MONASH University

The structure and basis of soil microbial biodiversity

Sean Keith Bay

BSc. (Hons. 1st class) University of Exeter

MSc. Kings College London

A thesis submitted for the degree of Doctor of Philosophy at Monash University in 2020

School of Biological Sciences

Copyright notice

‘I dedicate this thesis to my partner

Jodie F Lloyd,

an unwavering supporter,

and my granddad

Professor Wilhelm Scheuerpflug,

a father figure and academic inspiration’

Abstract

Soil bacteria are the most abundant and phylogenetically diverse cellular organisms. They underpin the health of the soil biosphere and mediate critical ecosystem services such as turnover of nutrients, carbon fixation and biogeochemical cycling of atmospheric gases. Most of these organisms lack cultured representatives and are not represented in existing genomic databases. This limitation has hindered the predictive power of microbial biodiversity studies and our understanding of their physiological functions and roles.

This thesis addresses three gaps in our understanding of soil microbial biodiversity: (i) ‘what drives microbial biogeographic patterns across environmental and distance gradients?’, (ii) ‘what processes support microbial primary production and energy generation across environmental gradients?’ and (iii) ‘how do these processes vary in abundance and activity across different ecosystem types?’. This was achieved primarily by combining ecogenomic approaches with biogeochemical measurements.

While studies have shown that soil microorganisms are biogeographically distributed, the patterns observed are typically much weaker than those observed for animals and plants. The first results chapter addresses whether this observation is a biological phenomenon or reflects methodological limitations. To do so, I performed soil microbial surveys across an aridity gradient in Israel and leveraged innovations in sampling design, sequence processing and diversity analysis. A key finding is that microbial communities exhibit stronger biogeographic patterns than previously reported. Moreover, it is shown that existing methods such as clustering and/or filter sequencing data underestimate spatial turnover of microorganisms. Concordant findings were made using local, regional and global datasets.

The second results chapter reports what energy sources support the microbial communities across this aridity gradient. Metagenomic and biogeochemical techniques were used to disentangle the relative importance of sunlight, organic compounds, and inorganic compounds as energy sources. These analyses demonstrated that biocrust and topsoil microbial communities harbour diverse metabolic capabilities. Whereas photosynthesis is a dominant primary production process in sub-humid and semi-arid soils, trace gases such as molecular hydrogen

are major energy sources for arid and hyper-arid soils. The most dominant taxa in these environments have the flexibility to use both organic carbon and trace gases to meet energy and carbon needs. Thus, multiple energy sources support desert microbial communities depending on resource availability.

Extending these findings, the third results chapter reveals the relationships between community composition and function across four different ecosystems. An in depth analysis of soil profiles across forest, wetland, grassland and dryland sites within Australia was conducted. Among these findings, I provide evidence from genome-resolved metagenomics that over 70% of the community have the potential to oxidise the trace gases molecular hydrogen and carbon monoxide. Various phyla were also shown to harbour the genetic capacity for atmospheric hydrogen, carbon monoxide, and methane oxidation for the first time. These findings are supported by *in situ* flux and *ex situ* activity measurements. In combination, this suggests that trace gas oxidisers are dominant, active, and widespread members of soil microbial communities.

This work presents a comprehensive advance in our understanding of soil microbial biodiversity and function. It highlights how current methodologies insufficiently capture microbial biogeographic distributions and advocate advances for future studies. Functional studies developed recent findings highlighting the importance of trace gas metabolism in energy conservation, while demonstrating the ecological significance of this processes across diverse ecosystems and along environmental gradients. The implications of these findings confirm that trace gases such as atmospheric H_2 and CO support the biodiversity of soil microorganisms and highlight how these overlooked energy inputs are important components of ecosystem services and function.

Publications during enrolment

Published

Islam ZF , Cordero PR, Feng J, Chen YJ, **Bay SK**, Jirapanjawat T, Gleadow RM, Carere CR, Stott MB, Chiri E, Greening C (2019). Two Chloroflexi classes independently evolved the ability to persist on atmospheric hydrogen and carbon monoxide. *The ISME Journal* 13 1801-1813.

Bay SK, Ferrari BC, Greening C* (2018). Life without water: how do bacteria generate biomass in desert ecosystems? *Microbiology Australia* 39, 28-32.

Ji M & Greening C, Carere CR, van Wonterghem I, **Bay S**, Steen J, Montgomery K, Lines T, Beardall J, Snape I, Stott MB, Hugenholtz P, Ferrari B* (2017). Atmospheric trace gases support primary production in Antarctic desert surface soil. *Nature* 552, 400-403.

Leung PM*, **Bay SK**, Meier DV, Chiri E, Cowan DA, Gillor O, Woebken D, Greening C* Energetic and trophic basis of microbial persistence in desert ecosystems. *mSystems* 5, e00495-19.

In review

Bay SK, Dong X, Bradley J, Leung PM, Jirapanjawat T, Arndt SA, Cook PLM, La Rowe D, Nauer P, Chiri E*, Greening C*. Atmospheric trace gas scavengers are dominant and active members of the soil biosphere. In revision: *Nature Microbiology*

Chen YJ, Leung PM, **Bay SK**, Hugenholtz P, Kessler AJ, Shelley G, Waite DW, Cook PLM, Greening C*. Metabolic flexibility allows generalist bacteria to become dominant in a frequently disturbed ecosystem. In revision: *Nature Communications*

Bay SK*, McGeoch MA, Gillor O, Wieler N, Palmer JP, Baker DJ, Chown SL, Greening C* Soil bacterial communities exhibit strong biogeographic patterns at fine taxonomic resolution. In revision: *mSystems*

Martinez C, Zhao Z, Lappan RJ, **Bay SK**, De Corte D, Hulbe C, Ohneiser C, Stevens C, Thompson B, Stepanauskas R, González JM, Logares R, Herndl GJ, Morales SE, Greening C, Baltar F* Dark carbon fixation sustains diverse microbial communities below the Ross Ice Shelf. In review: *Science*

In preparation

Bay SK, Waite DW, Gillor O, Hugenholtz P, Greening C* Reciprocal activities of chemosynthetic and photosynthetic bacteria across a steep desert aridity gradient. Target: *Nature Communications*

Thesis including published works declaration

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

This thesis includes two submitted publications and one in preparation. The core theme of the thesis is on the structure and basis of soil microbial biodiversity. The ideas, development and writing up of all the papers in the thesis were the principal responsibility of myself, the student, working within the School of Biological Sciences under the supervision of Associate Professor Chris Greening.

The inclusion of co-authors reflects the fact that the work came from active collaboration between researchers and acknowledges input into team-based research.

In the case of three results chapters my contribution to the work involved the following:

Thesis Chapter	Publication Title	Status	Nature and % of student contribution	Co-author name(s) Nature and % of Co-author's contribution*	Co-author(s), Monash student Y/N*
2	**Soil bacterial communities exhibit strong biogeographic patterns at fine taxonomic resolution**	In review	Conception, Design, Field sampling, DNA extraction, Sequence processing, Data analysis, Writing and editing (65%)	1. McGeoch MA, Design, Data analysis (5%) 2. Gillor O, Field sampling (2%) 3. Wieler N, Field sampling (1%) 4. Palmer JP, Data analysis (1%)	1-7 N

				5. Baker DJ, Data analysis (1%) 6. Chown SL, Conception, Design (5%) 7. Greening C, Conception, Design, Data analysis, Writing and editing (20%)	
3	**Reciprocal activities of chemosynthetic and photosynthetic bacteria across a steep desert aridity gradient**	In preparation	Design, Field sampling, DNA extraction, Sequence processing, Data analysis, Writing and editing (70%)	1. Waite DW, Metagenomic binning (6%) 2. Gillor O, Field sampling (2%) 3. Hugenholz P, Logistic support (2%) 4. Greening C, Conception, Design, Data analysis, Writing and editing (20%)	1-4 N
4	**Trace gas oxidizers are globally dominant and active members of soil biomes**	In review	Design, Field sampling, Lab work, Thermodynamic modelling, Data analysis, Writing and editing (55%)	1. Dong X, Binning (5%) 2. Bradley JA, Modelling (4%) 3. Leung PM, Field sampling (2%) 4. Jirapanjawat T, Lab work (2%)	1, 2, 4-10 N 3 Y

				5. Arndt SK, Logistic support (1%) 6. Cook PLM, Logistic support (1%) 7. La Rowe D, Modelling (1%) 8. Nauer PA, Thermodynamic modelling, Logistic support (2%) 9. Chiri E, Design, Data analysis, Field sampling, Thermodynamic modelling (7%) 10. Greening C, Conception, Design, Data analysis, Field sampling, Writing and editing (20%)	

I have renumbered sections of submitted or published papers in order to generate a consistent presentation within the thesis.

Student name: Sean Bay

Student signature: **Date: 23/06/2020**

I hereby certify that the above declaration correctly reflects the nature and extent of the student's and co-authors' contributions to this work. In instances where I am not

the responsible author I have consulted with the responsible author to agree on the respective contributions of the authors.

Main Supervisor name: Associate Professor Chris Greening

Main Supervisor signature: **Date: 23/6/2020**

Acknowledgements

I would like to thank my supervisor Associate Professor Chris Greening for introducing me to the world of microbiology, guiding my thinking and providing me with a warm, creative and collaborative research environment. From the very beginning Chris has been an exceptional mentor and supported me across all aspects of my PhD journey. He has provided a great deal of guidance, whilst giving me the intellectual freedom to develop my own ideas. I am incredibly grateful for the opportunities given to me, such as collaborating on diverse projects, travelling to the Negev desert and sharing my research at national and international meetings.

I also like to thank my co-supervisors Prof. Melodie McGeoch and Prof. Steven L Chown for their support. Discussions from personal meetings and attending weekly journal clubs of the McGeoch group have given me a true cross disciplinary perspective, helped me to structure my thinking, and develop ideas around my projects. I also like to thank Prof. Osnat Gillor for hosting me during my stay at the Jacob Blaustein Institute for Desert Research, Israel. Her logistic support, as well as expertise on the microbial ecology of deserts, has been of great help to my work. I also want to thank my PhD panel chair Prof. Paul Sunnucks, who has overseen all of my milestone meetings and provided kind guidance and feedback, as well as panel members A/Prof. Damian Dowling and A/Prof. Anne Peters.

For financial support I would like to thank Monash University for awarding me a Faculty of Science Dean's Postgraduate Research Scholarship (DPRS) and Faculty of Science Dean's International Postgraduate Research Scholarship (DIPRS). A Monash University and Ben Gurion University of the Negev Seed Fund (awarded to Chris Greening, Osnat Gillor and Steven L Chown) supported my travels to Israel and covered project costs of results chapters one and two. The Ecological Society of Australia for the Holsworth Wildlife Research Endowment (awarded to Sean K Bay) supported my final results chapter. I am also thankful for the Monash University Post Graduate Travel Award which has allowed me to present my work at the 17th International Symposium on Microbial Ecology in Leipzig, Germany.

For technical support I want to thank our lab manager Tent Jirapanjawat who has been an exceptional guide in the lab providing vital assistance including induction, training and logistical support. I collectively want to thank all members of the Greening lab past and present for contributing to a cohesive, creative and warm research environment, which has supported me throughout my PhD. I also want to thank specific members including Pok Man Leung, Dr Rachael Lappan, Ya-Jou Chen, Dr Eleonora Chiri and Guy Shelley for discussions and exchange of ideas on the themes of ecology, microbiology, bioinformatics and statistics. I am also grateful for the positive and helpful interactions with members from the McGeoch lab including Dr David Baker, Dr David Palmer, Sarah Reeve, Catherin Dickson and David Clarke. Finally, a special thanks to my partner Jodie F Lloyd who has been my unwavering supporter and kept me grounded throughout the ups and downs of my PhD.

Table of contents

List of figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Abbreviations

General

ASV	Amplicon Sequence Variant
ATP	Adenosine triphosphate
BLAST	Basic local alignment search tool
bp	Base pair
cm	Centimetre
Cp	Crossing point
ECEC	Effective Cation Exchange Capacity
ΔG	Gibbs free energy
dw	Dry weight
GC	Gas chromatography
GTDB	Genome Taxonomy Database
GPS	Global Positioning System
g^{-1}	Per gram
h	Hour
kb	Kilobase pairs
m	Metre
m^2	Metre squared
mm	Millimetre
MAG	Metagenome Assembled Genome
MAP	Mean Annual Precipitation
NCBI	National Center for Biotechnology Information
ORF	Operating Reading Frame
OTU	Operational Taxonomic Unit
PET	Potential Evapotranspiration
PCR	Polymerase chain reaction
ppbv	Parts per billion by volume
ppmv	Parts per million by volume

PSI	Photosystem I
PSII	Photosystem II
PVC	Polyvinyl chloride
qPCR	Quantitative real-time RT-PCR
RNA	Ribonucleic acid
rRNA	Ribosomal ribonucleic acid
rDNA	Ribosomal deoxyribonucleic acid
TUSC	Tropical Upland Soil Cluster
v/v	Volume per volume
W	Watt
yr^{-1}	Per year

Statistics & Bioinformatics

AIC	Akaike Information Criterion
ANOSIM	Analysis of similarities
ANOVA	Analysis of variance
CV	Coefficient of Variation
GLM	Generalized Linear Model
HMM	Hidden Markov Model
JJT	Jones-Taylor-Thornton model
LRT	Likelihood ratio test
MEGA 7	Molecular Evolutionary Genetics Analysis 7
MDS	Multidimensional scaling
MS-GDM	Multi-site generalized dissimilarity modelling
PCoA	Principal Coordinate Analysis
PERMANOVA	Permutational multivariate analysis of variance
PERMDISP	Permutational analysis of dispersion
RDA	Redundancy Analysis
RPKM	Reads Per Kilobase Per Million

Equations

Equation 1 $$P = \frac{r \cdot \Delta G_r}{B}$$

Equation 2 $$\Delta G_r = \Delta G_r^0 + RT \ln Q_r$$

Equation 3 $$Q_r = \prod a_g^{v_i}$$

Equation 4 $$a_g = \frac{f_g}{f_g^0}$$

Chapter 1

Introduction

1.1. Soil microbial communities

1.1.1. Soil structure

The pedosphere is Earth's outmost envelope where soils occur and soil forming processes are active [1,2] . Soil formation is driven by the underlying geological parent material, climate and organisms, as well as topographic features including elevation, slope and orientation. Along their profiles, soils comprise all three states of matter [1,3]. The solid phase comprises inorganic primary and weathered minerals, which account for the majority of soil dry weight, as well as organic matter that supports soil fertility [4–6]. The gaseous phase consists of atmospheric gases diffusing and interacting with the air spaces within and between soil aggregates, pore spaces and rock fractures [7]. The liquid phase consists of surface and pore water that regulates gaseous exchange and contains dissolved gases, minerals and organic matter. [8]. All three phases interact and provide the medium which supports all terrestrial life on earth and provide numerous services for humanity [1]. Soils and the functions that they provide are nevertheless highly vulnerable to anthropogenic pressures. These include the effects of local change, such as intense agricultural practices, and global change such as increasing frequency of drought events leading to desertification [9].

1.1.2. Community structure

Microorganisms are critical for the formation, structure, and function of soils. They play key roles in various ecosystem functions, which support the delivery of important ecosystem services [10,11]. Their functions control fluxes of energy, nutrients, and organic matter through the environment, while also supporting animal and plant life through symbiotic, trophic, and other interactions [12–15]. Soil microbial ecosystem services primarily fall within three types: (i) Supporting services such as primary production, nutrient cycling and plant symbiosis [16–18], (ii) Regulating services including disease control, biodegradation of pollutants and biogeochemical cycling [19–23] and (iii) Provisioning services which include products directly obtainable from ecosystems such as antibiotic production and genetic resources [12,13,24–28]. Links between microbial biodiversity and ecosystem functions are increasingly demonstrated, such as positive

relationships with ecosystem multi-functionality, terrestrial ecosystem productivity, nutrient cycling and litter decomposition [24,29–33].

Soil microorganisms include members from all three domains of life: bacteria, archaea and eukaryotes [34]. Global and regional surveys show that soil ecosystems are dominated by nine bacterial phyla namely the Proteobacteria, Actinobacteriota, Acidobacteriota, Planctomycetota, Chloroflexota, Verrucomicrobiota, Bacteroidota, Gemmatimonadota and Firmicutes [35–37]. Actinobacteriota are a particularly dominant group in soil ecosystems and are well known for their high stress tolerance and prevalence across many different habitats, including oligotrophic ecosystems such as deserts [38–41]. Archaea are generally less abundant, but can have critical roles; most notably, ammonia oxidising archaea such as *Nitrososphaera viennensis* belonging to Crenarchaeota play an important role in nitrification [42] and have been reported in many temperate soils [37,43]. The third domain of soil microorganisms are eukaryotes, which include fungi, protists, and microalgae. Fungi are particularly abundant soil microorganisms [44] with diverse roles as decomposers, mutualists, and pathogens. Each gram of soil can harbour up to 10^{10} bacterial cells, which range in richness between 10^{3}-10^{4} taxa [45]. However, not all taxa are equally distributed. Just like higher animal and plant communities, microbial communities show positively skewed occupancy frequency distributions. This means that in most ecosystems a low number of dominant taxa occupy the majority of sites and co-exist alongside a high number of taxa that have low occupancy, i.e. they are members of the community [46–48].

1.1.3. Methodologies

Pure culture studies of microorganisms are responsible for most of what we know about the structure, function, and diversity of soil microorganisms. However, the vast majority of soil bacteria are challenging to culture using growth media. Relying on cultured organisms alone thus limit our perspectives. Advances in culture-independent methods, including 16S rRNA gene amplicon sequencing and shotgun metagenomic sequencing, have transformed our capacity to profile the taxonomic and functional composition of soil microorganisms [49].

A mainstay of taxonomic studies has been sequencing of hyper-variable regions (V1-V9) along the 16S ribosomal rRNA gene sequence [50,51]. This gene encodes the RNA component of the 30S subunit of the bacterial and archaeal ribosome and provides a suitable phylogenetic marker, given it is universally distributed and relatively conserved over evolutionary time [52]. The 16S rRNA gene sequence can therefore be used to infer the taxonomic composition and evolutionary relationships of microbial communities. Following 16S rRNA gene sequencing on Illumina platforms, PCR-amplified 16S sequence variants contain low frequency errors (~0.1% per nucleotide), which obscure taxonomic assignments [53]. Traditionally this has been addressed by clustering sequences based on an arbitrary identify threshold (95-99%) into operational taxonomic units (OTUs), which serve as a proxy for species-level classification [54]. Following this step, OTUs are often further processed by removing those with low relative abundances, usually ≤ 0.05%. The main limitation of clustering sequences into OTUs is the trade-off between reducing error frequency at the cost of losing phylogenetic resolution. Recent advances in 16S sequence processing tools are able to overcome this limitation. Instead of clustering based on sequence similarity, each 16S sequence variant is profiled using sophisticated error-models to discriminate between many million reads that differ by as little as one nucleotide across the gene sequences [55,56].

In shotgun metagenomics, all community DNA in a sample is sequenced [57]. This can reveal the entire genetic repertoire of the community, allowing inference of the metabolic potential and other traits of community members [57]. Metagenomic sequences from short-read Illumina technologies consist of multiple short fragments (shotgun sequences) of the total community DNA. Overlapping regions of these sequence fragments can be computationally assembled into larger consensus regions of DNA called contigs. These contigs can then be grouped into phylogenetic bins and assembled into metagenome-assembled genomes (MAGs). Metagenomic analysis can be gene-centric by focusing on the short unassembled reads, which provides a means of estimating the relative abundance of specific genes of interest. Alternatively, a genome-centric perspective gives insights into the physiology and metabolic potential of specific organisms by interrogating MAGs. Since it was first used to describe the simple communities of biofilms [58] and seawater [59], millions of taxa and microbial genes have been described which have reshaped understanding of the

evolution and ecology of prokaryotes [34,60]. Metatranscriptomic derivatives of this approach using reverse transcribed RNA sequences are also frequently used to study gene expression of microbial communities [61].

The ongoing development of cultivation-independent technologies such as 16S rRNA gene amplicon, metagenomic and metatranscriptomic sequencing continues to transform our understanding of microbial communities but a more integrated methodological approach is needed to validate active metabolic functions [62]. While molecular sequence data provide a robust means of inferring functions likely to be involved in various pathways and processes, biochemical measurements are need to establish proof that they are active in the community. For example, radioactive carbon isotope studies can be used to trace the assimilation of inorganic carbon dioxide (CO_2) through biochemical reactions and confirm metabolic functions such as photosynthetic and chemosynthetic carbon fixation. ^{14}C studies were first developed to measure photosynthetic primary production in marine phytoplankton [63,64]. Since then, this approach has been widely used to highlight the importance of bacteria and archaea in global carbon cycling [65]. For example, this is exemplified by culture-based and culture-independent work demonstrating the capacity for hydrogenotrophic growth of Actinobacteriota [6667]. Likewise, gas chromatography studies can be used to confirm the activity of soil microbial mediated uptake of atmospheric gases such as molecular hydrogen (H_2), carbon monoxide (CO) and methane (CH_4). This method has been used to complement physiological and genomic studies confirming the upregulation, activity and kinetics of enzymes mediating the uptake of H_2 and CO [68–72]. Gas chromatography has also been used to validate metagenomic field studies by validating that the inferred metabolic potential to oxidise trace gases was active in soils [28]. Increasingly sensitive field measurements are also able to detect these biogeochemical processes *in situ* and can be used to confirm observations from *ex situ* laboratory studies. For example, static flux chamber methods have be widely used to measure local fluxes of the microbial soil sink of H_2, CO and CH_4 and estimate global trace gas budgets [73–75].

An integrated approach coupling molecular techniques with biochemical activity measurements provides a powerful basis of validating sequence data and confirming the functional basis of microbial biodiversity. Cultivation-dependent approaches also

remain extremely valuable and have recently been used to gain a deeper understanding of microbial lineages first identified through cultivation-dependent approaches [76–78]. In the following section **1.2.** I introduce the ecological processes that shape microbial community assembly and the conceptual framework I utilise to refer to these drivers throughout this thesis.

1.2. Community structure

1.2.1. Biogeographic patterns

A central aim in microbial ecology is to understand the ecological processes that determine community assembly. Biogeography measures the spatiotemporal variation of communities across geographic space and geological time and can reflect underlying community assembly processes [79]. The first law of geography states that "everything is related to everything else, but near things are more closely related than distant things" [80]. The decline in similarity between communities or taxa, as the distance between them increases, is known as the distance decay relationship and is a fundamental biogeographic pattern observed across all domains of life [81]. Distance decay patterns can reveal differences in community composition but also highlights how these differences are autocorrelated with spatial distance. Another universal biogeographic pattern is the taxon-area relationship, which describes the area of a habitat and the number of species it harbours, with larger areas generally containing a greater species richness [82,83].

It is now well established that soil bacteria and archaea display non-random biogeography at local, regional and global scales [10,81,84–89]. This realisation has led to the questions of what ecological processes underpin microbial community assembly processes. Historically, niche theory has been used as the dominant framework to describe how communities of macroorganisms and microorganisms assemble [79,81,90–92]. An ecological niche describes the spatiotemporal position a population or species can occupy under a certain set or resources and conditions [93,94]. Niche theory predicts that distinct species can co-exist because of their functional trait differences. In turn these differences govern specialization for different fundamental and realized niches that allow species to co-exist [94]. Niche theory describes the interactions between species and their environment (e.g. temperature, pH, organic carbon), and other species (e.g. predation, competition, and symbiosis) in relation to species' traits (e.g. metabolism, morphology and life history characteristics) [95].

Ecologists have also recognised that some changes in community structure are independent of species traits and inherently stochastic. Hubbel's neutral theory proposed that trophically similar species are ecologically equivalent in terms of birth, death, speciation and extinction rates, as well as dispersal limitation and colonization [96]. Thus, ecological processes which structure communities are predicted to be indistinguishable from those that arise by chance such as dispersal, speciation, extinction and drift [96]. Despite its fundamental challenge to the importance of niche-based processes, neutral theory has been widely used to describe some fundamental ecological processes structuring communities [86,97]. It is now recognised that the deterministic (non-random) processes described by niche theory and the stochastic (random) processes described by neutral theory are not mutually exclusive, but are intertwined and co-responsible for microbial community assembly processes [81,86,98–103].

More recently, these inherently different theories have been unified by Vellend into a framework of four high order ecological processes which structure communities: selection, diversification, drift and dispersal **(Fig. 1a)** [95,104]. Here I adopt Vellend's framework when referring to the deterministic and stochastic processes that structure microbial communities and briefly describe each process within its ecological context and its effect on biogeographic patterns such as distance decay **(Fig. 1a-b)**.

1.2.2. Selection

Selection generally occurs when individuals in a population vary due to fitness differences [95]. In ecological selection, differential growth or survival of microorganisms due to deterministic abiotic (e.g. pH, salinity, temperature, water content) and biotic (e.g. competition, predation, commensalism, mutualism) factors structures communities [95]. The strength of ecological selection is also predicted to co-vary with environmental heterogeneity (both abiotic and biotic) [95,105]. The strength of selection can be differentiated into 'homogeneous' and 'variable' selection [105]. Homogeneous selection generally occurs under conditions where there is little spatiotemporal environmental heterogeneity, therefore the selective pressure is constant and compositional turnover is predicted to be low [105,81,100]. Under this scenario, distance

decay relationships of soil microbial communities are expected to be weak **(Fig. 1b)**. In contrast, variable selection occurs under conditions where high environmental heterogeneity causes variation in selective pressures among taxa with fitness differences [81,95,100,105]. This is predicted to cause high rates of compositional turnover and increase the rate of distance decay **(Fig. 1b)**. The majority of studies to date have found that environmental selection is the dominant force structuring soil microbial communities [81,86]. This is unsurprising given that soils are highly heterogeneous in their physicochemistry with gradients in pH, organic carbon content, soil redox status, moisture availability, salinity and temperature. These gradients have been shown to structure microbial communities at the microscale extending to large distances across local, regional and global scales [106–111].

1.2.3. Drift

Ecological drift causes random fluctuations in species composition and abundances, because rates of birth, death and reproduction are inherently stochastic [95,96]. Under neutral theory and the absence of deterministic selection, demographically equivalent species are predicted to drift to extinction, with the exception of one dominant species [95]. Assuming that deterministic and stochastic processes interact and the environmental effect is strong, the effects of drift are generally expected to be outweighed by selection [86,95]. However, empirical evidence suggests that when microbial communities are under weak selection, such as during the early stages of colonisation, ecological drift could be an important determinant of community structure [81,95,100,112–114]. Similarly, rare taxa which generally make up the majority of the community are also predicted to be particularly vulnerable to stochastic changes in demographics [104], as well as stochastic changes in environmental conditions [115]. Like selection, drift is expected to structure microbial communities across spatial distance **(Fig. 1b)**.

1.2.4. Dispersal

Dispersal generally refers to the movement and successful establishment of taxa across space and time [81,116]. It is a means of connecting local, regional and global

communities and has the capacity to alter community diversity and composition [95]. Dispersal can be active, whereby soil bacteria propel themselves, or passive such as through aeolian or hydration controlled processes [117–119]. While active dispersal is thought to make a greater contribution to microbial community composition in aquatic ecosystems, this process is predicted to play a minor role in terrestrial ecosystems. However, both active and passive modes can result in dispersal limitation which can influence microbial community structure [95]. Dispersal limitation generally occurs when the movement or successful establishment of taxa is impeded [86]. These limiting effects can be due to physical constraints (e.g. spatial configuration of habitats, weak aeolian and aquatic processes), abiotic filters (e.g. pH, salinity, and moisture), biotic effects (e.g. cell morphology, metabolic strategy) as well as priority effects, whereby the first colonizers to a new habitat affect the successful establishment of taxa arriving at a later stage [100,120]. While passive aeolian dispersal might be seen as a stochastic process, differences in fitness traits might favour certain taxa to withstand selective pressures in the atmosphere such as lack of nutrients and radiation [118,121]. Thus dispersal encompasses both stochastic and deterministic components [122]. Dispersal interacts with selection and drift by weakening their effect [81]. As microbial dispersal increases, communities increasingly reflect the composition of the new colonizers rather than those shaped by selection and drift over time. Thus, in habitats undergoing high rates of dispersal, compositional differences are predicted to decline, slowing turnover and leading to a shallowing in the distance decay relationship **(Fig. 1b)**. Given the inherently different characteristics of prokaryotes when compared to higher eukaryotes such as size, generation time and abundance, processes such as dispersal are poorly understood and difficult to quantify.

1.2.5. Diversification

Despite the prominent role of bacteria in evolutionary history, relatively little is known about the dynamics by which their diversity arose in the first place. Diversification refers to the evolutionary processes that generate new genetic variation through mutations and balance the rates of speciation and extinction [86]. Diversification leaves phylogenomic footprints which can be used to infer past extinction and diversification rates [89,123]. However, estimates of speciation and extinction rates across evolutionary

time remain understudied in microbial organisms [86]. Studies examining the effects of diversification on community composition are also hampered by the fact many lineages are short-lived relative to geological timescales, which makes it difficult to estimate rates of past extinction and diversification events. Recent work suggests that, over the past billion years, bacterial diversity has been continuously increasing with a small fraction of bacteria that ever existed present in contemporary populations [124]. However, it remains difficult to estimate the effects of diversification on contemporary soil microbial community structure. Unlike selection and drift, which increase the rate of distance decay and dispersal which counteract it, diversification is theoretically predicted to modify its magnitude **(Fig. 1b)**. This is because mutations are predicted to affect all taxa and thus do not autocorrelate with distance.

Following the discussion about high order ecological drivers underpinning soil microbial community assembly, I will move towards introducing the functional capabilities of microorganisms, focusing on their modes of energy conservation and carbon acquisition strategies in section **1.3.**

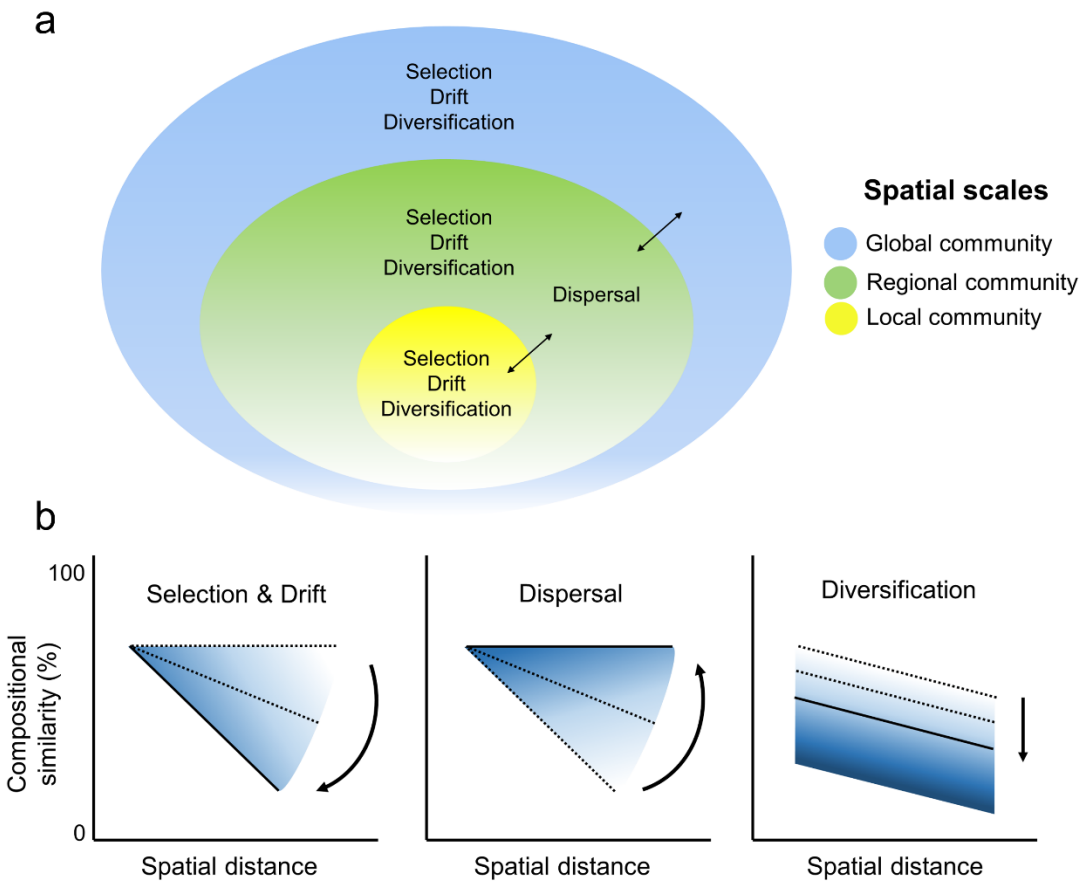


Figure 1. Adapted schematic showing the conceptual interaction between the four ecological processes structuring microbial communities (a) [95] and the biogeographic pattern of distance decay (b) [81]. These four high order ecological processes are not mutually exclusive but are thought to interact simultaneously in shaping microbial community structure.

1.3. Metabolic diversity

1.3.1. Microbial metabolism

Through metabolism, microorganisms convert chemical or light energy from the environment into usable cellular forms. Cells allocate this energy into growth, reproduction, and cellular maintenance processes, such as macromolecular repair, metabolite transport and membrane potential generation [125]. The major source of chemical energy within cells is the energy storage molecule ATP [126]. In most organisms the vast majority of ATP is generated by enzyme complexes bound to energy-converting membranes [127]. Energy-converting membranes are found across all domains of life and include the plasma membrane of bacteria and archaea, the inner membrane of mitochondria and the thylakoid of chloroplasts [127]. In respiratory organisms, electrons are generally transferred from reducing equivalents along an electron transport chain to acceptors such as oxygen. This energy transfer is coupled to proton translocation, leading to the generation of an electrochemical gradient (proton-motive force) [126]. This gradient is then used to power ATP synthase to synthesize ADP to ATP through a chemiosmotic mechanism [126]. Most animals are relatively metabolically inflexible, because the energy-transducing membranes of their mitochondria require electrons derived from organic substrates to be transferred to the electron acceptor O_2 [127]. However, bacteria and archaea are highly flexible in their metabolic strategies; they are able to use a variety of organic and inorganic electron donors, which can be transferred to various aerobic and anaerobic electron acceptors.

Reflecting this flexibility, microorganisms adopt a range of trophic strategies in nature. Organisms differ in how they acquire carbon for synthesis of biomass (autotrophy, heterotrophy), reducing equivalents for energy conservation or biosynthesis (lithotrophy, organotrophy) and energy for growth (chemotrophy, phototrophy) [126]. To determine the metabolic strategies of various microbial taxa, these trophic descriptors can be used combinatorially. For example, autotrophic organisms have the capacity to fix CO_2 but are differentiated in how they obtain energy from either light (photoautotrophy) or from inorganic compounds (chemoautotrophy). The same principle applies to heterotrophic organisms which rely on organic compounds.

Photoheterotrophs use light as energy sources, but require organic carbon rather than CO_2 as their carbon source, whereas chemoheterotrophs use chemical energy sources and organic carbon sources. A third distinction is whether reducing equivalents are obtained from either organic (organotrophy) or inorganic (lithotrophy) compounds.

1.3.2. Photosynthetic strategies

Phototrophs carry out one of the most fundamental biological processes, the conversion of light to chemical energy [128]. Photosynthesis relies on light-harvesting pigments such as chlorophylls, which absorb light at various wavelengths. The derived energy is transferred to photosynthetic reaction centres that produce a proton-motive force to energise ATP production [129]. Common photoautotrophs in the eukaryotic domain include green plants, mosses, lichens, and diverse algae. In addition, bacteria have evolved diverse modes of photosynthesis [26,130]. Cyanobacteria are the only phylum which carry out photosynthesis using an internal thylakoid membrane system in a similar manner to green plants [131,132]. All chloroplast organelles found in contemporary photoautotrophs are thought to originate from an ancient endosymbiosis event whereby a cyanobacterium was engulfed by a eukaryote [133]. In this mode of photoautotrophy, both ATP production and CO_2 reduction into organic compounds is driven using light as energy source and water (H_2O) as the reducing agent. This mode is known as oxygenic photosynthesis as the by-product of these reactions is oxygen (O_2). In soil microbial communities, Cyanobacteria play important roles in supporting primary production through photosynthetic carbon fixation and forming biological soil crusts through extracellular secretions [134–138].

Other photosynthetic bacteria obtain reducing power from inorganic sources such as sulfide, iron, or hydrogen to carry out this process. They lack the ability to use water as an electron donor and, because no O_2 is produced, this mode is known as anoxygenic photosynthesis [26]. These organisms are further differentiated from Cyanobacteria in that they use bacteriochlorophylls, which capture a longer wavelength than the plant-like chlorophyll *a*. Anoxygenic phototrophs have been found in various aquatic ecosystems but are generally minor members in terrestrial soils [132].

Unlike oxygenic photosynthesis which is only found in Cyanobacteria, anoxygenic photosynthesis has been found in various phyla, namely some Acidobacteriota, Chloroflexota, Bacteroidota, Gemmatimonadota, Firmicutes, and Proteobacteria [139,140]. It is likely that oxygenic photosynthesis arose from anoxygenic precursors that supported the Archaean ecosystems [26,132]. Cyanobacteria are predicted to have played a key role in the great oxidation event ~2.4 billion years ago, which changed earth atmosphere from a reduced to an oxygenated environment.

Both oxygenic and anoxygenic photosynthesis rely on photochemical reaction centres to transduce light energy into ATP synthesis. However, other photoheterotrophic bacteria employ a minimalistic form of light energy capture by using retinal-binding proteins called rhodopsins [26]. These can be differentiated into multiple groups, including energy-conserving transmembrane proton pumps, transmembrane chloride pumps, and light sensors [141–143]. Bacteriorhodopsins and proteorhodopsins are light-driven proton pumps which generate an ion-motive force to drive ATP synthesis. The ecological significance and role of these processes is well-established in marine ecosystems, but their roles in soils is poorly understood. However, metagenomic analysis of marine communities at the global scale suggest that the genes for rhodopsin-based light harvesting are widespread [144]. There is also evidence that these genes are abundant in the oligotrophic dry valleys of Antarctica, where they may enable energy generation during long-term survival [145].

1.3.3. Chemosynthetic strategies

In the absence of light harvesting mechanisms, some bacteria and archaea are able to obtain reducing equivalents for energy conservation and growth by using inorganic electron donors from a variety anthropogenic, geological and biological sources [146]. These include trace gases such as hydrogen (H_2) and carbon monoxide (CO), as well reduced sulfur compounds (e.g. sulfide, thiosulfate), nitrogen compounds (e.g. ammonium, nitrite), and metals (e.g. iron(II)) [147,148]. Obligate lithotrophs, for example most nitrifying bacteria and archaea, have a highly specialized metabolism for a particular substrate and entirely depend on autotrophic carbon fixation for growth [146,148]. In contrast, facultative lithotrophs are typically preferential organoheterotrophs

that can grow mixotrophically, for example by using electrons derived from H_2 and CO to drive aerobic respiration and sometimes carbon fixation [149]. Despite their broad phylogenetic and ecological diversity, many lithotrophs have the metabolic capacity to chemosynthetically assimilate CO_2. In aerobic ecosystems, most lithotrophs fix CO_2 using type I ribulose 1,5-bisphosphate carboxylase/oxygenase (RubisCO) *via* the Calvin-Benson-Bassham pathway [146,149–151].

In soil ecosystems, obligate and facultative lithotrophs play an important role in driving key biogeochemical processes such as nitrogen, carbon and trace gas cycling. Obligate lithotrophs have primarily studied for their role in mediating nitrification [152,153] such as the gammaproteobacterium *Nitrosomonas europea* [154–156] and crenarchaeote *Nitrososphaera viennensis* [42,43], which derive their energy for growth from ammonium oxidation. Other model examples include the gammaproteobacterium *Acidithiobacillus ferrooxidans*, which derives its energy from the oxidation of iron, sulfur, or hydrogen in acidic soils [157–159]. However, much less is known about the role of facultative lithotrophs, for example those mediating H_2 and CO oxidation [149,150,160]. Recent culture-based and culture-independent studies have identified bacteria within several dominant soil phyla that are surprisingly metabolically flexible. They are capable of switching from heterotrophic growth to lithotrophic H_2 and CO scavenging to obtain reducing equivalents for energy conservation, biosynthesis and in some cases growth [67,70,72,161–163]. In the following section **1.4**., I will introduce trace gas metabolism, the enzymes mediating this process, and their role in energy conservation and carbon acquisition.

1.4. Atmospheric trace gases

1.4.1. Trace gas metabolism

Atmospheric trace gases are globally ubiquitous and permeate the aerated layer of most surface soils. In the lower troposphere, hydrogen (H_2), carbon monoxide (CO) and methane (CH_4) occur in trace amounts at average global concentrations of 0.53 and 0.10 and 1.85 parts per million (ppm) respectively **(Fig. 2)** [164–166]. The turnover of these gases is driven by biological, geochemical and anthropogenic processes. Soil

microorganisms play a major role in the consumption, modification and production of atmospheric trace gases [20,167]. However, with the exception of CH_4, anthropogenic and biogeochemical sources of these gases are counterbalanced by sinks of similar magnitude [165,166,168]. Soil microorganisms are a major sink of all three gases **(Fig. 2)**, accounting for approximately 76% of annual losses for H_2, 10% for CO and 5% for CH_4 [20,165,168]. The remaining losses of trace gases are due to photochemical oxidising reactions [169]. Overall, we have a relative strong understanding of the processes and organisms responsible for CH_4 cycling. In contrast, while it has long been recognised that soils mediate H_2 and CO oxidation [170,171], the microorganisms and enzymatic processes responsible for the uptake of H_2 and CO are a relatively recent discovery.

The microbial oxidation of H_2 is catalysed by hydrogenase enzymes. Based on the metallic core of their active site the enzyme can be classified into three distinct groups, namely the [NiFe], [FeFe] and [Fe]-hydrogenases [160,172]. These metalloenzymes catalyse the interconversion of H_2 to protons and electrons in a reversible reaction $H_2 \rightleftharpoons 2H^+ + 2e^-$ [160]. The direction of this reaction is dependent on the reduction potential of the component interacting with the enzyme. Thus hydrogenases can oxidise H_2 and input the derived electrons into aerobic or anaerobic respiratory chains when respiratory electron acceptors are available. Alternatively, evolve H_2 through the reduction of protons to disperse excess reductants generated during fermentation [173].

Historically, H_2 metabolism was thought to primarily occur in anaerobic hydrogen-rich environments such as marine sediments, gastrointestinal tracts, and hydrothermal systems [160,167,174]. However, since the isolation of the *Streptomyces sp. PCB7*, the first organisms demonstrating high affinity uptake of H_2 [175], evidence is growing that H_2 is a universal energy source even in aerated ecosystems. Recent studies have demonstrated that soil bacteria can aerobically respire hydrogen at atmospheric concentrations to provide electrons for respiration and carbon fixation. H_2 is a highly desirable electron donor for soil microorganisms given its low redox potential, low activation energy, and high diffusivity across cellular membranes [19,176]. The underlying kinetics by which soils oxidise H_2 was discovered using gas chromatography and tritium exchange assays demonstrating the biphasic kinetics of H_2 in the upper soil layer via fast acting low affinity (K_m > 1000 nM ppmv) and slow acting high affinity (K_m < 10-70 nM ppmv) activity [177,178].

To harness energy from this trace gas, bacteria express high-affinity, oxygen-tolerant [NiFe]-hydrogenases that associate with the cell membrane. These enzymes comprise a large subunit containing the bi-metallic nickel-iron catalytic site and a small subunit made of three iron-sulphur clusters [179] Four phylogenetically divergent groups of these enzymes have been shown to oxidise atmospheric H_2, namely the group 1h, 1l, 1f, and 2a [NiFe]-hydrogenases [180–182]. These enzymes input H_2-derived electrons into the aerobic respiratory chain and are biochemically adapted to function in the presence of O_2. Studies on Actinobacteriota [163], Acidobacteriota [70] and Chloroflexota [72] pure cultures have shown that atmospheric H_2 oxidation is induced under carbon starvation; these organisms switch from organotrophic growth to mixotrophic persistence, by upregulating a group 1h [NiFe]-hydrogenases (HhyLS) that support maintenance needs [182,183]. Other aerobic bacteria, for example *Ralstonia eutropha*, can grow lithoautotrophically using H_2 and CO_2 as the sole energy and carbon source; however, this metabolism is restricted to H_2-enriched environments and depends on a distinct hydrogenase (group 1d [NiFe]-hydrogenases) [184–186].

A range of aerobic bacteria and archaea are also able to use CO as an energy source for growth and persistence [187,188]. These organisms possess a carbon monoxide dehydrogenases (CODH) that catalyse the reaction $CO + H_2O \rightarrow CO_2 + 2\ H^+ + 2e^-$ [189]. In aerobic CO oxidising bacteria this uptake is mediated by type I [MoCu]-CO-dehydrogenases, which are trimeric enzymes (CoxLSM) of the molybdenum-containing hydroxylase superfamily [190,191]. Two groups of CO oxidising bacteria have been identified based on their physiology. Carboxydotrophs, such as *Oligotropha carboxydivorans*, can grow chemolithoautotrophically using CO as the sole carbon and energy source [192–194]. In contrast, carboxydovores adopt a mixotrophic metabolic strategy; they rely on organic substrates as carbon sources, but can oxidise CO at atmospheric concentrations to enhance growth or survival [187]. Two recent pure culture studies have shown that atmospheric CO oxidation supported bacterial survival under carbon starvation in two dominant soil phyla Actinobacteria [161] and Chloroflexi [72]. There is now evidence that at least four soil phyla are able to aerobically scavenge carbon monoxide: Actinobacteriota [71,195], Chloroflexota [72,196], Proteobacteria [191,197] and Euryarchaeota [198].

The potent greenhouse gas CH_4 can be consumed by aerobic methanotrophic bacteria and anaerobic methanotrophic archaea. These use CH_4 as both an energy source and, with exception of the chemoorganoautotrophic verrucomicrobial methanotrophs, carbon source. In aerated soil ecosystems, methanotrophic bacteria mainly affiliate with various genera within the Alphaproteobacteria and Gammaproteobacteria [199–202, 203]. Two other phyla are also known to mediate CH_4 oxidation in specific soils, namely Verrucomicrobiota in acidic soils and *Candidatus* Methylomirabilota (NC10) in anoxic soils [204]. The key enzyme mediating the uptake of CH_4 is a methane monooxygenase which catalyses the initial conversion step of methane to methanol. Two distinct form of this enzyme have been identified, a soluble cytosolic methane monooxygenase which uses a diiron active site (sMMO) and a particulate membrane-associated form with a catalytic copper centre (pMMO) [205]. While particulate methane monooxygenases are present among the majority of aerobic methanotrophs [204], soluble methane monooxygenases are more variable in their distribution among methanotrophs [206]. While methanotrophs are of global significance given their role in the CH_4 cycle, they generally have a low abundance and phylogenetic diversity in many soil ecosystems [166,199]. Other organic trace gases are also known to be used as energy sources by microorganisms, for example isoprene and short-chain alkanes [207,208].

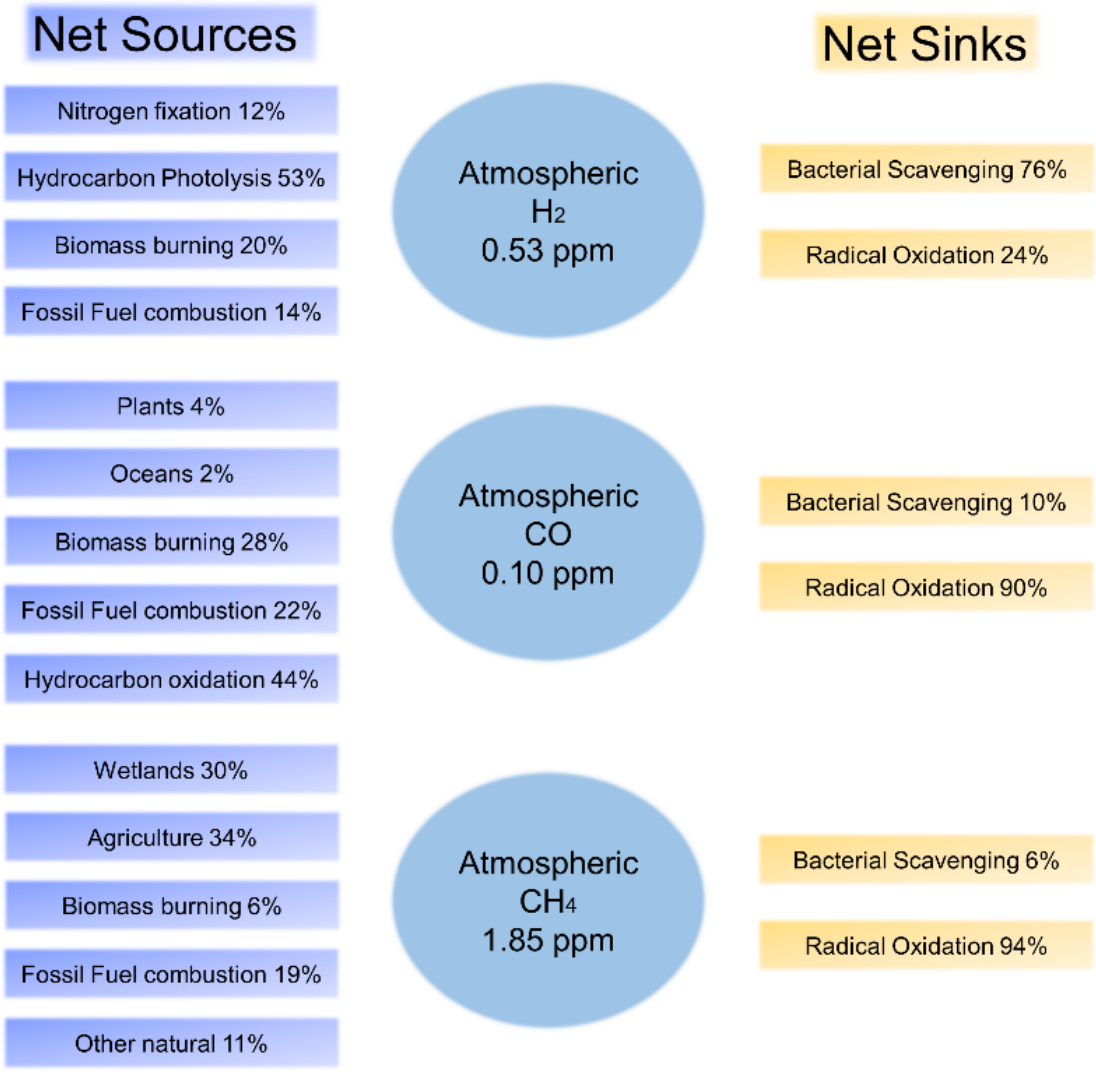


Figure 2. Schematic showing the main anthropogenic and natural sources and sinks of the three atmospheric trace gases H_2, CO and CH_4. Global average tropospheric mixing ratios are shown in parts per million (ppm) [164–166].

1.4.2. Physiological significance

It is well established that the majority of bacteria and archaea persist in various states of dormancy [209,210]. Only a small fraction (1-10%) are predicted to be actively growing in most soils [209]. Dormancy is a reversible survival strategy used by bacteria to avoid starvation and is thought to be induced under sustained energy limitation. To ensure survival, microorganisms rely on maintenance energy which is the sum of energetic costs of activities that do not involve reproduction [209]. This energy can then be used to carry out vital cellular functions such as biomolecular repair. Historically, dormancy

has been associated with ‘inactivity’, but recent studies suggest a more prominent role of microorganisms in this state in contributing to important ecosystem services such as biogeochemical cycling, antibiotic production, and biodiversity maintenance [62,163,211,212].

Metabolic flexibility enables dormant bacteria to survive in environments in which resources are limited. Metabolically flexible organoheterotrophs can switch from growth on organic carbon to persistence on atmospheric trace gases such as H_2 to enhance their chances of survival. For example, under carbon starvation, the model strain *Mycobacterium smegmatis* upregulates the expression of two high-affinity hydrogenases and persists by oxidising H_2 below atmospheric levels. Mutant strains, lacking the genes that encode the structural subunits of the group 1h [NiFe]-hydrogenase, have a 40% reduction in survival in carbon-limited batch and continuous cultures [69,163]. Similar observations have been made in Acidobacteriota and Chloroflexota cultures. Following the transition from exponential to stationary phase due to carbon starvation, *Pyrinomonas methylaliphatogenes*, *Thermomicrobium roseum* and the sporulator *Thermogemmatispora* T81 all upregulate the expression of group 1h [NiFe]-hydrogenases and mediate atmospheric H_2 oxidation [70,72]. Similarly to the upregulation of the [NiFe]-hydrogenase, the aerobic respiration of carbon monoxide is induced during carbon starvation. This has been demonstrated in two recent pure culture studies which have shown that atmospheric CO oxidation supported bacterial survival under carbon starvation in two dominant soil phyla Actinobacteriota [71] and Chloroflexota [72]. It is also known that this atmospheric CO can be oxidised by Chloroflexota [72], Proteobacteria [191,197] and Euryarchaeota [198] in a possible persistence-linked process.

1.4.3. Ecological significance

Evidence from physiological, biochemical and genomic studies has increasingly shown that bacteria can survive carbon starvation by trace gas scavenging [70–72,163,182]. Genomic and metagenomic studies highlight that a considerable phylogenetic diversity of bacterial taxa can carry out this function [161,182]. Given the ubiquity of atmospheric trace gases and their suitability for microbial uptake, trace gas

scavenging is a form of resource generalism which provides a competitive advantage in resource-limited ecosystems. Besides ensuring the survival of carbon starvation in energy limited environments, dormancy is also predicted to counteract stochastic processes such as drift [213]. This is because dormant microorganisms are not affected by stochastic changes in birth, death or reproduction rates. Thus, in line with the principles of seed banks observed in plant communities, dormancy is predicted to maintain or increase the genetic diversity in a population [214]. Entering dormant states under these conditions is therefore predicted to be advantageous [213]. While some have considered the cues that determine microbial dormancy stochastic [215], it is generally thought that microorganisms sense environmental conditions such as carbon availability and enter dormant states when conditions are unfavourable for growth [216].

Trace gas metabolism may play a particularly important role in oligotrophic ecosystems where organic carbon substrates are limited. Deserts represent one of the largest biomes and can be found from the tropics to the poles [217]. These are extreme environments with steep aridity gradients, limited organic substrates, and elevated pH, salinity and radiation [9,41]. Despite these pressures, 16S rRNA gene soil surveys have repeatedly found that both hot and cold desert ecosystems harbour diverse microbial communities [9,41]. At the phylum level, these communities are often similar in composition to temperate soils, with a dominance of Actinobacteriota, Proteobacteria and Chloroflexota [35,36]. The majority of these taxa are predicted to be preferential heterotrophs which rely on organic substrates for growth. However, due to the abiotic extremes of arid and hyper arid deserts, photosynthetic organisms are primarily restricted to lithic niches and biological soil crusts [138,218–222]. Photoautotrophs such as cyanobacteria and some algae occupy these environmental refugia and in presence of moisture can provide sufficient carbon inputs to maintain diverse heterotrophic communities [223,224]. However, these phototrophic communities are increasingly sparse and spatially fragmented as aridity increases [218,222,225,226]. They are therefore unlikely to account for the diverse microbial communities found in the interior of many arid and hyper-arid deserts. A longstanding enigma has been how these communities sustain themselves and maintain their diversity under these extreme energy-limiting conditions. Through recent studies, we have provided evidence that some desert surface soil communities are structured by a minimalistic mode of primary production, where atmospheric gases, not sunlight, serve as the main energy source [67].

