Breathing in Health: Assessing the potential of outdoor environments to transfer health-beneficial butyrate-producing bacteria to people

By

Joel E. Brame

N.D., M.Ed., B.A.

Thesis
Submitted to Flinders University
for the degree of

Doctor of Philosophy

College of Science and Engineering

16th of October 2024

TABLE OF CONTENTS

ABSTRACT

Background

Butyrate-producing bacteria ferment organic matter including fibres and amino acids into butyrate, a short-chain fatty acid that has important roles in human health and soils. Human gut bacterial assemblages include butyrate-producing bacteria that thrive in the anaerobic gut conditions. However, many physical and mental health conditions have been associated with a reduction in colonic butyrate-producing bacteria. Outdoor greenspaces may be rich reservoirs of butyrate-producing bacteria that could transfer to humans and supplement their abundances, especially for urban residents with less exposure to nature and biodiversity. Yet, outdoor butyrate producer abundances and the ecological characteristics that associate with them remain poorly resolved. In this thesis, I identify and describe the abundances of butyrate-producing bacteria in outdoor environments, particularly urban greenspaces, and the ecological conditions that associate with their abundances.

Methods and Results

I performed a literature search and found that the sources of gut colonisation of butyrate-producing bacteria after birth, particularly outdoor sources, remain largely unknown (= Chapter 1). I then used *in-silico* database interrogation and analyses of global shotgun metagenomic samples and Australian 16S rRNA soil samples to associate environmental characteristics with read abundances of human-accessible butyrate-producing bacteria. Butyrate-producing bacterial reads were widespread in environments such as soils and aquatic biomes throughout the world. In soils across continental Australia, butyrate producer read abundances were highest in temperate urban hinterlands and seasonally productive sandy croplands and associated strongly with iron (= Chapter 2). I then obtained soil shotgun metagenomic data and air 16S rRNA data from replicates (total $n = 33$) of two types of greenspaces (sports fields and nature parks) across metropolitan Adelaide, Australia, to describe the effects of ecological characteristics on aerobiome bacterial diversity and composition,

soil butyrate-producing bacterial read abundances, butyrate production terminal genes, and soil butyrate concentrations. The aerobiomes of sports fields at a 0.5 m sampling height were more influenced by soil iron and pH and rainfall than by the surrounding woody plant species diversity (= Chapter 3). Sports fields had more genes for butyrate production and higher soil butyrate concentrations than nature parks. The aerobiomes of sports fields at a 0.5 m sampling height were more influenced by soil iron and pH and rainfall than by the surrounding woody plant species diversity (= Chapter 3; Figure 1). Soil conditions, especially iron, had strong positive effects on both butyrate-related genes and soil butyrate concentrations (= Chapter 4). Finally, in a Māori cohort human exposure trial, I obtained and analysed air and nasal 16S rRNA taxonomic read abundances to quantify the transfer of airborne bacteria onto people taking a walk in an amenity grassland park (= Chapter 5). I found that 30-minute exposure periods to urban greenspaces did not create consistent or notable changes in the nasal microbiome of visitors.

Overall, we describe the influence of several ecological conditions (i.e., greenspace type, woody plant species diversity, and soil physicochemical parameters) on the abundances and activity of butyrate-producing bacteria in soils and aerobiome diversity. Our findings contribute important new insights into the role of sports fields as key exposure reservoirs of butyrate and butyrate producing bacteria. These findings provide opportunities for landscape designers, urban planners, ecologists, and public health experts to work together on new ways to support human health via urban greenspaces.

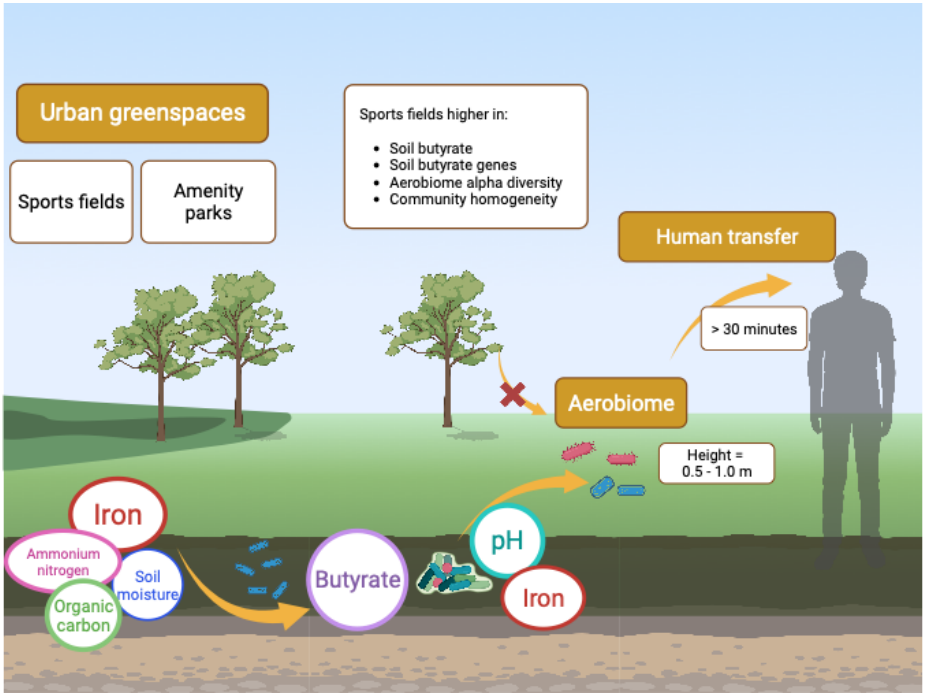


Figure 1. Graphical abstract showing overall results from Chapters 3-5.

DECLARATION

I certify that this thesis:

1. does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university

2. and the research within will not be submitted for any other future degree or diploma without the permission of Flinders University and

3. to the best of my knowledge and belief, does not contain any material previously published or written by another person except where due reference is made in the text.

Joel E. Brame

January, 2024

LIST OF SUPERVISORS

Primary supervisor: Associate Professor Martin Breed

Flinders University

College of Science and Engineering

Associate supervisors: Dr. Craig Liddicoat

Research Fellow

Flinders University

College of Science and Engineering

Professor Catherine Abbott

Flinders University

College of Science and Engineering

ACKNOWLEDGEMENTS

I am grateful for the funding from the Flinders Foundation, Australian Research Council [grants LP190100051], the New Zealand Ministry of Business Innovation and Employment [grant UOWX2101], the Royal Society of South Australia, and a Project Grant from the Health Research Council of New Zealand. Their generosity made this research possible.

I would like to acknowledge the dedicated supervision by my primary supervisor Associate Professor Martin Breed. From start to finish, his supervision has been exceptional and made a substantial difference in the success of the projects.

Tremendous gratitude goes to my associate supervisors Dr. Craig Liddicoat and Professor Cathy Abbott. Their involvement in this PhD has steadily pushed it toward success, and I am grateful for their countless contributions in writing, coding, laboratory work, and more.

Many thanks go to additional folks who have made substantial contributions during the PhD: Christian Cando-Dumancela, Dr. Jake Robinson, Professor Rob Edwards, Professor Liz Dinsdale, Dr. Isaac Warbrick, Dr. Deborah Heke, Dr. Helen Harrison, Dr. Russell Fuller, Dr. Michael Doane, Dr. Roger Yazbek, Dr. Nicolas Gauthier, Dr. Sunita Ramesh, and more.

I would like to thank my family and friends for their endless patience and support during this long and challenging PhD. Special thanks go to the incredible members my research group Frontiers of Restoration Ecology (the FORE Lab), including but not limited to: Chris, Carl, Shawn, Riley, Tarryn, Nicole, and Natalie. I could not have asked for a more dedicated group of people to co-endure the everyday rigours of PhD life. I wish them all the success in the world.

Published and submitted works

Brame JE, Liddicoat C, Abbott CA, Breed MF. The potential of outdoor environments to supply beneficial butyrate-producing bacteria to humans. 2021. *Science of The Total Environment*. doi: 10.1016/j.scitotenv.2021.146063.

Brame JE, Liddicoat C, Abbott CA, Edwards RE, Gauthier NE, Robinson JA, Breed MF. Towards the biogeography of butyrate-producing bacteria. 2023. *bioRxiv* (under review in *Ecology and Evolution*). doi: 10.1101/2022.10.07.510278.

Brame JE, Liddicoat C, Abbott CA, Cando-Dumancela C, Robinson JA, Breed MF. Urban greenspace aerobiomes are shaped by soil conditions and land cover type. 2024. *bioRxiv* (under review in *Microbial Ecology*), doi: 10.1101/2024.01.12.575340.

Brame JE, Liddicoat C, Abbott CA, Cando-Dumancela C, Robinson JA, Breed MF. Urban sports fields support higher levels of soil butyrate and butyrate-producing bacteria than urban nature parks. (in final stages of co-author review prior to submission to *ISME Journal*).

Brame JE, Warbrick I, Heke D, Liddicoat, C, Breed MF. Short-term passive greenspace exposures have little effect on nasal microbiomes: a cross-over exposure study of a Māori cohort. 2024. *bioRxiv* (under review in *Environmental Research*). doi: 10.1101/2024.01.17.576148.

GLOSSARY OF TERMINOLOGY

ANCOMBC: Analysis of Compositions of Microbiomes with Bias Correction

ANOVA: Analysis of Variance

ASV: amplicon sequence variant

BPC: butyrate production capacity

BPC_{16S}: butyrate production capacity for samples with 16S rRNA amplicon data

BPC_{meta}: butyrate production capacity for samples with shotgun metagenomic sequence data

CCA: Canonical correlation analysis

CoA: Coenzyme A

CSBP: Cuming Smith British Petroleum

DNA: deoxyribonucleic acid

EC: enzyme commission

eDNA: environmental DNA

EPC: Ethanol Production Capacity

GC-MS: gas chromatography – mass spectrometry

GIS: geographical information system

GTDB: Genome Taxonomy Database

IgA: immunoglobulin A

IL: interleukin

IMG/M: Integrated Microbial Genomics/Metagenomics

μM: micromolar

mM: millimolar

NCBI: National Center for Biotechnology Information

PC: principal component

PCA: principal components analysis

PERMANOVA: Permutational Multivariate Analysis of Variance

pH: potential of hydrogen

pM: picomolar

RCF: relative centrifugal force

rRNA: ribosomal ribonucleic acid

SCORPAN variables: S=soil; C=climate; O=organisms; R=relief; P=parent material; A=age; N=spatial location

TGF: transforming growth factor

Treg: regulatory T cell

tRNA: transfer ribonucleic acid

zOTU: zero-radius operational taxonomic unit

GENERAL INTRODUCTION

Establishing the interlinkages between exposure to a biodiverse environment, and the associated health-promoting butyrate-producing bacteria – with future development and research – should help to improve global public health initiatives and have a lasting positive impact on public health and biodiversity conservation.

Brame et al, 2021. Science of the Total Environment.

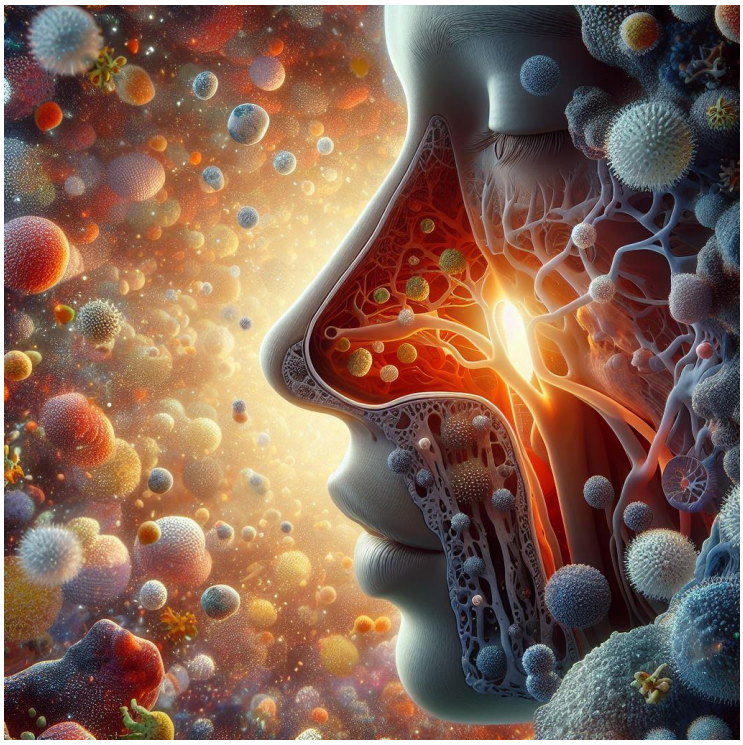


Image generated via Bing AI

General introduction

Next-generation sequencing technology has created new opportunities for health-associated microbiome research, including associations of diseases with human microbiome profiles. With an interest in autoimmune diseases, I reviewed the literature to identify which microbiome

characteristics had the greatest potential to modulate diseases. Butyrate-producing bacterial abundances had one of the strongest associations with autoimmune diseases. Even more, a reduction in abundances of butyrate-producing bacteria, and butyrate in gut (faecal) samples, has been linked with physical and mental health and a wide range of diseases, and that list continues to grow each year. Thus, with an interest in discovering ways to supplement butyrate producers and potentially restore health, I endeavoured to learn as much as possible about the sources of exposure to butyrate-producing bacteria to answer the question, "Where do they come from"?

The first chapter of this thesis provides much of the background on butyrate-producing bacteria. I synthesised a literature search about butyrate-producing bacteria and the potential for outdoor environments to be sources of human exposure. I found that substantial ongoing research is examining the molecular mechanisms of the intestinal uptake and utilisation of butyrate, as well as the role of dietary fibre as a feedstock for the butyrate-producing bacteria. However, the sources of human exposure and assemblage of gut butyrate-producing bacteria were substantial knowledge gaps. Given that (a) butyrate-producing bacterial abundances are reduced in many health conditions, (b) no cause of that reduction has yet been elucidated, and (c) humans obtain these bacteria during routine life events (i.e., no special inoculations are required), it became plausible that focused environmental exposures to these bacteria could potentially supplement their abundances in the human gut (**Figure 1**). Therefore, in subsequent thesis chapters, I sought to characterise the abundances of butyrate-producing bacteria in the environment, particularly the natural outdoor environment, investigate which ecological characteristics associate with their abundances, and evaluate the potential for airborne bacteria to transfer onto humans during urban greenspace exposures. This new knowledge could provide insight on the contributions of outdoor environments to the gut abundances and assemblages of butyrate-producing bacteria.

Figure 1. Graphical abstract from Brame et al (2021).

Thesis aims and objectives

With the background knowledge synthesised, I planned the next data chapters to address the following aims:

(1) characterise the abundances of butyrate-producing bacteria in outdoor environments (Chapter 2),

(2) identify the ecological characteristics that associate with their abundances (Chapters 2, 3, and 4), and

(3) evaluate the potential for airborne bacteria to transfer onto humans visiting urban greenspaces (Chapters 3 and 5).

To facilitate these aims, two additional research method objectives were required:

(1) create formulas for a Butyrate Production Capacity score that could be calculated for each sample to facilitate a rapid comparison of estimated potential to house butyrate-producing bacteria, and

(2) develop a method to extract and measure short-chain fatty acids from soils.

Thesis data chapters

In Chapter 2, we studied 14,000 global shotgun metagenomic and 1,200 Australia-wide 16S rRNA amplicon datasets to better understand the distribution of butyrate-producing bacteria. The sample sources included humans, non-human animals, plants, soils, aquatic, and agro-industrial sources. By casting such a broad exploratory net, we gained insights into the potential for human exposure to butyrate-producing bacteria across a wide range of settings. Such insights could inform microbial ecologists and health researchers who are interested in butyrate-producing bacteria, as well as provide guidance for my subsequent data chapters.

In Chapter 3, we focused on urban greenspaces. Urbanisation is rapidly expanding and can limit the opportunities of city residents to gain exposure to natural outdoor settings. Urban greenspaces provide those opportunities, and they can be designed and managed to maximise the health benefits for visitors including exposure to health-beneficial microbiota. Urban greenspace visitors breathe in the airborne bacterial communities (aerobiomes). Therefore, we studied aerobiomes in two commonly-found types of urban greenspaces: sports fields and amenity grassland parks. To learn whether manageable greenspace attributes (i.e., vegetation and soil characteristics) affect the aerobiome, we examined the effects of the surrounding woody plant species diversity and soil physicochemical parameters on the aerobiome diversity and community composition. These results could further our understanding of human exposure to key health-associated aerobiome characteristics, such as bacterial diversity and butyrate-producing bacteria.

In Chapter 4 we focused on soil butyrate and butyrate-producing bacteria because soils (a) are one of the largest repositories of bacteria on the planet, (b) have anaerobic microsites, (c) are human-accessible and modifiable, and (d) are commonly found in urban greenspaces. We studied shotgun metagenomic sequences from the same sports field and amenity grassland park sites from Chapter 3

and quantified the read abundances of known and putative butyrate-producing bacterial species. We also developed a new method to extract short-chain fatty acids from soil samples to determine the quantity of butyrate produced by soil butyrate-producing bacteria, and then associated ecological conditions (e.g., greenspace type, soil physicochemical parameters) with butyrate concentrations and read abundances of butyrate-producing bacteria. The results could identify environmental hotspots of butyrate-producing bacteria, along with their ecological drivers. This data would give key insights into where humans could gain greater exposure to butyrate-producing bacteria. Such information would be valuable to urban greenspace designers, landscape managers, and public health researchers and policy makers.

With the aerobiome and soil microbiomes evaluated in Chapters 3 and 4, for Chapter 5 we ran a human exposure trial. Our Flinders University team collaborated with a group of Māori researchers in New Zealand who share a similar interest in promoting outdoor activity for health benefits. We recruited 35 participants, and each participant underwent an 8-day trial period that included two outdoor exposures (i.e., 30-minute walks in a park) and two indoor exposures (i.e., spending time in their office). We obtained nasal swabs (total $n = 238$) before and after the exposures. By comparing the bacterial read abundances present in the nasal samples with air samples ($n = 7$) obtained at both exposure locations, we would be able to identify whether the nasal microbiomes change with 30-minute exposures. We could also quantify the taxa that transfer from the aerobiome into the nose. This exposure trial could provide crucial insights on whether aerobiomes, including bacterial groups with health-beneficial characteristics (e.g., butyrate-producing bacteria, Gammaproteobacteria), could transfer onto people.

Together, these five chapters could provide new insights into the following questions: (a) where do butyrate-producing bacteria assemble in the environment; (b) what ecological characteristics influence their abundances and activity in urban greenspaces; and (c) do outdoor exposures elicit the transfer of airborne bacteria onto people?

CHAPTER 1. The potential of outdoor environments to supply beneficial butyrate-producing bacteria to humans

---

This work appears in the following publication:

Brame JE, Liddicoat C, Abbott CA, Breed MF. The potential of outdoor environments to supply beneficial butyrate-producing bacteria to humans. 2021. *Science of the Total Environment*. doi: 10.1016/j.scitotenv.2021.146063.

Context

Butyrate-producing bacteria are associated with host organisms and free-living in the environment. In soils, butyrate is associated with the suppression of soil-borne plant pathogens (Poret-Peterson et al., 2019). In humans, butyrate is the primary fuel source for colonocytes, thereby contributing to gut barrier maintenance that separates gut lumen bacteria and contents from the body's interior (Rivière et al., 2016). Butyrate also contributes to anti-inflammatory blood profiles (Pandiyan et al., 2019), induces apoptotic responses during dysplasia (Chen, Zhao, & Vitetta, 2019), and dampens local and systemic immune responses (Sivaprakasam et al., 2017). Infant gut bacterial assemblages can be influenced by the outdoor environment (Nielsen et al., 2020). Thus, identifying environmental sources of butyrate-producing bacteria could elucidate new possibilities of supplementing gut abundances when they are reduced in disease states. However, the abundances and environmental drivers of butyrate producers have rarely been investigated.

This chapter often uses the word "commensal" to describe the bacteria being investigated. However, it is important to clarify how this term is used. Many butyrate producers examined here are not

known to play any role in humans; they could be considered “environment-associated” rather than “host-associated”. Still, these taxa may be found in faecal samples due to inadvertent human ingestion. Exposure to these bacteria could simply result in transit through the human and back into the environment, with no discernible colonisation or human-beneficial effects along the way. However, the transit dynamics of such taxa have rarely been studied. As such, the term “commensal” is used here to broadly describe taxa that have not shown pathogenicity in humans, regardless of their potential for colonisation or beneficial effects during their time in the human digestive tract.

In Chapter 1, I aimed to answer the following research question: What are the gaps in the state of knowledge about abundances and ecological associations with butyrate-producing bacteria in the outdoor environment, particularly in relation to environmental biodiversity?

Author contributions

J.B. and M.B. contributed to the conception and design of the article. J.B. performed the literature search and wrote the first draft of the manuscript. All authors edited and approved the submitted manuscript version.

The potential of outdoor environments to supply beneficial butyrate-producing bacteria to humans

Joel E. Brame[1]*, Craig Liddicoat[1,2], Catherine A. Abbott[1], Martin F. Breed[1]

[1] College of Science and Engineering, Flinders University, Bedford Park, SA 5042, Australia

[2] School of Public Health, The University of Adelaide, SA 5005, Australia

Graphical abstract

ABSTRACT

Butyrate is an important mediator of human health and disease. The mechanisms of action of butyrate are becoming increasingly well-known. Many commensal bacteria that inhabit the human gut can synthesise butyrate, which is then absorbed into the human host. Simultaneously, several immune- and inflammatory-mediated diseases are being linked to insufficient exposure to beneficial microbes from our environment, including butyrate-producing bacteria. However, the role of outdoor environmental exposure to butyrate-producing bacteria remains poorly understood. Here we review

the literature on the human exposure pathways to butyrate-producing bacteria, with a particular focus on outdoor environmental sources (e.g. associated with plants, plant-based residues, and soil), and the health implications of exposure to them. Emerging evidence suggests that environmental butyrate-producers may help supplement the human gut microbiota and represent an important component of the Biodiversity and Old Friends hypotheses. Improving our understanding of potential sources, precursors, and exposure pathways of environmental butyrate-producers that influence the gut microbiota and butyrate production offers promise to advance multiple disciplines of health and environmental science. We outline research priorities to address knowledge gaps in the outdoor environment-butyrate-health nexus and build knowledge of the potential pathways to help optimise exposure to human-beneficial butyrate-producing bacteria from the outdoor environment during childhood and adulthood.

Keywords: butyrate-producing bacteria; butyrate; microbiome; short-chain fatty acids; exposure pathways; ecosystem services;

1. Introduction

Short-chain fatty acids are end products of bacterial fermentation, the anaerobic breakdown mostly of carbohydrates yielding energy (Baxter et al., 2019). These fatty acids are produced in the human colon, as well as in non-human environments (e.g. soil; Levine et al., 2013). The human colon contains commensal bacteria that produce three major short-chain fatty acids - butyrate, propionate, and acetate. Of these short-chain fatty acids, butyrate in particular is associated with important roles in molecular signaling between gut microbiota and host, providing major fluxes of carbon from the diet to the host, and possibly regulating local and peripheral metabolism (Morrison & Preston, 2016). Butyrate also has important roles in supporting immune tolerance and maintenance of the gut barrier (Parada Venegas et al., 2019).

Researchers have documented >170 bacterial taxa that produce butyrate in animals and the outdoor environment (Li & Li, 2014; Vital, Howe, & Tiedje, 2014), many of which are found in the human gut. The majority of the human gut-associated butyrate-producers are in the phylum Firmicutes. Many are strict anaerobes found in the Clostridial groups IV and XIVa, and to a lesser extent groups I and XVI (Li & Li, 2014). The most abundant commensal taxa in the human gut are *Faecalibacterium prausnitzii* (Ruminococcaceae family) and *Eubacterium rectale (*Eubacteriaceae family*)*, with *Roseburia* spp. (Lachnospiraceae family), *Anaerostipes* spp. (Lachnospiraceae family) and other members of the Eubacteriaceae family also frequently present (Louis & Flint, 2017; Rivière et al., 2016). Not all human gut taxa that produce butyrate are commensal; for example, *Clostridioides difficile* is a known human pathogen that produces butyrate. However, there is a weight of evidence in the literature (e.g. Parada Venegas et al., 2019; Sanna et al., 2019; Valles-Colomer et al., 2019) suggesting that butyrate production by gut bacteria helps support human health, so we view these taxa as largely beneficial unless otherwise stated.

Many outdoor environments also have conditions capable of promoting the growth of butyrate-producing bacteria, such as the presence of degradable organic materials and permanently or intermittently anaerobic settings, including within bulk soil and within the gut of above- and below-ground animals. Potential exists for overlap of butyrate-producing species between outdoor environments and human gut, but there is an absence of studies that have examined ecosystems for this overlap. There is limited evidence of only a few species isolated from both outdoor and human gut environments such as *Anaerofustis stercorihominis* (Finegold et al., 2004) and *Clostridium butyricum* (Meng et al., 1999; Tran et al., 2020). In addition, perhaps the rare taxa encountered from outdoor airborne microbial (aerobiome) exposures, which can account for almost half the total bacterial abundance in low biomass aerobiomes (Selway et al., 2020) but are not given much consideration due to low individual abundances, could be a source of butyrate-producing bacteria (Liddicoat et al., 2020). After maternal transmission, the human gut microbiota derives mostly from the outdoor environment and the diet (Gilbert et al., 2018; Rothschild et al., 2018). Because the

outdoor environment influences human health outcomes through microbial exposure pathways (Chong, D'amato, & Rosário Filho, 2022), specifically exploring the outdoor environmental origins of non-pathogenic butyrate-producing bacteria, both human- and non-human-associated, may be of particular importance for human health.

In this discussion article, we review the known pathways of human exposure to butyrate-producing bacteria, with particular focus on the potential transfer of these bacteria from the outdoor environment onto the human body (Fig. 1). We include a brief overview of the synthesis and beneficial effects of butyrate in the human body. Finally, we outline future research priorities that will help strengthen the understanding of the linkages between the outdoor environment, microbial biodiversity, butyrate-producing bacteria, butyrate, and human health.

2. Synthesis of butyrate in the human body

Butyrate is a 4-carbon carboxylic acid. Its production in bacteria is derived from four main fermentation metabolic pathways: acetyl-CoA, glutarate, lysine, and 4-aminobutyrate (Vital, Howe, & Tiedje, 2014). Each pathway predominately originates from key feedstocks (most commonly dietary fibres) or bacterial cross-feeding mechanisms (Fu et al., 2019). The specific pathway utilised depends in part on the dietary substrate ingested, such as resistant starch from potatoes (Baxter et al., 2019) as well as inulin-type fructans (including oligofructose) and arabinoxylan oligosaccharides (Rivière et al., 2016). In some cases, colonic *Bifidobacteria* and *Ruminococcus* strains take in the dietary fibres and produce intermediates (including acetate and lactate). These intermediates are taken up by butyrate-producing bacteria and are metabolised into butyrate using either butyryl-CoA:acetate CoA transferase or butyrate kinase as a final enzymatic step (Louis & Flint, 2017). Resulting butyrate concentrations within the colon and rectum range from 30-240 mM (Li & Li, 2014).

Once released over the mucosa layer of the colon, butyrate is then absorbed into colonic mucosal cells through several dedicated transporters (e.g. MCT1 and SMCT1; Parada Venegas et al., 2019). As butyrate is metabolised by colonic epithelial cells for energy, substantially less butyrate passes through the epithelium and into the portal circulation. Butyrate is further metabolised in the liver, leaving very low concentrations (in the µM range) in peripheral circulation (Stilling et al., 2016).

While most butyrate is absorbed and metabolised by colonic cells, some ends up excreted in the faeces at a concentration range of 5-25 mM/kg (Baxter et al., 2019; Hald et al., 2016). Some studies have found that faecal butyrate levels positively associate with the richness of colonic butyrate-producing bacterial species (Baxter et al., 2019; Nilsen et al., 2020; Yamamura et al., 2019). However, faecal butyrate levels may vary due to the influences of bowel transit time on the colonic butyrate absorption, which should be considered when using faecal butyrate measurements as a proxy for butyrate production by gut bacteria (Brahe, Astrup, & Larsen, 2013).

3. Effects of butyrate in the human body

Butyrate has been associated with multiple local effects in the colon. It is the preferred energy source for colonic epithelial cells and helps maintain the gut barrier through tight junction stability and regulation of epithelial proliferation (Rivière et al., 2016). Butyrate regulates mucus secretion (Stilling et al., 2016) and supports IgA synthesis in the colon (Rivière et al., 2016).

Butyrate also has anti-inflammatory and immunomodulating effects that occur through a variety of mechanisms. One of the best-studied mechanisms is the butyrate-triggered accumulation of regulatory T cells (Treg) in the colonic mucosa. After butyrate passes through the epithelial cells, it can induce differentiation of peripheral naïve T cells into Tregs by acting as an inhibitor of histone deacetylase 1 and 3 (Sivaprakasam et al., 2017) and a ligand to the GPR109a and FFA2 cellular receptors on colonic immune cells (Kaisar et al., 2017; Takahashi et al., 2020). Tregs play a critical role in immunoregulation by dampening or suppressing excessive immune and inflammatory

responses, accomplished in part through production of cytokines IL-10 (which act on $CD4^+$ cells) and TGF-beta1, along with other immunomodulatory molecules (Pandiyan et al., 2019; Wasko, Nichols, & Clark, 2020; Zhao et al., 2018). It can also enhance polarisation of macrophages toward an anti-inflammatory profile (Parada Venegas et al., 2019) and the release of anti-inflammatory prostaglandin E2 (Ohira et al., 2016). Through these anti-inflammatory and immunomodulating mechanisms, butyrate can lead to commensal tolerance (Pandiyan et al., 2019) and protection from immune and allergic reactions.

Additional extraintestinal benefits include maintenance of the blood-brain barrier (Braniste et al., 2014), protection from metabolic diseases (Sanna et al., 2019), and reduction of depression and anxiety (Valles-Colomer et al., 2019). These benefits stem not only from the anti-inflammatory effects, but also through the pathways included in the microbiome-gut-brain axis, as well as the effects on gut hormones such as ghrelin (Torres-Fuentes et al., 2019) and glucagon-like peptide-1 (Liu et al., 2018). The microbiome-gut-brain axis is a prominent linkage between butyrate-producing bacteria and mental health effects through the nervous, immune, and endocrine systems (Foster, Rinaman, & Cryan, 2017). Psychobiotics including species of *Bifidobacteria* (Sarkar et al., 2016) play a role as intermediaries in the synthesis of butyrate, thus suggesting butyrate as an important contributor to mental health.

Several medical conditions (e.g. inflammatory bowel disease, multiple sclerosis) have been associated with a reduction in abundance of butyrate-producing bacterial taxa. For example, a reduction in *Faecalibacterium prausnitzii*, *Roseburia hominis* and *Roseburia intestinalis* has been found in patients with inflammatory bowel disease (Parada Venegas et al., 2019). Miyake et al. (2015) demonstrated that in multiple sclerosis there is an associated decrease in taxa from Clostridia clusters IV and XIVa. While studies show associations between dysbiosis and medical conditions, additional research is necessary to identify whether the condition has created the changes in the microbiota, or whether the changes in microbiota contribute to the condition.

Currently, the majority of studies indicate that butyrate is beneficial for the human body. However, researchers have reported some conflicting data on butyrate's connection with obesity. For example, a higher ratio of Firmicutes (the phylum with the majority of butyrate producers) to Bacteroidetes has been observed in obese subjects (Brahe, Astrup, & Larsen, 2013), although follow-up studies have found no such association (Liu et al., 2018). Indeed, further research is necessary to identify risks associated with butyrate and butyrate-producing bacteria in the human body.

4. Sanitation and hygiene

Not all bacteria that produce butyrate have beneficial effects in the human body. For example, the butyrate-producer *Acidaminococcus fermentans* was found in greater abundance in the gut of people with ankylosing spondylitis (Zhou et al., 2020), and *Clostridioides difficile* is commonly found in healthcare-related infections (Guh & Kutty, 2018). Advances in sanitation have helped reduce exposure to pathogenic bacteria (Engel & Susilo, 2014). However, because faecal matter exposure is part of human history as well as a source of gut-associated bacteria, modern sanitation measures also deplete human exposure to commensal butyrate-producing bacteria (Schmidt et al., 2011). Thus, a healthy level of exposure to environmental butyrate-producing bacteria, while avoiding pathogenic bacteria, is an important challenge to be addressed. Perhaps healthy exposure may be consistent with the U-shaped dose-response curve proposed by Liddicoat et al. (2016), in which neither an excess nor deficient level of exposure to butyrate-producing bacteria is ideal, but an intermediate level of exposure to diverse bacteria may be most beneficial for human health.

5. Pathways of exposure to butyrate-producing bacteria

5.1 Birth and neonatal period

The dominant bacterial genera in a vaginally-delivered newborn are facultative anaerobes, including *Lactobacillus* sp. and *Streptococcus* sp. (Martin & Sela, 2013). Obligate anaerobic butyrate-producing bacteria are typically not present at this stage, likely because the gut lumen has not yet

been depleted of oxygen (Appert et al., 2020). Hospital delivery by Caesarean section shows greater initial colonisation by microbiota found on maternal skin and the ambient hospital environment, such as *Clostridia* species, along with reduced *Bifidobacteria* colonisation when compared to vaginal delivery (Dominguez-Bello et al., 2010). While the difference in *Bifidobacteria* between vaginal and Caesarian births seems to resolve with breast milk intake within two months (Liu et al., 2019), the delay could have long-standing health implications by increasing the risk of many immune-related diseases such as asthma and inflammatory bowel disease (Sevelsted et al., 2015), as *Bifidobacteria* are involved in cross-feeding mechanisms of butyrate production. It bears to mention that the usage of intrapartum maternal antibiotics in either birth mode may be a confounding factor in neonatal microbiome assessments because it can lead to decreased diversity of infant gut microbiota (Dierikx et al., 2020).

5.2 Breast feeding

Breast milk confers a wide variety of beneficial nutrients and immune-related molecules, including butyrate at concentrations of 0.75 mM (Paparo et al., 2020). Research has shown that breast milk contains *Bifidobacterium*, *Bacteroides*, and *Clostridium*, which can colonise the gastrointestinal tract of infants (Martin & Sela, 2013). Obligate anaerobic butyrate-producing bacteria from the *Coprococcus*, *Roseburia*, *Faecalibacterium*, and *Subdoligranulum* genera were found shared between maternal faeces and breast milk, suggesting a possible vertical transmission pathway from mother to child (Jost et al., 2014). In addition, pathogenic bacteria can be inhibited by breast milk components such as antimicrobial molecules (e.g. secretory IgA), acids attached to milk oligosaccharides that lower the gut pH, and lactoferrin (a protein in the whey group; Andreas, Kampmann, & Le-Doare, 2015) which can disrupt the membrane of gram-negative bacteria (Martin & Sela, 2013).

Human breast milk also supplies oligosaccharides that have a wide variation of lengths and saccharide units and are indigestible by human infants, leaving these nutrients intact for the gut

bacteria to ferment (Martin & Sela, 2013). Human milk oligosaccharides, the third-largest component in breast milk (Andreas, Kampmann, & Le-Doare, 2015), can feed butyrate-producing Clostridiales bacteria (e.g. Eubacterium and Roseburia spp.; Pichler et al., 2020). They can also support cross-feeding pathways that lead to butyrate production by providing for preferential growth of certain butyrate-related species, including *Bifidobacterium* sp. and *Bacteroides* sp. (Baxter et al., 2019). *Bifidobacterium* in particular ferment the oligosaccharides and produce byproducts including acetate and lactate, which then can be taken up by the butyrate-producing bacteria (e.g. Eubaceterium hallii; Bunesova, Lacroix, & Schwab, 2018). *Bifidobacterium* can also utilise human-derived mucin as an energy source, which contributes to the butyrate-production potential of the gut microbial community. In contrast, formula-fed infants have been shown to have greater microbial diversity (Liu et al., 2019) but less butyrate production (Brink et al., 2020). Thus, breast milk can have a strong influence on the production of butyrate during infancy, and can help restore a Caesarean-section microbial profile to more closely resemble that of vaginal birth (Liu et al., 2019).

5.3 Ingestion of plant-based solid foods and fibres

The introduction of solid foods during and after weaning creates a major shift in gut microbiota toward a more mature, adult-like profile (Tanaka & Nakayama, 2017). For example, a recent study in rabbits showed solid food ingestion corresponded with increases in both gut alpha-diversity and the abundance of butyrate-associated Firmicutes (Beaumont et al., 2020). In humans, faecal butyrate levels increased 4-fold between six months and one year of age (Nilsen et al., 2020). While some animal products are partly composed of butyrate, such as butter (2.7% butyrate) and yogurt (0.1% butyrate) (Roduit et al., 2019), plant-based foods have the potential to feed gut bacteria that produce butyrate. Plant fibres and fermentable carbohydrates are the most common feedstocks for butyrate-producing bacteria in the gastrointestinal tract, and their ability to increase butyrate production and abundance of butyrate-producing species is well-documented in both in-vivo and in-vitro experiments (Fu et al., 2019). Plant-based nondigestible carbohydrates that have been shown to have butyrogenic potential, either directly via butyrate-producing bacteria or through primary and

secondary degrader cross-feeding pathways, include cereals (e.g. wheat bran, germinated barley), psyllium husk, resistant starch, guar gum, inulin, seaweeds (e.g. alginate) and pectins from fruits and vegetables (Fu et al., 2019). The different fermentable carbohydrates ingested can have their own distinct degradation pathways and outcomes on production of short-chain fatty acids (Baxter et al., 2019). Increasing diversity of plants in the diet has also been linked to a more diverse gut microbiome, and increases in several putative butyrate producers (McDonald et al., 2018). The impact of diverse plants in the diet may impact the gut microbiome through a diversity of dietary fibres and resistant starches, and possibly ingestion of diverse plant endophytic bacteria (see below).

Ingestion of plant-based foods also directly confers a variety of bacteria, possibly including butyrate-producing taxa. Bacteria can be found on soil residues on the food, inside plant organs, roots, leaves, and fruits, and on the exterior and interior of seeds (Flandroy et al., 2018). Lang et al. (2014) analysed the bacterial content of a variety of meals in three major dietary patterns (average American, USDA recommended, and vegan) and found members of Clostridiaceae, Lachnospiraceae and Ruminococcaceae families in all three dietary patterns' foods. The authors also found significant positive correlations between sodium and energy content within food and the presence of bacteria involved in butyrate production, suggesting that certain foods may confer greater amounts of butyrate-producing bacteria to the human gut. Whether food-borne butyrate-producing bacteria are transient or colonise the gut is under investigation.

5.4 The outdoor environment

The outdoor environment can influence the microbiomes of humans, with subsequent health effects (Chong, D'amato, & Rosário Filho, 2022). Indoor environmental bacteria originate from a variety of indoor sources including children and dogs (Hickman et al., 2022) and are influenced by outdoor bioaerosol diversity and building ventilation strategies (Meadow et al., 2014). Outdoor bioaerosols (e.g. soil, organic debris) are commonly carried into indoor environments, providing household members with continual exposure to outdoor-related microbiota while in the home. In addition,

exposure to bacteria while outdoors can alter the skin and nasal microbiome of the human, including through changes in rare taxa profiles (Selway et al., 2020), and potentially the gut microbiome (Nurminen et al., 2018). The general transfer of bacteria from the outdoor environment to humans represents an area of increasing research interest; however, the specific transfer of commensal butyrate-producing bacteria remains under-investigated. In this section we explore the potential transfer pathways of butyrate-producing bacteria that may be considered relevant for promoting human health.

Many of the anaerobic Gram-positive butyrate producers, including commensal *Clostridium* and *Bacillus* bacteria, create endospores that allow survival in a dormant state in aerobic conditions (Al-Hinai, Jones, & Papoutsakis, 2015). Human exposure to endospores has been constant during our evolutionary history and represents a normal aspect of exposure to natural biodiversity. Soil has approximately 10^6 bacterial endospores per gram, and analysis of human faeces shows on average 10^4 endospores per gram of faeces (Hong et al., 2009). Indeed, endospore-forming bacteria dominate the adult human gut microbiota, comprising 50–60% of bacterial genera, and they show greater turnover in abundance and species over time (compared to non-endospore-formers; Browne et al., 2016). Appert et al. (2020) showed that endospore-forming bacteria significantly increase in abundance from age 3 months to 1 year, possibly due to the transformation of the gut into an anaerobic compartment. In addition Appert et al. (2020) demonstrated that Clostridiaceae remain in high abundance during the first year of life, while members of the Peptostreptococcaceae, Erysipelococcaceae, and Lachnospiraceae families (each of which have spore-forming, butyrate-producing taxa) increased significantly in abundance. The ability of the endospores to persist in aerobic external environments allows them to spread widely (Kearney et al., 2018). Therefore, endospore-forming bacteria may potentially play a key role in the transfer of butyrate-producing capacity into the human gut, where the gut conditions can trigger spore germination. Future research on endospores as a potential transfer mechanism may help narrow the knowledge gap on the origin of human gut-associated butyrate-producing bacteria from the outdoor environment.

In a randomised controlled mouse study, Liddicoat et al. (2020) showed a soil-associated butyrate-producer (putatively identified as the anaerobic spore-former *Kineothrix alysoides*), which was found particularly in biodiverse soils, could be transferred via trace-levels of airborne soil dust to inhabit and influence the gastrointestinal tract of mice. *K. alysoides* was found to supplement the gut microbiota of mice to the greatest extent when sourced from a soil with high microbial diversity (extracted from a high plant macro-diversity setting), compared to low biodiversity or no soil (control) treatments. Increasing relative abundance of *K. alysoides* in caecal samples was subsequently linked to reduced anxiety-like behaviour in the most anxious mice. *K. alysoides*, which has also been isolated from the human gut (Chen et al., 2020; Noh & Lee, 2020), is in the family Lachnospiraceae, which is one of the most abundant in the human gut microbiome and contains many known plant degraders and most of the butyrate-producers in the gut (Haas & Blanchard, 2017). These results suggest that *K. alysoides*-like organisms may transfer through outdoor environmental pathways, and particularly through exposure to biodiverse plant-soil systems.

The microbiota found in air inside the home are likely sourced primarily from the outdoor air and human occupancy (Adams et al., 2015). Hickman et al. (2022) associated indoor microbial communities with the presence of children and dogs, as well as nearby outdoor environmental characteristics such as land use and. The taxa of airborne bacteria from the outdoor environment depend in part on the surrounding land cover, as well as other spatiotemporal dynamics, such as regional air mass movement and human activity including agricultural harvesting (Mhuireach et al., 2019) and height above the urban green spaces (Robinson, Dumancela, et al., 2020). Indeed, studies have examined the environmental influences on dust in homes. The PARSIFAL study (Alfven et al., 2007), for example, examined children's mattresses, and the GABRIELA study (Ege et al., 2011) used electrostatic dust collectors placed in the homes to study dust microbiota sources and associations with environmental characteristics. Both studies found greater microbial diversity in the

homes of children who lived on farms rather than in cities. To the best of our knowledge, no aerobiome studies have had a particular focus on butyrate-producing bacteria.

While the aerobiome enters homes through open doors and windows, outdoor bacteria may also be transferred into the home through clothing, shoes and pets. Parajuli et al (2018) found Firmicutes belonging to the *Roseburia* and *Faecalibacterium* genera, both known butyrate-producers, in doormat debris. Exposure to household mammalian pets has been shown to be associated with an increase in Firmicutes, particularly the genus *Ruminococcus*, in the faecal samples of infants (Tun et al., 2017). Within this genus, *R. torques* is a known butyrate-producer (La Reau & Suen, 2018), and *R. bromii* participates in a cross-feeding pathway between bacterial species that leads to butyrate production (Baxter et al., 2019).

Beyond potential exposure via the aerobiome, cutaneous exposure to butyrate-producing bacteria may serve as another pathway of transfer of outdoor environmental butyrate-producing bacteria to people. Both brushing up against leaves and interacting with dirt (to replicate children's outdoor activities) were shown to cause an increase in rare taxa (which may contain butyrate-producers) associated with the outdoor environment on the skin after 15 minutes to 1 hour of exposure (Selway et al., 2020). Skin application of soil/dung mixtures, even if quickly washed off, can have an impact on gut microbiota. A human study showed that daily topical applications of a biodiverse mixture (including soils and dung) led to greater alpha diversity of faecal gut microbiota, as well as increases in *Bacteroides* (Nurminen et al., 2018). In this study, Firmicutes showed significant increases in relative abundance, but to a lower degree than other phyla. However, the changes were transient and disappeared after three weeks. Both Firmicutes and *Bacteroides* have butyrate-producing member species, thus it may be possible to influence gastrointestinal butyrate production through cutaneous exposure to soils.

Living in proximity to agricultural land, particularly with livestock and poultry, has been shown to be inversely associated with levels of atopic disease when compared to living in urban land cover (Von Hertzen & Haahtela, 2004). For example, Steiman et al. (2018) showed that children growing up on dairy farms have less atopic dermatitis and greater skin microbial diversity than children raised in non-farm locations. The authors found these effects starting at two months of age. Even among agricultural settings, the level of biodiversity exposure partly depends on the farming practices utilised, such as conventional vs. organic farming (Flandroy et al., 2018), the use of horses vs. modern equipment for field work, and the physical proximity of farm animals to the farm housing structures (Ober et al., 2017; Stein et al., 2016). A study of piglets showed that exposure to agricultural topsoil during lactation phase resulted in significant changes in the microbiota of their gut at weaning, including enriched Firmicutes and *Prevotella* (Vo et al., 2017). Consistent knowledge gaps exist on whether farm environments also have a greater percentage of butyrate-producing bacteria, which warrants further investigation due to its potential human and animal health linkages.

