THE UNIVERSITY of ADELAIDE

Biodiversity, Environmental Microbiomes and Human Health

Craig Anthony Liddicoat

A thesis submitted in fulfilment of the

requirements for the degree of Doctor of Philosophy

School of Biological Sciences

The University of Adelaide, Australia

July 2019

Table of Contents

Abstract

Emerging evidence suggests that exposure to microbial diversity and key species (Old Friends) from biodiverse, natural environments may provide critical immune training and regulation. Conversely, reduced contact with the right kind of environmental microbial communities (microbiota) and their genetic material (microbiomes) may contribute to the modern growth in immunoregulatory disorders with potential to impact both infectious and non-communicable diseases.

However, possible connections between biodiversity, environmental microbiomes and human health remain understudied due to their multidisciplinary nature. There is limited knowledge of the composition, modes of action, and environmental distribution of Old Friends. We do not know if it is microbial diversity *per se*, key species, or a combination of both, that may have protective effects. Yet, importantly, the environment-host microbiota pathway offers promise for cost-effective population health interventions (e.g. through restoring biodiverse green space in cities) at a time when health care systems around the world are seeing unsustainable growth in utilisation and budget demands. Therefore, a primary goal of this thesis was to build knowledge of potential connections, with a view to informing policy. Specifically, the aims were to:

1. Test whether the notion of beneficial biodiversity-health relationships are supported in existing Australia-wide datasets
2. Examine what types or attributes of environments might be most associated with health benefits, to focus more detailed study
3. Identify microbial taxa that associate with natural vs. degraded environments and potential links to human health
4. Gather controlled experimental evidence of microbiota transfer from biodiverse environments to hosts, to explore potential mechanistic links.

I employed multidisciplinary methods reflecting the nature of the research topic. From continent-wide environmental mapping and hospital admission datasets, I found that landscape-scale measures of biodiversity correlated with reduced rates of respiratory disease and ranked highly among known predictors. Also, I found that populations living near soils with high cation exchange capacity—a proxy for soil microbial diversity—experienced lower rates of infectious and parasitic disease.

I developed a new merged-sample bootstrap resampling technique enabling deep analysis of soil bacterial 16S rRNA microbiome data from a grassy woodland restoration chronosequence, from which key indicator groups were identified. Human-associated opportunistic and pathogen-containing taxa were found to be favoured in disturbed environments, yet reduced in mature, biodiverse environments.

Finally, in a randomised controlled experiment with mice exposed to airborne dust from soil spanning a biodiversity gradient, I found changes to gut microbiota and reduced anxiety-like behaviour in females corresponding to the high biodiversity treatment. Among bacterial taxa that increased in the gut of high treatment mice, I identified a putative spore-forming, anaerobic environmental microbe capable of producing a key metabolite, butyrate, linked to mammalian gut health and mental health.

These findings suggest naturally-diverse soil microbial communities may provide a health protecting role due to: ecological controls on potential opportunistic pathogens, increased immunomodulatory microbial diversity, and enhanced capacity to support beneficial key species and metabolite production pathways, for example via the gut-brain-microbiome axis. Implications of this work include opportunities to improve public health through increased exposure to biodiverse green space and soils.

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

In addition, I certify that no part of this work will, in the future, be used in a submission in my name for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint award of this degree.

The author acknowledges that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

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

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

Signed __ _____ Date: 26 July 2019

Published and submitted works

Works included in thesis

The following works (1-5) are included in this thesis.

1. **Liddicoat, C.**, Waycott, M., and Weinstein, P. (2016). Environmental change and human health: Can environmental proxies inform the biodiversity hypothesis for protective microbial–human contact? *BioScience* 66, 1023-1034. https://doi.org/10.1093/biosci/biw127
2. **Liddicoat, C.**, Bi, P., Waycott, M., Glover, J., Lowe, A. J., and Weinstein, P. (2018). Landscape biodiversity correlates with respiratory health in Australia. *Journal of Environmental Management* 206, 113-122. https://doi.org/10.1016/j.jenvman.2017.10.007
3. **Liddicoat, C.**, Bi, P., Waycott, M., Glover, J., Breed, M., and Weinstein, P. (2018). Ambient soil cation exchange capacity inversely associates with infectious and parasitic disease risk in regional Australia. *Science of The Total Environment* 626, 117-125. https://doi.org/10.1016/j.scitotenv.2018.01.077
4. **Liddicoat, C.**, Weinstein, P., Bissett, A., Gellie, N., Mills, J., Waycott, M., and Breed, M. (2019). Can bacterial indicators of a grassy woodland restoration inform ecosystem assessment and microbiota-mediated human health? *Environment International* 129, 105-117. https://doi.org/10.1016/j.envint.2019.05.011
5. **Liddicoat, C.**, Sydnor, H., Cando-Dumancela, C., Dresken, R., Liu, J., Gellie, N., Mills, J., Young, J., Weyrich, L., Hutchinson, M., Weinstein, P. and Breed, M. (In Review). Naturally-diverse airborne environmental microbial exposures modulate the gut microbiome and reduce anxiety-like behaviour in mice. [Submitted to *Proceedings of the Royal Society B*]

Additional collaborative works during PhD candidature

The following published and submitted works (6-7) represent collaborative research activity that I have contributed to alongside the research presented in this thesis.

6. Mills, J.G., Brookes, J.D., Gellie, N.J.C., **Liddicoat, C.**, Lowe, A.J., Sydnor, H.R., Thomas, T., Weinstein, P., Weyrich, L.S., and Breed, M.F. (2019). Relating urban biodiversity to human health with the 'holobiont' concept. *Frontiers in Microbiology* 10, 550. https://doi.org/10.3389/fmicb.2019.00550
7. Baruch, Z., **Liddicoat, C.**, Laws, M., Marker, K., Morelli, H., Yan, D., Young, J.M., Breed, M.F. (In Review) Characterising the soil fungal microbiome in metropolitan green spaces across a vegetation biodiversity gradient. [Submitted to *Fungal Ecology*]

Acknowledgements

I sincerely thank my supervisors—Professor Philip Weinstein and Professor Michelle Waycott from the School of Biological Sciences, and Professor Peng Bi from the School of Public Health—for their generous support and guidance during my PhD candidature. Phil, your unwavering enthusiasm kept me going. I would like to especially thank key collaborator Martin Breed who helped guide the microbiome-related components of my research and who was a key supporter of the mouse-soil exposure study described here.

I would like to thank my colleagues and friends Nick Gellie, Duncan Jardine, Jacob Mills, Harrison Sydnor, Christian Cando-Dumancela, Romy Dresken, Colette Blyth, James Loader, Leslie Marker, Celeste Hill, and Zdravko Baruch for their comradery and insightful discussions. I also thank members of the Advanced DNA Identification and Forensic Facility including Jennifer Young, Korjent van Dijk, Eleanor Dormontt, Caleb Coish, and Andrew Lowe for their helpful support and advice.

We performed an ambitious mouse model study that would not have been possible without contributions from many people. Special thanks go to colleagues Harrison, Chris and Romy who were critical to the delivery of this project. I must acknowledge the ever helpful support provided by Laboratory Animal Services staff including Denise Noonan, Gail Anderson, Jaimee Spurr, Tiffany Boehm, and Pacita Wissell. Colleagues from the Adelaide Medical School, Jiajun Liu, Jonathan Jacobsen and Mark Hutchinson were also instrumental to the success of the mouse study. Thanks also to Shaun Kennedy and SA Water for their support during the field work to gather soils that underpinned the study.

Also, I thank colleagues from the Healthy Urban Microbiome Initiative (HUMI): Martin, Phil, Jacob, Chris Skelly, Emily Flies and David Phillips for their shared passion and most exciting adventures, advocating for healthy urban microbiomes. A real highlight of my PhD was our trip to Sharm El Sheikh, Egypt, where we were hosted by the Secretariat of the

Convention on Biological Diversity to present our research integrating biodiversity with human health to the United Nations delegates at their conference of the parties (COP14).

My PhD would not have been possible without support provided by an Australian Government Research Training Program Scholarship; for this, I am most grateful. Most importantly, I sincerely thank my family—Kristy, Sarah, Keira and Macey—for allowing me the time to dedicate to this research.

Chapter 1. Introduction

Thesis structure

The body of this thesis comprises five papers that have either been published or submitted for publication. They are presented as the published version or in the format of the relevant journal, preceded by a title page and statement of authorship. Supplementary information is provided at the end of each chapter where relevant.

Here in *Chapter 1*, I outline the composition of my thesis on the topic of 'Biodiversity, environmental microbiomes and human health'. I present the aims of my research, provide background knowledge in the form of a literature review (see later this chapter), and briefly summarise the contents and flow of ideas for each of the subsequent chapters. Following the literature review, I present two environmental epidemiology studies (*Chapter 2* and *3*) which strengthen support for my hypotheses and provide focus for more detailed investigation. Detailed environment-microbiome relationships are then examined in *Chapter 4* before culminating in an experimental approach to test my hypothesis in a randomized controlled mouse study (*Chapter 5*). The mouse study was designed to examine potential mechanistic links suggested and informed by the preceding work. Lastly, *Chapter 6* presents a synthesis of my work, including highlighting key developments in methods and knowledge, discussing limitations of the approaches taken and identifying areas for future research. I also make recommendations and discuss implications from the work.

The literature review has been published in *BioScience* and starts by introducing the key ideas of the Biodiversity hypothesis (von Hertzen *et al.*, 2011) and Old Friends mechanism (Rook, 2013). Briefly, emerging evidence suggests that exposure to microbial diversity and perhaps key species (Old Friends) from biodiverse, natural environments may provide critical immune system training and regulation. Conversely, reduced contact with the right kind of environmental microbial communities (microbiota) and their genetic material

(microbiomes) may contribute to the modern growth in immunoregulatory disorders with potential to impact both infectious and non-communicable diseases. However, possible connections between biodiversity, environmental microbiomes and human health remain understudied due to their multidisciplinary nature. There is limited knowledge of the composition, modes of action, and environmental distribution of Old Friends. We do not know if it is microbial diversity *per se*, key species, or a combination of both, that may provide a protective effect for human health. Yet, importantly, the environment-host microbiota pathway offers promise for cost-effective population health interventions (e.g. through restoring biodiverse green space in cities) at a time when health care systems around the world are seeing unsustainable growth in utilisation and budget demands.

In the review, we set out a list of important research questions, reflecting existing knowledge gaps. These are listed below in an order that reflects the progress of my research:

a) Is landscape-scale biodiversity associated with human health outcomes?
b) Are different types, or conditions (qualities), of environment potentially more beneficial than others?
c) How might protective environmental influences compare with recognized drivers of human health such as socioeconomic status, diet, and lifestyle risk factors?
d) Can we identify and prioritize particular environment (or environmental change) and health associations to target subsequent detailed research?
e) What are the effects of macro- to landscape-scale environmental change and biodiversity loss on environmental microbiota?
f) Can we characterise environments through their microbiota?
g) Under what circumstances might environmental microbiota (or other microscale bioactive agents) be associated with health benefits?

The review also advocated the use of environmental proxies as a pragmatic investigation tool to express the multi-faceted nature of environments in order to investigate potential environment-health relationships among other known predictors. In this publication we suggested that soils have been under-represented in studies to-date as a source of potentially beneficial immunomodulatory environmental microbial diversity. Also, we proposed a unification of the biodiversity hypothesis and Old Friends mechanism. That is, key species might be expected to play an optimum role when complemented by a diverse community with greater capacity to maintain ecological control of potential pathogenic behaviour.

The research gaps and questions flagged in the review established the direction and focus for my thesis. Questions a), b) and c) were investigated in *Chapters 2* and *3*. The findings from *Chapters 2* and *3*, informed my response to question d). That is, I identified that further examination of the potential health impacts from exposure to biodiversity and soils, or more specifically biodiverse vs. non-biodiverse soils, was warranted—which led to the formulation of the study in *Chapter 5*. Questions e) and f) were investigated in *Chapter 4*. Question g) is broad and flagged our interest to better understand potential mechanistic links between environments, environmental microbiota and health outcomes. As my research was indicating a possible beneficial influence from biodiverse soils, testing for such a link via passive environmental exposure became a focus for designing the experiment in *Chapter 5*.

Chapter 2 is a paper published in *The Journal of Environmental Management* that, in essence, establishes a competition between a large number of candidate variables (including measures of landscape biodiversity) to prove their value as a predictor of respiratory health. I used least absolute shrinkage and selection operator (LASSO) penalised regression, a tool designed to select key variables, while discarding less important and correlated variables. I also quantified uncertainty by performing the modelling within a 10-fold resampling and cross-validation framework. In the results, predictors could be readily identified and ranked for comparison of effect size with other variables, from the size and sign of standardised

regression coefficients. We observed the variable ‘diversity of mapped vegetation classes’ was associated with improved respiratory health and ranked highly among known health predictors such as socio-economic status.

Chapter 3 is a paper published in *Science of the Total Environment*. In this work, I found that populations living near soils with high soil cation exchange capacity (CEC)—a proxy for potential immunomodulatory soil microbial diversity—experienced lower rates of infectious and parasitic disease. Populations living near higher CEC soils also experienced reduced health inequality that might otherwise be expected for low socioeconomic groups. The findings from *Chapters 2* and *3* supported the contention that exposure to biodiverse environments, and in particular biodiverse soils, is associated with health benefits.

Chapter 4 is a paper published in *Environment International*. This study explores the use of genus-level bacterial indicators from a chronosequence of ecological restoration (i.e. cleared land > revegetation > reference / remnant native vegetation) to inform ecosystem assessment and potential implications for human health. I developed a novel method—termed merged-sample bootstrap resampling—to re-analyse soil bacterial 16S rRNA microbiome data from a grassy woodland restoration study (Gellie *et al.* 2017). The new analysis method provided a deeper view of the available microbiome data than has previously been possible using conventional rarefying techniques (used to normalise sampling effort). I was able to identify key bacterial indicator groups of 'opportunistic' taxa that decreased with ecological restoration, and 'niche-adapted' taxa that increased. These diverging indicator groups also showed consistent differential abundance patterns across a comparative set of natural vs. human-altered soil samples from elsewhere across Australia. A number of human-associated opportunistic and pathogen-containing taxa were also found to be favoured in disturbed environments, yet reduced in mature, biodiverse environments. From this I theorised that ecological restoration—or restoring more biodiverse and natural ecosystems—may benefit human health through reducing exposure to potential pathogens.

Then in *Chapter 5*, presented as a manuscript submitted to *Proceedings of the Royal Society B*, I tested the role of environmental microbiomes as a causative linkage mechanism that may be underpinning the type of beneficial health associations as found in my earlier studies. Leading a project team, I undertook a randomised controlled mouse model experiment in the laboratory. Mice were exposed to airborne dust from high biodiversity soil, low biodiversity soil, or no soil (control). By monitoring changes in caecal and faecal microbiomes with time, and by carrying out behavioural experiments at the endpoint, we found significant changes to gut microbiota and reduced anxiety-like behaviour in females in the high biodiversity treatment. Among bacterial taxa that increased in the gut of high biodiversity treatment mice, I identified a putative spore-forming, anaerobic environmental microbe capable of producing a key metabolite, butyrate, which is associated with both gut health and mental health in humans. These findings provide experimental evidence supporting the hypothesis that environmental microbes can provide a mechanistic link between exposure to biodiverse environments and beneficial health outcomes.

In my conclusion, *Chapter 6*, I reflect on the contribution that *Chapters 1* to *5* have made to available knowledge and methods in this emerging field of multidisciplinary research. I also discuss problems and limitations encountered, highlight areas worthy of future research, make recommendations and discuss the implications arising from the synthesis of evidence contributed by my work.

Thesis aims and objectives

The primary aim of my thesis was to assess support and build evidence for proposed beneficial connections between biodiversity, environmental microbiomes and human health. In the event that support and evidence could be found, the secondary aim was to synthesise these findings to build on the knowledge base and inform policy that may underpin cost-

effective public health interventions with concurrent benefits for biodiversity conservation and restoration.

From these aims a series of more specific objectives were developed:

1. Evaluate support for beneficial biodiversity, environmental microbiome, and human health connections in existing national datasets.
2. From environmental epidemiology studies (above), provide direction to guide more detailed study of possible sources and mechanisms.

This work pointed to the potential beneficial immunomodulatory influence of biodiverse soils, which informed further objectives:

3. Determine if environments can be characterised by their soil microbiota, and if environmental-soil-microbiota patterns exist, what implications they have for human health?
4. Examine experimentally the potential transfer of ambient airborne soil environmental microbiota to host microbiota, evidence of health outcomes, and potential mechanistic agents.

Literature review

Statement of Authorship

Title of Paper	Environmental change and human health: Can environmental proxies inform the biodiversity hypothesis for protective microbial–human contact?
Publication Status	☑ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	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, 1023-1034

Principal Author

Name of Principal Author (Candidate)	Craig Liddicoat		
Contribution to the Paper	I led the conceptualisation of this review paper, reviewed and analysed the literature, drafted the descriptions and visualisations of existing and new emergent theories, wrote the first draft and led the revision of the final manuscript.		
Overall percentage (%)	85		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	6 May 2019

Co-Author Contributions

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

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

Name of Co-Author	Michelle Waycott		
Contribution to the Paper	Contributed to the conceptualisation, review structure, analysis and interpretation of existing and new emergent theories, paper writing and revisions.		
Signature		Date	13/6/19

Name of Co-Author	Philip Weinstein		
Contribution to the Paper	Contributed to the conceptualisation, review structure, analysis and interpretation of existing and new emergent theories, paper writing and revisions.		
Signature		Date	21/5/19

Please cut and paste additional co-author panels here as required.

Overview Articles

Environmental Change and Human Health: Can Environmental Proxies Inform the Biodiversity Hypothesis for Protective Microbial–Human Contact?

CRAIG LIDDICOAT, MICHELLE WAYCOTT, AND PHILIP WEINSTEIN

Microbiota from environmental sources overlap and interact with human microbiota, contribute to human microbial diversity, and provide beneficial immunomodulatory stimuli. Meanwhile, reduced diversity in human microbiota and immune dysregulation have been associated with a range of diseases. Emerging evidence suggests landscape-scale drivers of microbial diversity may influence our health, but the area remains understudied because of its multidisciplinary nature. Here, we attempt to widen the view on this subject by offering an environmental researcher's viewpoint, proposing a unifying conceptual framework to stimulate multidisciplinary interest. To focus research in this challenging area, we propose greater emphasis on multiscale ecological links and that landscape-scale proxies for potential underlying microbial mechanisms be investigated to identify key environmental attributes and health relationships worthy of subsequent detailed examination. Wherever possible, ecological epidemiological studies should account for the temporal nature of environmental microbiota exposures, especially with respect to the early development of the human commensal microbiota.

Keywords: environmental microbiota, immunoregulation, biodiversity, dysbiosis, microbial old friends

People often express an intuitive sense that being in nature is good for their health. In addition to well-established risk-exposure scenarios in environmental health, modern scientific approaches are increasingly discovering that there are a range of nontrivial mechanisms and co-benefits linking natural surroundings, biodiversity, and human health influence (box 1). Better understanding these relationships may have important implications for developing cost-effective and mutually beneficial outcomes to help address simultaneous challenges in public health and biodiversity conservation (von Hertzen et al. 2011, WHO and SCBD 2015).

The last mentioned mechanism in box 1 is among the least understood while also having wide potential to influence human health because of the largely hidden but ubiquitous nature of microbes (or microorganisms). Microbes have dominated the evolution of life and constitute a dominant portion of the Earth's living biomass and its genetic diversity (Whitman et al. 1998). Microbes feature in every habitat where life is possible. The various human *microbiotas*, or communities of microbes (e.g., gut, skin, airway, oral cavity, genito-urinary), exist in interdependent symbioses performing much of our metabolism (Wikoff et al. 2009). Their importance to human physiology is reflected in current knowledge that the combined human microbial genome (or microbiome) expresses over 100 times more genes than the human genome (Belizario and Napolitano 2015). Beneficial connections between microbiota and host health—influencing bodily development, mood, and stress responses—have been observed in both humans and animal models (Round and Mazmanian 2009, Rook et al. 2013, Belizario and Napolitano 2015). The human microbiota is believed to play an important role in normal human development (of organs, gut, immune system, bone, and brain) and actively participate in the homeostasis of the human body (McFall-Ngai et al. 2013). With important metabolic, immune, and nutritional roles, the human intestinal microbiota has been described as a "super-organism" (Purchiaroni et al. 2013).

BioScience 66: 1023–1034.
doi:10.1093/biosci/biw127 Advance Access publication 12 October 2016

Box 1. The broad mechanisms of environmental and environmental-change impacts on human health (WHO and SCBD 2015 and references therein).

(1) Increased exposure to anthropogenic hazards and environmental pollution
(2) Increased exposure to natural hazards, including
- harmful biotic agents: emerging infectious disease arising from land-use change, encroachment, biodiversity loss, and altered human–animal–environment dynamics
- physical hazards: due to reduced buffering from extreme weather and other natural disasters

(3) Declining food security and nutritional deficiency
(4) Lifestyle: health benefits from exercise and sunlight influenced by surrounding natural environments
(5) Mental health and social and cultural well-being: linked to natural surroundings and sense of place
(6) Global change: including climate change, globalization, and conflicts over depleting natural resources
(7) Biomedicines: loss of biodiversity-related potential new pharmaceuticals and traditional biomedicines
(8) Reduced contact with protective environmental microbial diversity

Dysbiosis of the human microbiota (i.e., reduced diversity or changes in composition, often with an increase in the ratio of pathogenic to commensal organisms) has been associated with a range of immunological, gastrointestinal, metabolic, psychiatric, and behavioral disorders observed in humans and animal models, as has been reviewed elsewhere (Round and Mazmanian 2009, Clemente et al. 2012, Parker and Ollerton 2013, Rook et al. 2013, Belizario and Napolitano 2015). Ongoing research into particular microbiota–disease associations is supporting the increasing recognition of host microbiota–mediated mechanisms across diverse disease outcomes, such as in obesity (Ridaura et al. 2013), type 2 diabetes (Forslund et al. 2015), rheumatoid arthritis (Zhang et al. 2015), stroke (Yin et al. 2015), depression (Zheng et al. 2016), and some cancers (Sivan et al. 2015, Vétizou et al. 2015).

Multifactorial influences are known to drive the composition and diversity of the human microbiota, including diet, genetics, antibiotic use, age, birth mode of delivery (natural or caesarean), and geographic location (Clemente et al. 2012, Voreades et al. 2014, Belizario and Napolitano 2015). However, at least a portion of the human microbiota is in dynamic exchange with microbes from the surrounding environment; therefore, natural microbial diversity is now appreciated as an important contributor to normal (healthy) human immunological (and potentially other aspects of homeostatic) functioning (von Hertzen et al. 2011, WHO and SCBD 2015). Emerging experimental evidence also supports this line of thinking. For example, mice exposed to soil, house dust, and decaying plants had enhanced gut microbial diversity and innate immunity when all other variables (diet, age, genetic background, physiological status, and original gut microbiota) were controlled for (Zhou et al. 2016). In a separate study, mice exposed to dog-associated house dust experienced changes in gut microbiome that were associated with protective immune responses against airway allergens and virus infection (Fujimura et al. 2014). Rook (2013, figure 3) suggested several potential pathways through which environmental microbiota might affect the human microbiota and/or provide immunomodulatory stimuli. These pathways may include transient contact or colonization, with either direct recognition by immune receptors or indirect responses following interactions that alter the host microbiota.

Having emerged, in an evolutionary sense, from largely natural and biologically diverse surroundings, a growing proportion of the global population are now surrounded by relatively depauperate (low biodiversity) urban, industrialized, or highly managed and largely monocultural agroecological landscapes. As we discuss later, these macroscale changes can translate to microscale changes in biodiversity and ecosystem composition (Adams and Wall 2000, Bulgarelli et al. 2013, Turner et al. 2013). Meanwhile, the science of aerobiology (e.g., Womack et al. 2010, Polymenakou 2012, Bowers et al. 2013) shows that human populations have a real biological connection to their ambient surroundings (in addition to any direct physical environmental contact). These lines of evidence suggest that different sources and compositions of environmental microbiota—through interactions with the human microbiota and other immunomodulatory pathways (von Hertzen et al. 2011, Rook 2013)—may inadvertently provide protective or adverse background influences on human health.

Indeed, many medical researchers, including the World Allergy Organization, now suggest that microbiota-mediated mechanisms—and disruption to these, arising from environmental change—at least partly explain the pandemic of allergic, auto-immune, and chronic inflammatory diseases (AACIDs, discussed later) occurring across developed nations in recent decades (Haahtela et al. 2013). Described variously as the microbial old friends (MOF) mechanism (Rook et al. 2013), the high microbial turnover hypothesis (Matricardi and Bonini 2000), the biodiversity hypothesis (von Hertzen et al. 2011), or the evolutionary mismatch of "biome depletion" (Parker and Ollerton 2013), it is suggested that a lack of microbial diversity—or the reduced contact with the right type of microbes (or MOF, as we discuss later)—in our modern surroundings is an

important contributor to the rising incidence of immune dysregulation underlying AACIDs and possibly a range of other diseases, including some cancers (Rook and Dalgleish 2011). Highlighting concerns (shared by von Hertzen et al. 2011) for the impacts of biodiversity loss leading to reduced immunoregulation from natural environments, Rook (2013) proposed that environmental microbiota (as supplements to MOF) may provide an unappreciated ecosystem service that is essential to our well-being, and that "this insight will allow green spaces to be designed to optimize health benefits and will provide impetus from health systems for the preservation of ecosystem biodiversity."

However, large gaps remain in our knowledge: "Hardly anything is known about the interactions between environmental and indigenous [host commensal] microbiotas" (Haahtela et al. 2013). There are still many unknowns concerning the protective roles and membership of MOF, their possible modes of action, and broader relationships with biodiversity and the surrounding environment (Stanwell-Smith et al. 2012, WHO and SCBD 2015). Important research questions in the context of potential environmental microbiota–mediated influences on human health include the following: (a) Can we characterize environments through their microbiota? (b) What are the effects of macro- to landscape-scale environmental change and biodiversity loss on environmental microbiota? (c) Is landscape-scale biodiversity associated with human health outcomes? (d) Are different types, or conditions (qualities), of environment potentially more beneficial than others? (e) How might protective environmental influences compare with recognized drivers of human health such as socioeconomic status, diet, and lifestyle risk factors? (f) Can we identify and prioritize particular environment (or environmental change) and health associations to target subsequent detailed research? (g) Under what circumstances might environmental microbiota (or other microscale bioactive agents) be associated with health benefits? Answers to these questions may help to build insight and hypotheses and prioritize research opportunities before tackling more detailed investigations of possible environmental microbiota–mediated mechanisms.

To date, the MOF mechanism has principally been investigated from a medical research focus, with limited emphasis placed on the potential role and analysis of broadscale ecology or environmental change. This is despite a call to "bridge the chasm between ecology and medicine/immunology" (Rook 2013). Here, we further the argument for greater integration of ecological insight and environmental analyses into studying potential protective environmental microbiota–mediated mechanisms. In particular, we suggest that a comprehensive examination of broadscale, spatially variable environmental attributes in the context of spatially distributed public-health data may advance knowledge in this area. If we adopt the view that microbial diversity in the environment should be viewed as an inherent ecosystem service that is essential to our well-being (Rook 2013) and that this can be related to environmental biodiversity (von Hertzen et al. 2011), then we suggest that protective health effects should be observable and associable with recognizable environmental attributes (e.g., land use, vegetation, soil types, and their diversity), acting as proxies for as-yet-unknown microbial agents and mechanisms.

Immunomodulation, "old friends," and the biodiversity hypothesis

Microbes play a key role in educating and regulating the immune system (Purchiaroni et al. 2013, Belizario and Napolitano 2015, WHO and SCBD 2015). Having co-evolved with a diverse range of microbes (and their metabolic and decay products) in the surrounding environment, the human immune system has needed to develop defense mechanisms against harmful pathogens, as well as tolerance mechanisms to other commonly encountered, and mostly harmless, microbial agents. As developed societies around the world have improved standards of sanitation, we have witnessed a decline in infectious diseases. However, in recent decades, this has been paralleled by a corresponding increase in AACIDs (Haahtela et al. 2013).

Initial attempts to explain this trend gained most attention via the hygiene hypothesis (Strachan 1989). However, this has since been revised and expanded and is perhaps most notably described in terms of the MOF mechanism (Rook et al. 2013). Alternatively, WHO and SCBD (2015) use the terminology "supplements to the human symbiotic microbiota from the natural environment." The MOF mechanism suggests that following prolonged microbial exposure over evolutionary timescales, a dependence evolved between the immune system of mammals and some microorganisms. Possibly, this involved ancestral humans losing the need for gene expression associated with essential functions that could be readily performed by these partner microorganisms. In particular, this concerns a key function of the immune system in recognizing when *not* to activate in order to avoid unwarranted and potentially self-harming inflammatory responses to the body's own cells and normally harmless microbes from the surrounding environment.

Exposure to a broad diversity of microorganisms following birth (e.g., from vaginal delivery, diet, human contact, and the environment) provides important training inputs to the human immune system (O'Hara and Shanahan 2006, Wopereis et al. 2014). Microbes are sampled by immune cells associated with mucosal barrier tissues, prompting the establishment of complex immunoregulatory circuits that balance inflammatory responses (to suppress dangerous pathogens) with tolerance mechanisms that induce, for example, anti-inflammatory cytokines (signaling proteins) and regulatory T cells (T_{reg}) in order to avoid undue responses to common antigens (O'Hara and Shanahan 2006, von Hertzen et al. 2011, Purchiaroni et al. 2013).

In contrast, AACIDs have been associated with immune dysfunction, dysbiosis, and inappropriate inflammatory

responses to (a) our own tissues, manifesting as autoimmune diseases such as type 1 diabetes, multiple sclerosis, and rheumatoid arthritis; (b) normally harmless allergens and foods, manifesting as allergic disorders, eczema, asthma, and hay fever; and (c) gut contents including commensals, manifesting as inflammatory bowel diseases such as ulcerative colitis and Crohn's disease. These associations are reviewed in detail elsewhere (e.g., Round and Mazmanian 2009, Clemente et al. 2012, Parker and Ollerton 2013, Purchiaroni et al. 2013, Belizario and Napolitano 2015). The risk of AACIDs may be further enhanced by lack of physical activity and sunlight, poor diet, pollution and other factors, which may act in synergy with dysbiosis of the gut flora (Stanwell-Smith et al. 2012, Haahtela et al. 2013). Haahtela and colleagues (2013) also reviewed and speculated on possible connections between dysbiosis and AACIDs. They suggested that it is possible that some common members of the normal (healthy) commensal microbiota may play an active role in the development of T_{reg} cells, responsible for mediating suppression of T-cell mediated inflammatory responses. They speculated that altered environmental microbiota may play a role in the development of dysbiosis, such as through the reduced signaling of pattern recognition receptors (used by the innate immune system to identify particular microbes and thereby amplify or suppress responses). Reduced immune signaling may then lead to immune dysfunction, which enhances the colonization and growth of a biased microbiota, thereby reinforcing the host–microbe interaction toward an unhealthy state (Haahtela et al. 2013).

Failing immunoregulatory mechanisms can also lead to continuous background inflammation, even without a specific chronic inflammatory disorder. Persistent raised levels of inflammatory mediators have been associated with increased susceptibility to a range of diseases including insulin resistance, metabolic syndrome, type 2 diabetes, obesity, cardiovascular disease, reduced stress resilience, and psychiatric disorders such as depression (Parker and Ollerton 2013, Rook et al. 2013, Belizario and Napolitano 2015). Several forms of cancer are also associated with increases in AACIDs, which may be explained because chronic inflammation provides growth factors and mediators that stimulate the vascularization and metastasis of tumors (Rook and Dalgleish 2011).

Temporal dimensions of human and environmental microbiota interactions also require consideration. Early stimulation is viewed as particularly crucial for supporting the maturation of immunoregulatory mechanisms (Wopereis et al. 2014) and dysbiosis during early developmental periods may have lasting adverse health impacts (Cox et al. 2014). However, immunoregulatory effects associated with dysbiosis (Parker and Ollerton 2013, Belizario and Napolitano 2015) and helminth infections (Versini et al. 2015) are also observed in later childhood and in adults, while the immune-boosting effects of mycobacteria (a suggested old friend) are known to be transient (Matthews and Jenks 2013), suggesting that ongoing diverse exposures are also important (Matricardi and Bonini 2000). Temporal effects are also discussed later in relation to the variability of environmental microbiota exposures and addressing confounders in more detailed work.

Drawing on multiple lines of evidence, von Hertzen and colleagues (2011) extended the notion of a MOF mechanism to suggest that "declining biodiversity might actually increase the risk to humanity from chronic diseases." This idea arises because transient beneficial members of the human microbiota overlap with environmental microbiota, suggesting a dynamic interaction with the environment. As has been reviewed elsewhere (von Hertzen et al. 2011, Stanwell-Smith et al. 2012, Haahtela et al. 2013, Rook 2013, WHO and SCBD 2015, and references therein), the grounds for the notion of a wider association among dysbiosis, AACIDs, and a lack of biodiverse microbial stimuli from the surrounding environment come from (a) metagenomic studies of the microbiota in the gut and other sites; (b) epidemiological studies on immigrants moving to more affluent but more depauperate countries; (c) urban–rural AACID comparative studies; and (d) studies of immunomodulatory effects due to epigenetic mechanisms, farm and livestock exposures, proximity to agricultural land, and exposure to pets. Reduced exposure to biodiverse environments and urban green space is also suggested to partly explain the higher incidence of AACIDs associated with lower socioeconomic status (Rook et al. 2014).

Emerging evidence lends support to von Hertzen and colleagues' (2011) biodiversity hypothesis. Hanski and colleagues (2012) found associations between atopic sensitization (allergic disposition), skin microbiota and surrounding land-use types in a random sample of 118 adolescents living in a heterogeneous 100-kilometer (km) by 150-km region of Finland. Atopic individuals had reduced generic diversity of gammaproteobacteria on the skin compared with healthy individuals. In contrast, healthy individuals showed a significant correlation between the relative abundance of the gammaproteobacterial genus *Acinetobacter* and expression of interleukin (IL)-10, a key anti-inflammatory cytokine in immune tolerance. Atopic sensitization was significantly explained by land use, decreasing with the amount of forested and agricultural land within 3 km of the study subjects' homes. In cohort studies from Finland and Estonia, Ruokolainen and colleagues (2015) found that land-use patterns explained 20% of the variation in the relative abundance of proteobacteria on the skin of healthy individuals, and the amount of green environment (forest and agricultural land had similar effects) was inversely associated with the risk of atopic sensitization in children. They concluded that "the environmental effect may be mediated via the effect of environmental microbiota on the commensal microbiota influencing immunotolerance." There are, however, limited studies of this type, and more research to test the biodiversity hypothesis in different environments is needed.

Why focus on environmental proxies?

Sources of microbial diversity in the natural environment include soil, vegetation, animals, and aquatic and marine environments. The environment is highly multifaceted, and here, we discuss landscape-scale attributes as potential drivers of environmental microbiota diversity and therefore health. We might expect macroscale environments to be linked to microscale environments through the provision of characteristic feedstocks and microhabitats. Changes to aboveground macroscale features can affect microscale ecosystem dynamics of terrestrial, freshwater, and marine systems through (a) changes in resource supply, (b) physical and structural habitat heterogeneity, (c) biotic (ecological) interactions, and (d) cross-surface migration of above- and belowground organisms (Adams and Wall 2000). Broadscale geographic variation may also contribute to variation in human microbiota; for example, Suzuki and Worobey (2014) suggested that higher latitude colder climates are associated with changing proportions of dominant bacterial phyla linked to increased body mass. The key environmental themes linked to sources of microbiota are highlighted below.

Vegetation and land use. Different plant species are associated with different microbiota of the phyllosphere and rhizosphere (i.e., microbial habitats of aerial vegetation and belowground roots respectively; Bulgarelli et al. 2013, Turner et al. 2013). The composition of and similarities between plant microbiota are driven by factors including (a) biochemically induced mutualism between particular plant and microbial species, (b) genetic relatedness between plants, (c) climate, (d) anthropogenic influences (e.g., pesticide use), and (e) spatial proximity (Bulgarelli et al. 2013, Bringel and Couée 2015). The connection between aboveground (macroscale) and belowground (microscale) components within terrestrial ecosystems typically results from powerful mutual feedback mechanisms. For example, plant characteristics will dictate organic matter inputs to soil microbiota while soil microbiota will dictate the breakdown and re-supply of nutrients to plants. These feedbacks will vary depending on the natural fertility and productivity of an ecosystem (Wardle et al. 2004) and also with anthropogenic changes in land use and management (Coleman et al. 2004).

It is also possible that a range of (nonmicrobe) microscale bioactive agents may provide immunomodulatory influences. For example, Li and colleagues (2006) found immune-boosting effects from phytoncides (wood essential oils), and Stanwell-Smith and colleagues (2012) suggested that protective agents may extend beyond the living MOF themselves, to include their cellular components (e.g., endotoxin), decay products, and metabolites. In view of this potential wider context for health influences from the environment, plants are also known to emit pollens, aerosols, and a wide variety of volatile organic compounds (VOCs; Bulgarelli et al. 2013). VOCs can promote or inhibit (and thus shape) adjacent microbial communities, whereas phyllosphere microbiota are also active in the production, interception, and alteration of various plant-related VOC emissions (Bringel and Couée 2015).

Animals. Through interactions with their surrounding environment, animals may inadvertently sample and collect a wide variety of environmental microbiota. The characteristic microbial sources most relevant to human interactions will likely include fur or hides and fecal matter. At the landscape scale, different vegetation and land-use types are often associated with different animals (e.g., livestock grazing or feedlots on agricultural land, native species in conservation areas, and pest species in poorly managed areas). Human exposure to animal microbiota will be influenced by proximity, the amount and volatility of source material, as well as prevailing winds for airborne microbiota. Bowers and colleagues (2013) measured airborne bacterial signatures of cow fecal microbiota in a rural city surrounded by agricultural land containing cattle feedlots. Exposure to pet dogs in early infancy has been shown to reduce the risk of childhood allergic disease development, and dog-associated house dust has been found to be associated with beneficial immunomodulatory effects (Fujimura et al. 2014). The microbiota associated with animals and farm exposures are further reviewed elsewhere (Stanwell-Smith et al. 2012, Rook 2013).

Soils. Soils are the most complicated biomaterial on the planet (Young and Crawford 2004). They support an immense diversity of microbes that remain largely unexplored, with drivers of variability in soil microbiota including variation in soil types and microhabitats (arising from environmental conditions, anthropogenic and organic inputs, and soil texture or clay content; Torsvik and Øvreås 2002). Microbes from soils have produced many of the most important medicinal drugs, including the majority of antibiotics and many anti-cancer compounds (Charlop-Powers et al. 2015). Soil eating (geophagy) is widespread in vertebrates and many human cultures, typically targeting clay-rich soils and suggested to provide protective health benefits (Young et al. 2011); this is consistent with the mechanisms discussed here. Particular soil constituents may have biological effects and seasonal mobilization patterns, such as has been shown in studies of coccidioidomycosis (valley fever) caused by a soil-dwelling fungus (Kolivras et al. 2001). Loss of contact with soil (and associated microbiota) has been suggested as a possible contributor to the rise in AACIDs arising from broadscale sealing of soils in urban developments (von Hertzen and Haahtela 2006).

Aside from soil itself, biological soil crusts can constitute up to 70% of the living groundcover across many diverse natural environments (Belnap and Lange 2001). These crusts form an aggregation of soil particles and cyanobacteria, algae, microfungi, lichens, and bryophytes that live in or on the top few millimeters of soil and may also be important contributors to beneficial human–environmental microbiota contact.

Coastal and marine environments. The marine microbiome is also diverse and largely unexplored, has biomass (cell densities) concentrated in near surface layers, and shares over 70% of microbial gene functionality with the human gut microbiome (Sunagawa et al. 2015). Mobilization, via aerosols, of bioactive substances associated with marine microorganisms—thus influencing the health of near-coastal human populations—is demonstrated through the adverse example of harmful algal blooms or red tides (Weinstein 2013).

Air. Aerobiology demonstrates there is a real biological connection between humans and ambient environmental microbiota. The air is alive with microbial diversity—including bacteria, viruses, fungi, pollen, and algae—and acts as a source of both pathogenic and beneficial microbes to humans (Womack et al. 2010, Polymenakou 2012). Spatial and temporal variability may be expected in the composition of airborne microbiota. From sampling the near-surface atmosphere across three distinct land-use types (agricultural fields, suburban areas, and forests), Bowers and colleagues (2011) found that the composition of airborne microbiota was significantly related to land-use type and that differences were likely driven by shifts in the sources of bacteria rather than by local meteorological conditions. Also, Bowers and colleagues (2013) observed seasonal fluctuations in the composition and sources of near-surface airborne microbes, with soils and leaves representing important microbial sources across both urban and rural sites and cow fecal bacteria (associated with neighboring feedlots) also featuring in the rural location on a seasonal basis. They observed that microbial sources varied in prominence under seasonal conditions, potentially explained by climatic conditions, deciduous plant growth and senescence, and seasonal soil disturbance from surrounding agricultural land-use practices. Continental-scale patterns in the distribution of dust-associated bacteria and fungi have also been observed (Barberán et al. 2015) where geographic patterns were associated with climatic and soil variables. That work also found that urban areas were exposed to more homogenized airborne microbiota compared with the geographic variability found across rural areas.

Given the preceding evidence of relationships connecting various ecosystems (or ecosystem components), environmental microbiota, and potential human health effects, it may be possible to improve human health outcomes through environmental management specifically targeting microbiota-mediated linking mechanisms. We envisage theoretical links among landscape-scale environmental change, environmental microbiota, and human health, as we show in figure 1. However, in order to prioritize where more detailed study of underlying mechanisms and/or public-health interventions might be most cost effectively targeted, it is first necessary to quantify the strength of associations between environmental exposures and health outcomes. In the absence of temporal change data, we might examine these relationships using spatial analogues (environmental mapping) for differences in the landscape. We suggest that the links depicted in figure 1, while not comprehensive, may provide a useful conceptual framework for multidisciplinary research. In the next section, we briefly discuss methodological approaches for establishing priorities and addressing confounders, and we outline subsequent more detailed approaches required to advance knowledge in this emerging field of study.

There are a number of factors supporting the use of environmental proxies to investigate the MOF mechanism and related biodiversity hypothesis. Knowledge of the membership of MOF is incomplete and inconsistent (Stanwell-Smith et al. 2012); it is not known whether biodiversity, total biomass, or the particular source or species of environmental microbe(s) is important (Rook 2013), and there are still considerable computational challenges and base-knowledge limitations in trying to characterize and understand the genetic makeup and biological function of complex natural environments such as soil (Howe et al. 2014). Knowledge is still building on the microbiota of different environments, such as through the Earth Microbiome Project (Gilbert et al. 2014). A range of previously unappreciated (nonmicrobial) microscale bioactive agents from the environment may also be contributing to human health. There is a growing consensus that living in close proximity to the natural environment can provide a broad range of health benefits (WHO and SCBD 2015), but what type of natural environment? And are some environments better than others? Using a spatial analogue approach may help answer this question.

A focus on identifying particular environmental microbiota–mediated mechanisms affecting human health may also be hampered by redundancy that is likely to be found both in the microbial agents providing immunomodulatory stimuli and human immune system pathways (Stanwell-Smith et al. 2012). Microbiota can drive epigenetic responses (Shenderov 2012), bringing further potential complexity and requisite expertise to the examination of underlying mechanisms. Required doses are unknown, and it may be that subclinical (asymptomatic) exposures are all that is required to deliver protective health benefits (Stanwell-Smith et al. 2012). If so, this poses a challenge as subclinical exposure is much harder to detect in epidemiological studies. This means for subsequent detailed study into underlying mechanisms, immunological markers will be important, not just disease outcomes.

When it comes to analyzing environmental exposures, the question of required dose is worth exploring further. This is because emerging knowledge of the predominantly beneficial role of microbes (as we discuss here) is at odds with the traditional focus of microbiology, concerned with the negative role of microbes in driving infectious disease. Here, we raise the possibility that hormetic, U-, or J-shaped dose–response relationships (i.e., characterized by low-dose stimulation and high-dose inhibition; Calabrese et al. 2007) may provide a bridging paradigm between protective MOF and the traditional toxicological view of common pathogenic microbes by spanning divergent health

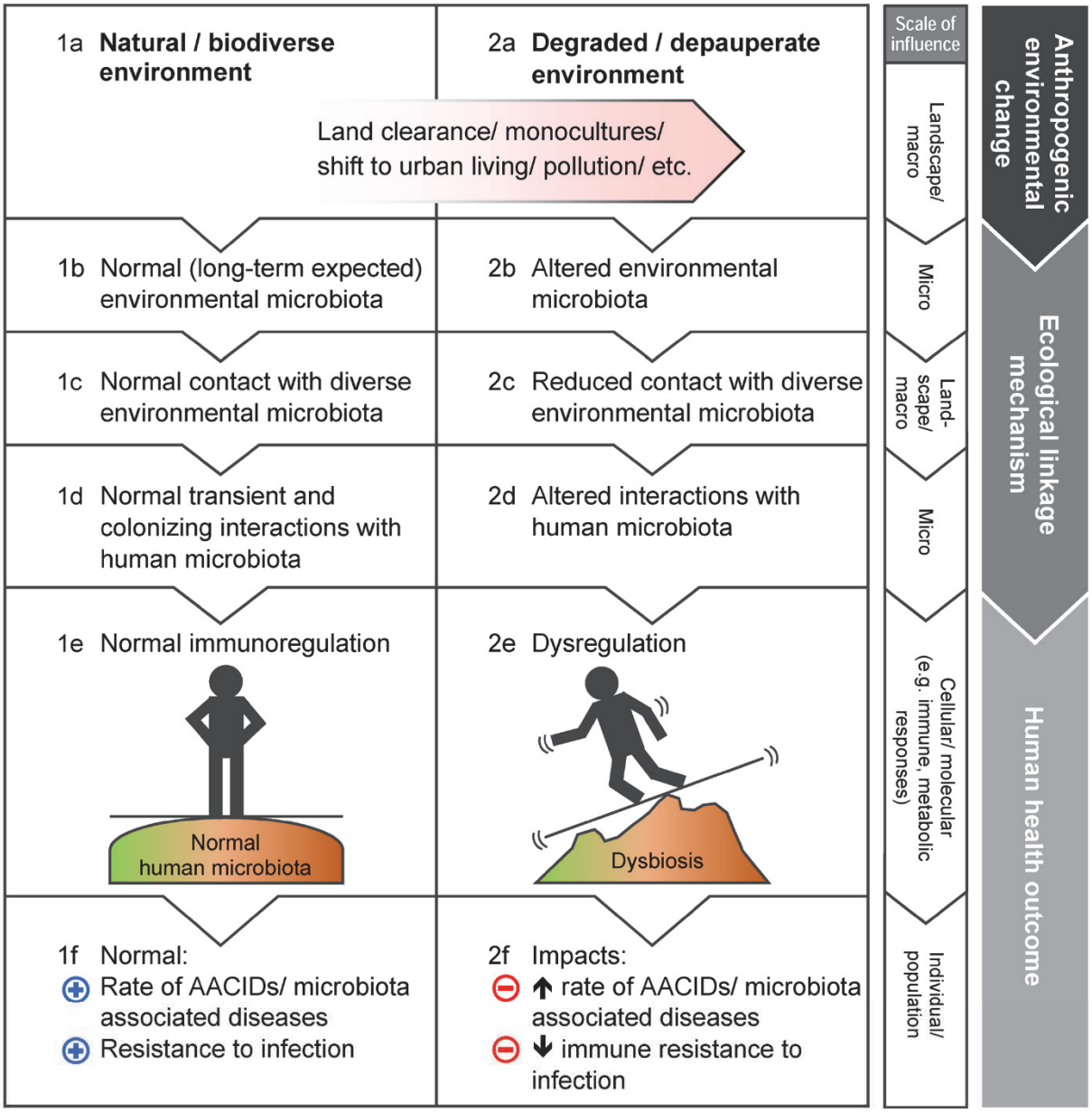


Figure 1. Theoretical multiscale links between environmental change, protective environmental microbiota, and human health. Environmental degradation (1a > 2a) alters feedstocks and microbial habitats (e.g., phyllosphere, rhizosphere, soil, and animals), altering microbial ecosystem dynamics and therefore the composition of environmental microbiota (2b). Consequently, exposure to environmental microbiota (2c) will differ from more natural and biodiverse surroundings (1c). Long-term expected (immunologically normal) interactions with human microbiota (1d) are therefore lacking in degraded environments (2d) and contribute to immune dysregulation (2e), autoimmune, allergic, and chronic inflammatory diseases (AACIDs) and other microbiota-associated diseases (2f). Environmental supplements to human microbiota are indicated in green shading; normally, these form part of a balanced human microbiota (1e) but are deficient or imbalanced in the case of dysbiosis (2e). The temporal effects (e.g., the timing and duration of environmental exposures) and other potential drivers of human microbiota (e.g., diet, genetics, and age) will also be important, but for simplicity, they are not included in this diagram.

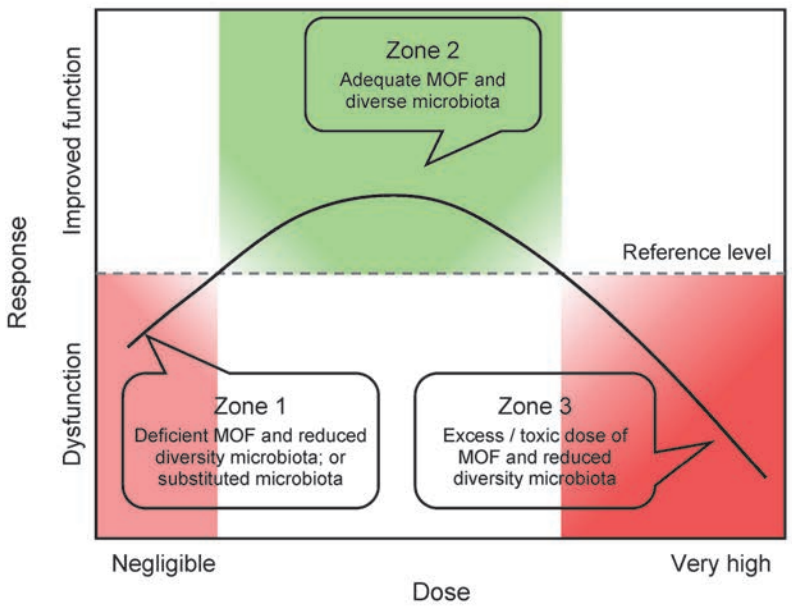


Figure 2. A theoretical dose–response curve for a generic microbial old friend (MOF), with inferred ecological context (as we discuss in the main text). Immune dysregulation and disease are associated with the absence of an MOF (zone 1). At some low to moderate dose, appropriate immune stimulation is provided to establish and maintain immunoregulatory circuits (zone 2). Pathogenic effects would be expected at increasingly high doses (zone 3). This theoretical curve follows a generalizable hormetic response proposed by Calabrese and colelagues (2007), except for a shift in the reference level that creates three response zones instead of two.

outcomes inferrable from varying microbial dosage rates. For example, known pathogens including *Escherichia coli*, *Helicobacter pylori*, species of *Salmonella* and *Staphylococcus*, enteroviruses, and parasitic helminth worms are among those microorganisms suggested to have protective roles (Stanwell-Smith et al. 2012). Calabrese and colleagues (2007) suggested that hormetic responses are generalizable and commonly encountered across a range of biological systems.

In figure 2, we conceptualize an idealized dose–response curve for a generic MOF. When otherwise expected, missing or very low doses of a MOF are associated with greater risk of AACIDs and related adverse health outcomes (zone 1). Nominally low to moderate doses are associated with protective benefits (zone 2; due to appropriate stimulation of immunoregulatory circuits). Increasingly elevated microbial doses are expected to be associated with disease (zone 3). Figure 2 mirrors Calabrese and colleagues' (2007) biphasic response curve except that instead of setting the reference response level at 100% of control (zero dose), by setting the reference response at some low–moderate dose range (perhaps corresponding to an evolutionary norm), three response zones instead of two are depicted. In this context, an evolutionary norm would correspond to long-term exposures to diverse environmental microbiota and natural allergens consistent with expected normal immunomodulation via the MOF mechanism. Such a curve also parallels the triphasic deficiency–adequacy–toxicity concept familiar in plant nutrition (Smith and Loneragan 1997), further supporting Calabrese and colleagues' (2007) claim that such hormetic curves are generalizable across many biological systems. Selecting spatial environmental attributes (proxies) that might mimic varying amplitudes of microbial exposure (e.g., soil erodibility and soil microbial activity indices) and using natural experiments may provide a means to test (or at least build support for) this hypothesis of a hormetic relationship.

Ecological interactions (e.g., competition, predation, mutualism, and commensalism) operate at the microbial scale (Coleman et al. 2004), so applying general principles, we might speculate on the ecological context corresponding to the three zones in figure 2. In zone 2, where some low–moderate concentration of the MOF is present, this would correspond to an evolutionary norm or long-term steady-state microbiota—consistent with the establishment of immunoregulatory norms. Such a long-term, well-established microbiota is also suggestive of a balanced composition with maximal biodiversity (and therefore buffering to change) compared with that of the other zones. In zone 1, we might envisage that environmental conditions or microbial ecosystem dynamics have reduced the populations of the particular MOF. Such a shift in environmental conditions (e.g., feedstocks, temperature, air, and moisture) will likely favor the proliferation of another microbial species to fill the vacant, or newly emergent, ecological niche. The loss of the MOF with a rise in some other remaining species would correspond to an overall reduction in biodiversity of the microbiota. Alternatively, the original environmental microbiota could have been largely substituted, for example, where people have moved from rural to urban areas. In zone 3, we might speculate that it is actually the particular MOF that has been favored by a shift in environmental conditions. This would be at the expense of reduced numbers or loss of other species, also corresponding to a loss of biodiversity. In this hypothetical scenario, it is interesting to note that appropriate, protective doses (and exposures) to a particular MOF might be entirely consistent with exposure to a high diversity environmental microbiota. However, a chain of evidence would be required to test this hypothesis in detail, such as we have outlined in figure 1.

Other known microbial ecology mechanisms also underlie the importance of microbiota composition and diversity. Greater

diversity suggests greater redundancy in gene functionality, as well as genetic adaptability (including horizontal gene transfer). Quorum sensing will also play a role, referring to intra- and interspecies signaling used to synchronize gene expression among bacterial groups to control production of, for example, antibacterial substances, disease-causing virulence factors, and immune-system suppressors (Belizario and Napolitano 2015). Alcock and colleagues (2014) suggested that through mechanisms such as quorum sensing, more abundant microbial species can coordinate their secretions to influence host mood and behavior and even manipulate host eating habits to increase their survival.

At the landscape scale, the highly faceted nature of the environment is reflected in the growing availability of diverse large-area environmental mapping data sets. Geographic information systems (GIS) are being increasingly used in epidemiology studies, including the use of spatial association (e.g., proximity analysis) to design surrogate exposure metrics to better understand environmental influences on disease (Nuckols et al. 2004). Environmental proxies to investigate possible relationships between environmental microbiota (and other microscale bioactive agents) and human health could include spatial measures (a) of relative exposure to particular environmental features or attributes (e.g., via GIS focal statistics calculations of proportions of different classes of vegetation, land cover, land use, or other themes within a predetermined neighborhood) that might subsequently be related to changes in airborne microbiota; (b) of biodiversity where we might expect to find positive correlations with human health outcomes; and (c) that might mimic ambient exposure to particular MOF (e.g., soil erodibility and soil microbial activity indices) in which we may find nonlinear (e.g., hormetic or U-shaped) relationships with health outcomes.

Spatial environmental mapping data vary from expert-assessed polygon-based thematic mapping to raster-based remote sensing data (with varying levels of processing and interpretability) and statistically based spatial predictive modeling or mapping for all manner of environmental attributes. Following the approach of McBratney and colleagues (2003), the predictive mapping of soil microbiota, for example, may be developed using a wide array of environmental variables as potential predictors. Predictors can be chosen to span various soil-influencing themes such as previously measured soil attributes, climate, organisms (including vegetation and land use), topography and terrain attributes, lithology, age, and spatial or geographic position. By extension, we might also consider a wide array of environmental variables as potential predictors, or proxies, for as-yet-undefined potential protective environmental microbiota and nonmicrobial influences, in a broadscale environmental correlation analysis with spatially defined public-health outcomes (or ecological epidemiological study). Such correlative studies could potentially involve tens to hundreds of environmental variables where many of these variables are often correlated. Traditional multivariate approaches such as principal-components analysis can deal with correlation in predictor variables; however, this may be at the expense of ease of interpretation, such as when attempting to compare the relative importance of environmental variables (which may have lower effect size) among other known public-health predictors (e.g., socioeconomic status and lifestyle risk factors).

Contemporary machine learning methods such as the least absolute shrinkage and selection operator (LASSO) penalized regression (Tibshirani 1996) are designed to tackle high dimensional problems with large numbers of (including often correlated) potential explanatory variables, and yield interpretable results. Using LASSO penalized regression modeling in the environmental correlation analysis of Liddicoat and colleagues (2015) enabled direct interpretation of the relative effect and direction of important environmental predictors from the size and sign of standardized regression coefficients. Using alternative methods such as the LASSO in ecological epidemiological studies may complement traditional multivariate approaches to highlight key environmental attributes to assist in hypothesis building and establishing priorities for subsequent work. Therefore, the availability of diverse environmental spatial mapping data sets, coupled with natural experiments that influence human health, can provide a wealth of data from which to draw key associations and thus point the way for subsequent studies to investigate causal links.

Limitations, confounders, and more detailed work

A complex interplay of factors can influence human health, including known confounders (e.g., socioeconomic status, diet, lifestyle risk factors, exercise, health support services, genetics, age, and sex) and environmental influences (box 1). A broad human health–environmental correlation analysis will obviously not restrict findings to environmental microbiota mechanisms. Follow-up work will be needed in those environments of interest to test connections (see figure 1) through characterizing environmental features and their related environmental microbiota, human exposures, interactions with human microbiota, immunomodulatory responses, and consequent human health responses. Natural experiments, whereby particular population groups can be found that provide inherent controls for other important confounding factors (e.g., diet, lifestyle, and antibiotic use) will assist this subsequent detailed work. Focusing analysis on lower socioeconomic groups—reflecting their stronger association with AACIDs (Rook et al. 2014)—or children—because of the important role of early immune stimulation (Wopereis et al. 2014)—may also assist in the identification of environmental microbiota–mediated health mechanisms.

Individual responses to environmental microbiota are expected to vary because of differences in host commensal microbiota. Studies in mice (Seedorf et al. 2014) investigating the colonization of host microbiota have shown that established indigenous host microbiotas are resilient to perturbation and resist colonization by foreign microbiota.

However, they also found that in the case of germfree or gnotobiotic (with no or limited known microbiota) mice, with a limited suite of environmental microbiota sources, there are reproducible selective processes that can drive initially disparate host microbiota compositions of separate co-housed animals to converge to similar phylogenetic structures. This included colonization of host gut microbiota by foreign microbes from highly divergent environmental habitats (e.g., soil microbiota). From this, we might speculate that the immunomodulatory influence of environmental microbiota could be greatest on individuals with immature or compromised (dysbiotic) commensal microbiota. Voreades and colleagues (2014) found that short-term diet interventions may transiently alter the gut microbiota composition but that long-term diet changes are required to shift to a new steady state. If we were to extrapolate these results more broadly, this could suggest that the lasting protective influences of environmental microbiota may depend on long-term exposures. Important temporal factors would need to be accounted for in any subsequent detailed work, such as (a) the timing and duration of exposure to potential beneficial environmental microbiota, (b) seasonal variations in environmental microbiota sources, and (c) short-term fluctuations, succession, and maturation in host commensal microbiota (Clemente et al. 2012, Wopereis et al. 2014).

Recognized health drivers may also be correlated with underlying environmental variables. For example, biodiversity and landscape productivity can be drivers of local economic activity, which in turn can drive higher socioeconomic status of communities. Investigating a large number of (including often correlated) environmental variables presents obvious challenges in attempting to identify links between health outcomes and microbiota-associated environmental proxies. Despite this challenge, medical researchers are calling for new approaches that invest ecological knowledge (Rook 2013) and investigate multiple interacting environmental influences that may potentially act across multiple health outcomes (Myers et al. 2013).

In spatial epidemiology, there are typically trade-offs between the availability and spatial resolution of health and key contextual data. Often, some level of spatial aggregation may occur for privacy or data-reliability purposes; therefore, environmental parameters will also need to be summarized to match the available area-based health data. In these situations, there can be difficulty in separating influences because of the scale and availability of data as well as spatial variability versus differences in ecological processes. Also, Ruokolainen and colleauges (2015) found that the spatial scale of land-use description affects the ability to detect a significant relationship between land-use gradients and allergic disorder; statistically significant relationships were observed at intermediate scales from 2 km to 5 km. The potential for ecological bias and ecological fallacy (Elliot et al. 2000) also needs to be recognized. When possible environmental proxies for MOF via spatial mapping are examined, these will at best represent potential exposure (not dose). Mapping data for environmental variables is often extrapolated from limited field-truthed sites to provide exhaustive spatial coverages. Such data often carry uncertainty that is unquantified but may represent the best available knowledge. These limitations need to be borne in mind but should not be seen as roadblocks for the purpose of hypothesis building and pointing to areas where more detailed research is required.

A sequence of progressively detailed studies is envisaged (in the context of potential links in figure 1). As we outline here, we recommend that broadscale environmental correlation analyses be examined to firstly identify particular environments and health outcome scenarios of interest. Where possible, this may take advantage of existing environmental and public-health data sets. In areas of interest, prospective epidemiological cohort studies are then recommended when possible to further establish possible associations between environmental exposures and possible protective health outcomes. Studies will need to account for recognized confounders (e.g., demographics, diet, social indicators, lifestyle risk factors such as smoking status, and environmental pollution); temporal factors including the timing, duration, and seasonality of environmental exposures; and incorporate immune biomarkers to track asymptomatic (or subclinical) exposures. As we discussed earlier, a focus on children (reflecting the importance of early immune system development) and/or lower socioeconomic groups (reflecting a stronger association with AACIDs) may also assist in the identification of environmental microbiota–health mechanisms.

More detailed ecological epidemiological studies based on environmental proxies may lead to hypotheses that can subsequently be tested with experimental studies on animal models. This has been demonstrated elsewhere; for example, Hanski and colleagues (2012) reported a special role for the gammaprotebacterial genus *Acinetobacter* in enhancing immunotolerance and in increasing the expression of anti-inflammatory cytokine IL-10. Subsequently, Fyhrquist and colleagues (2014) reported strong support for this hypothesis with a mouse model. Further research to understand relationships between potentially beneficial environmental microbiota and corresponding recognizable environmental features (e.g., plant species and soil types) will also benefit subsequent implementation of public-health policy, such as in translating new knowledge of protective environmental microbiota–mediated mechanisms into new urban green-space design.

Conclusions

Microbes and other microscale bioactive agents provide a real biological connection to our surrounding environment and represent an understudied influence on human health. Emerging evidence suggests that microbial old friends and/or diverse environmental microbiota supplement human microbiota and may provide protective

background immunomodulatory stimuli, whereas their absence may play a role in dysbiosis, immune dysregulation, and disease. To advance understanding, we advocate the use of environmental proxies as a pragmatic investigation tool. In this, we suggest that soils have been underrepresented in studies to date as a source of environmental microbial diversity with the potential for a protective role in microbiota-mediated human health. Similarly, the influence of different types of vegetation, land cover, and land use (among other themes) also remain largely untested. We suggest that comprehensive environmental correlation analyses examining recognizable environmental attributes and allergic, autoimmune, and chronic inflammatory diseases (as well as other dysbiosis-associated diseases) could help build understanding and provide greater focus for subsequent detailed studies of potential underlying microbiota-mediated mechanisms. In this way, we can advance the "eating of the elephant"—that is, we can provide a first step. The timing and duration of environmental microbiota exposures also require consideration, with respect to the establishment, maturation, and long-term stability of the human commensal microbiota. Knowledge gaps regarding potential sources of microbial old friends and their relationship with recognizable features in the environment need to be addressed, for example, to prescribe new urban design (green-space) health treatments. In short, it remains to be demonstrated convincingly that landscape-scale environmental influences can affect our human microbiota and health. However, the public-health implications of such a connection warrant further research into this area. Such work will ultimately inform concurrent improvements in environmental stewardship, biodiversity conservation, and human health.

Acknowledgments

We thank the anonymous reviewers and editors from *BioScience* and Professor Peng Bi from the University of Adelaide School of Public Health for their thoughtful comments and suggestions on earlier versions of this manuscript.

References cited

Adams GA, Wall DH. 2000. Biodiversity above and below the surface of soils and sediments: Linkages and implications for global change. BioScience 50: 1043–1048.

Alcock J, Maley CC, Aktipis CA. 2014. Is eating behavior manipulated by the gastrointestinal microbiota? Evolutionary pressures and potential mechanisms. Bioessays 36: 940–949.

Barberán A, Ladau J, Leff JW, Pollard KS, Menninger HL, Dunn RR. 2015. Continental-scale distributions of dust-associated bacteria and fungi. Proceedings of the National Academy of Sciences 112: 5756–5761.

Belizario JE, Napolitano M. 2015. Human microbiomes and their roles in dysbiosis, common diseases, and novel therapeutic approaches. Frontiers in Microbiology 6 (art. 1050).

Belnap J, Lange OL, eds. 2001. Biological Soil Crusts: Structure, Function, and Management. Springer.

Bowers RM, 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. ISME Journal 5: 601–612.

Bowers RM, Clements N, Emerson JB, Wiedinmyer C, Hannigan MP, Fierer N. 2013. Seasonal variability in bacterial and fungal diversity of the near-surface atmosphere. Environmental Science and Technology 47: 12097–12106.

Bringel F, Couée I. 2015. Pivotal roles of phyllosphere microorganisms at the interface between plant functioning and atmospheric trace gas dynamics. Frontiers in Microbiology 6 (art. 486).

Bulgarelli D, Schlaeppi K, Spaepen S, Themaat EVLv, Schulze-Lefert P. 2013. Structure and functions of the bacterial microbiota of plants. Annual Review of Plant Biology 64: 807–838.

Calabrese EJ, et al. 2007. Biological stress response terminology: Integrating the concepts of adaptive response and preconditioning stress within a hormetic dose–response framework. Toxicology and Applied Pharmacology 222: 122–128.

Charlop-Powers Z, Owen JG, Reddy BVB, Ternei MA, Guimarães DO, de Frias UA, Pupo MT, Seepe P, Feng Z, Brady SF. 2015. Global biogeographic sampling of bacterial secondary metabolism. eLife 4 (art. e05048).

Clemente JC, Ursell LK, Parfrey LW, Knight R. 2012. The impact of the gut microbiota on human health: An integrative view. Cell 148: 1258–1270.

Coleman D, Crossley D, Hendrix P. 2004. Fundamentals of Soil Ecology, 2nd ed. Elsevier.

Cox LM, et al. 2014. Altering the intestinal microbiota during a critical developmental window has lasting metabolic consequences. Cell 158: 705–721.

Elliot P, Wakefield J, Best N, Briggs D, eds. 2000. Spatial Epidemiology: Methods and Applications. Oxford University Press.

Forslund K, et al. 2015. Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. Nature 528: 262–266.

Fujimura KE, et al. 2014. House dust exposure mediates gut microbiome *Lactobacillus* enrichment and airway immune defense against allergens and virus infection. Proceedings of the National Academy of Sciences 111: 805–810.

Fyhrquist N, Ruokolainen L, Suomalainen A, Lehtimaki S, Veckman V, Vendelin J. 2014. Acinetobacter species in the skin microbiota protects from allergic sensitization and inflammation. Journal of Allergy and Clinical Immunology 134: 1301–1309.

Gilbert JA, Jansson JK, Knight R. 2014. The Earth Microbiome project: Successes and aspirations. BMC Biology 12 (art. 69).

Haahtela T, Holgate S, Pawankar R, Akdis CA, Benjaponpitak S, Caraballo L. 2013. The biodiversity hypothesis and allergic disease: World allergy organization position statement. World Allergy Organ Journal 6: 3.

Hanski I, et al. 2012. Environmental biodiversity, human microbiota, and allergy are interrelated Proceedings of the National Academy of Sciences 109: 8334–8339.

Howe AC, Jansson JK, Malfatti SA, Tringe SG, Tiedje JM, Brown CT. 2014. Tackling soil diversity with the assembly of large, complex metagenomes. Proceedings of the National Academy of Sciences 111: 4904–4909.

Kolivras KN, Johnson PS, Comrie AC, Yool SR. 2001. Environmental variability and coccidioidomycosis (valley fever). Aerobiologia 17: 31–42.

Li Q, Nakadai A, Matsushima H, Miyazaki Y, Krensky A, Kawada T, Morimoto K. 2006. Phytoncides (wood essential oils) induce human natural killer cell activity. Immunopharmacology and Immunotoxicology 28: 319–333.

Liddicoat C, Maschmedt D, Clifford D, Searle R, Herrmann T, Macdonald LM, Baldock J. 2015. Predictive mapping of soil organic carbon stocks in South Australia's agricultural zone. Soil Research 53: 956–973.

Matricardi P, Bonini S. 2000. High microbial turnover rate preventing atopy: A solution to inconsistencies impinging on the hygiene hypothesis? Clinical And Experimental Allergy 30: 1506–1510.

Matthews DM, Jenks SM. 2013. Ingestion of Mycobacterium vaccae decreases anxiety-related behavior and improves learning in mice. Behavioural Processes 96: 27–35.

McBratney AB, Mendonça Santos ML, Minasny B. 2003. On digital soil mapping. Geoderma 117: 3–52.

Overview Articles

McFall-Ngai M, et al. 2013. Animals in a bacterial world, a new imperative for the life sciences. Proceedings of the National Academy of Sciences 110: 3229–3236.

Myers SS, Gaffikin L, Golden CD, Ostfeld RS, H. Redford K, H. Ricketts T, Turner WR, Osofsky SA. 2013. Human health impacts of ecosystem alteration. Proceedings of the National Academy of Sciences 110: 18753–18760.

Nuckols JR, Ward MH, Jarup L. 2004. Using geographic information systems for exposure assessment in environmental epidemiology studies. Environmental Health Perspectives 112: 1007–1015.

O'Hara AM, Shanahan F. 2006. The gut flora as a forgotten organ. EMBO Reports 7: 688–693.

Parker W, Ollerton J. 2013. Evolutionary biology and anthropology suggest biome reconstitution as a necessary approach toward dealing with immune disorders. Evolution, Medicine, and Public Health 2013: 89–103.

Polymenakou PN. 2012. Atmosphere: A source of pathogenic or beneficial microbes? Atmosphere 3: 87–102.

Purchiaroni F, Tortora A, Gabrielli M, Bertucci F, Gigante G, Ianiro G, Ojetti V, Scarpellini E, Gasbarrini A. 2013. The role of intestinal microbiota and the immune system. European Review for Medical and Pharmacological Sciences 17: 323–333.

Ridaura VK, et al. 2013. Cultured gut microbiota from twins discordant for obesity modulate adiposity and metabolic phenotypes in mice. Science 341 (art. 1241214).

Rook G. 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: 18360–18367.

Rook G, Dalgleish A. 2011. Infection, immunoregulation, and cancer. Immunological Reviews 240: 141–159.

Rook G, Lowry C, Raison C. 2013. Microbial "old friends," immunoregulation, and stress resilience. Evolution, Medicine, and Public Health 2013: 46–64.

Rook G, Raison C, Lowry C. 2014. Microbial "old friends," immunoregulation, and socioeconomic status. Clinical and Experimental Immunology 177: 1–12.

Round JL, Mazmanian SK. 2009. The gut microbiome shapes intestinal immune responses during health and disease. Nature Reviews: Immunology 9: 313–323.

Ruokolainen L, et al. 2015. Green areas around homes reduce atopic sensitization in children. Allergy 70: 195–202.

Seedorf H, et al. 2014. Bacteria from diverse habitats colonize and compete in the mouse gut. Cell 159: 253–266.

Shenderov BA. 2012. Gut indigenous microbiota and epigenetics. Microbial Ecology in Health and Disease 2012: 23.

Sivan A, et al. 2015. Commensal Bifidobacterium promotes antitumor immunity and facilitates anti–PD-L1 efficacy. Science 350: 1084–1089.

Smith F, Loneragan J. 1997. Interpretation of plant analysis: Concepts and principles. Pages 3–33 in Reuter D, Robinson J, eds. Plant Analysis: An Interpretation Manual, 2nd ed. CSIRO.

Stanwell-Smith R, Bloomfield S, Rook G. 2012. The hygiene hypothesis and its implications for home hygiene, lifestyle and public health. International Scientific Forum on Home Hygiene. (30 August 2016; *www.ifh-homehygiene.org*)

Strachan DP. 1989. Hay fever, hygiene, and household size. British Medical Journal 299: 1259–1260.

Sunagawa S, et al. 2015. Structure and function of the global ocean microbiome. Science 348 (art. 1261359).

Suzuki TA, Worobey M. 2014. Geographical variation of human gut microbial composition. Biology Letters 10 (art. 20131037).

Tibshirani R. 1996. Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society B 58: 267–288.

Torsvik V, Øvreås L. 2002. Microbial diversity and function in soil: From genes to ecosystems. Current Opinion in Microbiology 5: 240–245.

Turner TR, Ramakrishnan K, Walshaw J, Heavens D, Alston M, Swarbreck D, Osbourn A, Grant A, Poole PS. 2013. Comparative metatranscriptomics reveals kingdom level changes in the rhizosphere microbiome of plants. ISME Journal 7: 2248–2258.

Versini M, Jeandel P-Y, Bashi T, Bizzaro G, Blank M, Shoenfeld Y. 2015. Unraveling the hygiene hypothesis of helminthes and autoimmunity: Origins, pathophysiology, and clinical applications. BMC Medicine 13 (art. 81).

Vétizou M, et al. 2015. Anticancer immunotherapy by CTLA-4 blockade relies on the gut microbiota. Science 350: 1079–1084.

Von Hertzen L, Haahtela T. 2006. Disconnection of man and the soil: Reason for the asthma and atopy epidemic? Journal of Allergy and Clinical Immunology 117: 334–344.

Von Hertzen L, Hanski I, Haahtela T. 2011. Natural immunity: Biodiversity loss and inflammatory diseases are two global megatrends that might be related. EMBO Reports 12: 1089–1093.

Voreades N, Kozil A, Weir T. 2014. Diet and the development of the human intestinal microbiome. Frontiers in Microbiology 5 (art. 494).

Wardle DA, Bardgett RD, Klironomos JN, Heikki S, van der Putten WH, Wall DH. 2004. Ecological linkages between aboveground and belowground biota. Science 304: 1629–1633.

Weinstein P. 2013. Red tides. Page 826 in Bobrowsky PT, ed. Encyclopaedia of Natural Hazards. Springer.

Whitman WB, Coleman DC, Wiebe WJ. 1998. Prokaryotes: The unseen majority. Proceedings of the National Academy of Sciences 95: 6578–6583.

[WHO] World Health Organization, [SCBD] Secretariat of the Convention on Biological Diversity. 2015. Connecting Global Priorities: Biodiversity and Human Health: A State of Knowledge Review. WHO.

Wikoff WR, Anfora AT, Liu J, Schultz PG, Lesley SA, Peters EC, Siuzdak G. 2009. Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites. Proceedings of the National Academy of Sciences 106: 3698–3703.

Womack AM, Bohannan BJM, Green JL. 2010. Biodiversity and biogeography of the atmosphere. Philosophical Transactions: Biological Sciences 365: 3645–3653.

Wopereis H, Oozeer R, Knipping K, Belzer C, Knol J. 2014. The first thousand days—Intestinal microbiology of early life: Establishing a symbiosis. Pediatric Allergy and Immunology 25: 428–438.

Yin J, et al. 2015. Dysbiosis of gut microbiota with reduced trimethylamine-N-oxide level in patients with large-artery atherosclerotic stroke or transient ischemic attack. Journal of the American Heart Association 4 (art. e002699).

Young IM, Crawford JW. 2004. Interactions and self-organization in the soil–microbe complex. Science 304: 1634–1637.

Young SL, Sherman PW, Lucks JB, Pelto GH. 2011. Why on earth? Evaluating hypotheses about the physiological functions of human geophagy. Quarterly Review of Biology 86: 97–120.

Zhang X, et al. 2015. The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. Nature Medicine 21: 895–905.

Zheng P, et al. 2016. Gut microbiome remodeling induces depressive-like behaviors through a pathway mediated by the host's metabolism. Molecular Psychiatry 21: 786–796.

Zhou D, et al. 2016. Exposure to soil, house dust, and decaying plants increases gut microbial diversity and decreases serum immunoglobulin E levels in BALB/c mice. Environmental Microbiology 18: 1326–1337.

Craig Liddicoat (craig.liddicoat@adelaide.edu.au) is a postgraduate researcher with the School of Biological Sciences at the University of Adelaide, in Australia, and a senior natural resource management scientist with the Department of Environment, Water, and Natural Resources. Michelle Waycott is a professor of plant systematics with the School of Biological Sciences at the University of Adelaide and chief botanist with the State Herbarium of South Australia at the Department of Environment, Water, and Natural Resources. Philip Weinstein is head of the School of Biological Sciences at the University of Adelaide.

Statement of Authorship

Title of Paper	Landscape biodiversity correlates with respiratory health in Australia
Publication Status	☑ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted w ork w ritten in manuscript style
Publication Details	Liddicoat, C., Bi, P., Waycott, M., Glover, J., Lowe, A.J., Weinstein, P., 2018. Landscape biodiversity correlates with respiratory health in Australia. *Journal of Environmental Management*. 206, 113-122

Principal Author

Name of Principal Author (Candidate)	Craig Liddicoat		
Contribution to the Paper	I led the conceptualisation, data collection and preparation, performed the modelling, led the analysis and interpretation, wrote the first draft, and led the revisions of the final manuscript.		
Overall percentage (%)	85		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	6 May 2019

Co-Author Contributions

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

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

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

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

Name of Co-Author	Peng Bi		
Contribution to the Paper	Contributed to the conceptualisation, analysis and interpretation of results, paper writing and revisions.		
Signature		Date	24-05-2019

Name of Co-Author	Michelle Waycott		
Contribution to the Paper	Contributed to the conceptualisation, analysis and interpretation of results, paper writing and revisions.		
Signature		Date	13/6/19

Name of Co-Author	John Glover		
Contribution to the Paper	Contributed to the data collection and preparation, analysis and interpretation of results, paper writing and revisions.		
Signature		Date	23/5/19

V

Name of Co-Author	Andrew Lowe		
Contribution to the Paper	Contributed to the conceptualisation, analysis and interpretation of results, paper writing and revisions.		
Signature		Date	8-6-19

Name of Co-Author	Philip Weinstein		
Contribution to the Paper	Contributed to the conceptualisation, analysis and interpretation of results, paper writing and revisions.		
Signature		Date	21/5/19

Please cut and paste additional co-author panels here as required.

Liddicoat, C., Bi, P., Waycott, M., Glover, J., Lowe, A. J., & Weinstein, P. (2018). Landscape biodiversity correlates with respiratory health in Australia. *Journal of Environmental Management*, 206, 113-122. https://doi.org/10.1016/j.jenvman.2017.10.007

Journal of Environmental Management 206 (2018) 113–122

Contents lists available at ScienceDirect

Journal of Environmental Management

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

Research article

Landscape biodiversity correlates with respiratory health in Australia

Craig Liddicoat [a, b, *], Peng Bi [c], Michelle Waycott [a, b], John Glover [d], Andrew J. Lowe [a], Philip Weinstein [a]

[a] School of Biological Sciences and the Environment Institute, The University of Adelaide, North Terrace, Adelaide, SA 5005, Australia
[b] Department of Environment, Water and Natural Resources, GPO Box 1047, Adelaide, SA 5001, Australia
[c] School of Public Health, The University of Adelaide, North Terrace, Adelaide, SA 5005, Australia
[d] Public Health Information Development Unit, Torrens University Australia, Level 1, 200 Victoria Square, Adelaide, SA 5000, Australia

ARTICLE INFO

Article history:
Received 8 April 2017
Received in revised form
12 September 2017
Accepted 6 October 2017

Keywords:
Biodiversity
Ecological epidemiology
Biodiversity hypothesis
Lasso

ABSTRACT

Megatrends of urbanisation and reducing contact with natural environments may pose a largely unappreciated risk to human health, particularly in children, through declining normal (healthy) immunomodulatory environmental exposures. On the other hand, building knowledge of connections between environments, biodiversity and human health may offer new integrated ways of addressing global challenges of rising population health costs and declining biodiversity. In this study we are motivated to build insight and provide context and priority for emerging research into potential protective (e.g. immunomodulatory) environmental exposures. We use respiratory health as a test case to explore whether some types and qualities of environment may be more beneficial than others, and how such exposures may compare to known respiratory health influences, via a cross-sectional ecological epidemiology study for the continent of Australia. Using Lasso penalized regression (to interpret key predictors from many candidate variables) and 10-fold cross-validation modelling (to indicate reproducibility and uncertainty), within different socio-geographic settings, our results show surrogate measures of landscape biodiversity correlate with respiratory health, and rank amongst known predictors. A range of possible drivers for this relationship are discussed. Perhaps most novel and interesting of these is the possibility of protective immunomodulatory influence from microbial diversity (suggested by the understudied 'biodiversity hypothesis') and other bioactive agents associated with biodiverse environments. If beneficial influences can be demonstrated from biodiverse environments on immunomodulation and human health, there may be potential to design new cost-effective nature-based health intervention programs to reduce the risk of immune-related disease at a population level. Our approach and findings are also likely to have use in the evaluation of environment and health associations elsewhere.

