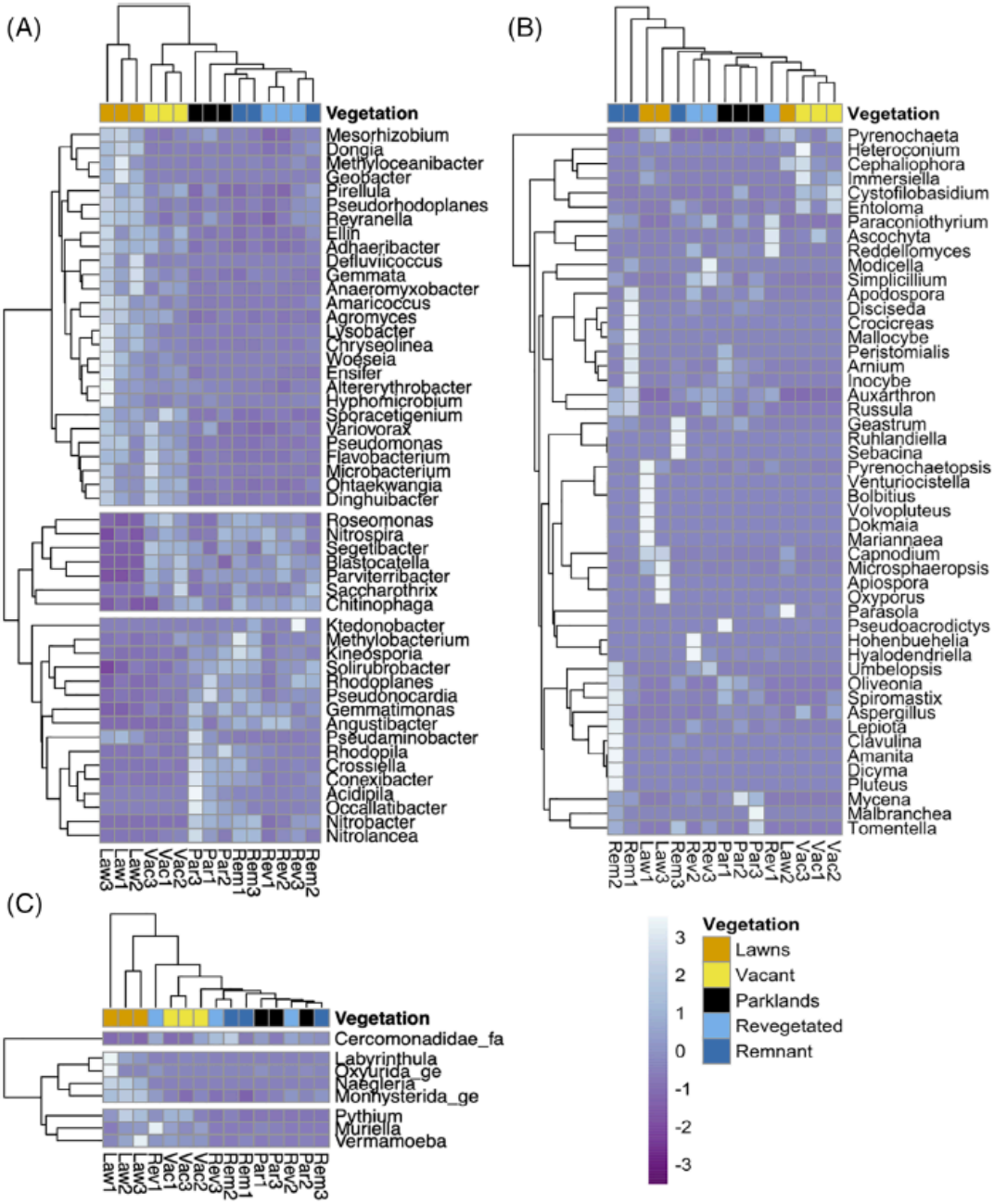

Figure 2. Differentially abundant genera of (A) bacteria, (B) fungi, and (C) eukaryotes in the urban green space vegetation types measured by log-2 fold-change with p value < 0.01. Extreme ends of the heat color scale represent 3 standard deviations from the mean rarefied abundance for each genus across samples. Clustering of genera (rows) and vegetation replicates (columns) are both by Manhattan distance.

ASVs in revegetated woodlands were significantly lower than in remnant woodland patches (Table 2 and Fig. S2). Faith's phylogenetic diversity and Shannon's diversity index was not significantly different between revegetated woodlands and all other vegetation types for fungi and archaea (Table 2 and Fig. S2). However, Faith's phylogenetic diversity of eukaryotes was significantly lower in revegetated woodlands than all other vegetation types. Further, Faith's phylogenetic diversity of bacteria was significantly higher in remnant woodlands than revegetated woodlands.

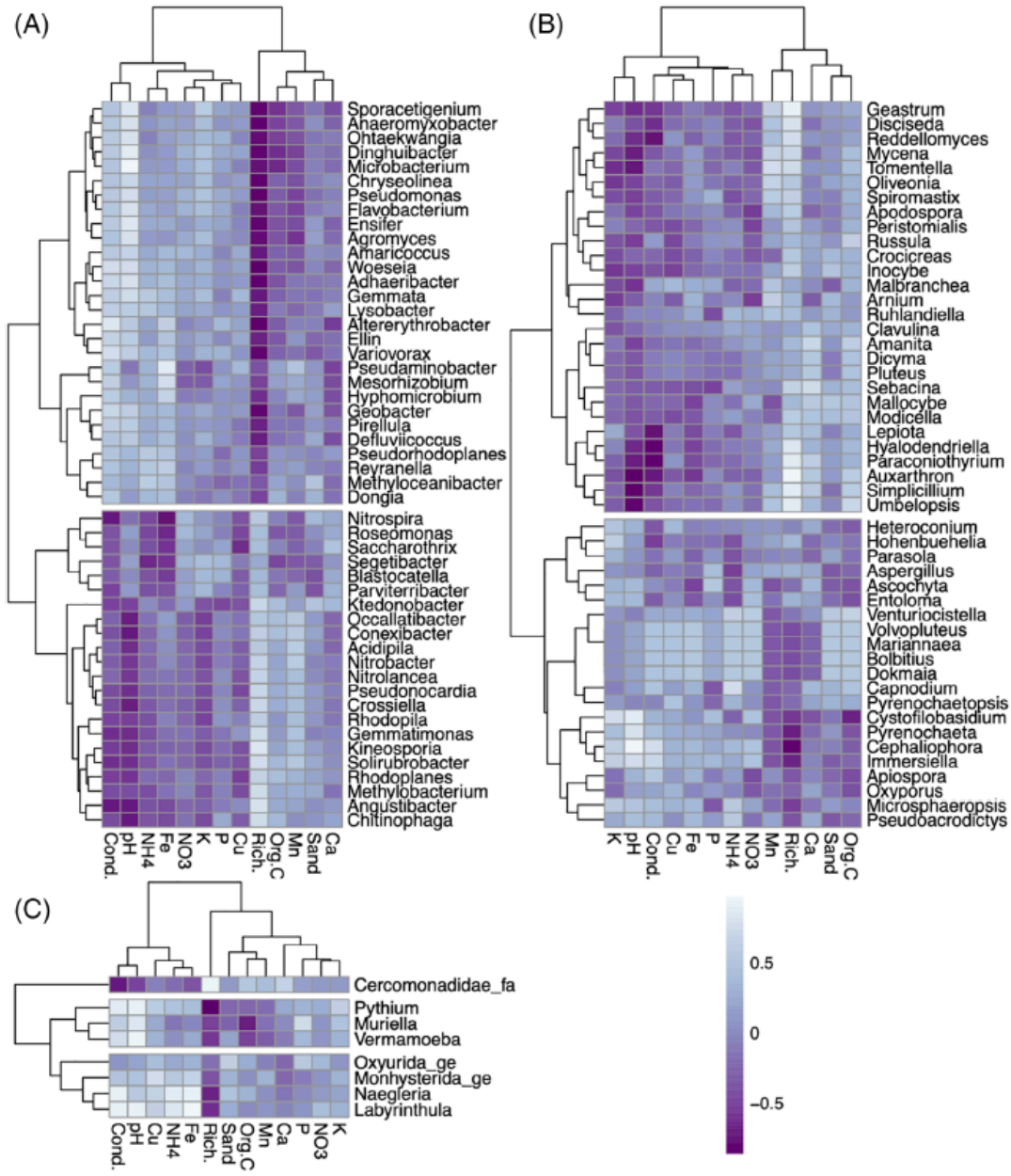

Figure 3. Spearman's correlations between differentially abundant genera of (A) bacteria, (B) fungi, and (C) eukaryotes and environmental variables (pH; sand; ammonium, NH^4; nitrate, NO_3; conductivity, Cond.; calcium, Ca; copper, cu; iron, Fe; manganese, Mn; potassium, K; phosphorus, P; organic carbon, org. C; plant species richness, Rich.). Clustering of genera (rows) and vegetation replicates (columns) are both by Manhattan distance. Scale is Spearman's correlation coefficient.

Environmental Characteristics Relating to Soil Microbial Community Composition

The constrained microbial community structures of bacteria, fungi, and archaea in revegetated woodland green space soils were significantly different from all other urban vegetation types—lawns, vacant lots, parklands, and remnant woodlands (Fig. 1 and Table 3). Eukaryotic communities of revegetated woodlands were significantly different from lawns, vacant lots, and parklands but not remnant woodlands. Overall, revegetated woodlands had bacterial, fungal, archaeal, and eukaryotic

Table 1. Vegetation functional (growth form, NVIS*gf*; NVIS is the National Vegetation Information System) and structural diversity (height class, NVIS*ht*) and species diversity (richness; Shannon's, H′; and effective species number [ESN], exp.[H′]) between the five urban green space vegetation types. Values are "mean (SD)" on the first line, and "*z* score, *p* value" from Poisson GLMs with "Revegetated" as the comparison group on the second line. Degrees of freedom were 14 (null deviance) and 10 (residual deviance). *p* Values: $p < 0.001$ '***', $p < 0.01$ '**', $p < 0.05$ '*', $p < 0.10$ '°', $p > 0.10$ 'ns'.