We analysed the surface soil microbial communities in two coastal ice-free desert sites in Antarctica [67]. Soils had limited capacity for photosynthesis and were extremely low in organic carbon content. Despite this, they harboured diverse communities of bacteria belonging to the superphylum Terrabacteria, including Actinobacteriota, Chloroflexota, *Candidatus* Eremiobacterota and *Candidatus* Dormibacterota. Metagenomic analysis revealed that genes supporting energy conservation were widespread, with the majority of the bacteria encoding high-affinity lineages of the enzymes [NiFe]-hydrogenase and a carbon monoxide dehydrogenase [67]. The capacity to oxidise these trace gases to below atmospheric concentrations was confirmed using gas chromatography measurements. Furthermore, we detected that Actinobacteriota, *Candidatus* Eremiobacterota, and *Candidatus* Dormibacterota clades encoded the genes for autotrophic CO_2 fixation via the Calvin Benson-Bassham (CBB) cycle [67]. They encoded and expressed the type IE RuBisCO enzyme, a recently discovered clade of the CO_2-fixing enzyme that supports hydrogenotrophic growth in some Actinobacteriota but is absent from phototrophs [227,228]. The co-occurrence of these genes with high-affinity hydrogenases and carbon monoxide dehydrogenases suggested that these communities were able to fix CO_2 into biomass using atmospheric trace gases, rather than solely relying on exogenous inputs from photosynthetic organisms. This was validated by tracing assimilation of ^{14}C-labelled CO_2 using microcosm experiments. We were able to demonstrate that, under H_2-enriched conditions, chemosynthetic CO_2 fixation increased up to tenfold. In contrast, no significant stimulation was observed following light illumination [67].

Based on these findings, we propose that, in desert ecosystems where photosynthetic organisms are often excluded due to aridity, dormant bacterial communities are sustained by atmospheric chemosynthesis. Community members may maintain energy and carbon needs by aerobically respiring atmospheric H_2 and CO and, in some cases, using these gases to fix CO_2 into biomass. We hypothesise that these metabolic strategies are differentiated along aridity gradients traversing humid to hyper-arid climates, with this switch being driven by the availability of organic carbon in the environment **(Fig. 3)**. Given the prevalence of dormant soil microorganisms in temperate soils, we also hypothesise that a significant proportion of community members may be capable of atmospheric H_2 and CO oxidation even in more organic soils.

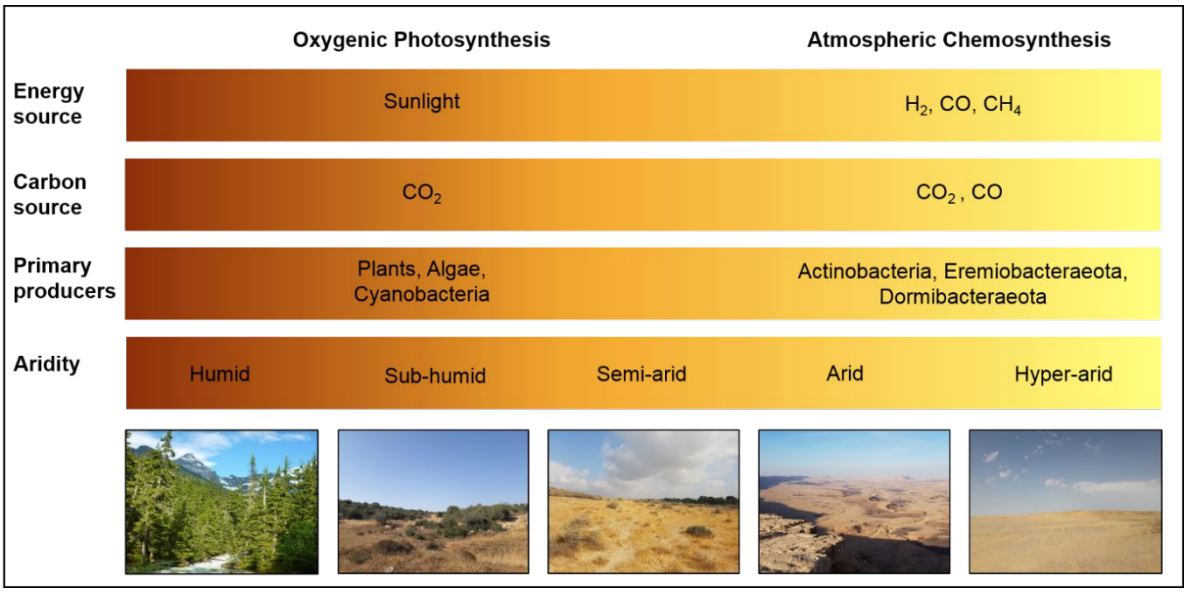


Figure 3. Schematic showing the predicted interactions between photosynthetic and chemosynthetic primary production strategies along an aridity gradient. As aridity increases, photosynthetic primary producers become less abundant relative to specialised bacteria that use atmospheric trace gases to generate biomass [229].

1.5. Thesis aims

Following our findings that atmospheric trace gases support primary production in Antarctica, the central aim of this thesis is to determine how soil microbial communities are spatially structured and what functional traits form the basis of soil microbial biodiversity. To address this aim, I use an integrated approach utilizing culture independent molecular technologies with biochemical measurements and statistical models. Three results chapters will address the following questions:

1. Determine what environmental and spatial drivers structure the biogeography of soil microbial communities.
2. Identify which metabolic processes support microbial primary production and energy conservation with increasing aridity.
3. Elucidate how primary production and energy conservation strategies vary in abundance and activity across different ecosystems.

Chapter 2

Soil bacterial communities exhibit strong biogeographic patterns at fine taxonomic resolution

Sean K. Bay[1,2 *], Melodie A. McGeoch[1], Osnat Gillor[3], Nimrod Wieler[3], David J. Palmer[1], David J. Baker[1], Steven L. Chown[1], Chris Greening[1,2 *]

[1] School of Biological Sciences, Monash University, Clayton, VIC 3800, Australia
[2] Department of Microbiology, Biomedicine Discovery Institute, Clayton, VIC 3800, Australia
[3] Department of Environmental Hydrology and Microbiology, Ben Gurion University of the Negev, Sde Boker, Israel

***** Correspondence can be addressed to:
Associate Professor Chris Greening (chris.greening@monash.edu), Department of Microbiology, Monash University, Clayton, VIC 3800, Australia
Sean Bay (sean.bay@monash.edu), Department of Microbiology, Monash University, Clayton, VIC 3800, Australia

In revision: *mSystems*

2.1. Abstract

Bacteria have been inferred to exhibit relatively weak biogeographic patterns. To what extent such findings reflect true biological phenomena or methodological artefacts remains unclear. Here, we addressed this question by analysing the turnover of soil bacterial communities from three datasets. We applied three methodological innovations: (i) design of a hierarchical sampling scheme to disentangle environmental from spatial factors driving turnover; (ii) resolution of 16S rRNA gene amplicon sequence variants to enable higher resolution community profiling; and (iii) application of the new metric zeta diversity to analyze multisite turnover and drivers. At fine taxonomic resolution, rapid compositional turnover was observed across multiple spatial scales. Turnover was overwhelmingly driven by deterministic processes and influenced by the rare biosphere. The communities also exhibited strong distance decay patterns and taxon-area relationships, with *z* values within the interquartile range reported for macroorganisms. These biogeographical patterns were weakened upon applying two standard approaches to process community sequencing data: clustering sequences at 97% identity threshold and/or filtering the rare biosphere (sequences lower than 0.05% relative abundance). Comparable findings were made across local, regional, and global datasets and when using shotgun metagenomic markers. Altogether, these findings suggest that bacteria exhibit strong biogeographic patterns, but these signals can be obscured by methodological limitations. We advocate various innovations, including using zeta diversity, to advance the study of microbial biogeography.

2.2. Introduction

A central goal of microbial ecology is to link microbial distribution patterns to underlying ecological processes. Developing such links is important both for fundamental science and applied outcomes, for example to make accurate global biodiversity assessments and prioritize management goals in the face of both local and global change [230,231]. However, achieving this critically depends on our abilities to adequately characterise biodiversity at the first stage, with various methodological and theoretical challenges

limiting our understanding of microbial distribution patterns and their underlying ecological drivers.

Several principles have nevertheless become established in soil microbial ecology through cultivation-independent studies over the last two decades. First, it is appreciated that most soils harbour rich and abundant bacterial communities [106,110,232,233]. In most soils, a small number of taxa are abundant and prevalent, while the remaining taxa have low abundance and frequency (the 'rare' biosphere) [35,37] . In common with macroorganisms [234], four key ecological processes control microbial assembly across space and time: environmental selection, diversification, dispersal, and drift [81,86,104]. While much work has emphasized the role of deterministic environmental selection in driving bacterial niche differentiation, especially edaphic factors such as pH [235–239], some studies have also inferred stochastic patterns of community structure, for example due to dispersal limitation or historical diversification [102,239–242]. The relative strength of these factors can vary across time, for example with dispersal controlling recruitment and selection affecting retention during initial stages of primary succession [239,243–245]. As is also the case in the field of macroecology, the relative importance of deterministic and stochastic processes in shaping contemporary distributions of microorganisms continues to be debated and there is a large body of often divergent literature in this area. A major methodological challenge is to perform sampling and analysis that sufficiently disentangles the autocorrelation between environmental and spatial factors in soil ecosystems [81,246–248].

Also controversial is the extent to which microbial communities vary across space. Soil bacteria are generally thought to exhibit weaker biogeographic patterns than macroorganisms [237,249]. Most empirical studies have reported low exponents for taxa-area relationships [237,250–253] and low regression coefficients in distance decay curves [249,253–256], though exceptions have been reported [33,257–259]. Several hypotheses have been put forward to explain these observations [28,260]. Primarily, bacteria are thought to be able to maintain wide geographic ranges in the face of environmental variation by entering dormant states [28,62], leading to limited geographic turnover and shallow taxon-area curves [237,249,261]. However, methodological artifacts may also account for some observations of weak spatial differences [249]. Microbial biogeographic patterns are known to be sensitive to various factors, including spatial scale [262,263], sampling

effort [82,249,264,265], and taxonomic resolution [82,237,249,266–268]. Communities are inherently prone to being undersampled, whether through insufficient sampling effort, low sequencing depth, or rarefying data [269,270]. In addition, the processing of 16S rRNA gene amplicon sequencing data typically used to profile communities can reduce dataset resolution; reads are usually clustered into operational taxonomic units (OTUs) based on an arbitrary identity threshold (usually 97%) and the rare biosphere is regularly removed [54,271]. Compounding these issues, the pairwise analyses generally used to quantify community turnover inadequately partition variation from all community members: incidence-based measures are highly sensitive to the rare biosphere and abundance-based measures focus on the common few [272,273].

In this study, we employed three methodological innovations to address these common limitations of microbial biogeographic surveys and reassess patterns of bacterial community turnover. Firstly, we adopted a hierarchical sampling scheme commonly used in macroecological surveys [274,275]; this enabled us to detect changes in community structure across multiple spatial scales and, in light of controversies in the literature, better distinguish the contributions of environmental and spatial drivers to community assembly processes [248]. Secondly, we profiled community composition using high-resolution 16S rRNA gene amplicon sequence variants (ASVs), leveraging a new generation of processing tools [53,56,276]. We compared the effects of the commonly used approaches of filtering and clustering sequences on calculated community turnover; this is important given that clustering sequences reduces taxonomic resolution and thus may increase the overall similarity of the community, thereby weakening biogeographic patterns [258,268]. Finally, we used the multi-site diversity metric zeta diversity to analyze spatial community turnover and predict the strength of underlying deterministic and stochastic drivers [273]. Unlike the commonly used beta diversity that is calculated from pairwise comparisons, zeta diversity describes the number of taxa shared across multiple sites. As a result, this parameter can discriminate diversity patterns across the spectrum of common, intermediate, and rare taxa [273,277–279], and infer deterministic and stochastic drivers of community assembly. On this basis, we provide evidence that at the level of exact sequence variants, biogeographic patterns of microorganisms are exceptionally stronger than previously reported.

2.3. Materials and Methods

2.3.1. Soil survey design

Topsoil samples were collected along perpendicular latitudinal and longitudinal transects in the Judea Hills and Negev Desert regions, Israel. The latitudinal transect, which was designed to capture a high level of environmental heterogeneity, extended for 160 km in a north/south direction along a steep aridity gradient. This transect traversed four climatic zones that were differentiated by mean annual precipitation patterns: sub-humid shrubland (300-400 mm/yr), semi-arid grassland (~200-250 mm/yr), arid desert (~50-90 mm/yr), and hyper-arid desert (<20 mm/yr). The longitudinal transect, sampled within the arid zone across a relatively homogenous climate, extended perpendicular to the latitudinal transect for 20 km in an east/west direction.

A hierarchical sampling scheme was used to capture biogeographic patterns across multiple spatial scales and provide sufficient spatial resolution to cover the majority of distance classes between sites (**Fig. S3**). Three spatial hierarchies were within each climatic zone: (i) site level (two representative sites of ~1000 m^2), (ii) plot level (three representative plots of ~100 m^2) and (iii) sample level (random triplicates of ~100 cm^2) (**Fig. S2**). Site selection was based on four criteria: (i) soil type (wind-deposited loessic soils in the sub-humid, semi-arid, and arid zone, and gypsic soils in the hyper-arid zone), (ii) presence of soil crust to indicate no recent disturbance, (iii) vegetation-free soil to minimise a vegetation effect, and (iv) a buffer of 100 m to roads, slopes, and seasonal run-off water channels. No statistical methods were used to predetermine sample size.

2.3.2. Soil sampling and analysis

In total, 99 topsoil samples were collected across both transects over a ten-day period in May 2017. Prior to sampling, GPS coordinates and site metadata were recorded. Soil samples of approximately 50 g were collected in triplicate, using sterile techniques, by removing the soil crust (0 – 2 cm depth) and then sampling the

underlying topsoil (2 – 10 cm depth). Samples were placed into individual 50 mL screw top falcon tubes and stored at 4°C until downstream analysis. Within 24 hours of sampling, all soils were homogenized by sieving (500 µm) and soil water content (%) was measured gravimetrically in duplicate. All samples were then shipped to quarantine approved facilities at the School of Biological Sciences, Monash University. For soil chemistry analysis, samples were pooled to form one representative sample per plot and sent to the Environmental Analysis Laboratory, Southern Cross University. In total, 21 separate soil chemical parameters were selected for analysis, based on commonly reported drivers of soil microbial communities globally and those reported by previous studies in the Judea Hills and Negev Desert [280,281]. These included: soil acidity (pH), electrical conductivity (EC), effective cation exchange capacity (ECEC), total organic carbon, total nitrogen, sodium (Na), sulfur (S), phosphate (P), potassium (K), nitrate (NO_3^-), and ammonium (NH_4^+), as well as bioavailable minerals including manganese (Mn), copper (Cu), zinc (Zn), boron (B), aluminium (Al), iron (Fe), and silicon (Si). Each chemical parameter was calculated following Rayment and Lyons methods [282]. Aridity data for each site was obtained from a global geospatial dataset [283] mapping the aridity index (MAP/PET, where MAP = mean annual precipitation, and PET = potential evapotranspiration) at a resolution of 90 arcseconds (approximately 1 km at the equator) using a climatic time series from 1950 to 2000 [284].

2.3.3. Community DNA extraction and sequencing

For all samples, total community DNA was extracted from 0.25 g of soil using the modified Griffiths' protocol [285]. We confirmed the DNA yield, purity, and integrity for each extraction using a Qubit Fluorometer, Nanodrop 1000 Spectrophotometer, and agarose gel electrophoresis. For each sample [286], the hypervariable V4 region of the 16S rRNA gene was amplified using the universal Earth Microbiome Project primer pairs F515 and R806 [287]. The amplicons were subject to Illumina paired-end sequencing at the Australian Centre for Ecogenomics, University of Queensland. Twelve samples were also subject to shotgun metagenomics sequencing (SH.1.A3, SH.1.C2, SH.1.C3, SA.2.B3, SA.1.C3, SA.1.B1, AR.2.A3, AR.2.A1, AR.1.C2, HA.2.C2, HA.1.B1, HA.1.C2). DNA was extracted from 0.25 g of soil using the MoBio

PowerSoil Isolation Kit according to the manufacturer's instructions. Metagenomic shotgun libraries were prepared for the 12 samples using the Nextera XT DNA Sample Preparation Kit (Illumina Inc., San Diego, CA, USA). Sequencing was performed on an Illumina NextSeq500 platform with 2 × 150 bp High Output run chemistry. For analysis of the previously published global dataset [35], the raw 16S rRNA gene amplicon sequences were downloaded from Figshare (https://figshare.com/s/82a2d3f5d38ace925492). This includes samples from six continents, Africa, Europe, Asia, Australia, North America, and South America.

2.3.4. Amplicon-based community profiling

Raw sequences from the Israel and global datasets were processed on the QIIME 2 platform [288] using the deblur pipeline [276] to resolve exact amplicon sequence variants (ASV). In contrast to operational taxonomic unit (OTU)-based approaches that cluster sequences to a fixed identity threshold (usually 97%), deblur controls error rates (typically 0.1% per nucleotide) to resolve single-nucleotide differences over the sequenced gene region [276]. Paired-end raw reads were demultiplexed and adapter sequences were trimmed, yielding 3,989,659 reads across all samples. Forward and reverse reads were joined using the q2-vsearch plugin [289]. A quality filtering step was applied using a sliding window of four bases with an average base call accuracy of 99% (Phred score 20). Low quality reads were removed and sequences were truncated at 250 base pairs before de-noising using deblur [276]. For downstream analysis, three samples with low read counts (<1000 reads) were excluded (SH.1.B2, AR.1.B1, AR.1.B2). In addition, singletons missed by deblur were manually removed, resulting in the loss of 414 ASVs. The final dataset contained 96 samples and 11,335 ASVs **(Table S1)**. In order to compare biogeographic patterns across different taxonomic resolutions, a second dataset was created by clustering the ASVs at a 97% identity threshold using open reference OTU picking via q2-vsearch [289]. The third and fourth datasets were created by removing reads with lower than 0.05% relative abundance from the 100% and 97% identity threshold datasets using the Phyloseq filter_taxa function. Taxonomic assignment was performed as per a previously described approach (https://osf.io/25djp/wiki/home/). Briefly, all reference reads that matched the 515F/806R primer pair were extracted from the Genome Taxonomy

Database (GTDB) [290] and used to train a naïve Bayes classifier [291] by using the fit-classifier-naive-bayes function with default parameters. The classifier was then used to assign the taxonomy to the ASV feature table. Representative sequences were aligned using the multiple sequence alignment program MAFFT [292] and a phylogenetic tree was constructed using the fast-tree method in QIIME 2.

2.3.5. Metagenome-based community profiling

The 16S rRNA gene amplicon sequence is commonly used as a marker to profile microbial communities. However, a major limitation of this approach is the intragenomic and intergenomic variation in copy number of the 16S rRNA gene [293,294]. We conducted a comparative metagenomic analysis on a subset of 12 samples (biological triplicate from within each climatic zone) along the latitudinal transect using a single copy ribosomal marker, in order to test whether our observations of community turnover by 16S rRNA gene amplicon sequencing were affected by this variation. Raw sequence reads in each sample were stripped of adapter and barcode sequences, then contaminating PhiX sequences were identified and removed using the BBDuk function of BBTools v. 36.92 (https://sourceforge.net/projects/bbmap/) with a kmer size of 31 and hamming distance of 1. Retained read pairs were then quality trimmed using BBDuk with Q >20. A total of 318,420,199 reads were obtained from metagenomic sequencing across the 12 samples. In contrast, the read counts for the negative controls were 6,547 (extraction control) and 1,360 (library preparation control). We then used SingleM [295], which uses hidden Markov models (HMMs) searches of single copy ribosomal markers, to generate *de novo* operational taxonomic units (OTUs). In total, 28 HMM searches were performed against 14 single ribosomal single copy marker genes. GraftM was used for taxonomic annotation of OTUs by searching sequences using hmmsearch (HMMER) [296]. For downstream analysis, one single copy marker gene was used for comparison, encoding ribosomal protein L16/L10E (*rplP*). This marker was previously identified as a robust means of distinguishing between both closely and distantly related genomes [297]. Sequences were then clustered *de novo* into OTUs using a sequence identity threshold of 97%. Taxonomic assignment was carried out using the GTDB taxonomy. Due to large differences in sequence depth between the single copy ribosomal marker and 16S

sequence, the amplicon and metagenomic sequences analyzed were both rarefied at 300, which was the minimum number of sequences observed for *rplP*. Rarefied datasets were only used in the supplementary analysis shown in **Figure S15**, whereas the rest of the study used unrarefied datasets.

2.3.6. Richness analysis

Statistical analysis and visualizations were performed in R version 3.4.4 (2018-03-15) using the packages ggplot2 [298], phyloseq [299], vegan [300] and zetadiv [273]. Occupancy frequency distributions [301] were used to visualize the distributions of the numbers of taxa occupying different numbers of areas and examine the distributional shifts at lower identity thresholds and after filtering rare taxa. Taxa accumulation curves were used to compare alpha diversity properties between sites and confirm adequate sampling of the microbial community. A sample-based rarefaction method was used to find the expected curve, namely the Mao Tau estimate, and a moment-based standard deviation was estimated from the extrapolated number of ASVs surveyed (gamma diversity) using the 'exact' method of the *specaccum* function [Vegan | R] [300]. Observed richness and estimated richness (Chao1 and ACE methods) were calculated using the *estimate_richness* function [Phyloseq | R] [299]. To test for significant differences in the mean observed and estimated richness at the site level, an analysis of variance (ANOVA) with a Shapiro-Wilk test to confirm normality was used (**Table S2**).

2.3.7. Turnover analysis

The multi-site diversity metric zeta diversity (ζ) was used [Zetadiv | R] [273] to examine incidence-based turnover in community composition **(Fig. S10)**. Pairwise metrics of incidence-based turnover (e.g. Jaccard, Simpson index) are biased towards detecting turnover that is driven predominantly by the loss and addition of taxa from the rare biosphere, as by definition rare taxa are not shared by many sites. Zeta diversity overcomes this limitation by enabling discrimination between turnover of rare, intermediate, and common taxa. With increasing orders of zeta, the average number of taxa shared between sites declines and the contribution of increasingly more

common taxa to the value of zeta diversity increases. Variation in the rate and form of zeta decline provides information on community structure and inference of the processes driving community assembly. If the zeta decline follows an exponential form (the ratio between ζ_i and ζ_{i-1} is constant), there is a similar probability of finding a common or rare taxon with the addition of a site, suggesting that turnover is predominantly stochastic or dispersal limited. However, if zeta decline follows a power-law form (the ratio between ζ_i and ζ_{i-1} increases at higher orders), then the chance of detecting a common taxon is greater than detecting a rare one with increasing orders, demonstrating structure in the community and suggesting that turnover is driven primarily by deterministic processes such as selection due to edaphic or climatic factors [273]. Zeta decline using Monte Carlo sampling was calculated via the *zeta.decline.mc* function [Zetadiv | R] [273]. Zeta diversity was calculated on non-weighted presence-absence data for ζ orders $\zeta_1 – \zeta_6$; this captured the extent at which the community was structured across each transect, as ζ values within each dataset approached zero. To account for differences in richness between sites, all ζ_i values were normalized by using a Jaccard normalization with subsampling set to 1000 permutations for each analysis. Power-law and exponential models were fitted to ζ_i decline curves and Akaike Information Criterion (AIC) were used to estimate the likelihood of either exponential or power-law model describing the relationship between ζ diversity and order *i*.

2.3.8. Biogeographic analysis

We calculated the distance decay of similarity across both transects to quantify the number of shared ASVs over geographic distance and to explore turnover within the context of geographic distance. Pairwise distance decay was calculated using normalised ζ_2, with sampling set to 1000 using the function *zeta.ddecay* [Zetadiv | R] [273]. To quantify the contribution of rare and common ASVs to turnover, distance decay was calculated for orders $\zeta_1 – \zeta_6$ by using the mean distances between pairs of *n* sites via the *zeta.decays* function [Zetadiv | R] [273]. Spatially explicit taxa-area relationships [302] were calculated by estimating richness as a function of the sample, plot and site level spatial hierarchies (**Fig. S2**) using the *specnumber* function [Vegan | R] [300]. The taxa-area curve was fitted using the Arrhenius model with the expression kA^z, where

k is the average number of taxa, A is the area (spatial hierarchy), and *z* is the steepness of the curve. For comparison, turnover rates from this this study were compared against a total of 655 datasets including bacteria (**Table S7**) and higher eukaryotes [303].

2.3.9. Community structure analysis

Principal coordinate analyses (PCoA) were used on both weighted and unweighted distance matrices. Read counts were normalized to relative abundance and a square root transformation was applied prior to calculating distances between samples using Bray Curtis. For non-weighted analysis, read counts were transformed to incidence (presence-absence) and distances were calculated using the Jaccard index. A multivariate model-based framework was used to test for significant differences in community structure among spatial hierarchies and identify the subset of environmental drivers that best explain spatial patterns in community structure [MVAbund | R] [304]. Microbial abundance and incidence data typically show a mean-variance relationship, which standard approaches such as PERMANOVA, ANOSIM and RDA fail to account for. Instead they rely on pairwise distance-matrices which convert multivariate datasets to univariate ones which has been shown to reduce statistical power. MVAbund solves this problem for non-normal data by fitting a single generalized linear model (GLM) to each ASV separately and performing re-sampling of *p* values to determine significance of a shared predictor variable. In this study, ASV incidence data were modelled using generalized linear models. Mean variance-relationships of the data were confirmed by visually inspecting scatterplots showing mean variance as a function of ASV incidence. Model assumptions were validated by inspecting Dunn-Smyth residuals as a function of each predictor variable and significance was established using a likelihood ratio test (LRT) with PIT-trap bootstrapping [305]. To obtain the subset of predictor variables which best explain a multivariate response, significant predictor variables were passed through a forward selection in a multivariate linear model using the top ten independent variables with the highest average R^2. A variation partitioning analysis was performed to disentangle the autocorrelation between environmental and geographic distance and partition variation in community structure into its spatial and environmental components. Multi-

site generalized dissimilarity modelling (MS-GDM) was used to identify the importance of correlates of turnover, by regressing ζ_2 against the sub-set of identified predictor variables for each plot and each taxonomic resolution using *zeta.msgdm* function [Zetadiv | R] [273]. Then a variation partitioning analysis was performed using the *zeta.varpart* function [ZetadivR] [273], which partitioned the variation into (a) variation explained by distance alone, (b) variation explained by either distance or environment, (c) variation explained by environment alone, and (d) unexplained variation.

2.4. Results

2.4.1. Most community members have a low to moderate occupancy across soil transects

We analyzed 96 topsoil samples along two perpendicular transects **(Fig. S1)**: a 160 km latitudinal transect (north/south) spanning four climatic zones (sub-humid, semi-arid, arid, hyper-arid; 69 samples) and a 20 km longitudinal transect (east/west) in the arid zone (27 samples). Within each transect, samples were collected according to a hierarchical design (2 sites per zone × 3 plots per site × 3 samples per plot) **(Fig. S2)**. This sampling scheme was designed to enable the analysis of microbial community turnover at multiple spatial scales, capture a wide spectrum of distance classes **(Fig. S3),** and discriminate underlying spatial and environmental drivers. The bacterial and archaeal communities in each sample were profiled using both new and standard approaches for processing 16S rRNA gene amplicon sequencing data. Rarefaction curves **(Fig. S4 & S5)** and richness estimators **(Fig. S6; Table S2 & S3)** confirmed that sequencing and sampling efforts sufficiently captured the diversity of taxa within and across samples. A high-resolution community profile was generated by processing reads using the deblur pipeline [276] to resolve 16S amplicon sequence variants (ASVs) at the single-nucleotide level (singletons removed) **(Table S1; Fig. S7 & S8)**. Most sequences were from the nine dominant soil phyla [35], especially Actinobacteriota, Chloroflexota, and Proteobacteria, as well as putatively ammonia-oxidizing archaea **(Fig. S8)**. The occupancy frequency distribution [301] of the 11,335 taxa (ASVs) detected was positively skewed; ~67% of 7602 taxa were detected in fewer than 10% of samples **(Fig. 1a; Fig. S7)**.

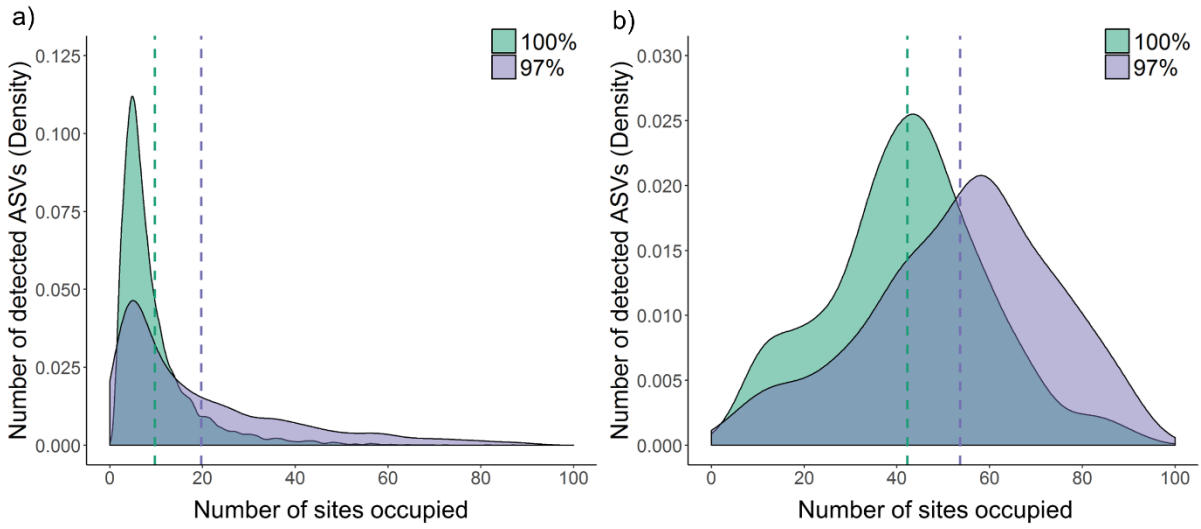


Figure 1. Occupancy frequency distribution of amplicon sequence variants at different taxonomic resolutions. The Kernel-smoothed density plot shows the number of sites that each taxon (amplicon sequence variants, ASVs) was detected in across the dataset. **(a)** Effect of clustering taxa at either 100% or 97% identity threshold. **(b)** Effect of either including or removing taxa with lower than 0.05% relative abundance. Vertical dotted lines show distribution means. Stacked histograms representing this data are shown in **Figure S7**.

We then compared the effects of applying two standard approaches used to process sequencing data into OTUs: (i) clustering, i.e. combining sequences with an identity threshold of 97%, and (ii) filtering, i.e. removing sequences with lower than 0.05% relative abundance. There was a sharp decrease in the number of taxa retained (2943 clustered, 222 filtered, 403 clustered then filtered). Though clustering inevitably reduced richness **(Fig. S6)**, as well as the frequency of intermediate taxa**,** it did not affect the skew of the occupancy frequency distribution **(Fig. 1a)**. However, when less abundant taxa were filtered from the datasets, occupancy frequency shifted from a positive skew to a modal distribution **(Fig. 1b)**. These findings suggest that the prevalence of most community members is low to moderate; standard clustering and filtering approaches not only affect the 'rare' biosphere, but a large percentage of community members with moderate range sizes. In turn, changing occupancy properties may underestimate ecological heterogeneity and markedly bias biogeographic interpretations.

2.4.2. Deterministic factors drive differences in community composition between soil samples

We subsequently used pairwise metrics (beta diversity) to analyze community composition between samples. We detected significant differences in community structure down to the plot level **(Fig. S8 & S9)**. The extent of compositional differences observed between sites depended on both the community property used (incidence vs abundance, taxonomic vs phylogenetic) and the taxonomic resolution of the dataset. MDS ordinations showed prominent "V" patterns **(Fig. S9)**; this pattern, also known as the horseshoe effect, has been shown to indicate the presence of niche differentiation along environmental gradients [306]. In line with the high environmental heterogeneity along the latitudinal transect, differentiation was more pronounced for the latitudinal than longitudinal transect **(Fig. S9)**.

A variation partitioning analysis was used to delineate the measured environmental and spatial predictor variables that account for the greatest amount of variation in pairwise community structure. Across the latitudinal gradient, 45% of the community variation of the high-resolution dataset was explained by measured edaphic factors **(Fig. S11)** with pH, C:N ratio, aridity, and salinity explaining the greatest amount of variation **(Table S4)**. These results broadly reflect other studies in the Negev region and along aridity gradients globally [107,280,307]. Less variation was explained for the more homogeneous longitudinal transect (35%) **(Fig. S11)**. Altogether, these results suggest environmental effects predominate over distance effects in driving community composition. In common with other biogeographic studies [253,308,309], a large proportion of variation was unexplained by the measured variables. A combination of factors could contribute to this unexplained variation, including deterministic processes driven by unmeasured abiotic and biotic factors, as well as neutral ecological drift and potentially sampling effects. In both the MDS and variation partitioning analyses, less variation in community composition could be explained and partitioned for the high-resolution dataset compared to filtered ones **(Fig. S9 & S11)**. The rank importance and weight of environmental predictors also shifted depending on taxonomic resolution for both transects **(Table S5 & S6)**. In support of recent findings [111], these results suggest that different environmental drivers structure common and rare microbial taxa.

2.4.3. Soil microbial communities exhibit rapid deterministically-driven multisite turnover

We also analyzed spatial turnover in the community using the recently developed metric zeta diversity. As depicted in the infographic in **Fig. S10**, zeta diversity describes the number of taxa shared by multiple combinations of sites; whereas beta diversity (which it encompasses) is predisposed to detecting turnover of rare taxa, zeta diversity discriminates patterns and drivers of turnover across the spectrum of common, intermediate, and rare taxa [273,277]. For the high-resolution dataset, zeta diversity rapidly declined towards zero within four orders in the latitudinal transect (ζ_4 = 0.0068) **(Fig. 2a)**. This means that the average number of taxa shared across any four plots was 0.68% of 10,826, indicating very rapid turnover. Similar patterns were observed across both transects and within each climatic zone; somewhat lower turnover was observed along the longitudinal transect (ζ_4 = 0.010) and hyper-arid samples **(Fig. S12)**. Reducing taxonomic resolution markedly slowed compositional turnover **(Fig. 2a)**; for the clustered and filtered dataset, up to 30% of the community were shared across any four plots (ζ_4 = 0.18 and 0.30 for the latitudinal and longitudinal transects respectively). Such findings reflect that, given common, intermediate, and rare community members show different distribution patterns, lowering taxonomic resolution distorts detection of microbial turnover and underlying drivers.

Derivations show that zeta decline most often follows either a power-law or an exponential form, which are respectively associated with either deterministic or stochastic community assembly processes [273]. Zeta decline much better fitted a power-law form for both transects and within each climatic zone **(Fig. 2c & Fig. S12)**, suggesting deterministic processes drive turnover. While power-law support was overwhelming for the high-resolution dataset, there was some support for exponential models in the low-resolution datasets; filtering microbial datasets, by obscuring biogeographic structure, may therefore cause false signals of stochastic assembly processes **(Fig. S13)**.

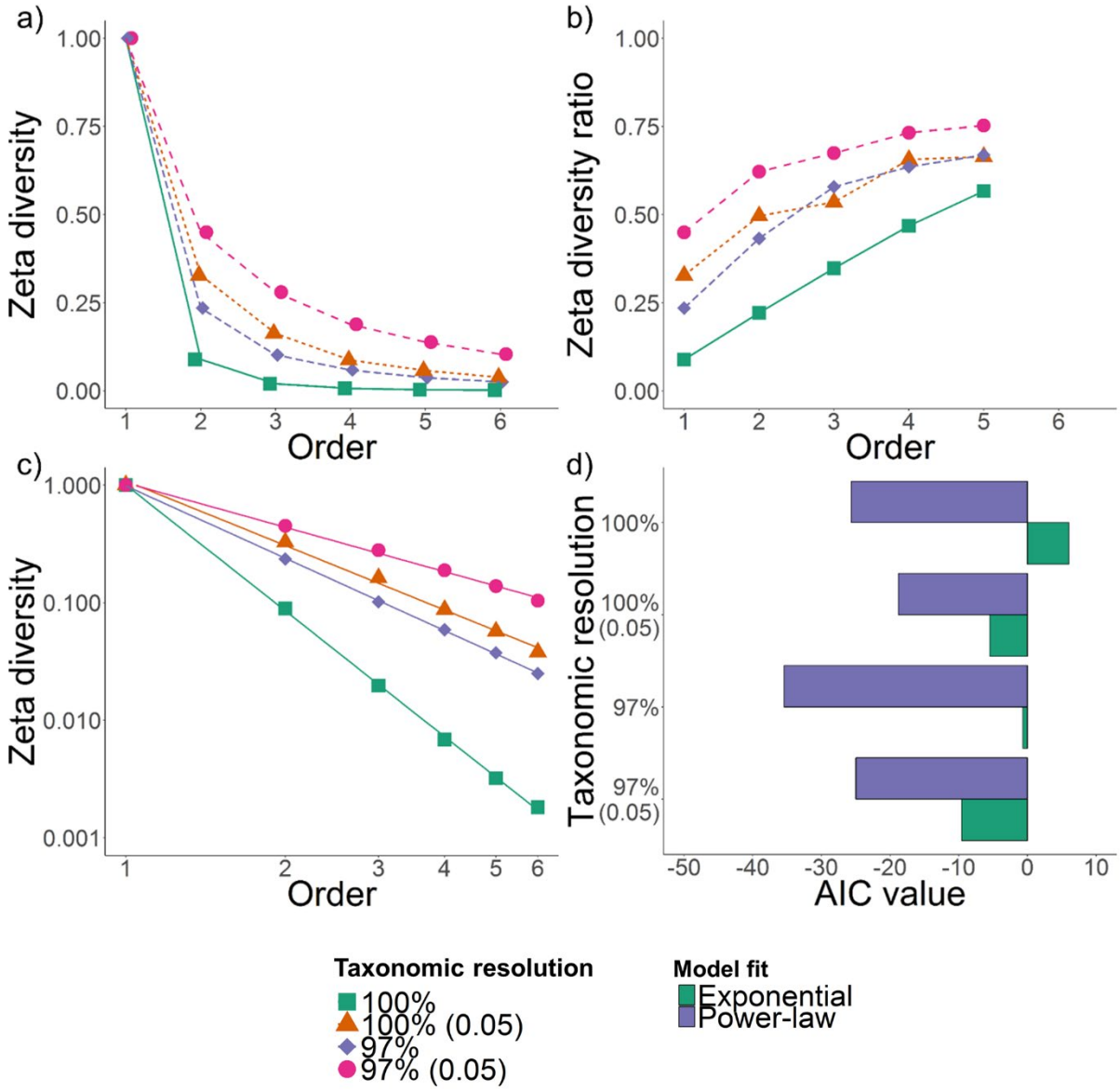


Figure 2. Community turnover and assembly processes along the latitudinal transect at different taxonomic resolutions. Zeta decline **(a)**, showing how the number of shared taxa (ASVs) decline with the addition of sites to the comparison (Order). The taxon retention rate using the zeta diversity ratio (**b**), which demonstrates the probability of retaining common over rare taxa at any particular order with the addition of an extra site. In all cases, the decline followed a power law form (**c**), which is associated with deterministic processes driving structure in the community (versus an exponential form). However, the relative support for the power law form varies and depended on the taxonomic resolution (**d**).

2.4.4. Soil microbial communities exhibit strong distance decay and taxon-area relationships

We subsequently measured distance decay using a combination of pairwise (beta decay) and multisite (zeta decay) metrics. Based on pairwise comparisons, a strong decay of shared taxa was also detected across transects ($p < 0.0001$) **(Fig. 3a & 3b; Table S6)**. Lowering taxonomic resolution caused a large increase in community similarity, a steeper distance decay coefficient, and a lower rate of community turnover overall; across the 160 km latitudinal transect, there was a 82% reduction in community similarity for the high-resolution dataset compared to 50% to 60% reductions for the clustered and/or filtered datasets. Given the concordant support for deterministic drivers, based on the zeta diversity **(Fig. 2),** variation partitioning analysis **(Fig. S12)**, and MDS analysis **(Fig. S9)**, these decay patterns likely reflect environmental filtering rather than dispersal limitation.

To quantify how distance decay compares between rare, intermediate, and common taxa, distance decay was calculated for up to six zeta orders by using the mean distance between pairs for up to six plots. For both transects at high-resolution, the gradient of the distance decay curve rapidly and significantly decreased with increasing zeta order (**Fig. 3b & Fig. S14)**. This provides additional evidence that these microbial communities are highly structured and that turnover is driven by loss of rare to intermediate members. In contrast, there were no significant changes in distance decay rates with zeta order for the less resolved datasets, further demonstrating that clustering and/or filtering obscures biogeographic patterns.

Given these outcomes, we revisited the controversial taxa-area relationship for bacterial communities [82,237] using these datasets. This universal relationship in ecology describes the increase in taxon richness with area sampled, i.e. $S = cA^z$ (where S is number of species, A is area sampled, c and z are fitted constants), and its exponent z is a normalized measure of turnover rates that can be compared between organismal groups [237]. A strong taxa-area relationship was also observed for both transects ($p < 0.001$) **(Fig. 3c; Fig. S14; Table S7)**. The z exponents were 0.39 (latitudinal transect) and 0.4 (longitudinal transect) for the original high-resolution

datasets, and decreased to 0.13 and 0.09 in the clustered then filtered datasets **(Table S7)**. Such *z* exponents greatly exceed those reported for bacterial communities in most previous studies (median 0.04), but are congruent with four studies [257–259,310], two of which also performed hierarchical sampling. These exponents are of the same order of magnitude to those previously reported for animal and plant datasets (median 0.27) **(Fig. 3d & Table S7)**, indicating biogeographic patterns of bacteria and macroorganisms may not profoundly differ. However, more broad and detailed side-by-side sampling is required to compare scaling relationships between bacteria and macroorganisms.

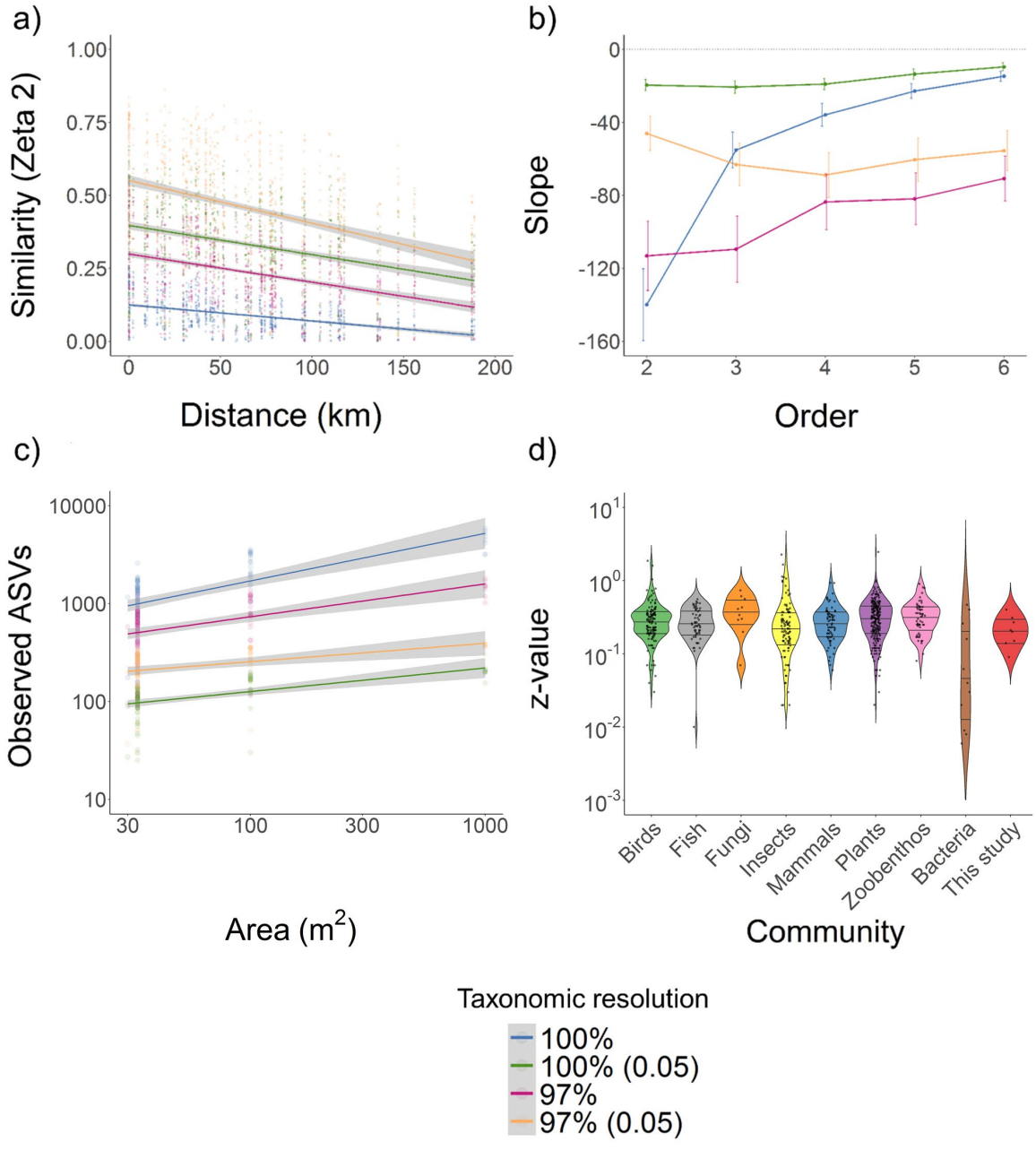


Figure 3. Distance decay in community similarity and the taxa-area relationship at different taxonomic resolutions. **(a)** Zeta distance decay relationship showing community turnover with increasing geographic distance based on pairwise comparisons (ζ_2) of sites along the latitudinal transect. **(b)** Differences in the slope (coefficient) of distance decay between pairwise and higher orders of zeta (>2) using the average distance between sites. **(c)** Taxa-area relationships of the increase in richness with area sampled along the latitudinal transect, both showing clear consequences of taxonomic resolution for understanding how compositional heterogeneity scales with distance. **(d)** Violin plots showing the density-distribution and interquartile range of the exponent *z* of the taxa-area slope reported here with those from other studies for bacteria and eukaryotes (**Table S7**). Results are compared at four different taxonomic resolutions, whereby (i) taxa were clustered at either 100% or 97% identity threshold and (ii) taxa with lower than 0.05% relative abundance were either included or removed.

2.4.5. Similar biogeographic patterns are observed using metagenomic sequences and global datasets

This study relies on 16S rRNA gene amplicon sequencing to profile the soil microbial communities. This approach remains standard practice for biogeographic studies, given the alternative of metagenomic profiling requires much higher sequencing depths and yields either less information-rich short reads or more error-prone long reads [311]. However, limitations of 16S rRNA gene sequencing include potential for amplification and sequencing errors, biases in the primer sets, and genome variability in 16S copy number [294]. While it is possible that the dataset includes some spurious sequences introduced through this approach, these are unlikely to account for the surprising observations made here. First, a range of accuracy measures suggest deblur efficiently denoises sequencing data and that a 100% identity threshold resolved using the deblur denoising pipeline is optimal for community profiling with the V4 region [53,276]. Secondly, similar but weaker patterns of rapid deterministically-driven community turnover was observed for the clustered (but not filtered) datasets, in which most spurious sequences should be removed **(Fig. 1a, 2, 3)**.

To test the reproducibility of our findings, we performed short-read metagenomic sequencing of a subset of 12 samples across the latitudinal transect and analyzed a single-copy ribosomal marker gene (L10e/L16). Similar to the 16S amplicon data, samples showed a high estimated richness, comparable taxonomic composition, and rapid community turnover **(Fig. S15)**. Zeta decline approached zero after three orders (ζ_3 = 0.06) using a rarefied dataset (**Fig. S15**). In combination, this suggests that 16S ASVs are sufficient to estimate community turnover, whereas standard methods of clustering and filtering data obscure biogeographic patterns and inflate signals of taxon commonness.

Having detected these patterns at local and regional scales, we analyzed whether similar patterns were observable at the continental scale. To do so, we analysed a previously published 16S profiles of 237 soil samples collected from six continents [35]. As with our original dataset, we processed the 16S rRNA amplicon sequencing data into ASVs and analyzed the effects of clustering and/or filtering. The occupancy frequency distribution of the taxa showed a similar skew to the Israel dataset **(Fig. S16)**. Concordant with our previous observations, zeta diversity rapidly declined across the first few orders and followed a power-law relationship with strong model support **(Fig. S16)**. Clustering and filtering altered the occupancy frequency distribution, resulting in ~10% to 30% of taxa being retained at six zeta orders **(Fig. S16)**. Thus, our key result that soil bacterial communities exhibit strong biogeographic patterns are reproducible in datasets at local (longitudinal transect), regional (latitudinal transect), and global scales.

2.5. Discussion

In this study, we analyze patterns and drivers of soil microbial composition across multiple scales. We overcame common limitations in microbial biogeographical studies by leveraging innovations in sampling design, amplicon processing, and diversity metrics. We found that (i) soil bacterial communities exhibit strong biogeographic patterns, (ii) spatial turnover is rapid as most taxa have low to moderate levels of occupancy, and (iii) community structure is influenced more by niche differentiation due to environmental variation rather than dispersal limitation. Our

findings agree with previous literature that reported the uneven distribution of bacteria across communities and the strong influence of deterministic drivers [81,106]. However, we observed much stronger spatial turnover than reported in most, though not all, previous literature [237,249,253]. This is reflected by the concordant findings of four independent analyses using the original high-resolution dataset. Occupancy frequency distributions revealed most taxa were shared across less than 10% of samples **(Fig. 1)**. Through zeta decline analysis, we detected a logarithmic decrease in the number of taxa shared as number of sites increased **(Fig. 2)**. In addition, we observed strong distance decay **(Fig. 3a)** and taxon-area relationships **(Fig. 3c)**, with *z* values one to two orders of magnitude higher than most previous observations [250,253,257–259].

Multiple factors may explain why we observed high environmentally-driven turnover. These potentially include the choices of sampling site, sampling scheme, sequence processing, and downstream analyses. It is notable that our desert sampling sites contained loessial soils that facilitate dispersal and the regional transect contained high environmental heterogeneity, which is known to be associated with increased bacterial turnover [252,280,308]; however, this is unlikely to primarily account for most discrepancies with previous literature, given rapid turnover was also observed in the local transect where physicochemical variation was lower and similar findings were also made in the global analysis. A more significant factor may be that our study adopted a hierarchical sampling design in order to quantify microbial variation across multiple spatial scales. In this regard, it is well-recognised that sampling design and sample size are critical determinants of taxa-area relationships [82,302]; this reflects that the detection of rare taxa largely determine species evenness and spatial structure, which in turn affects the exponent *z* [258]. Methodological advances that improve the detection and inclusion of rare taxa are therefore predicted to align microbial *z* values more closely with those reported for animal and plant communities [82,267]. it is notable that other studies reporting high taxon-area exponents also used spatially explicit hierarchical designs [257,312].

However, the biggest factor likely underlying these discrepancies is the treatment of sequencing data. A pervasive feature of 16S rRNA amplicon gene surveys is the clustering of similar sequences to remove potential ‘noise’ and, less commonly, the

filtering or undersampling of low frequency sequences that constitute the rare biosphere. As summarised in **Fig. 4**, this greatly reduces and distorts the information in datasets, obscuring patterns in occupancy, turnover, and drivers. We avoided such downfalls by using a recently developed denoising algorithm to resolve sequence variants [276], while confirming through rarefaction curves that our sequencing efforts captured most rare taxa within and between samples. Through simulating sequencing processing, we observed major differences in occupancy frequency, zeta diversity, distance decay, and taxon-area relationships upon filtering rare taxa and, to a lesser extent, clustering similar sequences **(Fig. 4)**. It should be noted that these observations may appear to conflict with those of a recent study that reported clustering did not "change the rate of microbial taxonomic turnover" [249]. However, this may be an issue of interpretation of distance decay curves. In common with this study [249], we also observed that the distance decay coefficient of bacteria and archaea remains similar between taxonomic resolutions, reflecting similar observations reported in fungal [313] and plant [314] communities. However, as the community similarity (y-intercept) is lower at higher resolution, a higher proportion of taxa are lost overall in unclustered compared to clustered datasets. Thus, it is reasonable to conclude that clustering masks microbial taxonomic turnover and broader biogeographic patterns.

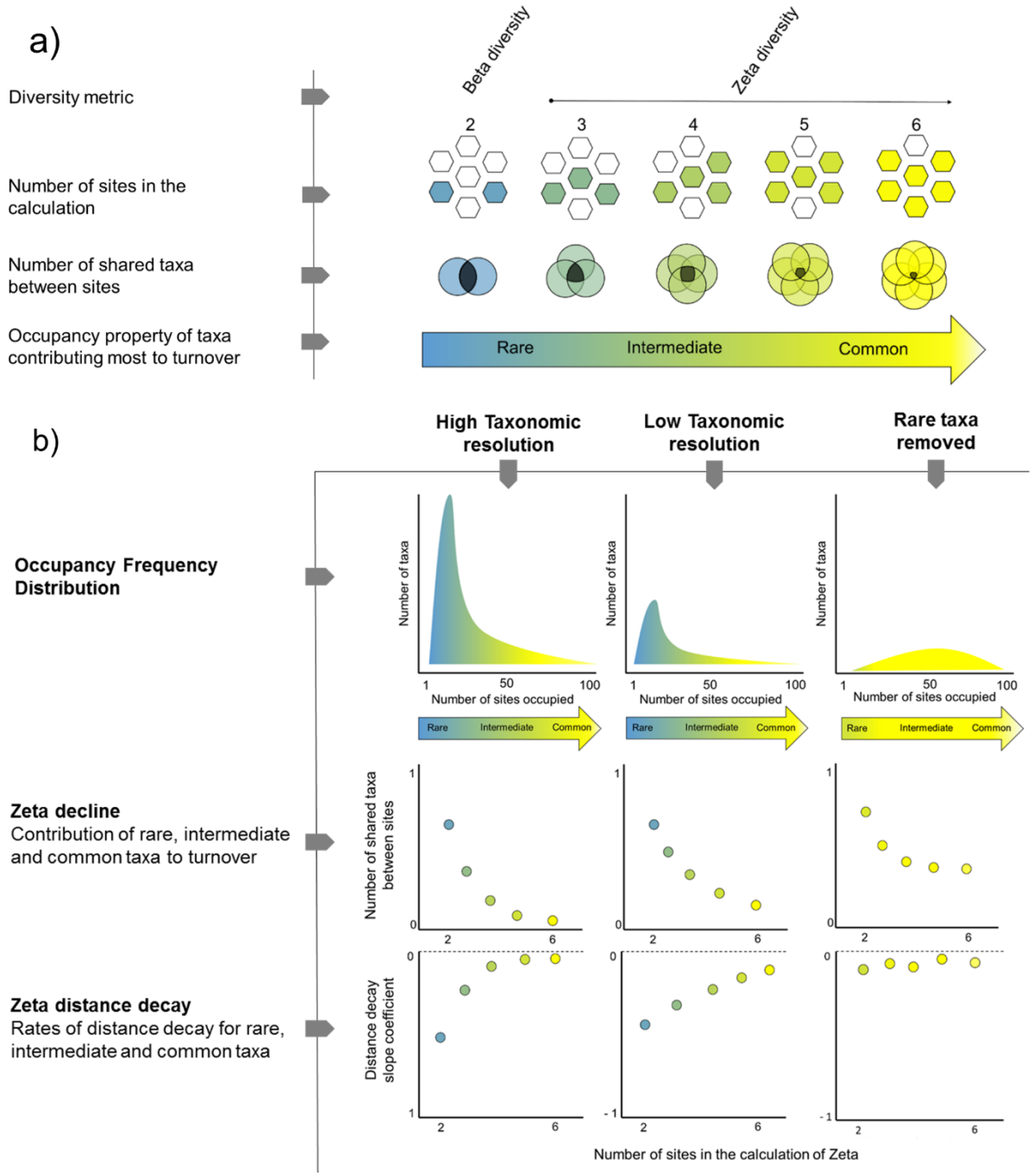


Figure 4. Summary of biogeographic patterns of soil microbial communities at different taxonomic resolutions. **(a)** Principle of how zeta diversity encompasses turnover of rare, intermediate, and common community members. **(b)** Comparison of patterns of occupancy frequency, zeta decline, and distance decay for rare, intermediate, and common community members. In addition, the figure demonstrates how the common approaches of clustering and filtering can bias biogeographic interpretations of microbial communities.