1. Introduction

There is growing awareness of the numerous mechanisms and co-benefits that link natural and biodiverse environments with human health. It is important to understand such connections given the global challenges of escalating population health costs and declining biodiversity (WHO and SCBD, 2015). As reviewed elsewhere (Craig et al., 2016; Keniger et al., 2013; Myers et al., 2013; Sandifer et al., 2015; WHO and SCBD, 2015), there are broad and interacting mechanisms of human health impact from environmental change, many of which are well-studied. However, some potentially important environmental influences on human health remain understudied due to their multidisciplinary nature. A key example is described by the biodiversity hypothesis (von Hertzen et al., 2011), and related microbial 'old friends' mechanism (Rook, 2013), as recognized by the World Health Organization (WHO and SCBD, 2015) and World Allergy Organization (Haahtela et al., 2013), which highlights that environmental microbiota (or communities of microorganisms from the surrounding environment) overlap and interact with human commensal microbiota, contribute to human microbial diversity and may provide important beneficial immunomodulatory roles. As discussed later, there are potential links between different environments, their

* Corresponding author. School of Biological Sciences, The University of Adelaide, North Terrace, Adelaide, SA 5005, Australia.
E-mail address: craig.liddicoat@adelaide.edu.au (C. Liddicoat).

https://doi.org/10.1016/j.jenvman.2017.10.007
0301-4797/

microbiotas and other possible bioactive agents (e.g. volatile organic compounds, VOCs; air ions), and via direct and aerobiological exposures, possible immunomodulatory effects and human health influences (e.g. see Liddicoat et al. (2016), their Fig. 1).

The possibility of populations receiving some level of inadvertent ambient beneficial or adverse immunomodulatory influence associated with different types and qualities of environment, highlights a potentially important gap in our awareness of possible links between environments and human health. Megatrends of urbanisation and reducing contact with natural environments may pose a largely unappreciated risk to human health through declining normal (healthy) immunomodulatory environmental exposures. Children may be particularly impacted by inadequate exposures during the critical early period of immune system development (Wopereis et al., 2014) where immunoregulatory commensal microbiota can be acquired through random environmental encounters (Artis, 2008). For example, children who grow up in environments with diverse microbial exposures such as traditional farms, are less prone to developing asthma and atopy (allergic sensitization) (Ege et al., 2011; Stein et al., 2016). Beneficial immunomodulatory environmental exposures are also suggested into adulthood (Douwes et al., 2007; Rottem et al., 2015; von Hertzen and Haahtela, 2006). Moreover, a lack of appropriate environmental exposures and deficient immune training and regulation may also impact other areas of immune-related human health including susceptibility to infectious disease (as below).

Respiratory health provides a conspicuous test case to explore environment-human health associations that have plausible microbiota-mediated linkages, because breathing offers a primary mode of exposure for people interacting passively with the environment. Throughout life, millions of litres of air move through the human respiratory tract, which provides one of the first points of contact with environmental contaminants and bioaerosols (airborne microbiota and bioactive agents). Airway epithelial tissues provide a protective arsenal of physical barriers, niche-occupying commensal microbiota, antimicrobial compounds, and receptors ready to orchestrate immune responses (Parker and Prince, 2011; Whitsett and Alenghat, 2014). Environmental microbiota can interact directly with respiratory mucosal immune receptors or via ecological interactions with host commensal microbiota. Similar interactions can also occur in the gut, influencing immune- and health-status, after environmental microbiota deposit in the airways and are transported by cilia to be swallowed (Rook, 2013). Importantly, dysregulation of the airway epithelial innate immune system can be associated with compromised immunity and chronic inflammation (Parker and Prince, 2011). Immune dysfunction may involve adverse feedbacks that reinforce imbalance (or dysbiosis) of host microbiota (Haahtela et al., 2013), which in turn may favour pathogenic microbes and increase susceptibility to infectious disease. As discussed later, there is often not a clear distinction between infectious and non-infectious respiratory disease outcomes, for example, where one type of disease (e.g. cold or influenza) can exacerbate symptoms of another (e.g. asthma). This means there is potential for environments and their microbiota to impact immune status, which in turn may have potential broad underlying (and population-level) influence on multiple infectious and non-infectious respiratory diseases.

If particular macro- and landscape-scale features of the environment (e.g. types of vegetation, soil, land use, and their diversity) can be associated with human health benefits (and ultimately supported by new knowledge of underlying causal mechanisms), it may be possible in the future to design new cost-effective, landscape and urban green space interventions with concurrent benefits for public health and biodiversity conservation. Informing such outcomes would require a large body of multidisciplinary research. The work presented here represents an early step.

Our motivation for this study is to build insight and provide context and priority for further research into these types of potential beneficial environmental exposures, through building on existing, inexpensive data. Given the possible abovementioned links between environments, immune development or dysfunction, and infectious and non-infectious respiratory disease, we examine available aggregated respiratory health outcome data in a cross-sectional ecological epidemiology study spanning the continent of Australia. Our aim is to test whether some types and qualities of environment may be more beneficial than others, and how such exposures may compare to known respiratory health influences. We appreciate that using aggregated health response data represents a limitation in terms of loss of specificity to link environmental influence with any particular disease. On the other hand, this approach may offer greater sensitivity to detect possible broad environmental influence on multiple respiratory disease outcomes.

Due to many unknowns (e.g. possible agents, behaviourally- and temporally-mediated exposures, requisite exposures, immunomodulatory and other possible physiological pathways), the use of environmental proxies is warranted as a pragmatic investigation tool (Liddicoat et al., 2016). Proxies allow us to consider a variety of (including possible beneficial immunomodulatory) environmental influences on human health. We might expect to see correlative signals between health outcomes and environmental exposures that are consistent with sources of microbial diversity, such as biodiverse environments, diversity in land use, and soils high in clay and/or organic matter content (Liddicoat et al., 2016; Rook, 2013; von Hertzen and Haahtela, 2006).

We use a data-intensive approach suited to this emerging area of scientific inquiry where it is important to gain an early understanding of key relationships among many variables, and note that models will improve iteratively over time (Elliott et al., 2016). Our modelling approach reflects the highly faceted nature of environments and is adept at handling large numbers of (included potentially correlated) candidate predictors. To guide future work, we provide clear interpretation and ranking of previously unaccounted environmental influences among important predictors of our respiratory health data.

2. Methods

We use an array of specially-prepared environmental covariates to estimate environmental exposures, each allowing for a potential surrounding zone of influence (section 2.2). To cater for geographically-variable environmental and multimodal social predictors, we stratify our analysis into three different socio-geographic groups spanning the Australian continent (section 2.3). We develop an automated screening algorithm to filter out extraneous variables and assess selected transformations of candidate predictors to help optimize linear relationships in subsequent modelling (section 2.4). We use least absolute shrinkage and selection operator (or Lasso) penalized regression (a contemporary machine-learning algorithm (Tibshirani, 1996)) to interpret key predictors from large numbers of candidate variables, and purpose-built 10-fold cross-validation (CV) modelling to indicate reproducibility and uncertainty in our results (section 2.5). This approach puts the onus on variables to compete and display strength and consistency of predictive value.

2.1. Public health and contextual data

We use merged 2011/12 and 2012/13 Social Health Atlas of Australia (PHIDU, 2015, 2016) data for respiratory disease public

hospital admissions (a breakdown of disease codes is provided in Appendix A, Table S1), together with accompanying contextual data (Appendix A, Table S2). For reliability and privacy purposes, these respiratory disease and associated contextual data are only available in aggregated form. In this study data are aggregated spatially, by Australian local government areas (LGAs), and thematically, grouped under principal diagnoses of diseases of the respiratory system. In each LGA, the 2011–13 normalized mean cumulative incidence of respiratory disease public hospital admissions was calculated using:

$$ASR_{2011-13} = \left(N_{11/12} + N_{12/13}\right) \Big/ \left(\frac{N_{11/12}}{ASR_{11/12}} + \frac{N_{12/13}}{ASR_{12/13}}\right) \quad (2.1)$$

where *ASR* is the respective age standardized rate per 100,000, and *N* is the respective raw number of annual admissions recorded. We used the LGA-based data because we considered they offered a balanced coverage of health reporting areas and a spatial framework suited to capturing environmental variability across both urban and regional Australia. Numbers of LGAs used in the modelling are shown in Table 1.

2.2. Environmental data

We prepared and collated an as large as practical number of environmental covariate layers to reflect possible direct and indirect influences on respiratory health. Many environmental layers were included to allow detection of as-yet-unexplained possible microbiota-mediated and other influences, as introduced earlier and discussed elsewhere (Liddicoat et al., 2016; Rook, 2013). Our data-intensive approach (i.e. including many variables and later use of Lasso machine-learning) also aimed to reduce modelling bias that might arise through pre-selection of only a small number of candidate environmental variables. We represent the diversity, and multi-faceted nature of environments using an array of climatic, soil, landscape and vegetation-based variables based on exhaustive Australia-wide gridded mapping datasets (Appendix A, Table S3). For consistent and pragmatic data-handling we adopt a common 250 m resolution grid system, as used elsewhere in continent-wide, landscape-scale mapping of land cover (Geoscience Australia, 2014). Where necessary to match the common grid system, resampling was performed bilinearly for numeric data layers, and using the nearest neighbour method for categorical layers.

Environmental map layers were then re-expressed using focal neighbourhood statistic calculations so that environmental data (at any cell location, or if averaged over an area) provide an estimate of exposure corresponding to a surrounding zone of influence. Such a zone reflects potential movement of populations within their surroundings and also the possibility of airborne dispersal of environmental microbiota and bioactive agents. We do not know how far populations or bioaerosols disperse, however we chose a nominal 3 km radius area as our representative environmental zone of influence. This area was consistent with previous studies examining possible links between land use and human immunomodulatory influence (Hanski et al., 2012) and green space and self-reported health (Maas et al., 2006). We were also guided by Ruokolainen et al. (2015), who found the spatial scale of land-use description affected the detection of statistically significant relationships between land-use and atopy (allergic sensitization), which they observed in the range 2–5 km.

This meant for all of the Australia-wide gridded environmental data layers, in every grid cell, we calculated a measure summarising a particular aspect of the surrounding environment over a 3 km-radius area. Numeric variables were averaged (including climatic, soil, landscape, and remotely-sensed vegetation parameters), while class proportions and Shannon diversity indices were calculated for categorical themes (i.e. land use, land cover, ecological land units and major vegetation groups). The conversion of categorical data to numeric data (using the focal neighbourhood calculations) also enabled a simpler linear modelling approach as all environmental data were ultimately expressed in numeric form. Additional information and formulae for the calculation of environmental layers are provided in Appendix A. A total of 176 gridded environmental-variable map layers were prepared.

To join with available health outcome data in LGAs, environmental data were averaged within each LGA boundary. (For future studies where finer resolution health data are available, we suggest these specially-prepared focal neighbourhood environmental layers could be sampled at point locations or averaged over smaller areas.)

Point-based air pollution data were also considered, however these were handled differently to the gridded data layers (due to reasons discussed below). Estimated total industrial emissions of inhalable particulate matter (of 10 μm or less in diameter, or PM10) for 2011/12 and 2012/13 were spatially intersected and summed in each LGA, expressed as area-based rates ($kg.km^{-2}.yr^{-1}$), then averaged to estimate the mean 2011–13 emissions.

2.3. Socio-geographic clustering

Environments vary greatly across the Australian continent (approx. area of 7.7 million km^2), with populations spanning urban, peri-urban, rural and remote locations. Therefore, we might expect any environmental influences on health (where present) to vary geographically. There are known health inequalities associated with socioeconomic status, remoteness (AIHW, 2007), and Aboriginal populations (Gubhaju et al., 2013); and these social

Table 1
Performance of 10-fold CV Lasso modelling. Mean performance statistics from explicit 10-fold CV Lasso modelling of respiratory disease public hospital admissions (see Appendix A for further details).

CV statistics (mean of 10-fold validation sets)	Local government area cluster		
	Moderate majority	Major cities	Remote disadvantaged
Response transformation	log10	–	log10
No. of LGAs used in modelling (n)	364	62	24
No. of predictors (p) chosen by the Lasso	20	11	10
SD (obs)	0.1479	242.2	0.1824
Root mean square error	0.0999	125.3	0.1463
Mean error (bias)	−0.0003	−2.833	0.0199
Skewness (residuals)	0.1363	−0.1932	−0.0341
Concordance correlation coefficient	0.6795	0.7768	0.4070
R^2	0.5557	0.7698	0.9244

factors also vary geographically. During preliminary data analysis we observed a negative skew in the Australia-wide distribution of Socioeconomic index, suggestive of a second minor mode of lower socioeconomic status LGAs (Appendix A, Fig. S1). To allow for likely varying environmental and multimodal socio-geographic influences across an expansive dataset, and to improve linear modelling of the respiratory health outcome, we stratify the data into three socio-geographic groups using k-means clustering based on all available data for Socioeconomic index, Population density and Percent Aboriginal persons (Fig. 1; Appendix A, Table S4). We broadly interpret the resulting groups (LGA clusters) as the 'Moderate majority', 'Major cities', and 'Remote disadvantaged'. Separate analyses were performed using these socio-geographic clusters.

2.4. Additional data preparation for modelling

Due to missing data, we excluded some LGAs from the health response modelling. We also developed an automated screening algorithm to objectively exclude extraneous variables (e.g. irrelevant environment class proportions in particular LGA clusters), and to consider a limited set of candidate predictor transformations (designed to improve linear relationships in subsequent modelling). The screening algorithm was fully coded in R script as a means to transparently and consistently inspect variables and consider preparatory steps that might otherwise be done manually one variable at a time, prior to multiple linear regression modelling. In each LGA cluster, pragmatic threshold criteria were used to exclude (original or transformed) variables if less than 50% of LGAs contained non-zero values, if variable skewness exceeded 1.5, kurtosis exceeded 4, or if stand-alone explanatory value (R^2) was less than 2.5%. Only logit transformations were considered for proportion or percentage data, otherwise square root, log10, square and cube root were considered, subject to certain disqualifications (e.g. log10 cannot be used on zero or negative values). Selection of a final representative candidate (whether original or transformed) was made according to pre-determined rules that attempted to balance the trade off between interpretability and optimising normality of predictor distributions. Further description of the data preparation steps are provided in Appendix A (including R code).

Within LGA clusters, all screening and subsequent analysis of candidate predictors was performed on 95% Winsorized data, designed to objectively eliminate the influence of extreme and potentially outlying values (Friedman and Popescu, 2008). The distributions of respective health response variables were inspected, and for the 'Moderate majority' and 'Remote disadvantaged' clusters, log10 variance-stabilising transformations were applied (Appendix A, Fig. S2). All predictor data were centred and scaled before input for Lasso modelling.

2.5. Lasso modelling

We use Lasso penalized regression as our primary tool for correlation analysis and health response modelling, as implemented in the R *glmnet* package (Friedman et al., 2010). Use of the Lasso is relatively new in epidemiology, but has been suggested as a credible alternative to more conventional stepwise multiple regression approaches when identifying key predictors from large datasets (Mansiaux and Carrat, 2014). As the Lasso can be prone to inconsistency in variable selection (Leng et al., 2006), within each LGA cluster, we apply explicit randomized 10-fold resampling to generate variation in input data for modelling. This was designed to help interpret the levels of variation, reproducibility and uncertainty in our modelling results. In each fold, internally within the Lasso software we ran leave-one-out CV to identify the optimal penalisation parameter (we used the parsimonious option corresponding to s = 'lambda.1se' when calling the *cv.glmnet* function, with the Lasso model corresponding to the default setting a = 1). Details of the statistical analyses and R scripts used for the modelling are provided in Appendix A.

We interpret the relative importance and direction of predictors identified by the Lasso from the size and sign of standardized regression coefficients (this is facilitated by centring and scaling of data as discussed). Identified predictors and their standardized coefficients were harvested from the respective 10-fold Lasso modelling outputs and plotted, as below.

3. Results and discussion

3.1. Key predictors of respiratory health

We show results for the 'Moderate majority' in Fig. 2, ordered by the mean absolute size of standardized regression coefficients. Important predictors (toward the top of Fig. 2) are consistently identified. Beneficial respiratory health outcomes were associated

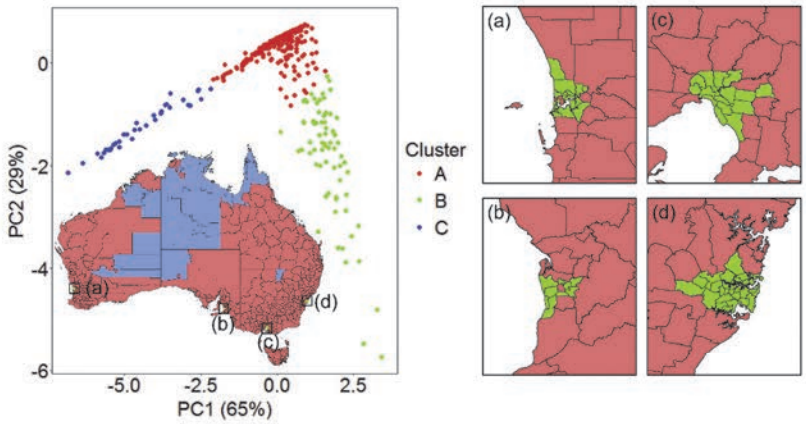


Fig. 1. Socio-geographic clusters of Australian LGAs. Three clusters were identified: the 'Moderate majority' (A, n = 451), 'Major cities' (B, n = 72), and 'Remote disadvantaged' (C, n = 35). Numbers indicate all available LGAs used to determine clusters, whereas a reduced set were available for health response modelling (see Table 1). Inset maps show LGAs associated with (a) Perth, (b) Adelaide, (c) Melbourne, and (d) Sydney.

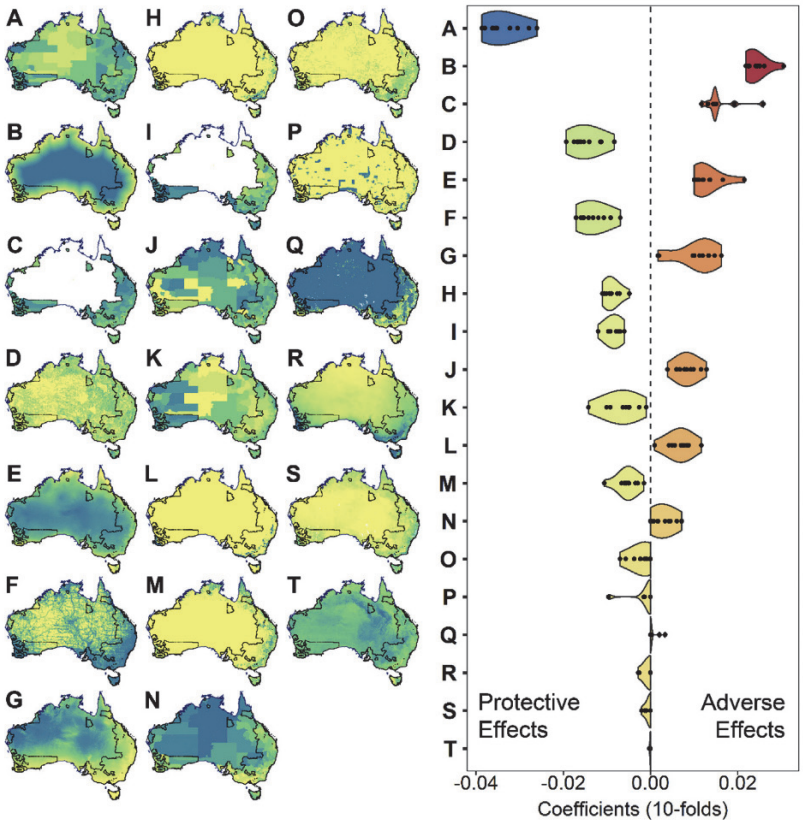


Fig. 2. Important predictors for the 'Moderate majority' socio-geographic cluster. The right panel shows density and point plots of standardized regression coefficients from 10-fold CV Lasso modelling of log10 (respiratory disease public hospital admissions). Shading of density plots indicates negative (blue) to positive (red) coefficients. Variables with negative coefficients associate with decreased hospital admissions. Twenty predictors were identified across the 10-fold Lasso models: (A) Socioeconomic index, (B) Distance to coast, (C) Percent obese persons, (D) Diversity of major vegetation groups, (E) Mean temperature annual range, (F) Species richness (log10), (G) Maximum temperature of warmest month, (H) Proportion of eucalypt forests 10–30 m (logit), (I) Percent overweight persons (logit), (J) Percent smoking during pregnancy, (K) Percent English-speaking immigrants (logit), (L) Proportion of warm wet plains (logit), (M) Proportion of open trees (logit), (N) Percent Aboriginal persons (logit), (O) Diversity of land use, (P) Proportion of nature conservation (logit), (Q) Vegetation fractional cover minimum nonphotosynthetic (logit), (R) Mean precipitation of the coldest quarter, (S) Vegetation fractional cover minimum photosynthetic (logit), (T) Soil cation exchange capacity * erodible fraction (geometric mean). Maps display all available predictor data for Australia (low values are shown in yellow, high values in blue), however only a portion of local government areas (LGAs) occur in the 'Moderate majority' socio-geographic cluster (Fig. 1). Due to reduced data availability among candidate predictors, a smaller subset of LGAs (n = 364, see maps C and I) were used in the modelling. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

with (in order of decreasing importance) Socioeconomic index, Diversity of major vegetation groups, Species richness (log10), Proportion of eucalypt forests 10–30 m (logit), Percent overweight persons (logit), Percent English-speaking immigrants (logit), Proportion of open [i.e. 30–70% canopy cover] trees (logit), Diversity of land use, and Proportion of nature conservation (logit). Adverse respiratory health associations were identified with Distance to coast, Percent obese persons, Mean temperature annual range, Maximum temperature of warmest month, Percent smoking during pregnancy, Proportion of warm wet plains (logit), Percent Aboriginal persons (logit), and Vegetation fractional cover minimum nonphotosynthetic (logit). Predictors decrease in importance down the y-axis and we suggest variables towards the bottom should be viewed with caution (e.g. Mean precipitation of the coldest quarter, Vegetation fractional cover minimum photosynthetic (logit), and Soil cation exchange capacity * erodible fraction (geometric mean)). Predictors that are rarely or not consistently selected may not be generally applicable (e.g. localized influence) and, possibly, the lowest ranked variables may be spurious (e.g. analogous to near-zero coefficient variables in a multiple linear regression). Many variables were not selected at all by the Lasso in any of the 10 folds (see Appendix A, Table S5).

The nature of correlations between predictors identified across the 10-fold Lasso modelling is shown, by way of example, for the 'Moderate majority' cluster (Appendix A, Fig. S3). Important predictors for the remaining clusters are shown in Appendix A, Figs. S4–S5. Summary performance statistics (Table 1) and CV plots (Appendix A, Fig. S6) for the 10-fold Lasso modelling, indicate that moderate levels of prediction success were achieved. Based on concordance correlation coefficients (which indicate how closely observed and predicted values adhere to a 1:1 relationship), moderate agreement was found between predicted and observed health responses in the 'Moderate majority' and 'Major cities'. However, poorer agreement was found for the 'Remote disadvantaged' LGAs, possibly due to their sparsely settled nature, coarse scale of data and small sample size. The high mean R^2 value for the 'Remote disadvantaged' cluster (Table 1) suggests good linear agreement within each of the CV folds. However, we view this with

caution due to small sample numbers and poor 1:1 alignment of observed and predicted values indicated by the low concordance correlation coefficient and Fig. S6 (c) (Appendix A).

Our approach provides a side-by-side comparison of the potential influence of (including previously unaccounted) environmental variables against recognized population health predictors of respiratory disease (Fig. 2). Reassuringly, a number of key predictors in our results match expectation, or align with findings elsewhere for lifestyle and environmental influences. For example, higher socioeconomic status (Socioeconomic index) associates with reduced hospital admissions and smoking associates with increased hospital admissions. Noting that our dependent variable is aggregated, we draw tentative support from literature relating to both general and disease-specific influences linked to respiratory health outcomes. For example, obesity has been negatively associated with respiratory health (Zammit et al., 2010), and lesser indications of a counter-intuitive health advantage attributed to being overweight are not without precedent (Flegal and Kalantar-Zadeh, 2013). Benefits from closeness to sea air (or disadvantage with distance from coast) might be expected with a number of respiratory conditions, for example in non-cystic fibrosis bronchiectasis (Kellett and Robert, 2011). Although, the variable of Distance to coast provides an example for possible parallel interpretations, such as respiratory conditions that could be linked to increasing dust in drier inland areas. High temperatures are reported to have an impact on respiratory admissions, particularly in the elderly (Michelozzi et al., 2009). Warm and wet (humid) environments also contribute to population risk of non-tuberculous mycobacterial pulmonary infection (Prevots and Marras, 2015).

Our results highlight a number of health-correlated environmental variables that are worthy of further investigation (refer to beneficial respiratory health associations listed earlier). In particular, Diversity of major vegetation groups featured in our results—and provides a measure of differentiation among habitats (analogous to beta diversity). The higher ranking of biodiversity surrogates (e.g. Diversity of major vegetation groups, Species richness (log10), Proportion of eucalypt forests 10–30 m (logit), Proportion of nature conservation), compared to remote sensing of vegetation greenness (e.g. fraction of photosynthetically active radiation and fractional cover layers) and land use class proportions for broad-acre agricultural uses (which were considered but not recognized in the modelling), is consistent with the notion that the quality of environments is important to respiratory health. From this, we speculate that optimal health benefit may not be merely associated with any type of green space and 'clean country air', but may be promoted by as-yet-unknown attributes of the green space itself. However, air pollution has not been fully accounted for, as discussed below. We also undertook extended analyses for the 'Moderate majority' LGA cluster to better understand the significance of variables selected by the Lasso (noting this represents an evolving area of statistical science), and this is described in Appendix A.

We saw a notably different pattern for 'Major cities' (Appendix A, Fig. S4). Health benefits were primarily associated with socioeconomic status (consistent with the notion of social 'insulating layers' (Myers et al., 2013)), while other social, lifestyle, and ambient environmental influences appeared to be generally detractive. For example, possibly, rainfed pasture and peak levels of remotely-sensed living vegetation could in this case be related to excessive levels of airway allergens from productive but low-biodiversity neighbouring farmland. It is likely the scale of our data has limited the prospects of detecting positive environment-health associations in this cluster.

The 'Remote disadvantaged' cluster was only represented by a small and coarse-scale dataset, so results may be misleading, however it was interesting to see wetlands, *Acacia* forests and woodlands, and swampy vegetation feature positively—also consistent with the notion that biodiversity may provide an as-yet-undetermined beneficial influence on respiratory health.

Previous modelling of Australia-wide asthma and chronic obstructive pulmonary disease (COPD) hospitalisation rates (AIHW et al., 2014) found significant associations with socioeconomic status, remoteness and the Indigenous proportion of the population; with generally higher hospitalisation rates for both asthma and COPD in inland and rural Australia. In general terms, our findings are consistent with previous studies examining environmental influences on respiratory health. However our approach in considering an array of environmental attributes (including landscape-scale biodiversity surrogates), provides potential new insight and priority research targets to further investigate causal mechanisms.

3.2. Air pollution

We lacked appropriate exhaustive, continent-wide mapping data to fully represent the potential influence of air pollution in this study. We did test the influence of point-based mean 2011–13 estimated total industrial PM10 emissions in each LGA, however these data did not register as a key predictor in our analyses. Closer inspection of the PM10 data showed many LGAs contained zero values which made the data poorly distributed in the context of our modelling approach (due to inherent skewness and also disqualification of log10 transformation due to zeros). We note these PM10 data represent modelled estimates (not actual measured data) and don't account for natural emissions (e.g. from windblown dust, sea salt aerosols and biological aerosol particles), which elsewhere (Liora et al., 2015) are estimated to be of a similar order to the lower range of estimated industrial PM10 emissions analysed here. Without wider data or knowledge of natural background PM10 levels we did not consider it valid for this study to interpolate a continent-wide map layer based on incomplete and spatially-limited data.

We suggest the lack of appropriate air pollution data doesn't invalidate our broader results; for example, we are encouraged by the recognition of known respiratory health predictors. Also, Australia's ambient air quality is generally good, typically meeting national health-based standards, with many key air pollutants having declined or remaining stable over the assessment period 1999–2008 (State of the Environment 2011 Committee, 2011). Particulate pollution from localized extreme events such as bushfires has been linked to increased respiratory hospital admissions (Chen et al., 2006), although bushfire frequency mapping was included but not identified as a useful predictor in our study. Urban air pollution is another localized concern for respiratory health (Simpson et al., 2005) that is not accounted for due to a lack of exhaustive data. To some extent, the separation of 'Major cities' may largely quarantine this factor in our analyses. We expect air pollution may contribute to respiratory health outcomes, however our data did not support findings of an association in this study.

3.3. Possible links between biodiverse environments and respiratory health

Our finding of an association between beneficial respiratory health outcomes and surrogate measures of landscape biodiversity (in particular, Diversity of major vegetation groups, and other measures listed earlier) warrants consideration of possible underlying ecological linkage mechanisms. As a megadiverse continent (Groombridge, 1992), Australia provides an interesting study area for exploring such links.

Simple indirect explanations may exist, for example, due to absence of air pollution (or 'clean country air'). Seasonal exposures to airborne airway allergens such as grass pollen, a major trigger for allergic rhinitis and asthma (Davies et al., 2015), may be indirectly related to landscape biodiversity. For example, where landscapes and land uses elsewhere may be characterized by a low biodiversity of plant species and a corresponding concentration of (often introduced) grass and tree species producing large amounts of allergenic pollen (Cariñanos and Casares-Porcel, 2011). In Australia, knowledge of seasonal and biogeographical patterns of pollen exposures and relationships with allergic respiratory disease is growing (Davies et al., 2015). Human health results from a complex interplay of factors, so a number of (including environmental) influences may possibly contribute. Viral infections are known to exacerbate other respiratory conditions, and are implicated in 50–80% of all hospitalisations for asthma (Bardin, 2004). This highlights the importance of underlying immune status and possible connections between common communicable and non-communicable respiratory disease.

In terms of general health outcomes, cross-sectional spatial epidemiology studies have previously found positive associations between neighbourhood green space and self-reported health status (e.g. de Vries et al., 2003; Maas et al., 2006). Exposure to green space and natural environments has been associated with reduced all-cause mortality (Mitchell and Popham, 2008). Mitchell and Popham also found that increased exposure to green space reduced health inequalities related to socioeconomic status. Common beneficial influences suggested from enhanced visual and/or physical access to green space include greater physical activity, reduced stress, enhanced restoration from mental fatigue, improved mood and self-esteem, and provision of psychological and mental health benefits (reviewed elsewhere e.g. (Keniger et al., 2013; WHO and SCBD, 2015)). These influences may act synergistically to further enhance benefits from 'green exercise' (exercise in natural surroundings) (Gladwell et al., 2013).

A number of reviews also suggest the possibility of beneficial environmental microbiota- and bioactive-mediated immunomodulatory influences on human health (Craig et al., 2016; Haahtela et al., 2013; Rook, 2013; Sandifer et al., 2015; WHO and SCBD, 2015). Intuitively, there may be plausible mechanisms for microbiome-respiratory system health influence including competitive exclusion and/or regulation of airway pathogens and immune priming. There is a known role for host commensal microbiota in normal (healthy) immune functioning (e.g. in the gut (Artis, 2008; Molloy et al., 2012)) and increasing awareness of associations between dysbiosis (i.e. altered composition of the human microbiota, often with an imbalance between pathogenic and commensal organisms) and a range of diseases (Belizario and Napolitano, 2015; Clemente et al., 2012), including respiratory diseases (Dickson et al., 2013; Noval Rivas et al., 2016). Certain commensals, particularly in the gut, may play important roles in the signalling and control of inflammatory and regulatory immune processes (Artis, 2008; Molloy et al., 2012). Emerging knowledge of a 'gut-lung' microbial axis (Fujimura and Lynch, 2015; Noval Rivas et al., 2016), or connection between host gut and airway microbiota, has been demonstrated in mouse models where changes in gut microbiota composition can influence airway immune responses (Fujimura et al., 2014; Ichinohe et al., 2011; Kim et al., 2014).

Importantly, at least a portion of the human microbiota is in dynamic exchange with environmental microbiota and, therefore, natural microbial diversity is recognized as an important potential contributor to healthy immune functioning (Haahtela et al., 2013; WHO and SCBD, 2015). Aerobiology studies tell us that the air is alive with all manner of biogenic and bioactive particles (Després et al., 2012; Polymenakou, 2012). Bioaerosols often comprise microorganisms from soils, plant surfaces, water bodies, rocks and built structures that are mobilized by wind and splashing water and can be transported considerable distances (Polymenakou, 2012). Bioaerosols can be deposited in the upper airways, carried up the trachea by the action of cilia and then swallowed—therefore, airborne microorganisms can end up on the skin, in the airways, and in the gut where they can perform immunomodulatory roles (Rook, 2013). Pulmonary neuroendocrine cells in the airway can also directly translate environmental cues into physiological and immune responses (Branchfield et al., 2016). Possible connections between environmental features (e.g. vegetation, soils, water bodies, land use, etc.) and environmental microbiota are discussed elsewhere (Liddicoat et al., 2016; Rook, 2013), and such knowledge is steadily building (e.g. via Bissett et al., 2016; Gilbert et al., 2014). Neighbouring land use and landscape composition have been shown to influence bioaerosol composition (Bowers et al., 2011; Després et al., 2012; Mhuireach et al., 2016). However, more work is needed to investigate possible links between aerobiological exposures and surrounding environments.

Although not explored here, temporal aspects of environmental microbiota-human contact are likely to be important, in relation to commensal microbiota and immune system development (Wopereis et al., 2014). In immature hosts, commensals appear to be acquired through random environmental exposures (Artis, 2008). Once established, the mature host commensal microbiota is more resilient to colonisation (Seedorf et al., 2014). Although, if we extrapolate from studies of short-versus long-term dietary change (Voreades et al., 2014), possibly, long-lasting environmental change may influence a shift in composition even for mature commensal microbiota. Poor western diets may cause irreversible loss of potential key beneficial commensal taxa, with suggestions that a rewilding of the human microbiota may be needed (Sonnenburg et al., 2016). Recent findings that the human intestinal microbiota may be dominated by spore-forming bacteria (Browne et al., 2016) which are designed to persist in the environment, aligns with the idea that critical early and perhaps 'ongoing maintenance' environmental exposures may help replenish key spore-forming bacteria to the human commensal microbiota. In turn, as discussed, such interactions may influence airway immune status and respiratory health.

Health benefits may also be linked to abiotic bioactive agents from the environment such as phytoncides (wood essential oils or VOCs) and negative air ions (Craig et al., 2016). Negative air ions from waterfall aerosol have been shown to produce lasting beneficial effects on asthma symptoms, lung function and airway inflammation (Gaisberger et al., 2012). Plants and phyllosphere microbiota are known to emit and influence a wide variety of VOCs (Bulgarelli et al., 2013). Such plant-based VOCs can promote or inhibit (and thus shape) adjacent microbial communities (Bringel and Couée, 2015), which may in turn have varying human health influences.

Australia has among the highest rates of allergies among developed countries, affecting almost 20% of the population, and this prevalence is increasing (Access Economics, 2007). According to the 2014-15 National Health Survey (ABS, 2015), long-term respiratory disease is estimated to affect nearly 31% of the population; with the prevalence of hayfever and allergic rhinitis, asthma, and chronic sinusitis estimated at 19.4%, 10.8%, and 8.4% respectively. If links between macro- and landscape-scale environmental biodiversity and beneficial immunomodulatory influences can be substantiated and quantified, significant reductions in disease burden and public health expenditure might be possible through greater consideration of biodiversity in public health programs and new landscape and urban green space design.

3.4. Limitations

The environmental mapping products used here have been produced from a variety of projects and methods, independently from routine Australia-wide public health reporting. Consequently, the trade-off for representing a wide diversity of potential environmental influences, is a loss of exact temporal coincidence of surrogate environmental exposures (derived from environmental mapping) with the health response data. We continued on the basis that the rate of environmental change across the continent should be sufficiently slow, or health impacts possibly integrated over time, that available environmental mapping datasets from recent decades should sufficiently characterize the type of environments and potential exposures relevant to our recent (2011–2013) health response data. We note this approach allows us to consider many expert-interpreted mapping layers for different classifications of the environment. Our approach is vindicated where a number of these expert-interpreted layers (and derived biodiversity surrogates) offer greater explanatory value than less-interpreted remote sensing-derived layers (e.g. based on greenness indices) also included in the modelling.

We recognize there will be environmental and social/lifestyle influences, and potential lurking variables and interactions, that we have not specifically accounted for (e.g. inadequate air pollution data, as discussed). Ecological epidemiology studies will always contain confounders, and cross-sectional studies are generally limited as they cannot infer temporal directionality or causality. However this type of study is suited to inexpensive examination of population-wide disease associations with multiple potential influences, as well as further hypothesis generation and informing more detailed studies. As our public health data in this study were aggregated into LGAs we note the potential for erroneous findings due to ecological bias and fallacy (Elliot et al., 2000), and the modifiable areal unit problem (Parenteau and Sawada, 2011). The scale of data is generally finest to coarsest in order from 'Major cities', 'Moderate majority', then 'Remote disadvantaged'—where coarser datasets are more prone to errors of the type suggested. To some extent, the use of 10-fold resampling and CV modelling, moderates concern due to the modifiable areal unit problem, particularly in 'Moderate majority' cluster, where top-ranking predictors are consistently identified from different (10-fold) subselections of LGAs.

The use of aggregated respiratory disease outcome data poses a limitation for interpretation of possible specific environment-respiratory disease relationships. On the other hand, our approach may offer insights to possible broad underlying influences (through as-yet-unknown mechanisms) that may impact a range of associated diseases. Also we note that the use of aggregated health outcome data is common in population health studies (e.g. (de Vries et al., 2003; Maas et al., 2006; Mitchell and Popham, 2008)). On the positive side, with this inexpensive but expansive continent-wide dataset, we have provided a side-by-side comparison of known population-level respiratory health predictors and previously untested potential environmental influences.

4. Conclusions

Using continent-wide datasets, we have explored previously unaccounted potential environmental influences on respiratory health. Our results are generally suggestive of a potential beneficial or protective health influence associated with natural and biodiverse environments. Across different socio-geographic settings, we highlight environmental features and attributes worthy of more targeted investigation—in particular, high biodiversity areas, natural forests through to open woodlands, wetlands, high diversity in land use, and proximity to coastal areas. The validity of these findings is supported by the parallel associations we found with well-recognized social predictors of health included in the modelling. These findings provide additional motivation to investigate the various potential connections between biodiverse environments and human health. Among these, we suggest possible beneficial immunomodulatory influences from environmental microbiota and bioactive agents (perhaps associated with various environmental components such as types of vegetation, soils, land use, and their diversity) represent a worthy research focus (e.g. via study of environmental and human microbiomes and human health biomarkers). Our results provide additional support and context to emerging research into the potential role of natural green space exposures in reducing the risk of immune-related disease at a population level. Our approach can be readily adapted to explore potential health and environmental associations elsewhere.

Acknowledgements

We thank Graham Rook and Geraint Rogers for helpful discussions and suggestions. This research used high-performance computing services provided by eRSA (www.ersa.edu.au), and mapping datasets produced through Australia's Terrestrial Ecosystem Research Network (TERN, www.tern.org.au).

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.jenvman.2017.10.007.

References

ABS, 2015. 2015 National Health Survey: First Results, 2014–15 (Catalogue No. 4364.0.55.001). Latest ISSUE Released at 11:30 AM (CANBERRA TIME) 08/12/2015. Australian Bureau of Statistics, Canberra, Australia.

Access Economics, 2007. The economic impact of allergic disease in Australia: not to be sneezed at. Report by Access Economics Pty Limited for the Australasian Society of Clinical Immunology and Allergy (ASCIA).

AIHW, 2007. Rural, Regional and Remote Health: a Study on Mortality. Rural health series, second ed. Australian Institute of Health and Welfare, Canberra. p. x + 351.

AIHW, Poulos, L.M., Cooper, S.J., Ampon, R., Reddel, H.K., Marks, G.B., 2014. Mortality from Asthma and COPD in Australia. Australian Institute of Health and Welfare, Canberra, Australia.

Artis, D., 2008. Epithelial-cell recognition of commensal bacteria and maintenance of immune homeostasis in the gut. Nat. Rev. Immunol. 8, 411–420.

Bardin, P.G., 2004. Vaccination for asthma exacerbations. Intern. Med. J. 34, 358–360.

Belizario, J.E., Napolitano, M., 2015. Human microbiomes and their roles in dysbiosis, common diseases, and novel therapeutic approaches. Front. Microbiol. 6.

Bissett, A., Fitzgerald, A., Meintjes, T., Mele, P.M., Reith, F., Dennis, P.G., Breed, M.F., Brown, B., Brown, M.V., Brugger, J., Byrne, M., Caddy-Retalic, S., Carmody, B., Coates, D.J., Correa, C., Ferrari, B.C., Gupta, V.V.S.R., Hamonts, K., Haslem, A., Hugenholtz, P., Karan, M., Koval, J., Lowe, A.J., Macdonald, S., McGrath, L., Martin, D., Morgan, M., North, K.I., Paungfoo-Lonhienne, C., Pendall, E., Phillips, L., Pirzl, R., Powell, J.R., Ragan, M.A., Schmidt, S., Seymour, N., Snape, I., Stephen, J.R., Stevens, M., Tinning, M., Williams, K., Yeoh, Y.K., Zammit, C.M., Young, A., 2016. Introducing BASE: the biomes of Australian soil environments soil microbial diversity database. GigaScience 5, 21.

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. ISME J. 5, 601–612.

Branchfield, K., Nantie, L., Verheyden, J.M., Sui, P., Wienhold, M.D., Sun, X., 2016. Pulmonary neuroendocrine cells function as airway sensors to control lung immune response. Science 351, 707.

Bringel, F., Couée, I., 2015. Pivotal roles of phyllosphere microorganisms at the interface between plant functioning and atmospheric trace gas dynamics. Front. Microbiol. 6.

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, 543–546.

Bulgarelli, D., Schlaeppi, K., Spaepen, S., Themaat, E.V.L.v., Schulze-Lefert, P., 2013. Structure and functions of the bacterial microbiota of plants. Annu. Rev. Plant Biol. 64, 807–838.

Cariñanos, P., Casares-Porcel, M., 2011. Urban green zones and related pollen allergy: a review. Some guidelines for designing spaces with low allergy impact. Landsc. Urban Plan. 101, 205–214.
Chen, L., Verrall, K., Tong, S., 2006. Air particulate pollution due to bushfires and respiratory hospital admissions in Brisbane, Australia. Int. J. Environ. Health Res. 16, 181–191.
Clemente, Jose C., Ursell, Luke K., Parfrey, Laura W., Knight, R., 2012. The impact of the gut microbiota on human health: an integrative view. Cell 148, 1258–1270.
Craig, J.M., Logan, A.C., Prescott, S.L., 2016. Natural environments, nature relatedness and the ecological theater: connecting satellites and sequencing to shinrin-yoku. J. Physiol. Anthropol. 35, 1–10.
Davies, J.M., Beggs, P.J., Medek, D.E., Newnham, R.M., Erbas, B., Thibaudon, M., Katelaris, C.H., Haberle, S.G., Newbigin, E.J., Huete, A.R., 2015. Trans-disciplinary research in synthesis of grass pollen aerobiology and its importance for respiratory health in Australasia. Sci. Total Environ. 534, 85–96.
de Vries, S., Verheij, R.A., Groenewegen, P.P., Spreeuwenberg, P., 2003. Natural environments—healthy environments? An exploratory analysis of the relationship between greenspace and health. Environ. Plan. A 35, 1717–1731.
Després, V., Huffman, J.A., Burrows, S., Hoose, C., Safatov, A., Buryak, G., Fröhlich-Nowoisky, J., Elbert, W., Andreae, M., Pöschl, U., Jaenicke, R., 2012. Primary biological aerosol particles in the atmosphere: a review. Tellus B Chem. Phys. Meteorol. 64, 15598.
Dickson, R.P., Erb-Downward, J.R., Huffnagle, G.B., 2013. The role of the bacterial microbiome in lung disease. Expert Rev. Respir. Med. 7, 245–257.
Douwes, J., Travier, N., Huang, K., Cheng, S., McKenzie, J., Le Gros, G., Von Mutius, E., Pearce, N., 2007. Lifelong farm exposure may strongly reduce the risk of asthma in adults. Allergy 62, 1158–1165.
Ege, M.J., Mayer, M., Normand, A.-C., Genuneit, J., Cookson, W.O.C.M., Braun-Fahrländer, C., Heederik, D., Piarroux, R., von Mutius, E., 2011. Exposure to environmental microorganisms and childhood asthma. N. Engl. J. Med. 364, 701–709.
Elliot, P., Wakefield, J., Best, N., Briggs, D., 2000. Spatial Epidemiology: Methods and Applications. Oxford University Press, New York, p. 475.
Elliott, K.C., Cheruvelil, K.S., Montgomery, G.M., Soranno, P.A., 2016. Conceptions of good science in our data-rich world. BioScience 66, 880–889.
Flegal, K.M., Kalantar-Zadeh, K., 2013. Overweight, mortality and survival. Obesity 21, 1744–1745.
Friedman, J., Hastie, T., Tibshirani, R., 2010. Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. 33, 1–22.
Friedman, J.H., Popescu, B.E., 2008. Predictive learning via rule ensembles. Ann. Appl. Stat. 2, 916–954.
Fujimura, K.E., Demoor, T., Rauch, M., Faruqi, A.A., Jang, S., Johnson, C.C., Boushey, H.A., Zoratti, E., Ownby, D., Lukacs, N.W., Lynch, S.V., 2014. House dust exposure mediates gut microbiome Lactobacillus enrichment and airway immune defense against allergens and virus infection. PNAS 111, 805–810.
Fujimura, K.E., Lynch, S.V., 2015. Microbiota in allergy and asthma and the emerging relationship with the gut microbiome. Cell Host Microbe 17, 592–602.
Gaisberger, M., Šanović, R., Dobias, H., Kolarž, P., Moder, A., Thalhamer, J., Selimović, A., Huttegger, I., Ritter, M., Hartl, A., 2012. Effects of ionized waterfall aerosol on pediatric allergic asthma. J. Asthma 49, 830–838.
Geoscience Australia, 2014. Dynamic Land Cover Dataset V1.0, 27 May 2014.
Gilbert, J.A., Jansson, J.K., Knight, R., 2014. The Earth Microbiome project: successes and aspirations. BMC Biol. 12, 69.
Gladwell, V.F., Brown, D.K., Wood, C., Sandercock, G.R., Barton, J.L., 2013. The great outdoors: how a green exercise environment can benefit all. Extreme Physiol. Med. 2, 3.
Groombridge, B. (Ed.), 1992. Global Biodiversity: Status of the Earth's Living Resources. A Report Compiled by the World Conservation Monitoring Centre. Chapman and Hall, London.
Gubhaju, L., McNamara, B.J., Banks, E., Joshy, G., Raphael, B., Williamson, A., Eades, S.J., 2013. The overall health and risk factor profile of Australian Aboriginal and Torres Strait Islander participants from the 45 and up study. BMC Public Health 13, 1–14.
Haahtela, T., Holgate, S., Pawankar, R., Akdis, C.A., Benjaponpitak, S., Caraballo, L., 2013. The biodiversity hypothesis and allergic disease: World Allergy Organization position statement. World Allergy Org. J. 6.
Hanski, I., von Hertzen, L., Fyhrquist, N., Koskinen, K., Torppa, K., Laatikainen, T., Karisola, P., Auvinen, P., Paulin, L., Mäkelä, M.J., Vartiainen, E., Kosunen, T.U., Alenius, H., Haahtela, T., 2012. Environmental biodiversity, human microbiota, and allergy are interrelated. PNAS 109, 8334–8339.
Ichinohe, T., Pang, I.K., Kumamoto, Y., Peaper, D.R., Ho, J.H., Murray, T.S., Iwasaki, A., 2011. Microbiota regulates immune defense against respiratory tract influenza A virus infection. PNAS 108, 5354–5359.
Kellett, F., Robert, N.M., 2011. Nebulised 7% hypertonic saline improves lung function and quality of life in bronchiectasis. Respir. Med. 105, 1831–1835.
Keniger, L., Gaston, K., Irvine, K., Fuller, R., 2013. What are the benefits of interacting with nature? Int. J. Environ. Res. Public Health 10, 913.
Kim, Y.-G., Udayanga, Kankanam Gamage S., Totsuka, N., Weinberg, Jason B., Núñez, G., Shibuya, A., 2014. Gut dysbiosis promotes M2 macrophage polarization and allergic airway inflammation via fungi-induced PGE2. Cell Host Microbe 15, 95–102.
Leng, C., Lin, Y., Wahba, G., 2006. A note on the LASSO and related procedures in model selection. Stat. Sin. 16, 1273–1284.
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, 1023–1034.
Liora, N., Markakis, K., Poupkou, A., Giannaros, T.M., Melas, D., 2015. The natural emissions model (NEMO): description, application and model evaluation. Atmos. Environ. 122, 493–504.
Maas, J., Verheij, R.A., Groenewegen, P.P., de Vries, S., Spreeuwenberg, P., 2006. Green space, urbanity, and health: how strong is the relation? J. Epidemiol. Community Health 60, 587–592.
Mansiaux, Y., Carrat, F., 2014. Detection of independent associations in a large epidemiologic dataset: a comparison of random forests, boosted regression trees, conventional and penalized logistic regression for identifying independent factors associated with H1N1pdm influenza infections. BMC Med. Res. Methodol. 14, 99.
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.
Michelozzi, P., Accetta, G., De Sario, M., D'Ippoliti, D., Marino, C., Baccini, M., Biggeri, A., Anderson, H.R., Katsouyanni, K., Ballester, F., Bisanti, L., Cadum, E., Forsberg, B., Forastiere, F., Goodman, P.G., Hojs, A., Kirchmayer, U., Medina, S., Paldy, A., Schindler, C., Sunyer, J., Perucci, C.A., 2009. High temperature and hospitalizations for cardiovascular and respiratory causes in 12 european cities. Am. J. Respir. Crit. Care Med. 179, 383–389.
Mitchell, R., Popham, F., 2008. Effect of exposure to natural environment on health inequalities: an observational population study. The Lancet 372, 1655–1660.
Molloy, M.J., Bouladoux, N., Belkaid, Y., 2012. Intestinal microbiota: shaping local and systemic immune responses. Semin. Immunol. 24, 58–66.
Myers, S.S., Gaffikin, L., Golden, C.D., Ostfeld, R.S., Redford, K.H., Ricketts, T.H., Turner, W.R., Osofsky, S.A., 2013. Human health impacts of ecosystem alteration. PNAS 110, 18753–18760.
Noval Rivas, M., Crother, T.R., Arditi, M., 2016. The microbiome in asthma. Curr. Opin. Pediatr. 28, 764–771.
Parenteau, M.-P., Sawada, M.C., 2011. The modifiable areal unit problem (MAUP) in the relationship between exposure to NO2 and respiratory health. Int. J. Health Geogr. 10, 58.
Parker, D., Prince, A., 2011. Innate immunity in the respiratory epithelium. Am. J. Respir. Cell Mol. Biol. 45, 189–201.
PHIDU, 2015. Social Health Atlas of Australia: Data by Local Government Area. March 2015 release. Public Health Information Development Unit. Torrens University Australia, Adelaide.
PHIDU, 2016. Social Health Atlas of Australia: Data by Local Government Area. May 2016 release. Public Health Information Development Unit. Torrens University Australia, Adelaide.
Polymenakou, P.N., 2012. Atmosphere: a source of pathogenic or beneficial microbes? Atmosphere 3, 87–102.
Prevots, D.R., Marras, T.K., 2015. Epidemiology of human pulmonary infection with non-tuberculous mycobacteria: a review. Clin. Chest Med. 36, 13–34.
Rook, G., 2013. Regulation of the immune system by biodiversity from the natural environment: an ecosystem service essential to health. PNAS 110, 18360–18367.
Rottem, M., Geller-Bernstein, C., Shoenfeld, Y., 2015. Atopy and asthma in migrants: the function of parasites. Int. Arch. Allergy Immunol. 167, 41–46.
Ruokolainen, L., von Hertzen, L., Fyhrquist, N., Laatikainen, T., Lehtomäki, J., Auvinen, P., Karvonen, A.M., Hyvärinen, A., Tillmann, V., Niemelä, O., Knip, M., Haahtela, T., Pekkanen, J., Hanski, I., 2015. Green areas around homes reduce atopic sensitization in children. Allergy 70, 195–202.
Sandifer, P.A., Sutton-Grier, A.E., Ward, B.P., 2015. Exploring connections among nature, biodiversity, ecosystem services, and human health and well-being: opportunities to enhance health and biodiversity conservation. Ecosyst. Serv. 12, 1–15.
Seedorf, H., Griffin, Nicholas W., Ridaura, Vanessa K., Reyes, A., Cheng, J., Rey, Federico E., Smith, Michelle I., Simon, Gabriel M., Scheffrahn, Rudolf H., Woebken, D., Spormann, Alfred M., Van Treuren, W., Ursell, Luke K., Pirrung, M., Robbins-Pianka, A., Cantarel, Brandi L., Lombard, V., Henrissat, B., Knight, R., Gordon, Jeffrey I., 2014. Bacteria from diverse habitats colonize and compete in the mouse gut. Cell 159, 253–266.
Simpson, R., Williams, G., Petroeschevsky, A., Best, T., Morgan, G., Denison, L., Hinwood, A., Neville, G., Neller, A., 2005. The short-term effects of air pollution on daily mortality in four Australian cities. Aust. N. Z. J. Public Health 29, 205–212.
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, 212–215.
State of the Environment 2011 Committee, 2011. Australia state of the Environment 2011. Independent Report to the Australian Government Minister for Sustainability, Environment, Water, Population and Communities. DSEWPaC, Canberra.
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., Neilson, J.W., Maier, R.M., Gilbert, J.A., Holbreich, M., Thorne, P.S., Martinez, F.D., von Mutius, E., Vercelli, D., Ober, C., Sperling, A.I., 2016. Innate immunity and asthma risk in amish and hutterite farm children. N. Engl. J. Med. 375, 411–421.
Tibshirani, R., 1996. Regression shrinkage and selection via the Lasso. J. R. Stat. Soc. Ser. B (Methodological) 58, 267–288.
von Hertzen, L., Haahtela, T., 2006. Disconnection of man and the soil: reason for the asthma and atopy epidemic? J. Allergy Clin. Immunol. 117, 334–344.
von Hertzen, L., Hanski, I., Haahtela, T., 2011. Natural immunity: biodiversity loss and inflammatory diseases are two global megatrends that might be related. EMBO Rep. 12, 1089–1093.

Voreades, N., Kozil, A., Weir, T., 2014. Diet and the development of the human intestinal microbiome. Front. Microbiol. 5, 494.
Whitsett, J.A., Alenghat, T., 2014. Respiratory epithelial cells orchestrate pulmonary innate immunity. Nat. Immunol. 16, 27–35.
WHO, SCBD, 2015. Connecting Global Priorities: Biodiversity and Human Health. A State of Knowledge Review. World Health Organisation, Geneva, Switzerland.
Wopereis, H., Oozeer, R., Knipping, K., Belzer, C., Knol, J., 2014. The first thousand days – intestinal microbiology of early life: establishing a symbiosis. Pediatr. Allergy Immunol. 25, 428–438.
Zammit, C., Liddicoat, H., Moonsie, I., Makker, H., 2010. Obesity and respiratory diseases. Int. J. General Med. 3, 335–343.

Supplementary Material for

Landscape biodiversity correlates with respiratory health in Australia

Craig Liddicoat[a,b,*], Peng Bi[c], Michelle Waycott[a,b], John Glover[d], Andrew J. Lowe[a], Philip Weinstein[a]

[a]School of Biological Sciences and The Environment Institute, The University of Adelaide, North Terrace, Adelaide SA 5005, Australia.

[b]Department of Environment, Water and Natural Resources, GPO Box 1047, Adelaide SA 5001, Australia.

[c]School of Public Health, The University of Adelaide, North Terrace, Adelaide SA 5005, Australia.

[d]Public Health Information Development Unit, Torrens University Australia, Level 1, 200 Victoria Square, Adelaide SA 5000, Australia.

*Author for correspondence: Craig Liddicoat. e-mail: craig.liddicoat@adelaide.edu.au

Contents:

Methods (Supplementary Information)

Environmental covariate preparation

We collated and prepared a wide array of environmental covariates from various sources (Table S3). A standard grid system was adopted matching the approx. 250 m cell size of Geoscience Australia dynamic land cover mapping data [1]. Where necessary, numeric layers were bilinearly resampled and categorical layers were resampled using the nearest neighbor method, in order to match the common grid system.

Focal neighborhood calculations using a 3 km radius were then undertaken for all of the Australia-wide map layers, such that each grid cell subsequently contained a summary value characterizing the surrounding 3 km-radius area. For numeric layers, each grid cell was recalculated to contain the average value of the surrounding area. Categorical layers were converted into numeric layers via (a) iteratively evaluating the proportion of each class within the 3 km-radius area, and (b) calculating the diversity of classes using the Shannon diversity index (H):

$$H = -\sum_{i=1}^{k} p_i \ln(p_i) \qquad \text{(S1.1)}$$

where p_i is the proportion in class k, and k is the number of classes found in each respective 3km-radius area.

A derived layer for the erodible fraction (*EF*) of soils was calculated using the formula proposed by Fryrear *et al.* [2]:

$$EF = [\,29.09 + 0.31 * sand + 0.17 * silt + 0.33 * \left(\frac{sand}{clay}\right) - 4.66 * OC - 0.95 * CaCO_3\,]/100 \qquad \text{(S1.2)}$$

where: *sand* = sand content (%), *silt* = silt content (%); *OC* = organic carbon content (%), $CaCO_3$ = calcium carbonate content (%); and sand, silt and *OC* content layers were sourced from Viscarra Rossel *et al.* [3], while the $CaCO_3$ content layer was sourced from Wilford *et al.* [4].

The majority of spatial data preparation and analysis was performed using the *R* software environment [5], in particular using tools adapted from the *R raster* package [6]. Australia-wide focal neighborhood calculations for class proportions and Shannon diversity indices were implemented using *R* software on high-performance computing facilities provided by eRSA (www.ersa.edu.au).

Environmental variables were finally summarized by averaging values to match Local Government Area (LGA) boundaries corresponding to the available public health and contextual data, using the zonal statistics tool in ESRI ArcGIS 10.2 [7].

Data cleaning due to missing data

The Social Health Atlas of Australia has missing data for some LGAs across the response and candidate predictor variables. To consider the widest possible range of potential explanatory variables, in the 'Moderate majority' and 'Major cities' clusters we first excluded all rows (LGAs) with missing data. This reduced the number of LGAs used (from 451 to 364 for the 'Moderate majority'; and from 72 to 62 for 'Major cities') in the subsequent modelling and analysis.

In the 'Remote disadvantaged' cluster, we first excluded four variables with a high frequency of missing data (Percent smokers, Percent high alcohol consumption, Percent overweight persons, Percent obese persons) and then excluded any remaining rows with missing data. This process reduced the number of LGAs available for analysis from 35 to 24.

Automated screening of candidate predictors

Because the Lasso tool is effectively a machine-learning adaptation of multiple linear regression, we considered the need to screen and potentially transform candidate predictors to help improve linear relationships while balancing the need for interpretability (of predictors) and transparency in any transformation process. To achieve this in a transparent, objective fashion, we developed an automated screening algorithm (using *R* script) to scrutinise and where necessary transform or eliminate candidate predictors prior to input to the Lasso. The algorithm was run separately for each LGA cluster and comprised the following steps:

Step 1: Variables were excluded if there were less than 50% of the total LGAs with non-zero observations from which to infer a relationship. This was designed to eliminate predictors that were present in the Australia-wide data but were considered non-representative for the particular socio-geographic cluster under consideration.

Step 2: The following transformations were calculated for each candidate predictor for later comparison against each other and the original (untransformed) data:

- Logit (log odds) transformations were the only alternative considered for proportion and percentage-based data, with lower and upper bounds (i.e. 0, 100%) remapped to the 2.5th and 97.5th percentile respectively.
- Square root (sqrt), which was disqualified if negative values were present.
- Log10, which was disqualified if negative or zero values were present.
- Square
- Cube root, which was only considered where negative values were present.

Step 3: Pragmatic threshold acceptance criteria were set for the original and transformed variables to be included in the modelling. Any candidate predictors not meeting these criteria were excluded:

- absolute value of skewness <= 1.5
- kurtosis <= 4

- coefficient of determination (R^2) between the candidate predictor and response variable >= 0.025 (i.e. at least 2.5% stand-alone explanatory value).

Step 4: Representatives for each variable were then chosen. If only one option remained (whether original or transformed), that was chosen. Then, for proportion and percentage-based variables, the most normally distributed data (comparing only original and logit transforms) were chosen, as judged by the maximum Shapiro-Wilk normality test p-value. For non-proportion-based variables, if the original form passed the acceptance criteria then that was chosen as a priority for ease of interpretation. If the original was not eligible, then the most normally-distributed data from the remaining transformations was chosen (again, as judged by the maximum Shapiro-Wilk normality test p-value). For a number of variables, no options passed the acceptance criteria, and they were eliminated from the modelling.

All of the above analyses, and subsequent modelling, were performed on candidate predictor variables that had been subject to 95% Winsorization (based on data within each cluster), which was designed to eliminate the influence of extreme and potentially outlying values [8]. This involved all extreme low values below the 2.5th percentile being replaced by the 2.5th percentile, and all extreme high values above the 97.5th percentile being replaced by the 97.5th percentile.

Predictor data were then centered and scaled (based on data within each cluster) prior to the 10-fold Lasso modelling.

Statistical analysis

The performance of the 10-fold Lasso modelling was assessed using mean cross-validation statistics for: root mean-square error, mean error (or bias), skewness of model residuals, coefficient of determination (R^2), and concordance correlation coefficient [9]. That is, health response observations from LGAs in each respective validation set (not used to develop the respective k-fold Lasso model) were compared against respective predictions based on the validation set predictor data. These results were then averaged over the 10 validation sets. This process was repeated for each socio-geographic cluster (see Table 1 of main article). The R^2 values were computed from the square of the Pearson correlation coefficient and indicate how well predictions and observations adhere to a straight line (least-squares linear fit). The concordance correlation coefficient indicates how well predictions and observations adhere to a 1:1 relationship, with values closer to 1 indicating greater agreement. The skewness of residuals indicates how balanced the model is in terms of under- and over-predictions. Table 1 also reports the standard deviation of observations (SD (obs)). Performance measures were calculated using the same form of health response data as used in modelling, i.e. using log10-transformed response data for the 'Moderate majority' and 'Remote disadvantaged', and untransformed data for the 'Major cities'. Validation set predictions and observations are plotted in Fig. S6.

To examine the significance of Diversity of major vegetation groups within quartiles of Socioeconomic index (using 'Moderate majority' data only, Fig. S7) we constructed standard multiple linear regression models for the health response (log10(respiratory disease public hospital admissions)) using only Socioeconomic index and Diversity of major

vegetation groups as explanatory variables. Variable significance was assessed using p-values for the regression coefficients from the respective multiple linear regression models. Throughout this study, where standard linear regression models are applied, these use base R software [5]. Testing of Lasso-identified predictor significance (i.e. selection-adjusted p-values, Tables S6-S7) use the R SelectiveInference package [10].

Extended analysis: examining variable significance

Based on the 10-fold Lasso modelling results, and using the 'Moderate majority' data only, we undertook two further analyses to better understand variable significance. Firstly, we examined the influence of Diversity of major vegetation groups, while controlling for Socioeconomic influence (Fig. S7). Within quartiles of Socioeconomic index, we considered only Socioeconomic index and Diversity of major vegetation groups as explanatory variables in separate multiple linear regression models for respiratory health. In every quartile, Diversity of major vegetation groups was identified as significant predictor. However, only in the wealthiest quartile was Socioeconomic index also identified as a significant predictor.

Secondly, we wished to understand whether predictors identified by the Lasso could be viewed as significant. We noted that the assessment of significance using contemporary machine-learning tools is a developing field of statistical science, and new tools for investigating selective inference with the Lasso provide more conservative p-values than might be expected from traditional methods [11]. To examine the significance of results from the Lasso we re-analysed all 'Moderate majority' data, this time considering a two-part random split (each with n=182 LGAs). (The reason for splitting data is that the Lasso tool is able to provide important insights into a particular dataset, and can cherry-pick variables with the most explanatory value for that dataset. From a purist statistical science viewpoint, it is therefore not appropriate to reuse the same dataset and the cherry-picked (Lasso-identified) variables in a traditional multiple linear regression model, as this will not provide a fair and independent assessment of variable significance.)

On data-split one we ran a single iteration of the Lasso (Fig. S8), and determined the selection-adjusted p-values [10] for coefficients of identified non-zero predictors. Ordering these Lasso-identified predictors by the absolute size of standardized regression coefficients, we then inputted respective predictor data from data-split two into standard multiple linear regression software. This generated an independently-generated and contrasting set of regression coefficients and p-values (Table S6). For completeness, we also reversed these operations: running the Lasso on data-split two (Fig. S9), and then entering the corresponding identified predictor data from data-split one into a standard multiple linear regression (with results in Table S7). For this extended Lasso analysis we used the default internal 10-fold cross-validation setting when calling cv.glmnet to identify the optimal penalisation parameter (using the parsimonious option, s = 'lambda.1se') and corresponding Lasso regression coefficients.

Results show some consistency with similar themes of predictors identified in comparison to the 10-fold modelling. However, different results were obtained because of differences in input data. P-values provided by the selective inference software were

conservative, however many of the predictors ranked highly by the Lasso (when ordered by absolute size of standardized regression coefficients) also showed up as significant predictors in the complementary data-split (i.e. using data kept separate from the Lasso variable-selection process).

Associated data

The following data are available for download from the article web page on ScienceDirect:

- Database S1 – All data in raw form (558 local government areas x 192 candidate explanatory variables)
- Example R scripts for:
    - Continent-wide 3 km-radius focal neighbourhood calculation - Shannon diversity index
    - Continent-wide 3 km-radius focal neighbourhood calculation - class proportion #1
    - Modelling workflow for Australia-wide 10-fold Lasso modelling of respiratory health by local government areas

6

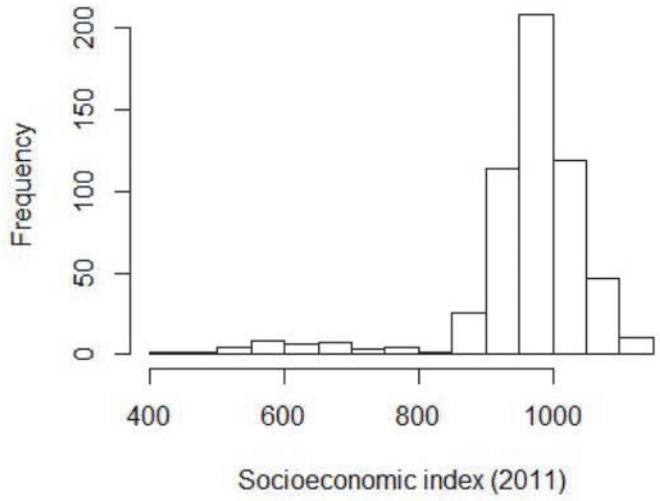


Fig. S1. Histogram of Australia-wide Socioeconomic index

Frequency distribution of Socioeconomic index (termed 'Index of relative socioeconomic disadvantage' in the Social Health Atlas of Australia [12, 13]) for n=558 local government areas across Australia. The left tail is suggestive of a second minor mode centered on communities of low socioeconomic status. Note that low values of Socioeconomic index correspond to greater disadvantage or deprivation, while high values correspond to higher socioeconomic status.

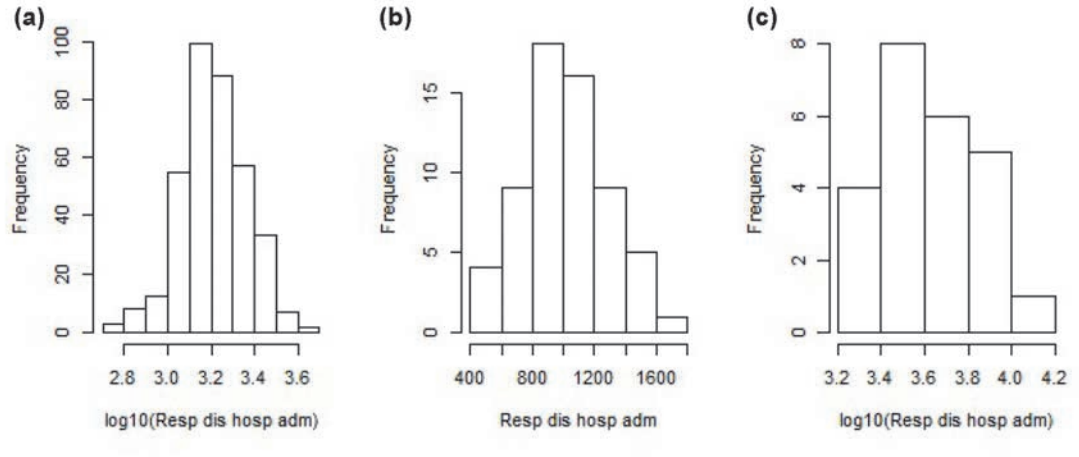


Fig. S2. Histograms of respiratory disease response variables

Frequency distributions of respiratory disease public hospital admissions (health response variable) for local government area clusters: (a) 'Moderate majority' [log10-transformed, n=364, mean ± sd: 3.210 ± 0.150], (b) 'Major cities' [n=62, mean ± sd: 1022 ± 267], and (c) 'Remote disadvantaged' [log10-transformed, n=24, mean ± sd: 3.612 ± 0.213]. Raw units are age standardized rate per 100,000.

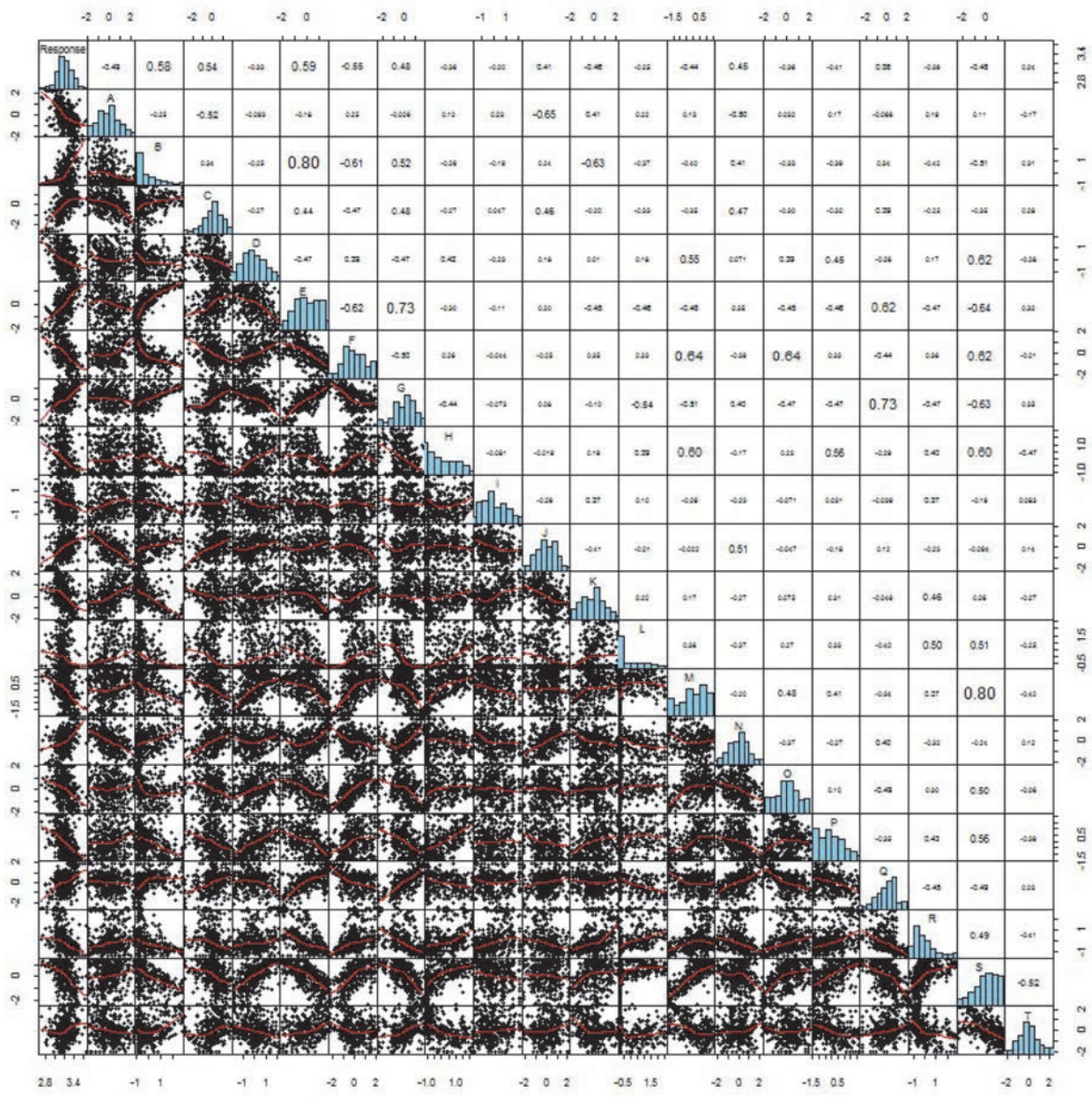


Fig. S3. Scatterplot matrix with correlation coefficients for response and predictor variables identified from 10-fold cross-validation Lasso modelling of the 'Moderate majority' cluster.

Variables labelled on the diagonal are 'Response', i.e. log10(respiratory disease public hospital admissions)', (A) Socioeconomic index, (B) Distance to coast, (C) Percent obese persons, (D) Diversity of major vegetation groups, (E) Mean temperature annual range, (F) Species richness (log10), (G) Maximum temperature of warmest month, (H) Proportion of eucalypt forests 10-30m (logit), (I) Percent overweight persons (logit), (J) Percent smoking during pregnancy, (K) Percent English-speaking immigrants (logit), (L) Proportion of warm wet plains (logit), (M) Proportion of open trees (logit), (N) Percent Aboriginal persons (logit), (O) Diversity of land use, (P) Proportion of nature conservation (logit), (Q) Vegetation fractional cover minimum nonphotosynthetic (logit), (R) Mean precipitation of the coldest quarter, (S) Vegetation fractional cover minimum photosynthetic (logit), (T) Soil cation exchange capacity * erodible fraction (geometric mean). Panels below the diagonal contain scatterplots with locally weighted scatterplot smoothing ('lowess') curves; and above the diagonal contain the calculated Pearson correlation coefficient (r). Variable histograms are included on the diagonal.

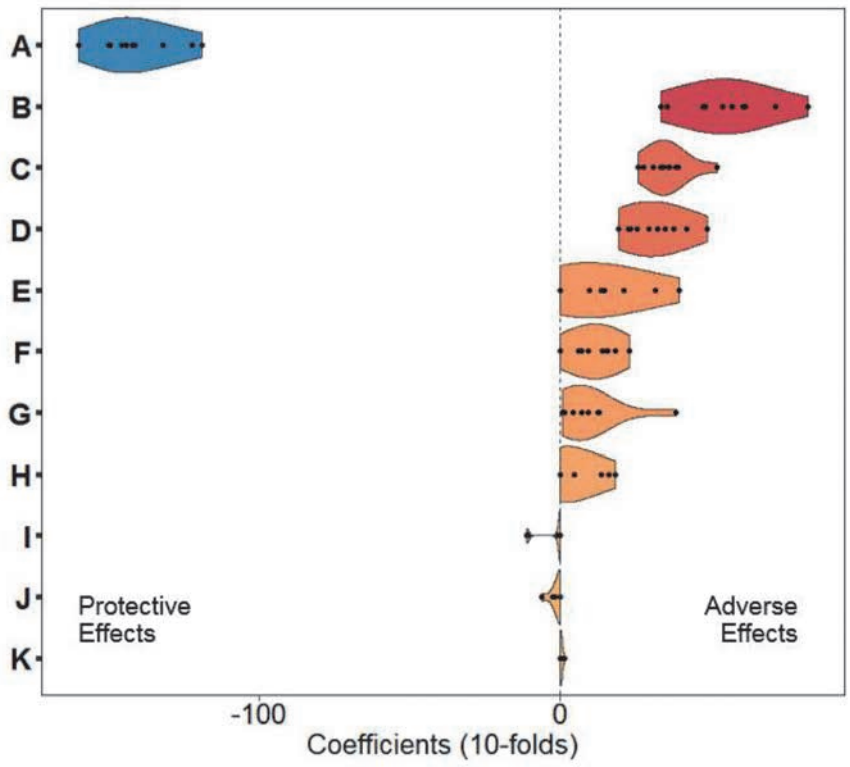


Fig. S4. Important predictors for the 'Major cities' cluster (n=62).

Density and point plots of standardized regression coefficients from 10-fold cross-validation Lasso modelling of respiratory disease public hospital admissions. Eleven predictors were identified across the 10-fold Lasso models: (A) Socioeconomic index, (B) Mean temperature of the wettest quarter, (C) Proportion of rainfed pasture (logit), (D) Percent Aboriginal persons (logit), (E) Percent of smokers, (F) Percent obese persons, (G) Percent smoking during pregnancy, (H) Vegetation fraction of photosynthetically-active radiation maximum (logit), (I) Total population, (J) Percent overseas-born from non-English-speaking countries resident less than 5 years (logit), (K) Proportion of tussock grass open (logit).

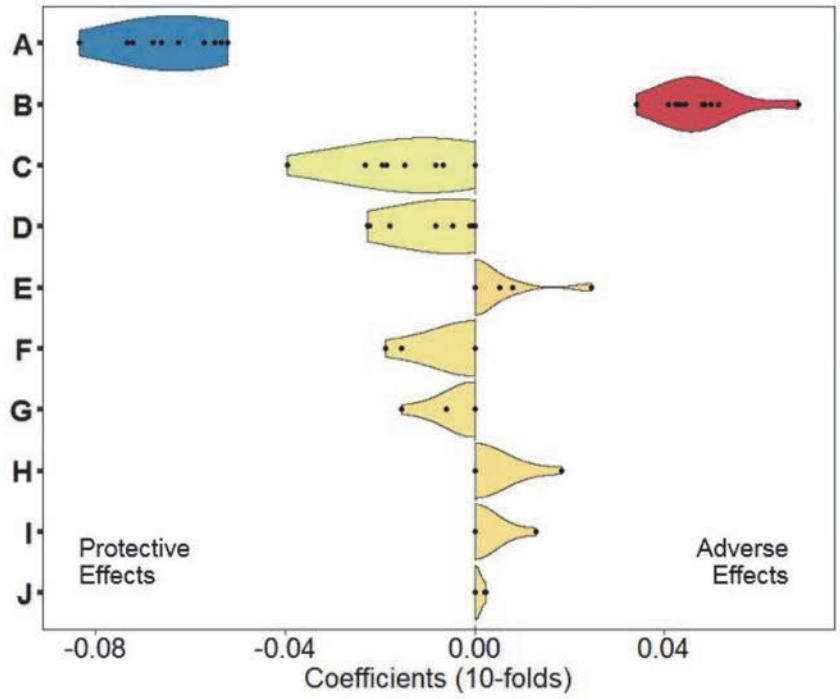


Fig. S5. Important predictors for the 'Remote disadvantaged' cluster (n=24).

Density and point plots of standardized regression coefficients from 10-fold cross-validation Lasso modelling of respiratory disease public hospital admissions. Ten predictors were identified across the 10-fold Lasso models: (A) Proportion of wetlands (logit), (B) Mean temperature of the wettest quarter, (C) Proportion of acacia forests and woodlands (logit), (D) Proportion of swampy grasses and sedges (logit), (E) Proportion of hummock grass sparse (logit), (F) Air temperature annual mean isothermality, (G) Proportion of eucalypt woodlands (logit), (H) Vegetation fractional cover standard deviation – bare soil, (I) Proportion of scattered trees (logit), (J) Vegetation fractional cover maximum – bare soil (logit).

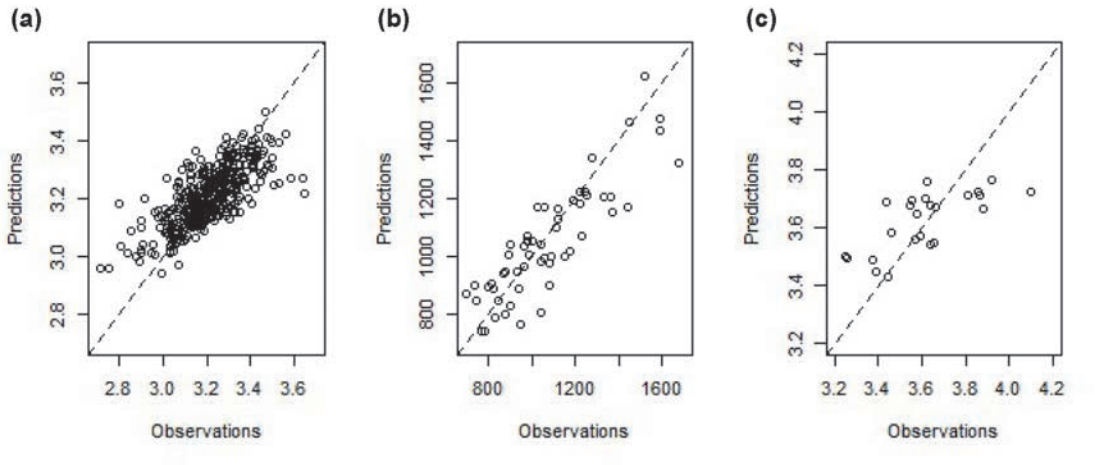
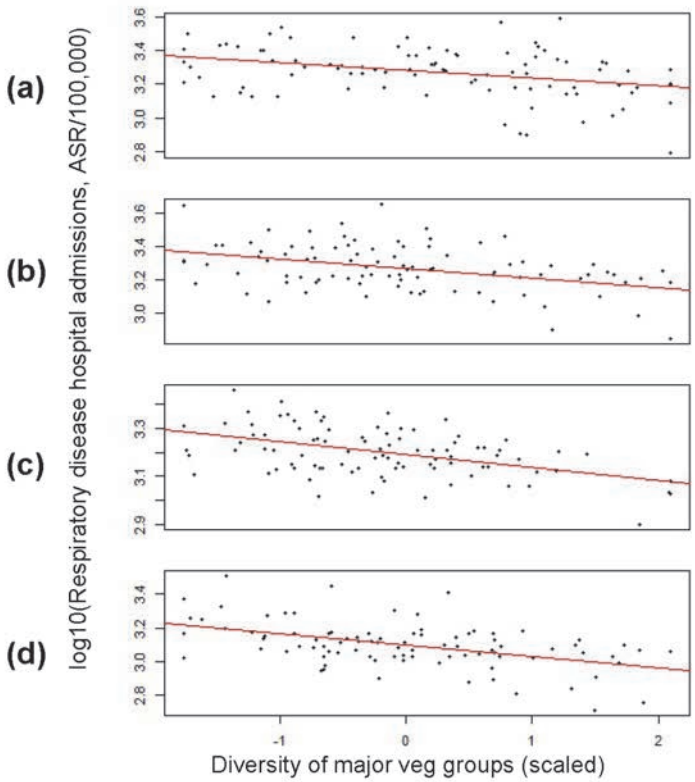


Fig. S6. Cross-validation plots

10-fold cross-validation predictions versus observations for Lasso penalized regression modelling of respiratory disease public hospital admissions, for local government area clusters: (a) 'Moderate majority', (b) 'Major cities', and (c) 'Remote disadvantaged'. Note: response variables for clusters (a) and (c) are log10-transformed.



	p-value for regression coefficient		Correlation (r)
	Socio-economic index	Div. of major veg. groups	Div. of major veg. groups & Socioeconomic index
(a)	0.3728 -	0.0005 ***	-0.2037
(b)	0.0437 *	0.0001 ***	0.0746
(c)	0.0665 .	9e-07 ***	-0.1350
(d)	1e-05 ***	2e-06 ***	0.1259

Fig. S7. Relationship between Diversity of major vegetation groups, Socioeconomic index, and respiratory health response for the 'Moderate majority' cluster—within quartiles of Socioeconomic index.