Green Space (n = *15*)	n *NVIS*gf	n *NVIS*ht	n *richness*	*H′*	*ESN*
Lawns ($n = 3$) *m* (SD)	3.0 (0.0)	1.0 (0.0)	8.7 (3.5)	1.0 (0.6)	3.0 (1.5)
	−1.70, °	−2.66, **	−5.04, ***	−1.31, ns	−3.63, ***
Vacant lots ($n = 3$)	2.0 (0.0)	1.7 (0.6)	17.0 (3.6)	1.8 (0.2)	6.1 (1.1)
	−2.33, *	−2.27, *	−2.59, **	−0.53, ns	−2.24, *
Parklands ($n = 3$)	3.0 (0.0)	3.7 (0.6)	21.3 (3.1)	2.3 (0.2)	10.0 (2.1)
	−1.70, °	−0.96, ns	−1.41, ns	−0.10, ns	−0.61, ns
Revegetated ($n = 3$)	6.0 (1.0)	5.3 (0.6)	27.0 (4.6)	2.4 (0.3)	11.7 (3.7)
(Intercept)	7.60, ***	6.70, ***	29.66, ***	2.39, *	14.54, ***
Remnant ($n = 3$)	6.3 (1.2)	4.3 (0.6)	29.3 (3.8)	2.4 (0.3)	11.0 (3.1)
	0.16, ns	−0.56, ns	0.54, ns	−0.05, ns	0.80, ns

communities that were most similar to remnant woodland microbiotas and most dissimilar to those of lawns and vacant lots. Community structure of all four microbial groups studied here were somewhat similar between constrained (Fig. 1) and unconstrained ordinations (Fig. S3).

We correlated environmental predictor variables with the primary and secondary axes of the constrained ordinations to identify potential drivers of the patterns that we observed in the microbial community structure between the vegetation types (Fig. 1 and Table S3). However, the primary axes of all four ordinations were the only axes in each model significantly explaining the variation (Table S3, permuted ANOVA, $F = 14.2_{\text{bacteria}}, 8.4_{\text{fungi}}, 16.3_{\text{archaea}}, 6.7_{\text{eukaryotes}}$; $df = 1, 5$; $p = 0.03_{\text{bacteria}}, 0.07_{\text{fungi}}, 0.01_{\text{archaea}}, 0.003_{\text{eukaryotes}}$). Plant species richness, soil conductivity, and pH were major predictors of the bacterial community composition according to Spearman's rank correlations ($p < 0.05$; Fig. 1 and Table S3). Further, plant species richness, soil pH, and ammonium were major predictors of the fungal community composition ($p < 0.05$; Fig. 1 and Table S3). Meanwhile, plant species richness, conductivity, and calcium were major predictors of the archaeal community composition ($p < 0.05$; Fig. 1 and Table S3). Plant species richness, soil pH, and conductivity were major predictors of the eukaryotic community composition ($p < 0.05$; Fig. 1 and Table S3). Plant species richness, a significant predictor of all four taxonomic groups represented here, was significantly, positively correlated with all other vegetation community characteristics—number of growth forms and height classes, Shannon's diversity, and effective species number (rho >0.7; Table S2). Overall, plant species richness, soil pH, and conductivity were the strongest predictors of soil microbiotas across the green space vegetation types.

Differential Abundance of Genera

Bacteria, fungi, archaea, and eukaryotes had 50, 50, 1, and 8 genera, respectively, that were differentially abundant across the vegetation types (based on a log-2 fold-change compared to revegetated woodlands with an adjusted *p* value threshold of <0.01). Differentially abundant bacterial genera generally fell into three groups by vegetation types as seen in Figure 2a—lawns and vacant lots (e.g. *Flavobacterium*, *Pseudomonas*); vacant lots, parklands, revegetated, and remnants (e.g. *Nitrospira*, *Parviterribacter*); and parklands, revegetated, and remnants (e.g. *Rhodoplanes*, *Solirubrobacter*). Most differentially abundant fungal genera were highly abundant in one replicate or another (Fig. 2b). The only identified archaeal genera that differed was *Haloterrigena*, which was found in one lawn replicate. Differentially abundant eukaryotic genera also formed three groups by vegetation type. These eukaryote groups were more abundant in lawns (e.g. *Naegleria*, *Labyrinthula*), or lawns and vacant lots (e.g. *Pythium*, *Vermamoeba*), while the family Cercomonadidae were higher in abundance in remnant and revegetated woodlands (Fig. 2c).

We found that several environmental variables correlated strongly with the differentially abundant genera. The bacteria and fungi split into two groups, those correlated with soil conductivity, pH, ammonium, nitrate, potassium, phosphorus, and copper (e.g. bacteria *Sporacetigenium* and fungi *Heteroconium*, Fig. 3), and those that correlated with plant species richness, organic carbon, manganese, sand, and calcium (e.g. bacteria *Nitrospira* and fungi *Geastrum*, Fig. 3). The eukaryotes split into three groups; the Cercomonadidae that correlated positively with plant species richness, sand, organic carbon, manganese, calcium, phosphorus, nitrate, and potassium; those that strongly, negatively correlated with plant species richness, sand, organic carbon, and manganese (e.g. *Pythium*); and those strongly, negatively correlated with plant species richness (e.g. *Naegleria*). Overall, the vegetation types had differentially abundant genera driving the differences in their soil microbiota community structure. The differential abundance of these genera was likely driven strongly by environmental variables such as soil pH and conductivity, and plant species richness.

Functional Predictions of Trophic Modes

Functional predictions of fungi revealed high-level trophic patterns across the urban vegetation types (Fig. 4). There was a significantly greater relative abundance of pathotrophic fungi in vacant lots than revegetated woodlands (GLM_{Binomial},

Table 2. Alpha-diversity of bacteria, fungi, archaea, and eukaryotes measured by ASV richness, Faith's phylogenetic diversity (PD), and Shannon's diversity (H'). Values are "mean (SD)" on the first line, followed by "z score, *p* value" from negative-binomial GLMs with "Revegetated" as the comparison group on the second line. Degrees of freedom were 14 (null deviance) and 10 (residual deviance). *p* Values: $p < 0.001$ '***', $p < 0.01$ '**', $p < 0.05$ '*', $p < 0.10$ '°', $p > 0.10$ 'ns'.

	Bacteria			*Fungi*			*Archaea*			*Eukaryotes*	
Green space (n = 15)	*Richness*	*PD*	*H'*	*Richness*	*PD*	*H'*	*Richness*	*PD*	*H'*	*Richness*	*PD*	*H'*
Lawns (3)	4,022 (508)	198 (10)	7.3 (0.1)	985 (223)	514 (147)	4.4 (0.4)	2,388 (331)	162 (34)	4.9 (0.4)	1,422 (255)	295 (28)	6.1 (0.1)
	−2.42, *	−0.35, ns	0.06, ns	−1.75, °	−0.26, ns	−0.05, ns	−0.74, ns	1.19, ns	−0.08, ns	2.03, *	4.90, ***	0.08, ns
Vacant (3)	4,804 (542)	211 (13)	7.5 (0.2)	1,346 (175)	614 (101)	5.0 (0.4)	2,667 (331)	127 (11)	5.2 (0.4)	1,307 (276)	256 (27)	6.0 (0.4)
	−0.29, ns	0.75, ns	0.16, ns	1.19, ns	0.1.21, ns	0.33, ns	0.43, ns	−0.39, ns	0.07, ns	1.14, ns	2.69, **	0.03, ns
Parklands (3)	5,313 (739)	221 (15)	7.7 (0.1)	1,259 (245)	547 (99)	4.6 (0.5)	2,643 (502)	130 (45)	5.5 (0.2)	1,237 (21)	264 (18)	5.8 (0.3)
	0.91, ns	1.62, ns	0.25, ns	0.56. ns	0.26, ns	0.08, ns	0.33, ns	−0.26, ns	0.24, ns	0.56, ns	3.17, **	−0.04, ns
Revegetated (3)	4,923 (860)	202 (23)	7.2 (0.5)	1,186 (62)	530 (31)	4.4 (0.4)	2,561 (427)	135 (40)	5.0 (0.4)	1,173 (64)	214 (21)	5.9 (0.1)
(Intercept)	143.98, ***	130.62, ***	9.13, ***	94.31, ***	73.14, ***	5.44, ***	117.00, ***	45.39, ***	6.28, ***	104.86, ***	110.40, ***	7.47, ***
Remnant (3)	5,707 (127)	226 (5)	7.7 (0.0)	1,445 (226)	610 (102)	4.9 (0.5)	3,260 (161)	163 (9)	5.8 (0.2)	1,497 (206)	285 (32)	6.1 (0.3)
	1.77, °	2.01, *	0.26, ns	1.86, °	1.16, ns	0.25, ns	2.55, *	1.25, ns	0.39, ns	2.57, *	4.35, ***	0.09, ns

$t_{14,10} = 3.56$, $p < 0.01$). Remnant woodlands had a significantly higher relative abundance of pathotrophic-saprotrophic fungi than revegetated woodlands ($GLM_{Binomial}$, $t_{14,10} = 2.44$, $p < 0.05$). Pathotrophic-symbiotrophic fungi were significantly higher in relative abundance in lawns ($GLM_{Binomial}$, $t_{14,10} = 2.33$, $p < 0.05$) and remnant woodlands ($GLM_{Binomial}$, $t_{14,10} = 2.22$, $p < 0.10$) than revegetated woodlands. Lawns ($GLM_{Binomial}$, $t_{14,10} = 8.28$, $p < 0.001$) and remnant woodlands ($GLM_{Binomial}$, $t_{14,10} = 3.32$, $p < 0.01$) also had a significantly higher relative abundance of saprotrophic fungi than revegetated woodlands. Relative abundance of saprotrophic-symbiotrophic fungi was significantly higher in revegetated woodland soils than lawns ($GLM_{Binomial}$, $t_{14,10} = -2.97$, $p < 0.05$) and vacant lots ($GLM_{Binomial}$, $t_{14,10} = -3.03$, $p < 0.05$). Symbiotrophic fungal relative abundance was also significantly higher in revegetated soils than lawns ($GLM_{Binomial}$, $t_{14,10} = -3.13$, $p < 0.05$) and vacant lots ($GLM_{Binomial}$, $t_{14,10} = -2.04$, $p < 0.10$). Overall, the functional profile of the fungal soil community was varied by vegetation type.

Discussion

We report evidence that urban ecological restoration may rewild the soil microbiota in green space vegetation types commonly found in cities, such as parks. We observed that the structure of soil microbiota between urban vegetation types changed with levels of macro-biodiversity, from low vegetation-diversity lawns and vacant lots, increasing to parklands, to the higher vegetation-diversity revegetated and remnant woodlands. Plant species richness, soil pH, and electrical conductivity were strong predictors of the observed differentiation in microbiotas between urban vegetation types. These results indicate that revegetated urban green spaces have microbiota that are similar to areas of biodiverse, natural remnant vegetation. This similarity, therefore, highlights the potential to align urban green spaces with rural spaces to become beneficial to human, animal, and plant health. However, more work is needed to determine the direct interactions between microbiome rewilding and human health.