This study also highlights the different patterns and drivers of community turnover between rare, common, and intermediate community members. As demonstrated by the occupancy frequency distribution, filtering sequences removes most rare species and retains most common ones. Based on the beta diversity analysis, we observed significant differences in the proportion of variation assigned to environmental, spatial, shared, or unexplained components at different taxonomic resolutions. This agrees with recent reports that environmental and spatial drivers differentially act on common and rare taxa [47,315]. Abundant generalists and rare specialists have been shown to differentially respond to environmental change, reflecting differences in niche breadth [86,316,317]. Beyond these pairwise observations, we used zeta diversity to demonstrate that the turnover patterns reflect those typically observed in deterministically structured communities. Zeta decline consistently follows a power-law, which indicates that communities are non-randomly structured such as those with clear niche or range differentiation. However, upon lowering taxonomic resolution, these patterns degrade and increasingly resemble stochastic patterns such as seen in habitats with strong aeolian or aquatic flows (**Fig. 4**). These findings suggest that at lower taxonomic resolutions [268] or when rare taxa are removed [258], the community structure becomes more similar and thus predicted assembly processes switch from deterministic to stochastic. Through incorporating a multi-site distance decay model, significant differences in the spatial structure of rare, intermediate, and common taxa could also be detected.

Looking forward, this work demonstrates how microbial biogeography can be advanced using readily implementable approaches. There is scope to use the methodological and theoretical innovations shown here to investigate these patterns across a broader range of environments, and temporal scales. Detailed studies are needed to better capture the biotic and abiotic subsets of drivers responsible for changes in community turnover across all occupancy classes; this has been achieved in plant ecology [277,318], but remains understudied in the microbial literature. Likewise, it is critical to compare the patterns and drivers of community turnover in parallel for microorganisms and macroorganisms. Indeed, a key observation of our study is that *z* exponent for the taxon-area relationships microbial communities falls within the interquartile range of higher animal and plant communities, suggesting microorganisms and macroorganisms exhibit similarly strong spatial structure.

However, given these exponents are highly sensitive to factors such as sampling design, sample size, and taxonomic resolution [82,302], a rigorous comparison of turnover between domains requires side-by-side sampling. Finally, emerging advances in long-read 16S rRNA gene amplicon sequencing and processing may enable resolution of biogeographic patterns of microorganisms at both the species and strain levels [293].

2.6. Footnotes

Author contribution: C.G., S.L.C., M.A.M., and O.G. conceived and supervised this study. S.B., C.G., M.A.M., S.L.C., and O.G. designed the study. S.B., O.G., and N.W. were responsible for field sampling. S.B. carried out all DNA extractions, sequence processing, and statistical analyses. S.B., C.G., M.A.M., S.L.C., D.J.B., and D.J.P. analyzed data. S.B. and C.G. wrote and edited the paper with input from all authors.

Acknowledgements: This work was primarily funded by a Monash University & Ben Gurion University of the Negev Seed Fund (awarded to C.G., O.G., and S.L.C.). It was supported by an ARC DECRA Fellowship (DE170100310; awarded to C.G.), an ARC Discovery Project Grant (DP170101046; awarded to S.L.C. and M.A.M.), a Holsworth Wildlife Research Endowment (awarded to S.B.), a Monash University PhD Scholarship (awarded to S.B.), and an NHMRC EL2 Fellowship (APP1178715; salary for C.G.). We thank Ya-Jou Chen and Thanavit Jirapanjawat for technical advice, Capucine Baubin, Dimitri Meier, and Stefanie Imminger for field assistance, and Philip Hugenholtz and David Waite for metagenome sequencing.

Data availability: The amplicon and shotgun sequencing datasets generated for the current study will be deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive prior to acceptance for publication.

Supplementary material: Supplementary figures and tables are found in the Appendix A.

Ethics Declaration: The authors declare no conflict of interest

Chapter 3

Reciprocal activities of chemosynthetic and photosynthetic bacteria across a steep desert aridity gradient

Sean K. Bay[1,2], David W. Waite[3], Osnat Gillor[4], Philip Hugenholtz[5], Chris Greening*[1,2]

[1] School of Biological Sciences, Monash University, Clayton, VIC 3800, Australia
[2] Department of Microbiology, Biomedicine Discovery Institute, Monash University, Clayton, VIC 3800, Australia
[3] School of Biological Sciences, University of Auckland, Auckland 1010, New Zealand
[4] Department of Environmental Hydrology and Microbiology, Ben Gurion University of the Negev, Sde Boker, Israel
[5] Australian Centre for Ecogenomics, School of Chemistry and Molecular Biosciences, St Lucia, QLD 4072, Australia

***** Correspondence can be addressed to:
Associate Professor Chris Greening (chris.greening@monash.edu), Department of Microbiology, Monash University, Clayton, VIC 3800, Australia

In preparation: *Nature Communications*

3.1. Abstract

Desert soils harbor diverse communities of heterotrophic bacteria despite lacking organic carbon inputs from vegetation. A major question is therefore how these communities maintain biodiversity and biomass in such resource-limited soils. We addressed this question by investigating desert topsoils and biological soil crusts collected along an aridity gradient traversing four climatic regions (sub-humid, semi-arid, arid and hyper-arid). Metagenomic analysis showed that these communities harbored a variable potential to utilize sunlight, organic compounds, and inorganic compounds as energy sources. Thermoleophilia and Actinobacteria were the most abundant and prevalent classes across the aridity gradient; genome-resolved analysis suggested these taxa are metabolically flexible, capable of mediating aerobic organoheterotrophic growth, as well as conserving energy and fixing carbon using atmospheric H_2 as an energy source. In contrast, the abundance of Cyanobacteria was variable and often low across the aridity gradient. We subsequently performed biogeochemical measurements to measure how two key metabolic processes interact with aridity: (i) chemosynthetic H_2 oxidation and (ii) photosynthetic CO_2 fixation. Gas chromatography analysis revealed biomass-normalized rates of H_2 consumption increased 500-fold along the aridity gradient, correlating with increased abundance of high-affinity hydrogenases. Radiolabelled carbon fixation assays confirmed that photosynthetic processes exhibited the inverse relationship, with reduced photosynthetic capacity in arid and hyper-arid soils. Altogether, this suggests that the dominant bacterial lineages inhabiting hot deserts use different strategies for energy and carbon acquisition depending on resource availability. Moreover, these findings suggest trace gases are critical energy sources supporting the productivity and resilience of desert biocrust and topsoil communities.

3.2. Introduction

Photosynthetic primary producers are in low abundance in the desert and dryland ecosystems that span 40% of the earth's surface [319]. Whereas most terrestrial ecosystems are driven by plant-derived organic matter, plant biomass declines with aridity and flora is particularly sparse in arid and hyper-arid deserts [217]. Some

cyanobacteria and microalgae are nevertheless able to persist even in hyper-arid deserts by retreating to environmental refugia such as biological soil crusts (biocrusts) and lithic niches [138,218–222]; such environments provide desiccation buffers, physical stability, and protection from ultraviolet radiation [222,320–323]. As oxygenic phototrophs, these microorganisms use photosystems to capture light and transduce energy, and use either type IA or IB RuBisCO (ribulose 1,5-bisphosphate carboxylase / oxygenase) to fix carbon dioxide (CO_2) into organic carbon [223,224]. In the interior of arid and hyper-arid deserts, the abundance of phototrophic communities become increasingly rare and spatially fragmented [218,222,225,226]. Nevertheless, diverse communities of microorganisms can be found in open desert soils and must survive the cumulative pressures of low water and carbon availability, elevated temperatures, salinity, and ultraviolet radiation [9,41,324,325]. The most abundant microorganisms in these environments are members of dominant bacterial soil phyla such as Actinobacteriota, Proteobacteria and Chloroflexota, most of which are thought to be aerobic organoheterotrophs [9,41,326]. A major question is how these bacteria maintain their energy and carbon needs in these environments given their multiple physicochemical pressures and the dearth of photosynthetic primary producers.

It is thought that bacteria reduce their energy requirements in desert ecosystems by entering dormant states [209,210]. Dormancy is a life history strategy in which cells enter a reversible state of reduced metabolic activity and increased environmental resilience in response to pressures such as resource limitation [210]. Dormant bacterial seed banks, by allowing bacteria to persist under conditions which favour survival over growth, have in turn been shown to act as reservoirs of microbial biodiversity [211,261,327]. While the energy needs of dormant cells are usually three orders of magnitude lower than growing cells, some energy expenditure is nevertheless required for cells to maintain basic functions, allowing an eventual return to active states [209,210]. It is generally thought that desert bacteria primarily survive in dormant states by using macromolecular reserves, which are synthesized when organic carbon becomes transiently available following hydration events [328,329]. However, recent culture-based studies have demonstrated that some aerobic organoheterotrophs can in fact broaden their repertoire of exogenous substrates during carbon starvation. Most notably, various bacterial isolates are known to use the atmospheric trace gases hydrogen (H_2) and carbon monoxide (CO) as alternative electron donors to sustain aerobic

respiration [70–72,330,331]. Genetic studies focused on Actinobacteriota have shown that trace gas oxidation significantly increases long-term survival under energy starvation [332–335]. Although these studies did not focus on desert isolates, it is plausible that Actinobacteriota and other taxa in desert ecosystems also meet their energy needs by scavenging trace gases in their dormant states.

In this regard, atmospheric H_2 may be a particularly important energy source driving aerobic respiration and carbon fixation in desert environments. This gas is thought to be highly dependable for bacteria for four key reasons: (i) it is ubiquitous throughout the earth's lower atmosphere (mixing ratio 0.53 ppmv), (ii) it readily diffuses through cell membranes, (iii) it has a low activation energy, and (iv) its combustion yields a high amount of free energy [19,176,336]. Bacteria oxidize atmospheric H_2 using high-affinity, oxygen-tolerant [NiFe]-hydrogenases; these bacteria transfer electrons derived from H_2 through the quinone pool to terminal oxidases, resulting in the generation of proton-motive force [185,331,337]. Various hydrogenase lineages are known to support aerobic H_2 oxidation, including the group 1h, 1d, 1f, 1l, and 2a [NiFe]-hydrogenases [162,179,180,331,338,339], the first of which seems to be principally responsible for atmospheric H_2 oxidation in soil ecosystems [330,331,340]. Some aerobic bacteria can use electrons derived from H_2 to fix CO_2 into biomass [160,228,341]. It was conventionally thought that aerobic hydrogenotrophic growth was restricted to H_2-enriched environments such as root nodules and geothermal systems [160]. However, our recent studies suggested that Actinobacteriota in Antarctic desert soils can use atmospheric H_2 as an energy source to support carbon fixation. Genome-resolved metagenomic analysis demonstrated that bacteria from phyla, including Actinobacteriota, co-encoded a group 1h [NiFe]-hydrogenase together with a type IE RuBisCO linked to the Calvin-Benson-Bassham cycle [28]. Consistently, microbial communities within these desert soils rapidly oxidised atmospheric H_2 and fixed CO_2 into biomass [28]. This minimalistic mode of primary production may be critical for maintaining energy and carbon levels in extreme desert environments [229,326].

In this study, we used metagenomic and biogeochemical approaches to determine how primary production strategies vary along an aridity gradient. Based on the above findings, we predicted that oxygenic photosynthesis would predominate in more moist soils, whereas trace gas oxidation would be most active in drier soils. To test this, we

investigated topsoils and biocrusts sampled along a steep aridity gradient in the Negev Desert, Israel traversing sub-humid, semi-arid, arid and hyper-arid climates. We show that, while photosynthetic primary production strategies are dominant processes in semi-arid soils, oxidation of H_2 to conserve energy may be a dominant strategy in arid and hyper-arid soils where photosynthesis is inhibited by water availability. In turn, these findings add to growing evidence that hidden metabolic flexibility of bacteria contributes to the resilience and productivity of oligotrophic ecosystems.

3.3. Materials and Methods

3.3.1. Field sampling

The sampling transect extended for 160 km in a north/south direction across the Judea Hills and Negev Desert regions of Israel. Samples were collected from four climatic zones differentiated by mean annual precipitation patterns and soil physicochemical properties: sub-humid shrubland (300-400 mm yr^{-1}), semi-arid grassland (~200-250 mm yr^{-1}), arid desert (~50-90 mm yr^{-1}), and hyper-arid desert (<20 mm yr^{-1}) [342]. Samples were collected according to a previously described hierarchical sampling design (4 zones × 2 sites × 3 plots × 3 subsamples) [342]. To minimise the effects of non-climatic variables, sampling was restricted to wind-deposited loessic soils in the sub-humid, semi-arid and arid zone, and gypsic soils in the hyper-arid zone. In addition, all samples were colected with a minimum 2 m buffer to vegetation, contained visible biocrusts, and were at least 100 m from roads and slopes. Biocrust samples (~1-2cm) were extracted whole using a stainless steel spatula to separate the biocrust layer from the underlying soil. Crusts varied in their physical appearance and colour along the gradient. Reflecting high organic matter and carbon content, sub-humid sites harboured dark brown crusts, which were replaced by lighter colours with increasing aridity. In semi-arid and arid sites some crusts had dark brown speckles, indicating the dominance of phototrophic communities dominated by Cyanobacteria. Samples were transferred to a petri dish padded with cotton wool and sealed using parafilm. Underlying topsoils (~2-10 cm) were also sampled, processed, and subject to physicochemical analysis as previously described [342]. Sampling was conducted over a 10 day period in May 2017.

3.3.2. Community DNA extraction

Total community DNA was extracted from 24 samples (12 topsoil, 12 biocrust) representing a biological triplicate from each climatic zone. DNA was extracted from 0.25 g of sample using the MoBio PowerSoil Isolation kit according to the manufacturer's instructions. Samples were eluted in DNase- and RNase-free UltraPure Water (ThermoFisher). A sample-free negative control was also run. Nucleic acid purity and yield were confirmed using a Nanodrop 1000 and Qubit Fluorometer. All DNA extractions were performed within two weeks of completing the sampling campaign.

3.3.3. Quantitative PCR

Quantitative polymerase chain reactions (qPCR) were used to estimate total bacterial and archaeal biomass of biocrust and topsoil samples. The 16S rRNA gene was amplified using the degenerate primer pair (515F 5'-154 GTGYCAGCMGCCGCGGTAA-3' and 806R 5'-GGACTACNVGGGTWTCTAAT-3'). A synthetic *E. coli* 16S rRNA gene sequence in a pUC-like cloning vector (pMA plasmid; GeneArt, ThermoFisher Scientific) was used as a standard. PCR reactions were set up in each well of a 96-well plate using LightCycler® 480 SYBR Green I Master Mix. Each sample was run in triplicate and standards in duplicate on a LightCycler® 480 Instrument II (Roche). The qPCR conditions were as follows: pre-incubation at 95°C for 3 min and 45 cycles of denaturation 95°C for 30 s, annealing at 54°C for 30 s, and extension at 72°C for 24 s. 16S rRNA gene copy numbers were calculated based on a standard curve constructed by plotting average Cp values of a serial dilution of the plasmid-borne standard against their copy numbers.

3.3.4. Metagenome sequencing

Metagenomic shotgun libraries were prepared for 12 biocrust samples using the Nextera XT DNA Sample Preparation Kit (Illumina Inc., San Diego, CA, USA). Sequencing was performed on an Illumina NextSeq500 platform with 2 × 150 bp High Output run chemistry. Raw sequence reads in each sample were stripped of adapter

and barcode sequences, then contaminating PhiX sequences were identified and removed using the BBDuk function of BBTools v. 36.92 (https://sourceforge.net/projects/bbmap/) with a kmer size of 31 and hamming distance of 1. Retained read pairs were then quality trimmed using BBDuk with Q >20. After quality filtering and trimming, 283,218,293 paired and 17,363,732 unpaired reads were obtained across the 12 samples. 96% of original sequences passed quality control. Metagenomic sequencing of the 12 topsoil samples was conducted as previously described [342].

3.3.5. Assembly, and binning

Metagenomes were normalised using BBTools (default parameter; bbnorm.sh) and co-assembled using SPAdes v3.13.0 [343] metagenomic mode (--meta) with error correction disabled (--only-assembler). Each assembly was binned using BamM to map reads and MetaBAT2 [344] to produce bins using default parameters. Completeness, contamination, and heterogeneity of each bin were estimated using CheckM [345], with medium- and high-quality bins (completeness >50%, contamination <10%) [346] retained for further analysis. After dereplication, a total of 13 metagenome-assembled genomes (MAGs) were obtained. Each bin was taxonomically assigned according to the Genome Taxonomy Database (GTDB) [290] using GTDB-tk [347].

3.3.6. Community profiling

Community composition profiles were obtained by retrieving single copy ribosomal marker genes from the biocrust and topsoil metagenomes using SingleM [295]. The tool uses Hidden Markov Models (HMMs) searches against unassembled metagenomic reads to generate *de novo* operational taxonomic units. In total, 28 HMM searches were performed against 14 single ribosomal single copy marker genes. The single copy marker gene *rplP* was selected for downstream analysis. This marker was previously identified as a robust means of distinguishing between both closely and distantly related genomes [297]. Sequences were then clustered *de novo* into operational taxonomic units (OTUs) using a sequence identity threshold of 97%. Taxonomic assignment was assigned based on the Genome Taxonomy Database (GTDB) [348].

Community richness and beta diversity was calculated using Phyloseq [299] and R package VEGAN [349]. To account for differences in richness between samples, all sequences were rarefied to the minimum sequencing depth. Observed richness and abundance-based estimated richness (Chao1) were calculated. A permutational analysis of variance (PERMANOVA) was performed to test for significant differences in community structure between soil types and climatic zones. Beta diversity (Bray-Curtis dissimilarity) was calculated and visualised using a multidimensional scaling ordination (MDS) and beta dispersion tests (PERMDISP) were used to ascertain if observed differences were influenced by dispersion.

3.3.7. Metagenomic contig annotation

Functional annotation of metagenomic unbinned contigs and assembled genomes was performed using a combined approach of homology-based searches against custom protein databases and HMM searches. Open reading frames (ORFs) were first predicted using Prodigal v.2.6.3 [350]. Homology-based searches were performed against 29 manually curated protein databases. These encompassed the genes encoding ATP-citrate lyase (AclB), acetyl-CoA synthase (AcsB), ammonia monooxygenase (AmoA), anaerobic sulfite reductase (AsrA), anaerobic carbon monoxide dehydrogenase (CooS), aerobic carbon monoxide dehydrogenases (CoxL), dissimilatory sulfite reductase (DsrA), flavocytochrome *c* sulfide dehydrogenase (FCC), 4-hydroxybutyryl-CoA synthase of Crenarchaeota (HbsC), 4-hydroxybutyryl-CoA synthase of Thaumarchaeota (HbsT), hydrazine synthase (HzsA), malonyl-CoA reductase of Chloroflexi (Mcr), methyl/alkyl-CoM reductase (McrA), soluble methane monooxygenase (MmoA), periplasmic nitrate reductase (NapA), dissimilatory nitrate reductase (NarG), nitrogenase (NifH), copper-containing nitrite reductase (NirK), cytochrome cd_1 nitrite reductase (NirS), nitrous oxide reductase (NosZ), ammonia-forming nitrite reductase (NrfA), nitrite oxidoreductase (NxrA), particulate methane monooxygenase (PmoA), ribulose 1,5-bisphosphate carboxylase/oxygenase (RbcL), succinate dehydrogenase / fumarate reductase (SdhA / FrdA), sulfur oxygenase/reductase (Sor), thiosulfohydrolase (SoxB), sulfide-quinone oxidoreductase (Sqr), and three hydrogenase classes (NiFe-hydrogenase large subunit, FeFe-hydrogenase catalytic domain, Fe-hydrogenase). DIAMOND mapping

was performed with a query coverage threshold of 80% for all databases, and a percentage identity threshold of 60% (AmoA, PmoA, MmoX, CoxL, HbsT, NxrA, RbcL) or 50% (all other databases) and e-value thresholds of 10^{-20}. HMM searches were performed against Pfam and Tigrfam databases using the *annotate* function of EnrichM v.0.5.0 (https://github.com/geronimp/enrichM), with domain noise-cut-off scores as previously described [351]. Eleven genes encoding subunits of ATP synthase (AtpA), two NADH dehydrogenases (NuoF, NqrF), four terminal oxidases (CcoN, CoxA, CydA, CyoA), two photosystems (PsaA, PsbA), formate dehydrogenase (FdhA), and reductive dehalogenase (RdhA) were searched.

3.3.8. Metagenomic short read annotation

For the functional annotation of short reads, paired-end reads in each sample were stripped of adapter and barcode sequences, then contaminating PhiX and low quality sequences were removed (minimum quality score 20) using the BBDuk function of BBTools v. 36.92 (https://sourceforge.net/projects/bbmap/). Resultant quality-filtered forward reads with lengths of at least 100 bp were searched for the presence of the 43 metabolic marker genes described above using DIAMOND *blastx* algorithm [352]. Specifically, reads were searched against the 32 custom-made reference databases and hits from the 11 HMM searches, using a query coverage of 80% and an identity threshold of either 60% (AmoA, PmoA, MmoX, CoxL, HbsT, NxrA, RbcL) or 50% (all other databases) and a maximum e-value threshold of 10^{-10}. Read counts were normalized to reads per kilobase per million (RPKM) and further normalised against a mean RPKM value estimated from 14 single copy ribosomal marker genes to infer the percentage of the community encoding the gene.

3.3.9. Phylogenetic analysis

Maximum-likelihood phylogenetic trees were constructed to visualize the evolutionary relationships of unbinned and binned contigs of the catalytic subunits of [NiFe]-hydrogenase and RuBisCO (RbcL) compared to reference sequences. Retrieved sequences were aligned to custom databases using ClustalW in MEGA7 [353]. For phylogenetic tree construction, initial trees for the heuristic search were obtained

automatically by applying Neighbour-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. All residues were used and trees were bootstrapped with 50 replicates.

3.3.10. Soil wetting

We simulated rainfall conditions to determine the effects of soil moisture on H_2 oxidation and photosynthetic carbon fixation rates. To do this, we used a custom Perspex collar fitted with a water-draining stainless steel woven mesh (0.17 mm) and a water-catching tray. Collars were sterilised using ethanol. Topsoil and biocrust samples of 5 g were placed in the centre of the mesh surface. Soils were then watered until fully saturated by repeated addition of 1 mL MilliQ water. Once fully saturated, each collar was sealed at the top using cling film to avoid evaporation and left to drain for 24 hrs in the dark.

3.3.11. Gas chromatography

Rates of atmospheric H_2 oxidation by biocrusts and topsoils was measured by gas chromatography. Samples of 5 g were suspended in 120 mL serum vials and left to equilibrate with ambient air (12 h). Vials were then sealed with a butyl rubber septum and amended with H_2 (*via* 1% v/v H_2 in N_2 gas cylinder, 99.999% pure) to achieve headspace concentrations of ~10 ppmv. Sampling commenced immediately after sealing the vial to measure the initial uptake rates. Headspace H_2 mixing ratios in samples were measured by gas chromatography using a pulsed discharge helium ionization detector (model TGA-6791-W-4U-2, Valco Instruments Company Inc.) as previously described [72]. Rates of uptake were measured for all 72 biocrust and 72 topsoil subsamples under both dry and wet conditions. Heat killed samples (two autoclave cycles at 120°C) and blank measurements (empty serum vials) were used as controls to confirm that oxidation occurred due to biotic processes. Concentrations of H_2 in each sample were regularly calibrated against ultra-pure gas standards of known concentrations.

3.3.12. ^{14}C isotope labelling

A radiolabelled carbon dioxide ($^{14}CO_2$) incubation assay was used to measure the photosynthetic capacity of biocrusts and topsoils (dry and wetted). Six biological replicates were pooled for each climatic zone and technical triplicates of 0.25 g were weighed and transferred to 4 ml glass vials sealed with rubber septum lids. For each replicate, a heat killed control (two autoclave cycles at 120°C) was used. Gaseous $^{14}CO_2$ (1% mol.) gas stocks were generated by adding 75 µl of sodium bicarbonate solution ($NaH^{14}CO_3$, Perkin Elmer, 53.1 mCi $nmol^{-1}$) to 75 µl of 10% hydrochloric acid (HCl) solution inside a 4 ml glass vial, which was sealed with a rubber septum lid and incubated for two hours at room temperature. 160 µl of $^{14}CO_2$ gas (1% v/v) was added to each biocrust or topsoil sample using a 1 ml gas tight syringe (SGE Analytical Science), obtaining initial headspace mixing ratios of 400 ppmv $^{14}CO_2$. Both treatment groups were then incubated under either light (40 µmol photons m^{-2} s^{-1} under constant illumination) or dark conditions (covered in aluminium foil) for 96 hours at ~20 °C. To remove any unfixed $^{14}CO_2$, incubated soils were transferred to 12 ml scintillation vials and suspended in 2 ml of 1 M HCl and left to dry in an oven at 50°C. Once dried, 10 ml of scintillation cocktail (EcoLume ™) was added and radioisotope analysis was carried out using a liquid scintillation spectrometer (Tri-Carb 2810 TR, Perkin Elmer precisely) operating at ~95% efficiency. Background luminescence and chemiluminescence were corrected through internal calibration standards.

3.3.13. Statistical analysis

All statistical analysis was carried out in R-studio v3.5.3. Data manipulation and visualization was carried out using the package Tidyverse [354]. For H_2 and ^{14}C oxidation rates, normality of the distribution was confirmed using Shapiro-Wilk test. For normally distributed data, ANOVAs were used to test for significant differences of rates between climatic zones followed by Tukey post-hoc tests to determine significant pairs. A non-parametric Wilcox signed rank test was used to determine differences between dark and light CO_2 fixation rates. To identify significant predictors of H_2 oxidation, a Pearson's correlation matrix was calculated against all H_2 oxidation rates and visualised using a heatmap. Significant predictor variables were selected for the final

model and tested for co-linearity using a variance inflation factor cut-off of 10. Collinear predictors were removed to determine the final subset.

3.4. Results

3.4.1. Biocrusts and topsoils harbour diverse microbial communities which are structured by aridity

We analyzed biocrusts and previously reported topsoils [342] sampled along a 160 km aridity gradient in a north-south direction in Israel. Soil physicochemical analysis confirmed expected environmental variation along the transect, with soil water and organic carbon content respectively dropping from an average of 8.2% and 3.4% in the sub-humid northern zones to 1.1% and 0.11% in the hyper-arid southern zones [342]. We profiled the abundance, richness, and composition of the sampled microbial communities using quantitative PCR and metagenomic sequencing. Bacterial and archaeal cell numbers (estimated from copy numbers of the 16S rRNA gene) sharply decreased across the aridity gradient, with a 200-fold higher cell count in sub-humid compared to hyper-arid samples **(Table S1)**. In contrast, richness (Chao1 based on the single-copy ribosomal marker gene *rplP*) between sites was variable as indicated by the coefficient of variation (CV = 46.39%; Soil % CV= 45.67%), but did not significantly decline with aridity **(Fig. S1; Table S1)**. Beta diversity (measured by non-metric multidimensional scaling analysis of Bray-Curtis distances) revealed that communities were significantly differentially structure between climatic zones ($F = 1.16$, $p < 0.001$), but did not significantly differ between biocrusts and topsoils **(Fig. S1; Table S1)**. Altogether, this suggests that aridity significantly influences the abundance and composition of microbial communities in desert ecosystems, but diverse microbial communities can co-exist even in the most arid sites.

At the phylum-level, the microbial community composition of the samples was similar to that described in most other desert soils [41,326]. Most sequences affiliated with Actinobacteriota (59%), particularly classes Thermoleophilia, Actinobacteria, Acidimicrobiia, and Rubrobacteria, with significant proportions of Proteobacteria (13%), Chloroflexota (7%), Bacteroidota (3%), and Planctomycetota (2%) also detected **(Fig. 1a; Table S1)**. Most of these bacteria are likely to be aerobic chemoheterotrophs, but may also harbour hidden metabolic flexibility [67]. Archaea were less abundant (1.1%) and were primarily affiliated with the ammonia-oxidizing

thaumarchaeota *Nitrosocosmicus* [355]. Cyanobacteria were twofold more abundant in biocrusts (1.7%) compared to topsoils (0.6%) and achieved abundances above 3% in specific biocrust samples from in the semi-arid, arid, and hyper-arid zones (most notably SA.1.B1, 8.1%). Whereas biocrust cyanobacterial communities were dominated by oxygenic cyanobacteria (Oxyphotobacteria), including cosmopolitan biocrust genus *Microcoleus* [321,356], approximately half of the cyanobacterial community in topsoils affiliated with the recently discovered heterotrophic class Melainabacteria [357,358] **(Fig. S2; Table S1)**. In contrast to the other zones, grasses were abundant in the sub-humid zone and are likely to predominate primary production.

3.4.2. Genes encoding chemosynthetic and photosynthetic enzymes are differentially distributed along the aridity

We performed homology-based searches of metagenomes to determine the abundance of different energy and carbon acquisition processes across the desert samples. Consistent with our inferences from the community composition profile, most of the bacteria are predicted to be aerobic organoheterotrophs; reflecting this, NADH dehydrogenases, succinate dehydrogenase, and terminal oxidases were highly abundant **(Fig. 2)**. A large proportion of the community were also predicted to oxidize atmospheric trace gases, with an average of 66% and 30% of community members encoding uptake hydrogenases and carbon monoxide dehydrogenases respectively across the aridity gradient. The abundance of these genes remained relatively consistent between biocrusts and topsoils, suggesting that the potential for these processes is dominant throughout the upper 10 cm of these soils **(Table S2)**. By contrast, fewer microorganisms were predicted to oxidize sulfide (6.5%), thiosulfate (2.2%), ammonia (1.4%), nitrite (0.80%), or methane (0.38%). A significant proportion of the community were also capable of using nitrate as an electron acceptor (12%), whereas capacities for other anaerobic respiration or fermentation processes were negligible **(Fig. 1b; Table S2)**.

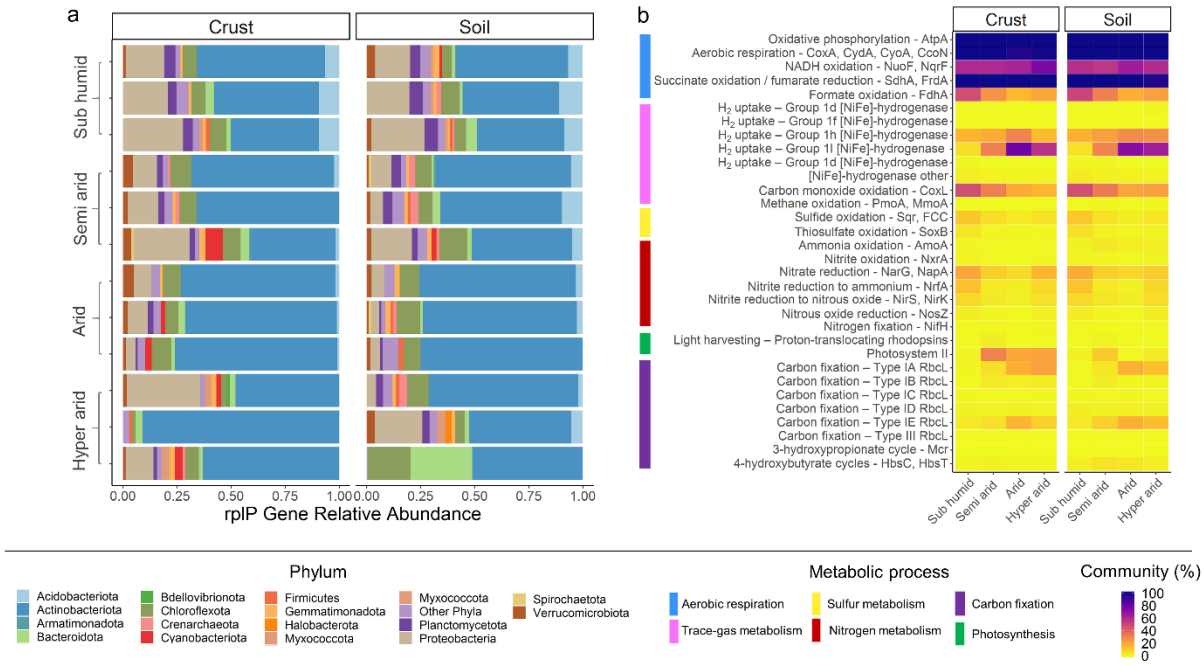


Figure 1. Community composition and metabolic capabilities of biocrust and topsoil microbial communities sampled along the aridity gradient. **(a)** Stacked barchart showing phylum-level bacterial and archaeal community composition. Taxonomic classification is based on the relative abundance of the single-copy ribosomal marker gene *rplP* and follows GTDB taxonomy. Results are shown for each of biological triplicate samples collected per zone. **(b)** Heatmap showing the abundance of genes in the metagenomic short reads. Community percentages were calculated by dividing reads per kilobase millions (RPKM) of metabolic genes to those of single-copy ribosomal protein marker genes. Results are averaged for the biological triplicate samples collected per zone. Undetected genes included those associated with methanogenesis (McrA), anammox (HzsA), Wood-Ljundahl pathway (AcsB), and reductive TCA cycle (AclB).

Of the 38 genes surveyed, those encoding hydrogenases and RuBisCO exhibited the greatest variations in relative abundance across the aridity gradient. All four hydrogenase subgroups known to support atmospheric H_2 oxidation were detected, namely the group 1h, 1l, 2a, and 1f [NiFe]-hydrogenases **(Fig. 1b; Table S2)**, with the first two most abundant. The group 1h enzymes, which are the main clade thought to mediate atmospheric H_2 oxidation in global soils [19,330,331], were abundant in all climatic zones and were encoded by an average of 22% of community members. However, in common with recent observations made in the Mackay Glacier ecotone [339], the

recently discovered group 1l [NiFe]-hydrogenase was the most abundant lineage in oligotrophic soils. Its relative abundance increased between climatic zones (sub-humid 5.9%, semi-arid 30.8%, arid 75.9%, hyper-arid 61.0%), with most bacteria in the arid and hyper-arid zones predicted to encode it. There was also a concomitant increase in the relative abundance of two RuBisCO lineages, Type IA (sub-humid 0.63%, semi-arid 5.4%, arid 17.8%, hyper-arid 17.9%) and Type IE (sub-humid 3.8%, semi-arid 8.0%, arid 16.9%, hyper-arid 12.0%), with the hits from both top genes most closely related to actinobacterial lineages known to encode high-affinity hydrogenases [162,359]. The hydrogenase and RuBisCO results together suggest that, in common with Antarctic desert soils [28], Negev desert bacteria also use electrons derived from atmospheric H_2 to catalyse CO_2 fixation. The abundance of CO dehydrogenases showed the opposite trend, declining from 48% in the sub-humid zones to 19% in the hyper-arid zone, in agreement with previous studies showing CO oxidation rates are strongly correlated with soil organic content [360,361] **(Fig. 1b, Table S3)**.

Genes encoding photosystems and RuBisCO lineages associated with oxygenic photosynthesis were also detected in the metagenomes. Most notably, various type IB RuBisCO lineages were detected that are homologous to cyanobacteria (*Microcoleus* spp., *Leptolyngbya* spp.), chlorophytes (*Myrmecia israeliensis*), and bryophytes (*Pseudocrossidium* spp.) known to be abundant in Israel desert biocrusts [362–364]. Reflecting observed cyanobacterial distributions **(Fig. 1a)**, type IB RuBisCO was in moderate abundance across all climatic zones (1.8%) and enriched in biocrusts (2.3%) compared to topsoils (1%) **(Fig. 1b; Table S3)**. Isolated increases in abundance of 3.1% and 9.1% corresponded to sites with a high count of cyanobacterial sequences (AR1.C2 and SA1.B1) in the arid and semi-arid zone, respectively **(Fig. S2; Table S1)**. Together with the community analysis, these results highlight that there is some potential for photosynthesis even in arid and hyper-arid regions, but phototrophs have a lower and more variable distribution than hydrogenotrophs.

3.4.3. Actinobacteria encode diverse uptake hydrogenase and RuBisCO enzymes across the aridity gradient

To gain a more detailed perspective of the mediators of photosynthesis and chemosynthesis, we co-assembled and binned the metagenomes of the biocrusts and topsoils sequenced from each climatic zone. This resulted in the recovery of 13 medium-quality [346] metagenome-assembled genomes (MAGs) that affiliated with the phyla Actinobacteriota (8), Cyanobacteria (3), Proteobacteria (1), and Thermoplasmatota (1) **(Table S3)**. In line with expectations, genome annotations suggested that the actinobacterial and proteobacterial MAGs are capable of aerobic organotrophic respiration, whereas the cyanobacterial MAGs are oxygenic phototrophs. Three MAGs from the arid zone, from classes Actinobacteria and Thermoleophila, encoded a group 1l [NiFe]-hydrogenase; one of these MAGs (bin001) recovered from the arid zone co-encoded this hydrogenase with a type IE RubisCO, suggesting it can mediate hydrogenotrophic carbon fixation. Photosynthetic type IB RuBisCO was also detected in a MAG (bin007) from a *Microcoleus* species recovered from the sub-humid zone **(Table S3)**. In a further indication of the metabolic versatility of desert actinobacteria, we recovered a *Mycobacterium* MAG (bin003) from the semi-arid zone that encoded a key enzyme for aerobic sulfide oxidation (sulfide-quinone oxidoreductase).

Maximum-likelihood phylogenetic trees were constructed of contigs encoding catalytic subunit sequences of [NiFe]-hydrogenase **(Fig. 2a)** and RuBisCO **(Fig. 2b)**. Across the metagenomes, large subunit genes encoding 24 group 1h and 87 group 1l [NiFe]-hydrogenases were recovered from the unbinned and binned contigs **(Fig. 2a)**. Most of these sequences were most closely related to actinobacterial reference genomes, though three of the five major clades of the group 1l [NiFe]-hydrogenases encoded in the metagenomes lacked any pure culture relatives; one of these clades was represented by a Thermophilia MAG (bin004) and hence likely originates from the most abundant class-level lineage in the Negev desert **(Table S1)**, whereas two other deep-branching clades were most closely related to halotolerant archaea and are of uncertain affiliation **(Fig. 2a)**. Altogether, this adds further evidence that the actinobacterial majority in Negev desert soils and biocrusts are capable of H_2 oxidation. Other hydrogenase groups associated with aerobic uptake (group 2a), sensing (group 2c), and carbon fixation (group 3d) were sparsely represented with a single sequence each.

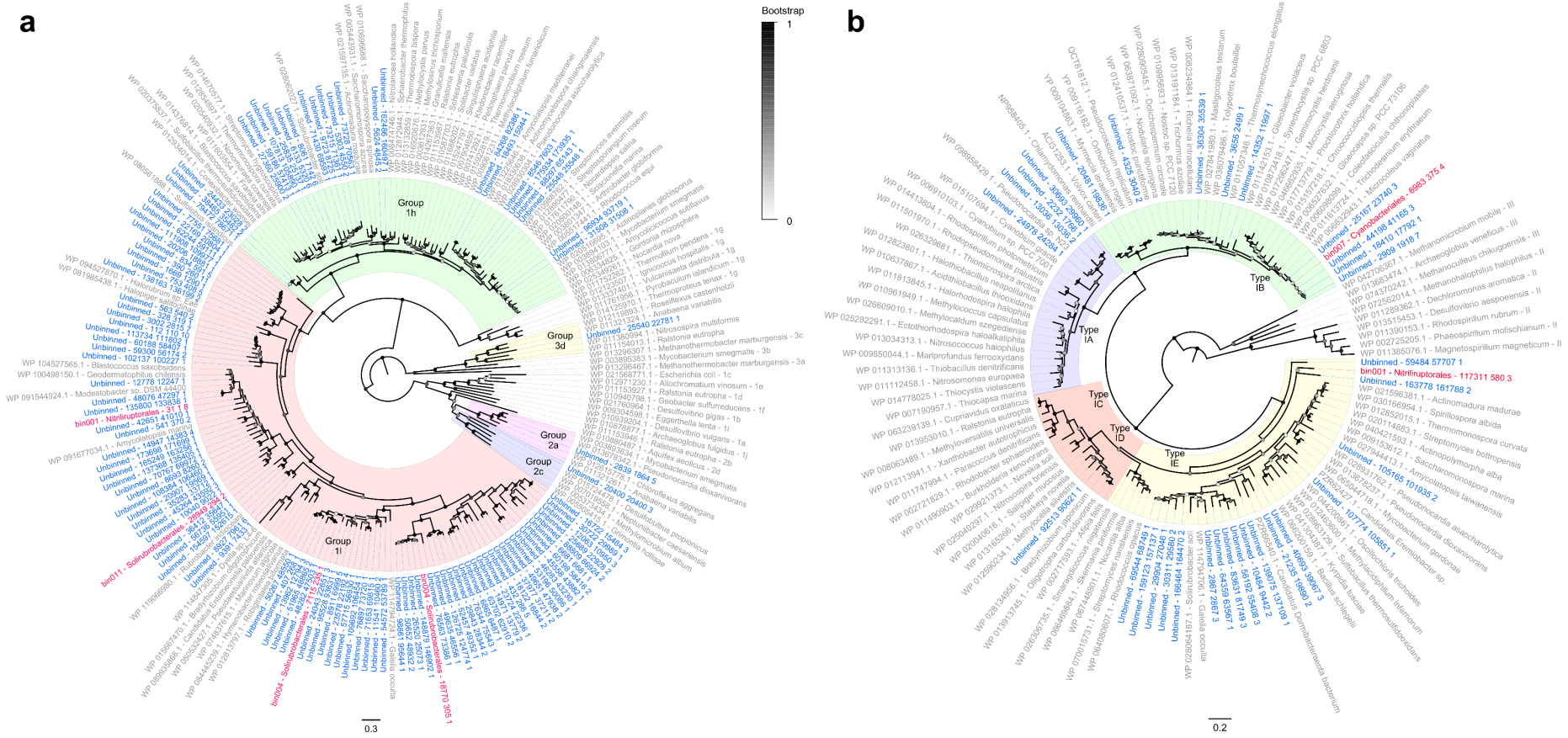


Figure 2. Maximum likelihood radial phylogenetic trees showing sequence diversity and taxonomic distribution of enzymes responsible for H_2 oxidation and carbon fixation. **(a)** Phylogenetic tree of [NiFe]-hydrogenases, with a focus on the group 1h and 1l high-affinity uptake hydrogenases to which most binned and unbinned sequences affiliated with. **(b)** Phylogenetic tree of RuBisCO, with a focus on the Type IB (phototroph-type) and Type IE (lithotroph-type) enzymes that the majority of binned and unbinned

sequences affiliated with. Trees show hits to genome bins (red) and unbinned contigs (black) relative to reference amino acid sequences (grey).

The RuBisCO phylogenetic tree confirmed the presence of unbinned sequences related to actinobacterial type IE sequences and cyanobacterial, microalgal, and plant type IB sequences **(Fig. 2b)**. Relatively few type IA RuBisCO sequences were detected in the assembled reads compared to unassembled reads, and all were closely related (88-95% sequence identity) to reference sequences from the facultatively hydrogenotrophic genus *Pseudonocardia* [365]. Altogether, these findings suggest that phototrophs and hydrogenotrophs are the dominant primary producers across the aridity gradient.

3.4.4. Differential activities of chemosynthetic and photosynthetic microorganisms across the aridity gradient

Having confirmed the metabolic potential for atmospheric H_2 oxidation *via* the group 1h and 1l [NiFe]-hydrogenase and photosynthetic carbon fixation, we used two biogeochemical measurements to validate that microorganisms mediate these processes in biocrusts and topsoils: (i) gas chromatography to measure H_2 oxidation and (ii) ^{14}C incorporation to measure CO_2 fixation rates. These rates were measured under dry conditions (soils as collected) and wet conditions (24 hours after simulated rainfall).

All biocrust and topsoil samples consumed H_2 to sub-atmospheric levels **(Fig. S3; Table S4)**. This occurred at relatively slow rates under dry conditions and increased 45-fold upon hydration **(Fig. 3a & 3b)**. H_2 oxidation rates per gram of sample significantly increased across the aridity gradient for biocrust samples, but were more variable for topsoil samples **(Fig. 3a & 3b)**. Given the large decrease in microbial biomass across the aridity gradient, oxidation rates were subsequently normalised to 16S rRNA gene copy number **(Table S1)**. Biomass-normalized H_2 oxidation rates increased by 500-fold across the aridity gradient for both biocrusts ($F = 3.14$, $p < 0.0001$) and topsoils ($F = 55.58$, $p < 0.0001$) (**Table S4, Fig. S4**). Simulated rainfall conditions amplified this response. On average, oxidation rates increased from 1.3 and 1.4 nmol hr^{-1} g^{-1} 16S $copies^{-1}$ for crusts and topsoils respectively under dry conditions, to 31.6 and 30.0 nmol hr^{-1} g^{-1} under wet conditions. Linear models were used to determine the relationships between edaphic characteristics and H_2 oxidation

rates. After accounting for co-linearity, a subset of predictor variables included the aridity index, total organic carbon, carbon / nitrogen ratio, and sodium concentration **(Fig. S5)**. The relationship between H_2 oxidation rate and total organic carbon content was especially strong (crust R^2 = 0.80, topsoil R^2 = 0.55; **Fig. S5)**.

We additionally measured CO_2 fixation rates of both crust and topsoil under light and dark conditions. We found that both light and dark CO_2 fixation was virtually absent under dry conditions **(Fig. 3c & 3d; Table S5)**. Following simulated rainfall, light and dark CO_2 fixation was observed across all samples. On average, CO_2 fixation occurred at significantly higher rates under light compared to dark conditions (p < 0.001 for crusts, p < 0.01 for soils), and photosynthetic capacity of crusts was approximately 12-fold higher (0.23 nmol g^{-1} day^{-1}) than topsoils (0.0015 nmol g^{-1} day^{-1}) **(Fig. 3c & 3d; Fig. S4)**. There was a significant decline in topsoil photosynthetic CO_2 fixation rates with increasing aridity (F = 71.04, p < 0.0001) from sub-humid (0.04 nmol g^{-1} day^{-1}) to semi-arid (0.01 nmol g^{-1} day^{-1}), arid (0.005 nmol g^{-1} day^{-1}) and hyper-arid (0.002 nmol g^{-1} day^{-1}) **(Fig. 3d)**. In contrast, biocrust CO_2 fixation rates did not consistently vary with aridity and peaked in the semi-arid zone (0.62 nmol g^{-1} day^{-1}) **(Table S5)**. This likely reflects the patchy distribution of Cyanobacteria in crust samples inferred from the metagenomic analyses of community composition and function **(Fig. 1)**.

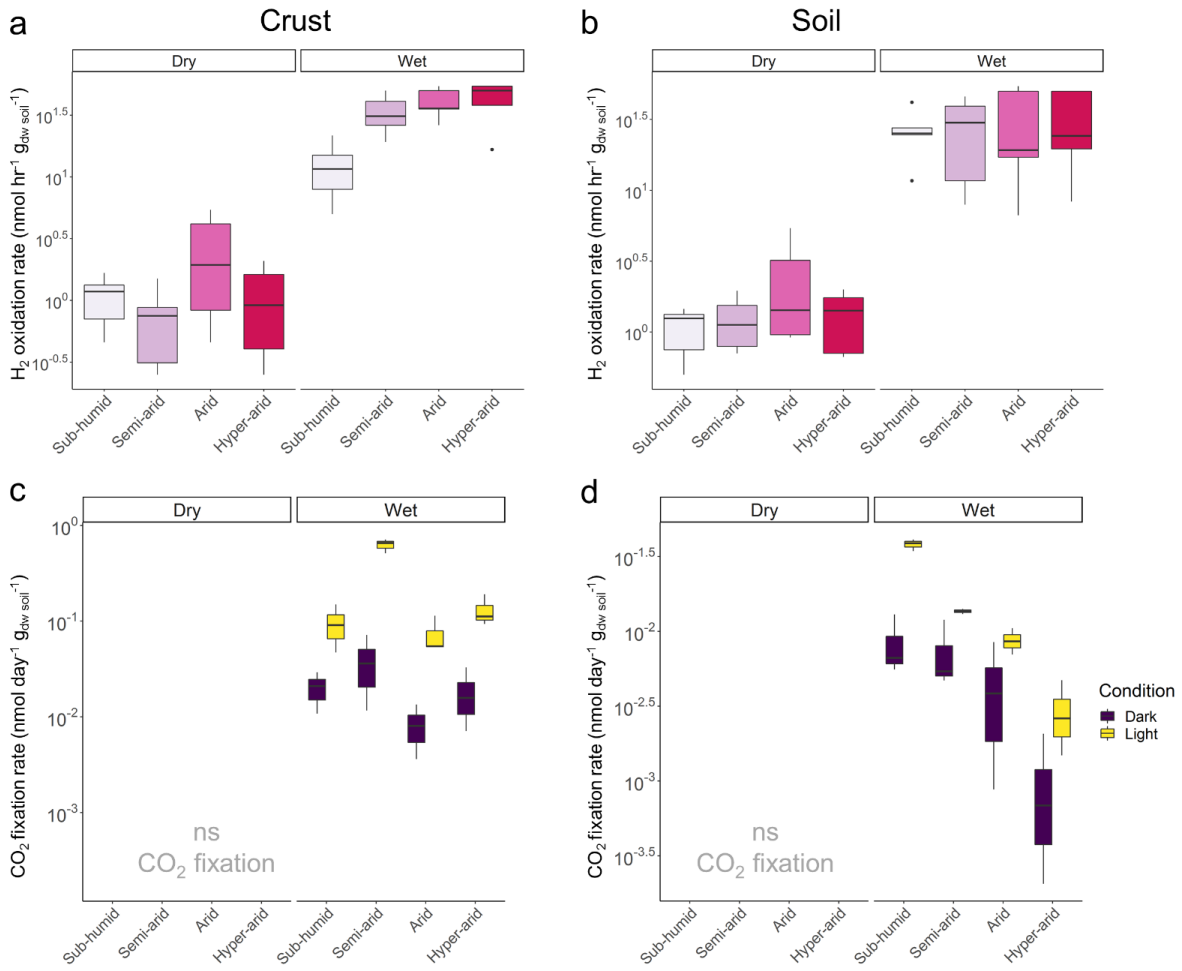


Figure 3. Rates of chemosynthetic and photosynthetic processes of biocrusts and topsoils collected along the aridity gradient. **(a & b)** Rates of H_2 oxidation measured by gas chromatography. **(c & d)** Rates of carbon fixation measured by incorporation of ^{14}C-labelled CO_2. Activities were measured in *ex situ* microcosms under dry and wet conditions. Centre values show median, boxes upper and lower quartiles and whiskers minimum and maximum values. Panel **a-b** represent biological triplicates from two sites per climatic zone and panel **b-c** are technical triplicates from a pooled sample from each zone. Biomass-normalized rates are shown in **Figure S5**.

3.5. Discussion

In this work we demonstrate how two key microbial energy conservation strategies, photosynthesis and chemosynthesis, interact with aridity. Building on previous findings in Antarctic deserts [28,339], we combine metagenomic evidence and biogeochemical measurements to demonstrate that atmospheric H_2 oxidation is a key microbial process mediated by dominant taxa in hot desert soils. We found that the determinants of atmospheric chemosynthesis were widespread across the aridity gradient and were particularly abundant in the oligotrophic interior of arid and hyper-arid regions. This is evident from the abundance and diversity of genes associated with aerobic H_2 oxidation (group 1h and 1l [NiFe]-hydrogenases) and chemosynthetic carbon fixation (type IA and IE RuBisCO genes). We provide phylogenetic evidence that the most dominant bacterial classes in desert soils, Thermoleophilia and Actinobacteria, encode these genes. Biogeochemical studies confirmed that these communities actively consume atmospheric H_2, with biomass-normalised rates greatly increasing across the aridity gradient. Measurable activity occurred even under dry conditions, though hydration accelerated rates. Altogether, this suggests that atmospheric H_2 oxidation is a critical trait mediated by the dominant bacteria in oligotrophic desert ecosystems.

Metagenomic and biogeochemical measurements revealed some potential for oxygenic photosynthesis across the aridity gradient. In line with expectations [222,366], various cyanobacteria were detected, such as the keystone crust-forming taxon *Microcoleus*, with abundance peaking in biocrust samples particularly from the semi-arid zone [222,367]. Correlating with their presence were genes supporting the light reactions (photosystems) and dark reactions (type IB RuBisCO) of photosynthesis. However, radiolabelling studies confirmed that photosynthetic processes are virtually absent under dry conditions and are only activated once the electron donor water is introduced. Thus, in contrast to the hydrogenotrophic community, it is likely that these primary producers only become active when the desert is hydrated. Moisture is provided in the Negev desert on a regular basis by dewfall and more occasionally from rainfall [368,369]. The significance of dewfall in activating photosynthesis is unclear; studies on cyanobacterial biocrusts suggest that photosystem II is only activated by

heavy fog events, which account for ~5% of the total annual dewfall events in the region [356,370]. However, our studies agree with a large body work showing simulated rainfall greatly stimulates activities of phototrophs in biocrusts and topsoils alike [356,370]. Altogether, these findings suggest that phototroph abundance and activity in the Negev desert is highly variable across spatial and temporal scales.

These findings have unexpected ramifications for understanding the microbial composition and function of biological soil crusts. It is generally thought that the dominant primary producers in biocrusts are cyanobacteria and phototrophic eukaryotes, which supply organic carbon to organoheterotrophs. However, we show here that the abundance of these phototrophs and their functional genes is variable and often low, in contrast to the consistently abundant hydrogenotrophic actinobacteria. Reflecting this, the exclusive chemosynthetic type IE RuBisCO was on average 3.9-fold more abundant in the biocrusts than the exclusively photosynthetic type IB RuBisCO. Moreover, contrary to the paradigm that only biocrusts are only active when wet [321,366], we observed substantial levels of H_2 oxidation but not photosynthesis even in dry biocrusts. It should be noted that the abundance of the cyanobacterial community is much lower than in many previously reported biocrusts [371–373]. This reflects that the dry loess soils of the Negev arid and hyper-arid zones are covered by biocrusts that are relatively thin and lightly hued compared to other regions. Nevertheless, Actinobacteriota generally have high abundances in desert biocrusts and are also the dominant taxa in certain Mojave and Tengger biocrusts [374–376]. Altogether, these novel findings justify further studies to resolve the relative contributions of photosynthetic and chemosynthetic microorganisms in the establishment, maintenance, and productivity of biocrusts.

Integrating these considerations, our findings metabolic flexibility underlies the dominance of the actinobacterial lineages in desert biocrusts and topsoils. These organoheterotrophs also take advantage of transient hydration events, likely using exudates released by phototrophs and necromass released through osmotic shock to increase respiration rates and accumulate macromolecular stores [377–383]. However, their capacity to conserve energy through trace gas oxidation independently of organic inputs will confer a survival advantage during subsequent desiccation and starvation. While atmospheric H_2 is likely to be the main energy source sustaining these bacteria,

our metagenomic analysis indicates carbon monoxide and sulfur compounds could also be significant energy sources. These inferences are well-supported by pure culture studies showing sporulating and non-sporulating actinobacterial species alike can survive carbon starvation by utilising atmospheric trace gases [71,330–332,334,384]. Moreover, the capacity to use trace gases to fix CO_2 is likely to enable cells to maintain biomass levels and even potentially sustain slow growth. Altogether, it can be inferred that trace gas oxidation confers a selective advantage for metabolically flexible organoheterotrophs, by providing means of acquiring alternative energy donors to sustain basal energy requirements during dormancy and in some cases a mixotrophic means of acquiring biomass. Given the widespread taxonomic distribution of this metabolism [70–72,162,191,202], it is likely that other bacterial and archaeal lineages in these desert soils also possess hidden metabolic flexibility.

3.6. Footnotes

Data availability statement: All new sequencing data and MAGs will be deposited to the Sequence Read Archive under BioProject accession number PRJNAxxxxxx.

Code availability statement: No custom code was used for this study. All analyses were carried out using publicly available resources.

Acknowledgements: This study was funded by a Monash University & Ben Gurion University of the Negev Seed Fund (awarded to C.G., O.G.). It was supported by an ARC DECRA Fellowship (DE170100310; awarded to C.G.), and a Monash University PhD Scholarship (awarded to S.B.). We thank Ya-Jou Chen and Thanavit Jirapanjawat for technical advice, Capucine Baubin, Dimitri Meier, and Stefanie Imminger for field assistance.

Author contributions: C.G. conceived and supervised this study. S.K.B. and O.G. performed field work. S.K.B. performed field sampling, laboratory work, and bioinformatic analysis. D.W.W. performed assembly and binning of metagenomes. O.G. and P.H. provided logistical and sequencing support. S.K.B. and C.G. wrote the paper with input from all authors.