Plots show data for local government areas split by: (a) 0-25th, (b) 25-50th, (c) 50-75th, and (d) 75-100th percentiles (i.e. quartiles) of Socioeconomic index. For each socioeconomic quartile (n=91), the corresponding table row on the right contains p-values for regression coefficients from a multiple linear regression model for the health response—log10(respiratory disease public hospital admissions)—with only Socioeconomic index and Diversity of major vegetation groups as explanatory variables. Significance codes are: 0–0.001: '***', 0.001–0.01: '**', 0.01–0.05: '*', 0.05– 0.1: '.' The third column in the table shows the respective Pearson correlation coefficient (r) calculated between Diversity of major vegetation groups and Socioeconomic index (indicating they have low levels of correlation).

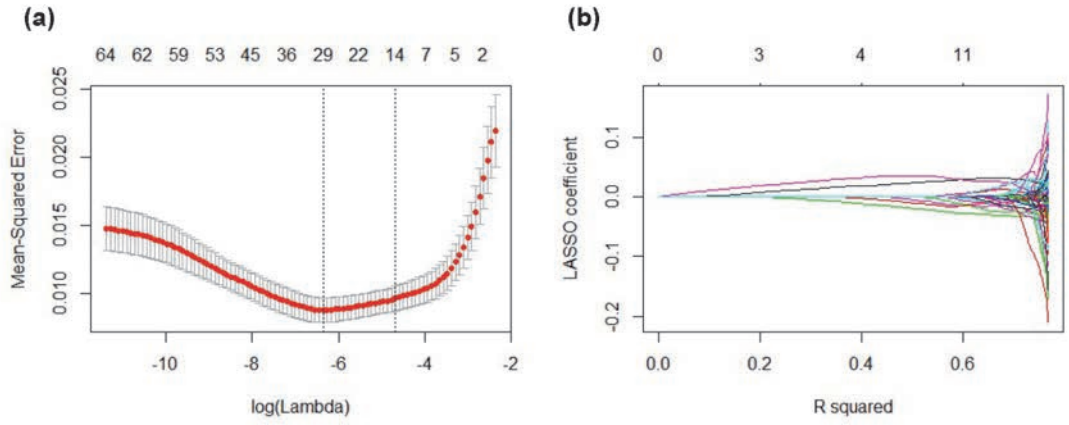


Fig. S8. Alternate Lasso model parameters for the 'Moderate majority' cluster – based on *data-split one* training data.

Panel (a) shows mean square error on the y-axis, the log of the penalisation parameter is shown on the lower x-axis, and the increasing number of non-zero predictors introduced by alternate Lasso models is shown on the top axis. Panel (b) shows increasing R^2 on the lower x-axis (eventually indicating over-fitting to the training data) for alternate regression models where the number of predictors (upper x-axis) is allowed to increase. Both panels indicate the 'sweet spot' of low mean square error, moderately high R^2 value, and parsimony in the number of predictors selected.

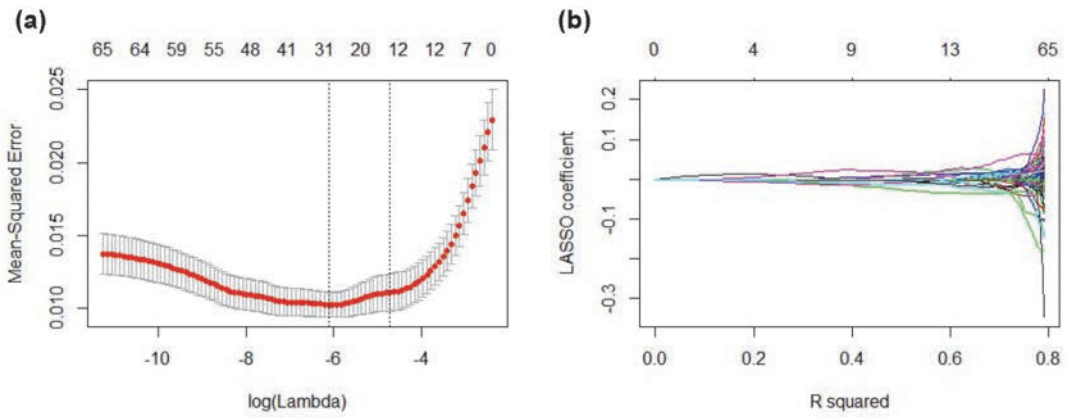


Fig. S9. Alternate Lasso model parameters for the 'Moderate majority' cluster – based on internal cross-validation of *data-split <u>two</u>* training data.

Panel (a) shows mean square error on the y-axis, the log of the penalisation parameter is shown on the lower x-axis, and the increasing number of non-zero predictors introduced by alternate Lasso models is shown on the top axis. Panel (b) shows increasing R^2 on the lower x-axis (eventually indicating over-fitting to the training data) for alternate regression models where the number of predictors (upper x-axis) is allowed to increase. Both panels indicate the 'sweet spot' of low mean square error, moderately high R^2 value, and parsimony in the number of predictors selected.

Table S1. International classification of diseases 10th revision Australian modification (ICD-10-AM) codes for diseases of the respiratory system

This study has used aggregated data for public hospital admissions with principal diagnoses of diseases of the respiratory system (ICD-10-AM codes J00-J99 [14]), listed below, as reported by the Social Health Atlas of Australia [12, 13]. To provide an indicative breakdown of individual disease cases, separations from all hospitals (public + private) are shown in the right-hand columns [15]. Note: at the time of writing, separation data were not available for public hospitals only.

Sub-category	Disease	Separations (% of total)	
		2011-12	2012-13
Acute upper respiratory infections (J00-J06)			
	Acute nasopharyngitis [common cold] (J00)	186 (<0.05%)	164 (<0.05%)
	Acute sinusitis (J01)	1039 (0.3%)	900 (0.2%)
	Acute pharyngitis (J02)	2364 (0.6%)	2145 (0.5%)
	Acute tonsillitis (J03)	12195 (3%)	11617 (2.9%)
	Acute laryngitis and tracheitis (J04)	447 (0.1%)	406 (0.1%)
	Acute obstructive laryngitis [croup] and epiglottitis (J05)	7308 (1.8%)	5799 (1.5%)
	Acute upper respiratory infections of multiple and unspecified sites (J06)	**14463 (3.6%)**	**13883 (3.5%)**
Influenza and pneumonia (J09-J18)			
	Influenza due to certain identified influenza virus (J09)	687 (0.2%)	184 (<0.05%)
	Influenza due to other identified influenza virus (J10)	2400 (0.6%)	4846 (1.2%)
	Influenza, virus not identified (J11)	1178 (0.3%)	1316 (0.3%)
	Viral pneumonia, not elsewhere classified (J12)	3508 (0.9%)	3798 (1%)
	Pneumonia due to Streptococcus pneumoniae (J13)	1775 (0.4%)	2017 (0.5%)
	Pneumonia due to Haemophilus influenzae (J14)	1151 (0.3%)	1237 (0.3%)
	Bacterial pneumonia, not elsewhere classified (J15)	3862 (1%)	3908 (1%)
	Pneumonia due to other infectious organisms, not elsewhere classified (J16)	No data	No data
	Pneumonia, organism unspecified (J18)	**67507 (16.8%)**	**63791 (16%)**
Other acute lower respiratory infections (J20-J22)			
	Acute bronchitis (J20)	2642 (0.7%)	2338 (0.6%)
	Acute bronchiolitis (J21)	**18347 (4.6%)**	**19152 (4.8%)**
	Unspecified acute lower respiratory infection (J22)	**24099 (6%)**	**23631 (5.9%)**
Other diseases of upper respiratory tract (J30-J39)			
	Vasomotor and allergic rhinitis (J30)	681 (0.2%)	645 (0.2%)
	Chronic rhinitis, nasopharyngitis and pharyngitis (J31)	1115 (0.3%)	1297 (0.3%)
	Chronic sinusitis (J32)	10970 (2.7%)	11065 (2.8%)
	Nasal polyp (J33)	3419 (0.9%)	3313 (0.8%)
	Other disorders of nose and nasal sinuses (J34)	**25627 (6.4%)**	**26089 (6.5%)**
	Chronic diseases of tonsils and adenoids (J35)	**40813 (10.2%)**	**41384 (10.4%)**
	Peritonsillar abscess (J36)	3251 (0.8%)	3283 (0.8%)
	Chronic laryngitis and laryngotracheitis (J37)	84 (<0.05%)	78 (<0.05%)
	Diseases of vocal cords and larynx, not elsewhere classified (J38)	4436 (1.1%)	4516 (1.1%)
	Other diseases of upper respiratory tract (J39)	1752 (0.4%)	1677 (0.4%)
Chronic lower respiratory diseases (J40-J47)			
	Bronchitis, not specified as acute or chronic (J40)	2451 (0.6%)	2115 (0.5%)
	Simple and mucopurulent chronic bronchitis (J41)	41 (<0.05%)	57 (<0.05%)
	Unspecified chronic bronchitis (J42)	420 (0.1%)	429 (0.1%)
	Emphysema (J43)	419 (0.1%)	349 (0.1%)
	Other chronic obstructive pulmonary disease (J44)	**61893 (15.4%)**	**62151 (15.6%)**
	Asthma (J45)	**36176 (9%)**	**35257 (8.8%)**
	Status asthmaticus (J46)	2505 (0.6%)	2267 (0.6%)

Bronchiectasis (J47)	5185 (1.3%)	5388 (1.4%)
Lung diseases due to external agents (J60-J70)		
Coalworker's pneumoconiosis (J60)	25 (<0.05%)	27 (<0.05%)
Pneumoconiosis due to asbestos and other mineral fibres (J61)	132 (<0.05%)	124 (<0.05%)
Pneumoconiosis due to dust containing silica (J62)	35 (<0.05%)	25 (<0.05%)
Pneumoconiosis due to other inorganic dusts (J63)	No data	No data
Unspecified pneumoconiosis (J64)	5 (<0.05%)	6 (<0.05%)
Pneumoconiosis associated with tuberculosis (J65)	No data	No data
Airway disease due to specific organic dust (J66)	No data	No data
Hypersensitivity pneumonitis due to organic dust (J67)	150 (<0.05%)	106 (<0.05%)
Respiratory conditions due to inhalation of chemicals, gases, fumes and vapours (J68)	20 (<0.05%)	16 (<0.05%)
Pneumonitis due to solids and liquids (J69)	9857 (2.5%)	9894 (2.5%)
Respiratory conditions due to other external agents (J70)	236 (0.1%)	230 (0.1%)
Other respiratory diseases principally affecting the interstitium (J80-J84)		
Adult respiratory distress syndrome (J80)	156 (<0.05%)	148 (<0.05%)
Pulmonary oedema (J81)	689 (0.2%)	609 (0.2%)
Pulmonary eosinophilia, not elsewhere classified (J82)	277 (0.1%)	271 (0.1%)
Other interstitial pulmonary diseases (J84)	3594 (0.9%)	3638 (0.9%)
Suppurative and necrotic conditions of lower respiratory tract (J85-J86)		
Abscess of lung and mediastinum (J85)	531 (0.1%)	532 (0.1%)
Pyothorax (J86)	722 (0.2%)	733 (0.2%)
Other diseases of pleura (J90-J94)		
Pleural effusion, not elsewhere classified (J90)	6696 (1.7%)	6773 (1.7%)
Pleural plaque (J92)		
Pneumothorax (J93)	3189 (0.8%)	3274 (0.8%)
Other pleural conditions (J94)	326 (0.1%)	345 (0.1%)
Other diseases of the respiratory system (J95-J99)		
Postprocedural respiratory disorders, not elsewhere classified (J95)	No data	No data
Respiratory failure, not elsewhere classified (J96)	3476 (0.9%)	3683 (0.9%)
Other respiratory disorders (J98)	5514 (1.4%)	5958 (1.5%)

(Separations > 3% annual total are highlighted in bold)

Table S2. Socio-geographic variables

The following socio-geographic variables were used as candidate predictor data, sourced from the Social Health Atlas of Australia dataset [13], except Area of local government area and Population density which were derived from spatial mapping of LGAs [16].

Variable	Units	Reference period
Total population	No. persons	2013
Area of local government area	km^2	N/A
Population density	Persons/km^2	2013
Percent Aboriginal persons	%	2013
Percent Australian-born persons	%	2011
Percent born overseas in predominantly English-speaking countries	%	2011
Percent born overseas in a predominantly non-English-speaking country	%	2011
Percent born overseas in a predominantly non-English-speaking country, resident in Australia for 5 years or more	%	2011
Percent born overseas in a predominantly non-English-speaking country, resident in Australia for less than 5 years	%	2011
*Socioeconomic index	No units	2011
Percent smoking during pregnancy	%	2008 to 2010 (NSW, Qld, SA and ACT), 2009 to 2011 (Vic, WA, Tas), 2006 to 2008 (NT)
Percent smokers (estimated population rates, 18 years and over)	ASR /100	2011-13
Percent high alcohol consumption (estimated population rates, 18 years and over)	ASR/100	2011-13
#Percent overweight persons (estimated population rates, 18 years and over)	ASR/100	2011-13
#Percent obese persons (estimated population rates, 18 years and over)	ASR/100	2011-13

N/A = not applicable, ASR = age standardized rate

*Socioeconomic index was originally called Index of relative socioeconomic disadvantage in the Social Health Atlas of Australia. Note that low values correspond to greater disadvantage or deprivation, while high values correspond to reduced disadvantage (or higher socioeconomic status).

#In our data, obese denotes body mass index > 30, while overweight denotes body mass index 25 to <30.

Table S3. Environmental variables

The following environmental variables provided the basis for candidate predictor data (all listed in Table S5). Broad themes of environmental data are: climate (C), ecological (E), geographic (G), land use (LU), pollution (P), soil parameters (S), vegetation or land cover classes (VLC), and vegetation indices from remote sensing (VRS). Units are displayed for numeric variables, whereas the number of classes (in brackets) are displayed for categorical† variables. Further preparatory steps for gridded environmental variables (i.e. 3 km radius focal calculations for numeric average, and conversion of categorical layers to class proportions and Shannon diversity indices), and subsequent selective transformations, are described in the *Methods* (refer to main article). #Focal calculations were not performed for point-based PM10 emissions (air pollution) data, as no exhaustive mapping was available; instead PM10 sites were intersected within each LGA, expressed as total/area, and averaged for the 2 year period. Additional abbreviations are explained in the Table footnote.

Variable	Theme	Units / (classes†)	Reference period	Data source
#Air pollution: Total industry PM10 emissions (mean 2011-12 and 2012-13)/LGA area	P	kg/km^2	2011-2013	[17, 18]
Air temperature annual mean diurnal range	C	°C	1970-2012	[19]
Air temperature annual mean isothermality	C	°C*100	1970-2012	[19]
Annual mean precipitation	C	mm	1970-2012	[19]
Annual mean temperature	C	°C	1970-2012	[19]
Annual precipitation seasonality (coefficient of variation)	C	N/A	1970-2012	[19]
Annual temperature seasonality (standard deviation*100)	C	°C*100	1970-2012	[19]
Distance to coast	G	Decimal degrees	N/A	Calc.
Ecological land units (combinations of bioclimate and landform)†	E	(39)	2012-2013	[20]
Fire frequency mapping for Australia	P	No. years burnt	1997-2010	[21]
Land cover†	VLC	(34)	2000-2008	[1]
Land use†	LU	(15)	1997-2014	[22]
Major vegetation groups†	VLC	(32)	2012	[23]
Maximum temperature of the warmest month	C	°C	1970-2012	[19]
Mean precipitation of the coldest quarter	C	mm	1970-2012	[19]
Mean precipitation of the driest month	C	mm	1970-2012	[19]
Mean precipitation of the driest quarter	C	mm	1970-2012	[19]

Variable	Theme	Units / (classes†)	Reference period	Data source
Mean precipitation of the warmest quarter	C	mm	1970-2012	[19]
Mean precipitation of the wettest month	C	mm	1970-2012	[19]
Mean precipitation of the wettest quarter	C	mm	1970-2012	[19]
Mean temperature annual range	C	°C	1970-2012	[19]
Mean temperature of the coldest quarter	C	°C	1970-2012	[19]
Minimum temperature of the coldest month	C	°C	1970-2012	[19]
Mean temperature of the driest quarter	C	°C	1970-2012	[19]
Mean temperature of the warmest quarter	C	°C	1970-2012	[19]
Mean temperature of the wettest quarter	C	°C	1970-2012	[19]
Prescott Index	C	N/A	1981-2006	[24]
Soil clay content, in the < 2 mm fraction (0-5 cm)	S	%	2014	[3]
Soil effective cation exchange capacity (0-5 cm)	S	meq/100g	2014	[3]
Soil effective cation exchange capacity (0-5 cm) * Soil erodible fraction * Vegetation fractional cover – mean BS; calculated as a geometric mean (intended to represent possible exposures to high clay and organic matter content soils, weighted by erodibility and soil cover)	S	N/A	2000-2012	Calc.
Soil effective cation exchange capacity (0-5 cm) * Soil erodible fraction; calculated as a geometric mean (intended to represent possible exposures to high clay and organic matter content soils, weighted by erodibility)	S	N/A	2014	Calc.
Soil effective cation exchange capacity (0-5 cm) * Vegetation fractional cover – mean BS; calculated as a geometric mean (intended to represent possible exposures to high clay and organic matter content soils, weighted by soil cover)	S	N/A	2000-2012	Calc.
Soil erodible fraction; calculated using the method of Fryrear *et al.* [2]	S	%	2014	Calc.
Soil erodible fraction * Vegetation fractional cover – mean BS; calculated as a geometric mean (intended to represent possible soil exposures due to erodible and bare soils)	S	N/A	2000-2012	Calc.
Soil organic carbon content, in the < 2 mm fraction (0-5 cm)	S	%	2014	[3]
Soil organic carbon stocks in the top 30 cm	S	$t.ha^{-1}$	2010	[25]
Soil pH (in $CaCl_2$) (0-5 cm)	S	pH	2014	[3]
Soil sand content, in the < 2 mm fraction (0-5 cm)	S	%	2014	[3]
Soil silt content, in the < 2 mm fraction (0-5 cm)	S	%	2014	[3]

Variable	Theme	Units / (classes†)	Reference period	Data source
Species richness (average of all biological species occurrences in 9 pane moving window, where each pane is 0.01 degrees latitude/ longitude, or ~1km^2)	E	No. of species	N/A	[26]
Vegetation FPAR maximum	VRS	Fraction	1981-2011	[27]
Vegetation FPAR mean	VRS	Fraction	1981-2011	[27]
Vegetation FPAR median	VRS	Fraction	1981-2011	[27]
Vegetation FPAR minimum	VRS	Fraction	1981-2011	[27]
Vegetation FPAR standard deviation	VRS	Fraction	1981-2011	[27]
Vegetation fractional cover – maximum BS	VRS	%	2000-2012	[28]
Vegetation fractional cover – maximum NPV	VRS	%	2000-2012	[28]
Vegetation fractional cover – maximum PV	VRS	%	2000-2012	[28]
Vegetation fractional cover – mean BS	VRS	%	2000-2012	[28]
Vegetation fractional cover – mean NPV	VRS	%	2000-2012	[28]
Vegetation fractional cover – mean PV	VRS	%	2000-2012	[28]
Vegetation fractional cover – min BS	VRS	%	2000-2012	[28]
Vegetation fractional cover – min NPV	VRS	%	2000-2012	[28]
Vegetation fractional cover – min PV	VRS	%	2000-2012	[28]
Vegetation fractional cover – standard deviation BS	VRS	%	2000-2012	[28]
Vegetation fractional cover – standard deviation NPV	VRS	%	2000-2012	[28]
Vegetation fractional cover – standard deviation PV	VRS	%	2000-2012	[28]

Abbreviations used: N/A = not applicable; Calc. = calculated; FPAR = Fraction of photosynthetically active radiation (i.e. radiation signal due to living green vegetation); PV = photosynthetic (living green) vegetation; NPV = non-photosynthetic vegetation (e.g. non-living straw, stubble, plant remnants), BS = bare soil.

Note: Summary statistical layers (minimum, mean, median, maximum, standard deviation) for time-series vegetation FPAR and fractional cover products were previously calculated by CSIRO Land and Water for collaborative Australia-wide predictive soil and landscape mapping via the Terrestrial Ecosystem Research Network 'Soil and Landscape Grid of Australia' [29].

Table S4. Socio-geographic clustering of Australian local government areas

The total available number of local government areas (LGAs), as indicated, were used in k-means clustering, however a reduced number [in brackets] were used in subsequent modelling due to reduced availability of public health response and candidate predictor data. Cluster means and interquartile ranges (in brackets) are displayed, with scaled cluster centers beneath[#].

Description	n	Cluster characteristic values		
		Socioeconomic index	Percent Aboriginal persons (%)	Population density (persons/km^2)
Cluster A. 'Moderate majority': The majority of Australian LGAs with moderate socioeconomic status and low to moderate population density	451 [364]	974 (945-1003) 0.1164[#]	6.5 (2.0-7.3) -0.1815[#]	95.8 (0.474-22.5) -0.3306[#]
Cluster B. 'Major cities': Highest population density, highest average wealth LGAs concentrated around major capital cities	72 [62]	1041 (1010-1126) 0.7765[#]	0.8 (0.4-1.0) -0.5066[#]	2978 (2087-3531) 2.272[#]
Cluster C. 'Remote disadvantaged': Lowest population density, highest percent Aboriginal persons, and lowest average socioeconomic status LGAs spanning the arid inland to high rainfall northern Australia.	35 [24]	649 (583-696) -3.097[#]	69.6 (54.0-87.4) 3.380[#]	3.1 (0.029-0.856) -0.4142[#]

Notes:

1. Cluster B 'Major cities' contain LGAs associated with capitals Perth, Adelaide, Melbourne, and Sydney. The capital of Brisbane was not included in this cluster due to larger LGA boundaries being defined in the official LGA dataset in that region, which resulted in lower population densities.

2. Summary descriptive statistics for areas of LGAs ultimately used in the modelling are (note these area distributions are positively skewed):

- Cluster A. 'Moderate majority' (n=364): median 2779 km^2; interquartile range 921-5177 km^2; range 9-93211 km^2
- Cluster B. 'Major cities' (n=62): median 27 km^2; interquartile range 14-54 km^2; range 3-103 km^2
- Cluster C. 'Remote disadvantaged' (n=24): median 43047 km^2; interquartile range 1693-159867 km^2; range 70-320706 km^2

Table S5. Candidate predictors selected (and not selected) from Lasso modelling

From the 10-fold Lasso modelling, mean positive association, mean inverse association, low coefficients of questionable reliability (?), and variables not selected by the Lasso (blank) are denoted below. Variables with mean positive coefficients (+, shaded orange) associate with increased hospital admissions, and those with mean negative coefficients (–, shaded light blue) associate with decreased hospital admissions.

Candidate predictors	Sign of mean association identified from 10-fold Lasso modelling, by LGA cluster		
	'Moderate majority'	'Major cities'	'Remote disadvantaged'
Social variables			
Total population		–?	
Area of local government area			
Population density			
Percent Aboriginal persons (logit)	+	+	
Percent Australian-born persons			
Percent born overseas in predominantly English-speaking countries (logit)	–		
Percent born overseas in a predominantly non-English-speaking country			
Percent born overseas in a predominantly non-English-speaking country, resident in Australia for 5 years or more			
Percent born overseas in a predominantly non-English-speaking country, resident in Australia for less than 5 years (logit)		–?	
Socioeconomic index	–	–	
Percent smoking during pregnancy	+	+	
Percent smokers (estimated population rates, 18 years and over)		+	
Percent high alcohol consumption (estimated population rates, 18 years and over)			
Percent overweight persons (estimated population rates, 18 years and over) (logit)	–		
Percent obese persons (estimated population rates, 18 years and over)	+	+	
Environmental variables (3 km focal class proportion^, Shannon diversity index†, mean‡)			
Species richness (log10)^	–		
Diversity of land cover†			
Diversity of land use†	–		
Diversity of ecological land units†			
Diversity of major vegetation groups†	–		
Land cover #1: Proportion of extraction sites^			
Land cover #10: Proportion of rainfed sugar^			
Land cover #11: Proportion of wetlands (logit)^			–
Land cover #12: Proportion of forbs open^			
Land cover #13: Proportion of forbs sparse^			
Land cover #14: Proportion of tussock grass closed^			
Land cover #15: Proportion of alpine grass open^			
Land cover #16: Proportion of hummock grass open^			

Land cover #17: Proportion of sedges open^			
Land cover #18: Proportion of tussock grass open (logit)^		+?	
Land cover #19: Proportion of grassland scattered^			
Land cover #2: Proportion of bare areas^			
Land cover #20: Proportion of tussock grass scattered^			
Land cover #21: Proportion of grassland sparse^			
Land cover #22: Proportion of hummock grass sparse (logit)^			+
Land cover #23: Proportion of tussock grass sparse^			
Land cover #24: Proportion of shrubs closed^			
Land cover #25: Proportion of shrubs open^			
Land cover #26: Proportion of chenopod shrubs open^			
Land cover #27: Proportion of shrubs scattered^			
Land cover #28: Proportion of chenopod shrubs scattered^			
Land cover #29: Proportion of shrubs sparse^			
Land cover #3: Proportion of inland waterbodies^			
Land cover #30: Proportion of chenopod shrubs sparse^			
Land cover #31: Proportion of trees closed^			
Land cover #32: Proportion of trees open (logit)^	−		
Land cover #33: Proportion of trees scattered (logit)^			+
Land cover #34: Proportion of trees sparse^			
Land cover #4: Proportion of salt lakes^			
Land cover #5: Proportion of irrigated cropping^			
Land cover #6: Proportion of irrigated pasture^			
Land cover #7: Proportion of irrigated sugar^			
Land cover #8: Proportion of rainfed cropping^			
Land cover #9: Proportion of rainfed pasture (logit)^		+	
Land use #1: Proportion of nature conservation (logit)^	−		
Land use #11: Proportion of irrigated pastures^			
Land use #12: Proportion of irrigated cropping^			
Land use #13: Proportion of irrigated horticulture^			
Land use #14: Proportion of urban intensive use^			
Land use #15: Proportion of intensive plant and animal production^			
Land use #16: Proportion of rural residential farm infrastructure^			
Land use #17: Proportion of mining and waste^			
Land use #18: Proportion of water^			
Land use #4: Proportion of grazing native vegetation^			
Land use #5: Proportion of production forestry^			
Land use #6: Proportion of grazing modified pastures^			
Land use #7: Proportion of plantation forestry^			
Land use #8: Proportion of dryland cropping^			
Land use #9: Proportion of dryland horticulture^			
Climate: Air temperature mean diurnal range‡			
Climate: Air temperature mean isothermality‡			−
Climate: Maximum temperature of the warmest month‡	+		
Climate: Mean annual precipitation‡			

Climate: Mean annual precipitation seasonality‡			
Climate: Mean annual temperature‡			
Climate: Mean annual temperature seasonality‡			
Climate: Mean precipitation of the coldest quarter‡	–?		
Climate: Mean precipitation of the driest month‡			
Climate: Mean precipitation of the driest quarter‡			
Climate: Mean precipitation of the warmest quarter‡			
Climate: Mean precipitation of the wettest month‡			
Climate: Mean precipitation of the wettest quarter‡			
Climate: Mean temperature annual range‡	+		
Climate: Mean temperature of the coldest quarter‡			
Climate: Mean temperature of the driest quarter‡			
Climate: Mean temperature of the warmest quarter‡			
Climate: Mean temperature of the wettest quarter‡		+	+
Climate: Minimum temperature of the coldest month‡			
Climate: Prescott index‡			
Major vegetation groups (MVG) #1: Proportion of rainforests^			
MVG #10: Proportion of other forests and woodlands^			
MVG #11: Proportion of sparse eucalypt woodlands^			
MVG #12: Proportion of tropical eucalypt woodlands with annual grasses > 2 m^			
MVG #13: Proportion of sparse acacia woodlands^			
MVG #14: Proportion of mallee eucalypt woodlands and shrublands^			
MVG #15: Proportion of tall dense thickets^			
MVG #16: Proportion of acacia shrublands^			
MVG # 17: Proportion of other shrublands^			
MVG # 18: Proportion of heathlands^			
MVG # 19: Proportion of grasslands^			
MVG # 2: Proportion of 2 tall eucalypt forests > 30 m^			
MVG # 20: Proportion of arid spinifex grasslands^			
MVG # 21: Proportion of swampy grasses and sedges (logit)^			–
MVG # 22: Proportion of saltbushes and salt marshes^			
MVG # 23: Proportion of mangroves^			
MVG # 24: Proportion of water^			
MVG # 25: Proportion of cleared vegetation^			
MVG # 26: Proportion of unclassified native vegetation^			
MVG # 27: Proportion of naturally bare^			
MVG # 28: Proportion of sea^			
MVG # 29: Proportion of regrowth^			
MVG # 3: Proportion of eucalypt forests 10-30 m (logit)^	–		
MVG # 30: Proportion of unclassified forest^			
MVG # 31: Proportion of other sparse woodlands^			
MVG # 32: Proportion of sparse mallee eucalypt woodlands and shrublands^			
MVG # 4: Proportion of low eucalypt forests < 10 m^			
MVG # 5: Proportion of eucalypt woodlands (logit)^			–
MVG # 6: Proportion of acacia forests and woodlands (logit)^			–

MVG # 7: Proportion of cypress pine forests and woodlands^			
MVG # 8: Proportion of sheoak forests and woodlands^			
MVG # 9: Proportion of paperbark forests and woodlands^			
Soil effective cation exchange capacity (0-5 cm)‡			
Soil clay content, in the < 2 mm fraction (0-5 cm)‡			
Soil sand content, in the < 2 mm fraction (0-5 cm)‡			
Soil silt content, in the < 2 mm fraction (0-5 cm)‡			
Soil organic carbon content, in the < 2 mm fraction (0-5 cm)‡			
Soil pH (CaCl2) (0-5 cm)‡			
Soil organic carbon stocks in the top 30 cm‡			
Soil effective cation exchange capacity (0-5 cm) * Soil erodible fraction; calculated as a geometric mean‡	–?		
Soil effective cation exchange capacity (0-5 cm) * Soil erodible fraction * Vegetation fractional cover – mean BS; calculated as a geometric mean‡			
Soil effective cation exchange capacity (0-5 cm) * Vegetation fractional cover – mean BS; calculated as a geometric mean‡			
Soil erodible fraction * Vegetation fractional cover – mean BS; calculated as a geometric mean‡			
Soil erodible fraction‡			
Distance to coast‡	+		
Fire frequency ‡			
Vegetation Fraction of photosynthetically active radiation (FPAR) maximum (logit)‡		+	
Vegetation FPAR mean‡			
Vegetation FPAR median‡			
Vegetation FPAR minimum‡			
Vegetation FPAR standard deviation‡			
Vegetation fractional cover – maximum BS (logit)‡			+?
Vegetation fractional cover – maximum NPV‡			
Vegetation fractional cover – maximum PV‡			
Vegetation fractional cover – mean BS‡			
Vegetation fractional cover – mean NPV‡			
Vegetation fractional cover – mean PV‡			
Vegetation fractional cover – minimum BS‡			
Vegetation fractional cover – minimum NPV (logit)‡	+		
Vegetation fractional cover – minimum PV (logit)‡	–?		
Vegetation fractional cover – standard deviation BS‡			+
Vegetation fractional cover – standard deviation NPV‡			
Vegetation fractional cover – standard deviation PV‡			
Ecological land unit (ELU) #1: Proportion of artificial or urban area^			
ELU # 10: Proportion of cool semi dry plains^			
ELU # 11: Proportion of cool wet hills^			
ELU # 12: Proportion of cool wet mountains^			
ELU # 13: Proportion of cool wet plains^			
ELU # 14: Proportion of hot dry hills^			
ELU # 15: Proportion of hot dry mountains^			
ELU # 16: Proportion of hot dry plains^			
ELU # 17: Proportion of hot moist hills^			

ELU # 18: Proportion of hot moist mountains^			
ELU # 19: Proportion of hot moist plains^			
ELU # 21: Proportion of hot semi dry mountains^			
ELU # 22: Proportion of hot semi dry plains^			
ELU # 2: Proportion of cold wet hills^			
ELU # 20: Proportion of hot semi dry hills^			
ELU # 23: Proportion of hot wet hills^			
ELU # 24: Proportion of hot wet mountains^			
ELU # 25: Proportion of hot wet plains^			
ELU # 26: Proportion of snow and ice^			
ELU # 27: Proportion of undefined^			
ELU # 28: Proportion of warm dry hills^			
ELU # 29: Proportion of warm dry mountains^			
ELU # 3: Proportion of cold wet mountains^			
ELU # 30: Proportion of warm dry plains^			
ELU # 31: Proportion of warm moist hills^			
ELU # 32: Proportion of warm moist mountains^			
ELU # 33: Proportion of warm moist plains^			
ELU # 34: Proportion of warm semi dry hills^			
ELU # 36: Proportion of warm semi dry plains^			
ELU # 37: Proportion of warm wet hills^			
ELU # 38: Proportion of warm wet mountains^			
ELU # 39: Proportion of warm wet plains (logit)^	+		
ELU # 4: Proportion of cold wet plains^			
ELU # 40: Proportion of water body^			
ELU # 5: Proportion of cool moist hills^			
ELU # 6: Proportion of cool moist mountains^			
ELU # 7: Proportion of cool moist plains^			
ELU # 8: Proportion of cool semi dry hills^			
ELU # 9: Proportion of cool semi dry mountains^			
Air pollution - total industry PM10 emissions (mean 2011-13)			

Abbreviations: MVG = Major vegetation groups, FPAR = Fraction of Photosynthetically Active Radiation, PV = photosynthetic (living) vegetation, NPV = non-photosynthetic (non-living) vegetation, BS = bare soil, ELU = Ecological land unit.

Table S6. Significance testing of regression coefficients for the 'Moderate majority' cluster—based on *data-split one*.

Lasso regression coefficients and selective inference were evaluated using *data-split one*[#]. Then data for top-ranking predictors (ordered by absolute size of standardized regression coefficients) from *data-split two* were input to standard multiple linear regression software, with results shown†.

Predictor	Lasso[#]		Standard†	
	Coef	p-value	Coef	p-value
Proportion of eucalypt forests 10-30m (logit)	-0.0386	0.3363	-0.0270	0.0055 **
Percent obese persons	0.0366	0.2680	0.0098	0.3528
Socioeconomic index	-0.0325	0.2899	-0.0382	0.0008 ***
Distance to coast	0.0256	0.3944	0.0411	0.0076 **
Proportion of warm wet plains (logit)	0.0241	0.0149 *	0.0190	0.0455 *
Proportion of dryland cropping (logit)	-0.0236	0.9195	-0.0046	0.6692
Mean temperature annual range	0.0217	0.5958	0.0293	0.0583 .
Soil cation exchange capacity x erodible fraction – geometric mean	-0.0204	0.5875	-0.0101	0.2607
Diversity of major vegetation groups	-0.0191	0.6542	-0.0324	0.0033 **
Area of local government area (log10)	0.0187	0.3082	-0.0158	0.2515
Percent English-speaking immigrants (logit)	-0.0185	0.1033	0.0133	0.2415
Percent overweight persons (logit)	-0.0096	0.0486 *	-0.0256	0.0124 *
Percent smoking during pregnancy	0.0051	0.6161	0.0233	0.0218 *
Species richness (log10)	-0.0044	0.5686	-0.0376	0.0239 *

Significance codes: 0–0.001: '***', 0.001–0.01: '**', 0.01–0.05: '*', 0.05– 0.1: '.'

Table S7. Significance testing of regression coefficients for the 'Moderate majority' cluster—based on *data-split two*.

Lasso regression coefficients and selective inference were evaluated using *data-split two*[#]. Then data for top-ranking predictors (ordered by absolute size of standardized regression coefficients) from *data-split one* were input to standard multiple linear regression software, with results shown†.

Predictor	Lasso[#]		Standard†	
	Coef	p-value	Coef	p-value
Maximum temperature of the warmest month	0.0468	0.1105	0.0173	0.2579
Socioeconomic index	-0.0438	0.1748	-0.0511	5e-6 ***
Distance to coast	0.0381	0.1142	0.0355	0.0180 *
Diversity of major vegetation groups	-0.0318	0.0869 .	-0.0134	0.2499
Mean temperature annual range	-0.0295	0.0954 .	0.0229	0.2160
Proportion of warm wet hills (logit)	0.0239	0.6361	-0.0364	0.0099 **
Vegetation fractional cover minimum photosynthetic (logit)	-0.0231	0.5251	0.0205	0.2822
Proportion of open trees (logit)	-0.0202	0.5697	-0.0121	0.3738
Percent smoking during pregnancy	0.0201	0.0206 *	0.0171	0.1345
Percent overweight persons (logit)	-0.0197	0.5545	-0.0126	0.2142
Percent Aboriginal persons (logit)	0.0183	0.3977	-0.0030	0.7843
Proportion of warm wet plains (logit)	0.0179	0.2707	0.0454	0.0002 ***
Mean precipitation of the coldest quarter	-0.0139	0.0930 .	-0.0010	0.9295
Air temperature annual mean isothermality	-0.0135	0.2846	-0.0057	0.6077
Diversity of land use	-0.0077	0.1901	-0.0138	0.1933

Significance codes: 0–0.001: '***', 0.001–0.01: '**', 0.01–0.05: '*', 0.05– 0.1: '.'

Supplementary References

1. Geoscience Australia. 2014 Dynamic Land Cover Dataset V1.0, 27 May 2014.
2. Fryrear DW, Krammes CA, Williamson DL, Zobeck TM. 1994 Computing the wind erodible fraction of soils. *Journal of Soil and Water Conservation* **49**, 183-188.
3. Viscarra Rossel et al. 2014 Soil and Landscape Grid National Soil Attribute Maps (3" resolution) - Release 1. v1. (ed. CSIRO) Data collection. http://www.clw.csiro.au/aclep/soilandlandscapegrid/index.html
4. Wilford J, de Caritat P, Bui E. 2015 Modelling the abundance of soil calcium carbonate across Australia using geochemical survey data and environmental predictors. *Geoderma* **259–260**, 81-92. (doi:http://dx.doi.org/10.1016/j.geoderma.2015.05.003)
5. R Core Team. 2015 R: A language and environment for statistical computing. (Vienna, Austria, R Foundation for Statistical Computing.
6. Hijmans RJ. 2015 Raster: Geographic Data Analysis and Modeling. R package version 2.4-20.
7. ESRI. 2014 ArcGIS Desktop: Release 10.2. (Redlands, CA, Environmental Systems Research Institute).
8. Friedman JH, Popescu. 2008 Predictive learning via rule ensembles. The Annals of Applied Statistics 2, 916–954. (doi:10.1214/07-AOAS148)
9. Lin L. 1989 A Concordance Correlation Coefficient to Evaluate Reproducibility. *Biometrics* **45**, 255-268. (doi:10.2307/2532051)
10. Tibshirani R, Tibshirani R, Taylor J, Loftus J, Reid S. 2016 selectiveInference: Tools for post-selection inference. R package version 1.2.0.
11. Taylor J, Tibshirani RJ. 2015 Statistical learning and selective inference. *Proceedings of the National Academy of Sciences* **112**, 7629-7634. (doi:10.1073/pnas.1507583112)
12. PHIDU. 2015 Social Health Atlas of Australia: Data by Local Government Area, March 2015 release. (Adelaide, Public Health Information Development Unit, Torrens University Australia).
13. PHIDU. 2016 Social Health Atlas of Australia: Data by Local Government Area, May 2016 release. (Adelaide, Public Health Information Development Unit, Torrens University Australia).
14. Australian Consortium for Classification Development. 2015 International Statistical Classification of Diseases and Related Health Problems, Tenth Revision, Australian Modification. https://www.accd.net.au/Icd10.aspx
15. AIHW. 2017 AIHW National Hospital Morbidity Database: Separation statistics by principal diagnosis in ICD-10-AM, Australia, 2011−12 to 2012–13. (Canberra, Australian Institute of Health and Welfare). URL: http://www.aihw.gov.au/hospitals-data/principal-diagnosis-data-cubes/ [Accessed 27-08-2017]
16. ABS. 2011 Local Government Area ASGC Ed 2011 Digital Boundaries in ESRI Shapefile Format (Canberra, Australian Bureau of Statistics).
17. NPI. 2012 2011/12 data within Australia - Particulate Matter 10.0 um from All Sources. (National Pollutant Inventory, Australia). http://www.npi.gov.au/npidata/action/load/download-result/criteria/substance/70/destination/AIR/source-type/ALL/substance-name/Particulate%2BMatter%2B10.0%2Bum/subthreshold-data/Yes/year/2012

18. NPI. 2013 2012/2013 data within Australia - Particulate Matter 10.0 um from All Sources. (National Pollutant Inventory, Australia). http://www.npi.gov.au/npidata/action/load/download-result/criteria/substance/70/destination/AIR/source-type/ALL/substance-name/Particulate%2BMatter%2B10.0%2Bum/subthreshold-data/Yes/year/2013
19. Whitley R, Evans B, Pauwels J, Hutchinson M, Xu T, Han W. 2014 Bioclimatic surfaces: eMAST-R-Package 2.0, 0.01 degree, Australian Coverage Obtained from http://dap.nci.org.au, made available by the Ecosystem Modelling and Scaling Infrastructure (eMAST, http://www.emast.org.au) Facility of the Terrestrial Ecosystem Research Network (TERN, http://www.tern.org.au). (Macquarie University, Sydney, Australia).
20. Sayre et al. 2014 A New Map of Global Ecological Land Units — An Ecophysiographic Stratification Approach. (Association of American Geographers, Washington DC).
21. WA Landgate, North Australia Fire Information Service. 2012 1km 1997-2010 Fire frequency mapping for Australia. http://138.80.128.151/firehistory/
22. ABARES. 2014 Catchment scale land use of Australia March 2014. (Australian Bureau of Agricultural and Resource Economics and Sciences, Department of Agriculture and Water Resources). http://www.agriculture.gov.au/abares/aclump/land-use/data-download
23. Department of the Environment and Energy, A. 2012 Major Vegetation Groups - NVIS Version 4.1 (Albers 100m analysis product). http://www.environment.gov.au/land/native-vegetation/national-vegetation-information-system/data-products#mvg41
24. Gallant J, Austin J. 2012 Prescott Index derived from 1" SRTM DEM-S. v2. (ed. CSIRO). Data collection. (doi:10.4225/08/53EB2D0EAE377)
25. Viscarra Rossel R, Webster R, Bui E, Baldock J. 2014 Baseline map of Australian soil organic carbon stocks and their uncertainty. v2. (ed. CSIRO) Data collection (doi:10.4225/08/556BCD6A38737)
26. Atlas of Living Australia. 2016 Species Richness download. Data collection. http://spatial.ala.org.au/
27. Donohue R, McVicar T, Roderick M. 2013 Fraction of Photosynthetically Active Radiation (FPAR) - derived from Advanced Very High Resolution Radiometer (AVHRR), CSIRO Land and Water algorithm, Australia coverage. (ed. CSIRO). (doi:10.4225/08/50FE0CBE0DD06)
28. CSIRO Land and Water. 2012 Fractional cover - MODIS, CSIRO Land and Water algorithm, Australia coverage. Data collection. http://www.auscover.org.au/xwiki/bin/view/Product+pages/Fractional+Cover+MODIS+CLW+to+v2
29. Grundy et al. 2015 Soil and Landscape Grid of Australia. *Soil Research* **53**, 835-844. (doi:http://dx.doi.org/10.1071/SR15191)

Statement of Authorship

Title of Paper	Ambient soil cation exchange capacity inversely associates with infectious and parasitic disease risk in regional Australia
Publication Status	☑ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted w ork w ritten in manuscript style
Publication Details	Liddicoat, C., Bi, P., Waycott, M., Glover, J., Breed, M., Weinstein, P., 2018. Ambient soil cation exchange capacity inversely associates with infectious and parasitic disease risk in regional Australia. Science of The Total Environment. 626, 117-125

Principal Author

Name of Principal Author (Candidate)	Craig Liddicoat
Contribution to the Paper	I led the conceptualisation, data collection and preparation, performed the modelling, led the analysis and interpretation, wrote the first draft, and led the revisions of the final manuscript.
Overall percentage (%)	85
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature	Date: 6 May 2019

Co-Author Contributions

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

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

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

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

Name of Co-Author	Peng Bi
Contribution to the Paper	Contributed to the conceptualisation, analysis and interpretation of results, paper writing and revisions.
Signature	Date: 24-05-2019

Name of Co-Author	Michelle Waycott
Contribution to the Paper	Contributed to the conceptualisation, analysis and interpretation of results, paper writing and revisions.
Signature	Date: 13/6/19

Name of Co-Author	John Glover		
Contribution to the Paper	Contributed to the data collection and preparation, analysis and interpretation of results, paper writing and revisions.		
Signature		Date	23/5/19

Name of Co-Author	Martin Breed		
Contribution to the Paper	Contributed to the conceptualisation, analysis and interpretation of results, paper writing and revisions.		
Signature		Date	7/5/19

Name of Co-Author	Philip Weinstein		
Contribution to the Paper	Contributed to the conceptualisation, analysis and interpretation of results, paper writing and revisions.		
Signature		Date	21/5/19

Please cut and paste additional co-author panels here as required.

Liddicoat, C., Bi, P., Waycott, M., Glover, J., Breed, M., & Weinstein, P. (2018). Ambient soil cation exchange capacity inversely associates with infectious and parasitic disease risk in regional Australia. *Science of the Total Environment*, 626, 117-125. https://doi.org/10.1016/j.scitotenv.2018.01.077

Science of the Total Environment 626 (2018) 117–125

Contents lists available at ScienceDirect

Science of the Total Environment

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

Ambient soil cation exchange capacity inversely associates with infectious and parasitic disease risk in regional Australia

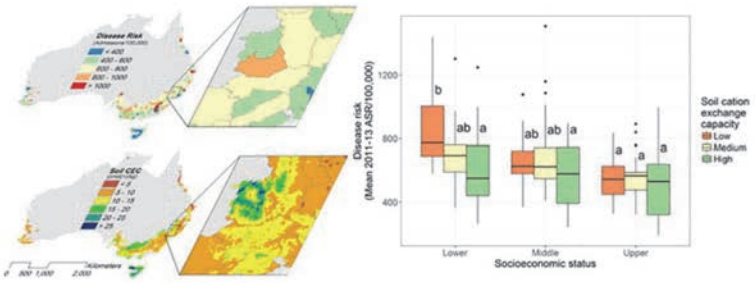
Craig Liddicoat [a,*], Peng Bi [b], Michelle Waycott [a,c], John Glover [d], Martin Breed [a], Philip Weinstein [a]

[a] School of Biological Sciences and The Environment Institute, The University of Adelaide, North Terrace, Adelaide, SA 5005, Australia
[b] School of Public Health, The University of Adelaide, North Terrace, Adelaide, SA 5005, Australia
[c] Department of Environment, Water and Natural Resources, GPO Box 1047, Adelaide, SA 5001, Australia
[d] Public Health Information Development Unit, Torrens University Australia, Level 1, 200 Victoria Square, Adelaide, SA 5000, Australia

HIGHLIGHTS

- Soil exposures may benefit human health but effects need to be explored.
- We use soil cation exchange capacity (CEC) as a proxy to test links to human health.
- We compared soils with infectious and parasitic disease risk in regional Australia.
- Effects of ambient soil quality are comparable to increasing socioeconomic status.
- Considering soil significantly improves disease risk prediction in unseen test areas.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 28 November 2017
Received in revised form 8 January 2018
Accepted 8 January 2018
Available online xxxx

Editor: D. Barcelo

Keywords:
Biodiversity hypothesis
Soil microbial diversity
Cation exchange capacity
Infectious and parasitic disease risk
Environmental epidemiology
Health inequality

ABSTRACT

Human contact with soil may be important for building and maintaining normal healthy immune defence mechanisms, however this idea remains untested at the population-level. In this continent-wide, cross-sectional study we examine the possible public health benefit of ambient exposures to soil of high cation exchange capacity (CEC), a surrogate for potential immunomodulatory soil microbial diversity. We compare distributions of normalized mean 2011/12–2012/13 age-standardized public hospital admission rates (cumulative incidence) for infectious and parasitic diseases across regional Australia (representing an average of 29,516 patients/year in 228 local government areas), within tertiles of socioeconomic status and soil exposure. To test the significance of soil CEC, we use probabilistic individual-level environmental exposure data (with or without soil), and group-level variables, in robust non-parametric multilevel modelling to predict disease rates in unseen groups. Our results show that in socioeconomically-deprived areas with high CEC soils, rates of infectious and parasitic disease are significantly lower than areas with low CEC soils. Also, health inequality (relative risk) due to socioeconomic status is significantly lower in areas with high CEC soils compared to low CEC soils (Δ relative risk = 0.47; 95% CI: 0.13, 0.82). Including soil exposure when modelling rates of infectious and parasitic disease significantly improves prediction performance, explaining an additional 7.5% ($\Delta r^2 = 0.075$; 95% CI: 0.05, 0.10) of variation in disease risk, in local government areas that were not used for model building. Our findings suggest that exposure to high CEC soils (typically high soil biodiversity) associates with reduced risk of infectious and parasitic diseases, particularly in lower socioeconomic areas.

* Corresponding author.
E-mail address: craig.liddicoat@adelaide.edu.au (C. Liddicoat).

https://doi.org/10.1016/j.scitotenv.2018.01.077
0048-9697/

1. Introduction

Increasing urbanisation, declining natural biodiversity, and consequent declines in population contact with environmental sources of microbial diversity, including soils, may compromise normal healthy immune system development and regulation (Rook, 2013; von Hertzen and Haahtela, 2006). Meanwhile, exposure to environmental microbial communities (microbiota) can help shape commensal microbiota in the gut and other sites, and consequently, also the development of immune status and predisposition to both infectious and non-infectious diseases (Ichinohe et al., 2011; Stein et al., 2016).

The notion of natural immunity, or enhanced protection from immune-related disease provided by exposure to environmental microbial diversity, is described by the Biodiversity Hypothesis (Haahtela et al., 2013; von Hertzen et al., 2011). Similarly, the related Old Friends Mechanism (Rook, 2013) suggests a protective immunomodulatory role for key microbial species that is lost or reduced with less exposure to natural and biodiverse environments. Employing beneficial microbiota-mediated immunomodulatory mechanisms from soil biodiversity may represent an underutilized resource for protecting and improving human health (Wall et al., 2015), and possibly a new cost-effective public health intervention (Mills et al., 2017). However, the notion that exposure to microbially-diverse soil may provide benefits to human health remains untested at the population-level.

Soils are a known reservoir of high biological diversity (Coleman et al., 2004) and catalogues of soils and their microbiota are growing due to the use of modern DNA sequencing technology (Bissett et al., 2016). However, to our knowledge, soil microbiota data are not yet available at the scale and coverage needed to compare with human health outcomes, suggesting the need for an intermediary, or proxy measure of soil microbiota with greater spatial coverage suited to epidemiological analysis. Liddicoat et al. (2016) justify the use of proxies as a pragmatic tool for investigating links between environmental microbial diversity and human health. Emerging research is demonstrating connections between soil microbiota, soil properties, and natural and anthropogenic influences including aboveground plant biomass and diversity (Delgado-Baquerizo et al., 2017; Gellie et al., 2017; Yan et al., 2018). Also, recent continent-wide soil mapping for Australia (Grundy et al., 2015) may offer possible candidate proxies, and below we highlight a key measure used in this study.

Soil microbes are variously involved in, and depend on, the development, turnover and stabilisation of soil organic matter (Kallenbach et al., 2016). Soil clay content supports water and nutrient retention, and together with plant interactions, moisture inputs, and organic matter, clay also supports soil aggregation and the resulting diversity of microbial habitats (Jastrow and Miller, 1998; Young and Crawford, 2004). As such, high soil clay content and increasing quantity and diversity of organic matter content generally contribute to greater soil microbial abundance and diversity (Torsvik and Øvreås, 2002). Soil organic matter and clay content also dominate measures of soil cation exchange capacity (CEC) (Peverill et al., 1999), since cations associate with the negatively charged sites in clay minerals and organic matter. CEC indicates the soil's storage capacity for nutrients and is estimated from the sum of exchangeable major cations (Ca^{2+}, Mg^{2+}, K^{+}, Na^{+}) expressed in centimoles of positive charge per kilogram of soil (cmol(+)/kg). Consequently, soil CEC should provide a useful proxy for soil microbial diversity, which we explore here using soil microbiome data relevant to our study area (refer to Methods and Fig. 2). Furthermore, Docherty et al. (2015) identified CEC as a key factor explaining variation in soil microbial community structure.

We can anticipate exposure pathways between ambient soils and people, through direct contact and diffuse airborne microbiota. In the latter case, soil microbiota form part of the so-called aerobiology (Polymenakou, 2012) that derives from the surrounding environment. In areas containing similar soils, and through diffuse aerobiology, it is plausible that soils may contribute to immunomodulatory effects (discussed later) in individuals and potentially in local populations, including possibly through naturally-acquired herd immunocompetence.

In epidemiology, health influences and outcomes are typically studied at the level of individuals or populations (groups), or in combination in multilevel scenarios (i.e. with individual and group-level influences and outcomes) (Susser, 1994). At this early research stage, it is impractical and cost-prohibitive to undertake a large-area study of soil exposures and infectious disease outcomes in individuals. Instead, we maximise value from existing nationwide health and social context reporting data through analysis of group-level (i.e. area-based or ecological) datasets.

We also perform more advanced analyses to explore multilevel models of disease risk and the significance of soil CEC by combining group-level predictor and outcome data with pseudo individual-level probabilistic sampling of environmental exposures. Such multilevel analyses are recognized in public health (Koopman and Longini Jr, 1994; Susser, 1994), and here we add to the rare examples in the literature of so-called micro-macro studies (Croon and van Veldhoven, 2007) by offering a robust non-parametric modelling approach. Our motivation for this work is to better understand possible connections between soils and human health, and explore the previously untested influence of natural ambient soil exposures on population health.

In this study we present a cross-sectional analysis of infectious and parasitic disease and ambient environmental exposures—including soil CEC, a proxy of environmental microbial diversity—for regional Australia in 2011–13, to explore possible beneficial soil-associated influences on immune-related human health. We specifically test the hypotheses that (a) soil CEC may provide an indicative proxy of soil microbial diversity; (b) cumulative incidence of infectious and parasitic disease (hereafter termed disease risk) and the relative risk of disease or health inequality (defined below) differ with ambient soil CEC, and; (c) soil CEC is a significant predictor of infectious and parasitic disease risk.

2. Methods

2.1. Soil CEC and microbial diversity

We extracted representative soil sample data, relevant to our study, from the Biomes of Australian Soil Environments (BASE; Bissett et al., 2016; https://data.bioplatforms.com/organization/bpa-base) by matching available: (a) sample locations within our study area (defined below), plus a 20 km buffer zone; (b) data for exchangeable major cations (Ca^{2+}, Mg^{2+}, Na^{+}, K^{+}), which were summed to estimate soil CEC; and (c) community composition (operational taxonomic unit, OTU) data for bacteria, fungi and eukaryotes identified from soil environmental DNA (eDNA). Respective sample locations used are shown in Fig. S1 B–D. Bissett et al. (2016) and Delgado-Baquerizo et al. (2017) describe the BASE methods used for sample collection; soil chemical analyses; eDNA extraction; sequencing of the bacterial 16S rRNA gene, eukaryotic 18S rRNA gene, and fungal internal transcribed spacer (ITS) region; and the bioinformatic analyses to derive OTU abundance tables.

We filtered out samples with extreme, outlying soil conditions due to their potential to exert undue influence on our assessment of soil CEC-microbial diversity relationships. Samples were filtered out that were strongly acid ($pH_{H2O} < 4.5$), strongly alkaline ($pH_{H2O} > 9$), highly saline (electrical conductivity >8 dS/m), very high (>50%) clay content, or deep (900 cm soil depth). These samples were considered unrepresentative of population-soil exposures in the study area as they are uncommon in rural and regional settings (Grundy et al., 2015; McKenzie et al., 2004; Peverill et al., 1999), and represent known or potential limitations to biological activity, microbial food webs, or soil structure, and therefore may excessively impact the activity and diversity of certain microbial taxa (Barrett et al., 2004; Fierer and Jackson, 2006; Zahran, 1997). Samples that lacked major cation data required to estimate CEC were also excluded.

We matched the remaining BASE soil sample data (with calculated CEC values) to the respective OTU tables for bacteria, fungi and eukaryotes. We then performed subsequent analyses using the R *phyloseq* package (McMurdie and Holmes, 2013), employing methods similar to previous analyses of BASE data (Gellie et al., 2017; Yan et al., 2018). Within each taxonomic group we respectively removed rare OTUs (i.e. OTUs with <100 reads across all samples) and normalized the data to an even sampling depth by rarefying to the lowest number of reads found across all samples. In summary, bacterial 16S amplicons (from n = 353 soil samples) were rarefied to 5477 reads per sample, fungal ITS amplicons (from n = 338 soil samples) were rarefied to 2316 reads per sample, and eukaryote 18S amplicons (from n = 345 soil samples) were rarefied to 5323 reads per sample. Shannon index values were calculated and then exponentially transformed to derive the effective number of species (Jost, 2006) to estimate microbial diversity in each sample. Resulting plots of soil eDNA-based microbial diversity versus log10-transformed CEC are shown in Fig. 2. We also calculated 95% confidence intervals for slope coefficients of the respective linear regressions of effective number of species on log10(soil CEC), using bootstrap (B = 10,000) estimation of standard errors. Analyses are documented in R script in Appendix A.

2.2. Health and environmental data

We use aggregated hospitalisation data from the Social Health Atlas of Australia (PHIDU, 2015; PHIDU, 2016). For reliability and privacy reasons these data have been grouped: (a) spatially by Australian local government areas (LGAs; Fig. 1); and (b) thematically under principal diagnoses of infectious and parasitic diseases (A00-B99, ICD-10-AM 7th edition; see Table S1, Appendix A). Specifically, in this study we analyse the normalized mean 2011–13 (i.e. merged 2011/12 and 2012/13) cumulative incidence data for infectious and parasitic disease public hospital admissions, calculated in each LGA using:

$$ASR_{2011-13} = (N_{11/12} + N_{12/13}) / \left(\frac{N_{11/12}}{ASR_{11/12}} + \frac{N_{12/13}}{ASR_{12/13}}\right) \quad (1)$$

where ASR_i is the respective age-standardized rate of hospital admissions per 100,000 population, and N_i is the respective raw number of annual admissions recorded. Subscripts 11/12 and 12/13 refer to financial years 2011/12 and 2012/13 respectively. Public hospital only data (excluding private hospitals) were used on the basis that they should more closely reflect local environmental influences.

Environmental mapping datasets used in the study are described in Table S2, Appendix A.

2.3. Spatial framework

LGA-based data were used to offer a balanced coverage of health reporting areas and a spatial framework suited to capturing soil and other environmental variability across regional Australia. (Alternative 'Population Health Area' data in the Social Health Atlas provide higher spatial resolution in major cities and urban areas, at the expense of poor regional resolution.) We focus on regional Australia, as the soil mapping (Grundy et al., 2015) has been largely based on land resource assessment data from agricultural lands, and will be most reliable in these areas. Any soil effect is also more likely in regional areas because contact with soil is greater here than in cities.

Regional LGAs used in the study (Fig. 1) were defined by dominant remoteness class (ABS, 2011), but excluded very large LGAs (>10,000 km^2), as we would expect increasing variation in environmental and social influences that would not be adequately represented by available group-level data in larger LGAs.

We adopted a common 250 m resolution grid system for environmental mapping variables to improve spatial consistency and for pragmatic data handling reasons. Where necessary (to match the common grid system), resampling of spatial rasters was performed bilinearly for numeric data layers, and using the nearest neighbour method for the categorical land use layer.

2.4. Comparing risks and relative risks of disease: group-level analyses

The mean Soil CEC in each LGA was calculated using zonal statistics in ArcMap 10.3.1 geographic information system (ESRI, 2015). Mean 2011–13 infectious and parasitic disease risk data were compared after splitting LGAs into tertiles of Socioeconomic index and Soil CEC (see Table 1 for summary descriptive statistics). We tested for differences in the distribution of tertile groups using the Kruskal-Wallis rank sum test in base R (R-Core-Team, 2017). Where significant, we used the Dunn test for multiple comparisons to determine which levels of the predictor variable differed from each other level, with Bonferroni adjusted threshold *P* values (Zar, 2010). The multiple comparison testing used default two-sided *P* values and alpha = 0.05 nominal level of significance.

To examine possible spatial clustering, we performed hot spot analyses in ArcMap 10.3.1 using the Getis-Ord Gi* statistic (ESRI, 2015; Ord and Getis, 1995) with input data as the mean 2011–13 disease risk mapping, projected in the GDA 1994 Geoscience Australia Lambert coordinate reference system. We undertook separate hot spot analyses using both inverse distance weighting and fixed distance band conceptualization of spatial relationships, with and without applying the optional false discovery rate correction. Using default settings, the ArcMap hot spot analyses used a neighbourhood search threshold of approximately 217 km.

Disease risk ratios (relative risks) were estimated using bootstrap methods to examine potential health inequalities. Relative risk was calculated by dividing the disease risk in the most deprived socioeconomic areas by the disease risk in the least deprived areas, for both low Soil CEC and high Soil CEC (where soil CEC categories of low/medium/high were assigned to the respective tertiles). In each of B = 10,000 bootstrap samples, we randomly sampled with replacement up to the mean sample size of the respective LGA groups. We estimated standard errors to calculate 95% confidence intervals (CI) for relative risks from the standard deviation of the bootstrap sample statistics. Similarly, bootstrap methods (and an appropriate sample statistic) were used to estimate a 95% CI for the difference in relative risk between low Soil CEC and high Soil CEC areas.

2.5. Testing the effect of soil on disease risk: multilevel modelling

To facilitate sampling of pseudo individual-level environmental exposures, candidate environmental predictors (listed below) were re-expressed using focal neighbourhood statistic calculations. This meant environmental data at any cell location provided an exposure estimate corresponding to a surrounding zone of influence. Such a zone reflects potential movement of populations within their surroundings and also the possibility of airborne dispersal of environmental microbiota and bioactive agents (i.e. bioaerosols). We do not know how far populations or bioaerosols disperse, however we chose a nominal 3 km radius area as our representative environmental zone of influence, which is consistent with previous studies examining possible links between land use and human immunomodulatory influence (Hanski et al., 2012; Ruokolainen et al., 2015). That is, for all gridded environmental layers used, in every grid cell, we calculated a summary measure for a 3 km-radius area.

Environmental conditions were calculated within the 3 km-radius area for Mean rainfall, Mean temperature, Diurnal temperature range, and Soil CEC. Temperature and rainfall are considered key environmental determinants of infectious disease (Eisenberg et al., 2007). We considered Humidity and Latitude during preliminary data exploration, but did not include them in the analysis due to high correlation with Diurnal temperature range and Mean temperature respectively. From

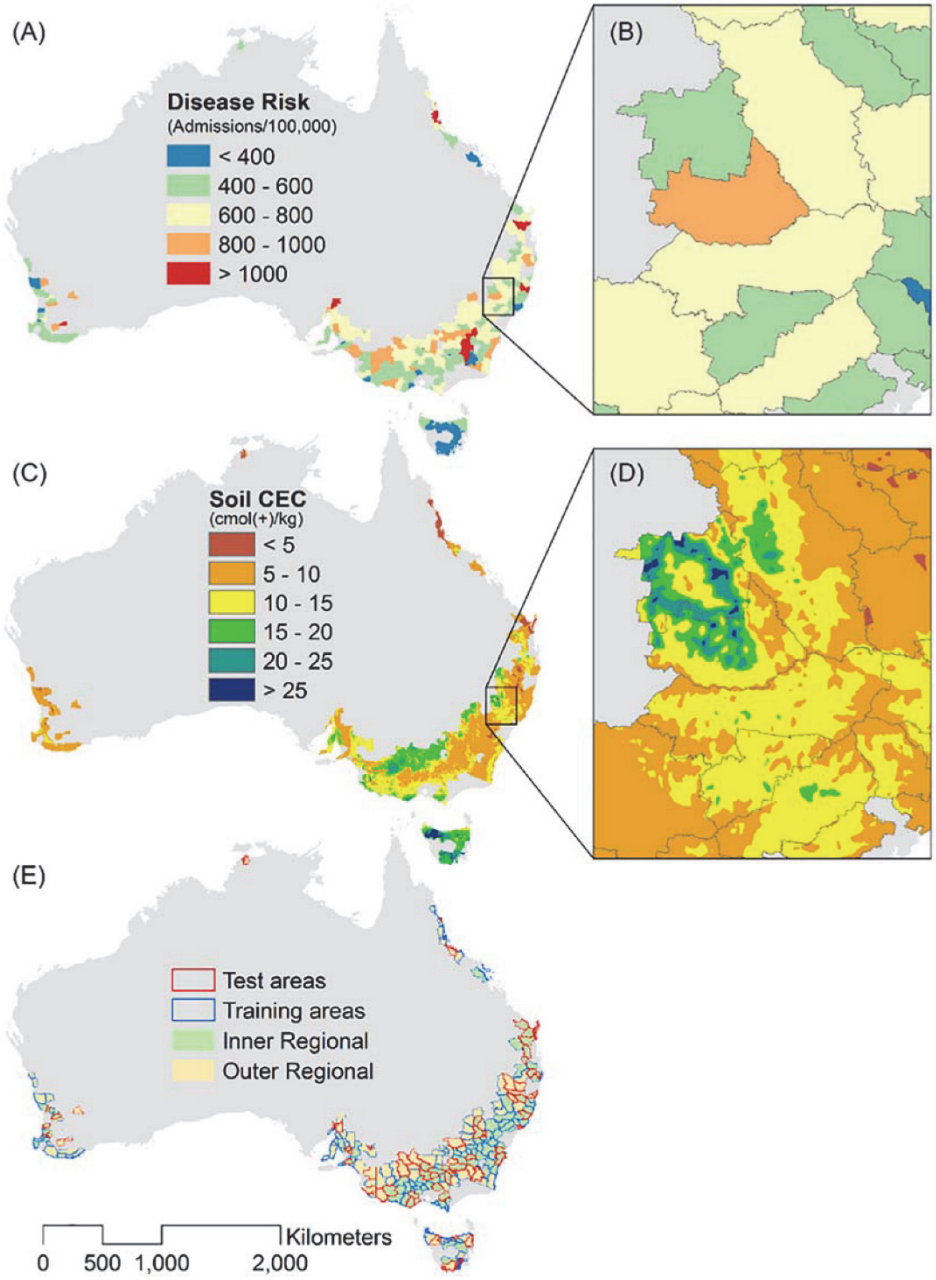


Fig. 1. Map of regional Australian local government areas (LGAs) used in the study. These LGAs have a range of areas (9–9686 km^2; median = 3180 km^2; interquartile range = 1506–4701 km^2; n = 228). Our analysis excluded LGAs in major cities, remote and very remote areas, and those with missing data or areas >10,000 km^2. Infectious and parasitic disease risk (mean 2011–13 age-standardized rate of public hospital admissions per 100,000 persons) is shown in (A) and inset area (B). Focal 3 km radius-area mean Soil CEC (cmol(+)/kg) is shown in (C), and inset area (D). LGAs were divided into training and test groups for multilevel predictive modelling (E).

rasterized land use mapping, we calculated the Proportion of conservation land (i.e. protected areas) and Proportion of natural land (i.e. protected areas plus productive natural areas) to represent possible influence from surrounding biodiverse green space. From rasterized land cover mapping, we calculated the Proportion of inland water bodies and Proportion of salt lakes to represent possible exposure to water-borne disease and associated vectors (e.g. mosquitoes; O'Sullivan et al., 2008). Social variables included Socioeconomic index (or status) and Population density, based on LGA (group-level) data. We log10-transformed Population density to reduce skewness and improve correlation with the health outcome data. Further information on social and environmental predictor data used is provided in Table S2, Appendix A.

Rasterized land use mapping was also used to identify potential areas in each LGA where people might live or work. Raster cells with land uses of conservation, protected, minimal use, or water, were disqualified. Then 20 cells per LGA were randomly sampled to extract

Table 1
Summary statistics for regional Australian local government areas (LGAs) used in the study, split into tertile groups by Socioeconomic index and Soil CEC.

Socioeconomic index		Mean soil CEC (cmol(+)/kg)		Mean 2011–13 disease risk (public hospital admission rates; ASR/100,000)	No of LGAs
	Median (IQR)		Median (IQR)	Median (IQR)	(n)
Lower	931 (917, 942)	Low	8.6 (7.1, 9.6)	774 (690, 1004)	20
	930 (906, 942)	Medium	11.6 (10.9, 12.5)	695 (592, 761)	22
	928 (917, 940)	High	18.2 (16.4, 23.5)	552 (442, 756)	34
Middle	975 (958, 981)	Low	9.2 (8.1, 9.5)	628 (581, 723)	25
	968 (955, 973)	Medium	12.3 (11.1, 12.8)	625 (549, 744)	31
	963 (955, 968)	High	19.9 (17.6, 22.0)	580 (394, 746)	20
Upper	1003 (991, 1025)	Low	8.2 (7.5, 9.0)	546 (450, 627)	31
	998 (992, 1013)	Medium	11.8 (10.9, 12.7)	567 (477, 588)	23
	1000 (991, 1037)	High	18.3 (16.6, 20.6)	532 (322, 639)	22

CEC = cation exchange capacity; IQR = interquartile range; ASR = age standardized rate.

pseudo individual-level environmental exposure data (n = 4544 after removal of 16 cases with missing data due to minor gaps in environmental mapping). Group-level social variables of Socioeconomic index and log10(Population density) were assigned to pseudo-individuals based on their respective LGA. We tested for, and did *not* find high levels of correlation within any pairs in the final set of candidate predictor data (Fig. S2, Appendix A; all correlations were <0.75).

Generalized boosting models (GBM) and random forest (RF) models were trained to predict disease risk in unseen LGAs. To do this, we created a stratified random 70:30 training:test data split based on LGA codes. We stratified data by remoteness class to distribute inner regional and outer regional LGAs evenly across the training (n = 3192) and test (n = 1352) data sets. Since data were split at the group-level using LGAs, all pseudo-individual data in the test set corresponded to LGAs that were not seen during training of the prediction models. We used the R *caret* package (Kuhn et al., 2017) for data-splitting and subsequent cross-validation optimisation (model tuning) of GBM and RF models. Training data were further divided into 10 folds using the *groupKfold* function in *caret* to ensure that LGA groups could be iteratively withheld during the cross-validation model optimisation process. 10-fold cross-validation involves using 9 folds (or parts) of the data to estimate model parameters which were then validated on the 10th fold (in our case, containing pseudo-individuals from LGAs not modelled in the earlier 9 folds). This process continues until each fold takes a turn as the validation set. We used 10-fold cross-validation repeated 5 times, with centred and scaled data, and an automated suite of tuneable model parameters (i.e. tune length = 10) to determine optimum model parameters for GBM and RF that maximise the coefficient of determination (r^2) between predictions and observations. Finally, prediction performance on withheld test data was examined separately for both GBM and RF, with and without soil. We used bootstrap probability assessments for measures of r^2, root mean square error (RMSE), and 95% confidence intervals (CIs) to test if the addition of soil CEC significantly improved model performance. Statistical analyses were performed using R software (R-Core-Team, 2017). R scripts and additional packages used are detailed in Appendix A (online supplement).

3. Results

3.1. Soil CEC: an indicative proxy for microbial diversity

We found significant positive correlations between soil CEC and the diversity of OTUs for three important taxonomic groups: bacteria, fungi and eukaryotes (Fig. 2).

3.2. Risk and relative risk of disease

Infectious and parasitic disease risk differed significantly across LGAs when split into tertiles of Socioeconomic index ($\chi^2 = 19.02$; $df = 2$; $P < .001$), Soil CEC ($\chi^2 = 7.39$; $df = 2$; $P = .025$), and the combined tertile groups of Socioeconomic index and Soil CEC ($\chi^2 = 38.42$; $df = 8$; $P < .001$; Fig. 3). Infectious and parasitic disease risk generally decreased with decreasing deprivation, and with increasing Soil CEC. This relationship was most prominent in the most deprived areas, with significantly lower disease risk in areas with high compared to low CEC soils. Health inequality (relative risk) associated with socioeconomic status was significantly lower in areas with high CEC soils (1.13; 95% CI: 0.88, 1.37) compared to low CEC soils (1.60; 95% CI: 1.36, 1.84; Fig. 4). The mean difference in relative risk was estimated to be 0.47 (95% CI: 0.13, 0.82). Our results indicate populations living in areas of high CEC soils have a similar risk of infectious and parasitic disease regardless of wealth. By contrast, those living in areas that are socioeconomically deprived and have low CEC soils experience health inequality of 1.6 times the risk of infectious and parasitic disease (compared to the least deprived areas).

3.3. Hot spot analyses

Results from the hot spot (cluster) analyses (Fig. S3) were inconclusive. The output map from the fixed distance band analysis with false discovery rate correction is suggestive of disease risk hot spots in southern New South Wales and northern Queensland, and cold spots in Tasmania. However, these results should be viewed with caution as they were markedly different from the lack of significant clustering (hot spots or cold spots) found using the inverse distance weighting approach (Fig. S3 A–B), and appear to overemphasise disease risk in southern New South Wales compared to the actual mean 2011–13 disease risk mapping.

3.4. Multilevel modelling of disease risk

Including Soil CEC when modelling infectious and parasitic disease risk in population groups significantly improves prediction performance in unseen areas (i.e. withheld test data), compared to not including Soil CEC (Table 2). GBM models with Soil CEC have significantly lower RMSE (mean difference = −10.0; 95% CI: −12.9, −7.1) and significantly higher r^2 (mean difference = 0.075; 95% CI 0.047, 0.102). This means that including Soil CEC in GBM models helps explain 7.5% more variation in population disease risk in previously unvisited (with-held test) areas. Similarly, RF models with Soil CEC have significantly lower RMSE (mean difference = −4.3; 95% CI: -5.7, −2.8) and significantly higher r^2 (mean difference = 0.032; 95% CI: 0.021, 0.044). Measures of variable importance (Fig. S4, Appendix A) suggest soil CEC ranks among other known social and environmental predictors of infectious disease risk. Partial dependence plots for GBM and RF models (Fig. S5–S6 Appendix A) show that as Soil CEC increases, the marginal effect is for almost monotonically decreasing disease risk, analogous to the beneficial effect expected from increasing socioeconomic status. Plots of predicted versus observed group-level disease risk in withheld test

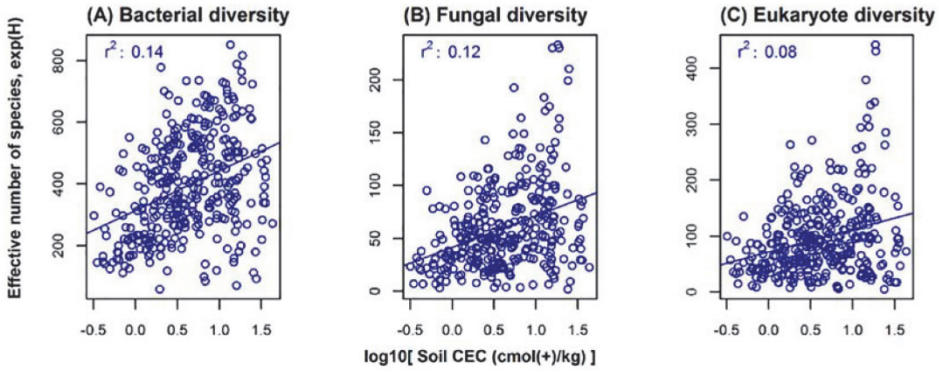


Fig. 2. Significant positive correlations between log10-transformed soil CEC and (A) bacterial diversity ($r^2 = 0.14$; $P < .001$; $\beta = 128$, 95% CI: 93, 163; $n = 353$), (B) fungal diversity ($r^2 = 0.12$; $P < .001$; $\beta = 30$, 95% CI: 21, 40; $n = 338$), and (C) eukaryote diversity ($r^2 = 0.08$; $P < .001$; $\beta = 40$, 95% CI: 24, 57; $n = 345$). Slope coefficient (β) units represent the additional effective number of species per unit rise of log10(soil CEC). Plots are based on our analysis of OTU tables and exchangeable cation data extracted from the Biomes of Australian Soil Environments (Bissett et al., 2016). Effective number of species were calculated from the exponential transform of the Shannon diversity index (H) (Jost, 2006).

areas for GBM and RF models, with and without Soil CEC, are shown in Fig. S7, Appendix A.

4. Discussion

We found consistent and significant trends of generally increasing biological complexity with increasing soil CEC for three important taxonomic groups of microorganisms found in soils (Fig. 2), which supports our idea to use soil CEC as an indicative proxy for soil microbial diversity. Bacteria, fungi and eukaryotes (including protozoa, helminths and other organisms) represent key groups under active research for possible microbial Old Friends (Rook, 2013). We show that soil CEC may provide a potentially important influence on infectious and parasitic disease that warrants further investigation. We acknowledge that soil CEC represents a simplistic proxy for microbial diversity, and expect that more sophisticated soil quality measures may be examined in connection to beneficial influence on human health in the future.

There are plausible mechanisms for microbial diversity associated with higher CEC surface soils to provide beneficial ambient immunomodulatory exposures to local populations. Human commensal microbiota play a role in normal (healthy) immune function (Molloy et al., 2012) and are in dynamic exchange with environmental microbiota (Hanski et al., 2012; von Hertzen et al., 2011). Bioaerosols can be deposited in the upper airways, transported by cilia or coughing, and swallowed. Therefore, airborne microorganisms can deposit on the skin, in the airways, and in the gut where they can perform immunomodulatory roles (Rook, 2013). Mechanisms whereby environmental microbiota may benefit immune status and human health include competitive exclusion of pathogenic microbes (where environmental microbiota contribute to commensal microbiota) and through molecular patterns (sampled by mucosal barrier immune cells) which contribute to immune-signalling pathways, in-turn regulating both pathogens and inflammatory immune responses (Molloy et al., 2012; von Hertzen et al., 2011).

The human intestinal microbiota may be dominated by spore-forming bacteria (Browne et al., 2016) designed to persist in the environment, which aligns well with the idea that the environment may resupply key spore-forming and other (including possibly immunomodulatory) bacteria to the human commensal microbiota. Differences in soil quality, including moisture- and nutrient-retention, will influence survival and persistence of microbes, including pathogenic species. As pathogens and parasites of humans represent a minority of the microbial species living in soils (Wall et al., 2015), it is plausible that there may be a host of harmless microorganisms and/or potential beneficial microbial Old Friends that could be supplied via soils.

Due to many unknowns in the topic of possible beneficial environmental immunomodulation (e.g. specific beneficial microbes or Old

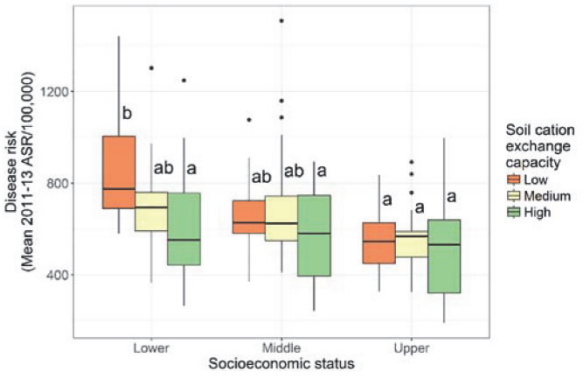


Fig. 3. Boxplot distributions of mean 2011–13 infectious and parasitic disease risk in regional Australian LGAs ($n = 228$) stratified by tertiles of Socioeconomic index (i.e. status) and Soil CEC. Y-axis units are age-standardized rate of public hospital admissions per 100,000. Groups that do not share a letter are significantly different.

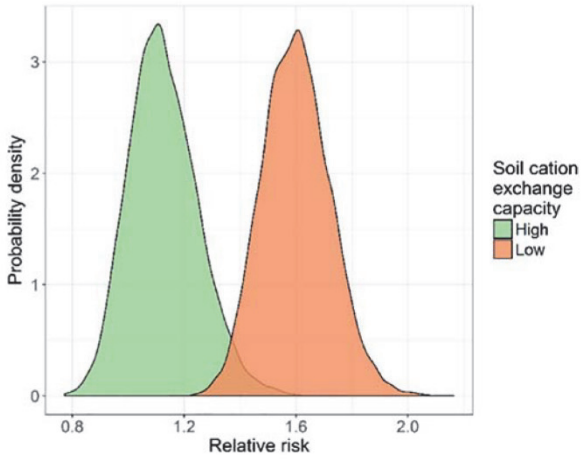


Fig. 4. Bootstrap sample ($B = 10{,}000$) density plots for health inequality or relative risk of infectious and parasitic disease for areas of high versus low Soil CEC. Each distribution represents a ratio of disease risk in low socioeconomic LGAs to disease risk in high socioeconomic LGAs.