Diversifying Green Space Microbiota With Diverse Plant Communities

Ecological restoration of urban green spaces could lead to changes in the microbial diversity of urban soils. In our study, bacterial and eukaryotic microbiotas had significant differences in alpha-diversity between the revegetated and other vegetation types. However, such microbial diversity differences are neither consistently observed in ecological restoration projects nor generally predicted by changes to vegetation diversity (Prober et al. 2015; Gellie et al. 2017; Hamonts et al. 2017). This inconsistency across studies is likely due to the nuances of spatial heterogeneity such as land-use history and soil types interacting with the restoration intervention. Nevertheless, what is generally consistent across studies is the relationship between plant

Table 3. Revegetated urban woodland soil microbiotas significantly differ to other urban soil microbiotas, particularly lawns, vacant lots, and parklands. Negative binomial multivariate GLM, $n = 15$ with "Revegetated" as the comparison group. Values are "Wald statistic (W), p value." p Values: $p < 0.001$ '***', $p < 0.01$ '**', $p < 0.05$ '*', $p < 0.10$ '°', $p > 0.10$ 'ns'.

Green space (n = *15*)	*Bacteria*	*Fungi*	*Archaea*	*Eukaryotes*
Overall W, p value	285.4, ***	177.9, ***	235.4, ***	156.6, ***
Lawns (3)	124.1, ***	80.4, ***	112.4, **	71.05, ***
Vacant lots (3)	132.4, ***	73.6, **	113.6, **	63.5, *
Parklands (3)	118.4, ***	61.9, **	136.5, ***	61.0, *
Revegetated (3) (Intercept)	293.3, ***	134.4, *	389.7, ***	165.1, **
Remnant (3)	101.1, *	64.3, *	90.4, °	57.9, ns

community turnover and soil microbiota structure (Banning et al. 2011; Prober et al. 2015; Liddicoat et al. 12019*a*).

Our beta-diversity results show that revegetated soil microbiotas were significantly different to all other urban vegetation types, though they were most similar to remnant woodland sites—potentially leaving room for inoculations with remnant soils to close the gap and improve the restoration outcome (Wubs et al. 2016). Additionally, the importance of plant species richness to bacterial, fungal, archaeal, and eukaryotic community structure in this study follows trends in previous studies. For example, the revegetation of managed pasture sites to woodlands has been shown to return soil bacterial and fungal communities to near-remnant reference states after several years (Gellie et al. 2017; Yan et al. 2018). Further, plant functional groups (e.g. growth forms in this study) and diversity have been shown to influence soil microbiota structure in unique ways (Chen et al. 2017; Hui et al. 2017; McGee et al. 2019).

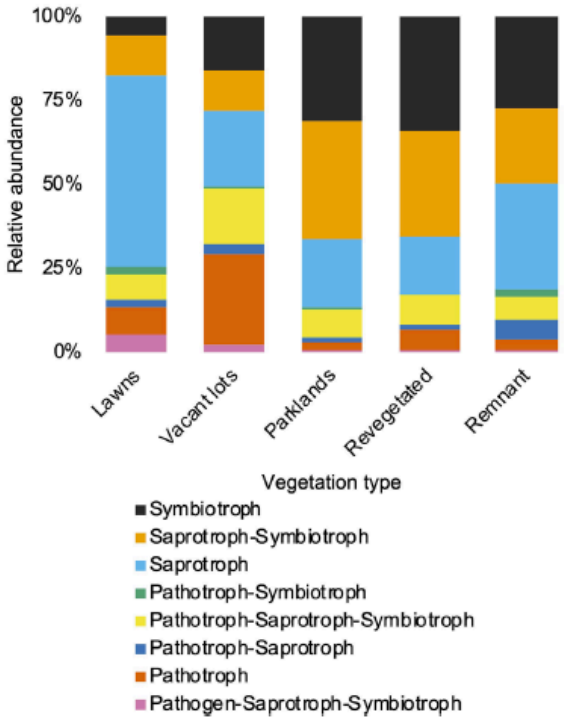

Figure 4. Proportions of soil fungi trophic mode predictions of the urban green space vegetation types. Binomial GLM, $t_{14,10}$ was used to test for significant differences between revegetated woodland soil microbiotas and those of lawns, vacant lots, parklands, and remnant woodlands.

Taken together, the results of this and the other studies mentioned here suggest that more heterogeneous, or "nature-like," green spaces—such as botanic gardens that combine many vegetation types in one location—could be used to increase the overall microbial diversity of an individual green space. This heterogeneity should increase the diversity of microbes that people using that space are exposed to considering that microbial diversity is important for health outcomes associated with the human microbiome (Stein et al. 2016; Lehtimäki et al. 12018*b*). Furthermore, "nature-like" green spaces are more likely to be sought out by people compared with traditional low-diversity local parks (Lin et al. 2014); therefore, such spaces will improve the likelihood of exposure. However, more research is needed on the influence of urban green space heterogeneity on the human microbiome.

Associations Between Soil Physicochemistry and Microbiota

Soil pH, well known to influence soil microorganisms (Fierer 2017), was strongly correlated with the structure of bacterial, fungal, archaeal, and eukaryotic microbiotas as well as certain genera in our study and was higher in the lawns and vacant lots. As such, the structure of the lawn microbiota observed in our study may be influenced by applications of mineral ammonium fertilizer that can temporarily increase the soil pH (Geisseler & Scow 2014). Eventually, this fertilization can decrease soil pH as nitrification of ammonium to nitrite and then nitrate produces protons that acidify the soil. However, nitrification appeared less active in lawn soils as some nitrifying bacteria (i.e. *Nitrolancea*, *Nitrobacter*, *Nitrospira*) were less abundant in the lawns of our study where ammonium and pH were higher, relative to other vegetation types (Sorokin et al. 2014; Spieck & Bock 2015; Daims & Wagner 2018). These abundances suggest that the lawn soils were under the temporary increase in soil pH due to some interaction between mineral ammonium fertilization and reduced nitrifying capacity. Moreover, the higher abundance of the aforementioned nitrifiers in the more floristically diverse green spaces potentially led to more active nitrification, theoretically leading to the observed lower pH and ammonium levels in

those soils. Furthermore, the nitrogen-fixers *Mesorhizobium* and *Ensifer* (Dhaoui et al. 2016; Beukes et al. 2019) that were more abundant in lawns and vacant lots were correlated with ammonium, suggesting that ammonification, as well as fertilization, was leading to the higher ammonium and pH levels in the lawn soils.

Additionally, eukaryotic differences along pH gradients have previously been shown to be marked by differences in parasites and saprotrophs (Dupont et al. 2016). These differences are supported by our results as we found several eukaryotic taxa associated with parasitism, pathotrophy, or saprotrophy to be significantly more abundant in lawn soils and strongly correlated with soil pH. These taxa included *Pythium* (plant and animal parasite), *Labyrinthula* (wasting disease in plants), fungal saprotrophs, and a genus of soil amoeba, *Naegleria* (found mostly in lawn soils of our study). Within the *Naegleria* genus, the species *N. fowleri* can be found in water and soil and can infect humans causing brain damage and possibly death (SAHealth 2019). The length of our ASV sequence did not allow us to unambiguously identify the species of *Naegleria* in the lawn soils; however, the general abundance of this genus indicates that conditions may be suitable for *N. fowleri*.

Previous work has identified that degraded and depauperate soils potentially harbor more human pathogens (Liddicoat et al. 12019*a*) and this is tentatively supported by our results. While lawns certainly have value for sports and recreational purposes that come with their own health benefits, they are arguably not an ecologically "healthy" urban green space. Lawns negatively impact the environment because they are extensive in land area, high in fertilizer and pesticide use, have excessive carbon emissions from maintenance and lost standing-carbon storage capacity (Milesi et al. 2005; Wheeler & Nauright 2006), and now appear to potentially harbor more opportunistic pathogens compared to biodiverse green spaces. Furthermore, previous studies have shown that people are more inclined to use biodiverse green spaces over lawns and that these provide greater physiological and mental well-being benefits (Lin et al. 2014; Shanahan et al. 2016). Weight should therefore be given to these factors during planning of green space land-use proportions in urban designs.

The importance of heterogeneity of plant communities and soil components—key characteristics of biodiverse ecosystems—on soil microbiota are highlighted by our results. Therefore, such heterogeneity will be generally important for practitioners to consider implementing during microbiome rewilding. While further empirical research is needed, our study provides a footing for urban planners and designers to place the environmental microbiome and access to diverse green spaces in their design principles when developing and rejuvenating urban areas. This placement may be important because the rapid rise of human noncommunicable diseases in industrialized urban areas across the globe has been linked to reduced biodiversity exposure (von Hertzen et al. 2011). One hypothesis for this rise is that urban children receive inadequate interactions with microbiotas from natural ecosystems, which are important for immune training during early life (Rook et al. 2003; Gilbert et al. 2018). Microbiome rewilding has potential to help reduce the rising noncommunicable disease burden in cities across the world by improving biodiversity exposure, particularly to natural microbiotas, in a cost-effective and socio-economically transcendent manner (Prescott et al. 2017; Robinson et al. 2018; Mills et al. 2019).

Our study suffered from several limitations as follows. More replicates per treatment would have improved the accuracy of our differential abundance testing (Weiss et al. 2017); however, the City of Playford does not contain many revegetated and remnant woodlands. We also suspect that may be the case in most cities across the globe. Furthermore, our study was cross-sectional and would have benefited from temporal sampling before and after revegetation and across seasons and years. Also, testing our hypotheses across various biomes and climatic zones would be of great interest to this field. Additionally, while 16S rRNA gene databases are extensive, there remains a current lack of knowledge and microbial annotation relative to the diversity of microorganisms in the world. Our results must be considered in the light of these limitations.

Acknowledgments

We are grateful for the contribution of the Biomes of Australian Soil Environments (BASE) consortium (https://data.bioplatforms.com/organization/pages/bpa-base/acknowledgements) in the generation of data used in this publication. The BASE project is supported by funding from Bioplatforms Australia through the Australian Government National Collaborative Research Infrastructure Strategy. We are also grateful for the contributions of Jo Park, Kaarina Sarac, and Ken Daniel from the City of Playford for their support in bringing this project about and to fruition. We would also like to acknowledge Mark Laws, Christopher Fulton, and Jacob Maher for their assistance with fieldwork and plant species identification. This work was supported by Australian Research Council funding to A.L. and M.B. (DE150100542; DP150103414). We declare no conflict of interest.