Together, these studies on outdoor environmental bacteria indicate that certain soil and airborne microbiota have the potential to transfer butyrate-producing bacteria into homes via, for example, air flow, pet activity, shoes and clothing. They also suggest that butyrate-producing bacteria can then enter the gut, modify the gut microbiota, and provide potential for health benefits. Of course, such benefits would only accrue in concert with a healthy lifestyle (e.g. diet, which provides important feedstocks for the microbes and butyrate-producing pathways). (Table 1)

6. Can biodiverse ecosystems promote human health via increased exposure to butyrate-producing bacteria?

The combination of microbial diversity, complex organic matter residues, and periodic conditions of reduced redox potential which can be found within many non-arid biodiverse plant-soil systems may represent important factors contributing to the presence of soil-associated butyrate-producing

bacteria and, therefore, their potential transfer onto humans. A diverse suite of micro-organisms are involved in the decomposition of organic matter, with many pathways and taxa contributing to butyrate production (Vital, Karch, & Pieper, 2017). Aboveground vegetation diversity correlates with increased soil bacterial diversity (Coleman, 2004; Delgado-Baquerizo et al., 2018) and increased soil organic matter (Chen et al., 2018) – therefore biodiverse ecosystems are more likely to provide the feedstocks and microbiota required to support outdoor environmental butyrate production. Butyrate derives from the fermentative breakdown of organic matter under low redox potential conditions, and we can expect such conditions to occur in periodically wet soils, and within the gut of aboveground or belowground animals (e.g. soil invertebrates).

Because biodiversity (microbial and other) is potentially correlated with protection from immune-based diseases (e.g. asthma and allergy; Haahtela, 2019), research examining the soil characteristics of human-surrounding ecosystems (including microbial biodiversity and substrate content), may give insights about the potential for transfer of butyrate-producing bacteria within the ecosystem. In the aforementioned study by Liddicoat et al. (2020), the researchers found the putative butyrate-producer *K. alysoides* in soil with high biodiversity, but not in soil with low biodiversity. A daycare study by Roslund et al. (2020) found that covering yard gravel with biodiverse forest soil and sod was associated with an increase in the gut abundance of Ruminococcaceae (a family with known butyrate-producers). These results suggest that soils higher in biodiversity may associate with greater abundance of gut butyrate-producers. Insufficient exposure to these biodiverse microbiota as a result of urbanisation may be connected with the increasing rise in many immune- and inflammatory-mediated diseases (Rook, Lowry, & Raison, 2013). This is amplified by the finding that the loss of key gut microbiota (including butyrate-producers) from lifestyle factors such as antibiotics and low-fibre diet can compound over generations (Sonnenburg et al., 2016).

Such linkages substantiate the Biodiversity hypothesis and Old Friends hypothesis, which offer possible connections between exposure to outdoor environmental microbiota and positive benefits on

human health. For example, the Old Friends hypothesis postulates an evolutionary role of microbial 'Old Friends' that have co-evolved with humans. Rook includes the butyrate-producers *F. prausnitzii* and *Clostridia* species among the key 'Old Friends' (Rook, 2012). However, the exact mechanisms of the links between outdoor environment, biodiversity, butyrate-producing bacterial species, and human health are still under investigation (Lai et al., 2019; Liddicoat et al., 2020; Robinson & Breed, 2019, 2020; Stanhope, Breed, & Weinstein, 2020). Because the infant gut microbial communities can be influenced by the outdoor environment (Nielsen et al., 2020), identifying how biodiversity in the environment can impact the abundance of butyrate-producing bacteria in the gut, especially during the first year of life, is worthy of further examination. In alignment with the Biodiversity hypothesis, public health policies might someday help restore greater biodiversity in urban areas so people can be regularly exposed to health-promoting bacteria (e.g. accessible urban green spaces, commuter pathways; (Breed et al., 2020; Robinson & Breed, 2019, 2020; Robinson, Mills, & Breed, 2018; Watkins et al., 2020) while keeping health risks to a minimum.

7. Methods to assess the butyrate production capacity of the microbial ecosystem

DNA sequencing methods can be used to identify both taxonomic and functional gene presence in samples, which could be used to determine the presence of butyrate-producing bacteria and their transfer to humans. For example, 16S rRNA gene sequencing is frequently used to assess the presence and potentially also the abundance of butyrate-producing bacteria in a sample, such as soil (Poret-Peterson et al., 2019) and human faeces (Nurminen et al., 2018). Software such as Piphillin and PICRUSt2 can be used to predict functions including butyrate production from 16S marker sequences, although caution is required with these approaches in less-studied environments such as soils (Sun, Jones, & Fodor, 2020). Alternately, whole genome shotgun sequencing can identify both taxonomic and functional genes, including coverage of genes found in metabolic pathways. However, whole genome shotgun sequencing is also more expensive than 16S rRNA gene sequencing.

Metatranscriptomics methods, such as RNA-seq, have been used to demonstrate the expression of genes in the butyrate production pathways. For example, RNA-seq data from David et al. (2014) indicated that ingested foodborne bacteria were alive and metabolically active in the gut. In addition, using gas chromatography to measure short-chain fatty acids levels in stool, they were able to correlate presence of bacteria that contain genes for butyate production with butyrate levels. RNA-seq can reveal the functional capacity of the microbiome, even when taxonomic data is unavailable. Another approach to assessment of butyrate production capacity is to use a combination of 16S rRNA gene sequencing, metabolomics and metaproteomics. Maier et al. (2017) used this approach to demonstrate that a diet rich in resistant starch led to the increase of the amounts of bacterial species associated with butyrate production, a slight increase of butyrate levels, and the identification of several proteins involved in butyrate metabolism in stool samples. Such multi-omics approaches can provide more thorough information for both taxonomic and functional assessments of butyrate-producing bacteria that transfer from environmental sources to the human body.

8. Conclusion

Humans are exposed to butyrate-producing bacteria, as well as bacteria that cross-feed and interact with the butyrate-producers, through a variety of pathways. Many of these exposure pathways occur early in life (e.g. birth mode, breast milk, solid food introduction, household pet exposure), and most of these pathways have considerable evidence to support beneficial influences of exposure on the gut microbiome and on human health. However, the pathways involving exposure to outdoor environmental butyrate-producing bacteria remain understudied.

Clear knowledge gaps have arisen from this review, including: (a) identifying outdoor environmental sources of butyrate-producing bacteria to which humans might be exposed; (b) determining whether non-pathogenic, outdoor-associated butyrate-producing bacteria may transfer onto humans and produce butyrate in the gut; (c) understanding links between exposure and doses of incident bacteria

with different environments and activities; (d) understanding mechanisms of action, such as how respiratory and cutaneous exposure may lead to changes in gut microbial communities; and (e) determining the dose-response relationship between external dose of butyrate-producing bacteria and changes in butyrate production.

Future research should include surveys of a variety of soil and air sources of butyrate-producing bacteria, with metadata on the specific ecological characteristics of soils and sediments that house butyrate-producing bacteria. Further exposure assessments with adults and children can also help build insights to immune system training and longitudinal health research in the context of the Biodiversity and Old Friends hypotheses. Exposure and colonisation studies can add understanding to the role of endospores and non-human-associated butyrate-producers in human gut communities. Such data would help inform the creation, design and/or restoration of outdoor environments that safely increase exposure to butyrate-producing bacteria, specifically in urban environments depleted of microbial biodiversity. Establishing the interlinkages between exposure to a biodiverse environment, and the associated health-promoting butyrate-producing bacteria – with future development and research – should help to improve global public health initiatives and have a lasting positive impact on public health and biodiversity conservation.

CHAPTER 2. Towards the biogeography of butyrate-producing bacteria

---

This work appears in the following publication:

Brame JE, Liddicoat C, Abbott CA, Edwards RE, Gauthier NE, Robinson JA, Breed MF. Towards the biogeography of butyrate-producing bacteria. 2023. *bioRxiv* (under review in *Ecology and Evolution*). doi: 10.1101/2022.10.07.510278.

Context

Butyrate-producing bacteria are found in many outdoor ecosystems and host organisms, including humans, and are vital to ecosystem functionality and human health. These bacteria ferment organic matter, producing the short-chain fatty acid butyrate. However, the macroecological influences on their biogeographical distribution remain poorly resolved. Here we aimed to characterise their global distribution together with key explanatory climatic, geographic, and physicochemical variables. We developed new normalised butyrate production capacity (BPC) indices derived from global metagenomic (*n*=13,078) and Australia-wide soil 16S rRNA (*n*=1,331) data, using Geographic Information System (GIS) and modelling techniques to detail their ecological and biogeographical associations. The resulting new macroecological insights further our understanding of the ecological patterns of outdoor butyrate-producing bacteria, with implications for emerging microbially-focused ecological and human health policies

In Chapter 2, I aimed to answer the following research question:

Which global environmental characteristics associate with the abundances of human-accessible butyrate-producing bacteria?

Author contributions

J.B., C.L., R.E., and M.B. conceived and designed the study. J.B. performed data analysis, with input from C.A., R.E., N.G., C.L., and M.B. J.B. and J.R. designed and developed the figures. J.B. wrote the manuscript, with input from all authors.

Towards the biogeography of butyrate-producing bacteria

*Joel E. Brame[a], Craig Liddicoat[a,b], Catherine A. Abbott[a], Robert A. Edwards[a], Jake M. Robinson[a], Nicolas E. Gauthier[c], Martin F. Breed[a]

[a]College of Science and Engineering, Flinders University, Bedford Park, SA 5042, Australia

[b]School of Public Health, The University of Adelaide, SA 5005, Australia

[c]Florida Museum of Natural History, University of Florida, Gainesville, FL 32611, USA

ABSTRACT

Butyrate-producing bacteria are found in many outdoor ecosystems and host organisms, including humans, and are vital to ecosystem functionality and human health. These bacteria ferment organic matter, producing the short-chain fatty acid butyrate. However, the macroecological influences on their biogeographical distribution remain poorly resolved. Here we aimed to characterise their global distribution together with key explanatory climatic, geographic, and physicochemical variables. We developed new normalised butyrate production capacity (BPC) indices derived from global metagenomic (*n*=13,078) and Australia-wide soil 16S rRNA (*n*=1,331) data, using Geographic Information System (GIS) and modelling techniques to detail their ecological and biogeographical associations. The highest BPC scores were found in anoxic and fermentative environments, including the human and non-human animal gut, and in some plant-soil systems. Within plant-soil systems, roots and rhizospheres had the highest BPC scores. Among soil samples, geographic and climatic variables had the strongest overall influence on BPC scores, with human population density also making a considerable contribution. Higher BPC scores were in soils from seasonally productive sandy rangelands, temperate rural residential areas, and sites with moderate-to-high soil iron concentrations. Abundances of butyrate-producing bacteria in outdoor soils followed complex ecological patterns influenced by geography, climate, soil chemistry, and hydrological fluctuations.

These new macroecological insights further our understanding of the ecological patterns of outdoor butyrate-producing bacteria, with implications for emerging microbially-focused ecological and human health policies.

Keywords: butyrate, ecosystem services, gut microbiome, metagenomics, microbial ecology, soil microbiota

INTRODUCTION

Butyrate-producing bacteria are both associated with host organisms and are free-living in outdoor ecosystems. They have critical roles in breaking down organic products including fibres (Baxter et al., 2019) and cellulose (Goldfarb et al., 2011). Given suitable organic substrates and anaerobic conditions, these bacteria can produce butyrate, a short-chain fatty acid, as a metabolic by-product of fermentation. In soils, butyrate is associated with the suppression of soil-borne plant pathogens (Poret-Peterson et al., 2019). In humans and non-human animals, gut-associated butyrate provides energy to colonic epithelial cells, maintains gut barrier stability, has anti-inflammatory and immunomodulatory effects, and strengthens the integrity of the blood-brain barrier (Bedford & Gong, 2018; Brame et al., 2021; Knox et al., 2022; Rivière et al., 2016). As such, butyrate production by bacteria is now understood to be a vital part of host-microbe symbioses.

Efforts to quantify the taxonomic and functional abundances of butyrate-producing bacteria through metagenomic analyses have realised multiple challenges. For example, in the SEED functional genome annotation system (Overbeek et al., 2005; https://pubseed.theseed.org/), the subsystem "Acetyl-CoA fermentation to butyrate" (https://pubseed.theseed.org/) includes 20 genes involved in the butyrate production pathway, and individual bacterial genome isolates show varying copies and synteny of the genes for the enzymes (Vital, Howe, & Tiedje, 2014). Quantifying complete functional pathway abundances within samples can therefore be computationally time-consuming and expensive, limiting large-scale sample comparisons. Alternately, molecular methods that directly

measure butyrate can provide insights into the functional potential of bacterial communities in samples. However, to our knowledge, few studies have focused on butyrate in outdoor ecosystems (Brame et al., 2021; Liu, Qiu, & Lu, 2011), and no studies to date have matched soil bacterial short-chain fatty acids and metagenomics on a global biogeographical scale. Therefore, we aimed to develop a novel method to estimate a sample's potential for butyrate production in a computationally agile way. This workflow could then be scaled up to examine large numbers of samples from a broad spectrum of sources, including outdoor ecosystems.

Soils represent a key reservoir for microbial diversity in ecosystems (Fierer & Jackson, 2006) and provide a focus for us to examine conditions that may support butyrate producers. Soil butyrate producers may be involved in soil-plant-organic matter cycling processes and be ingested and expelled from the digestive tract of soil fauna (e.g., earthworms; Thakuria et al., 2010). At the same time, non-human animal gut-associated bacteria are regularly dispersed into outdoor ecosystems, where they can settle into soils (Blum, Zechmeister-Boltenstern, & Keiblinger, 2019). Moreover, bacteria in the superficial layer of soils can be dispersed into the air (Polymenakou, 2012; Robinson, Cando-Dumancela, et al., 2020) and transmitted to hosts (Liddicoat et al., 2020). However, almost no existing studies have quantified outdoor abundances of butyrate-producing bacteria. As such, examining the distribution of butyrate-producing bacteria among the ecological components of their transmission cycle could provide more detailed insight into epidemiological and ecological knowledge of these bacteria, including conditions and mechanisms that support persistence and transmission. Macroecological knowledge of butyrate-producing bacteria could allow urban landscape architects and managers to understand how certain environmental conditions contribute to the human and non-human exposome. Moreover, it could also allow researchers to develop ways of selecting for potentially health-promoting bacterial assemblages in the environment (Brame et al., 2021).

Here, we examined a broad variety of global sample sources and ecological conditions (e.g., climate, geography, and soil physicochemical characteristics) to identify associations with butyrate production abundances. We developed two new normalised indices to estimate and compare the butyrate production capacity (BPC) of representative samples from both metagenomic (BPC_{meta}) and 16S rRNA amplicon (BPC_{16S}) data. We combined global metagenomic datasets with continent-wide Australian 16S rRNA amplicon data to provide insight into the global macroecology of butyrate-producing bacteria. Specifically, we aimed to: **(a)** determine the butyrate production capacities across a range of hosts and ecological conditions using our novel BPC_{meta} scores; **(b)** provide a more detailed comparison of BPC_{meta} scores across subcategories within each broader source group; **(c)** conduct an in-depth analysis of Australian soil samples using BPC_{16S} scores to describe ecological associations with soil butyrate production capacity; and **(d)** examine the potential influence of human population density on abundances of soil butyrate producers. By developing novel targeted formulae and collating multiple types of data and analyses, we were able to synthesise a novel and in-depth view into the global macroecology of butyrate-producing bacteria and the interlinkages between their hosts and ecosystems.

MATERIALS AND METHODS

Method overview

To compare the bacterial butyrate production capacity of samples across a broad spectrum of hosts and ecosystems, we chose to analyse samples based on two types of bacterial sequencing data: shotgun metagenomic sequences and 16S rRNA amplicons. Each data type required the interrogation of separate databases, using separate statistical tools for analysis and visualisation. We developed novel formulae for each data type to estimate the butyrate production capacity (BPC) for each sample (detailed below). Each formula included a normalisation step to allow BPC scores between samples (within a data type) to be directly compared (**Table 1**).

Global shotgun metagenomic sample analysis

Gene selection and metagenome database interrogation

To characterise the global distribution of butyrate-producing bacteria, we analysed shotgun metagenomic datasets. To begin, the butanoate (butyrate) synthesis pathways were reviewed using Seed viewer subsystems (https://pubseed.theseed.org/) and the KEGG pathway (Kanehisa et al., 2016; https://www.genome.jp/kegg/pathway.html)
to determine the genes that code for enzymes that are part of the butyrate production pathway. Based on these pathways, the following two genes were chosen for further analysis: *buk* (butyrate kinase) and *atoA* (acetate-CoA:acetoacetyl-CoA transferase subunit beta). *ACADS/bcd* (butyryl-CoA dehydrogenase) and *ptb* (phosphate butyryltransferase) were analysed but excluded (all gene decisions explained in **Supplementary Table 1** in Supporting Information). The included genes *atoA* and *buk* participate in each of the two terminal pathways. The Integrated Microbial Genomes & Microbiomes (IMG/M) database (Chen et al., 2021) was then searched for each butyrate-related gene to obtain their mean counts among genomes with at least one copy of either gene (**Supplementary Table 2** in Supporting Information). At the time of download, IMG/M utilised Annotation Pipeline v.5.0 protocols (release date October 2018) for functional annotations (Chen et al., 2019).

We next searched global metagenomic databases for *atoA* and *buk* to find metagenomes that suggest the potential presence of butyrate-producing bacteria. Initial gene and translated gene searches of metagenomics data via the 'Searching SRA' web facility (https://www.searchsra.org) using bowtie2 and diamond yielded low numbers of samples and/or high E-values. The largest datasets came from searching IMG/M using EC numbers for each butyrate-production enzyme (butyrate kinase = EC 2.7.2.7, acetate-CoA:acetoacetyl-CoA transferase subunit beta = EC 2.8.3.8) as well as three enzymes with single-copy genes used for subsequent normalisation (phenylalanine—rRNA ligase = EC 6.1.1.20; guanylate kinase = EC 2.7.4.8; alanine—tRNA ligase = EC 6.1.1.7). Sample datasets with the genes *atoA* (*n*=21,147) and *buk* (*n*=16,263) were downloaded as our starting point for metagenomics analysis. Among the downloaded datasets, 14,407 datasets had both genes, and many datasets had one gene but not the other (*atoA* only *n*=6,330 and *buk* only *n*=1,856). We then collated

these datasets to create an initial set of 22,593 unique metagenomic samples (**Supplementary Table 3** in Supporting Information).

Counts for each butyrate-production gene were normalised by dividing by counts of the single-copy gene *pheS*, which codes for the protein phenylalanine—tRNA ligase alpha subunit and was used as a proxy for total genome count. Counts for two other single-copy genes (*GUK1*: guanylate kinase and *alaS*: alanine—tRNA ligase) were also inspected, but they were not used because *GUK1* searches showed low counts, and *alaS* showed slightly different but proportional counts to *pheS*, which validated the usage of *pheS* to normalise estimates of total genomes in the samples. However, 115 samples did not include *pheS* count data and were subsequently removed from our analysis. Based on examination of data distribution (distribution shown in **Supplementary Figure 1** in Supporting Information), samples with a low *pheS* count <100 (*n*=5,479) and samples with a (*buk*+*atoA)*/*pheS* ratio >30% (*n*=804) were removed from our analysis to minimise bias from samples with a low normalising pheS count. Outlier samples with pheS count >50,000 (*n*=19) were also removed from our analysis. In addition, samples that did not fall within the scope of our research question (*n*=3,098), such as deep subsurface, contaminated (e.g., uranium-contaminated sites), and experimentally altered samples, were excluded from analysis. The remaining 13,078 samples were retained and analysed for our project. The samples originated from a broad range of sources, including soils and sediments, marine samples, human and non-human animal faecal samples, and wastewater samples.

BPC scores for shotgun metagenomic samples

To derive the Butyrate Production Capacity (BPC_{meta}) score for each sample with metagenomic data, the following formula was developed:

$$\text{Sample } BPC_{meta} \text{ score} = \log10\left(\left(\frac{\frac{CountGene1}{MeanGene1Copies}}{CountSCG}\right) + \left(\frac{\frac{CountGene2}{MeanGene2Copies}}{CountSCG}\right)\right) * 10{,}000$$

where: SCG = single copy gene (*pheS*)

Gene1 = *buk*, Gene 2 = *atoA*

CountGene1, CountSCG are from global metagenomics sample datasets

MeanGeneXCopies = mean count of copies of gene X among all genomes found from searches of gene X within the IMG/M genome database.

Once BPC_{meta} scores were computed and added to the spreadsheet using Excel formulae, the samples were sorted into six categories: soil and terrestrial sediments, aquatic, human, non-human animal, plant, and agro-industrial (**Supplementary Table 4** in Supporting Information). Samples were then grouped by subcategories for further analysis: human samples were grouped by body compartment; non-human animal samples were grouped by vertebrate/invertebrate and by phylum; plants were grouped by compartment; soil samples were grouped by anthropogenic biome classification (i.e., anthromes, which are "terrestrial biomes based on global patterns of sustained, direct human interaction with ecosystems"; Ellis & Ramankutty, 2008); aquatic samples were grouped by source subcategory; and agro-industrial samples were grouped by source site. To reduce bias within the anthrome "Class" categories, two studies whose samples accounted for >50% of the total class samples were removed from the analysis.

To identify whether our BPC_{meta} formula was estimating butyrate production rather than general anaerobicity, we adapted it to estimate ethanol production, which also requires anaerobic conditions. The butyrate synthesis genes were replaced with the terminal gene for alcohol dehydrogenase (*ADH*, EC 1.1.1.1) to derive an Ethanol Production Capacity (EPC) score. Comparing the resulting EPC scores of the soil metagenomic samples with their BPC_{meta} scores showed a negligible correlation (**Supplementary Figure 2** in Supporting Information). This further validated that our BPC_{meta} scores were specific to butyrate production.

Statistics

Statistics were done in R (version 4.0.2; R Core Team, 2020). The Shapiro-Wilk test was used to determine the normality of distribution. In each category the BPC_{meta} scores did not fit a normal

distribution, and either the non-parametric Kruskal-Wallis test or the Wilcoxon rank-sum test was used to test for between-group variation. Due to a high *n* in some subgroups, a post hoc Dunn test with Bonferroni correction was used to compare subgroup pairwise differences at α=0.05. The R package 'ggplot2' (version 3.3.5; Wickham, 2016) was used for data visualisation. Mapping of samples using the pseudocylindric 'Robinson' projection in QGIS (v 3.2.2; QGIS Development Team, 2020) was performed on 2,850 sample metadata coordinates after excluding 360 samples with coordinates with less than two decimal points and 153 samples with no coordinates.

Australian 16S rRNA amplicon sample analysis

BPC scores for 16S rRNA amplicon samples

To assemble 16S rRNA gene abundance data in Australian soil samples, the Australian Microbiome Initiative (Bissett et al., 2016) database was queried for the following parameters: Amplicon = "27F519R", Kingdom = "bacteria", Environment = "is soil", Depth = "between 1 and 10" (cm). We downloaded the abundances of sequences with 100% identity threshold (zero-radius operational taxonomic units or zOTU) and metadata for each resulting sample (*n*=3,023). We used the phyloseq package (McMurdie & Holmes, 2013) for managing and cleaning the 16S rRNA data. We removed all "chloroplast" and "mitochondria" data. We removed low abundance zOTUs that did not occur in at least two samples and had total counts of <20, as these may have arisen from processing errors. In addition, we kept only samples with total number of sequences between 30,000 and 500,000 to remove outliers and samples with low sequence depth. The resulting sample count was *n*=2,795. To normalise the data and reduce bias in comparing samples, we transformed the zOTU abundances into relative abundances.

Because 16S rRNA data can have a relatively poor resolution at the species level (Jovel et al., 2016) as well as many unclassified genera, we focused our BPC_{16S} derivation on family-level data. Using the Genome Taxonomy Database (GTDB) website interface and the set of putative butyrate-producing species (*n*=118) from Vital et al. (2014), we collated the families that included members

in this list of butyrate-producers (*n*=54, **Supplementary Table 5** in Supporting Information). This family list was matched with the Australian Microbiome Initiative taxonomy listings for each downloaded sample. Of the 54 taxonomic families with butyrate-producing bacteria analysed, 31 families had no representative (zOTUs) in any sample. The proportion of butyrate producers in each butyrate-producing family (with *n*=23 such families represented in our dataset) was then used to estimate the abundance of butyrate-producing taxa within each sample and a corresponding BPC_{16S} score, as follows:

Sample BPC_{16S} score =

$$\left(\sum_{i=1}^{n=54}\left[\left(RelAbundZOTUs\ from\ family\ 1\right)\left(\frac{\#\ butyrate\ producing\ species\ in\ family\ 1}{\#\ species\ in\ family\ 1}\right)+\right.\right.$$

$$\left(RelAbundZOTUs\ from\ family\ 2\right)\left(\frac{\#\ butyrate\ producing\ species\ in\ family\ 2}{\#\ species\ in\ family\ 2}\right)+\cdots+$$

$$\left.\left.\left(RelAbundZOTUs\ from\ family\ n\right)\left(\frac{\#\ butyrate\ producing\ species\ in\ family\ n}{\#\ species\ in\ family\ n}\right)\right]\right)$$

Where:family1= Acetonemaceae, family 2 = Acidaminococcaceae, … (see **Supplementary Table 5** for a list of all 54 families). Count zOTUs in each butyrate-producing family are from Australian Microbiome Initiative datasets. # butyrate-producing species (and total binomial species) in each family were evaluated from the entire GTDB.

Modelling ecological associations of BPC_{16S} scores

To provide further macroecological context to butyrate-producing bacteria in soils, BPC_{16S} scores were associated with geographically-paired ecological metadata. We chose 16S rRNA amplicon-based studies for this analysis because many soil studies in Australia have utilised 16S rRNA data, and the Australian Microbiome Initiative facilitated access to a continental coverage of data collected via consistent sampling and bioinformatic protocols. By selecting a substantial yet manageable spatial scale (i.e., continental Australia), we efficiently examined associations of a larger pool of ecological characteristics with BPC_{16S} scores. We also used ecological metadata from sources focusing solely on Australia (e.g., Atlas of Living Australia), which differs from the metadata sources utilised in our global analyses (e.g., anthropogenic biomes).

To associate ecological data with BPC scores, we first removed any non-continental Australia samples (*n*=342). Covariate data were then collated from a variety of sources and reflected a range of soil-forming factors (i.e., SCORPAN variables: S=soil; C=climate; O=organisms; R=relief; P=parent material; A=age; N=spatial location; McBratney, Santos, & Minasny, 2003) (see **Supplementary Table 6** in Supporting Information for a list and description of all ecological SCORPAN variables). We identified 49 predictor variables (43 continuous, 6 categorical) as being relevant to our study, for which data sets were downloaded from the following sources: Australian Microbiome Initiative (e.g., organic carbon, clay content %, conductivity; Bissett et al., 2016), Atlas of Living Australia (e.g., annual temperature range, aridity index annual mean; Belbin, 2011; Williams et al., 2010), Soil and Landscape Grid of Australia (e.g., Prescott index, topographic wetness index; O'Brien, 2021), and Geoscience Australia (silica index; Cudahy, Caccetta, & Thomas, 2012). We used the best available resolution of source data as supplied to avoid introducing additional noise or bias into our analyses. For example, certain analytical test results were available from sample metadata corresponding to 16S rRNA amplicon data, while other ecological covariate data were extracted from gridded spatial ecological layers at points corresponding to the site locations.

Analysis of the predictor variables showed multiple instances of collinearity (e.g., $r > 0.80$ or Variable Inflation Factor scores >15), and some scatterplots generated showed a curvilinear relationship with BPC_{16S} scores (scatterplots shown in **Supplementary Figure 3** in Supporting Information). Therefore, we used two approaches (detailed below) that were less influenced by collinearity for subsequent analyses: principal components analysis into *k*-means clustering and decision tree modelling via Random Forest (Breiman, 2001). Because many samples did not include soil physicochemical data, removing incomplete cases (*n*=1,122) left 1,331 samples for our analyses (**Supplementary Table 7** in Supporting Information). A further five samples were also removed as they were characterised by outlying values of continuous variables.

We scaled and analysed the 43 continuous predictor variables using principal component analysis (PCA) to reduce the dimensionality of the variables. PC1 and PC2 explained 27.6% and 14.4% of variance, respectively, and were selected for data visualisations (**Supplementary Figure 4** in Supporting Information). We then used *k*-means clustering on scaled original data to assign the samples into clusters. The optimal number of clusters was examined using the "elbow" method, silhouette method, and gap statistic method. While four was considered an optimal number of clusters, we examined both four and five clusters. We found that the additional fifth cluster more distinctly separated land types. Thus, the five-cluster approach was selected for analysis. The resulting cluster data was collated, and BPC_{16S} scores were then matched and returned to the dataset. Median values were calculated for each variable in each cluster, revealing ecological trends distinct to each cluster. Between-cluster significance was examined using Kruskal-Wallis tests. We gave each cluster a generalised description and plotted the sample geospatial coordinates into maps using 'ggmap' (Kahle & Wickham, 2013) and Google maps to visualise their geographical distributions.

We then used Random Forest regression modelling (Breiman, 2001) via the R package 'spatial RF' (Benito, 2021) to discern variable importance results and obtain partial dependence plots for each variable against BPC_{16S} scores. Only the 43 continuous variables were included in this analysis due to 'spatial RF' package limitations. The model fit was estimated using out-of-bag error from the bootstrap. To reduce multicollinearity, highly correlated predictor variables ($r > 0.8$ or VIF > 12) were removed (n=9). Tuning the hyperparameters of the model (mtry=24, num.trees=500, min.node size=5) improved its performance (R^2) by 0.006. Spatial autocorrelation of the residuals was then minimised while fitting the spatial regression model. The resulting Random Forest decision tree model explained 46.5% of the variance in our BPC_{16S} dataset. The variable importance plot was created using random permutations for each predictor variable's values in out-of-bag data, then calculating the mean decrease in node impurity. Thirty model repetitions were used to create the plot of variable importance. Partial dependence plots were then generated and confirmed the non-linear

relationship of most variables with BPC_{16S} scores (**Supplementary Figure 5** in Supporting Information).

RESULTS

Global distribution of butyrate-producing bacteria

Metagenomes with genes for butyrate production were found on every continent, in every ocean, and in 89 countries (**Figure 1A**). Overall highest median BPC_{meta} scores were found in human host-associated (2.99, *n*=1,553) and non-human animal host-associated samples (2.91, *n*=771), with the lowest median BPC_{meta} scores in aquatic samples (1.93, *n*=6,017) (**Figure 1B**).

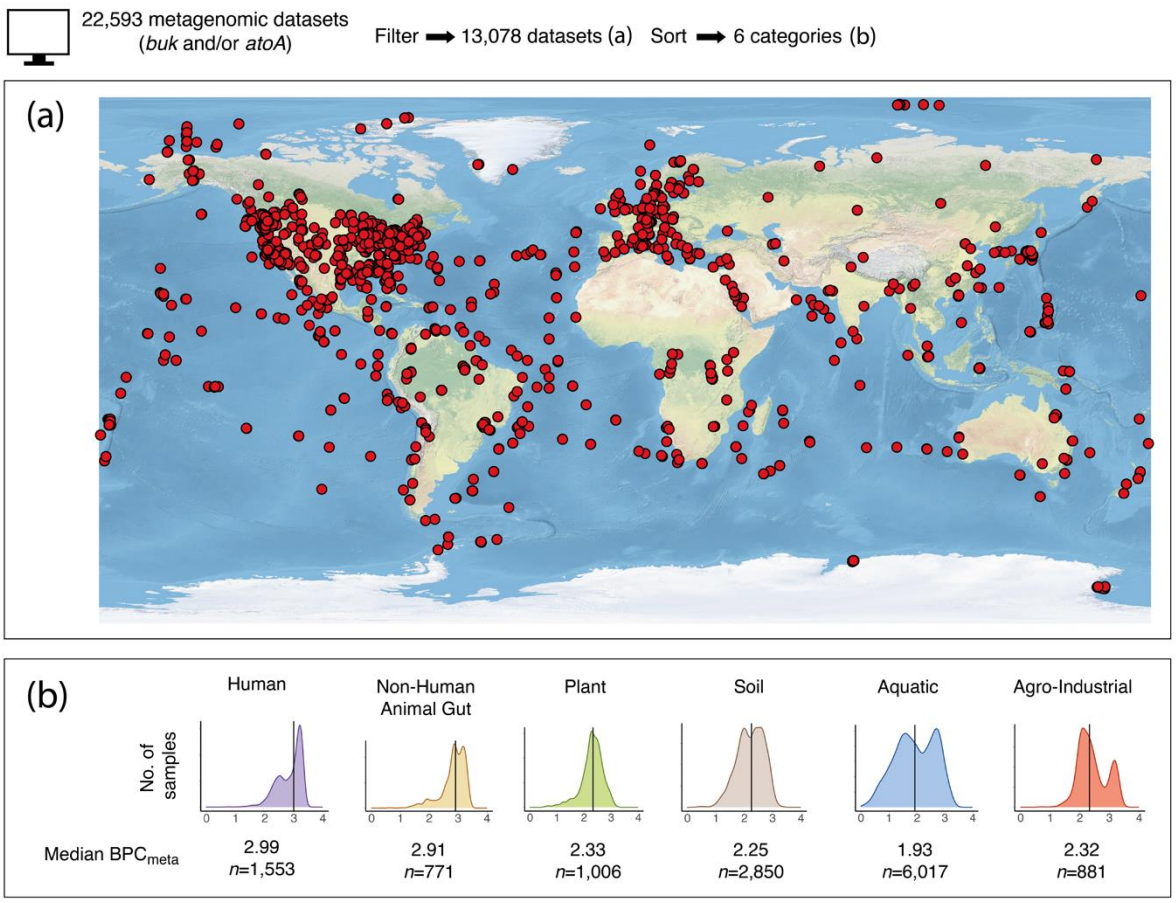


Figure 1: Butyrate-producing bacteria are found on every continent, in every ocean, and in 89 countries. (a) Map showing study locations of samples with *buk* and/or *atoA* genes. (b) Density plots showing frequency distributions of sample Butyrate Production Capacity (BPC_{meta}) scores in the six highest-level groupings (x-axis=BPC_{meta}). BPC_{meta} score medians rather than means are presented due to non-normal BPC_{meta} score distributions. The range of sample BPC_{meta} scores was from 0.02 to 3.39. Bimodal peaks in five of the six categories may represent divergence between environments supportive and unsupportive of fermentative activity (discussed below). *n* is the number of samples.

Butyrate production capacity of different ecosystems

Human hosts

Human samples were sorted into five body compartments: skin, nasal, oral, genital, and gut. The highest median BPC_{meta} score came from the gut (3.19, *n*=800), with faecal samples expected to be acting as a proxy for the anaerobic gut environment. The lowest median BPC_{meta} score came from the skin (1.86, *n*=17), which is exposed to aerobic conditions. Between-group differences were statistically significant (Kruskal-Wallis test: H=1136, 4 d.f., *P*=<0.001; **Figure 2A**).

Non-human animal hosts

Non-human animal host-associated samples included in our analysis (*n*=771) were either direct or proxy (e.g., faecal) measures of animal gut microbiota (*n*=448) or were non-gut but host-associated samples (e.g., attine ant fungus gardens, gutless marine worms, *n*=323). We first compared non-human animal groupings by vertebrates (median BPC_{meta} score=3.11, *n*=389) and invertebrates (median BPC_{meta} score=2.76, *n*=382) (between-group differences were statistically significant: Wilcoxon rank sum test: W=22,592, *P*<0.001). We then compared non-human animal samples by taxonomic phylum (between-group differences were statistically significant: Kruskal-Wallis test: H=331, 4 d.f., *P*<0.001; **Figure 2B**), where Chordata had the highest median BPC_{meta} score (3.11, *n*=389) and Porifera (sponges), which lack a gut, had the lowest BPC_{meta} scores (1.87, *n*=34). A further comparison showed that the human gut median BPC score (3.19, *n*=800) was similar to the primate gut (3.12, *n*=26).

Plant hosts

Our dataset included 1,006 plant-associated samples. These were subcategorised into four groups by plant compartment: leaf surface, plant litter, rhizosphere, and root. Root samples had the highest median BPC_{meta} score (2.50, *n*=123). Leaf surface samples had the lowest median BPC_{meta} score (1.76, *n*=30; between-group differences were statistically significant: Kruskal-Wallis test: H=105, 3 d.f., *P*<0.001; **Figure 2C**).

Soil ecosystems

Soil samples (*n*=2,850) were sorted using the anthropogenic biome (anthrome) categories (Ellis et al., 2021; Gauthier, Ellis, & Goldewijk, 2021), representing varying densities of human population and land use (anthrome classes and world map shown in **Supplementary Figure 6** in Supporting Information). At the "Level" category of anthromes, the highest median BPC_{meta} scores came from both "Dense settlements" (includes classes "urban" and "mixed settlements"; median BPC_{meta} score=2.38, *n*=297) and "Cultured" (includes "woodlands" classes and the "inhabited drylands" class; median BPC_{meta} score=2.36, *n*=1076). The lowest median BPC_{meta} score (1.94, *n*=167) came from the anthrome level "Wildlands", which has the lowest human influence (between-group differences were statistically significant: Kruskal-Wallis test: H=186, 5 d.f., *P*<0.001; **Figure 2D**).

Aquatic ecosystems

Aquatic samples (*n*=6,017) were sub-grouped into five categories: marine, freshwater, brackish water and estuary, springs, and inland saltwater. The highest median BPC_{meta} score (2.52, *n*=911) was found in inland saltwater samples, and marine samples had the lowest median BPC_{meta} score (1.67, *n*=2047) (between-group differences were statistically significant: Kruskal-Wallis test: H=530, 4 d.f., *P*<0.001; **Figure 2E**).

Agricultural and industrial samples

Agricultural and industrial ("agro-industrial") samples (*n*=881) were from a wide variety of sources and materials. We grouped them into seven source types, which include two sample types from wastewater treatment plants (i.e., activated sludge from aeration tanks and anaerobic digesters). The highest median BPC_{meta} scores (3.16, *n*=120) were from anaerobic digester samples. The lowest median BPC_{meta} scores were from the agricultural soils (2.16, *n*=486) and activated sludge (2.23, *n*=50) (between-group differences were statistically significant: Kruskal-Wallis test: H=431, 6 d.f.,

P<0.001; **Figure 2F**). Activated sludge is a bacteria-rich product formed in aeration tanks with aerobic conditions.

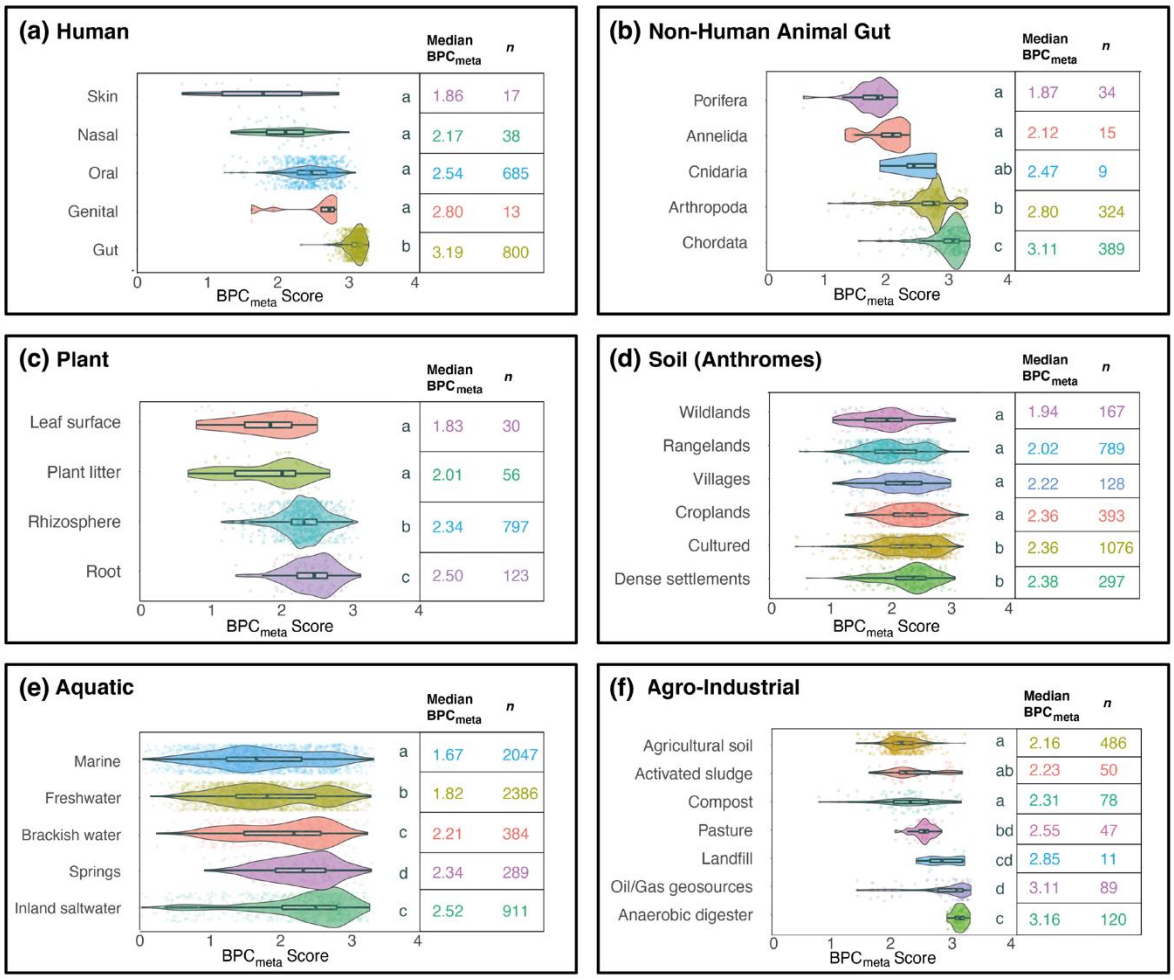


Figure 2: Butyrate Production Capacity (BPC_{meta}) scores vary between host communities and ecological sources. BPC_{meta} score density plots by group subcategories. (a) BPC_{meta} scores of humans, sorted by body compartment. (b) BPC_{meta} scores of non-human animal-associated microbial communities, sorted by class. Note that Porifera do not possess a gut. (c) BPC_{meta} scores of plant-associated samples, grouped into compartments. (d) BPC_{meta} scores of soil samples, grouped into anthropogenic biomes (anthromes) levels that represent human influence on land use. The level "Cultured" includes woodlands and inhabited drylands. (e) BPC_{meta} scores of aquatic ecosystem samples, grouped into source site categories. (f) BPC_{meta} scores of agricultural and industrial samples, grouped by source site. Activated sludge and anaerobic digesters are common components of wastewater treatment plants. In each of (a)-(f), Kruskal-Wallis tests show that between-group differences were significant at P<0.001. Medians sharing a letter are not significantly different by the adjusted Dunn test at the 5% significance level. Boxes show the interquartile range.

Ecological characteristics associate with soil BPC scores

When clustering the Australian 16S rRNA soil samples on ecological metadata, each cluster aligned with distinct representative land types, which were given the following descriptive titles: ‘arid inland clay plains’ (cluster 1), ‘seasonally productive sandy soils’ (cluster 2), ‘sandy inland deserts’ (cluster 3), ‘temperate urban hinterland’ (cluster 4), and ‘wet, cold, acidic, vegetated montane’ (cluster 5). Mapping the geographical locations of the sample sites showed consistency with the clustered land type descriptions (**Figure 3A**). Median BPC_{16S} scores varied significantly between the clusters (Kruskal-Wallis test: H=164, 4 d.f., $P<0.001$; **Figure 3B**). The highest median BPC_{16S} scores came from the ‘seasonally productive sandy soils’ (0.93, $n=383$) and ‘temperate urban hinterland’ cluster (0.86, $n=440$). The lowest median BPC_{16S} score came from the ‘wet, cold, acidic, vegetated montane’ cluster (0.49, $n=208$). Principal components analysis of all continuous ecological predictor variables showed two distinct patterns of axial distribution, ecological wetness and greenness and soil fertility, and dimensions 1 and 2 explained 27.6% and 14.4% of the variation in the data, respectively (**Figure 3C**).

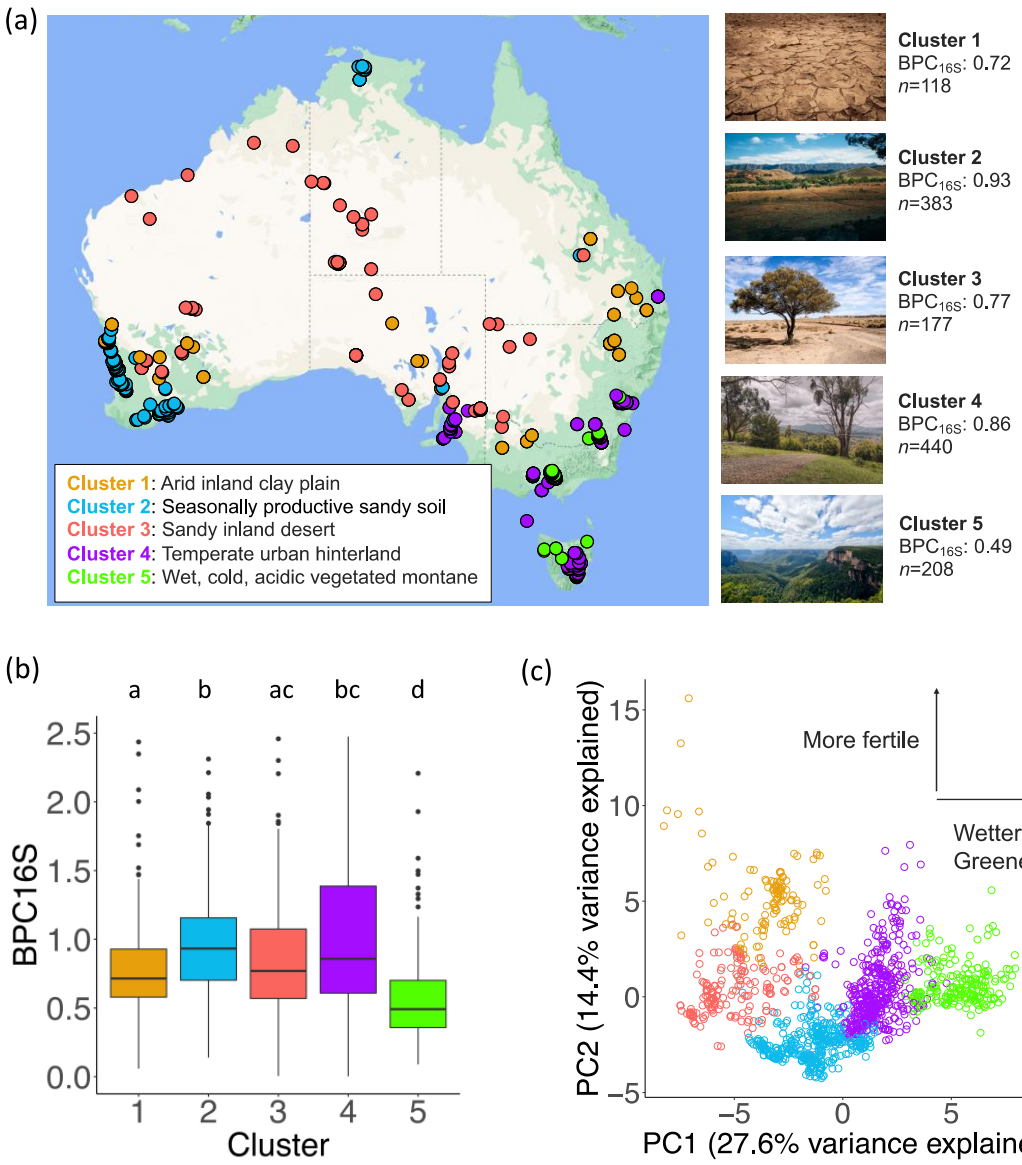


Figure 3: Clustering of ecological data shows five distinct land types. (a) Map of Australian soil samples clustered on 43 continuous ecological variables, five cluster distribution, mapped using R package ggmap and Google maps. Photographs were downloaded from Unsplash.com under CC0 license. (b) Boxplots of median Butyrate Production Capacity (BPC_{16S}) scores across each of the five clusters. Medians sharing a letter are not significantly different by the adjusted Dunn test at the 5% significance level. Boxes show the interquartile range. (c) The first two principal components coloured by k-means clusters. The x-axis can be broadly interpreted as ecological wetness and greenness and associated variables (e.g., vegetation cover). The y-axis can be broadly interpreted as soil fertility and the presence of cations.