Table 2
Prediction performance on withheld test data (n = 1352), for multilevel machine-learning models of infectious and parasitic disease risk (with and without soil CEC): regional Australia, 2011–2013. Confidence intervals were derived using bootstrap (B = 10,000) estimation of standard errors for each respective statistic.

Model	Included Soil CEC?	RMSE Mean (95% CI)	r^2 Mean (95% CI)
GBM	With CEC	190.3 (95% CI: 179, 202)	0.468 (95% CI: 0.423, 0.512)
	Without CEC	200.3 (95% CI: 189, 212)	0.393 (95% CI: 0.339, 0.446)
	Difference (with - without)	ΔRMSE: −10.0 (95% CI: −12.9, −7.1)	Δr^2: 0.075 (95% CI: 0.047, 0.102)
RF	With CEC	204.1 (95% CI: 192, 217)	0.373 (95% CI: 0.337, 0.408)
	Without CEC	208.3 (95% CI: 196, 221)	0.340 (95% CI: 0.306, 0.374)
	Difference (with - without)	ΔRMSE: −4.3 (95% CI: −5.7, −2.8)	Δr^2: 0.032 (95% CI: 0.021, 0.044)

GBM = generalized boosting model; RF = random forest.

Friends and antigenic materials, behaviourally- and temporally-mediated exposures, possible sub-clinical requisite exposures, possible integrated influence over time, and potential for multiple and redundant immunomodulatory pathways), our use of soil CEC as a proxy for environmental microbial diversity is justified as a pragmatic investigation tool (Liddicoat et al., 2016). For example, environmental proxies may have potential to reveal health associations that might develop over long periods of exposure, that may not be evident from analysis of one-off measurements of temporally-fluctuating (e.g. seasonally-variable) soil biology.

The growing number of studies that link potentially beneficial environmental microbiota exposures and human health (e.g., Hanski et al., 2012; Ichinohe et al., 2011; Ruokolainen et al., 2015; Stefka et al., 2014; Stein et al., 2016) give limited attention to infectious disease. However, we can conceptualise such links. A lack of formative diverse environmental exposures (Wopereis et al., 2014) may result in maladapted composition and diversity of niche-occupying commensal microbiota, required for competitive exclusion of potentially harmful microbial invaders. Also, adverse feedbacks in a poorly trained, dysfunctional immune system may reinforce imbalance (or dysbiosis) of host microbiota (Haahtela et al., 2013), in turn possibly favouring pathogenic microbes and increasing susceptibility to infectious disease. Perhaps more intuitively, from vaccine development we know that exposure to microbes and associated antigenic material (in certain forms and concentrations) has potential to induce protective immune responses against infectious disease. However, the potential for natural and biodiverse environments, and their microbiomes, to impart this type of protective immune training is not well understood.

We note that our study is not the first to suggest that there may be as-yet-unexplained beneficial linkage mechanisms between natural exposures and human health. For example, Mitchell and Popham (2008) found health inequalities related to income deprivation in England were the lowest in populations living in areas with the highest green space exposure.

Alternative explanations may also exist and are possibly complementary to the ideas above. We might expect low CEC soils to support less soil biodiversity and have reduced ecological controls on microbiota composition. If poorly managed by landholders, a lack of microbial ecological 'buffering' in low CEC soils might render them more susceptible to build-up of opportunistic pathogenic microbes. Sandy, low-organic matter (i.e. low CEC) soils lack structure and are more prone to wind erosion and adverse impacts to human populations. Populations may also experience more frequent exposures to wind-blown soil-borne contaminants (such as pesticide residues) from these lower quality soils. We note that soil-borne contaminant and pathogen exposures will be familiar in environmental epidemiology. However the possibility of lost immunomodulatory benefit (in the case where higher quality soils were maintained) may represent an unappreciated health impact.

Differing soil exposures and inadvertent ingestion of biologically-essential trace elements (Kabata-Pendias and Mukherjee, 2007), and other bioactive agents in wind-blown dust may also influence nutrition, immune status and resistance to disease. Any of the possible beneficial soil exposures discussed here may have sub-clinical effects (e.g. inward effects on host microbiota, immune signalling and regulation; Rook, 2013; von Hertzen et al., 2011) and, if so, are likely to go unnoticed. Therefore, future detailed studies of health effects from soil-associated bioaerosols should ideally include immune biomarkers.

There are important limitations in our study. We note this is a cross-sectional (correlational) study and more rigorous (e.g. longitudinal or experimental) studies are needed to build evidence of possible microbiota-mediated causal links. Caution is needed in interpreting this work, especially involving immunocompromised individuals or situations where soil-borne human pathogens may accumulate (e.g. in contaminated or poorly managed land) as soil exposures can also lead to infectious disease (Baumgardner, 2012; Kolivras et al., 2001). Indeed, individual exposure and dose-response relationships involving soil CEC and associated microbiota composition and diversity remain to be investigated. We note that relationships between soil physicochemical properties (such as CEC) and microbiological attributes are still under active research. Soil CEC will not always indicate soil microbial diversity, for example when biological growth is impeded in poorly structured soils, or due to salinity or pH constraints. Mapping of soil CEC will include errors and uncertainty (Grundy et al., 2015). To reduce complexity, we have not incorporated uncertainty in soil CEC and other environmental mapping in the scope of this analysis. We do not attempt to investigate or interpret potential complex and multilevel interactions between environmental determinants and socio-ecological disease transmission dynamics (Eisenberg et al., 2007). Using aggregated health response data represents a limitation in terms of loss of specificity to link environmental influence with any particular disease. However, this approach may offer greater sensitivity to detect possible broad environmental influence on multiple infectious and parasitic disease outcomes (e.g. possibly via underlying immunocompetence). Spatially aggregated data are subject to well-known potential errors of ecological bias and fallacy (Susser, 1994), and the modifiable unit area problem (i.e. where arbitrary spatial boundaries may bias analyses).

On the other hand, we aimed to moderate these shortfalls through probabilistic pseudo-sampling of individual-level environmental exposures. Our use of group-level (ecological) and pseudo-individual data with GBM and RF models is consistent with previous suggestions that superior disease prediction frameworks may require multilevel non-linear modelling (Koopman and Longini Jr, 1994). In probabilistic terms, we demonstrate that soil exposures significantly contribute to population health outcomes for infectious and parasitic disease risk.

Population movement will not be a substantial source of error in this study as we expect migration rates to be low in regional Australia. For example, the Social Health Atlas of Australia (PHIDU, 2015) contains estimates for the percent of populations born overseas in non-English speaking countries and resident in Australia less than five years. For the LGAs used, these data show a median of 0.5% (interquartile range 0.3–0.9%; range 0–6%), indicating a very non-mobile population.

Our findings do not support perverse recommendations for exposing large areas of soil with the goal of achieving public health benefits. Such an approach would increase the risk of soil degradation through erosion of top soil and decomposition of organic matter (both of which would likely cause reductions in surface soil CEC). Possibly, occasional inadvertent exposure may be all that is required. Hands-on activities such as gardening, or children playing in the right kind of soil (e.g.

moderate to high CEC, with balanced and diverse microbial composition) may be more beneficial than passive exposures—however these ideas remain to be tested. In agricultural landscapes, maintaining and enhancing soil cover (to prevent erosion) and re-establishing biodiversity and perenniality of plants where possible (which also enhances soil organic matter and promotes soil microbial diversity and function; Wall et al., 2015; Zak et al., 2003), are cornerstone principles of sustainable land management. In urban areas, ecological restoration may help drive the rewilding of environmental microbiomes (Mills et al., 2017). These activities to enhance aboveground environments, soil organic matter and other aspects of soil health may simultaneously protect and enhance aboveground biodiversity, soil CEC, microbial diversity, and, importantly, human health.

5. Public health implications

Measures to conserve and enhance soil quality and biodiversity in natural, agricultural, and urban environments may offer ambient cost-effective benefits in reducing population risk of infectious and parasitic diseases. Health intervention programs and activities that increase population exposure to biodiverse green space and associated microbially-diverse living soils, should be jointly considered by health and environmental management agencies, local councils and the wider community. Our results are consistent with growing evidence for the health benefits of exposure to high biodiversity environments (WHO and SCBD, 2015), and support the need for experimental studies to better quantify the relationship, and confirm the causal pathways underlying it. Only with such an evidence base can innovative new policy be formed to protect the environment and human health concurrently.

6. Conclusions

Ambient soil exposures may provide an important environmental influence on immune-related human health. Soils are a recognized source of microbial diversity, and our results show that soil CEC provides an indicative proxy for soil biodiversity in regional Australia. With the context of possible varying immunomodulatory influence, we examine associations between population health outcomes and ambient soil CEC (among other environmental and social variables) at the continental-scale. Our results are suggestive of a beneficial health influence from high CEC soils, with reduced infectious and parasitic disease risk, particularly in lower socioeconomic areas. We also demonstrate that knowledge of the spatial distribution of soil CEC can improve the prediction of infectious and parasitic disease risk in unseen test areas.

Human participant protection

This research was exempt from institutional review board approval because the health outcome data were previously compiled by the Australian Institute of Health and Welfare and the Public Health Information Development Unit, Torrens University Australia, and did not include clinical or personal patient information.

Acknowledgements

We thank Nicholas Gellie, Jacob Mills, and Andrew Lowe for their helpful discussions and suggestions. Craig Liddicoat received support while undertaking this study through an Australian Government Research Training Program Scholarship.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2018.01.077.

References

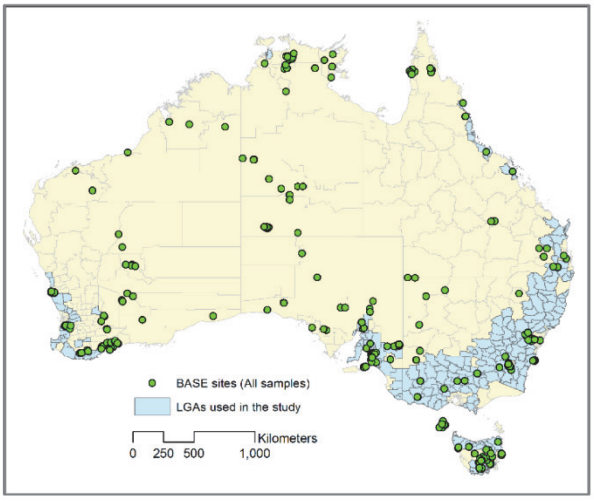
ABS, 2011. 1270.0.55.006 - Australian Statistical Geography Standard (ASGS): Correspondences, July 2011 - Local Government Area 2011 to Remoteness Area 2011. Australian Bureau of Statistics.
Barrett, J.E., Virginia, R.A., Wall, D.H., Parsons, A.N., Powers, L.E., Burkins, M.B., 2004. Variation in biogeochemistry and soil biodiversity across spatial scales in a polar desert ecosystem. Ecology 85, 3105–3118.
Baumgardner, D.J., 2012. Soil-related bacterial and fungal infections. J. Am. Board Fam. Med. 25, 734–744.
Bissett, A., Fitzgerald, A., Meintjes, T., Mele, P.M., Reith, F., Dennis, P.G., et al., 2016. Introducing BASE: the Biomes of Australian Soil Environments soil microbial diversity database. GigaScience 5, 21.
Browne, H.P., Forster, S.C., Anonye, B.O., Kumar, N., Neville, B.A., Stares, M.D., et al., 2016. Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation. Nature 533, 543–546.
Coleman, D., Crossley, D., Hendrix, P., 2004. Fundamentals of Soil Ecology. 2nd ed. Elsevier Academic Press, Burlington USA.
Croon, M.A., van Veldhoven, M.J.P.M., 2007. Predicting group-level outcome variables from variables measured at the individual level: a latent variable multilevel model. Psychol. Methods 12, 45–57.
Delgado-Baquerizo, M., Powell, J.R., Hamonts, K., Reith, F., Mele, P., Brown, M.V., et al., 2017. Circular linkages between soil biodiversity, fertility and plant productivity are limited to topsoil at the continental scale. New Phytol. 215, 1186–1196.
Docherty, K.M., Borton, H.M., Espinosa, N., Gebhardt, M., Gil-Loaiza, J., Gutknecht, J.L.M., et al., 2015. Key edaphic properties largely explain temporal and geographic variation in soil microbial communities across four biomes. PLoS One 10, e0135352.
Eisenberg, J.N.S., Desai, M.A., Levy, K., Bates, S.J., Liang, S., Naumoff, K., et al., 2007. Environmental determinants of infectious disease: a framework for tracking causal links and guiding public health research. Environ. Health Perspect. 115, 1216–1223.
ESRI, 2015. ArcGIS Desktop: Release 10.3.1. Environmental Systems Research Institute, Redlands, CA.
Fierer, N., Jackson, R.B., 2006. The diversity and biogeography of soil bacterial communities. Proc. Natl. Acad. Sci. 103, 626–631.
Gellie, N.J.C., Mills, J.G., Breed, M.F., Lowe, A.J., 2017. Revegetation rewilds the soil bacterial microbiome of an old field. Mol. Ecol. 26, 2895–2904.
Grundy, M.J., Rossel, R.A.V., Searle, R.D., Wilson, P.L., Chen, C., Gregory, L.J., 2015. Soil and Landscape Grid of Australia. Soil Re. 53, 835–844.
Haahtela, T., Holgate, S., Pawankar, R., Akdis, C.A., Benjaponpitak, S., Caraballo, L., 2013. The biodiversity hypothesis and allergic disease: World Allergy Organization position statement. World Allergy Organ. J. 6.
Hanski, I., von Hertzen, L., Fyhrquist, N., Koskinen, K., Torppa, K., Laatikainen, T., et al., 2012. Environmental biodiversity, human microbiota, and allergy are interrelated. Proc. Natl. Acad. Sci. 109, 8334–8339.
Ichinohe, T., Pang, I.K., Kumamoto, Y., Peaper, D.R., Ho, J.H., Murray, T.S., et al., 2011. Microbiota regulates immune defense against respiratory tract influenza A virus infection. Proc. Natl. Acad. Sci. 108, 5354–5359.
Jastrow, J., Miller, R., 1998. Soil aggregate stabilization and carbon sequestration: feedbacks through organomineral associations. In: Lal, R., Kimble, J.M., Follett, R.F., Stewart, B.A. (Eds.), Soil Processes and the Carbon Cycle. CRC Press, Boca Raton.
Jost, L., 2006. Entropy and diversity. Oikos 113, 363–375.
Kabata-Pendias, A., Mukherjee, A.B., 2007. Trace Elements from Soil to Human. Springer-Verlag, Berlin.
Kallenbach, C.M., Frey, S.D., Grandy, A.S., 2016. Direct evidence for microbial-derived soil organic matter formation and its ecophysiological controls. Nat. Commun. 7, 13630.
Kolivras, K.N., Johnson, P.S., Comrie, A.C., Yool, S.R., 2001. Environmental variability and coccidioidomycosis (valley fever). Aerobiologia 17, 31–42.
Koopman, J.S., Longini Jr., I.M., 1994. The ecological effects of individual exposures and nonlinear disease dynamics in populations. Am. J. Public Health 84, 836.
Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., et al., 2017. Caret: Classification and Regression Training. R Package Version 6.0-77.
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, 1023–1034.
McKenzie, N., Jacquier, D., Isbell, R., Brown, K., 2004. Australian Soils and Landscapes: An Illustrated Compendium. CSIRO Publishing, Melbourne, Australia.
McMurdie, P.J., Holmes, S., 2013. Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One 8, e61217.
Mills, J.G., Weinstein, P., Gellie, N.J.C., Weyrich, L.S., Lowe, A.J., Breed, M.F., 2017. Urban habitat restoration provides a human health benefit through microbiome rewilding: the microbiome rewilding hypothesis. Restor. Ecol. 25, 866–872.
Mitchell, R., Popham, F., 2008. Effect of exposure to natural environment on health inequalities: an observational population study. Lancet 372, 1655–1660.
Molloy, M.J., Bouladoux, N., Belkaid, Y., 2012. Intestinal microbiota: shaping local and systemic immune responses. Semin. Immunol. 24, 58–66.
Ord, J.K., Getis, A., 1995. Local spatial autocorrelation statistics: distributional issues and an application. Geogr. Anal. 27, 286–306.
O'Sullivan, L., Jardine, A., Cook, A., Weinstein, P., 2008. Deforestation, mosquitoes, and ancient Rome: lessons for today. Bioscience 58, 756–760.
Peverill, K.I., Sparrow, L.A., Reuter, D.J., 1999. Soil Analysis: An Interpretation Manual. CSIRO Publishing, Collingwood, Vic.
PHIDU, 2015. Social Health Atlas of Australia: Data by Local Government Area, March 2015 Release. Public Health Information Development Unit. Torrens University Australia, Adelaide.

PHIDU, 2016. Social Health Atlas of Australia: Data by Local Government Area, May 2016 Release. Public Health Information Development Unit. Torrens University Australia, Adelaide.
Polymenakou, P.N., 2012. Atmosphere: a source of pathogenic or beneficial microbes? Atmosphere 3, 87–102.
R-Core-Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Rook, G., 2013. Regulation of the immune system by biodiversity from the natural environment: an ecosystem service essential to health. Proc. Natl. Acad. Sci. 110, 18360–18367.
Ruokolainen, L., von Hertzen, L., Fyhrquist, N., Laatikainen, T., Lehtomäki, J., Auvinen, P., et al., 2015. Green areas around homes reduce atopic sensitization in children. Allergy 70, 195–202.
Stefka, A.T., Feehley, T., Tripathi, P., Qiu, J., McCoy, K., Mazmanian, S.K., et al., 2014. Commensal bacteria protect against food allergen sensitization. Proc. Natl. Acad. Sci. 111, 13145–13150.
Stein, M.M., Hrusch, C.L., Gozdz, J., Igartua, C., Pivniouk, V., Murray, S.E., et al., 2016. Innate immunity and asthma risk in Amish and Hutterite farm children. N. Engl. J. Med. 375, 411–421.
Susser, M., 1994. The logic in ecological: I. The logic of analysis. Am. J. Public Health 84, 825–829.
Torsvik, V., Øvreås, L., 2002. Microbial diversity and function in soil: from genes to ecosystems. Curr. Opin. Microbiol. 5, 240–245.
von Hertzen, L., Haahtela, T., 2006. Disconnection of man and the soil: reason for the asthma and atopy epidemic? J. Allergy Clin. Immunol. 117, 334–344.
von Hertzen, L., Hanski, I., Haahtela, T., 2011. Natural immunity: biodiversity loss and inflammatory diseases are two global megatrends that might be related. EMBO Rep. 12, 1089–1093.
Wall, D.H., Nielsen, U.N., Six, J., 2015. Soil biodiversity and human health. Nature 528, 69–76.
WHO, SCBD, 2015. Connecting Global Priorities: Biodiversity and Human Health. A State of Knowledge Review. World Health Organisation, Geneva, Switzerland.
Wopereis, H., Oozeer, R., Knipping, K., Belzer, C., Knol, J., 2014. The first thousand days - intestinal microbiology of early life: establishing a symbiosis. Pediatr. Allergy Immunol. 25, 428–438.
Yan, D., Mills, J.G., Gellie, N.J.C., Bissett, A., Lowe, A.J., Breed, M.F., 2018. High-throughput eDNA monitoring of fungi to track functional recovery in ecological restoration. Biol. Conserv. 217, 113–120.
Young, I.M., Crawford, J.W., 2004. Interactions and self-organization in the soil-microbe complex. Science 304, 1634–1637.
Zahran, H.H., 1997. Diversity, adaptation and activity of the bacterial flora in saline environments. Biol. Fertil. Soils 25, 211–223.
Zak, D.R., Holmes, W.E., White, D.C., Peacock, A.D., Tilman, D., 2003. Plant diversity, soil microbial communities, and ecosystem function: are there any links? Ecology 84, 2042–2050.
Zar, J.H., 2010. Biostatistical Analysis. Pearson Prentice Hall, Upper Saddle River, NJ.

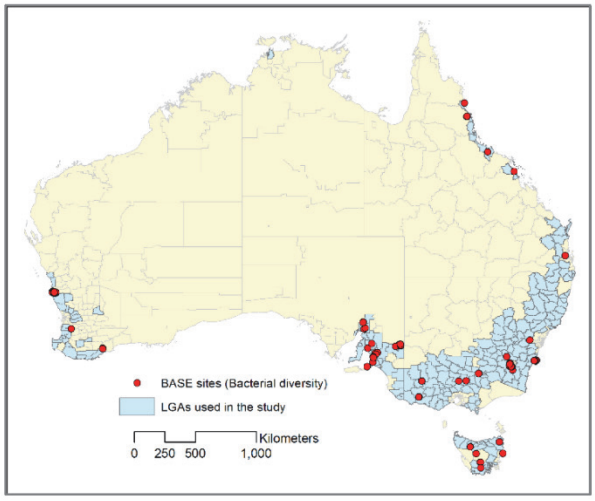
Appendix A – Supplementary Data

Ambient soil cation exchange capacity inversely associates with infectious and parasitic disease risk in regional Australia

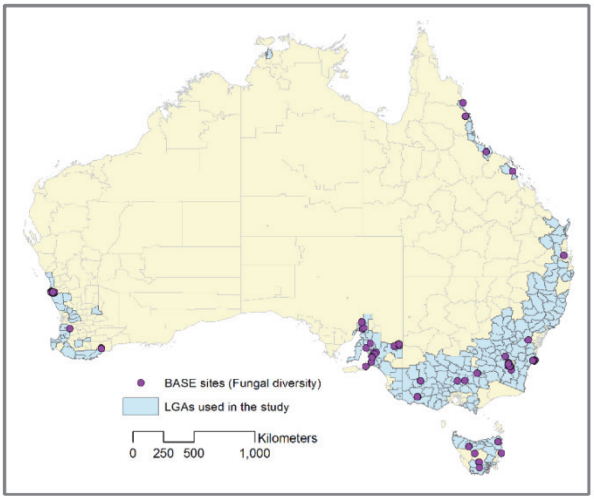
Craig Liddicoat, Peng Bi, Michelle Waycott, John Glover, Martin Breed, Philip Weinstein



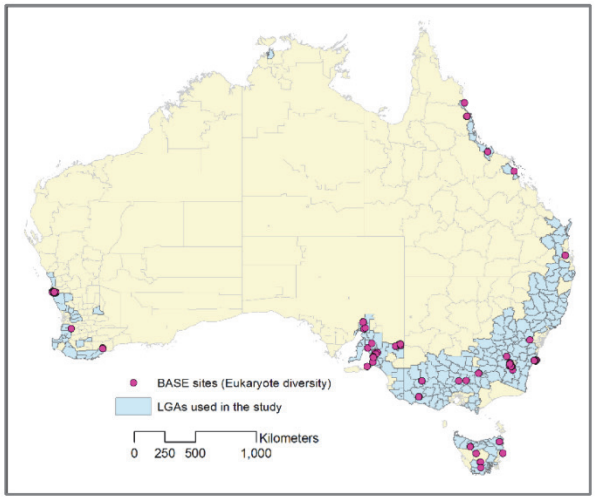
(A) All BASE sample locations (from contextual data published online 19/12/2017)



(B) Bacterial (16S) samples used (n = 353)



(C) Fungal (ITS) samples used (n = 338)



(D) Eukaryote (18S) samples used (n = 345)

Fig. S1. Locations of Biomes of Australian Soil Environments soil samples (BASE; Bisset et al., 2016; https://data.bioplatforms.com/organization/bpa-base, data extracted on 19/12/2017). Panel (A) shows all BASE sample locations, however at the time of writing not all of these were relevant to our study as some samples were variously missing data for major cations (needed to estimate soil CEC) or missing bacterial, fungal, or eukaryote community composition data. Maps (B) to (D) show the locations of BASE samples that we used to examine soil CEC-microbial diversity relationships, due to proximity to our study area, availability of major cation data, and availability of (B) bacterial (16S), (C) fungal (ITS), and (D) eukaryote (18S) operational taxonomic unit abundance data respectively.

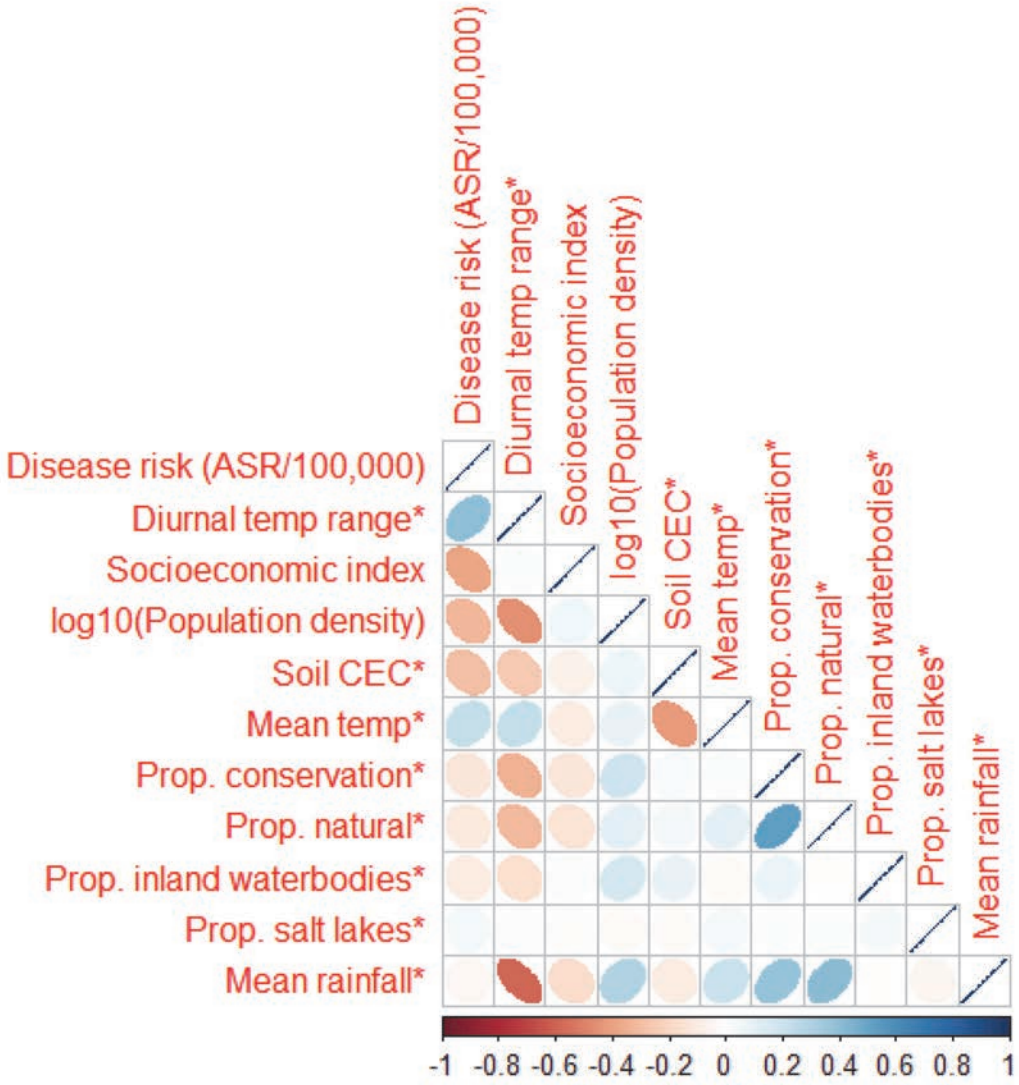


Fig. S2. Correlation plot of health outcome and predictor data used in multilevel modelling. Disease risk refers to the mean 2011-2013 age-standardized rate of public hospital admissions per 100,000 population (cumulative incidence) of infectious and parasitic diseases in Australian local government areas (LGAs). Environmental predictors marked with * are from probabilistic pseudo individual-level exposure sampling (n = 4544), while the outcome and remaining predictor data are derived from group-level LGA-based data (n = 228). Variables are ordered by the magnitude of correlation with the disease risk outcome data.

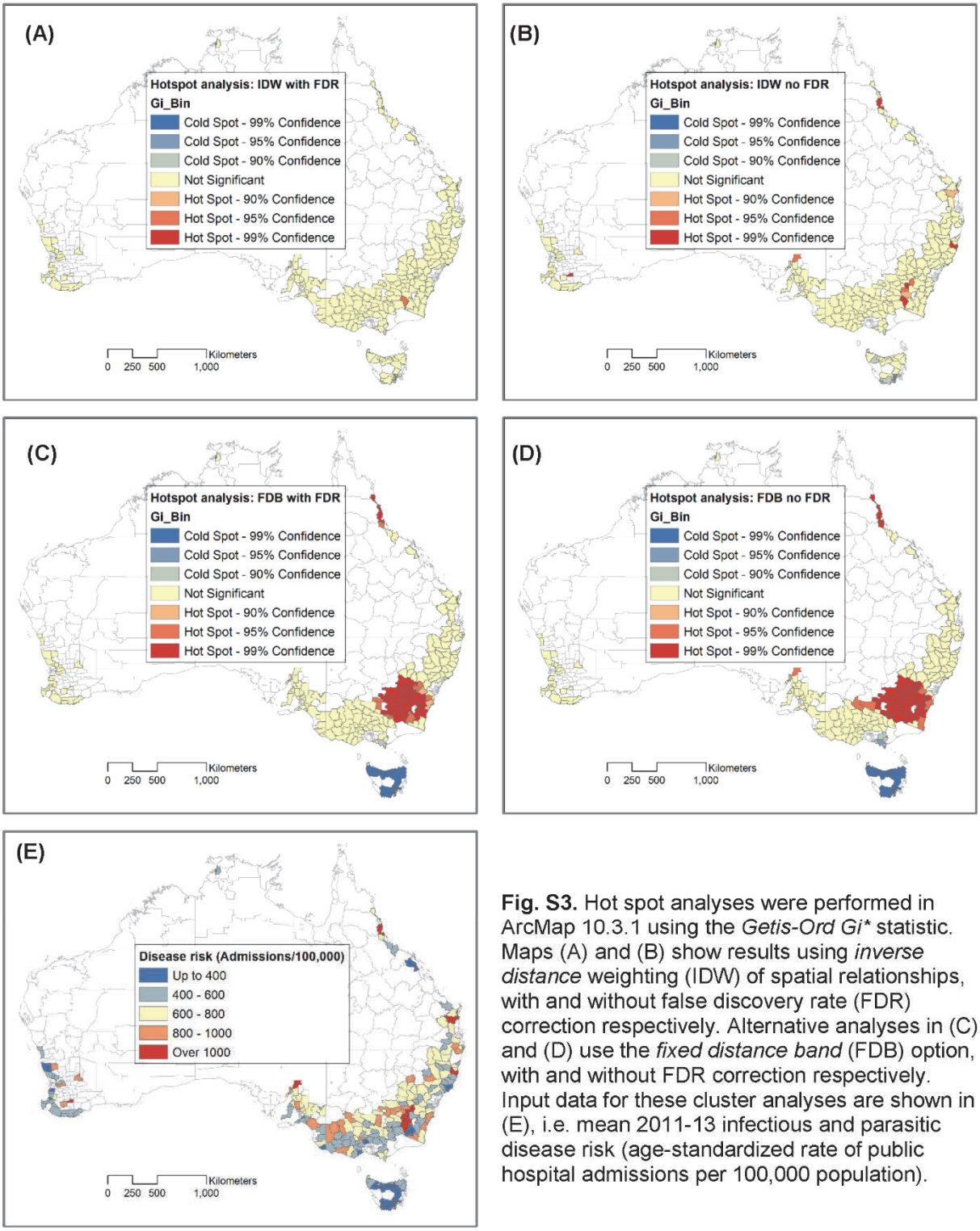


Fig. S3. Hot spot analyses were performed in ArcMap 10.3.1 using the *Getis-Ord Gi** statistic. Maps (A) and (B) show results using *inverse distance* weighting (IDW) of spatial relationships, with and without false discovery rate (FDR) correction respectively. Alternative analyses in (C) and (D) use the *fixed distance band* (FDB) option, with and without FDR correction respectively. Input data for these cluster analyses are shown in (E), i.e. mean 2011-13 infectious and parasitic disease risk (age-standardized rate of public hospital admissions per 100,000 population).

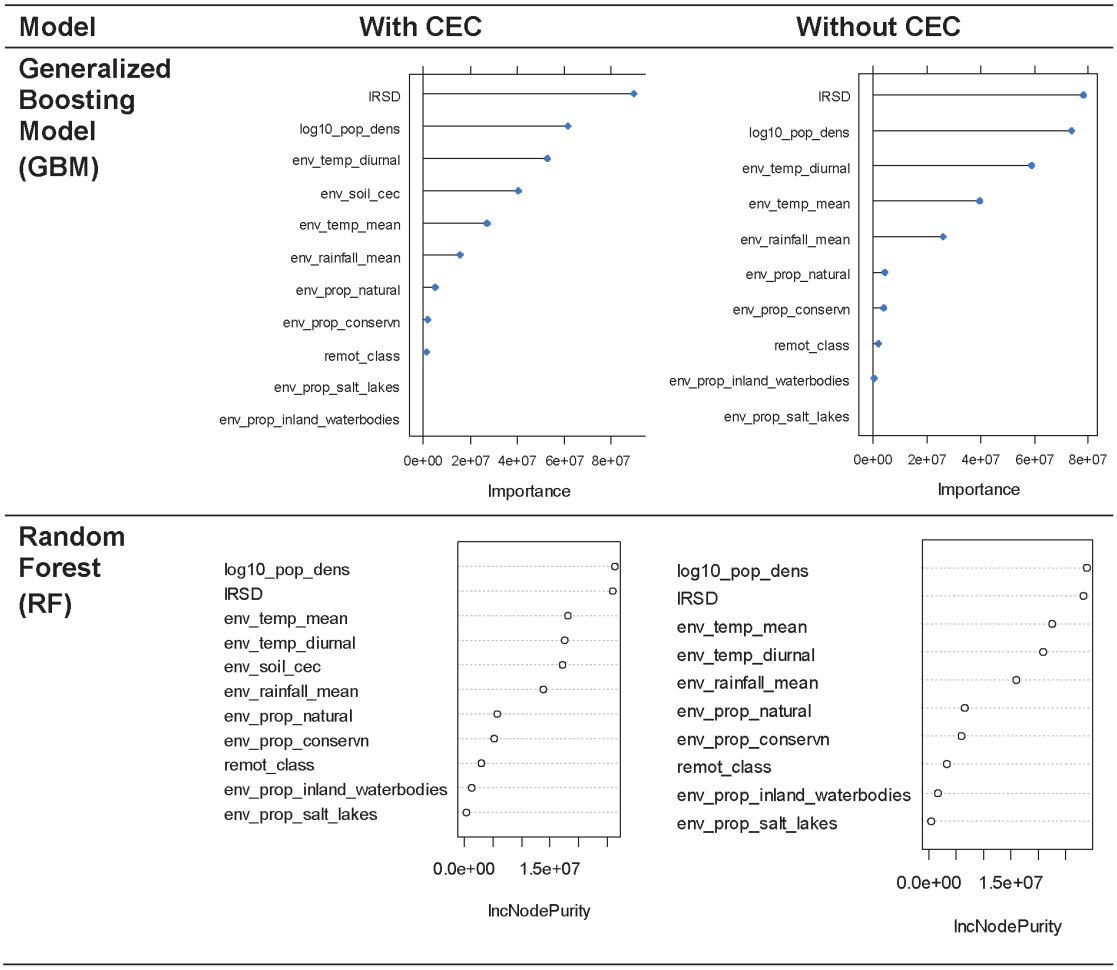


Fig. S4. Variable importance plots for GBM and RF models, with and without soil. Note: IRSD is Socioeconomic index (or status). Soil CEC ranks among other known environmental predictors of infectious and parasitic disease risk such as temperature and rainfall.

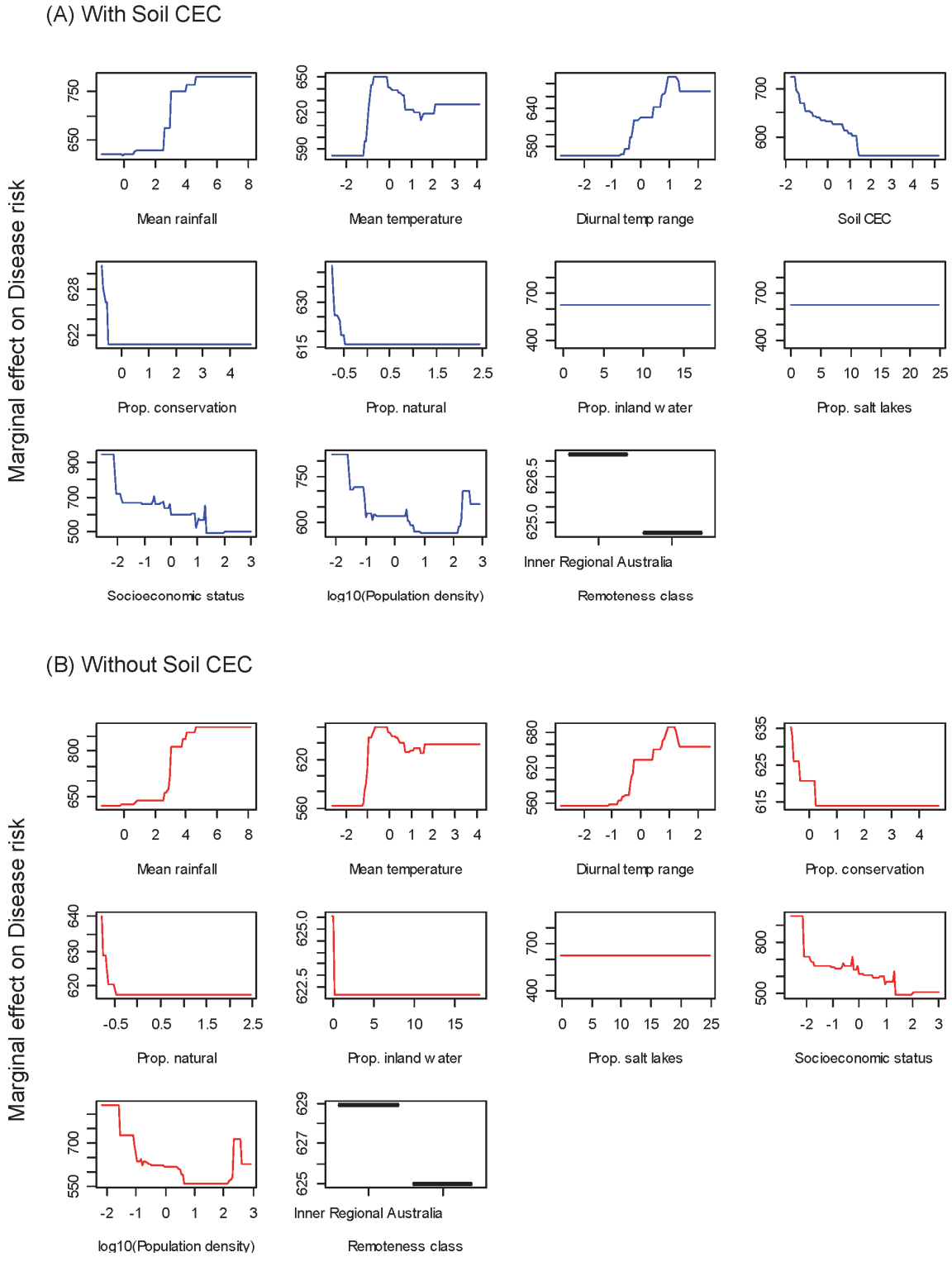


Fig. S5. Partial dependence plots for GBM models, (A) with, and (B) without inclusion of Soil CEC. X-axis variables represent centred and scaled data as used by the model.

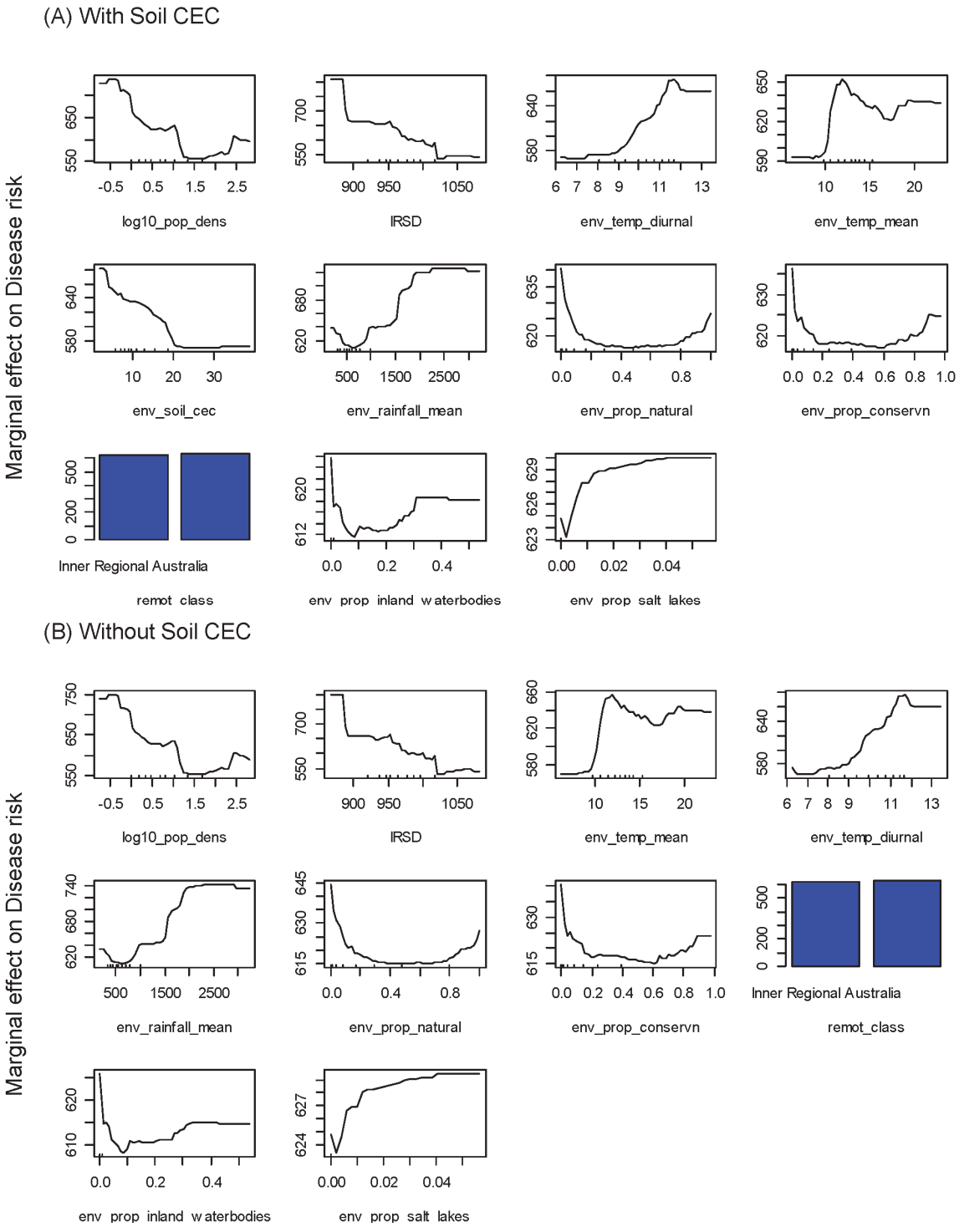


Fig. S6. Partial dependence plots for RF models, (A) with, and (B) without inclusion of Soil CEC. X-axis variables are shown in raw (unscaled) units.

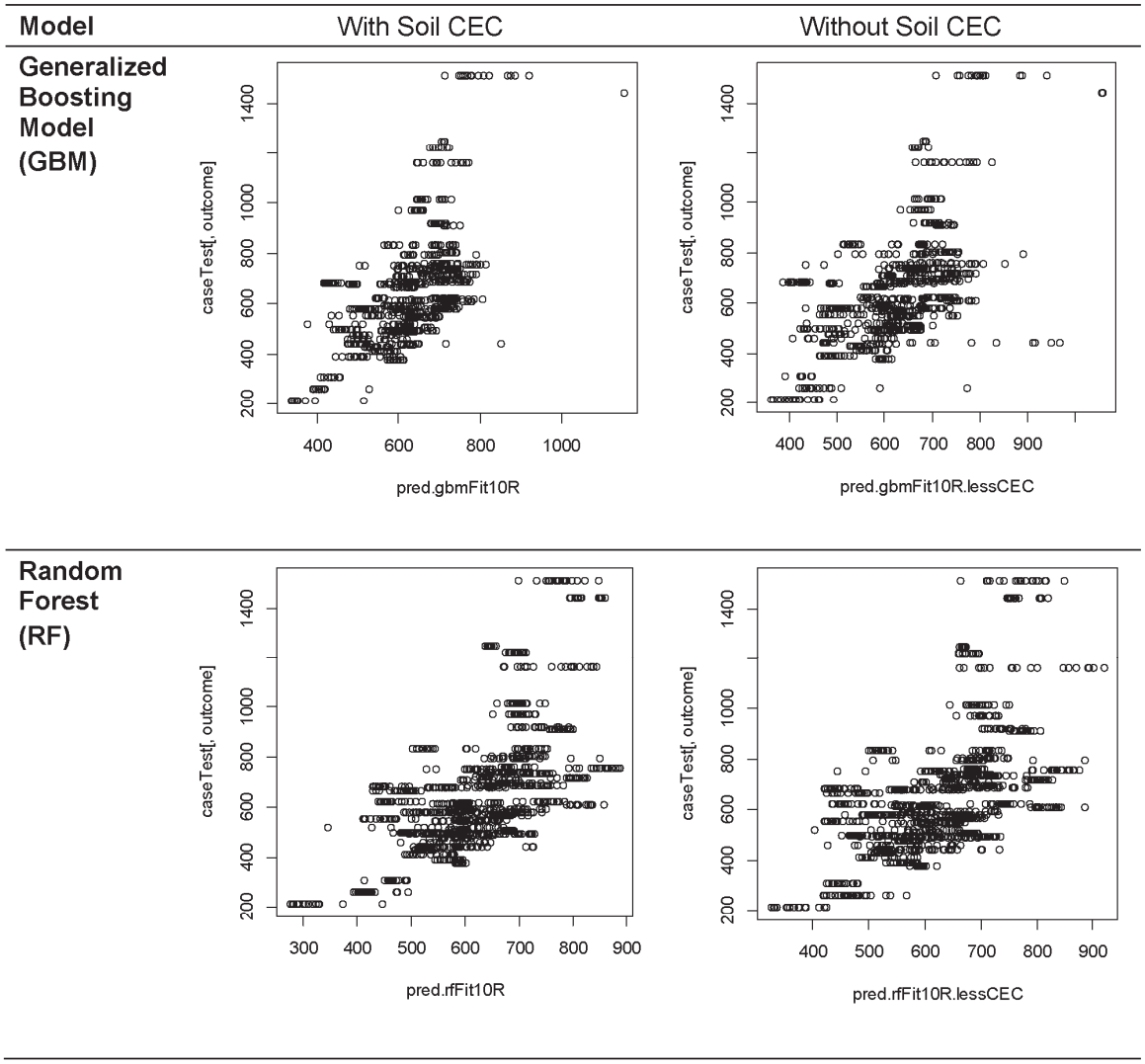


Fig. S7. Predictions (x-axes) versus observations (y-axes) of infectious and parasitic disease risk in withheld test areas (i.e. local government areas not used in modelling) for GBM and RF models, with and without soil.

Table S1. International classification of diseases 10th revision Australian modification (ICD-10-AM 7th edition) codes for infectious and parasitic diseases. This study has used aggregated data for public hospital admissions with principal diagnoses of infectious and parasitic diseases (ICD-10-AM codes A00-B99) listed below, as reported by the Social Health Atlas of Australia (PHIDU, 2015; PHIDU, 2016). To provide an indicative breakdown of disease types, separations from all hospitals (public + private) are shown in the right-hand columns. Note: at the time of writing, separation data were not available for public hospitals only. Further information is available from the AIHW National Hospital Morbidity Database (AIHW, 2017).

Subchapter	Separations (% of total)	
	2011-12	2012-13
0 Intestinal infectious diseases (A00–A09)	**76,160 (53.9%)**	**74,798 (52.8%)**
1 Tuberculosis (A15–A19)	1585 (1.1%)	999 (0.7%)
2 Certain zoonotic bacterial diseases (A20–A28)	390 (0.3%)	340 (0.2%)
3 Other bacterial diseases (A30–A49)	**23,698 (16.8%)**	**25,594 (18.1%)**
4 Infections with a predominantly sexual mode of transmission (A50–A64)	2412 (1.7%)	2422 (1.7%)
5 Other spirochaetal diseases (A65–A69)	170 (0.1%)	81 (0.1%)
6 Other diseases caused by chlamydiae (A70–A74)	67 (<0.05%)	69 (<0.05%)
7 Rickettsioses (A75–A79)	217 (0.2%)	222 (0.2%)
8 Viral infections of the central nervous system (A80–A89)	2318 (1.6%)	3912 (2.8%)
9 Arthropod–borne viral fevers and viral haemorrhagic fevers (A90–A99)	407 (0.3%)	478 (0.3%)
10 Viral infections characterized by skin and mucous membrane lesions (B00–B09)	6614 (4.7%)	6552 (4.6%)
11 Viral hepatitis (B15–B19)	1763 (1.2%)	1757 (1.2%)
12 Human immunodeficiency virus (HIV) disease (B20–B24)	174 (0.1%)	268 (0.2%)
13 Other viral diseases (B25–B34)	**20,336 (14.4%)**	**19208 (13.5%)**
14 Mycoses (B35–B49)	2987 (2.1%)	2997 (2.1%)
15 Protozoal diseases (B50–B64)	782 (0.6%)	882 (0.6%)
16 Helminthiases (B65–B83)	277 (0.2%)	230 (0.2%)
17 Pediculosis, acariasis and other infestations (B85–B89)	615 (0.4%)	598 (0.4%)
18 Sequelae of infectious and parasitic diseases (B90–B94)	No data	1 (<0.05%)
19 Bacterial, viral and other infectious agents (B95–B97)	11 (<0.05%)	48 (<0.05%)
20 Other infectious diseases (B99)	253 (0.2%)	306 (0.2%)
TOTAL	141,236	141,762

(Separations > 10% annual total are highlighted in bold)

Table S2. Reference information for social and environmental variables (candidate predictors) used in the study.

Social variables (group-level only, by LGAs)	Units	Analysis type^	Reference period	Data source(s)
Socioeconomic index#	No units	G, M	2011	(PHIDU, 2015)
log10(Population density)	log10(Persons/km^2)	M	2013	(ABS, 2011; PHIDU, 2015)
Environmental variables (raster mapping)				
Mean rainfall	mm	M	1976-2005	(Whitley et al., 2014)
Mean temperature	°C	M	1970-2012	(Whitley et al., 2014)
Diurnal temperature range	°C	M	1970-2012	(Whitley et al., 2014)
Soil CEC	meq/100g*	G, M	2014	(Grundy et al., 2015)
Proportion of conservation land (this was a composite of 'nature conservation', 'managed resource protection', and 'other minimal use' land use classes)	% (of 3 km radius area)	M	2014	Calculated based on (ABARES, 2014)
Proportion of natural land (in addition to the conservation land use classes above, this also included 'grazing native vegetation', and 'production forestry' areas)	% (of 3 km radius area)	M	2014	Calculated based on (ABARES, 2014)
Proportion of inland water bodies	% (of 3 km radius area)	M	2000-2008	Calculated based on (Geoscience-Australia, 2014)
Proportion of salt lakes	% (of 3 km radius area)	M	2000-2008	Calculated based on (Geoscience-Australia, 2014)

Note: #The Socioeconomic index variable was originally termed Index of relative socioeconomic disadvantage in the Social Health Atlas of Australia (PHIDU, 2015), however the scale and direction of this variable is that low values correspond to greater disadvantage or deprivation while high values correspond to reduced disadvantage. Accordingly, in this study we refer to Socioeconomic index and Socioeconomic status interchangeably. *National mapping of 0-5 cm soil cation exchange capacity has units of milliequivalents per 100 grams of soil (meq/100g) which is equivalent to the units of centimoles of positive charge per kg (cmol(+)/kg) as discussed in the context of Biomes of Australian Soil Environments site chemical data in the main article Fig. 1. ^Analysis type G refers to group-level analyses of risk and relative risks of disease (using Socioeconomic index and Soil CEC summarized within LGAs). Analysis type M refers to the multilevel analysis including probabilistic pseudo individual-level exposure estimates. For the multilevel analysis, each of the cells in the environmental raster data layers were re-expressed as a summary value of the 3 km radius surrounding area then 20 cells per LGA were sampled at random. LGAs are Australian local government areas.

Supplementary References

ABARES. Catchment scale land use of Australia March 2014. Australian Bureau of Agricultural and Resource Economics and Sciences, Department of Agriculture and Water Resources, 2014.

ABS. Local Government Area ASGC Ed 2011 Digital Boundaries in ESRI Shapefile Format. Australian Bureau of Statistics, 2011.

AIHW. AIHW National Hospital Morbidity Database. Australian Institute of Health and Welfare. Separation statistics by principal diagnosis (ICD-10-AM 7th edition), Australia, 2011-12 to 2012-13, 2017.

Bissett A, Fitzgerald A, Meintjes T, Mele PM, Reith F, Dennis PG, et al. Introducing BASE: the Biomes of Australian Soil Environments soil microbial diversity database. GigaScience 2016; 5: 21.

Geoscience-Australia. Dynamic Land Cover Dataset V1.0, 27 May 2014. Geoscience Australia, 2014.

Grundy MJ, Rossel RAV, Searle RD, Wilson PL, Chen C, Gregory LJ. Soil and Landscape Grid of Australia. Soil Research 2015; 53: 835-844.

PHIDU. Social Health Atlas of Australia: Data by Local Government Area, March 2015 release. Public Health Information Development Unit. Torrens University Australia, Adelaide, 2015.

PHIDU. Social Health Atlas of Australia: Data by Local Government Area, May 2016 release. Public Health Information Development Unit. Torrens University Australia, Adelaide, 2016.

Whitley R, Evans B, Pauwels J, Hutchinson M, Xu T, Han W. Bioclimatic surfaces: eMAST-R-Package 2.0, 0.01 degree, Australian Coverage obtained from http://dap.nci.org.au, made available by the Ecosystem Modelling and Scaling Infrastructure (eMAST, http://www.emast.org.au) Facility of the Terrestrial Ecosystem Research Network (TERN, http://www.tern.org.au). Macquarie University, Sydney, Australia, 2014.

Chapter 4: Bacterial indicators of restoration and implications for human health

Statement of Authorship

Title of Paper	Can bacterial indicators of a grassy woodland restoration inform ecosystem assessment and microbiota-mediated human health?
Publication Status	☐ Published ☑ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted w ork w ritten in manuscript style
Publication Details	Liddicoat, C., Weinstein, P., Bissett, A., Gellie, N., Mills, J., Waycott, M., Breed, M., 2019. Can bacterial indicators of a grassy woodland restoration inform ecosystem assessment and microbiota-mediated human health? *Environment International*. [Accepted 6 May 2019]

Principal Author

Name of Principal Author (Candidate)	Craig Liddicoat		
Contribution to the Paper	I led the conceptualisation and data integration, developed the new merged-sample bootstrap resampling framework, performed the modelling, led the analysis and interpretation, wrote the first draft, and led the revisions of the final manuscript.		
Overall percentage (%)	80		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	6 May 2019

Co-Author Contributions

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

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

Name of Co-Author	Philip Weinstein		
Contribution to the Paper	Contributed to the conceptualisation, analysis and interpretation of results, paper writing and revisions.		
Signature		Date	21/5/19

Name of Co-Author	Andrew Bissett		
Contribution to the Paper	Contributed to the conceptualisation, data collection and preparation, analysis and interpretation of results, paper writing and revisions.		
Signature		Date	24/5/2019

Name of Co-Author	Nick Gellie		
Contribution to the Paper	Contributed to the conceptualisation, data collection, analysis and interpretation of results, paper writing and revisions.		
Signature		Date	17/05/2019

Name of Co-Author	Jacob Mills		
Contribution to the Paper	Contributed to the conceptualisation, data collection, analysis and interpretation of results, paper writing and revisions.		
Signature		Date	14/06/2019

Name of Co-Author	Michellle Waycott		
Contribution to the Paper	Contributed to the analysis and interpretation of results, paper writing and revisions.		
Signature		Date	13/6/19

Name of Co-Author	Martin Breed		
Contribution to the Paper	Contributed to the conceptualisation, analysis and interpretation of results, paper writing and revisions.		
Signature		Date	6/6/19

Please cut and paste additional co-author panels here as required.

Environment International 129 (2019) 105–117

Contents lists available at ScienceDirect

Environment International

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

Can bacterial indicators of a grassy woodland restoration inform ecosystem assessment and microbiota-mediated human health?

Craig Liddicoat[a,*], Philip Weinstein[a], Andrew Bissett[b], Nicholas J.C. Gellie[a], Jacob G. Mills[a], Michelle Waycott[a], Martin F. Breed[a]

[a] School of Biological Sciences and the Environment Institute, The University of Adelaide, SA 5005, Australia
[b] CSIRO Oceans and Atmosphere, Hobart, TAS, 7000, Australia

ARTICLE INFO

Handling Editor: Yong-Guan Zhu

Keywords:
Bacterial indicators
Biodiversity hypothesis
Ecological restoration
Environmental health
Merged-sample bootstrap
Restoration genomics

ABSTRACT

Understanding how microbial communities change with environmental degradation and restoration may offer new insights into the understudied ecology that connects humans, microbiota, and the natural world. Immunomodulatory microbial diversity and 'Old Friends' are thought to be supplemented from biodiverse natural environments, yet deficient in anthropogenically disturbed or degraded environments. However, few studies have compared the microbiomes of natural vs. human-altered environments and there is little knowledge of which microbial taxa are representative of ecological restoration—i.e. the assisted recovery of degraded ecosystems typically towards a more natural, biodiverse state. Here we use novel bootstrap-style resampling of site-level soil bacterial 16S rRNA gene environmental DNA data to identify genus-level indicators of restoration from a 10-year grassy eucalypt woodland restoration chronosequence at Mt Bold, South Australia. We found two key indicator groups emerged: 'opportunistic taxa' that decreased in relative abundance with restoration and more stable and specialist, 'niche-adapted taxa' that increased. We validated these results, finding seven of the top ten opportunists and eight of the top ten niche-adapted taxa displayed consistent differential abundance patterns between human-altered vs. natural samples elsewhere across Australia. Extending this, we propose a two-dimensional mapping for ecosystem condition based on the proportions of these divergent indicator groups. We also show that restoring a more biodiverse ecosystem at Mt Bold has increased the potentially immune-boosting environmental microbial diversity. Furthermore, environmental opportunists including the pathogen-containing genera *Bacillus*, *Clostridium*, *Enterobacter*, *Legionella* and *Pseudomonas* associated with disturbed ecosystems. Our approach is generalizable with potential to inform DNA-based methods for ecosystem assessment and help target environmental interventions that may promote microbiota-mediated human health gains.

1. Introduction

People are losing contact with nature due to rapid urbanization (Rydin et al., 2012), biodiversity loss (Mace et al., 2018), and environmental degradation (Navarro et al., 2017). At the same time, rates of immune-related disease are increasing (von Hertzen et al., 2011). Many immunologists and medical researchers now believe these trends are linked (WHO and SCBD, 2015). Microbial diversity and perhaps key species (microbial 'Old Friends') from biodiverse and natural environments are thought to play a critical beneficial role in the development and maintenance of human immune fitness (Rook, 2013; von Hertzen et al., 2011). Environmental microbiota might supplement our own protective human microbiota (e.g. in the skin, airway, gut), participate in immune signalling (triggering defence or tolerance responses), and help build immune memory (Flandroy et al., 2018; Rook, 2013). These interactions underpin many aspects of our health, and can help regulate both infectious and non-infectious disease (Ichinohe et al., 2011; Ottman et al., 2018; Stein et al., 2016).

Meanwhile, human societies unwittingly affect landscape-scale drivers of environmental microbiota. For example, land use and management, biomass production, plant species and diversity, and soil quality each contribute to the development of characteristic plant and soil microbiota (Bulgarelli et al., 2013; Delgado-Baquerizo et al., 2017; Delgado-Baquerizo et al., 2018; Gellie et al., 2017a; Turner et al.,

* Corresponding author.
E-mail addresses: craig.liddicoat@adelaide.edu.au (C. Liddicoat), philip.weinstein@adelaide.edu.au (P. Weinstein), andrew.bisset@csiro.au (A. Bissett), nick.gellie@adelaide.edu.au (N.J.C. Gellie), jacob.mills@adelaide.edu.au (J.G. Mills), michelle.waycott@adelaide.edu.au (M. Waycott), martin.breed@adelaide.edu.au (M.F. Breed).

https://doi.org/10.1016/j.envint.2019.05.011
Received 19 December 2018; Received in revised form 5 May 2019; Accepted 6 May 2019

2013). Differing land use and management can also contribute distinct microbial signatures to airborne microbiota (i.e. aerobiology) dispersed among exposed neighbouring human populations (Bowers et al., 2013; Bowers et al., 2011; Mhuireach et al., 2016). With estimates of up to 47% of the global ice-free land area considered degraded and up to 76% impacted by human activities (Navarro et al., 2017), there is considerable scope for the anthropogenic modification of environmental microbiota to represent a potentially important but largely unappreciated feedback mechanism, linking ecosystems and human health. Equally, restoring biodiverse ecosystems where people live may offer promise to improve immune fitness and human health (Mills et al., 2017; Robinson et al., 2018). Ecological restoration—i.e. the assisted recovery of an ecosystem towards an ecological reference state (SER International Science and Policy Working Group, 2004)—also has key objectives to support native biodiversity, ecosystem functions, genetic resources, resilience against environmental stressors, productive potential and sustainable use.

However, large knowledge gaps remain concerning the immunomodulatory potential of environmental microbiomes. There are few studies that compare the microbiomes (i.e. genetic material of microbiota) of natural vs. human-altered outdoor environments from a human health perspective. Microbiomes of indoor only environments have received greater attention (Gilbert and Stephens, 2018). Suspected Old Friends include environmental microorganisms that we have co-evolved with, and for which we needed to develop tolerance (e.g. *Salmonella*, helminths, *Mycobacterium vaccae*; Rook, 2012). However the membership and modes of action of microbial Old Friends remain largely unknown (Rook, 2013), in part due to the uncharacterized 'microbial dark matter' that dominates environmental microbiomes (Lloyd et al., 2018). We do not yet know if it is microbial diversity alone, microbial diversity with Old Friends, or possibly some other functional component(s) of environmental microbiomes that may provide protective immune-training or other physiological benefits. However, if we can identify indicator taxa for natural and biodiverse environments, these might represent possible new candidate taxa to investigate further for immunomodulatory potential and to otherwise help inform the assessment, design and restoration of new nature-based public health interventions.

To progress, there is a need to build the knowledge base of microbiomes that associate with different environment types. Soil environmental DNA (eDNA) may provide a useful medium for tracking changes in ecosystem condition (Gellie et al., 2017a; Yan et al., 2018) because soils can retain a biological imprint from recent land use and management (Janzen, 2016). With the reducing costs of DNA sequencing technology, soil bacterial 16S rRNA marker gene survey data are increasingly used to help characterize soil microbiota. Also, soil-associated microbes often represent a significant component of the aerobiology derived from environments (Polymenakou, 2012), which provides ambient biological connectivity to exposed human populations. Soils have also been highlighted elsewhere due to their often high microbial diversity and immunomodulatory potential (Liddicoat et al., 2018; von Hertzen and Haahtela, 2006).

The task of exploring potential human exposures to microbiota from different characteristic environments suggests the need for microbiota profiling methods that reflect a human-scale or site-level exposure. To estimate the potential accumulative human-microbial exposure from environmental samples alone, without human test subjects, it may be necessary to consider multiple samples for each site. Such an approach would provide a broader picture of the microbial diversity, composition, and key taxa within each particular environment type. For example, we might expect the overall alpha diversity of microbiota encountered by a person interacting with a biodiverse site (represented by a number of samples) to be greater than the maximum alpha diversity of individual microbiota samples considered in isolation, due to natural heterogeneity and particular taxa not occurring everywhere. However, customary approaches to microbiome data analysis contain unexplored uncertainty and have potential to overlook less abundant taxa. It is common to rarefy marker gene survey data to a minimum sequence read depth in order to normalize sampling effort (Weiss et al., 2017), however this approach has been criticized for discarding valuable information about microbial communities (McMurdie and Holmes, 2014). These issues point to the opportunity for employing general purpose bootstrap resampling techniques on merged samples from across a site, combined with rarefying to normalize sampling effort, as a means to preserve survey data and obtain a deeper understanding of site-level (multiple sample) microbiota characteristics. Through bootstrap-style resampling we can also better understand the uncertainty or distribution of equally likely outcomes for the analyses undertaken.

Here we sought to answer the following research questions: (1) Can we employ innovative bootstrap-style resampling methods for site-level analysis of microbiome survey data to improve understanding of different environment types, potential human-environmental microbiome exposures, and their uncertainty? (2) What are the key bacterial genera (i.e. indicator taxa) that show directional trends where restoration has been undertaken, and are there generalizable patterns in contrasting human-altered vs. natural land uses at a continental scale? (3) Are there trends in human-associated genera (e.g. commensals, potential pathogens, potential Old Friends) with restoration? Our findings highlight characteristics of the trending microbial taxa from a localized restoration study site that display generalizability to microbiota changes between human-altered and natural samples elsewhere across Australia.

2. Methods

2.1. Overview

We first consider existing soil microbiome data from a historic chronosequence of restoration of a *Eucalyptus leucoxylon*-dominated grassy woodland vegetation community at Mt Bold, South Australia (Gellie et al., 2017a, 2017b, 2017c). Grassy woodland communities typically comprise groundcover of grasses, herbs and shrubs between scattered medium to large trees. Bacterial 16S rRNA marker gene survey data from sites spanning a gradient from cleared land, various revegetation ages (6, 7, 8, 10 years), and three reference patches of native vegetation (Remnants A, B, C) provide a rare dataset to investigate microbial signatures of ecological restoration. We extend the broad community and phylum-level trends described in Gellie et al. (2017a) (Web Appendix, Figs. S1–S2) to highlight genus-level indicators. To identify key trending taxa with restoration, and their uncertainty, we propose an innovative *merged-sample bootstrap resampling* framework (described below) based on triplicate 16S microbiome samples that characterize each site, as an alternative to conventional one-off rarefying of individual samples. We then looked for support for the generalizability of patterns identified from Mt Bold within publicly-available Australia-wide soil microbiome data from the Biomes of Australian Soil Environments (BASE) dataset (https://data.bioplatforms.com/organization/bpa-base; Bissett et al., 2016).

2.2. Study data

Methods for sample collection, soil chemical and physical analysis, eDNA extraction, sequencing of the bacterial 16S rRNA gene, and bioinformatic analyses to derive operational taxonomic unit (OTU) abundance tables and taxonomic classifications are described in Gellie et al. (2017a) and Bissett et al. (2016). Briefly, OTUs were based on clustered sequences of ≥97% similarity and derived separately for the Mt Bold and BASE datasets. In both cases, open reference OTU picking and sequence abundance assignment workflows were used, with taxonomy mapped to the Greengenes (13–5) reference database, as described in Bissett et al. (2016).

From Mt Bold, we used the three replicates of 16S data collected for each restoration treatment (or sample type). We focussed our analyses

on surface soil (0–10 cm; n = 24), although subsurface (20–30 cm; n = 24) data are also included in Web Appendix plots (described later) to provide confidence in our observed trends. From BASE, after excluding the Mt Bold samples, we included surface soil 16S data using the following selection steps. We excluded BASE data from off-shore islands and external territories, and filtered samples with extreme, outlying soil conditions to avoid undue influence from data associated with atypical and limiting conditions for soil biology. Specifically, we excluded samples that were strongly acid (pH_{H2O} < 4.5), strongly alkaline (pH_{H2O} > 9), saline (electrical conductivity > 2 dS/m), and very high clay content (> 50%). From available 16S data, we assigned samples with the following land uses to classes of either 'natural' (including national park, nature conservation, strict nature reserves, wilderness area), or 'human-altered' (including cereals–wheat, cotton, irrigated seasonal horticulture, pasture legume/grass mixtures, rehabilitation – i.e. once degraded land that is being restored towards a reference state, sown grasses, sugar, tree fruits–apple). These assignments were guided by interpreting Australian land use classification guidelines (ABARES, 2016). Due to ambiguity, we excluded land uses of grazing native vegetation, native/exotic pasture mosaic, natural feature protection, other conserved area, protected landscape, reserve, and residual native cover. Lastly, we did not wish to compare human-altered and natural samples separated by large geographic distances, so we used nearest neighbour calculations based on latitude and longitude to exclude samples with > 5 degrees of geographic separation from their nearest complementary sample type (i.e. each natural sample is ≤5 degrees from a human-altered sample, and vice versa), to arrive at the final human-altered (n = 78) and natural (n = 139) samples (locations shown in Web Appendix, Fig. S3). The nearest neighbour filtering eliminated the sample type of wilderness area.

Within the respective Mt Bold and Australia-wide BASE microbiome data we only used taxa assigned as Bacteria, and excluded all taxa assigned as chloroplast or mitochondria and any taxa not assigned at the bacterial phylum level. The two datasets were filtered separately to remove taxa with < 100 sequence reads across all samples (Gellie et al., 2017a), or that did not occur in at least two samples.

2.3. Data analysis

We used R software (R Core Team, 2018) for the majority of analyses, with purpose-built scripts (see Web Appendix) employing the microbiome data analysis framework of the R *phyloseq* package (McMurdie and Holmes, 2013). *Phyloseq* uses microbiome data objects that facilitate linked analysis of OTU abundance, taxonomy and sample contextual data. As described below, we used both rarefied and non-rarefied OTU abundance data, reflecting different input and sample normalization requirements for particular analyses. Briefly, rarefied OTU abundance data were used in visualising beta diversity and permutational multivariate analysis of variance (PERMANOVA; Anderson, 2017) testing of differences in microbiota groups, while non-rarefied OTU abundances were used when comparing OTU relative abundance (%) data and for differential abundance testing using the R *DESeq2* package (Love et al., 2014). We implemented the new *merged-sample bootstrap resampling* (described below) to examine mean responses and uncertainty in OTU- and functional-alpha diversity, and associations between OTU relative abundance and restoration at Mt Bold. We visualized the sequence depth of samples using rarefaction curves (Web Appendix, Fig. S4). OTU alpha diversity in each sample was estimated using the exponential transform of Shannon Index values to derive the effective number of OTUs (Jost, 2006). We visualized differences between sample microbiota (beta diversity) using non-metric multidimensional scaling (NMDS) ordination of Bray-Curtis distances based on rarefied OTU abundances, at the minimum sequence read depth of the samples concerned. PERMANOVA was used to examine the statistical significance of compositional differences between rarefied OTU abundances of sample groups, implemented using the *adonis()* function, followed by the *betadisper()* function to test for homogeneity of group dispersions, where both functions are from the R *vegan* package (Oksanen et al., 2018). Distance-based redundancy analysis was used in preliminary exploration of relationships between microbiota and soil conditions at Mt Bold. Further analyses are described in more detail below.

2.4. Merged-sample bootstrap resampling

To preserve microbiome survey data while examining site-level (i.e. multiple sample) characteristics and their uncertainty, we implemented the following resampling framework (also see Fig. 1):

1. Prepare OTU abundance data in non-rarefied form with varying sequence depths across multiple sites. Here the full OTU table is expressed in a single *phyloseq* microbiome data object with the multiple samples per site given a respective site label in the *phyloseq* sample contextual data table. In our case, the Mt Bold data have triplicate samples at each site (i.e. treatment or sample type).
2. Set seed for pseudo-random number generator.
3. Rarefy all samples to an even sampling depth; equal to the minimum sequence depth across all samples, using sampling with replacement.
4. With the rarefied data from step 3, merge samples (i.e. group by site label) and sum taxa counts by site. Transfer results to a new site-level *phyloseq* microbiome data object.
5. Set seed for pseudo-random number generator.
6. Rarefy the merged data from step 4 to an even sampling depth, using the same minimum sequence depth value from step 3, again using sampling with replacement.
7. Evaluate the measure of interest based on the merged and rarefied site-level data from step 6 (e.g. alpha diversity, evaluate taxa relative abundance and correlation with revegetation age) and store results as an element of an output list.
8. Repeat steps 2 to 7, up to the number of predefined iterations (e.g. B = 100).

Sampling with replacement was chosen for consistency with general purpose bootstrap methods, and particularly in step 3 to avoid generating repeated data for the sample with the minimum sequence size. We did, however, trial combinations of sampling with- and without-replacement and achieved similar results (Web Appendix, Fig. S5). We note that two rounds of rarefying (as above) achieved normalization of sampling effort that contributes to site-level data and enables the *merged-sample bootstrap* density distributions (for equally likely site-level outcomes) to be plotted on the same scale as data derived from one-off rarefied samples.

2.5. Assessing trends: Mt Bold restoration

We examined trends in taxa relative abundance with restoration by considering metrics for effect size and significance. In a strict sense, the restoration chronosequence represents an ordinal independent variable. Using ordinal regression analysis of variance (AOV) (Gertheiss, 2015) for each taxon, we tested the null hypothesis of no ordinal trend by revegetation age, assuming ordinal values (cleared = 1; 6 years = 2; 7 years = 3; 8 years = 4; 10 years = 5; Remnants A, B, C = 6). However, treating the data in this way does not provide information on effect size. So, we assessed effect size or strength of relationship (Shinichi and Innes, 2007) by assuming pseudo-continuous values for the missing revegetation ages (cleared = 0 years; Remnants A, B, C = 20 years) to calculate the Pearson correlation coefficient with each distribution of taxon relative abundance. These metrics were calculated using the *merged-sample bootstrap resampling* framework, as above. The top trending taxa were identified based on the magnitude and sign (positive = increasing, negative = decreasing) of the mean correlation

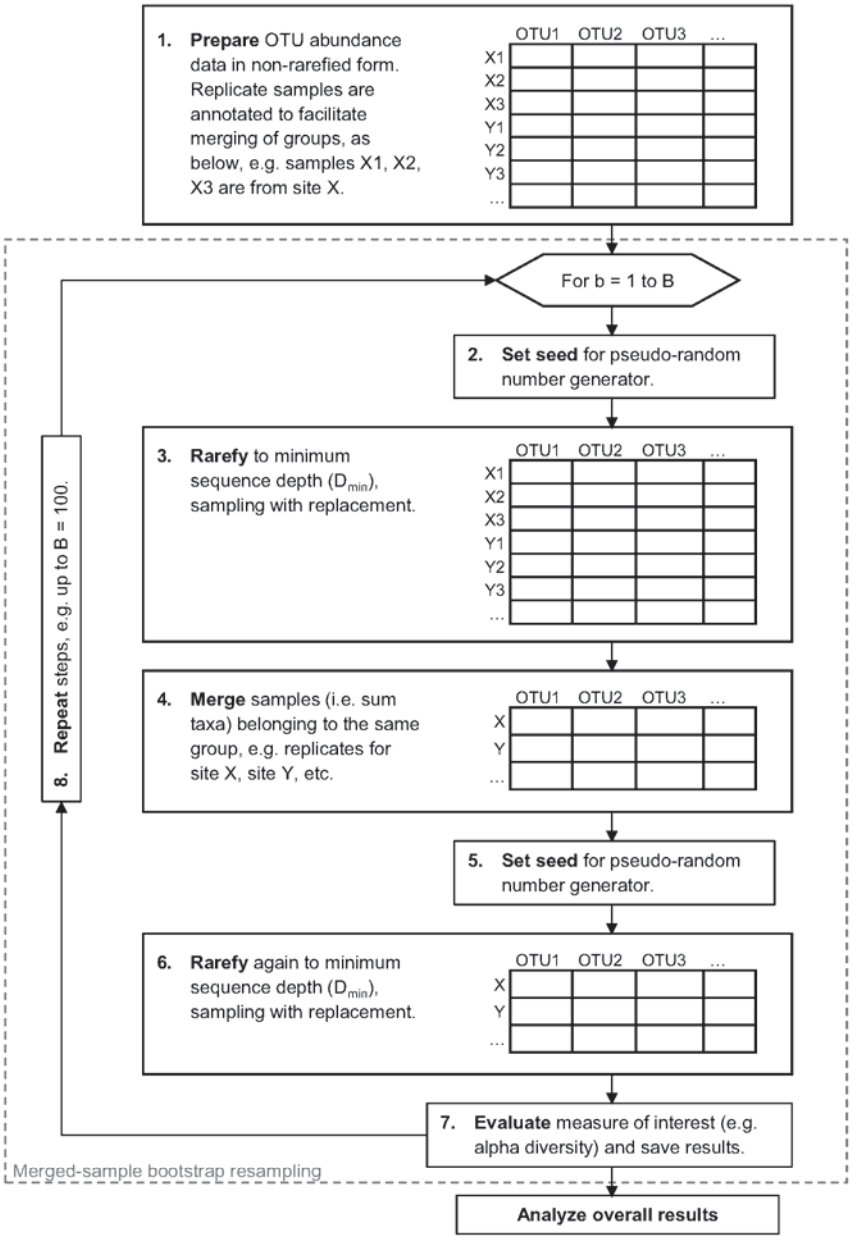


Fig. 1. Flowchart of the *merged-sample bootstrap resampling* framework.

coefficient of taxon OTU relative abundance with revegetation age from the *merged-sample bootstrap* results.

To minimize spurious results, taxa were excluded if they did not appear in at least 30% of the resamples, or if they were not represented across at least three different sample types. Significant *P*-values from the ordinal regression AOV were adjusted using the Benjamini-Hochberg method to control for false discovery rate at an alpha level of 0.05. For the trend analyses, we focussed on genera, except for unclassified genera which were aggregated to the next available classified taxonomic group. We also defined taxa that were missing in cleared or remnant samples using a threshold of 95% absence from the *merged-sample bootstrap* results. We tested for differences in the distribution of OTU alpha diversity, derived from *merged-sample bootstrap resampling*, among Mt Bold samples grouped by revegetation age using the 95% confidence interval (CI) for the mean difference between groups, based on the bootstrap results.

2.6. Analysing community shifts: Mt Bold vs. Australia-wide

DESeq2 differential abundance testing was used to estimate fold changes in OTU abundance between the Australia-wide human-altered and natural groups. The *DESeq2* algorithm internally adjusts for testing

multiple hypotheses by applying the Benjamini-Hochberg method to control for false discoveries. To test whether microbiota shifts observed at Mt Bold might be generalizable, we examined the increasing and decreasing taxa most correlated with restoration to see if similar patterns of differential abundance were found across the Australia-wide samples. We plotted correlation with revegetation age against the log2-fold-change results from differential abundance testing and highlighted two quadrants where similar patterns were evident. In the first quadrant, defined by a positive correlation with revegetation age and positive log2-fold-change between human-altered and natural groups, taxa are associated with a shift towards more natural or restored ecosystems. Conversely, in the second quadrant (negative correlation and negative log2-fold-change), taxa are associated with more altered or unnatural ecosystems. Plots of taxa correlation with revegetation age (Mt Bold) against log2-fold-change from human-altered to natural for Australia-wide data were developed for the top-trending taxa identified from Mt Bold and for selected human-associated bacterial genera identified in the literature (Baumgardner, 2012; Berg et al., 2005; Bultman et al., 2013; Jeffery and van der Putten, 2011).

To further validate results for the top-trending taxa from Mt Bold in the Australia-wide samples, we compared taxon relative abundances between human-altered and natural samples and performed one-sided Wilcoxon rank-sum tests to detect significant differences between the groups in the direction predicted by the respective trend seen at Mt Bold.

2.7. Functional assignments and case study comparisons

To explore functional similarities and differences between sample types we used PICRUSt v1.1.1, which infers functional gene abundance from 16S data (Langille et al., 2013), as implemented in the R *themetagenomics* package (https://github.com/eesi/themetagenomics; Woloszynek et al., 2017). Greengenes sequence identifiers were needed to make PICRUSt functional assignments, so we used VSEARCH (Rognes et al., 2016) to seek matches for our study OTUs (represented by fasta-format sequences) with the closest available Greengenes (13–5) fasta sequences (see Web Appendix, Methods S1). Greengenes sequences were substituted for study OTUs for use in PICRUSt only where ≥97% sequence similarity was achieved and where the Nearest Sequenced Taxon Index (NSTI) was ≤0.15. Beyond this NSTI threshold, PICRUSt user documentation suggests functional predictions will be of low quality (http://picrust.github.io/picrust/tutorials/quality_control.html). We recorded the loss of functional representation for our study OTUs where they could not be aligned to representative Greengenes sequences, and then where function could not be reliably inferred due to large phylogenetic distance (NSTI > 0.15) between representative sequences and available functionally described sequences in the PICRUSt reference genome database. When using PICRUSt we followed the developers' recommendation to employ 16S copy number correction, which aims to account for different gene copy numbers in different organisms. We expressed functional data using the Kyoto Encyclopedia of Genes and Genomes (KEGG) orthology option in PICRUSt. KEGG orthologous gene data, or orthologs, refer to equivalent genes and gene products (e.g. RNA and proteins) that arise in different organisms, and are organized under categories including: metabolism, genetic information processing, environmental information processing, cellular processes and human diseases (Kanehisa et al., 2012).

For comparison to the Mt Bold sites that are each represented by triplicate samples, functional data were prepared for three randomly selected samples from all the Australia-wide natural and human-altered sites that contained at least three samples. Functional alpha diversity was estimated using both *merged-sample bootstrap resampling* and individual rarefied sample data (based on the minimum sequence depth across Mt Bold and Australia-wide samples). We also visualized bacterial functional profiles corresponding to rarefied OTU abundance data using a heat map. Additional subsamples, representing the top 30 increasing taxa (isolated from the 10 year samples), and the top 30 decreasing taxa (isolated from the cleared samples) were included to characterize functional trends associated with the Mt Bold restoration. KEGG function relative abundance data from the three representative samples per site for all Mt Bold and Australia-wide samples were averaged to indicate site-level functional profiles. Lastly, for display in the functional heat map, we ran row (i.e. sample-wise) and column (i.e. function-wise) normalization by subtracting the mean from the observed values and then dividing by the standard deviation.