Supplementary material: Supplementary figures and tables are found in Appendix B.

Ethics declaration: The authors declare no conflict of interest.

Chapter 4

Trace gas oxidizers are globally dominant and active members of soil biomes

Sean K. Bay[1,2], Xiyang Dong[3], James A. Bradley[4,5], Pok Man Leung[1,2], Thanavit Jirapanjawat[1,2], Stefan K. Arndt[6], Perran L.M. Cook[7], Douglas LaRowe[8], Philipp A. Nauer[7], Eleonora Chiri[1,2 *], Chris Greening[1,2 *]

[1]School of Biological Sciences, Monash University, Clayton, VIC 3800, Australia
[2]Department of Microbiology, Biomedicine Discovery Institute, Monash University, Clayton, VIC 3800, Australia
[3]School of Marine Sciences, Sun Yat-Sen University, Zhuhai 51900, China
[4]School of Geography, Queen Mary University of London, Bethnal Green, London E1 4NS, United Kingdom
[5]Interface Geochemistry, GFZ German Research Centre for Geosciences, 14473 Potsdam, Germany.
[6]School of Ecosystem and Forest Sciences, University of Melbourne, Richmond, VIC 3121, Australia
[7]School of Chemistry, Monash University, Clayton VIC 3800, Australia
[8]Department of Earth Sciences, University of Southern California, Los Angeles, CA 90089, United States

Correspondence:
* Assoc Prof Chris Greening, Monash University, Department of Microbiology, Innovation Walk, Clayton, VIC 3800, Australia
Email: chris.greening@monash.edu, Ph: +61 451 085 339, ORCID: 0000-0001-7616-0594
* Dr Eleonora Chiri, Monash University, Department of Microbiology, Innovation Walk, Clayton, VIC 3800, Australia
Email: eleonora.chiri@monash.edu

In Revision: *Nature Microbiology*

4.1. Abstract

Soil microorganisms globally are thought to be sustained primarily by organic carbon sources [106]. Certain bacteria also consume inorganic energy sources such as trace gases [72,191,330,385], but they are presumed to be rare community members [361,386,387], except within extreme desert ecosystems [28,388]. Here we combined metagenomic, biogeochemical, and modelling approaches to determine how soil microbial communities conserve energy and acquire carbon. Analysis of 40 metagenomes and 757 derived genomes indicated over 70% of soil bacterial taxa can consume inorganic energy sources. Bacteria from 19 phyla encoded enzymes to aerobically respire the trace gases hydrogen and carbon monoxide. We validated *in situ* and *ex situ* that communities within soil profiles from diverse biomes rapidly oxidized both gases below atmospheric concentrations. Thermodynamic modelling indicated that bacteria consume trace gases at sufficient rates to meet their maintenance needs, with a diverse subset also capable of autotrophic growth. In a further demonstration of the hidden diversity of trace gas oxidizers, we also identified a fourth phylum [204,389] capable of aerobic methanotrophy (Gemmatimonadota). Trace gas oxidation confers a selective advantage in soil ecosystems, where availability of preferred organic substrates limits microbial growth [13]. The observation that inorganic energy sources sustain most soil bacteria has broad implications for understanding atmospheric chemistry and microbial biodiversity in a changing world.

4.2. Introduction

Bacteria mediate key supporting and regulatory services in soil ecosystems worldwide [106,390]. Culture-independent surveys have shown that soils harbour abundant and diverse bacterial communities [10,35], with most cells thought to be in dormant states due to pressures such as resource limitation [210,211]. Most soil bacteria use organic carbon derived from vegetation and other inputs as energy and carbon sources [106]; reflecting this, some isolates from all nine of the most abundant phyla in soils are aerobic organotrophs [36,391]. However, various bacteria have been characterised that use inorganic energy sources, including molecular hydrogen (H_2), carbon monoxide (CO), sulfide, thiosulfate, ammonia, and nitrite [173,392]. Such organisms use electrons derived

from these compounds to support aerobic respiration and, in the case of autotrophs, carbon dioxide (CO_2) fixation through various pathways. While these lithotrophs are ecologically and biogeochemically important [106,393], they are generally thought to be minor community members compared to organotrophs [106,386].

Emerging evidence suggests that the trace gases H_2 and CO are particularly important energy sources for bacterial growth and persistence. Both gases are ubiquitously available in the atmosphere at average global concentrations of 0.53 and 0.10 parts per million (ppmv) respectively, with local CO levels greatly varying due to natural and anthropogenic processes [165,394,395]. These compounds are also produced within soils: H_2 through biological nitrogen fixation and fermentation [335,395,396], CO primarily through abiotic thermal and photochemical processes [394,397]. Microorganisms have evolved specialised metalloenzymes, called hydrogenases and carbon monoxide dehydrogenases, to oxidize these gases, including below atmospheric concentrations [71,330,331]. Some bacteria sustain autotrophic or mixotrophic growth by using electrons derived from H_2 and CO to drive aerobic respiration and carbon fixation [76,162,398]. In addition, bacterial cultures from four phyla (Actinobacteria, Proteobacteria, Acidobacteria, Chloroflexi) have been shown to switch from growth on organic carbon to persistence on these trace gases in response to carbon starvation [71,72,191,330,331,385]. This metabolic flexibility is advantageous in environments where resource availability is low or variable [71,72,331,335].

Previous soil surveys reported that atmospheric H_2 and CO oxidizers are members of the rare biosphere, each comprising just 1% in abundance of the total bacterial community, similarly to methane-oxidizing bacteria (methanotrophs) [361,386,387]. Extreme environments, such as Antarctic deserts, are notable exceptions where trace gas oxidisers are major primary producers in soils with otherwise low photosynthetic input [28,388,399]. However, a holistic understanding of how global soil communities meet energy and carbon needs is lacking [62]. Here, we combined metagenomic, biogeochemical, and modelling approaches to understand the metabolic adaptations of soil bacteria, including the role of different strategies for energy conservation (organotrophy, lithotrophy, phototrophy) and carbon acquisition (heterotrophy vs autotrophy). We hypothesise that, unlike specialist processes such as methanotrophy,

aerobic H_2 and CO metabolism is a broad metabolic strategy utilised by much of the community.

4.3. Materials and Methods

4.3.1. Site description and sampling

Soil sampling was conducted in four sites within Australia that differed in climate, vegetation cover and soil type: (i) wetland (-37.908°, 145.139°; Jock Marshall Reserve, Clayton, VIC; JMR), (ii) forest (-37.446°, 144.470°; Wombat State Forest, VIC; WSF), (iii) grassland (-37.927°, 145.313°; Don Bosco Grassland, Lysterfield, VIC; DBG), and (iv) dryland (-23.874°, 133.967°; 25.5 km south of Alice Springs, NT; ASD). The sites were sampled on December 19 2018, January 5 2019, January 21 2019, and January 29 2019 respectively. At each site (~100 m^2), four sampling plots (~1 m^2) were selected with ~50m buffers to structures such as roads and foot paths. At each plot, *in situ* depth-resolved gas concentrations profiles and soil-atmosphere fluxes were measured. In addition, a core was used to collect soil samples at four depths (0-5, 5-10, 15-20, and 25-30 cm) for *ex situ* oxidation measurements, physicochemical analysis, and DNA extractions to perform quantitative polymerase chain reactions (qPCR) and metagenomic sequencing. All sampling occurred during daylight hours, gas samples were processed within 48 h of collection, and soil samples were incubated for *ex situ* oxidation measurements within 24 h of collection.

4.3.2. Sampling and measurement of soil gas profiles

Depth-resolved *in situ* gas concentrations of H_2, CO, and CH_4 were measured using stainless steel capillaries fitted with a Luer Lock and Discofix three-way Stopcock. Prior to sampling, each capillary was inserted into the soil at depth intervals of 2, 4, 6, 8, 10, and 16 cm. All samplers were installed simultaneously, arranged in a hexagonal grid (diameter ~20 cm), and left to equilibrate for ~30 min. A gas sample was also collected at the soil-atmosphere interface (0 cm). All gas samples were stored in evacuated 12 mL glass exetainers sealed with rubber septum lids and analyzed using gas chromatography. Gas concentrations in samples were measured by gas

chromatography using a pulsed discharge helium ionization detector (model TGA-6791-W-4U-2, Valco Instruments Company Inc.) as previously described [72]. Samples were calibrated against H_2, CO and CH_4 standards which were prepared using ultra-pure concentrations of each gas (1% in N_2 gas cylinder, 99.999% pure, Air Liquide Australia) down to the limit of quantification (H_2 20 ppbv, CO 90 ppbv, CH_4 500 ppbv. Pressurized air (Air Liquide Australia) with known trace gas concentrations was used as an internal reference standard.

4.3.3. Measurement of soil-atmosphere gas fluxes

In situ soil-atmosphere fluxes of H_2, CO, and CH_4 were measured using static flux chambers. The chamber consisted of a 20 × 15 cm polyvinylchloride (PVC) pipe with a threaded access cap. The cap was fitted with a gastight O-ring, two butyl rubber septa (one for air sampling and one for a thermometer), and an axial fan on the inside to promote internal mixing. At each plot, a 10 × 14.8 cm PVC base collar was inserted ~5 cm into the soil and left to equilibrate for ~30 minutes prior to sampling to reduce lateral gas fluxes. Once the chamber was fitted over the collar, the cap was closed and the axial fan was started. Three consecutive gas measurements were taken at approximately one-minute intervals, followed by either four or five measurements at approximately five-minute intervals. For each measurement, 15 mL of gas was collected using a gas tight 5 mL Terumo Syringe fitted with a Luer Lock and Discofix three-way Stopcock and measured by gas chromatography as described above. Control gas measurements of ambient air were taken directly before, during, and after sampling. The temperature of the chamber, ambient air, and soil were monitored throughout. Concentrations were then converted to nmol m^{-3} at ambient pressure and temperature using the ideal gas law. Atmospheric flux (J_{atm}) was calculated from the concentration gradient at chamber deployment using a linear and an exponential model fit [400] for each chamber measurement; the best model was chosen according to the lower Akaike information criterion (AIC). Conservative flux detection limits incorporating errors of sample handling and storage were calculated using mean and standard deviation of air samples [401].

4.3.4. Soil sampling and physicochemical analysis

Four soil cores of 30 cm depth were collected from each site (16 total). Cores (*diameter* = 5cm) were carefully extracted and immediately segmented at 5 cm depth intervals before being transferred into 50 mL Falcon tubes. At four different depth intervals (0-5, 5-10, 15-20, 25-30 cm), a subset of 10 g of soil was frozen at -20°C for community DNA extraction and a subset of 5 g was used within 24 hours for gas chromatography studies. Additional surface soil was collected using a 10 × 5 cm bulk density ring to estimate bulk density and soil water content, which was measured gravimetrically using a drying oven at 140°C. For soil chemistry analysis, surface soils samples (0-5 cm) from each plot were pooled to form one representative composite sample per site and sent to the Environmental Analysis Laboratory, Southern Cross University. In total, 21 separate soil chemical parameters were selected for analysis, based on commonly reported drivers of soil microbial composition globally. These included: soil acidity (pH), electrical conductivity (EC), effective cation exchange capacity (ECEC), total organic carbon, total nitrogen, sodium (Na), sulfur (S), phosphate (P), potassium (K), nitrate (NO_3^-), and ammonium (NH_4^+), as well as bioavailable manganese (Mn), copper (Cu), zinc (Zn), boron (B), aluminium (Al), iron (Fe), and silicon (Si). In addition, particle size analysis by hydrometry was performed to estimate the percentage of gravel, sand, silt, loam, and clay **(Table S1)**. Each chemical parameter was calculated following Rayment and Lyons methods [282].

4.3.5. *Ex situ* oxidation rates

To determine the capacity of the soils to oxidise atmospheric trace gases, 5 g soil from the four depth intervals (0-5, 5-10, 15-20, 25-30 cm) were placed in 120 mL serum vials. The headspace was repeatedly flushed with air from a pressurized cylinder (Air Liquide, Australia) to achieve headspace mixing ratios reflecting atmospheric levels (~0.5 ppmv H_2, ~0.6 ppmv CO, ~1.8 ppmv CH_4). Sampling commenced immediately after sealing the vial and headspace samples of 2 mL were taken at 10 min intervals for 40 minutes. While this timecourse sufficiently capture oxidation rates for most samples, additional gas samples were collected every 24 hrs for up to nine days to capture oxidation rates of soils that mediated slow gas consumption. Gas concentrations were measured by gas chromatography as described above. Heat killed soils (two 30-minute autoclave cycles at 121°C) and blank measurements (empty serum vials) were used as controls, confirming trace gas oxidation occurred due to

biotic processes. Given the low capacity of dryland soils to oxidize atmospheric trace gases, we simulated a rainfall event to determine whether soil hydration enhances activity. To do this, we used a custom Perspex collar fitted with a water-draining stainless steel woven mesh (0.17 mm) and a water-catching tray. Collars were sterilised using ethanol and soils (5 g) were placed in the centre of the mesh surface. Soils were watered until fully saturated by repeated addition of MilliQ water. Once fully saturated, each collar was sealed at the top using clingfilm to avoid evaporation and left to drain for 24 hrs in the dark. Once drained, soil samples were transferred using sterile techniques into a 120 mL serum vial and gas oxidation was measured as described in the previous section. Rates were calculated as described above using the initial four times points of each measurement. The data was tested for normality using a Shapiro-Wilk test. To test for significant difference in oxidation rates between ecosystems and depths, a non-parametric Kruskal-Wallis test was used. This was followed by a pairwise Wilcox Rank Sum test to test significant relationships between pairs.

4.3.6. Community DNA extraction

At each soil depth sampled per site, soils from the four plots were pooled together. For each of the 16 resultant samples, total community DNA was extracted using 0.25 g soil. Extractions were performed using the MoBio PowerSoil Isolation kit according to the manufacturer's instructions. Samples were eluted in DNase- and RNase-free UltraPure Water (ThermoFisher). A sample-free negative control was also run. Nucleic acid purity and yield were measured using a NanoDrop ND-1000 spectrophotometer and a Qubit Fluorometer 2.0.

4.3.7. Quantitative PCR

Quantitative polymerase chain reactions (qPCR) were used to estimate total bacterial and archaeal biomass. The 16S rRNA gene was amplified using the degenerate primer pairs 515F (GTGYCAGCMGCCGCGGTAA) and 806R (GGACTACNVGGGTWTCTAAT) [402]. A synthetic *E. coli* 16S rRNA gene sequence in a pUC-like cloning vector (pMA plasmid; GeneArt, ThermoFisher Scientific) was used as a standard. PCR reactions were set up in each well of a 96-well plate using a

LightCycler 480 SYBR Green I Master Mix. Each sample was run in triplicate and standards in duplicate on a LightCycler 480 Instrument II (Roche). The qPCR conditions were as follows: pre-incubation at 95°C for 3 min and 45 cycles of denaturation 95°C for 30 s, annealing at 54°C for 30 s, and extension at 72°C for 24 s. 16S rRNA gene copy numbers were calculated based on a standard curve constructed by plotting average Cp values of a serial dilution of the plasmid-borne standard against their copy numbers.

4.3.8. Sequencing, assembly, and binning of Australian metagenomes

Metagenomic shotgun libraries were prepared for the 16 samples using the Nextera XT DNA Sample Preparation Kit (Illumina Inc., San Diego, CA, USA). Sequencing was performed on an Illumina NextSeq500 platform with a 2 × 150 bp High Output run. Raw reads derived from the 16 metagenome libraries were quality-controlled by clipping off primers and adapters then filtering out artifacts and low-quality reads using Read_QC module within the metaWRAP pipeline [403]. For each ecosystem, the four quality-controlled metagenomes were co-assembled using MEGAHIT v1.1.3 [404] (default parameters) and individually assembled using SPAdes v3.13.0 [343] (metaSPAdes mode, default parameters), producing five assemblies for each ecosystem. Short contigs (<1000 bp) were removed. Each assembly was binned using the binning module within the metaWRAP [403] pipeline (MetaBAT [405], MetaBAT2 [344] and MaxBin2 [406]). For each assembly, the three bin sets were then consolidated into a final bin set with the bin_refinement module of metaWRAP [403]. For each ecosystem, the final bin sets were aggregated and de-replicated using dRep [407] (-comp 50 -con 10 options). Completeness, contamination, and heterogeneity of each bin were estimated using CheckM [345], with medium- and high-quality bins (completeness >50%, contamination <10%) [346] retained for further analysis. After dereplication, a total of 93 metagenome-assembled genomes (MAGs) were obtained for the four ecosystems. Each bin was taxonomically assigned according to the Genome Taxonomy Database (GTDB; release 04-RS89) [290] using GTDB-tk [347].

4.3.9. Assembly and binning of global public soil metagenomes

A total of 24 previously sequenced metagenomes from eight different soil ecosystems (three metagenomes each ecosystem) **(Table S2)** were downloaded from the Integrated Microbial Genomes database [408] and the NCBI Sequence Read Archive (SRA) [409]. These comprised: Barrow Environmental Observatory site, Barrow, Alaska, USA (Arctic Tundra, BEO); St. Claude, Quebec, Canada (Agricultural Land – Crop Rotation, SCQ); Kellogg Biological Station, Michigan, USA (Agricultural Land – Switchgrass, KBS); Algoma, Ontario, Canada (Coniferous Forest, ALO); Anza Borrego Desert, California, USA (Hot Desert, ABD); Department of Meta, Colombia (Tropical Peatland, DMC); National Park of Serra do Cipo, Brazil (Rupestrian Grassland, NPS); and Luquillo Experimental Forest, Rio Grande, Puerto Rico (Tropical Rainforest, LEF). This cohort of metagenomes was chosen to capture a diverse range of ecosystem types and land use classes. Raw reads derived from the 24 metagenome libraries were quality-controlled by clipping off primers and adapters and filtering out artefacts and low-quality reads using Read_QC module in the metaWRAP pipeline [403]. For each ecosystem, the three quality-controlled metagenomes were both co-assembled and individually assembled using MEGAHIT v1.1.3 [404], producing four assemblies for each ecosystem. For the assembly process, all of them used default parameters except co-assembly of metagenomes for Kellogg Biological Station (--k-min 27). Short contigs (<1000 bp) were removed. Each assembly was binned using the binning module within the metaWRAP [403] options (MetaBAT [405], MetaBAT2 [344] and MaxBin2 [406]) except assemblies derived from Kellogg Biological Station and Luquillo Experimental Forest where only MetaBAT2 [344] was used. Dereplication and quality-control of produced bins were performed as above. After dereplication, a total of 664 high- or medium-quality MAGs were obtained for the eight ecosystems. Each bin was taxonomically assigned as above.

4.3.10. Functional annotation of binned and unbinned contigs

The sequences of 43 marker genes representing energy conservation and carbon acquisition processes were retrieved from binned and unbinned contigs. Open reading frames (ORFs) were first predicted using Prodigal v.2.6.3 [350] and genes were annotated using a combination of homology-based searches and hidden Markov

model (HMM) searches. For homology-based searches, predicted ORFs were searched using DIAMOND *blastp* [410] against 32 custom protein databases described below. These encompassed the genes encoding ATP-citrate lyase (AclB), acetyl-CoA synthase (AcsB), ammonia monooxygenase (AmoA), anaerobic sulfite reductase (AsrA), anaerobic carbon monoxide dehydrogenase (CooS), aerobic carbon monoxide dehydrogenase (CoxL), dissimilatory sulfite reductase (DsrA), flavocytochrome *c* sulfide dehydrogenase (FCC), 4-hydroxybutyryl-CoA synthase of Crenarchaeota (HbsC), 4-hydroxybutyryl-CoA synthase of Thaumarchaeota (HbsT), hydrazine synthase (HzsA), malonyl-CoA reductase of Chloroflexota (Mcr), methyl/alkyl-CoM reductase (McrA), soluble methane monooxygenase (MmoX), periplasmic nitrate reductase (NapA), dissimilatory nitrate reductase (NarG), nitrogenase (NifH), copper-containing nitrite reductase (NirK), cytochrome *cd*$_1$ nitrite reductase (NirS), nitrous oxide reductase (NosZ), ammonia-forming nitrite reductase (NrfA), nitrite oxidoreductase (NxrA), particulate methane monooxygenase (PmoA), ribulose 1,5-bisphosphate carboxylase/oxygenase (RbcL), succinate dehydrogenase / fumarate reductase (SdhA / FrdA), sulfur oxygenase/reductase (Sor), thiosulfohydrolase (SoxB), sulfide-quinone oxidoreductase (Sqr), and three hydrogenase classes (NiFe-hydrogenase large subunit, FeFe-hydrogenase catalytic domain, Fe-hydrogenase). DIAMOND mapping was performed with a query coverage threshold of 80% for all databases, and a percentage identity threshold of 60% (AmoA, PmoA, MmoX, CoxL, HbsT, NxrA, RbcL) or 50% (all other databases) and e-value thresholds of 10^{-20}. HMM searches were performed against Pfam and Tigrfam databases using the *annotate* function of EnrichM v.0.5.0 (https://github.com/geronimp/enrichM), with domain noise-cut-off scores as previously described [351]. Eleven genes encoding subunits of ATP synthase (AtpA), two NADH dehydrogenases (NuoF, NqrF), four terminal oxidases (CcoN, CoxA, CydA, CyoA), two photosystems (PsaA, PsbA), formate dehydrogenase (FdhA), and reductive dehalogenase (RdhA) were searched.

4.3.11. Phylogenetic analysis

Phylogenetic trees were constructed to understand the distribution and diversity of bacteria and archaea consuming inorganic energy sources. Trees were constructed for the group 1 and 2 [NiFe]-hydrogenase large subunits, group 3 [NiFe]-hydrogenase large subunits, CoxL, PmoA, RbcL, AmoA, NxrA, Sqr, SoxB, and DsrA. In all cases,

protein sequences retrieved from the MAGs reads by homology-based searches were aligned against a subset of reference sequences from the custom protein databases using ClustalW in MEGA7 [353]. Evolutionary relationships were visualised by constructing a maximum-likelihood phylogenetic tree; specifically, initial trees for the heuristic search were obtained automatically by applying Neighbour-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. All residues were used and trees were bootstrapped with 50 replicates. To further visualize the diversity of trace gas oxidizers, neighbor-joining trees were constructed using the binned and unbinned sequences for group 1 and 2 [NiFe]-hydrogenase large subunits and CoxL; these trees were constructed using the Poisson model with gaps treated with pairwise deletion and bootstrapped with 50 replicates.

4.3.12. Analysis of community composition and diversity

Bacterial and archaeal community composition was determined from the pre-processed metagenomic reads with SingleM v0.12.1 (https://github.com/wwood/singlem). In total, 28 HMM searches were performed against 14 single-copy ribosomal marker genes. The gene for single-copy ribosomal protein L16/L10E (*rplP*) was selected for downstream analysis and sequences were clustered *de novo* into operational taxonomic units at a sequence identity threshold of 97%. Taxonomic assignment was carried out using the Genome Taxonomy Database [290]. Community richness and beta diversity were calculated using the phyloseq [299] and R package VEGAN [349]. To account for differences in richness between samples, all sequences were rarefied to within 90% of the minimum sequence count. Observed richness and estimated richness (Chao1) were calculated. First, beta diversity (Bray-Curtis) was calculated and visualised using a multidimensional scaling ordination (MDS), then significance testing was carried out using a permutational analysis of variance (PERMANOVA) to test for significant differences in community structure between ecosystems and between depth profiles. A beta dispersion test (PERMDISP) was used to ascertain if significant differences in community structure were due to data dispersion.

4.3.13. Metabolic analysis of short reads

For the functional annotation of short reads, paired-end reads in each sample were stripped of adapter and barcode sequences, then contaminating PhiX and low quality sequences were removed (minimum quality score 20) using the BBDuk function of BBTools v. 36.92 (https://sourceforge.net/projects/bbmap/). Resultant quality-filtered forward reads with lengths of at least 100 bp were searched for the presence of the 43 metabolic marker genes described above using DIAMOND *blastx* algorithm [352]. Specifically, reads were searched against the 32 custom-made reference databases and hits from the 11 HMM searches, using a query coverage of 80% and an identity threshold of either 60% (AmoA, PmoA, MmoX, CoxL, HbsT, NxrA, RbcL) or 50% (all other databases) and a maximum e-value threshold of 10^{-10}. Read counts were normalized to reads per kilobase per million (RPKM) and further normalised against a mean RPKM value estimated from 14 single copy ribosomal marker genes to infer the percentage of the community encoding the gene.

4.3.14. Thermodynamic modelling

Cell-specific power was calculated (i.e. Gibbs energy per unit time per microbial cell), *P* (W) according to:

$$P = \frac{r \cdot \Delta G_r}{B} \qquad \text{(EQ. 1)}$$

where *r* denotes the rate of reaction (mol s^{-1} $g_{dry\ soil}^{-1}$), ΔG_r represents the Gibbs energy of the reaction (J mol^{-1}), and *B* (cells cm^{-3}) is the number of microbial cells carrying out each of the following the reactions: 2 H_2 + O_2 → 2 H_2O (dihydrogen oxidation); 2 CO + O_2 → CO_2 + 2 H_2O (carbon monoxide oxidation); CH_4 + 2 O_2 → CO_2 +2 H_2O (methane oxidation).Values of ΔG_r are calculated using:

$$\Delta G_r = \Delta G_r^0 + RT \ln Q_r \qquad \text{(EQ. 2)}$$

where ΔG_r^0 and Q_r refer to the standard Gibbs energy and the reaction quotient of the indicated reaction, respectively, *R* represents the gas constant, and *T* denotes

temperature in Kelvin. Values of ΔG_r^0 were calculated using the revised-HKF equations of state [411–413], the SUPCRT92 software package [414], and thermodynamic data taken from refs [415–419].

Values of Q_r are calculated for each reaction using:

$$Q_r = \prod a_g^{v_i} \quad \text{(EQ. 3)}$$

where a_i stands for the activity of the i^{th} species and v_i corresponds to the stoichiometric coefficient of the i^{th} species in the reaction of interest. Activities of gas-phase compounds, a_g, were calculated using:

$$a_g = \frac{f_g}{f_g^0} \quad \text{(EQ. 4)}$$

where f_g and f_g^0 designate the fugacity and standard state fugacity of the respective gas. Due to the low temperatures and pressures of soil ecosystems, fugacity coefficients for all gases are one (see [420]). Therefore, partial pressures are equivalent to fugacity since f_g^0 was taken to be 1 bar. Gibbs energy calculations were carried out at 20°C and 1 bar. The concentrations of reactants in the catabolic reactions and reaction rates were measured by gas chromatography as described above.

Estimates of the number of microbial cells carrying out each reaction (*B*) were obtained by calculating the proportion of cells performing a specific catabolic reaction in our 16S rRNA gene copy number dataset, assuming all cells to be active. The proportion of a specific functional community was retrieved from the relative abundance of the individual biomarker genes (encoding the large subunits of group 1c, 1d, 1f, 1h, 2a [NiFe]-hydrogenase, carbon monoxide dehydrogenase (CoxL), particulate methane monooxygenase (PmoA), and soluble methane monooxygenase (MmoX) in our metagenomics dataset. An average of 4.2 16S rRNA copies per cell was assumed for cell number estimation [294].

4.4. Results

4.4.1. Diverse bacterial phyla are capable of oxidising H_2, CO, and CH_4 in soil ecosystems

First, we sequenced metagenomes from depth profiles of four Australian soil biomes (wetland, grassland, forest, dryland) **(Table S1 & S2)** and analyzed metagenomes from eight further global sites **(Table S2)**. Assembly and binning yielded 757 high- or medium-quality metagenome-assembled genomes (MAGs) spanning 27 bacterial and three archaeal phyla **(Table S3)**. Consistent with the community composition of the sites **(Fig. S1; Table S4)**, half of the genomes affiliate with globally dominant soil phyla Actinobacteriota, Proteobacteria, and Acidobacteriota [35]. We used comprehensive reference databases to search key metabolic genes in metagenomic short reads and derived genomes **(Fig. 1; Table S5 & S6)**. In agreement with established paradigms [106], most bacterial MAGs in this study have the capability to generate energy through aerobic respiration of organic compounds. More surprisingly, many of these bacteria also encoded the key enzymes (uptake hydrogenases and CO dehydrogenases) to consume trace gases H_2 and CO: an average of 39% and 47% of bacteria based on short reads, and 31% and 26% of assembled genomes. Also widespread was the capacity for sulfide (28% reads / 22% genomes) and thiosulfate (6.3% / 11%) oxidation, whereas nitrite (4.9% / 3.7%), ammonia (1.4% / 2.4%), and methane (1.1% / 0.37%) oxidation and photosynthesis (1.3% / 1.5%) were more restricted traits **(Fig. 1)**. Altogether, on average 72% of bacterial genomes were predicted to consume inorganic energy sources. For electron acceptor utilization, most bacteria encoded terminal oxidases for aerobic respiration and many were predicted to conserve energy by nitrate reduction (25% / 14%), nitrite reduction (21% / 21%), and hydrogenogenic fermentation (29% / 21%) **(Fig. 1)**.

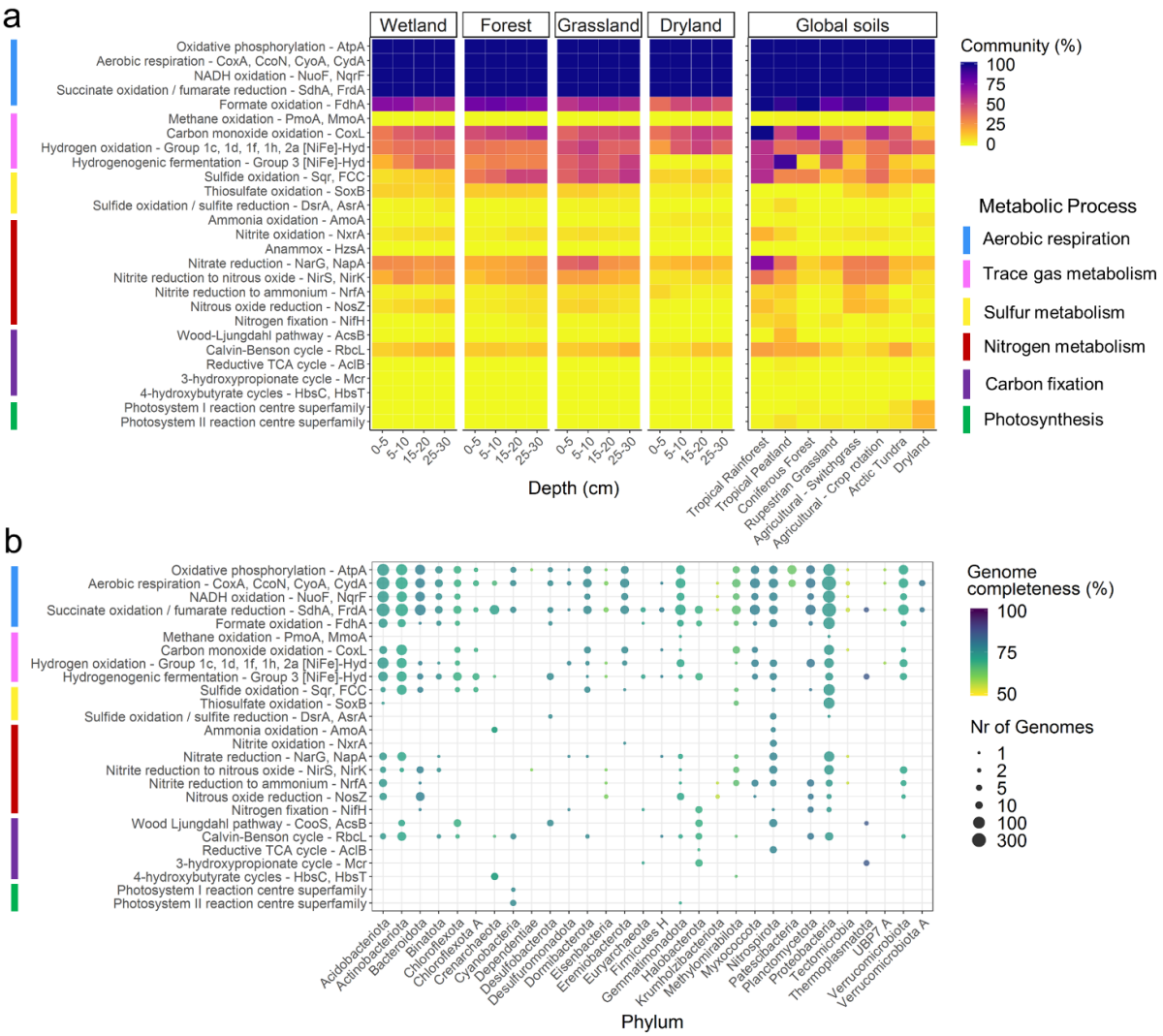


Fig. 1. Energy conservation and carbon acquisition strategies of global soil bacteria. Homology-based searches were used to detect marker genes for key metabolic processes in the metagenomic short reads and metagenome-assembled genomes. **(a)** Heatmaps showing the abundance of each gene in the metagenomic short reads of the four Australian depth profiles and eight global sites. The percentage of the total community predicted to encode at least one of each gene for a process is shown, based on normalization to single-copy marker genes. The genes detected are usually present in single copies in genomes and, where genes performing similar functions are collapsed together, the values are summed up to 100%. **(b)** Dot plot showing the metabolic potential of the 757 metagenome-assembled genomes (MAGs). The size of each point represents the number of genomes in each phylum that encode the gene of interest and the shading represents the average genome completeness. Taxonomy assignment is based on the Genome Taxonomy Database (GTDB) [290].

To gain insight into phylogenetic diversity of trace gas oxidizers, we generated phylogenetic trees of the uptake hydrogenases and CO dehydrogenases from binned and unbinned assembled sequences **(Table S7)**. 19 phyla encoded one or both enzymes, including all nine dominant soil phyla (Acidobacteriota, Actinobacteriota, Bacteroidota, Chloroflexota, Firmicutes, Gemmatimonadota, Planctomycetota, Proteobacteria, Verrucomicrobiota) and six candidate phyla (Binatota, Dormibacterota, Eremiobacterota, Methylomirabilota, Tectomicrobia, UBP7). Among the uptake hydrogenases, 49% of the 1522 hits affiliated with group 1h [NiFe]-hydrogenases. This relatively recently discovered high-affinity enzyme supports persistence of organotrophic bacteria by oxidizing atmospheric H_2 [72,162,330,331,385,421]. The phylogenetic tree revealed an unprecedented sequence diversity and broad taxonomic distribution of this subgroup, with sequences affiliated with 13 phyla **(Fig. 2a; Fig. S2)**. Also widespread were various radiations within the group 1d (8.9%), 1f (19%), and 2a (3.6%) [NiFe]-hydrogenases known to support aerobic hydrogen oxidation [162,180,331,422], as well as a novel subclade of the group 1c hydrogenases (13%) encoded by Acidobacteriota, Planctomycetota, and Gemmatimonadota **(Fig. 2a; Fig. S3 & S4)**. CO dehydrogenases were similarly prevalent. The tree shows large actinobacterial and proteobacterial clades flanking a central mixed clade containing sequences from 11 different phyla **(Fig. 2b; Fig. S5)**. In line with previous inferences [71,162], both trees suggest enzymes for trace gas oxidation were horizontally disseminated across soil bacteria on multiple occasions.

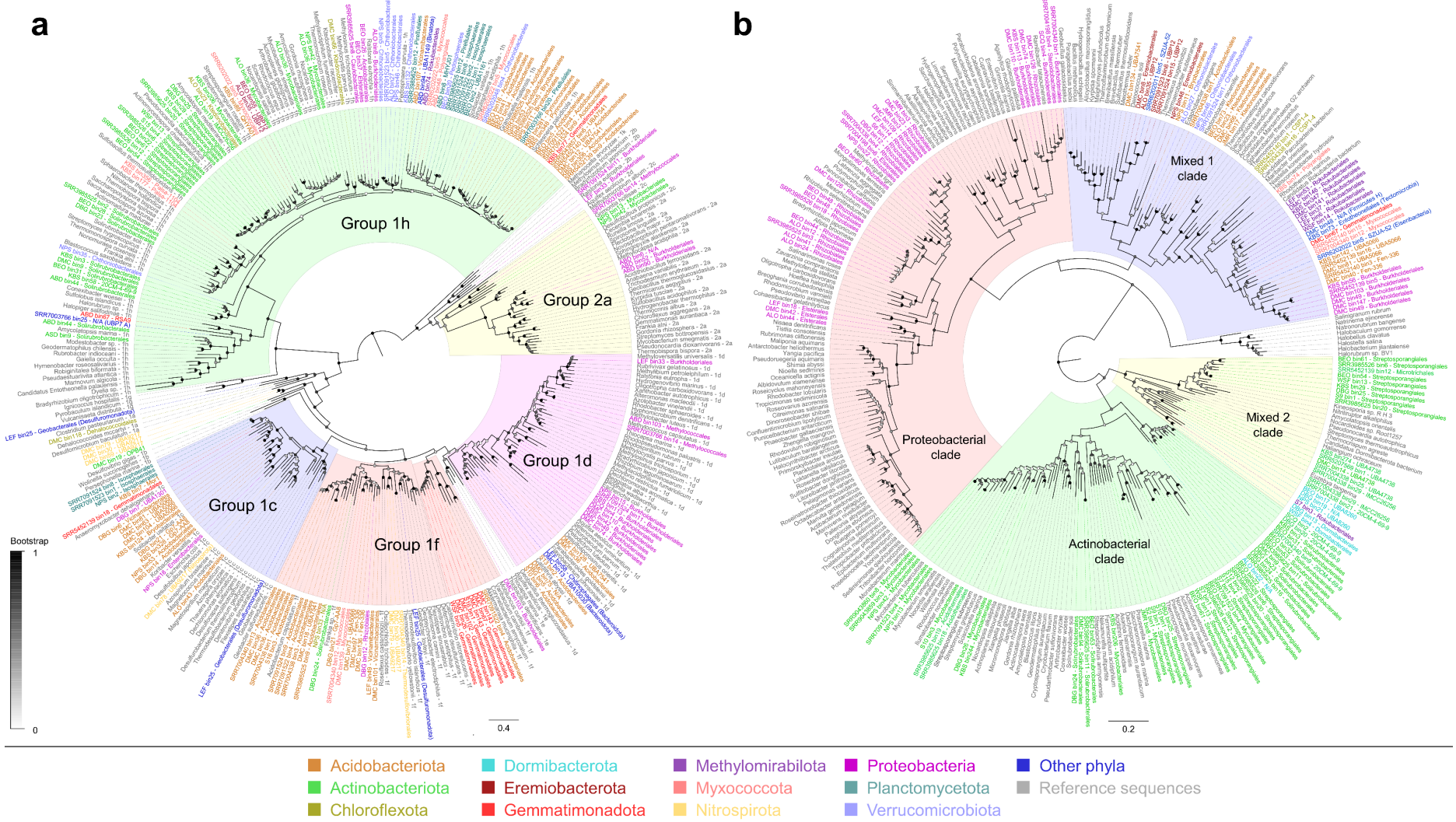

a
Group 1h
Group 2a
Group 1d
Group 1f
Group 1c
Bootstrap
1
0
0.4
b
Mixed 1 clade
Mixed 2 clade
Proteobacterial clade
Actinobacterial clade
0.2
Acidobacteriota
Actinobacteriota
Chloroflexota
Dormibacterota
Eremiobacterota
Gemmatimonadota
Methylomirabilota
Myxococcota
Nitrospirota
Proteobacteria
Planctomycetota
Verrucomicrobiota
Other phyla
Reference sequences

Fig. 2. Radial maximum-likelihood phylogenetic trees showing the sequence diversity and taxonomic distribution of key enzymes associated with trace gas oxidation. The trees show the amino acid sequences of **(a)** group 1 and 2 [NiFe]-hydrogenase large subunits and **(b)** carbon monoxide dehydrogenase (CoxL). Binned reads (taxa names colored by phylum) are shown alongside reference amino acid sequences (taxa names shown in grey). The hydrogenase subgroups and carbon monoxide dehydrogenase clades predicted to support trace gas oxidation in the sampled soils are highlighted. Both trees were constructed using the JTT matrix-based model, used all sites, and were bootstrapped with 50 replicates.

Phylogenetic analysis also validated the presence and affiliations of sulfide, thiosulfate, ammonia, nitrite, and methane oxidizers **(Figs. S6 to S12)**. These included the first genomes of soil comammox bacteria [423,424] and a potential nitrite-oxidizing bacterium within recently reported candidate phylum Eremiobacterota [28]. A major finding was the recovery of a high-quality Gemmatimonadota bin predicted to use methane, hydrogen, and acetate as energy sources **(Fig. S13)**. Its particulate methane monooxygenase is distantly related to those of sequenced methanotrophs and instead affiliates with an uncultivated methanotrophic lineage (Tropical Upland Soil Cluster, TUSC) that has been detected by amplicon sequencing in diverse soils worldwide **(Fig. S12)**. These results suggest Gemmatimonadota is a fourth phylum [204,389] capable of aerobic methanotrophy, though cultivation studies are required to confirm activity. We propose this bacterium is named *Candidatus* Methylotropicum kingii **(Etymological Information)**.

4.4.2. Trace gas oxidizers are active across soil types and depths

Reflecting their abundance and diversity, we confirmed that trace gas oxidizers are highly active in soil communities. *In situ* concentrations and soil-atmosphere fluxes of H_2, CO, and, as a well-studied reference gas, CH_4 were measured across the four Australian biomes in biological quadruplicate using an ultra-sensitive gas chromatograph **(Table S8)**. All three gases were present within typical mixing ratios [394,395,425] at the soil-atmosphere interface and top 20 cm of each soil (av. 0.84 ± 0.46 ppmv H_2, 0.91 ± 0.33 ppmv CO, 1.67 ± 0.17 ppmv CH_4) **(Fig. 3a)**; gas concentrations

decreased with depth at some sites **(Fig. S14)**, indicating microbial consumption, whereas H_2 concentrations were elevated in deeper wetland soils likely due to activity of the numerous hydrogenogenic fermenters (33% reads / 32% genomes; **Fig. S15**) in hypoxic zones [335]. In line with the dominance of H_2 oxidizers, net *in situ* H_2 uptake was observed at all sites and consumption was particularly rapid at the grassland, forest, and wetland (exceeding 20 nmol m^{-2} s^{-1}) **(Fig. 3b)**; most ambient H_2 in the flux chambers was oxidized by underlying soil within two minutes. In contrast, no significant fluxes of CO were observed; the high rates of photochemical CO production known to occur under daylight potentially obscured underlying microbial consumption [397]. Reflecting global patterns [426], CH_4 fluxes were low with the exception of forest sites.

We subsequently measured *ex situ* oxidation rates by incubating samples from soil depth profiles in serum vials containing ambient air headspaces **(Table S8)**. Under these conditions, biological gas consumption could be accurately quantified given no significant abiotic production of these gases was detected **(Fig. S16)**. Atmospheric H_2 and CO were both rapidly oxidized by forest, grassland, and wetland soils. Rates were significantly higher for topsoils (0-5 and 5-10 cm) compared to deeper soils (15-20 and 25-30 cm) **(Table S9)**. However, cell-specific rates (normalized to 16S rRNA gene copy number; **Fig. S17**) did not significantly vary with depth **(Fig. 3c; Table S9)**. Atmospheric CH_4 oxidation occurred at 60-fold lower rates than H_2 and CO on average, with most rapid consumption again occurring at the forest site. Such observations are consistent with the much higher levels of uptake hydrogenases and CO dehydrogenases compared to methane monooxygenases in the metagenomes and derived genomes. In line with the *in situ* observations, trace gas oxidation was slowest for dryland soils. This possibly reflects, in line with previous laboratory studies, that low water content inhibits trace gas uptake [427,428]. Contrary to the pattern observed for other biomes, deeper dryland soils consumed trace gases more rapidly, perhaps reflecting their higher measured water content **(Table S1)**. Experimentally wetting dryland soils, in order to simulate rainfall events, enhanced rates by six-fold **(Fig. 3c; Fig. S16)**. Overall, these *in situ* and *ex situ* measurements complement the metagenomic analysis by confirming soil communities contain highly active H_2 and CO oxidizers.

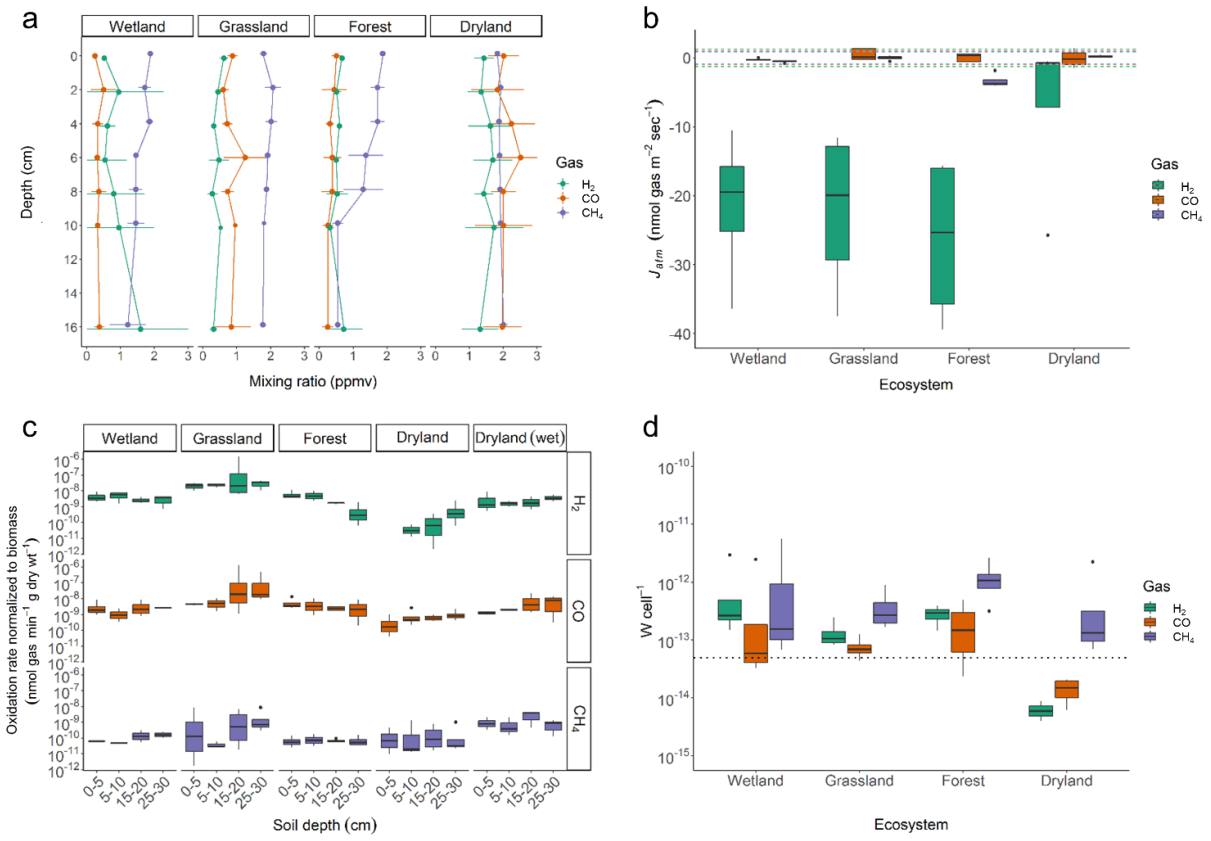


Fig. 3. Measurement of oxidation of the trace gases H_2, CO, and CH_4 across four Australian biomes. **(a)** Depth-resolved *in situ* gas concentrations from 0 to 16 cm depth. **(b)** *In situ* soil-atmosphere gas fluxes (J_{atm}; positive values indicate net gas production, negative values indicate net gas consumption). For **a** and **b**, measurements were performed in four separate soils per biome and error bars showing standard deviations. **(c)** Depth-resolved *ex situ* oxidation rates for each ecosystem using core samples from four different depths (0-5, 5-10, 15-20, 25-30 cm). Measurements were performed with four separate soils per biome and values are normalized to biomass based on 16S rRNA gene counts. Dashed lines represent simulated flux detection limits. **(d)** Amount of power per cell derived from the oxidation of each trace gas. These were calculated using thermodynamic models based on *ex situ* rates measured across the four depths per biome, with the dotted line showing the minimal maintenance energy requirements measured for aerobic heterotroph pure cultures (4.9×10^{-14} W per cell) [429]. For **c** and **d**, boxplots show means, lower and upper quartile, and minimum and maximum values.

4.4.3. Trace gas oxidation can theoretically sustain maintenance of entire community and growth of some autotrophs

It is probable that trace gas oxidation primarily supports persistence of organotrophic bacteria. Most sequenced soil bacteria encoded uptake hydrogenases and CO dehydrogenases together with the genes for organotrophy. Based on observations from pure culture studies, bacteria consume trace gases to conserve energy for cellular maintenance when organic carbon supplies are limiting for growth [72,330,331,385]. We used thermodynamic modelling to predict the amount of power per cell **(Table S10)** that could be generated based on the *ex situ* oxidation rates **(Fig. S16)** and the number of trace gas oxidizers detected per gram of soil **(Fig. 1 & Fig. S17)**. On average, oxidation rates were sufficient to generate 3.3×10^{-13} W per H_2-oxidizing cell, 2.4×10^{-13} W per CO-oxidizing cell, and 1.0×10^{-12} W per CH_4-oxidizing cell, and trends were similar across biomes and depths with the exception of the dryland soils **(Fig. 3d)**. Such values are higher than the minimal maintenance energy requirements measured for aerobic heterotroph pure cultures (4.9×10^{-14} W per cell) [429] and greatly exceed the requirements calculated for highly energy-limited ecosystems (10^{-17} to 10^{-19} W per cell) [430,431]. Thus, trace gases can theoretically sustain the persistence of the entire gas-consuming community.

Trace gases are also likely to sustain autotrophic growth of a significant proportion of bacteria across these soil biomes. Our analysis indicates that most community members, including trace gas oxidisers can acquire carbon heterotrophically. However, some bacteria (13% reads / 12% genomes) encoded the capacity to assimilate CO_2 through the Calvin-Benson cycle **(Fig. 1)**. Genome-resolved analysis indicated most of these autotrophs were capable of oxidising H_2 (79%) and/or CO (63%), with some mediating sulfide oxidation (32%), thiosulfate oxidation (17%), or photosynthesis (11%) **(Table S6)**. Putative autotrophic H_2 and CO oxidizers were patchily distributed among seven dominant soil phyla and candidate phylum Dormibacterota, suggesting autotrophy has been acquired multiple times **(Fig. 1)**. Consistently, phylogenetic analysis showed that the 93% of the binned ribulose 1,5-bisphosphate carboxylase (RuBisCO / RbcL) hits were from clades known to support chemosynthesis rather than photosynthesis **(Fig. S18)**. Reads for other CO_2 fixation

pathways were also detected, but with exception of the peatland sample, were encoded by less than 1% of the total community **(Fig. 1)**. The above power calculations suggest that sufficient energy is generated from trace gas oxidation for a subset of bacteria to allocate to growth in addition to maintenance at some sites. Autotrophic growth may be particularly favorable in environments where there are significant rates of edaphic gas production, for example due to hydrogenogenic fermentation **(Fig. S15)** during hypoxia [335]. Remarkably, this suggests that the process of atmospheric chemosynthesis recently discovered in Antarctic deserts [28] extends to other biomes.

4.5. Discussion

Overall, bacteria with the metabolic flexibility to use both organic and inorganic energy sources are likely to have a selective advantage in soil environments. In most soils, organic carbon is the main factor limiting microbial growth [432,433]; this reflects the inherent spatiotemporal variability in organic carbon availability of soils, together with the recalcitrance of many organic polymers and the intense competition for more degradable compounds [106]. Thus, the ability to consume alternative energy sources is likely to be critical for adaptation and resilience of many taxa. H_2 and CO are ideal compounds in this regard given they are readily available from both atmospheric and edaphic sources; likewise, their high energy content, low activation energy, and diffusibility into microbial cells make these gases dependable for survival [71,176].

Consistently, our findings suggest that oxidation of these trace gases is a generalist process, rather than a specialist one as previously suggested [386]. We provide multiple lines of evidence that H_2 and CO oxidizers are abundant, diverse, and active across different soil biomes and depths. These findings strikingly contrast with previous reports that trace gas oxidizers comprise just 1% of the community. Such discrepancies reflect that previous work on atmospheric H_2 oxidizers relied on non-degenerate quantitative PCR primers that only capture a small proportion of the total diversity of soil uptake hydrogenases; indeed, we observed high proportions of high-affinity hydrogenases (30% reads / 40% genomes) in the Canadian cropland metagenome where H_2 oxidizers were previously inferred to be minor community members [386,387]. Moreover, while previous genome surveys indicate Actinobacteriota

predominantly mediate atmospheric H_2 oxidation [162,421], our analysis indicates all major bacterial soil phyla can mediate this process. These metabolically flexible generalists co-exist with more specialist taxa that use niche substrates, including methane, ammonia, and nitrite. The finding that some Gemmatimonadota are inferred to be capable of aerobic methane oxidation further highlights the diversity of trace gas oxidizers in soils. Our results also reveal an unexpected diversity and abundance of chemoautotrophs in soils, while hinting at roles for sulfur-based compounds as further supplementary energy sources.

More broadly, the extensive soil-atmosphere interaction described here appears to be a key regulator of soil biodiversity and atmospheric chemistry. It is well-established that dormancy contributes to the maintenance of microbial biodiversity and the resilience of soil communities [211]; given trace gases are major energy sources sustaining the dormant soil majority, their oxidation will in turn influence wider community ecology. Biogeochemically, soil bacteria are major sinks of trace gases, accounting for the net loss of approximately 75% atmospheric H_2 (70 megatonnes per year) [395,396] and 10% atmospheric CO (250 megatonnes per year) [165,394]. The observation that the bacteria responsible are more numerous and diverse than previously thought suggests these sinks are relatively resilient; this may explain why, despite high anthropogenic emissions of H_2, global mixing ratios have remained stable [395]. Nevertheless, various human activities could undermine this soil-atmosphere interaction, for example through changing soil properties (e.g. via agricultural practices and desertification) and atmospheric composition (e.g. through urban pollution or a hydrogen economy [434]), with potential ecological and biogeochemical ramifications. Further studies are required to understand what physicochemical factors influence the abundance and activities of trace gas oxidizers, and how they respond to local and global change.

4.6. Footnotes

Etymological Information: *Candidatus* Methylotropicum (Me.thy.lo.tro'pi.cum.) N.L. n. *methylum* (from French *méthyle* back-formation from French *méthylène*, coined from Gr. n. *methu*, wine and Gr. n. *hulê*, wood), the methyl group; N.L. pref. *methylo*-, pertaining to the methyl radical; L. masc. adj. *tropicus* pertaining to tropical zone of the

Earth; N.L. neut. n. *Methylotropicum* referring to a methyl-using bacterium from the tropics. *Candidatus* Methylotropicum kingii (ki'ngi.i.) N.L. gen. n. *kingii* in honour of Prof. Gary King, who has pioneered research into the microbial oxidation of methane, CO, and H_2.

Acknowledgements: This research was supported by an Australian Research Council DECRA Fellowship (DE170100310; C.G.), a Discovery Project grant (DP180101762; C.G. and P.L.M.C.), a Swiss National Foundation Early Mobility Postdoctoral Fellowship (P2EZP3_178421; E.C.), a NSFC grant (41906076; X.D.), a NERC grant (NE/T010967/1; J.B.), a Humboldt Foundation Fellowship (J.B.), Monash University PhD scholarships (S.K.B. and P.M.L.), and an NHMRC EL2 Fellowship (APP1178715; salary for C.G.). We thank Maria Chuvochina for etymological advice and Steven Chown, Sergio Morales, Matthew Stott, and Gregory Cook for helpful discussions.

Author contributions: C.G. conceived and supervised this study. S.K.B., C.G., and E.C. designed experiments and analysed data. Different authors were responsible for performing fieldwork (S.K.B., E.C., P.M.L., C.G.), laboratory work (S.K.B., T.J.), metagenome assembly (X.D.), metagenome analysis (S.K.B., C.G., X.D., E.C.), and thermodynamic modelling (J.A.B., D.L., E.C., S.K.B., P.A.N.). P.A.N., S.K.A., and P.L.M.C. provided logistical and theoretical support. S.K.B. and C.G. wrote the paper with input from all authors.