References

Amaral-Zettler LA, McCliment EA, Ducklow HW, Huse SM (2009) A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. PLoS One 4:e6372

Banning NC, Gleeson DB, Grigg AH, Grant CD, Andersen GL, Brodie EL, Murphy DV (2011) Soil microbial community successional patterns during forest ecosystem restoration. Applied and Environmental Microbiology 77: 6158–6164

Bengtsson-Palme J, Ryberg M, Hartmann M, Branco S, Wang Z, Godhe A, et al. (2013) Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. Methods in Ecology and Evolution 4:914–919

Beukes CW, Boshoff FS, Phalane FL, Hassen AI, le Roux MM, Stepkowski T, et al. (2019) Both alpha-and beta-rhizobia occupy the root nodules of *Vachellia karroo* in South Africa. Frontiers in Microbiology 10:1195

Bissett A, Fitzgerald A, Meintjes T, Mele PM, Reith F, Dennis PG, et al. (2016) Introducing BASE: the biomes of Australian soil environments soil microbial diversity database. GigaScience 5. https://doi.org/10.1186/s13742-016-0126-5

Bodenhofer U, Bonatesta E, Horejš-Kainrath C, Hochreiter S (2015) Msa: an R package for multiple sequence alignment. Bioinformatics 31:3997–3999

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME Journal 6:1621–1624

Chen Y-L, Xu T-L, Veresoglou SD, Hu H-W, Hao Z-P, Hu Y-J, Liu L, Deng Y, Rillig MC, Chen BD (2017) Plant diversity represents the prevalent determinant of soil fungal community structure across temperate grasslands in northern China. Soil Biology and Biochemistry 110:12–21

Daims H, Wagner M (2018) Nitrospira. Trends in Microbiology 26:462–463

Dhaoui S, Rejili M, Mergaert P, Mars M (2016) *Rhizobium leguminosarum* symbiovar *trifolii*, *Ensifer numidicus* and *Mesorhizobium amorphae* symbiovar *ciceri* (or *Mesorhizobium loti*) are new endosymbiotic bacteria of *Lens culinaris* Medik. FEMS Microbiology Ecology 92. https://doi.org/10.1093/femsec/fiw118

Dupont A, Griffiths R, Bell T, Bass D (2016) Differences in soil micro-eukaryotic communities over soil pH gradients are strongly driven by parasites and saprotrophs. Environmental Microbiology 18:2010–2024

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460–2461

Edgar RC, Flyvbjerg H (2015) Error filtering, pair assembly and error correction for next-generation sequencing reads. Bioinformatics 31:3476–3482

ESCAVI (2003) *Australian Vegetation Attribute Manual: National Vegetation Information System*. Department of Environment and Heritage, Canberra, ACT, Australia

Fierer N (2017) Embracing the unknown: disentangling the complexities of the soil microbiome. Nature Reviews Microbiology 15:579–590

Flies EJ, Mavoa S, Zosky GR, Mantzioris E, Williams C, Eri R, Brook BW, Buettel JC (2019) Urban-associated diseases: candidate diseases, environmental risk factors, and a pathway forward. Environment International 133:105187

Gardes M, Bruns TD (1993) ITS primers with enhanced specificity for basidiomycetes-application to the identification of mycorrhizae and rusts. Molecular Ecology 2:113–118

Geisseler D, Scow KM (2014) Long-term effects of mineral fertilizers on soil microorganisms–a review. Soil Biology and Biochemistry 75:54–63

Gellie NJ, Mills JG, Breed MF, Lowe AJ (2017) Revegetation rewilds the soil bacterial microbiome of an old field. Molecular Ecology 26:2895–2904

Gilbert JA, Blaser MJ, Caporaso JG, Jansson JK, Lynch SV, Knight R (2018) Current understanding of the human microbiome. Nature Medicine 24:392–400

Haahtela T, Laatikainen T, Alenius H, Auvinen P, Fyhrquist N, Hanski I, et al. (2015) Hunt for the origin of allergy–comparing the Finnish and Russian Karelia. Clinical & Experimental Allergy 45:891–901

Hamonts K, Bissett A, Macdonald BC, Barton PS, Manning AD, Young A (2017) Effects of ecological restoration on soil microbial diversity in a temperate grassy woodland. Applied Soil Ecology 117:117–128

von Hertzen L, Hanski I, Haahtela T (2011) Natural immunity. Biodiversity loss and inflammatory diseases are two global megatrends that might be related. EMBO Reports 12:1089–1093

Hui N, Jumpponen A, Francini G, Kotze DJ, Liu X, Romantschuk M, et al. (2017) Soil microbial communities are shaped by vegetation type and park age in cities under cold climate. Environmental Microbiology 9:1281–1295

Hui N, Grönroos M, Roslund MI, Parajuli A, Vari HK, Soininen L, Laitinen OH, Sinkkonen A, The ADELE Research Group (2019) Diverse environmental microbiota as a tool to augment biodiversity in urban landscaping materials. Frontiers in Microbiology 10. https://doi.org/10.3389/fmicb.2019.00536

Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD, Blomberg SP, Webb CO (2010) Picante: R tools for integrating phylogenies and ecology. Bioinformatics 26:1463–1464

Kembel SW, Jones E, Kline J, Northcutt D, Stenson J, Womack AM, Bohannan BJ, Brown GZ, Green JL (2012) Architectural design influences the diversity and structure of the built environment microbiome. The ISME Journal 6:1469–1479

Kirjavainen PV, Karvonen AM, Adams RI, Täubel M, Roponen M, Tuoresmäki P, et al. (2019) Farm-like indoor microbiota in non-farm homes protects children from asthma development. Nature Medicine 25:1089–1095

Kolde R, Kolde MR (2015) Package 'pheatmap'. R Package 1

Lane D (1991) 16S/23S rRNA sequencing. Pages 115–175. In: Stackebrandt E, Goodfellow M (eds) Nucleic acid techniques in bacterial systematics. John Wiley and Sons, Chichester, United Kingdom

Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, Pace NR (1985) Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proceedings of the National Academy of Sciences 82:6955–6959

Lehtimäki J, Karkman A, Laatikainen T, Paalanen L, Von Hertzen L, Haahtela T, Hanski Ilkka, Ruokolainen Lasse (2017) Patterns in the skin microbiota differ in children and teenagers between rural and urban environments. Scientific Reports 7:45651

Lehtimäki J, Laatikainen T, Karkman A, von Hertzen L, Haahtela T, Hanski I, Ruokolainen L (2018a) Nature-oriented daycare diversifies skin microbiota in children—no robust association with allergies. Pediatric Allergy and Immunology 29:318–321

Lehtimäki J, Sinkko H, Hielm-Björkman A, Salmela E, Tiira K, Laatikainen T, et al. (2018b) Skin microbiota and allergic symptoms associate with exposure to environmental microbes. Proceedings of the National Academy of Sciences 115:4897–4902

Liddicoat C, Bi P, Waycott M, Glover J, Lowe AJ, Weinstein P (2018) Landscape biodiversity correlates with respiratory health in Australia. Journal of Environmental Management 206:113–122

Liddicoat C, Weinstein P, Bissett A, Gellie NJ, Mills JG, Waycott M, Breed Martin F (2019a) Can bacterial indicators of a grassy woodland restoration inform ecosystem assessment and microbiota-mediated human health? Environment International 129:105–117

Liddicoat C, Sydnor H, Cando-Dumancela C, Dresken R, Liu J, Gellie NJ, et al. (2019b) Naturally-diverse airborne environmental microbial exposures modulate the gut microbiome and may provide anxiolytic benefits in mice. Science of the Total Environment 701:134684

Lin BB, Fuller RA, Bush R, Gaston KJ, Shanahan DF (2014) Opportunity or orientation? Who uses urban parks and why. PLoS One 9:e87422

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

Magoč T, Salzberg SL (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27:2957–2963

Marchesi JR, Ravel J (2015) The vocabulary of microbiome research: a proposal. Microbiome 3. https://doi.org/10.1186/s40168-015-0094-5

McGee KM, Eaton WD, Porter TM, Shokralla S, Hajibabaei M (2019) Soil microbiomes associated with two dominant Costa Rican tree species, and implications for remediation: a case study from a Costa Rican conservation area. Applied Soil Ecology 137:139–153

McMurdie PJ, Holmes S (2013) Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One 8:e61217

Milesi C, Running SW, Elvidge CD, Dietz JB, Tuttle BT, Nemani RR (2005) Mapping and modeling the biogeochemical cycling of turf grasses in the United States. Environmental Management 36:426–438

Mills JG, Weinstein P, Gellie NJC, Weyrich LS, Lowe AJ, Breed MF (2017) Urban habitat restoration provides a health benefit through microbiome engineering: the microbiome rewilding hypothesis. Restoration Ecology 25:866–872

Mills JG, Brookes JD, Gellie NJ, Liddicoat C, Lowe AJ, Sydnor HR, Thomas Torsten, Weinstein Philip, Weyrich Laura S, Breed Martin F (2019) Relating urban biodiversity to human health with the 'holobiont' concept. Frontiers in Microbiology 10. https://doi.org/10.3389/fmicb.2019.00550

Nguyen NH, Song Z, Bates ST, Branco S, Tedersoo L, Menke J, Schilling JS, Kennedy PG (2016) FUNGuild: an open annotation tool for parsing fungal community datasets by ecological guild. Fungal Ecology 20:241–248

Nilsson RH, Larsson K-H, Taylor AFS, Bengtsson-Palme J, Jeppesen TS, Schigel D, et al. (2018) The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. Nucleic Acids Research 47:D259–D264

Prescott SL, Larcombe D-L, Logan AC, West C, Burks W, Caraballo L, et al. (2017) The skin microbiome: impact of modern environments on skin ecology, barrier integrity, and systemic immune programming. World Allergy Organization Journal 10:29

Prober SM, Leff JW, Bates ST, Borer ET, Firn J, Harpole WS, et al. (2015) Plant diversity predicts beta but not alpha diversity of soil microbes across grasslands worldwide. Ecology Letters 18:85–95

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO (2012) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Research 41:D590–D596

R Core Team (2019) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Revelle W (2015) Package 'psych'. The Comprehensive R Archive Network

Ripley B, Venables B, Bates DM, Hornik K, Gebhardt A, Firth D, et al. (2013) Package 'mass'. Cran R

Robinson J, Mills JG, Breed MF (2018) Walking ecosystems in microbiome-inspired green infrastructure: an ecological perspective on enhancing personal and planetary health. Challenges 9:40

Rook GAW, Martinelli R, Brunet LR (2003) Innate immune responses to mycobacteria and the downregulation of atopic responses. Current Opinion in Allergy and Clinical Immunology 3:337–342

SAHealth (2019) *Naegleria fowleri*. SA Health, South Australia, Australia. sahealth.sa.gov.au (accessed 15 Oct 2019)

Samuel-Rosa A (2020) Pedometrics: miscellaneous pedometrics tools. R package version 0.7.0. https://CRAN.R-project.org/package=pedometrics (accessed 21 Feb 2020)

Sbihi H, Boutin RCT, Cutler C, Suen M, Finlay BB, Turvey SE (2019) Thinking bigger: how early life environmental exposures shape the gut microbiome and influence the development of asthma and allergic disease. Allergy 74:2103–2115

Schliep KP (2010) Phangorn: phylogenetic analysis in R. Bioinformatics 27:592–593

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and Environmental Microbiology 75:7537–7541

Shanahan DF, Bush R, Gaston KJ, Lin BB, Dean J, Barber E, Fuller RA (2016) Health benefits from nature experiences depend on dose. Scientific Reports 6:28551