3. Results

3.1. Shifting soil microbiota with restoration vs. Australia-wide patterns

Soil bacterial communities from the Mt Bold restoration chronosequence displayed compositional shifts that were consistent with a transition between the broader groups of human-altered and natural microbiota from elsewhere in Australia (Fig. 2). The Australia-wide, human-altered and natural rarefied OTU abundance data (sequence depth 6377) showed significantly different compositional centroids (PERMANOVA based on the Bray-Curtis distance matrix: $F = 22.7$, $P = 0.001$), not due to differences in beta dispersions of the two groups ($F = 3.7$, $P = 0.062$).

We identified key trending taxa including bacterial genera and unclassified groups associated with restoration at Mt Bold (Web Appendix, Fig. S6, Table S1). To provide visual confirmation of trending taxa identified using the *merged-sample bootstrap resampling* framework and to compare with conventional rarefied data, we plotted OTU relative abundance against revegetation age for the top 10 increasing and top 10 decreasing taxa associated with restoration at Mt Bold (Web Appendix, Figs. S8–S9). We did not observe patterns that might indicate undue bias from soil conditions on restoration treatments or microbiome samples at Mt Bold (Web Appendix, Fig. S10). Eight out of the top 10 increasing taxa associated with restoration at Mt Bold (comprising *DA101*, *Candidatus Xiphinematobacter*, *Bradyrhizobium*, *Candidatus Solibacter*, *Candidatus Koribacter*, unclassified (family: *Rhodospirillaceae*), *Rhodopila*, and unclassified (family: *[Leptospirillaceae]*)), and seven out of the top 10 decreasing taxa (comprising *Bacillus*, unclassified (order: *Ellin5290*), *Sporosarcina*, unclassified (family: *Ellin5301*), *Ammoniphilus*, *Flavisolibacter*, unclassified (class: *C0119*)), showed consistent trends when comparing OTU differential abundance from the human-altered to natural samples elsewhere in Australia (Table 1; Web Appendix, Figs. S11–S12). The Australia-wide human-altered and natural samples separate into two distinguishable clusters when mapped in two dimensions corresponding to the cumulative OTU relative abundance of the top 10 increasing and top 10 decreasing taxa that trend with restoration at Mt Bold (Fig. 3; PERMANOVA on Euclidean distances: $F = 119.33$, $P = 0.001$; although dispersions of the two groups were different: $F = 9.14$, $P = 0.003$). Eight taxa were missing from cleared samples, increased in OTU relative abundance with restoration, and were found in remnants (Web Appendix, Table S2); while 14 taxa were found in cleared samples, decreased with restoration, and were missing in remnants at Mt Bold (Web Appendix, Table S3).

A range of taxa displayed outlying differential abundance in the Australia-wide human-altered vs. natural samples (Web Appendix, Figs. S13–S14; the top 30 increasing and top 30 decreasing taxa based on fold change between human-altered to natural samples are listed in Web Appendix, Table S4). However, the top-trending taxa with restoration at Mt Bold show consistent differential abundance patterns in the Australia-wide data (Web Appendix, Fig. S13, Fig. S15). This association is strongest for the top few taxa most correlated with revegetation age at Mt Bold, and declined as more taxa were considered (Web Appendix, Fig. S15).

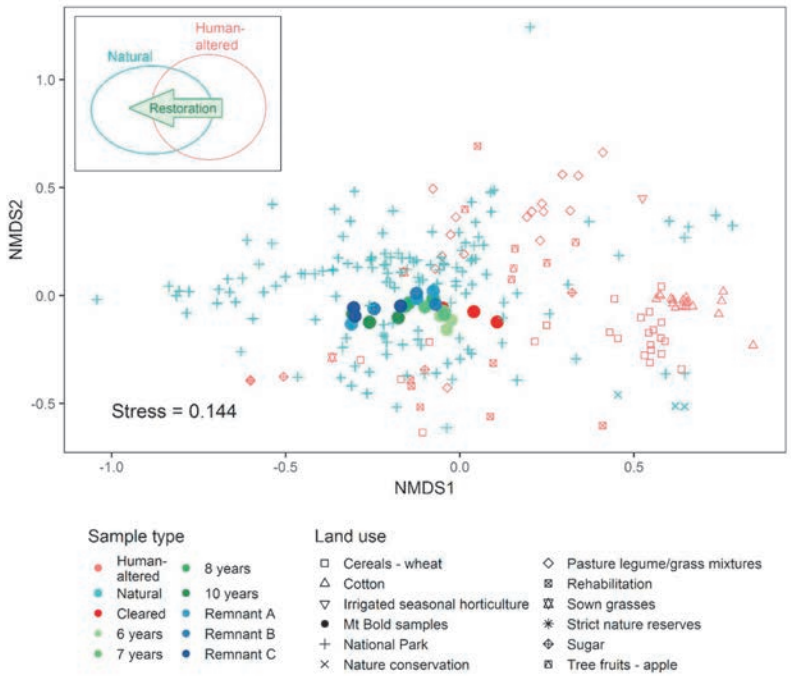


Fig. 2. Surface soil bacterial communities from the Mt Bold restoration chronosequence (large filled circles of red, greens, and blues; n = 24) traverse the broader groupings of natural (cyan, n = 139) and human-altered (pale red, n = 78) bacterial communities from elsewhere in Australia, as stylized in the inset. Microbiota visualization is based on rarefied (sequence depth = 6377) bacterial 16S OTU abundance of taxa aggregated at the genus or next available classified level, using NMDS ordination of Bray-Curtis distances. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.2. Patterns in human-associated bacteria

Human-associated genera that increased with restoration at Mt Bold and had higher differential abundance in natural samples included *Actinomadura*, *Burkholderia*, and *Mycobacterium*. Genera that decreased with restoration and had higher differential abundance in human-altered samples included *Achromobacter*, *Bacillus*, *Bacteroides*, *Chryseobacterium*, *Clostridium*, *Enterobacter*, *Flavobacterium*, *Legionella*, *Pseudomonas*, *Rhodococcus*, *Sphingobacterium* and *Streptomyces* (Fig. 4; Web Appendix, Table S5, Figs. S16–S17).

3.3. Alpha diversity, functional diversity, and functional clustering

There was no apparent trend in OTU alpha diversity with restoration at Mt Bold when only considering the rarefied OTU abundance data. However, a different signal emerged when using site-level density distributions from the *merged-sample bootstrap* results (Web Appendix, Fig. S18a). We observed a significant rising trend in OTU alpha diversity with restoration at Mt Bold as determined from 95% CIs of differences between mean bootstrap results across groupings of revegetation age (Fig. 5). Estimated functional alpha diversity was, however, highest for cleared samples and remained steady or declined with revegetation age, although only 20–30% of OTUs were represented when inferring functions (Web Appendix, Fig. S18b–c). We observed no apparent trends in either OTU alpha diversity or functional alpha diversity for the Australia-wide human-altered vs. natural case study samples considering both rarefied OTU abundance data and the *merged-sample bootstrap* density plots (Web Appendix, Fig. S19); nor was there any relationship apparent between OTU alpha diversity and functional alpha diversity in the Australia-wide data. Only 10–30% of OTUs were represented when inferring functions (Web Appendix, Fig. S19c).

Functional profiles for the top 30 increasing taxa with restoration at Mt Bold clustered together with the majority of other Mt Bold samples, except for the cleared samples whose mean functional profile was most different to the other samples considered (Fig. 6). The top 30 decreasing taxa with restoration at Mt Bold clustered with human-altered samples from annual croplands of cotton, sugar and wheat; although this cluster also contained samples from two national parks. The majority of four national parks clustered together with human-altered samples from perennial land uses including horticulture (apples) and pastures.

4. Discussion

4.1. Bacterial indicators of restoration

Our results broadly indicate a core microbiome that shifts with ecological restoration. We demonstrate support for the relative abundance patterns in the majority of top 10 increasing and top 10 decreasing, and human-associated taxa identified from Mt Bold, by showing consistent differential abundance patterns within human-altered and natural sites across Australia. We might expect restoration to bring increased stability, perenniality and diversity of aboveground plants, and increased rhizosphere interactions involving different plant species and combinations of root exudates. Soils under restoration would also accumulate organic matter including plant debris and fungal hyphae networks due to reduced disturbance. With increasing complexity and stability of microbial habitats and feedstocks (Adams and Wall, 2000), it is unsurprising that soil microbiota would shift, in comparison to highly disturbed or regularly cleared land. In disturbed environments, where microbial habitats and feedstocks may undergo large fluctuations and possibly collapse, there is potential for fast-growing, adaptable environmental opportunists to thrive as vacant ecological niches arise. Broad shifts in soil microbiota composition with restoration have been previously identified (Gellie et al., 2017a), however, here we reveal in fine taxonomic resolution the key genera that are increasing and decreasing in these shifting communities.

The eight taxa we observed from the top 10 increasing with restoration at Mt Bold, that also showed higher differential abundance in natural samples Australia-wide, include K-selected organisms associated with stable, late-successional ecosystems (MacArthur and Wilson, 1967) and organisms with particular resource requirements.

Table 1
Summary statistics for the top 10 increasing and top 10 decreasing taxa associated with restoration at Mt Bold, with corresponding differential abundance data in human-altered samples (n = 78) vs. natural samples (n = 139) elsewhere in Australia.

Genus or group	Phylum	Mt Bold correlation with revegetation age: Mean (95% CI)	Australia-wide Mean log2-fold-change in OTU abundance: human-altered to natural	Australia-wide human-altered samples OTU relative abundance: Median (IQR) (%)	Australia-wide natural samples OTU relative abundance: Median (IQR) (%)	Australia-wide trend between human-altered vs. natural	Wilcoxon one-sided rank-sum test *P*-value[a]
Top 10 increasing (Mt Bold)							
DA101	Verrucomicrobia	0.95 (0.94, 0.96)	1.65	0.122 (0.008, 0.801)	1.188 (0.371, 3.544)	Increasing	< 0.001
Candidatus_Xiphinematobacter	Verrucomicrobia	0.94 (0.91, 0.95)	2.36	0.01 (0, 0.121)	0.856 (0.227, 1.946)	Increasing	< 0.001
Bradyrhizobium	Proteobacteria	0.92 (0.88, 0.95)	1.34	1.237 (0.386, 2.969)	5.305 (3.224, 7.954)	Increasing	< 0.001
Candidatus_Solibacter	Acidobacteria	0.9 (0.87, 0.94)	0.59	0.89 (0.558, 2.202)	2.102 (1.567, 2.545)	Increasing	< 0.001
Candidatus_Koribacter	Acidobacteria	0.88 (0.84, 0.92)	0.31	0.689 (0.215, 2.137)	1.401 (0.929, 2.373)	Increasing	< 0.001
Unclassified (family: Rhodospirillaceae)	Proteobacteria	0.87 (0.85, 0.9)	0.77	1.349 (0.981, 1.587)	2.689 (2.149, 3.648)	Increasing	< 0.001
Rhodopila	Proteobacteria	0.79 (0.49, 0.94)	0.96	0.002 (0, 0.019)	0.026 (0.003, 0.139)	Increasing	< 0.001
Edaphobacter	Acidobacteria	0.77 (0.58, 0.91)	0.08	0 (0, 0.001)	0 (0, 0.002)	NS	0.116
Unclassified (order: Solibacterales)	Acidobacteria	0.77 (0.72, 0.82)	−0.42	0.799 (0.506, 1.2)	0.569 (0.399, 0.783)	NS	1
Unclassified (family: [Leptospirillaceae])	Nitrospirae	0.76 (0.41, 0.91)	1.57	0 (0, 0)	0.008 (0, 0.028)	Increasing	< 0.001
Top 10 decreasing (Mt Bold)							
Bacillus	Firmicutes	−0.95 (−0.95, −0.94)	−1.04	4.499 (2.622, 9.865)	1.706 (0.52, 2.956)	Decreasing	< 0.001
Rummeliibacillus	Firmicutes	−0.87 (−0.92, −0.82)	−0.82	0 (0, 0.019)	0 (0, 0.009)	NS	0.051
Unclassified (family: Actinospicaceae)	Actinobacteria	−0.86 (−0.92, −0.8)	0.32	0 (0, 0.02)	0.01 (0, 0.039)	NS	0.997
Unclassified (order: Ellin5290)	Gemmatimonadetes	−0.86 (−0.89, −0.82)	−1.12	0.465 (0.199, 0.75)	0.298 (0.123, 0.545)	Decreasing	0.001
Sporosarcina	Firmicutes	−0.85 (−0.87, −0.84)	−4.41	0.323 (0.019, 1.449)	0.009 (0, 0.07)	Decreasing	< 0.001
Cytophagales	Bacteroidetes	−0.85 (−0.9, −0.81)	1.7	0 (0, 0.001)	0.004 (0, 0.025)	NS	1
unclassified (family: Ellin5301)	Gemmatimonadetes	−0.85 (−0.94, −0.72)	−2.73	0.063 (0.014, 0.211)	0.012 (0, 0.048)	Decreasing	< 0.001
Ammoniphilus	Firmicutes	−0.85 (−0.88, −0.83)	−1.27	0.067 (0.016, 0.409)	0.006 (0, 0.03)	Decreasing	< 0.001
Flavisolibacter	Bacteroidetes	−0.85 (−0.92, −0.75)	−2.53	0.21 (0.044, 1.001)	0.017 (0.001, 0.103)	Decreasing	< 0.001
Unclassified (class: C0119)	Chloroflexi	−0.83 (−0.86, −0.79)	−1.11	0.129 (0.067, 0.462)	0.031 (0.004, 0.127)	Decreasing	< 0.001

[a] Wilcoxon one-sided rank sum tests for difference in OTU relative abundance between Australia-wide human-altered vs. natural samples, set in the direction suggested by results from restoration at Mt Bold. NS = no significant difference.

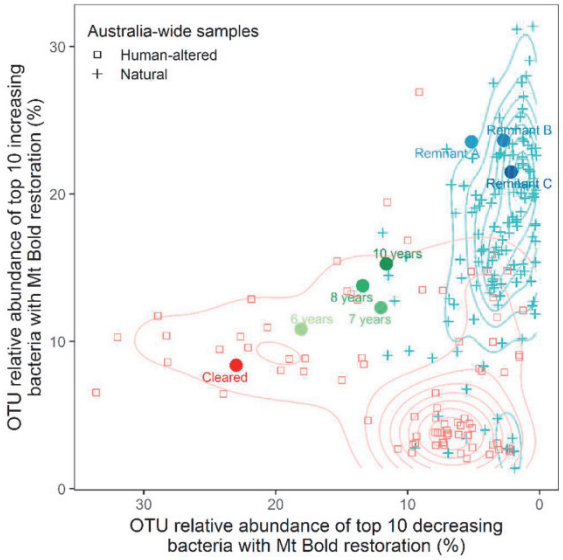


Fig. 3. Two-dimensional mapping of ecosystem condition based on soil bacterial 16S data. Natural (cyan crosses, n = 139) and human-altered (pale red squares, n = 78) samples from across Australia form distinguishable groups when characterized by the cumulative OTU relative abundance of the top 10 increasing (y-axis) and top 10 decreasing (x-axis) bacterial taxa that trend with restoration at Mt Bold. Kernel density contours emphasize the weight of samples for each group. Mean site-level data for Mt Bold (annotated circles; n = 8) indicate a shift in key microbiota from disturbance-adapted opportunistic taxa (lower left of plot) to mature niche-adapted taxa (upper right of plot). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

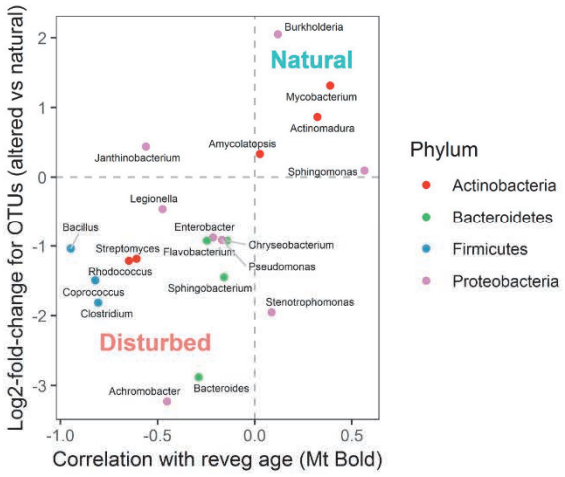


Fig. 4. Mean correlation of human-associated taxa OTU relative abundance with revegetation age from Mt Bold (x-axis) plotted against the mean differential OTU abundance between human-altered and natural samples elsewhere in Australia. The quadrant labelled 'Natural' corresponds to taxa that increase with restoration and have higher differential abundance in natural sites, while the quadrant labelled 'Disturbed' corresponds to taxa that decrease with restoration and have higher differential abundance in human-altered sites.

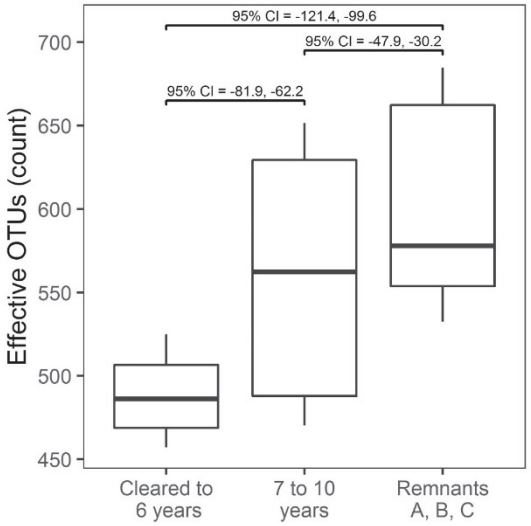


Fig. 5. Boxplots of OTU alpha diversity for Mt Bold sites derived from *merged-sample bootstrap resampling* (B = 100) then grouping by revegetation age (*Cleared to 6 years*, n = 200; *7 to 10 years*, n = 300; *Remnants A, B, C*, n = 300). Bootstrap statistics for the mean difference in effective OTU count between groups are: *Cleared to 6 years – 7 to 10 years*, mean Δ = −72.3 (95% CI = −81.9, −62.2); *7 to 10 years – Remnants A, B, C*, mean Δ = −38.6 (95% CI = −47.9, −30.2); *Cleared to 6 years – Remnants A, B, C*, mean Δ = −110.9 (95% CI = −121.4, −99.6), as illustrated.

DA101 is one of the most abundant bacteria found in soil, particularly in grasslands (Brewer et al., 2016). *Candidatus Xiphinematobacter* species are maternally-transmitted endosymbionts of *Xiphinema americanum*-group nematodes, which have cosmopolitan distribution (Archidona-Yuste et al., 2016) and field data suggest these nematodes are K-selected organisms with a long lifespan and low reproduction rate (Jaffee et al., 1987). *Bradyrhizobium* is another ubiquitous and abundant genus of bacteria that dominates forest soils (VanInsberghe et al., 2015) and tends not to overlap where *DA101* dominates (Brewer et al., 2016). Genomic investigation of *Candidatus Solibacter* and *Candidatus Koribacter* suggest that these organisms are able to decompose complex substrates such as plant litter in soils and are slow-growing with a K-selected lifestyle (Ward et al., 2009). The family *Rhodospirillaceae* contains genera mostly with a preferred photoheterotrophic growth mode under anoxic conditions in light, or they grow chemotrophically in the dark (Garrity et al., 2005). *Rhodopila* is an acid-loving member of the *Rhodospirillaceae* family. Photoheterotrophs cannot use carbon dioxide as their sole carbon source, so they also use organic compounds from the environment. The character of unclassified taxa from the family *Leptospirillaceae* is uncertain; this family appears to be renamed as *Nitrospiraceae*, and contains iron-oxidising *Leptospirillum* (Hippe, 2000) and nitrate-oxidising *Nitrospira* (Watson et al., 1986).

The seven taxa from the top 10 decreasing with restoration at Mt Bold, that also showed higher differential abundance in human-altered samples Australia-wide, include fast-growing and opportunistic species. *Bacillus* species are ubiquitous in nature, are often fast-growing, live in aerobic or anaerobic conditions, and are capable of forming resistant endospores to survive stressful environmental conditions for long periods of time. They include medically significant *B. anthracis* (anthrax) and *B. cereus* associated with food spoilage and poisoning, as well as many normally harmless species. Unclassified taxa in the order

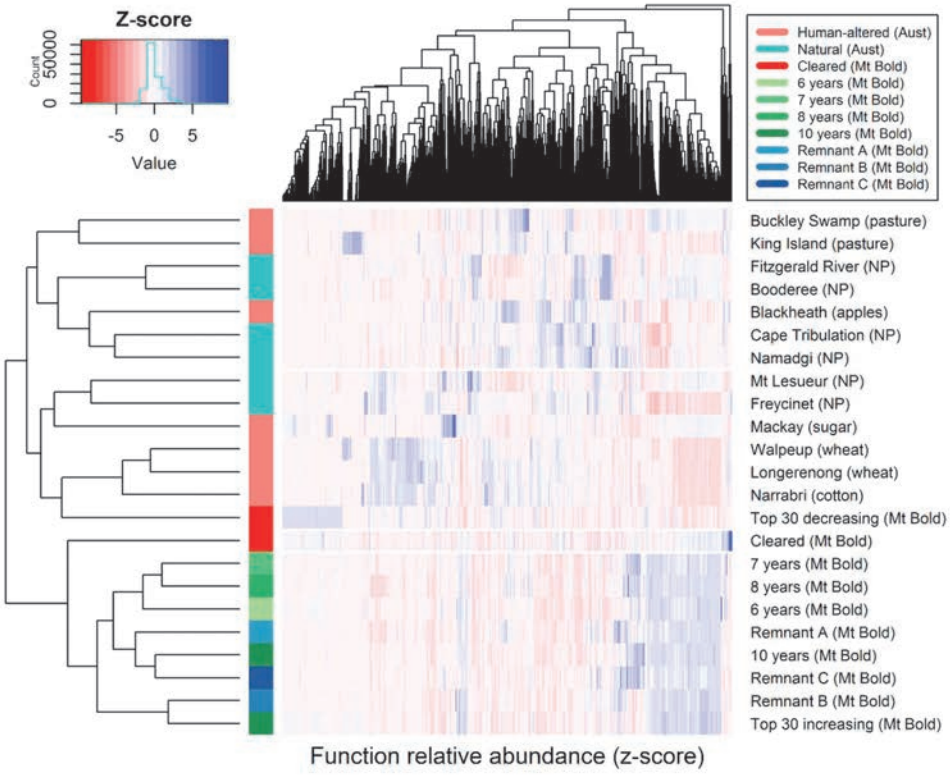


Fig. 6. Inferred microbial function relative abundance based on site (row) and function (column) z-scores, derived from rarefied OTU abundance data (sequence depth = 16,704). The method used to infer gene functions could only represent 20–30% of sequences from the rarefied sample microbiota data. Row groupings highlight sites with similar functional profiles based on second-level branching of the row dendrogram. Row side colors indicate sample types. This plot illustrates similarities in functional profiles—i.e. columns display variation in 5434 orthologous genes—however it is beyond the scope of this study to examine these functions in detail. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Ellin5290, and in the family *Ellin5301*, are all from the phylum *Gemmatimonadetes* which display cosmopolitan distribution and versatile, generalist ecological strategies, adapting to a wide variety of environments including low moisture conditions (DeBruyn et al., 2011). With similar features to *Bacillus*, *Sporosarcina* are facultatively anaerobic or strictly aerobic heterotrophs, capable of forming endospores that can persist in the environment (Yoon et al., 2001). *Ammoniphilus* species are aerobic, spore-forming bacteria, dedicated to the use of oxalate (a common constituent of fresh plant tissues; Libert and Franceschi, 1987) for carbon and energy (De Vos et al., 2009). *Flavisolibacter* species are aerobic, non-motile, non-spore-forming rods, which appear to be environmental opportunists, having been isolated from various soils, fresh water, and a biofilm coating parts of an automotive air conditioning system (Kim et al., 2018). No information was available on the character of unclassified taxa from the class *C0119* (phylum *Chloroflexi*).

We suggest that microbial indicators of ecosystem condition should be detectable across a range of environments and provide a meaningful association with a limited range of organisms. Also, we expect that the choice of taxonomic rank for such indicators will represent a trade-off between sensitivity to detect effects, and specificity to particular organisms. Our bacterial 16S data are appropriate for genus-level observations (Fox et al., 1992) and reflect a common and cost-effective mode of environmental microbiome survey data. Using correlation coefficients as a measure of effect size (so that our results emphasize the tightness of relationship with restoration) enabled us to identify trends in abundant and rare taxa alike. This approach acknowledges that rare taxa can play important ecological roles (Hol et al., 2010). Our two-dimensional mapping of soil microbiome data, with axes highlighting proportions of opportunistic vs. mature niche-adapted taxa, may have use for ecosystem monitoring and management, as discussed further below.

4.2. Patterns in alpha diversity and function

We observed an increasing trend in alpha diversity with restoration at Mt Bold facilitated by the *merged-sample bootstrap resampling* approach. However, microbiome data relating to restoration treatments were not available for any other samples considered in our study. Considering just the one-off microbiome survey data for natural vs. human-altered samples, we found no generalizable patterns in OTU alpha diversity or functional alpha diversity between the two groups. In other words, the variability in OTU and functional alpha diversity was largely driven by site-specific factors rather than whether the samples were classed as natural or human-altered. On the other hand, coherent patterns in functional profiles did emerge. The top 30 increasing taxa clustered with the majority of Mt Bold samples, perhaps reflecting local adaptation. Perennial agriculture and the majority of natural samples clustered together, possibly reflecting generally more stable and mature soil ecosystems. Samples from annual crops (i.e. the most disturbed soils) also clustered together. The functional data also suggest not all natural samples behave the same, as two national park samples clustered with annual crop samples.

4.3. Implications for ecosystem restoration and management

We propose a two-dimensional index of ecosystem condition based on soil bacterial 16S survey data, where soil eDNA samples might be used to map study sites into zones ranging from degraded ecosystems, through intermediate stages of restoration, to mature restored and reference ecosystems. Such an index could have value in tracking the progress of restoration activities, assessing the condition of land, and prioritising areas for restoration investments. The axes of Fig. 3 reflect shifting proportions of often fast-growing, environmental generalists and opportunists (x-axis) vs. more stable niche-holders adapted to mature and reference ecosystems (y-axis). We offer a first-cut approach relevant to our temperate grassy woodland ecosystem which may be improved by considering trending taxa from a wider diversity of environments. For example, similar microbiome datasets could be collated from restoration sites (e.g. chronosequences like Mt Bold) across more dissimilar environments with the objective to build a universal indicator set. Alternatively, locally representative trajectories of restoration or land improvement, for a given soil type and land use, could provide a localized frame of reference for bacterial shifts in particular ecosystems (e.g. low to medium to high sustainable production agricultural land), against which neighbouring land could be evaluated.

Interestingly, we observed some natural environments can have characteristics similar to disturbed and developing ecosystems, while a minority of human-altered environments can resemble mature reference ecosystems. Presumably, human-altered environments that are close to mimicking natural systems would include stable, perennial land cover, as suggested by the functional clustering of many natural samples with perennial horticulture and pastures. On the other hand, natural environments with low in-situ build-up of organic matter (e.g. due to sandy or infertile soils) and therefore stunted soil biological activity and buffering capacity, might cause the soil microbiota to resemble a developing or degraded ecosystem. We acknowledge that soil microbiota may not always be representative of current land use and management. It is possible for soils to have legacy influences, and soil microbiota may undergo multi-decadal shifts as microbial habitats and feedstocks move towards new equilibria following land use change (Bell and Lawrence, 2009). Issues such as grazing pressure and isolation can also cause native vegetation-based ecosystems to not necessarily function in a natural or reference state.

Biological indicators are used elsewhere in ecosystem monitoring (Stanford and Spacie, 1994) due to the ability of organisms to reflect cumulative or integrative responses across physical and chemical parameters that may undergo short-term fluctuations. That is, physicochemical monitoring often cannot fully represent the condition of ecosystems without expensive multi-parameter, fine temporal resolution monitoring. Additionally, soil inoculation can be a powerful tool to help facilitate restoration of disturbed terrestrial ecosystems and steer plant community development (Wubs et al., 2016). Therefore identifying beneficial microbial signatures may help in targeting soil microbiota for inoculations or recognizing sites with inherent remediation potential.

4.4. Implications for microbiota-mediated human health

Our results suggest that there is potential for disturbed soils to harbour environmental opportunists with potential pathogenic character. We found a number of often fast-growing, environmental opportunists that were associated with human-altered soil samples and inversely correlated with restoration. Notable environmental taxa with importance for human health in this category included the genera *Bacillus*, *Clostridium*, *Enterobacter*, *Legionella* and *Pseudomonas* which include opportunistic pathogens, often associated with nosocomial infections. Through direct contact or wind-blown aerobiology these soil-borne bacteria may impact susceptible individuals, neighbouring buildings, and even highly sterilized (i.e. ecologically vacant) health care facilities such as hospitals. Aerobiological access is likely via high traffic areas for patients, staff and visitors. Such a prospect is of particular concern where environmental and/or exposed human microbiomes suffer declining microbial diversity, thus increasing the potential for pathogenic microbes from the environment to overpower the defence mechanisms of resident environmental and/or exposed human microbiota. That is, pathogenicity should be considered in the context of the host-microbe system, as discussed later. To illustrate this possibility, environmental sources were suspected as the most likely origin in a recent *Escherichia coli* O55:H7 outbreak in Dorset England (Public Health England, 2017). Strains of *E. coli* can become naturalized to live in soils (Ishii and Sadowsky, 2008). It is also concerning that environmental opportunists such as *Klebsiella pneumoniae* may be implicated in spreading anti-microbial resistance genes to clinically-important pathogens (Wyres and Holt, 2018). However, it may be possible to reduce the threat from opportunistic potential environmental pathogens through ecological restoration and preserving and enhancing natural and biodiverse vegetation in urban areas. It may also be possible to reduce the risk of hospital-acquired infections, which may in part be due to airborne opportunistic microbes from environmental sources. Although not all natural areas will be the same, nor have the same immune-priming attributes, we expect that the maintenance of nature-based microbial diversity, together with more slow-growing microbiota adapted to restored and reference ecosystems, will have population health benefits through building immune fitness and suppressing opportunistic pathogens.

Our analyses show that presumptions should not be made about the level of microbial diversity, nor functional diversity, for natural sites in comparison to human-altered sites, nor generalizations about sites of the same class, based on single-time-point eDNA surveys. Variation in 16S OTU alpha diversity and functional alpha diversity appear to be driven by site-specific factors, overshadowing any effect linked to the natural or human-altered classification. On the other hand, where we had eDNA survey data characterizing restoration from cleared (human-altered) land towards a natural state at the single location of Mt Bold, we saw an increase in microbial diversity (although not functional diversity). We did not have further data available to apply our new methods more widely and test whether restoring native biodiversity might lead to increased microbial diversity within particular locations elsewhere. In short, our study provides early supporting evidence that restoring a more biodiverse ecosystem may boost the microbial diversity within that location, which should then become available for beneficial immunomodulation (Mills et al., 2017).

It is beyond the scope of our study to identify microbial Old Friends, however we highlight example taxa that associate with ecological restoration and natural reference soil microbiomes vs. human-altered or disturbed soil microbiomes (Table 1, Fig. 4; Web Appendix, Tables S1–S5). These results offer potential targets for subsequent research into possible immunomodulatory capabilities. From the human-associated taxa that associated with natural samples and restoration, we observed the often slow-growing *Mycobacterium*. This genus includes the common non-pathogenic soil-dwelling *M. vaccae* which has been associated with reduced anxiety-like behaviour in mice (Matthews and Jenks, 2013). Although, *Mycobacterium* also contains pathogenic species associated with infrequent but serious diseases including tuberculosis (*M. tuberculosis*) and leprosy (*M. leprae*). We note that new conceptualisations of what makes a pathogen are relevant to this discussion. Casadevall and Pirofski (2000) suggest the definition of a pathogen should be based on the potential for host damage arising from the host-microbe relationship, and not attributes of the microbe alone. Examples where pathogenicity depends on both the microbe and host environment include cell-to-cell signalling, termed quorum sensing, where high cell densities of genetically-related bacteria can produce a positive feedback and self-promotion of growth factors and higher virulence (Rumbaugh et al., 2012). Conversely, through ecological mechanisms, greater microbial diversity in soils can resist the invasion

and establishment of potentially pathogenic species (van Elsas et al., 2012). In earlier work (Liddicoat et al., 2016), we sought to unite the Biodiversity Hypothesis and microbial Old Friends concepts, suggesting that optimum beneficial immunomodulation from a particular Old Friend is likely to occur in the context of microbial diversity, required to keep any potential pathogenic behaviour in check.

4.5. Limitations

Our analysis of bacterial indicators reflected a limited microbiome dataset associated with restoration towards a grassy woodland ecosystem. Therefore, the top trending taxa we identified may not be applicable for very different environments. Also, as we analysed genera, our study cannot inform the implications for particular and often rare taxa. Instead, our study helps to build understanding of the distribution patterns of larger groups of related organisms that may share similar ecological niches. It is speculative to suggest that increased abundance of genera such as *Bacillus*, *Clostridium*, *Enterobacter*, *Legionella* and *Pseudomonas* may contribute to increased rates of human disease. Even so, it is informative to appreciate that increased exposure to these taxa is likely in disturbed environments, and as a consequence there may be an increased likelihood of health impacts in susceptible individuals due to their generally fast-growing and opportunistic nature (Benenson and APHA, 1995; Ristuccia and Cunha, 1985).

Our study lacked data on the actual biomass of environmental microbiota, as well as actual human exposures. Also, amplicon-based OTU abundance data are subject to taxon-specific biases (e.g. during DNA extraction and polymerase chain reaction amplification of DNA). However, despite such biases, using relative sequence abundance information, as we have done, has potential to provide more accurate insights to actual biomass proportions compared to alternative analyses using presence-absence only data (Deagle et al., 2018). Although eDNA analyses do not permit true insight into actual bacterial exposures or biomass, the increasing knowledge we gain into microbiota diversity and the relative abundance of key taxa in different environments is relevant to human exposures and microbial ecological interactions between environmental and host microbiotas.

We experienced limitations using PICRUSt v1.1.1 to infer functional profiles for the different environmental microbiomes, in the conversion of our study OTUs into Greengenes (13–5) sequence identifiers, and in finding suitable nearest available taxa in the PICRUSt database. We used copy number correction only for the functional analysis within the PICRUSt software as this followed developers' recommendations. However, in the remainder of our microbiome data analyses we did not seek to correct for gene copy numbers, as this is routinely not included elsewhere due to a lack of available knowledge (Louca et al., 2018). Despite these issues, we believe our coherent findings from the analysis of functional differences between the different soil communities has provided useful insight. We acknowledge that attributing functions to 16S data and refining microbiome data analyses methods more broadly represent areas of active research and development.

4.6. Conclusions

Using our new *merged-sample bootstrap resampling* framework, which preserves microbiome data and permits more detailed study of site-level information, we discovered emergent patterns that were not apparent using conventional one-off rarefying of individual samples. Specifically, we found patterns in OTU relative abundance and function for ecologically-relevant and human-associated key trending soil bacteria from a localized restoration chronosequence at Mt Bold (South Australia), which aligned with differences between human-altered and natural soil microbiome samples from across Australia. Our results help build knowledge towards using microbial indicators from soil eDNA to assess, manage and restore ecosystem condition. We also show that environmental opportunists, including potential human pathogens, increase in OTU relative abundance in disturbed ecosystems; a finding that may have important implications for microbiota-mediated human health and possible new ecologically-based health improvement and pathogen control interventions.

Acknowledgement and funding information

We acknowledge the contribution of the BASE project partners (DOI: https://doi.org/10.4227/71/561c9bc670099), an initiative supported by Bioplatforms Australia with funds provided by the Australian Commonwealth Government through the National Collaborative Research Infrastructure Strategy. We thank David Phillips, Director of Public Health at Public Health Dorset for helpful discussions and suggestions. Craig Liddicoat received support while undertaking this study through an Australian Government Research Training Program Scholarship. Martin Breed is supported by the Australian Research Council and Australian Smart Cities Consortium. We also thank the two anonymous reviewers whose comments have helped improve the manuscript.

Declaration of Competing Interest

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envint.2019.05.011.

References

ABARES, 2016. The Australian Land Use and Management Classification Version 8. Australian Bureau of Agricultural and Resource Economics and Sciences, Canberra.
Adams, G.A., Wall, D.H., 2000. Biodiversity above and below the surface of soils and sediments: linkages and implications for global change. (Cover story). BioScience 50, 1043–1048.
Anderson, M.J., 2017. Permutational Multivariate Analysis of Variance (PERMANOVA). In: Balakrishnan T.C., N., Everitt, B., Piegorsch, W., Ruggeri, F., Teugels, J.L. (Eds.), Wiley StatsRef: Statistics Reference Online.
Archidona-Yuste, A., Navas-Cortés, J.A., Cantalapiedra-Navarrete, C., Palomares-Rius, J.E., Castillo, P., 2016. Cryptic diversity and species delimitation in the Xiphinema americanum-group complex (Nematoda: Longidoridae) as inferred from morphometrics and molecular markers. Zool. J. Linnean Soc. 176, 231–265.
Baumgardner, D.J., 2012. Soil-related bacterial and fungal infections. The Journal of the American Board of Family Medicine 25, 734–744.
Bell, M., Lawrence, D., 2009. Soil carbon sequestration - myths and mysteries. Tropical Grasslands 43, 227–231.
Benenson, A.S., American Public Health Association, 1995. Control of Communicable Diseases Manual: An Official Report of the American Public Health Association, 16th ed. American Public Health Association, Washington, D.C.
Berg, G., Eberl, L., Hartmann, A., 2005. The rhizosphere as a reservoir for opportunistic human pathogenic bacteria. Environ. Microbiol. 7, 1673–1685.
Bissett, A., Fitzgerald, A., Meintjes, T., Mele, P.M., Reith, F., Dennis, P.G., Breed, M.F., Brown, B., Brown, M.V., Brugger, J., Byrne, M., Caddy-Retalic, S., Carmody, B., Coates, D.J., Correa, C., Ferrari, B.C., Gupta, V.V.S.R., Hamonts, K., Haslem, A., Hugenholtz, P., Karan, M., Koval, J., Lowe, A.J., Macdonald, S., McGrath, L., Martin, D., Morgan, M., North, K.I., Paungfoo-Lonhienne, C., Pendall, E., Phillips, L., Pirzl, R., Powell, J.R., Ragan, M.A., Schmidt, S., Seymour, N., Snape, I., Stephen, J.R., Stevens, M., Tinning, M., Williams, K., Yeoh, Y.K., Zammit, C.M., Young, A., 2016. Introducing BASE: the Biomes of Australian Soil Environments soil microbial diversity database. GigaScience 5, 21.
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, 601–612.
Bowers, R.M., Clements, N., Emerson, J.B., Wiedinmyer, C., Hannigan, M.P., Fierer, N., 2013. Seasonal variability in bacterial and fungal diversity of the near-surface atmosphere. Environmental Science & Technology 47, 12097–12106.
Brewer, T.E., Handley, K.M., Carini, P., Gilbert, J.A., Fierer, N., 2016. Genome reduction in an abundant and ubiquitous soil bacterium 'Candidatus Udaeobacter copiosus'. Nat. Microbiol. 2, 16198.
Bulgarelli, D., Schlaeppi, K., Spaepen, S., Themaat, E.V.L.v., Schulze-Lefert, P., 2013. Structure and functions of the bacterial microbiota of plants. Annu. Rev. Plant Biol. 64, 807–838.
Bultman, M.W., Fisher, F.S., Pappagianis, D., 2013. The ecology of soil-borne human pathogens. In: Selinus, O. (Ed.), Essentials of Medical Geology, revised edition. Springer Netherlands, Dordrecht.

Casadevall, A., Pirofski, L.A., 2000. Host-pathogen interactions: basic concepts of microbial commensalism, colonization, infection, and disease. Infect. Immun. 68, 6511–6518.
De Vos, P., Garrity, G.M., Jones, D., Krieg, N.R., Ludwig, W., Rainey, F.A., Schleifer, K.-H., Whitman, W.B., 2009. Bergey's Manual of Systematic Bacteriology Vol. 3 The Firmicutes. Springer.
Deagle, B.E., Thomas, A.C., McInnes, J.C., Clarke, L.J., Vesterinen, E.J., Clare, E.L., Kartzinel, T.R., Eveson, J.P., 2018. Counting with DNA in metabarcoding studies: how should we convert sequence reads to dietary data? Mol. Ecol. https://doi.org/10.1111/mec.14734.
DeBruyn, J.M., Nixon, L.T., Fawaz, M.N., Johnson, A.M., Radosevich, M., 2011. Global biogeography and quantitative seasonal dynamics of Gemmatimonadetes in soil. Appl. Environ. Microbiol. 77, 6295–6300.
Delgado-Baquerizo, M., Powell, J.R., Hamonts, K., Reith, F., Mele, P., Brown, M.V., Dennis, P.G., Ferrari, B.C., Fitzgerald, A., Young, A., Singh, B.K., Bissett, A., 2017. Circular linkages between soil biodiversity, fertility and plant productivity are limited to topsoil at the continental scale. New Phytol. 215, 1186–1196.
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, 583–596.
van Elsas, J.D., Chiurazzi, M., Mallon, C.A., Elhottovā, D., Krištůfek, V., Salles, J.F., 2012. Microbial diversity determines the invasion of soil by a bacterial pathogen. Proc. Natl. Acad. Sci. 109, 1159–1164.
Flandroy, L., Poutahidis, T., Berg, G., Clarke, G., Dao, M.-C., Decaestecker, E., Furman, E., Haahtela, T., Massart, S., Plovier, H., Sanz, Y., Rook, G., 2018. The impact of human activities and lifestyles on the interlinked microbiota and health of humans and of ecosystems. Sci. Total Environ. 627, 1018–1038.
Fox, G.E., Wisotzkey, J.D., Jurtshuk, P., 1992. How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. Int. J. Syst. Evol. Microbiol. 42, 166–170.
Garrity, G.M., Brenner, D.J., Krieg, N.R., Staley, J.T., 2005. Bergey's manual of systematic bacteriology. In: The Proteobacteria Part C: The Alpha-, Beta-, Delta-, and Epsilonproteabacteria, 2nd ed. 2 Springer.
Gellie, N.J.C., Mills, J.G., Breed, M.F., Lowe, A.J., 2017a. Revegetation rewilds the soil bacterial microbiome of an old field. Mol. Ecol. 26, 2895–2904.
Gellie, N.J.C., Mills, J.G., Breed, M.F., Lowe, A.J., 2017b. Revegetation Rewilds the Soil Bacterial Microbiome of an Old Field. Part 1: OTU Raw Data Matrix, Version 1.0. AEKOS Data Portal. https://doi.org/10.4227/05/5878480a91885.
Gellie, N.J.C., Mills, J.G., Breed, M.F., Lowe, A.J., 2017c. Revegetation Rewilds the Soil Bacterial Microbiome of an Old Field. Part 2: Soil Chemistry, Version 2.0. AEKOS Data Portal. https://doi.org/10.4227/05/587d63e2dd056.
Gertheiss, J., 2015. ordPens: Selection and/or Smoothing of Ordinal Predictors. R package version 0.3-1. https://CRAN.R-project.org/package=ordPens.
Gilbert, J.A., Stephens, B., 2018. Microbiology of the built environment. Nat. Rev. Microbiol. https://doi.org/10.1038/s41579-018-0065-5.
von Hertzen, L., Haahtela, T., 2006. Disconnection of man and the soil: reason for the asthma and atopy epidemic? J. Allergy Clin. Immunol. 117, 334–344.
von Hertzen, L., Hanski, I., Haahtela, T., 2011. Natural immunity: biodiversity loss and inflammatory diseases are two global megatrends that might be related. EMBO Rep. 12, 1089–1093.
Hippe, H., 2000. Leptospirillum gen. nov. (ex Markosyan 1972), nom. rev., including Leptospirillum ferrooxidans sp. nov. (ex Markosyan 1972), nom. rev. and Leptospirillum thermoferrooxidans sp. nov. (Golovacheva et al. 1992). Int. J. Syst. Evol. Microbiol. 50, 501–503.
Hol, W.H.G., De Boer, W., Termorshuizen, A.J., Meyer, K.M., Schneider, J.H.M., Van Dam, N.M., Van Veen, J.A., Van Der Putten, W.H., 2010. Reduction of rare soil microbes modifies plant–herbivore interactions. Ecol. Lett. 13, 292–301.
Ichinohe, T., Pang, I.K., Kumamoto, Y., Peaper, D.R., Ho, J.H., Murray, T.S., Iwasaki, A., 2011. Microbiota regulates immune defense against respiratory tract influenza A virus infection. Proc. Natl. Acad. Sci. 108, 5354–5359.
Ishii, S., Sadowsky, M.J., 2008. Escherichia coli in the environment: implications for water quality and human health. Microbes Environ. 23, 101–108.
Jaffee, B.A., Harrison, M.B., Shaffer, R.L., Strang, M.B., 1987. Seasonal population fluctuations of Xiphinema americanum and Xiphinema rivesi in New York and Pennsylvania orchards. J. Nematol. 19, 369–378.
Janzen, H.H., 2016. The soil remembers. Soil Sci. Soc. Am. J. 80, 1429–1432.
Jeffery, S., van der Putten, W.H., 2011. Soil borne human diseases. In: JRC Scientific and Technical Reports. Joint Research Centre - Institute for Environment and Sustainability, Luxembourg.
Jost, L., 2006. Entropy and diversity. Oikos 113, 363–375.
Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., Tanabe, M., 2012. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res. 40, D109–D114.
Kim, D.-U., Lee, H., Lee, S., Kim, S.-G., Park, A.-Y., Ahn, J.-H., Ka, J.-O., 2018. Flavisolibacter metallilatus sp. nov., isolated from an automotive air conditioning system and emended description of the genus Flavisolibacter. Int. J. Syst. Evol. Microbiol. 68, 917–923.
Langille, M.G.I., Zaneveld, J., Caporaso, J.G., McDonald, D., Knights, D., Reyes, J.A., Clemente, J.C., Burkepile, D.E., Vega Thurber, R.L., Knight, R., Beiko, R.G., Huttenhower, C., 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat. Biotechnol. 31, 814.
Libert, B., Franceschi, V.R., 1987. Oxalate in crop plants. J. Agric. Food Chem. 35, 926–938.
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, 1023–1034.
Liddicoat, C., Bi, P., Waycott, M., Glover, J., Breed, M., Weinstein, P., 2018. Ambient soil cation exchange capacity inversely associates with infectious and parasitic disease risk in regional Australia. Sci. Total Environ. 626, 117–125.
Lloyd, K.G., Steen, A.D., Ladau, J., Yin, J., Crosby, L., 2018. Phylogenetically novel uncultured microbial cells dominate earth microbiomes. mSystems 3.
Louca, S., Doebeli, M., Parfrey, L.W., 2018. Correcting for 16S rRNA gene copy numbers in microbiome surveys remains an unsolved problem. Microbiome 6, 41.
Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550.
MacArthur, R.H., Wilson, E.O., 1967. The Theory of Island Biogeography. Princeton University Press, Princeton, NJ.
Mace, G.M., Barrett, M., Burgess, N.D., Cornell, S.E., Freeman, R., Grooten, M., Purvis, A., 2018. Aiming higher to bend the curve of biodiversity loss. Nature Sustainability 1, 448–451.
Matthews, D.M., Jenks, S.M., 2013. Ingestion of Mycobacterium vaccae decreases anxiety-related behavior and improves learning in mice. Behav. Process. 96, 27–35.
McMurdie, P.J., Holmes, S., 2013. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One 8, e61217.
McMurdie, P.J., Holmes, S., 2014. Waste not, want not: why rarefying microbiome data is inadmissible. PLoS Comput. Biol. 10, e1003531.
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.
Mills, J.G., Weinstein, P., Gellie, N.J.C., Weyrich, L.S., Lowe, A.J., Breed, M.F., 2017. Urban habitat restoration provides a human health benefit through microbiome rewilding: the Microbiome Rewilding Hypothesis. Restor. Ecol. 25, 866–872.
Navarro, L.M., Marques, A., Proença, V., Ceauşu, S., Gonçalves, B., Capinha, C., Fernandez, M., Geldmann, J., Pereira, H.M., 2017. Restoring degraded land: contributing to Aichi Targets 14, 15, and beyond. Curr. Opin. Environ. Sustain. 29, 207–214.
Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H., 2018. vegan: Community Ecology Package. R Package Version 2.5-2. https://CRAN.R-project.org/package=vegan.
Ottman, N., Ruokolainen, L., Suomalainen, A., Sinkko, H., Karisola, P., Lehtimäki, J., Lehto, M., Hanski, I., Alenius, H., Fyhrquist, N., 2018. Soil exposure modifies the gut microbiota and supports immune tolerance in a mouse model. J. Allergy Clin. Immunol. https://doi.org/10.1016/j.jaci.2018.06.024.
Polymenakou, P.N., 2012. Atmosphere: a source of pathogenic or beneficial microbes? Atmosphere 3, 87–102.
Public-Health-England, 2017. *Escherichia coli* O55:H7 Outbreak in Dorset. July 2014 to December 2015. London.
R-Core-Team, 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Ristuccia, P.A., Cunha, B.A., 1985. Enterobacter. Infect. Control. 6, 124–128.
Robinson, J., Mills, J., Breed, M., 2018. Walking ecosystems in microbiome-inspired green infrastructure: an ecological perspective on enhancing personal and planetary health. Challenges 9, 40. https://doi.org/10.3390/challe9020040.
Rognes, T., Flouri, T., Nichols, B., Quince, C., Mahé, F., 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ 4, e2584.
Rook, G.A.W., 2012. Hygiene hypothesis and autoimmune diseases. Clin. Rev. Allergy Immunol. 42, 5–15.
Rook, G., 2013. Regulation of the immune system by biodiversity from the natural environment: an ecosystem service essential to health. Proc. Natl. Acad. Sci. 110, 18360–18367.
Rumbaugh, K.P., Trivedi, U., Watters, C., Burton-Chellew, M.N., Diggle, S.P., West, S.A., 2012. Kin selection, quorum sensing and virulence in pathogenic bacteria. Proc. R. Soc. B Biol. Sci. 279, 3584–3588.
Rydin, Y., Bleahu, A., Davies, M., Dávila, J.D., Friel, S., De Grandis, G., Groce, N., Hallal, P.C., Hamilton, I., Howden-Chapman, P., Lai, K.-M., Lim, C.J., Martins, J., Osrin, D., Ridley, I., Scott, I., Taylor, M., Wilkinson, P., Wilson, J., 2012. Shaping cities for health: complexity and the planning of urban environments in the 21st century. Lancet 379, 2079–2108.
SER International Science & Policy Working Group, 2004. The SER International Primer on Ecological Restoration. www.ser.orgSociety for Ecological Restoration International, Tucson, Arizona USA.
Shinichi, N., Innes, C., 2007. Effect size, confidence interval and statistical significance: a practical guide for biologists. Biol. Rev. 82, 591–605.
Stanford, L.L., Spacie, A., 1994. Biological Monitoring of Aquatic Systems. CRC Press, Boca Raton, Florida USA.
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., Neilson, J.W., Maier, R.M., Gilbert, J.A., Holbreich, M., Thorne, P.S., Martinez, F.D., von Mutius, E., Vercelli, D., Ober, C., Sperling, A.I., 2016. Innate immunity and asthma risk in Amish and Hutterite farm children. N. Engl. J. Med. 375, 411–421.
Turner, T.R., Ramakrishnan, K., Walshaw, J., Heavens, D., Alston, M., Swarbreck, D., Osbourn, A., Grant, A., Poole, P.S., 2013. Comparative metatranscriptomics reveals kingdom level changes in the rhizosphere microbiome of plants. The ISME Journal 7, 2248–2258.
VanInsberghe, D., Maas, K.R., Cardenas, E., Strachan, C.R., Hallam, S.J., Mohn, W.W., 2015. Non-symbiotic Bradyrhizobium ecotypes dominate North American forest soils. The ISME Journal 9, 2435–2441.
Ward, N.L., Challacombe, J.F., Janssen, P.H., Henrissat, B., Coutinho, P.M., Wu, M., Xie, G., Haft, D.H., Sait, M., Badger, J., Barabote, R.D., Bradley, B., Brettin, T.S., Brinkac, L.M., Bruce, D., Creasy, T., Daugherty, S.C., Davidsen, T.M., DeBoy, R.T., Detter, J.C., Dodson, R.J., Durkin, A.S., Ganapathy, A., Gwinn-Giglio, M., Han, C.S., Khouri, H.,

Kiss, H., Kothari, S.P., Madupu, R., Nelson, K.E., Nelson, W.C., Paulsen, I., Penn, K., Ren, Q., Rosovitz, M.J., Selengut, J.D., Shrivastava, S., Sullivan, S.A., Tapia, R., Thompson, L.S., Watkins, K.L., Yang, Q., Yu, C., Zafar, N., Zhou, L., Kuske, C.R., 2009. Three genomes from the phylum Acidobacteria provide insight into the lifestyles of these microorganisms in soils. Appl. Environ. Microbiol. 75, 2046.

Watson, S.W., Bock, E., Valois, F.W., Waterbury, J.B., Schlosser, U., 1986. Nitrospira marina gen. nov. sp. nov.: a chemolithotrophic nitrite-oxidizing bacterium. Arch. Microbiol. 144, 1–7.

Weiss, S., Xu, Z.Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., Lozupone, C., Zaneveld, J.R., Vázquez-Baeza, Y., Birmingham, A., Hyde, E.R., Knight, R., 2017. Normalization and microbial differential abundance strategies depend upon data characteristics. Microbiome 5, 27.

WHO, SCBD, 2015. Connecting Global Priorities: Biodiversity and Human Health. A State of Knowledge Review. World Health Organisation, Geneva, Switzerland.

Woloszynek, S., Mell, J.C., Simpson, G., Connor, M.P., Rosen, G.L., 2017. Uncovering thematic structure to link co-occurring taxa and predicted functional content in 16S rRNA marker gene surveys. bioRxiv 146126. https://doi.org/10.1101/146126.

Wubs, E.R.J., van der Putten, W.H., Bosch, M., Bezemer, T.M., 2016. Soil inoculation steers restoration of terrestrial ecosystems. Nature Plants 2, 16107.

Wyres, K.L., Holt, K.E., 2018. Klebsiella pneumoniae as a key trafficker of drug resistance genes from environmental to clinically important bacteria. Curr. Opin. Microbiol. 45, 131–139.

Yan, D., Mills, J.G., Gellie, N.J.C., Bissett, A., Lowe, A.J., Breed, M.F., 2018. High-throughput eDNA monitoring of fungi to track functional recovery in ecological restoration. Biol. Conserv. 217, 113–120.

Yoon, J.-H., Lee, K.-C., Weiss, N., Kho, Y.H., Kang, K.H., Park, Y.-H., 2001. Sporosarcina aquimarina sp. nov., a bacterium isolated from seawater in Korea, and transfer of Bacillus globisporus (Larkin and Stokes 1967), Bacillus psychrophilus (Nakamura 1984) and Bacillus pasteurii (Chester 1898) to the genus Sporosarcina as Sporosarcina globispora comb. nov., Sporosarcina psychrophila comb. nov. and Sporosarcina pasteurii comb. nov., and emended description of the genus Sporosarcina. Int. J. Syst. Evol. Microbiol. 51, 1079–1086.

Web Appendix – Supplementary Material for

Can bacterial indicators of a grassy woodland restoration inform ecosystem assessment and microbiota-mediated human health?

Craig Liddicoat*, Philip Weinstein, Andrew Bissett, Nicholas J.C. Gellie, Jacob G. Mills, Michelle Waycott, Martin F. Breed

*Corresponding author. E-mail: craig.liddicoat@adelaide.edu.au

The following Supplementary Material is available for this article:

- Methods S1
- Figures S1-S19
- Tables S1-S5
- Supplementary References

Data availability

Data and R code that support the findings of this study are available on *figshare* at: https://figshare.com/s/1acbc273dfc93da272be or DOI: 10.25909/5cbef4802c4d8

Methods S1

Alignment of study OTU representative sequences to Greengenes 13-5 IDs using VSEARCH

We used VSEARCH version 2.8.1, and followed user documentation (User Manual version 2.8.1; https://github.com/torognes/vsearch/releases/download/v2.8.1/vsearch_manual.pdf) to find the closest matching Greengenes sequences to the study 97% clustered OTUs. We ran the following VSEARCH commands in the Windows DOS Command Prompt (Microsoft Windows version 10.0.14393).

For Mt Bold surface samples:
C:\vsearch-2.8.1-win-x86_64\vsearch --usearch_global C:\Workspace\PROJ\PAPER-Trending-Taxa-Resto\modelling\study_selection_base_AMD_seq.fasta --db C:\Workspace\DATA\Greengenes_13_5\gg_13_5.fasta --id 0.6 --userout base_AMD_search_results2.txt --userfields query+target+id+alnlen+mism+opens+qlo+qhi+tlo+thi+evalue+bits

[VSEARCH found matching query sequences: 3227 of 3238 (99.66%) above a first-pass 60% identity threshold.]

For the selected Australia-wide BASE samples:
C:\vsearch-2.8.1-win-x86_64\vsearch --usearch_global C:\Workspace\PROJ\PAPER-Trending-Taxa-Resto\modelling\study_selection_base_seq.fasta --db C:\Workspace\DATA\Greengenes_13_5\gg_13_5.fasta --id 0.6 --userout base_search_results2.txt --userfields query+target+id+alnlen+mism+opens+qlo+qhi+tlo+thi+evalue+bits

[VSEARCH found matching query sequences: 12406 of 12413 (99.94%) above a first-pass 60% identity threshold.]

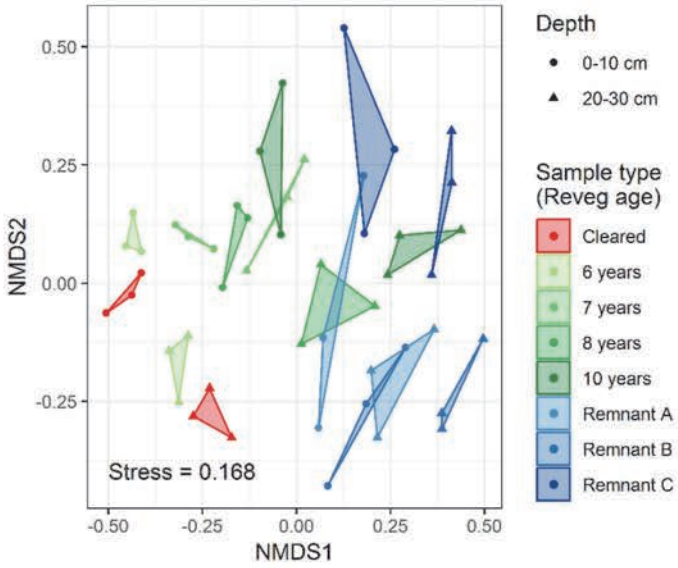


Fig. S1. Visualisation of differences between bacterial communities (beta diversity) for the Mt Bold restoration chronosequence, using NMDS ordination of Bray-Curtis distance matrices of OTU rarefied abundance (sequence depth = 33625). These are data from Gellie et al. (2017) repeated for here context.

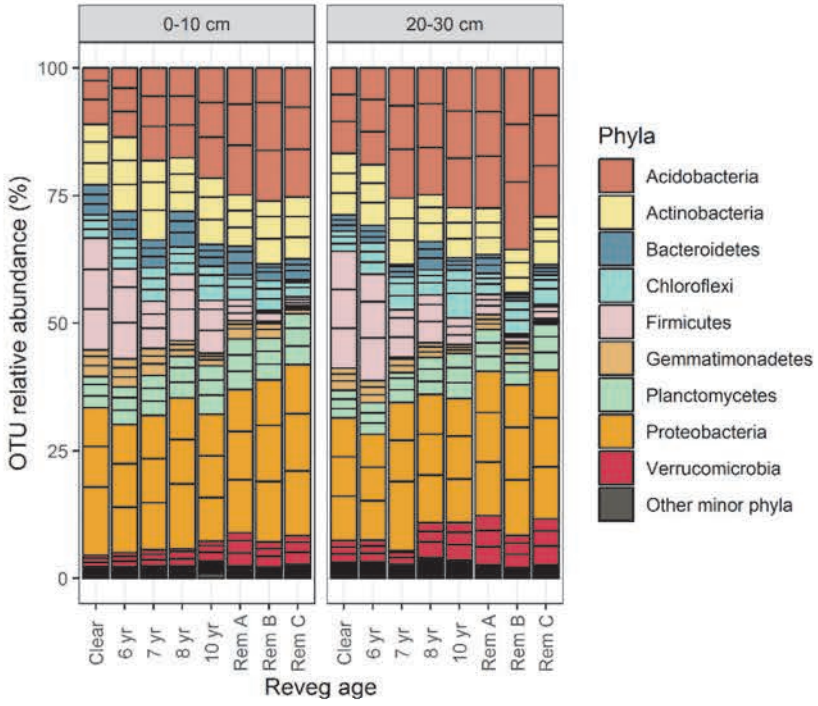


Fig. S2. Phyla relative abundance across the Mt Bold restoration chronosequence. These are data from Gellie et al. (2017) repeated for here context.

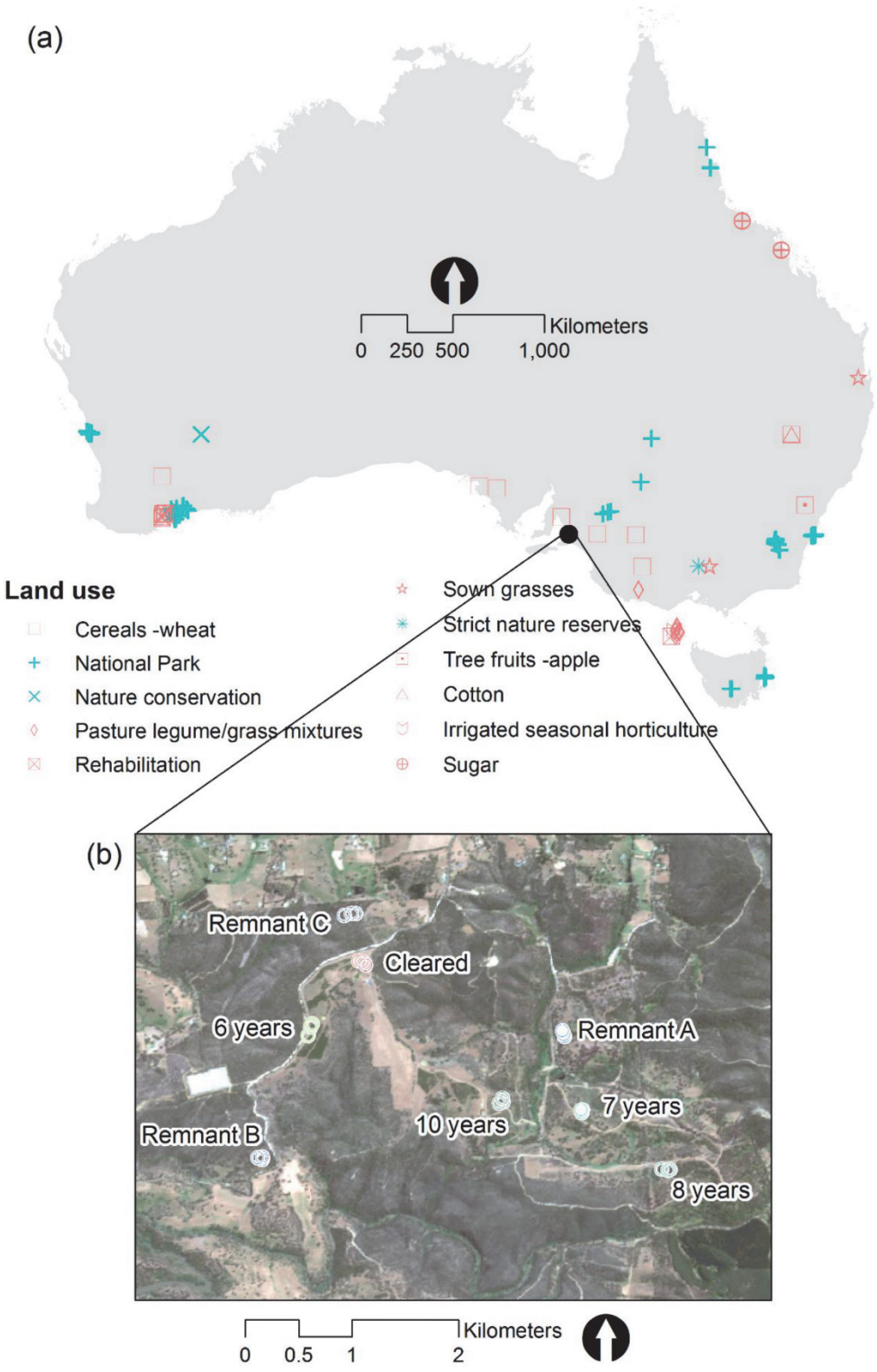


Fig. S3. Locations of soil microbiome comparison samples from the Biomes of Australian Soil Environments database (a) and study sites from the Mt Bold restoration chronosequence (b).

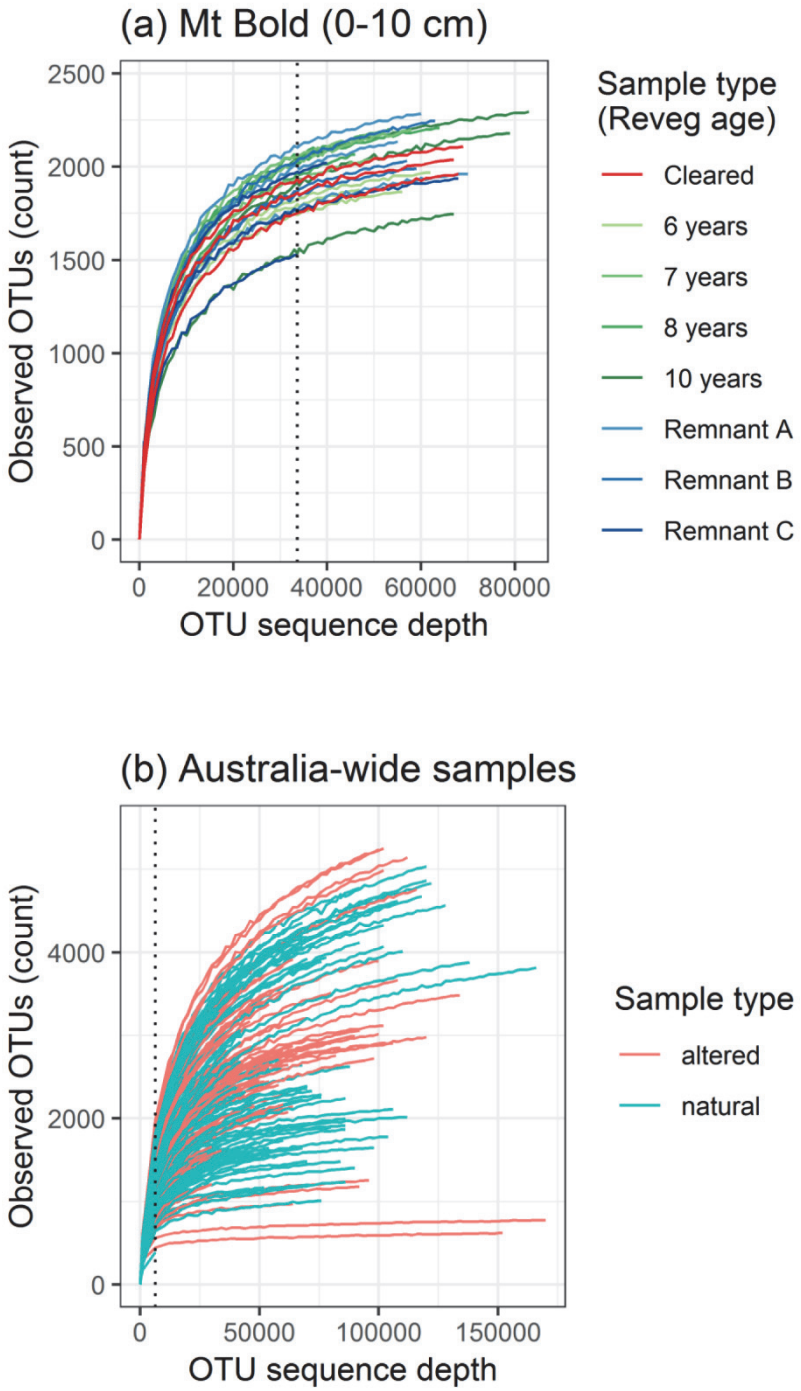


Fig. S4. Rarefaction curves for Mt Bold (a) and Australia-wide human-altered and natural samples (b). The minimum sample read depths were 33625 for Mt Bold 0-10 cm samples, and 6377 for the Australia-wide samples considered in our study.

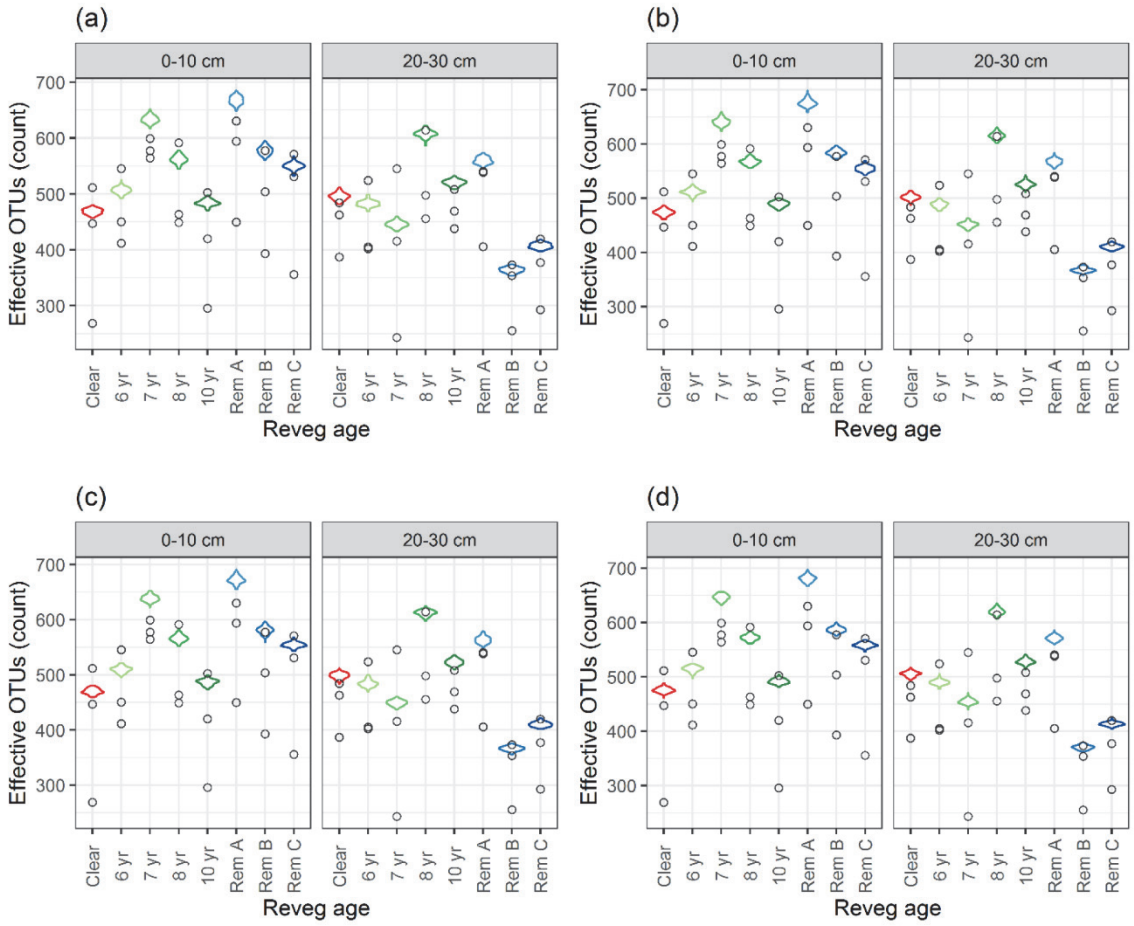


Fig. S5. Trial comparison outputs from site-level merged-sample resampling (B = 100) for calculations of alpha diversity, based on the exponential of Shannon's diversity index (Jost 2006). Circles display data from one-off rarefied (sequence depth = 33625) abundances for each triplicate sample, while density distributions indicate the weight of data derived from merged samples at each site. With reference to the *merged-sample bootstrap* workflow outlined in the main paper, these trials show:

(a) Rarefy#1 (with replacement) > Merge triplicate samples > Rarefy#2 (with replacement)

(b) Rarefy#1 (with replacement) > Merge triplicate samples > Rarefy#2 (without replacement)

(c) Rarefy#1 (without replacement) > Merge triplicate samples > Rarefy#2 (with replacement)

(d) Rarefy#1 (without replacement) > Merge triplicate samples > Rarefy#2 (without replacement)

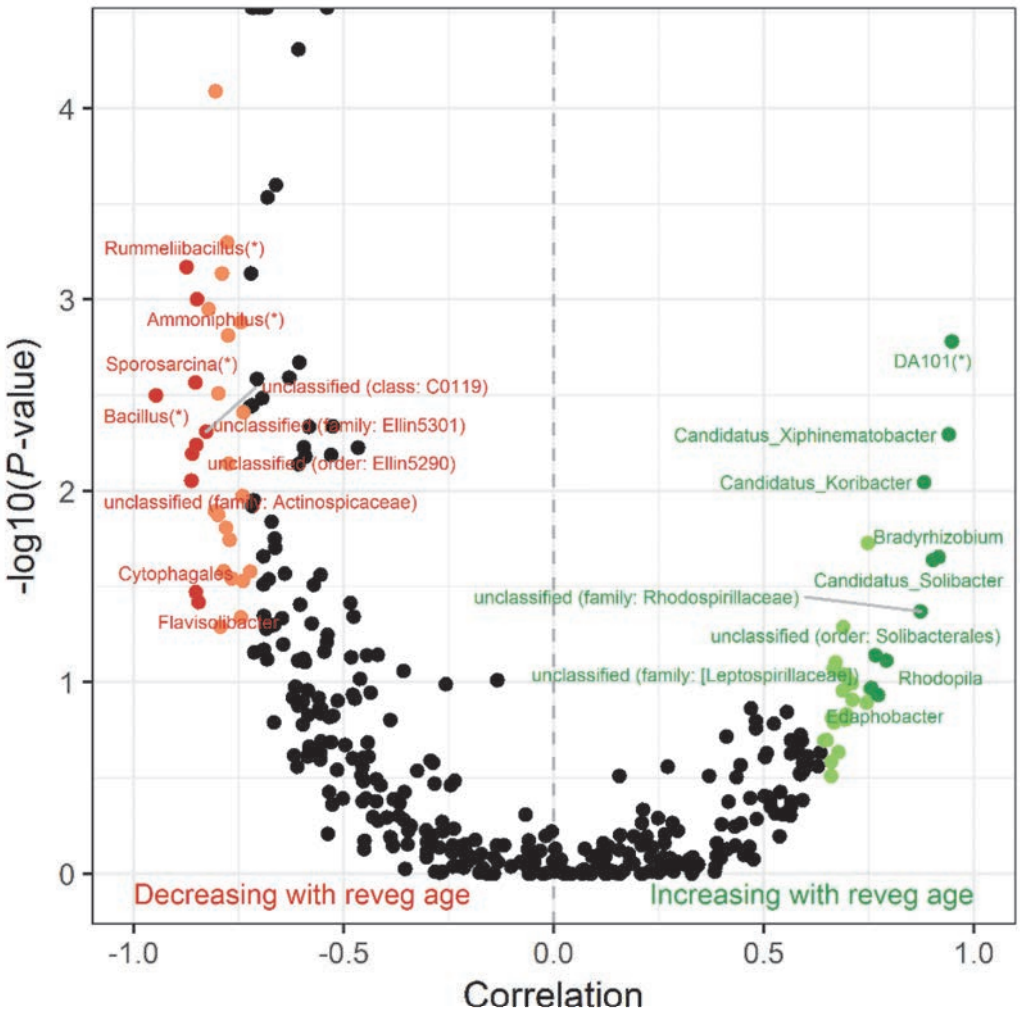


Fig. S6. Key trending bacteria with restoration in Mt Bold surface soils. Effect size is indicated by the correlation of OTU relative abundance (%) with revegetation age (x-axis). Only the top 10 increasing (dark green) and top 10 decreasing (red) taxa based on effect size are labelled. The top 30 increasing (dark green, green) and top 30 decreasing (orange, red) taxa are reported in Table S1. The significance of taxa displaying an ordinal trend is indicated by the *P*-value from a separate ordinal regression AOV test (y-axis). Significant *P*-values are denoted (*) using the Benjamini–Hochberg procedure to control the false discovery rate at alpha level 0.05 (refer to Fig. S7). Data represent mean values from *merged-sample bootstrap resampling* (B=100).

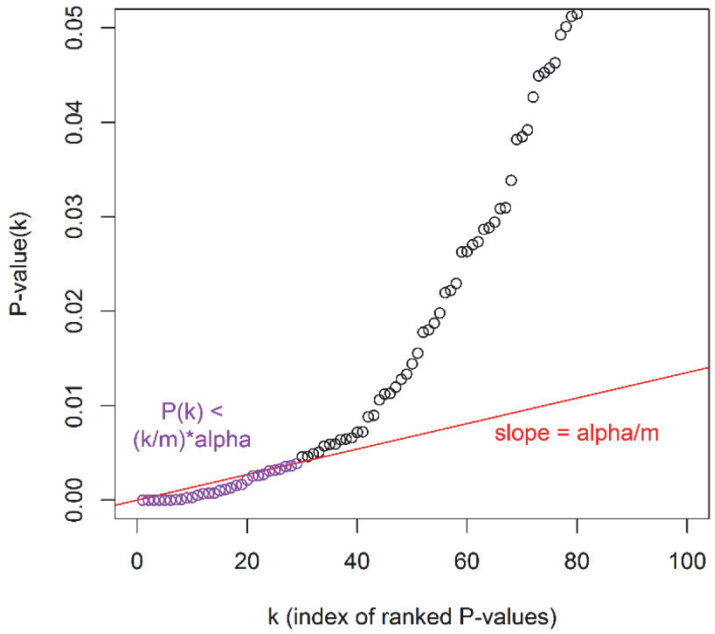


Fig. S7. Visualisation of Benjamini-Hochberg false discovery rate correction for significant *P*-values arising from the ordinal regression AOV for trending taxa in Mt Bold surface soils. Only the lowest 29 *P*-values were considered significant for alpha = 0.05.

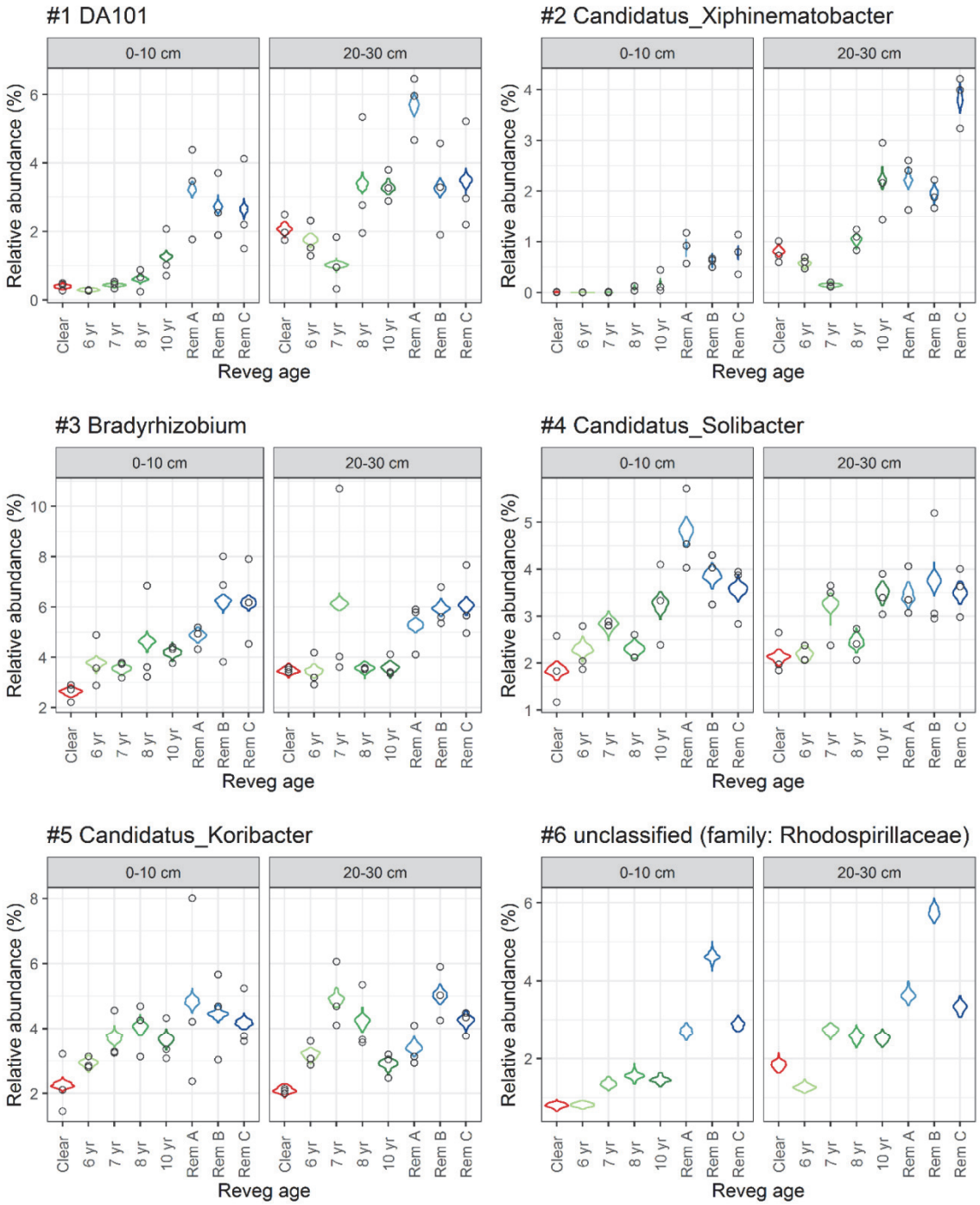


Fig. S8. Top 10 increasing bacteria with restoration at Mt Bold, ordered by the strength of correlation between OTU relative abundance (y-axis) and revegetation age in the surface soils (x-axis). Subsoil data are also presented (right panels) to provide assurance in the trends observed. Circles represent rarefied (sequence depth = 33625) triplicate samples, while the density distributions indicate the weight of equally likely outcomes from *merged-sample bootstrap resampling* (B=100). Where circles are absent, these taxa were omitted by the initial one-off rarefying process. Continued next page.