The cluster with the highest median BPC_{16S} score was the 'seasonally productive sandy soil', which has low clay content, low levels of cations, high precipitation seasonality and mean temperature, and generally moderate topographic relief and Prescott index (a measure of soil water balance based on average precipitation and potential evaporation). The cluster with the second-highest median BPC_{16S} score was 'temperate urban hinterland', which is generally moderate in elevation, annual rainfall,

topographic relief, clay, and soil fertility and has high levels of zinc and manganese. The cluster with the lowest median BPC_{16S} score, 'wet, cold, acidic, vegetated montane', had high elevation and topographic relief, colder mean annual temperature, high annual rainfall and aridity index, consistent rainfall levels throughout the year, high soil organic carbon content, and high soil iron and aluminium content (where soil aluminium content and pH are inversely correlated). The two additional clusters, 'arid inland clay plains' and 'sandy inland desert', also had distinct characteristics (**Supplementary Table 8** in Supporting Information). A separate analysis of categorical variables also showed the highest BPC_{16S} scores among specific land cover types ("built-up" and "plantation"), land use types ("grazing of native pastures" and "rural residential"), and anthromes ("mixed settlements" and "remote rangelands") (**Table 2** and **Supplementary Table 9** in Supporting Information).

Table 2. Categorical variable subcategories with the highest Butyrate Production Capacity (BPC_{16S}) scores†.

Category	Highest median BPC_{16S}	Second-highest median BPC_{16S}
Land cover	Built-up (1.39, *n*=53)	Plantation (softwood/mixed) (1.13, *n*=11)
Land use	Grazing of native pastures (1.39, *n*=439)	Rural residential (1.28, *n*=26)
Anthrome	Mixed settlements (1.07, *n*=28)	Remote rangelands (1.00, *n*=1,103)
Vegetation type	Eucalyptus open woodlands (1.48, *n*=147)	Hummock grasslands (1.34, *n*=77)
Major vegetation subgroup	Mitchell grass (Astrebla) tussock grasslands (1.90, *n*=38)	Eucalyptus open woodlands with a grassy understorey (1.80, *n*=28)

† Only subcategories with $n \geq 10$ are reported here.

The variable importance plot from Random Forest decision tree analysis showed that geography most closely associates with the BPC_{16S} scores in soils (**Figure 4A)**. The eight top predictor variables were longitude, elevation, iron, topographic relief, latitude, precipitation seasonality,

population density, and the Prescott index. Specifically, low-to-moderate elevation, higher soil iron, and moderate-to-high topographic relief showed a close relationship with higher BPC_{16S} scores (**Figure 4B**). Partial dependence plots for each variable are shown in **Supplementary Figure 5**.

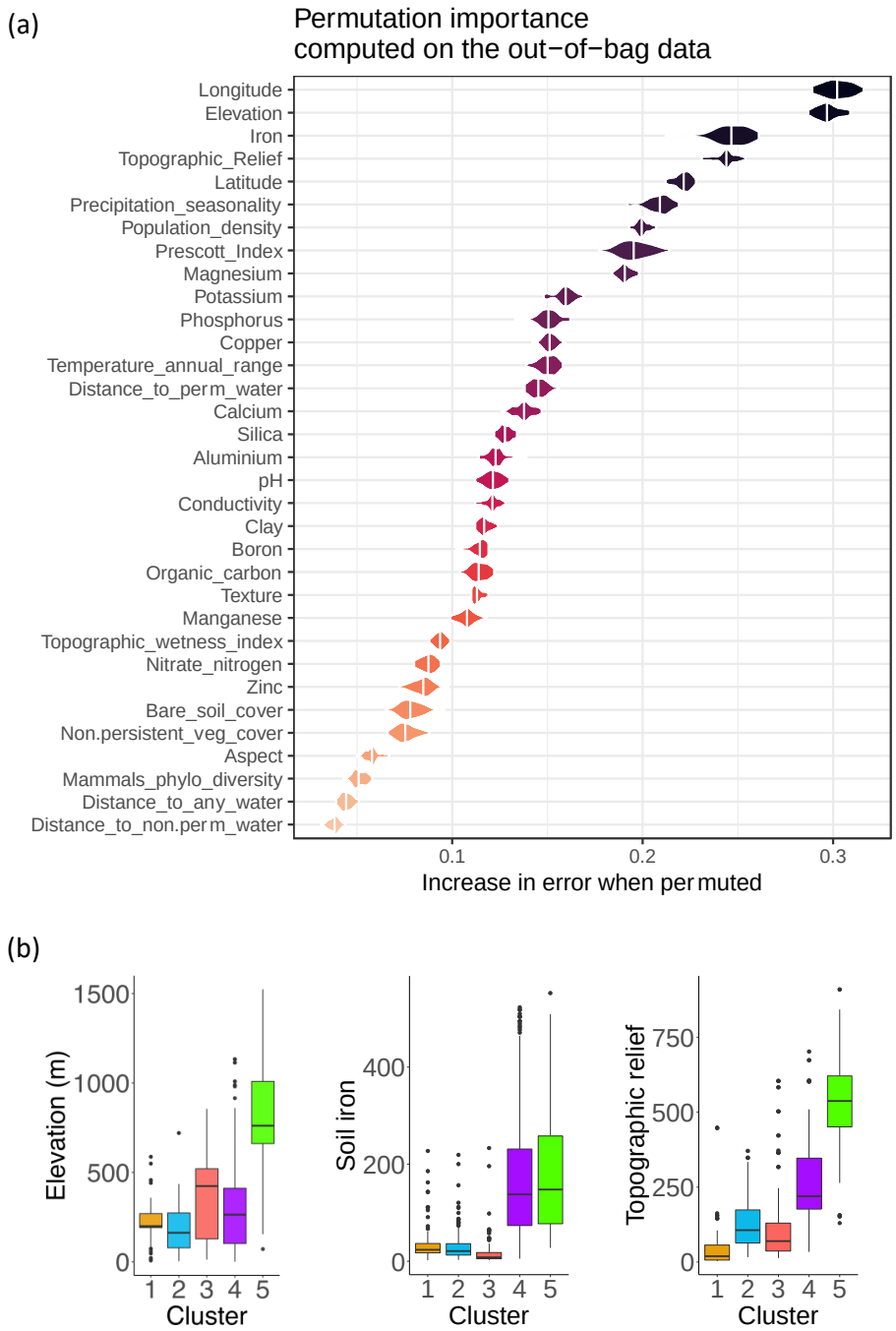


Figure 4: Variable importance results from Random Forest decision tree modelling. (a) Random Forest variable importance results from 43 continuous ecological predictor variables. The model was fitted using out-of-bag errors from the bootstrap. The variable importance was determined using random permutations of predictor variables and the mean decrease in node impurity. (b) Boxplots for selected top variables from (a) across each of the five clusters. In each of elevation, soil iron, and topographic relief, Kruskal-Wallis tests show that between-group differences were significant at $P<0.001$. Boxes show the interquartile range.

DISCUSSION

We report macroecological patterns of butyrate-producing bacteria via our novel BPC formulae that use both metagenomic and 16S rRNA amplicon bacterial data. Our BPC_{meta} data showed that samples from anaerobic and fermentative conditions, such as the animal gut compartment and soil rhizosphere, had increased genomic potential for butyrate production. Our BPC_{16S} results showed that geographical and climatic variables, soil iron, and human population density were key ecological covariates of soil butyrate production capacity. Soils from sandy, sparsely-populated rangelands with high precipitation seasonality, as well as temperate peri-urban sites with greater soil fertility, had higher BPC_{16S} scores. We discuss the novel and distinct macroecological patterns of butyrate production observed below.

Anaerobic conditions associated with higher butyrate-production gene abundances

We show that samples from the chordate gut and anaerobic digesters had the highest BPC_{meta} scores. These results are consistent with the knowledge that butyrate production occurs in anaerobic compartments, such as the human gut and anaerobic digesters (Conrad, 2020; Liu et al., 2016). Expectedly, air-exposed skin surface, animals without a digestive system (e.g., sponges), and activated sludge from aeration tanks had among the lowest BPC_{meta} scores for their respective categories. However, our results also show that this anaerobic requirement is maintained within plant-soil systems that are more susceptible to complex ecological dynamics. Underground compartments, such as roots and rhizospheres, had significantly higher median BPC_{meta} scores than air-exposed compartments such as leaf surfaces. Uteau et al. (2015) demonstrated within plant rhizospheres a spatial oxygen gradient based on the presence of oxygen-filled pores. They also found that episodic water-saturated conditions within the rhizosphere decreased the oxygen partial pressures substantially, which may help explain our finding that topographic relief, precipitation seasonality, and evapotranspiration rates play important roles in the butyrate production capacity within soils.

Anaerobicity emerged as a key pattern in the bacterial genetic potential for butyrate production across a range of substrates and environments. In some substrates and locations, environmental conditions may reflect human gut conditions (e.g., predominantly anaerobic, wet, and dark) and may thus harbour butyrate-producing bacteria as long-term "residents" through environmental selection processes. However, our finding of bacterial genes for butyrate production in environmental samples does not indicate whether these bacteria are residents of that location or are transient allochthonous taxa. This would be especially true with soil samples from locations where animals are present and faecal contamination is possible. As such, we examined samples from both anaerobic and aerobic processes at wastewater treatment plants, both of which would be expected to contain faecal matter. We found that samples from the anaerobic processes had higher levels of butyrate production capacity than from the aerobic processes. Thus, our evidence suggests that anaerobic conditions in soils could select for higher butyrate production capacity, raising the probability that these butyrate-producing bacteria could be long-term residents of the soil rather than resulting from temporary faecal contamination. Future research should examine the ecological forces that shape anaerobic bacterial assemblages within outdoor environmental substrates.

Influence of human population density on soil BPC_{16S}

We report an association between the population density of humans and soil BPC_{16S} scores. Soils from the 'built-up' land cover, 'mixed settlements' anthrome, and 'temperate urban hinterlands' cluster showed among the highest BPC_{16S} scores. In contrast, soils from the 'remote woodlands' and 'wild woodlands' anthromes had among the lowest median BPC_{16S} scores. The anthropogenic influence on terrestrial ecosystems is well-known (Ellis et al., 2021), and our findings show this influence extends to microbial abundances, particularly butyrate-producing bacteria.

We show that 'temperate urban hinterlands' are a substantial reservoir of butyrate-producing bacteria. Geographical and climatic data show that these Australian sample sites are near major cities

and are moderate in their elevation, mean temperature, topographic relief, and annual rainfall. Thus, the rainfall tends to run off rather than accumulate, which creates the potential for fluctuating hydrology. Additionally, the increased BPC_{16S} scores from these peri-urban locations may be influenced by inadvertent dissemination and artificial soil 'contamination' with gut-associated bacteria from people and their pets (Blum, Zechmeister-Boltenstern, & Keiblinger, 2019), possibly contributing into a transmission cycle of butyrate producers. On the other hand, Australia's major cities and hinterlands are typically coastal, often with river-floodplain systems and areas of fertile soils that were attractive to European settlers. Thus, our BPC_{16S} results suggest an association between high human population density and high BPC scores. However, the direction of influence remains a compelling question that may be of future research interest.

Hot, seasonally productive sandy soils

Our cluster 'seasonally productive sandy soils' had the highest BPC scores. These soils associated with high precipitation seasonality, which was the climatic variable with the strongest importance toward BPC scores. Hydrological fluctuations in soils have been shown to induce changes in both bacterial community structure and exometabolites (extracellular metabolites). For example, the wetting process in dry soil biocrusts can induce a shift from a cyanobacteria-dominated to a Firmicutes-dominated community within 49.5 hours (Swenson et al., 2017). RoyChowdhury et al. (2022) also reported that soil abundances of Firmicutes increased in a dry-to-wet transition, even more so than in fixed saturated conditions. Because most butyrate producers are within the phylum Firmicutes, hydrological fluctuations may play a key role in butyrate production.

The 'seasonally productive sandy soils' cluster also had a high mean temperature, low soil fertility, and low population density, which differs from the above patterns of higher BPC scores around temperate residential areas. Furthermore, these soils had high non-persistent vegetation cover (**Supplementary Table 8**). Seasonal rainfall could lead to flushes of green growth, followed by a suitable climate to enable microbial breakdown. Sandy soils would indicate less protection of

organic matter associated with clay minerals, which could enhance microbial decomposition of existing organic matter (Johnston, Poulton, & Coleman, 2009). This suggests a high turnover system, and indeed this cluster contains 67% of our samples characterised by land use as 'cropping lands' (**Supplementary Table 7**).

Soil iron associated with BPC

We observed that iron had the strongest association with BPC scores. Our partial dependency plot shows moderate soil iron levels associated with decreased BPC scores, but BPC scores then increased with higher iron levels. This result is consistent with findings from Dostal et al. (2015), where high iron levels enhanced butyrate production in *Roseburia intestinalis* cultures. Additionally, the authors' *in vitro* child gut fermentation model showed that a strong iron deficiency significantly decreased butyrate production. It is worth noting that the butyrate production pathway requires several ferredoxins and ferredoxin-like proteins, which are proteins structured with iron, during the reduction of crotonyl-CoA into butyryl-CoA (Chowdhury, Kahnt, & Buckel, 2015). On the other hand, our results also show that 'wet, cold, acidic montane', the cluster with the lowest median BPC_{16S} score, also had a high concentration of iron. Lee, Miyahara and Noike (2001) found that butyrate production from a sucrose solution began to decrease as iron concentration increased above 20 mg/l. Iron availability is also pH dependent, increasing in acidic conditions. Thus, the relationship between iron concentrations and butyrate production is not linear and may rely on a suite of conditions, possibly including the presence of wetting/drying cycles, that work in tandem to create the potential for butyrate production.

Future research directions and limitations

During the development of our methods, several limitations of our study became apparent. Analyses of shotgun metagenomic sequences and bacterial 16S rRNA amplicons rely on reference databases that are continually being developed but are incomplete. Missed or incomplete sequence identification could affect the reliability of our formulae. Likewise, taxonomy databases are regularly

updated, but their updates are not uniform across databases. We used the Genome Taxonomy Database to classify our list of butyrate-producing bacteria, but it showed occasional discrepancies with the classification system applied to the Australian soil 16S rRNA data. Thus, utilising multiple taxonomic classification systems likely means that some butyrate producers were not identified in our data. This could affect the reliability of our results and a future sensitivity analysis is warranted. In addition, 16S rRNA amplicon bioinformatic methods offer poor resolution at the species level, and our reliance on proportions of butyrate producers in family level data is only an estimation. Greater precision could result from utilising relative abundances of shotgun metagenomic reads of butyrate-producing species across other studies as a method of estimating proportions of butyrate producing bacterial species.

To maximise the precision of our butyrate producer database, we chose to use species-level classifications via GTDB representative species. However, this may have inadvertently created inconsistent data from species with multiple strains (sometimes hundreds of strains are present within a species), among which some may be butyrate producers and others not. Analysis at the strain level could provide a higher resolution of data, which should be a future research priority. In addition, our data were genomic only and did not include direct measurements of butyrate concentrations; this would require different data types (e.g., metabolomic data) and pipelines. Future studies that examine butyrate concentrations in relation to butyrate-producing taxonomic and functional gene abundances could more precisely reveal conditions that promote active butyrate production beyond our estimations. Furthermore, because our 16S rRNA data came only from Australia, our modelling may not be generalisable to global conditions that exceed the ranges of our ecological covariate data. For example, the height of mountains on the Australian mainland does not exceed 2,228 m; thus, our mountain cluster modelling may not fit other countries with higher mountains.

In our examination of soil shotgun metagenomic data using anthromes, the sampling methods were not disclosed and might not have been uniform. Metadata that includes sampling depth or indicates whether the soil is horizontally layered or horizontally homogenised would be necessary to draw more comprehensive conclusions. There is also a possibility that urban soils may also be influenced by either the use of compost, which may still bear bacterial spores or DNA (Subirats, Sharpe, & Topp, 2022), or direct contamination by animal faeces. Thus, we have presented the soil BPC_{meta} results but have refrained from drawing conclusions from them; future studies should include details of the sampling methods, which would reduce such biases.

Finally, our data depends on the capacity of laboratory DNA extraction methods to open endospores. Sampling methods that expose the samples to air may inadvertently cause the sporulation of bacteria. Such methods may subsequently reduce the quantities of DNA extracted from spore-formers, several of which could be butyrate-producing bacteria (Browne et al., 2016). Consistency across sampling and DNA extraction methods in future studies could help improve butyrate-producing bacterial abundance data reliability.

Conclusions

Butyrate-producing bacteria provide critical ecosystem services for hosts and environments, including humans, and soils. Our study focused on the distribution and ecological associations of these bacteria, building new knowledge of their roles in human-plant-soil ecosystem dynamics. We present new evidence that many outdoor ecosystems influenced by human-associated processes may represent key reservoirs of butyrate producers. Because nearly 60% of the world's population now lives in urban areas (Güneralp et al., 2020), understanding the influence of dense populations of humans on outdoor urban microbiomes is essential to biodiversity research, informing urban landscape design, and the study of biodiversity-human health linkages (Delgado-Baquerizo et al., 2018; Kondo et al., 2018; Watkins et al., 2020). Our study helps advance these research areas. Future assessment of butyrate-production capacity across fine spatial scales (i.e., below global and

continental, as used here) will help provide greater detail for city infrastructure planning and further microbiome-based public health and ecology research.

Chapter 3. Urban greenspace aerobiomes are shaped by soil conditions and land cover type

---

This work appears in the following publication:

Brame JE, Liddicoat C, Abbott CA, Cando-Dumancela C, Robinson JA, Breed MF. Urban greenspace aerobiomes are shaped by soil conditions and land cover type. 2024. *bioRxiv* (under review in *Microbial Ecology*), doi: 10.1101/2024.01.12.575340.

Context

While Chapter 2 examined environmental influences on bacterial abundances across global and continent-wide scales, Chapter 3 uses a metropolitan-scale focus to examine environmental influences on aerobiomes. Urban greenspaces harbour airborne bacterial communities (aerobiomes) with the potential to transfer beneficial bacteria to humans. However, limited studies have examined the ecological influences of soil, vegetation, and rainfall on aerobiomes in urban greenspaces. Here, we utilised 16S rRNA amplicon sequence data to analyse the effects of land cover, soil abiotic characteristics, surrounding vegetation diversity, and rainfall on aerobiome diversity and composition from 33 urban greenspace sites in Adelaide, South Australia. We sampled air and soil from two urban greenspace land cover types: highly-managed sports fields (n = 11) and minimally-managed nature parks (n = 22).

In Chapter 3, I aimed to answer the following research question: What are the effects of land cover and soil and vegetation characteristics on aerobiome bacterial diversity and composition in two types of urban greenspaces?

Author contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by J.B., C.L, and M.B.. The first draft of the manuscript was written by J.B. and all authors commented on further versions of the manuscript. All authors read and approved the final manuscript.

Original Article

Urban greenspace aerobiomes are shaped by soil conditions and land cover type

Joel E. Brame[a,b], Craig Liddicoat[a,b], Catherine A. Abbott[a], Christian Cando-Dumancela[a,b], Jake M. Robinson[a,b], Martin F. Breed[a,b]

[a]College of Science and Engineering, Flinders University, Bedford Park, SA 5042, Australia

[b]The Aerobiome Innovation and Research Hub (The AIR Hub), College of Science and Engineering, Flinders University, Bedford Park, SA 5042, Australia

Corresponding author email address:

Joel E. Brame: joel.brame@flinders.edu.au

Joel Brame ORCID: 0000-0002-1852-9987

Craig Liddicoat ORCID: 0000-0002-4812-7524

Catherine Abbott ORCID: 0000-0001-8215-4406

Christian Cando-Dumancela ORCID: 0000-0003-0186-0056

Jake Robinson ORCID: 0000-0001-8108-3271

Martin Breed ORCID: 0000-0001-7810-9696

ABSTRACT

Growing evidence suggests that exposure to microbial biodiversity is important for human immunoregulation and health. Urban greenspaces harbour airborne bacterial communities (aerobiomes) with the potential to transfer beneficial bacteria to humans. However, limited studies have examined the ecological influences of soil, vegetation, and rainfall on aerobiomes in urban greenspaces. Here, we utilised 16S rRNA amplicon sequence data to analyse the effects of land cover, soil abiotic characteristics, surrounding vegetation diversity, and rainfall on aerobiome diversity and composition from 33 urban greenspace sites in Adelaide, South Australia. We sampled air and soil from two urban greenspace land cover types: highly-managed sports fields (n = 11) and minimally-managed nature parks (n = 22). Sports field aerobiomes had a distinct aerobiome community composition and higher alpha diversity than nature parks. Aerobiome alpha diversity was shaped more by soil abiotic characteristics, particularly soil pH and iron levels, than woody plant species diversity. Rainfall prior to sampling also had strong effects on the aerobiome community composition and associated with decreased alpha diversity. These findings point toward soil iron and pH management as pathways to increase aerobiome bacterial diversity. Our study shows that, with additional research, there is potential for greenspace managers and urban planners to target specific soil abiotic characteristics in urban greenspaces to improve microbiome-mediated urban health.

INTRODUCTION

Biodiverse microbial exposure via the outdoor environment is interconnected with human health through historical and co-evolutionary mechanisms, including immunoregulation (Jimenez et al., 2021; Robinson, Mills, & Breed, 2018; Rook, 2013). However, the rapid expansion of modern urbanisation is leading to a consequential change in everyday environments (Von Hertzen, Hanski, & Haahtela, 2011). For city residents, urban greenspaces maintain opportunities for exposure to biodiversity (Wintle et al., 2019) and have been associated with improvement in physical, social, and psychological health (Kondo et al., 2018; Lai et al., 2019; Lee, Jordan, & Horsley, 2015). There are various pathways linking biodiversity to human health. These include social (e.g., greenspaces

provide areas for convivial activities), biological (e.g., exposure to stimuli that trigger physiological effects), psychological (e.g., exposure to stimuli that change moods and emotions), physical activity (facilitating exercise), and environmental buffering pathways (e.g., trees providing cooling that reduces heat stress; (Marselle et al., 2021). The biological pathway linking biodiversity to health is thought to encompass exposure to greater microbial biodiversity and specific microbial taxa (also known as "old friends") that regulate the innate immune system, as described by the biodiversity hypothesis (Haahtela, 2019). Growing evidence links greater environmental microbial diversity exposures with increases in immunomodulatory cells, including regulatory T-cells, which provide protection from allergies and some chronic inflammatory diseases (Alfven et al., 2007; Roslund et al., 2022). Thus, evaluations of urban greenspace soils, vegetation, and air for microbial diversity and composition, alongside the analyses of pathogens, can provide insights into their potential linkages with human health.

Airborne bacterial community (aerobiome) biodiversity has been examined across greenspace land covers, but the number of such studies is limited. The alpha diversity of aerobiomes has been compared across several urban land cover types, including parks versus parking lots (Mhuireach et al., 2016), grassy versus forested areas (Mhuireach, Wilson, & Johnson, 2021), and bare ground versus grasslands and scrub habitat (Robinson et al., 2021). Dispersal from leaf surfaces is recognised as an important factor in shaping aerobiome communities (Bowers et al., 2011; Sessitsch et al., 2023). It is well-established that vegetation composition directly impacts – and is impacted by – microbial communities in outdoor urban ecosystems, especially within soils (Baruch et al., 2020). Urban greenspace vegetation can take many forms; however, two contrasting and commonly encountered land cover types include: intensely maintained grass sports fields, and minimally maintained and more natural vegetation systems such as parklands or woodlands. Despite the regular human usage of sports fields and the associated potential microbial exposure, to our knowledge, they have received minimal attention in aerobiome studies.

Furthermore, the ecological influences on aerobiome biodiversity extend beyond vegetation and land cover. Many ecological variables influence the urban aerobiome, including anthropogenic activity (Delgado-Baquerizo et al., 2018), wind-carried bacterial contributions from allochthonous (distant) sites (Robinson & Breed, 2023), and soil physicochemical conditions (Brodie et al., 2007). Temporal variation (Woo et al., 2013), vertical microbial stratification dynamics (Robinson, Cando-Dumancela, et al., 2020), and air pollution (Franchitti et al., 2022), among other ecological and meteorological variables (Bowers et al., 2011), also modulate aerobiome composition. Analysis of these complex ecological interactions from the perspective of beneficial bacterial exposure has only recently gained momentum, and significant knowledge gaps remain. Soil abiotic/physicochemical characteristics (e.g., pH, nutrients) are known to shape the composition and stochastic and deterministic assembly of soil microbial communities (Tripathi et al., 2018). However, their association with aerobiome composition has received limited attention.

Here, we characterised aerobiome community profiles along with vegetation and soil physicochemical parameters from 33 urban greenspaces across metropolitan Adelaide, South Australia, to determine the influence of local plant and soil conditions in shaping aerobiomes. Our urban greenspaces were grouped into two types: highly-maintained grassy sports fields (n = 11) and minimally-maintained nature parks (n = 22). To understand ecological influences on aerobiomes in urban greenspaces, we aimed to evaluate the effects of land cover type, woody plant species diversity, rainfall, and soil physicochemical parameters on **(1)** the aerobiome alpha diversity and **(2)** the aerobiome community composition, including specific commensal taxa.

MATERIALS AND METHODS

Study locations

We sampled spatially independent 25 x 25 m replicates of sports fields (n = 11) and nature parks (n = 22) (33 sites total) in greater metropolitan Adelaide, South Australia (**Figure 1a**). The 25 x 25 m area is shown to be suitable to describe vegetation and microbial variation (Baruch et al., 2020; Mills

et al., 2020). A panoramic photo was taken at the centre of each replicate (**Figure 1b-1c** shows examples of greenspace types). Sites were chosen so that: (1) all sites were at a distance of > 5 km from the coast to avoid coastal effects; (2) all sites were within the low-elevation plains metropolitan Adelaide region to avoid the different (colder, wetter) climatic conditions of the hills and mountains bordering to the east; and (3) nature park sites represented a range of woody plant complexity.

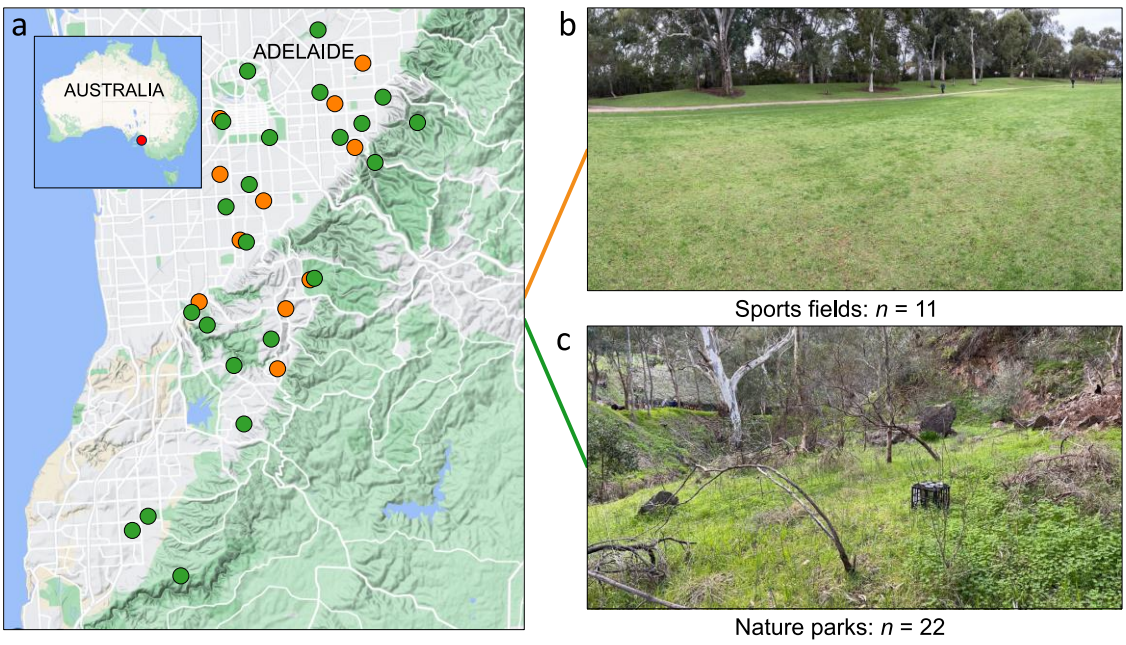


Fig. 1 (a) Map of Australia and sampling sites across greater metropolitan Adelaide. Orange dots denote sports fields, and green dots denote parks. (b) Representative panoramic photo of an Adelaide sports field site. (c) Representative panoramic photo of an Adelaide nature park site.

Vegetation surveys

We surveyed vegetation at all sites between August 14 to 25, 2022, using established methods from White et al. (2012). In brief, this included assessing 26 points at 1 m intervals across six north-south transects separated by 5 m within each replicate site (6 x 26 = 156 points per site). At each point, we used the plant growth forms “graminoid”, “herb”, “shrub”, and “tree” to record the species richness and proportion of growth forms from ground to canopy, with differentiation at the species level whenever possible. The presence of bare ground (i.e., no vegetation cover including grasses or herbs), litter, and canopy cover were also recorded at each point, due to their potential influence on

soil abiotic factors and aerobiome characteristics. The number of points with each of bare ground, litter, and canopy cover were divided by the total points (n = 156) to obtain their proportion.

Rainfall

Rainfall data (in mm) from the recording station in closest proximity to each site was downloaded from the Bureau of Meteorology website (Bureau of Meteorology, 2022). The Bureau of Meteorology utilises > 45 stations across greater metropolitan Adelaide to record daily rainfall. The total rainfall for the seven days prior to the sampling day was aggregated for analysis

Air sampling

At each urban greenspace site, air samples were collected over an approximately 8-hour period during the days of May 21-24 and June 11-14, 2022, following the method described in Mhuireach et al. (2016). We sampled seven sports fields and 14 parks in May, and we sampled four sports fields and eight parks in June. The aerobiome sampling stations were set up on site between 07:00 and 09:00 hours and collected between 15:00 and 17:00.

Each urban greenspace had a single sampling station, a plastic box at a height of 0.3 m from the ground. We opened and placed three sterile, clear plastic Petri dish bases and lids on each station, providing six collection surfaces per site (**Figure 2**). This method of passive aerobiome sampling has been shown to be as effective as active sampling methods (Mhuireach et al., 2016). A field control for each day was generated by holding an additional Petri dish open for 30 seconds at the equipment box. Immediately after the field sampling activity, each Petri dish was sealed, labelled, transported on ice, and then frozen at -20°C until DNA extraction (described below).

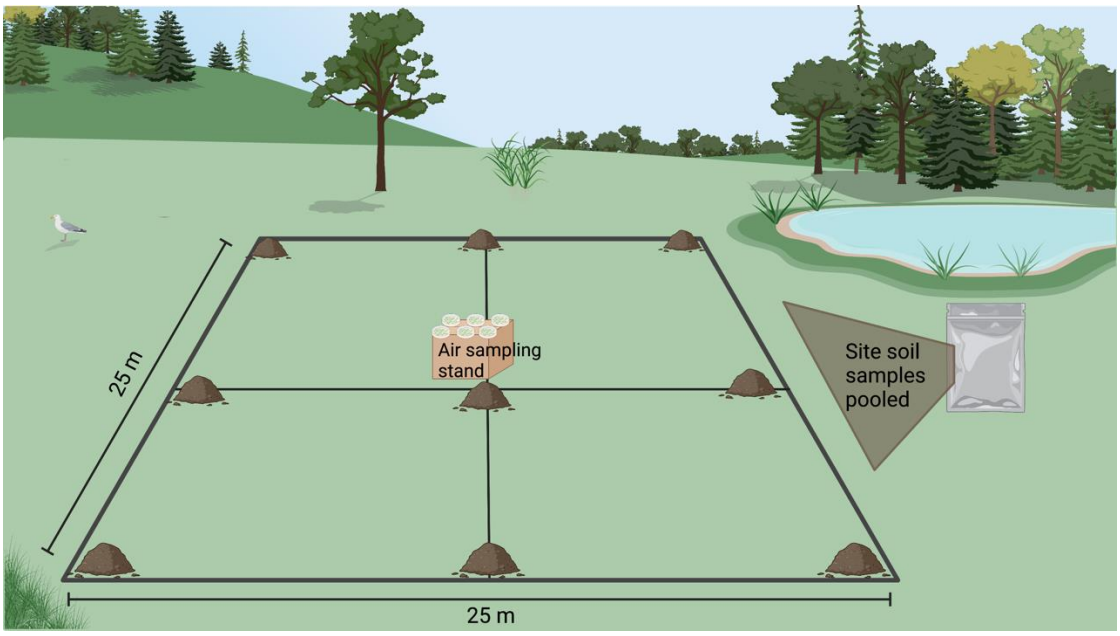


Fig. 2 Diagram of sampling area for air and soil sample collection. Soil mounds indicate sample collection points, with all soil samples at one site pooled into a single bag. The air sampling stand with six Petri dish surfaces was placed at the central point. Figure created with Biorender.

Soil sampling

Each 25 x 25 m site area was formatted into nine grid points (**Figure 2**). We employed Australian Microbiome Initiative (Bissett et al., 2016) sampling protocols. Briefly, this involved using a trowel decontaminated with ethanol and 5% Decon 90 (Decon Laboratories Ltd, Pennsylvania, USA) to collect approximately 50 g of soil from 0-5 cm soil depth at each grid point. The soil samples were then pooled into a sterile plastic bag, homogenised, and the bag sealed. Upon completion of all sampling sites, a minimum of 180 g subsample of each homogenised composite soil sample was placed into new bags and sent to CSBP Soil and Plant Analysis Laboratory (Bibra Lake, Western Australia) for analysis of 23 physical and chemical parameters, including pH, organic carbon, nitrate nitrogen, and cation concentrations (see **Table S1** for all parameter data). Soil moisture (%) was calculated in-house using an oven-drying process as follows: 40 g from each soil sample was transferred to an unsealed metal container, weighed, and placed in an oven at 105°C for 24 hours. Containers with the oven-dried soils were then re-weighed, and the weight lost (= weight of water) as a percentage of total dry mass was calculated.

eDNA extraction, PCR, sequencing

eDNA extractions and quantifications were performed in the dedicated Biological Sciences laboratory at Flinders University, South Australia. The Petri dishes for each site were opened and swabbed with sterile nylon-flocked swab tips (FLOQSwabs Lot 2011490, Copan Flock Technologies, Bescia, Italy) inside a laminar flow cabinet (Aura PCR PC10000, EuroClone, Milan, Italy). One swab and 40 uL of added sterile phosphate-buffered saline was used for swabbing all six surfaces, except for surfaces that showed visual signs of damage or environmental contamination, for approximately four minutes total using a consistent pattern of swabbing. The tips were cut directly into 2 mL sterilised tubes (QIAGEN, Hilden, Germany). During laboratory preparation for DNA extraction, an extraction blank control was creating using sterile water and processed using the same DNA extraction process as the field site samples.

For air sample DNA extractions, we used the QIAamp DNA Mini Kit (QIAGEN, Hilden, Germany) and followed the manufacturer's instructions. The extraction concentrations were then quantified using the Quantus fluorometer (Promega, Madison, WI, USA). Samples were sent to South Australia Genomics Centre (SAGC; Adelaide, South Australia) for library preparation via their 16S Metagenomic Sequencing Library Preparation protocol. DNA libraries were QC'd by Tapestation 2200 for size and Qubit for quantity. Equimolar pools were prepared and denatured with a final concentration of 4nM. Pooled libraries were diluted to 8pM (including 10%PhiX) and used for cluster generation. Denaturing and on-board clustering was performed using SAGC's Illumina protocol v05. Amplification of the bacterial 16S rRNA V3-V4 regions was performed using 314F-806R primers for library preparation that included 25+8 cycles of amplification, and sequencing of amplicon sequence variants was completed on the Illumina Miseq v3 at South Australia Genomics Centre.

Bioinformatics

From the aerobiome 16s rRNA raw sequence data, amplicon sequence variants (ASVs) were trimmed and filtered using an established Qiime2 pipeline (version 2022.2; Bolyen et al., 2019).

Seven samples had insufficient read counts and were removed from subsequent aerobiome analyses. For the remaining 26 samples, taxonomy was assigned using the onboard Naïve Bayes taxonomic classifier and Greengenes 13-8 database. Sequences were then cleaned using custom code on the R phyloseq package (version 1.42.0; McMurdie & Holmes, 2013) by removing the following: sequences from mitochondria and chloroplasts, taxa that did not occur in at least two samples to reduce artifacts, and ASVs with sums <30. Sequences that were likely of contamination origin were identified and removed using the R decontam package (version 1.18.0; Davis et al., 2018) using the function "isNotContaminant" suited for use with low biomass samples.

Statistical analysis

All statistical analyses were performed using R (version 4.2.3; R Core Team, 2023). To maintain consistency with prior aerobiome studies, statistical significance was set at alpha = 0.05. For each sample we generated statistical data on alpha diversity using Shannon's Diversity index and Faith's phylogenetic diversity, beta diversity ordinations using both Aitchison distances and weighted Unifrac, and distance-to-centroid. To prepare phylogenetic trees, sequences were rarefied at the sequencing depth of 12,054, multiple sequence alignment was performed using the program MAFFT (Katoh and Standley, 2013) with post-cleaning using GBlocks (Talavera and Castresana, 2007), and phylogenetic trees were created using IQTree2 (Nguyen et al., 2015) using the "Generalized time-reversible with Gamma rate variation" parameter. Two outlying low values of alpha diversity among sports fields (Kingswood Oval: Shannon index = 2.82, and Daly Oval: Shannon index = 4.64) were omitted from alpha diversity analyses because these outliers were considered unrepresentative at more than two standard deviations from the sports fields group mean. Maps were created using the R ggmap package (version 3.0.2; Kahle and Wickham, 2013).

To prepare for compositional beta diversity tests, the sequence abundance data was evaluated using R *zCompositions* package (version 1.4.0.1; Palarea-Albaladejo and Martín-Fernández, 2015) and zeros were imputed using R *scImpute* package (version 0.0.9; Li and Li, 2018). The resultant

abundances were then transformed with centred-log ratio using the R *compositions* package (version 2.0.6; Van den Boogaart and Tolosana-Delgado, 2008), followed by ordination with principal components analysis based on Aitchison distances obtained with the R *vegan* package (version 2.6.4; Oksanen et al., 2022). In addition, weighted Unifrac distances were calculated with R *vegan* package using the phylogenetic trees described above. Distance-to-centroid analyses were performed using the *betadisper* function in the R *vegan* package. Analysis of Compositions of Microbiomes with Bias Correction using the *ancombc2* function in the R *ANCOMBC* package (version 2.0.3; Lin and Peddada, 2020) was performed on untransformed amplicon data for differential abundance analyses. The ANCOMBC algorithm has been shown to minimise bias due to sampling fractions and reduce false discovery rates. Canonical correspondence analyses were performed using the R *CCA* package (version 1.2.1; González et al., 2008). The step-based variable selection model was performed using the *ordistep* function in the R *vegan* package. R *ggplot2* package (version 3.4.2; Wickham, 2016) was used for data visualisations.

RESULTS

We obtained 17.8 million raw reads with 73.6% >Q30 from 33 air samples. After quality control, data from 26 samples with 6,205 bacterial amplicon sequence variants across 33 phyla were utilised in subsequent analyses.

Aerobiome alpha diversity

Land cover had a weak but non-significant effect on aerobiome alpha diversity, with sports fields ($\bar{x}$ = 6.26 ± 0.26) higher than nature parks ($\bar{x}$ = 5.24 ± 1.27) (Welch Two Sample t-test: t = -2.998, p = 0.065; **Fig. 3a, Fig. S1a**). Volume of rainfall in the seven days prior to sampling had a negative effect on the aerobiome alpha diversity (F = 5.265, R^2 = 0.15, df = 1 and 23, p = 0.03; **Fig. 3b, Fig. S1b**). Soil iron negatively influenced the alpha diversity in nature parks (F = 22.38, R^2 = 0.60, df = 1 and 13, p < 0.001; **Fig. 3c, Fig. S1c**), and nature parks had a substantially lower mean concentration of soil iron ($\bar{x}$ = 60.4 mg/kg ± 39.8) than sports fields ($\bar{x}$ = 116 mg/kg ± 104). Soil pH had a positive

effect on the alpha diversity of both sports fields (F = 6.155, R^2 = 0.42, df = 1 and 6, p = 0.048) and nature parks (F = 6.156, R^2 = 0.27, df = 1 and 13, p = 0.03; **Fig. 3d, Fig. S1d**) (all correlations between soil physicochemical parameters and alpha diversity indices are shown in **Fig. S2 and S3**). Among parks, the diversity of woody plant species had no effect on the aerobiome alpha diversity (F = 0.88, R^2 = - 0.01, df = 1 and 13, p = 0.37; **Fig. S4a-b**).

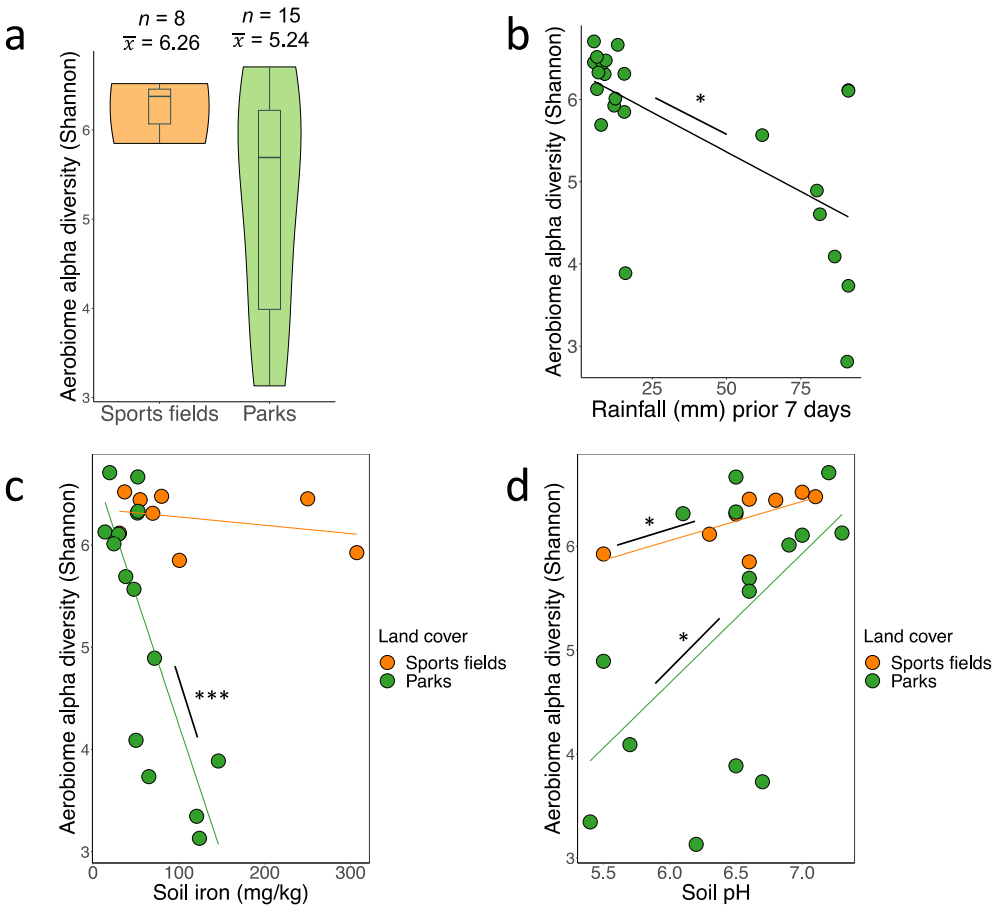


Fig. 3 (a) Boxplots of aerobiome alpha diversity by land cover. The y-axis shows the aerobiome alpha diversity calculated by the Shannon diversity index. Boxes show the median and interquartile range, while whiskers extend to the remaining range of data. (b) Relationship of aerobiome bacterial alpha diversity and total rainfall during seven days prior to sampling. (c) Relationship of soil iron with aerobiome alpha diversity. Orange and green lines separately show regressions for sports fields and nature parks, respectively. (d) Relationship of soil pH with aerobiome alpha diversity. Orange and green lines show regressions separately for sports fields and nature parks, respectively.