Sorokin DY, Vejmelkova D, Lücker S, Streshinskaya GM, Rijpstra WIC, Damste JSS, et al. (2014) *Nitrolancea hollandica* gen. nov., sp. nov., a chemolithoautotrophic nitrite-oxidizing bacterium isolated from a bioreactor belonging to the phylum Chloroflexi. International Journal of Systematic and Evolutionary Microbiology 64:1859–1865

Spieck E, Bock E (2015) Nitrobacter. Bergey's manual of systematics of archaea and bacteria. John Wiley, Hoboken, New Jersey. https://doi.org/10.1002/9781118960608.gbm00803

Stein MM, Hrusch CL, Gozdz J, Igartua C, Pivniouk V, Murray SE, et al. (2016) Innate immunity and asthma risk in Amish and Hutterite farm children. New England Journal of Medicine 375:411–421

Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Applied and Environmental Microbiology 73:5261–5267

Wang Y, Naumann U, Wright ST, Warton DI (2012) Mvabund–an R package for model-based analysis of multivariate abundance data. Methods in Ecology and Evolution 3:471–474

Weiss S, Xu ZZ, Peddada S, Amir A, Bittinger K, Gonzalez A, et al. (2017) Normalization and microbial differential abundance strategies depend upon data characteristics. Microbiome 5:27

Wheeler K, Naurright J (2006) A global perspective on the environmental impact of golf. Sport in Society 9:427–443

White TJ, Bruns T, Lee S, Taylor J (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. Pages 315–322. In: PCR protocols: a guide to methods and applications. Vol 18. Academic Press, New York

White A, Sparrow B, Leitch E, Foulkes J, Flitton R, Lowe AJ, Caddy-Retalic S (2012) AusPlots rangelands survey protocols manual. The University of Adelaide Press, Adelaide, South Australia, Australia

Wubs E, van der Putten W, Bosch M, Bezemer TB (2016) Soil inoculation steers restoration of terrestrial ecosystems. Nature Plants 2:16107

Yan D, Mills JG, Gellie NJ, Bissett A, Lowe AJ, Breed MF (2018) High-throughput eDNA monitoring of fungi to track functional recovery in ecological restoration. Biological Conservation 217:113–120

Supporting Information

The following information may be found in the online version of this article:

Table S1. Soil physicochemical properties of the urban green space types at City of Playford.

Table S2. Correlated predictor variables were removed from the ordination model based on Spearman's correlation tests with a cut off of $|r| > 0.7$ or < -0.7 and variance inflation factor analysis (VIF).

Table S3. Two-tailed Spearman's correlation (rho) of predictor variables to CAP1 and CAP2 of the constrained analysis of principle coordinates (CAP).

Figure S1. Vegetation functional (growth form, NVISgf; NVIS is the National Vegetation Information System) and structural diversity (height class, NVISht) and species diversity (Richness; Shannon's, H'; and Effective Species Number (ESN), calculated as exp(H')) between the five urban green space vegetation types.

Figure S2. Alpha diversity of bacteria, fungi, archaea, and eukaryotes measured by ASV richness, Faith's phylogenetic diversity (PD), and Shannon's diversity (H').

Figure S3. Sampling locations and photos of urban green space vegetation types in the City of Playford, South Australia.

Supplement S1. R code script

Guest Coordinating Editor: Jeffrey Fehmi

Received: 29 November, 2019; First decision: 27 January, 2020; Accepted: 8 April, 2020

Appendix VI

Early markers of periodontal disease and altered oral microbiota are associated with glycemic control in children with type 1 diabetes (Jensen *et al.* 2020)

Received: 5 July 2020 | Revised: 16 November 2020 | Accepted: 1 December 2020
DOI: 10.1111/pedi.13170

ORIGINAL ARTICLE

ISPAD WILEY

Early markers of periodontal disease and altered oral microbiota are associated with glycemic control in children with type 1 diabetes

Emilija D. Jensen[1,2] | Caitlin A. Selway[3] | Gabrielle Allen[1,2] | Jana Bednarz[4] | Laura S. Weyrich[3,5] | Sam Gue[6] | Alexia S. Peña[7,8] | Jennifer Couper[7,8]

[1]Discipline of Paediatric Dentistry, Adelaide Dental School, University of Adelaide, Adelaide, South Australia, Australia

[2]Department of Paediatric Dentistry, Women's and Children's Hospital, North Adelaide, South Australia, Australia

[3]Department of Molecular and Biomedical Science, University of Adelaide, Adelaide, South Australia, Australia

[4]Adelaide Health Technology Assessment, School of Public Health, University of Adelaide, Adelaide, South Australia, Australia

[5]Department of Anthropology and Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, Pennsylvania, USA

[6]Private Practice, Dulwich, Dulwich, South Australia, Australia

[7]Discipline of Paediatrics and Robinson Research Institute, University of Adelaide, Adelaide, South Australia, Australia

[8]Department of Diabetes and Endocrinology, Women's and Children's Hospital, North Adelaide, South Australia, Australia

Correspondence
Emilija D Jensen, Discipline of Paediatric Dentistry University of Adelaide Level 10 Adelaide Health and Medical Sciences Building Corner of North Terrace and George Street Adelaide, SA 5000, Australia.
Email: emilija.jensen@adelaide.edu.au

Funding information
Australian Government Research Training Program Scholarship; Women's and Children's Hospital Foundation, Grant/Award Numbers: G0077S, G00801

Abstract

Objectives: To determine the relationship between periodontal disease and glycemic control in children with type 1 diabetes and to characterize the diversity and composition of their oral microbiota.

Methods: Cross-sectional study including children with type 1 diabetes recruited from clinics at the Women's and Children's Hospital (Australia). Participants had a comprehensive dental assessment, periodontal examination, and buccal and gingival samples collected for 16S rRNA sequencing.

Results: Seventy-seven participants (age 13.3 ± 2.6 years, 38 males, BMI z-score 0.81 ± 0.75) had a diabetes duration of 5.6 ± 3.9 years and median HbA1c of 8.5% (range 5.8–13.3), 69.4 mmol/mol (range 39.9–121.9). Thirty-eight (49%) had early markers of periodontal disease. HbA1c was positively correlated with plaque index (Rho = 0.34, P = 0.002), gingival index (Rho = 0.30, P = 0.009), bleeding on probing (Rho = 0.44, P = 0.0001) and periodontal pocket depth >3 mm (Rho = 0.21, P = 0.06). A 1% increase in HbA1c was independently associated with an average increase in bleeding on probing of 25% (P = 0.002) and with an increase in the rate of sites with pocket depth >3 mm of 54% (P = 0.003). Higher HbA1c was independently related to increased phylogenetic alpha diversity (P = 0.008) and increased compositional variation (beta diversity P = 0.02) in gingival, but not buccal, microbiota. Brushing frequency, plaque index, and gingival index had a significant effect on microbiota composition, independent of HbA1c.

Conclusions: Children with type 1 diabetes showed a continuous relationship between less favorable glycemic control and increased early markers of periodontal disease. Glycemic control was also related to the complexity and richness of the plaque microbiota, with diversity increasing as HbA1c levels increase.

KEYWORDS
adolescent, child, diabetes mellitus, type 1, microbiota, periodontal disease

1 | INTRODUCTION

Periodontal disease is a frequent complication in adults with type 2 diabetes, and both conditions are related to systemic inflammatory states. There is a bidirectional relationship between periodontal disease and diabetes; periodontal inflammation adversely affects glycemic control, which, in turn, increases the risk of periodontal disease.[1] The degree of chronic hyperglycemia is positively related to the prevalence and severity of periodontal disease in adults.[2] This relationship is not well researched in adults with type 1 diabetes, and there is minimal data in children.[3]

Chronic hyperglycemia augments the pro-inflammatory response in the periodontal environment. It affects key cellular functions in leukocytes, fibroblasts and osteoclasts, and increases proinflammatory cytokines, as previously described in individuals with diabetes and periodontal disease.[4] Elevated levels of advanced glycation end products (AGEs) and their receptor, RAGE, are detected in the gingival tissues of individuals with diabetes, which further contributes to increased severity of periodontal disease.[4]

The extent to which type 1 diabetes affects the composition of the periodontal microbiota is yet to be well-established. Studies characterizing oral microbiota in adults with type 1 diabetes and type 2 diabetes including controls and using 16S rRNA sequencing have found evidence of altered microbial composition with increased subgingival biodiversity.[5-7] To the authors' knowledge, there have been no published studies that characterize the oral microbiota in children with type 1 diabetes using 16S rRNA sequencing.

The present study had four main aims: (a) to characterize periodontal risk markers (plaque index (PI), gingival index (GI), of bleeding on probing (BOP) and of a pocket depth (PD) greater than 3 mm) in children and adolescents with type 1 diabetes; (b) to determine the relationship between glycemic control and periodontal risk markers (PI, GI, BOP, PD>3 mm); (c) to characterize the oral (buccal and gingival) microbiota diversity and composition; and (d) to determine the relationships between the oral microbiota and both glycemic control and periodontal risk markers. We hypothesized that higher levels of HbA1c would increase the frequency of early periodontal disease risk markers and alter the diversity and composition of the oral (buccal and gingival) microbiota.

2 | METHODS

2.1 | Subjects

Seventy-seven participants (39 females) were consecutively recruited from the Pediatric Diabetes Clinics at the Women's and Children's Hospital (Adelaide, South Australia). Recruitment occurred from February 2018 to February 2019 and was time limited and therefore a convenience sample; the last clinical examination was completed in March 2019. Inclusion criteria were those aged between 8 and 18 years at recruitment who had been previously diagnosed with type 1 diabetes by detectable islet cell autoantibodies. Exclusion criteria were those with a diagnosis of diabetes other than type 1 diabetes or inadequate English language skills to understand the information sheet. Participants who had an intercurrent fever or infection (n = 0), diabetic ketosis (n = 0) or who were taking a course of antibiotics (n = 0) on the scheduled day of the dental examination had their appointment rescheduled.

Informed written consent was obtained from all parents or guardians of participants younger than 16 years and from all participants who were 16 years or older on the day of dental examination at the Department of Pediatric Dentistry, Women's and Children's Hospital (Adelaide, South Australia). All participants under 16 years gave assent. The study was approved by the Women's and Children's Health Network Human Research Ethics Committee (HREC/17/WCHN/165).

2.2 | Oral examination protocol

2.2.1 | Clinical data

HbA1c values were obtained from the most recent measurement available from the participants' medical records, performed within 3 months of the dental examination date. The HbA1c levels of all participants were measured using a Vantage analyzer (Siemens Diagnostics, Camberley, UK), which shows very high correlation with DCCT standardized sample controls (r = 0.98). Details regarding type 1 diabetes diagnosis were obtained from medical records. Height, measured in a standardized manner using a calibrated stadiometer, and weight were measured at the time of the dental examination.