Data Availability Statement: All metagenomes sequenced for this project and all metagenome-assembled genomes will be deposited in the Sequence Read Archive before publication.

Supplentary material: Supplementary figures and tables are found in the Appendix C.

Etics declarations: The authors declare no competing financial interests.

Chapter 5

Discussion and Outlook

5.1. Summary

5.1.1. Community composition

The biogeographic analysis presented in chapter two suggests that the spatial turnover in soil bacterial and archaeal communities is higher than previously recognised. Our 16S soil survey, covering local and regional scales in the Negev region, found that the majority of taxa are members of the rare biosphere and only a small minority are cosmopolitan. This was highlighted in the positively skewed occupancy frequency distribution, which showed the full spectrum of occupancy classes, extending from highly frequent rare members to intermediate members to a long tail of low frequency common members. While these findings broadly agree with the occupancy frequency distribution of microbial taxa across the majority of terrestrial ecosystems [35,48], our findings suggest that the magnitude in frequency should be higher, given the positive correlation between taxonomic resolution and observed richness. The occupancy trends observed in our analysis are of global relevance, given similar observations were made using a previously published global scale dataset [35].

Our findings also shed new light on how aridity influences the richness, abundance, and composition of microbial communities. We found that, with increasing aridity, community abundance decreased, whereas richness was variable but remained relatively constant between climatic regions. This was evident from the similar alpha diversity observed between sub-humid and hyper-arid regions. In the Negev aridity transect, the dominant phyla detected were Actinobacteriota, Acidobacteriota, Chloroflexota and Proteobacteria. These phyla are among the nine dominant soil phyla frequently cited in global and regional scale soil surveys [35–37] and are known to thrive in desert environments[326]. Compositional profiling at the phylum levels revealed that the relative abundance and diversity of these taxa was similar between climatic zones. This suggests that a broad range of taxa have the metabolic capacity to acquire carbon and energy, even in hyper-arid soils, thereby enabling them to withstand aridity-associated environmental filters and maintain structured communities. In contrast, there were major variations in the abundance of Cyanobacteria between topsoils and

biocrusts, as well as between specific samples, suggesting there are relatively sensitive to variations in water availability and other physicochemical composition. The four above-mentioned phyla were also observed to be the most abundant across metagenomes from Australian and global soils, though metagenome-assembled genomes were also recovered for a wide variety of other cultured and candidate phyla.

5.1.2. Community turnover

Increasingly studies have shown that deterministic and stochastic assembly processes variably interact to structure soil microbial communities [98,99,102,103,244]. Beta diversity studies, using both abundance- and incidence-based metrics, showed that the taxonomic and phylogenetic composition of the Negev desert soil communities was significantly different between soils in different plots and zones. To determine the underlying drivers of this variation, we used a variation partitioning analysis. This enabled us to parse the relative contribution of drivers associated with abiotic environmental conditions, and spatial drivers which are independent of environmental heterogeneity. The analysis revealed that the majority of explained variation could be attributed to abiotic drivers associated with increasing aridity such as soil water content, organic carbon, pH and salinity. This suggests that deterministic environmental drivers are major determinants of microbial community structure along aridity gradients, with spatial effects playing a minor role. Nevertheless, we observed a marginal increase in the proportion of spatial effects along our local transect, which extended along a single climatic region and had much weaker environmental heterogeneity. It is therefore likely that spatial effects play a greater role under conditions where environmental selection is predicted to be weak, in agreement with other biogeographic studies [86,98,100,435]. Overall, these findings concur with the majority of studies examining environmental and spatial drivers, which report that environmental not spatial effects are the most important determinants of contemporary community structure [81,86].

Beyond examining a particular subset of environmental and spatial drivers to predict drivers of community assembly, we measured the contribution of rare, intermediate and common taxa to turnover. Using the multi-site diversity metric zeta, this analysis

enables a statistical prediction to ascertain if community structure follows stochastic or deterministic assembly processes [273,278]. Zeta decline revealed that the proportion of compositional change in community structure was typical of deterministic niche differentiation and overwhelmingly driven by rare taxa. Concordant observations were made using the pairwise and multi-site distance decay models of zeta, which confirmed high compositional turnover along the regional transect. Furthermore, it confirmed that the majority of turnover was driven by rare taxa, which are known to be important determinants of community structure [46,48,278]. Given that microbial community turnover rates were much higher than anticipated, taxa area curves were constructed and compared against the wider microbial and eukaryotic literature. This analysis enabled comparison of turnover rates between bacteria and higher animals and plants, which revealed that turnover of soil bacteria and archaea exceeded the majority of previous estimates by one to two orders of magnitude [250,253,257–259]. These findings support an increasing body of work demonstrating that biodiversity estimates and biogeographic patterns are significantly degraded when rare taxa are excluded [46,48,436]. They also demonstrate that universal distribution hypothesis of microorganisms are outdated and highlight that commonly used methods are likely to underestimate microbial community turnover and biogeographic patterns in general [249,261,437,438].

Our results are concordant with previous literature suggesting dormancy increases richness of microbial communities, but suggest the role of this trait in reducing community turnover may have been overestimated. Broadly, dormancy is predicted to buffer microbial communities against environmental filters such as energy limitation and adverse biotic effects such as competition [67,260,439]. Metabolic strategies such as dormancy have been shown to sustain the minimum energy requirements of bacteria, allowing them to survive under conditions which favour survival over growth [210]. These effects could lead to overlapping generations as well as facilitating the persistence of competitors that would otherwise not coexist via processes such as the ‘storage effect’ [440]. This has led to the assumption that microbial community structure is much weaker than those observed in animal and plant communities [249,327,441]. For example, recent studies comparing active and dormant microbial communities suggest that distance decay slopes are moderately shallower in dormant communities, because the selective effects of environmental filtering are weakened [261]. Indeed, our findings

suggest that the relative consistent alpha diversity between climatic regions is likely to be maintained by processes such as dormancy. However, we also found that, despite increasing aridity, community structure and turnover remained high and was predicted to be driven by environmental niche processes. This suggests that environmental pressures such as carbon and water limitation continue to select for metabolic strategies which sustain the energy and carbon requirements of these dormant communities. We therefore predict that metabolic strategies supporting the energy and carbon requirements of dormant soil microorganisms play a key role in maintaining community structure along aridity gradients.

5.1.3. Metabolic function

The metagenomic and biochemical analysis presented in chapters three and four suggest that most soil bacteria are mixotrophs. They can grow organoheterotrophically, utilise various inorganic electron donors to support cellular maintenance, and in some cases, potentially fix CO_2 to maintain carbon requirements or potentially sustain growth. Consistent with the aerated niche of surface soils and established paradigms of soil microbial metabolism [110], the majority of the community were inferred to be capable of aerobic respiration of organic compounds. More surprisingly, most taxa encoded genes supporting inorganic energy acquisition, including through trace gas metabolism of H_2 and CO via uptake hydrogenases and CO dehydrogenases. Based on short read averages, the proportion of communities capable of consuming atmospheric H_2 and CO was: 38% and 47 % in Australian soils, 40% and 46% in Global soils and 43% and 29% in the Negev region soils. The high abundance of these genes coincided with high rates of trace gas uptake across various ecosystems. Strong correlations were observed between H_2 oxidation rates, increasing aridity and the abundance of H_2 uptake hydrogenases (group 1h and 1l [NiFe]-hydrogenase) along the aridity gradient in the Negev. *In situ* flux, and *ex situ* measurements confirmed similar biomass-normalised rates with depth in wetland, grassland and forest soils. Here H_2 and CO were rapidly oxidised to below atmospheric levels. This suggests that the metabolic capacity to utilize inorganic electron donors such as H_2 and CO confers a major selective advantage for soil bacteria. These findings are supported by an increasing body of culture-based genetic and biochemical

studies suggesting that these atmospheric trace gases serve as an ubiquitous and reliable energy sources for dormant soil bacteria [70–72,163,182].

The analysis of 757 MAGs presented in chapter four further demonstrate the prevalence of trace gas oxidation in surface soils. While the dominant soil phylum Actinobacteriota was primarily thought to mediate H_2 scavenging [162,421], we provide evidence from metagenomic and biochemical studies showing that diverse soil bacteria can mediate this process. Phylogenetic analysis revealed that 17 phyla including all nine dominant soil phyla and six candidate phyla encoded the capacity to oxidise H_2 *via* uptake hydrogenases. Furthermore, phylogenetic trees showed a previously unrecognised sequence diversity of these enzymes, with various lineages of group 1h, 1d, 1f, 1l, 2a, and 1c [NiFe]-hydrogenases distributed among soil bacteria. This suggests that, beyond the well-studied persistence-supporting 1h and growth-supporting 1d hydrogenases, other understudied lineages are also important mediators of aerobic hydrogen metabolism. Similarly, CO dehydrogenases were found to be diverse and distributed across 11 phyla. Overall, these trees suggest that the enzymes supporting H_2 and CO oxidation were most likely horizontally disseminated, over multiple occasions [72,182]. Given the co-occurrence of genes supporting aerobic respiration of organic compounds and trace gas metabolism, it is likely that atmospheric trace gases serve as a reliable energy source to sustain energy acquisition during dormancy, when organic substrates are insufficiently available to support growth [161,163,385]. To further understand whether trace gas metabolism releases sufficient energy to sustain dormancy, we used a thermodynamic model to predict the amount of power per cell generated by the oxidation of H_2, CO and CH_4. Based on *ex situ* oxidation rate measurements presented in chapter four, average oxidation rates of all three gases were found to exceed the theoretical minimal maintenance energy requirements predicted for aerobic heterotroph pure cultures and oligotrophic ecosystems [429–431]. This is further evidence that trace gas metabolism is likely to sustain the majority of dormant soil bacteria.

Our results suggest that, whereas oxidation of H_2 and CO oxidation are broad and generalist processes, other traits like CH_4 are relatively specialised ones. However, while results broadly concur with previous studies suggesting that aerobic CH_4 metabolism is a relatively rare trait in soil microbial communities [361,386,387], they

nevertheless show that this process may be more broadly distributed than previously recognized. This was evident given our discovery of a potential novel methanotroph belonging to the phylum Gemmatimonadota, which encoded a particular methane monooxygenase affiliated with an uncultivated methanotrophic lineage (Tropical Upland Soil Cluster, TUSC). Given that all three atmospheric gases are ubiquitous, diffusible, and yield sufficient energy, a major question is why there are such large discrepancies in the prevalence of CH_4 metabolism compared to H_2 and CO metabolism. It is likely CH_4 metabolism is a more specialist process because it requires a high amount of activation energy, additional oxygen expenditure, and requires organisms to carry out a complex metabolic balancing act given the requirements for oxygen, reductant, and substrate [341].

5.1.4. Primary production

Our findings suggest that atmospheric H_2 and CO may also support primary production in global soil ecosystems. Our metagenomic analysis presented in chapter three shows that genes encoding ribulose 1,5-bisphosphate carboxylase (RuBisCO) which support CO_2 fixation via the Calvin Benson-Bassam cycle (CBB) were encoded by an average of ~10% of the community throughout the aridity gradient, though this proportion was higher in arid and hyper-arid zones. While RuBisCO lineages associated with Cyanobacteria and phototrophic eukaryotes were found along the aridity gradient, the most widespread RuBisCO variants were type IA and IE that were affiliated with hydrogenotrophs [228,442]. These variants were an order of magnitude more abundant than the photosynthetic type IB RuBisCO and correlated with increasing aridity, suggesting that in arid and hyper-arid some members assimilate CO_2 via using atmospheric reduced gases as energy sources. Similar observations were made in our global analysis presented in chapter four, which found that ~12% of the community had the capacity for autotrophic CO_2 assimilation via the CBB cycle. Binned RuBisCO hits showed that 93% were affiliated with chemosynthetic rather than photosynthetic microorganisms. Our thermodynamic models revealed that the energy yields acquired by trace gas oxidation exceed the minimum energy required for maintenance. Given the prevalence of type IE RuBisCO in these soils, it is likely that

some residual energy remaining from cellular maintenance is allocated to chemoautotrophic growth by some members of the gas-scavenging community.

These findings are further supported by biochemical radiolabeling and gas chromatography studies presented in chapter three. These demonstrate photosynthetic processes are highly dependent on soil moisture and gradually inhibited with increasing aridity. In contrast, the positive correlation between H_2 oxidation and aridity suggest that chemosynthetic processes increasingly replace photosynthetic energy and carbon acquisition along aridity gradients. Nevertheless, as highlighted by the relatively low oxidation rates observed in Australian drylands, low water content does inhibit trace gas uptake albeit not as much as photosynthesis [427,428]. Reflecting this, we show how photosynthetic and chemosynthetic process are not mutually exclusive but co-occur when sufficient moisture is available. The high prevalence of genes supporting atmospheric chemosynthesis in soil biomes globally suggests that chemosynthesis provides an alternative metabolic strategy for cells, given intense competition from more competitive polysaccharide degraders or to sustain maintenance following the switch from growth to persistence. These findings are of global significance given they suggest that the atmospheric chemosynthesis, which has been shown to support a minimalistic mode of microbial primary production in Antarctic deserts [67], is also prevalent in other soil biomes.

The realization that many taxa are involved in the consumption of atmospheric H_2 and CO also has biogeochemical implications. Soils are a major biogeochemical sink for H_2 and CO produced by anthropogenic and natural processes, with 75% of atmospheric H_2 and 10% of atmospheric CO consumed by soil bacteria [164,165]. While it was previously thought that these losses are driven by <1% of the community, which is consistent with the abundance of methanotrophs [361,386,387], our findings suggest that the majority of taxa mediate the atmospheric turnover of these gases. This further highlights that the biogeochemical sink of H_2 and CO is highly resilient to changes in the atmospheric composition, which might explain why atmospheric mixing ratios of H_2 have remained stable despite increasing anthropogenic emissions [395].

5.2. Outlook

5.2.1. Integrating turnover of macro and micro organisms

Our findings add to a growing body of work suggesting that the similarities in the biogeographic scaling relationships between macro and microbial communities are likely to exceed any fundamental differences [234]. Universal biogeographic patterns can be defined as an equivalent process underpinning the same pattern regardless of the domain. Universal distribution patterns are usefully because the same frameworks, tools and metrics can be utilised to predict and extract meaningful results [95]. In turn these can be used in an applied or managed response to address challenges posed by global change such as biodiversity assessments of habitats, measuring the effects of disturbance or assessing the impact of biological invasion. Here we have taken some of the latest innovations to demonstrate how compositional turnover is more similar to higher animals and plant communities than previously recognised. Additionally, we have advanced methodologies and provided evidence on how microbial turnover patterns can be used to predict underlying niche and neutral assembly processes.

To adequately detect compositional changes in community structure and make predictions about underlying assembly processes, the contribution across the whole spectrum of occupancy classes is required [273]. Biases in counting individuals is not an exclusive issue for microbial studies. Arbitrary decisions to include or exclude individuals are also prevalent in field studies of higher animals and plants. Tree surveys exclude individuals that fall below a certain diameter, benthic communities are sieved using various mesh sizes and variable decisions are made when to start counting bird species returning from annual migrations [443–446]. Uncertainty in estimating richness and diversity is arguably inherent to all ecological studies, regardless the domain or size of the individual.

Future studies should incorporate multi-site diversity metrics to address differences in turnover and underlying assembly processes between macro and microbial communities using side-by-side sampling protocols [273,318]. These differences could be teased out by addressing some of the following questions: “Are there differences in

turnover between domains and their occupancy classes”? “Do deterministic and stochastic processes differentially influence community assembly between domains”? “If deterministic niche based processes dominate, which subset of abiotic drivers contribute to turnover”? “How do drivers differ between occupancy classes”? These questions should be tested across multiple spatial and temporal scales and use sampling designs which specifically capture spatial and environmental transects. Disentangling these interdomain relationships and assembly processes would aid our understanding of the relationship between microbial biodiversity and ecosystem function by differentiating the contribution to diversity between different trophic levels.

5.2.2. Drivers of trace gas metabolism

All ecosystems are exposed to various press and pulse dynamics which determine the spatiotemporal trajectory of environmental parameters [447]. While press changes such as anthropogenic climate change, desertification and eutrophication act in the long term, pulse changes such as rainfall, bushfires can rapidly change the physicochemical structure of soils in the short term. Both press and pulse dynamics are predicted to influence community composition and structure as well as the extent to which trace gas scavengers interact with the atmosphere. In this work we have shown how the activity and abundance of trace gas scavengers is sensitive to aridity gradients and differentiated across various ecosystems. Beyond that, we have also demonstrated that trace gas metabolism of H_2 and CO are active and dominant processes across the majority of the community and do not present a niche process like aerobic CH_4 oxidation. However, it is now important to determine how these metabolic strategies respond to different press and pulse changes in the environment and which physicochemical drivers determine the abundance and activity of these organisms.

In chapter two we have shown that with increasing aridity, chemosynthetic processes are increasingly important and photosynthetic processes generally cease to function. However, biochemical measurements of dryland soils in chapter three showed that trace gas oxidation was generally higher at lower depths and influenced by soil water content. Simulated precipitation pulses were not only found to amplify this trend but also initiated photosynthesis. This suggests that when water is available these

processes operate in unison and only become differentiated as soils dry up and communities become energy limited. Studies are now needed to address (i) how these processes operate on temporal scales, for example during and after precipitation events in desert ecosystems, and (ii) along aridity gradients capturing soils ranging from fully saturated (~30% soil water content) to completely dry (<1%). This is important because previous studies have described variable optima for trace gas uptake in soils ranging from 1.7% [448] to 10 % [427] for H_2 and 10 - 20% for CO [360]. This could be addressed by determining at which soil saturation level H_2 and CO oxidation rates peak and photoautotrophic / chemoautotrophic CO_2 fixation is activated and deceived. These finer scale measurements would provide answers under which environmental conditions microbial communities initiate a 'switch' between photo and chemosynthetic metabolic strategies or mixotrophically function to obtain carbon and energy.

These question could be extended to other parameters such as soil organic carbon content soil temperature and pH, which have also been shown to influence the activity of microbial H_2 and CO fluxes [360,448]. For example, the high abundance of genes supporting aerobic CO oxidation in forest soils and relative low abundance in deserts and dryland coincides with previous laboratory and field studies showing strong relationships between soil organic carbon and aerobic CO oxidation rates [190,360,361]. However, a major issue of *in situ* studies remain the strong CO emissions from photo and thermal degradation of organic compounds [397,449]. These emissions are particularly high in organic soils and they increase exponentially with temperature [448,449]. Given we observed high *in situ* CO emissions, which obscured our flux measurements, it will be important to accurately partition this abiotic contribution to ascertain the true microbial sink dynamics for CO. This would provide a more meaningful comparisons of *in situ* H_2 and CO fluxes to accurately quantify these important biogeochemical ecosystem services and their contribution to microbial energetics.

5.2.3. Environmental gradients

To determine the environmental conditions under which trace gas metabolisms is important, field studies should look towards ecosystems and environmental gradients which are predicted to select for traits supporting these processes. For example, aridity gradients described in this work provide a natural system to test hypotheses about the differentiation of photo and chemosynthetic processes. Similar gradients have been previously described in microbial soil surveys including the Negev desert [280] Atacama desert [450] and Namib desert [451], which potentially provide equivalent study systems. Future studies should include soils in humid and sub-humid climates alongside the semi-arid, arid and hyper-arid climates in order to capture wider range parameters such as organic carbon and soil water content. This would provide further evidence of how these processes are differentiated along various aridity gradients and provide answers about how microbial communities remain energised following desertification

Studies are also needed to examine the role of atmospheric trace gases in supporting the energy requirements of soil microbial communities during primary succession, to ascertain if trace gas oxidisers can be considered first colonizers. This could include natural systems which have been shown to harbour diverse microbial communities including glacial forefields, volcanic flows and meteorites [452,453] as well as anthropogenic systems such as open-pit mine tailings [454–456]. Following succession, microbial community assembly processes have been shown to differ over time [103,244]. Initially, communities are predicted to stochastically assemble showing little community structure due to weak competition and selection [86,114]. This should favour habitat generalists which can utilize a wide range of energy sources. Given our observation that trace gases such as H_2 and CO serve as a ubiquitous and dependable energy source for the majority of the community it is likely that metabolically flexible taxa are dominant during the initial stages of succession. It can also be predicted that in newly colonized habitats these organisms play a key role in the priority effect, whereby they change the abiotic environment to suit other taxa including those that are rare and highly specialised taxa [101]. Answers to these questions could provide further evidence that trace gas metabolism plays an important role in maintaining biodiversity and resilience of soil microbial communities.

Finally, an often overlooked terrestrial ecosystem are karst landforms or cave systems, which cover 10-20 % of terrestrial ecosystems and form as acidic water flows dissolve limestone rock [457]. Other processes such as volcanic lava flows and tidal erosion are among other processes driving cave formation [458]. These ecosystems provide a conduit between the atmosphere and a vast surface area of below ground chambers, tunnels and rock fractures. Studies have found that caves harbour diverse microbial communities which at the phylum level reflect those found in soils, dominated by Actinobacteriota, Acidobacteria and Proteobacteria [459,460]. They inhabit carbon poor mineral soils as well as form biofilms along mineral surfaces. Given complete absence of light energy, oxygenic photosynthesis is inhibited and communities have to rely on inorganic energy sources. Studies have found that genes supporting all six known CO_2 fixation pathways are widespread, including the chemosynthetic type IE RuBisCO implicated in the CBB cycle [67,227,460]. However, more recently studies suggest that caves in fact are a major sink for atmospheric trace gases such as CH_4 [461]. Given that many caves have similar atmospheric compositions to those above ground, these ecosystems could be an overlooked microbial sink for atmospheric trace gases such as H_2 and CO. Indeed, these gases may support the energy requirement for survival and growth in these permanently dark conditions and communities may harbour as of yet undiscovered lineages of genes supporting these processes.

5.2.4. Isolation of trace gas oxidisers

Following the isolation of the first high affinity H_2 oxidiser *Streptomyces sp PCB7* [175], the determinant for aerobic H_2 scavenging have become much clearer. Subsequent pure culture studies of other model organisms capable of this process have resolved many physiological, phylogenetic and enzymatic questions and demonstrated that this trait is widespread among dominant soil phyla [70,72,163]. Further isolation studies are needed to address this growing realization. Recent discoveries, based on culture independent approaches, have identified new H_2 and CO trace gas oxidisers in Antarctic soils such as the phyla *Candidatus* Dormibacterota and *Candidatus* Eremibacterota. Similarly, this work has described a putative methanotroph *Candidatus* Methylotropicum kingii from the phylum Gemmatimonadota, which highlights the need for isolation of these organisms. However, the isolation of many trace gas oxidisers may not be trivial, given that they are likely to grow mixotrophically,

by using electrons derived from H_2 and CO to drive aerobic respiration and in some cases carbon fixation [149]. In addition to some of the approaches described previously for H_2 oxidisers such as dynamic microcosm chamber which enriched for high affinity H_2 oxidisers [175] and methanotrophs [76], techniques based on cell size fractionation and minimal agar media have also shown promise [462,463]. Future studies are needed to develop techniques to enrich for and isolate such organisms into pure culture, which will no doubt advance our understanding about their physiology, metabolism and ecological significance.

6. Appendices

6.1. Appendix A: Chapter 2

Soil bacterial communities exhibit strong biogeographic patterns at fine taxonomic resolution

Sean K. Bay[1,2 *], Melodie A. McGeoch[1], Osnat Gillor[3], Nimrod Wieler[3], David J. Palmer[1], David J. Baker[1], Steven L. Chown[1], Chris Greening[1,2 *]

[1] School of Biological Sciences, Monash University, Clayton, VIC 3800, Australia

[2] Department of Microbiology, Biomedicine Discovery Institute, Clayton, VIC 3800, Australia

[3] Department of Environmental Hydrology and Microbiology, Ben Gurion University of the Negev, Sde Boker, Israel

***** Correspondence can be addressed to:

A/Prof Chris Greening (chris.greening@monash.edu), Department of Microbiology, Monash University, Clayton, VIC 3800, Australia

Sean Bay (sean.bay@monash.edu), Department of Microbiology, Monash University, Clayton, VIC 3800, Australia

6.1.1. Supplementary Figures

Figure S1. Study site in the Judea Hills and Negev region of Israel. The top left panel shows the study site (blue enclosed circle) within a world map shaded by the aridity index. The right panel shows the locations of the sampling sites. Samples were collected across a 160 km latitudinal (north/south) transect and a 20 km longitudinal (east/west) transect. The latitudinal transect occurred along a steep aridity gradient and samples were collected from climatic zones (sub-humid, semi-arid, arid, hyper-arid). The photographs on the right were taken during the sampling campaign, showing typical vegetation and geographic features of each climatic zone. For the longitudinal transect, all samples were collected in the arid zone.

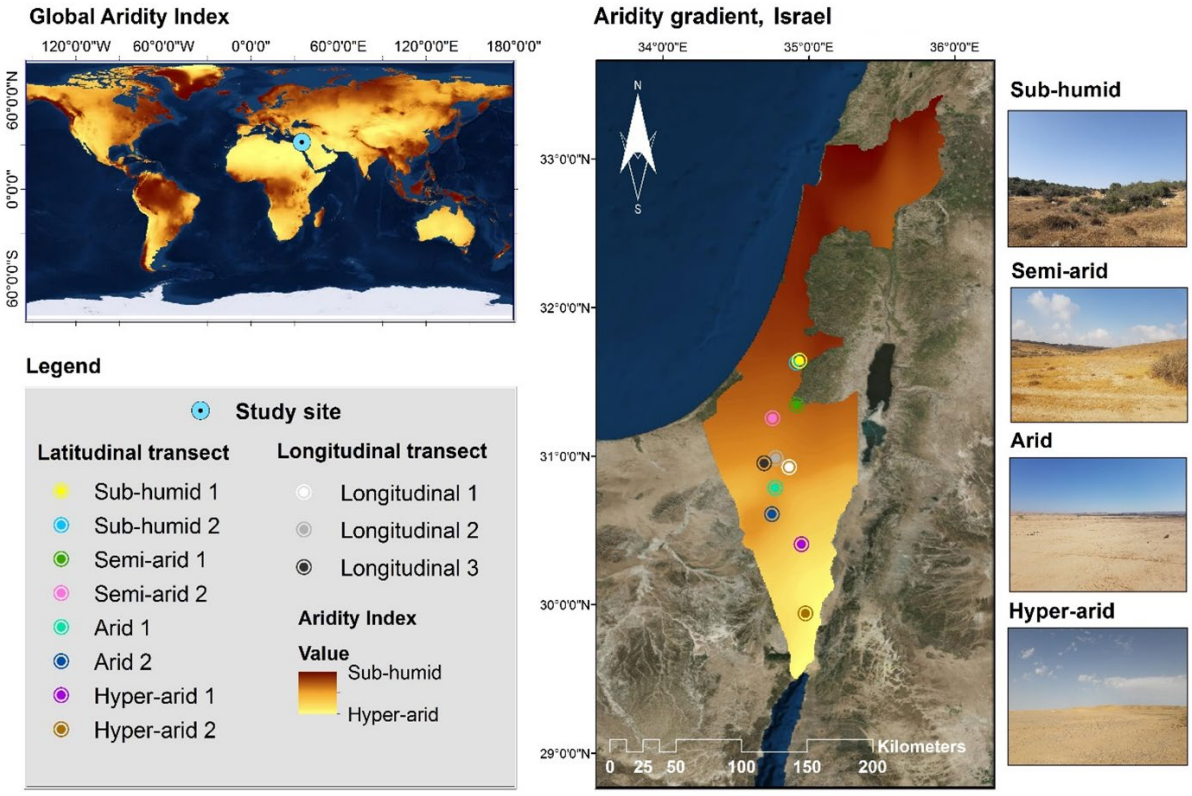


Figure S2. Details of the hierarchically nested sampling design. Samples were collected from four climatic zones along the aridity gradient (sub-humid, semi-arid, arid, hyper-arid). There were three hierarchies of spatial sampling within each climatic zone: (i) two sites were sampled at each zone (Site 1, Site 2); (ii) three different plots were sampled at each site (Plot A, Plot B, Plot C); and (iii) three random soil samples were collected from each plot (Sample 1, Sample 2, Sample 3).

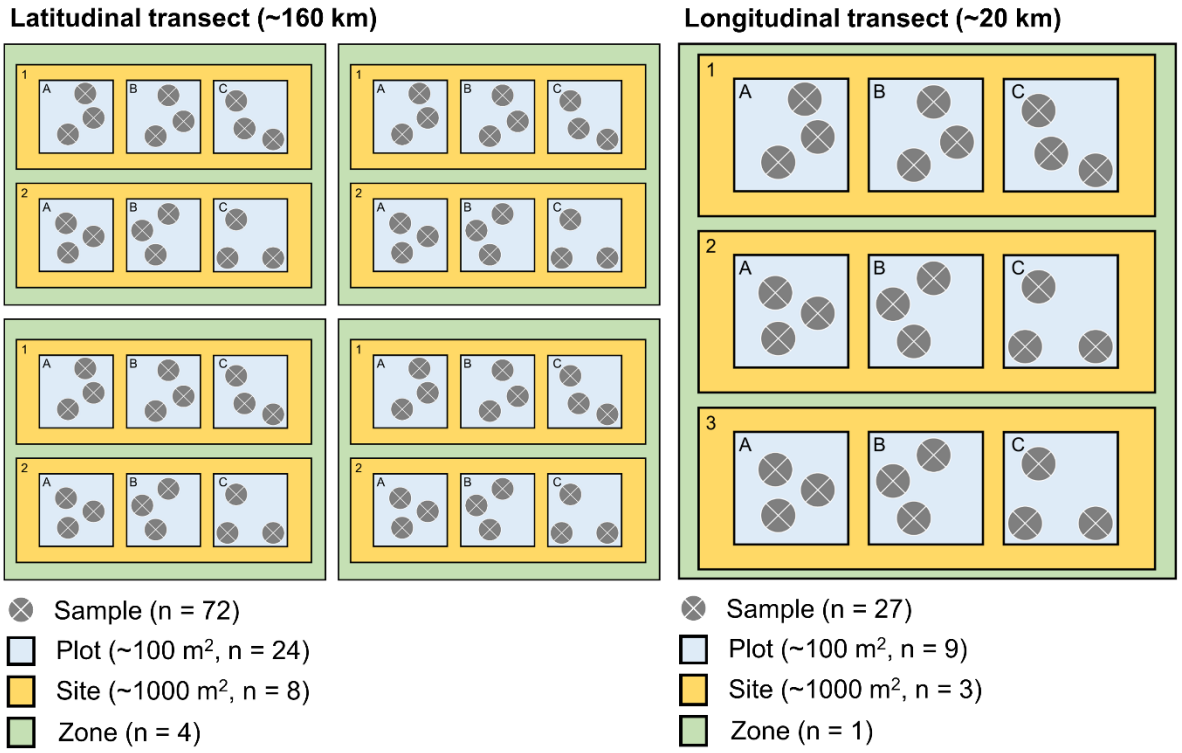


Figure S3. Frequency distribution of sampling distances between the 96 sampling sites. Pairwise distances are showed at a resolution of 10 km intervals. This analysis confirms that most distance classes were represented (unrepresented bins: 110 km and 140 km) across the study site and that all represented distance classes were associated with >100 site pairs (min = zero, 1st q = 24.6 km, median = 42.5 km, mean = 50.5 km, 3rd q = 68.8 km, max = 155.5 km).

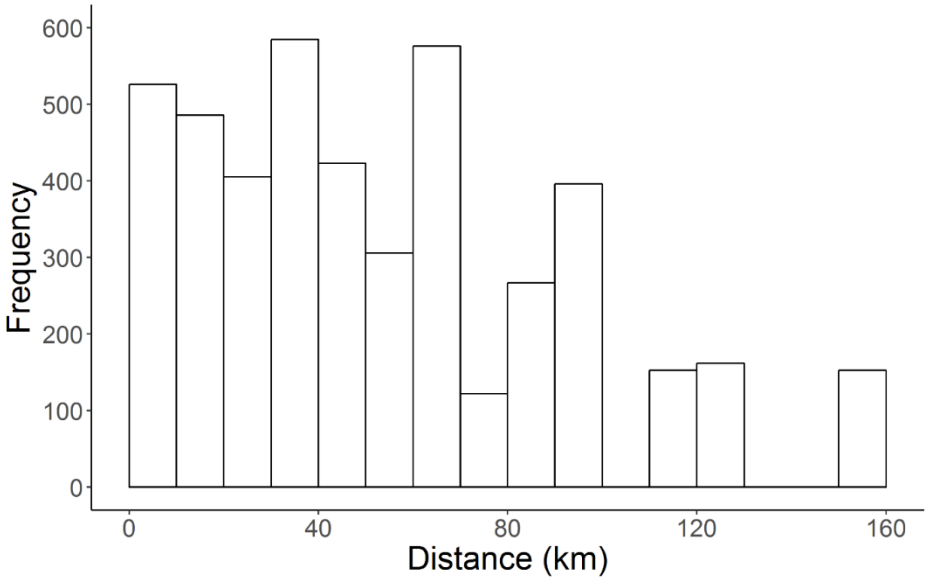


Figure S4. Sample-based alpha rarefaction curves. The curves show number of taxa (ASVs) detected relative to number of sequencing reads at identity thresholds of 100% **(a)** and 97% **(b)**.

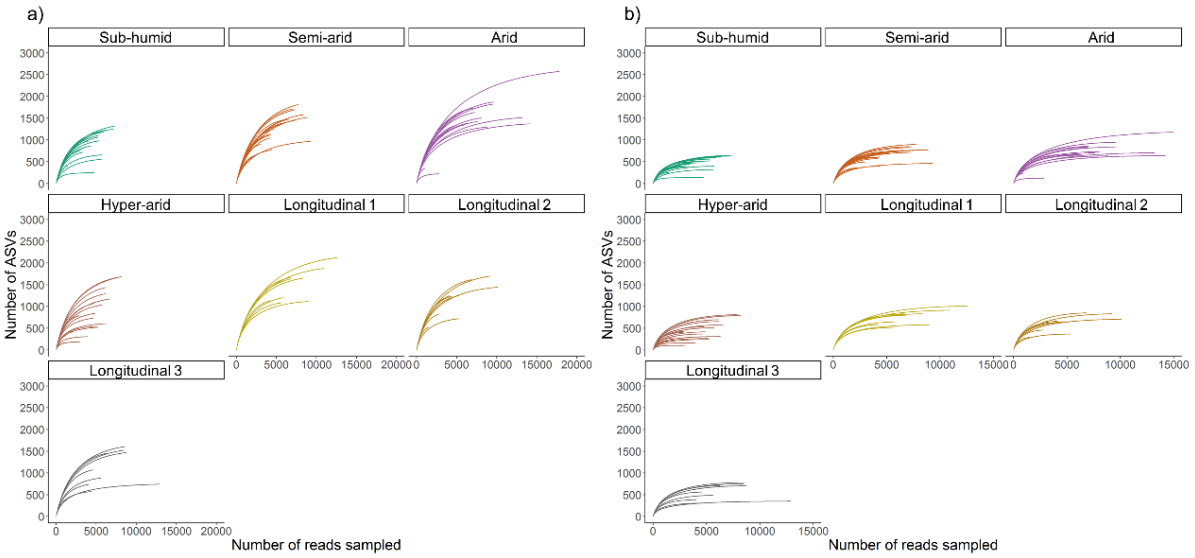


Figure S5. Taxa accumulation curve showing number of taxa (ASVs) detected across all sites. Points show cumulative sample richness and error-bars show the estimated standard deviation at identity thresholds of 100% and 97%.

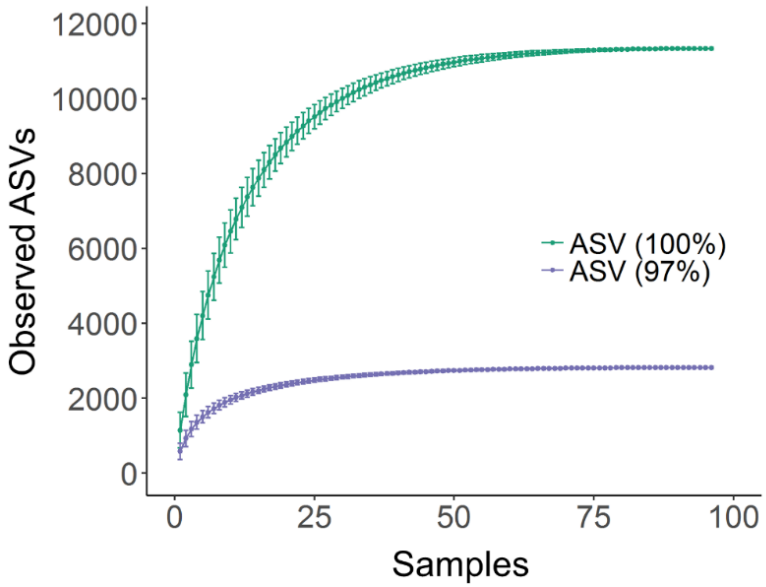


Figure S6. Observed and estimated community richness. The site-level (1 – 2) observed and estimated richness of taxa (ASVs) is shown across four climatic zone, sub-humid (SH), semi-arid (SA), arid (AR), hyper-arid (HA) and three sites along the longitudinal transet (LO 1-3). Estimated richness was calculated using the Chao1 method and abundance coverage estimate (ACE). Box plots show the median, upper and lower quartiles at two taxonomic resolutions. **Table S2** summarizes analysis of variance (ANOVA) results testing for significant differences among site means.

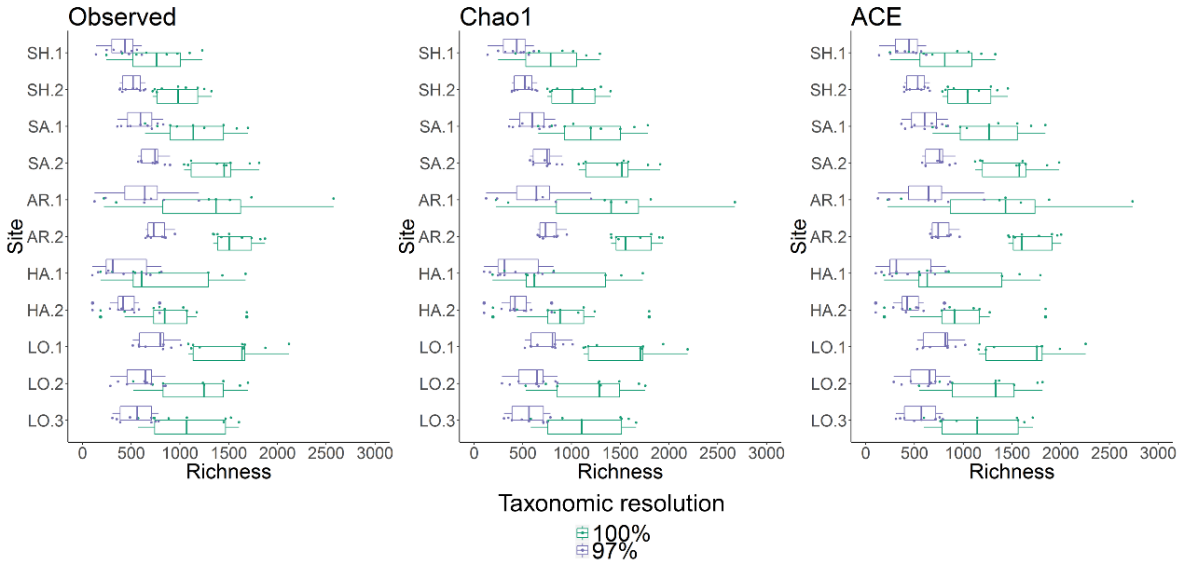


Figure S7. Histograms showing occupancy frequency distribution of taxa. Stacked histograms show the number of sites that each taxa (ASV) was detected in across the dataset, with **(a)** effect of clustering taxa at either 100% or 97% identity threshold and **(b)** effect of either including or removing taxa with lower than 0.05% relative abundance. This data was used to produce the Kernel-smoothed density plots shown in **Figure 1**.

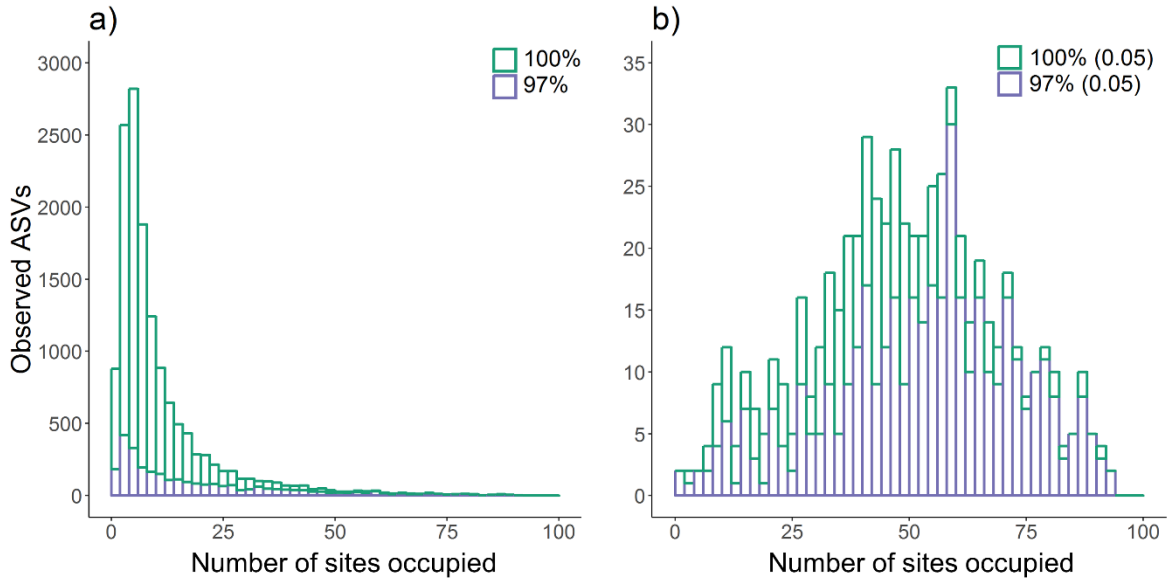


Figure S8. Phylum-level community structure. Bars represent relative abundance of bacteria and archaea detected by amplicon sequencing of 16S rRNA genes (100% ASVs). Sample identity are given along with corresponding climatic zone and sampling transect (for climate zones and sampling transect refer to **Figure S1**). Sequences were assigned based on Genome Taxonomy Database (GTDB) taxonomy.

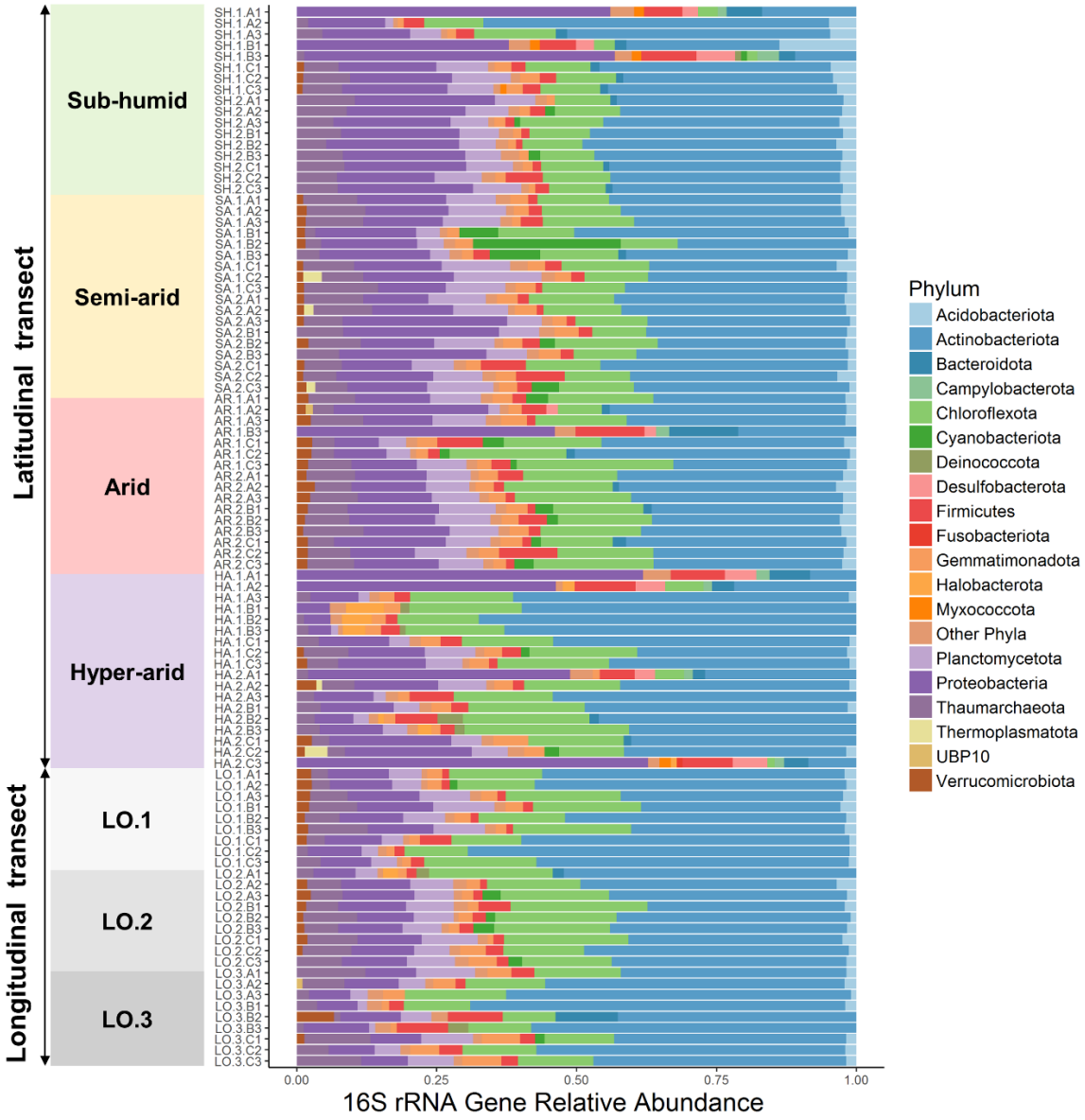
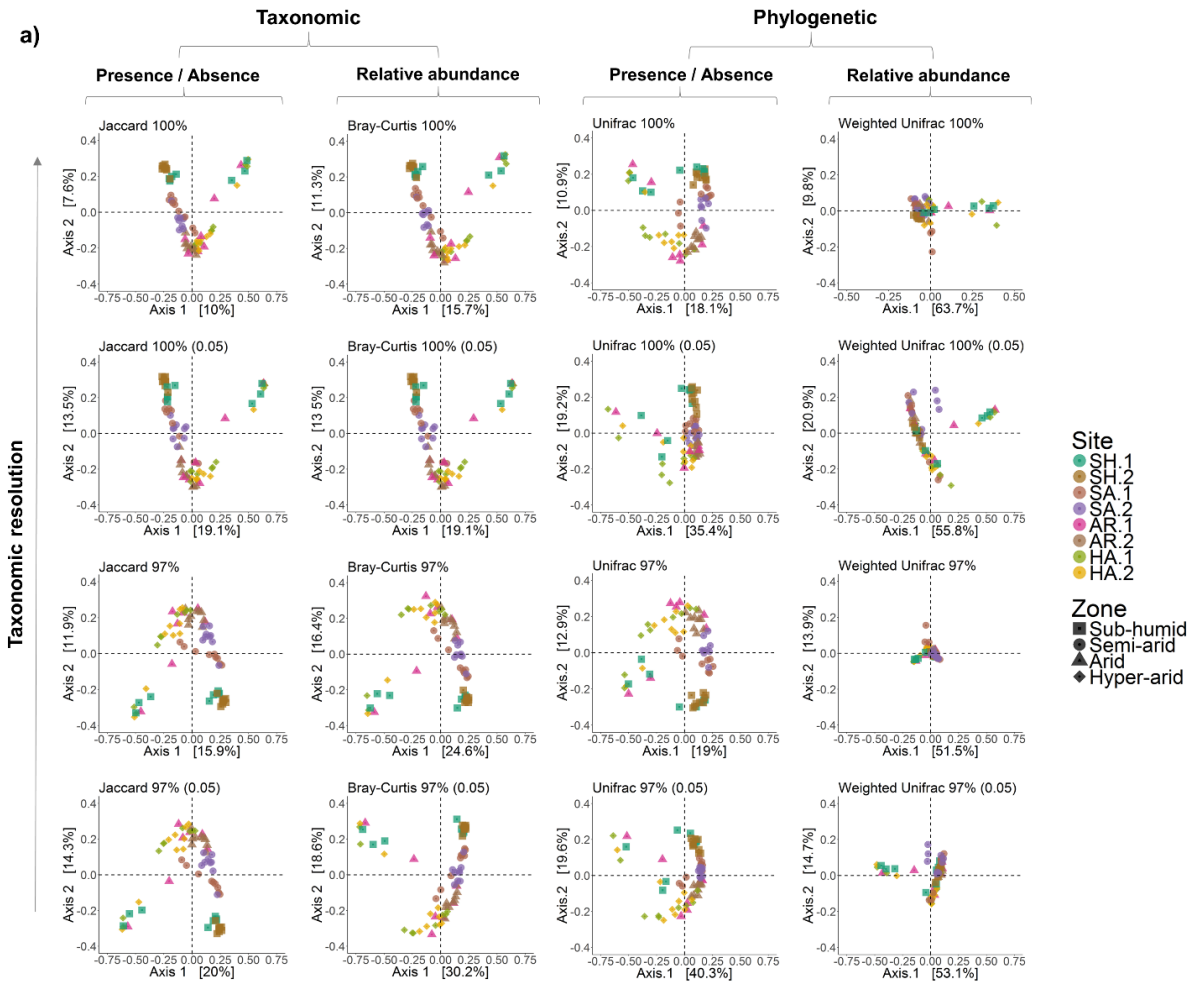


Figure S9. Multidimensional scaling visualizing taxonomic and phylogenetic pairwise incidence and abundance dissimilarity of microbial communities. Results are shown across the latitudinal transect **(a)** and longitudinal transect **(b)**. Each data point shows an individual sample. The axes show the explained variation of taxa (ASVs) between samples using four different dissimilarity metrics: Jaccard (taxonomic incidence-based) and Bray-Curtis (taxonomic abundance-based) and unweighted Unifrac (phylogenetic incidence-based) and weighted Unifrac (phylogenetic abundance-based). The MDS ordination is compared at four different taxonomic resolutions (taxa clustered at 100% or 97% identity; taxa with <0.05% relative abundance included or removed). Statistical tests showing significant differences in community composition at the zone, site, and plot levels are shown in **Table S3.**



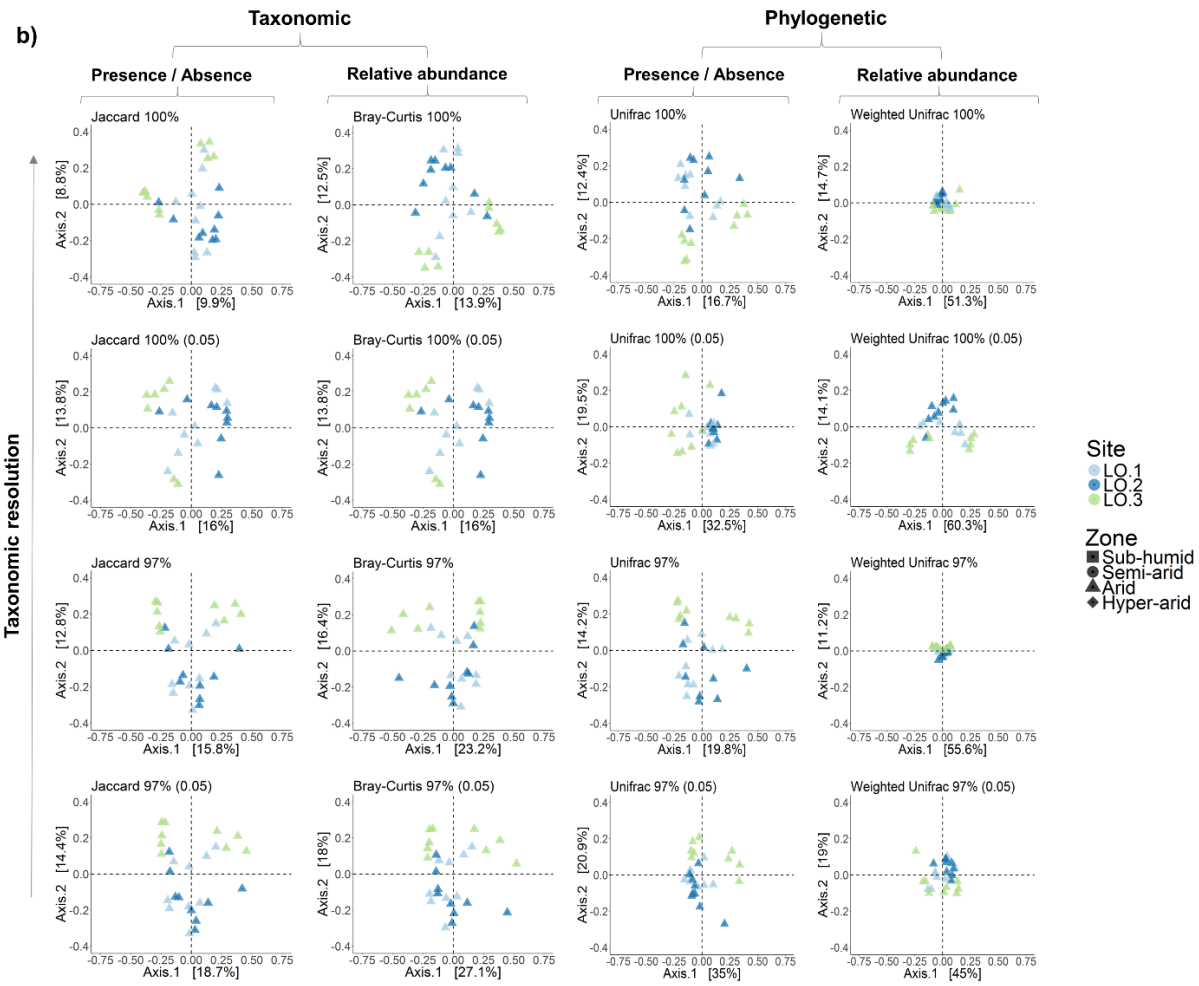

b)
Taxonomic
Phylogenetic
Presence / Absence
Relative abundance
Presence / Absence
Relative abundance
Taxonomic resolution
Jaccard 100%
Bray-Curtis 100%
Unifrac 100%
Weighted Unifrac 100%
Jaccard 100% (0.05)
Bray-Curtis 100% (0.05)
Unifrac 100% (0.05)
Weighted Unifrac 100% (0.05)
Jaccard 97%
Bray-Curtis 97%
Unifrac 97%
Weighted Unifrac 97%
Jaccard 97% (0.05)
Bray-Curtis 97% (0.05)
Unifrac 97% (0.05)
Weighted Unifrac 97% (0.05)
Axis.1 [9.9%]
Axis.2 [8.8%]
Axis.1 [13.9%]
Axis.2 [12.5%]
Axis.1 [16.7%]
Axis.2 [12.4%]
Axis.1 [51.3%]
Axis.2 [14.7%]
Axis.1 [16%]
Axis.2 [13.8%]
Axis.1 [16%]
Axis.2 [13.8%]
Axis.1 [32.5%]
Axis.2 [19.5%]
Axis.1 [60.3%]
Axis.2 [14.1%]
Axis.1 [15.8%]
Axis.2 [12.8%]
Axis.1 [23.2%]
Axis.2 [16.4%]
Axis.1 [19.8%]
Axis.2 [14.2%]
Axis.1 [55.6%]
Axis.2 [11.2%]
Axis.1 [18.7%]
Axis.2 [14.4%]
Axis.1 [27.1%]
Axis.2 [18%]
Axis.1 [35%]
Axis.2 [20.9%]
Axis.1 [45%]
Axis.2 [19%]
Site
LO.1
LO.2
LO.3
Zone
Sub-humid
Semi-arid
Arid
Hyper-arid

Figure S10. Infographic describing zeta diversity. This shows how zeta diversity, unlike beta diversity, can provide information on the contribution of rare, intermediate and common taxa to community turnover. A complementary infographic showing the effects of clustering and filtering on zeta decline and zeta distance decay, within the context of this study, is shown in **Figure 4**.