Aerobiome community composition

Land cover type had a significant influence on the aerobiome community composition. Sports fields had a distinct aerobiome compared to nature parks (PERMANOVA Adonis test: 999 permutations, df = 1, F = 3.01, R^2 = 0.12 p = 0.001; **Fig. 4a, Fig. S5a**). Sports fields (distance to centroid = 122)

and nature parks (distance to centroid = 118.5) had similar levels of homogeneity of aerobiome compositions (ANOVA: F = 0.28, df = 1, p = 0.60). Rainfall in the week prior to sampling significantly influenced aerobiome community composition across both land cover types (PERMANOVA Adonis test: 999 permutations, df = 1, F = 1.84, R^2 = 0.07, p = 0.016; **Fig. 4b, S5b**). Among nature parks, where the woody plant species diversity (Shannon's Diversity index) ranged from 0.60 to 1.90, woody plant species diversity had no effect on aerobiome composition (PERMANOVA Adonis test: 999 permutations, df = 1, F = 0.80, R^2 = 0.05, p = 0.82; **Fig. S6a-b**).

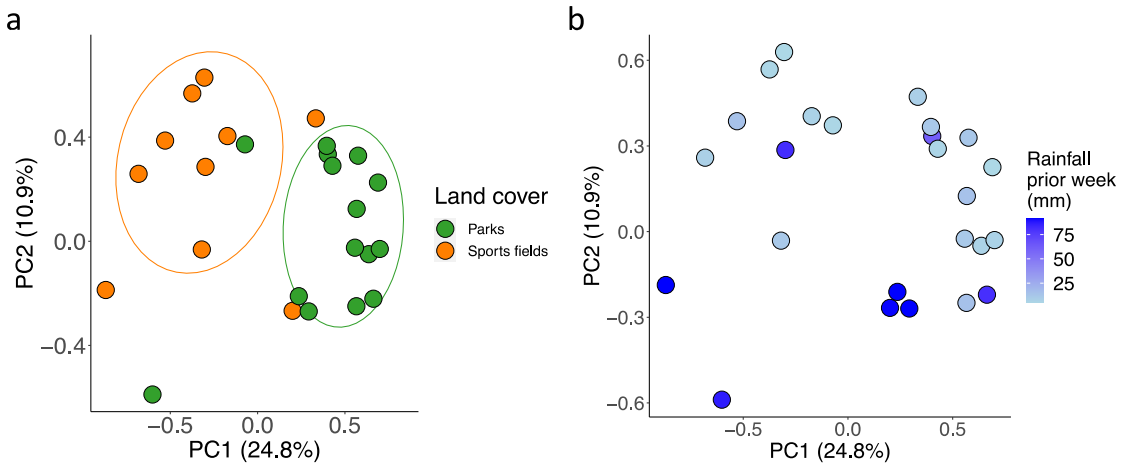


Fig. 4 (a) Principal components analysis based on Aitchison distances displaying significant variation in aerobiome community composition between land cover types (PERMANOVA Adonis test: 999 permutations, df = 1, F = 3.01, R^2 = 0.12 p = 0.001, n = 25 sites). (b) Principal components analysis based on Aitchison distances displaying variation in aerobiome community composition by total rainfall volume in the week prior to sampling (PERMANOVA Adonis test: 999 permutations, df = 1, F = 1.84, R^2 = 0.07, p = 0.016, n = 25 sites).

The ten phyla with the highest abundances from both land covers showed similar proportions between sports fields and nature parks (**Fig. 5a-b**). Proteobacteria was the dominant phylum in samples from both land covers. Land cover had a strong effect on Desulfobacterota abundances (ANCOM-BC log linear model: log fold change from parks = 2.51, W = 3.89, adjusted p = 0.002) and a weak effect on Nitrospirota abundances (ANCOM-BC log linear model: log fold change from parks = 2.06, W = 2.85, adjusted p = 0.052; **Fig. 5c)**, both of which were higher in sports fields aerobiomes. While these phyla were differentially abundant, they were not among the ten phyla with the highest abundances. Among the 353 genera in the dataset, only *Thermomonas* (ANCOM-BC log

linear model: log fold change from parks = 2.85, W = 3.74, adjusted p = 0.03) and an uncultured genus (ANCOM-BC log linear model: log fold change from parks = 3.18, W = 3.93, adjusted p = 0.03) had significantly greater abundances in sports fields aerobiomes versus nature parks (all genus differential abundances listed **Table S2**).

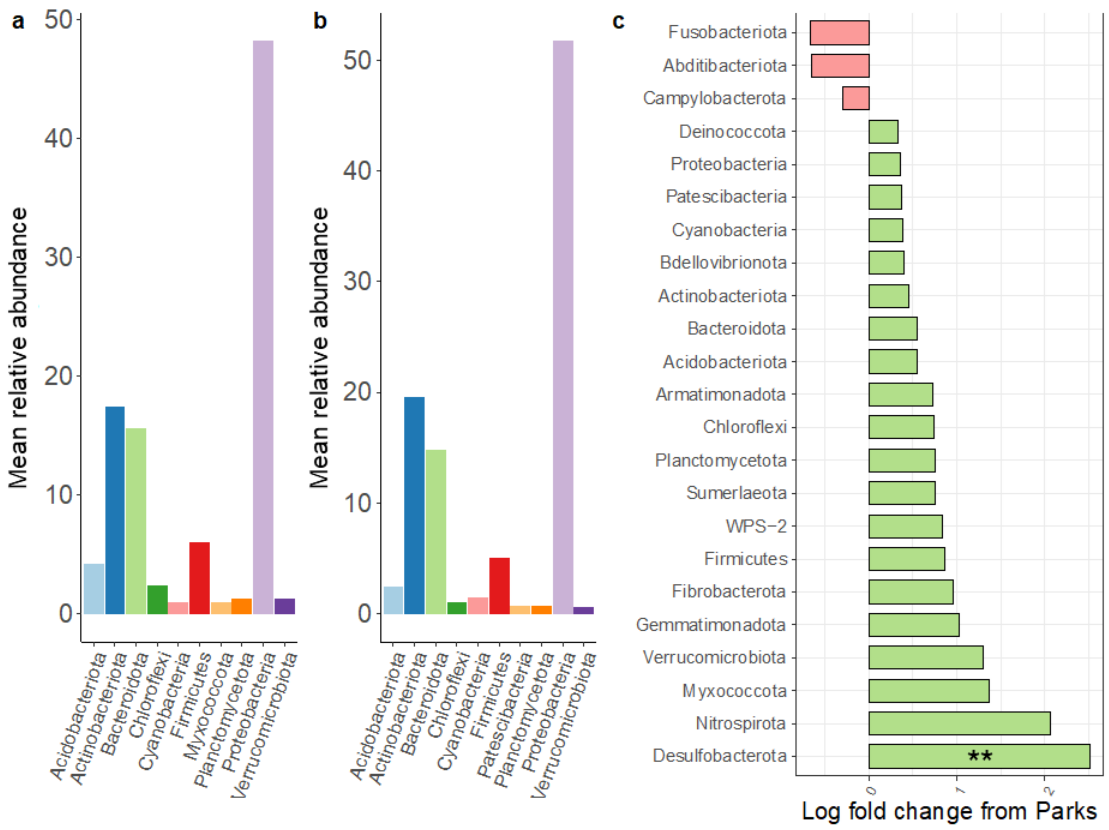


Fig. 5 (a) Bar plot showing the phyla with the ten highest mean relative abundances among air samples at sports field sites. Y axis shows the means of the ASV relative abundances from each phylum. (b) Bar plot showing the phyla with the ten highest mean relative abundances among air samples at all park sites. Y axis shows the means of the ASV relative abundances from each phylum. (c) Differential abundance of phyla in aerobiomes by land cover type. The x axis shows the log fold change from parks to sports fields. Red bars indicate a decrease in log fold change, and green bars indicate an increase in log fold change. Significance indicator: ** <0.01.

Influences of soil physicochemical parameters aerobiome composition

With data from both land cover types merged, soil physicochemical variables were overall weak predictors of aerobiome composition, and both canonical correspondence analysis (CCA) axes described < 7% (**Fig. 6a**). However, repeating the CCA only in nature parks (n = 15) showed that soil iron had a strong effect on the aerobiome composition (ANOVA: F = 1.308, df = 1, p = 0.04;

Fig. 6b). Among sports fields (*n* = 10), soil physicochemical variables showed no significant effect on aerobiome composition among sports fields (**Fig. 6c**).

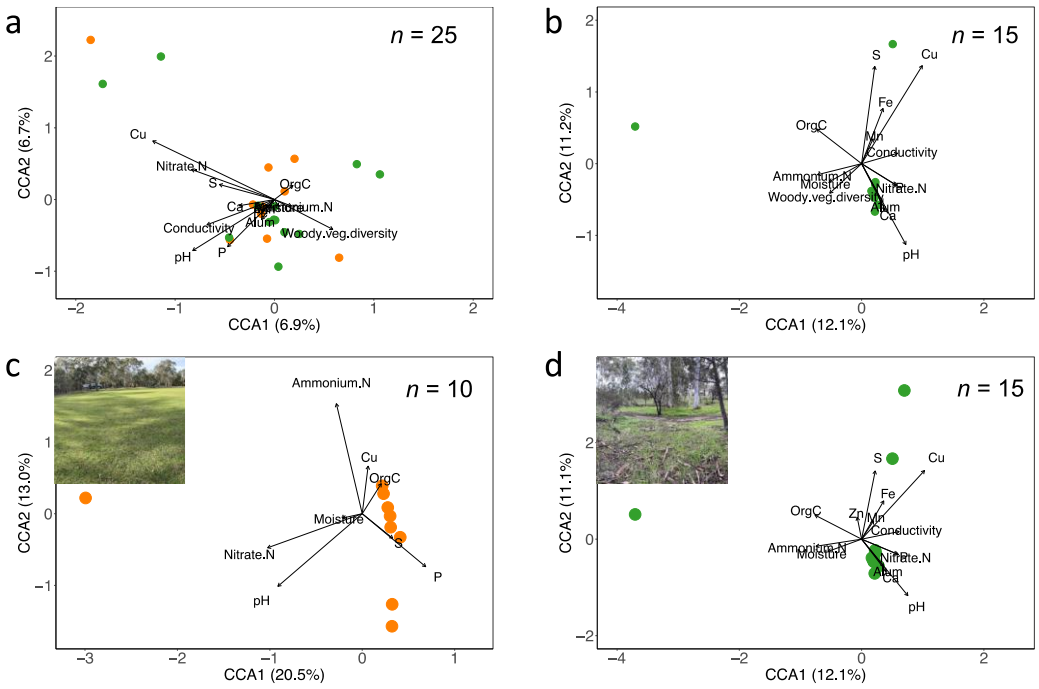


Fig. 6 (a) Canonical correspondence analysis (CCA) of aerobiome community composition constrained by soil physicochemical variables across all sites. (b) CCA of aerobiome community composition of nature parks constrained by soil physicochemical variables. (c) CCA of aerobiome community composition of sports fields constrained by soil physicochemical variables. In (a-c), arrow lengths indicate strength of constraint by each explanatory variable, and arrowheads show the direction of increase. Orange points indicate sports fields and green points indicate nature parks; *n* is the number of sites.

DISCUSSION

Urban greenspaces provide city residents with opportunities to gain access to nature and exposure to potentially health-supporting biodiverse airborne bacterial communities (aerobiomes). Therefore, describing the characteristics of aerobiomes in urban greenspaces, and identifying the ecological factors that influence them, can support greenspace management and future planning. Here, we found that near-surface aerobiomes of sports fields had distinctly different bacterial communities than nature parks. However, levels of bacterial diversity were influenced more by soil physicochemical characteristics (primarily soil pH and iron) and rainfall than by the land cover type and the woody vegetation diversity. The vegetation and soils of urban greenspaces, especially sports

fields, appear to be managed in such a way that results in consistent characteristics of the airborne microbial communities. Because such characteristics are important for urban greenspace development, we examined the influence of four key ecological variables on urban greenspace aerobiomes, each of which is discussed below.

Effects of soil physicochemical parameters on aerobiomes

We show that soil physicochemical parameters had a greater influence on the alpha diversity of aerobiomes than on the community composition. Though iron played the strongest role among the parameters in shaping the aerobiome community, its effects were moderate and were limited to nature parks. On the other hand, soil pH and iron concentrations had strong effects on the aerobiome alpha diversity. Previous research has demonstrated that abiotic features (e.g., pH) are important drivers in shaping microbial community composition in soils, sometimes taking precedence over plant-based factors (Fierer, 2017). The importance of soil abiotic factors in shaping bacterial alpha diversity is recognised in soils, but to our knowledge has not previously been translated to aerobiomes. Although here we did not analyse soil microbiomes, the bacteria in soils are well-known contributors to aerobiomes (Brodie et al., 2007; Robinson and Breed, 2023). Our results suggest that maintaining soil pH close to neutral (e.g., via liming) and lowering (or otherwise counteracting) elevated soil iron levels could increase the aerobiome alpha diversity. Notably, pH associated with aerobiome alpha diversity regardless of the land cover. Thus, human exposure to salutogenic microbial diversity could potentially be modulated through landscape maintenance strategies that focus on soil pH across both sports fields and nature parks.

It is important to note that manually altering soil pH could affect wider ecosystem functions, For instance, landscape maintenance efforts to substantially shift soil pH could affect the growth and distribution of plant species in the ecosystem (Seaton et al., 2023) and alter soil carbon cycling (Malik et al., 2018) potentially impacting the entire food web. Furthermore, exposure to greater alpha diversity might not invariably produce salutogenic effects; the community composition and

their associated functional traits may also play a role. For instance, species richness (and evenness) might be 100-fold higher in one sample compared to another but, in theory, could contain 100-fold more pathogens with high relative abundances. In such a scenario, alpha diversity will be considerably higher, but so will the sample's pathogenic potential. Nonetheless, altering soil properties to enhance the aerobiome in highly-managed anthropogenic environments, such as sports fields, could contribute to human immunoregulation, whilst minimising adverse ecological impacts.

Effects of land cover type on aerobiomes

We found that greenspace land cover weakly influenced aerobiome alpha diversity, with intensely-managed grassy sports fields having higher alpha diversity than less-managed nature parks. While related aerobiome studies are limited, our findings are consistent with Delgado-Baquerizo et al. (2018) who observed higher soil bacterial alpha diversity in urban greenspaces with more intense land management than in adjacent, more natural sites. It may be possible that sports field management techniques, such as regular fertilisation, introduce soil conditions that select for a wider range of bacterial inhabitants. Because soil bacterial communities are shaped by a complex network of biotic and abiotic factors, future research can examine the unique properties of sports field management that may influence the soil and airborne bacterial communities.

Furthermore, we found that the aerobiome community composition of sports fields was distinct from nature parks. Differential abundance analyses showed that the top ten phyla had slightly higher abundances in sports fields, which may have contributed to the composition distinctions found in PERMANOVA tests, although these differences were non-significant for nine of the top ten phyla. The only significant differential abundance was observed in the phylum Desulfobacterota, which includes taxonomic classes with a preference for anaerobic environments (Murphy et al., 2021) and the capacity to use sulphur and iron as terminal electron acceptors. Sports field soils had almost double the concentration of both sulphur and iron than nature parks. As such, sports fields soil

conditions may be selecting for bacteria such as Desulfobacterota with the capacity to utilise sulphur and iron in metabolic processes.

Effects of woody plant species diversity on aerobiomes

Previous studies have shown that wooded urban habitats with more complex vegetation had higher aerobiome alpha diversity than grassy habitats (Mhuireach et al., 2021; Robinson et al., 2021). Leaf surfaces can offer key sources of airborne bacteria (Bowers et al., 2011b), and greater vegetation complexity can provide additional leaf surface. Among our nature park sites, however, the diversity of woody plant species did not influence the aerobiome alpha diversity. The seasonal conditions prior to sampling (May-June 2022) with late autumn to winter rainfall may have been a factor in 'washing off' and reducing the contribution of phyllosphere microbiota to aerobiomes.

In addition, our finding that soil abiotic characteristics had a stronger effect than nearby vegetation could have been influenced by our sampling height. While we sampled at 0.3 m height, Mhuireach et al. (2021), who found that vegetation affected aerobiomes, sampled at 2 m height. Aerobiomes, including pathogenic taxa, are known to stratify vertically, with greater alpha diversity closer to the ground (Robinson et al., 2021; Robinson et al., 2020a). It is, therefore, possible that vegetation may have a greater effect on aerobiomes above 0.3 m. Future studies should evaluate how ecological features, including detailed soil and vegetation properties, impact aerobiomes at varying heights.

Effects of rainfall on aerobiomes

We observed that increased rainfall in the week prior to sampling associated with a reduction in aerobiome alpha diversity. Rainfall is known to reduce the volume of bioaerosols (Després et al., 2012; Li et al., 2017). Whether bioaerosol volume reduction links with a decrease in bacterial diversity likely depends on factors such as the typical aerosol particle size present to which bacteria can attach, which can influence which specific taxa may be removed by rainfall (Jang et al., 2018). Simultaneously, rainfall can disperse surface soil bacteria into the air through the impact of raindrops

on soils (Joung et al., 2017). When previously airborne bacteria are removed by rain, and soil-borne bacteria are dispersed into the air, the aerobiome can experience a shift in composition (Jang et al., 2018). Indeed, we observed that increased rainfall associated with a difference in aerobiome composition. Future studies should employ repeated sampling methods at urban greenspaces to gain further insight into specific compositional changes associated with rainfall, and whether those changes could be beneficial or harmful to human health.

Limitations

Our study examined the bacteria only among air samples. Because soils make key contributions to adjacent aerobiomes, future studies that directly compare aerobiome and soil microbiome compositions could provide further insights into quantity and mechanistic theory of bacterial transfer, and subsequent human exposure, in urban greenspaces. Although we focused on bacteria, future assessment of fungi among urban greenspace aerobiome datasets could also give further insights into the contributions of greenspace exposure to human health and disease. In addition, wind can carry distant airshed bacteria and have a substantial influence on the aerobiome; further studies could examine the effects of wind on urban greenspace aerobiomes to which humans may be exposed.

Conclusions

We performed a novel analysis of airborne microbial communities across two types of urban greenspaces: sports fields and nature parks. We show that the near-surface aerobiomes at 0.3 m sampling height were influenced by land cover type, recent rainfall, and soil physicochemical characteristics (i.e., soil pH and iron) but not by woody plant species diversity. We also show that soil pH strongly affects the bacterial diversity in the air regardless of greenspace land cover type. Greenspaces are an integral part of urban design, providing opportunities for city residents to gain greater exposure to natural biodiversity. Urban greenspace planners and restoration ecologists rely on data to manage ecological variables such as woody plant diversity and soil physicochemical

parameters. Our findings should further assist the development of greenspace design initiatives that aim to harness the health-promoting effects of biodiverse microbial exposures.

Chapter 4. Urban sports fields support higher levels of soil butyrate and butyrate-producing bacteria than urban nature parks

This manuscript is in the final stages of co-author review prior to being submitted for publication.

Context

While Chapter 2 described the global and regional ecological characteristics that associate with butyrate producer abundances, Chapter 4 examines such ecological associations in a metropolitan-scale environmental study. Aerobiomes in urban greenspaces are important reservoirs of butyrate-producing bacteria as they supplement the human microbiome, but soil butyrate producer communities have rarely been examined in detail. Here, we studied soil metagenome taxonomic and functional profiles and soil physicochemical data from two urban greenspace types: sports fields (n = 11) and nature parks (n = 22). We also developed a novel method to quantify soil butyrate and used this to characterise the activity of butyrate-producing bacteria in greenspace soils.

In Chapter 4, I aimed to answer the following research question: What are the effects of greenspace type and soil physicochemical parameters on butyrate-producing bacterial read abundances, butyrate production terminal genes, and soil butyrate concentrations?

Author contributions

All authors contributed to the study conception and design. Soil short-chain fatty acid extraction method was developed by C.A. and J.B.. J.B. wrote the first manuscript draft, and all authors commented on further versions of the manuscript. All authors read and approved the final manuscript.

Article

Urban sports fields support higher levels of soil butyrate and butyrate-producing bacteria than urban nature parks

Joel E. Brame[a,b], Craig Liddicoat[a,b], Catherine A. Abbott[a], Christian Cando-Dumancela[a,b], Jake M. Robinson[a,b], Martin F. Breed[a,b]

[a]College of Science and Engineering, Flinders University, Bedford Park, SA 5042, Australia

[b]The Aerobiome Innovation and Research Hub (The AIR Hub), College of Science and Engineering, Flinders University, Bedford Park, SA 5042, Australia

Corresponding author: Joel E. Brame: joel.brame@flinders.edu.au

Competing interests

The authors have no competing interests to declare that are relevant to the content of this article. This work was supported by funding from the Flinders Foundation and the New Zealand Ministry of Business Innovation and Employment (grant UOWX2101).

ABSTRACT

Butyrate-producing bacteria colonise the gut of humans and non-human animals, where they produce butyrate, a short-chain fatty acid with known health benefits. Butyrate-producing bacteria also reside in soils, and soil bacteria can drive the assembly of airborne bacterial communities (the aerobiome). Aerobiomes in urban greenspaces are important reservoirs of butyrate-producing bacteria as they supplement the human microbiome, but soil butyrate producer communities have rarely been

examined in detail. Here, we studied soil metagenome taxonomic and functional profiles and soil physicochemical data from two urban greenspace types: sports fields (n = 11) and nature parks (n = 22). We also developed a novel method to quantify soil butyrate and characterised the activity of butyrate-producing bacteria in these greenspace soils. We show that soil butyrate was higher in sports fields than nature parks, and that sports fields also had significantly higher relative abundances of the terminal butyrate production genes *buk* and *butCoAT* than nature parks. Soil butyrate positively correlated with *buk* gene abundance (but not *butCoAT*). Soil moisture (r = 0.50), calcium (r = -0.62), iron (rho = 0.54), ammonium nitrogen (rho = 0.58), and organic carbon (r = 0.45) had the strongest soil abiotic effects on soil butyrate concentrations, and iron (rho = 0.56) and calcium (rho = -0.57) had the strongest soil abiotic effects on *buk* read abundances. Overall, our findings contribute important new insights into the role of sports fields as key exposure reservoirs of butyrate and butyrate producing bacteria, with important implications for the provision of microbiome-mediated human health benefits via butyrate.

INTRODUCTION

The environment has a role in shaping human commensal bacterial communities (Chong, D'amato, & Rosário Filho, 2022; Nielsen et al., 2020). When humans spend time outdoors, environmental microbiota such as air- and soil-borne bacteria can transfer to the body and influence the human microbiome (Roslund et al., 2022; Sessitsch et al., 2023). While much health-related research has examined this dynamic with a focus on mitigating exposure to pathogenic environmental bacteria, outdoor environments can also house bacterial communities with potential to support human health, including diversity (Spragge et al., 2023) and the presence of specific taxonomic groups (e.g., probiotic and butyrate-producing bacteria; Brame et al., 2022; Roslund et al., 2020; Zhang et al., 2023). Thus, environmental bacteria – particularly soil bacteria – are key sources of the human microbiome, with immunoregulation and downstream health implications (Roslund et al., 2022; Roslund et al., 2020).

Humans are exposed to soil bacteria via direct contact with soil (Selway et al., 2020) and ingestion of plants with soil residues (Flandroy et al., 2018). In addition, airborne bacterial communities (aerobiomes) can be inhaled into the respiratory tract and oral surfaces (Flies et al., 2020). These exposures can alter the skin (Grönroos et al., 2019), nose (Selway et al., 2020), and potentially gut bacterial communities (Nurminen et al., 2018). Growing evidence shows that exposure to bacterial diversity helps train the adaptive immune system and may regulate the innate immune system, often described as the Biodiversity Hypothesis (Haahtela, 2019). As such, outdoor air and soil are important reservoirs of bacteria to which humans may be exposed via outdoor activities. Therefore, understanding the microbial compositional variation and its ecological drivers in outdoor environments, particularly urban greenspaces, is a critical step in managing environmental microbiome exposures.

The ecology of urban greenspaces influences the composition of the microbial communities in these areas. Urban vegetation influences the diversity and abundance of soil bacteria and fungi (Baruch et al., 2020). Furthermore, urban plant-soil systems provide key inputs into aerobiomes (Robinson et al., 2020a), as surface bacteria disperse into the air (Bowers et al., 2011a). Highly vegetated urban parks have distinct aerobiome compositions from non-vegetated and nearby parking lots (Mhuireach et al., 2016). Tree density, proximity, and canopy coverage modulate urban aerobiome alpha diversity (Robinson et al., 2021). Urban forest aerobiomes are also compositionally different from grassland aerobiomes (Mhuireach et al., 2021). Thus, soils and vegetation influence the greenspace microbial communities to which humans are exposed. However, rapid urbanisation creates an urgent need to better understand the ecological influences on urban greenspace microbial communities, particularly for specific health-associated microbiota such as butyrate-producing bacteria.

Butyrate-producing bacteria are key human gut taxa with important health implications. Butyrate is produced in anaerobic conditions through a fermentative enzymatic pathway that requires iron and has multiple, sometimes reversible, steps (Vital et al., 2014; Figure S1). Butyrate provides numerous

health benefits for humans, including metabolic energy for gut epithelial cells (Rivière et al., 2016), maintenance of gut homeostasis (Parada Venegas et al., 2019), and inhibition of the enzyme histone deacetylase, resulting in epigenetic modifications with anti-inflammatory and immunoregulatory outcomes such as an increase in regulatory T-cells (Pandiyan et al., 2019; Sivaprakasam et al., 2017). However, a range of human health conditions are associated with a reduction in gut butyrate-producing bacteria including asthma (Demirci et al., 2019), atopic dermatitis (Lee et al., 2022), and inflammatory bowel disease (Parada Venegas et al., 2019). Given that the outdoor environment influences the human bacterial communities (Chong, D'amato, & Rosário Filho, 2022; Nielsen et al., 2020) urban greenspaces are compelling potential sources of commensal bacteria that could supplement the human microbiome. Thus, there is an unfilled need to further examine how the ecology of urban greenspaces – e.g., soil biotic and abiotic factors, vegetation – influences the abundance and exposure to butyrate-producing bacteria (Brame et al., 2021).

Here, we investigated the effects of greenspace type and soil abiotic factors on soil butyrate levels and butyrate-producing bacteria read abundances in greater metropolitan Adelaide, South Australia. We did this by developing a new method to quantify soil butyrate levels and combined this dataset with soil shotgun metagenomic and comprehensive soil physicochemical data. We generated these data from two greenspace types: intensively-managed grassy sports fields (n = 11 sites) and minimally-managed amenity grassland parks with more natural vegetation systems (n = 22 sites). We asked the following research question: what effects do greenspace type and soil abiotic conditions have on soil butyrate concentrations, butyrate-producing bacteria read abundances, and functional abundances of genes for terminal butyrate-production enzymes? Our study provides a new view on greenspaces by focussing on their soil butyrate and butyrate-producing bacterial levels.

METHODS

Study sites

We sampled spatially independent sports field (n = 11) and nature park (n = 22) (33 sites total) sites in greater metropolitan Adelaide, South Australia, using a 25 x 25 m sampling area that is considered appropriate for characterising vegetation and microbial communities at the site-scale (Baruch et al., 2020; Mills et al., 2020; **Figure 1A-C**). Sites were chosen so that: (1) all sites were > 5 km from the coast to avoid coastal effects; (2) all sites were within the low-elevation metropolitan Adelaide plains to minimise climatic variation across sites; and (3) nature park sites represented a range of woody vegetation complexity.

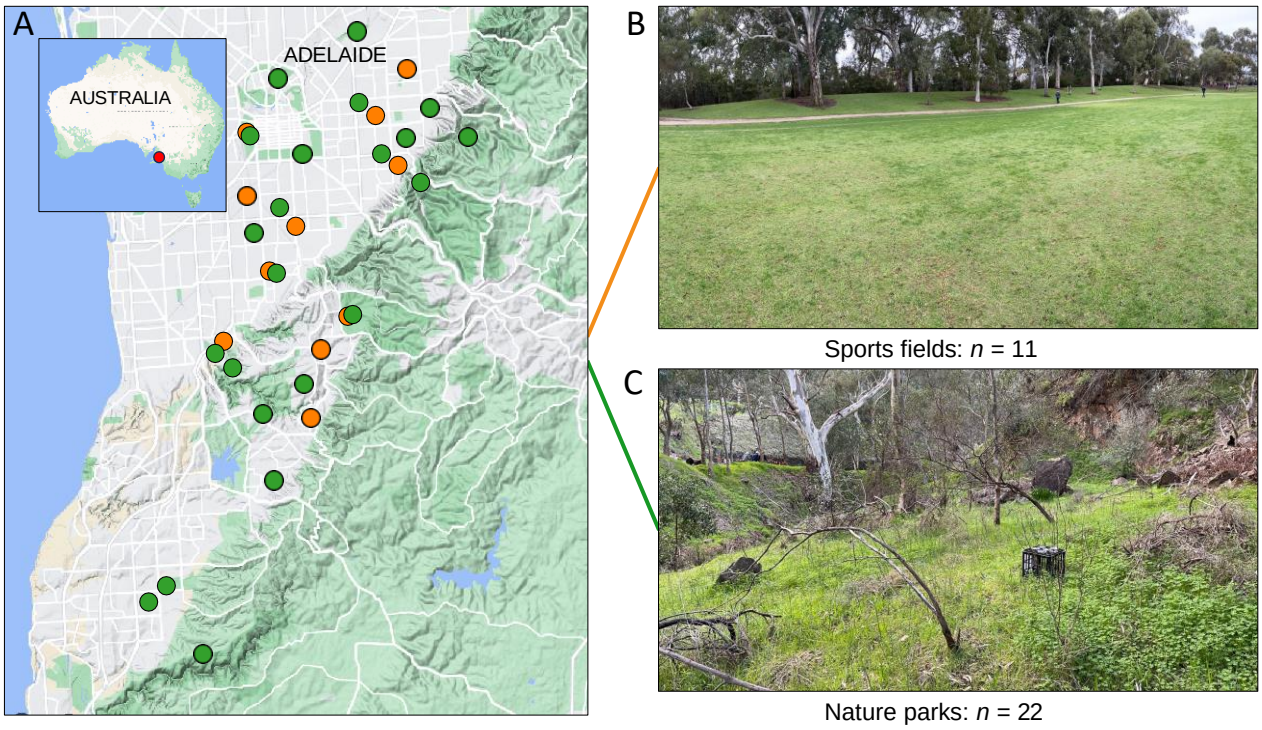


Fig. 1 (A) Map showing sample sites across greater metropolitan Adelaide, Australia (orange dots = sports fields, and blue dots = nature parks). (B) Representative photo of a sports field site. (C) Representative photo of a nature park site.

Greenspace type classification

We utilised woody species diversity to classify our sites into greenspace types as sports fields or nature parks. We surveyed vegetation at all sites between August 14 and 25, 2022, using established methods from White et al. (2012). In brief, this included assessing 26 points at 1 m intervals across six north-south transects separated by 5 m within each replicate site (6 x 26 = 156 points per site). At each point, we used the plant growth forms “graminoid”, “herb”, “shrub”, and “tree” to record the

species richness and proportion of growth forms from ground to canopy, with differentiation at the species level whenever possible. Greenspace type classification was performed by calculating Shannon's diversity index on the woody plant (i.e., trees and shrubs) species, where sports fields had Shannon = 0 and nature parks had Shannon > 0 (**Figure S2**), with greenspace types also showing distinct soil nutrient profiles (**Table S1**).

Soil sampling

Soil samples were collected between May 21 and 24 and June 11 and 14, 2022, from nine grid points at each 25 x 25 m site using an adapted Australian Microbiome Initiative sampling protocol (Bissett et al., 2016). A trowel decontaminated with ethanol and 5% Decon 90 (Decon Laboratories Ltd, Pennsylvania, USA) was used to collect approximately 50 g of soil from 0-5 cm depth at each grid point. Sterile nitrile gloves were worn during all sample collection steps to minimise contamination (Cando-Dumancela et al., 2023). The soil samples were then pooled and homogenised in a sterile plastic bag. A 50 mL subsample of soil for DNA analysis was placed into a separate 50 mL sterile falcon tube and immediately put onto ice. Upon completion of field sampling activity, a ca.180 g subsample of each homogenised composite soil sample was placed into new bags and sent to CSBP Soil and Plant Analysis Laboratory (Bibra Lake, Western Australia) for analysis of 19 physical and chemical parameters, including pH, organic carbon, nitrate nitrogen, and cation concentrations (see **Table S2** for all parameter data). Soil moisture (%) was calculated in-house using an oven-drying process as follows: 40 g from each soil sample was transferred to an unsealed metal container, weighed, and placed in an oven at 105°C for 24 hours. Containers with the oven-dried soils were then re-weighed, and the weight lost (= weight of water) as a percentage of total dry mass was calculated.

Soil short-chain fatty acid sampling and quantification

No method was available to quantify short-chain fatty acid concentrations in soils. Thus, we adapted a method from García-Villalba (2012) who examined short-chain fatty acids in human faeces. This

method used phosphoric acid for stabilising the short-chain fatty acids, which for our purposes was preferable to snap-freezing in liquid nitrogen due to long field days distant from the lab.

At each site, we collected soil from each grid point in the 25 x 25 m area using a decontaminated trowel (described above). We used a scale in the field to weigh 1.1 g from each of eight site grid points and 1.2 g from one grid point, to equal 10.0 g total soil weight per site. The depth of soil collection was 3-5 cm. To minimise the loss of volatiles, we collected short-chain fatty acid samples first and placed these samples directly into a 50 mL tube pre-filled with 16 mL of 0.5% phosphoric acid (Sigma Laboratory, Osterode am Harz, Germany). Each tube was then immediately placed on ice and stored in a -20°C laboratory freezer until short-chain fatty acid extraction.

For short-chain fatty acid extractions, sample tubes were removed from the freezer, thawed, vortexed for 20 seconds, and centrifuged using a Sigma 3-16KL centrifuge (Sigma Laboratory, Osterode am Harz, Germany) at 1,000 relative centrifugal force (RCF). Then, to move the short-chain fatty acids from the tube with phosphoric acid into the solvent ethyl acetate (100% hypergrade, Supelco Analytical, Bellefonte, PA, USA), 600 µL of supernatant was pipetted into a 2 mL polypropylene centrifuge tube with 600 µL of ethyl acetate. Each 2 mL tube was then vortexed for 20 seconds and centrifuged at 18,000 RCF on an Eppendorf Centrifuge 5425 (Eppendorf, Hamburg, Germany). Afterwards, 250 µL of the supernatant with the ethyl acetate was pipetted into 2 mL glass vials with 250 µL of the internal standard 4-methylvaleric acid (Sigma-Aldrich, St. Louis, MO, USA). Standards were created with butyric acid (Sigma-Aldrich, St. Louis, MO, USA), propionic acid (Sigma-Aldrich, St. Louis, MO, USA), and acetic acid (Sigma-Aldrich, St. Louis, MO, USA) diluted in ethyl acetate to 0.5 µM, 1 µM, 2 µM, 5 µM, and 10 µM concentrations each, and the standards were stored in a -20°C freezer until gas chromatography analysis at Flinders Analytical (Flinders University).

The gas chromatography-mass spectrometry (GC-MS) set up included a SGE BP20 PEG WAX bromoform column (30m x 0.25mm x 0.25 μm; Trajan, Ringwood, VIC, Australia), fitted with an Agilent 7683 automatic liquid sampler autoinjector (G4513A), in tandem with an Agilent 7890 mass spectrometer (Agilent Technologies, Palo Alto, CA, USA). Samples were injected using a pulsed splitless injection of 1uL. Column temperature was initially at 40°C, then increased to 250°C. Nitrogen was used as the carrier gas. A solvent delay was set at two minutes.

To create calibration curves, new standards were prepared on the day before GC-MS analysis and stored in a -20°C freezer until analysis (García-Villalba et al., 2012). Calibration curves for butyric acid and propionic acid were obtained (R^2 = 0.998 and 0.994, respectively). Acetic acid was also quantified but could not achieve reliable calibration curves using our methods. After the calibration curves were created, the samples were injected with a hexane rinse, and then again after every ten samples. Data acquisition was performed using MassHunter Quantitative Analysis Software (Agilent Technologies, Palo Alto, CA, USA). Selected ion mass technique was chosen to quantify standard compounds with optimal sensitivity. The peaks with optimal height and shape were shown at the following *m/z* values: butyric acid = 60 and propionic acid = 72. This ion for butyrate was chosen for subsequent analyses, as our study focused only on butyrate.

Soil DNA extraction, PCR, and sequencing

DNA extractions and quantifications were performed in a dedicated DNA extraction laboratory at Flinders University. For soil DNA extractions, we used the Qiagen Power Soil kit (QIAGEN, Hilden, Germany) and followed the manufacturer's instructions. The extraction concentrations were then quantified using the Quantus fluorometer (Promega, Madison, WI, USA) and samples were sent to the South Australia Genomics Centre (Adelaide, South Australia) for library preparation via the Nextera XT DNA library prep kit for Illumina (Part No. FC131-1024), protocol Version 05_05/19, and included 12 cycles of amplification. Libraries were all similar size and quantity. Equimolar pools were prepared and denatured with a final concentration of 3 pM and used for template

generation. This resulted in 1,334.56 million reads with 84.76% > Q30. Denaturing and on-board clustering was performed using the MGI protocol. 150bp paired end read sequencing was completed using the MGI DNBSEQ-G400 at South Australia Genomics Centre.

Bioinformatics

We performed quality control on the raw shotgun metagenomic sequence data of each sample using PRINSEQ++ (Cantu, Sadural, & Edwards, 2019). Adapter sequences were removed using Cutadapt (Martin, 2011). Short read datasets were taxonomically classified using Kraken2 (Wood et al., 2019). Read sums were then normalised by dividing by the expected genome lengths obtained from the NCBI (Nayfach and Pollard, 2015; Sayers et al., 2023) and then collated using Kraken2-output-manipulation. Next, relative abundance estimations were obtained using Bracken (Lu et al., 2017). The resulting data were filtered for butyrate-producing bacteria using a list of 118 putative butyrate-producing species derived from Vital et al. (2014) and NCBI using current classifications from Genome Taxonomy Database (see Supplementary Table 3 for full butyrate producer list; Parks et al., 2021). Functional gene profiles were obtained using SUPER-FOCUS (Silva et al., 2016), which reports the Seed subsystems (and corresponding functions) present in the datasets and profiles their abundances.

As our interest was on butyrate-producing bacteria, reads were filtered in R (version 4.2.3; R Core Team, 2023) for annotations to the subsystem “Acetyl-CoA fermentation to Butyrate", and additionally for "tRNA aminoacylation, Phe" for the gene *pheS*, a single-copy gene which was utilised for normalisation. The datasets were then filtered for the following terminal butyrate synthesis genes, based on Vital et al (2014): Butyrate kinase (EC_2.7.2.7; gene *buk*), AcylCoA-acetate CoA-transferase, alpha subunit (EC_2.8.3.8; gene *butCoAT*), Butyrate-acetoacetate CoA-transferase subunit_A (EC_2.8.3.9; gene *Ctf*), and Phenylalanyl-tRNA synthetase alpha chain (EC_6.1.1.20; gene *pheS*). It is important to note that the enzymes EC_2.8.3.8 and EC_2.8.3.9 have many synonyms other than the ones listed above, including the gene name *butCoAT*, due to broad

substrate specificity. Kraken2 classification data without the Bracken processing step was collated into a phyloseq object using the phyloseq package (version 1.42.0; McMurdie and Holmes, 2013) for differential abundance testing with ANCOMBC (version 2.0.3; Lin and Peddada, 2020).

Statistics

All statistics were done in R (version 4.2.3; R Core Team, 2023). To maintain consistency with prior soil microbiome studies, statistical significance was set at alpha = 0.05. Spatial maps were created using the ggmap package (version 3.0.2; Kahle and Wickham, 2013). Analysis of Compositions of Microbiomes with Bias Correction using the *ancombc2* function in the ANCOMBC package (version 2.0.3; Lin and Peddada, 2020) was performed on Kraken2 output data for differential abundance analyses. The ANCOMBC algorithm has been shown to minimise bias due to sampling fractions and reduces false discovery rates. The ggplot2 package (version 3.4.2; Wickham, 2016) was used for data visualisations. PERMANOVA (Adonis) tests and ordinations using principal coordinates analysis based on centred-log-ratio transformation were done using the vegan package (version 2.6.4; Oksanen et al., 2022). Random Forest regression modelling (Breiman, 2001) via the package ranger (Wright and Ziegler, 2015) was used to obtain variable importance results. The model fit was estimated using out-of-bag error from the bootstrap with mtry = 6, ntree = 500. The resulting Random Forest decision tree model explained 34.1% of the variance in our datasets. The variable importance plot was created using random permutations for each predictor variable's values in out-of-bag data, then calculating the mean decrease in node impurity.

RESULTS

We obtained 1,334,560,000 raw sequence reads from our soil samples, of which 84.76% had a Phred score >Q30. After quality control and removal of adapters, 1,179,992,332 paired reads remained. Taxonomic annotation of the reads identified the presence of 107 putative butyrate-producing bacterial species, which covered 90.7% of the initial target set of 118 putative butyrate producers (**Table S3**). This working list of 107 species was utilised in our taxonomic abundance analyses.

Effects of land cover on soil butyrate-producing bacterial species

The total relative abundances of butyrate-producing bacteria were similar between sports fields (mean total relative abundance = 0.172) and nature parks (mean total relative abundance = 0.18) (t = -1.096, df = 17, p = 0.29; **Figure 2A**). Sports fields had significantly higher relative abundances of three butyrate-producing bacterial species: *Geobacter metallireducens* (lfc = 0.396, adj p = 0.004), *Anaerotignum propionicum* (lfc = 0.304, adj p = 0.043), and *Clostridium kluyveri* (lfc = 0.289, adj p = 0.024), compared to nature parks (**Figure 2B**). However, sports fields and nature parks shared the same ten most abundant butyrate-producing bacterial species (**Figure 2C**). In sports fields, the butyrate-producing bacterial species with the two highest abundances were *Sorangium cellulosum* and *Micromonospora aurantiaca*. In nature parks, the butyrate-producing bacterial species with the two highest abundances were *Sorangium cellulosum* and *Kribbella flavida*. Butyrate-producing bacterial community compositions were also similar between sports fields and nature parks (Adonis PERMANOVA: F = 1.905, R^2 = 0.06, df = 1, p = 0.10; **Figure S3**).

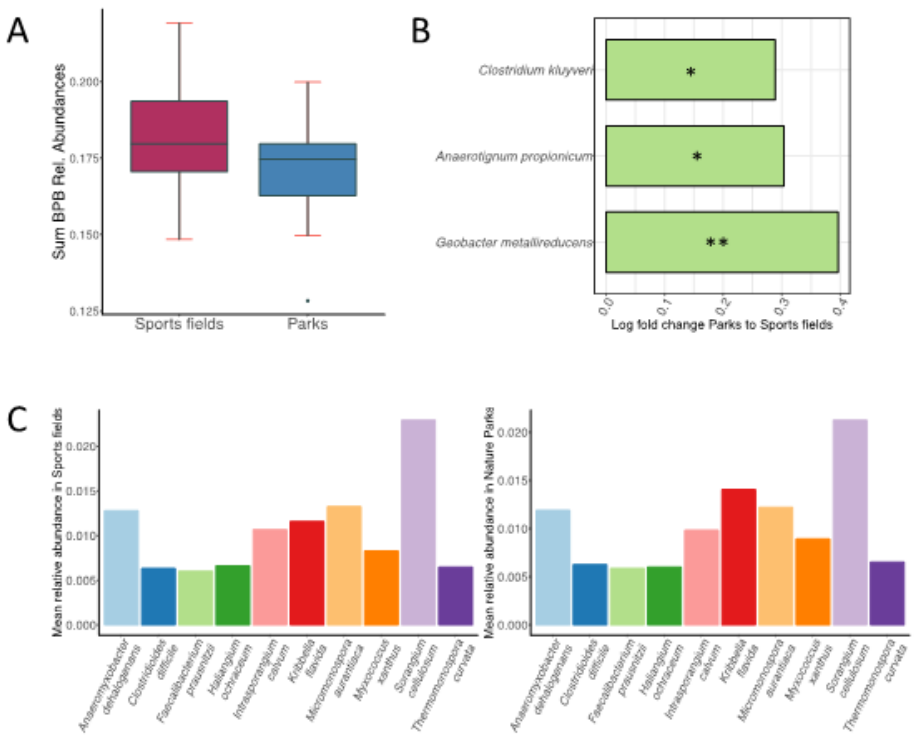


Figure 2. (A) Boxplots of total relative abundances of soil butyrate-producing bacteria by greenspace type (not significantly different). Boxes show the median and interquartile range, while

whiskers extend to the remaining range of data. (B) Log fold change (from nature parks to sports fields) in significantly differentially abundant soil butyrate-producing bacterial species. Note these species shown have lower mean relative abundances than the ten most common species highlighted in (C). (C) Bar plot showing the mean relative abundances of the ten most abundant soil butyrate-producing bacterial species by greenspace type. Y axis shows the mean relative abundances.

Butyrate synthesis terminal genes in soils

Sports fields had a significantly higher abundance of terminal butyrate production genes *buk* (Wilcoxon test: W = 30, $p < 0.001$; **Figure 3A**) and *butCoAT* (Welch t-test: t = -2.673, df = 16.3, $p = 0.016$; **Figure 3B**) than nature parks.