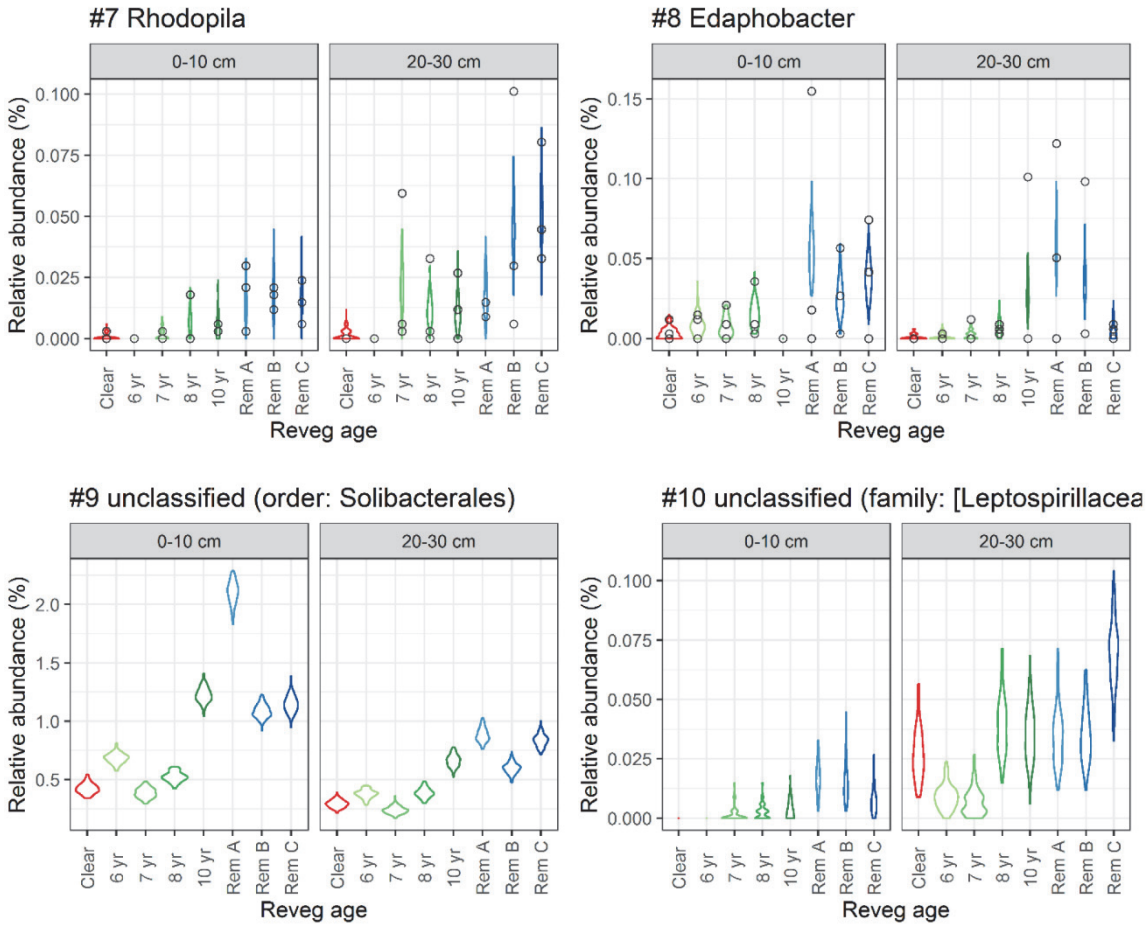


Fig. S8. (continued)

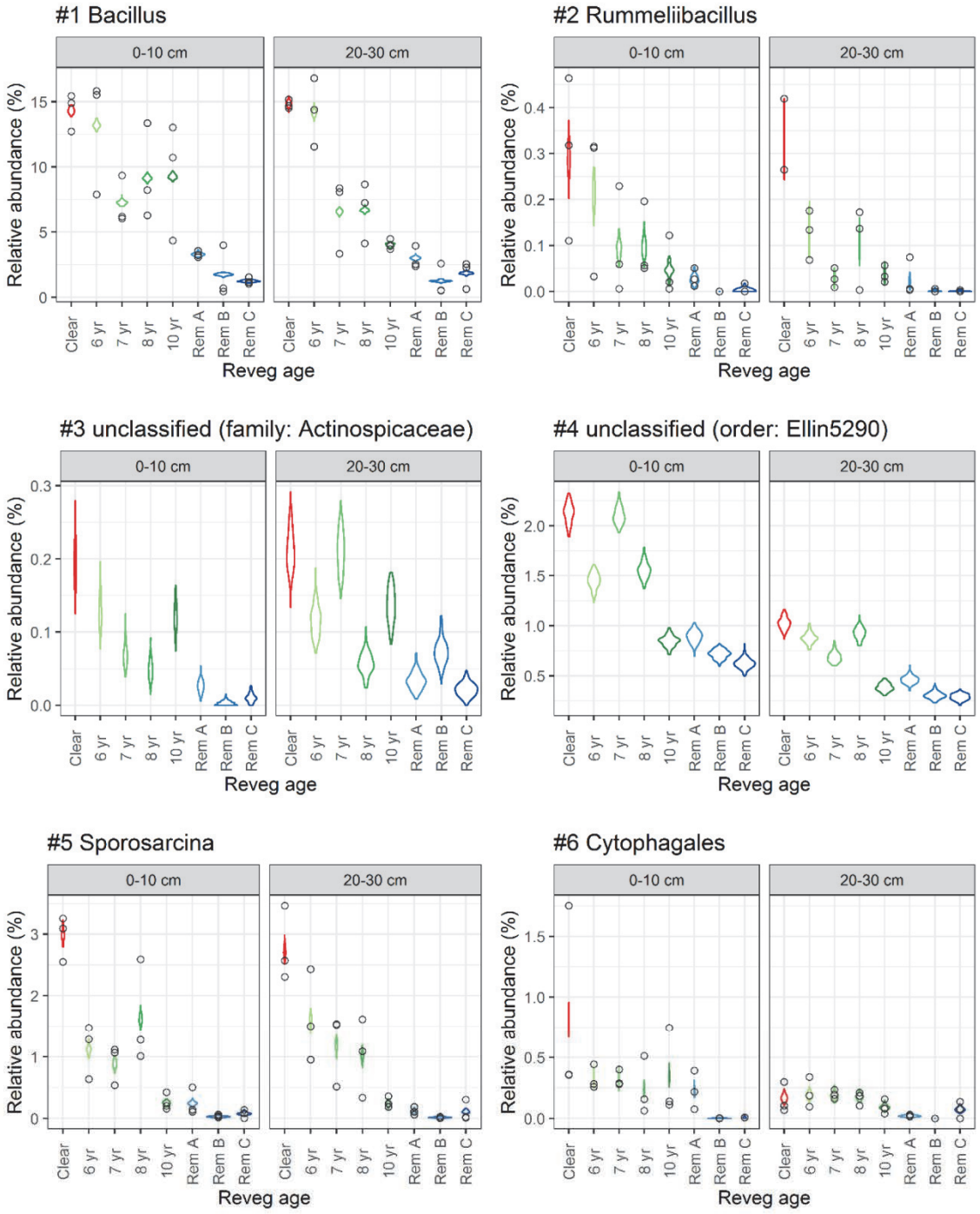


Fig. S9. Top 10 decreasing bacteria with restoration at Mt Bold, ordered by the strength of correlation between OTU relative abundance (y-axis) and revegetation age in the surface soils (x-axis). Subsoil (20-30 cm) data are also presented (right panels) to provide assurance in the trends observed. Circles represent rarefied (sequence depth = 33625) triplicate samples, while the density distributions indicate the weight of data from equally likely outcomes from *merged-sample bootstrap resampling* (B=100). Where circles are absent, these taxa were omitted by the initial one-off rarefying process. Continued next page.

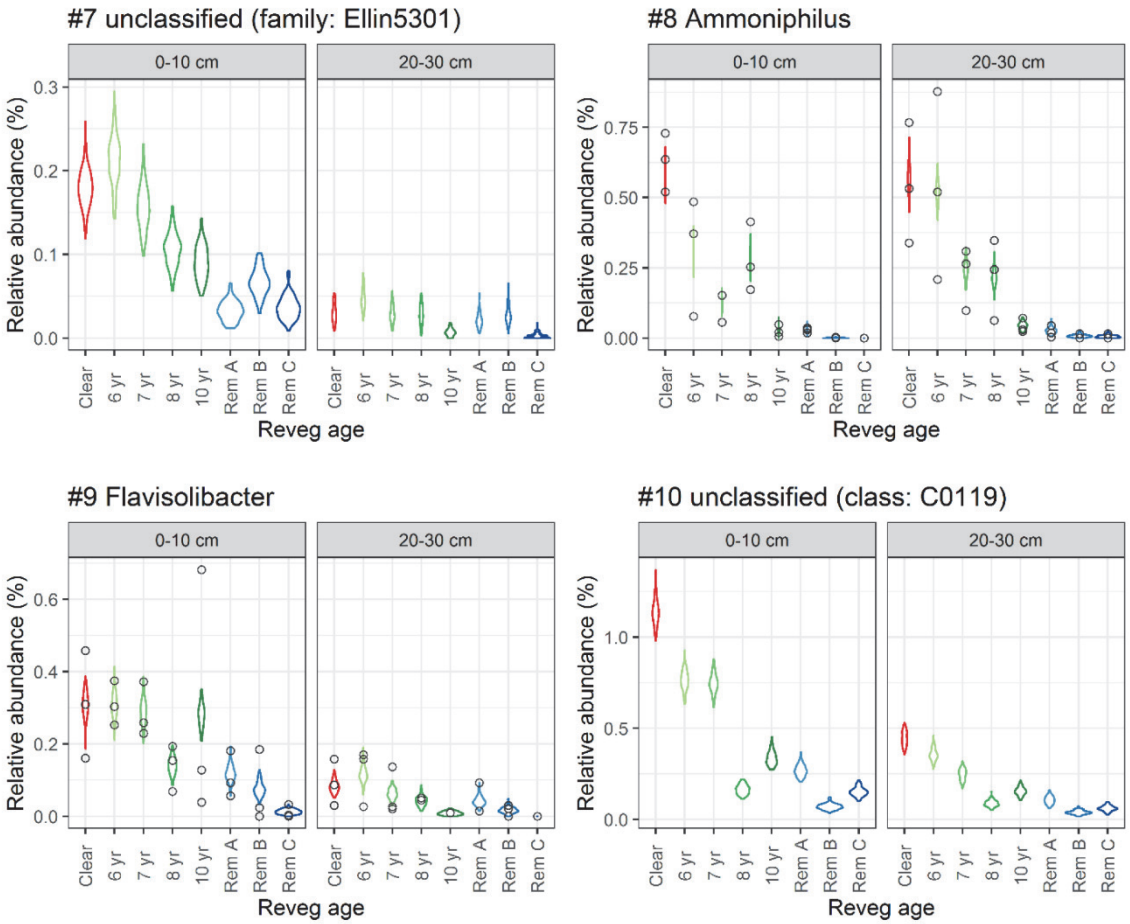


Fig. S9. (Continued)

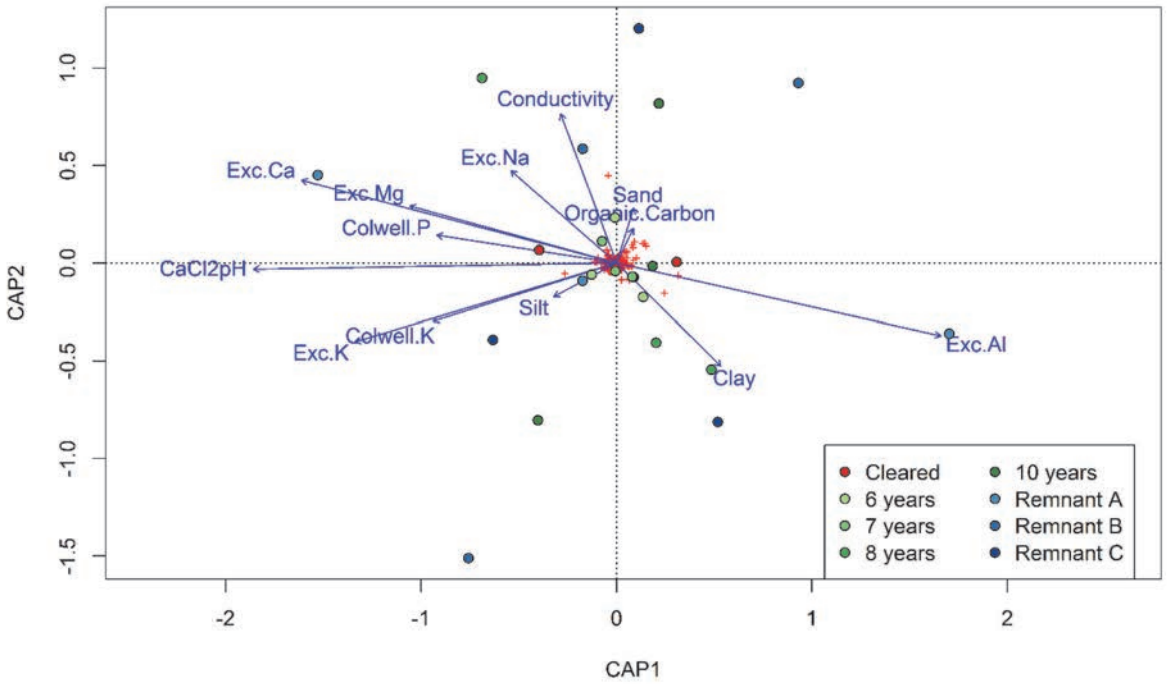


Fig. S10. Constrained analysis of principal coordinates or (distance-based redundancy analysis), based on OTU rarefied abundance (sequence depth = 33625), showing relationships between samples (filled circles) and OTUs (red crosses) for Mt Bold surface soils. Relationships with variables were examined using the model formula: Sand + Silt + Clay + Colwell.P + Colwell.K + Organic.Carbon + Conductivity + CaCl2pH + Exc.Al + Exc.Ca + Exc.Mg + Exc.K + Exc.Na + CEC + Condition(Reveg_age).

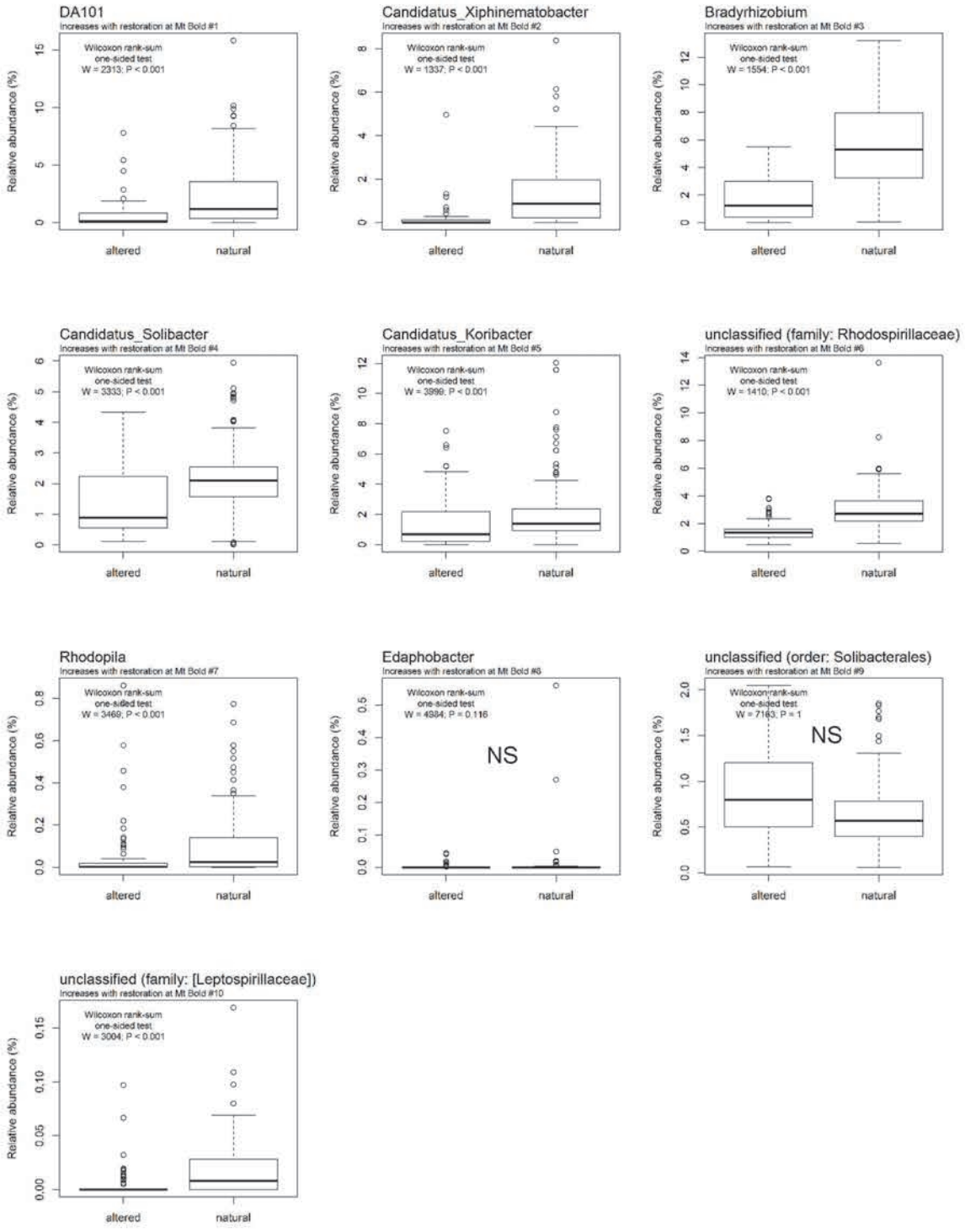


Fig. S11. OTU relative abundance comparisons between Australia-wide human-altered and natural samples for the top 10 <u>increasing</u> taxa with restoration at Mt Bold. Trends in the taxa are consistent with Mt Bold restoration trends, except where results from a one-sided Wilcoxon rank-sum test report no significant difference (NS).

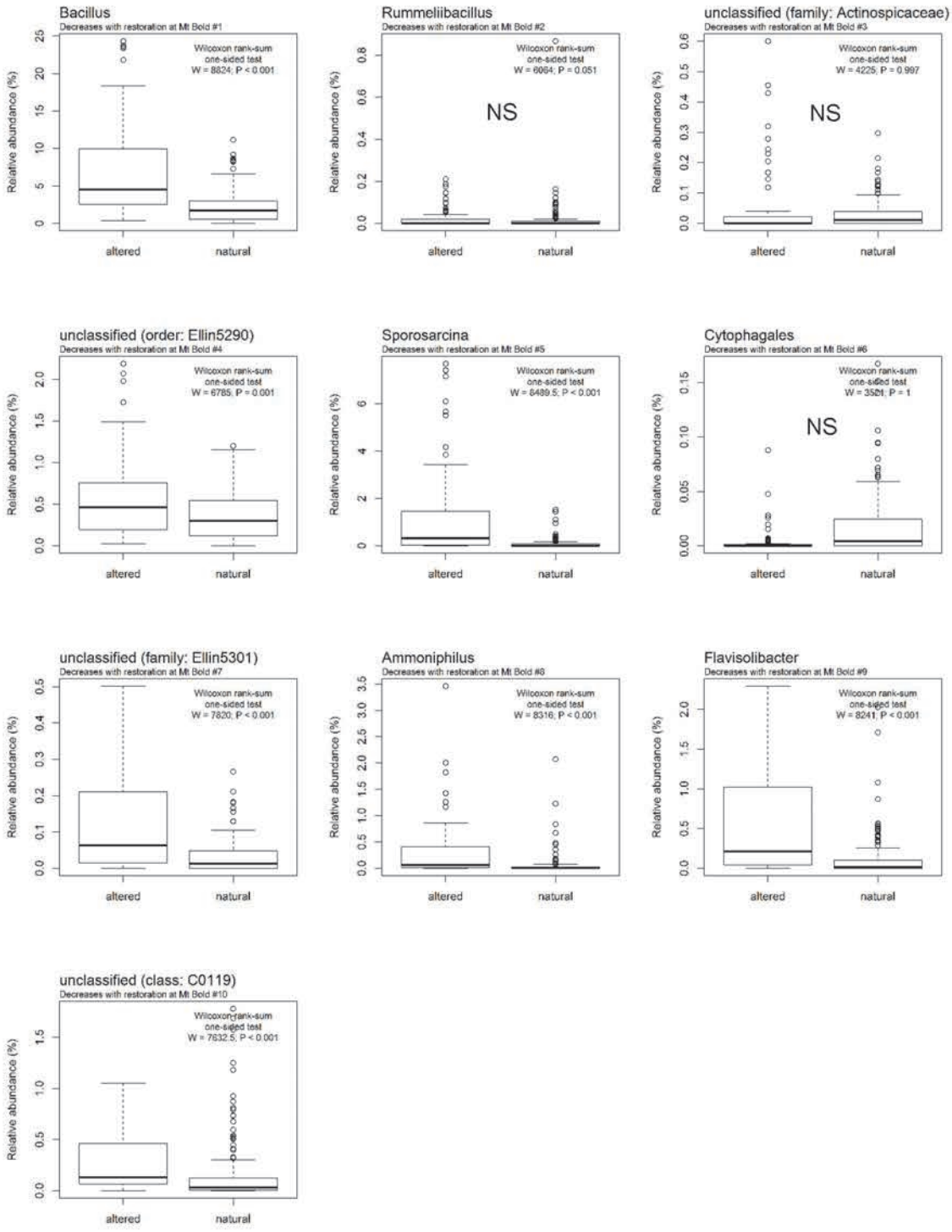


Fig. S12. OTU relative abundance comparisons between Australia-wide human-altered and natural samples for the top 10 decreasing taxa with restoration at Mt Bold. Trends in the taxa are consistent with Mt Bold restoration trends, except where results from a one-sided Wilcoxon rank-sum test report no significant difference (NS).

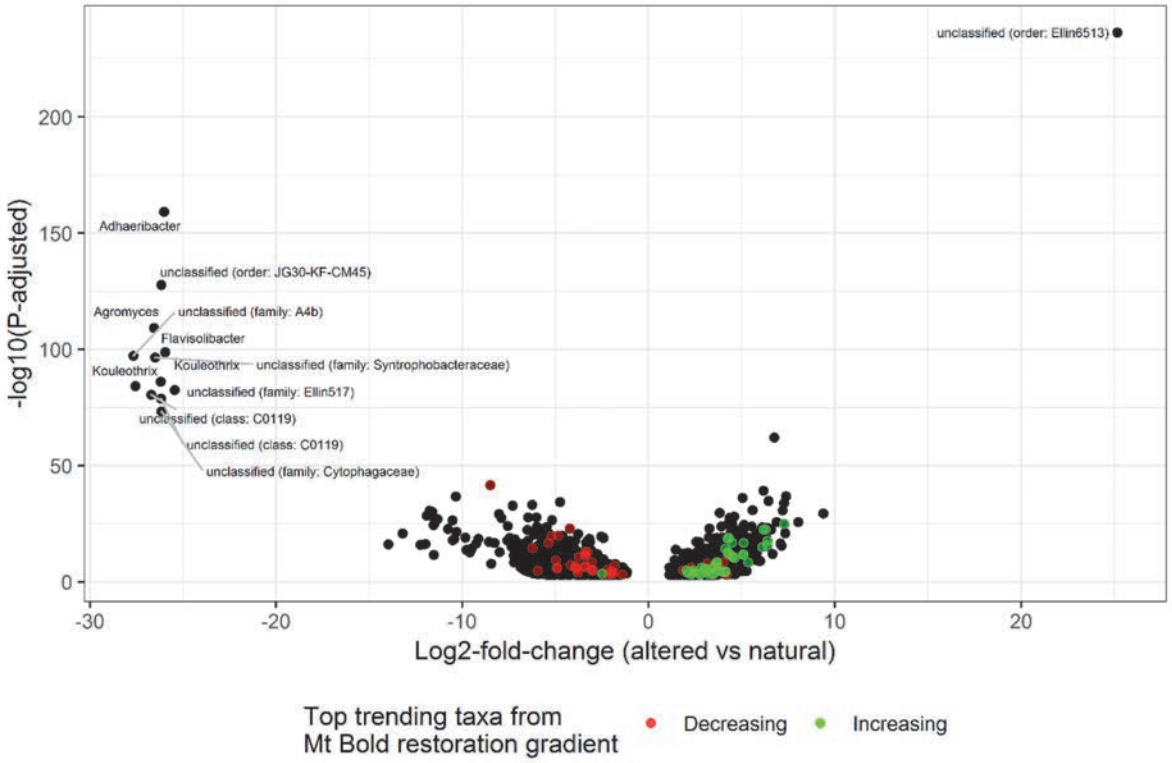


Fig. S13. Volcano plot of results from *DESeq2* differential abundance testing (alpha = 0.001), showing log-2-fold-change between Australia-wide human-altered and natural samples (x-axis) and the corresponding –log10(adjusted *P*-values) (y-axis). The top two increasing taxa (green, genera *DA101* and *Candidatus Xiphinematobacter*) and top decreasing taxa (red, genus *Bacillus*) that trend with restoration at Mt Bold are highlighted. Note the genus *Rummeliibacillus* was in the top two decreasing taxa from Mt Bold, however it is absent in the Australia-wide study samples.

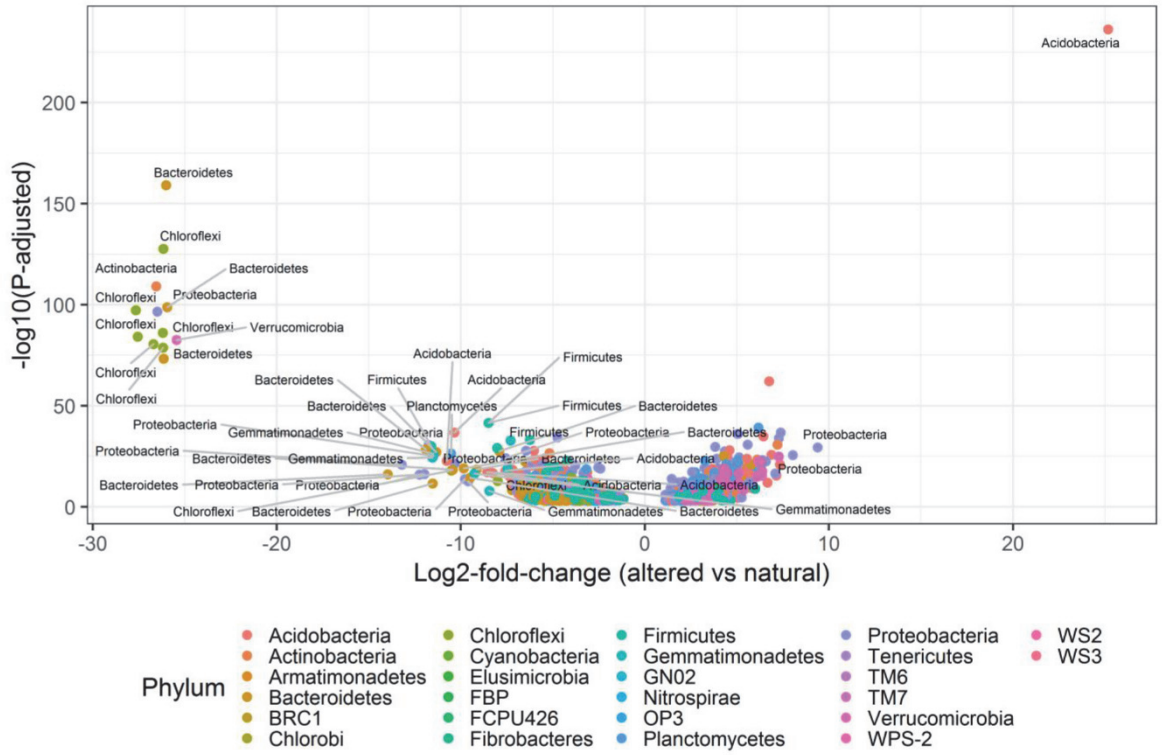


Fig. S14. Plot as per Fig. S13 above, but with phyla highlighted.

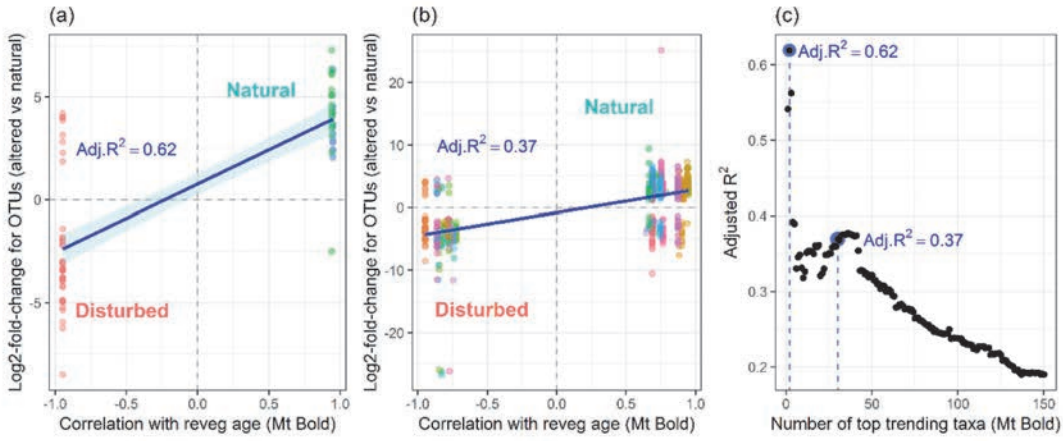


Fig. S15. The top trending (increasing/decreasing) genera with restoration at Mt Bold show similar differential abundance between human-altered and natural samples across Australia. Correlation of OTU relative abundance with revegetation age is shown on the x-axis, and differential abundance of the respective OTUs between Australia-wide human-altered and natural samples (from *DESeq2* results with alpha level = 0.001) is on the y-axis, for the top two increasing and top two decreasing genera in (a), and the top 30 (increasing/ decreasing) genera in (b). The strength of association is indicated by the linear regression (blue line) and adjusted R^2 value. In (a) and (b), the quadrant labelled 'Natural' corresponds to taxa that increase with restoration and have higher differential abundance in natural sites, while the quadrant labelled 'Disturbed' corresponds to taxa that decrease with restoration and have higher differential abundance in human-altered sites. The association breaks down as more taxa are considered, as shown in (c). Note that each genera (or group classified at the next available level) identified from the top-trending taxa at Mt Bold corresponds to a number of OTUs identified in the Australia-wide differential abundance testing, i.e. multiple points in y-domain in (a) and (b).

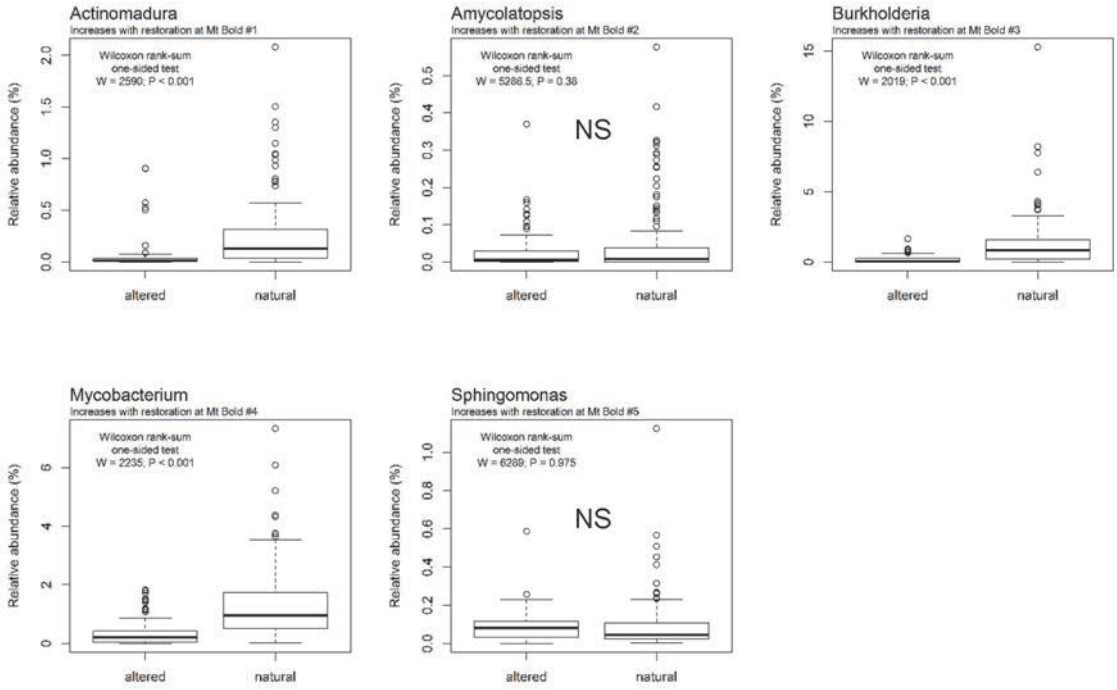


Fig. S16. OTU Relative abundance comparisons between Australia-wide human-altered and natural samples for human-associated bacteria that increase in relative abundance with restoration at Mt Bold. Trends in the taxa are consistent with Mt Bold restoration trends, except where results from a one-sided Wilcoxon rank-sum test report no significant difference (NS).

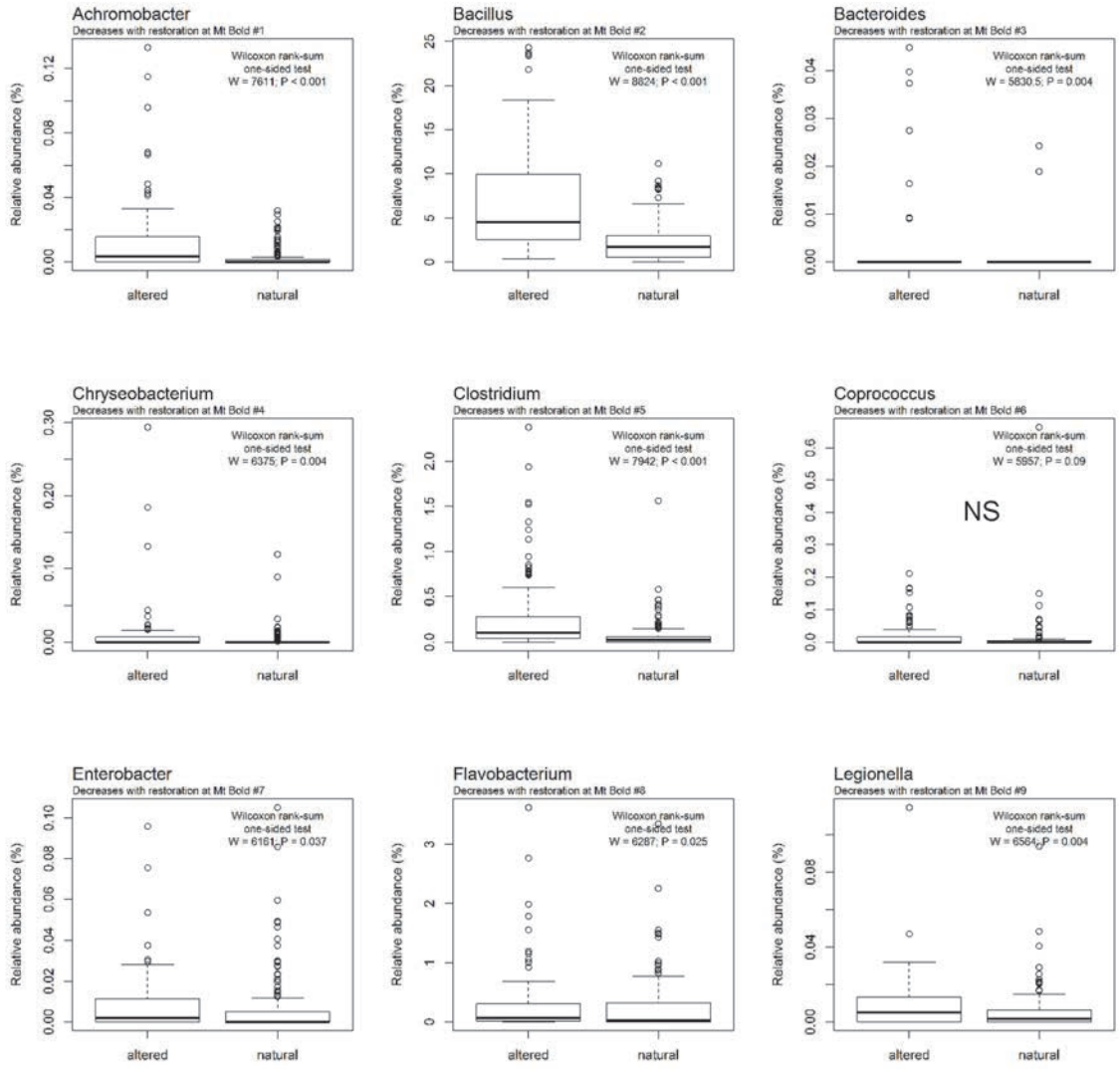


Fig. S17. OTU relative abundance comparisons between Australia-wide human-altered and natural samples for human-associated bacteria that decrease in relative abundance with restoration at Mt Bold. Trends in the taxa are consistent with Mt Bold restoration trends, except where results from a one-sided Wilcoxon rank-sum test report no significant difference (NS). Continued next page.

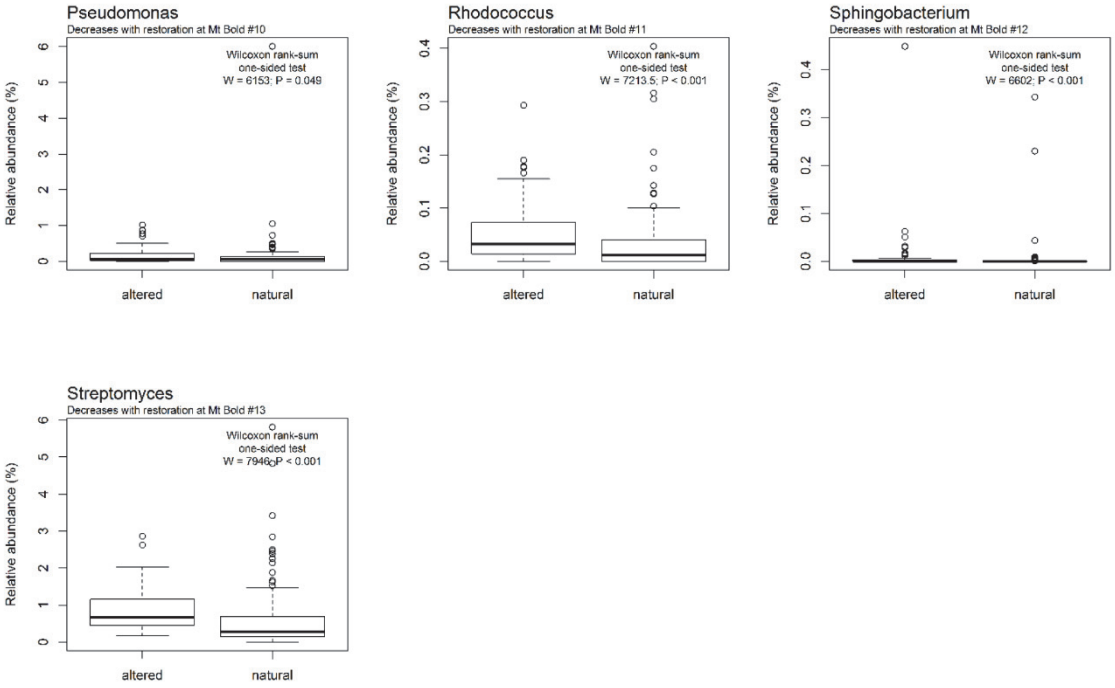


Fig. S17. (Continued)

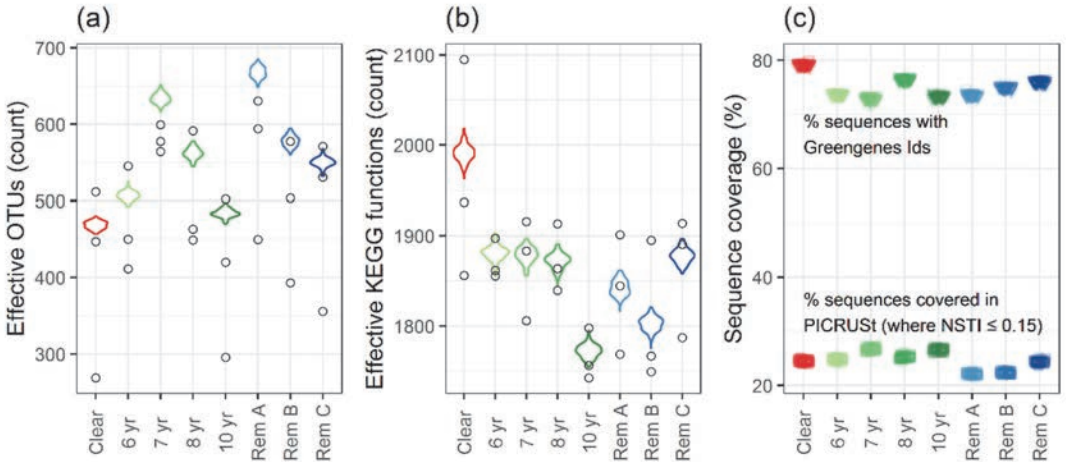


Fig. S18. Microbial diversity (a), inferred functional diversity (b), and coverage of sequences used in functional assignments (c), with uncertainty, for samples from the Mt Bold restoration gradient. Only sequences with NSTI ≤ 0.15 were used for functional assignments, resulting in a median weighted NSTI across all samples of 0.090 (range 0.076-0.103). Circles correspond to rarefied OTU abundance data (sequence depth= 33625 in (a), and sequence depth= 17614 in (b) and (c) for comparison with Australia-wide functional data) from each sample, while the density distributions indicate the weight of site-level data from *merged-sample bootstrap resampling* (B=100).

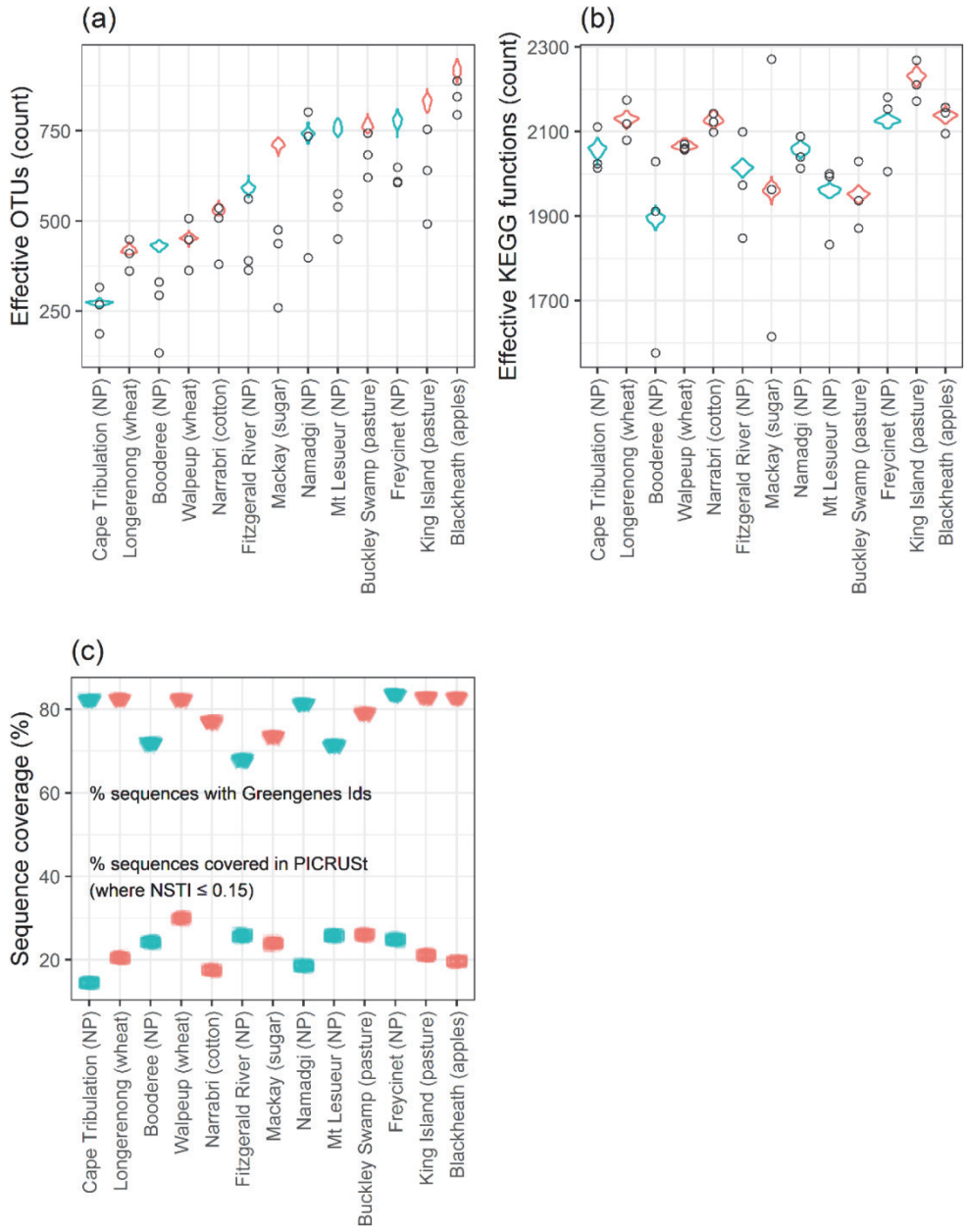


Fig. S19. Microbial diversity (a), estimated functional diversity (b), and coverage of sequences used in functional assignments (c) with uncertainty, for Australia-wide natural and human-altered samples. Only sequences with NSTI ≤ 0.15 were used for functional assignments, resulting in a median weighted NSTI across all samples of median 0.084 (range 0.060-0.102). Circles correspond to rarefied (sequence depth = 17614) sample data from each site, while the density distributions (cyan for natural, pale red for human-altered) indicate the weight of site-level data from *merged-sample bootstrap resampling* (B=100).

Table S1. Top 30 increasing and top 30 decreasing taxa from Mt Bold restoration.

Genus (or nearest classified group)	Relative abundance (%); mean (range)	Correlation with revegetation age; mean (95% CI)	P-value for ordinal AOV; mean (95% CI)	Missing in cleared	Missing in remnants	% bootstraps with data
Increasing with restoration						
DA101(*)	2.23 (0.22-6.06)	0.95 (0.94,0.96)	0 (0,0)	No	No	100
Candidatus_Xiphinematobacter	0.96 (0-4.14)	0.94 (0.91,0.95)	0.01 (0,0.01)	No	No	100
Bradyrhizobium	4.6 (2.41-6.54)	0.92 (0.88,0.95)	0.02 (0.01,0.04)	No	No	100
Candidatus_Solibacter	3.06 (1.64-5.12)	0.9 (0.87,0.94)	0.02 (0.01,0.04)	No	No	100
Candidatus_Koribacter	3.76 (1.87-5.37)	0.88 (0.84,0.92)	0.01 (0,0.02)	No	No	100
unclassified (family: Rhodospirillaceae)	2.49 (0.66-6.12)	0.87 (0.85,0.9)	0.04 (0.03,0.06)	No	No	100
Rhodopila	0.01 (0-0.09)	0.79 (0.49,0.94)	0.08 (0,0.25)	No	No	100
Edaphobacter	0.02 (0-0.1)	0.77 (0.58,0.91)	0.12 (0,0.3)	No	No	100
unclassified (order: Solibacterales)	0.74 (0.17-2.29)	0.77 (0.72,0.82)	0.07 (0.04,0.1)	No	No	100
unclassified (family: [Leptospirillaceae])	0.02 (0-0.1)	0.76 (0.41,0.91)	0.11 (0.01,0.26)	Yes	No	98
unclassified (order: Ellin6513)	1.12 (0.04-4.97)	0.75 (0.71,0.8)	0.11 (0.07,0.13)	No	No	100
unclassified (family: Gemmataceae)	1.68 (0.8-2.86)	0.75 (0.66,0.82)	0.02 (0,0.05)	No	No	100
unclassified (order: Acidobacteriales)	0.07 (0-0.4)	0.74 (0.65,0.85)	0.13 (0.05,0.21)	Yes	No	100
Rhodomicrobium	0.14 (0-0.67)	0.71 (0.66,0.76)	0.12 (0.09,0.16)	No	No	100
Singulisphaera	0.02 (0-0.1)	0.71 (0.45,0.9)	0.1 (0.01,0.24)	No	No	100
unclassified (order: Acidimicrobiales)	0.4 (0.12-0.77)	0.7 (0.48,0.85)	0.09 (0.02,0.19)	No	No	100
unclassified (family: Acetobacteraceae)	1.35 (0.52-2.36)	0.7 (0.58,0.8)	0.15 (0.07,0.25)	No	No	100
unclassified (family: Pseudonocardiaceae)	0.01 (0-0.11)	0.69 (0.51,0.92)	0.16 (0.01,0.31)	Yes	No	58
unclassified (family: Isosphaeraceae)	0.67 (0.27-1.47)	0.69 (0.59,0.77)	0.09 (0.04,0.16)	No	No	100
unclassified (order: WD2101)	2.67 (1.51-4.17)	0.69 (0.58,0.78)	0.05 (0,0.15)	No	No	100
unclassified (family: Koribacteraceae)	8.05 (2.72-19.58)	0.69 (0.65,0.72)	0.11 (0.08,0.14)	No	No	100
unclassified (order: Phycisphaerales)	0.09 (0.01-0.23)	0.68 (0.41,0.85)	0.23 (0.04,1)	No	No	100
unclassified (family: Methylocystaceae)	0.73 (0.08-2.11)	0.67 (0.6,0.73)	0.08 (0.05,0.12)	No	No	100
unclassified (class: EC1113)	0.09 (0.01-0.23)	0.67 (0.28,0.92)	0.16 (0.01,1)	No	No	100
Rhodoplanes	3.35 (0.91-5.69)	0.67 (0.61,0.71)	0.09 (0.06,0.11)	No	No	100
Kibdelosporangium	0.05 (0-0.66)	0.66 (0.47,0.81)	0.15 (0.06,0.28)	No	No	100
unclassified (order: CV90)	0.01 (0-0.06)	0.66 (0.34,0.85)	0.26 (0.02,1)	No	No	100
unclassified (class: P2-11E)	0.03 (0-0.25)	0.66 (0.3,0.88)	0.31 (0.03,1)	No	No	70
unclassified (family: Nitrosomonadaceae)	0.01 (0-0.11)	0.65 (0.55,0.76)	0.2 (0.11,0.28)	No	No	100
unclassified (family: Myxococcaceae)	0.02 (0-0.12)	0.64 (0.48,0.77)	0.2 (0.1,0.33)	No	No	100
Decreasing with restoration						
Nocardioides	0.25 (0.02-0.7)	-0.72 (-0.84,-0.63)	0.03 (0.01,0.07)	No	No	100
[Clostridium](*)	0 (0-0.05)	-0.74 (-0.86,-0.48)	0 (0,0)	No	Yes	73
Turicibacter	0.08 (0-0.31)	-0.74 (-0.88,-0.56)	0.03 (0,0.12)	No	No	100
unclassified (family: Dolo_23)	0.08 (0-0.29)	-0.74 (-0.86,-0.61)	0.01 (0,0.03)	No	No	100
Pimelobacter	0.02 (0-0.11)	-0.75 (-0.85,-0.63)	0.05 (0,0.35)	No	Yes	100
unclassified (family: Peptostreptococcaceae)(*)	0.01 (0-0.09)	-0.75 (-0.86,-0.66)	0 (0,0.01)	No	No	99
unclassified (order: Sphaerobacterales)	0.11 (0-0.41)	-0.77 (-0.85,-0.66)	0.03 (0.01,0.07)	No	No	100
unclassified (order: JG30-KF-CM45)	0.05 (0-0.22)	-0.77 (-0.88,-0.61)	0.02 (0,0.06)	No	No	100

Genus (or nearest classified group)	Relative abundance (%); mean (range)	Correlation with revegetation age; mean (95% CI)	P-value for ordinal AOV; mean (95% CI)	Missing in cleared	Missing in remnants	% bootstraps with data
Geobacter	0.12 (0-0.49)	-0.78 (-0.85,-0.69)	0.01 (0,0.03)	No	No	100
Pseudonocardia(*)	0.08 (0-0.34)	-0.78 (-0.88,-0.66)	0 (0,0.01)	No	No	100
Solibacillus(*)	0.31 (0-1.53)	-0.78 (-0.8,-0.75)	0 (0,0)	No	No	100
unclassified (order: Bacillales)	0.02 (0-0.12)	-0.78 (-0.9,-0.66)	0.02 (0,0.11)	No	No	100
Pelosinus	0.01 (0-0.06)	-0.79 (-0.92,-0.61)	0.03 (0,0.21)	No	No	100
Arthrobacter(*)	0.06 (0-0.32)	-0.79 (-0.85,-0.71)	0 (0,0)	No	No	100
Caloramator	0.03 (0-0.14)	-0.79 (-0.93,-0.64)	0.05 (0,0.22)	No	No	100
unclassified (family: Nitrospiraceae)(*)	0.08 (0-0.36)	-0.8 (-0.86,-0.74)	0 (0,0.01)	No	No	100
unclassified (family: Nocardioidaceae)	0.04 (0-0.17)	-0.8 (-0.92,-0.64)	0.01 (0,0.06)	No	No	100
SMB53(*)	0.07 (0-0.37)	-0.81 (-0.85,-0.76)	0 (0,0)	No	No	100
Clostridium	0.27 (0.01-0.79)	-0.81 (-0.87,-0.72)	0.01 (0,0.03)	No	No	100
Coprococcus(*)	0.02 (0-0.15)	-0.82 (-0.91,-0.74)	0 (0,0.01)	No	No	100
unclassified (class: C0119)	0.32 (0.02-1.37)	-0.83 (-0.86,-0.79)	0 (0,0.01)	No	No	100
Flavisolibacter	0.12 (0-0.41)	-0.85 (-0.92,-0.75)	0.04 (0.01,0.08)	No	No	100
Ammoniphilus(*)	0.19 (0-0.71)	-0.85 (-0.88,-0.83)	0 (0,0)	No	No	100
unclassified (family: Ellin5301)	0.07 (0-0.29)	-0.85 (-0.94,-0.72)	0.01 (0,0.02)	No	No	100
Cytophagales	0.2 (0-0.96)	-0.85 (-0.9,-0.81)	0.03 (0.02,0.06)	No	No	100
Sporosarcina(*)	0.89 (0-3.24)	-0.85 (-0.87,-0.84)	0 (0,0)	No	No	100
unclassified (order: Ellin5290)	0.95 (0.21-2.34)	-0.86 (-0.89,-0.82)	0.01 (0,0.02)	No	No	100
unclassified (family: Actinospicaceae)	0.09 (0-0.29)	-0.86 (-0.92,-0.8)	0.01 (0,0.03)	No	No	100
Rummeliibacillus(*)	0.09 (0-0.42)	-0.87 (-0.92,-0.82)	0 (0,0)	No	No	100
Bacillus(*)	6.99 (1.07-15.42)	-0.95 (-0.95,-0.94)	0 (0,0)	No	No	100

Table S2. Taxa missing in cleared samples, increasing with restoration, and found in remnants at Mt Bold.

Genus (or nearest classified group)	Relative abundance (%); mean (range)	Correlation with revegetation age; mean (95% CI)	P-value for ordinal AOV; mean (95% CI)	Missing in cleared	Missing in remnants	% bootstraps with data
unclassified (family: [Leptospirillaceae])	0.019 (0-0.104)	0.756 (0.415,0.913)	0.107 (0.009,0.26)	Yes	No	97
unclassified (order: Acidobacteriales)	0.071 (0-0.399)	0.745 (0.649,0.848)	0.128 (0.048,0.211)	Yes	No	100
unclassified (family: Pseudonocardiaceae)	0.006 (0-0.107)	0.694 (0.506,0.919)	0.157 (0.005,0.315)	Yes	No	49
unclassified (family: Trebouxiophyceae)	0.008 (0-0.086)	0.593 (0.502,0.707)	0.265 (0.157,0.356)	Yes	No	99
FFCH10602	0.062 (0-0.497)	0.475 (0.429,0.537)	0.843 (0.32,1)	Yes	No	100
Beijerinckia	0.005 (0-0.101)	0.466 (0.413,0.559)	0.724 (0.263,1)	Yes	No	75
unclassified (phylum: Firmicutes)	0.01 (0-0.065)	0.212 (-0.083,0.515)	0.464 (0.033,1)	Yes	No	100
unclassified (family: [Chthoniobacteraceae])	0.013 (0-0.137)	0.131 (-0.016,0.341)	0.725 (0.22,1)	Yes	No	47

Table S3. Taxa found in cleared samples, decreasing with restoration, and missing in remnants at Mt Bold.

Genus (or nearest classified group)	Relative abundance (%); mean (range)	Correlation with revegetation age; mean (95% CI)	P-value for ordinal AOV; mean (95% CI)	Missing in cleared	Missing in remnants	% bootstraps with data
unclassified (family: Peptostreptococcaceae)(*)	0.015 (0-0.107)	-0.745 (-0.852,-0.633)	0.046 (0,0.352)	No	Yes	98
Pimelobacter	0.004 (0-0.048)	-0.74 (-0.864,-0.476)	0.004 (0,0)	No	Yes	100
[Clostridium](*)	0.018 (0-0.134)	-0.724 (-0.815,-0.593)	0.004 (0,0)	No	Yes	69
Segetibacter(*)	0.012 (0-0.101)	-0.72 (-0.764,-0.68)	0 (0,0)	No	Yes	83
Rhodanobacter	0.02 (0-0.181)	-0.714 (-0.794,-0.607)	0 (0,0)	No	Yes	33
unclassified (class: Thermomicrobia)(*)	0.003 (0-0.054)	-0.707 (-0.83,-0.506)	0.003 (0,0)	No	Yes	98
Alkaliphilus	0.005 (0-0.057)	-0.699 (-0.799,-0.443)	0 (0,0)	No	Yes	100
unclassified (family: Gracilibacteraceae)(*)	0.004 (0-0.051)	-0.691 (-0.744,-0.64)	0 (0,0)	No	Yes	35
Tissierella_Soehngenia(*)	0.004 (0-0.048)	-0.684 (-0.823,-0.509)	0 (0,0)	No	Yes	39
Tepidibacter(*)	0.004 (0-0.048)	-0.679 (-0.774,-0.48)	0.029 (0,0.134)	No	Yes	41
unclassified (order: Clostridiales)	0.009 (0-0.104)	-0.606 (-0.789,-0.401)	0.002 (0,0)	No	Yes	48
Terracoccus(*)	0.004 (0-0.045)	-0.557 (-0.8,-0.252)	0.253 (0,1)	No	Yes	97
Methylotenera(*)	0.003 (0-0.039)	-0.541 (-0.766,-0.384)	0 (0,0)	No	Yes	93
unclassified (family: Ruminococcaceae)	0.01 (0-0.14)	-0.527 (-0.667,-0.415)	0.005 (0,0.032)	No	Yes	79

Table S4. Top 30 increasing and top 30 decreasing taxa based on fold change between human-altered to natural samples (Australia-wide), from *DESeq2* analysis.

Genus (or nearest classified group) [OTU]	Phylum	Log2-fold-change	-Log10(P_{adj})
Increasing from human-altered to natural			
unclassified (order: Ellin6513)[16S_OTUa_974]	Acidobacteria	25.17	236.35
Rhodoplanes[16S_OTUa_1185]	Proteobacteria	9.39	29.42
Acidocella[16S_OTUa_99]	Proteobacteria	8.03	25.65
Acidocella[16S_OTUa_3518]	Proteobacteria	7.39	36.81
unclassified (order: Ellin6513)[16S_OTUa_2310]	Acidobacteria	7.33	20.93
Candidatus_Xiphinematobacter[16S_OTUa_970]	Verrucomicrobia	7.28	24.88
Acidocella[16S_OTUa_6357]	Proteobacteria	7.27	33.84
Mycobacterium[16S_OTUa_767]	Actinobacteria	7.2	30.73
unclassified (order: Ellin6513)[16S_OTUa_1996]	Acidobacteria	7.14	15.56
Rhodoplanes[16S_OTUa_1713]	Proteobacteria	7.08	16.46
unclassified (order: Ellin6513)[16S_OTUa_291]	Acidobacteria	6.87	25.66
unclassified (family: Acidobacteriaceae)[16S_OTUa_55]	Acidobacteria	6.75	62.11
unclassified (family: Koribacteraceae)[16S_OTUa_614]	Acidobacteria	6.66	11.97
unclassified (family: Rhodospirillaceae)[16S_OTUa_93]	Proteobacteria	6.58	22.41
unclassified (order: Ellin6513)[16S_OTUa_313]	Acidobacteria	6.43	34.8
unclassified (family: Rhodospirillaceae)[16S_OTUa_569]	Proteobacteria	6.39	23.16
Candidatus_Xiphinematobacter[16S_OTUa_1342]	Verrucomicrobia	6.37	17.5
DA101[16S_OTUa_678]	Verrucomicrobia	6.37	15.72
unclassified (order: Ellin329)[16S_OTUa_1462]	Proteobacteria	6.35	15.35
unclassified (family: auto67_4W)[16S_OTUa_3078]	Verrucomicrobia	6.32	15.43
Devosia[16S_OTUa_993]	Proteobacteria	6.32	18.62
DA101[16S_OTUa_198]	Verrucomicrobia	6.28	22.47
unclassified (family: Isosphaeraceae)[16S_OTUa_265]	Planctomycetes	6.17	39.26
Candidatus_Xiphinematobacter[16S_OTUa_2140]	Verrucomicrobia	6.17	22.46
Actinomadura[16S_OTUa_638]	Actinobacteria	6.11	23.46
unclassified (order: Ellin6513)[16S_OTUa_2477]	Acidobacteria	6.1	15.22
Candidatus_Xiphinematobacter[16S_OTUa_332]	Verrucomicrobia	6.09	14.88
unclassified (order: Ellin6513)[16S_OTUa_1419]	Acidobacteria	6	16.59
Alicyclobacillus[16S_OTUa_1942]	Firmicutes	5.99	8.99
unclassified (family: Isosphaeraceae)[16S_OTUa_3810]	Planctomycetes	5.94	19.99

Decreasing from human-altered to natural			
unclassified (family: [Entotheonellaceae])[16S_OTUa_736]	Proteobacteria	-10.32	21.51
unclassified (family: Ellin6075)[16S_OTUa_91612]	Acidobacteria	-10.36	36.72
Adhaeribacter[16S_OTUa_3542]	Bacteroidetes	-10.45	18.38
unclassified (family: A4b)[16S_OTUa_1930]	Chloroflexi	-10.52	17.97
unclassified (order: WD2101)[16S_OTUa_935]	Planctomycetes	-10.53	26.42
unclassified (order: iii1-15)[16S_OTUa_1598]	Acidobacteria	-10.77	22.68
unclassified (family: Chitinophagaceae)[16S_OTUa_955]	Bacteroidetes	-11.35	27.06
unclassified (family: Syntrophobacteraceae)[16S_OTUa_1737]	Proteobacteria	-11.48	25.21
unclassified (order: Ellin5290)[16S_OTUa_1412]	Gemmatimonadetes	-11.49	26.33
Pontibacter[16S_OTUa_3777]	Bacteroidetes	-11.53	11.54
unclassified (family: Ellin5301)[16S_OTUa_1476]	Gemmatimonadetes	-11.55	24.38
unclassified (family: Peptostreptococcaceae)[16S_OTUa_2064]	Firmicutes	-11.61	29.96
unclassified (family: Ectothiorhodospiraceae)[16S_OTUa_659]	Proteobacteria	-11.75	30.62
Pontibacter[16S_OTUa_2658]	Bacteroidetes	-11.9	28.44
Aquicella[16S_OTUa_1264]	Proteobacteria	-11.99	16.23
Aquicella[16S_OTUa_624]	Proteobacteria	-12.25	15.94
Aquicella[16S_OTUa_1082]	Proteobacteria	-13.21	20.83
Salinimicrobium[16S_OTUa_612]	Bacteroidetes	-13.97	16.08
unclassified (family: Ellin517)[16S_OTUa_2924]	Verrucomicrobia	-25.44	82.56
Flavisolibacter[16S_OTUa_2059]	Bacteroidetes	-25.95	98.8
Adhaeribacter[16S_OTUa_1884]	Bacteroidetes	-26.01	159.12
unclassified (family: Cytophagaceae)[16S_OTUa_3145]	Bacteroidetes	-26.15	73.26
unclassified (order: JG30-KF-CM45)[16S_OTUa_1267]	Chloroflexi	-26.16	127.72
unclassified (class: C0119)[16S_OTUa_1637]	Chloroflexi	-26.18	78.79
Kouleothrix[16S_OTUa_1325]	Chloroflexi	-26.2	86.09
unclassified (family: Syntrophobacteraceae)[16S_OTUa_1686]	Proteobacteria	-26.49	96.56
Agromyces[16S_OTUa_3809]	Actinobacteria	-26.56	109.21
unclassified (class: C0119)[16S_OTUa_1094]	Chloroflexi	-26.7	80.46
Kouleothrix[16S_OTUa_576]	Chloroflexi	-27.56	84.15
unclassified (family: A4b)[16S_OTUa_803]	Chloroflexi	-27.67	97.25

Table S5. Summary statistics for human-associated taxa identified from the Mt Bold restoration gradient, with corresponding differential abundance data in human-altered samples (n = 78) versus natural samples (n = 139) elsewhere in Australia.

Genus or group	Phylum	Mt Bold correlation with reveg age: Mean (95%CI)	Australia-wide Mean log2-fold-change in OTU abundance: human-altered to natural	Australia-wide human-altered samples OTU relative abundance: Median (IQR) (%)	Australia-wide natural samples OTU relative abundance: Median (IQR) (%)	Mt Bold trend with restoration/ Australia-wide trend human-altered vs natural	Wilcoxon one-sided rank-sum test *P*-value†
Sphingomonas	Proteobacteria	0.56 (0.3, 0.77)	0.1	0.081 (0.033, 0.116)	0.046 (0.023, 0.107)	Increasing / NS	0.975
Mycobacterium	Actinobacteria	0.39 (0.29, 0.46)	1.31	0.208 (0.046, 0.42)	0.963 (0.513, 1.729)	Increasing / Increasing	< 0.001
Actinomadura	Actinobacteria	0.32 (-0.07, 0.58)	0.87	0.016 (0.007, 0.036)	0.129 (0.037, 0.315)	Increasing / Increasing	< 0.001
Burkholderia	Proteobacteria	0.12 (0.02, 0.2)	2.05	0.047 (0.006, 0.251)	0.838 (0.203, 1.606)	Increasing / Increasing	< 0.001
Amycolatopsis	Actinobacteria	0.03 (-0.36, 0.37)	0.33	0.007 (0.002, 0.03)	0.008 (0, 0.038)	Increasing / NS	0.38
Bacillus	Firmicutes	-0.95 (-0.95, -0.94)	-1.04	4.499 (2.622, 9.865)	1.706 (0.52, 2.956)	Decreasing / Decreasing	< 0.001
Coprococcus	Firmicutes	-0.82 (-0.91, -0.74)	-1.49	0 (0, 0.016)	0 (0, 0.004)	Decreasing / NS	0.09
Clostridium	Firmicutes	-0.81 (-0.87, -0.72)	-1.81	0.104 (0.039, 0.27)	0.023 (0.002, 0.061)	Decreasing / Decreasing	< 0.001
Rhodococcus	Actinobacteria	-0.65 (-0.84, -0.44)	-1.21	0.032 (0.014, 0.073)	0.011 (0, 0.04)	Decreasing / Decreasing	< 0.001
Streptomyces	Actinobacteria	-0.61 (-0.66, -0.55)	-1.18	0.674 (0.446, 1.133)	0.283 (0.152, 0.694)	Decreasing / Decreasing	< 0.001
Legionella	Proteobacteria	-0.47 (-0.58, -0.37)	-0.46	0.005 (0, 0.013)	0.002 (0, 0.006)	Decreasing / Decreasing	0.004
Achromobacter	Proteobacteria	-0.45 (-0.61, -0.28)	-3.23	0.003 (0, 0.015)	0 (0, 0.001)	Decreasing / Decreasing	< 0.001
Bacteroides	Bacteroidetes	-0.29 (-0.3, -0.28)	-2.88	0 (0, 0)	0 (0, 0)	Decreasing / Decreasing	0.004
Flavobacterium	Bacteroidetes	-0.25 (-0.32, -0.17)	-0.92	0.07 (0.014, 0.307)	0.029 (0, 0.323)	Decreasing / Decreasing	0.025
Enterobacter	Proteobacteria	-0.21 (-0.33, -0.11)	-0.88	0.002 (0, 0.011)	0 (0, 0.005)	Decreasing / Decreasing	0.037
Pseudomonas	Proteobacteria	-0.17 (-0.28, -0.07)	-0.91	0.071 (0.021, 0.222)	0.053 (0.006, 0.129)	Decreasing / Decreasing	0.049
Sphingobacterium	Bacteroidetes	-0.16 (-0.22, -0.11)	-1.44	0 (0, 0.004)	0 (0, 0)	Decreasing / Decreasing	< 0.001
Chryseobacterium	Bacteroidetes	-0.14 (-0.2, -0.08)	-0.92	0 (0, 0.007)	0 (0, 0)	Decreasing / Decreasing	0.004

†Wilcoxon one-sided rank sum tests for difference in OTU relative abundance between Australia-wide human-altered vs. natural samples, set in the direction suggested by results from restoration at Mt Bold (also see Fig. S16-S17). NS = no significant difference.

Supplementary References

Gellie, N.J.C.; Mills, J.G.; Breed, M.F.; Lowe, A.J., 2017. Revegetation rewilds the soil bacterial microbiome of an old field. *Molecular Ecology*. 26, 2895-2904

Jost, L., 2006. Entropy and diversity. *Oikos*. 113, 363-375

Statement of Authorship

Title of Paper	Naturally-diverse airborne environmental microbial exposures modulate the gut microbiome and reduce anxiety-like behaviour in mice
Publication Status	☐ Published ☐ Accepted for Publication ☑ Submitted for Publication ☐ Unpublished and Unsubmitted w ork w ritten in manuscript style
Publication Details	Liddicoat, C., Sydnor, H., Cando-Dumancela, C., Dresken, R., Liu, J., Gellie, N., Mills, J., Young, J., Weyrich, L., Hutchinson, M., Weinstein, P., Breed, M., [In Review]. Naturally-diverse airborne environmental microbial exposures modulate the gut microbiome and reduce anxiety-like behaviour in mice.

Principal Author

Name of Principal Author (Candidate)	Craig Liddicoat		
Contribution to the Paper	I led the conceptualisation, project design, experimental design and setup, sample and data collection, contributed to DNA laboratory sample processing, led the data analysis and interpretation, wrote the first draft, and led the revisions of the final manuscript.		
Overall percentage (%)	75		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	7/5/2019.

Co-Author Contributions

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

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

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

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

Name of Co-Author	Harrison Sydnor		
Contribution to the Paper	Contributed to the conceptualisation, experimental design and setup, sample and data collection, DNA laboratory sample processing, analysis and interpretation of results, paper writing and revisions.		
Signature		Date	17/5/2019

Name of Co-Author	Christian Cando-Dumancela		
Contribution to the Paper	Contributed to the experimental design and setup, sample and data collection, DNA laboratory sample processing, analysis and interpretation of results, paper writing and revisions.		
Signature		Date	17/5/2019

Name of Co-Author	Romy Dresken		
Contribution to the Paper	Contributed to the experimental design and setup, sample and data collection, DNA laboratory sample processing, analysis and interpretation of results, paper writing and revisions.		
Signature		Date	24-05-2019

Name of Co-Author	Jiajun Liu		
Contribution to the Paper	Supervised elements of the experimental design and setup (i.e. animal behaviour testing, surgical procedures and sample processing), contributed to sample and data collection, analysis and interpretation of results, paper writing and revisions.		
Signature		Date	17/06/2019

Name of Co-Author	Nick Gellie		
Contribution to the Paper	Contributed to the sample and data collection, analysis and interpretation of results, paper writing and revisions.		
Signature		Date	17/05/2019

Name of Co-Author	Jacob Mills		
Contribution to the Paper	Contributed to the sample and data collection, analysis and interpretation of results, paper writing and revisions.		
Signature		Date	14/06/2019

Name of Co-Author	Jennifer Young		
Contribution to the Paper	Supervised elements of the experimental design and setup (i.e. experimental hygiene, minimising contamination), supervised DNA laboratory sample processing, contributed to analysis and interpretation of results, paper writing and revisions.		
Signature		Date	24-05-19

Name of Co-Author	Laura Weyrich		
Contribution to the Paper	Supervised elements of the experimental design and setup (i.e. experimental hygiene, minimising contamination, surgical procedures), contributed to analysis and interpretation of results, paper writing and revisions.		
Signature		Date	29/5/19

Name of Co-Author	Mark Hutchinson
Contribution to the Paper	Contributed to the conceptualisation, experimental design, analysis and interpretation of results, paper writing and revisions.
Signature	Date: 20 May 2019

Name of Co-Author	Philip Weinstein
Contribution to the Paper	Contributed to the conceptualisation, experimental design, analysis and interpretation of results, paper writing and revisions.
Signature	Date: 21/5/19

Name of Co-Author	Martin Breed
Contribution to the Paper	Contributed to the conceptualisation, experimental design, analysis and interpretation of results, paper writing and revisions.
Signature	Date: 7/5/19

Please cut and paste additional co-author panels here as required.

Title: Naturally-diverse airborne environmental microbial exposures modulate the gut microbiome and reduce anxiety-like behaviour in mice

Authors: Craig Liddicoat[1,a], Harrison Sydnor[1], Christian Cando-Dumancela[1], Romy Dresken[1], Jiajun Liu[2,3], Nicholas J.C. Gellie[1], Jacob G. Mills[1], Jennifer M. Young[1,4], Laura S. Weyrich[5,6], Mark R. Hutchinson[2,3], Philip Weinstein[1], Martin F. Breed[1,a]

Affiliations:

[1]School of Biological Sciences and the Environment Institute, The University of Adelaide, Adelaide, South Australia 5005, Australia

[2]Adelaide Medical School, The University of Adelaide, Adelaide, South Australia 5005, Australia

[3]Australian Research Council Centre of Excellence for Nanoscale BioPhotonics, The University of Adelaide, Adelaide, South Australia 5005, Australia

[4]College of Science and Engineering, Flinders University, Bedford Park, South Australia 5042, Australia

[5]Australian Centre for Ancient DNA, The University of Adelaide, Adelaide, South Australia 5005, Australia

[6]Australian Research Council Centre of Excellence for Australian Biodiversity and Heritage, The University of Adelaide, Adelaide, South Australia 5005, Australia

[a]Authors for correspondence: Craig Liddicoat, craig.liddicoat@adelaide.edu.au; and Martin Breed, martin.breed@adelaide.edu.au

Open Researcher and Contributor IDs (ORCIDs): CL, 0000-0002-4812-7524; JL, 0000-0003-1887-0218; NJCG, 0000-0001-9761-8832; JGM, 0000-0001-6713-0035; LSW, 0000-0001-5243-4634; MRH, 0000-0003-2154-5950; PW, 0000-0001-9860-7166; MFB, 0000-0001-7810-9696

Keywords: microbiome, environmental health, biodiversity hypothesis, microbial old friends, mental health, butyrate

Abstract

Human populations are losing contact with nature. Yet, growing epidemiological evidence links natural green space exposure with a range of health benefits, including for mental health. Indeed, greater urbanisation associates with increased risk of mental health disorders. Microbiomes are proposed as an important but understudied link that may help explain many beneficial natural green space-human health associations. However, there remains a lack of controlled experimental evidence testing possible beneficial effects from passive exposure to natural biodiversity via airborne microbes. Previous mouse model studies have used unrealistic environmental microbial exposures—including excessive soil and organic matter contact, feed supplements and injections—to demonstrate host microbiota, immune biomarker, and behavioural changes. Here we show changes to mouse gut microbiota and reduced anxiety-like behaviour in female mice due to trace-level dust exposures from high vs. low biodiversity soils vs. no soil (control). We provide evidence of a potential beneficial mechanistic link between natural biodiversity, gut health and mental health, via soil-derived butyrate-producing bacteria. These bacteria increased with high biodiversity treatments and appeared to moderate anxiety-like behaviour in female mice. Our results will have implications for cost-effective population health interventions through microbiome-conscious green space design, and the potential mainstreaming of biodiversity into health care.

1. Introduction

The influence of environmental microbial communities (microbiotas) and associated genetic material (microbiomes) on human health represents an important knowledge gap with potentially far-reaching implications for cost-effective public health interventions, and the management, design and use of our natural and built environments [1, 2]. Environments are a key factor in shaping our human (e.g., skin, airway, gut) microbiota [3] and immune system [4], particularly from an early age [5]. Gut microbiota are connected to mammalian health and disease at distant body locations (e.g., via immune signalling) [6], and also influence brain development and behaviour [7]. Because ambient airborne environmental microbiota may interact with and supplement our gut microbiota [2, 8], differing environments and their characteristic microbiota (e.g., [9]) have potential to influence neurodevelopment and mental health via the bidirectional brain-gut-microbiome axis [10].

It is important to build knowledge of possible mechanisms that may underpin associations found between green space exposure and mental health (e.g., [11-16]), given the rapid rates of global urbanisation [17]. Beyond mental health, the diversity of health benefits associated with nature contact suggests that a broad, nonspecific physiological pathway of action, a multiplicity of pathways, or a combination of these, may be present [18]. Due to their ubiquitous and immunomodulatory nature, environmental microbiota are suggested to be part of the causal connection [1, 2], and offer promise for cost-effective solutions [19, 20], to the rapidly increasing prevalence of many modern diseases. Indeed, health benefits from nature contact, biodiversity and microbial diversity exposures have been found in humans [21-23] and animal models [24-26]. However, support for a tangible biological link between nature contact and mental health is limited [27-29].

In particular, there is little controlled experimental evidence testing for changes in host gut microbiota and mental health outcomes from normal passive exposure to ambient natural

biodiversity. Previous studies have used unrealistic environmental microbiota exposures—including excessive soil and organic matter contact (e.g., ~30-50 g.mouse^{-1}.week^{-1} [24, 30]; electronic supplementary material, Supplementary Methods), forced feeding [25] and injections [26]—to demonstrate changes in host microbiota, immune biomarkers, and anxiety-related behaviour.

Here we sought evidence of potential mechanistic links between passive exposure to ambient natural biodiversity, environmental and gut microbiota, and the intriguing outcome of mental health. Specifically, we tested the hypothesis that varying airborne microbial exposures, spanning a gradient from low to high natural biodiversity, may have differential impact on the gut microbiome and anxiety-like behaviour in mice. To achieve this, we ran a randomised controlled study in mice subjected to trace-level soil dust exposures from high biodiversity soil, low biodiversity soil, or no soil (hereafter referred to as *high*, *low* and *control*). Mice were housed in open wire-top cages, with each cage residing in a single environmental enclosure. After a 7 week exposure, changes to gut microbiota were characterised by bacterial 16S rRNA gene sequences and aspects of anxiety-like behaviour were assessed using Open Field and Elevated Plus Maze apparatus [31].

2. Methods

(a) Study design

We subjected 54 weaned, inbred 3-5 week old specific-pathogen-free BALB/C mice to a 7-week exposure of trace-level airborne dust from soils that reflected both a macro- and micro-biodiversity gradient. Same-sex groups of three mice were housed in open wire-top cages, with each cage residing in a larger individually-isolated environmental treatment enclosure (electronic supplementary material, figure S1 and S2). Each soil biodiversity treatment (i.e. *high, low*, *control*) was replicated in 6 enclosures, comprising 3 all-female and 3 all-male

enclosures. *Low* soils came from a low plant macro-diversity setting (electronic supplementary material, figure S3 and S4) and had low microbial diversity (figure 1). Likewise, *high* soils came from a high plant macro-diversity setting and had high microbial diversity. Source soils were homogenised and 1.75 kg was spread over a shallow tray in each soil treatment enclosure. We chose a no-soil *control* to mimic regularly sanitised built environments. Ten cm diameter USB fans adjacent to the trays were used on a 2-hr on/2-hr off cycle to generate the light dust exposures. To focus on the potential influence of airborne microbes (or aerobiology), we controlled for known microbiota effects including genetics, birth mode, breeding facility, diet, age, light and sleep cycles, and non-treatment environmental parameters. Littermates were randomised across treatments to help normalise the starting microbiotas (electronic supplementary material, figure S5). Possible gender effects were unknown, so we included an equal number of female and male mice across all treatments.