2.2.2 | Periodontal measurements

Familial, medical, and dental history were provided by the parents, guardians or participants, and included information regarding current home oral hygiene practices (including frequency of brushing, type of toothbrush, type of toothpaste, regularity of flossing). Dental and periodontal examination was performed by a trained and calibrated dentist in pediatric specialist training (E.J.) using a dental mirror, dental explorer and periodontal probe with dental chair lighting. Body mass index (BMI) z-score was calculated using the Children's Hospital of Philadelphia Research Institute pediatric z-score calculator (https://zscore.research.chop.edu/).

Periodontal risk markers were measured on six teeth (fully erupted first permanent molars in each quadrant, the right maxillary central incisor and left mandibular central incisor). Each periodontal parameter was recorded on six points per tooth (mesio-buccal, buccal, disto-buccal, mesio-lingual, lingual, disto-lingual) by the same operator (E.J). For each participant, values for each parameter were averaged across the six points and six teeth. The following periodontal risk markers were evaluated:

a. Plaque index was assessed using the system described by Silness and Loe[8] with a corresponding score from 0 to 3, scored as 0, no

plaque; 1, a film on the tooth surface observed using the probe; 2, moderate accumulation of soft deposits; 3, abundance of soft matter.

b. Gingival index score was assessed using the Loe and Silness Gingival Index System,[9] which was coded as 0, absence of inflammation; 1, mild inflammation; 2, moderate inflammation; 3, severe inflammation.

c. Bleeding on probing was defined as any bleeding in the coronal marginal gingiva after a World Health Organization (WHO) 621 probe was used on the bottom of the sulcus. The percentage BOP was calculated as the number of positive sites relative to the total measured sites.

d. Pocket depth was measured using a standardized manual WHO 621 probe with a controlled force in a walking step. The depth was measured from the base of the pocket to the coronal marginal gingiva. Participants were recorded as having no periodontal PD >3 mm or a count of pockets with PD values >3 mm.

2.3 | Microbiota protocol

2.3.1 | Sample collection

Two samples were collected from each participant by a practitioner (E.J.) wearing a clean mask and gloves; the first was collected from the right buccal mucosa, while the second was from the buccal gingival margin of the lower left first permanent molar. Microbiota samples were collected for 76 of the 77 participants, as sample collection vials were unavailable at the time of examination for one subject who was subsequently excluded from microbiota analyses only. In total, 152 samples were collected in sterile screw-cap vials and stored within 15 minute in a −80°C alarmed freezer. Frozen samples were thawed when required for DNA extraction in May 2019. A total of 147/152 samples (73 gingival, 74 buccal) were used for analysis, following omission of one gingival sample due to contamination, two samples from one individual due to unavailable data on periodontal markers and two samples (one gingival and one buccal) due to low sequencing depth.

2.3.2 | DNA extractions, 16S RNA amplification, and sequencing preparation

DNA extractions and amplification preparations were carried out in a laboratory designed to process low-biomass samples at the University of Adelaide, and the researcher (C.S.) wore a face mask, gloves, laboratory coat, and booties during all laboratory work. DNA was extracted from buccal and gingival samples using the Qiagen DNeasy PowerSoil extraction kit (Hilden, North Rhine-Westphalia, Germany) with specific modifications. Samples were added to the PowerBead tubes with C1 solution prior to attaching onto a vertex bed for 10 minute of mechanical lysis. Tubes were centrifuged and the supernatant was added to a new tube with the C2 solution. Removal of humic acids and inhibitors, binding of DNA to silica, cleaning of DNA and DNA elution was carried out as per manufacturer's instructions. Extraction blank controls (EBC; an extraction with no biological sample added) were also included to monitor background levels of DNA from the laboratory and reagents.

Extracts were then subjected to polymerase chain reaction (PCR) amplification, targeting the V4 region of the 16S ribosomal RNA (rRNA) gene (515F- AATGATACGGCGACCACCGAGATCTACAC TAT GGTAATT GT GTGCCAGCMGCCGCGGTAA; barcoded 806R-CAAG CAGAAGACGGCATACGAGAT XXXXXXXXXXXX AGTCAGTCAG CC GGACTACHVGGGTWTCTAAT, where 'X' represents the location of the unique 12 nucleotide barcode). Each 25 μL PCR reaction was performed in triplicate and contained 18.05 μL DNA-free water, 2.5 μL 10× High Fidelity PCR Buffer (Invitrogen, CA, USA), 1.0 μL $MgSO_4$ (50 mM), 0.2 μL dNTPs (25 mM Invitrogen), 0.25 μL Platinum® Taq DNA Polymerase High Fidelity (Invitrogen), 1.0 μL of each V4 primer (10 μM) and 1.0 μL of DNA extract. Cycling conditions consisted of 95°C for 6 minute (denaturation); 38 cycles of 95°C for 30 second, 50°C for 30 second, and 72°C for 90 s (annealing and extension); and a final extension at 60°C for 10 minute. No template controls (PCR reaction without sample) were included in each PCR batch.

Data for each triplicate reaction were pooled, and the presence of the 16S rRNA gene was verified by gel electrophoresis (2.5% agarose gel). Successful PCR products were quantified using a Qubit dsDNA HS assay (Thermo Fisher Scientific, MA) and pooled at equimolar concentrations into groups of approximately 30 samples. Pools were then cleaned (AxyPrep, Fisher Biotec Australia, WA, Australia) and quantified again (Tapestation 2200, Aligent, CA). These pools were combined at equimolar concentrations to create a final sequencing pool. Two sequencing runs were performed at Australian Cancer Research Foundation, Cancer Genomics Facility (Adelaide, Australia), using 2 × 150 bp MiSeq Reagent Kits v2 (Illumina, CA).

2.3.3 | Bioinformatics and biostatistics

a. 16S rRNA sequencing information.

Using the QIIME2 software,[10] DNA sequences were demultiplexed, cleaned, and the taxonomy assigned. First, sequences were demultiplexed based on the unique barcode identifier fused to DNA in PCR amplifications. Next, paired-end sequences were joined using vectorized search tool. After quality assessment, sequences were trimmed to 250 bp and denoised using Deblur.[11] Following this, a SEPP insertion tree was created, and sequences were assigned using the SILVA database (v132; 16S rRNA gene 515–806). Using Decontam,[12] contaminant amplicon sequence variants (ASVs; Appendix 1) were identified from EBCs (prevalence threshold set to 0.5) and removed from biological samples. Low abundance sequences (<11 total sequences per ASV) were also removed.

b. Diversity and composition analyses.

Diversity and compositional analyses were performed in QIIME2. At a rarefied depth of 5000 sequences, diversity (alpha diversity, the number and richness of species within the oral microbiota) was calculated using observed species (microbial richness) and Faith's

phylogenetic diversity metrics. Similarly, composition (beta diversity, the community structure and how the oral microbiota differs between other sites) was calculated using Bray-Curtis (dissimilarity measure of abundance and difference in species populations between sites) and unweighted UniFrac (dissimilarity measure of phylogenetic distances between observed organisms and relative relatedness of the microbiota community) metrics at a rarefied depth of 5000 sequences.

2.4 | Statistical analysis

Participant characteristics and periodontal disease markers were summarized using mean and standard deviation (SD) or median and range for continuously measured variables, and frequencies and percentages for categorical variables. Univariable monotonic relationships between HbA1c (measured continuously) and each periodontal disease marker were described using Spearman's Rank Correlations. The effects of the HbA1c (measured continuously) on PI and GI were assessed using multivariable linear regression, controlling for the individual's age and gender, diabetes duration, BMI, mixed dentition status, and self-reported frequency of tooth brushing. Predicted values for each marker at varying levels of HbA1c were estimated post-hoc, with all other predictors held at their mean values. Model adequacy was assessed by visual inspection of residual plots and leverage plots were used to identify potential outliers. Sensitivity analyses involved examining the effect of the removal of disproportionately influential observations on the regression results. Such observations were defined as those with Cook's distance values above the threshold of 4 divided by the number of observations in the model ($4/77 \approx 0.05$).[13]

There was overdispersion in the number of sites exhibiting BOP and in the number of sites with PD > 3 mm. Therefore, the effect of HbA1c on the number of sites exhibiting BOP and on the number of sites with PD > 3 mm was assessed using negative binomial regression. Adjustment factors were the same as those used for PI and GI, and included the individual's age, gender, diabetes duration, BMI, mixed dentition status and self-reported frequency of brushing. The predicted number of affected sites across varying levels of HbA1c were estimated post-hoc, with all other predictors held at their mean values. All analyses were performed using Stata (Version 15, Stata Corp, TX), and the level of statistical significance was set at 0.05.

Biostatistical analyses for categorical and continuous microbiome data were performed using the QIIME2 (version 2019.7) and R (version 3.6.1, Vienna, Austria), respectively. Significant associations between diversity and categorical metadata were found using a Kruskal Wallis pairwise statistical test, whereby false discovery rate adjusted *P*-values (q-values) were reported. General linear models were performed for correlations between diversity and continuous metadata. For composition, adonis was used to measure the amount of variation explained by categorical and continuous variables. Underlying and confounding factors were assessed throughout.

3 | RESULTS

3.1 | Glycemic control and periodontal risk markers

A total of 77 participants were enrolled for oral examination. Participant demographics and clinical dental and periodontal parameters are presented in Table 1. 38 participants (49%) had early markers periodontal disease indicated by having at least one site with a PD > 3 mm.

HbA1c was positively associated with PI (Rho = 0.34, $P = 0.002$). Controlling for the effects of age, gender, diabetes duration, BMI, mixed dentition status and self-reported frequency of tooth brushing, a 1% increase in HbA1c was associated with an average increase of 0.1 units in PI (95% CI [0.025, 0.167]; $P = 0.008$). There were seven observations with Cook's distance >4/77. When these potentially influential data points were omitted from the model, the effect of HbA1c on PI was unchanged and the *P*-value remained significant. HbA1c was also positively associated with GI (Rho = 0.30, $P = 0.009$). The adjusted effect of a 1% increase in HbA1c was an average increase in GI of 0.1 units (95% CI [−0.003, 0.122]; $P = 0.061$) given all other factors are held constant.

TABLE 1 Summary of study participant characteristics

Characteristic	Participants (n = 77)
Age (years)	13.3 ± 2.6
Male gender (n, %)	37 (48.1)
Duration of T1D (years)	5.62 ± 3.89
HbA1c (%, median, range)	8.5 (5.8–13.3)
HbA1c (mmol/mol, median, range)	69.41 (39.89–121.87)
BMI z-score	0.82 ± 0.75
Dentition status (n, %)	
Mixed	28 (36.4)
Permanent	49 (63.6)
Self-reported frequency of brushing	
No tooth brushing	2 (2.6)
Brushes sporadically	16 (20.8)
Brushes once a day	31 (40.3)
Brushes twice a day	28 (36.4)
Plaque index	0.93 ± 0.51
Gingival index	0.67 ± 0.42
Bleeding on probing (%)[a]	21.39 ± 19.44
Pocket depth >3 mm (count; n, %)	38 (49.35)
Pocket depth >3 mm (count; n, %)	
0 sites	39 (50.7)
1 site	12 (15.6)
2 sites	4 (5.2)
3 sites	7 (9.1)
≥ 4 sites	15 (19.5)

Note: Mean ± SD unless specified. BMI, body mass index; T1D, type 1 diabetes.
[a]Percentage BOP calculated from the number of positive sites from the total measured sites.