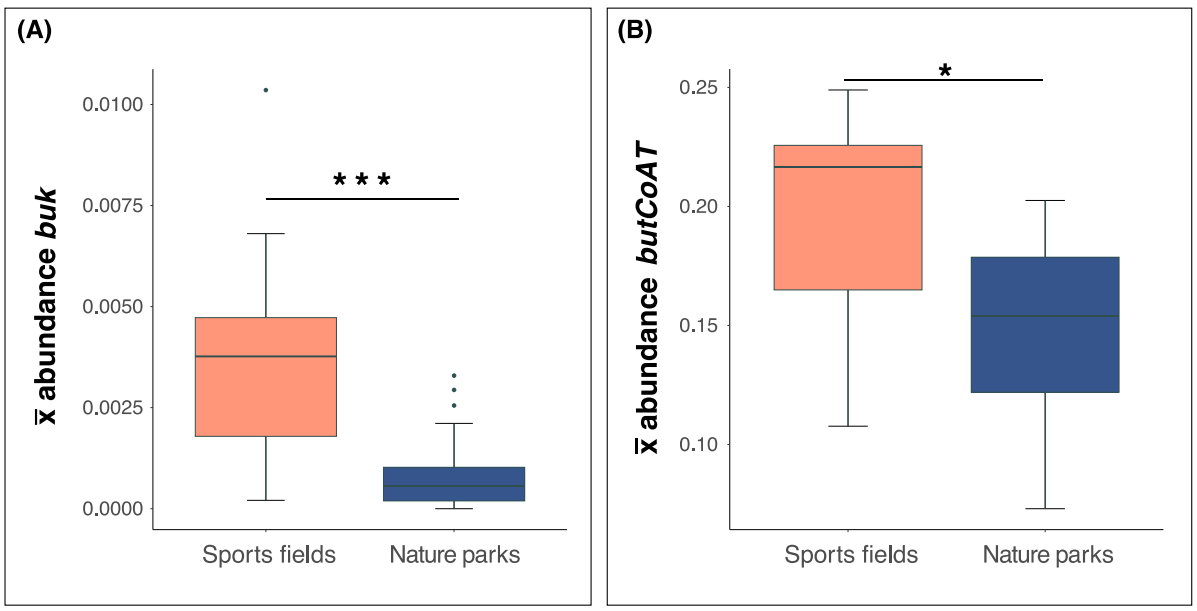


Figure 3. (A) Boxplots of abundances of butyrate metabolism genes *buk* by greenspace type (Wilcoxon test: W = 30, $p < 0.001$; sports fields $n = 11$, nature parks $n = 21$). (B) Boxplots of abundances of butyrate metabolism gene *butCoAT* by greenspace type (Welch t-test: t = -2.673, df = 16.3, $p = 0.016$). The y-axis shows the mean count of the genes across the sites, normalised by the count of single copy gene *pheS*. Boxes show the median and interquartile range, while whiskers extend to the remaining range of data.

Butyrate concentrations in soils

Greenspace type had a strong effect on the soil butyrate concentration (Welch t-test: t = -3.20, df = 22.9, $p = 0.004$; **Figure 4A**), with sports fields having a higher concentration ($\bar{x} = 0.194$ µM, $n = 10$) than nature parks ($\bar{x} = 0.135$ µM, $n = 22$). Soil butyrate associated positively with *buk* read abundances (F = 5.295, df = 1 and 29, adj $R^2 = 0.13$, $p = 0.029$; **Figure 4B**), but not with *butCoAT* (F = 2.568, df = 1 and 29, adj $R^2 = 0.05$, $p = 0.12$). Soil butyrate also did not associate with the sum of

butyrate-producing bacterial relative read abundances F = 0.14, df = 1 and 29, adj R^2 = -0.03, *p* = 0.71). Interestingly, soil butyrate had no association with the gene *Ctf*, which encodes an enzyme that routes butyryl-CoA into other non-butyrate pathways (e.g., lysine degradation; F = 0.002, df = 1 and 29, adj R^2 = -0.03, *p* = 0.97).

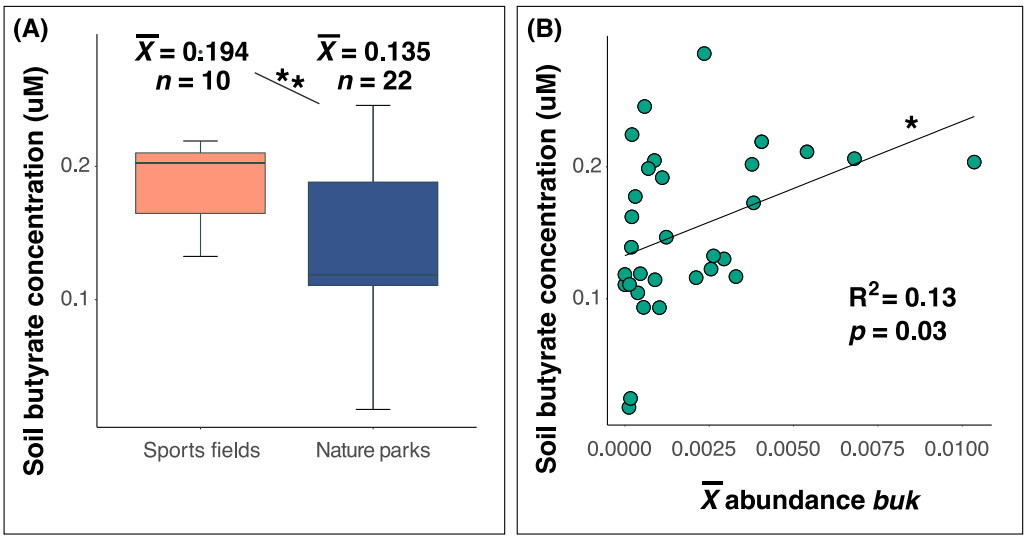


Figure 4. (A) Boxplots of soil butyrate concentrations by greenspace type (Welch t-test: t = -3.20, df = 22.9, *p* = 0.004), with sports fields having a higher concentration ($\bar{x}$ = 0.194 µM, *n* = 10) than nature parks ($\overline{x}$ = 0.135 µM, *n* = 22). The y-axis shows the soil butyrate concentration. Boxes show the median and interquartile range, while whiskers extend to the remaining range of data. (B) Relationship of the normalised abundance of butyrate metabolism gene *buk* with soil butyrate concentration (F = 5.295, df = 1 and 29, adj R^2 = 0.13, *p* = 0.029). The X-axis shows the mean count of *buk* across the sites, normalised by the count of the single copy gene *pheS*.

Effects of soil abiotic parameters on butyrate-producing bacteria

A Random Forest regression model with the soil parameters explained 33.3% of the variation in soil butyrate. Soil moisture (variable importance = 0.0116; Pearson r = 0.50, *p* = 0.004), calcium (variable importance = 0.0111; Pearson r = -0.62, *p* < 0.001), iron (variable importance = 0.0110, Spearman rho = 0.54, *p* = 0.002), ammonium nitrogen (variable importance = 0.0107; Spearman rho = 0.58, *p* < 0.001), and organic carbon (variable importance = 0.009; Pearson r = 0.45, *p* = 0.013) had the five highest influences on soil butyrate concentration **(Figure 5**). We further examined the effects of these five parameters on terminal butyrate production genes. Iron had a strong effect on *buk* read abundances (Spearman rho = 0.56, *p* = 0.001) and calcium strongly and negatively affected *buk* read abundances (Spearman rho = -0.57, *p* = 0.001; **Table 1**). All other effects were negligible.

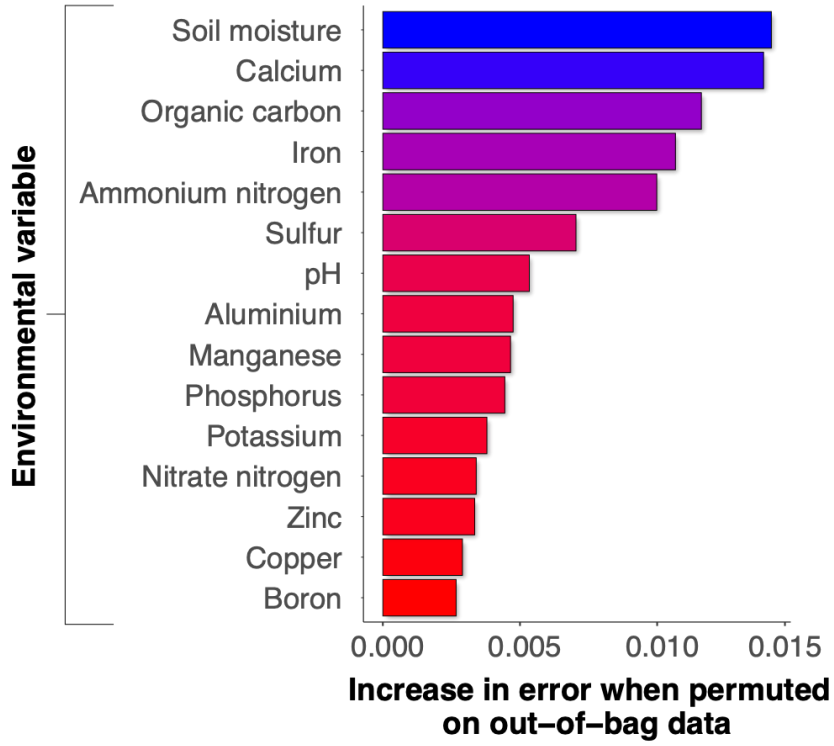


Figure 5. Random forest variable importance of soil physicochemical variables on soil butyrate concentration. Model performance R^2 = 0.34, out-of-bag MSE = 0.0024.

Table 1. Soil physicochemical values and correlation coefficients with soil butyrate concentrations and butyrate production terminal gene abundances.

Soil parameter	Sports fields (mean ± SD)	Nature parks (mean ± SD)	Correlation coefficient with soil butyrate conc.	Correlation coefficient with *buk*	Correlation coefficient with *butCoAT*
Moisture (%)	25.4 ± 5.40	18.9 ± 7.98	.50**	.15	.13
Calcium (meq/100 g)	11.2 ± 5.52	21 ± 7.53	−.62***	−.57**	−.28
Organic carbon (%)	4.05 ± 0.99	4.14 ± 0.38	.45*	.02	−.12
Iron (mg/kg)	117 ± 94.8	59.7 ± 46.9	.54**	.56**	.32
Ammonium nitrogen (mg/kg)	7.36 ± 2.20	5.91 ± 3.28	.58**	.12	.18

DISCUSSION

We examined soils in two common types of urban greenspaces – sports fields and nature parks – and show that sports fields had higher butyrate concentrations and butyrate production gene abundances. Soil abiotic conditions, specifically moisture, calcium, iron, ammonium nitrogen, and organic carbon, strongly influenced soil butyrate concentrations, and soil iron and calcium influenced soil butyrate production gene abundances. These findings suggest that urban greenspace management that impacts on these soil abiotic conditions can influence the activity of butyrate producing bacteria. Our work shows that urban greenspaces, particularly sports fields, have the potential to provide exposure to health-associated butyrate-producing bacteria to humans. These findings have important implications to the designing and planning of urban greenspaces, plus to public health and human microbiome research.

Soil moisture affects the production of butyrate

We show that soil moisture levels strongly influenced soil butyrate concentrations. Butyrate production requires anaerobic conditions (Baxter et al., 2019), and wetting of soils (e.g., via rainfall, irrigation) can deplete oxygen and induce redox heterogeneity within soils, thereby establishing anoxic microsites with conditions conducive to butyrate production (Lacroix et al., 2023; Lentini et al., 2012). Our finding runs parallel with Brame et al. (2022) who found that hydrological fluctuations with wetting-and-drying cycles associated with greater butyrate-producing bacterial read abundances in outdoor soils in the Australia continent. Thus, intermittent irrigation and rainfall events on urban greenspaces should provide the anoxic conditions required for fermentative butyrate synthesis. Future similar projects that assess absolute abundance via qPCR could provide further insights into whether greater butyrate production is due to greater absolute abundances of butyrate producers or to their increased activity.

We derived a new method to directly quantify soil butyrate concentrations. This has not been done previously, probably because butyrate concentrations in soil tend to be low. Across our samples, we observed butyrate concentrations <100 µM, which is far lower than typically found in human gut

samples (e.g., faecal butyrate concentrations tend to be 2-70 mM; Baxter et al., 2019; Kircher et al., 2022). Due to the volatility of butyrate, even with our methods that aimed to minimise butyrate loss (e.g., by obtaining each butyrate-related soil sample before the bulk soil sample, and by immediately fixing the soil sample in phosphoric acid), our measured butyrate concentrations may have been impacted by butyrate loss. Refining how butyrate is directly measured in soil would thus be useful.

Bidirectional relationship between soil iron and butyrate producing bacterial activity

We observed a strong positive association between soil iron levels and the abundances of the terminal butyrate production gene *buk*. Iron's role in the butyrate metabolic pathway is not fully understood, and the reversibility of the enzyme butyrate kinase, encoded by the *buk* gene, suggests that a bidirectional relationship between iron and *buk* may be possible. This bidirectional relationship could characterised as follows: (A) iron facilitates butyrate production and, in reverse, B) iron-reducing bacteria oxidise butyrate and produce ferrous Fe(II) iron. We describe the evidence for these possible two explanations and discuss their implications below.

First, iron has been positively associated with measured butyrate concentrations. Dostal et al. (2015) showed in faecal samples that iron levels modulated butyrate-related bacterial communities by using a polyfermenter model inoculated with the colonic microbiota from a child. Adjusting the iron levels (i.e., simulating different conditions within the proximal colon of a child) elicited substantial changes in the butyrate-producing bacterial communities (e.g., Lachnospiraceae and Ruminococcaceae). However, the mechanisms of such an effect are not yet known. Furthermore, they observed that the abundances of the butyrate producer *Roseburia* were diminished by high levels of iron. *Roseburia* spp. utilise an energetically-favourable CoA transferase with acetate to form butyrate (Hartmanis and Gatenbeck, 1984; Hillman et al., 2020). This CoA transfer employs a ping pong bi-bi kinetic mechanism that bears double competitive substrate inhibition, where high concentrations of one substrate creates an inhibitory effect (Gheshlaghi et al., 2009).

On the other hand, our finding may reflect the role of iron as an electron acceptor in an anaerobic energy production pathway. Iron-reducing bacteria can create energy by oxidising butyrate using the same enzymes from butyrate production, but in reverse direction – they couple this oxidation with the reduction of iron from ferric Fe(III) to the more soluble ferrous Fe(II) state (Lentini et al., 2012). Indeed, coastal paddy soil research (Jiang et al., 2023) showed that, under anoxic soil conditions, butyrate enhances abundances of dissimilatory iron-reducing bacteria such as *Geobacter*, a genus we found in higher read abundances in sports fields.

Thus, the strong relationship between iron levels and butyrate-producing bacterial activity levels appears to be bidirectional and may be constrained to a particular concentration range of iron. Future research could investigate the influence of varied iron concentrations on butyrate-producing bacteria in urban greenspace soils, with implications on landscape management practices that aim to modulate butyrate-producing bacteria.

Abundances of genes buk and butCoAT could reflect carbon substrate availability

We show that sports fields had higher abundances of the terminal genes *buk* and *butCoAT* than nature parks. The enzymes for both *buk* and *butCoAT* are reversible (Chang et al., 2021; Huang et al., 2000), and the direction of movement in their metabolic pathways could be related to soil conditions, such as the carbon substrate availability. *Geobacter metallireducens* utilises the *buk* gene and had the highest differential abundance between sports fields and nature parks. *G. metallireducens* is known to grow vigorously when acetate is the only carbon source (Hartmanis and Gatenbeck, 1984; Lentini et al., 2012). *Anaerotignum propionicum* and *Clostridium kluyveri* utilise *butCoAT* rather than *buk* (Hillman et al., 2020) and also had higher read abundances in sports fields. *C. kluyveri* is a known butyrate producer and uses acetate and ethanol as carbon sources (Seedorf et al., 2008). Therefore, acetate could be an important substrate for butyrate metabolism in sports field soils via both the *buk* and *butCoAT* metabolic pathways. This aligns with findings in Liddicoat et al., (2023), who reported increased potential metabolism of acetate in more highly disturbed (compared

to more mature and natural) plant-soil systems. Though we did not evaluate acetate in this study, future research could further examine the role of acetate in shaping urban greenspace soil microbial communities.

Soil organic carbon and ammonium nitrogen associate with butyrate producing bacterial activity

We report that soil organic carbon and ammonium nitrogen correlated positively with soil butyrate concentrations. Carbon-containing soil organic matter consists of plant, animal, and microbial matter in varying stages of decomposition (Lal et al., 1997). In anoxic conditions, butyrate-producing bacteria generate energy by fermenting organic carbon substrates into butyrate (Buckel, 2021). While increased organic carbon associated with enhanced butyrate production, we found similar levels of organic carbon between sports fields and nature parks but higher levels of butyrate in sports fields. Thus, sports fields appear to have additional conditions that enhance the effects of organic carbon on butyrate producer activity. These findings show that greenspace management strategies to increase soil organic carbon alone may have limited effects on the activity of butyrate-producing bacteria.

Ammonium nitrogen also associated with soil butyrate concentration. Nitrogen, together with carbon, is a key element present in soil organic matter, which acts as a store of nutrients, including cations such as ammonium. As such, ammonium nitrogen levels may reflect both nutrient storage (i.e., attached to soil organic matter) and cycling (i.e., from soil organic matter decomposition), such that higher ammonium nitrogen reflects increased fermentative pathways toward butyrate production. Alternately, increased soil ammonium nitrogen could reflect remnant levels from reduced ammonium breakdown. Soil ammonium is degraded by the feammox reaction. Jiang et al. (2023) observed that an increase in volatile fatty acids (e.g., butyrate) in coastal paddy soil from fertiliser degradation promoted the abundances of butyrate-oxidising bacteria such as *Geobacter* They proposed that iron reduction competes with the feammox reaction. In this way an increase in butyrate oxidation could competitively inhibit the breakdown of ammonium, resulting in sustained

levels of soil ammonium. Thus, future studies investigating how soil butyrate and ammonium are linked could clarify key roles of short-chain fatty acids in plant-soil systems.

Temporal incongruence

It is worth noting that the soil samples were collected in May and June, 2022, and the vegetation surveys were completed in August, 2022. This temporal incongruence may have had an unknown effect on our results. However, we expect that such an effect would be minimal for two reasons. First, our vegetation surveys only included woody plants, which generally grow more slowly than seasonal weeds. Second, the soil samples and vegetation survyes were both performed during a single winter season, which could minimise seasonal effects.

Conclusions

We show that urban greenspaces are reservoirs of butyrate-producing bacteria, which were more actively producing butyrate in sports fields than nature parks. Soil conditions such as moisture, iron, ammonium nitrogen, and organic carbon enhanced butyrate-related activity, but we found evidence of bidirectional movement of enzymatic steps on those pathways, warranting further investigation. Urban greenspace management practices therefore appear to play important roles in shaping soil butyrate-producing bacterial activity. Our study shows that sports fields could offer greater potential than nature parks to expose and supply health-associated environmental butyrate-producing bacteria to people. These results suggest that commonly employed urban greenspace management practices (e.g., irrigation, fertiliser addition) could strongly influence people's exposure to butyrate-producing bacteria. These findings provide opportunities for landscape designers, urban planners, ecologists, and public health experts to work together on new ways to support human health via urban greenspaces.

Chapter 5. Short-term passive greenspace exposures have little effect on nasal microbiomes: a cross-over exposure study of a Māori cohort

---

This work appears in the following publication:

Brame JE, Warbrick I, Heke D, Liddicoat, C, Breed MF. Short-term passive greenspace exposures have little effect on nasal microbiomes: a cross-over exposure study of a Māori cohort. 2024. *bioRxiv* (under review in *Environmental Research*). doi: 10.1101/2024.01.17.576148.

Context

Chapter 5 was a collaborative effort between our research group, researchers at Auckland University of Technology, and a group of Māori participants. Indigenous health interventions have emerged in New Zealand aimed at increasing people's interactions with and exposure to macro and microbial diversity. Urban greenspaces provide opportunities for people to gain such exposures. However, the dynamics and pathways of microbial transfer from natural environments onto a person remain poorly understood. Here, we analysed bacterial 16S rRNA amplicons in air samples (n = 7) and pre- and post-exposure nasal samples (n = 238) from 35 participants who had 30-minute exposures in an outdoor park. The participants were organised into two groups: over eight days each group had two outdoor park exposures and two indoor office exposures, with a cross-over study design and washout days between exposure days.

In Chapter 5, I aimed to answer the following research question: What are the effects of participant group, location (outdoor park vs. indoor office), and exposures (pre vs. post) on the nasal bacterial community composition and three key suspected health-associated bacterial indicators (i.e., alpha diversity, generic diversity of Gammaproteobacteria, and read abundances of butyrate-producing bacteria)?

Author contributions

J.B., I.W., D.H., C.L., and M.B. contributed to the research design and logistical planning. J.B., C.L., and M.B. analysed the data, with input from I.W. and D.H. J.B. wrote the initial manuscript, with subsequent revision input from all authors.

Research Paper

Short-term passive greenspace exposures have little effect on nasal microbiomes: a cross-over exposure study of a Māori cohort

Joel E. Brame[a], Isaac Warbrick[b], Deborah Heke[b], Craig Liddicoat[a], Martin F. Breed[a]

[a]College of Science and Engineering, Flinders University, Bedford Park, SA 5042, Australia

[b]Taupua Waiora Māori Research Centre, Auckland University of Technology, Auckland, New Zealand

Author email addresses:

Joel E. Brame: joel.brame@flinders.edu.au

Isaac Warbrick: isaac.warbrick@aut.ac.nz

Deborah Heke: deborah.heke@aut.ac.nz

Craig Liddicoat: craig.liddicoat@flinders.edu.au

Martin F. Breed: martin.breed@flinders.edu.au

Corresponding author:

Joel E. Brame

Mailing address: College of Science and Engineering, Flinders University, Bedford Park, SA 5042, Australia

ABSTRACT

Indigenous health interventions have emerged in New Zealand aimed at increasing people's interactions with and exposure to macro and microbial diversity. Urban greenspaces provide opportunities for people to gain such exposures. However, the dynamics and pathways of microbial transfer from natural environments onto a person remain poorly understood. Here, we analysed bacterial 16S rRNA amplicons in air samples (n = 7) and pre- and post-exposure nasal samples (n = 238) from 35 participants who had 30-minute exposures in an outdoor park. The participants were organised into two groups: over eight days each group had two outdoor park exposures and two indoor office exposures, with a cross-over study design and washout days between exposure days. We investigated the effects of participant group, location (outdoor park vs. indoor office), and exposures (pre vs. post) on the nasal bacterial community composition and three key suspected health-associated bacterial indicators (alpha diversity, generic diversity of Gammaproteobacteria, and read abundances of butyrate-producing bacteria). The participants had distinct nasal bacterial communities, but these communities did not display notable shifts in composition following exposures. The community composition and key health bacterial indicators were stable throughout the trial period, with no clear or consistent effects of group, location, or exposure. We conclude that 30-minute exposure periods to urban greenspaces are unlikely to create notable changes in the nasal microbiome of visitors, which contrasts with previous research. Our results suggest that longer exposures or activities that involves closer interaction with microbial rich ecological components (e.g., soil) are required for greenspace exposures to result in noteworthy changes in the nasal microbiome.

Funding sources

This work was supported by funding from the Flinders Foundation; and a Project Grant from the Health Research Council of New Zealand.

Human ethics approval

Ethical approval for this study was obtained on 7 Dec 2021 from AUTEC – Auckland University of Technology Ethics Committee (Application 21/414). All procedures were performed in compliance with relevant laws and institutional guidelines. Informed consent for experimentation was obtained by all human subjects.

INTRODUCTION

Disconnection from natural environments is a characteristic of urban lifestyles and one which is associated with poorer health outcomes (Robinson et al., 2024; Sibthorpe & Brymer, 2020). For Indigenous Peoples, whose identity, culture, and health are intertwined with the natural environment (Durie, 2004; Warbrick et al., 2016), the disconnection from ancestral lands and natural environments generally, is particularly concerning. Warbrick et al (2023) recently proposed that the relationship between environmental microbiomes and health has important implications for the health of Indigenous Peoples, despite Indigenous people rarely being represented in studies of the microbiome. With the majority of people now living in cities (United Nations, 2018), urban greenspaces and their accompanying aerobiomes are key points of exposure to natural environmental microbiomes (Robinson et al., 2023).

Bacterial colonisation of the human body occurs during and after birth, with post-birth bacterial communities shaped by people's environments (Chong, D'amato, & Rosário Filho, 2022; Gilbert et al., 2018; Nielsen et al., 2020). Pathways of exposure to environmental bacteria include ingested and inhaled substances, either directly or indirectly (e.g., via hand-to-face transfer). Air is a well-understood transmission medium for microbiota, which triggers health conditions such as allergies and infectious disease (Kim, Kabir, & Jahan, 2018). However, the transmission pathway of health-supporting airborne bacteria has received much less attention (Robinson et al., 2023). Airborne bacterial communities (aerobiomes) of built indoor environments are highly variable due to a wide range of possible conditions (Ghosh, Lal, & Srivastava, 2015). Outdoor environments are also rich aerobiome reservoirs (Robinson et al., 2021). Because airborne dispersal of microbiota is a key pathway of bacterial exposure and transfer, air transfer dynamics can be studied via sampling nasal bacterial communities (Robinson et al., 2023). Nasal microbiome changes may reflect the characteristics of aerobiomes of recent exposure, suggesting that the study of outdoor aerobiomes can provide critical insights into human microbiome assemblages (Selway et al., 2020). However, few studies have examined how nasal microbiomes change after exposure to outdoor air.

Greenspace aerobiomes originate from leaf surfaces and soil, with modulating effects from vegetation complexity and height above the ground (i.e., vertical stratification; Robinson et al., 2021), air pollution (Franchitti et al., 2022), and wind-carried airshed influences (Robinson et al., 2023). In urban settings, land cover has a strong influence on the composition of aerobiomes. For example, the aerobiomes of parks have different community compositions than adjoining parking lots (Mhuireach et al., 2019). Among greenspaces, amenity grassland aerobiomes have different compositions to remnant native vegetation aerobiomes and possess consistent alpha diversity at heights up to 2 m (Robinson et al., 2021). Thus, urban amenity grasslands should have distinct aerobiomes compared to indoor offices and provide useful locations to study the transfer of aerobiomes into the airways of people. Yet, the use of amenity grassland aerobiomes in bacterial transfer studies is limited.

The Biodiversity Hypothesis describes how exposure to a greater amount of microbial diversity in the natural environment may be required to promote innate immune training and immunoregulation (Haahtela, 2019). In a complex network of interactions, exposure to bacterial diversity can modulate immune responses and reduce pro-inflammatory and allergenic antibodies and cytokines (Haahtela, 2019). For example, the diversity of Gammaproteobacterial genera on the skin has been associated with increased plasma transforming growth factor beta 1 (TGF-β1) levels, decreased interleukin-17 (pro-inflammatory cytokines), and increased relative abundance of regulatory T-cells (Roslund et al., 2020). Increased TGF-β1 and decreased interleukin-17 are associated with an anti-inflammatory molecular profile, and regulatory T-cells are critical for immunotolerance, including tolerance of commensal taxa (Roslund et al., 2020). Thus, exposure to higher alpha diversity of bacteria within outdoor aerobiomes with a low level of pathogenic taxa could potentially support human health (Spragge et al., 2023).

Butyrate-producing bacteria are key members of the human and animal gut with numerous health benefits, and after birth they are primarily supplied by the environment with nutritional support via ingestion of fibre (Brame et al., 2021). Certain outdoor environments are reservoirs of butyrate producers that could disperse into the aerobiome and transfer onto people visiting those environments (Brame et al., 2022). Thus, butyrate-producing bacterial read abundances could provide indicators of human health-associated benefits of aerobiome exposure.

Here we studied the changes in 16S rRNA amplicons in pre- and post-exposure nasal microbiome samples from 35 Māori (Indigenous New Zealand) participants, divided into groups (A and B), who spent two repeated 30-minute exposure periods in each of two locations: an indoor office and an outdoor park (amenity grassland). We utilised a cross-over study design to control for effects of group and day, with two exposure days in one location (Days 1 and 3), followed by a two-day washout period, then two further exposure days in the other location (Days 6 and 8). To understand the influences of exposures on the nasal microbiomes, we examined the effects of location, individual, group, single exposures, and repeated exposures on (1) the nasal bacterial alpha diversity, (2) nasal bacterial community composition, and (3) specific bacterial taxonomic groups with known health associations (Gammaproteobacterial diversity and butyrate-producing bacterial abundances).

MATERIALS AND METHODS

Experimental design

We utilised a crossover trial design (**Figure 1A**). We recruited 35 participants into the trial, which took place March 15-22, 2023. Te Arawa Whānau Ora is an Indigenous community health organisation, and all participants in this study were adult employees of the organisation and identify as Māori. Roughly half of the participants worked during the day in the office used for the study, while the others worked in a nearby office building and came to the study office each day of exposure. For cultural sensitivity, no other data about the participants were obtained for this study, although participants were encouraged not to participate if they were experiencing an active

infection or nasal lesions. The participants were divided into two groups: outdoor and indoor for exposure days 1 and 3, with crossover for exposure days 6 and 8. Exposure days were on March 15 and 17, then on March 20 and 22, allowing for a single washout day between testing days and two washout days before the crossover.

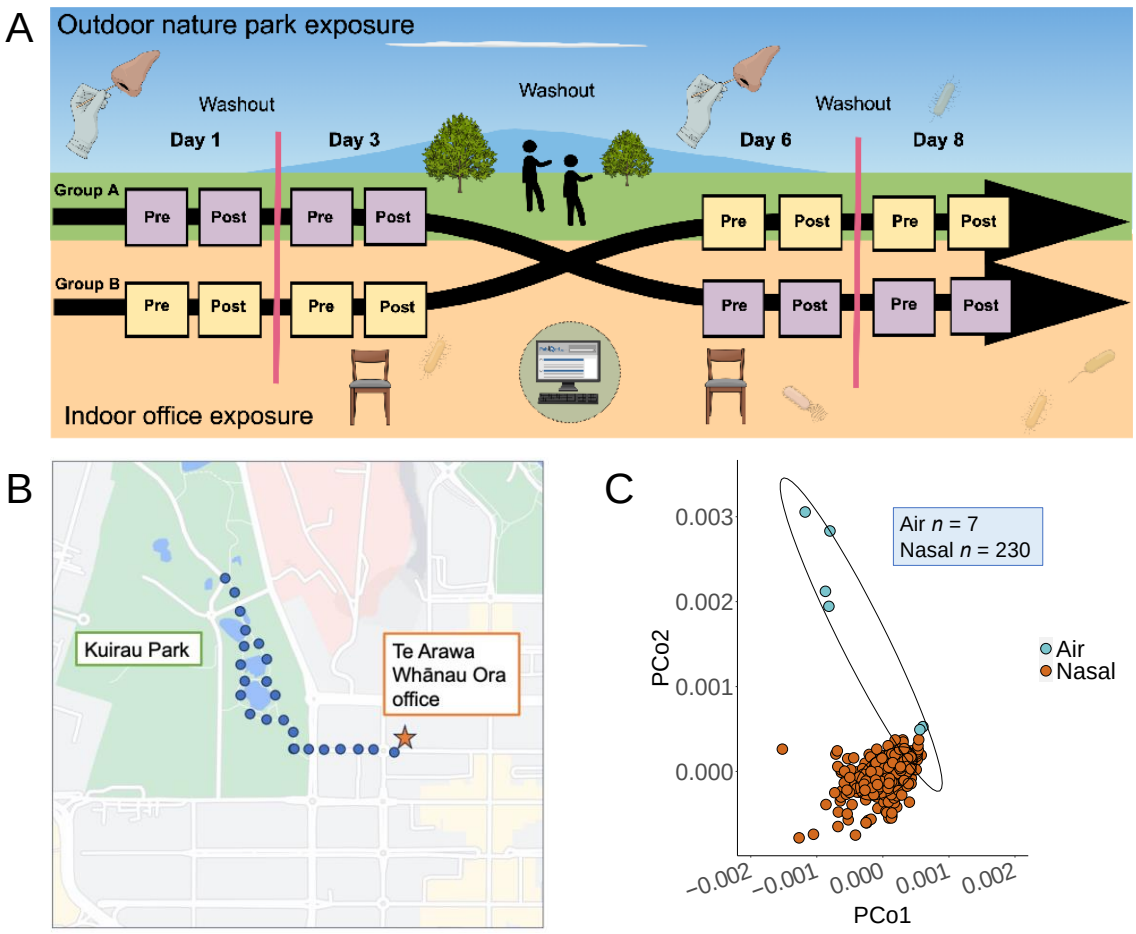


Figure 1. (A) Overview of the cross-over experimental design. (B) Walking path map of the outdoor treatment group in Rotorua, New Zealand. Map generated with Google maps. (C) Principal coordinates analysis based on centred-log ratio compositional abundance data displaying variation in community composition by sample type (Adonis PERMANOVA: F = 5.515, R^2 = 0.023, p = 0.001).

The outdoor treatment group met at the Te Arawa Whānau Ora office in Rotorua, New Zealand, at approximately 8:30am. Participants' noses were swabbed pre-exposure (hereafter referred to as "Pre", see description below), and they then went for a walk to Kuirau Park, approximately 600 m from the office, for 30 minutes (**Figure 1B**). Upon their return, before entering the office, they were re-tested with a second nasal swab (hereafter referred to as "Post").

The indoor treatment group met at the same office at the same time and day as the outdoor treatment group. Their noses were swabbed using the same methods. However, during the 30-minute exposure period, they remained in the office.

Nasal swabbing

Nasal swab samples were obtained by inserting a sterile nylon-flocked swab tip (FLOQSwabs Lot 2011490, Copan Flock Technologies, Bescia, Italy) into the anterior nares and rotating in a circular motion for 3-5 seconds per naris, then repeated in the opposite naris using the same swab. The swab tip was then immediately snipped into a sterile 15 mL falcon tube, sealed with the lid, wrapped with parafilm, and placed in a -20°C freezer in the office.

Air sampling

Air samples were obtained at an outdoor park site along the same walking path where participants walked during their outdoor period and at a central indoor location in the Te Arawa Whānau Ora office. The Kuirau Park site in Rotorua, New Zealand is predominantly amenity grassland with interspersed geothermal springs. At Kuirau Park, air samples were collected over an approximately 8-hour period during each testing day, following the method described in Mhuireach et al. (2016). The aerobiome sampling stations were set up on site between 0800 and 0830 hours and collected between 1500 and 1530 hours. At the Te Arawa Whānau Ora office control site, air samples were collected following the same procedures and the same times. On one day, March 17, the weather was rainy and the air stand assembly using protective umbrellas was vandalised, thus an outdoor air sample was not obtained for that day.

The outdoor park air sampling station was made of plastic boxes and achieved a height of 1.2 m. Sampling at this height should be representative of aerobiome exposure potential for children and adults alike, and is within the 2 m height range of similar alpha diversity as measured elsewhere in amenity grassland aerobiomes (Robinson et al., 2021). The indoor office sampling station was a

single plastic box placed on a table, achieving a height of approximately 1.5 m. On the top of each station, we opened and placed three sterile clear plastic petri dish bases and lids, which provided six collection surfaces per site. This method of passive aerobiome sampling has been shown to be as effective as active sampling methods (Mhuireach et al., 2016). On two days, a field control was generated by holding open an additional petri dish for 30 seconds at the equipment box. Immediately after the air sampling activity, each petri dish was sealed, labelled, and placed in the office freezer at -20°C until DNA extraction (described below).

DNA extraction, PCR, sequencing

Within one week of obtaining all samples, DNA extractions and quantifications were performed in a PC2 laboratory at Auckland University of Technology, Auckland, New Zealand. To transport samples from Rotorua to Auckland, samples were removed from the office freezer, placed onto ice in a sealed insulated container, and transported by vehicle to the lab. Upon arrival at the lab, they were immediately placed into a -20 °C freezer.

The petri dishes for each site were opened and swabbed with sterile nylon-flocked swab tips (FLOQSwabs) inside a laminar flow cabinet. One swab and 40 µL of added sterile phosphate-buffered saline was used for swabbing all six surfaces, except for surfaces that showed visual signs of damage or contamination, for approximately four minutes total using a consistent pattern of swabbing. The tips were cut directly into 15 mL sterile falcon tubes. We obtained an extraction blank control for each extraction batch using the same process as samples but without a swab tip.

We used the QIAamp DNA Mini Kit (QIAGEN) for all samples and followed the manufacturer's instructions with two modifications to increase final concentration: the incubation step was extended from 10 min to 15 min, and the final elution buffer volume was reduced from 80 µL to 60 µL. The extraction concentrations were then quantified using a Qubit High Sensitivity dsDNA assay (ThermoFisher Scientific). Once DNA concentrations were verified, PCR amplification of the

bacterial 16S rRNA V3-V4 regions was performed in the lab at Auckland University of Technology using Kappa HiFi Taq mix with 341F-805R primers (Kapa Biosystems) via PCR on an Eppendorf Vapo.Protect Mastercycler Pro thermocycler. The first PCR round included 38 amplification cycles. Plate clean-up was performed via AMPure XP reagent. To normalise clean PCR products to 1 ng/µL, samples below 1 ng/µL were concentrated using the Eppendorf Concentration and using the following conditions: D-AQ, 30 C, 18 min. Second round PCR used the Nextera XT Index Kit to index samples, with eight cycles of amplification. Samples were then pooled, cleaned with AMPure XP reagent, and quantified using Qubit High Sensitivity. The Bioanalyzer 2100 expert High Sensitivity DNA assay was performed to check library quality and molarity, and libraries were pooled for equal molarity. Upon completion of library preparation, sequencing of amplicon sequence variants was completed on the Illumina Miseq V3 using the Illumina MiSeq Reagent Kit v3 (600 cycle). Four PCR negative blanks were generated during the library preparation steps for quality control.

Bioinformatics

From the 16S rRNA raw sequence data, amplicon sequence variants (ASVs) were trimmed and filtered using an established Qiime2 pipeline (version 2023.5), with forward reads truncated at 260 bp and reverse reads truncated at 198 bp. Taxonomy was assigned using the onboard Naïve Bayes taxonomic classifier and Silva database v 138.1. Sequences were then cleaned using scripts utilising the R *phyloseq* package (version 1.42.0; McMurdie & Holmes, 2013) by removing the following sequences: those assigned to mitochondria and chloroplasts, taxa that did not occur in at least two samples, and ASVs with total sums < 20 reads. Sequences that were likely of contamination origin were identified and removed using the R *decontam* package (version 1.18.0; Davis et al., 2018) using the function “isNotContaminant” suited for low biomass samples.

Statistical analysis

All statistical analyses were performed using R (version 4.2.3; R Core Team, 2023). To maintain consistency with prior aerobiome studies, statistical significance was set at alpha = 0.05. Sample alpha diversity based on Hill numbers was examined using R *hillR* package (Li, 2018), which integrates sample size and coverage. We set the q parameter for Hill numbers at 0.80 for reduced sensitivity to relative abundances compared with Shannon index.

To prepare for beta diversity tests, the read abundance data were evaluated using R *zCompositions* package (version 1.4.0.1; Palarea-Albaladejo & Martín-Fernández, 2015), zeros were imputed using the R *scImpute* package (version 0.0.9; Li & Li, 2018), and eight low total read abundance samples were discarded to reduce data sparsity. The resultant read abundances were then transformed with centred-log ratio using the R *compositions* package (verion 2.0.6; Van den Boogaart & Tolosana-Delgado, 2008), followed by ordination with principal coordinates analysis using R *ecodist* package (version 2.0.9; Goslee & Urban, 2007), based on Aitchison distances obtained with the R *vegan* package (version 2.6.4; Oksanen et al., 2022); statistics were generated using PERMANOVA (Adonis) tests via the R *vegan* package. Distance-to-centroid analyses were performed using the R *vegan* package. Maps were created using the R *ggmap* package (version 3.0.2; Kahle & Wickham, 2013). Differential abundance analysis using the Analysis of Compositions of Microbiomes with Bias Correction (ANCOM-BC) method was performed on untransformed amplicon data with the *ancombc2* function in the R *ANCOMBC* package (version 2.0.3; Lin & Peddada, 2020). Participant was set as a random effect (rand_formula) for mixed effects modelling. The p-value adjustment was set as "fdr", and prv_cut and lib_cut were set at "0". The ANCOMBC algorithm has been shown to minimise bias due to sampling fractions and reduces false discovery rates. We downloaded a comprehensive list of pathogens from Bartlett et al. (2022) to examine pathogenic read abundances in the samples. Time-series analyses were performed using repeated-ANOVAS with R *rstatix* package (version 0.7.2; Kassambara, 2023). R *ggplot2* package (version 3.4.2; Wickham, 2016) was used for data visualisations.

RESULTS

Aerobiomes were different from nasal microbiomes

Aerobiomes had a higher alpha diversity (hill number = 82.3 ± 64.4 SD, n = 7) than nasal microbiomes (hill number = 19.5 ± 10.6 SD, n = 238; W = 1378, p = 0.003). Aerobiome location had no effect on alpha diversity between outdoor (hill number = 60.4 ± 70.3 SD, n = 3) and indoor (hill number = 98.8 ± 64.6 SD, n = 4) samples (t = 0.740, df = 4.2, p = 0.50). Overall, aerobiomes and nasal microbiomes had different community compositions (Adonis PERMANOVA: F = 5.515, R^2 = 0.023, p = 0.001; **Figure 1C**), and outdoor aerobiomes were compositionally similar to indoor aerobiomes (Adonis PERMANOVA: F = 1.268, R^2 = 0.20, p = 0.17). Additionally, all alpha diversity results using q = 0 and q = 1 for Hill number statistics are reported in **Table S1**.

Exposure effect on composition, diversity and differential ASV abundances

The 30-minute outdoor exposures did not change the nasal bacterial community composition for either group A (Adonis PERMANOVA: F = 0.686 R^2 = 0.013, p = 0.99) or group B (Adonis PERMANOVA: F = 0.726, R^2 = 0.013, p = 0.98) (**Figure 2A**). The 30-minute indoor exposures also did not change the community composition for either group A (Adonis PERMANOVA: F = 0.809, R^2 = 0.014, p = 0.89) or group B (Adonis PERMANOVA: F = 0.675, R^2 = 0.012, p = 0.99) (**Figure 2D**).

There was no effect of group on changes in nasal bacterial alpha diversity after 30-minute exposures among both outdoor exposures (Wilcox: W = 433, p = 0.83) and indoor exposures (Wilcox: W = 358, p = 0.18), even though the groups visited the locations on separate days. When the two groups were combined, there was no effect on the alpha diversity by the outdoor exposures (W = 1388, p = 0.11; **Figure 2B)** and the alpha diversity remained similar after the indoor exposures (W = 1818, p = 0.93; **Figure 2E**).

For group A, the 30-minute outdoor exposure had no effect on the read abundance of any genus on day 1. However, on day 3, the outdoor treatment resulted in a significant decrease in the genera *Escherichia-Shigella* (ANCOMBC: log fold change (lfc) = -1.91, adjusted-*p* (*q*) < 0.001) and *Pseudomonas* (ANCOMBC: lfc = -1.72, *q* < 0.001) (**Figure 2C**). For group B, on day 6, the outdoor treatment resulted in five taxa with significantly decreased read abundances: *Rheinheimera* (ANCOMBC: lfc = -3.37, *q* < 0.001), *Massilia* (ANCOMBC: lfc = -3.22, *q* < 0.001), *Acinetobacter* (ANCOMBC: lfc = -3.16, *q* < 0.001), *Flavobacterium* (ANCOMBC: lfc = -3.16, *q* < 0.001), and family Comomonadaceae (ANCOMBC: lfc = -2.10, *q* = 0.004; **Figure 2F**). The outdoor treatment had no effect on any genus on day 8 for group B (all data are in **Table S2**).

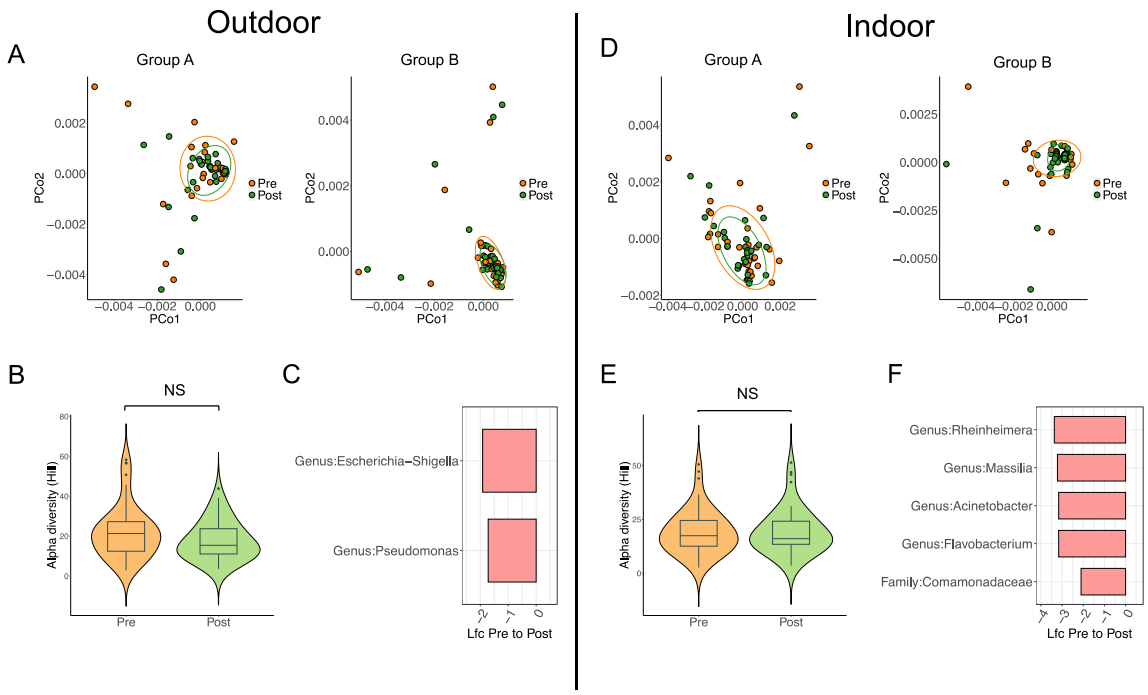


Figure 2. (A) Principal coordinates analysis based on centred-log ratio compositional abundance data displaying variation in community composition before (Pre) and after (Post) outdoor exposure for groups A and B. (B) Boxplots of changes in alpha diversity from before (Pre) and after (Post) outdoor exposure. The y-axis shows the alpha diversity based on Hill numbers. Boxes show the median and interquartile range, while whiskers extend to the remaining range of data. (C) Significantly differentially abundant genera in nasal microbiomes after outdoor exposure. The x axis shows the log fold change from before pre-exposure to post-exposure. Red bars indicate a decrease in log fold change. (D) Principal coordinates analysis based on centred-log ratio compositional abundance data displaying variation in community composition before (Pre) and after (Post) indoor exposure for groups A and B. (E) Boxplots of changes in alpha diversity from before (Pre) and after (Post) indoor exposure. The y-axis shows the alpha diversity based on Hill numbers. Boxes show the median and interquartile range, while whiskers extend to the remaining range of data. (F)

Significantly differentially abundant genera in nasal microbiomes after indoor exposure. The x axis shows the log fold change from pre-exposure to post-exposure. Red bars indicate a decrease in log fold change.