(b) Data collection

Soil, air (dust), faecal and caecal samples were taken to assess microbiota changes over the 7-week study (electronic supplementary material, figure S6). Microbiota samples were assessed using bacterial 16S rRNA (V3-V4) marker gene survey data, clustered into operational taxonomic units (OTUs) with ≥ 97% sequence similarity. Anxiety-like behaviour was assessed by analysing time spent in the anxiety-provoking zones of the Open Field and Elevated Plus Maze apparatus. We considered the centre zone in the Open Field, and the centre and open arms in the Elevated Plus Maze, as the anxiety-provoking zones. Details of the enclosure design, mice strain and husbandry, sampling procedures, DNA extraction, PCR and sequencing are provided in electronic supplementary material, Supplementary Methods.

(c) Bioinformatic and statistical analyses

The processing of 16S rRNA gene sequences, formation of OTUs, taxonomic assignments, data exclusions and decontamination are described in electronic supplementary material, Supplementary Methods. Microbiota compositions (beta diversity) were visualised using non-metric multidimensional scaling (NMDS) ordination of Bray-Curtis distances based on rarefied OTU abundances. To test for compositional differences between microbiota sample groups we used permutational multivariate analysis of variance (PERMANOVA), followed by testing for homogeneity of group dispersions. We estimated OTU alpha diversity based on rarefied abundances and used the exponential transform of Shannon Index values to derive the effective number of OTUs [32]. Merged-sample bootstrap resampling [33] was also used to visualise alpha diversity across related samples. We used the Kruskal-Wallis rank sum test (no. groups ≥ 3) or the Wilcoxon rank sum test (no. groups = 2) for testing of statistical differences between groups for microbiome and behaviour data. All such tests excluded outliers, identified as any data points beyond 1.5 times the interquartile range above the upper quartile or below the lower quartile, as defined in the default R *boxplot()* function in base R [34]. We assessed differentially abundant OTUs between week 0 and week 7 faecal samples within each treatment using the *DESeq2* algorithm [35], which applies the Benjamini-Hochberg method to control false discoveries. OTUs that were identified as increasing within each treatment were assigned to a putative species based on the closest match in the NCBI 16S database. Heatmaps were used to visualise the relative enrichment of increasing OTUs within treatments, and to visualise microbiota variance across the ‘1/3 most anxious’ and ‘1/3 least anxious’ groupings of female mice based on the bottom third and top third of Open Field centre time results. Further details are provided in the electronic supplementary material, Supplementary Methods.

3. Results

(a) Biodiverse soil dust exposure modulates gut microbiota

The environmental treatments influenced the composition of mice gut microbiota (figure 2), despite very low soil dust exposures (~0.0034 g soil.mouse^{-1}.week^{-1}; electronic supplementary material, figure S7); several thousand times less than previous studies. Treatment soil and air samples displayed distinct bacterial signatures (figure 1; electronic supplementary material, figure S8*a*,*b*). The treatments explained 5.5% and 4.2% of the variation in week 7 faecal and caecal microbiota respectively (figure 2). As expected, we found large cage/enclosure effects on gut microbiota, explaining 55% and 58% of variation in week 7 faecal and caecal microbiota respectively (figure 2). Here, cage and enclosure represent the same level of co-housing. Mouse gender did not help explain the composition of week 7 faecal or caecal microbiota and was removed from PERMANOVA models. We found that all *high* females showed an increase ($\Delta > 0$) in faecal alpha diversity within individual animals between week 0 to week 7, in contrast to *control* and *low* females (electronic supplementary material, figure S9). Also, there was a rising pattern in week 7 faecal alpha diversity towards the *high* females that bordered on significance (Kruskal-Wallis $\chi^2 = 5.58$, df = 2, $P = 0.06$, n = 8 to 9 per group; electronic supplementary material, figure S10). No trend was found in caecal alpha diversity (electronic supplementary material, figure S11).

(b) High biodiversity exposure reduced anxiety-like behaviour in females

We observed reduced anxiety-like behaviour in females only with the higher biodiversity treatments, as assessed by time spent in the centre of the Open Field (Kruskal-Wallis $\chi^2 = 8.08$, df = 2, $P = 0.018$, n = 7 to 9 per group; figure 3). In a different assessment of anxiety-related behaviour in the Elevated Plus Maze, we found no overall trend by gender or

treatment (electronic supplementary material, figure S12). Total distances travelled and counts of entries into anxiety-provoking zones showed no trend by gender or treatment (electronic supplementary material, figure S13 and S14), indicating that the Open Field reduced anxiety-like behavioural response in females to our *high* biodiversity treatment was not due to underlying differences in activity levels.

(c) Anxiety-like behaviour associated with gut microbiota

We found a connection between anxiety-like behaviour and faecal microbiota at the microbial community level. We compared females with the top third of Open Field centre times ('1/3 least anxious') to the bottom third ('1/3 most anxious'), and found their week 7 faecal microbiota were different (figure 4*a*). Then using the log relative abundance of week 7 faecal OTUs across these individuals to drive a clustering process, we saw three microbiota-behaviour groupings emerge: (i) a predominantly less anxious-*high* treatment grouping, (ii) an anxious-*control* dominated grouping, and (iii) a mixed behaviour-*control* and *low* grouping (figure 4*b*). No such differentiation in microbial community compositions by the anxiety groupings was observed in the caecal samples.

(d) Putative butyrate-producing environmental bacteria increased most in high biodiversity mice

We found evidence supporting a microbial link between airborne exposure to high biodiversity soils, subsequent changes to gut microbiota and reductions in anxiety-like behaviour, via soil-derived butyrate-producing bacteria. We identified the differentially abundant taxa between week 0 and week 7 faecal microbiota separately within each treatment (electronic supplementary material, figure S15 and S16, Tables S1–S3). Among many

recognised intestinal bacteria, we found an environmentally-derived, putative butyrate-producer, OTU 37, in *high* biodiversity soils and in corresponding air, faecal and caecal samples, which showed increasing relative abundance and prevalence among *high* faecal samples between week 0 and week 7 (electronic supplementary material, Table S3); the treatment that produced less anxious female mice. This OTU shared 97% sequence similarity to the closest match in the NCBI 16S database, identified as the spore-forming, anaerobic, butyrate-producing environmental bacteria *Kineothrix alysoides* [36]. Three other taxa (OTU 60, OTU 5520, OTU 5790), with 95-96% identity to the closest NCBI match of *K. alysoides*, also exhibited similar increasing relative abundance and prevalence within *high* faecal samples.

Additional taxa with the closest NCBI match to the soil-inhabiting environmental microbe *K. alysoides* were also present, and increased to a lesser degree in *control* and *low* faecal samples (electronic supplementary material, Tables S1 and S2). However, the strongest increase of this putative species was found in the *high* treatment, where they were largely detected at higher mean % OTU relative abundance (compared to *low* or *control* samples) and more often than not in 100% of post-exposure animals (figure 5).

(e) Putative butyrate-producing environmental bacteria may moderate the most anxious behaviour in female mice

We examined associations between anxiety-like behaviour in the female mice and the total % OTU relative abundance (in week 7 faecal and caecal samples) of all *K. alysoides*-like OTUs that increased in faecal samples during the experiment. From the caecal samples, we found strong positive correlations (e.g., $r = 0.90$, for $n = 2$ mice per treatment; $r = 0.76$ for $n = 4$ mice per treatment; figure 6) between the total % OTU relative abundance of *K. alysoides*-like OTUs and Open Field centre times for the most anxious mice across the treatments. Similarly,

a moderate correlation ($r = 0.58$, $n = 9$) was observed for the '1/3 most anxious' females as described in figure 4. Weaker correlations were observed as more animals, or less anxious animals, were considered, i.e., as the focus moved away from animals with the most anxious behaviours (figure 6). No such relationships were observed for this taxon of interest in the week 7 faecal samples.

4. Discussion

Our results provide evidence for a plausible biological link between exposure to trace-levels of natural biodiversity, gut health and mental health, via soil-derived butyrate-producing bacteria. This result is consistent with the large number of human health benefits, particularly to mental health, associated with nature contact (e.g., [11-16]). We suggest that such a beneficial biological linkage may be provided by airborne soil-derived butyrate-producing bacteria in combination with a naturally-diverse environmental microbial community. Butyrate is a short-chain fatty acid that is essential to gut health [37] and linked to immunological homeostasis [38], protection from metabolic diseases [39], and improved quality of life and reduced depression [40]. There are multiple pathways to butyrate production [41], so an array of supporting microbes is likely to be important. Indeed, having greater diversity of butyrate-producing bacteria may support functional stability through life disturbances, such as periods of antibiotic treatment and disease [41]. Butyrate is also a key intermediate in the breakdown of organic matter in anaerobic soils [42]. Anaerobic conditions can impact well-drained soils to some extent, particularly under conditions of high moisture and organic (oxygen-consuming) content [43]. We might expect natural spaces with high aboveground diversity to associate with increased soil bacterial diversity [44, 45] and increased soil organic matter [46]. Also, soil-associated microbes represent a large component of the aerobiology derived from environments [47]. Therefore, it is plausible that normal

passive exposures to biodiverse, natural environments may often involve contact with diverse airborne soil microbiota including butyrate-producers. Such contact might be enhanced through active interactions, such as environmental volunteering (e.g., planting, weeding) or working in biodiverse, organic-rich gardens or farms.

Spore-forming bacteria (such as *K. alysoides*) can persist and be transported in aerobic environments and then activate under anaerobic conditions. Spore-formers also dominate the human gut, comprising 50-60% of bacterial genera [48], and their capacity to survive in aerobic external environments allows them to be shared widely [49]. Some spore-forming bacteria might be capable of performing key roles across different habitats, such as interchangeably within soils and within the mammalian gut. In our study, we showed that the *K. alysoides*-like OTUs increased to varying degrees in the faecal samples from all treatments, suggesting that gut microbiota may have an affinity for these nominally soil-associated organisms. We know that poor diets can induce individual and intergenerational loss of key gut microbiota [50], with flow-on impacts to health. Our results suggest there may be an opportunity to supply key bacteria, such as butyrate-producers, from exposure to natural biodiversity.

The microbiota-behaviour relationships we observed suggest that contact with biodiverse environmental microbiota including butyrate-producers may play a role in protecting mental health, particularly in female mice that are most susceptible to anxiety-like behaviour. Gender differences in anxiety-related behaviour in BALB/C mice are not reported elsewhere [51, 52], so the Open Field reduced anxiety-like behavioural response within females to our *high* biodiversity treatment is novel and intriguing. This is a particularly interesting result given the consistent large-scale epidemiological evidence for higher prevalence rates of anxiety disorders in women compared to men [53], and points to opportunities for nature-based public health interventions.

Our results add to known links between gut microbiota, brain function and behavioural outcomes [7]. For example, early life exposure to normal gut microbiota has been shown to affect brain development and behaviour [54]. Particular gut microbiome profiles have been linked to autism spectrum disorder-like behaviours in mice [55]. Also, certain probiotic bacteria have been shown to reduce anxiety in mice [56]. Our study contrasts with previous work (e.g. [54]) where sterile germ-free mice displayed reduced anxiety compared to specific-pathogen-free mice with normal gut microbiota. Here, we started with specific-pathogen-free mice and found reduced anxiety (in females) following exposure to naturally-diverse environmental microbiota. We used a specific-pathogen-free mouse model to avoid the artificially stunted host microbiome of sterile (e.g., germ-free) animals, while providing greater control over known key microbiome-influencing factors than is possible in human studies.

Our study also has some limitations. The biodiversity treatment explained a relatively low percentage of variation in the gut microbiota. However, rare taxa can play important ecological roles [57], and our results will help to build knowledge towards the construction of environmental dose-health outcome response relationships. We observed a large influence on gut microbiota from our cohousing variable of cages/enclosures. This result reflects existing knowledge that cage effects are a known major influence on gut microbiota in mice studies [58]. It was beyond the scope of our study to examine causal mechanisms that may underpin the gender differences observed in the Open Field behaviour test results. Our Elevated Plus Maze and Open Field results did not align, however these apparatus often display discordant results and may encompass multiple dimensions of anxiety-related behaviour [31]. We only analysed bacterial microbiomes in this study, however, other agents (e.g., fungi, viruses, volatile organic compounds) may have a modulating influence on the gut microbiome and host health. Also, spore-forming environmental taxa have historically been underrepresented

in culture-independent studies due to greater resistance to universal DNA extraction techniques [59], and the generally low OTU relative abundances for the putative *K. alysoides* found here may reflect this limitation. We observed patterns of interest in both faecal and caecal samples. For example, microbiota composition and OTU alpha diversity patterns emerged in faecal samples, while correlations between anxiety-like behaviour and *K. alysoides* were apparent in caecal samples. We speculate that varying chemistry, redox and other conditions in faecal vs. caecal samples [60] may contribute to the observed differences in results for community-level vs. key taxa analyses.

Overall, our findings support the notion that microbiota-mediated health benefits from biodiverse nature contact may arise from a microbial community influence in which a range of beneficial and complementary microbes may be present. We note that more biodiverse microbiota offer a greater variety of organisms that may support: ecological control of potential pathogens, resilience to microbiota disturbances, immune priming, and key functional pathways such as butyrate production. We have shown links between exposure to the aerobiology from biodiverse green spaces (via soil microbiota) and changes to gut microbiota and anxiety-related behaviour in mice, with a resulting increase in gut microbiota of putative spore-forming soil-derived butyrate-producing bacteria. Such butyrate-producing bacteria provide a plausible explanation for the reduced anxiety-like behaviour observed here, and may help explain beneficial biodiverse green space-human health and mental health associations seen elsewhere.

Data availability. All data and code used in this study are available on *figshare* at http://doi.org/10.25909/5caaf18c3450d. All 16S rRNA gene sequences have been deposited in the European Nucleotide Archive (accession no. PRJEB31983).

Note (to be deleted): For review purposes, the supporting data and code can be reviewed at: https://adelaide.figshare.com/s/cef9e2e25c17aa7daf19 . Sequence data in the ENA study accession no. PRJEB31983 will be made freely available with the publication of the study.

Ethics Statement. All animal procedures performed in this study were approved by the University of Adelaide Animal Ethics Committee (Approval no. S-2017-112).

Acknowledgments. We thank Lab Animal Services, the Advanced DNA Identification and Forensic Facility, and the Environment Institute from the University of Adelaide. Craig Liddicoat received support from an Australian Government Research Training Program Scholarship. Martin Breed is supported by the Australian Research Council. We also thank Leanne McGrath, Naga Kasinadhuni and Christopher Noune from AGRF; Gail Anderson, Jaimee Spurr, Tiffany Boehm, Jonathan Jacobsen, Stefan Musolino, Kelsi Dodds, Ashley Grant, Korjent van Dijk, Andrew Lowe and Michelle Waycott for helpful discussions and input; and Shaun Kennedy and SA Water for access and guidance for source soil locations.

Author contributions. C.L., H.S., P.W. and M.B. conceived the research; C.L., H.S., C. C-D., R.D., J.L., N.G., J.M., J.Y., L.W., M.H., P.W. and M.B. designed the study; C.L., H.S., C. C-D., R.D., J.L., N.G., and J.M. conducted experiments and collected data; C.L., H.S., M.H., P.W., M.B. analyzed data; C.L. drafted the manuscript. All authors edited the manuscript and gave final approval for publication.

Competing interests. The authors declare no competing interests.

Figures

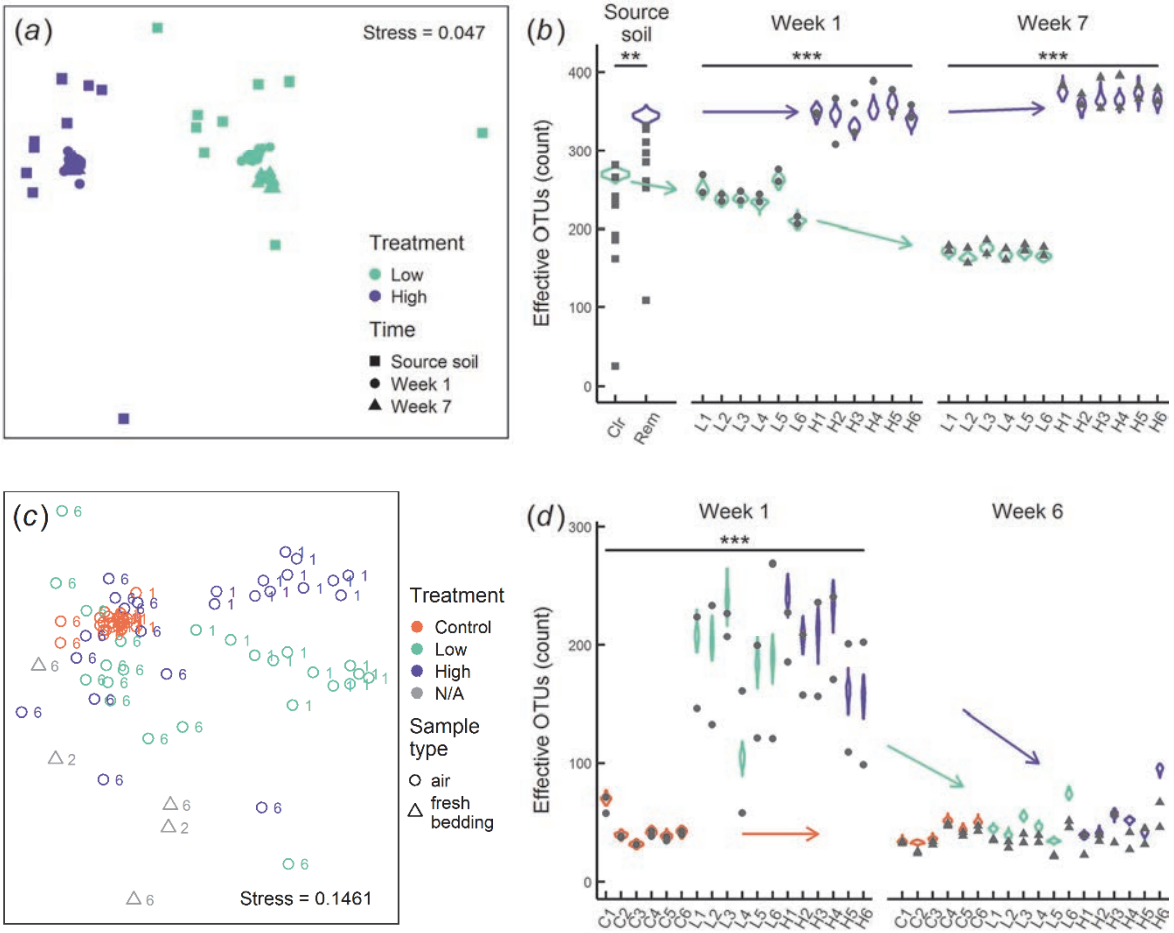


Figure 1. Characteristics of the soil and air bacteria exposures. (*a*) NMDS visualisation of microbiota in treatment soils from week 1, week 7, and source soils (cleared vs. remnant). Rarefied OTU abundance data (sequence depth 12524) show different compositional centroids by treatment (PERMANOVA df = 1, F = 75.4, R^2 = 0.545, *P* = 0.001, n = 32 and 33 per group). Beta dispersions of treatment groups are comparable (df = 1, F = 0.072, *P* = 0.78). (*b*) Alpha diversity of soil bacteria, showing individual subsamples (points) and overall diversity density distributions. *High* soils have higher alpha diversity than *low* soils at source (Wilcoxon W = 6, *P* = 0.0047, n = 8 per group), week 1 (Wilcoxon W = 120, *P* < 0.001, n = 10 and 12 per group) and week 7 (Wilcoxon W = 144, *P* < 0.001, n = 12 per group). (*c*) NMDS visualisation of low biomass air (dust) and fresh bedding microbiota, labelled by week

of collection. Rarefied OTU abundance data (sequence depth 839) showed different compositional centroids by treatment (PERMANOVA df = 2, F = 7.21, R^2 = 0.162, *P* = 0.001, n = 24 per group) and time (df = 1, F = 6.84, R^2 = 0.077, *P* = 0.001, n = 36 per group), although beta dispersions of treatment groups were different (df = 2, F = 47.5, *P* = 0.001). Final air samples came from week 6. Fresh bedding was excluded from testing of compositional differences (N/A = not applicable). (*d*) Alpha diversity of air microbiota samples, showing subsamples (points) and overall diversity density distributions. Week 1 *low* and *high* air samples are more diverse than *controls* (Kruskal-Wallis χ^2 = 20.6, df = 2, *P* < 0.001, n = 10 to 12 per group), however diversity is comparable by week 6 (Kruskal-Wallis χ^2 = 0.5, df = 2, *P* = 0.78, n = 11 to 12 per group). Alpha diversity density distributions are estimated from merged-sample bootstrap resampling (B = 100; electronic supplementary material, Supplementary Methods).

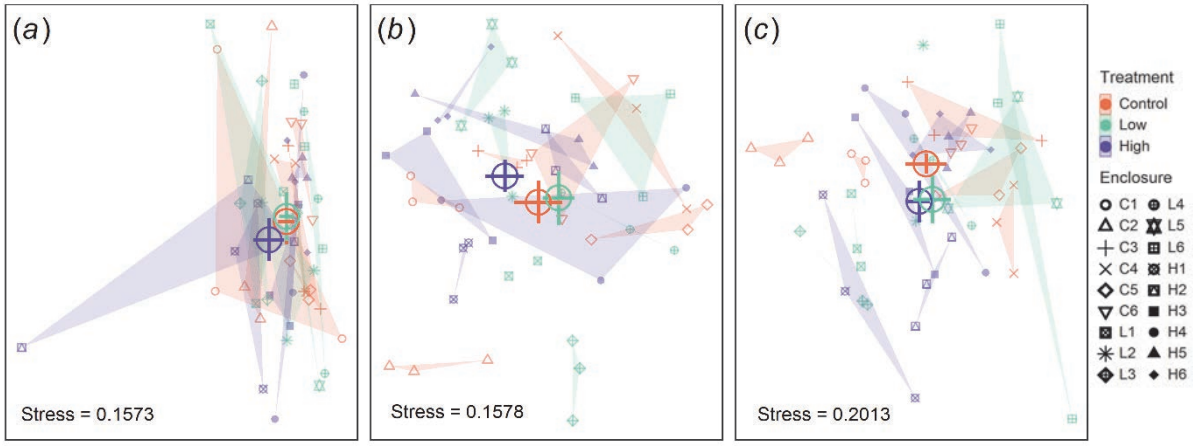


Figure 2. Changes in mouse gut microbiota. Plots show individual mouse gut microbiota and linked cagemates, superimposed with treatment-based centroids (large circles) and two-dimensional NMDS-dissimilarity standard errors. (*a*) The treatment centroids were not significantly different in week 0 faecal samples (PERMANOVA df = 2, F = 1.50, R^2 = 0.034, *P* = 0.072, n = 18 per group), but showed significant separation in (*b*) week 7 faecal and (*c*) caecal samples. Mice faecal bacteria in week 0 associated with litters (electronic supplementary material, figure S5). Week 7 faecal bacteria associated with enclosure/cage (PERMANOVA df = 15, F = 3.28, R^2 = 0.552, *P* = 0.001, n = 2 to 3 per group) and treatment (df =2, F = 2.43, R^2 = 0.055, *P* = 0.018, n = 17 to 18 per group), with similar beta distributions across treatments (df = 2, F = 0.42, *P* = 0.66). Similarly, caecal bacteria associated with enclosure/cage (PERMANOVA df = 15, F = 3.63, R^2 = 0.583, *P* = 0.001, n = 2 to 3 per group) and treatment (df = 2, F = 1.96, R^2 = 0.042, *P* = 0.027, n = 17 to 18 per group), with similar beta distributions across treatments (df = 2, F = 1.97, *P* = 0.14). NMDS visualisations are based on rarefied OTU abundances with respective sequence depths: faecal week 0 = 11450, faecal week 7 = 14195, caecal = 14632.

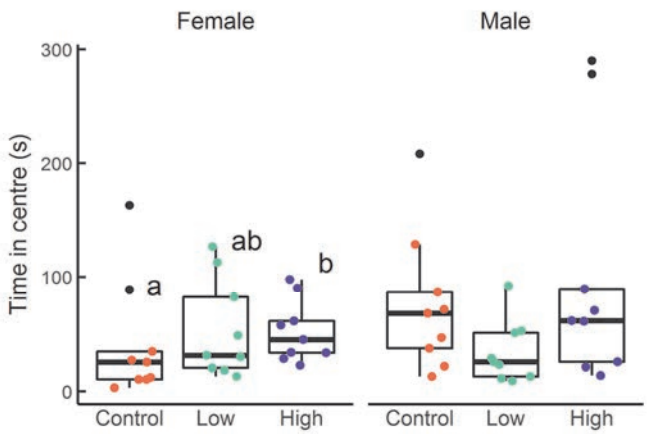


Figure 3. Post-treatment Open Field behaviour test results. Boxplots (with outliers in black) indicate that treatments involving higher soil biodiversity exposure associated with reduced anxiety-like behaviour in females only, in the form of increasing time spent in the centre of the Open Field apparatus. Groups not sharing a letter were significantly different. No trend was observed in males.

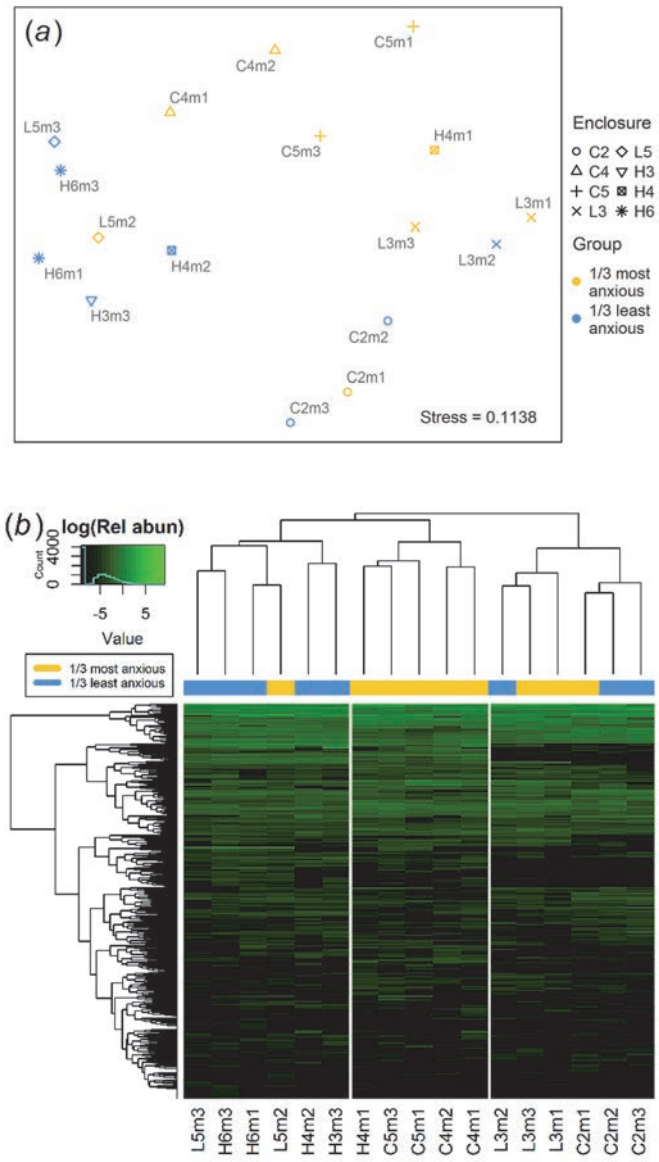


Figure 4. Faecal microbiota and anxiety-like behaviour. (*a*) NMDS visualisation of week 7 female mice faecal microbiota showing individuals in the top third (*1/3 least anxious*) and lower third (*1/3 most anxious*) of results for time spent in the centre of the Open Field. Rarefied OTU abundance data (sequence depth 17606) showed different compositional centroids accounting for anxiety group (PERMANOVA df = 1, F = 3.38, R^2 = 0.137, *P* = 0.017, n = 8 and 9 per group) and enclosure/cage (df = 7, F = 1.89, R^2 = 0.538, *P* = 0.018, n = 1 to 3 per group). Beta dispersions of the anxiety groups were comparable (df = 1, F = 0.516, *P* = 0.49). (*b*) Clustering of OTUs in the same microbiota samples (in *a*) by their log relative abundance yielded three microbiota-behaviour groupings, as described in the text.

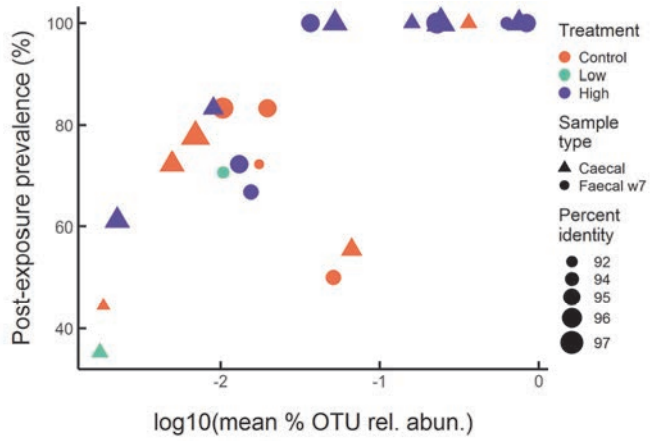


Figure 5. High biodiversity treatment animals showed the greatest post-exposure expression of putative butyrate-producing *K. alysoides*-like OTUs. The plot considers all *K. alysoides*-like OTUs (in week 7 faecal and caecal samples) that significantly increased during the experiment (i.e. between week 0 and week 7 faecal samples). *High* biodiversity samples dominate the top-right portion of the plot, indicating high mean % OTU relative abundance and high percent prevalence in post-exposure animals.

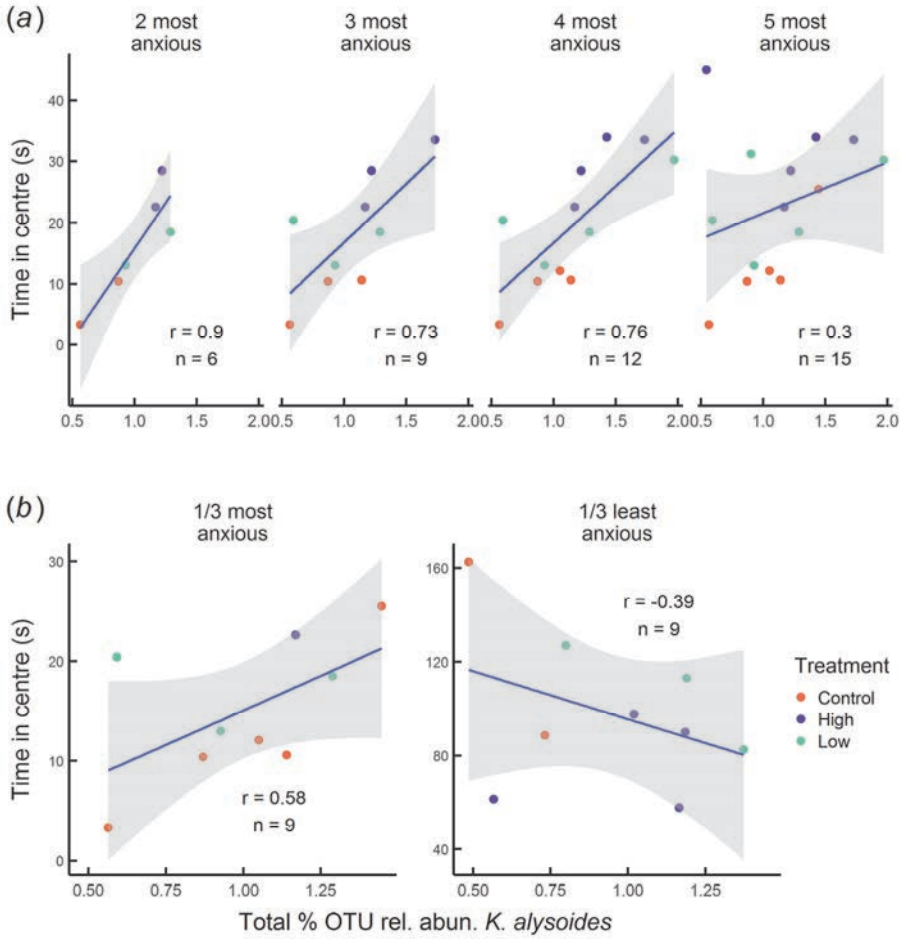


Figure 6. Putative butyrate-producing *K. alysoides* appears to moderate the most anxious behaviour in female mice. Plots show total % OTU relative abundance within caecal samples of all *K. alysoides*-like OTUs that increase during the experiment (x-axes) versus Open Field centre times (y-axes), considering: (*a*) the 'x' most anxious female mice from each treatment, and (*b*) the third most anxious and third least anxious groups based on Open Field centre times (as per figure 4). Pearson correlation coefficients (r) are noted. Shading indicates the 95% confidence intervals for each linear regression.

References

1. von Hertzen, L., Hanski, I. & Haahtela, T. 2011 Natural immunity: Biodiversity loss and inflammatory diseases are two global megatrends that might be related. *EMBO reports* **12**, 1089-1093. (doi:10.1038/embor.2011.195).

2. Rook, G. 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 of the United States of America* **110**, 18360-18367. (doi:10.1073/pnas.1313731110).

3. 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**, 392. (doi:10.1038/nm.4517).

4. Carr, E.J., Dooley, J., Garcia-Perez, J.E., Lagou, V., Lee, J.C., Wouters, C., Meyts, I., Goris, A., Boeckxstaens, G., Linterman, M.A., *et al.* 2016 The cellular composition of the human immune system is shaped by age and cohabitation. *Nature Immunology* **17**, 461-468. (doi:10.1038/ni.3371).

5. Wopereis, H., Oozeer, R., Knipping, K., Belzer, C. & Knol, J. 2014 The first thousand days – intestinal microbiology of early life: establishing a symbiosis. *Pediatric Allergy and Immunology* **25**, 428-438. (doi:10.1111/pai.12232).

6. Molloy, M.J., Bouladoux, N. & Belkaid, Y. 2012 Intestinal microbiota: Shaping local and systemic immune responses. *Seminars in Immunology* **24**, 58-66. (doi:10.1016/j.smim.2011.11.008).

7. Foster, J.A. & McVey Neufeld, K.-A. 2013 Gut–brain axis: how the microbiome influences anxiety and depression. *Trends in Neurosciences* **36**, 305-312. (doi:10.1016/j.tins.2013.01.005).

8. Flandroy, L., Poutahidis, T., Berg, G., Clarke, G., Dao, M.-C., Decaestecker, E., Furman, E., Haahtela, T., Massart, S., Plovier, H., *et al.* 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. (doi:10.1016/j.scitotenv.2018.01.288).

9. 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. *Science of The Total Environment* **571**, 680-687. (doi:10.1016/j.scitotenv.2016.07.037).

10. Martin, C.R., Osadchiy, V., Kalani, A. & Mayer, E.A. 2018 The Brain-Gut-Microbiome Axis. *Cellular and molecular gastroenterology and hepatology* **6**, 133-148. (doi:10.1016/j.jcmgh.2018.04.003).

11. Sarkar, C., Webster, C. & Gallacher, J. 2018 Residential greenness and prevalence of major depressive disorders: a cross-sectional, observational, associational study of 94879 adult UK Biobank participants. *The Lancet Planetary Health* **2**, e162-e173. (doi:10.1016/S2542-5196(18)30051-2).

12. Beyer, K.M., Kaltenbach, A., Szabo, A., Bogar, S., Nieto, F.J. & Malecki, K.M. 2014 Exposure to neighborhood green space and mental health: evidence from the survey of the health of Wisconsin. *Int J Environ Res Public Health* **11**, 3453-3472. (doi:10.3390/ijerph110303453).

13. Tillmann, S., Tobin, D., Avison, W. & Gilliland, J. 2018 Mental health benefits of interactions with nature in children and teenagers: a systematic review. *Journal of Epidemiology and Community Health* **72**, 958-966. (doi:10.1136/jech-2018-210436).

14. Newbury, J., Arseneault, L., Caspi, A., Moffitt, T.E., Odgers, C.L. & Fisher, H.L. 2016 Why Are Children in Urban Neighborhoods at Increased Risk for Psychotic Symptoms? Findings From a UK Longitudinal Cohort Study. *Schizophrenia Bulletin* **42**, 1372-1383. (doi:10.1093/schbul/sbw052).

15. Sundquist, K., Frank, G. & Sundquist, J. 2018 Urbanisation and incidence of psychosis and depression: Follow-up study of 4.4 million women and men in Sweden. *British Journal of Psychiatry* **184**, 293-298. (doi:10.1192/bjp.184.4.293).

16. Engemann, K., Pedersen, C.B., Arge, L., Tsirogiannis, C., Mortensen, P.B. & Svenning, J.-C. 2019 Residential green space in childhood is associated with lower risk of psychiatric disorders from adolescence into adulthood. *Proceedings of the National Academy of Sciences of the United States of America* **116**, 5188-5193. (doi:10.1073/pnas.1807504116).

17. Rydin, Y., Bleahu, A., Davies, M., Dávila, J.D., Friel, S., De Grandis, G., Groce, N., Hallal, P.C., Hamilton, I., Howden-Chapman, P., *et al.* 2012 Shaping cities for health: complexity and the planning of urban environments in the 21st century. *The Lancet* **379**, 2079-2108. (doi:10.1016/S0140-6736(12)60435-8).

18. Frumkin, H., Bratman, G.N., Breslow, S.J., Cochran, B., Kahn, P.H., Jr., Lawler, J.J., Levin, P.S., Tandon, P.S., Varanasi, U., Wolf, K.L., *et al.* 2017 Nature Contact and Human Health: A Research Agenda. *Environmental health perspectives* **125**, 075001-075001. (doi:10.1289/EHP1663).

19. Flies, E.J., Skelly, C., Negi, S.S., Prabhakaran, P., Liu, Q., Liu, K., Goldizen, F.C., Lease, C. & Weinstein, P. 2017 Biodiverse green spaces: a prescription for global urban health. *Frontiers in Ecology and the Environment* **15**, 510-516. (doi:10.1002/fee.1630).

20. Mills, J.G., Weinstein, P., Gellie, N.J.C., Weyrich, L.S., Lowe, A.J. & Breed, M.F. 2017 Urban habitat restoration provides a human health benefit through microbiome rewilding: the Microbiome Rewilding Hypothesis. *Restoration Ecology* **25**, 866-872. (doi:10.1111/rec.12610).

21. Hanski, I., von Hertzen, L., Fyhrquist, N., Koskinen, K., Torppa, K., Laatikainen, T., Karisola, P., Auvinen, P., Paulin, L., Mäkelä, M.J., *et al.* 2012 Environmental biodiversity, human microbiota, and allergy are interrelated *Proceedings of the National Academy of Sciences of the United States of America* **109**, 8334-8339. (doi:10.1073/pnas.1205624109).

22. 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., *et al.* 2016 Innate Immunity and Asthma Risk in Amish and Hutterite Farm Children. *New England Journal of Medicine* **375**, 411-421. (doi:10.1056/NEJMoa1508749).

23. Ege , M.J., Mayer , M., Normand , A.-C., Genuneit , J., Cookson , W.O.C.M., Braun-Fahrländer , C., Heederik , D., Piarroux , R. & von Mutius , E. 2011 Exposure to Environmental Microorganisms and Childhood Asthma. *New England Journal of Medicine* **364**, 701-709. (doi:10.1056/NEJMoa1007302).

24. Ottman, N., Ruokolainen, L., Suomalainen, A., Sinkko, H., Karisola, P., Lehtimäki, J., Lehto, M., Hanski, I., Alenius, H. & Fyhrquist, N. 2018 Soil exposure modifies the gut microbiota and supports immune tolerance in a mouse model. *Journal of Allergy and Clinical Immunology* **143**, 1198-1206.e1112. (doi:10.1016/j.jaci.2018.06.024).

25. Matthews, D.M. & Jenks, S.M. 2013 Ingestion of Mycobacterium vaccae decreases anxiety-related behavior and improves learning in mice. *Behavioural Processes* **96**, 27-35.

26. Schuijs, M.J., Willart, M.A., Vergote, K., Gras, D., Deswarte, K., Ege, M.J., Madeira, F.B., Beyaert, R., van Loo, G., Bracher, F., *et al.* 2015 Farm dust and endotoxin protect against allergy through A20 induction in lung epithelial cells. *Science* **349**, 1106-1110. (doi:10.1126/science.aac6623).

27. Reber, S.O., Siebler, P.H., Donner, N.C., Morton, J.T., Smith, D.G., Kopelman, J.M., Lowe, K.R., Wheeler, K.J., Fox, J.H., Hassell, J.E., *et al.* 2016 Immunization with a heat-killed preparation of the environmental bacterium *Mycobacterium vaccae* promotes stress resilience in mice. *Proceedings of the National Academy of Sciences of the United States of America* **113**, E3130–E3139. (doi:10.1073/pnas.1600324113).

28. Rook, G., Lowry, C. & Raison, C. 2013 Microbial 'Old Friends', immunoregulation and stress resilience. *Evolution, medicine, and public health* **2013**, 46-64. (doi:10.1093/emph/eot004).

29. Gascon, M., Triguero-Mas, M., Martínez, D., Dadvand, P., Forns, J., Plasència, A. & Nieuwenhuijsen, M.J. 2015 Mental Health Benefits of Long-Term Exposure to Residential Green and Blue Spaces: A Systematic Review. *International Journal of Environmental Research and Public Health* **12**, 4354-4379.

30. Zhou, D., Zhang, H., Bai, Z., Zhang, A., Bai, F., Luo, X., Hou, Y., Ding, X., Sun, B., Sun, X., *et al.* 2016 Exposure to soil, house dust and decaying plants increases gut microbial diversity and decreases serum immunoglobulin E levels in BALB/c mice. *Environmental Microbiology* **18**, 1326-1337. (doi:10.1111/1462-2920.12895).

31. Carola, V., D'Olimpio, F., Brunamonti, E., Mangia, F. & Renzi, P. 2002 Evaluation of the elevated plus-maze and open-field tests for the assessment of anxiety-related behaviour in inbred mice. *Behavioural Brain Research* **134**, 49-57. (doi:10.1016/S0166-4328(01)00452-1).

32. Jost, L. 2006 Entropy and diversity. *Oikos* **113**, 363-375. (doi:10.1111/j.2006.0030-1299.14714.x).

33. Liddicoat, C., Weinstein, P., Bissett, A., Gellie, N., Mills, J., Waycott, M. & Breed, M. 2019 Can bacterial indicators of a grassy woodland restoration inform ecosystem assessment and microbiota-mediated human health? *Environment International* **129**, 105-117. (doi:10.1016/j.envint.2019.05.011).

34. R-Core-Team. 2018 R: A language and environment for statistical computing. (Vienna, Austria, R Foundation for Statistical Computing.

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

36. 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**, 402-410. (doi:10.1099/ijsem.0.001643).

37. Nicholson, J.K., Holmes, E., Kinross, J., Burcelin, R., Gibson, G., Jia, W. & Pettersson, S. 2012 Host-Gut Microbiota Metabolic Interactions. *Science* **336**, 1262-1267. (doi:10.1126/science.1223813).

38. Furusawa, Y., Obata, Y., Fukuda, S., Endo, T.A., Nakato, G., Takahashi, D., Nakanishi, Y., Uetake, C., Kato, K., Kato, T., *et al.* 2013 Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. *Nature* **504**, 446-450. (doi:10.1038/nature12721).

39. Sanna, S., van Zuydam, N.R., Mahajan, A., Kurilshikov, A., Vich Vila, A., Võsa, U., Mujagic, Z., Masclee, A.A.M., Jonkers, D.M.A.E., Oosting, M., *et al.* 2019 Causal

relationships among the gut microbiome, short-chain fatty acids and metabolic diseases. *Nature Genetics* **51**, 600-605. (doi:10.1038/s41588-019-0350-x).

40. Valles-Colomer, M., Falony, G., Darzi, Y., Tigchelaar, E.F., Wang, J., Tito, R.Y., Schiweck, C., Kurilshikov, A., Joossens, M., Wijmenga, C., *et al.* 2019 The neuroactive potential of the human gut microbiota in quality of life and depression. *Nature Microbiology* **4**, 623-632. (doi:10.1038/s41564-018-0337-x).

41. Vital, M., Karch, A. & Pieper, D.H. 2017 Colonic Butyrate-Producing Communities in Humans: an Overview Using Omics Data. *mSystems* **2**, e00130-00117. (doi:10.1128/mSystems.00130-17).

42. Liu, P., Qiu, Q. & Lu, Y. 2011 Syntrophomonadaceae-affiliated species as active butyrate-utilizing syntrophs in paddy field soil. *Applied and environmental microbiology* **77**, 3884-3887. (doi:10.1128/AEM.00190-11).

43. Tiedje, J.M., Sexstone, A.J., Parkin, T.B. & Revsbech, N.P. 1984 Anaerobic processes in soil. *Plant and Soil* **76**, 197-212. (doi:10.1007/BF02205580).

44. 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**, 583-596. (doi:doi:10.1002/ecy.2137).

45. Coleman, D., Crossley, D. & Hendrix, P. 2004 *Fundamentals of soil ecology (2nd ed.)*. Burlington USA, Elsevier Academic Press; 386 pp.

46. Chen, S., Wang, W., Xu, W., Wang, Y., Wan, H., Chen, D., Tang, Z., Tang, X., Zhou, G., Xie, Z., *et al.* 2018 Plant diversity enhances productivity and soil carbon storage. *Proceedings of the National Academy of Sciences of the United States of America* **115**, 4027. (doi:10.1073/pnas.1700298114).

47. Polymenakou, P.N. 2012 Atmosphere: A Source of Pathogenic or Beneficial Microbes? *Atmosphere* **3**, 87-102.

48. 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**, 543-546. (doi:10.1038/nature17645).

49. 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**, 2403-2416. (doi:10.1038/s41396-018-0192-z).

50. 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**, 212-215. (doi:10.1038/nature16504).

51. Kim, S., Lee, S., Ryu, S., Suk, J.-g. & Park, C. 2002 Comparative analysis of the anxiety-related behaviors in four inbred mice. *Behavioural Processes* **60**, 181-190. (doi:10.1016/S0376-6357(02)00085-2).

52. An, X.-L., Zou, J.-X., Wu, R.-Y., Yang, Y., Tai, F.-D., Zeng, S.-Y., Jia, R., Zhang, X., Liu, E.-Q. & Broders, H. 2011 Strain and Sex Differences in Anxiety-Like and Social Behaviors in C57BL/6J and BALB/cJ Mice. *Experimental Animals* **60**, 111-123. (doi:10.1538/expanim.60.111).

53. McLean, C.P., Asnaani, A., Litz, B.T. & Hofmann, S.G. 2011 Gender differences in anxiety disorders: Prevalence, course of illness, comorbidity and burden of illness. *Journal of Psychiatric Research* **45**, 1027-1035. (doi:10.1016/j.jpsychires.2011.03.006).

54. Heijtz, R.D., Wang, S., Anuar, F., Qian, Y., Björkholm, B., Samuelsson, A., Hibberd, M.L., Forssberg, H. & Pettersson, S. 2011 Normal gut microbiota modulates brain development and behavior. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 3047-3052. (doi:10.1073/pnas.1010529108).

55. Sharon, G., Cruz, N.J., Kang, D.-W., Gandal, M.J., Wang, B., Kim, Y.-M., Zink, E.M., Casey, C.P., Taylor, B.C., Lane, C.J., *et al.* 2019 Human Gut Microbiota from Autism Spectrum Disorder Promote Behavioral Symptoms in Mice. *Cell* **177**, 1600-1618.e1617. (doi:10.1016/j.cell.2019.05.004).

56. Bravo, J.A., Forsythe, P., Chew, M.V., Escaravage, E., Savignac, H.M., Dinan, T.G., Bienenstock, J. & Cryan, J.F. 2011 Ingestion of Lactobacillus strain regulates emotional behavior and central GABA receptor expression in a mouse via the vagus nerve. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 16050. (doi:10.1073/pnas.1102999108).

57. Hol, W.H.G., De Boer, W., Termorshuizen, A.J., Meyer, K.M., Schneider, J.H.M., Van Dam, N.M., Van Veen, J.A. & Van Der Putten, W.H. 2010 Reduction of rare soil microbes modifies plant–herbivore interactions. *Ecology Letters* **13**, 292-301. (doi:10.1111/j.1461-0248.2009.01424.x).

58. Laukens, D., Brinkman, B.M., Raes, J., De Vos, M. & Vandenabeele, P. 2016 Heterogeneity of the gut microbiome in mice: guidelines for optimizing experimental design. *FEMS Microbiology Reviews* **40**, 117-132. (doi:10.1093/femsre/fuv036).

59. Wunderlin, T., Junier, T., Roussel-Delif, L., Jeanneret, N. & Junier, P. 2014 Endospore-enriched sequencing approach reveals unprecedented diversity of Firmicutes in sediments. *Environmental Microbiology Reports* **6**, 631-639. (doi:doi:10.1111/1758-2229.12179).

60. Tanca, A., Manghina, V., Fraumene, C., Palomba, A., Abbondio, M., Deligios, M., Silverman, M. & Uzzau, S. 2017 Metaproteogenomics Reveals Taxonomic and Functional Changes between Cecal and Fecal Microbiota in Mouse. *Frontiers in Microbiology* **8**. (doi:10.3389/fmicb.2017.00391).

Supplementary Information for

Naturally-diverse airborne environmental microbial exposures modulate the gut microbiome and reduce anxiety-like behaviour in mice

Craig Liddicoat, Harrison Sydnor, Christian Cando-Dumancela, Romy Dresken, Jiajun Liu, Nicholas J.C. Gellie, Jacob G. Mills, Jennifer M. Young, Laura S. Weyrich, Mark R. Hutchinson, Philip Weinstein, Martin F. Breed

Correspondence to: craig.liddicoat@adelaide.edu.au or martin.breed@adelaide.edu.au

This PDF file includes:

Supplementary Methods
Figures S1–S16
Tables S1–S5
Supplementary References

Data availability:

Data and code that support this study are available on *figshare* at:
http://doi.org/10.25909/5caaf18c3450d
or https://adelaide.figshare.com/s/cef9e2e25c17aa7daf19
Sequence read data are deposited in the European Nucleotide Archive study accession no. PRJEB31983

Supplementary Methods

Overview of experimental apparatus and sampling. Experimental data characterising starting and finish conditions were collected at the enclosure level for soil and air microbiota, and at the individual level for mouse weights, faecal and caecal samples, and anxiety-like behaviour. Three enclosures were constructed per one conventional metal mouse rack, and all 6 racks required for the experiment were contained within a single conventional climate-controlled room of the University of Adelaide Helen Mayo Animal House research facility. Further details of the enclosure design and room layout are provided below. The arrangement of enclosures was designed to: provide an equal number of each treatment, and each gender, at each rack height; ensure all treatments had equivalent within-rack adjacency to other treatments; cover all combinations of within-rack (vertical) ordering of treatments; and cover all combinations of gender, rack height, and treatment at least once. The rack and enclosure order (figure S1*b*) were used to set the sampling order that was used routinely throughout the study, for the purpose of avoiding any treatment-related biases. All activities with the animals (including weighing, faecal sampling, behavioural tests, daily checks, and caecal sampling) were performed in this order from rack 1 to rack 6, and from top to bottom within each rack (i.e. C1, L1, H1, H2, C2, L2, L3, H3, C3, H4, L4, C4, C5, H5, L5, L6, C6, H6).

Mice. We used 54 (27 male, 27 female) inbred BALB/C strain A,b,c:H-2d mice, from 15 litters, sourced from the University of Adelaide Laboratory Animal Services Helen Mayo Animal House specific-pathogen-free barrier facility. This breeding colony was re-derived into the specific-pathogen-free barrier facility in December 2008 from generation F25 Laboratory Animal Services stock received from the Animal Resource Centre (http://www.arc.wa.gov.au/) in 1996. Mice were 3-5 weeks old on commencement of the soil exposure treatments and 10-12 weeks old at completion. All animal procedures performed in this study were approved by the University of Adelaide Animal Ethics Committee (Approval number S-2017-112).

Source soils. Two candidate soil-vegetation associations were chosen *a priori* to display contrasting levels of microbial diversity (i.e. *low* biodiversity and *high* biodiversity, based on the vegetation-bacterial associations found in ref. [1]. On 13/8/2018 (2 weeks prior to starting the soil exposures) we collected soils from 6 sites reflecting *high* and *low* biodiversity soils. The soils were sourced from Mount Bold Reservoir Reserve, Kangarilla, South Australia (see figures S3, S4 and Table S4). Sampling occurred in late winter and the condition of soils (e.g. ~20-30% moisture content; Table S5) reflected natural field conditions, including light rain on the day of sampling. Each site was represented by a 25 × 25 m plot. The *low* and *high* biodiversity sites were selected to be ecogeographically matched (e.g. less than 1 km apart and equivalent in aspect, soil texture, rainfall and altitude). *Low* soil samples (3 sites: Clr1, Clr2, Clr3) were sourced from grass-dominated land that had been historically cleared of trees and

grazed by livestock from the early 1900s. However since 2003, when South Australia's water utility (SA Water) took over management of the area as a surface water catchment, grazing ceased and the land has been subject to a regular mowing and opportunistic hay-cutting regime for fire hazard management purposes. *High* soil samples (3 sites: Rem1, Rem2, Rem3) were sourced nearby in a similar landscape position and soil texture, but under remnant native vegetation consisting of open eucalypt woodland with comparatively high plant macro-diversity. The *low* and *high* biodiversity soil was collected for experimental treatments (20 L bucket per site, n = 6 total), physical and chemical soil analysis (0.5 kg sealed bag per site, n = 6 total) and eDNA sequencing (9 ×2 mL Eppendorf tubes at each site, n = 54 total).

At each site (25 × 25 m plot) we undertook the following sampling protocol:

1. We chose a representative 25 × 25 m plot set out in N-S-E-W orientation.
2. We marked and georeferenced the SW corner, and recorded the GPS coordinates.
3. From the SW corner we took a 360^{o} panoramic photo.
4. We chose 9 areas for sub-sampling (i.e. to take a 2-3 kg sod from the 0-10 cm soil horizon cleared of litter) that represented the heterogeneity or strata across the plot.
5. An unused disposable plastic ground sheet was laid out in a central location of each 25 × 25 m plot to aggregate the soil and hygienically house the soil containers before use (e.g. an unused 20 L plastic bucket, sample bags containing 9 × 2 mL Eppendorf tubes and 1 × soil physical sample test bag per site.
6. To reduce cross-contamination between sites we cleaned all sampling utensils with 3% bleach (e.g. 4 mm aperture soil laboratory test sieve (Endecotts, London, UK), a stainless steel hand trowel, long handled spade, plastic sampling spoons).
7. At each of the 9 sub-sampling locations:
    a. We cleaned the long-handle shovel and an unused plastic spoon for eDNA sampling, both with 3% bleach.
    b. Surface leaf litter was cleared away over an approx. 40 × 40 cm area and the surface 0-10 cm soil was loosened using the long handle shovel.
    c. A representative sample of the loose 0-10 cm soil was taken with the plastic spoon and transferred into a 2 mL Eppendorf tube for eDNA sequencing. Each set of 9 eDNA sub-samples for each site (in 2 mL Eppendorf tubes) were placed in a bag then transferred in dry ice for transport back to the lab.
    d. Then we collected approx. 2-3 kg of the surface 0-10 cm soil with the long handle shovel, depositing it onto the central plastic sheet.
8. After completing the 9 sub-samples:
    a. We homogenised the 9 × 2-3 kg samples on the plastic sheet using the cleaned hand trowel.
    b. We sieved and collected approx. 15 L of soil into the 20 L plastic bucket.

c. From the homogenised soil we then isolated approx. 500 g of soil and transferred it into the soil physical sample test bag for soil chemistry, particle size analysis and soil moisture tests.

On returning to the lab, the 6 sets × 9 eDNA samples were subsequently stored at -80ºC. The 6 × 20 L plastic buckets (i.e. 3 × *low* soils, 3 × *high* soils) were stored in a walk-in refrigerator at 4 ºC for two weeks prior to starting the soil exposures. Three eDNA samples from each site were subsequently sequenced and analysed (as described below) to characterise the microbiota from source soils. Physico-chemical characterisation tests for the source soils were undertaken at CSBP Soil & Plant Analysis Laboratory (Perth, Western Australia), with results provided in Table S5. Details of CSBP lab methods (as at November 2018) used are available online at: https://www.csbp.com.au/docs/default-source/csbp-lab/csbp-lab-methods-1118.pdf.

Enclosure design. Enclosures consisted of a plastic-lined containment volume fitted with dust-proof filters and resealable access panels. Each enclosure was built using 200 µm thick clear polythene film wrapped around 2 shelves of a conventional 6-shelf metal mouse rack, i.e. three individually-isolated enclosures were fitted per rack. Clear adhesive polypropylene tape was used to secure the film-wrap in place and seal air gaps in the structures. Dust filter panels were installed at each end, using autoclaved commercial air conditioning filter media (Filtrair TP 45 EN 779:G4, Filtrair B.V., Netherlands) secured using polyvinyl chloride right-angle frames and zinc-plated fasteners. Two 45 cm × 45 cm access panels were cut adjacent to the soil tray and cage, and overlapping 50 cm × 50 cm clear polythene film panels were fixed along the top edge and then secured on the sides and lower edges using two-part magnetic strip seals (Parts 59057A and 59057B, AMF Magnetics, Sydney, Australia).

Each enclosure housed one large white opaque plastic open cage (41 cm long × 25 cm wide × 12.5 cm high), with a stainless steel wire-top, containing three mice. Flooring in the enclosure consisted of 0.8 mm thick polypropylene sheeting. Ten cm diameter 2.5 W USB-powered fans (Product Code: 749901, R. Weatherdon & Co. Pty. Ltd, Sydney, Australia) sat adjacent to shallow plastic soil trays (high-impact polystyrene, internal dimensions: 42 cm long × 30.5 cm wide × 2.5 cm high; Product AIH008BLK, AdMerch, Adelaide, Australia).

Petri dishes for weekly passive air / dust sampling for microbiota were placed either side of the cage. Petri dishes for weekly dust weight quantification were also placed either side of the cage but ultimately not used due to low quantities of dust settling on a weekly basis. Further design aspects and dimensions of the enclosures, and details of the arrangement of enclosures within racks, and layout of racks within the study room are provided in figure S1. Example images of the enclosures and the study room are shown in figure S2.

Animal husbandry and handling. Mice were ear-notched by Laboratory Animal Services staff prior to receipt by the study team, and we tail-marked them with permanent marker every 1-2 days for identification purposes. Food and acidified water (pH 2.5 – 3) were provided *ad libitum*. The study room ran an automatic 12-hr light (7 am – 7

4

pm)/12-hr dark (7 pm – 7 am) cycle. All food (Meat Free Rat and Mouse Diet, Specialty Feeds, Australia), water, and cages with bedding materials (comprising corn cob and rolled paper; Shepherd's Cob Plus, Shepherd Specialty Papers, US), were autoclaved prior to supply. Following autoclaving, water bottles were covered in sterilised containers and other supplies covered in sterilised cloth drapes for intra-facility transport from the autoclave to the study room. Cages, bedding, and water were changed weekly on Tuesdays. Individual animal health checks were performed daily and the study room floor was cleaned weekly on Thursdays. There were no noteworthy animal health issues, except that we added autoclaved marbles (× 5) to all cages in week 7 (10/10/2018) for additional enrichment after observing over-grooming and loss of whiskers in two *high* cages, affecting animals H5m1, H5m3, H3m1, and H3m2. All researcher interactions with the animals were performed on a cage-by-cage basis following the predefined sampling order, described above. To inspect and weigh the mice, and to collect faecal samples, we promptly transferred cages from their enclosures to the cleaned bench of a laminar flow changing station (Model CS5, Techniplast, Italy). We transferred the mice to individual white opaque plastic tubs for weighing each mouse and collecting fresh faecal samples. Nine tubs were used in total, labelled by treatment and mouse number 1-3 (i.e. 3 × *control* tubs, 3 × *low* tubs, 3 × *high* tubs). Between each use, these tubs were cleaned with 5% Decon 90 (Decon Laboratories Ltd) and stacked separately by treatment.

Experimental hygiene. We routinely used 5% Decon 90 for decontamination of all laboratory surfaces and experimental equipment. This included equipment used to hold or test the mice, where strong odours (e.g. from bleach or ethanol) might adversely impact on mice behaviour or health. Prior to setting up the mouse study room, the floor and walls were cleaned with 3% bleach-demineralised water solution. Enclosure filter panels were autoclaved prior to fitting. Following construction of the enclosures, all internal surfaces as well as the enclosure floor sheets and soil trays were thoroughly cleaned with 5% Decon 90. External surfaces of the fans were also, as far as practicable, cleaned with 5% Decon 90. All researchers and animal facility staff wore laboratory gowns, nitrile gloves and face masks at all times on entry to the mouse study room. Gowns used were either freshly sterilised cloth or new clean disposable items. During all experimental work, researchers maintained an inner pair of gloves, sealed with masking tape to the sleeves of the laboratory gown, and a second outer pair of gloves that were routinely changed after each contact with an animal, a cage, a soil tray, an enclosure, or after performing a distinct cleaning task. Similarly, gowns, face masks, and regularly changed double nitrile gloves were routinely used in all laboratory work associated with sample preparation and DNA extraction. All 1.5 mL and 2 mL Eppendorf Safe-Lock tubes used for sample collection were prepared from bulk supply bags by closing, labelling, and placing tubes into clean unused large clear plastic bags corresponding to weekly sampling batches, which were then laid out flat and UV-sterilised for 30 min. All DNA extractions were performed within a laminar flow hood (S@FEMATE Eco 1.8 ABC, Euroclone, Italy) after thorough cleaning with 5% Decon 90 and UV-sterilisation for 20 min before each use. Additional hygiene measures specific to each procedure are described elsewhere in these methods.

5

Soil treatments and sampling. On the morning of starting the soil exposures, we transferred and mixed soil in equal parts from the 3 × 20 L buckets corresponding to either the *low*, or *high,* soil treatment, to make 1.75 kg of homogenised soil per tray. Using a small stainless steel garden fork (cleaned between use with 5% Decon 90) we distributed the soil evenly and then raked the soil using 10 strokes across (5 up, 5 back) then 10 strokes lengthways (5 up, 5 back). This raking procedure was repeated each week after soil sampling to facilitate new exposure of fine soil particles. Soil sampling was undertaken on Wednesday morning each week, and involved ¾ filling 2 × 2 mL Eppendorf tubes with randomly located sub-samples from the soil tray. One new disposable plastic spoon, cleaned with 5% Decon 90, was used per enclosure soil tray sampling event. Gloves were changed between enclosures. Samples were promptly placed into -80 ºC storage. In this study we only report on soil microbiota analyses from the week 1 and week 7 sampling.

Enclosure air sampling. Weekly air samples were collected for each enclosure via passive deposition of dust onto 9 cm diameter sterile Petri dishes. Both the lid and base from two Petri dishes were opened out onto the enclosure floor either side of the cage (figure S1*a*). The lid and base for sample 'a1' were placed on the side closest to the soil tray, while the lid and base for sample 'a2' were placed furthest from the soil tray. Additional samples for weekly dust weight measurement were also collected either side of the cage (i.e. lid and base labelled 'aw'). However, the 'aw' samples were ultimately not used due to practical difficulties in measuring the negligible amount of dust that settled on a weekly basis. Instead, we quantified soil dust weights using an alternative approach described below. After one week of dust accumulation, eDNA samples were obtained by swabbing the Petri dishes with nylon tip FLOQSwabs (Cat. No. 501CS01, Copan Diagnostics Inc., CA, USA), treating 'a1' and 'a2' samples separately. Each half of the Petri dish was swabbed for 30 s using a consistent protocol. This involved rolling the swab while making parallel strokes back and forth, holding the swab at a slight angle and with moderate pressure, while also slowly rotating the Petri dish, to systematically cover its entire area. We performed this swabbing under the shelter of each enclosure, via the access panels, to minimise possible contamination from external dust fall in the study room. While still under cover from the enclosure access panel, swabs were cut with decontaminated scissors (cleaned with 5% Decon 90 between use) directly into labelled 2 mL Eppendorf tubes. Air sampling blanks, comprising unused swabs, were opened and cut into sample tubes out in the open air of the study room. Air swab samples were stored at -80 ºC. The air samples were routinely collected on Mondays to represent the preceding week. However, at the finish of the 7-week exposure period, this routine would have overlapped with two days of behaviour tests (figure S6). To avoid the possibility of any contaminating influence on the low biomass air samples due to animals participating in the behaviour tests, we used week 6 samples to represent the final air microbiotas.

Faecal sampling. Twice a week, predominantly on Tuesdays and Fridays, individual faecal samples were collected. Cages were brought to the changing station, and mice were placed in their respective individual plastic tub for weighing, after which we commenced faecal sample collection. Using one new disposable plastic spoon per mouse, cleaned with 5% Decon 90, several faecal boli were collected over 2-3 min. Occasionally, we waited up to 5-10 min to collect at least 2 faecal boli. After the required samples were secured from each mouse, they were tail marked with permanent marker if required, then returned to the cage. When all 3 mice were returned, the cage was promptly returned to its respective enclosure. Gloves were always changed between cages. On completion, samples were placed into -80 ºC storage. In this study we only report on faecal microbiota analyses from the initial pre-exposure (T01 or week 0) and final (T16 or week 7) faecal samples.

Euthanasia and whole cecum sampling. On completion of experimental work, mice were euthanised and a number of tissues were sampled including whole ceca. All 54 mice were euthanised on the same day (17/10/2018), on a cage-by-cage basis, using 2 stations of 3 researchers at each station to facilitate timely and sterile parallel processing of the 3 mice per cage. Animals were euthanised via overdose with a single intraperitoneal injection of sodium pentobarbital dosed at 60 mg/kg. Following terminal anaesthesia, each animal was presented on an individually labelled and cleaned (with 5% Decon 90) foam board that was overlaid with clean paper towel. At the first station, we sprayed the animal's fur with ethanol to minimise contamination of internal tissues. We collected blood via cardiac puncture, and administered a transcardial perfusion with sterile saline (0.9% Sodium Chloride Injection USP, Baxter Healthcare Corporation, USA) to facilitate later post-fixing of brain tissues. The animals were passed to the second station situated under a flow hood (animals #1 and #2) or a portable benchtop hood (animal #3) for removal of ceca. Immediately after opening the abdominal cavity, the spleen was first recovered, then the whole cecum was removed and transferred into a 1.5 mL Eppendorf tube and snap frozen in liquid nitrogen. From the liquid nitrogen, samples were later transferred to storage at -80 ºC. Between animals, surgical equipment was decontaminated by rinsing in a beaker of saline, then resting in a bleach bath for 3 min. Work areas were cleaned with 5% Decon 90 after each animal, a new surgical mat laid out, then surgical instruments were sprayed and wiped using 70% ethanol prior to setting them out ready for next use.

DNA extraction. We performed DNA extractions in the Advanced DNA Identification and Forensic Facility lab, at the University of Adelaide's North Terrace campus. DNA was extracted from soil, faecal, caecal and fresh bedding samples, together with extraction blank controls, using Qiagen's DNeasy PowerLyzer PowerSoil Kit (Cat. No. 12855-100) according to manufacturer's instructions, except for an additional homogenisation step for the whole cecum samples as described below. Per extraction, we used approx. 250 mg of soil, approx. 30 mg of faecal material for week 0, and approx. 50 mg faecal material for week 7, which typically corresponded to 2-4 faecal boli. Prior to DNA extraction from the caecal samples, we performed an additional homogenisation of the whole ceca,

with the objective that our analyses would better reflect both luminal and mucosa-associated microbes. We homogenised the caecal samples using custom prepared bead tubes, freezing in liquid nitrogen and powderisation using an Omni Bead Rupter 24 homogeniser. To prepare the custom bead tubes, we first UV-sterilised (for 30 min) 2 × 2 mL bead tubes per cecum, for 54 ceca. We oven-sterilised (6 hr at 125ºC) 20 × 1.4 mm diameter and 5 × 2.2 mm diameter beads per bead tube, then transferred these into the bead tubes in the laminar flow hood using sterilised tweezers. Whole ceca were taken from the -80 ºC freezer to partially defrost in the laminar flow hood. Each cecum was cut into 4 equal parts using a sterilised surgical blade (cleaned via a bleach bath then wiped with 70% ethanol) while working on a sterile Petri dish. The caecal material was divided equally into 2 × 2 mL bead tubes, marked as 'a' and 'b' sub-samples. The bead tubes with caecal sub-samples were submerged in a liquid nitrogen bath and the bead ruptor rotor was also rolled in liquid nitrogen. Then we loaded and ran the bead ruptor for 30 s at 5.5 m/s. We repeated this freezing and homogenisation process a second time, then returned the samples to the -80 ºC freezer to await DNA extraction. For DNA extraction, we transferred equal parts (including some of the inert beads) from the 'a' and 'b' cecum bead tubes, totalling 100-200 mg per extraction, into the standard extraction kit homogeniser tubes and continued with the standard DNeasy PowerLyzer PowerSoil extraction protocol.

We extracted DNA from air (i.e. dust swab) samples, together with extraction blank controls, using Qiagen's QIAamp DNA Mini Kit (Cat No. 51304). We followed the manufacturer's instruction for DNA purification from buccal swabs (spin protocol) (QIAamp DNA Mini and Blood Mini Handbook 05/2016) using volumes prescribed for the Omni Swab option, except we used a reduced final elution volume of 80 µL buffer AE to increase the final DNA concentration.

Extracted genomic DNA concentrations were quantified using a Quantus Fluorometer (Promega, USA). Genomic DNA samples and negative controls were transferred to 4 × Eppendorf LoBind 96 well plates using 20 µL volume per sample. Upon transfer, the higher biomass soil, faecal and caecal genomic DNA samples were normalised to approx. 10 ng/µL concentration, buffered with PCR grade water (Ultra Pure Water, Fisher Biotec, Australia). Lower biomass air, fresh bedding and negative controls were left undiluted to preserve DNA concentrations (typically 0.001-0.05 ng/µL). Plates were heat sealed (using Axygen Platemax Plate Sealer) prior to inter-lab transport. Storage was at -80 ºC, except for transport between facilities on dry ice.

PCR. Genomic DNA samples were PCR-amplified at the Australian Genome Research Facility (AGRF) located within the Victorian Comprehensive Cancer Centre in Melbourne, Victoria. We targeted the bacterial 16S rRNA V3-V4 gene region using forward primer 341F (CCTAYGGGRBGCASCAG) and reverse primer 806R (GGACTACNNGGGTATCTAAT). AGRF used a modified Illumina two-stage PCR library preparation technique adapted from ref. [2]. In the primary PCR stage, amplicons were generated using the region-specific primers coupled to forward primer overhang adaptors (5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3') and

reverse primer overhang adaptors (5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-3'), together with AmpliTaq Gold 360 mastermix (Life Technologies, Australia). Thermocycling was completed with an Applied Biosystems Veriti 384-well unit, consisting of an initial denaturation at 95 °C for 7 min, followed by 29 cycles of denaturation at 94 °C for 30 s, annealing at 50 °C for 60 s and extension at 72 °C for 60 s, and then a final extension of 72 °C for 7 min. The first stage PCR was cleaned using magnetic beads, and samples were visualised on 2% SYBR E-Gel (Thermo-Fisher). A secondary PCR was performed with TaKaRa PrimeSTAR Max DNA Polymerase (Clontech) to add sample indices (barcodes) and Illumina sequencing adaptors to the first stage amplicons using Illumina's Nextera XT Index kit. The second stage thermocycling comprised an initial denaturation at 98 °C for 10 s, followed by 8 cycles of denaturation at 98 °C for 10 s, annealing at 50 °C for 5 s and extension at 72 °C for 10 s, then holding at 10 °C. The resulting amplicons were cleaned again using magnetic beads, quantified by fluorometry (Promega QuantiFluor) and then normalised. The eqimolar pool was cleaned a final time using magnetic beads to concentrate the pool and then measured using a High-Sensitivity D1000 Tape on an Agilent 2200 TapeStation. The pool was diluted to 5nM and molarity was confirmed again using a High-Sensitivity D1000 Tape.

Sequence processing pipeline and OTU assignment. Sequencing was also performed at AGRF using Illumina's paired end chemistry on the MiSeq platform (San Diego, CA, USA) with a V3, 600 cycle kit (2 × 300 base pairs paired-end). All clonal libraries were processed through two flow cells in November 2018. Image analysis was performed in real time by the MiSeq Control Software (MCS) v2.6.2.1 and Real Time Analysis (RTA) v1.18.54, running on the instrument computer. RTA performs real-time base calling on the MiSeq instrument computer. The Illumina bcl2fastq 2.20.0.422 pipeline was used to generate the sequence data. Paired-end reads were assembled by aligning the forward and reverse reads using PEAR [3] (version 0.9.5). Primers were identified and trimmed. Trimmed sequences were processed using Quantitative Insights into Microbial Ecology (QIIME 1.8) [4], USEARCH [5] (version 8.0.1623, including UCLUST, UCHIME [6] and UPARSE [7]) software. Using USEARCH tools, sequences were quality filtered using a maximum expected errors threshold of 0.5, full length duplicate sequences were removed, and remaining sequences sorted by abundance. Singletons or unique reads in the data set were discarded. Sequences were clustered, followed by chimera filtering using the "rdp_gold" database as a reference. OTUs were formed from clustering sequences with ≥ 97% similarity. Using QIIME and USEARCH/UCLUST, taxonomy was assigned using the Greengenes database [8] (version 13_8, Aug 2013).

Microbiome data exclusions and decontamination. We discarded OTUs that were: not identified as belonging to bacteria, unidentified at the phylum level, assigned as chloroplast or mitochondria, did not occur in at least two samples, or had < 50 reads across all study samples. Three samples were excluded due to anomalously low sequence read counts (i.e. samples: 'L1m2T16' = week 7 faecal sample from mouse #2 in enclosure L1; 'H4s2w1' = week 1 soil sub-sample #2 from enclosure H4; 'L2m3Ce' = caecal sample from mouse #3 in enclosure L2). We

identified and removed contaminating sequences using the R *decontam* package [9]. Lower biomass samples (i.e. air, air sampling blanks, fresh bedding, and corresponding extraction blanks) were analysed using the *isNotContaminant()* function, as recommended in the *decontam* user guide, where contaminants are identified by increased prevalence in negative controls. Meanwhile, higher biomass samples (i.e. soil, faecal, caecal samples, and corresponding extraction blanks) were analysed using the *isContaminant()* function. For higher biomass samples using *isContaminant()*, contaminants were called using the 'either' setting, where contaminants are identified by frequency that varies inversely with sample DNA concentration, or by increased prevalence in negative controls. Otherwise, default settings were used throughout. Subsequently, all taxa identified as contaminants in either the lower or higher biomass samples were pooled and removed from further microbiome data analysis.

Microbiome visualisation and statistical analysis. We used R software [10] for the majority of analyses with purpose-built scripts (see data availability statement) employing the microbiome data analysis framework of the R *phyloseq* package [11]. We variously used rarefied OTU abundance data (at the minimum sequence read depth of the samples concerned), non-rarefied OTU abundance data, and non-rarefied OTU relative abundance (%) data, reflecting different input and sample normalisation requirements for particular analyses. We visualised differences between sample microbiota (beta diversity) using non-metric multidimensional scaling (NMDS) ordination of Bray-Curtis distances based on rarefied OTU abundances. For low biomass air and fresh bedding samples we applied a minimum threshold of 800 sequence reads to provide a balance between preserving samples and sufficiently representing microbiota composition in the rarefied OTU data. This step caused the elimination of a single sample 'bfre1T5', i.e. the fresh bedding sample from week 2 (time #5).

Permutational multivariate analysis of variance (PERMANOVA) was used to examine the statistical significance of compositional differences between rarefied OTU abundances of sample groups, as implemented in the *adonis()* function from the R *vegan* package [12], and followed by the *betadisper()* function to test for homogeneity of group dispersions (also from *vegan*).

To visualise the high level microbial community composition of samples we prepared OTU relative abundance bar plots using the *plot_bar()* function of the R *phyloseq* package, after grouping taxa at the phylum level. Phyla with < 0.5% OTU relative abundance were further grouped and categorised as 'other minor phyla'.

We estimated OTU alpha diversity in each sample based on rarefied abundances (to normalise sampling effort) and used the exponential transform of Shannon Index values to derive the effective number of OTUs [13]. We tested for differences in alpha diversity between groups of three or more using the Kruskal-Wallis rank sum test in base R (10). Where significant, we used the Dunn test for multiple comparisons to determine which groups differed from other groups, with Bonferroni-adjusted threshold *P* values [14]. The multiple comparison testing used default two-sided *P* values and alpha = 0.05 nominal level of significance. We tested for differences between groups of two using the Wilcoxon rank sum test (or *wilcox.test()* in the *stats* package of base R), also using the

default two-sided *P* values and alpha = 0.05. Where applicable, the following significance annotations were used: $P \leq 0.001$ = '***'; $0.001 < P \leq 0.01$ = '**'; $0.01 < P \leq 0.05$ = '*'.

All such statistical testing for differences between groups, for microbiome data and behaviour test data below, excluded outliers as identified by the default R *boxplot()* function in the *graphics* package of base R. Outliers were identified as any data points located outside 1.5 times the interquartile range above the upper quartile or below the lower quartile. All boxplots displayed here present the median and interquartile range. Boxplot whiskers extend to the most extreme data points that are no more than 1.5 times the interquartile range from the box (i.e. from the upper or lower quartile).

To help visualise the overall alpha diversity and associated uncertainty for soil and air microbiota where replicate sub-samples were available, we generated density distributions using a new *merged-sample bootstrap resampling* technique (B=100) [15]. Our *merged-sample bootstrap resampling* framework involved the following steps (adapted from ref. [15]):

1. OTU abundance data were prepared in non-rarefied form, i.e. with varying sequence depths across multiple sub-samples, spanning different sample origins. Here the full OTU table was expressed in a single *phyloseq* microbiome data object with sub-samples corresponding to a particular sample origin given a respective label in the *phyloseq* sample contextual data table. For example, there were duplicate sub-samples that characterised the soil and air from each enclosure at a given point in time.
2. Set seed for pseudo-random number generator.
3. All sub-samples were then rarefied to an even sampling depth, equal to the minimum sequence depth across all sub-samples, using sampling with replacement.
4. With the rarefied data from step 3, we merged sub-samples (i.e. grouped data by the sample origin label, e.g. all soil sub-samples from enclosure L1 in week 1) and summed taxa counts by the sample origin. We transferred the results to a new merged-sample *phyloseq* data object.
5. Set seed for the pseudo-random number generator.
6. Then we rarefied the merged data from step 4 to an even sampling depth, using the same minimum sequence depth value from step 3, again using sampling with replacement.
7. The measure of interest (e.g. for alpha diversity we used effective number of OTUs, as above) was then calculated based on the merged and rarefied data from step 6 and results were stored as an element of an output list.
8. Steps 2 to 7 were repeated, up to the number of predefined iterations (e.g. B = 100).

When testing for statistical differences between the alpha diversity of sample groups we considered only the original sub-sample data, i.e. the *merged-sample bootstrap* data were used only for visualisation purposes in this study.

Differential abundance testing using the R *DESeq2* package [16], which requires non-rarefied OTU abundance data, was used to assess the significantly increasing and significantly decreasing taxa between week 0 and week 7 faecal samples, within each treatment. We only considered mice with data available for both week 0 and week 7, i.e. mouse L1m2 was not included, due to excluded anomalous week 7 faecal data described earlier. The *DESeq2* algorithm internally adjusts for testing multiple hypotheses by applying the Benjamini-Hochberg method to control for false discoveries. We applied the default 'Wald' significance test, 'parametric' fitting, and considered significant differentially abundant results with adjusted P values < 0.05. We then quantified the number and overlap of significant increasing and decreasing taxa within treatments and visualised these using the R *eulerr* package [17]. Further, we assessed the relative abundance and prevalence of the significant differentially abundant increasing taxa across all sample types, i.e. in soil, air, week 0 and week 7 faecal, and caecal samples. These taxa were also assigned to a putative species based on the closest match in the National Centre for Biotechnology Information (NCBI) 16S database, as described below.

To illustrate at an OTU level the relative enrichment and overlap of significantly increasing taxa we isolated these taxa and their abundance data from treatment-specific week 7 faecal samples. With these data we formed a combined subset of week 7 faecal OTU data, which were then normalised by converting to relative abundance, and scaled (i.e. mean-centred and divided by standard deviation) firstly within taxa/OTUs and then scaled within samples to highlight the relative enrichment of taxa across the treatments. These data were plotted as a heatmap using the *heatmap.2()* function from the R *gplots* package [18].

We also used the *heatmap.2()* function to visualise microbiota similarities and differences across the *less anxious* and *anxious* groupings of female mice based on the top third and bottom third of results for time spent in the centre of the Open Field (see below). In this case, we displayed the log relative abundance (†see note below) of week 7 faecal OTUs, and used this OTU abundance data to drive the clustering of samples/individuals, from which dominant behaviour-treatment groupings emerged. †Note: we used a log relative abundance transformation function of the form: log((100*x/sum(x)) + 0.0001), where the nominal value of 0.0001 was used to avoid negative infinite results from a log(0) calculation. Prior to adding the nominal 0.0001, the smallest non-zero value for % OTU relative abundance was approx. 0.0013. In all heatmaps, dendrogram clustering of taxa and samples was based on the default Euclidean distances.

Assignment of putative species (closest NCBI match). We performed Basic Local Alignment Search Tool [19] – Nucleotide (BLASTN) searches of the latest available NCBI 16S Microbial reference database to assign putative species to OTUs based on their representative FASTA sequences. We downloaded the NCBI 16S database (ftp://ftp.ncbi.nlm.nih.gov/blast/db/16SMicrobial.tar.gz; dated 20/01/2019) and installed a local version of BLAST software (ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/ncbi-blast-2.8.1+-win64.exe; dated 27/11/2018). We used the closest match returned from the BLAST search as the putative species assigned to each

OTU, and recorded corresponding percent identity and expectation values. Expectation values were all well below 10^{-10} denoting significant matches [20] from the BLAST taxonomic assignments. R scripts used to manage the BLAST searches are provided in other Supplementary Materials (see data availability statement).