HbA1c was positively associated with BOP (Rho = 0.44, P = 0.0001) and PD > 3 mm (Rho = 0.21, P = 0.06). Controlling for potential confounding factors, a 1% increase in HbA1c was associated with an average increase in BOP of 25% (adjusted IRR = 1.25; 95% CI [1.08, 1.45]; P = 0.002) and an increase in the rate of sites with PD > 3 mm of 54% (adjusted IRR = 1.54; 95% CI [1.16, 2.04]; P = 0.003).

3.2 | Glycemic control and microbiota

HbA1c was significantly correlated with phylogenetic alpha diversity of the oral microbiota for gingival samples (Faith PD: t = 2.732, P = 0.008) but not for buccal samples (Figure 1). Similarly, HbA1c was significantly associated with beta diversity composition in gingival samples (Bray Curtis: R^2 = 0.026, P = 0.02; unweighted UniFrac: R^2 = 0.041, P = 0.006) but not in buccal samples (Figure 2).

For gingival samples, the beta diversity compositional differences explained by HbA1c were not confounded by age, gender, diabetes duration, BMI, PI, GI, BOP, PD, presence of mixed or permanent dentition, frequency of brushing, frequency of flossing or use of manual or electric toothbrush. This was also observed for both Bray-Curtis and unweighted UniFrac metrics.

Brushing frequency, PI and GI significantly influenced the oral microbiota composition, independent of HbA1c. Children who brushed less than once a day had a microbial composition that was dissimilar to children who brushed at least once per day. All children brushed with fluoridated toothpaste. Additionally, age contributed to composition as measured with Bray-Curtis, but not when measured with unweighted UniFrac. The presence of PD >3 mm contributed to composition, as measured with unweighted UniFrac but not Bray-Curtis, supporting species differences in the mouth of children with and without periodontal pockets (Table 2). Factors that showed significance were then tested together to identify any independent or dependent influences on the microbiota are shown in Appendix 2. Although no interactions were significant, there was a shared amount of compositional variation between PI and GI, suggesting that these categories may be dependent.

Correlation tests of Kendall, Spearman, and Pearson found no species nor genera that were strongly correlated with continuous HbA1c, PI, GI, or brushing frequency. An analysis of composition of microbiomes (ANCOM) correlation test was used for categorical factors of brushing and PD >3 mm. An uncultured *Alloprevotella* ASV was increased in children who had one or more pockets with PD >3 mm.

4 | DISCUSSION

We report for the first time a continuous relationship between glycemic control measured by HbA1c and the early markers of periodontal

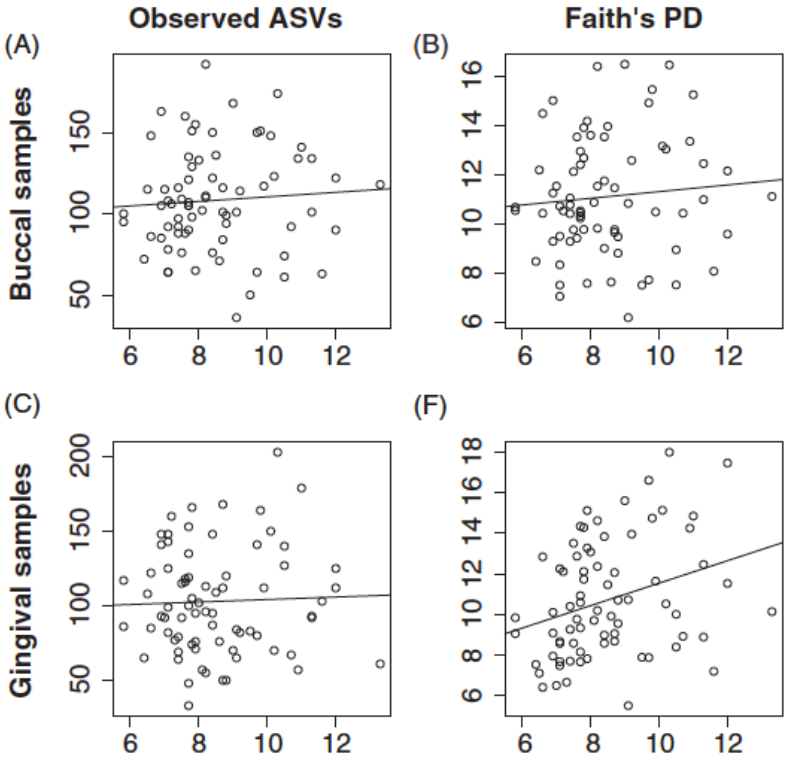

FIGURE 1 A positive correlation was observed between HbA1c and alpha diversity for gingival samples (Faith PD: t = 2.732, P = 0.008, n = 73) but not for buccal samples. Alpha diversity was measured using observed amplicon sequence variants (ASV) and Faith's phylogenetic diversity metrics for both (A-B) buccal samples and (C-D) gingival samples

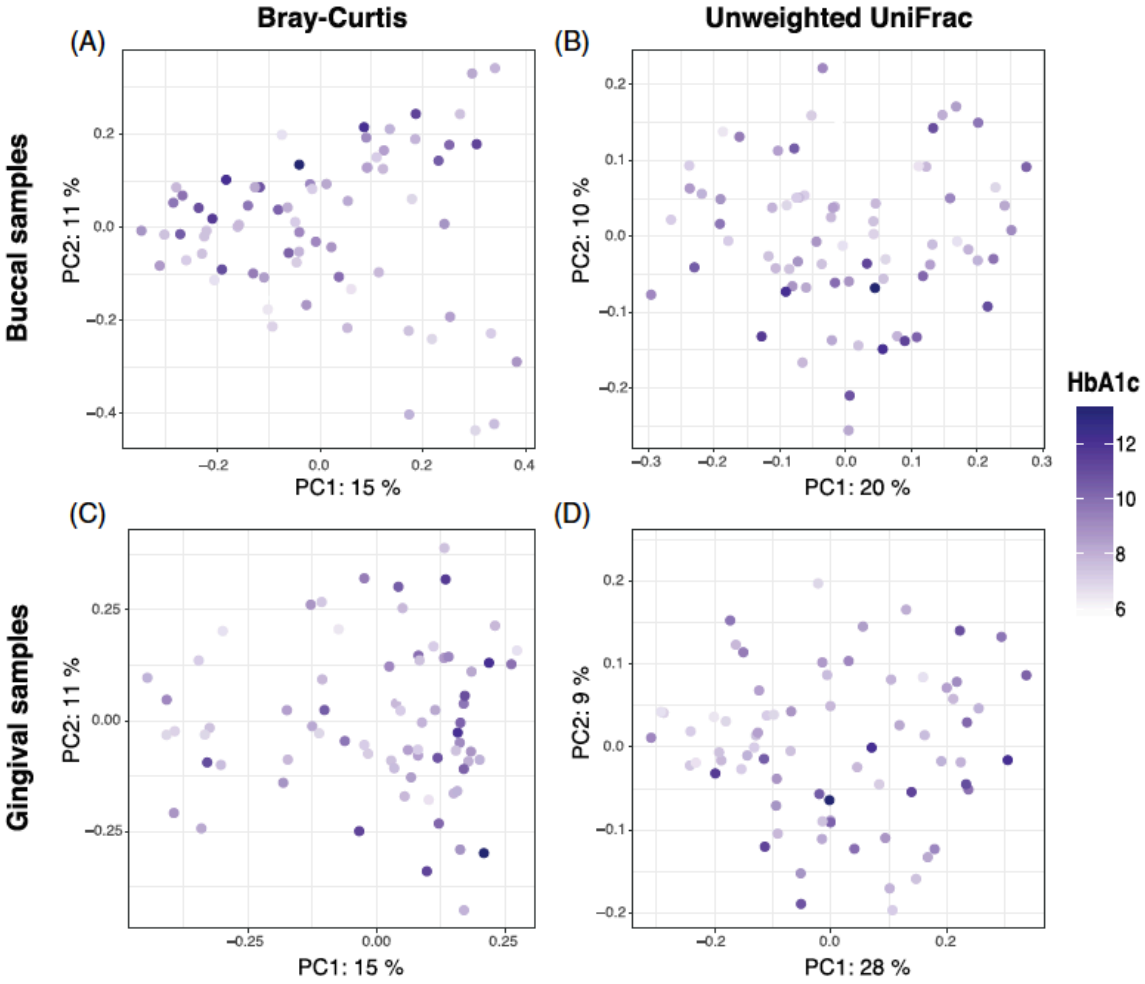

FIGURE 2 A correlation between HbA1c and beta diversity was observed for gingival samples (Bray Curtis: $R^2 = 0.026$, $P = 0.02$; unweighted UniFrac: $R^2 = 0.041$, $P = 0.006$) but not buccal samples. Distance matrices were calculated using Bray-Curtis and unweighted UniFrac metrics, and principle coordinate analysis plots were generated for both (A-B) buccal samples and (C-D) gingival samples. For each analysis, the first two principle components, PC1 and PC2, were displayed

disease, namely PI, GI, BOP and PD >3 mm, in children and adolescents with type 1 diabetes. Glycemic control was also related to the complexity and richness of the microbiota residing within the gingival plaque. Both alpha diversity and beta diversity composition differences changed with increasing HbA1c levels. Finally, lower brushing frequency, higher plaque index and higher gingival index were all related to increased variation in microbial composition, independent of HbA1c.

Consistent with our hypotheses, we found evidence that worsening glycemic control is associated with increased severity of early markers of periodontal disease in children and adolescents with type 1 diabetes, and to our knowledge, this is the first description of these continuous relationships in children. Periodontal literature commonly uses categorical values for HbA1c as seen in the new classification, staging, and grading of periodontitis.[14] A moderate progression of periodontitis is expected when HbA1c<7% in individuals with diabetes, and rapid progression of periodontitis is expected for those with HbA1c≥7%.[14] However, dichotomizing glycemic control using a cut-point is associated with a loss of information and statistical power that may underestimate the extent of the impact of changes in HbA1c on each periodontal risk marker.[15] The use of the continuous HbA1c variable in this study estimated the increase in each periodontal risk marker for every 1% increase in HbA1c, so that we could show that worsening glycemic control was associated with clinically relevant periodontal change.