Exposure effects on health-associated bacterial groups

30-minute exposures had different effects in groups A and B on the number of Gammaproteobacteria genera (t-test: t = -2.111, df = 115.12, $p = 0.036$), so we examined the two groups separately. Indoor exposure significantly decreased the Gammaproteobacteria diversity in group A (t-test: t = -2.221, df = 56.61, $p = 0.03$) but had no effect in group B (Wilcox: W = 358, $p = 0.91$). Outdoor exposure had no effect on Gammaproteobacteria diversity for group A (t-test: t = -1.015, df = 49.3, $p = 0.32$) but weakly decreased Gammaproteobacteria diversity for group B (t-test: t = -1.905, df = 54.99, $p = 0.062$).

There was no effect of group on changes in nasal butyrate-producing bacterial read abundances after 30-minute exposures among both outdoor exposures (Wilcox: W = 329, $p = 0.24$) and indoor exposures (Wilcox: W = 396.5, $p = 0.43$). With Groups A and B combined, we observed no effect of treatment location on butyrate producer read abundances (Wilcox: W = 1940, $p = 0.28$).

Aerobiome-associated taxa in nasal microbiomes

We identified 1098 bacterial taxa in the outdoor aerobiome samples and then constrained nasal microbiome analyses with only these taxa. 30-minute exposures had no effect on the percentage of aerobiome taxa in nasal samples in either outdoor (t-test: t = 0.331, df = 114.77, $p = 0.74$; **Figure 3A**) or indoor treatments (W = 1740, $p = 1$; **Figure 3D**). 30-minute outdoor exposures had no effect on the community composition of aerobiome taxa in nasal samples in either group A (Adonis PERMANOVA: F = 0.761, $R^2 = 0.014$, $p = 0.96$; **Figure 3B**) or group B (Adonis PERMANOVA: F = 0.861, $R^2 = 0.015$, $p = 0.77$; **Figure 3C**), and 30-minute indoor exposures had no effect on the community composition of aerobiome taxa in nasal samples taxa in either group A (Adonis PERMANOVA: F = 0.862, $R^2 = 0.015$, $p = 0.80$; **Figure 3E**) or group B (Adonis PERMANOVA: F = 0.704, $R^2 = 0.013$, $p = 0.99$; **Figure 3F**).

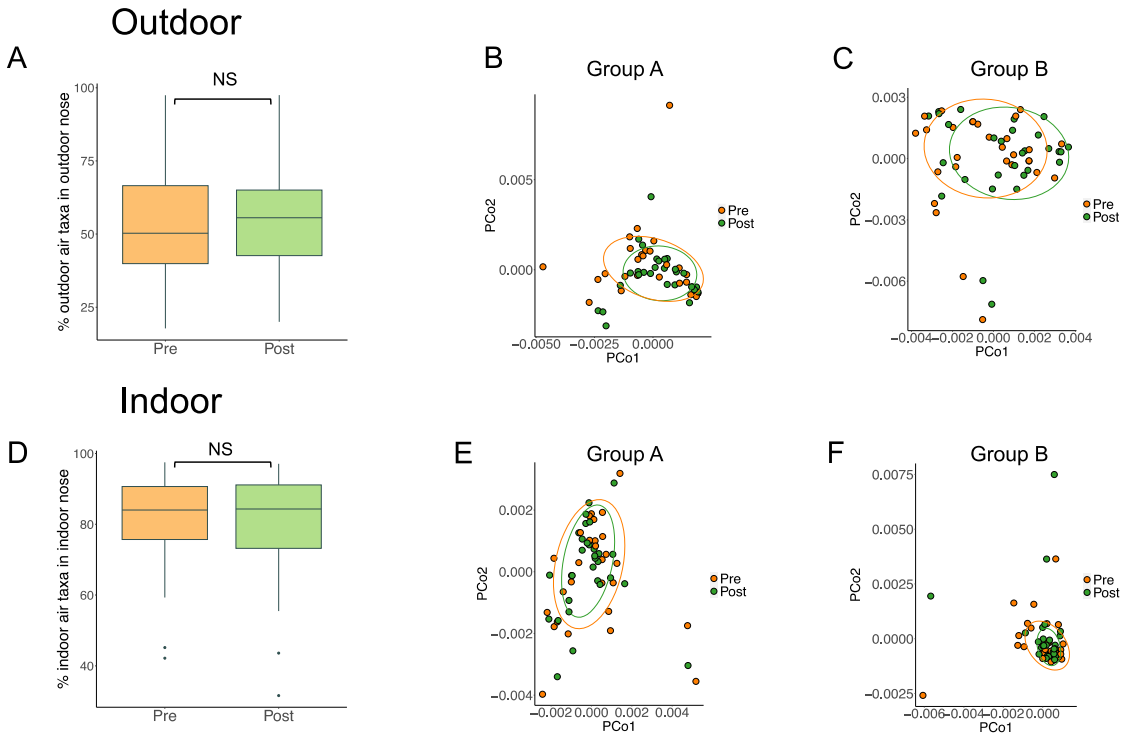


Figure 3. (A) Boxplots of the percentage of outdoor air taxa that were found in the nose (y-axis) before (Pre) and after (Post) outdoor exposure. Boxes show the median and interquartile range, while whiskers extend to the remaining range of data. (B-C) Principal coordinates analysis based on centred-log ratio compositional abundance data of only aerobiome-associated taxa found in the nose, displaying variation in community composition before (Pre) and after (Post) outdoor exposure for groups A (panel B) and B (panel C). (D) Boxplots of the percentage of outdoor air taxa that were found in the nose (y-axis) before (Pre) and after (Post) outdoor exposure. Boxes show the median and interquartile range, while whiskers extend to the remaining range of data. (E-F) Principal coordinates analysis based on centred-log ratio compositional abundance data of only aerobiome-associated taxa found in the nose, displaying variation in community composition before (Pre) and after (Post) outdoor exposure for groups A (panel E) and B (panel F).

Time-series effects on nasal microbiome characteristics

Participant had a strong effect on nasal bacterial communities from Day 1 to Day 8 (Adonis PERMANOVA: $F = 3.667$, $R^2 = 0.382$, $p = 0.001$; **Figure S1**). However, time had no effect on post-exposure group nasal bacterial community composition (**Figure 4A-D**). Group homogeneity (beta dispersion) also did not change from Day 1 to Day 8 (ANOVA: $F = 1.147$, $p = 0.29$).

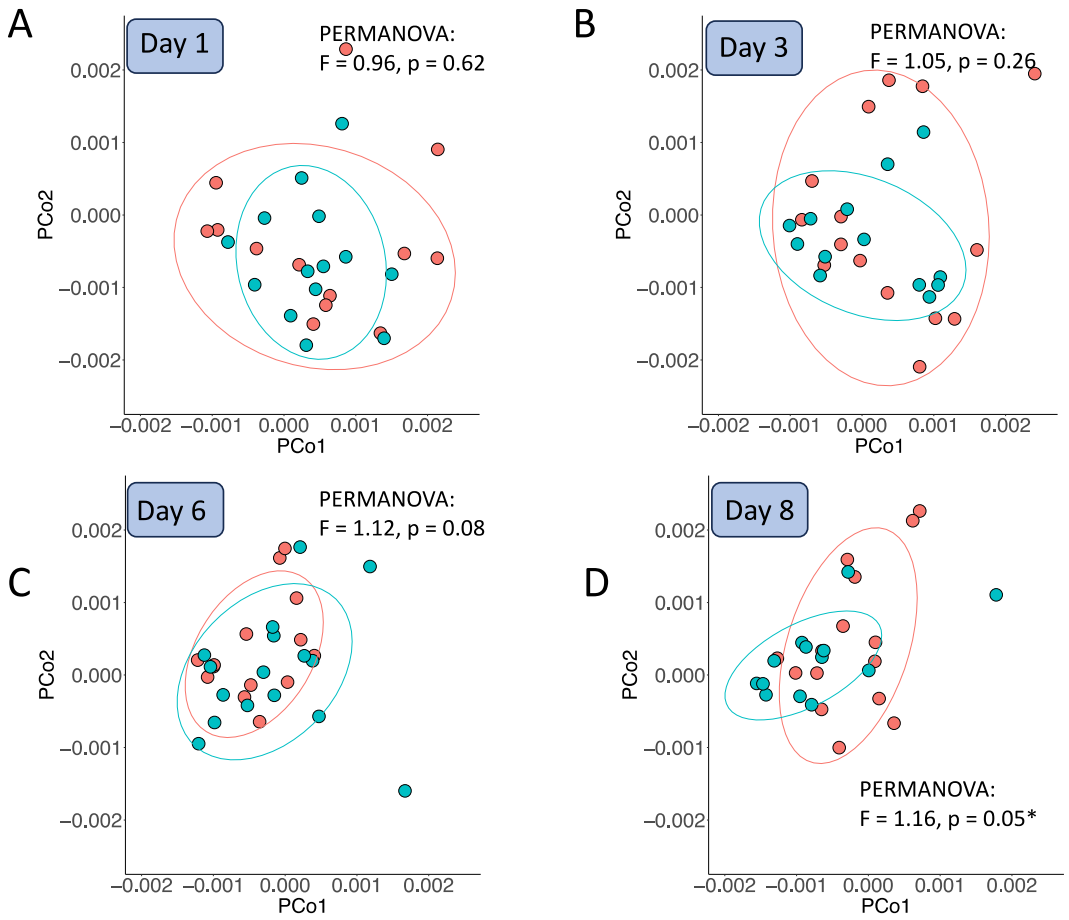


Figure 4. Display of time-series effects on post-exposure community composition between group A and group B, using principal coordinates analysis based on centred-log ratio compositional abundance data for days 1 (panel A), 3 (panel B), 6 (panel C), and 8 (panel D). Red points and ellipses are Group A. Blue points and ellipses are Group B. Outliers were removed on Days 1, 3, and 8.

Time had no effect on alpha diversity for group A (repeated measures ANOVA: ges = 0.084, *p* = 0.30) or group B (repeated measures ANOVA: ges = 0.192, *p* = 0.16; **Figure 5A**). Group B showed a time effect on Gammaproteobacteria diversity, with significantly reduced diversity from Day 1 post to Day 8 post (repeated measures ANOVA: ges = 0.344, *p* = 0.002; **Figure 5B**), but showed no effect on Group A (repeated measures ANOVA: ges = 0.082, *p* = 0.31). Time had no effect on the sum of relative abundances of butyrate-producing bacteria for group A (repeated measures ANOVA: ges = 0.119, *p* = 0.14) or B (repeated measures ANOVA: ges = 0.2, *p* = 0.24), although time had a weak effect on increasing read abundances of butyrate-producing bacteria from Day 1 post to Day 8 post in group B (repeated measures ANOVA: ges = 0.167, *p* = 0.058; **Figure 5C**).

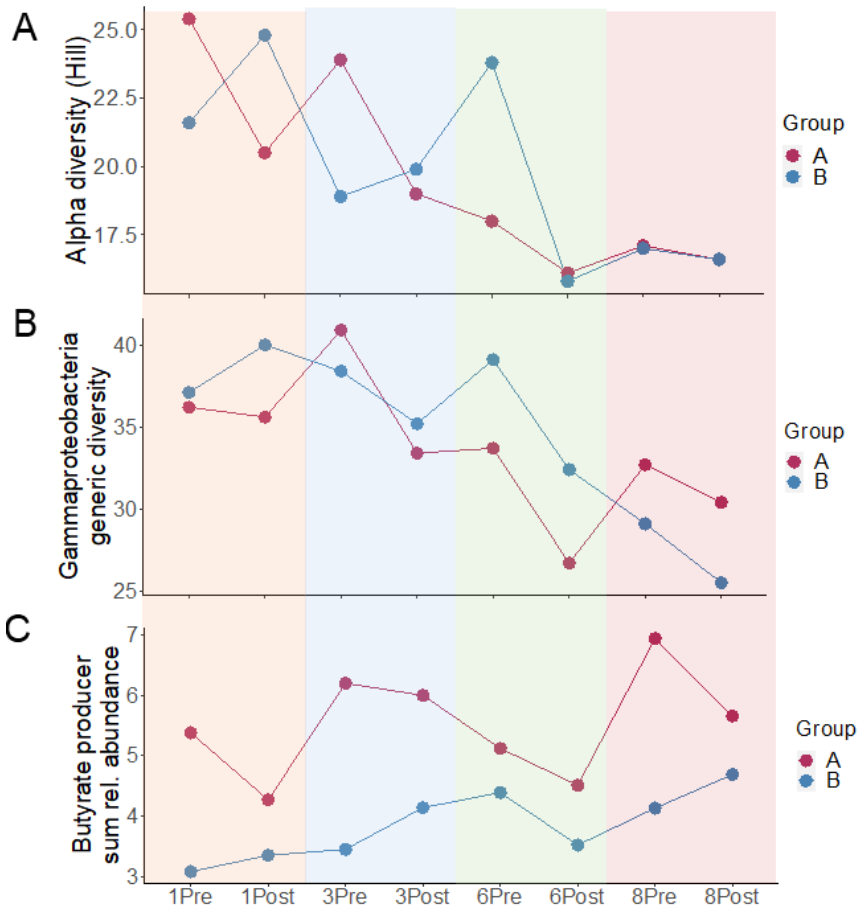


Figure 5. Line plots showing pre- versus post-exposure measures of human health-associated bacterial characteristics in nasal samples of groups A and B across the trial period: (A) alpha diversity (Hill numbers), (B) Gammaproteobacterial generic diversity, and (C) sums of relative abundances of butyrate-producing bacteria.

DISCUSSION

We ran a short-term greenspace cross-over exposure trial of a Māori cohort and showed that this exposure had little effect on nasal microbiomes. This low responsiveness of the nasal microbiome was following repeated 30-minute passive exposures to an outdoor nature park. Location, participant, and time had weak or no effect on the nasal microbiome alpha diversity, community composition, aerobiome taxa present in nasal samples, and health-associated bacterial groups. Overall, our results contrast with an earlier study that reported changes in nasal microbiomes after greenspace exposure (Selway et al., 2020). We suggest that nasal microbiomes are relatively stable over short periods of passive greenspace exposure, and 30 minutes of this passive exposure (i.e., walking in greenspaces) does not result in notable and/or consistent changes in the nasal bacterial

communities of participants. Our work raises important questions about the types of activities and duration of exposure to greenspaces required to result in meaningful changes to the nasal microbiome.

Indoor and outdoor aerobiomes were similar and had higher alpha diversity than nasal microbiomes

We found that overall aerobiomes had higher alpha diversity than nasal microbiomes. This is consistent with the findings from Selway et al. (2020), where outdoor air samples had higher alpha diversities than nasal samples. To our knowledge, no previous studies have compared the aerobiome alpha diversity of indoor office and urban greenspace environments. However, several studies have compared outdoor aerobiomes with indoor aerobiomes within other types of buildings. Nunez et al. (2022) found that 88% of bacterial and fungal bioaerosols were shared among both outdoor and indoor (small classroom) aerobiomes, and that the alpha diversity (Shannon index) was similar between the outdoor and indoor air samples. Chen et al. (2024) also found that the dominant bacterial species were similar across indoor and outdoor air at residential houses. Likewise, our findings showed no difference between office and amenity park aerobiome alpha diversity; however, we had only seven air samples (three outdoor and four indoor), which likely limited our power to detect an effect. Recent studies have placed value on urban greenspaces and natural outdoor locations as environmental reservoirs of immunoregulatory biodiversity for urban residents (Robinson et al., 2021; Roslund et al., 2020). However, future direct comparisons of indoor and outdoor aerobiomes across a range of built environments and outdoor settings are needed to establish the conditions that may drive potential health-promoting exposure effects.

Short greenspace exposures had little effect on nasal microbiomes

We found no clear effects of the 30-minute exposures on nasal microbiome alpha diversity or community composition. Aforementioned studies (Chen et al., 2024; Núñez & García, 2022) found that aerobiomes of adjacent indoor and outdoor locations tended to be similar, possibly due to mixing of air through mechanical ventilation systems. It may be plausible that the park selected for

the study had similar aerobiome characteristics to the air circulating into the office due to physical proximity. Thus, future studies could compare outdoor activities at a location with further distance fro the study's indoor location.

Even when filtering the microbial taxa to just particular health-associated bacterial groups (i.e., Gammaproteobacteria, butyrate-producing bacteria), the only notable effects were a reduction in generic diversity of Gammaproteobacteria and an increase in butyrate-producing bacterial read abundances in group B across the trial period. Roslund et al. (2020) recently found that generic diversity of Gammaproteobacteria on the skin of children associated with shifts in blood plasma markers TGF-β1 and interleukin-17 toward an anti-inflammatory profile. Our observed reduction in generic diversity of Gammaproteobacteria and an increase in butyrate-producing bacterial read abundances may be part of normal temporal bacterial variability (Vandeputte et al., 2021) or could have been driven by an unmeasured factor. However, since so few studies have generated data directly comparable to ours, the capacity to compare our findings with other studies is limited.

Exposure times

Our trial ran for eight days, with four 30-minute exposure events across these days. We found only minimal changes in nasal microbiome characteristics after each exposure. Our 30-minute exposure length was intended to represent a typical nature exposure of, for example, going for a walk in a park during a lunch break or walking a pet. Similar human exposure trials are limited, but some provide noteworthy discussion. In a study with two or three participants spending time in urban greenspaces, Selway et al. (2020) found skin and nasal microbiome changes, but participants performed activities that encouraged more direct interaction with soils and/or vegetation and utilised ca. 1.5 hour exposure periods. Roslund et al. (2020) added biodiverse forest floor and sod into daycare centres, then found changes in the skin and gut microbiomes of participant children (3-5 years old) over 28 days with approximately 1.5 hour daily exposure periods. Lai et al. (2017) examined the exposure impacts of academic mouse researchers working in the dirty cage wash area on nasal and skin

microbiomes. Their exposure period was a single 8-hour shift, and they found no significant change in the nasal microbiome between pre- and post-shift samples.

Longer exposure periods also show distinct changes in nasal microbiomes. The nasal microbiome of pig farmers who stopped occupational exposure to pig farms for 50 days shows a return toward non-farmer microbiome characteristics (Kraemer et al., 2021). Studies assessing the effects of land cover surrounding a person's home on their skin microbiome are able to integrate much longer exposure periods to show effects on residents' microbiomes. For example Hanski et al. (2012) assessed the influence of living near biodiversity and found notable effects on the bacterial classes in the skin. Thus, longer and/or repeated exposure periods plus more direct exposure (e.g., handling soils) may be required to elicit changes in nasal and skin microbiomes. Future urban greenspace research should further examine the effect of different activities (e.g., passive walking as in our study, direct handling of microbially-rich ecosystem components such as soil), durations (e.g., short 30-minute periods as in our study, longer and/or repeated short exposures) as well as adjacency and ecological quality of greenspaces on causing changes to human nasal microbiomes.

Individual participant nasal microbiome stability

We showed a relatively stable participant nasal microbiome over our study period, with strong between-subject effects found on all days. This finding corroborates with Costello et al. (2012) who described how the host shapes the microbiota through environmental selection processes. Indeed, Biswas et al. (2015) found that interpersonal differences accounted for more variation between participant nasal microbiomes than sampling location within the nasal cavity or disease status. We show that the composition of an individual's nasal microbiome appeared to change over the eight-day period, but not in ways that could be explained by our environmental exposure treatments. Our groups rotated through the same two sites, with similar exposures to the associated aerobiomes. The stability of between-subject microbiome diversity provides additional evidence that more direct,

longer, and/or more frequent exposure is necessary for environmental exposures to overcome other host selection pressures to modulate an individual's nasal microbiome.

Conclusions

Spending time in urban greenspaces can provide a person with exposure to outdoor aerobiomes that may have health-beneficial properties, such as by providing exposure to high bacterial diversity (Robinson et al., 2021; Roslund et al., 2020). Our study utilised pre- and post-exposure bacterial data to identify changes in the nasal microbiome following 30-minute walks in an outdoor urban park. We observed stability of the alpha diversity, community composition, and abundances of specific health-associated bacterial groups across exposure periods and across the trial period. Between-subject differences in nasal microbiomes were maintained during the trial period, although some evidence indicated a reduction in the diversity of Gammaproteobacteria and an increase in butyrate producing taxa. Our results suggest that 30 minutes of passive exposure to greenspaces provides insufficient aerobiome exposure to results in changes in nasal bacterial diversity and communities. Indigenous initiatives, which are driven by Indigenous knowledge and emphasise cultural connection as a motivator, could benefit from the expanding collection of microbiome data to better understand the complex (and holistic) relationship between health and the environment. Our study demonstrates the need for future human exposure trials investigating urban greenspace health benefits to examine the types of activity and duration of exposure.

Acknowledgements

We would like to acknowledge the contributions by Te Arawa Whānau Ora for their participation, enthusiasm, and on-site leadership. Special acknowledgement goes to the laboratory team members at Auckland University of Technology who lent their time, resources, and expertise during the field work and laboratory portions of the project. We would also like to acknowledge Christian Cando-Dumancela for his expertise and assistance in preparing field work resources. This work was

supported by funding from the Flinders Foundation and a Project Grant from the Health Research Council of New Zealand.

Data uploads

All study data and custom R code is available at Figshare at the following doi:

10.6084/m9.figshare.24993471

(Note to Reviewers: We have reserved the DOI mentioned above which will be published upon acceptance; however, for review purposes please see this private figshare link:

https://figshare.com/s/78d56add67e2a92db9fb)

Author contributions

J.B., I.W., D.H., C.L., and M.B. designed and planned the project. J.B., C.L., and M.B. analysed the data, with input from I.W. and D.H. J.B. wrote the initial manuscript, with subsequent revision input from all authors.

GENERAL DISCUSSION

Overview

Bacteria are essential to human life. They colonise the human body and perform a multitude of critical functions for health. Butyrate-producing bacteria have key roles in human health, but their abundances are decreased in a range of human health conditions and the sources of these gut-associated bacteria are unclear. After birth, bacterial colonisation of the human body comes primarily from the environment. Outdoor environments may be compelling potential sources of commensal bacteria that could supplement the human microbiome, but little is known about butyrate producer abundances in outdoor environments, nor about ecological conditions that influence their abundances. Thus, this thesis endeavours to elucidate (a) where and how people can be exposed exposure to health-associated butyrate-producing bacteria, and (b) how urban greenspace managers and public health experts can manage soil conditions to promote butyrate producer abundances.

Thesis synthesis

This thesis characterises outdoor environmental butyrate-producing bacteria and ecological conditions that influence their abundances and potential human exposures. My methodologies included in-silico analyses of global samples with shotgun metagenomic data and Australian soil samples with 16S rRNA amplicon data, examination of eDNA from soil and air samples from sports fields and nature parks across metropolitan Adelaide, South Australia together with soil physicochemical data and soil butyrate measurements from these same sites, and a human exposure trial. I show that butyrate-producing bacteria are found across the world in a broad variety of substrates, and that distinct ecological conditions influence their abundances. I also show that soils in sports fields had higher levels of soil butyrate and butyrate production genes than nature parks. I identified that soil physicochemical parameters can influence the airborne bacterial communities (aerobiome) at 0.5 m height. I also found that short-term outdoor exposures (i.e., walking through a park for 30 minutes) had little effect on the nasal microbiomes of participants. Thus, sports fields

appear to have conditions that increase the potential for transfer of butyrate-producing bacteria to people, and exposures of longer duration (e.g., >30 minutes of exposure) or more direct contact with ecological elements (e.g., putting hands in the soil) may be necessary for a notable shift in the nasal microbiome.

Abundances of butyrate-producing bacterial reads across environments

The literature review found little to no data about the abundances of butyrate-producing bacteria in outdoor environments. Instead, most butyrate producer research has characterised these bacteria inside the body of humans and non-human animals. Therefore, in Chapter 2 I studied their abundances throughout the world, examining as many substrates and outdoor locations as possible. Environments with anaerobicity (e.g., animal guts, industrial anaerobic digesters, and plant rhizospheres) associated with higher read abundances of butyrate producers than aerobic environments (e.g., human skin, industrial activated sludge, and leaf surfaces). Thus, the requirement of anaerobicity for butyrate production appears to be maintained across a range of substrates and locations. Soils in seasonally-productive sandy croplands and urban hinterlands had ecological conditions (e.g., geographical, meteorological, and biochemical) that supported greater butyrate producer read abundances. Identification of such differences in butyrate producer read abundances across locations shows that human exposure to these bacteria is potentially unequal. Social equity may become an important consideration as research continues to characterise butyrate-producing bacteria in outdoor microbial communities.

Ecological associations with butyrate-producing bacteria

Beyond quantifying butyrate producer read abundances, I also sought to understand the ecological influences on outdoor butyrate producers, an equally substantial knowledge gap. Because I have an interest in ecological restoration, I investigated the influence of woody plant species diversity in urban greenspaces on soil and air bacterial communities in several chapters. Woody plant diversity had less influence on aerobiomes at 0.5 m height and soil butyrate producers than other variables

such as rainfall and soil physicochemical parameters. Soil iron levels influenced the aerobiome and soil microbiome in three of the data chapters. This led to a more in-depth discussion on soil iron and butyrate producers in chapter 4, where I discussed the potential bidirectional association of soil iron and the butyrate production gene *buk*. Because soil iron appears to have a strong effect on soil butyrate producing bacteria, future ecological studies should closely examine their relationship. It may be possible that findings in soil butyrate producer research could also inform research on the dynamics of butyrate producing bacteria in the human body.

Soil butyrate production

I led the development of a novel method for quantifying soil butyrate concentrations to associate urban greenspace type and soil physicochemical parameters with soil butyrate levels. Several variables had a different level of influence on soil butyrate levels than on the read abundances of butyrate-producing bacteria. For example, soil moisture and organic carbon associated with soil butyrate concentrations but did not associate with butyrate production gene abundances. My findings suggest that butyrate producers may not be metabolically active until the necessary conditions are met. This could also have important implications for human health research, where the read abundances of butyrate producers in the faeces might not associate with faecal butyrate concentrations. Thus, future studies should continue examining whether butyrate producer abundances or faecal butyrate levels hold greater influences on health conditions.

Sports fields as a unique ecosystem

Chapters 3 and 4 compared sports fields with nature parks using vegetation as the stratification criterion. However, we found that sports fields had a broad range of ecological characteristics that substantially varied from nature parks. For example, sports fields had higher soil moisture, almost twice the level of soil iron, and higher aerobiome alpha diversity than nature parks, plus greater levels of soil butyrate. In addition, sports fields are generally more intensively managed, with regular watering and application of fertilisers to maintain a safe pitch for sports. Thus, sports fields

management techniques appear to lead to unique qualities from nature parks. As such, results from sports fields (e.g., the effect of soil moisture on butyrate concentrations) may not be generalisable to other types of urban greenspace land covers. We included only 11 sports field sites in our studies; thus future butyrate-related studies should include a greater number of sports field sites to discern detailed patterns and associations.

Limitations and future recommendations

The research in this thesis revealed four key limitations. First, while human health studies have linked diseases with a reduction in butyrate producers, it remains unclear whether the reduction contributed to the disease or were the result of the disease, directly (e.g. via inflammatory mediators) or indirectly (e.g., via changes in the diet due to the disease). Supplementation of butyrate producers through interventions could have value regardless of the direction of influence, but elucidating the direction of influence would be a worthwhile goal of future research.

Second, the classification of a particular bacterial taxon as a butyrate producer is challenging. Metagenomic analyses have revealed genes for butyrate production in a wide range of bacterial species, but culture studies to confirm butyrate production capacity have been performed on only a few species. In addition, the abundances of genes for butyrate metabolism vary widely among putative butyrate producer species. Indeed, with at least 20 enzymes in the butyrate production metabolic pathway, classification of a bacterium with <20 butyrate production genes poses challenges. Furthermore, terminal butyrate synthesis enzymes are often reversible, and the conditions that drive metabolic reactions toward butyrate production versus butyrate oxidation remain poorly resolved. Therefore, expansion of culture studies to include many more putative butyrate-producing bacteria would provide valuable contributions to the field.

Third, many metagenomic studies have associated a particular disease with both a decrease in several butyrate producer species and a concomitant increase in others. The interpretation of such

mixed treatment effects could be improved by quantifying the relative contributions of each species to the butyrate pool; however, without thorough culture studies, this level of resolution is not currently available. Thus, future research should include more extensive culture studies to better understand how specific butyrate-producing taxa influence the butyrate pool.

Fourth, microbiome analyses have inherent limitations that could have affected the outcomes. Butyrate-producing bacteria are usually strictly anaerobic, and many of them form protective endospores on exposure to air. Commercial DNA extraction kits may vary in their capacity to disrupt endospores. We utilised one particular kit brand, which could have introduced bias into our results. In addition, reference databases to which DNA sequences are matched vary in their annotations. Our Qiime2 pipeline uses Silva for 16S rRNA amplicon annotation, and Kraken2 uses NCBI taxonomy for shotgun metagenomic read annotation. Our choices of reference databases could have introduced bias into our results. Furthermore, DNA contamination poses a routine issue in microbiome studies. While we made efforts to minimise contamination at every step, including identifying and removing potential contaminants in silico, contamination could have occurred.

Our findings also included two negative results that also represent study limitations. First, we found that outdoor greenspace ecological components may have different relative contributions to the aerobiome at different heights. For example, several aerobiome studies have shown that vegetation complexity has a significant influence on the aerobiome. However, Chapter 3 utilised a sampling height at 0.5 m, which is lower than the other studies, and we found that soil physicochemical conditions influenced the aerobiome more than woody plant species diversity. It is possible that soil confers greater contributions to aerobiomes that are closer to the ground. Future studies should examine the influence of soil physicochemical parameters on the aerobiome diversity and composition at varying heights.

Second, in Chapter 5 I expected that people taking a 30-minute walk through an amenity grassland park would receive a substantial "dose" of airborne bacteria into their noses. This effect did not occur. It appeared that 30-minute passive exposures are not long enough to create notable change in the nasal microbiome. Future exposure trials should compare short-term (i.e., minutes), medium-term (i.e., hours or days), and long-term (i.e., weeks or months, such as living adjacent to an urban greenspace) exposures, along with comparing active (e.g., digging in soil) and passive (e.g., walking in a park) treatment activities, to determine how exposure affects nasal microbiome composition.

Full publication list

Brame JE, Liddicoat C, Abbott CA, Breed MF. The potential of outdoor environments to supply beneficial butyrate-producing bacteria to humans. 2021. *Science of the Total Environment*. doi: 10.1016/j.scitotenv.2021.146063.

Brame JE, Liddicoat C, Abbott CA, Edwards RE, Gauthier NE, Robinson JA, Breed MF. Towards the biogeography of butyrate-producing bacteria. 2023. *bioRxiv* (under review in *Ecology and Evolution*). doi: 10.1101/2022.10.07.510278.

Brame JE, Liddicoat C, Abbott CA, Cando-Dumancela C, Robinson JA, Breed MF. Urban greenspace aerobiomes are shaped by soil conditions and land cover type. 2024. *bioRxiv* (under review in *Microbial Ecology*), doi: 10.1101/2024.01.12.575340.

Brame JE, Liddicoat C, Abbott CA, Cando-Dumancela C, Robinson JA, Breed MF. Urban sports fields support higher levels of soil butyrate and butyrate-producing bacteria than urban nature parks. (in final stages of co-author review prior to submission to *ISME Journal*).

Brame JE, Warbrick I, Heke D, Liddicoat, C, Breed MF. Short-term passive greenspace exposures have little effect on nasal microbiomes: a cross-over exposure study of a Māori cohort. 2024. *bioRxiv* (under review in *Environmental Research*). doi: 10.1101/2024.01.17.576148

Fickling N, Abbott C, **Brame JE**, Cando-Dumancela C, Liddicoat C, Robinson JM, Breed MF. Light-dark cycles may influence in situ soil bacterial networks and diurnally-sensitive taxa. *Ecology & Evolution* (accepted December 2023).

Newman N, Abbott C, **Brame JE**, Cando-Dumancela C, Fickling N, Liddicoat C, Breed MF. Childcare centre soil microbiomes are influenced by substrate type and surrounding vegetation condition. 2024. SSRN (under review in *Science of the Total Environment*). doi: 10.2139/ssrn.4681385.

Liddicoat C, Edwards R, Roach M, Robinson J, Wallace K, Barnes A, **Brame JE**, Heintz-Buschart A, Cavagnaro T, Dinsdale E, Doane M, Eisenhauer N, Mitchell G, Rai B, Ramesh S, Breed MF. Bioenergetic mapping of 'healthy microbiomes' via compound processing potential imprinted in gut and soil metagenomes. 2023. *bioRxiv* (under review in *Proceedings of the National Academy of Sciences*). doi: 10.1101/2023.11.05.565728.

References from this document (exclusive of published and submitted article references)

Adams, R. I., Bateman, A. C., Bik, H. M., & Meadow, J. F. (2015). Microbiota of the indoor environment: a meta-analysis. *Microbiome*, *3*, 1-18.

Al-Hinai, M. A., Jones, S. W., & Papoutsakis, E. T. (2015). The Clostridium sporulation programs: diversity and preservation of endospore differentiation. *Microbiology and Molecular Biology Reviews*, *79*(1), 19-37. https://doi.org/10.1128/mmbr.00025-14

Alfven, T., Braun-Fahrlander, C., Brunekreef, B., von Mutius, E., Riedler, J., & Scheynius, A. (2007). Allergic diseases and atopic sensitisation in children related to farming and anthroposophic lifestyle - the PARSIFAL study [Correction]. *Allergy*, *62*(4), 455-455. https://doi.org/10.1111/j.1398-9995.2007.01361.x

Andreas, N. J., Kampmann, B., & Le-Doare, K. M. (2015). Human breast milk: A review on its composition and bioactivity. *Early Human Development*, *91*(11), 629-635.

Appert, O., Garcia, A. R., Frei, R., Roduit, C., Constancias, F., Neuzil-Bunesova, V., Ferstl, R., Zhang, J., Akdis, C., & Lauener, R. (2020). Initial butyrate producers during infant gut microbiota development are endospore formers. *Environmental Microbiology*, *22*(9), 3909-3921. https://doi.org/10.1111/1462-2920.15167

Bartlett, A., Padfield, D., Lear, L., Bendall, R., & Vos, M. (2022). A comprehensive list of bacterial pathogens infecting humans. *Microbiology*, *168*(12), 001269.

Baruch, Z., Liddicoat, C., Laws, M., Marker, L. K., Morelli, H., Yan, D., Young, J. M., & Breed, M. F. (2020). Characterising the soil fungal microbiome in metropolitan green spaces across a vegetation biodiversity gradient. *Fungal Ecology*, *47*, 100939.

Baxter, N. T., Schmidt, A. W., Venkataraman, A., Kim, K. S., Waldron, C., & Schmidt, T. M. (2019). Dynamics of Human Gut Microbiota and Short-Chain Fatty Acids in Response to Dietary Interventions with Three Fermentable Fibers [Article]. *Mbio*, *10*(1), 13, Article e02566-18. https://doi.org/10.1128/mBio.02566-18

Beaumont, M., Paes, C., Mussard, E., Knudsen, C., Cauquil, L., Aymard, P., Barilly, C., Gabinaud, B., Zemb, O., Fourre, S., Gautier, R., Lencina, C., Eutamene, H., Theodorou, V., Canlet, C., & Combes, S. (2020). Gut microbiota derived metabolites contribute to intestinal barrier maturation at the suckling-to-weaning transition [Article; Early Access]. *Gut Microbes*, *11*(5), 1268-1286. https://doi.org/10.1080/19490976.2020.1747335

Bedford, A., & Gong, J. (2018). Implications of butyrate and its derivatives for gut health and animal production. *Anim Nutr*, *4*(2), 151-159. https://doi.org/10.1016/j.aninu.2017.08.010

Belbin, L. (2011). The Atlas of Living Australia's spatial portal. Proceedings of the Environmental Information Management Conference (EIM 2011), Santa Barbara.

Benito, M. (2021). spatialRF: Easy spatial regression with random forest. *R package version*, *1*(0).

Bissett, A., Fitzgerald, A., Meintjes, T., Mele, P. M., Reith, F., Dennis, P. G., Breed, M. F., Brown, B., Brown, M. V., & Brugger, J. (2016). Introducing BASE: the Biomes of Australian Soil Environments soil microbial diversity database. *Gigascience*, *5*(1), s13742-13016-10126-13745.

Biswas, K., Hoggard, M., Jain, R., Taylor, M. W., & Douglas, R. G. (2015). The nasal microbiota in health and disease: variation within and between subjects. *Frontiers in Microbiology*, *6*, 134.

Blum, W. E., Zechmeister-Boltenstern, S., & Keiblinger, K. M. (2019). Does soil contribute to the human gut microbiome? *Microorganisms*, *7*(9), 287.

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., & Asnicar, F. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, *37*(8), 852-857.

Bowers, R. M., McLetchie, S., Knight, R., & Fierer, N. (2011). Spatial variability in airborne bacterial communities across land-use types and their relationship to the bacterial communities of potential source environments. *The ISME Journal*, *5*(4), 601-612.

Brahe, L. K., Astrup, A., & Larsen, L. H. (2013). Is butyrate the link between diet, intestinal microbiota and obesity-related metabolic diseases? [Review]. *Obesity Reviews*, *14*(12), 950-959. https://doi.org/10.1111/obr.12068

Brame, J. E., Liddicoat, C., Abbott, C. A., & Breed, M. F. (2021). The potential of outdoor environments to supply beneficial butyrate-producing bacteria to humans. *Science of the Total Environment*, *777*, 146063.

Brame, J. E., Liddicoat, C., Abbott, C. A., Edwards, R. A., Robinson, J. M., Gauthier, N. E., & Breed, M. F. (2022). Towards the biogeography of butyrate-producing bacteria. *bioRxiv*, 2022.2010. 2007.510278.

Braniste, V., Al-Asmakh, M., Kowal, C., Anuar, F., Abbaspour, A., Tóth, M., Korecka, A., Bakocevic, N., Ng, L. G., & Kundu, P. (2014). The gut microbiota influences blood-brain barrier permeability in mice. *Science Translational Medicine*, *6*(263), 263ra158-263ra158. https://doi.org/10.1126/scitranslmed.3009759

Breed, M. F., Cross, A. T., Wallace, K., Bradby, K., Flies, E., Goodwin, N., Jones, M., Orlando, L., Skelly, C., & Weinstein, P. (2020). Ecosystem Restoration: A Public Health Intervention. *EcoHealth*, 1-3. https://doi.org/10.1007/s10393-020-01480-1

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5-32.

Brink, L. R., Mercer, K. E., Piccolo, B. D., Chintapalli, S. V., Elolimy, A., Bowlin, A. K., Matazel, K. S., Pack, L., Adams, S. H., & Shankar, K. (2020). Neonatal diet alters fecal microbiota and metabolome profiles at different ages in infants fed breast milk or formula. *The American Journal of Clinical Nutrition*, *111*(6), 1190-1202.

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

Browne, H. P., Forster, S. C., Anonye, B. O., Kumar, N., Neville, B. A., Stares, M. D., Goulding, D., & Lawley, T. D. (2016). Culturing of 'unculturable'human microbiota reveals novel taxa and extensive sporulation. *Nature*, *533*(7604), 543-546. https://doi.org/10.1038/nature17645

Bunesova, V., Lacroix, C., & Schwab, C. (2018). Mucin Cross-Feeding of Infant Bifidobacteria and Eubacterium hallii [Article]. *Microbial Ecology*, *75*(1), 228-238. https://doi.org/10.1007/s00248-017-1037-4

Cantu, V. A., Sadural, J., & Edwards, R. (2019). PRINSEQ++, a multi-threaded tool for fast and efficient quality control and preprocessing of sequencing datasets. *PeerJ Preprints*, *7*, e27553v27551.

Chen, I.-M. A., Chu, K., Palaniappan, K., Pillay, M., Ratner, A., Huang, J., Huntemann, M., Varghese, N., White, J. R., & Seshadri, R. (2019). IMG/M v. 5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Research*, *47*(D1), D666-D677.

Chen, I.-M. A., Chu, K., Palaniappan, K., Ratner, A., Huang, J., Huntemann, M., Hajek, P., Ritter, S., Varghese, N., & Seshadri, R. (2021). The IMG/M data management and analysis system v. 6.0: new tools and advanced capabilities. *Nucleic Acids Research*, *49*(D1), D751-D763.

Chen, J., Zhao, K.-N., & Vitetta, L. (2019). Effects of intestinal microbial–elaborated butyrate on oncogenic signaling pathways. *Nutrients*, *11*(5), 1026.

Chen, S., Wang, W., Xu, W., Wang, Y., Wan, H., Chen, D., Tang, Z., Tang, X., Zhou, G., & Xie, Z. (2018). Plant diversity enhances productivity and soil carbon storage. *Proceedings of the National Academy of Sciences*, *115*(16), 4027-4032.

Chen, Y., Liang, Z., Li, G., & An, T. (2024). Indoor/Outdoor airborne microbiome characteristics in residential areas across four seasons and its indoor purification. *Environment International*, *190*, 108857.

Chen, Z., Xie, Y., Zhou, F., Zhang, B., Wu, J., Yang, L., Xu, S., Stedtfeld, R., Chen, Q., & Liu, J. (2020). Featured Gut Microbiomes Associated With the Progression of Chronic Hepatitis B Disease. *Frontiers in Microbiology*, *11*, 383.

Chong, H. J., D'amato, G., & Rosário Filho, N. A. (2022). Impact of the environment on the microbiome. *Jornal de Pediatria*, *98*(suppl 1), 32-37.