Behaviour testing and analysis. We used Open Field (OF) and Elevated Plus Maze (EPM) apparatus to test anxiety-like behaviour [21]. OF and EPM tests were undertaken on consecutive days, and for each test we used 3 apparatus to allow the 3 mice in each cage to be run in parallel. The OF consisted of a square arena (internal dimensions 35 cm × 35 cm floor, with 15 cm walls) constructed from black polypropylene board. Internal edges were sealed with black silicon sealant. The floor was divided into (4 × 4) 16 equal zones using digital test apparatus zone templates created in ANY-maze video tracking software (Version 4.63, Stoelting Co., USA). We considered the centre zone, comprising the central 4 squares, as the anxiety-provoking zone for analysis of the OF tests. Mice were individually placed in the centre of the OF at the commencement of the test and allowed to explore for 10 min.

The EPM consisted of four arms (with each arm 24 cm long × 5 cm wide) radiating from a centre zone (5 cm × 5 cm), elevated 0.5 m above the floor. Two arms were open and two were closed with internal walls 15 cm high. EPM were made from black acrylic plastic. The open arms, closed arms and centre zone were mapped digitally using the ANY-maze software. We considered the open arms and centre zone together as the anxiety-provoking zones for analysis in this study. Mice were individually placed in the centre of the maze at the commencement of the test and allowed to explore for 10 min.

Both OF and EPM tests were recorded and scored (based on the mouse body centre position) using the ANY-maze video tracking software, which also provided data files for statistical analysis. Behaviour tests were performed within a separate zone of the same study room used to house the animals. We used a black plastic drape across half the room width and over a suspended ceiling to reduce light levels in the behaviour test area, in order to achieve sufficient contrast between the white mice and dark test equipment to facilitate video tracking. After each subject completed its test session, faecal boli and urine were removed with clean paper towel, then surfaces were sprayed with 5% Decon 90 and wiped with clean paper towel. All test apparatus were visually inspected and confirmed to be completely dry before starting the next group of subjects. We assessed the times spent in the anxiety-provoking zones as a measure of anxiety-like behaviour; and total distance travelled and number of entries into the anxiety-provoking zones as measures of overall activity levels. To test for significant differences in these variables between groups of three or more (i.e. between treatments, sexes, or both) we used non-parametric Kruskal-Wallis rank sum tests with outlier exclusion as described earlier for microbial alpha diversity assessments. We ran pre-exposure (at 3-5 weeks old) and post-exposure (at 10-12 weeks old) OF and EPM testing on the mice, however we focus our analyses on the post-exposure results.

Quantifying dust exposures. At the completion of the experiment, we quantified the amount of airborne dust deposited within each enclosure by collecting and weighing material deposited over 3 × 20 cm lengths of 2.5 cm wide (i.e. 3 × 20 cm × 2.5 cm= 150 cm^2) exposed internal framework at the mid-height of the enclosures and directly above the mouse cage position. Dust was collected on approx. 4 cm × 4 cm cloth squares which had been pre-weighed together with a pre-labelled 5 mL tube for each sample, to +/- 0.05 mg accuracy on precision scales (ME5 Microbalance, Sartorius, Germany). After collecting dust, the cloth and tube were weighed again and the difference calculated. We assumed a similar rate of dust deposition, as measured on the exposed internal frame, would occur over each mouse cage which had a plan area of 41 cm × 25 cm = 1025 cm^2. Therefore, we estimated the total available soil dust exposure in $g.mouse^{-1}.week^{-1}$ by calculating the: total collected dust in each enclosure [mg] × (1/150)[rate per cm^2] × (1/3)[per mouse] × (1/7)[per week] × (1/1000)[g per mg] × 1025[cage area cm^2].

Comparison rates of bulk soil exposure used by ref. [22] were calculated based on their methods and soil data. They used 300 mL of soil (with a moist bulk density of 430 $g.L^{-1}$) per cage per week, with 4 mice per cage. Estimated soil exposures of ~ 32 $g.mouse^{-1}.week^{-1}$ were calculated from: 430 [g per L] × 300 [mL of soil per week] × (1/1000) [L per mL] × (1/4) [per mouse]. Also, ref. [23] used 50 g total per cage of house dust, organic matter and top soil in their treatments.

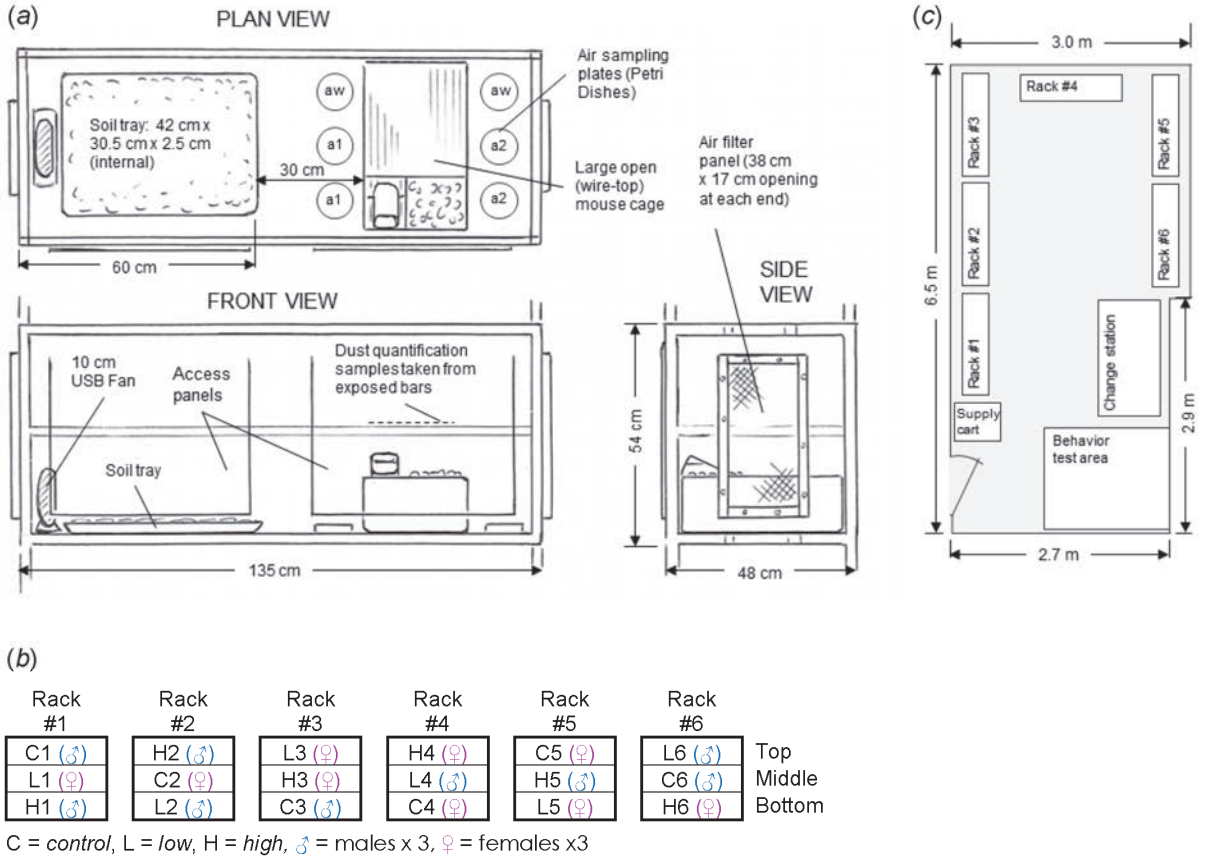


Figure S1. Enclosure design and layout. (*a*) Individually isolated treatment enclosures were constructed around 2 shelves of a conventional 6-shelf metal mouse rack, so that 3 enclosures occupied one full rack (also see figure S2). Each enclosure contained a large open wire-top cage with 3 mice. Airborne soil dust exposures were generated by a 10 cm USB-powered fan adjacent to a shallow tray of soil. Filter panels at each end allowed ventilation of gases while minimising external aerobiological influence. Petri dishes for weekly passive air / dust sampling for microbiota (a1, a2) were placed either side of the cage. Petri dishes for weekly dust weight quantification (aw) were ultimately not used due to negligible quantities of dust settling on a weekly basis. Instead, total accumulated dust was measured along 3 × 20 cm lengths of exposed internal bars, as indicated. Additional dimensions of the enclosure layout are indicated. Diagram not to scale. (*b*) Layout of treatment enclosures and sexes within racks. (*c*) Arrangement of racks within the study room.

Figure S2. Images of experimental apparatus used for the environmental microbiota treatments.

16

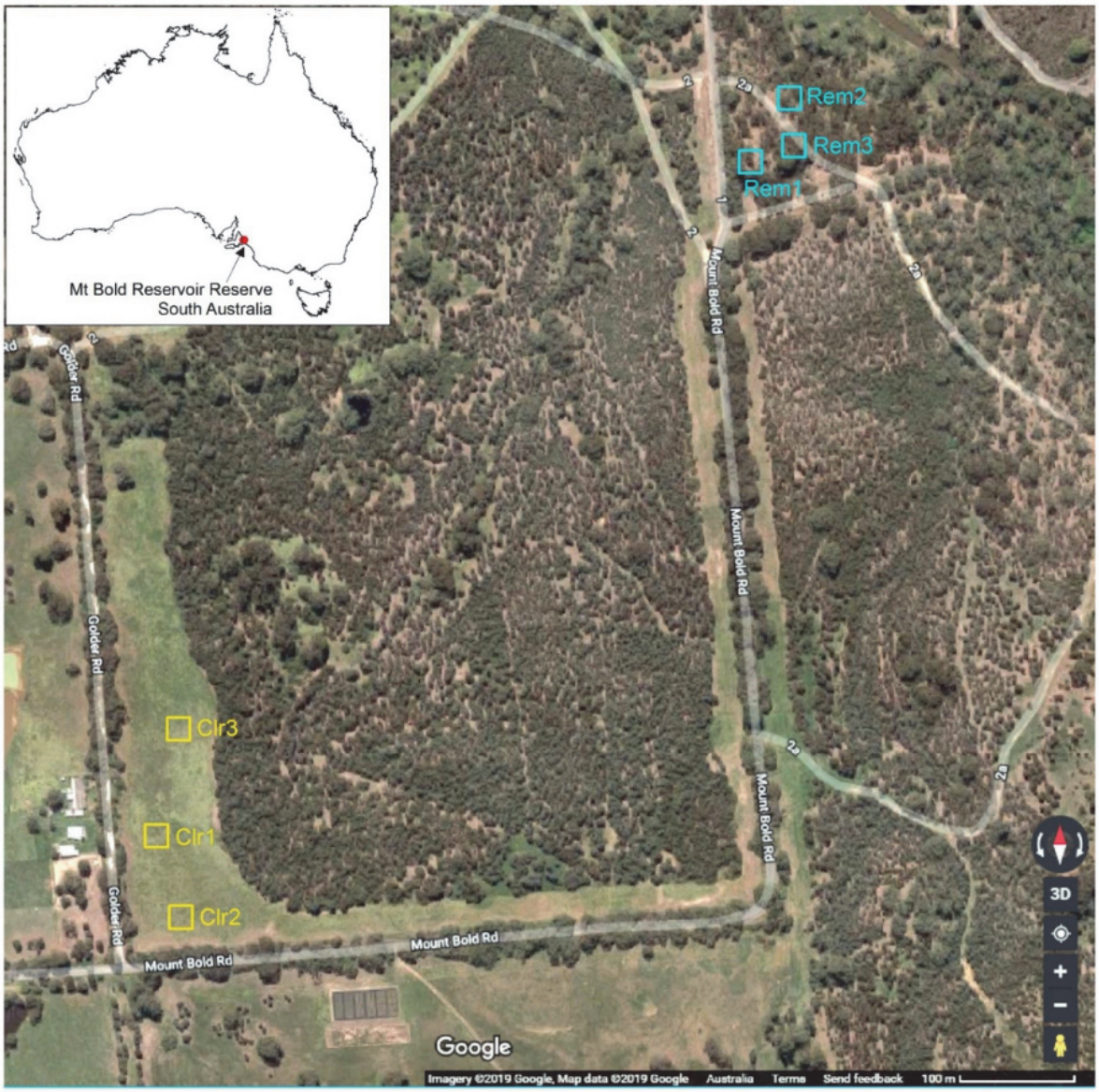


Figure S3. Source soil sampling locations. Source soils were sampled from Mount Bold Reservoir Reserve, Mount Bold Road, Kangarilla, South Australia (Map data: Google 2019). Clr1, Clr2, and Clr3 (used to make *low* soil treatments) are from cleared land and reflect low plant macro-diversity. Rem1, Rem2, and Rem3 (used to make *high* soil treatments) are from neighbouring remnant native vegetation and reflect high plant macro-diversity. Locations and representative soil test data are in Tables S4 and S5.

(*a*)

(*b*)

(*c*)

(*d*)

(*e*)

(*f*)

Figure S4. 360-degree panoramic photos of source soil sampling locations. Remnant vegetation sites comprised: (*a*) Rem1; (*b*) Rem2; and (*c*) Rem3. Cleared sites comprised: (*d*) Clr1; (*e*) Clr2; and (*f*) Clr3.

18

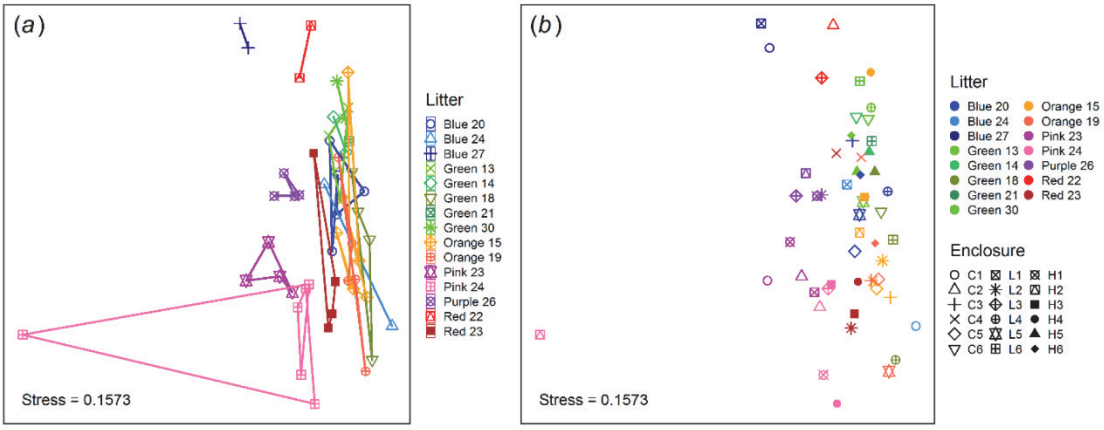


Figure S5. Mice faecal bacteria in Week 0. (*a*) Initial microbiota compositions were significantly associated with litters (PERMANOVA df = 14, F = 3.39, R^2 = 0.543, *P* = 0.001, n = median 4 with interquartile range: 2-4.5 per group), while beta dispersions across litters are comparable (df = 14, F = 1.05, *P* = 0.41). (*b*) Using the same data, the distribution of littermates across the treatment enclosures are indicated. NMDS visualisations are based on rarefied OTU abundances (sequence depth 11450). These plots present the same data as the main figure 1*a* but here emphasise litters and their distribution across enclosures.

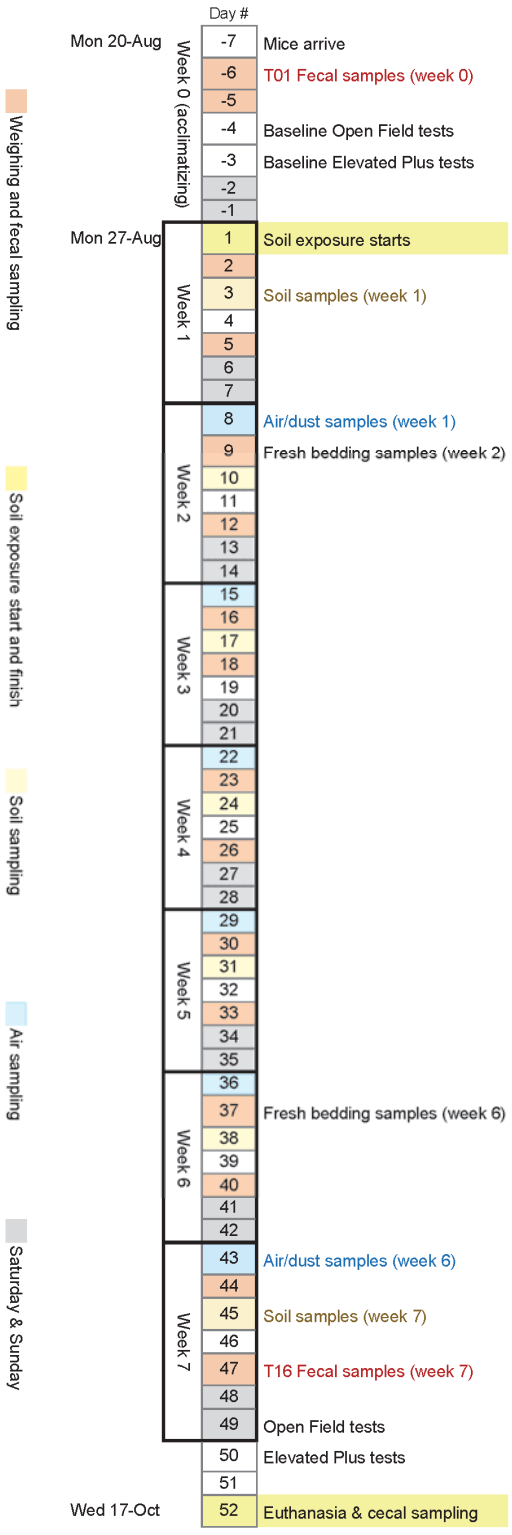


Figure S6. Experimental timeline.

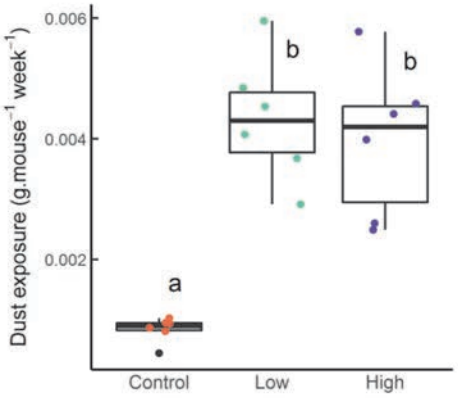


Figure S7. Dust exposures. *High* and *low* treatments were significantly different to background *controls* (Kruskal-Wallis $\chi^2 = 10.2$, df = 2, $P = 0.006$, n = 5 to 6 per group; groups not sharing a letter are different). Dust levels in *control* enclosures had a median = 0.0009 (interquartile range, IQR: 0.00082, 0.00095) g.mouse^{-1}week^{-1}; *low* enclosures had a median = 0.0043 (IQR: 0.0038, 0.0048) g.mouse^{-1}week^{-1}; and *high* enclosures had a median = 0.0042 (IQR: 0.0029, 0.0045) g.mouse^{-1}week^{-1}. Therefore, treatments represented median soil exposures of approx. 0.0034 g.mouse^{-1}week^{-1} above background *control* levels.

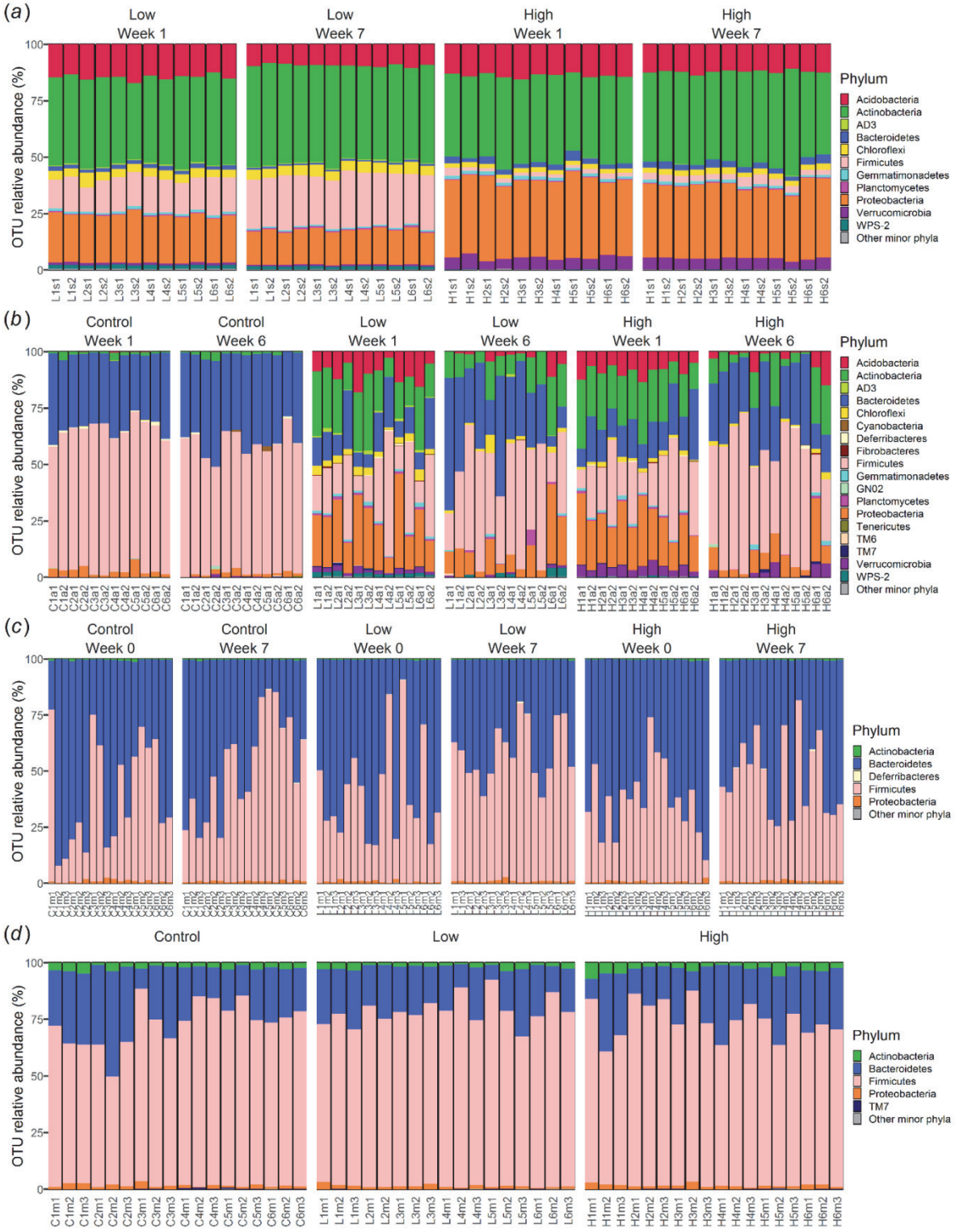


Figure S8. OTU relative abundance bar plots. (*a*) treatment soils; (*b*) air; (*c*) faecal; (*d*) caecal samples. Only phyla with ≥ 0.5% relative abundance are displayed.

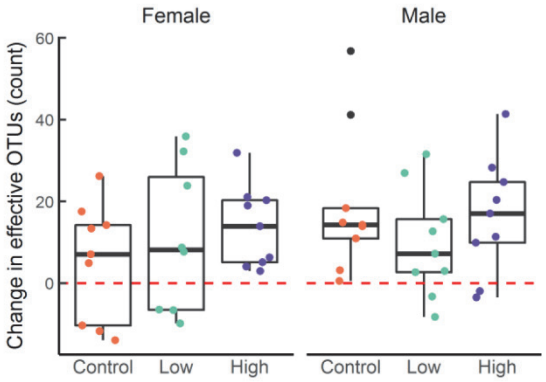


Figure S9. Change in faecal bacteria alpha diversity within animals. Plots show the difference in effective number of OTUs within each animal, between week 0 and week 7. All females in the *high* biodiversity treatment group experience an increase in faecal bacteria alpha diversity following the exposure.

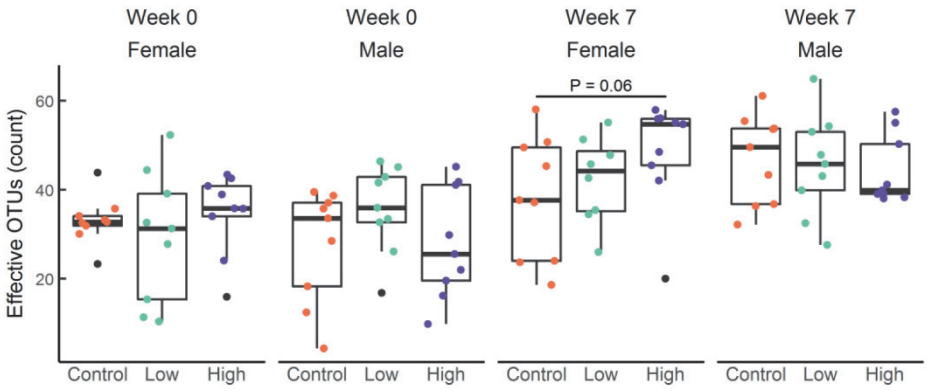


Figure S10. Treatment-wise comparisons of faecal bacteria alpha diversity. Plots show the effective number of OTUs in faecal samples in week 0 and week 7, across females and males. The week 7 female faecal samples bordered on a significant trend for increasing alpha diversity towards the higher biodiversity exposure (Kruskal-Wallis $\chi^2 = 5.58$, df = 2, $P = 0.06$, n = 8 to 9 per group).

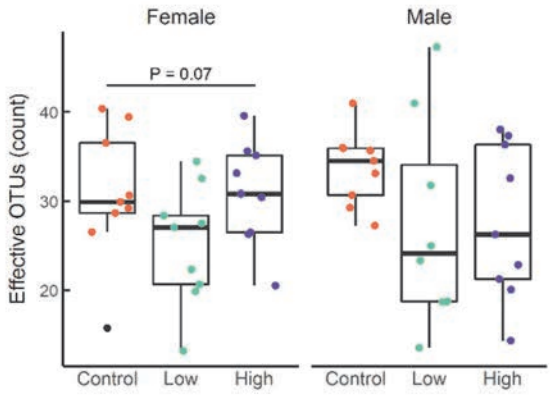


Figure S11. Alpha diversity of caecal bacteria. The plot shows the effective number of OTUs in caecal samples, separated by treatment and sex. Female caecal samples were not significantly different (Kruskal-Wallis $\chi^2 = 5.29$, df = 2, $P = 0.07$, n = 8 to 9 per group).

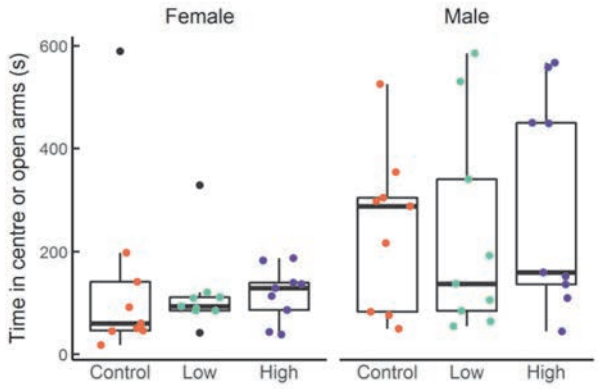


Figure S12. Elevated Plus Maze anxiety-like behaviour test results. The plots show time spent in the centre and open arms of the Elevated Plus Maze. No overall trends were observed by gender or treatment.

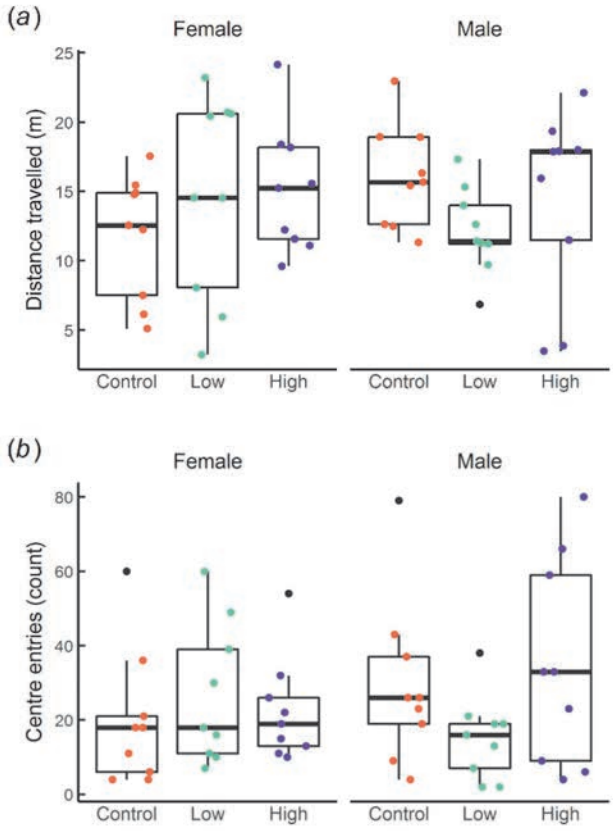


Figure S13. Open field activity levels. (*a*) total distance travelled, and (*b*) counts of centre entries. No significant differences were observed between treatment groups, both in overall terms, nor analyzing separately by gender, in both distance travelled and centre entries. These results suggest that the Open Field reduced anxiety-like behaviour seen in *high* biodiversity treatment females was not due to underlying differences in activity levels.

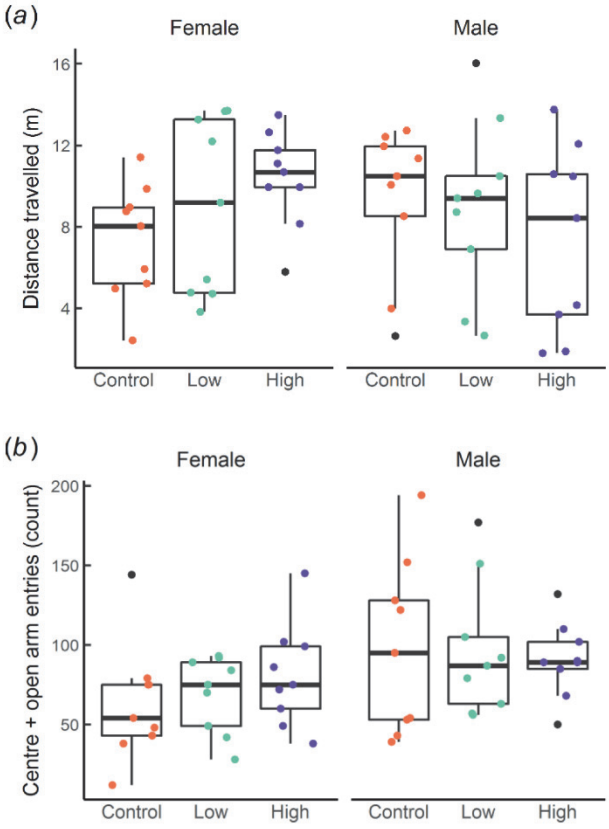


Figure S14. Elevated Plus maze activity levels. (*a*) total distance travelled, and (*b*) counts of entries into the centre and open arms. No significant differences were observed between treatment groups, both in overall terms, nor analyzing separately by gender, in both distance travelled and centre and open arm entries.

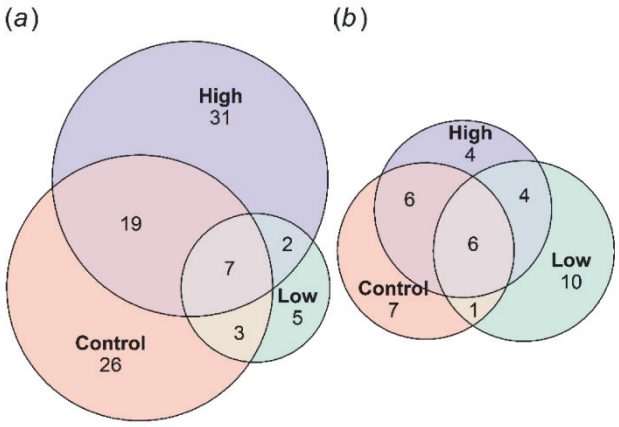


Figure S15. Counts of significant differentially abundant faecal bacterial OTUs and their overlap across treatments. Plots summarise the counts of OTUs that (*a*) increase, and (*b*) decrease, between week 0 and week 7.

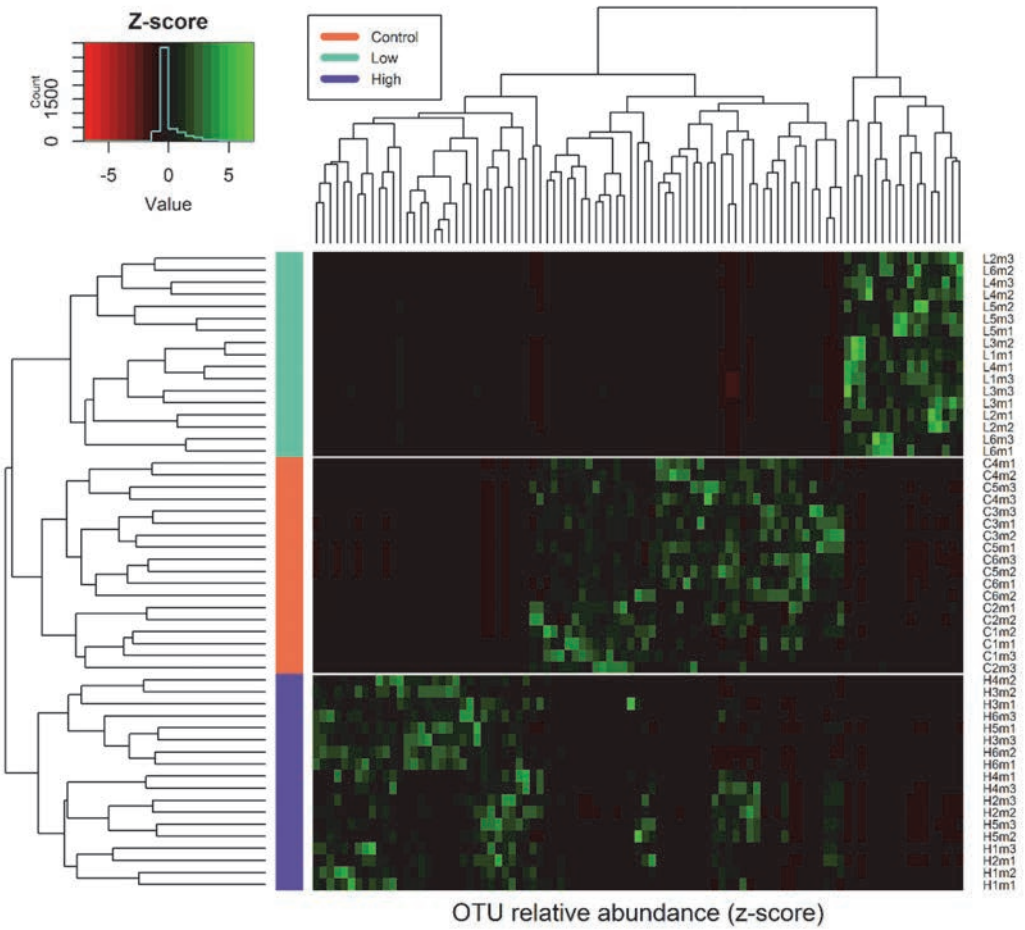


Figure S16. Heatmap of OTU relative abundance z-scores for treatment-specific increasing taxa in week 7 faecal microbiota. The figure highlights only taxa that are significantly increasing within each treatment from week 0 to week 7 based on differential abundance testing. OTU relative abundance data are scaled within taxa (columns) then within samples (rows) to highlight the relative enrichment of different taxa between treatments.

Table S1. Exclusively detected (unshared) significantly increasing taxa in *CONTROL* faecal samples from week 0 to week 7 (based on differential abundance testing)

OTU ID	Closest NCBI match (putative species)	NCBI Accession No	Expect	Percent identity (%)	Phylum	Air (dust)		Faecal week 0		Faecal week 7		Caecal	
						OTU relative abundance (%) Mean (range)	No. samples detected (total)	OTU relative abundance (%) Mean (range)	No. samples detected (total)	OTU relative abundance (%) Mean (range)	No. samples detected (total)	OTU relative abundance (%) Mean (range)	No. samples detected (total)
OTU_19	Bacteroides vulgatus	NR_074515.1	0	100	Bacteroidetes	1.1103 (0–3.6104)	23 (24)	1.4988 (0.0032–16.6367)	18 (18)	1.4691(0.086–3.8173)	18 (18)	0.4112(0.0083–1.4708)	18 (18)
OTU_57	[Eubacterium] tortuosum	NR_044648.2	3e-167	91	Firmicutes	0.3113 (0–3.8073)	16 (24)	0.0207(0–0.1407)	9 (18)	0.121(0.0069–0.2963)	18 (18)	0.3365(0.022–0.8903)	18 (18)
OTU_67	Oscillibacter valericigenes	NR_074793.2	0	95	Firmicutes	0.4603 (0–1.0335)	22 (24)	0.0874(0.0063–0.3653)	18 (18)	0.2378(0.0647–0.7915)	18 (18)	0.0978(0.0054–0.248)	18 (18)
OTU_130	Ihubacter massiliensis	NR_144749.1	0	98	Firmicutes	0.0297 (0–0.3822)	5 (24)	0.006(0–0.0329)	7 (18)	0.0371(0–0.1263)	12 (18)	0.0645(0–0.1458)	12 (18)
OTU_133	[Clostridium] polysaccharolyticum	NR_119085.1	0	95	Firmicutes	0.1255 (0–0.5949)	11 (24)	0.0137(0–0.0619)	15 (18)	0.0345(0.0062–0.0896)	18 (18)	0.0809(0.0062–0.3335)	18 (18)
OTU_143	Kineothrix alysoides	NR_156080.1	1e-160	95	Firmicutes	0.035 (0–0.4139)	5 (24)	0.0033(0–0.0447)	3 (18)	0.0511(0–0.2184)	9 (18)	0.0665(0–0.3543)	10 (18)
OTU_169	Tindallia californiensis	NR_025162.1	8e-69	79	TM7	0.028 (0–0.2728)	3 (24)	0.0014(0–0.0088)	4 (18)	0.0173(0–0.0756)	13 (18)	0.1278(0–0.6941)	14 (18)
OTU_201	Anaeroworax odorimutans	NR_028911.1	1e-161	92	Firmicutes	0.0013 (0–0.0305)	1 (24)	0.0048(0–0.0335)	5 (18)	0.0165(0–0.0356)	17 (18)	0.039(0.017–0.0876)	18 (18)
OTU_202	Paraeggerthella hongkongensis	NR_042825.1	1e-165	93	Actinobacteria	0.0081 (0–0.1661)	2 (24)	0.001(0–0.0116)	3 (18)	0.0074(0–0.0192)	12 (18)	0.0465(0–0.1626)	12 (18)
OTU_247	Roseburia inulinivorans	NR_042007.1	2e-169	93	Firmicutes	0.0273 (0–0.3639)	5 (24)	0.0064(0–0.054)	4 (18)	0.0522(0–0.1768)	13 (18)	0.0468(0–0.3636)	12 (18)
OTU_254	[Clostridium] leptum	NR_114789.1	3e-177	94	Firmicutes	0.0526 (0–0.2681)	8 (24)	0.0105(0–0.0494)	13 (18)	0.028(0.0033–0.0672)	18 (18)	0.0723(0.0178–0.1718)	18 (18)
OTU_258	Ruminiclostridium thermocellum	NR_074629.1	3e-147	90	Firmicutes	0.0104 (0–0.1662)	2 (24)	0.0067(0–0.1143)	3 (18)	0.0047(0–0.0312)	9 (18)	0.0146(0–0.0698)	9 (18)
OTU_402	Lachnotalea glycerini	NR_148620.1	0	96	Firmicutes	0.0386 (0–0.5815)	4 (24)	0.004(0–0.0476)	7 (18)	0.017(0–0.0529)	17 (18)	0.0021(0–0.0092)	9 (18)
OTU_419	Acetatifactor muris	NR_117905.1	1e-170	93	Firmicutes	Nd	nd	0.0016(0–0.0088)	6 (18)	0.0083(0–0.023)	10 (18)	0.0234(0–0.1101)	12 (18)
OTU_531	Kineothrix alysoides	NR_156081.1	0	97	Firmicutes	0.0071 (0–0.1467)	2 (24)	0.0029(0–0.0156)	7 (18)	0.0103(0–0.0311)	15 (18)	0.0069(0–0.0233)	14 (18)
OTU_2384	[Clostridium] indolis	NR_026493.1	1e-175	94	Firmicutes	Nd	nd	0.0222(0–0.1619)	9 (18)	0.1745(0.0033–0.8359)	18 (18)	0.0395(0.001–0.2237)	18 (18)
OTU_2934	[Ruminococcus] gnavus	NR_036800.1	7e-174	94	Firmicutes	3e-04 (0–0.0033)	3 (24)	0.0163(0–0.1816)	13 (18)	0.0674(0.0031–0.2979)	18 (18)	0.0114(0–0.0799)	12 (18)
OTU_3671	Ruminococcus faecis	NR_116747.1	7e-154	91	Firmicutes	4e-04 (0–0.0074)	2 (24)	0.0025(0–0.0216)	5 (18)	0.0283(0–0.0636)	17 (18)	0.0037(0–0.0165)	11 (18)
OTU_3795	Oscillibacter valericigenes	NR_074793.2	2e-168	93	Firmicutes	0.0037 (0–0.0238)	9 (24)	0.0021(0–0.0191)	6 (18)	0.0104(0–0.0489)	16 (18)	0.0015(0–0.0042)	10 (18)
OTU_4034	Faecalimonas umbilicata	NR_156907.1	3e-172	94	Firmicutes	nd	nd	0.0036(0–0.0167)	9 (18)	0.018(0–0.0538)	16 (18)	0.0053(0–0.0192)	11 (18)
OTU_4077	[Clostridium] indolis	NR_026493.1	7e-169	93	Firmicutes	nd	nd	2e-04(0–0.0026)	1 (18)	0.0064(0–0.0269)	12 (18)	0.0054(0–0.0162)	16 (18)
OTU_4702	Falcatimonas natans	NR_152688.1	1e-170	93	Firmicutes	nd	nd	7e-04(0–0.0106)	2 (18)	0.0063(0–0.0184)	16 (18)	0.003(0–0.0142)	11 (18)
OTU_5971	Kineothrix alysoides	NR_156080.1	3e-162	92	Firmicutes	nd	nd	0.0019(0–0.0168)	3 (18)	0.0174(0–0.0848)	13 (18)	0.0018(0–0.0083)	8 (18)
OTU_6689	[Ruminococcus] gnavus	NR_036800.1	3e-177	94	Firmicutes	nd	nd	7e-04(0–0.0064)	3 (18)	0.0082(0–0.0653)	12 (18)	8e-04(0–0.0047)	8 (18)
OTU_7119	Marvinbryantia formatexigens	NR_042152.1	1e-111	65	Bacteroidetes	nd	nd	3e-04(0–0.0033)	2 (18)	0.0038(0–0.0098)	12 (18)	6e-04(0–0.0027)	8 (18)
OTU_8231	[Clostridium] scindens	NR_028785.1	3e-108	91	Bacteroidetes	nd	nd	2e-04(0–0.0031)	1 (18)	0.0074(0–0.0332)	9 (18)	0.0012(0–0.0089)	6 (18)

nd = not detected

Table S2. Exclusively detected (unshared) significantly increasing taxa in *LOW* faecal samples from week 0 to week 7 (based on differential abundance testing)

OTU ID	Closest NCBI match (putative species)	NCBI Accession No	Expect	Percent identity (%)	Phylum	Soil		Air (dust)		Faecal week 0		Faecal week 7		Caecal	
						OTU relative abundance (%) Mean (range)	No. samples detected (total)	OTU relative abundance (%) Mean (range)	No. samples detected (total)	OTU relative abundance (%) Mean (range)	No. samples detected (total)	OTU relative abundance (%) Mean (range)	No. samples detected (total)	OTU relative abundance (%) Mean (range)	No. samples detected (total)
OTU_240	Cuneatibacter caecimuris	NR_144608.1	0	96	Firmicutes	nd	nd	0.0333(0–0.4321)	3 (24)	0.0112(0–0.0559)	10 (18)	0.0541(0.0092–0.193)	17 (17)	0.0749(0.0137–0.2486)	17 (17)
OTU_332	Flintibacter butyricus	NR_144611.1	2e-163	92	Firmicutes	nd	nd	0.0144(0–0.3244)	2 (24)	0.0055(0–0.0998)	1 (18)	0.0135(0–0.058)	9 (17)	0.0027(0–0.0346)	4 (17)
OTU_5095	Pseudoflavonifractor phocaeensis	NR_147370.1	0	97	Firmicutes	nd	nd	0.0665(0–1.2262)	5 (24)	0.008(0–0.027)	12 (18)	0.0467(0–0.2433)	15 (17)	0.0195(0–0.1376)	15 (17)
OTU_8615	Prevotellamassilia timonensis	NR_144750.1	3e-128	86	Bacteroidetes	nd	nd	nd	nd	0.0053(0–0.0835)	3 (18)	0.006(0–0.0232)	11 (17)	6e-04(0–0.0051)	3 (17)
OTU_9639	Kineothrix alysoides	NR_156081.1	3e-172	94	Firmicutes	nd	nd	nd	nd	0.0019(0–0.0143)	5 (18)	0.0104(0–0.0571)	12 (17)	0.0017(0–0.0105)	6 (17)

nd = not detected

Table S3. Exclusively detected (unshared) significantly increasing taxa in *HIGH* faecal samples from week 0 to week 7 (based on differential abundance testing)

OTU ID	Closest NCBI match (putative species)	NCBI Accession No	Expect	Percent identity (%)	Phylum	Soil		Air (dust)		Faecal week 0		Faecal week 7		Caecal	
						OTU relative abundance (%) Mean (range)	No. samples detected (total)	OTU relative abundance (%) Mean (range)	No. samples detected (total)	OTU relative abundance (%) Mean (range)	No. samples detected (total)	OTU relative abundance (%) Mean (range)	No. samples detected (total)	OTU relative abundance (%) Mean (range)	No. samples detected (total)
OTU_26	Muribaculum intestinale	NR_144616.1	4e-131	87	Bacteroidetes	2e-04(0–0.0048)	2 (32)	0.4411(0–2.4194)	12 (24)	0.1035(0–1.0357)	8 (18)	1.5057(0–3.3)	15 (18)	0.453(0–1.2249)	15 (18)
OTU_29	Muribaculum intestinale	NR_144616.1	6e-160	91	Bacteroidetes	3e-04(0–0.005)	3 (32)	0.175(0–1.2735)	6 (24)	0.3272(0–0.9177)	17 (18)	0.614(0.062–1.2088)	18 (18)	0.6381(0.0742–2.1106)	18 (18)
OTU_31	Murimonas intestini	NR_134772.1	0	96	Firmicutes	1e-04(0–0.0024)	1 (32)	0.7253(0–4.0606)	18 (24)	0.2895(0–0.843)	15 (18)	0.8614(0.251–3.6124)	18 (18)	0.2444(0.0455–0.8755)	18 (18)
OTU_35	Prevotella stercorea	NR_041364.1	1e-166	92	Bacteroidetes	nd	nd	0.2767(0–1.8395)	16 (24)	0.0569(0–0.2861)	9 (18)	0.9259(0–3.2227)	12 (18)	0.1036(0–0.3466)	11 (18)
OTU_37	Kineothrix alysoides	NR_156080.1	1e-147	97	Firmicutes	1e-04(0–0.0047)	1 (32)	0.3678(0–1.9866)	15 (24)	0.0466(0–0.2337)	9 (18)	0.23(0.0492–0.8507)	18 (18)	0.242(0.0145–0.9094)	18 (18)
OTU_38	Acetatifactor muris	NR_117905.1	1e-180	95	Firmicutes	1e-04(0–0.0047)	1 (32)	0.389(0–1.5542)	15 (24)	0.213(0–0.9803)	16 (18)	0.5578(0.0493–1.784)	18 (18)	0.2699(0.0449–0.9244)	18 (18)
OTU_60	Kineothrix alysoides	NR_156080.1	0	96	Firmicutes	2e-04(0–0.0048)	2 (32)	0.9457(0–4.9293)	18 (24)	0.444(0–1.3864)	17 (18)	0.8409(0.1409–2.2246)	18 (18)	0.7511(0.2296–1.591)	18 (18)
OTU_79	Prevotella oralis	NR_113117.1	6e-155	90	Bacteroidetes	2e-04(0–0.0061)	1 (32)	0.0881(0–0.593)	7 (24)	0.1253(0–1.7139)	2 (18)	0.2749(0–2.1016)	5 (18)	0.0492(0–0.3055)	4 (18)
OTU_82	Olsenella profuse	NR_116938.1	0	96	Actinobacteria	nd	nd	nd	nd	0.004(0–0.0506)	5 (18)	0.0603(0–0.3924)	17 (18)	0.4719(0.0087–3.0062)	18 (18)
OTU_91	Alistipes indistinctus	NR_113271.1	0	95	Bacteroidetes	1e-04(0–0.0024)	1 (32)	0.0089(0–0.2141)	1 (24)	0.0203(0–0.1181)	10 (18)	0.0724(0.0083–0.1764)	18 (18)	0.122(0.0061–0.2509)	18 (18)
OTU_105	Muribaculum intestinale	NR_144616.1	6e-150	89	Bacteroidetes	1e-04(0–0.0047)	1 (32)	0.014(0–0.1239)	4 (24)	0.0771(0–0.2321)	12 (18)	0.2549(0.0372–0.6221)	18 (18)	0.1597(0.0292–0.3974)	18 (18)
OTU_109	[Clostridium] indolis	NR_026483.1	7e-159	91	Firmicutes	1e-04(0–0.0024)	1 (32)	0.0736(0–0.4883)	9 (24)	0.0237(0–0.2831)	5 (18)	0.0894(0–0.5418)	17 (18)	0.058(0–0.2528)	15 (18)
OTU_115	Phocea massiliensis	NR_144748.1	2e-163	92	Firmicutes	nd	nd	0.0479(0–0.3678)	9 (24)	0.021(0–0.0666)	13 (18)	0.0898(0.0211–0.264)	18 (18)	0.0399(0.0087–0.1266)	18 (18)
OTU_139	Anaerotruncus colihominis	NR_027558.1	0	96	Firmicutes	nd	nd	0.0884(0–1.545)	4 (24)	0.0064(0–0.0179)	10 (18)	0.0484(0–0.1995)	15 (18)	0.0215(0–0.0599)	17 (18)
OTU_173	Intestinimonas butyriciproducens	NR_116554.1	7e-169	93	Firmicutes	nd	nd	0.0715(0–1.2989)	4 (24)	0.0015(0–0.0095)	4 (18)	0.037(0–0.1153)	14 (18)	0.0075(0–0.0274)	9 (18)
OTU_196	Saccharofermentans acetigenes	NR_115340.1	2e-115	85	Firmicutes	nd	nd	0.0554(0–0.7201)	6 (24)	0.0182(0–0.2373)	5 (18)	0.0402(0.0057–0.1796)	18 (18)	0.0034(0–0.0231)	9 (18)
OTU_556	Oscillibacter ruminantium	NR_118156.1	5e-175	94	Firmicutes	nd	nd	0.0425(0–0.3307)	8 (24)	0.0224(0–0.089)	10 (18)	0.058(0–0.2644)	17 (18)	0.0161(0–0.0519)	16 (18)
OTU_1737	Acetatifactor muris	NR_117905.1	1e-175	94	Firmicutes	nd	nd	0.1065(0–1.4537)	9 (24)	0.0038(0–0.0224)	7 (18)	0.0644(0.0127–0.1422)	18 (18)	0.0232(0–0.0697)	17 (18)
OTU_2750	Muribaculum intestinale	NR_144616.1	3e-158	90	Bacteroidetes	nd	nd	nd	nd	0.0152(0–0.1128)	8 (18)	0.0578(0–0.17)	17 (18)	0.0166(0–0.0661)	16 (18)
OTU_3408	Clostridium cavendishii	NR_115711.1	2e-114	85	Firmicutes	nd	nd	nd	nd	nd	nd	0.0075(0–0.0474)	12 (18)	3e-04(0–0.003)	2 (18)
OTU_3456	Muribaculum intestinale	NR_144616.1	3e-153	90	Bacteroidetes	nd	nd	nd	nd	0.006(0–0.0295)	10 (18)	0.0242(0–0.0962)	17 (18)	0.011(0–0.0459)	17 (18)
OTU_3669	Eisenbergiella massiliensis	NR_144731.1	3e-167	92	Firmicutes	nd	nd	nd	nd	0.0138(0–0.0627)	12 (18)	0.0855(0–0.341)	17 (18)	0.1505(0–0.3509)	17 (18)
OTU_3876	Muribaculum intestinale	NR_144616.1	6e-160	91	Bacteroidetes	nd	nd	nd	nd	0.005(0–0.0273)	7 (18)	0.0203(0–0.062)	15 (18)	0.0127(0–0.0559)	17 (18)
OTU_5375	[Ruminococcus] gnavus	NR_036800.1	1e-160	92	Firmicutes	nd	nd	nd	nd	0.0014(0–0.0128)	3 (18)	0.0089(0–0.0474)	11 (18)	3e-04(0–0.0032)	3 (18)
OTU_5520	Kineothrix alysoides	NR_156081.1	0	96	Firmicutes	nd	nd	0.0016(0–0.0247)	3 (24)	0.0076(0–0.0463)	11 (18)	0.0366(0.0065–0.1548)	18 (18)	0.0521(0.0012–0.1536)	18 (18)
OTU_5765	Bacteroides xylanolyticus	NR_104899.1	3e-162	92	Firmicutes	nd	nd	nd	nd	0.0021(0–0.011)	6 (18)	0.0132(0–0.0517)	13 (18)	0.0089(0–0.0357)	13 (18)
OTU_5790	Kineothrix alysoides	NR_156081.1	5e-136	95	Firmicutes	nd	nd	nd	nd	8e-04(0–0.0126)	2 (18)	0.0155(0–0.0649)	12 (18)	0.009(0–0.0413)	15 (18)
OTU_5915	Muribaculum intestinale	NR_144616.1	1e-166	92	Bacteroidetes	nd	nd	nd	nd	0.0015(0–0.0109)	4 (18)	0.0076(0–0.0325)	12 (18)	0.0032(0–0.0157)	12 (18)
OTU_6509	Muribaculum intestinale	NR_144616.1	9e-143	88	Bacteroidetes	nd	nd	nd	nd	0.009(0–0.092)	6 (18)	0.0413(0.0052–0.1761)	18 (18)	0.0079(0–0.0451)	14 (18)
OTU_7377	Muribaculum intestinale	NR_144616.1	9e-143	88	Bacteroidetes	nd	nd	nd	nd	0.0079(0–0.0613)	5 (18)	0.085(0–0.2536)	15 (18)	0.0154(0–0.0586)	13 (18)
OTU_9265	Muribaculum intestinale	NR_144616.1	1e-137	88	Bacteroidetes	2e-04(0–0.0048)	1 (32)	nd	nd	0.0108(0–0.075)	6 (18)	0.131(0–0.4146)	16 (18)	0.0254(0–0.0872)	14 (18)

nd = not detected

Table S4. Spatial coordinates of source soil sampling locations

Site	Treatment	Latitude	Longitude
Remnant 1 (Rem1)	*High*	-35.123464	138.670893
Remnant 2 (Rem2)	*High*	-35.122853	138.671073
Remnant 3 (Rem3)	*High*	-35.123440	138.671347
Cleared 1 (Clr1)	*Low*	-35.128965	138.665037
Cleared 2 (Clr2)	*Low*	-35.129555	138.665291
Cleared 3 (Clr3)	*Low*	-35.128029	138.665276

Note: coordinates are in WGS84 geographic coordinate reference system (units: decimal degrees).

Table S5. Source soil laboratory test data summary

	Remnant (*high*) soils				Cleared (*low*) soils			
	Rem 1	Rem2	Rem3	Mean (range)	Clr1	Clr2	Clr3	Mean (range)
Colour	Dark Gray	Dark Gray	Gray Brown		Gray Brown	Dark Gray	Dark Gray	
Gravel (%)	0	5	5-10		5-10	5	0	
Clay (%)	8.17	23.39	11.91	14.5 (8.17–23.39)	10.23	7.85	7.16	8.4 (7.16–10.23)
Course Sand (%)	21.80	29.94	23.21	25 (21.8–29.94)	32.25	34.96	43.82	37 (32.25–43.82)
Fine Sand (%)	55.65	39.53	48.94	48 (39.53–55.65)	51.36	51.26	40.78	47.8 (40.78–51.36)
Sand (%)	77.45	69.47	72.15	73 (69.47–77.45)	83.61	86.22	84.60	84.8 (83.61–86.22)
Silt (%)	14.38	7.14	15.94	12.5 (7.14–15.94)	6.17	5.93	8.24	6.8 (5.93–8.24)
Moisture (%)	34.10	28.34	22.26	28.2 (22.26–34.1)	18.15	17.38	22.43	19.3 (17.38–22.43)
Conductivity (dS/m)	0.065	0.065	0.063	0.064 (0.063–0.065)	0.023	0.019	0.028	0.023 (0.019–0.028)
pH (H2O)	6.3	6.7	6.2	6.4 (6.2–6.7)	5.6	5.6	5.4	5.5 (5.4–5.6)
pH (CaCl2)	5.1	5.7	5.0	5.3 (5–5.7)	4.4	4.3	4.2	4.3 (4.2–4.4)
Organic Carbon (%)	3.13	2.82	2.86	2.9 (2.82–3.13)	1.92	1.56	1.81	1.8 (1.56–1.92)
Ammonium Nitrogen (mg/kg)	9	6	5	6.7 (5–9)	6	2	3	3.7 (2–6)
Nitrate Nitrogen (mg/kg)	3	6	6	5 (3–6)	2	2	4	2.7 (2–4)
Phosphorus Colwell (mg/kg)	8	5	5	6 (5–8)	12	11	10	11 (10–12)
Potassium Colwell (mg/kg)	124	185	146	151.7 (124–185)	38	21	38	32.3 (21–38)
Sulfur (mg/kg)	4.9	3.9	4.8	4.5 (3.9–4.9)	2.2	2.0	2.1	2.1 (2–2.2)
DTPA Copper (mg/kg)	2.03	0.56	0.44	1 (0.44–2.03)	0.18	1.35	0.17	0.6 (0.17–1.35)
DTPA Iron (mg/kg)	133.53	55.04	88.25	92.3 (55.04–133.53)	235.48	293.04	271.92	266.8 (235.48–293.04)
DTPA Manganese (mg/kg)	12.50	9.03	9.55	10.4 (9.03–12.5)	2.56	2.50	1.99	2.4 (1.99–2.56)
DTPA Zinc (mg/kg)	1.73	2.12	3.38	2.4 (1.73–3.38)	0.50	0.58	0.57	0.6 (0.5–0.58)
Exc. Aluminium (meq/100g)	0.043	0.045	0.085	0.1 (0.043–0.085)	0.743	0.542	0.516	0.6 (0.516–0.743)
Exc. Calcium (meq/100g)	7.89	10.55	6.20	8.2 (6.2–10.55)	1.44	1.31	1.39	1.4 (1.31–1.44)
Exc. Magnesium (meq/100g)	2.15	3.46	1.81	2.5 (1.81–3.46)	0.40	0.33	0.42	0.4 (0.33–0.42)
Exc. Potassium (meq/100g)	0.27	0.37	0.25	0.3 (0.25–0.37)	0.07	0.05	0.08	0.07 (0.05–0.08)
Exc. Sodium (meq/100g)	0.24	0.20	0.19	0.2 (0.19–0.24)	0.08	0.05	0.06	0.1 (0.05–0.08)
Boron Hot CaCl2 (mg/kg)	0.76	1.18	0.72	0.9 (0.72–1.18)	0.37	0.28	0.29	0.3 (0.28–0.37)

Note: all samples are from 0-10 cm surface soil.

Supplementary References

1. Gellie, N.J.C., Mills, J.G., Breed, M.F. & Lowe, A.J. 2017 Revegetation rewilds the soil bacterial microbiome of an old field. *Molecular Ecology* **26**, 2895-2904. (doi:10.1111/mec.14081).
2. Illumina. 2014 16S Metagenomic Sequencing Library Preparation: Preparing 16S Ribosomal RNA Gene Amplicons for the Illumina MiSeq System. Part # 15044223 Rev. B. https://support.illumina.com/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf.
3. Stamatakis, A., Zhang, J., Kobert, K. & Flouri, T. 2013 PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30**, 614-620. (doi:10.1093/bioinformatics/btt593).
4. Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., et al. 2010 QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7**, 335-336. (doi:10.1038/nmeth.f.303).
5. Edgar, R.C. 2010 Search and clustering orders of magnitude faster than blast. *Bioinformatics* **26**, 2460-2461. (doi:10.1093/bioinformatics/btq461).
6. Edgar, R.C., Haas, B.J., Quince, C., Clemente, J.C. & Knight, R. 2011 UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**, 2194-2200. (doi:10.1093/bioinformatics/btr381).
7. Edgar, R.C. 2013 UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods* **10**, 996-998. (doi:10.1038/nmeth.2604).
8. DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P. & Andersen, G.L. 2006 Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. *Applied and Environmental Microbiology* **72**, 5069-5072. (doi:10.1128/AEM.03006-05).
9. Davis, N.M., Proctor, D.M., Holmes, S.P., Relman, D.A. & Callahan, B.J. 2018 Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome* **6**, 226. (doi:10.1186/s40168-018-0605-2).
10. R Core Team. 2018 R: A language and environment for statistical computing. (Vienna, Austria, R Foundation for Statistical Computing).
11. McMurdie, P.J. & Holmes, S. 2013 phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE* **8**, e61217. (doi:10.1371/journal.pone.0061217).
12. Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., et al. 2018 vegan: Community Ecology Package. R package version 2.5-2.
13. Jost, L. 2006 Entropy and diversity. *Oikos* **113**, 363-375. (doi:10.1111/j.2006.0030-1299.14714.x).
14. Zar, J.H. 2010 Biostatistical Analysis. 5th ed. Upper Saddle River, NJ, Pearson Prentice Hall.

15. Liddicoat, C., Weinstein, P., Bissett, A., Gellie, N., Mills, J., Waycott, M. & Breed, M. 2019 Can bacterial indicators of a grassy woodland restoration inform ecosystem assessment and microbiota-mediated human health? *Environment International* **129**, 105-117. (doi: 10.1016/j.envint.2019.05.011).

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

17. Larsson, J., Godfrey, J.R., Gustafsson, P., Eberly, D.H. & Huber, E. 2018 eulerr: Area-Proportional Euler and Venn Diagrams with Ellipses. R package version 5.0.0.

18. Warnes, G., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., et al. 2016 gplots: Various R Programming Tools for Plotting Data. R package version 3.0.1.

19. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. 1990 Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403-410. (doi: 10.1016/S0022-2836(05)80360-2).

20. Pearson, W.R. 2013 An Introduction to Sequence Similarity ("Homology") Searching. *Current Protocols in Bioinformatics* **42**, 3.1.1-3.1.8. (doi:10.1002/0471250953.bi0301s42).

21. Carola, V., D'Olimpio, F., Brunamonti, E., Mangia, F. & Renzi, P. 2002 Evaluation of the elevated plus-maze and open-field tests for the assessment of anxiety-related behaviour in inbred mice. *Behavioural Brain Research* **134**, 49-57. (doi:10.1016/S0166-4328(01)00452-1).

22. Ottman, N., Ruokolainen, L., Suomalainen, A., Sinkko, H., Karisola, P., Lehtimäki, J., Lehto, M., Hanski, I., Alenius, H. & Fyhrquist, N. 2018 Soil exposure modifies the gut microbiota and supports immune tolerance in a mouse model. *Journal of Allergy and Clinical Immunology* **143**, 1198-1206.e1112. (doi:10.1016/j.jaci.2018.06.024).

23. Zhou, D., Zhang, H., Bai, Z., Zhang, A., Bai, F., Luo, X., Hou, Y., Ding, X., Sun, B., Sun, X., et al. 2016 Exposure to soil, house dust and decaying plants increases gut microbial diversity and decreases serum immunoglobulin E levels in BALB/c mice. *Environmental Microbiology* **18**, 1326-1337. (doi:10.1111/1462-2920.12895).

Chapter 6. Discussion, conclusions, and recommendations

Outline of work

This thesis aimed to build evidence and knowledge of potential beneficial relationships between biodiversity, environmental microbiomes and human health—because knowledge of such relationships has the potential to drive cost-effective improvements to population health outcomes, as well as biodiversity benefits, if we can modify or manage the environment accordingly.

In order to achieve this, I sought evidence of associations, with plausible microbiota-mediated links, through a series of environmental exposure-health outcome studies. Also, I sought to characterise and understand relationships between environments and their microbiomes spanning a gradient of restoration—from disturbed to natural—to help understand potential changes in microbiota and implications for human health.

As proposed in my literature review (*Chapter 1*), my early studies used environmental proxies for microbial diversity to test support for the notion of beneficial biodiverse microbiome-health associations within existing Australian national environmental and public health datasets. Specifically, I first examined relationships between respiratory health hospital admissions across Australia and various social and environmental variables, including measures of landscape biodiversity (*Chapter 2*). Here, I found high biodiversity environments associate with positive respiratory health outcomes. Next, I examined relationships between infectious and parasitic disease risk in regional Australia and social and environmental variables, including mapping for soil cation exchange capacity; which I showed to provide a useful proxy indicator for potential immunomodulatory soil microbial diversity (*Chapter 3*). I found improved health outcomes in areas with higher cation exchange capacity soils.

These positive findings support my *Chapter 1* hypothesis, suggesting that environmental proxies can be used to explore and inform the biodiversity hypothesis for protective microbial-human contact. Indeed, the findings from these two environmental epidemiology studies (*Chapters 2* and *3*) provided direction for subsequent work, and led to progressively more detailed, microbiome-focussed investigations to move closer to examining a causative mechanism. In *Chapter 4*, I therefore examined relationships between aboveground biodiversity and soil microbial communities, to build understanding of how environmental condition (i.e. cleared land /disturbed soil through to restored and remnant native vegetation) can influence respective environmental microbiomes, and also, potentially, human health. These studies pointed to the notion that aboveground biodiversity associates with biodiverse soils, which may influence host microbiota and provide health benefits in exposed individuals.

Therefore, finally, in *Chapter 5* I sought to test these ideas experimentally in a mouse model. Leading a research team, I established a randomised controlled experiment testing different airborne microbial exposures from no soil, low biodiversity soil, and high biodiversity soil. We measured changes in mouse gut microbiota, transfer of microbes from soils, through the air, and into the gut, and also measured anxiety-like behavioural outcomes. My findings from this study supported the notion of both key species and microbial diversity combining to provide a beneficial microbiota-mediated health influence, as I initially proposed in *Chapter 1* (Figure 2).

Discussion

The Australian continent-wide environmental epidemiology studies in *Chapter 2* and *Chapter 3* found that proxy measures of biodiversity and soil microbial diversity associated with reductions in a range of non-communicable and communicable diseases. While the number of

green space-human health correlational studies continues to grow in the scientific literature, in *Chapter 2* I demonstrated that the 'quality of the green space' matters. Specifically, I found that the diversity of vegetation ranked among known beneficial health correlates, such as socioeconomic status, and by far outperformed other landscape measures based on greenness alone.

In *Chapter 3*, based on knowledge that aboveground biodiversity, soil organic matter, and soil microbial diversity are often associated (Chen *et al.*, 2018; Delgado Baquerizo *et al.*, 2018; Coleman *et al.*, 2004), I examined the explanatory value of candidate soil variables when comparing infectious and parasitic disease risk across groupings of socioeconomic and environmental classes, and in a probabilistic machine-learning predictive modelling framework. In these analyses I found a special role for soils in accounting for variation in the health outcomes. Areas with typically higher microbial diversity soils (indicated by high soil cation exchange capacity) were associated with reduced risk of infectious and parasitic disease. Also, including soils in the disease risk modelling boosted prediction performance by 7.5% (Δ r^2=0.075; 95% confidence interval: 0.05, 0.10) in unseen areas (i.e. not used in model building).

In *Chapter 4*, I discovered that opportunistic (e.g. fast-growing, generalist) bacteria, including genera containing noteworthy potential pathogens in humans, were favoured in disturbed soils. Presumably these organisms perform better in soil conditions that provide simpler and less reliable microbial habitat and feedstocks. However, the opportunists were displaced with the restoration of more mature, biodiverse aboveground vegetation. This change is likely due to the increased complexity and stability of microbial habitats and feedstocks, as we observed a corresponding increase in more slow-growing, specialist niche-adapted taxa. Remarkably, I found that the trending bacteria identified from one localized restoration study site in the Adelaide Hills, South Australia, displayed generalizable trends

between soil samples from ‘disturbed’ (e.g. cleared or low biodiversity land) and ‘natural’ (e.g. under native vegetation) sites across Australia generally.

In the mouse model study of *Chapter 5*, to our knowledge, we provided the first evidence supporting a potentially broadly-applicable beneficial microbial linkage between natural biodiversity, gut health and mental health, via airborne soil-derived butyrate-producing bacteria. These bacteria increased in the faecal samples of the high biodiversity treatment mice; they were found in respective soil, air and gut samples; and they appeared to moderate the most anxious behaviour in female mice. This result suggests aerobiome (airborne microbiome) exposure alone can provide a health benefit, a critical finding to link this work to human health outcomes—since a majority of city populations will receive the majority of their environmental microbiome exposure via this pathway.

Towards understanding of what is a ‘healthy environmental microbiome’

These studies offer new perspectives on the contentious question of what constitutes a healthy environmental microbiome? Based on my research, I suggest as a starting point that a healthy environmental microbiome brings together the protective qualities of microbial diversity and key species (Old Friends), and is capable of offering additional benefits beyond the mere sum of its parts.

It is already recognised that increased microbial diversity may have a health-protecting role due to greater capacity for ecological controls on potential opportunistic pathogens (e.g. via competition, predation, diluting quorum sensing signals); and also providing greater scope for immune system education and training (Rook, 2013; van Elsas *et al.*, 2012; von Hertzen *et al.*, 2011). Key species, or Old Friends, are also recognised as being supplied from natural systems such as soil; however examples are limited (e.g. Reber *et al.*,

2016) and potential applications for environmental management to deliver cost-effective public health gains are unclear.

My research suggests that restoring mature, biodiverse (e.g. towards naturally adapted) plant-soil systems can not only boost microbial diversity (e.g. as measured under ecological restoration at Mt Bold) but also provide additional benefits in the form of: (i) displacing opportunistic and potential pathogenic bacteria with slower growing, specialist niche-adapted taxa; and (ii) enhancing the supply and exposure to beneficial key species, in particular organisms closely resembling the butyrate-producer *Kineothrix alysoides* (Haas and Blanchard, 2017). *Kineothrix alysoides* is an anaerobic spore-former that appears to have an affinity with both biodiverse, organic-rich soils and the mammalian gut, and deserves further investigation as a candidate Old Friend. Being an anaerobic spore-former, this bacterium is capable of surviving and dispersing in a dormant state in the environment, and has the capacity to reactivate under anaerobic conditions. Meanwhile, the same naturally-diverse microbial community is likely to supply additional as-yet-unrecognised members involved in supporting complex butyrate production pathways (Vital *et al.*, 2017). When these butyrate producers colonise the mammalian gut (as we observed most in the high biodiversity treatment in our mouse model) they may ultimately provide broad health benefits through supporting gut health and via the gut-brain-microbiome axis (Sanna *et al.*, 2019; Valles-Colomer *et al.*, 2019; Furusawa *et al.*, 2013; Nicholson *et al.*, 2012). In particular, my research suggests that soil-associated components of the microbiome, and aerobiome, from biodiverse environments can provide an important link to human health. It is possible that this mechanism may help explain many of the beneficial biodiverse green space-human health and mental health associations seen elsewhere.

New scientific methods developed

In order to examine a number of hypotheses it was necessary to develop new methods and apparatus, either due to: a lack of suitable available options, adaptations to data limitations, and to maximise value from available data. In particular, I developed:

- A method for probabilistic randomised pseudo-individual level sampling to build multi-level machine-learning disease risk models (*Chapter 3*). This technique provided a means to explore the explanatory value of fine-resolution environmental data (e.g. soil quality) in predicting coarser area-level health outcome data.
- A method for merged-sample bootstrap resampling (*Chapter 4*) to provide deeper analyses of microbiome data. The technique provides normalization for sampling effort while preserving microbiome data (in contrast to conventional rarefying), and provides a means to measure the characteristics and uncertainty associated with site-level (multiple subsample) data.
- A method using bacterial indicator groups to compare different sites based on an established reference frame (e.g. grading from ‘disturbed’ to ‘natural’ condition). This technique may have application in comparing and evaluating soil- and ecosystem-condition in sites elsewhere (*Chapter 4*).
- Environmental microbiome exposure enclosures to provide aerobiome treatments (via light soil dust) while controlling other key influences on the microbiome. These enclosures aimed to model the influence of spending time in different types of environments—reflecting a potential primary pathway for varying microbiome exposures—from artificial sanitised surroundings (i.e. no soil dust) to biodiverse green space (i.e. high biodiversity soil dust).

Limitations

These studies contained important limitations. The environmental epidemiology studies were cross-sectional and based on area-level and aggregated disease class data. Aggregated health response data represents a limitation in terms of loss of specificity to link environmental influence with any particular disease. However, using such data may offer greater sensitivity to detect possible broad environmental influences on the underlying immune fitness in the population. In *Chapter 3* especially we aimed to moderate the limitations of area-level response data through modelling based on probabilistic pseudo-individual-level sampling of environmental exposures.

Our studies involving microbiome data (*Chapters 3* to *5*) could be improved on by including data on actual biomass of microbiota. Also, amplicon-based OTU abundance data are subject to taxon-specific biases (e.g. during DNA extraction and polymerase chain reaction amplification of DNA). Despite these recognised limitations, we applied appropriate analytic techniques, and found coherent patterns across a range of samples and ecosystems, including meaningful potential implications for human health. We acknowledge that microbiome data analytic methods represent an area of active research and development.

Finally, for *Chapter 5*, we do not expect our mouse model study to exactly reflect human physiological responses to environmental exposures, as is the case with any animal model study. However, mice do offer a widely accepted model system (Laukens *et al.*, 2016) and the capability to control for key variables influencing the microbiome (e.g. genetics, diet, etc.), which was critical in our study of the influence of naturally-diverse airborne environmental microbiota.

I minimised the potential effect of these limitations by moving from associative studies to experimental studies, and found reinforcement of my conclusions at each stage. Given that the limitations differ by study and that the studies all support the same notion of a

health benefit from biodiversity and biodiverse soil exposure, I am confident that the conclusions hold despite these limitations.

Significance

The implications of this research are considerable, with potential applications in delivering cost-effective public health interventions and potentially in clinical medicine. A significant proportion of the human disease burden could potentially be saved either by supplementing individuals with specific combinations of microbes (e.g. including natural butyrate-producers), or by increasing the exposure of populations to biodiversity through the inclusion of well-designed, microbiome-conscious green space in urban planning. Also, my studies suggest that establishing more mature, biodiverse environments may suppress soil-derived airborne pathogens. My research has applications in both environmental management and public health policy due to the potential for health benefits and pathogen suppression to follow from the inclusion of biodiverse environments in urban planning. There is potential for activities such as environmental volunteering, gardening, and other activities involving low-level soil exposure to be prescribed by health professionals for beneficial microbial exposures, provided that policy makes access to such activities feasible for all. Further research is warranted to understand exposure-dose-response relationships for the key microbial mechanism discussed in *Chapter 5*. Also, my findings suggest that areas of long-term disturbed and degraded soils should be minimised in urban environments to minimise public health risks from opportunistic bacteria.

The new techniques I have developed will have applications in enhancing microbiome analyses (via merged-sample bootstrap resampling) and help inform new DNA-based methods for assessing soil and ecosystem condition (using bacterial indicator groups). The original findings and methodological development both provide original contributions to a

rapidly growing field. Importantly, they both contribute at the intersection between environmental and clinical microbiome studies—arguably an understudied but important area for contributing to: a burgeoning environmental-biomedical prospecting sector; and the design of new cost-effective population health interventions with simultaneous promotion of biodiversity conservation and restoration.

Future work

Further research questions and recommendations have arisen from my work. My research suggests a special role for *K. alysoides*, and I believe this species warrants further investigation as a candidate Old Friend. Additionally, many bacterial taxa were identified to trend with ecological restoration (*Chapter 4*), and the potential immunomodulatory value of these taxa should be examined.

A further question is whether natural butyrate-producers are commonly found across a range of soils? I suggest that various types of biodiverse and non-biodiverse green space (e.g. natural areas, restored/revegetated areas, urban gardens, production agriculture) should be examined. Also, knowledge of environmental microbiome exposure-dose-health response relationships is still lacking. More thought is required on how this might be examined in a structured way that is relevant to environmental management for public health benefits.

I recommend that further research examines whether universal DNA extraction techniques should be supplemented with specialist techniques designed for lysis-resistant spores. This may address the apparent under-representation of suggested key spore-forming bacteria such as *K. alysoides*.

Results from the mouse model study pointed to the expression of key species relationships in the more redox-constrained conditions of the caecum, whereas treatment-

related patterns in microbial community and diversity emerged in faecal samples. These results may have implications for other studies investigating key species vs. microbiota composition, and should be investigated in further detail. If these patterns are consistent (e.g. faecal samples are more likely to display community profile effects vs. caecal samples are more likely to show key species effects) such information may benefit future research study design.

Also, from the mouse study, the behavioural response to our high biodiversity treatment was observed only in female mice. This was an intriguing result that warrants further investigation.

Lastly, a key knowledge gap that warrants investigation occurs at the intersection of environmental, health, and economics—to determine the cost-effectiveness of establishing microbiome-conscious urban green space interventions to help decrease the disease burden. As a hypothetical example, would investing $1 to establish biodiverse vegetation and soils in a high use public space translate to a saving of $5 in disease burden?

Bibliography (in addition to published and submitted chapters)

Chen, S., Wang, W., Xu, W., Wang, Y., Wan, H., Chen, D., Tang, Z., Tang, X., Zhou, G., Xie, Z., Zhou, D., Shangguan, Z., Huang, J., He, J.-S., Wang, Y., Sheng, J., Tang, L., Li, X., Dong, M., Wu, Y., Wang, Q., Wang, Z., Wu, J., Chapin, F.S., and Bai, Y. (2018). Plant diversity enhances productivity and soil carbon storage. *Proceedings of the National Academy of Sciences* 115, 4027.

Coleman, D., Crossley, D., and Hendrix, P. (2004). Fundamentals of soil ecology (2nd ed.). Burlington USA: Elsevier Academic Press.

Delgado Baquerizo, M., Reith, F., Dennis, P.G., Hamonts, K., Powell, J.R., Young, A., Singh, B.K., and Bissett, A. (2018). Ecological drivers of soil microbial diversity and soil biological networks in the Southern Hemisphere. *Ecology* 99, 583-596.

Furusawa, Y., Obata, Y., Fukuda, S., Endo, T.A., Nakato, G., Takahashi, D., Nakanishi, Y., Uetake, C., Kato, K., Kato, T., Takahashi, M., Fukuda, N.N., Murakami, S., Miyauchi, E., Hino, S., Atarashi, K., Onawa, S., Fujimura, Y., Lockett, T., Clarke, J.M., Topping, D.L., Tomita, M., Hori, S., Ohara, O., Morita, T., Koseki, H., Kikuchi, J., Honda, K., Hase, K., and Ohno, H. (2013). Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. *Nature* 504, 446-450.

Gellie, N.J.C., Mills, J.G., Breed, M.F., and Lowe, A.J. (2017). Revegetation rewilds the soil bacterial microbiome of an old field. *Molecular Ecology* 26, 2895-2904.

Haas, K.N., and 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, 402-410.

Laukens, D., Brinkman, B.M., Raes, J., De Vos, M., and Vandenabeele, P. (2016). Heterogeneity of the gut microbiome in mice: guidelines for optimizing experimental design. *FEMS Microbiology Reviews* 40, 117-132.

Nicholson, J.K., Holmes, E., Kinross, J., Burcelin, R., Gibson, G., Jia, W., and Pettersson, S. (2012). Host-Gut Microbiota Metabolic Interactions. *Science* 336, 1262-1267.

Reber, S.O., Siebler, P.H., Donner, N.C., Morton, J.T., Smith, D.G., Kopelman, J.M., Lowe, K.R., Wheeler, K.J., Fox, J.H., Hassell, J.E., Greenwood, B.N., Jansch, C., Lechner, A., Schmidt, D., Uschold-Schmidt, N., Füchsl, A.M., Langgartner, D., Walker, F.R., Hale, M.W., Lopez Perez, G., Van Treuren, W., González, A., Halweg-Edwards, A.L.,

Fleshner, M., Raison, C.L., Rook, G.A., Peddada, S.D., Knight, R., and Lowry, C.A. (2016). Immunization with a heat-killed preparation of the environmental bacterium *Mycobacterium vaccae* promotes stress resilience in mice. *Proceedings of the National Academy of Sciences* 113, E3130–E3139.

Rook, G. (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, 18360-18367.

Sanna, S., Van Zuydam, N.R., Mahajan, A., Kurilshikov, A., Vich Vila, A., Võsa, U., Mujagic, Z., Masclee, A.a.M., Jonkers, D.M.a.E., Oosting, M., Joosten, L.a.B., Netea, M.G., Franke, L., Zhernakova, A., Fu, J., Wijmenga, C., and Mccarthy, M.I. (2019). Causal relationships among the gut microbiome, short-chain fatty acids and metabolic diseases. *Nature Genetics* 51, 600-605.

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., and Raes, J. (2019). The neuroactive potential of the human gut microbiota in quality of life and depression. *Nature Microbiology* 4, 623-632.

Van Elsas, J.D., Chiurazzi, M., Mallon, C.A., Elhottovā, D., Krištůfek, V., and Salles, J.F. (2012). Microbial diversity determines the invasion of soil by a bacterial pathogen. *Proceedings of the National Academy of Sciences* 109, 1159-1164.

Vital, M., Karch, A., and Pieper, D.H. (2017). Colonic Butyrate-Producing Communities in Humans: an Overview Using Omics Data. *mSystems* 2, e00130-00117.

Von Hertzen, L., Hanski, I., and Haahtela, T. (2011). Natural immunity: Biodiversity loss and inflammatory diseases are two global megatrends that might be related. *EMBO Reports* 12, 1089-1093.