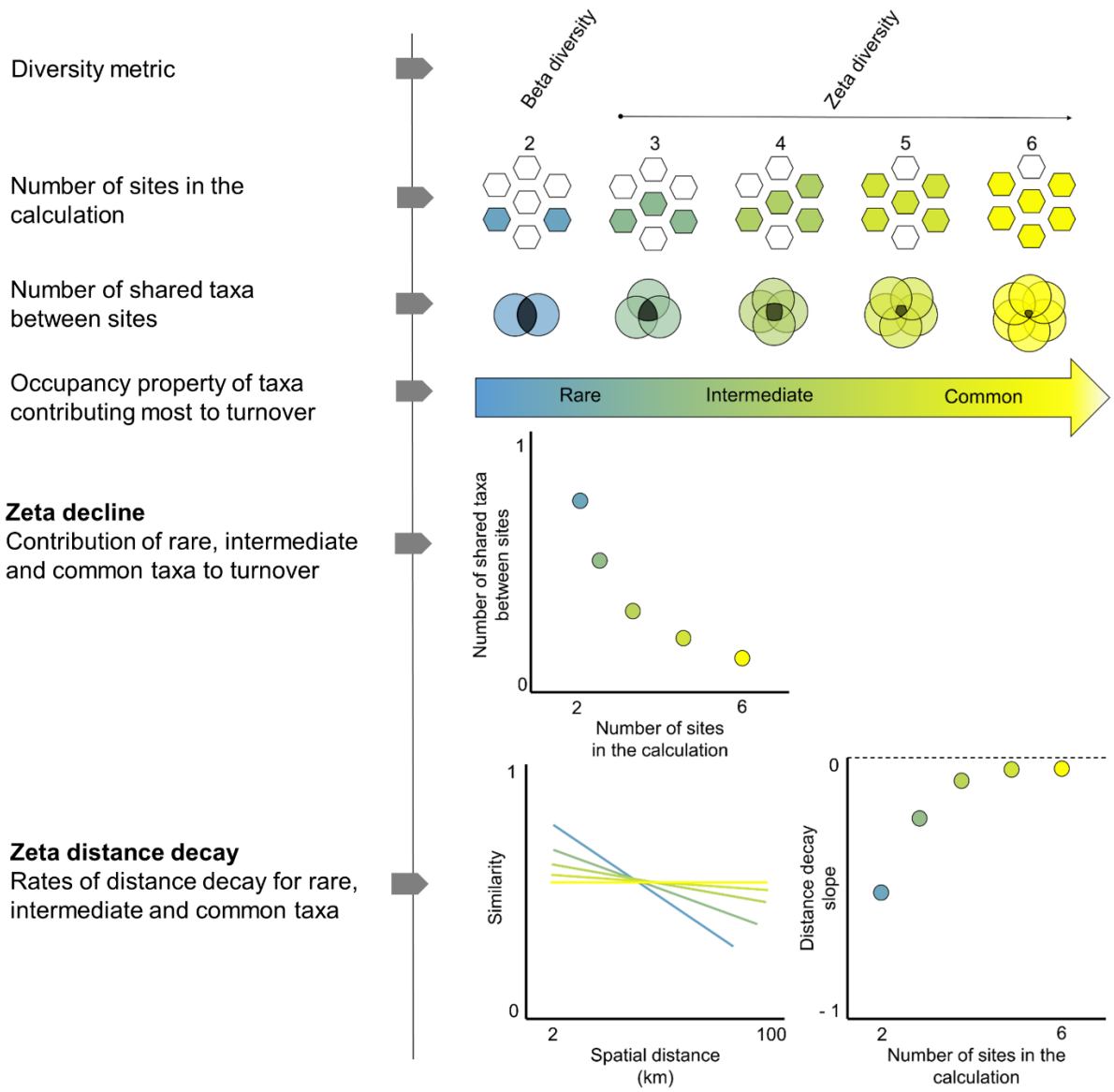


Figure S11. Variation partitioning analysis delineating the relative contributions of environmental and spatial sources of variation on microbial community structure. The analysis shows the proportion of variation in microbial incidence between sample pairs (zeta 2) as explained by environmental, spatial, overlapping, and unexplained sources of variation. These analyses were performed using data from each plot in the **(a)** latitudinal transect and **(b)** longitudinal transect. Results are compared at four different taxonomic resolutions (taxa clustered at 100% or 97% identity; taxa with <0.05% relative abundance included or removed). **Table S4** and **Table S5** summarize the environmental variables that best explain the variation along the two transects.

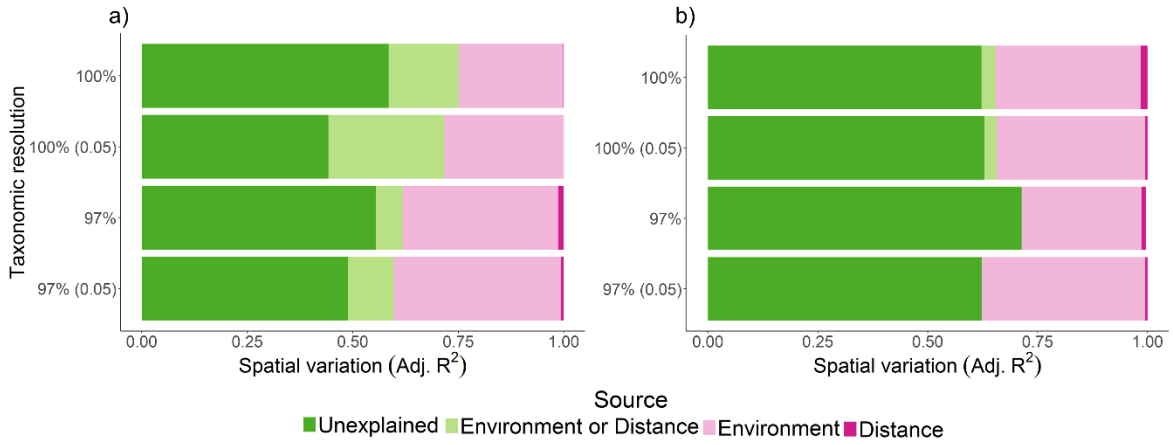


Figure S12. Normalised zeta diversity decline showing the compositional turnover in taxa (ASVs) across sites at different taxonomic resolutions. Zeta decline (**a, d, g, j**) quantifies how the number of shared taxa declines with increasing orders of zeta (number of sites included in the calculate of zeta). The functional decline frequently follows either an exponential (equal probability of taxa occurrence across sites) or a power law form (unequal probability of taxa occurrence across sites), which reflects turnover being driven largely by either stochastic (exponential) and deterministic (power-law) community assembly processes. In all cases, the decline followed a power law form (**b, e, h, k**), though the goodness of fit varied depending on taxonomic resolution **(Fig. S13)**. The taxa retention rate using zeta ratio (**c, f, i, l**) quantifies the probability of retaining common over rare taxa at any particular order with the addition an extra site.

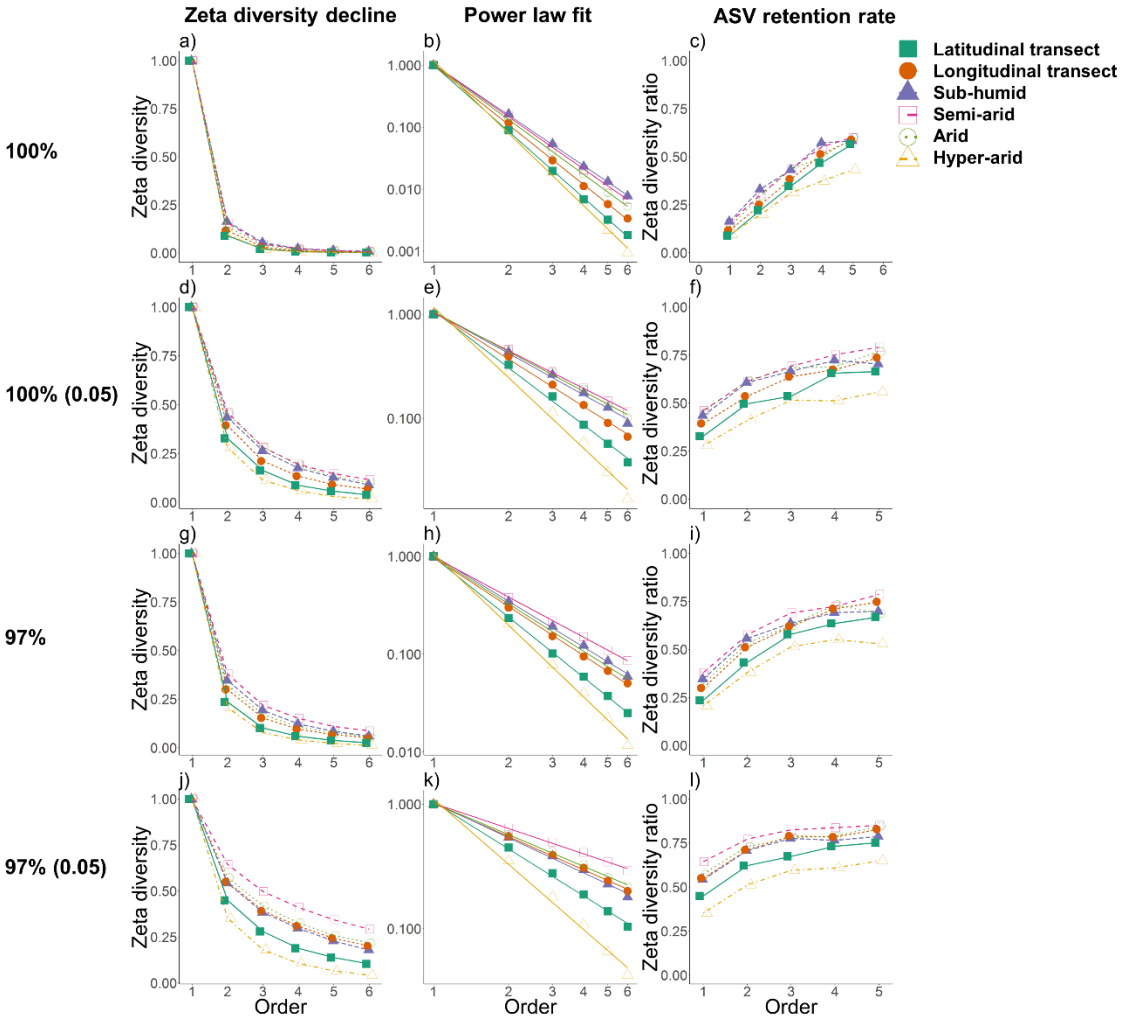


Figure S13. Statistical support for power law and exponential model fits of zeta diversity decline. The bar plots show AIC values of power law and exponential general linear model fits for the two transects (latitudinal and longitudinal) and four climatic zones (from sub-humid to hyper-arid). Results are compared at four different taxonomic resolutions (taxa clustered at 100% or 97% identity; taxa with <0.05% relative abundance included or removed).

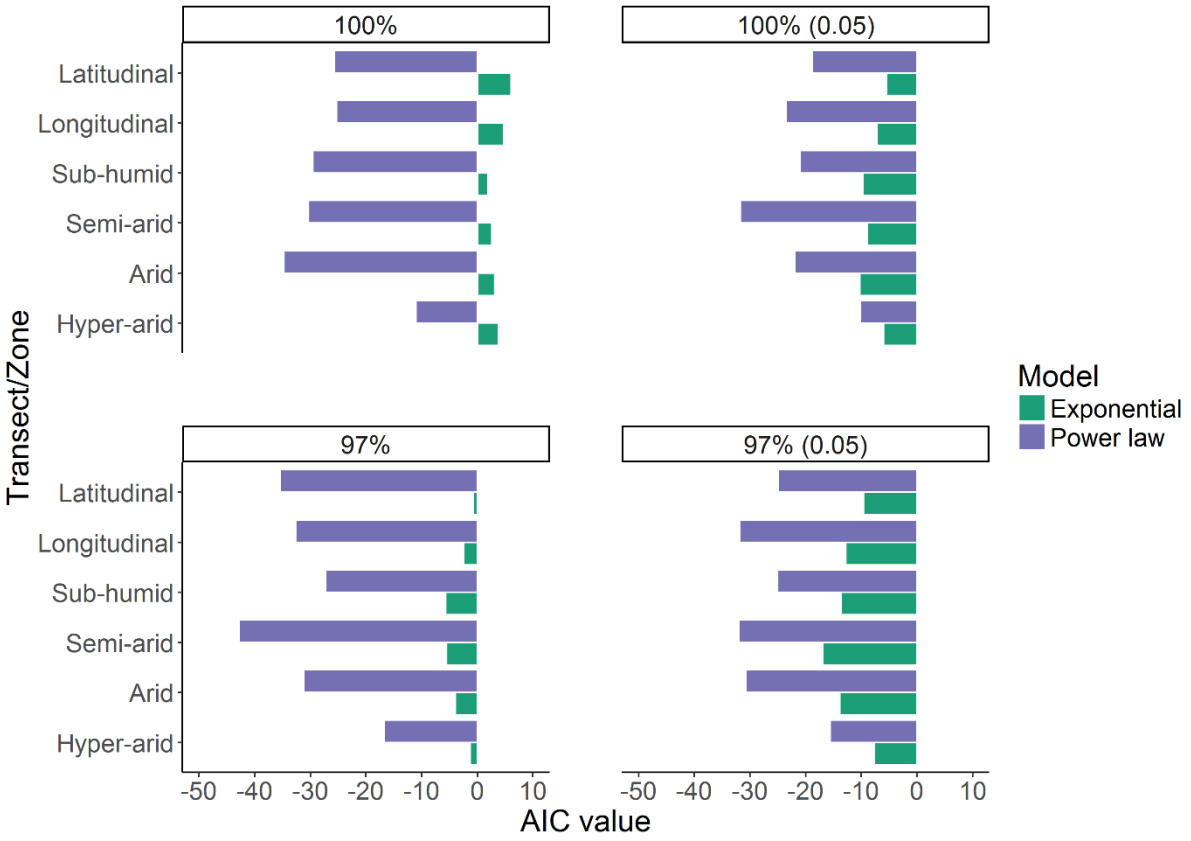


Figure S14. Measurement of distance decay relationship and taxa-area relationship at different taxonomic resolutions for longitudinal transect. Distance decay relationship showing **(a)** community turnover with increasing geographic distance based on pairwise comparisons and **(b)** differences in the slope (coefficient) of distance decay when moving from pairwise comparisons to higher orders of zeta (>2) using the average distance between higher orders of zeta. **(c)** Taxa-area relationship showing increase in richness with area sampled.

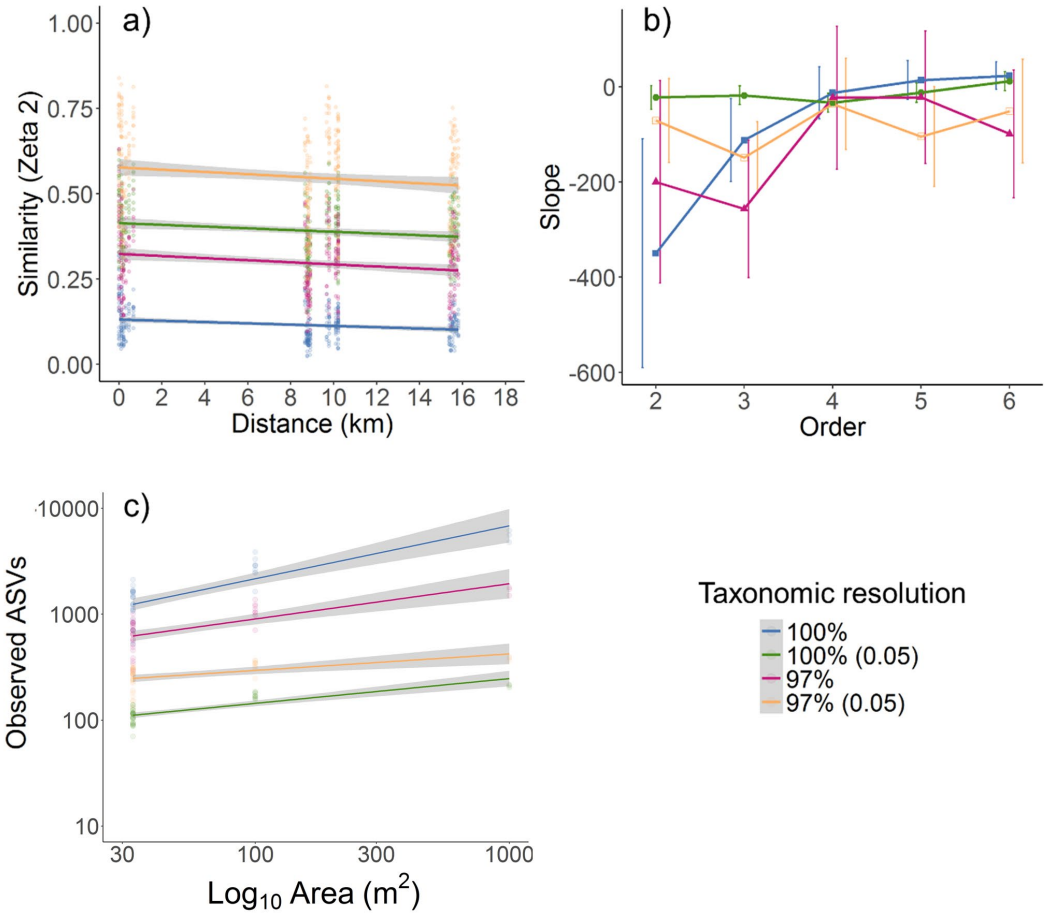


Figure S15. Comparison of community diversity, composition, and turnover based on amplicon and metagenomic sequencing. The bacterial and archaeal community composition of 12 sites across the latitudinal gradient was determined by either amplicon sequencing of the multi-copy 16S rRNA gene V4 region or shotgun metagenomic sequencing of the single-copy ribosomal protein gene L10e/L16. **(a)** Observed and estimated richness. Estimated richness was calculated using the Chao1 method and abundance coverage estimate (ACE). Box plots show the median, upper and lower quartiles at two taxonomic resolutions. **(b)** Phylum-level community composition across the sites for 16S and L10e/L16. Taxonomic assignment is based on the Genome Taxonomy Database (GTDB). **(c)** Normalized zeta diversity decline showing compositional turnover in taxa from 16S ASVs at two identity thresholds (100%, and 97%) and the single copy ribosomal marker gene L10e/L16 rpsP clustered at 97%.

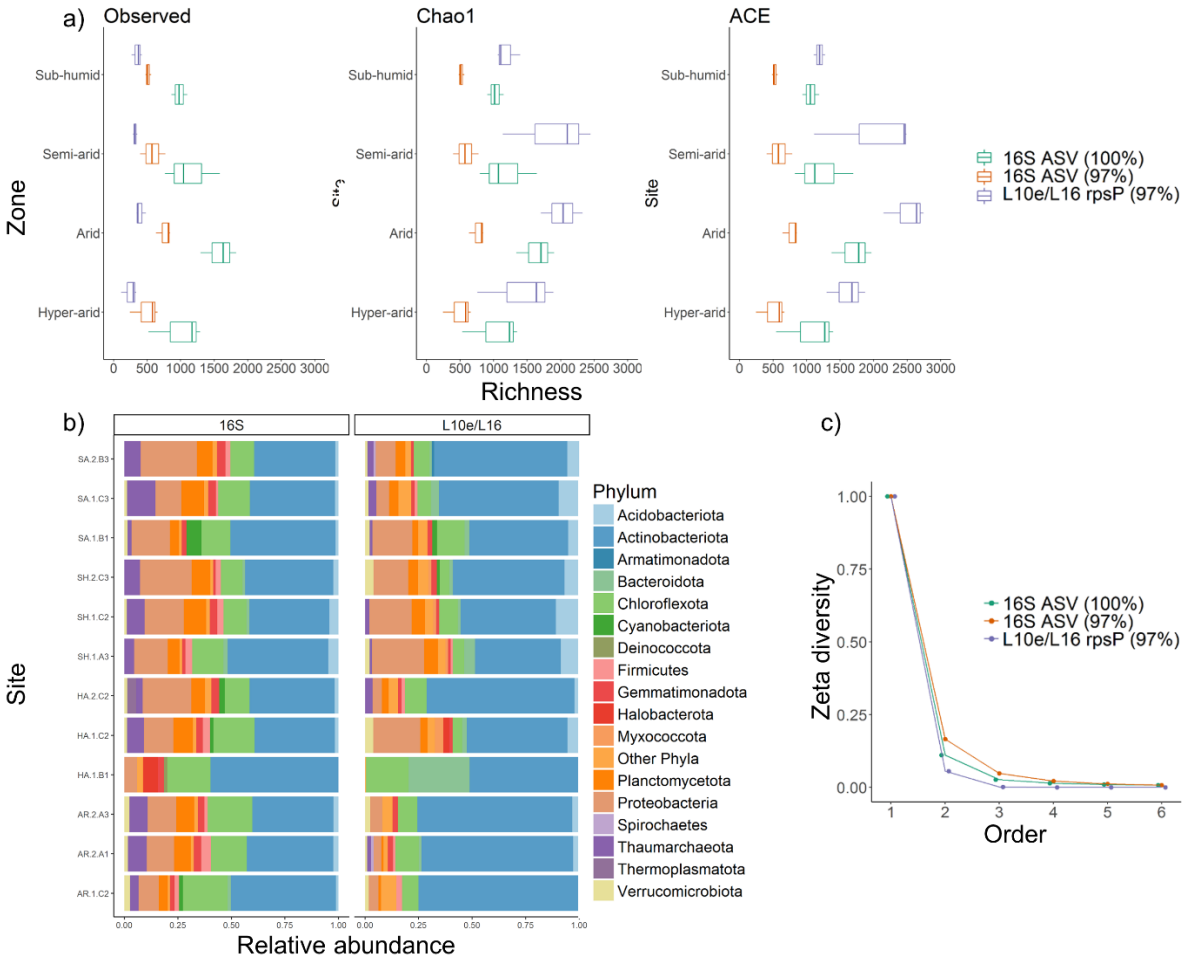
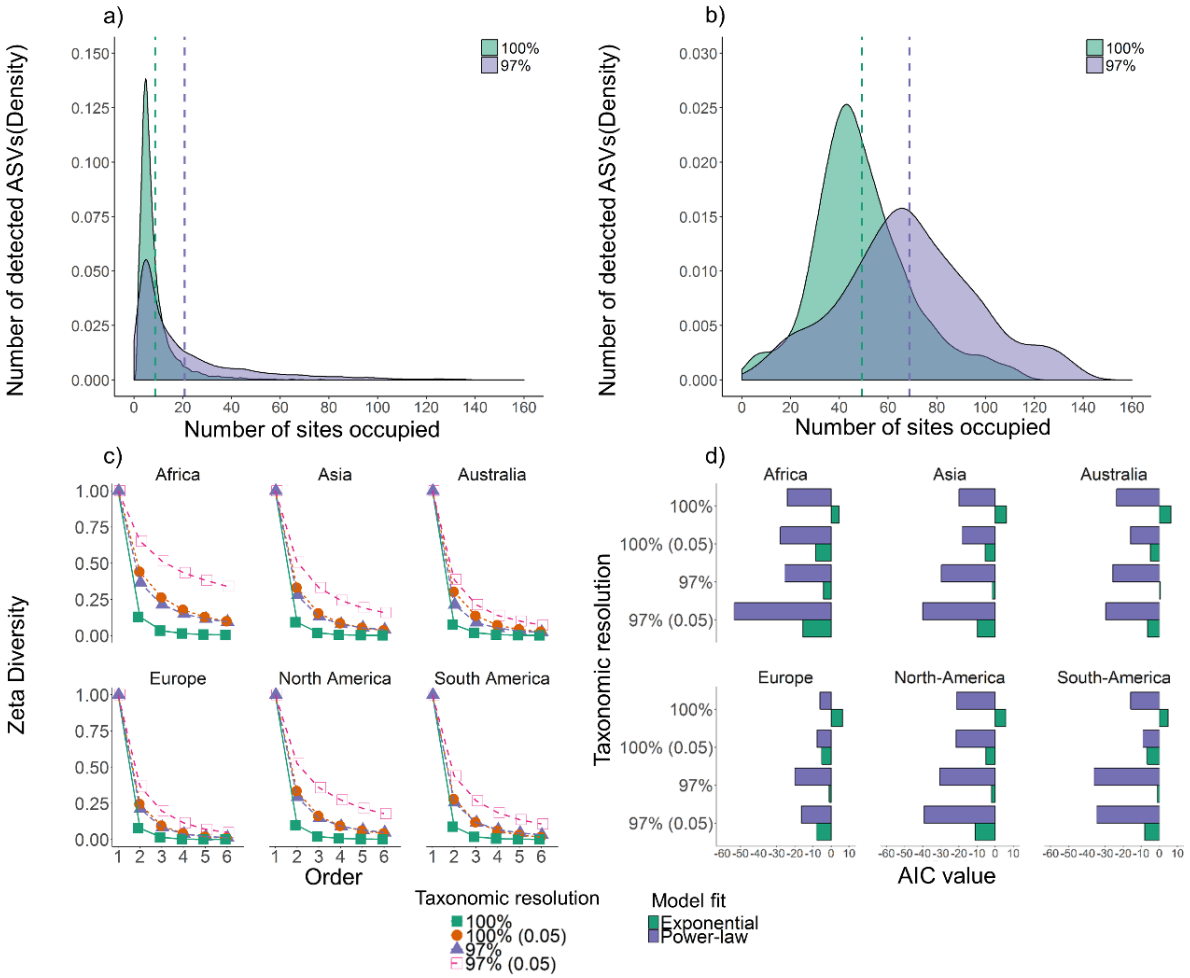


Figure S16. Compositional turnover analysis at the continental scale. (a-b) Global scale occupancy frequency at two identity thresholds (100% and 97%). Taxa (ASVs) with lower than 0.05% relative abundance were retained in **a** and removed in **b**. **(c)** Normalized zeta decline at the continental scale across two identity thresholds (100% and 97%) and with rare variants removed (0.05%). **(d)** AIC values for exponential and power-law model fits.



6.1.2. Supplementary Tables

Table S1 is available on the Google Drive folder under **Chapter 2**:
https://drive.google.com/open?id=16UnlZtajRSBZ_vSvJzw3Pl_XraDWUNV2

Table S1 (xlsx). Amplicon sequence variants detected in each sample. The sequences and read counts for each amplicon sequence variant detected is shown per sample.

Table S2. One-way ANOVA results. Showing between site differences (**Figure S6**) for observed and estimated alpha diversity at two taxonomic resolutions.

		Factor	Df	SS	MS	F-Value	P-value	Significance
100%	Observed	Between sites	10	6.507E+06	6.507E+05	3.764	3.220E-04	***
		Within sites	85	1.470E+07	1.729E+05			
	Chao1	Between sites	10	7.023E+06	7.023E+05	3.687	3.990E-04	***
		Within sites	85	1.619E+07	1.905E+05			
	ACE	Between sites	10	7.647E+06	7.647E+05	3.767	3.200E-04	***
		Within sites	85	1.726E+07	2.030E+05			
97%	Observed	Between sites	10	1.485E+06	1.485E+05	4.077	1.350E-04	***
		Within sites	85	3.096E+06	3.642E+04			
	Chao1	Between sites	10	1.497E+06	1.497E+05	4.07	1.380E-04	***
		Within sites	85	3.126E+06	3.677E+04			
	ACE	Between sites	10	1.533E+06	1.533E+05	4.061	1.410E-04	***
		Within sites	85	3.208E+06	3.774E+04			

Table S3. One-way ANOVA results. Showing between zone, site, plot and samples differences in community structure at four taxonomic resolutions using likelihood ratio test (LRT).

	100%			100% (0.05)			97%			97% (0.05)		
Factor	**LRT**	**P-value**	**Sig.**	**LRT**	**P-value**	**Sig.**	**LRT**	**P-value**	**Sig.**	**LRT**	**P-value**	**Sig.**
Latitudinal transect												
Between zones	87986	0.001	***	3315	0.001	***	32587	0.001	***	7254	0.001	***
Between sites	117500	0.001	***	6796	0.001	***	44150	0.001	***	10679	0.001	***
Between plots	140616	0.001	***	19792	0.001	***	70455	0.001	***	19792	0.001	***
Between samples	142019	0.368	ns	9837	0.337	ns	146414	0.188	ns	32794	0.341	ns
Longitudinal transect												
Between sites	9442	0.004	**	892	0.003	***	9442	0.003	**	1840	0.002	**
Between plots	23106	0.001	***	2737	0.005	**	23106	0.001	***	4992	0.001	***
Between samples	16375	0.078	ns	6712	0.048	*	54934	0.055	ns	10689	0.206	ns

Table S4. Subset of independent variables that best explain community variation along the latitudinal transect. Showing total R^2 obtained from a forward selection in a multivariate linear model. The results are shown for models of each of the four taxonomic resolutions.

100%		100% (0.05)		97%		97% (0.05)	
Predictor	**R^2**	**Predictor**	**R^2**	**Predictor**	**R^2**	**Predictor**	**R^2**
pH	0.05	Aridity (Aridity Index)	0.11	Soil water content (%)	0.16	Conductivity (dS/m)	0.15
Carbon/Nitrogen (ratio)	0.09	Conductivity (dS m^{-1})	0.20	pH	0.17	pH	0.25
Aridity (Aridity Index)	0.12	Soil water content (%)	0.24	Sodium (mg Kg^{-1})	0.20	Soil water content (%)	0.29
Conductivity (dS m^{-1})	0.14	Iron (mg Kg^{-1})	0.28	Iron (mg Kg^{-1})	0.22	Carbon/Nitrogen (ratio)	0.33
Total nitrogen (%)	0.16	Phosphate (mg Kg^{-1})	0.30	Total organic carbon (%)	0.25	Aridity (Aridity Index)	0.36
Boron (mg Kg^{-1})	0.18	Sodium (mg Kg^{-1})	0.34	Aridity (Aridity Index)	0.27	Total organic carbon (%)	0.38
Cation Exchange Capacity ($cmol^+$ Kg^{-1})	0.20	Sulphate (mg Kg^{-1})	0.36	Boron (mg Kg^{-1})	0.29	Manganese (mg Kg^{-1})	0.41
Copper (mg Kg^{-1})	0.22	Total organic carbon (%)	0.37	Cation Exchange Capacity ($cmol^+$/Kg)	0.31	P (mg Kg^{-1})	0.43
Calcium/Magnesium (ratio)	0.24	Manganese (mg Kg^{-1})	0.38	Potassium (mg Kg^{-1})	0.33	Ammonium (mg Kg^{-1})	0.45
Magnesium (mg Kg^{-1})	0.25	pH	0.40	Zink (mg Kg^{-1})	0.35	Nitrogen (mg Kg^{-1})	0.46

Table S5. Subset of independent variables that best explain community variation along the longitudinal transect. Results show total R^2 obtained from a forward selection in a multivariate linear model. The results are shown for models of each of the four taxonomic resolutions.

100%		**100% (0.05)**		**97%**		**97% (0.05)**	
Predictor	**R^2**	**Predictor**	**R^2**	**Predictor**	**R^2**	**Predictor**	**R^2**
Manganese (mg Kg^{-1})	0.24	Soil water content (%)	0.14	Ammonium (mg Kg^{-1})	0.30	Ammonium (mg Kg^{-1})	0.16
Carbon/Nitrogen (ratio)	0.29	Manganese (mg Kg^{-1})	0.20	Manganese (mg Kg^{-1})	0.35	Soil Water Content (%)	0.24
pH	0.33	Nitrate (mg Kg^{-1})	0.25	Nitrate (mg Kg^{-1})	0.39	Iron (mg Kg^{-1})	0.29
Total Nitrogen (%)	0.37	Silicon (mg Kg^{-1})	0.29	Iron (mg Kg^{-1})	0.43	Total Nitrogen (%)	0.34
Zinc (mg Kg^{-1})	0.41	Ammonium (mg Kg^{-1})	0.33	Soil Water Content (%)	0.46	Aluminum (mg Kg^{-1})	0.37
Conductivity (dS m^{-1})	0.44	Total Nitrogen (%)	0.35	Copper (mg Kg^{-1})	0.49	Silicon (mg Kg^{-1})	0.41
Potassium (mg Kg^{-1})	0.47	-		Calcium (mg Kg^{-1})	0.52	Calcium (mg Kg^{-1})	0.44
Total Organic Carbon (%)	0.49	-		Total Organic Carbon (%)	0.54	Conductivity (dS m^{-1})	0.47
Ammonium (mg Kg^{-1})	0.50	-		Sodium (mg Kg^{-1})	0.55	Manganese (mg Kg^{-1})	0.47
Nitrate (mg Kg^{-1})	0.50	-		pH	0.55	Nitrate. (mg Kg^{-1})	0.51

Table S6. Summary statistics of general linear models of zeta diversity distance decay. Results are shown for the latitudinal and longitudinal transect for each of the four taxonomic resolutions.

Taxonomic resolution	Transect	Zeta order	Model coeff.	SE	R^2	t	Pr(>|t|)	Significance
100%	Latitudinal	ζ_2	-5.51E-04	3.53E-05	0.196	-15.62	<2e-16	***
100%	Longitudinal	ζ_2	-1.87E-03	4.51E-04	0.044	-4.15	4.23E-05	***
97%	Latitudinal	ζ_2	-9.67E-04	7.24E-05	0.151	-13.37	<2e-16	***
97%	Longitudinal	ζ_2	-3.06E-03	9.16E-04	0.028	-3.35	9.08E-04	***
100% (0.05)	Latitudinal	ζ_2	-1.00E-03	9.30E-05	0.103	-10.75	<2e-16	***
100% (0.05)	Longitudinal	ζ_2	-2.55E-03	8.27E-04	0.023	-3.08	2.24E-03	**
97% (0.05)	Latitudinal	ζ_2	-1.46E-03	1.20E-04	0.129	-12.19	<2e-16	***
97% (0.05)	Longitudinal	ζ_2	-3.33E-03	1.26E-03	0.016	-2.64	8.71E-03	**

Table S7. Comparative analysis of values for exponent *z* of the species-area relationship previously reported for both eukaryotic and prokaryotic communities in comparison to this study. Studies ranked by z value.

z value	Community	Identity (%)	Habitat	Method	Contiguous?	Study	Significance	R^2
0.39	Bacteria	100	Desert	16S Seq.	N	This study (lat.)	***	0.4
0.4	Bacteria	100	Desert	16S Seq.	N	This study (long.)	****	0.7
0.47	Bacteria		Rainforest	T-RFLP	Y	[257]		
0.4	Bacteria		Sediment	16S Seq.	Y	[258]		
0.27	Bacteria	97	Desert	16S Seq.	N	This study (lat.)	***	0.3
0.21	Bacteria	97	Desert	16S Seq.	N	This study (long.)	****	0.5
0.26	Bacteria		Tree hole	DGGE	N	[250]		
0.19	Bacteria	100 (0.05)	Desert	16S Seq.	N	This study (lat.)	***	0.3
0.15	Bacteria	100 (0.05)	Desert	16S Seq.	N	This study (long.)	****	0.7
0.13	Bacteria	97 (0.05)	Desert	16S Seq.	N	This study (lat.)	***	0.3
0.09	Bacteria	97 (0.05)	Desert	16S Seq.	N	This study (long.)	****	0.4
0.16	Bacteria		Lake	DGGE	N	[259]		
0.062	Bacteria		Soil	DGGE	Y	[251]		
0.04	Bacteria	99	Salt-marsh	16S Seq.	Y	[237]		
0.03	Bacteria		Various	T-RFLP	N	[253]		
0.02	Bacteria	97	Salt-marsh	16S Seq.	Y	[237]		
0.009	Bacteria		Lake	T-RFLP	Y	[254]		
0.008	Bacteria	95	Salt-marsh	16S Seq.	Y	[237]		
0.006	Bacteria		Soil	T-RFLP	N	[252]		

Table S8. Geographic and chemical characteristics of soils collected along the latitudinal and longitudinal transects

Plot	Zone	Transect	Latitude	Longitude	Aridity Index	Calcium (mg Kg^{-1})	Magnesium (mg Kg^{-1})	Potassium (mg Kg^{-1})	Phosphate (mg Kg^{-1})
SH.1.A	Sub-humid	Latitudinal	31.642	34.935	0.287	8964.000	303.430	96.580	67.560
SH.1.B	Sub-humid	Latitudinal	31.642	34.934	0.287	0.000	283.530	34.180	25.280
SH.1.C	Sub-humid	Latitudinal	31.642	34.933	0.287	9216.000	304.950	178.630	84.950
SH.2.A	Sub-humid	Latitudinal	31.630	34.916	0.287	9184.500	249.470	117.680	102.230
SH.2.B	Sub-humid	Latitudinal	31.630	34.916	0.287	9198.000	250.230	73.830	71.230
SH.2.C	Sub-humid	Latitudinal	31.630	34.915	0.287	9292.500	240.640	103.430	57.520
SA.1.A	Semi-arid	Latitudinal	31.346	34.915	0.234	8500.500	320.390	78.130	28.860
SA.1.B	Semi-arid	Latitudinal	31.349	34.913	0.234	8703.000	199.120	68.380	5.020
SA.1.C	Semi-arid	Latitudinal	31.346	34.913	0.234	8586.000	318.060	145.980	36.240
SA.2.A	Semi-arid	Latitudinal	31.255	34.751	0.234	8599.500	0.000	106.130	27.290
SA.2.B	Semi-arid	Latitudinal	31.253	34.753	0.234	8338.500	324.950	108.080	30.430
SA.2.C	Semi-arid	Latitudinal	31.254	34.752	0.234	8311.500	345.470	190.630	29.880
AR.1.A	Arid	Latitudinal	30.786	34.767	0.068	8311.500	313.930	131.880	1.410
AR.1.B	Arid	Latitudinal	30.787	34.766	0.068	10926.000	351.980	148.280	1.770
AR.1.C	Arid	Latitudinal	30.785	34.767	0.068	8550.000	247.480	150.830	1.580
AR.2.A	Arid	Latitudinal	30.609	34.746	0.068	8563.500	265.760	152.930	1.650
AR.2.B	Arid	Latitudinal	30.609	34.747	0.068	8464.500	293.220	189.980	1.900
AR.2.C	Arid	Latitudinal	30.608	34.745	0.068	8689.500	302.530	144.580	1.880
HA.1.A	Hyper-arid	Latitudinal	30.406	34.951	0.031	9319.500	409.020	76.430	1.880
HA.1.B	Hyper-arid	Latitudinal	30.406	34.951	0.031	9396.000	464.840	54.380	1.710
HA.1.C	Hyper-arid	Latitudinal	30.405	34.949	0.031	8550.000	209.240	113.480	1.770
HA.2.A	Hyper-arid	Latitudinal	29.941	34.976	0.025	9702.000	411.020	99.280	1.830

HA.2.B	Hyper-arid	Latitudinal	29.941	34.975	0.025	8806.500	284.530	125.230	1.610
HA.2.C	Hyper-arid	Latitudinal	29.942	34.975	0.025	11142.000	456.620	96.680	1.740
LO.1.A	Arid	Longitudinal	30.926	34.852	0.080	8392.500	428.400	105.580	1.610
LO.1.B	Arid	Longitudinal	30.925	34.852	0.080	8185.500	359.150	115.130	54.650
LO.1.C	Arid	Longitudinal	30.927	34.854	0.080	8415.000	403.470	91.880	41.150
LO.2.A	Arid	Longitudinal	30.988	34.774	0.070	8473.500	473.910	100.230	1.290
LO.2.B	Arid	Longitudinal	30.987	34.774	0.070	8329.500	357.340	177.580	1.910
LO.2.C	Arid	Longitudinal	30.987	34.773	0.070	8280.000	320.060	169.130	1.840
LO.3.A	Arid	Longitudinal	30.951	34.691	0.070	8433.000	377.860	166.580	2.040
LO.3.B	Arid	Longitudinal	30.950	34.693	0.070	8388.000	304.430	113.780	2.030
LO.3.C	Arid	Longitudinal	30.952	34.692	0.070	8194.500	398.720	172.280	1.740

Plot	Nitrate (mg Kg^{-1})	Ammonium (mg Kg^{-1})	Sulfate (mg Kg^{-1})	pH	Conductivity (ds m^{-1})	Sodium (mg Kg^{-1})	Aluminum (mg Kg^{-1})	Cation Exchange Capacity ($cmol^{+}$ Kg^{1})	Calcium/ Magnesium ratio	Zink (mg Kg^{-1})
SH.1.A	3.060	11.750	13.560	7.900	0.190	44.330	2.330	53.970	12.880	0.860
SH.1.B	2.110	9.630	11.010	8.070	0.160	42.000	1.730	51.780	13.880	0.270
SH.1.C	8.740	17.610	19.900	7.740	0.260	40.660	2.360	65.660	17.540	1.820
SH.2.A	2.460	11.190	19.410	7.960	0.200	41.670	1.560	45.670	14.110	0.670
SH.2.B	2.370	12.240	20.570	7.960	0.200	38.370	1.450	47.500	15.400	0.560
SH.2.C	2.630	10.530	17.870	8.000	0.200	39.510	1.600	43.800	14.490	0.440
SA.1.A	1.050	3.400	10.660	8.550	0.130	52.570	1.420	33.620	9.000	0.140
SA.1.B	2.350	5.260	5.210	8.770	0.090	20.090	0.900	23.840	14.050	0.550
SA.1.C	6.690	4.800	22.710	8.490	0.200	106.300	1.520	32.720	8.760	0.340
SA.2.A	1.180	2.620	4.180	8.830	0.100	62.130	1.110	28.040	6.980	0.050

SA.2.B	1.150	2.340	6.470	8.800	0.110	59.410	0.840	27.720	8.360	0.060
SA.2.C	1.280	2.940	4.260	8.790	0.110	61.860	0.170	29.000	8.080	0.110
AR.1.A	2.450	1.940	8.620	8.740	0.120	54.890	0.250	28.260	8.380	0.150
AR.1.B	56.140	0.510	0.000	8.040	3.050	938.800	0.570	145.580	46.660	0.050
AR.1.C	38.320	1.140	85.000	8.850	1.370	1747.000	0.190	33.560	10.870	0.110
AR.2.A	16.670	1.750	28.270	8.900	0.550	696.100	0.200	33.790	11.270	0.140
AR.2.B	7.510	2.300	24.540	8.770	0.340	374.100	0.280	30.420	9.120	0.210
AR.2.C	4.480	3.090	9.570	8.870	0.180	194.900	0.260	29.100	8.710	0.170
HA.1.A	138.950	1.120	214.220	8.180	3.410	2799.000	0.260	44.250	6.780	0.230
HA.1.B	132.490	0.690	270.020	8.160	2.460	1884.000	0.180	41.290	6.020	0.230
HA.1.C	4.080	1.800	13.320	8.930	0.100	93.850	0.130	26.550	12.020	0.150
HA.2.A	158.930	1.670	63.650	8.070	1.920	467.900	0.200	33.640	6.890	0.300
HA.2.B	61.270	0.930	50.260	8.430	2.230	2261.000	0.120	37.430	9.420	0.260
HA.2.C	285.500	1.870	1445.920	7.920	3.590	626.200	0.320	46.040	9.170	0.460
LO.1.A	5.730	1.270	35.990	9.050	0.170	273.300	0.260	31.190	6.270	0.070
LO.1.B	12.210	1.430	16.340	8.960	0.620	891.700	0.250	32.520	7.450	0.290
LO.1.C	32.140	1.200	28.160	8.590	1.490	1775.000	0.190	36.570	6.690	0.040
LO.2.A	44.480	1.690	123.320	8.360	1.760	1604.000	0.280	36.970	5.690	0.130
LO.2.B	3.230	1.470	9.590	8.920	0.160	214.900	0.240	29.900	7.730	0.120
LO.2.C	3.370	1.620	14.390	8.960	0.240	306.900	0.330	29.670	8.290	0.080
LO.3.A	62.180	1.660	21.500	8.620	0.490	462.800	0.320	32.320	6.990	0.230
LO.3.B	76.650	2.910	43.290	8.470	1.500	1473.000	0.250	33.450	8.700	0.140
LO.3.C	3.520	1.080	9.560	8.870	0.130	154.800	0.390	30.740	6.310	0.100

Plot	Manganese (mg Kg^{-1})	Iron (mg Kg^{-1})	Copper (mg Kg^{-1})	Boron (mg Kg^{-1})	Silicon (mg Kg^{-1})	Total Nitrogen (%)	Total Carbon (%)	Carbon/ Nitrogen	Total Organic	Soil Water Content (%)

								ratio	Carbon (%)	
SH.1.A	30.440	5.370	1.040	1.090	37.600	0.340	7.160	20.870	3.950	4.000
SH.1.B	24.700	4.830	1.180	2.390	72.600	0.190	4.040	20.820	3.950	26.280
SH.1.C	32.160	7.070	0.560	1.240	58.280	0.650	18.300	28.240	3.950	8.068
SH.2.A	20.240	7.830	1.020	0.000	0.000	0.230	4.870	21.170	2.860	2.591
SH.2.B	23.280	7.220	0.780	0.000	0.000	0.310	5.490	17.480	2.860	4.970
SH.2.C	25.120	5.470	1.230	0.830	25.600	0.230	4.800	20.960	2.860	3.269
SA.1.A	18.240	2.360	0.460	0.610	19.400	0.070	3.310	48.680	0.690	1.714
SA.1.B	7.150	1.970	5.810	0.500	23.200	0.020	2.300	104.550	0.690	0.655
SA.1.C	17.750	2.510	6.090	0.560	18.920	0.080	3.110	37.470	0.690	2.780
SA.2.A	12.510	0.000	0.310	0.580	20.600	0.020	2.480	103.330	0.380	1.379
SA.2.B	12.190	1.410	2.380	0.620	17.640	0.030	2.410	92.690	0.380	2.514
SA.2.C	12.480	1.420	0.340	1.000	17.500	0.010	2.970	270.000	0.380	1.270
AR.1.A	12.010	1.600	0.380	1.000	20.400	0.040	4.600	112.200	0.461	8.408
AR.1.B	4.180	0.920	1.510	13.070	36.200	0.040	0.000	112.200	0.460	3.018
AR.1.C	3.050	0.980	0.390	5.950	32.000	0.010	4.300	358.330	0.460	0.973
AR.2.A	5.620	1.280	1.040	2.470	23.200	0.020	4.970	331.330	0.540	0.705
AR.2.B	7.610	1.670	0.820	1.700	21.400	0.040	5.360	127.620	0.540	1.709
AR.2.C	7.190	1.650	0.990	1.110	21.600	0.020	5.140	214.170	0.540	0.960
HA.1.A	1.170	0.970	0.290	3.060	24.200	0.020	7.170	478.000	0.090	1.324
HA.1.B	0.510	0.860	0.220	2.070	20.800	0.010	7.550	943.750	0.090	1.442
HA.1.C	3.060	1.220	0.240	0.920	29.200	0.020	8.850	465.790	0.090	0.896
HA.2.A	1.230	1.010	0.440	1.370	26.200	0.020	6.880	458.670	0.120	0.753
HA.2.B	1.670	0.990	0.320	1.480	24.000	0.010	6.340	528.330	0.120	1.123
HA.2.C	1.160	0.850	1.090	2.090	20.400	0.020	6.280	299.050	0.120	0.924

LO.1.A	7.000	1.290	0.320	0.990	36.000	0.010	3.190	245.380	0.270	0.855
LO.1.B	7.400	1.220	0.430	1.540	32.400	0.020	3.010	125.420	0.270	0.855
LO.1.C	5.020	0.890	0.780	1.190	27.400	0.020	2.720	181.330	0.270	0.855
LO.2.A	3.530	1.190	0.420	6.140	24.400	0.000	3.490	0.000	0.280	1.005
LO.2.B	6.460	1.340	0.530	1.570	24.800	0.000	3.230	3230.000	0.280	1.005
LO.2.C	8.210	1.160	0.530	1.370	25.000	0.030	3.110	124.400	0.280	1.005
LO.3.A	6.710	1.350	0.480	2.680	28.000	0.030	3.290	96.760	0.280	1.473
LO.3.B	3.540	0.970	0.420	1.270	20.400	0.020	2.370	118.500	0.280	1.473
LO.3.C	13.840	1.170	0.530	1.080	24.200	0.020	4.190	279.330	0.280	1.473

6.2. Appendix B: Chapter 3

Supplementary information for

'Reciprocal activities of chemosynthetic and photosynthetic bacteria across a steep desert aridity gradient'

Sean K. Bay[1,2], David W. Waite[3], Osnat Gillor[4], Philip Hugenholtz[5], Chris Greening*[1,2]

[1] School of Biological Sciences, Monash University, Clayton, VIC 3800, Australia

[2] Department of Microbiology, Biomedicine Discovery Institute, Monash University, Clayton, VIC 3800, Australia

[3] School of Biological Sciences, University of Auckland, Auckland 1010, New Zealand

[4] Department of Environmental Hydrology and Microbiology, Ben Gurion University of the Negev, Sde Boker, Israel

[5] Australian Centre for Ecogenomics, School of Chemistry and Molecular Biosciences, St Lucia, QLD 4072, Australia

6.2.1. Supplementary Figures

Figure S1. Alpha and beta diversity of biocrust and topsoil samples. (a) Boxplot showing estimated richness (Chao1). **(b)** Nonmetric multidimensional scaling plot showing Bray-Curtis community dissimilarity. Community profiling is based on the single copy marker *rplP*.

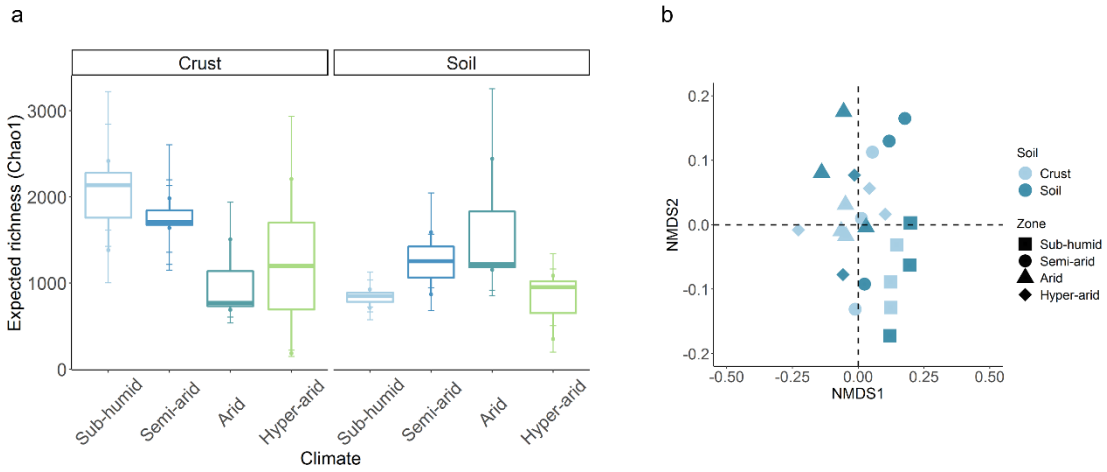


Figure S2. Stacked barchart showing the relative abundance of photosynthetic and non-photosynthetic Cyanobacteriota classes Melainabacteria and Oxyphotobacteria in crusts and topsoils.

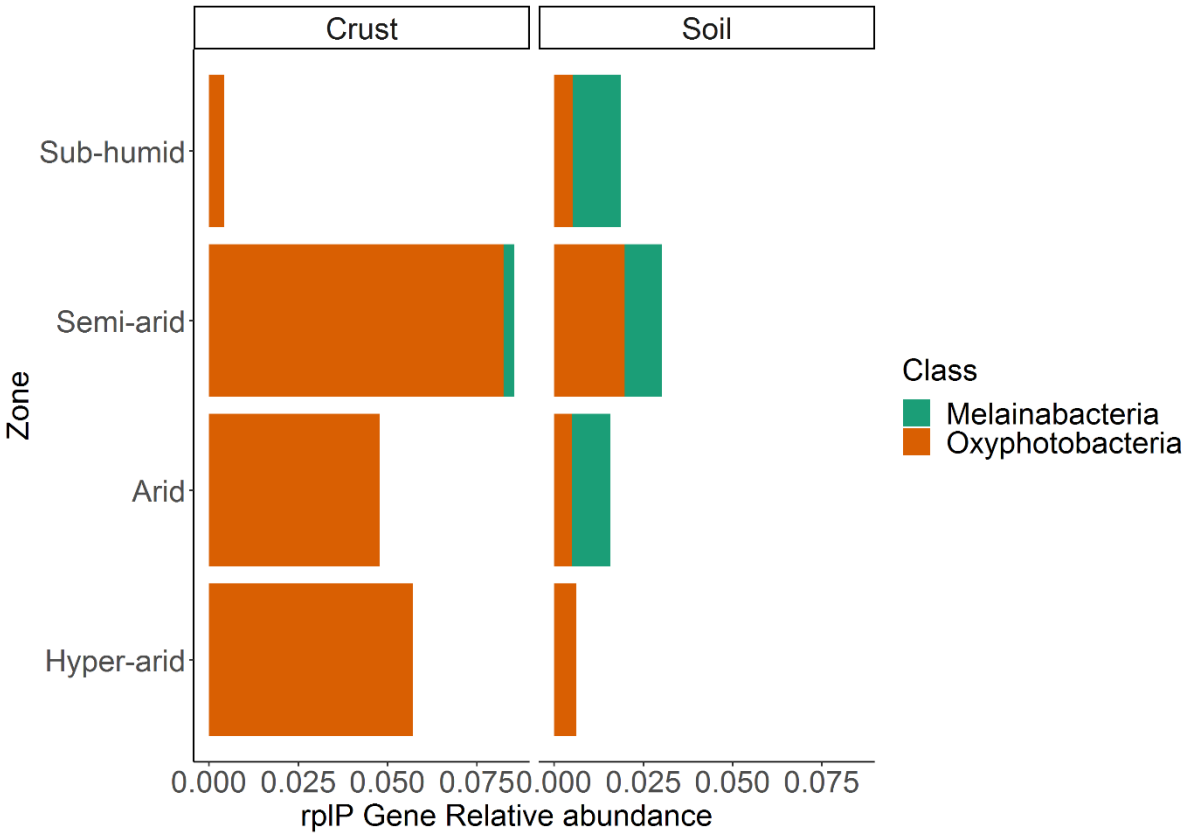


Figure S3. Headspace H_2 mixing ratios during dry and wet microcosm incubations of crust and soils. Error bars represent one standard deviation and the blue line indicates atmospheric H_2 concentrations.