Chowdhury, N. P., Kahnt, J., & Buckel, W. (2015). Reduction of ferredoxin or oxygen by flavin-based electron bifurcation in Megasphaera elsdenii. *The FEBS Journal*, *282*(16), 3149-3160.
Coleman, D., Crossley, D., Hendrix, P. (2004). *Fundamentals of Soil Ecology*. Elsevier Academic Press.
Conrad, R. (2020). Methane production in soil environments—anaerobic biogeochemistry and microbial life between flooding and desiccation. *Microorganisms*, *8*(6), 881.
Costello, E. K., Stagaman, K., Dethlefsen, L., Bohannan, B. J., & Relman, D. A. (2012). The application of ecological theory toward an understanding of the human microbiome. *Science*, *336*(6086), 1255-1262.
Cudahy, T., Caccetta, M., & Thomas, M. (2012). National ASTER Map of Australia. *Scale*, *1*, 1000000.
David, L. A., Maurice, C. F., Carmody, R. N., Gootenberg, D. B., Button, J. E., Wolfe, B. E., Ling, A. V., Devlin, A. S., Varma, Y., & Fischbach, M. A. (2014). Diet rapidly and reproducibly alters the human gut microbiome. *Nature*, *505*(7484), 559-563.
Davis, N. M., Proctor, D. M., Holmes, S. P., Relman, D. A., & Callahan, B. J. (2018). Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome*, *6*, 1-14.
Delgado-Baquerizo, M., Reith, F., Dennis, P. G., Hamonts, K., Powell, J. R., Young, A., Singh, B. K., & Bissett, A. (2018). Ecological drivers of soil microbial diversity and soil biological networks in the Southern Hemisphere. *Ecology*, *99*(3), 583-596.
Dierikx, T., Visser, D., Benninga, M., van Kaam, A., de Boer, N., de Vries, R., van Limbergen, J., & de Meij, T. (2020). The influence of prenatal and intrapartum antibiotics on intestinal microbiota colonisation in infants: A systematic review. *Journal of Infection*, *81*(2), 190-204. https://doi.org/10.1016/j.jinf.2020.05.002
Dominguez-Bello, M. G., Costello, E. K., Contreras, M., Magris, M., Hidalgo, G., Fierer, N., & Knight, R. (2010). Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proceedings of the National Academy of Sciences USA*, *107*(26), 11971-11975. https://doi.org/10.1073/pnas.1002601107
Dostal, A., Lacroix, C., Bircher, L., Pham, V. T., Follador, R., Zimmermann, M. B., & Chassard, C. (2015). Iron Modulates Butyrate Production by a Child Gut Microbiota In Vitro [Article]. *Mbio*, *6*(6), 12, Article e01453-15. https://doi.org/10.1128/mBio.01453-15
Durie, M. (2004). An indigenous model of health promotion. *Health Promotion Journal of Australia*, *15*(3), 181-185.
Ege, M. J., Mayer, M., Normand, A.-C., Genuneit, J., Cookson, W. O., Braun-Fahrländer, C., Heederik, D., Piarroux, R., & von Mutius, E. (2011). Exposure to environmental microorganisms and childhood asthma. *New England Journal of Medicine*, *364*(8), 701-709. https://doi.org/10.1056/NEJMoa1007302
Ellis, E. C., Gauthier, N., Goldewijk, K. K., Bird, R. B., Boivin, N., Díaz, S., Fuller, D. Q., Gill, J. L., Kaplan, J. O., & Kingston, N. (2021). People have shaped most of terrestrial nature for at least 12,000 years. *Proceedings of the National Academy of Sciences*, *118*(17).
Ellis, E. C., & Ramankutty, N. (2008). Putting people in the map: anthropogenic biomes of the world. *Frontiers in Ecology and the Environment*, *6*(8), 439-447.
Engel, S., & Susilo, A. (2014). Shaming and sanitation in Indonesia: a return to colonial public health practices? *Development and Change*, *45*(1), 157-178. https://doi.org/10.1111/dech.12075
Fierer, N., & Jackson, R. B. (2006). The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences*, *103*(3), 626-631.
Finegold, S. M., Lawson, P. A., Vaisanen, M.-L., Molitoris, D. R., Song, Y., Liu, C., & Collins, M. D. (2004). Anaerofustis stercorihominis gen. nov., sp. nov., from human feces. *Anaerobe*, *10*(1), 41-45.
Flandroy, L., Poutahidis, T., Berg, G., Clarke, G., Dao, M.-C., Decaestecker, E., Furman, E., Haahtela, T., Massart, S., & Plovier, H. (2018). The impact of human activities and lifestyles on the interlinked microbiota and health of humans and of ecosystems. *Science of the Total Environment*, *627*, 1018-1038. https://doi.org/10.1016/j.scitotenv.2018.01.288

Foster, J. A., Rinaman, L., & Cryan, J. F. (2017). Stress & the gut-brain axis: regulation by the microbiome. *Neurobiology of Stress*, *7*, 124-136.
Franchitti, E., Caredda, C., Anedda, E., & Traversi, D. (2022). Urban Aerobiome and Effects on Human Health: A Systematic Review and Missing Evidence. *Atmosphere*, *13*(7), 1148.
Fu, X., Liu, Z., Zhu, C., Mou, H., & Kong, Q. (2019). Nondigestible carbohydrates, butyrate, and butyrate-producing bacteria. *Critical Reviews in Food Science and Nutrition*, *59*(sup1), S130-S152. https://doi.org/10.1080/10408398.2018.1542587
García-Villalba, R., Giménez-Bastida, J. A., García-Conesa, M. T., Tomás-Barberán, F. A., Carlos Espin, J., & Larrosa, M. (2012). Alternative method for gas chromatography-mass spectrometry analysis of short-chain fatty acids in faecal samples. *Journal of Separation Science*, *35*(15), 1906-1913.
Gauthier, N., Ellis, E., & Goldewijk, K. K. (2021). Anthromes 12K DGG (V1) Full Dataset. *Harvard Dataverse*, *10*.
Ghosh, B., Lal, H., & Srivastava, A. (2015). Review of bioaerosols in indoor environment with special reference to sampling, analysis and control mechanisms. *Environment International*, *85*, 254-272.
Gilbert, J. A., Blaser, M. J., Caporaso, J. G., Jansson, J. K., Lynch, S. V., & Knight, R. (2018). Current understanding of the human microbiome. *Nature Medicine*, *24*(4), 392-400. https://doi.org/10.1038/nm.4517
Goldfarb, K. C., Karaoz, U., Hanson, C. A., Santee, C. A., Bradford, M. A., Treseder, K. K., Wallenstein, M. D., & Brodie, E. L. (2011). Differential growth responses of soil bacterial taxa to carbon substrates of varying chemical recalcitrance. *Frontiers in Microbiology*, *2*, 94.
Goslee, S. C., & Urban, D. L. (2007). The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, *22*, 1-19.
Guh, A. Y., & Kutty, P. K. (2018). Clostridioides difficile infection. *Annals of Internal Medicine*, *169*(7), ITC49-ITC64. https://doi.org/10.7326/aitc201810020
Güneralp, B., Reba, M., Hales, B. U., Wentz, E. A., & Seto, K. C. (2020). Trends in urban land expansion, density, and land transitions from 1970 to 2010: A global synthesis. *Environmental Research Letters*, *15*(4), 044015.
Haahtela, T. (2019). A biodiversity hypothesis. *Allergy*, *74*(8), 1445-1456. https://doi.org/10.1111/all.13763
Haas, K. N., & Blanchard, J. L. (2017). Kineothrix alysoides, gen. nov., sp. nov., a saccharolytic butyrate-producer within the family Lachnospiraceae. *International Journal of Systematic and Evolutionary Microbiology*, *67*(2), 402-410. https://doi.org/10.1099/ijsem.0.001643
Hald, S., Schioldan, A. G., Moore, M. E., Dige, A., Lærke, H. N., Agnholt, J., Bach Knudsen, K. E., Hermansen, K., Marco, M. L., & Gregersen, S. (2016). Effects of arabinoxylan and resistant starch on intestinal microbiota and short-chain fatty acids in subjects with metabolic syndrome: a randomised crossover study. *Plos One*, *11*(7), e0159223.
Hanski, I., von Hertzen, L., Fyhrquist, N., Koskinen, K., Torppa, K., Laatikainen, T., Karisola, P., Auvinen, P., Paulin, L., & Mäkelä, M. J. (2012). Environmental biodiversity, human microbiota, and allergy are interrelated. *Proceedings of the National Academy of Sciences*, *109*(21), 8334-8339.
Hickman, B., Kirjavainen, P. V., Täubel, M., De Vos, W. M., Salonen, A., & Korpela, K. (2022). Determinants of bacterial and fungal microbiota in Finnish home dust: Impact of environmental biodiversity, pets, and occupants. *Frontiers in Microbiology*, *13*, 1011521.
Hong, H. A., To, E., Fakhry, S., Baccigalupi, L., Ricca, E., & Cutting, S. M. (2009). Defining the natural habitat of Bacillus spore-formers. *Research in Microbiology*, *160*(6), 375-379. https://doi.org/10.1016/j.resmic.2009.06.006
Jimenez, M. P., DeVille, N. V., Elliott, E. G., Schiff, J. E., Wilt, G. E., Hart, J. E., & James, P. (2021). Associations between nature exposure and health: A review of the evidence. *International Journal of Environmental Research and Public Health*, *18*(9), 4790.
Johnston, A. E., Poulton, P. R., & Coleman, K. (2009). Soil organic matter: its importance in sustainable agriculture and carbon dioxide fluxes. *Advances in Agronomy*, *101*, 1-57.

Jost, T., Lacroix, C., Braegger, C. P., Rochat, F., & Chassard, C. (2014). Vertical mother-neonate transfer of maternal gut bacteria via breastfeeding [Article]. *Environmental Microbiology*, *16*(9), 2891-2904. https://doi.org/10.1111/1462-2920.12238

Jovel, J., Patterson, J., Wang, W., Hotte, N., O'Keefe, S., Mitchel, T., Perry, T., Kao, D., Mason, A. L., & Madsen, K. L. (2016). Characterization of the gut microbiome using 16S or shotgun metagenomics. *Frontiers in Microbiology*, *7*, 459.

Kahle, D. J., & Wickham, H. (2013). ggmap: spatial visualization with ggplot2. *R J.*, *5*(1), 144.

Kaisar, M. M., Pelgrom, L. R., van der Ham, A. J., Yazdanbakhsh, M., & Everts, B. (2017). Butyrate conditions human dendritic cells to prime type 1 regulatory T cells via both histone deacetylase inhibition and G protein-coupled receptor 109A signaling. *Frontiers in Immunology*, *8*, 1429.

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, *44*(D1), D457-D462.

Kassambara, A. (2023). *rstatix: Pipe-Friendly Framework for Basic Statistical Tests*. In https://rpkgs.datanovia.com/rstatix/

Kearney, S. M., Gibbons, S. M., Poyet, M., Gurry, T., Bullock, K., Allegretti, J. R., Clish, C. B., & Alm, E. J. (2018). Endospores and other lysis-resistant bacteria comprise a widely shared core community within the human microbiota. *The ISME Journal*, *12*(10), 2403-2416. https://doi.org/10.1038/s41396-018-0192-z

Kim, K.-H., Kabir, E., & Jahan, S. A. (2018). Airborne bioaerosols and their impact on human health. *Journal of Environmental Sciences*, *67*, 23-35.

Knox, E. G., Aburto, M. R., Tessier, C., Nagpal, J., Clarke, G., O'Driscoll, C. M., & Cryan, J. F. (2022). Microbial-Derived Metabolites Induce Actin Cytoskeletal Rearrangement and Protect Blood-Brain Barrier Function. *iScience*, 105648.

Kondo, M. C., Fluehr, J. M., McKeon, T., & Branas, C. C. (2018). Urban green space and its impact on human health. *International Journal of Environmental Research and Public Health*, *15*(3), 445.

Kraemer, J. G., Aebi, S., Hilty, M., & Oppliger, A. (2021). Nasal microbiota composition dynamics after occupational change in animal farmers suggest major shifts. *Science of the Total Environment*, *782*, 146842.

La Reau, A. J., & Suen, G. (2018). The Ruminococci: key symbionts of the gut ecosystem. *Journal of Microbiology*, *56*(3), 199-208. https://doi.org/10.1007/s12275-018-8024-4

Lai, H., Flies, E. J., Weinstein, P., & Woodward, A. (2019). The impact of green space and biodiversity on health. *Frontiers in Ecology and the Environment*, *17*(7), 383-390. https://doi.org/10.1002/fee.2077

Lai, P. S., Allen, J. G., Hutchinson, D. S., Ajami, N. J., Petrosino, J. F., Winters, T., Hug, C., Wartenberg, G. R., Vallarino, J., & Christiani, D. C. (2017). Impact of environmental microbiota on human microbiota of workers in academic mouse research facilities: An observational study. *Plos One*, *12*(7), e0180969.

Lang, J. M., Eisen, J. A., & Zivkovic, A. M. (2014). The microbes we eat: abundance and taxonomy of microbes consumed in a day's worth of meals for three diet types. *PeerJ*, *2*, e659.

Lee, A. C. K., Jordan, H. C., & Horsley, J. (2015). Value of urban green spaces in promoting healthy living and wellbeing: prospects for planning. *Risk Management and Healthcare Policy*, 131-137.

Lee, Y. J., Miyahara, T., & Noike, T. (2001). Effect of iron concentration on hydrogen fermentation. *Bioresource Technology*, *80*(3), 227-231.

Levine, U. Y., Looft, T., Allen, H. K., & Stanton, T. B. (2013). Butyrate-producing bacteria, including mucin degraders, from the swine intestinal tract. *Applied and Environmental Microbiology*, *79*(12), 3879-3881. https://doi.org/10.1128/aem.00589-13

Li, D. (2018). hillR: taxonomic, functional, and phylogenetic diversity and similarity through Hill Numbers. *J. Open Source Softw.*, *3*(31), 1041.

Li, R. W., & Li, C.-J. (2014). Enhancing butyrate biosynthesis in the gut for health benefits. In *Butyrate: Food Sources, Functions and Health Benefits* (pp. 1-23). Nova Science.

Li, W. V., & Li, J. J. (2018). An accurate and robust imputation method scImpute for single-cell RNA-seq data. *Nature Communications*, *9*(1), 997.

Liddicoat, C., Sydnor, H., Cando-Dumancela, C., Dresken, R., Liu, J., Gellie, N. J. C., Mills, J. G., Young, J. M., Weyrich, L. S., Hutchinson, M. R., Weinstein, P., & Breed, M. F. (2020). Naturally-diverse airborne environmental microbial exposures modulate the gut microbiome and may provide anxiolytic benefits in mice. *Science of the Total Environment, 701*, 134684. https://doi.org/10.1016/j.scitotenv.2019.134684

Liddicoat, C., Waycott, M., & Weinstein, P. (2016). Environmental Change and Human Health: Can Environmental Proxies Inform the Biodiversity Hypothesis for Protective Microbial–Human Contact? *BioScience*, *66*(12), 1023-1034. https://doi.org/10.1093/biosci/biw127

Lin, H., & Peddada, S. D. (2020). Analysis of compositions of microbiomes with bias correction. *Nature Communications*, *11*(1), 3514.

Liu, C., Li, H., Zhang, Y., & Chen, Q. (2016). Characterization of methanogenic activity during high-solids anaerobic digestion of sewage sludge. *Biochemical Engineering Journal*, *109*, 96-100.

Liu, H., Wang, J., He, T., Becker, S., Zhang, G., Li, D., & Ma, X. (2018). Butyrate: a double-edged sword for health? *Advances in Nutrition*, *9*(1), 21-29.

Liu, P., Qiu, Q., & Lu, Y. (2011). Syntrophomonadaceae-affiliated species as active butyrate-utilizing syntrophs in paddy field soil. *Applied and Environmental Microbiology*, *77*(11), 3884-3887.

Liu, Y., Qin, S., Song, Y., Feng, Y., Lv, N., Xue, Y., Liu, F., Wang, S., Zhu, B., & Ma, J. (2019). The perturbation of infant gut microbiota caused by cesarean delivery is partially restored by exclusive breastfeeding. *Frontiers in Microbiology*, *10*, 598. https://doi.org/10.3389/fmicb.2019.00598

Louis, P., & Flint, H. J. (2017). Formation of propionate and butyrate by the human colonic microbiota. *Environmental Microbiology*, *19*(1), 29-41. https://doi.org/10.1111/1462-2920.13589

Maier, T. V., Lucio, M., Lee, L. H., VerBerkmoes, N. C., Brislawn, C. J., Bernhardt, J., Lamendella, R., McDermott, J. E., Bergeron, N., & Heinzmann, S. S. (2017). Impact of dietary resistant starch on the human gut microbiome, metaproteome, and metabolome. *Mbio*, *8*(5). https://doi.org/10.1128/mBio.01343-17

Marselle, M. R., Lindley, S. J., Cook, P. A., & Bonn, A. (2021). Biodiversity and health in the urban environment. *Current Environmental Health Reports*, *8*(2), 146-156.

Martin, M. A., & Sela, D. A. (2013). Infant Gut Microbiota: Developmental Influences and Health Outcomes. *Building Babies* (pp. 233-256). https://doi.org/10.1007/978-1-4614-4060-4_11

McBratney, A. B., Santos, M. M., & Minasny, B. (2003). On digital soil mapping. *Geoderma*, *117*(1-2), 3-52.

McDonald, D., Hyde, E., Debelius, J. W., Morton, J. T., Gonzalez, A., Ackermann, G., Aksenov, A. A., Behsaz, B., Brennan, C., & Chen, Y. (2018). American gut: an open platform for citizen science microbiome research. *Msystems*, *3*(3).

McMurdie, P. J., & Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PloS One*, *8*(4), e61217.

Meadow, J. F., Altrichter, A. E., Kembel, S. W., Kline, J., Mhuireach, G., Moriyama, M., Northcutt, D., O'Connor, T. K., Womack, A. M., & Brown, G. (2014). Indoor airborne bacterial communities are influenced by ventilation, occupancy, and outdoor air source. *Indoor Air*, *24*(1), 41-48.

Meng, X., Yamakawa, K., Zou, K., Wang, X., Kuang, X., Lu, C., Wang, C., Karasawa, T., & Nakamura, S. (1999). Isolation and characterisation of neurotoxigenic Clostridium butyricum from soil in China. *Journal of Medical Microbiology*, *48*(2), 133-137.

Mhuireach, G., Johnson, B. R., Altrichter, A. E., Ladau, J., Meadow, J. F., Pollard, K. S., & Green, J. L. (2016). Urban greenness influences airborne bacterial community composition. *Sci Total Environ*, *571*, 680-687. https://doi.org/10.1016/j.scitotenv.2016.07.037

Mhuireach, G. Á., Betancourt-Román, C. M., Green, J. L., & Johnson, B. R. (2019). Spatiotemporal Controls on the Urban Aerobiome. *Frontiers in Ecology and Evolution*, *7*(43). https://doi.org/10.3389/fevo.2019.00043

Mhuireach, G. A., Wilson, H., & Johnson, B. R. (2021). Urban aerobiomes are influenced by season, vegetation, and individual site characteristics. *EcoHealth*, *18*, 331-344.

Mills, J. G., Bissett, A., Gellie, N. J., 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*, S322-S334.

Miyake, S., Kim, S., Suda, W., Oshima, K., Nakamura, M., Matsuoka, T., Chihara, N., Tomita, A., Sato, W., & Kim, S.-W. (2015). Dysbiosis in the gut microbiota of patients with multiple sclerosis, with a striking depletion of species belonging to clostridia XIVa and IV clusters. *Plos One*, *10*(9), e0137429. https://doi.org/10.1371/journal.pone.0137429

Morrison, D. J., & Preston, T. (2016). Formation of short chain fatty acids by the gut microbiota and their impact on human metabolism. *Gut Microbes*, *7*(3), 189-200. https://doi.org/10.1080/19490976.2015.1134082

Nielsen, C. C., Gascon, M., Osornio-Vargas, A. R., Shier, C., Guttman, D. S., Becker, A. B., Azad, M. B., Sears, M. R., Lefebvre, D. L., & Moraes, T. J. (2020). Natural environments in the urban context and gut microbiota in infants. *Environment International*, *142*, 105881.

Nilsen, M., Madelen Saunders, C., Leena Angell, I., Arntzen, M. Ø., Lødrup Carlsen, K. C., Carlsen, K.-H., Haugen, G., Hagen, L. H., Carlsen, M. H., & Hedlin, G. (2020). Butyrate Levels in the Transition from an Infant-to an Adult-Like Gut Microbiota Correlate with Bacterial Networks Associated with Eubacterium Rectale and Ruminococcus Gnavus. *Genes*, *11*(11), 1245.

Noh, C.-K., & Lee, K. J. (2020). Fecal Microbiota Alterations and Small Intestinal Bacterial Overgrowth in Functional Abdominal Bloating/Distention. *Journal of Neurogastroenterology and Motility*, *26*(4), 539.

Núñez, A., & García, A. M. (2022). Effect of the passive natural ventilation on the bioaerosol in a small room. *Building and Environment*, *207*, 108438.

Nurminen, N., Lin, J., Grönroos, M., Puhakka, R., Kramna, L., Vari, H. K., Viskari, H., Oikarinen, S., Roslund, M., & Parajuli, A. (2018). Nature-derived microbiota exposure as a novel immunomodulatory approach. *Future Microbiology*, *13*(07), 737-744. https://doi.org/10.2217/fmb-2017-0286

O'Brien, L. (2021). *slga: Data Access Tools for the Soil and Landscape Grid of Australia*. In (Version 1.2.0)

Ober, C., Sperling, A. I., von Mutius, E., & Vercelli, D. (2017). Immune development and environment: lessons from Amish and Hutterite children. *Current Opinion in Immunology*, *48*, 51-60. https://doi.org/10.1016/j.coi.2017.08.003

Ohira, H., Tsutsui, W., Mamoto, R., Yamaguchi, S., Nishida, M., Ito, M., & Fujioka, Y. (2016). Butyrate attenuates lipolysis in adipocytes co-cultured with macrophages through non-prostaglandin E2–mediated and prostaglandin E2–mediated pathways. *Lipids in Health and Disease*, *15*(1), 1-17. https://doi.org/10.1186/s12944-016-0387-0

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P., O'Hara, R., Simpson, G., & Solymos, P. (2022). vegan: Community Ecology Package. R package version 2.5-7. 2020.

Overbeek, R., Begley, T., Butler, R. M., Choudhuri, J. V., Chuang, H.-Y., Cohoon, M., de Crécy-Lagard, V., Diaz, N., Disz, T., & Edwards, R. (2005). The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Research*, *33*(17), 5691-5702.

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

Pandiyan, P., Bhaskaran, N., Zou, M., Schneider, E., Jayaraman, S., & Huehn, J. (2019). Microbiome dependent regulation of Tregs and Th17 cells in mucosa. *Frontiers in Immunology*, *10*, 426. https://doi.org/10.3389/fimmu.2019.00426

Paparo, L., Nocerino, R., Ciaglia, E., Di Scala, C., De Caro, C., Russo, R., Trinchese, G., Aitoro, R., Amoroso, A., & Bruno, C. (2020). Butyrate as bioactive human milk protective component against food allergy. *Allergy*.

Parada Venegas, D., De la Fuente, M. K., Landskron, G., González, M. J., Quera, R., Dijkstra, G., Harmsen, H. J., Faber, K. N., & Hermoso, M. A. (2019). Short chain fatty acids (SCFAs)-mediated gut epithelial and immune regulation and its relevance for inflammatory bowel diseases. *Frontiers in Immunology*, *10*, 277. https://doi.org/10.3389/fimmu.2019.00277

Parajuli, A., Grönroos, M., Siter, N., Puhakka, R., Vari, H. K., Roslund, M. I., Jumpponen, A., Nurminen, N., Laitinen, O. H., & Hyöty, H. (2018). Urbanization reduces transfer of diverse environmental microbiota indoors. *Frontiers in Microbiology*, *9*, 84. https://doi.org/10.3389/fmicb.2018.00084

Pichler, M. J., Yamada, C., Shuoker, B., Alvarez-Silva, C., Gotoh, A., Leth, M. L., Schoof, E., Katoh, T., Sakanaka, M., & Katayama, T. (2020). Butyrate producing colonic Clostridiales metabolise human milk oligosaccharides and cross feed on mucin via conserved pathways. *Nature communications*, *11*(1), 1-15.

Polymenakou, P. N. (2012). Atmosphere: a source of pathogenic or beneficial microbes? *Atmosphere*, *3*(1), 87-102.

Poret-Peterson, A. T., Albu, S., McClean, A. E., & Kluepfel, D. A. (2019). Shifts in soil bacterial communities as a function of carbon source used during anaerobic soil disinfestation. *Frontiers in Environmental Science*, *6*, 160.

QGIS Development Team. (2020). *QGIS Geographic Information System*. In QGIS Association. http://www.qgis.org

R Core Team. (2020). *R: A language and environment for statistical computing.* In R Foundation for Statistical Computing. https://www.R-project.org/

R Core Team. (2023). *R: A language and environment for statistical computing.* In R Foundation for Statistical Computing.

Rivière, A., Selak, M., Lantin, D., Leroy, F., & De Vuyst, L. (2016). Bifidobacteria and butyrate-producing colon bacteria: importance and strategies for their stimulation in the human gut. *Frontiers in Microbiology*, *7*, 979. https://doi.org/10.3389/fmicb.2016.00979

Robinson, J. M., Breed, A. C., Camargo, A., Redvers, N., & Breed, M. F. (2024). Biodiversity and human health: A scoping review and examples of underrepresented linkages. *Environmental Research*, 118115.

Robinson, J. M., & Breed, M. F. (2019). Green prescriptions and their co-benefits: Integrative strategies for public and environmental health. *Challenges*, *10*(1), 9. https://doi.org/10.3390/challe10010009

Robinson, J. M., & Breed, M. F. (2020). The Lovebug Effect: Is the human biophilic drive influenced by interactions between the host, the environment, and the microbiome? *Science of the Total Environment*, *720*, 137626. https://doi.org/10.1016/j.scitotenv.2020.137626

Robinson, J. M., & Breed, M. F. (2023). The aerobiome–health axis: a paradigm shift in bioaerosol thinking. *Trends in Microbiology*.

Robinson, J. M., Breed, M. F., Kümmerli, R., Frank, S. A., Xu, A., Zhou, J., Blank, L. M., Jiang, M., & Pilosof, S. (2023). The aerobiome–health axis: a paradigm shift in bioaerosol thinking. *Trends in Microbiology*, *31*(6), 550-551.

Robinson, J. M., Cando-Dumancela, C., Antwis, R. E., Cameron, R., Liddicoat, C., Poudel, R., Weinstein, P., & Breed, M. F. (2021). Exposure to airborne bacteria depends upon vertical stratification and vegetation complexity. *Scientific Reports*, *11*(1), 9516.

Robinson, J. M., Cando-Dumancela, C., Liddicoat, C., Weinstein, P., Cameron, R., & Breed, M. F. (2020). Vertical stratification in urban green space aerobiomes. *Environmental Health Perspectives*, *128*(11), 117008.

Robinson, J. M., Dumancela, C. C., Liddicoat, C., Weinstein, P., Cameron, R., & Breed, M. (2020). Vertical Stratification in Urban Green Space Aerobiomes. *bioRxiv*. https://doi.org/10.1101/2020.06.28.176743

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*(2), 40. https://doi.org/10.3390/challe9020040

Roduit, C., Frei, R., Ferstl, R., Loeliger, S., Westermann, P., Rhyner, C., Schiavi, E., Barcik, W., Rodriguez-Perez, N., & Wawrzyniak, M. (2019). High levels of butyrate and propionate in early life are associated with protection against atopy. *Allergy*, *74*(4), 799-809. https://doi.org/10.1111/all.13660

Rook, G. A. (2012). Hygiene hypothesis and autoimmune diseases. *Clinical Reviews in Allergy & Immunology*, *42*(1), 5-15.

Rook, G. A. (2013). Regulation of the immune system by biodiversity from the natural environment: an ecosystem service essential to health. *Proceedings of the National Academy of Sciences*, *110*(46), 18360-18367. https://doi.org/10.1073/pnas.1313731110

Rook, G. A., Lowry, C. A., & Raison, C. L. (2013). Microbial 'Old Friends', immunoregulation and stress resilience. *Evolution, Medicine, and Public Health*, *2013*(1), 46-64. https://doi.org/10.1093/emph/eot004

Roslund, M. I., Parajuli, A., Hui, N., Puhakka, R., Grönroos, M., Soininen, L., Nurminen, N., Oikarinen, S., Cinek, O., & Kramná, L. (2022). A Placebo-controlled double-blinded test of the biodiversity hypothesis of immune-mediated diseases: Environmental microbial diversity elicits changes in cytokines and increase in T regulatory cells in young children. *Ecotoxicology and Environmental Safety*, *242*, 113900.

Roslund, M. I., Puhakka, R., Grönroos, M., Nurminen, N., Oikarinen, S., Gazali, A. M., Cinek, O., Kramná, L., Siter, N., & Vari, H. K. (2020). Biodiversity intervention enhances immune regulation and health-associated commensal microbiota among daycare children. *Science Advances*, *6*(42), eaba2578.

Rothschild, D., Weissbrod, O., Barkan, E., Kurilshikov, A., Korem, T., Zeevi, D., Costea, P. I., Godneva, A., Kalka, I. N., & Bar, N. (2018). Environment dominates over host genetics in shaping human gut microbiota. *Nature*, *555*(7695), 210-215. https://doi.org/10.1038/nature25973

RoyChowdhury, T., Bramer, L. M., Brown, J., Kim, Y.-M., Zink, E., Metz, T. O., McCue, L. A., Diefenderfer, H. L., & Bailey, V. (2022). Soil Metabolomics Predict Microbial Taxa as Biomarkers of Moisture Status in Soils from a Tidal Wetland. *Microorganisms*, *10*(8), 1653.

Sanna, S., van Zuydam, N. R., Mahajan, A., Kurilshikov, A., Vich Vila, A., Vosa, U., Mujagic, Z., Masclee, A. A. M., Jonkers, D., Oosting, M., Joosten, L. A. B., Netea, M. G., Franke, L., Zhernakova, A., Fu, J., Wijmenga, C., & McCarthy, M. I. (2019). Causal relationships among the gut microbiome, short-chain fatty acids and metabolic diseases. *Nature Genetics*, *51*(4), 600-605. https://doi.org/10.1038/s41588-019-0350-x

Sarkar, A., Lehto, S. M., Harty, S., Dinan, T. G., Cryan, J. F., & Burnet, P. W. (2016). Psychobiotics and the manipulation of bacteria–gut–brain signals. *Trends in Neurosciences*, *39*(11), 763-781.

Schmidt, B., Mulder, I. E., Musk, C. C., Aminov, R. I., Lewis, M., Stokes, C. R., Bailey, M., Prosser, J. I., Gill, B. P., & Pluske, J. R. (2011). Establishment of normal gut microbiota is compromised under excessive hygiene conditions. *Plos One*, *6*(12), e28284. https://doi.org/10.1371/journal.pone.0028284

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

Sessitsch, A., Wakelin, S., Schloter, M., Maguin, E., Cernava, T., Champomier-Verges, M.-C., Charles, T. C., Cotter, P. D., Ferrocino, I., & Kriaa, A. (2023). Microbiome Interconnectedness throughout Environments with Major Consequences for Healthy People and a Healthy Planet. *Microbiology and Molecular Biology Reviews*, e00212-00222.

Sevelsted, A., Stokholm, J., Bønnelykke, K., & Bisgaard, H. (2015). Cesarean section and chronic immune disorders. *Pediatrics*, *135*(1), e92-e98. https://doi.org/10.1542/peds.2014-0596
Sibthorpe, R. L., & Brymer, E. (2020). Disconnected from nature: the lived experience of those disconnected from the natural world. *Innovations in a Changing World*, 59.
Sivaprakasam, S., Bhutia, Y. D., Ramachandran, S., & Ganapathy, V. (2017). Cell-surface and nuclear receptors in the colon as targets for bacterial metabolites and its relevance to colon health. *Nutrients*, *9*(8), 856. https://doi.org/10.3390/nu9080856
Sonnenburg, E. D., Smits, S. A., Tikhonov, M., Higginbottom, S. K., Wingreen, N. S., & Sonnenburg, J. L. (2016). Diet-induced extinctions in the gut microbiota compound over generations. *Nature*, *529*(7585), 212-215. https://doi.org/10.1038/nature16504
Spragge, F., Bakkeren, E., Jahn, M. T., BN Araujo, E., Pearson, C. F., Wang, X., Pankhurst, L., Cunrath, O., & Foster, K. R. (2023). Microbiome diversity protects against pathogens by nutrient blocking. *Science*, *382*(6676), eadj3502.
Stanhope, J., Breed, M. F., & Weinstein, P. (2020). Exposure to greenspaces could reduce the high global burden of pain. *Environmental Research*, *187*, 109641. https://doi.org/10.1016/j.envres.2020.109641
Steiman, C. A., Fadrosh, D., Evans, M. D., Vrtis, R. F., Olson, B., Barnes, K. L., Seroogy, C. M., Bendixsen, C. G., Lynch, S. V., & Gern, J. E. (2018). Farm Exposure Influences Skin Microbiota In Infancy. *Journal of Allergy and Clinical Immunology*, *141*(2), AB86. https://doi.org/10.1016/j.jaci.2017.12.278
Stein, M. M., Hrusch, C. L., Gozdz, J., Igartua, C., Pivniouk, V., Murray, S. E., Ledford, J. G., Marques dos Santos, M., Anderson, R. L., & Metwali, N. (2016). Innate immunity and asthma risk in Amish and Hutterite farm children. *New England Journal of Medicine*, *375*(5), 411-421. https://doi.org/10.1056/NEJMoa1508749
Stilling, R. M., van de Wouw, M., Clarke, G., Stanton, C., Dinan, T. G., & Cryan, J. F. (2016). The neuropharmacology of butyrate: The bread and butter of the microbiota-gut-brain axis? [Review]. *Neurochemistry International*, *99*, 110-132. https://doi.org/10.1016/j.neuint.2016.06.011
Subirats, J., Sharpe, H., & Topp, E. (2022). Fate of Clostridia and other spore-forming Firmicute bacteria during feedstock anaerobic digestion and aerobic composting. *Journal of Environmental Management*, *309*, 114643.
Sun, S., Jones, R. B., & Fodor, A. A. (2020). Inference-based accuracy of metagenome prediction tools varies across sample types and functional categories. *Microbiome*, *8*, 1-9.
Swenson, T. L., Karaoz, U., Swenson, J. M., Bowen, B. P., & Northen, T. (2017). Linking soil biology and chemistry using bacterial isolate exometabolite profiles. *bioRxiv*, 109330.
Takahashi, D., Hoshina, N., Kabumoto, Y., Maeda, Y., Suzuki, A., Tanabe, H., Isobe, J., Yamada, T., Muroi, K., & Yanagisawa, Y. (2020). Microbiota-derived butyrate limits the autoimmune response by promoting the differentiation of follicular regulatory T cells. *Ebiomedicine*, *58*, 102913.
Tanaka, M., & Nakayama, J. (2017). Development of the gut microbiota in infancy and its impact on health in later life. *Allergology International*, *66*(4), 515-522. https://doi.org/10.1016/j.alit.2017.07.010
Thakuria, D., Schmidt, O., Finan, D., Egan, D., & Doohan, F. M. (2010). Gut wall bacteria of earthworms: a natural selection process. *The ISME journal*, *4*(3), 357-366.
Torres-Fuentes, C., Golubeva, A. V., Zhdanov, A. V., Wallace, S., Arboleya, S., Papkovsky, D. B., Aidy, S. E., Ross, P., Roy, B. L., & Stanton, C. (2019). Short-chain fatty acids and microbiota metabolites attenuate ghrelin receptor signaling. *The FASEB Journal*, *33*(12), 13546-13559.
Tran, N. T., Li, Z., Ma, H., Zhang, Y., Zheng, H., Gong, Y., & Li, S. (2020). Clostridium butyricum: a promising probiotic confers positive health benefits in aquatic animals. *Reviews in Aquaculture*, *12*(4), 2573-2589.
Tripathi, B. M., Stegen, J. C., Kim, M., Dong, K., Adams, J. M., & Lee, Y. K. (2018). Soil pH mediates the balance between stochastic and deterministic assembly of bacteria. *The ISME Journal*, *12*(4), 1072-1083.

Tun, H. M., Konya, T., Takaro, T. K., Brook, J. R., Chari, R., Field, C. J., Guttman, D. S., Becker, A. B., Mandhane, P. J., & Turvey, S. E. (2017). Exposure to household furry pets influences the gut microbiota of infants at 3–4 months following various birth scenarios. *Microbiome*, *5*(1), 40. https://doi.org/10.1186/s40168-017-0254-x

United Nations. (2018). Revision of world urbanization prospects. *United Nations: New York, NY, USA, 799*.

Uteau, D., Hafner, S., Pagenkemper, S. K., Peth, S., Wiesenberg, G. L., Kuzyakov, Y., & Horn, R. (2015). Oxygen and redox potential gradients in the rhizosphere of alfalfa grown on a loamy soil. *Journal of Plant Nutrition and Soil Science*, *178*(2), 278-287.

Valles-Colomer, M., Falony, G., Darzi, Y., Tigchelaar, E. F., Wang, J., Tito, R. Y., Schiweck, C., Kurilshikov, A., Joossens, M., Wijmenga, C., Claes, S., Van Oudenhove, L., Zhernakova, A., Vieira-Silva, S., & Raes, J. (2019). The neuroactive potential of the human gut microbiota in quality of life and depression [Article]. *Nature Microbiology*, *4*(4), 623-632. https://doi.org/10.1038/s41564-018-0337-x

Van den Boogaart, K. G., & Tolosana-Delgado, R. (2008). "Compositions": a unified R package to analyze compositional data. *Computers & Geosciences*, *34*(4), 320-338.

Vandeputte, D., De Commer, L., Tito, R. Y., Kathagen, G., Sabino, J., Vermeire, S., Faust, K., & Raes, J. (2021). Temporal variability in quantitative human gut microbiome profiles and implications for clinical research. *Nature Communications*, *12*(1), 6740.

Vital, M., Howe, A. C., & Tiedje, J. M. (2014). Revealing the Bacterial Butyrate Synthesis Pathways by Analyzing (Meta) genomic Data [Article]. *Mbio*, *5*(2), 11, Article e00889-14. https://doi.org/10.1128/mBio.00889-14

Vital, M., Karch, A., & Pieper, D. H. (2017). Colonic Butyrate-Producing Communities in Humans: an Overview Using Omics Data [Article]. *Msystems*, *2*(6), 18, Article UNSP e00130-17. https://doi.org/10.1128/mSystems.00130-17

Vo, N., Tsai, T. C., Maxwell, C., & Carbonero, F. (2017). Early exposure to agricultural soil accelerates the maturation of the early-life pig gut microbiota. *Anaerobe*, *45*, 31-39. https://doi.org/10.1016/j.anaerobe.2017.02.022

Von Hertzen, L., & Haahtela, T. (2004). Asthma and atopy–the price of affluence? *Allergy*, *59*(2), 124-137. https://doi.org/10.1046/j.1398-9995.2003.00433.x

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*(11), 1089-1093. https://doi.org/10.1038/embor.2011.195

Warbrick, I., Dickson, A., Prince, R., & Heke, I. (2016). The biopolitics of Māori biomass: towards a new epistemology for Māori health in Aotearoa/New Zealand. *Critical Public Health*, *26*(4), 394-404.

Wasko, N. J., Nichols, F., & Clark, R. B. (2020). Multiple sclerosis, the microbiome, TLR2, and the hygiene hypothesis. *Autoimmunity Reviews*, *19*(1), 102430. https://doi.org/10.1016/j.autrev.2019.102430

Watkins, H., Robinson, J. M., Breed, M. F., Parker, B., & Weinstein, P. (2020). Microbiome-Inspired Green Infrastructure: A Toolkit for Multidisciplinary Landscape Design. *Trends in Biotechnology*. https://doi.org/10.1016/j.tibtech.2020.04.009

White, A., Sparrow, B., Leitch, E., Foulkes, J., Flitton, R., Lowe, A. J., & Caddy-Retalic, S. (2012). AUSPLOTS rangelands survey protocols manual.

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis (2nd ed)*. In *Springer-Verlag* Springer-Verlag.

Williams, K., Stein, J., Storey, R., Ferrier, S., Austin, M., Smyth, A., & Harwood, T. (2010). *0.01 degree stack of climate layers for continental analysis of biodiversity pattern, version 1.0*

Wintle, B. A., Kujala, H., Whitehead, A., Cameron, A., Veloz, S., Kukkala, A., Moilanen, A., Gordon, A., Lentini, P. E., & Cadenhead, N. C. (2019). Global synthesis of conservation studies reveals the importance of small habitat patches for biodiversity. *Proceedings of the National Academy of Sciences*, *116*(3), 909-914.

Woo, A. C., Brar, M. S., Chan, Y., Lau, M. C., Leung, F. C., Scott, J. A., Vrijmoed, L. L., Zawar-Reza, P., & Pointing, S. B. (2013). Temporal variation in airborne microbial populations and

microbially-derived allergens in a tropical urban landscape. *Atmospheric Environment*, *74*, 291-300.

Yamamura, R., Nakamura, K., Kitada, N., Aizawa, T., Shimizu, Y., Nakamura, K., Ayabe, T., Kimura, T., & Tamakoshi, A. (2019). Associations of gut microbiota, dietary intake, and serum short-chain fatty acids with fecal short-chain fatty acids. *Bioscience of Microbiota, Food and Health*, *39*, 11-17. https://doi.org/10.12938/bmfh.19-010

Zhao, Y., Chen, F., Wu, W., Sun, M., Bilotta, A. J., Yao, S., Xiao, Y., Huang, X., Eaves-Pyles, T. D., & Golovko, G. (2018). GPR43 mediates microbiota metabolite SCFA regulation of antimicrobial peptide expression in intestinal epithelial cells via activation of mTOR and STAT3. *Mucosal Immunology*, *11*(3), 752-762. https://doi.org/10.1038/mi.2017.118

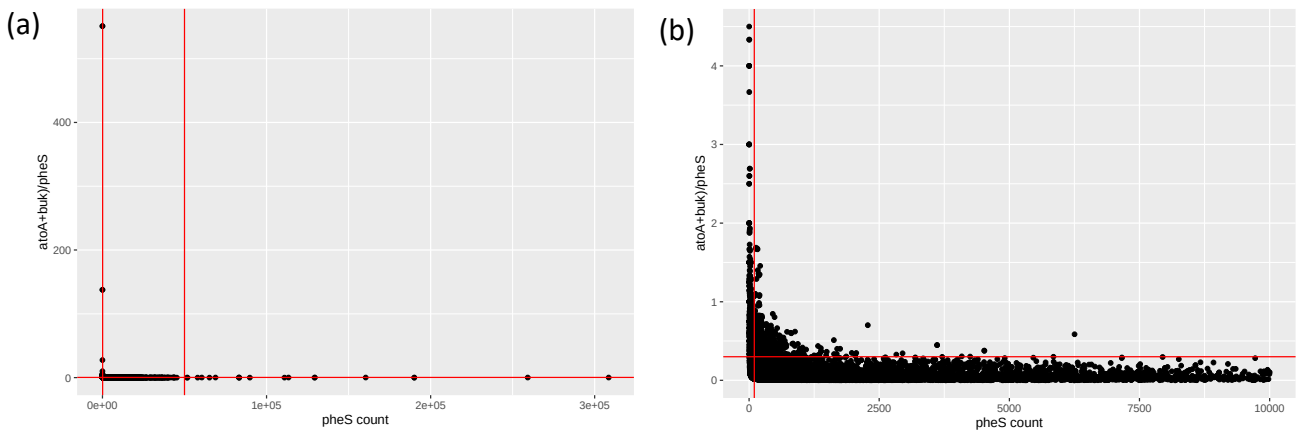
Zhou, C., Zhao, H., Xiao, X.-y., Guo, R.-j., Wang, Q., Chen, H., Zhao, L.-d., Zhang, C.-c., Jiao, Y.-h., & Ju, Y.-m. (2020). Metagenomic profiling of the pro-inflammatory gut microbiota in ankylosing spondylitis. *Journal of Autoimmunity*, *107*, 102360. https://doi.org/10.1016/j.jaut.2019.102360

APPENDICES

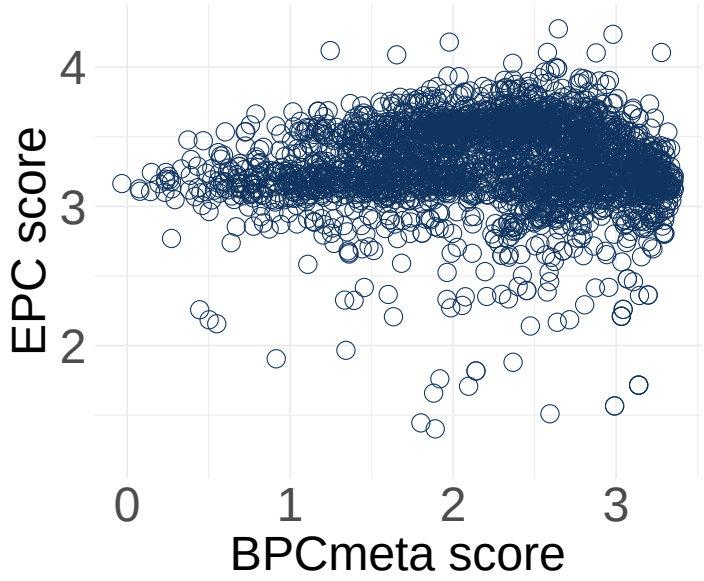
Appendix 1: Supplementary material from Chapter 2

The contents of this supplementary material appear in the following publication:

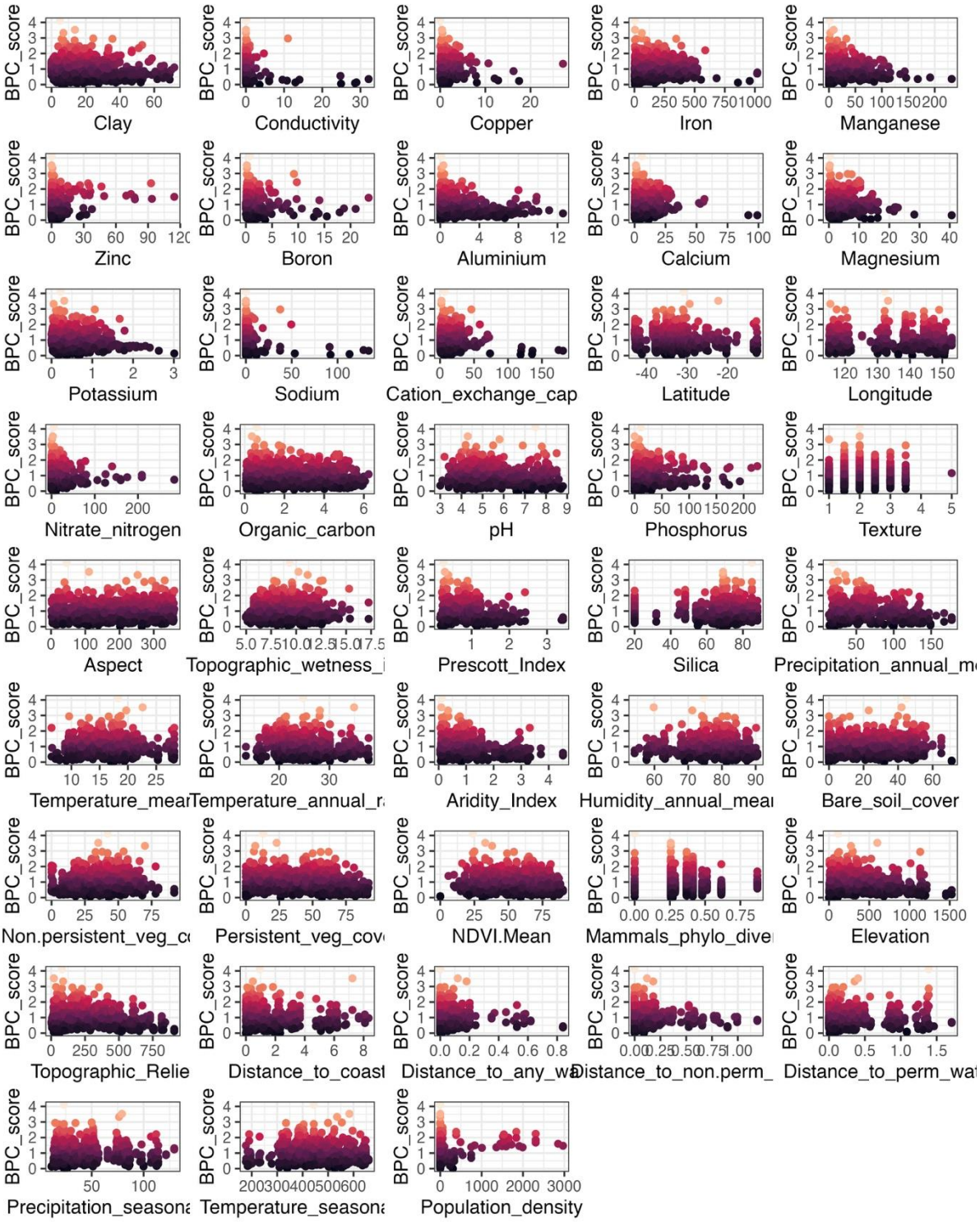
Brame JE, Liddicoat C, Abbott CA, Edwards RE, Gauthier NE, Robinson JA, Breed MF. Towards the biogeography of butyrate-producing bacteria. 2023. *bioRxiv* (under review in *Ecology and Evolution*). doi: 10.1101/2022.10.07.510278.