Alpha diversity (richness) within the gingival plaque increased as the participants' HbA1c levels increased, indicating a more complex

TABLE 2 Permutational multivariate analysis of variance using adonis to examine factors that contribute to the composition of oral microbiota after controlling for HbA1c

	Bray-Curtis		Unweighted UniFrac	
Factors	R2	Pr (>F)	R2	Pr (>F)
HbA1c	0.026	0.022	0.041	0.007
Age	0.022	0.049	0.018	0.128
HbA1c:age	0.010	0.721	0.019	0.158
HbA1c	0.026	0.011	0.041	0.003
Brushing	0.028	0.014	0.024	0.056
HbA1c:brushing	0.014	0.354	0.009	0.798
HbA1c	0.026	0.020	0.041	0.004
PI	0.035	0.001	0.020	0.136
HbA1c:PI	0.009	0.830	0.010	0.711
HbA1c	0.026	0.027	0.041	0.002
GI	0.031	0.003	0.047	0.001
HbA1c:GI	0.012	0.521	0.009	0.703
HbA1c	0.026	0.021	0.041	0.002
PD	0.018	0.136	0.029	0.037
HbA1c:PD	0.017	0.199	0.012	0.466

Abbreviations: GI, gingival index; PD, pocket depth; PI, plaque index.

microbiota. In contrast, a small study examining obese adults with type 2 diabetes with comparable glycemic control showed no correlation between HbA1c and salivary alpha diversity.[16] These changes were only observed in the gingival samples taken from the plaque on the lower left first molar, and were not detected in the buccal samples taken from the right buccal mucosa. This suggests that the alpha diversity changes in the gingival plaque relate specifically to periodontal risk markers. As HbA1c increased, a change in composition was observed for both species abundance and species differences. This is consistent with other microbiota 16S rRNA sequencing studies that demonstrate a microbial shift as a result of type 2 diabetes in adults, but is a unique finding among individuals with type 1 diabetes in all age groups.[5,6,17] Previous reports found an association with HbA1c and microorganisms that were associated with periodontal disease in children with type 1 diabetes using the DNA–DNA hybridization method of quantitative estimates, while a microbiologically healthy periodontium was associated with good glycemic control.[18]

Plaque accumulation in individuals is related to their frequency and efficacy of tooth brushing.[19] In turn, gingival inflammation is a response to plaque accumulation.[14] Consistent with this, brushing frequency, plaque index and gingival index were all related to increased variation in microbial composition, independent of HbA1c. Children who brushed less than once a day had a microbial composition that was dissimilar to children who brushed at least once per day. *Alloprevotella* was increased in children who had one or more PD>3 mm, which is consistent with studies of subgingival plaque of adults diagnosed with periodontitis.[20] Mean counts of microbiota in supragingival and subgingival plaque have been reported to be higher in individuals with periodontitis and without diabetes.[21] Our study extends these findings, where children with early periodontal pockets with PD >3 mm had species differences relative to children who had no PD >3 mm.

Our study has several strengths. First, we performed a comprehensive evaluation of multiple early risk markers of periodontal disease (PI, GI, BOP, and PD) on six sites from six teeth. Specifically, PI and GI have been validated as adequate representations of a child's amount of plaque and degree of gingival inflammation, respectively,[22] and BOP and PD are other important clinical measurements that are used to identify periodontal disease.[14] Second, all measurements were taken by a single experienced examiner (E.J.), and complete periodontal marker data were obtained from all but one participant in the study. Third, the analysis of periodontal markers was complemented by analysis of the microbiota of gingival and buccal sites, which was completed in 96.7% of samples.

Our study also has some limitations. Although 77 participants is a relatively large cohort for the investigation of the microbiota compared to other studies in adults with diabetes,[5,6] this cross sectional study lacked a pediatric control group. The cross sectional design precluded the study from showing any causality between glycemic control and periodontal risk markers. However, the range of HbA1c levels was wide and we were able to detect novel, independent relationships between glycemic control and periodontal disease markers and microbiota diversity. The majority of our participants did not achieve target HbA1c levels, with a higher median than that of the whole pediatric clinic (n = 750, 8.5 vs 8.0% [69 vs 64 mmol/mol]). The temporal relationship between HbA1c, periodontal risk markers, and microbiota diversity, and whether these early changes are reversible with improved glycemic control, are best explored with prospective longitudinal studies.

Periodontal disease increases the systemic inflammatory response, which may augment the accelerated risk of cardiovascular disease in children with type 1 diabetes.[14,23,24] Risk assessment in periodontal disease prevention and treatment is critical and can only be achieved by a dental practitioner. Review by a dental practitioner should occur at a minimum of 6–12 monthly or earlier according to the periodontal and dental caries risk profile. Our findings that an increase in HbA1c were associated with a moderate increase in PI, GI, BOP, and PD >3 mm and changes to the diversity and composition of the oral microbiota, underscoring the need for developing a personalized risk assessment strategy for these children and adolescents. Dental practitioners should recognize that periodontal risk factors are higher in this group of patients, thus preventative advice including oral hygiene instruction, education about periodontal risk management, and prophylactic management of plaque accumulation, gingival inflammation and early periodontal pockets is paramount. Pediatricians and endocrinologists should be encouraged to action early referral of children and adolescents with type 1 diabetes to a dental practitioner for initial and ongoing review of dental and periodontal health.

ACKNOWLEDGMENTS

This research was supported by Women's and Children's Hospital Foundation G00801, G0077S. CAS was supported by an Australian

Government Research Training Program Scholarship. We would also like to thank the staff at the Australian Cancer Research Foundation Cancer and Genomics Facility for troubleshooting and providing expertise with the sequencing data. We are grateful to the children who participated in this study and their families.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/pedi.13170.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Emilija D Jensen, Gabrielle Allen, Alexia S Peña, Jennifer Couper, Sam Gue, Caitlin A Selway and Laura S Weyrich designed the research study, Emilija D Jensen recruited participants, obtained consent and performed the research, Jana Bednarz analyzed the periodontal disease data, Caitlin A Selway performed the microbiota laboratory work and analyzed the microbiota data, Emilija D Jensen wrote the initial draft of the manuscript Jennifer Couper, Alexia S Peña, Gabrielle Allen, Jana Bednarz, Caitlin A Selway, Laura S Weyrich and Sam Gue. critically revised the manuscript. All authors read and approved the final manuscript.

ORCID

Emilija D. Jensen https://orcid.org/0000-0001-9462-2596
Alexia S. Peña https://orcid.org/0000-0002-6834-4876
Jennifer Couper https://orcid.org/0000-0003-4448-8629

REFERENCES

1. Preshaw PM, Alba AL, Herrera D, et al. Periodontitis and diabetes: a two-way relationship. *Diabetologia*. 2012;55(1):21-31.
2. Mealey BL, Ocampo GL. Diabetes mellitus and periodontal disease. *Periodontol 2000*. 2007;44(1):127-153.
3. Lappin DF, Robertson D, Hodge P, et al. The influence of glycated hemoglobin on the cross susceptibility between type 1 diabetes mellitus and periodontal disease. *J Periodontol*. 2015;86(11):1249-1259.
4. Taylor JJ, Preshaw PM, Lalla E. A review of the evidence for pathogenic mechanisms that may link periodontitis and diabetes. *J Clin Periodontol*. 2013;40:S113-S134.
5. Casarin R, Barbagallo A, Meulman T, et al. Subgingival biodiversity in subjects with uncontrolled type-2 diabetes and chronic periodontitis. *J Periodontal Res*. 2013;48(1):30-36.
6. Zhou M, Rong R, Munro D, et al. Investigation of the effect of type 2 diabetes mellitus on subgingival plaque microbiota by high-throughput 16S rDNA pyrosequencing. *PLoS One*. 2013;8(4):e61516.
7. De Groot PF, Belzer C, Aydin Ö, et al. Distinct fecal and oral microbiota composition in human type 1 diabetes, an observational study. *PLoS One*. 2017;12(12):e0188475.
8. Silness J, Loe H. Periodontal disease in pregnancy II. Correlation between oral hygiene and periodontal condition. *Acta Odontol Scand*. 1964;22:121-135.
9. Loe H, Silness J. Periodontal disease in pregnancy I. Prevalence and severity. *Acta Odontol Scand*. 1963;21:533-551.
10. Bolyen E, Rideout JR, Dillon MR, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol*. 2019;37(8):852-857.
11. Amir A, McDonald D, Navas-Molina JA, et al. Deblur rapidly resolves single-nucleotide community sequence patterns. *ASM*. 2017;2(2):e00191-e00116.
12. Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome*. 2018;6(1):226.
13. Bollen KA, Jackman RW. Regression diagnostics: an expository treatment of outliers and influential cases. *Sociol Methods Res*. 1985;13(4):510-542.
14. Tonetti MS, Greenwell H, Kornman KS. Staging and grading of periodontitis: framework and proposal of a new classification and case definition. *J Periodontol*. 2018;89:S159-S172.
15. Altman DG, Royston P. The cost of dichotomising continuous variables. *BMJ*. 2006;332(7549):1080.
16. Tam J, Hoffmann T, Fischer S, Bornstein S, Graessler J, Noack B. Obesity alters composition and diversity of the oral microbiota in patients with type 2 diabetes mellitus independently of glycemic control. *PLoS One*. 2018;13(10):e0204724.
17. Ganesan SM, Joshi V, Fellows M, et al. A tale of two risks: smoking, diabetes and the subgingival microbiome. *ISME J*. 2017;11(9):2075-2089.
18. Merchant AT, Nahhas GJ, Wadwa RP, et al. Periodontal microorganisms and cardiovascular risk markers in youth with type 1 diabetes and without diabetes. *J Periodontol*. 2016;87(4):376-384.
19. Lang NP, Cumming BR, Löe H. Toothbrushing frequency as it relates to plaque development and gingival health. *J Periodontol*. 1973;44(7):396-405.
20. Chen C, Hemme C, Beleno J, et al. Oral microbiota of periodontal health and disease and their changes after nonsurgical periodontal therapy. *ISME J*. 2018;12(5):1210-1224.
21. Socransky SS, Haffajee AD. Periodontal microbial ecology. *Periodontol 2000*. 2005;38(1):135-187.
22. Goldberg P, Matsson L, Anderson H. Partial recording of gingivitis and dental plaque in children of different ages and in young adults. *Community Dent Oral Epidemiol*. 1985;13(1):44-46.
23. Dregan A, Charlton J, Chowienczyk P, Gulliford MC. Chronic inflammatory disorders and risk of type 2 diabetes mellitus, coronary heart disease, and stroke: a population-based cohort study. *Circulation*. 2014;130(10):837-844.
24. Tonetti MS. periodontitis and systemic diseases: proceedings of a workshop jointly held by the European Federation of Periodontology and American Academy of periodontology; 11-14, Segovia, Spain, 2012

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Jensen ED, Selway CA, Allen G, et al. Early markers of periodontal disease and altered oral microbiota are associated with glycemic control in children with type 1 diabetes. *Pediatr Diabetes*. 2020;1–8. https://doi.org/10.1111/pedi.13170