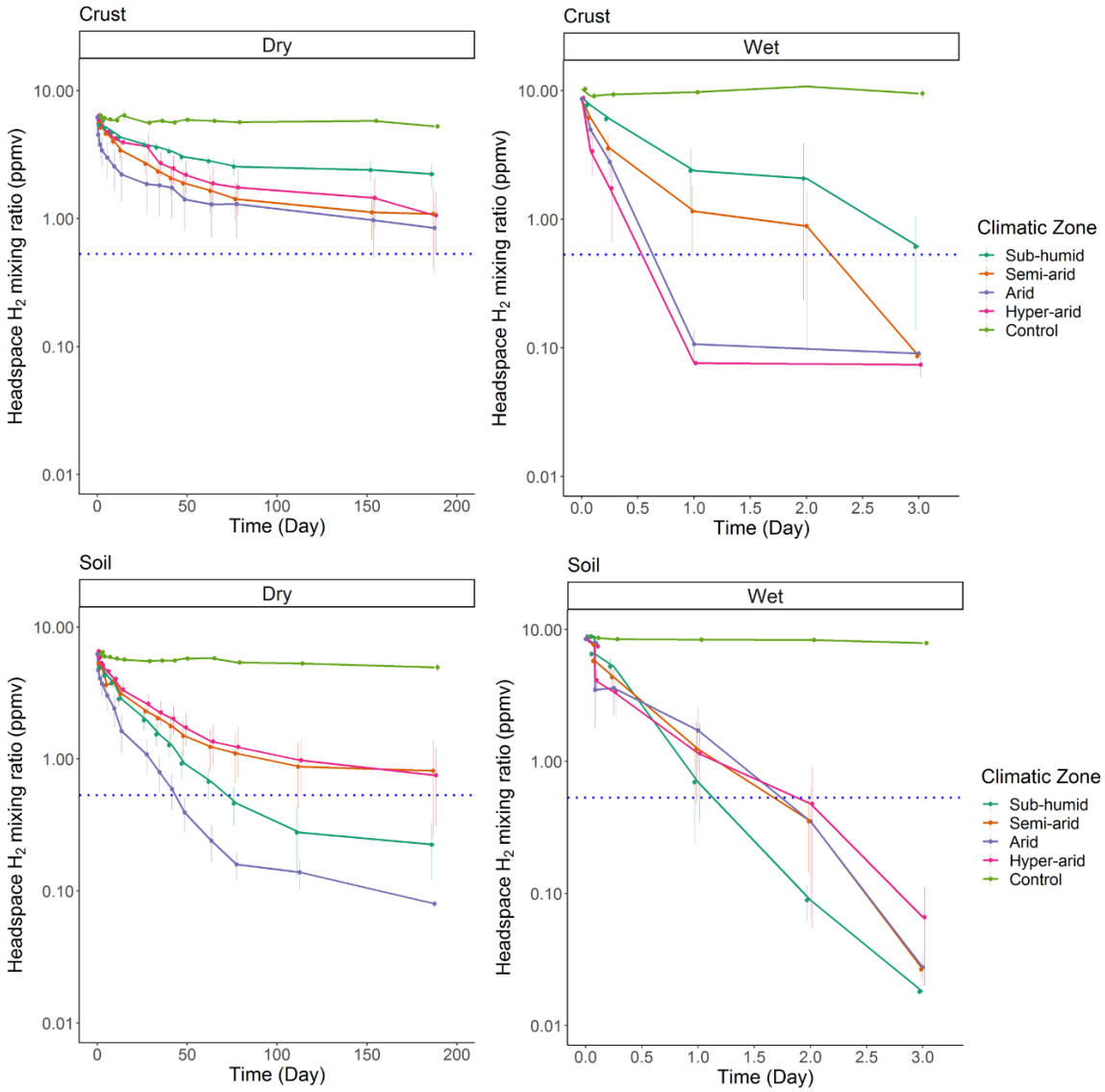


Figure S5. Biomass-normalized rates of chemosynthetic and photosynthetic processes of biocrusts and topsoils collected along the aridity gradient. **(a & b)** Rates of H_2 oxidation measured by gas chromatography. **(c & d)** Rates of carbon fixation measured by incorporation of ^{14}C-labelled CO_2. Activities were measured in *ex situ* microcosms under dry and wet conditions. Centre values show median, boxes upper and lower quartiles and whiskers minimum and maximum values. Panel **a-b** represent biological triplicates from two sites per climatic zone and panel **b-c** are technical triplicates from a pooled sample from each zone.

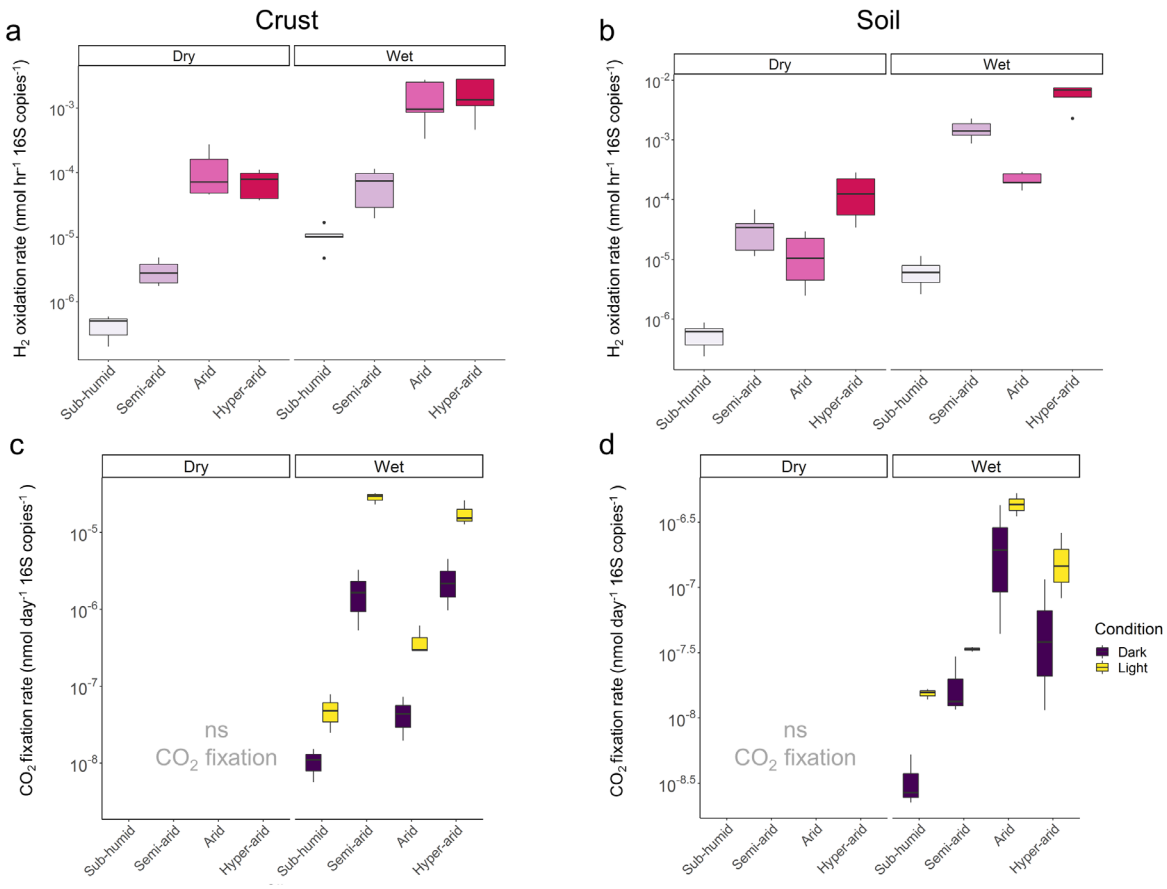
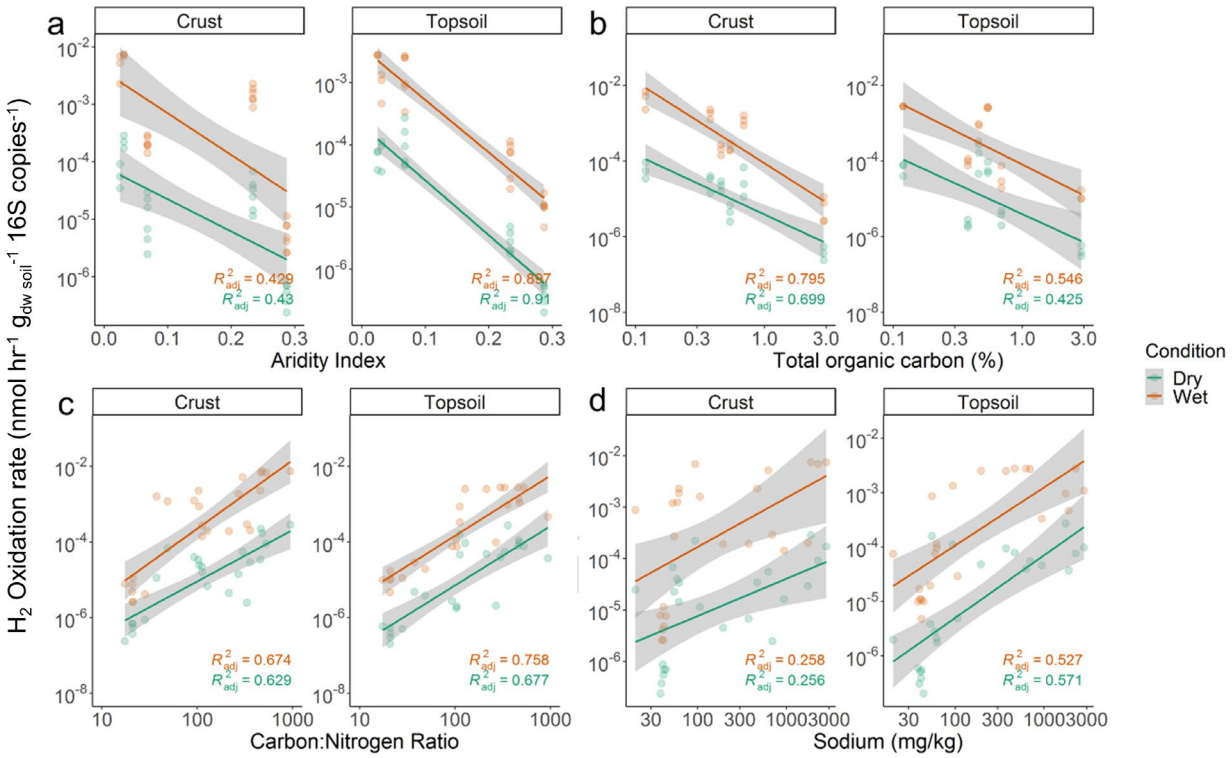


Figure S6. Linear models of predicted subset of edaphic drivers showing significant correlation with H_2 oxidation rates. Shaded ribbons represent standard error (95% confidence).



6.2.2. Supplementary Tables

Tables S1 – S5 are available on the Google Drive folder under **Chapter 3**:
https://drive.google.com/open?id=16UnlZtajRSBZ_vSvJzw3Pl_XraDWUNV2

Table S1 (xlsx). Abundance, diversity, and composition of microbial communities in sampled biocrusts and topsoils.

Table S2 (xlsx). Relative abundance of genes in metagenomic short reads.

Table S3 (xlsx). Assembly statistics, taxonomic classification, and metabolic traits of metagenome-assembled genomes.

Table S4 (xlsx). H_2 oxidation rates for biocrusts and topsoils under both dry and wet conditions, and associated statistical tests.

Table S5 (xlsx). ^{14}C fixation rates for biocrusts and topsoil under both dry and wet conditions, and associated statistical tests.

6.3. Appendix C: Chapter 4

Supplementary information for

Trace gas oxidizers are globally dominant and active members of soil biomes

Sean K. Bay[1,2], Xiyang Dong[3], James A. Bradley[4,5], Pok Man Leung[1,2], Thanavit Jirapanjawat[1,2], Stefan K. Arndt[6], Perran L.M. Cook[7], Douglas LaRowe[8], Philipp A. Nauer[7], Eleonora Chiri[1,2 *], Chris Greening[1,2 *]

[1]School of Biological Sciences, Monash University, Clayton, VIC 3800, Australia
[2]Department of Microbiology, Biomedicine Discovery Institute, Monash University, Clayton, VIC 3800, Australia
[3]School of Marine Sciences, Sun Yat-Sen University, Zhuhai 51900, China
[4]School of Geography, Queen Mary University of London, Bethnal Green, London E1 4NS, United Kingdom
[5]Interface Geochemistry, GFZ German Research Centre for Geosciences, 14473 Potsdam, Germany.
[6]School of Ecosystem and Forest Sciences, University of Melbourne, Richmond, VIC 3121, Australia
[7]School of Chemistry, Monash University, Clayton VIC 3800, Australia
[8]Department of Earth Sciences, University of Southern California, Los Angeles, CA 90089, United States

Correspondence:
* Assoc Prof Chris Greening, Monash University, Department of Microbiology, Innovation Walk, Clayton, VIC 3800, Australia
Email: chris.greening@monash.edu, Ph: +61 451 085 339, ORCID: 0000-0001-7616-0594
* Dr Eleonora Chiri, Monash University, Department of Microbiology, Innovation Walk, Clayton, VIC 3800, Australia
Email: eleonora.chiri@monash.edu, Ph: +61 3 9902 0123, ORCID: 0000-0002-6627-0762

6.3.1. Supplementary Figures

Figure S1. Composition of the bacterial and archaeal communities sequenced in each soil metagenome. Stacked barcharts showing the relative abundance of different phyla in **(a)** Australian soils and **(b)** global soils based on reads for single-copy ribosomal protein gene *rplP*. Alpha and beta diversity of **(c)** Australian soils and **(d)** global soils. For alpha diversity, observed richness and estimated richness (Chao1) are shown. For beta diversity, abundance-based Bray-Curtis diversity is visualized on a a multidimensional scaling plot.

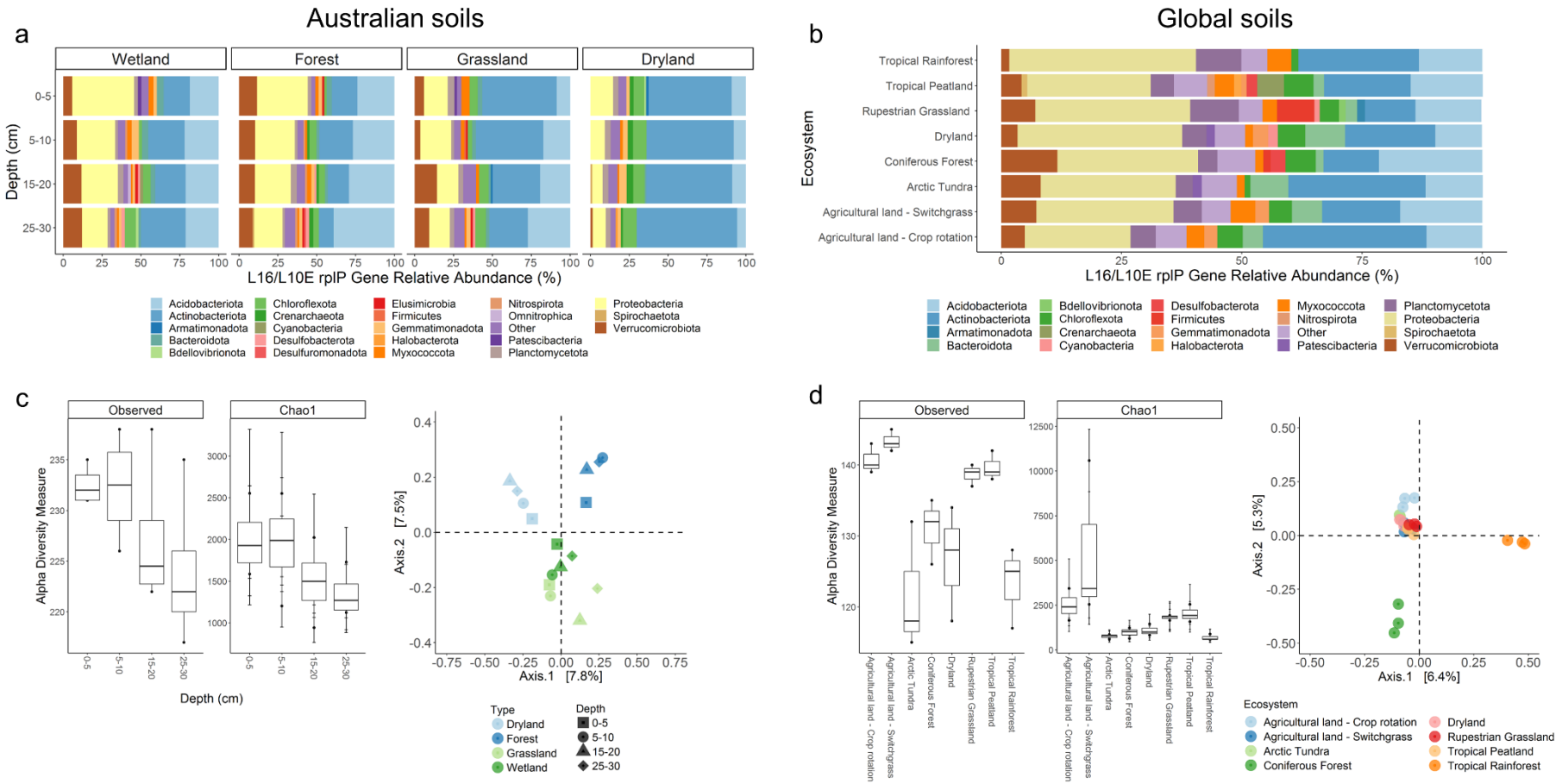

a
Australian soils
Wetland
Forest
Grassland
Dryland
Depth (cm)
0-5
5-10
15-20
25-30
L16/L10E rplP Gene Relative Abundance (%)
Acidobacteriota
Actinobacteriota
Armatimonadota
Bacteroidota
Bdellovibrionota
Chloroflexota
Crenarchaeota
Cyanobacteria
Desulfobacterota
Desulfuromonadota
Elusimicrobia
Firmicutes
Gemmatimonadota
Halobacterota
Myxococcota
Nitrospirota
Omnitrophica
Other
Patescibacteria
Planctomycetota
Proteobacteria
Spirochaetota
Verrucomicrobiota
b
Global soils
Ecosystem
Tropical Rainforest
Tropical Peatland
Rupestrian Grassland
Dryland
Coniferous Forest
Arctic Tundra
Agricultural land - Switchgrass
Agricultural land - Crop rotation
L16/L10E rplP Gene Relative Abundance (%)
Acidobacteriota
Actinobacteriota
Armatimonadota
Bacteroidota
Bdellovibrionota
Chloroflexota
Crenarchaeota
Cyanobacteria
Desulfobacterota
Firmicutes
Gemmatimonadota
Halobacterota
Myxococcota
Nitrospirota
Other
Patescibacteria
Planctomycetota
Proteobacteria
Spirochaetota
Verrucomicrobiota
c
Observed
Chao1
Alpha Diversity Measure
Depth (cm)
Axis.2 [7.5%]
Axis.1 [7.8%]
Type
Dryland
Forest
Grassland
Wetland
Depth
0-5
5-10
15-20
25-30
d
Observed
Chao1
Alpha Diversity Measure
Agricultural land - Crop rotation
Agricultural land - Switchgrass
Arctic Tundra
Coniferous Forest
Dryland
Rupestrian Grassland
Tropical Peatland
Tropical Rainforest
Axis.2 [5.3%]
Axis.1 [6.4%]
Ecosystem
Agricultural land - Crop rotation
Agricultural land - Switchgrass
Arctic Tundra
Coniferous Forest
Dryland
Rupestrian Grassland
Tropical Peatland
Tropical Rainforest

Figure S2. Neighbor-joining tree of amino acid sequences of group 1h [NiFe]-hydrogenase large subunits, a marker for atmospheric H_2 oxidation. The tree shows sequences from soil metagenome-assembled genomes (blue) and unbinned contigs (red) alongside representative reference sequences (black). The tree was constructed using the Poisson model with gaps treated with pairwise deletion, was bootstrapped with 50 replicates, and was rooted with group 1k [NiFe]-hydrogenase sequences (not shown). To enable neighbor-joining, all sequences shorter than 350 amino acids were omitted.

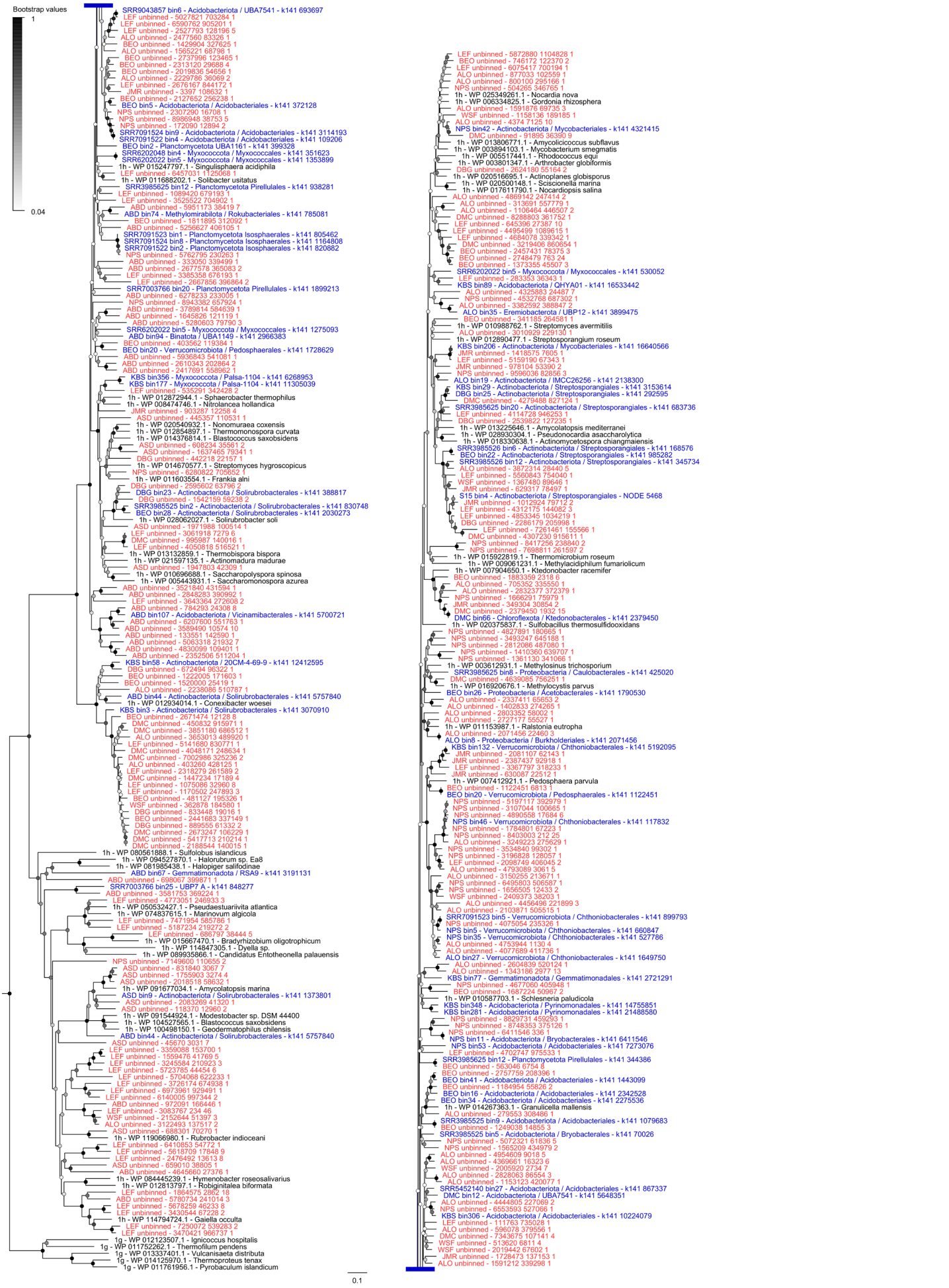

Bootstrap values
1
0.04
SRR9043857 bin6 - Acidobacteriota / UBA7541 - k141 693697
LEF unbinned - 5027821 703284 1
LEF unbinned - 6590762 905201 1
LEF unbinned - 2527793 128196 5
ALO unbinned - 2477560 83326 1
BEO unbinned - 1429904 327625 1
ALO unbinned - 1565221 68798 1
BEO unbinned - 2737996 123465 1
BEO unbinned - 2313120 29688 4
BEO unbinned - 2019836 54656 1
ALO unbinned - 2229786 36069 2
LEF unbinned - 2676167 844172 1
JMR unbinned - 3397 108632 1
BEO unbinned - 2127652 256238 1
BEO bin5 - Acidobacteriota / Acidobacteriales - k141 372128
NPS unbinned - 2307290 16708 1
NPS unbinned - 8986948 38753 5
NPS unbinned - 172090 12894 2
SRR7091524 bin9 - Acidobacteriota / Acidobacteriales - k141 3114193
SRR7091522 bin4 - Acidobacteriota / Acidobacteriales - k141 109206
BEO bin2 - Planctomycetota UBA1161 - k141 399328
SRR6202048 bin4 - Myxococcota / Myxococcales - k141 351623
SRR6202022 bin5 - Myxococcota / Myxococcales - k141 1353899
1h - WP 015247797.1 - Singulisphaera acidiphila
LEF unbinned - 6457031 1125068 1
1h - WP 011688202.1 - Solibacter usitatus
SRR3985625 bin12 - Planctomycetota Pirellulales - k141 938281
LEF unbinned - 1089420 679193 1
LEF unbinned - 3525522 704902 1
ABD unbinned - 5951173 384197
ABD bin74 - Methylomirabilota / Rokubacteriales - k141 785081
BEO unbinned - 1811895 312092 1
ABD unbinned - 5256627 406105 1
SRR7091523 bin1 - Planctomycetota Isosphaerales - k141 805462
SRR7091524 bin8 - Planctomycetota Isosphaerales - k141 1164808
SRR7091522 bin2 - Planctomycetota Isosphaerales - k141 820882
NPS unbinned - 5762795 230283 1
ABD unbinned - 333050 339499 1
ABD unbinned - 2677578 365083 2
LEF unbinned - 3385358 676193 1
LEF unbinned - 2667856 396864 2
SRR7003766 bin20 - Planctomycetota Pirellulales - k141 1899213
ABD unbinned - 6278233 233005 1
NPS unbinned - 8943382 657924 1
ABD unbinned - 3789814 584639 1
ABD unbinned - 1645826 121119 1
ABD unbinned - 5280603 79790 3
SRR6202022 bin5 - Myxococcota / Myxococcales - k141 1275093
ABD bin94 - Binatota / UBA1149 - k141 2966383
BEO unbinned - 403562 119384 1
BEO bin20 - Verrucomicrobiota / Pedosphaerales - k141 1728629
ABD unbinned - 5936843 541081 1
ABD unbinned - 2610343 202864 2
ABD unbinned - 2417691 558962 1
KBS bin356 - Myxococcota / Palsa-1104 - k141 6268953
KBS bin177 - Myxococcota / Palsa-1104 - k141 11305039
LEF unbinned - 535291 342428 2
1h - WP 012872944.1 - Sphaerobacter thermophilus
1h - WP 008474746.1 - Nitrolancea hollandica
JMR unbinned - 903287 12258 4
ASD unbinned - 445357 110531 1
1h - WP 020540932.1 - Nonomuraea coxensis
1h - WP 012854897.1 - Thermomonospora curvata
1h - WP 014376814.1 - Blastococcus saxobsidens
ASD unbinned - 608234 35561 2
ASD unbinned - 1637465 79341 1
DBG unbinned - 442218 22157 1
1h - WP 014670577.1 - Streptomyces hygroscopicus
NPS unbinned - 6280822 705652 1
1h - WP 011603554.1 - Frankia alni
DBG unbinned - 2595602 63796 2
DBG bin23 - Actinobacteriota / Solirubrobacterales - k141 388817
DBG unbinned - 1542159 59238 2
SRR3985525 bin2 - Actinobacteriota / Solirubrobacterales - k141 830748
BEO bin28 - Actinobacteriota / Solirubrobacterales - k141 2030273
1h - WP 028062027.1 - Solirubrobacter soli
ASD unbinned - 1971988 100514 1
LEF unbinned - 3061918 7279 6
DMC unbinned - 995987 140016 1
LEF unbinned - 4050818 516521 1
1h - WP 013132859.1 - Thermobispora bispora
1h - WP 021597135.1 - Actinomadura madurae
ASD unbinned - 1947803 42309 1
1h - WP 010696688.1 - Saccharopolyspora spinosa
1h - WP 005443931.1 - Saccharomonospora azurea
ABD unbinned - 3521840 431594 1
ABD unbinned - 2848283 390992 1
LEF unbinned - 3643364 272608 2
ABD unbinned - 784293 24308 8
ABD bin107 - Acidobacteriota / Vicinamibacterales - k141 5700721
ABD unbinned - 6207600 551763 1
ABD unbinned - 3589490 10574 10
ABD unbinned - 133551 142590 1
ABD unbinned - 5063318 21932 7
ABD unbinned - 4830099 109401 1
ABD unbinned - 2352506 511204 1
KBS bin58 - Actinobacteriota / 20CM-4-69-9 - k141 12412595
DBG unbinned - 672494 98322 1
BEO unbinned - 1222005 171603 1
BEO unbinned - 1520000 25419 1
ALO unbinned - 2238086 510787 1
ABD bin44 - Actinobacteriota / Solirubrobacterales - k141 5757840
1h - WP 012934014.1 - Conexibacter woesei
KBS bin3 - Actinobacteriota / Solirubrobacterales - k141 3070910
BEO unbinned - 2671474 12128 8
DMC unbinned - 450832 915971 1
DMC unbinned - 3851180 686512 1
ALO unbinned - 3653013 489920 1
LEF unbinned - 5141680 830771 1
DMC unbinned - 4048171 248634 1
DMC unbinned - 7002986 325236 2
ALO unbinned - 403260 428125 1
LEF unbinned - 2318279 261589 2
DMC unbinned - 1447234 17189 4
LEF unbinned - 1075086 32960 8
LEF unbinned - 1170502 247893 3
BEO unbinned - 481127 195326 1
WSF unbinned - 362878 184580 1
DBG unbinned - 833448 19016 1
BEO unbinned - 2441683 337149 1
DBG unbinned - 889555 61332 2
DMC unbinned - 2673247 106229 1
DMC unbinned - 5417713 210214 1
DMC unbinned - 2188544 140015 1
1h - WP 080561888.1 - Sulfolobus islandicus
1h - WP 094527870.1 - Halorubrum sp. Ea8
1h - WP 081985438.1 - Halopiger salifodinae
ABD bin67 - Gemmatimonadota / RSA9 - k141 3191131
ABD unbinned - 698067 399671 1
SRR7003766 bin25 - UBP7 A - k141 848277
ABD unbinned - 3581753 369224 1
LEF unbinned - 4773051 248933 3
1h - WP 050532427.1 - Pseudaestuariivita atlantica
1h - WP 074837615.1 - Marinovum algicola
LEF unbinned - 7471954 585786 1
LEF unbinned - 5187234 219272 2
LEF unbinned - 686797 38444 5
1h - WP 015667470.1 - Bradyrhizobium oligotrophicum
1h - WP 114847305.1 - Dyella sp.
1h - WP 089935866.1 - Candidatus Entotheonella palauensis
NPS unbinned - 7149600 110655 2
ASD unbinned - 831840 3067 7
ASD unbinned - 1755903 3274 4
ASD unbinned - 2018518 58632 1
1h - WP 091677034.1 - Amycolatopsis marina
ASD bin9 - Actinobacteriota / Solirubrobacterales - k141 1373801
ASD unbinned - 2083269 41320 1
ASD unbinned - 118370 12960 2
1h - WP 091544924.1 - Modestobacter sp. DSM 44400
1h - WP 104527565.1 - Blastococcus saxobsidens
1h - WP 100498150.1 - Geodermatophilus chilensis
ABD bin44 - Actinobacteriota / Solirubrobacterales - k141 5757840
ASD unbinned - 45670 3031 7
LEF unbinned - 3359088 153700 1
LEF unbinned - 1559476 41769 5
LEF unbinned - 3245584 210923 3
LEF unbinned - 5723785 44454 6
LEF unbinned - 5704068 6222233 1
LEF unbinned - 3726174 674938 1
LEF unbinned - 6973961 929491 1
LEF unbinned - 6140005 997344 2
ABD unbinned - 972091 166446 1
LEF unbinned - 3083767 234 46
WSF unbinned - 2152644 51397 3
ALO unbinned - 3122493 137517 2
ASD unbinned - 688301 70270 1
1h - WP 119066980.1 - Rubrobacter indioceani
LEF unbinned - 6410853 54772 1
LEF unbinned - 5618709 17848 9
LEF unbinned - 2476492 13613 8
ASD unbinned - 659010 38805 1
ABD unbinned - 4645660 27376 1
1h - WP 084445239.1 - Hymenobacter roseosalivarius
1h - WP 012813797.1 - Robiginitalea biformata
LEF unbinned - 1864575 2862 18
ABD unbinned - 5780734 241014 3
LEF unbinned - 5678259 46233 8
LEF unbinned - 3430544 67228 2
1h - WP 114794724.1 - Gaiella occulta
LEF unbinned - 7250072 539283 2
LEF unbinned - 3470421 966737 1
1g - WP 012123507.1 - Ignicoccus hospitalis
1g - WP 011752262.1 - Thermofilum pendens
1g - WP 013337401.1 - Vulcanisaeta distributa
1g - WP 014125970.1 - Thermoproteus tenax
1g - WP 011761956.1 - Pyrobaculum islandicum
0.1
LEF unbinned - 5872880 1104828 1
BEO unbinned - 746172 122370 2
LEF unbinned - 6075417 700194 1
ALO unbinned - 877033 102559 1
ALO unbinned - 800100 295166 1
NPS unbinned - 504265 346765 1
1h - WP 025349261.1 - Nocardia nova
1h - WP 006334825.1 - Gordonia rhizosphera
ALO unbinned - 1591876 69735 3
WSF unbinned - 1158136 189185 1
ALO unbinned - 4374 7125 10
NPS bin42 - Actinobacteriota / Mycobacteriales - k141 4321415
DMC unbinned - 91895 36390 9
1h - WP 013806771.1 - Amycolicicoccus subflavus
1h - WP 003894103.1 - Mycobacterium smegmatis
1h - WP 005517441.1 - Rhodococcus equi
1h - WP 003801347.1 - Arthrobacter globiformis
DBG unbinned - 2624180 55184 2
1h - WP 020516695.1 - Actinoplanes globisporus
1h - WP 020500148.1 - Scisicionella marina
1h - WP 017611790.1 - Nocardiopsis salina
ALO unbinned - 4869142 247414 2
ALO unbinned - 313691 557779 1
ALO unbinned - 1106464 446507 2
DMC unbinned - 8288803 361752 1
LEF unbinned - 645396 27387 10
LEF unbinned - 4495499 1089615 1
LEF unbinned - 4684078 339342 1
DMC unbinned - 3219406 860654 1
BEO unbinned - 2457431 78375 3
BEO unbinned - 2748479 763 24
BEO unbinned - 1373355 45507 3
SRR6202022 bin5 - Myxococcota / Myxococcales - k141 530052
LEF unbinned - 283353 36343 1
KBS bin89 - Acidobacteriota / QHYA01 - k141 16533442
ALO unbinned - 4325883 24487 7
NPS unbinned - 4532766 687302 1
ALO unbinned - 3382592 388847 2
ALO bin35 - Eremiobacterota / UBP12 - k141 3899475
BEO unbinned - 341185 264581 1
1h - WP 010988762.1 - Streptomyces avermitilis
ALO unbinned - 3010929 229130 1
1h - WP 012890477.1 - Streptosporangium roseum
KBS bin206 - Actinobacteriota / Mycobacteriales - k141 16640566
JMR unbinned - 1418575 7605 1
LEF unbinned - 5159190 67343 1
JMR unbinned - 978104 53390 2
NPS unbinned - 9596036 82856 3
ALO bin19 - Actinobacteriota / IMCC26256 - k141 2138300
KBS bin29 - Actinobacteriota / Streptosporangiales - k141 3153614
DBG bin25 - Actinobacteriota / Streptosporangiales - k141 292595
DMC unbinned - 4279488 827124 1
SRR3985625 bin20 - Actinobacteriota / Streptosporangiales - k141 683736
LEF unbinned - 4114728 946253 1
DBG unbinned - 2539822 127235 1
1h - WP 013225646.1 - Amycolatopsis mediterranei
1h - WP 028930304.1 - Pseudonocardia asaccharolytica
1h - WP 018330638.1 - Actinomycetospora chiangmaiensis
SRR3985526 bin6 - Actinobacteriota / Streptosporangiales - k141 168576
BEO bin22 - Actinobacteriota / Streptosporangiales - k141 985282
SRR3985526 bin12 - Actinobacteriota / Streptosporangiales - k141 345734
ALO unbinned - 3872314 28440 5
LEF unbinned - 5560843 754040 1
WSF unbinned - 1367480 89646 1
JMR unbinned - 629317 78497 1
S15 bin4 - Actinobacteriota / Streptosporangiales - NODE 5468
JMR unbinned - 1012924 79712 2
LEF unbinned - 4312175 144082 3
LEF unbinned - 4853345 1034219 1
DBG unbinned - 2286179 205998 1
LEF unbinned - 7261461 155566 1
DMC unbinned - 4307230 915611 1
NPS unbinned - 8417256 238840 2
NPS unbinned - 7698811 261597 2
1h - WP 015922819.1 - Thermomonospora roseum
1h - WP 009061231.1 - Methylacidiphilum fumariolicum
1h - WP 007904650.1 - Ktedonobacter racemifer
BEO unbinned - 1883359 2318 6
ALO unbinned - 705352 335550 1
ALO unbinned - 2832377 372379 1
NPS unbinned - 1666291 75979 1
JMR unbinned - 349304 30854 2
DMC unbinned - 2379450 1932 15
DMC bin66 - Chloroflexota / Ktedonobacterales - k141 2379450
1h - WP 020375837.1 - Sulfobacillus thermosulfidooxidans
NPS unbinned - 4827891 180665 1
NPS unbinned - 3493247 645188 1
NPS unbinned - 2812086 487080 1
NPS unbinned - 1410360 639707 1
NPS unbinned - 1361130 341066 1
1h - WP 003612931.1 - Methylosinus trichosporium
SRR3985625 bin8 - Proteobacteria / Caulobacterales - k141 425020
DMC unbinned - 4639085 756251 1
1h - WP 016920676.1 - Methylocystis parvus
BEO bin26 - Proteobacteria / Acetobacterales - k141 1790530
ALO unbinned - 2337411 65653 2
ALO unbinned - 1402833 274265 1
ALO unbinned - 2803352 58002 1
ALO unbinned - 2727177 55527 1
1h - WP 011153987.1 - Ralstonia eutropha
ALO unbinned - 2071456 22460 3
ALO bin8 - Proteobacteria / Burkholderiales - k141 2071456
KBS bin132 - Verrucomicrobiota / Chthoniobacterales - k141 5192095
JMR unbinned - 2081107 62143 1
JMR unbinned - 2387437 92918 1
LEF unbinned - 3367797 318233 1
JMR unbinned - 630087 22512 1
1h - WP 007412921.1 - Pedosphaera parvula
BEO unbinned - 1122451 6813 1
BEO bin20 - Verrucomicrobiota / Pedosphaerales - k141 1122451
NPS unbinned - 5197117 392979 1
NPS unbinned - 3107044 100665 1
NPS unbinned - 4890558 17684 6
NPS bin46 - Verrucomicrobiota / Chthoniobacterales - k141 117832
NPS unbinned - 1784801 67223 1
NPS unbinned - 8403003 212 25
ALO unbinned - 3249223 275629 1
NPS unbinned - 3534840 99302 1
NPS unbinned - 3196828 128057 1
LEF unbinned - 2098749 406045 2
ALO unbinned - 4793089 3061 5
ALO unbinned - 3150255 213671 1
NPS unbinned - 6495803 506587 1
NPS unbinned - 1656505 12433 2
WSF unbinned - 2409373 38203 1
ALO unbinned - 4456496 221899 3
ALO unbinned - 2103871 505515 1
SRR7091523 bin5 - Verrucomicrobiota / Chthoniobacterales - k141 899793
NPS unbinned - 4075054 235326 1
NPS bin5 - Verrucomicrobiota / Chthoniobacterales - k141 660847
NPS bin35 - Verrucomicrobiota / Chthoniobacterales - k141 527786
ALO unbinned - 4753944 1130 4
ALO unbinned - 4077689 411736 1
ALO bin27 - Verrucomicrobiota / Chthoniobacterales - k141 1649750
ALO unbinned - 2604839 520124 1
ALO unbinned - 1343186 2977 13
KBS bin77 - Gemmatimonadota / Gemmatimonadales - k141 2721291
NPS unbinned - 4677080 405948 1
BEO unbinned - 1687224 50967 2
1h - WP 010587703.1 - Schlesneria paludicola
KBS bin348 - Acidobacteriota / Pyrinomonadales - k141 14755851
KBS bin281 - Acidobacteriota / Pyrinomonadales - k141 21488580
NPS unbinned - 8829731 459293 1
NPS unbinned - 8748353 375126 1
NPS unbinned - 6411546 336 1
NPS bin11 - Acidobacteriota / Bryobacterales - k141 6411546
NPS bin53 - Acidobacteriota / Acidobacterales - k141 7273076
LEF unbinned - 4702747 975533 1
SRR3985625 bin12 - Planctomycetota Pirellulales - k141 344386
BEO unbinned - 563046 6754 8
BEO unbinned - 2757759 208396 1
BEO bin41 - Acidobacteriota / Acidobacteriales - k141 1443099
BEO unbinned - 1184954 55826 2
BEO bin16 - Acidobacteriota / Acidobacteriales - k141 2342528
BEO bin34 - Acidobacteriota / Acidobacteriales - k141 2275536
1h - WP 014267363.1 - Granulicella mallensis
ALO unbinned - 2795553 308486 1
SRR3985525 bin9 - Acidobacteriota / Acidobacteriales - k141 1079683
BEO unbinned - 1249038 14855 3
SRR3985525 bin5 - Acidobacteriota / Bryobacterales - k141 70026
NPS unbinned - 5072321 61836 5
NPS unbinned - 1565209 434979 2
ALO unbinned - 4954609 9018 5
ALO unbinned - 4369661 16323 6
WSF unbinned - 2005920 27347
ALO unbinned - 2828063 86554 3
ALO unbinned - 3153123 420077 1
SRR5452140 bin27 - Acidobacteriota / Acidobacteriales - k141 867337
DMC bin12 - Acidobacteriota / UBA7541 - k141 5648351
ALO unbinned - 4444805 227069 2
NPS unbinned - 6553593 527066 1
KBS bin306 - Acidobacteriota / Acidobacteriales - k141 10224079
LEF unbinned - 111763 735028 1
ALO unbinned - 596078 379556 1
DMC unbinned - 7343675 107141 4
WSF unbinned - 513620 6811 4
WSF unbinned - 2019442 67602 1
JMR unbinned - 1728473 137153 1
ALO unbinned - 1591212 339298 1

Figure S3. Neighbor-joining tree of amino acid sequences of other group 1 [NiFe]-hydrogenase large subunits. The tree shows sequences from soil metagenome-assembled genomes (blue) and unbinned contigs (red) alongside representative reference sequences (black). The subgroup of each reference sequence is denoted according to the HydDB classification scheme. Subgroups 1c, 1d, and 1f are predicted to support aerobic respiration, whereas the other subgroups primarily support anaerobic respiration. The tree was constructed using the Poisson model with gaps treated with pairwise deletion, was bootstrapped with 50 replicates, and was rooted with group 1k [NiFe]-hydrogenase sequences. To enable neighbour-joining, all sequences shorter than 350 amino acids were omitted.


Bootstrap values
1
0.04
0.1

Figure S4. Neighbor-joining tree of amino acid sequences of group 2 [NiFe]-hydrogenase large subunits. The tree shows sequences from soil metagenome-assembled genomes (blue) and unbinned contigs (red) alongside representative reference sequences (black). The subgroup of each reference sequence is denoted according to the HydDB classification scheme, with subgroup 2a predicted to support aerobic respiration and subgroups 2b and 2c predicted to support hydrogen sensing. The tree was constructed using the Poisson model with gaps treated with pairwise deletion, was bootstrapped with 50 replicates, and was was rooted with group 1k [NiFe]-hydrogenase sequences. To enable neighbour-joining, all sequences shorter than 350 amino acids were omitted.

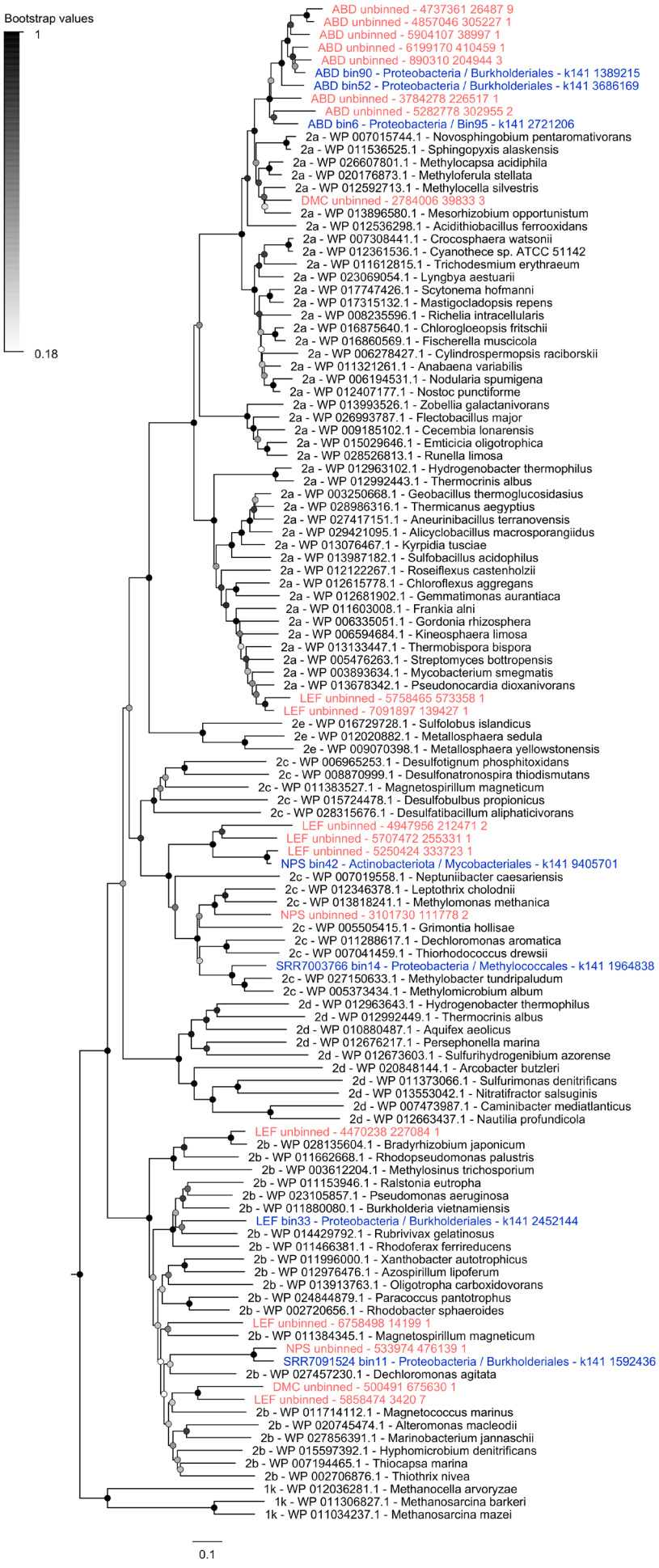

Bootstrap values
1
0.18
ABD unbinned - 4737361 26487 9
ABD unbinned - 4857046 305227 1
ABD unbinned - 5904107 38997 1
ABD unbinned - 6199170 410459 1
ABD unbinned - 890310 204944 3
ABD bin90 - Proteobacteria / Burkholderiales - k141 1389215
ABD bin52 - Proteobacteria / Burkholderiales - k141 3686169
ABD unbinned - 3784278 226517 1
ABD unbinned - 5282778 302955 2
ABD bin6 - Proteobacteria / Bin95 - k141 2721206
2a - WP 007015744.1 - Novosphingobium pentaromativorans
2a - WP 011536525.1 - Sphingopyxis alaskensis
2a - WP 026607801.1 - Methylocapsa acidiphila
2a - WP 020176873.1 - Methyloferula stellata
2a - WP 012592713.1 - Methylocella silvestris
DMC unbinned - 2784006 39833 3
2a - WP 013896580.1 - Mesorhizobium opportunistum
2a - WP 012536298.1 - Acidithiobacillus ferrooxidans
2a - WP 007308441.1 - Crocosphaera watsonii
2a - WP 012361536.1 - Cyanothece sp. ATCC 51142
2a - WP 011612815.1 - Trichodesmium erythraeum
2a - WP 023069054.1 - Lyngbya aestuarii
2a - WP 017747426.1 - Scytonema hofmanni
2a - WP 017315132.1 - Mastigocladopsis repens
2a - WP 008235596.1 - Richelia intracellularis
2a - WP 016875640.1 - Chlorogloeopsis fritschii
2a - WP 016860569.1 - Fischerella muscicola
2a - WP 006278427.1 - Cylindrospermopsis raciborskii
2a - WP 011321261.1 - Anabaena variabilis
2a - WP 006194531.1 - Nodularia spumigena
2a - WP 012407177.1 - Nostoc punctiforme
2a - WP 013993526.1 - Zobellia galactanivorans
2a - WP 026993787.1 - Flectobacillus major
2a - WP 009185102.1 - Cecembia lonarensis
2a - WP 015029646.1 - Emticicia oligotrophica
2a - WP 028526813.1 - Runella limosa
2a - WP 012963102.1 - Hydrogenobacter thermophilus
2a - WP 012992443.1 - Thermocrinis albus
2a - WP 003250668.1 - Geobacillus thermoglucosidasius
2a - WP 028986316.1 - Thermicanus aegyptius
2a - WP 027417151.1 - Aneurinibacillus terranovensis
2a - WP 029421095.1 - Alicyclobacillus macrosporangiidus
2a - WP 013076467.1 - Kyrpidia tusciae
2a - WP 013987182.1 - Sulfobacillus acidophilus
2a - WP 012122267.1 - Roseiflexus castenholzii
2a - WP 012615778.1 - Chloroflexus aggregans
2a - WP 012681902.1 - Gemmatimonas aurantiaca
2a - WP 011603008.1 - Frankia alni
2a - WP 006335051.1 - Gordonia rhizosphera
2a - WP 006594684.1 - Kineosphaera limosa
2a - WP 013133447.1 - Thermobispora bispora
2a - WP 005476263.1 - Streptomyces bottropensis
2a - WP 003893634.1 - Mycobacterium smegmatis
2a - WP 013678342.1 - Pseudonocardia dioxanivorans
LEF unbinned - 5758465 573358 1
LEF unbinned - 7091897 139427 1
2e - WP 016729728.1 - Sulfolobus islandicus
2e - WP 012020882.1 - Metallosphaera sedula
2e - WP 009070398.1 - Metallosphaera yellowstonensis
2c - WP 006965253.1 - Desulfotignum phosphitoxidans
2c - WP 008870999.1 - Desulfonatronospira thiodismutans
2c - WP 011383527.1 - Magnetospirillum magneticum
2c - WP 015724478.1 - Desulfobulbus propionicus
2c - WP 028315676.1 - Desulfatibacillum aliphaticivorans
LEF unbinned - 4947956 212471 2
LEF unbinned - 5707472 255331 1
LEF unbinned - 5250424 333723 1
NPS bin42 - Actinobacteriota / Mycobacteriales - k141 9405701
2c - WP 007019558.1 - Neptuniibacter caesariensis
2c - WP 012346378.1 - Leptothrix cholodnii
2c - WP 013818241.1 - Methylomonas methanica
NPS unbinned - 3101730 111778 2
2c - WP 005505415.1 - Grimontia hollisae
2c - WP 011288617.1 - Dechloromonas aromatica
2c - WP 007041459.1 - Thiorhodococcus drewsii
SRR7003766 bin14 - Proteobacteria / Methylococcales - k141 1964838
2c - WP 027150633.1 - Methylobacter tundripaludum
2c - WP 005373434.1 - Methylomicrobium album
2d - WP 012963643.1 - Hydrogenobacter thermophilus
2d - WP 012992449.1 - Thermocrinis albus
2d - WP 010880487.1 - Aquifex aeolicus
2d - WP 012676217.1 - Persephonella marina
2d - WP 012673603.1 - Sulfurihydrogenibium azorense
2d - WP 020848144.1 - Arcobacter butzleri
2d - WP 011373066.1 - Sulfurimonas denitrificans
2d - WP 013553042.1 - Nitratifractor salsuginis
2d - WP 007473987.1 - Caminibacter mediatlanticus
2d - WP 012663437.1 - Nautilia profundicola
LEF unbinned - 4470238 227084 1
2b - WP 028135604.1 - Bradyrhizobium japonicum
2b - WP 011662668.1 - Rhodopseudomonas palustris
2b - WP 003612204.1 - Methylosinus trichosporium
2b - WP 011153946.1 - Ralstonia eutropha
2b - WP 023105857.1 - Pseudomonas aeruginosa
2b - WP 011880080.1 - Burkholderia vietnamiensis
LEF bin33 - Proteobacteria / Burkholderiales - k141 2452144
2b - WP 014429792.1 - Rubrivivax gelatinosus
2b - WP 011466381.1 - Rhodoferax ferrireducens
2b - WP 011996000.1 - Xanthobacter autotrophicus
2b - WP 012976476.1 - Azospirillum lipoferum
2b - WP 013913763.1 - Oligotropha carboxidovorans
2b - WP 024844879.1 - Paracoccus pantotrophus
2b - WP 002720656.1 - Rhodobacter sphaeroides
LEF unbinned - 6758498 14199 1
2b - WP 011384345.1 - Magnetospirillum magneticum
NPS unbinned - 533974 476139 1
SRR7091524 bin11 - Proteobacteria / Burkholderiales - k141 1592436
2b - WP 027457230.1 - Dechloromonas agitata
DMC unbinned - 500491 675630 1
LEF unbinned - 5858474 3420 7
2b - WP 011714112.1 - Magnetococcus marinus
2b - WP 020745474.1 - Alteromonas macleodii
2b - WP 027856391.1 - Marinobacterium jannaschii
2b - WP 015597392.1 - Hyphomicrobium denitrificans
2b - WP 007194465.1 - Thiocapsa marina
2b - WP 002706876.1 - Thiothrix nivea
1k - WP 012036281.1 - Methanocella arvoryzae
1k - WP 011306827.1 - Methanosarcina barkeri
1k - WP 011034237.1 - Methanosarcina mazei
0.1

Figure S5. Neighbor-joining tree of amino acid sequences of carbon monoxide dehydrogenase large subunit (CoxL), a marker for aerobic carbon monoxide oxidation. The tree shows sequences from soil metagenome-assembled genomes (blue) and unbinned contigs (red) alongside representative reference sequences (black). The tree was constructed using the Poisson model with gaps treated with pairwise deletion, was bootstrapped with 50 replicates, and was midpoint-rooted. To enable neighbor-joining, all hits shorter than 500 amino acids were omitted.

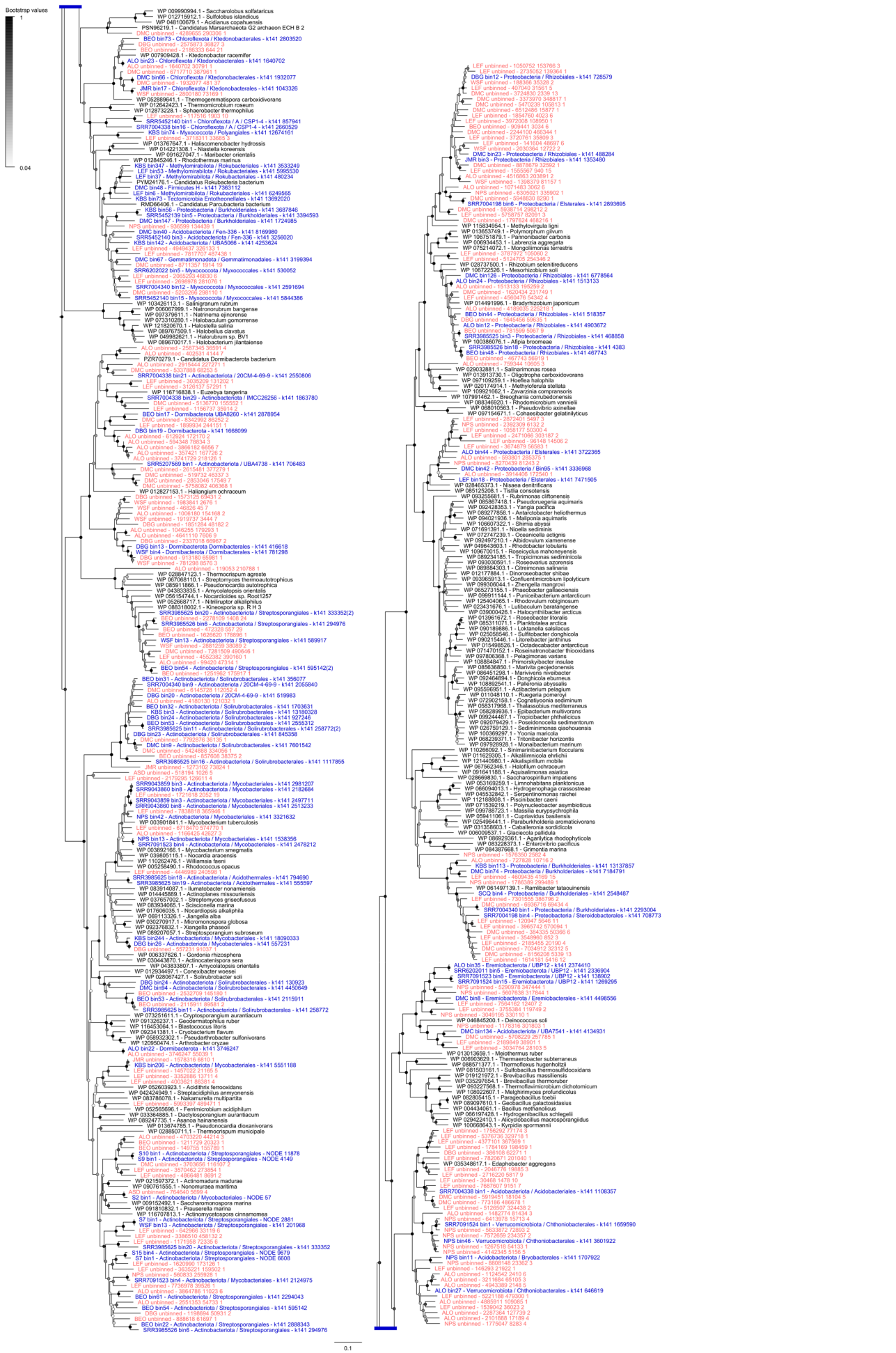

Bootstrap values
1
0.04
0.1

Figure S6. Maximum-likelihood tree of amino acid sequences of sulfide-quinone oxidoreductase (Sqr), a marker for aerobic sulfide oxidation. The tree shows sequences from soil metagenome-assembled genomes (blue) and unbinned contigs (red) alongside representative reference sequences (black). The tree was constructed using the JTT matrix-based model, used all sites, and was bootstrapped with 50 replicates and midpoint-rooted.