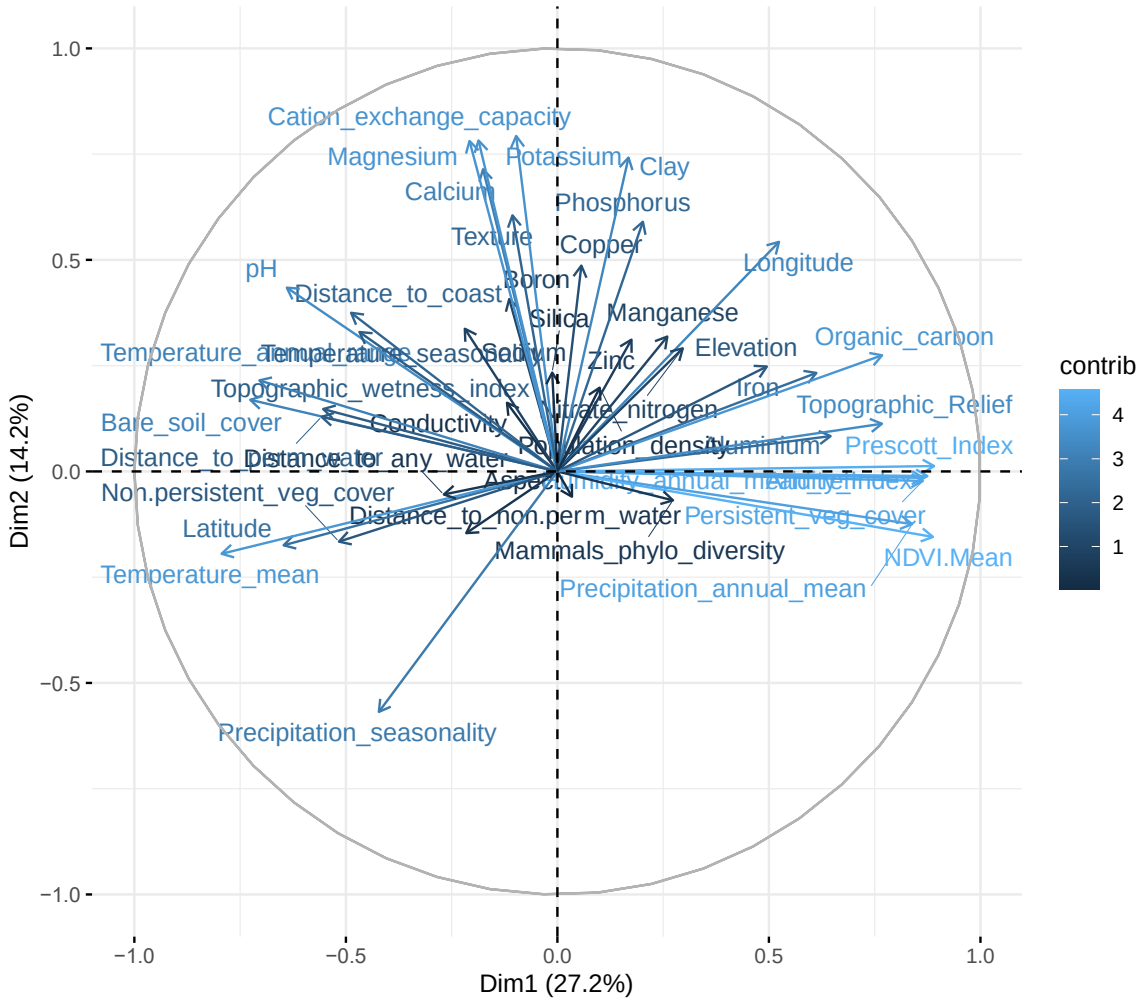
Supplementary Figure 1. Scatterplots showing thresholds of pheS counts and (atoA+buk)/pheS ratio. Scatterplots of pheS count samples against (atoA+buk)/pheS ratios. (a) The horizontal red line shows the threshold (y-intercept=0.3) of the (atoA+buk)/pheS ratio, above which samples were removed from analysis. The vertical red lines show the thresholds (x-intercepts 100 and 50,000) of the pheS counts, outside of which samples were removed from analysis. (b) Data from (a) shown with x-axis upper limit=10,000 and y-axis upper limit=5.0 set for a clearer visualisation of thresholds.



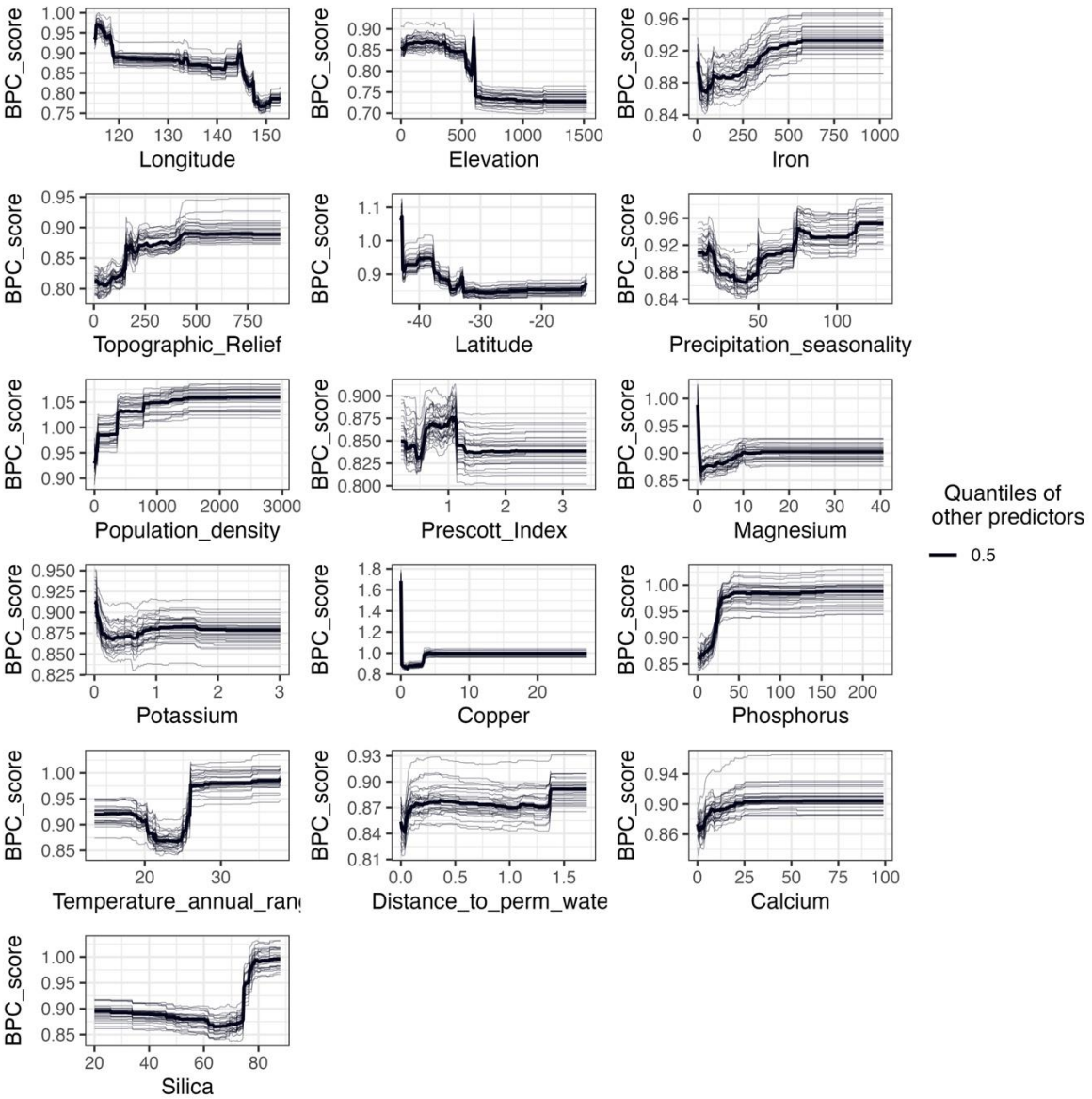
Supplementary Figure 2. Butyrate production capacity scores show specificity toward butyrate production rather than general anaerobicity.
Scatterplot of two anaerobic condition production capacities of soil bacteria, ethanol production capacity (EPC) and butyrate production capacity (BPC_{meta}) scores, among metagenomic soil samples (r^2=5.6e-05, p=0.70).



Supplementary Figure 3. Scatterplots of ecological covariates with BPC_{16S} scores.
Scatterplots of BPC_{16S} scores against 43 continuous ecological predictor variables.

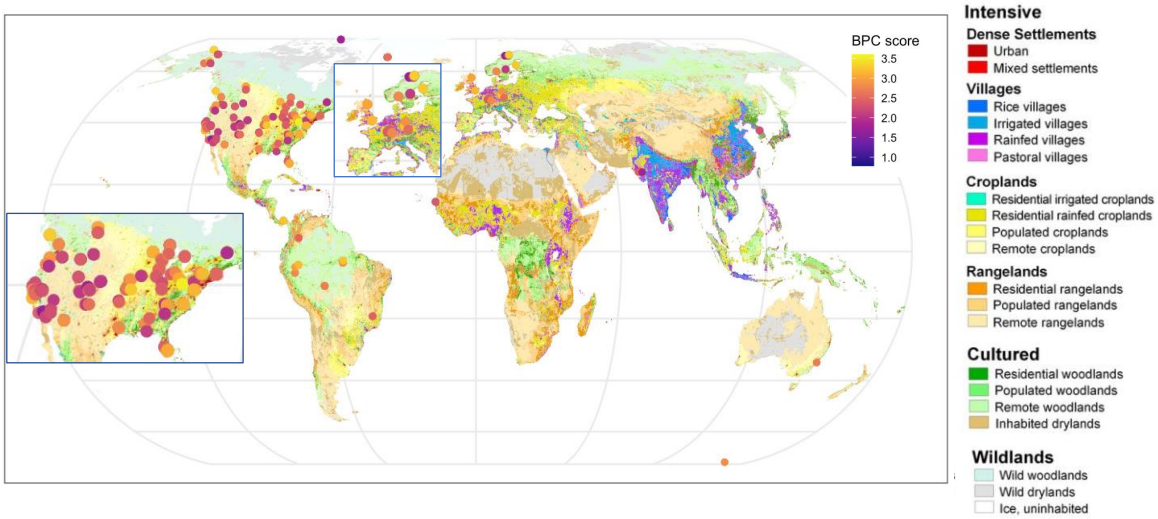


Supplementary Figure 4. PCA variables plot from analysis of Australian 16S rRNA soil samples. Contributions of 43 continuous predictor variables to the top two principal components from principal component analysis. Longer arrows indicate greater contribution. “Dim1” generally represents climatic influences, and “Dim2” generally represents soil fertility and cation content.



Supplementary Figure 5. Partial dependence plots from Random Forest modelling with Australian 16S rRNA soil samples.

Partial dependence plots from Random Forest modelling with Australian 16S rRNA soil samples and ecological covariates. Only covariates with a variable importance score above the median are shown. Trendlines indicate changes in BPC_{16S} score response while maintaining centred values of ecological covariates.



Supplementary Figure 6. Anthrome-based world map of downloaded soil and land-based sediment samples.
Visualization of variation in BPC_{meta} scores of global soil and land-based sediment samples using an anthrome-based world map. *n*=2850. BPC_{meta} scores indicated by point colors. Anthrome classification legend shows anthrome types ("Intensive", "Cultured", and "Wild"), levels ("Dense Settlements", "Villages", "Croplands", "Rangelands", "Cultured", and "Wildlands"), and classes (e.g., "Urban" and "Residential woodlands").

Supplementary Tables

The following datasets generated during and/or analysed in the current study are available on figshare at https://figshare.com/s/3ae2f1327a11f91793d8.

Supplementary Table 1. Descriptions of genes considered for analyses of global metagenomic samples.
This table lists and describes the genes within the butyrate synthesis pathway considered for analysis of global metagenomic samples. The table includes the EC number of the associated enzymes, the general function of each enzyme within the butyrate synthesis pathway, and reason for inclusion or exclusion in our analyses.

Supplementary Table 2. Gene counts of *buk* and *atoA* genes among metagenomic samples.
This table shows *buk* counts and *atoA* counts on two separate sheets for each bacterial genome from the IMG/M database with at least one of either gene. Mean gene count of *buk* was 1.180044. Mean gene count of *atoA* was 2.113735. "Status" column shows genome completion: "F"=Finished, "P"=Permanent draft, "D"=Draft.

Supplementary Table 3. Downloaded metagenomic data and metadata from 22,593 samples.
This table provides the source download data from 22,593 metagenomic samples. The first sheet is the combined data, and subsequent sheets show separately the gene count data for *atoA*, *buk*, and *pheS* and sample metadata.

Supplementary Table 4. Downloaded metagenomic sample data distributed into six general source categories.
This table shows the distribution of metagenomic sample data into six general source categories. Each sheet in the workbook has a separate category: Human, Non-human animal gut, Plant, Soil, Aquatic, and Agro-Industrial. Columns (A)-(K) are gene data and calculations of BPC_{meta} scores. Columns (L)-(BD) are sample metadata.

Supplementary Table 5. Set of 118 putative butyrate-producing bacteria species and their taxonomic data.
This table lists the set of 118 putative butyrate-producing bacterial species (Column F) and their taxonomic families (Column G), with the proportion of species within each family that are putative butyrate producers (Columns A-D). The species set was developed from Vital et al(Vital, Howe, & Tiedje, 2014), with taxonomic and proportion data derived from Genome Taxonomy Database (GTDB).

Supplementary Table 6. Ecological covariates used in the analyses of Australian 16S rRNA soil samples.
This table describes each of 49 ecological SCORPAN predictor variables. The predictor variables include 43 continuous variables and 6 categorical variables. For each variable, the table lists the download source, ecological range/units, type of data, download date, and general description.

Supplementary Table 7. Gene data, metadata, and covariate data for 1,331 Australian 16S rRNA soil samples.

This table provides the data used in analysis of 1,331 Australian 16S rRNA soil samples with 49 ecological SCORPAN covariates. Samples with incomplete data were removed.

Supplementary Table 8. Data for each of five land type clusters generated from Australian 16S rRNA soil sample analysis.
This table provides data for each land type cluster derived from *k*-means clustering. Data for the 43 numeric predictor variables and sample BPC_{16S} scores are included. All listed data are medians.

Supplementary Table 9. Data for categorical variables generated from Australian 16S rRNA soil sample analysis.
This table provides BPC_{16S} data for each category of the categorical predictor variables: land cover, land use, anthropogenic biome, vegetation types, and major vegetation subgroups. The table includes the count, mean BPC_{16S} score, median BPC_{16S} score, and standard deviation and is sorted by descending medians.

Appendix 2: Supplementary material from Chapter 3

The contents of this supplementary material appear in the following publication:

Brame JE, Liddicoat C, Abbott CA, Cando-Dumancela C, Robinson JA, Breed MF. Urban greenspace aerobiomes are shaped by soil conditions and land cover type. 2024. *bioRxiv* (under review in *Microbial Ecology*), doi: 10.1101/2024.01.12.575340

Table S1: Sample data of 23 soil physicochemical variables (separate file on Figshare, doi: 10.6084/m9.figshare.24112917)

Table S2: Genus-level bacterial differential abundances between sports fields and nature parks (separate file on Figshare, doi: 10.6084/m9.figshare.24112917)

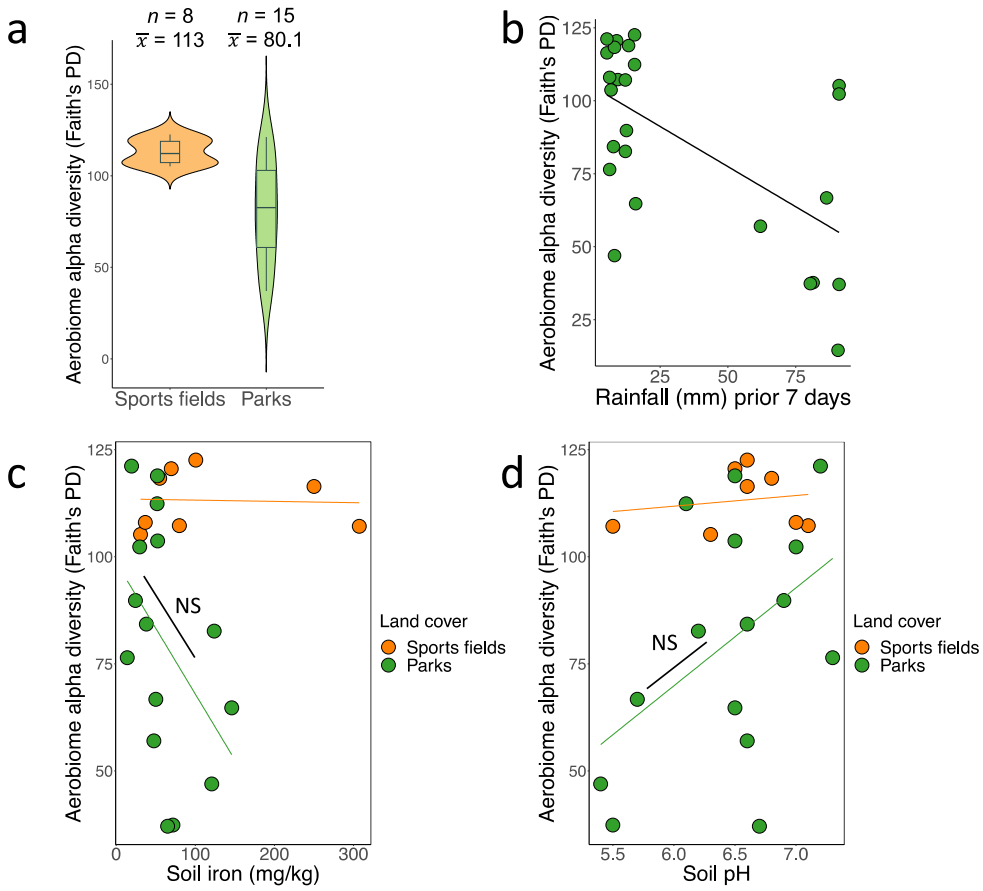


Fig. S1. (a) Boxplots of aerobiome alpha diversity by land cover. The y-axis shows the aerobiome alpha diversity calculated by Faith's phylogenetic diversity (PD). Boxes show the median and interquartile range, while whiskers extend to the remaining range of data. (b) Relationship of aerobiome bacterial alpha diversity (Faith's PD) and total rainfall volume during seven days prior to sampling. (c) Relationship of soil iron with aerobiome alpha diversity (Faith's PD). Orange and green lines separately show regressions for sports fields and nature parks, respectively. (d) Relationship of soil pH with aerobiome alpha diversity (Faith's PD). Orange and green lines show regressions separately for sports fields and nature parks, respectively.

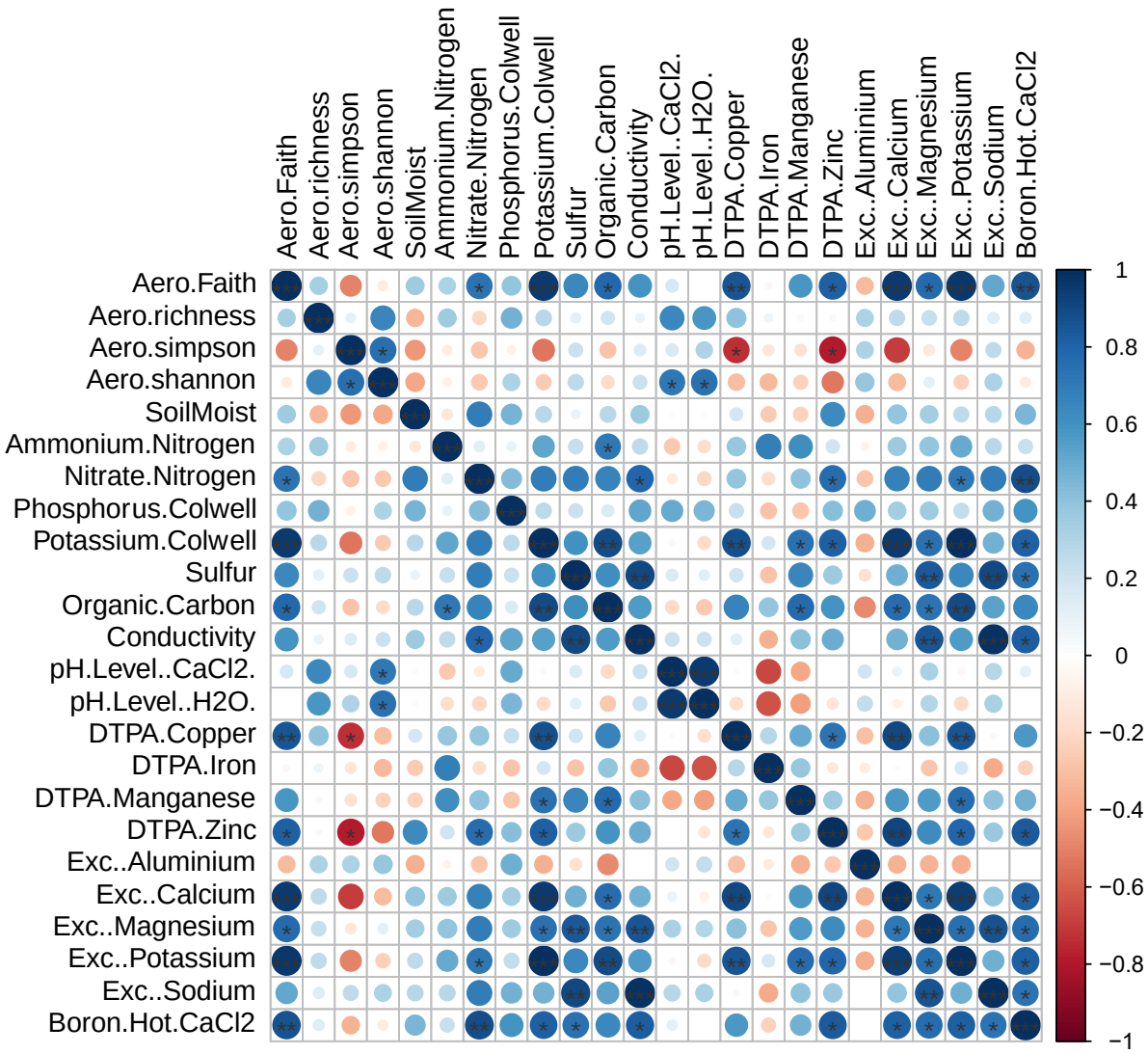


Fig. S2. Pearson correlation coefficients between four aerobiome alpha diversity indices and soil physicochemical parameters in sports fields. * indicates significance at $p < 0.05$, and ** indicates significance at $p < 0.01$.

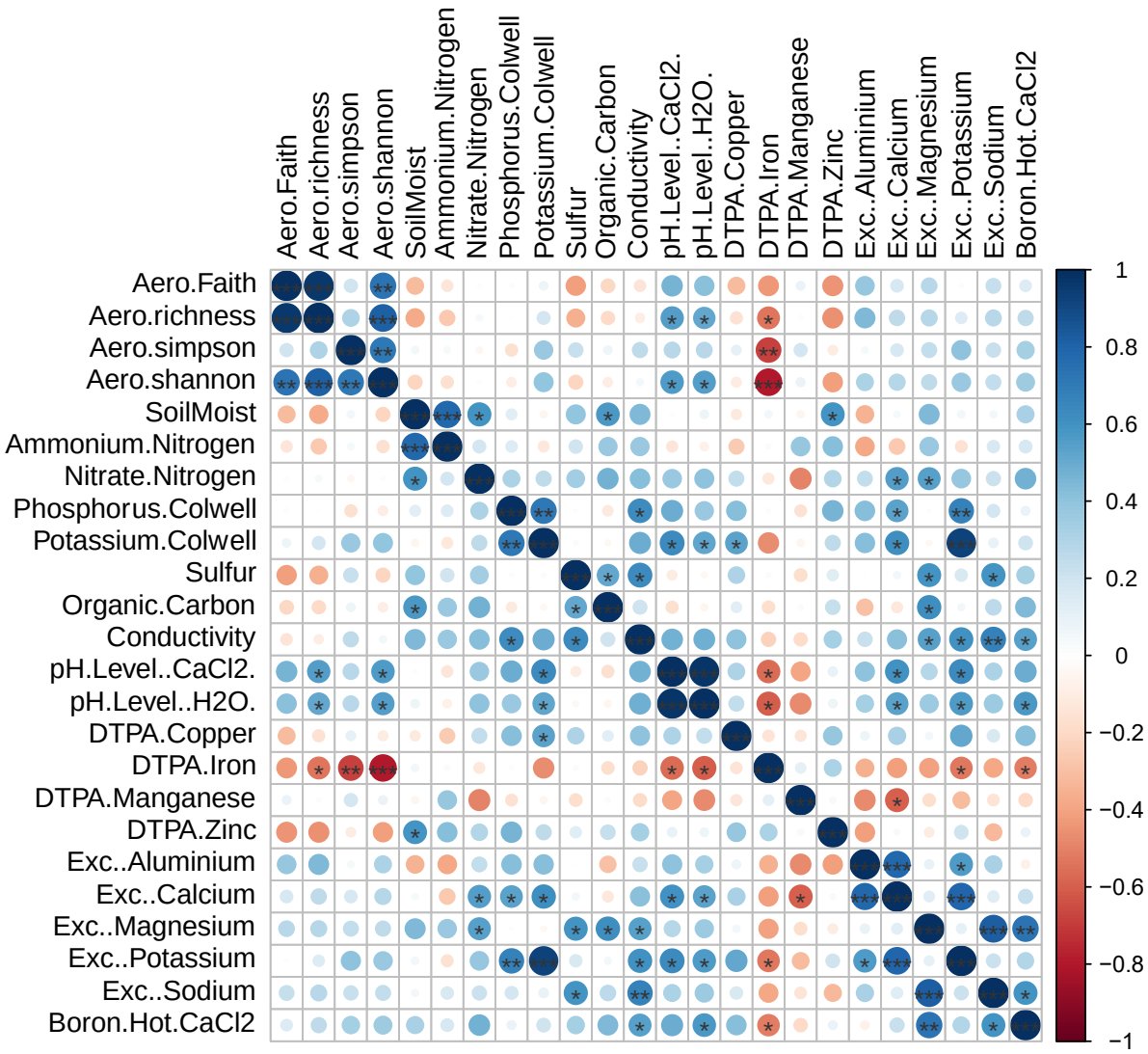


Fig. S3. Pearson correlation coefficients between four aerobiome alpha diversity indices and soil physicochemical parameters in nature parks. * indicates significance at $p < 0.05$, ** indicates significance at $p < 0.01$, and *** indicates significance at $p < 0.001$.

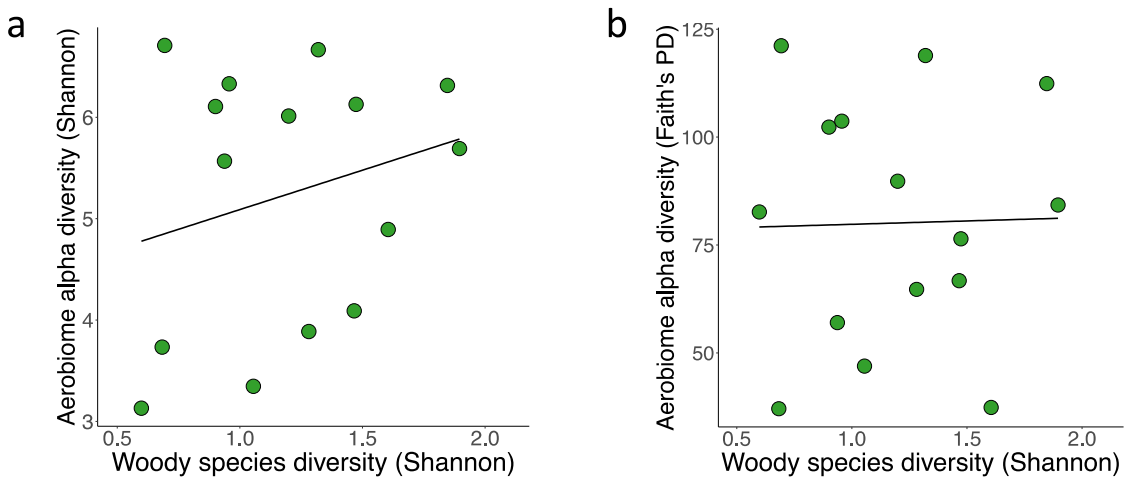


Fig. S4. (a) Relationship of woody plant species diversity (Shannon) with aerobiome alpha diversity using Shannon index. (b) Relationship of woody plant species diversity (Shannon) with aerobiome alpha diversity using Faith's phylogenetic diversity.

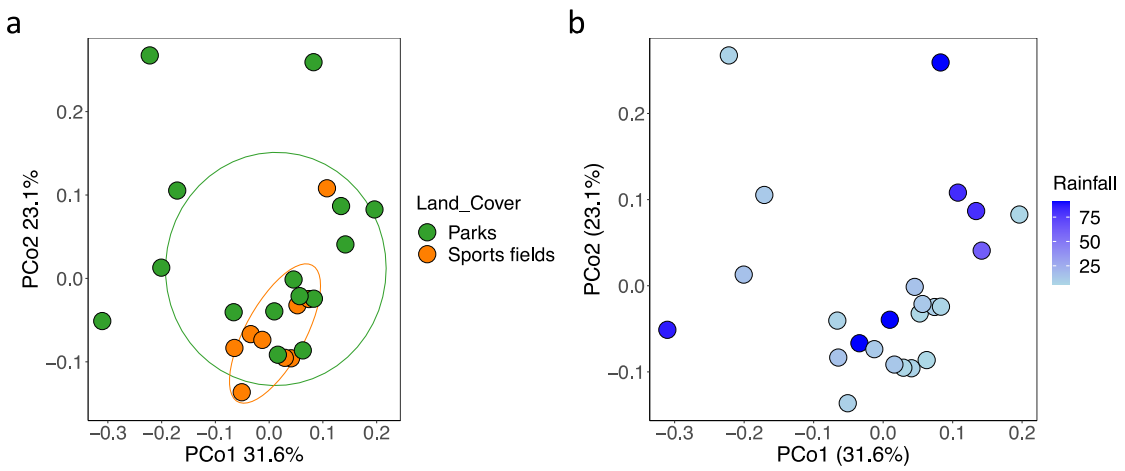


Fig. S5. (a) Principal coordinates analysis based on weighted Unifrac distances displaying variation in aerobiome community composition between land cover types (PERMANOVA Adonis test: 999 permutations, df = 1, F = 1.423, R^2 = 0.06 p = 0.168, n = 25 sites). (b) Principal coordinates analysis based on weighted Unifrac distances displaying variation in aerobiome community composition by total rainfall volume in the week prior to sampling (PERMANOVA Adonis test: 999 permutations, df = 1, F = 1.524, R^2 = 0.06 p = 0.12, n = 25 sites).

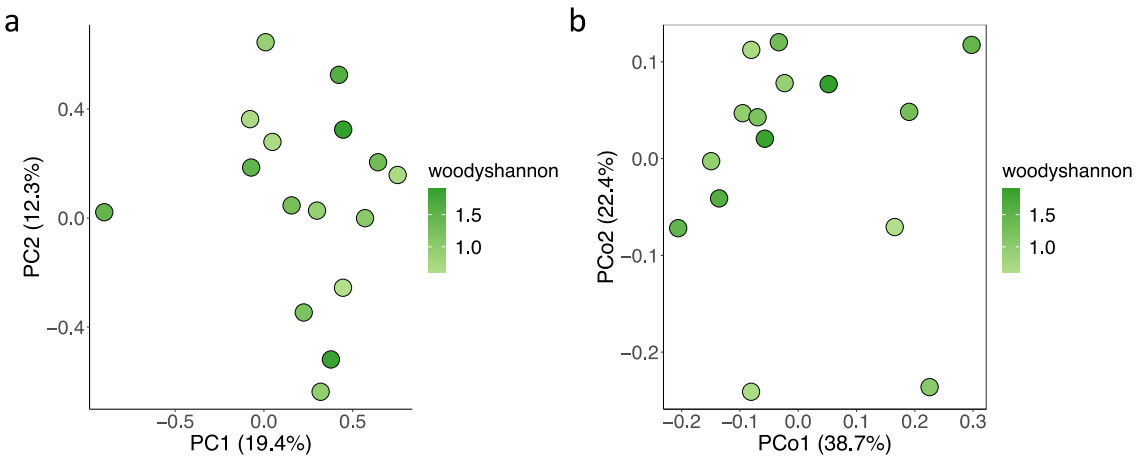


Fig. S6. (a) Principal components analysis based on Aitchison distances displaying variation in aerobiome community composition of parks with varying Shannon diversity of woody plant species (PERMANOVA df = 1, F = 0.80, R^2 = 0.05, p = 0.82, n = 14 sites). (b) Principal coordinates analysis based on weighted Unifrac distances displaying variation in aerobiome community composition of parks with varying Shannon diversity of woody plant species (PERMANOVA df = 1, F = 0.50, R^2 = 0.04, p = 0.86, n = 14 sites).

Appendix 3: Supplementary material from Chapter 4

The contents of this supplementary material appear in the following manuscript:

Brame JE, Liddicoat C, Abbott CA, Cando-Dumancela C, Robinson JA, Breed MF. Urban sports fields support higher levels of soil butyrate and butyrate-producing bacteria than urban nature parks. (in final stages of co-author review prior to submission to *ISME Journal).*

Table S1. Woody plant diversity and soil physicochemical parameters by land cover classification

	Sports fields	Parks
Woody plant diversity	0	1.260893
Soil moisture	25.38864	18.93182
Ammonium nitrogen	7.363636	5.909091
Nitrate nitrogen	20.54545	20.40909
Phosphorus	62.63636	33.86364
Sulfur	26.08182	14.93636
Organic carbon	4.047273	4.140909
Conductivity	0.2942727	0.2387727
pH	6.509091	6.468182
Copper	1.715455	1.672273
Iron	116.97273	59.72273
Manganese	8.062727	13.115455
Zinc	14.16727	20.98091
Aluminium	0.03636364	0.05954545
Calcium	11.19909	20.94136
Magnesium	4.663636	5.081364
Potassium	1.209091	1.415
Sodium	0.9754545	0.6231818
Boron	2.165455	2.984091

Table S2. Physicochemical parameters of soils by site.
See Figshare: doi: 10.6084/m9.figshare.24993345.

Table S3. Taxonomy of putative butyrate-producing bacteria.
See Figshare: doi: 10.6084/m9.figshare.24993345.

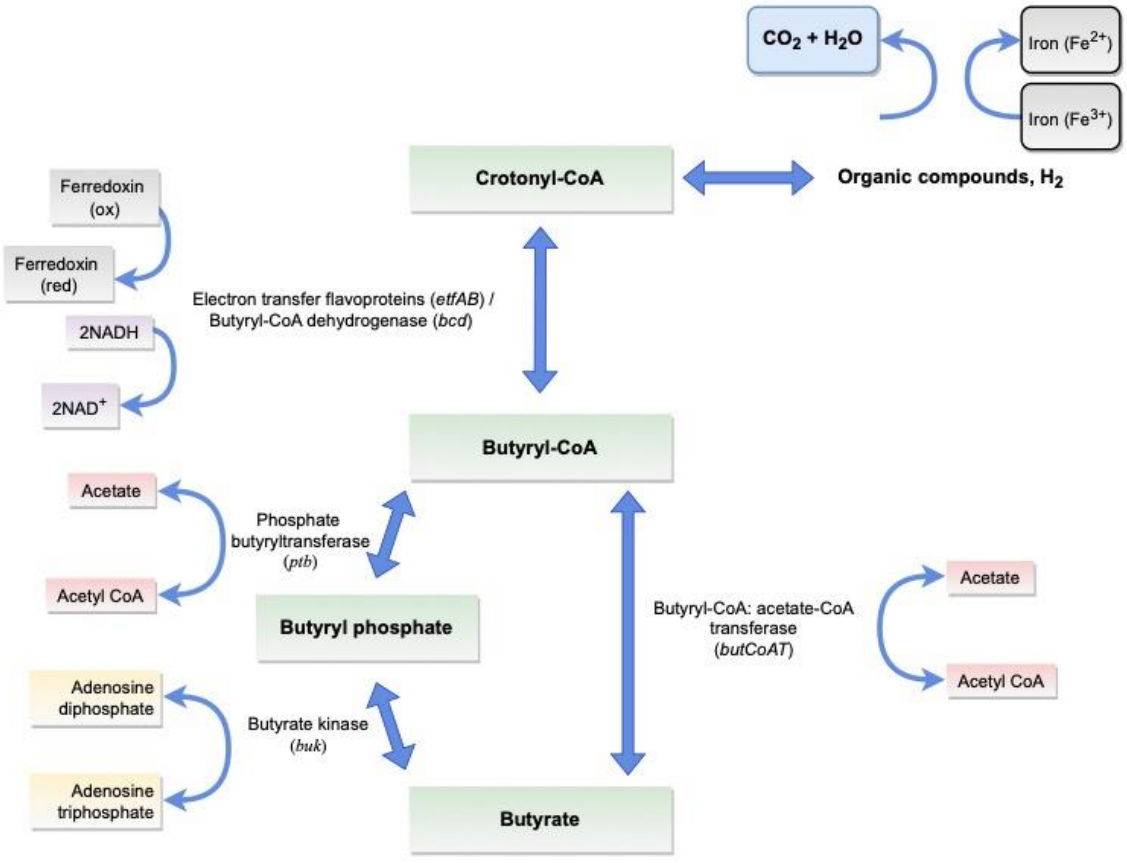


Figure S1. Biochemical pathway of butyrate production and butyrate oxidation.

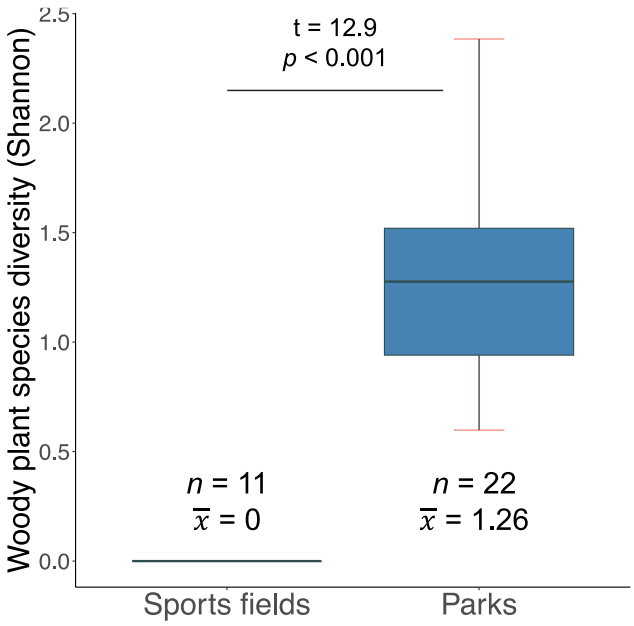


Figure S2. Boxplots of woody plant species diversity by land cover. The y-axis shows the species diversity calculated by Shannon index. Boxes show the median and interquartile range, while whiskers extend to the remaining range of data.

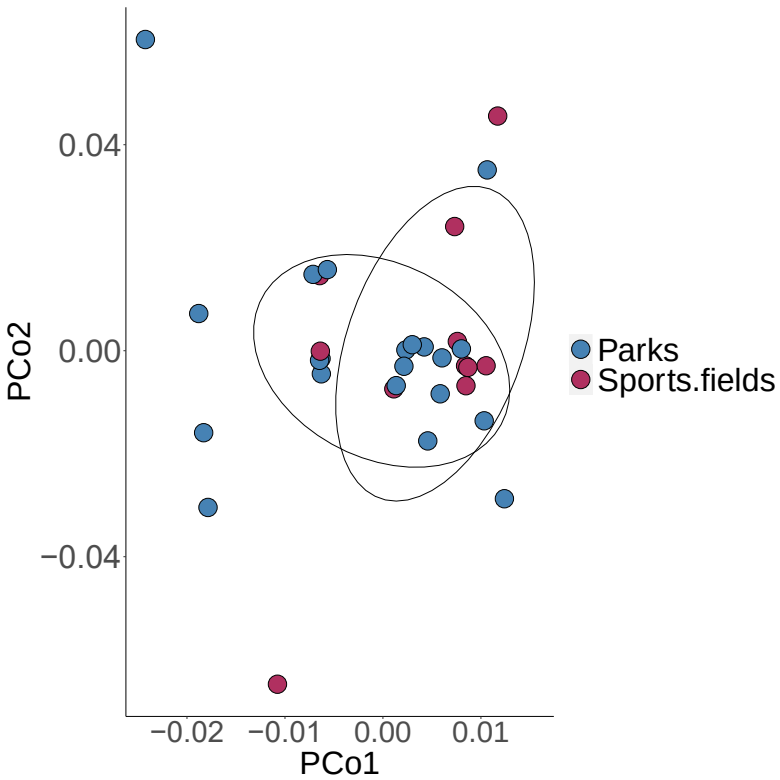


Figure S3. Principal coordinates analysis based on Aitchison distances displaying showing similar soil butyrate-producing bacterial community composition between sports fields and nature parks (Adonis PERMANOVA: F = 1.905, R^2 = 0.06, df = 1, p = 0.10).

Appendix 4: Supplementary material from Chapter 5

The contents of this supplementary material appears in the following publication:

Brame JE, Warbrick I, Heke D, Liddicoat, C, Breed MF. Short-term passive greenspace exposures have little effect on nasal microbiomes: a cross-over exposure study of a Māori cohort. 2024. *bioRxiv* (under review in *Environmental Research*). doi: 10.1101/2024.01.17.576148.

Table S1. Alpha diversity analysis results comparing Hill numbers generated with q = 0.8, q = 0, and q = 1.

	q = 0.8		**q = 0**		**q = 1.0**	
	Mean hill number	**Statistic**	**Mean hill number**	**Statistic**	**Mean hill number**	**Statistic**
Aerobiomes (*n* = 7)	82.3 ± 64.4 SD		224.6 ± 162.6 SD		59.1 ± 48.0 SD	
Nasal microbiomes (*n* = 238)	19.5 ± 10.6 SD	W = 1378, *p* = 0.003	80.1 ± 46.8 SD	W = 1204, *p* = 0.045	15.3 ± 8.1 SD	W = 1344, *p* = 0.006
Outdoor aerobiomes (*n* = 3)	60.4 ± 70.3 SD		153 ± 165 SD		45.3 ± 53.6 SD	
Indoor aerobiomes (*n* = 4)	98.8 ± 64.6 SD	t = 0.740, df = 4.2, *p* = 0.50	278 ± 161 SD	t = 1.01, df = 4.4, *p* = 0.37	69.4 ± 48.5 SD	t = 0.614, df = 4.2, *p* = 0.57
		Wilcox		**Wilcox**		**Wilcox**
Effect of group on changes in nasal bacterial alpha diversity after outdoor exposures		W = 433, *p* = 0.83		W = 361, *p* = 0.97		W = 377, *p* = 0.99
Effect of group on changes in nasal bacterial alpha diversity after indoor exposures		W = 358, *p* = 0.18		W = 310, *p* = 0.13		W = 342, *p* = 0.32
Effect on the nasal bacterial alpha diversity by the outdoor exposures –		W = 1388, *p* = 0.11		W = 1320, *p* = 0.20		W = 1237, *p* = 0.07

combined groups			
Effect on the nasal bacterial alpha diversity by the indoor exposures – combined groups	W = 1818, p = 0.93	W = 1512, p = 0.63	W = 1538, p = 0.74

	Repeated measures ANOVA	**Repeated measures ANOVA**	**Repeated measures ANOVA**
Time-series effect on nasal microbiome alpha diversity of group A	ges = 0.084, p = 0.30	ges = 0.089, p = 0.24	ges = 0.079, p = 0.35
Time-series effect on nasal microbiome alpha diversity of group B	ges = 0.192, p = 0.16	ges = 0.659, p = 0.008	ges = 0.201, p = 0.15

Table S2. Differential abundances of genera in groups A and B following outdoor exposures. See Figshare (doi: 10.6084/m9.figshare.24993471) for the Excel workbook, with data from each exposure day on a different sheet.

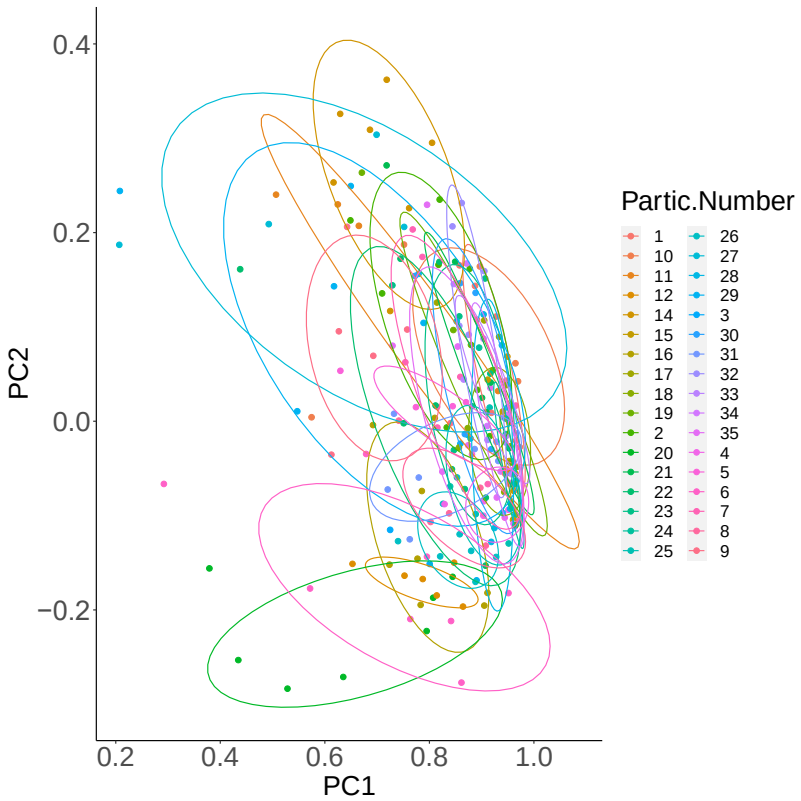


Figure S1. Principal components analysis based on centred-log ratio compositional abundance data showing strong effect of participant on the community composition of samples from Day 1 to Day 8 (Adonis PERMANOVA: F = 3.667, R^2 = 0.382, *p* = 0.001).