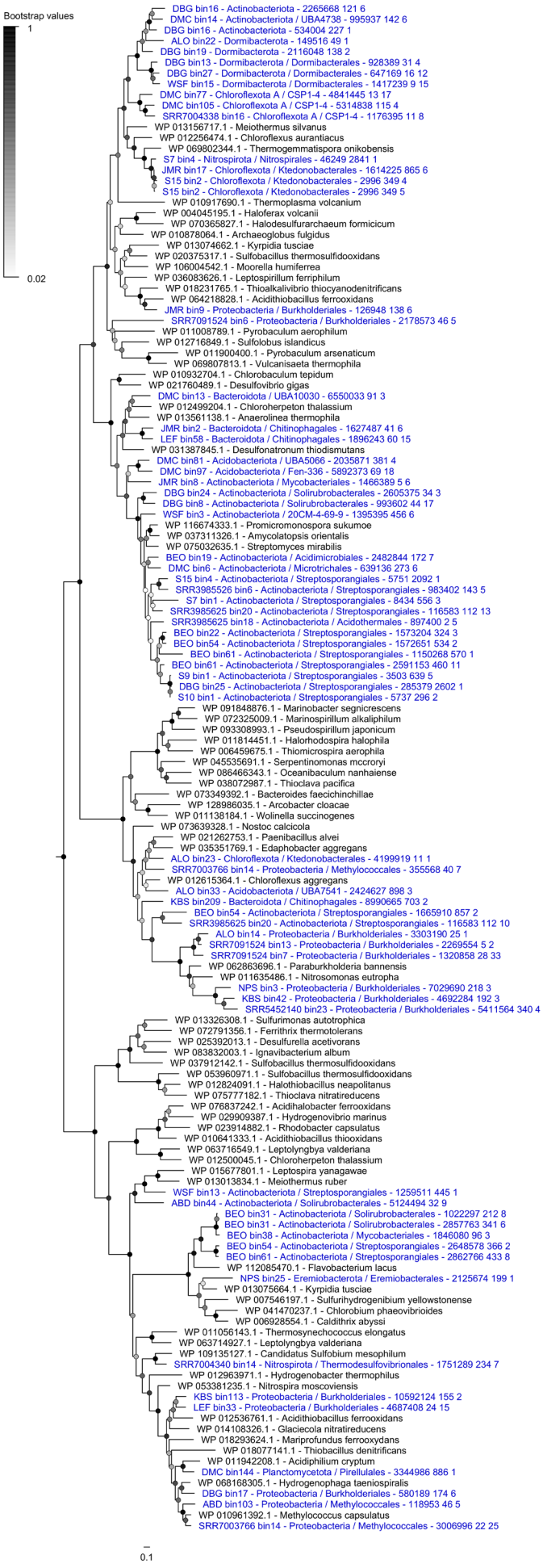

Bootstrap values
1
0.02
DBG bin16 - Actinobacteriota - 2265668 121 6
DMC bin14 - Actinobacteriota / UBA4738 - 995937 142 6
DBG bin16 - Actinobacteriota - 534004 227 1
ALO bin22 - Dormibacterota - 149516 49 1
DBG bin19 - Dormibacterota - 2116048 138 2
DBG bin13 - Dormibacterota / Dormibacterales - 928389 31 4
DBG bin27 - Dormibacterota / Dormibacterales - 647169 16 12
WSF bin15 - Dormibacterota / Dormibacterales - 1417239 9 15
DMC bin77 - Chloroflexota A / CSP1-4 - 4841445 13 17
DMC bin105 - Chloroflexota A / CSP1-4 - 5314838 115 4
SRR7004338 bin16 - Chloroflexota A / CSP1-4 - 1176395 11 8
WP 013156717.1 - Meiothermus silvanus
WP 012256474.1 - Chloroflexus aurantiacus
WP 069802344.1 - Thermogemmatispora onikobensis
S7 bin4 - Nitrospirota / Nitrospirales - 46249 2841 1
JMR bin17 - Chloroflexota / Ktedonobacterales - 1614225 865 6
S15 bin2 - Chloroflexota / Ktedonobacterales - 2996 349 4
S15 bin2 - Chloroflexota / Ktedonobacterales - 2996 349 5
WP 010917690.1 - Thermoplasma volcanium
WP 004045195.1 - Haloferax volcanii
WP 070365827.1 - Halodesulfurarchaeum formicicum
WP 010878064.1 - Archaeoglobus fulgidus
WP 013074662.1 - Kyrpidia tusciae
WP 020375317.1 - Sulfobacillus thermosulfidooxidans
WP 106004542.1 - Moorella humiferrea
WP 036083626.1 - Leptospirillum ferriphilum
WP 018231765.1 - Thioalkalivibrio thiocyanodenitrificans
WP 064218828.1 - Acidithiobacillus ferrooxidans
JMR bin9 - Proteobacteria / Burkholderiales - 126948 138 6
SRR7091524 bin6 - Proteobacteria / Burkholderiales - 2178573 46 5
WP 011008789.1 - Pyrobaculum aerophilum
WP 012716849.1 - Sulfolobus islandicus
WP 011900400.1 - Pyrobaculum arsenaticum
WP 069807813.1 - Vulcanisaeta thermophila
WP 010932704.1 - Chlorobaculum tepidum
WP 021760489.1 - Desulfovibrio gigas
DMC bin13 - Bacteroidota / UBA10030 - 6550033 91 3
WP 012499204.1 - Chloroherpeton thalassium
WP 013561138.1 - Anaerolinea thermophila
JMR bin2 - Bacteroidota / Chitinophagales - 1627487 41 6
LEF bin58 - Bacteroidota / Chitinophagales - 1896243 60 15
WP 031387845.1 - Desulfonatronum thiodismutans
DMC bin81 - Acidobacteriota / UBA5066 - 2035871 381 4
DMC bin97 - Acidobacteriota / Fen-336 - 5892373 69 18
JMR bin8 - Actinobacteriota / Mycobacteriales - 1466389 5 6
DBG bin24 - Actinobacteriota / Solirubrobacterales - 2605375 34 3
DBG bin8 - Actinobacteriota / Solirubrobacterales - 993602 44 17
WSF bin3 - Actinobacteriota / 20CM-4-69-9 - 1395395 456 6
WP 116674333.1 - Promicromonospora sukumoe
WP 037311326.1 - Amycolatopsis orientalis
WP 075032635.1 - Streptomyces mirabilis
BEO bin19 - Actinobacteriota / Acidimicrobiales - 2482844 172 7
DMC bin6 - Actinobacteriota / Microtrichales - 639136 273 6
S15 bin4 - Actinobacteriota / Streptosporangiales - 5751 2092 1
SRR3985526 bin6 - Actinobacteriota / Streptosporangiales - 983402 143 5
S7 bin1 - Actinobacteriota / Streptosporangiales - 8434 556 3
SRR3985625 bin20 - Actinobacteriota / Streptosporangiales - 116583 112 13
SRR3985625 bin18 - Actinobacteriota / Acidothermales - 897400 2 5
BEO bin22 - Actinobacteriota / Streptosporangiales - 1573204 324 3
BEO bin54 - Actinobacteriota / Streptosporangiales - 1572651 534 2
BEO bin61 - Actinobacteriota / Streptosporangiales - 1150268 570 1
BEO bin61 - Actinobacteriota / Streptosporangiales - 2591153 460 11
S9 bin1 - Actinobacteriota / Streptosporangiales - 3503 639 5
DBG bin25 - Actinobacteriota / Streptosporangiales - 285379 2602 1
S10 bin1 - Actinobacteriota / Streptosporangiales - 5737 296 2
WP 091848876.1 - Marinobacter segnicrescens
WP 072325009.1 - Marinospirillum alkaliphilum
WP 093308993.1 - Pseudospirillum japonicum
WP 011814451.1 - Halorhodospira halophila
WP 006459675.1 - Thiomicrospira aerophila
WP 045535691.1 - Serpentinomonas mccroryi
WP 086466343.1 - Oceanibaculum nanhaiense
WP 038072987.1 - Thioclava pacifica
WP 073349392.1 - Bacteroides faecichinchillae
WP 128986035.1 - Arcobacter cloacae
WP 011138184.1 - Wolinella succinogenes
WP 073639328.1 - Nostoc calcicola
WP 021262753.1 - Paenibacillus alvei
WP 035351769.1 - Edaphobacter aggregans
ALO bin23 - Chloroflexota / Ktedonobacterales - 4199919 11 1
SRR7003766 bin14 - Proteobacteria / Methylococcales - 355568 40 7
WP 012615364.1 - Chloroflexus aggregans
ALO bin33 - Acidobacteriota / UBA7541 - 2424627 898 3
KBS bin209 - Bacteroidota / Chitinophagales - 8990665 703 2
BEO bin54 - Actinobacteriota / Streptosporangiales - 1665910 857 2
SRR3985625 bin20 - Actinobacteriota / Streptosporangiales - 116583 112 10
ALO bin14 - Proteobacteria / Burkholderiales - 3303190 25 1
SRR7091524 bin13 - Proteobacteria / Burkholderiales - 2269554 5 2
SRR7091524 bin7 - Proteobacteria / Burkholderiales - 1320858 28 33
WP 062863696.1 - Paraburkholderia bannensis
WP 011635486.1 - Nitrosomonas eutropha
NPS bin3 - Proteobacteria / Burkholderiales - 7029690 218 3
KBS bin42 - Proteobacteria / Burkholderiales - 4692284 192 3
SRR5452140 bin23 - Proteobacteria / Burkholderiales - 5411564 340 4
WP 013326308.1 - Sulfurimonas autotrophica
WP 072791356.1 - Ferrithrix thermotolerans
WP 025392013.1 - Desulfurella acetivorans
WP 083832003.1 - Ignavibacterium album
WP 037912142.1 - Sulfobacillus thermosulfidooxidans
WP 053960971.1 - Sulfobacillus thermosulfidooxidans
WP 012824091.1 - Halothiobacillus neapolitanus
WP 075777182.1 - Thioclava nitratireducens
WP 076837242.1 - Acidihalobacter ferrooxidans
WP 029909387.1 - Hydrogenovibrio marinus
WP 023914882.1 - Rhodobacter capsulatus
WP 010641333.1 - Acidithiobacillus thiooxidans
WP 063716549.1 - Leptolyngbya valderiana
WP 012500045.1 - Chloroherpeton thalassium
WP 015677801.1 - Leptospira yanagawae
WP 013013834.1 - Meiothermus ruber
WSF bin13 - Actinobacteriota / Streptosporangiales - 1259511 445 1
ABD bin44 - Actinobacteriota / Solirubrobacterales - 5124494 32 9
BEO bin31 - Actinobacteriota / Solirubrobacterales - 1022297 212 8
BEO bin31 - Actinobacteriota / Solirubrobacterales - 2857763 341 6
BEO bin38 - Actinobacteriota / Mycobacteriales - 1846080 96 3
BEO bin54 - Actinobacteriota / Streptosporangiales - 2648578 366 2
BEO bin61 - Actinobacteriota / Streptosporangiales - 2862766 433 8
WP 112085470.1 - Flavobacterium lacus
NPS bin25 - Eremiobacterota / Eremiobacterales - 2125674 199 1
WP 013075664.1 - Kyrpidia tusciae
WP 007546197.1 - Sulfurihydrogenibium yellowstonense
WP 041470237.1 - Chlorobium phaeovibrioides
WP 006928554.1 - Caldithrix abyssi
WP 011056143.1 - Thermosynechococcus elongatus
WP 063714927.1 - Leptolyngbya valderiana
WP 109135127.1 - Candidatus Sulfobium mesophilum
SRR7004340 bin14 - Nitrospirota / Thermodesulfovibrionales - 1751289 234 7
WP 012963971.1 - Hydrogenobacter thermophilus
WP 053381235.1 - Nitrospira moscoviensis
KBS bin113 - Proteobacteria / Burkholderiales - 10592124 155 2
LEF bin33 - Proteobacteria / Burkholderiales - 4687408 24 15
WP 012536761.1 - Acidithiobacillus ferrooxidans
WP 014108326.1 - Glaciecola nitratireducens
WP 018293624.1 - Mariprofundus ferrooxydans
WP 018077141.1 - Thiobacillus denitrificans
WP 011942208.1 - Acidiphilium cryptum
DMC bin144 - Planctomycetota / Pirellulales - 3344986 886 1
WP 068168305.1 - Hydrogenophaga taeniospiralis
DBG bin17 - Proteobacteria / Burkholderiales - 580189 174 6
ABD bin103 - Proteobacteria / Methylococcales - 118953 46 5
WP 010961392.1 - Methylococcus capsulatus
SRR7003766 bin14 - Proteobacteria / Methylococcales - 3006996 22 25
0.1

Figure S7. Maximum-likelihood tree of amino acid sequences of flavocytochrome *c* sulfide dehydrogenase (FCC), a marker for aerobic sulfide oxidation. The tree shows sequences from soil metagenome-assembled genomes (blue) and unbinned contigs (red) alongside representative reference sequences (black). The tree was constructed using the JTT matrix-based model, used all sites, and was bootstrapped with 50 replicates and midpoint-rooted.

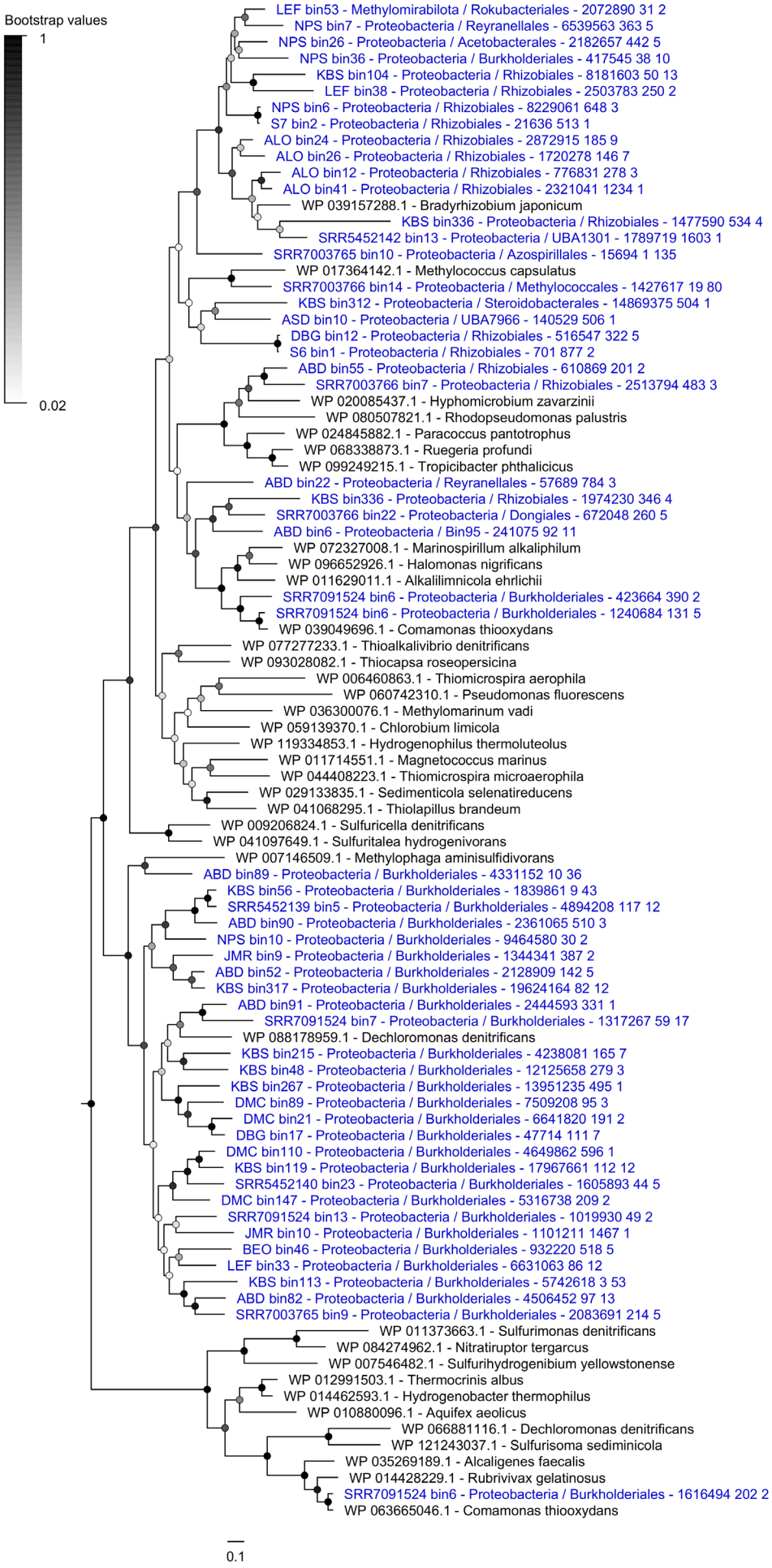

Bootstrap values
1
0.02
LEF bin53 - Methylomirabilota / Rokubacteriales - 2072890 31 2
NPS bin7 - Proteobacteria / Reyranellales - 6539563 363 5
NPS bin26 - Proteobacteria / Acetobacterales - 2182657 442 5
NPS bin36 - Proteobacteria / Burkholderiales - 417545 38 10
KBS bin104 - Proteobacteria / Rhizobiales - 8181603 50 13
LEF bin38 - Proteobacteria / Rhizobiales - 2503783 250 2
NPS bin6 - Proteobacteria / Rhizobiales - 8229061 648 3
S7 bin2 - Proteobacteria / Rhizobiales - 21636 513 1
ALO bin24 - Proteobacteria / Rhizobiales - 2872915 185 9
ALO bin26 - Proteobacteria / Rhizobiales - 1720278 146 7
ALO bin12 - Proteobacteria / Rhizobiales - 776831 278 3
ALO bin41 - Proteobacteria / Rhizobiales - 2321041 1234 1
WP 039157288.1 - Bradyrhizobium japonicum
KBS bin336 - Proteobacteria / Rhizobiales - 1477590 534 4
SRR5452142 bin13 - Proteobacteria / UBA1301 - 1789719 1603 1
SRR7003765 bin10 - Proteobacteria / Azospirillales - 15694 1 135
WP 017364142.1 - Methylococcus capsulatus
SRR7003766 bin14 - Proteobacteria / Methylococcales - 1427617 19 80
KBS bin312 - Proteobacteria / Steroidobacterales - 14869375 504 1
ASD bin10 - Proteobacteria / UBA7966 - 140529 506 1
DBG bin12 - Proteobacteria / Rhizobiales - 516547 322 5
S6 bin1 - Proteobacteria / Rhizobiales - 701 877 2
ABD bin55 - Proteobacteria / Rhizobiales - 610869 201 2
SRR7003766 bin7 - Proteobacteria / Rhizobiales - 2513794 483 3
WP 020085437.1 - Hyphomicrobium zavarzinii
WP 080507821.1 - Rhodopseudomonas palustris
WP 024845882.1 - Paracoccus pantotrophus
WP 068338873.1 - Ruegeria profundi
WP 099249215.1 - Tropicibacter phthalicicus
ABD bin22 - Proteobacteria / Reyranellales - 57689 784 3
KBS bin336 - Proteobacteria / Rhizobiales - 1974230 346 4
SRR7003766 bin22 - Proteobacteria / Dongiales - 672048 260 5
ABD bin6 - Proteobacteria / Bin95 - 241075 92 11
WP 072327008.1 - Marinospirillum alkaliphilum
WP 096652926.1 - Halomonas nigrificans
WP 011629011.1 - Alkalilimnicola ehrlichii
SRR7091524 bin6 - Proteobacteria / Burkholderiales - 423664 390 2
SRR7091524 bin6 - Proteobacteria / Burkholderiales - 1240684 131 5
WP 039049696.1 - Comamonas thiooxydans
WP 077277233.1 - Thioalkalivibrio denitrificans
WP 093028082.1 - Thiocapsa roseopersicina
WP 006460863.1 - Thiomicrospira aerophila
WP 060742310.1 - Pseudomonas fluorescens
WP 036300076.1 - Methylomarinum vadi
WP 059139370.1 - Chlorobium limicola
WP 119334853.1 - Hydrogenophilus thermoluteolus
WP 011714551.1 - Magnetococcus marinus
WP 044408223.1 - Thiomicrospira microaerophila
WP 029133835.1 - Sedimenticola selenatireducens
WP 041068295.1 - Thiolapillus brandeum
WP 009206824.1 - Sulfuricella denitrificans
WP 041097649.1 - Sulfuritalea hydrogenivorans
WP 007146509.1 - Methylophaga aminisulfidivorans
ABD bin89 - Proteobacteria / Burkholderiales - 4331152 10 36
KBS bin56 - Proteobacteria / Burkholderiales - 1839861 9 43
SRR5452139 bin5 - Proteobacteria / Burkholderiales - 4894208 117 12
ABD bin90 - Proteobacteria / Burkholderiales - 2361065 510 3
NPS bin10 - Proteobacteria / Burkholderiales - 9464580 30 2
JMR bin9 - Proteobacteria / Burkholderiales - 1344341 387 2
ABD bin52 - Proteobacteria / Burkholderiales - 2128909 142 5
KBS bin317 - Proteobacteria / Burkholderiales - 19624164 82 12
ABD bin91 - Proteobacteria / Burkholderiales - 2444593 331 1
SRR7091524 bin7 - Proteobacteria / Burkholderiales - 1317267 59 17
WP 088178959.1 - Dechloromonas denitrificans
KBS bin215 - Proteobacteria / Burkholderiales - 4238081 165 7
KBS bin48 - Proteobacteria / Burkholderiales - 12125658 279 3
KBS bin267 - Proteobacteria / Burkholderiales - 13951235 495 1
DMC bin89 - Proteobacteria / Burkholderiales - 7509208 95 3
DMC bin21 - Proteobacteria / Burkholderiales - 6641820 191 2
DBG bin17 - Proteobacteria / Burkholderiales - 47714 111 7
DMC bin110 - Proteobacteria / Burkholderiales - 4649862 596 1
KBS bin119 - Proteobacteria / Burkholderiales - 17967661 112 12
SRR5452140 bin23 - Proteobacteria / Burkholderiales - 1605893 44 5
DMC bin147 - Proteobacteria / Burkholderiales - 5316738 209 2
SRR7091524 bin13 - Proteobacteria / Burkholderiales - 1019930 49 2
JMR bin10 - Proteobacteria / Burkholderiales - 1101211 1467 1
BEO bin46 - Proteobacteria / Burkholderiales - 932220 518 5
LEF bin33 - Proteobacteria / Burkholderiales - 6631063 86 12
KBS bin113 - Proteobacteria / Burkholderiales - 5742618 3 53
ABD bin82 - Proteobacteria / Burkholderiales - 4506452 97 13
SRR7003765 bin9 - Proteobacteria / Burkholderiales - 2083691 214 5
WP 011373663.1 - Sulfurimonas denitrificans
WP 084274962.1 - Nitratiruptor tergarcus
WP 007546482.1 - Sulfurihydrogenibium yellowstonense
WP 012991503.1 - Thermocrinis albus
WP 014462593.1 - Hydrogenobacter thermophilus
WP 010880096.1 - Aquifex aeolicus
WP 066881116.1 - Dechloromonas denitrificans
WP 121243037.1 - Sulfurisoma sediminicola
WP 035269189.1 - Alcaligenes faecalis
WP 014428229.1 - Rubrivivax gelatinosus
SRR7091524 bin6 - Proteobacteria / Burkholderiales - 1616494 202 2
WP 063665046.1 - Comamonas thiooxydans
0.1

Figure S8. Maximum-likelihood tree of amino acid sequences of dissimilatory sulfite reductase A subunit (DsrA). The tree shows sequences from soil metagenome-assembled genomes (blue) alongside representative reference sequences (black). This enzyme is a marker for dissimilatory sulfite reduction (reductive and ancestral clades; Nitrospirota and Desulfobacterota bins) and sulfide oxidation (oxidative clade, r-DsrA; Proteobacteria bin). The tree was constructed using the JTT matrix-based model, used all sites, and was bootstrapped with 50 replicates and midpoint-rooted.

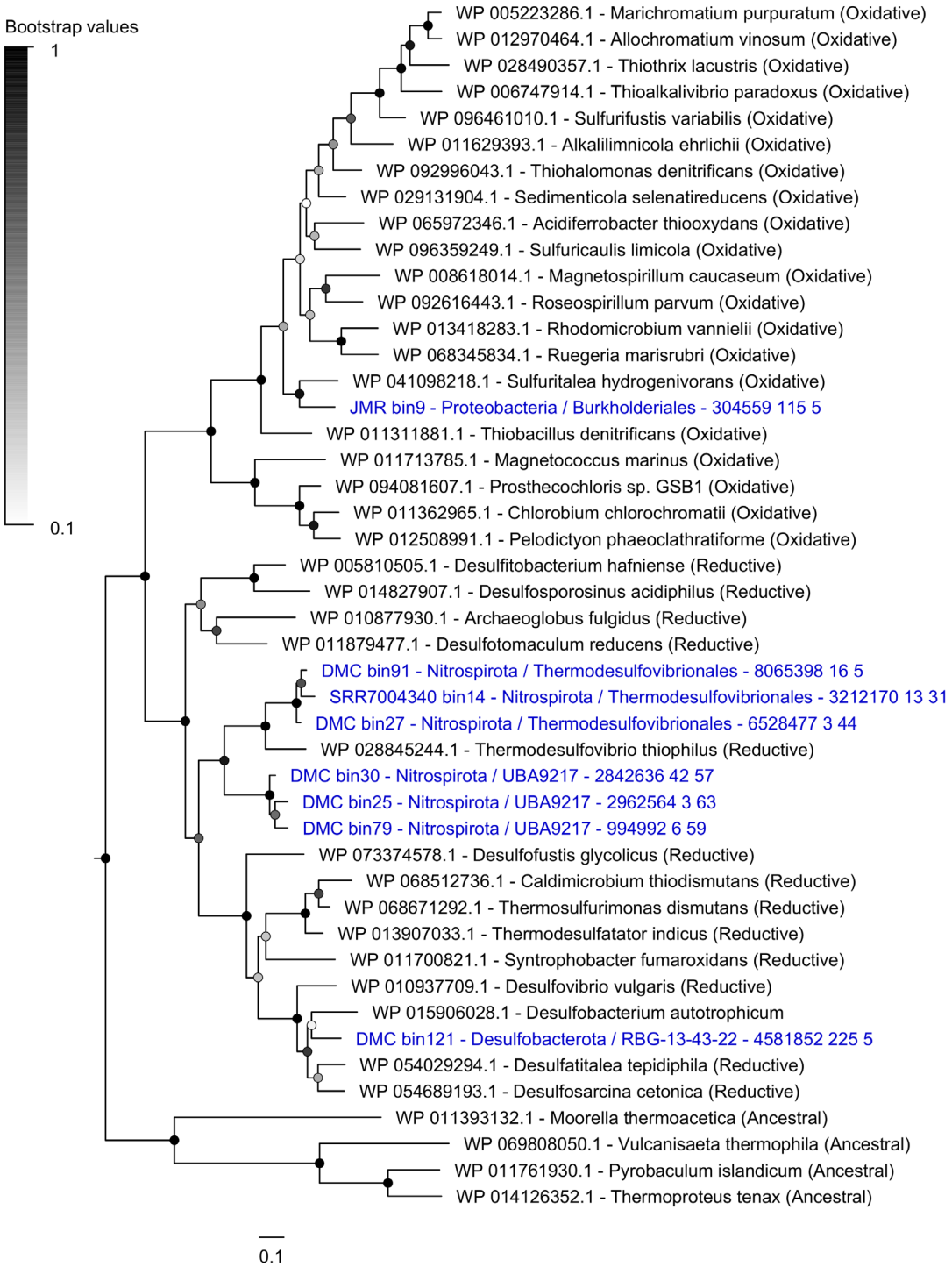


Figure S9. Maximum-likelihood tree of amino acid sequences of thiosulfohydrolase (SoxB), a marker for thiosulfate oxidation. The tree shows sequences from soil metagenome-assembled genomes (blue) and unbinned contigs (red) alongside representative reference sequences (black). The tree was constructed using the JTT matrix-based model, used all sites, and was bootstrapped with 50 replicates and midpoint-rooted.

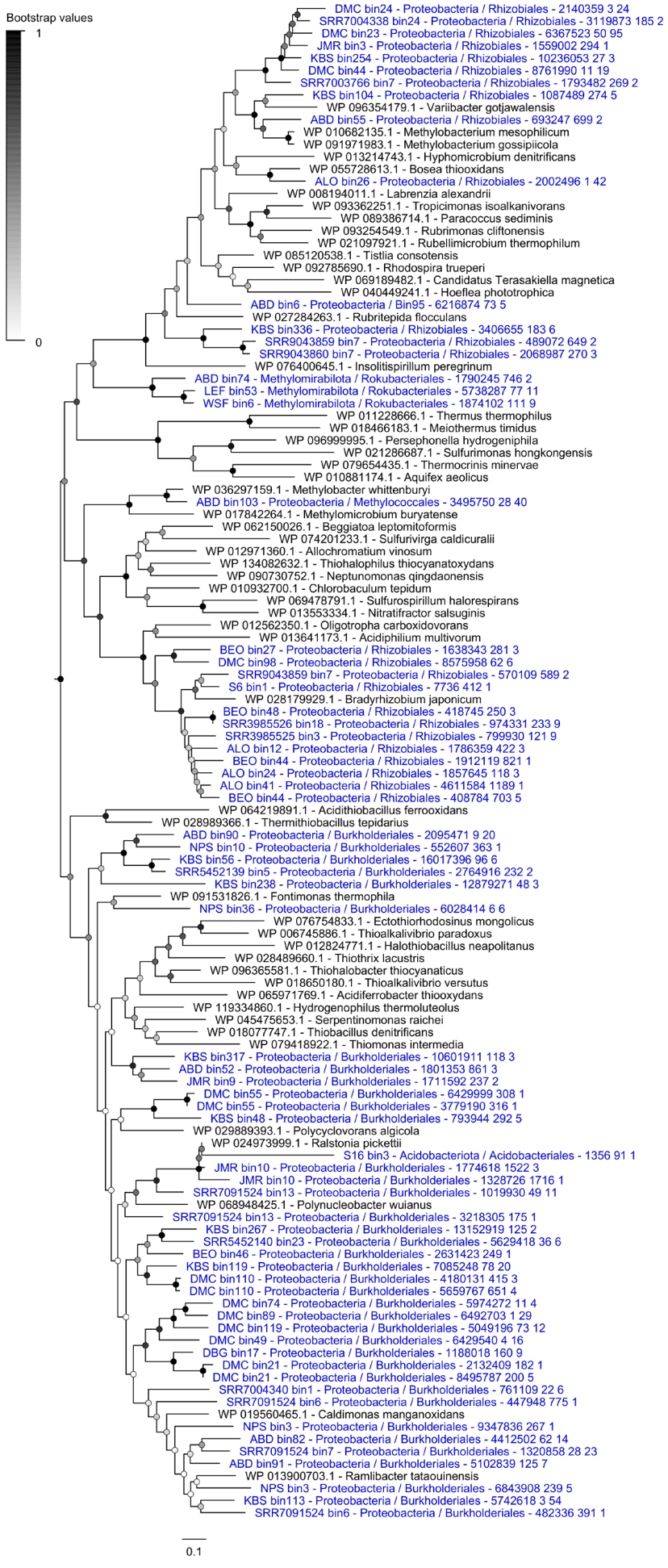

Bootstrap values
1
0
DMC bin24 - Proteobacteria / Rhizobiales - 2140359 3 24
SRR7004338 bin24 - Proteobacteria / Rhizobiales - 3119873 185 2
DMC bin23 - Proteobacteria / Rhizobiales - 6367523 50 95
JMR bin3 - Proteobacteria / Rhizobiales - 1559002 294 1
KBS bin254 - Proteobacteria / Rhizobiales - 10236053 27 3
DMC bin44 - Proteobacteria / Rhizobiales - 8761990 11 19
SRR7003766 bin7 - Proteobacteria / Rhizobiales - 1793482 269 2
KBS bin104 - Proteobacteria / Rhizobiales - 1087489 274 5
WP 096354179.1 - Variibacter gotjawalensis
ABD bin55 - Proteobacteria / Rhizobiales - 693247 699 2
WP 010682135.1 - Methylobacterium mesophilicum
WP 091971983.1 - Methylobacterium gossipiicola
WP 013214743.1 - Hyphomicrobium denitrificans
WP 055728613.1 - Bosea thiooxidans
ALO bin26 - Proteobacteria / Rhizobiales - 2002496 1 42
WP 008194011.1 - Labrenzia alexandrii
WP 093362251.1 - Tropicimonas isoalkanivorans
WP 089386714.1 - Paracoccus sediminis
WP 093254549.1 - Rubrimonas cliftonensis
WP 021097921.1 - Rubellimicrobium thermophilum
WP 085120538.1 - Tistlia consotensis
WP 092785690.1 - Rhodospira trueperi
WP 069189482.1 - Candidatus Terasakiella magnetica
WP 040449241.1 - Hoeflea phototrophica
ABD bin6 - Proteobacteria / Bin95 - 6216874 73 5
WP 027284263.1 - Rubritepida flocculans
KBS bin336 - Proteobacteria / Rhizobiales - 3406655 183 6
SRR9043859 bin7 - Proteobacteria / Rhizobiales - 489072 649 2
SRR9043860 bin7 - Proteobacteria / Rhizobiales - 2068987 270 3
WP 076400645.1 - Insolitispirillum peregrinum
ABD bin74 - Methylomirabilota / Rokubacteriales - 1790245 746 2
LEF bin53 - Methylomirabilota / Rokubacteriales - 5738287 77 11
WSF bin6 - Methylomirabilota / Rokubacteriales - 1874102 111 9
WP 011228666.1 - Thermus thermophilus
WP 018466183.1 - Meiothermus timidus
WP 096999995.1 - Persephonella hydrogeniphila
WP 021286687.1 - Sulfurimonas hongkongensis
WP 079654435.1 - Thermocrinis minervae
WP 010881174.1 - Aquifex aeolicus
WP 036297159.1 - Methylobacter whittenburyi
ABD bin103 - Proteobacteria / Methylococcales - 3495750 28 40
WP 017842264.1 - Methylomicrobium buryatense
WP 062150026.1 - Beggiatoa leptomitoformis
WP 074201233.1 - Sulfurivirga caldicuralii
WP 012971360.1 - Allochromatium vinosum
WP 134082632.1 - Thiohalophilus thiocyanatoxydans
WP 090730752.1 - Neptunomonas qingdaonensis
WP 010932700.1 - Chlorobaculum tepidum
WP 069478791.1 - Sulfurospirillum halorespirans
WP 013553334.1 - Nitratifractor salsuginis
WP 012562350.1 - Oligotropha carboxidovorans
WP 013641173.1 - Acidiphilium multivorum
BEO bin27 - Proteobacteria / Rhizobiales - 1638343 281 3
DMC bin98 - Proteobacteria / Rhizobiales - 8575958 62 6
SRR9043859 bin7 - Proteobacteria / Rhizobiales - 570109 589 2
S6 bin1 - Proteobacteria / Rhizobiales - 7736 412 1
WP 028179929.1 - Bradyrhizobium japonicum
BEO bin48 - Proteobacteria / Rhizobiales - 418745 250 3
SRR3985526 bin18 - Proteobacteria / Rhizobiales - 974331 233 9
SRR3985525 bin3 - Proteobacteria / Rhizobiales - 799930 121 9
ALO bin12 - Proteobacteria / Rhizobiales - 1786359 422 3
BEO bin44 - Proteobacteria / Rhizobiales - 1912119 821 1
ALO bin24 - Proteobacteria / Rhizobiales - 1857645 118 3
ALO bin41 - Proteobacteria / Rhizobiales - 4611584 1189 1
BEO bin44 - Proteobacteria / Rhizobiales - 408784 703 5
WP 064219891.1 - Acidithiobacillus ferrooxidans
WP 028989366.1 - Thermithiobacillus tepidarius
ABD bin90 - Proteobacteria / Burkholderiales - 2095471 9 20
NPS bin10 - Proteobacteria / Burkholderiales - 552607 363 1
KBS bin56 - Proteobacteria / Burkholderiales - 16017396 96 6
SRR5452139 bin5 - Proteobacteria / Burkholderiales - 2764916 232 2
KBS bin238 - Proteobacteria / Burkholderiales - 12879271 48 3
WP 091531826.1 - Fontimonas thermophila
NPS bin36 - Proteobacteria / Burkholderiales - 6028414 6 6
WP 076754833.1 - Ectothiorhodosinus mongolicus
WP 006745886.1 - Thioalkalivibrio paradoxus
WP 012824771.1 - Halothiobacillus neapolitanus
WP 028489660.1 - Thiothrix lacustris
WP 096365581.1 - Thiohalobacter thiocyanaticus
WP 018650180.1 - Thioalkalivibrio versutus
WP 065971769.1 - Acidiferrobacter thiooxydans
WP 119334860.1 - Hydrogenophilus thermoluteolus
WP 045475653.1 - Serpentinomonas raichei
WP 018077747.1 - Thiobacillus denitrificans
WP 079418922.1 - Thiomonas intermedia
KBS bin317 - Proteobacteria / Burkholderiales - 10601911 118 3
ABD bin52 - Proteobacteria / Burkholderiales - 1801353 861 3
JMR bin9 - Proteobacteria / Burkholderiales - 1711592 237 2
DMC bin55 - Proteobacteria / Burkholderiales - 6429999 308 1
DMC bin55 - Proteobacteria / Burkholderiales - 3779190 316 1
KBS bin48 - Proteobacteria / Burkholderiales - 793944 292 5
WP 029889393.1 - Polycyclovorans algicola
WP 024973999.1 - Ralstonia pickettii
S16 bin3 - Acidobacteriota / Acidobacteriales - 1356 91 1
JMR bin10 - Proteobacteria / Burkholderiales - 1774618 1522 3
JMR bin10 - Proteobacteria / Burkholderiales - 1328726 1716 1
SRR7091524 bin13 - Proteobacteria / Burkholderiales - 1019930 49 11
WP 068948425.1 - Polynucleobacter wuianus
SRR7091524 bin13 - Proteobacteria / Burkholderiales - 3218305 175 1
KBS bin267 - Proteobacteria / Burkholderiales - 13152919 125 2
SRR5452140 bin23 - Proteobacteria / Burkholderiales - 5629418 36 6
BEO bin46 - Proteobacteria / Burkholderiales - 2631423 249 1
KBS bin119 - Proteobacteria / Burkholderiales - 7085248 78 20
DMC bin110 - Proteobacteria / Burkholderiales - 4180131 415 3
DMC bin110 - Proteobacteria / Burkholderiales - 5659767 651 4
DMC bin74 - Proteobacteria / Burkholderiales - 5974272 11 4
DMC bin89 - Proteobacteria / Burkholderiales - 6492703 1 29
DMC bin119 - Proteobacteria / Burkholderiales - 5049196 73 12
DMC bin49 - Proteobacteria / Burkholderiales - 6429540 4 16
DBG bin17 - Proteobacteria / Burkholderiales - 1188018 160 9
DMC bin21 - Proteobacteria / Burkholderiales - 2132409 182 1
DMC bin21 - Proteobacteria / Burkholderiales - 8495787 200 5
SRR7004340 bin1 - Proteobacteria / Burkholderiales - 761109 22 6
SRR7091524 bin6 - Proteobacteria / Burkholderiales - 447948 775 1
WP 019560465.1 - Caldimonas manganoxidans
NPS bin3 - Proteobacteria / Burkholderiales - 9347836 267 1
ABD bin82 - Proteobacteria / Burkholderiales - 4412502 62 14
SRR7091524 bin7 - Proteobacteria / Burkholderiales - 1320858 28 23
ABD bin91 - Proteobacteria / Burkholderiales - 5102839 125 7
WP 013900703.1 - Ramlibacter tataouinensis
NPS bin3 - Proteobacteria / Burkholderiales - 6843908 239 5
KBS bin113 - Proteobacteria / Burkholderiales - 5742618 3 54
SRR7091524 bin6 - Proteobacteria / Burkholderiales - 482336 391 1
0.1

Figure S10. Maximum-likelihood tree of amino acid sequences of ammonia monooxygenase A subunit (AmoA), a marker for ammonia oxidation during nitrification. The tree shows sequences from soil metagenome-assembled genomes (blue) alongside representative reference sequences (black). The tree was constructed using the JTT matrix-based model, used all sites, and was bootstrapped with 50 replicates and midpoint-rooted.

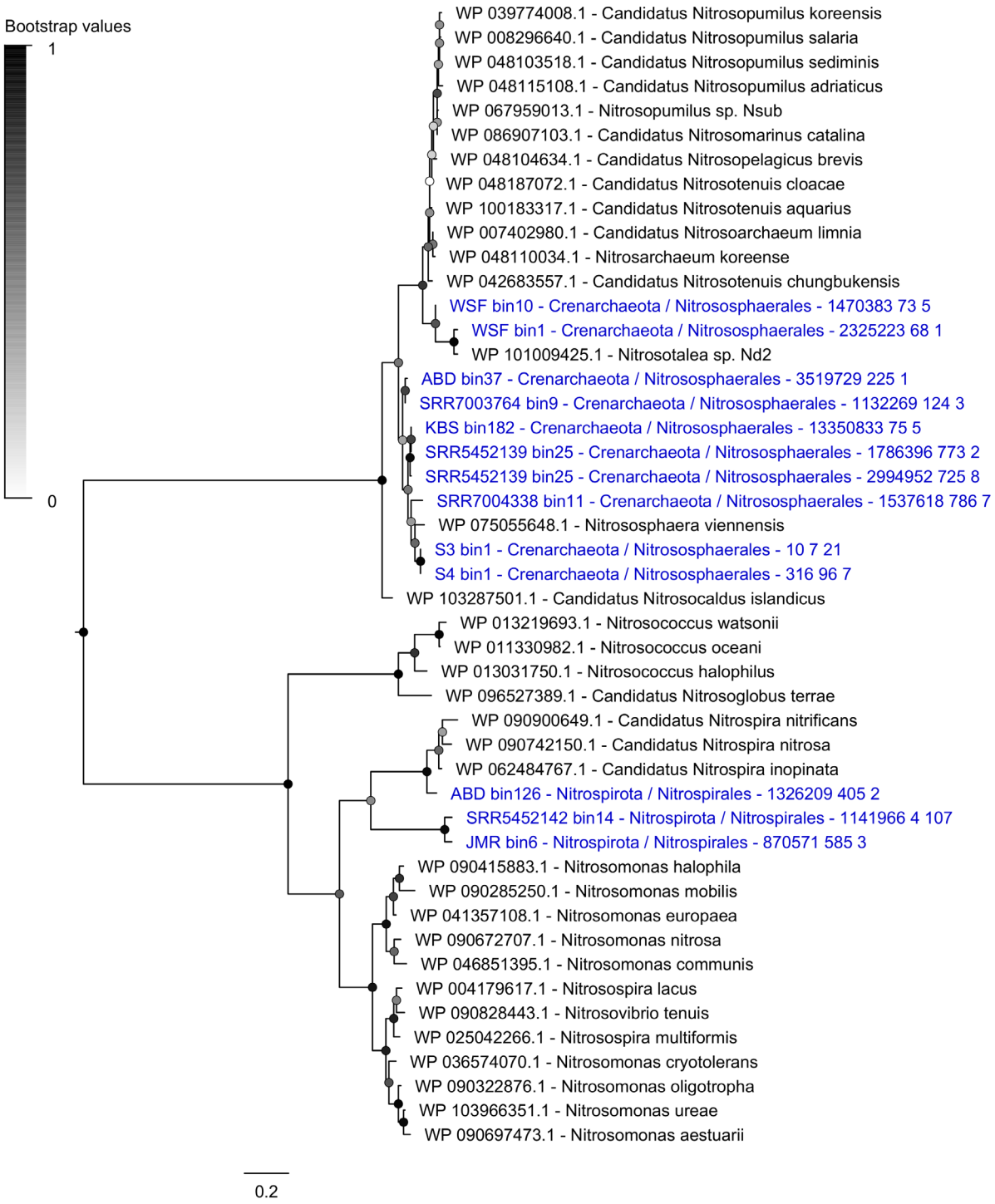


Figure S11. Maximum-likelihood tree of amino acid sequences of nitrite oxidoreductase A subunit (NxrA), a marker for nitrite oxidation during nitrification. The tree shows sequences from soil metagenome-assembled genomes (blue) alongside representative reference sequences (black). The tree was constructed using the JTT matrix-based model, used all sites, and was bootstrapped with 50 replicates and midpoint-rooted.

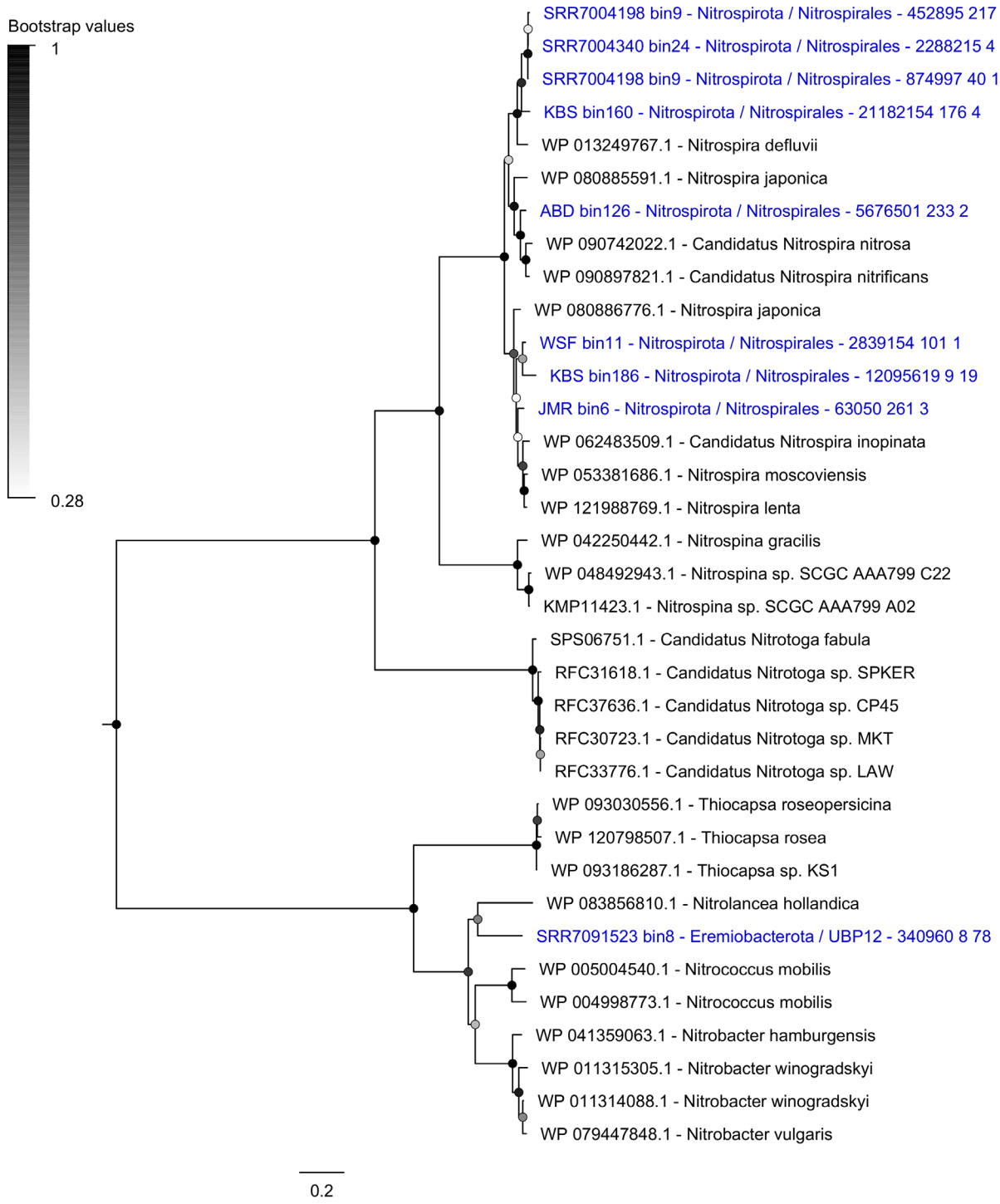


Figure S12. Maximum-likelihood tree of amino acid sequences of particulate methane monooxygenase A subunit (PmoA), a marker for aerobic methane oxidation. The tree shows sequences from soil metagenome-assembled genomes (blue) and unbinned contigs (red) alongside representative reference sequences (black). The tree shows the affiliation of the PmoA from *Candidatus* Methylotropicum kingii with those of amplicons of the tropical upland soil cluster (TUSC). The tree was constructed using the JTT matrix-based model, used all sites, and was bootstrapped with 50 replicates and midpoint-rooted.

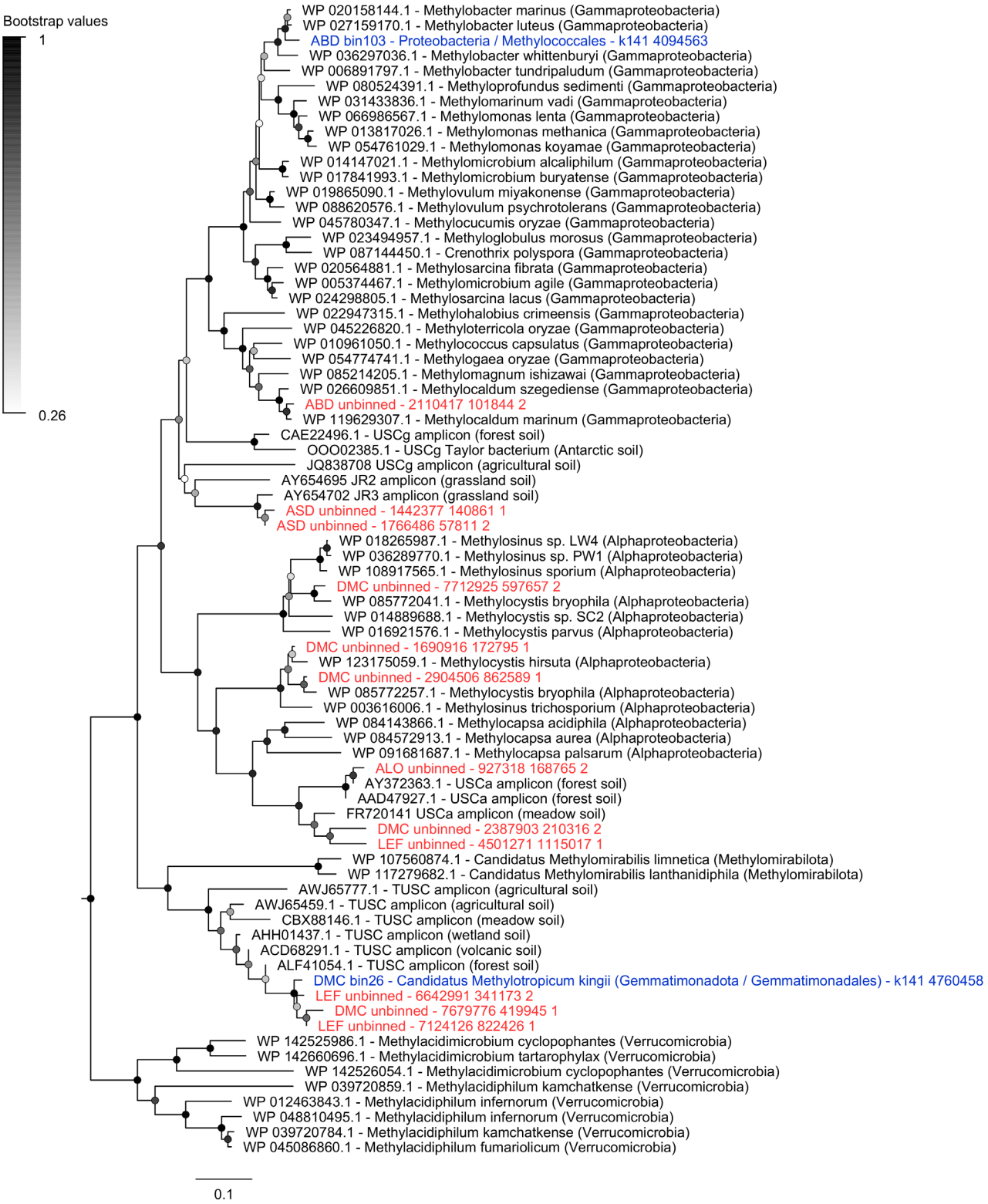

Bootstrap values
1
0.26
WP 020158144.1 - Methylobacter marinus (Gammaproteobacteria)
WP 027159170.1 - Methylobacter luteus (Gammaproteobacteria)
ABD bin103 - Proteobacteria / Methylococcales - k141 4094563
WP 036297036.1 - Methylobacter whittenburyi (Gammaproteobacteria)
WP 006891797.1 - Methylobacter tundripaludum (Gammaproteobacteria)
WP 080524391.1 - Methyloprofundus sedimenti (Gammaproteobacteria)
WP 031433836.1 - Methylomarinum vadi (Gammaproteobacteria)
WP 066986567.1 - Methylomonas lenta (Gammaproteobacteria)
WP 013817026.1 - Methylomonas methanica (Gammaproteobacteria)
WP 054761029.1 - Methylomonas koyamae (Gammaproteobacteria)
WP 014147021.1 - Methylomicrobium alcaliphilum (Gammaproteobacteria)
WP 017841993.1 - Methylomicrobium buryatense (Gammaproteobacteria)
WP 019865090.1 - Methylovulum miyakonense (Gammaproteobacteria)
WP 088620576.1 - Methylovulum psychrotolerans (Gammaproteobacteria)
WP 045780347.1 - Methylocucumis oryzae (Gammaproteobacteria)
WP 023494957.1 - Methyloglobulus morosus (Gammaproteobacteria)
WP 087144450.1 - Crenothrix polyspora (Gammaproteobacteria)
WP 020564881.1 - Methylosarcina fibrata (Gammaproteobacteria)
WP 005374467.1 - Methylomicrobium agile (Gammaproteobacteria)
WP 024298805.1 - Methylosarcina lacus (Gammaproteobacteria)
WP 022947315.1 - Methylohalobius crimeensis (Gammaproteobacteria)
WP 045226820.1 - Methyloterricola oryzae (Gammaproteobacteria)
WP 010961050.1 - Methylococcus capsulatus (Gammaproteobacteria)
WP 054774741.1 - Methylogaea oryzae (Gammaproteobacteria)
WP 085214205.1 - Methylomagnum ishizawai (Gammaproteobacteria)
WP 026609851.1 - Methylocaldum szegediense (Gammaproteobacteria)
ABD unbinned - 2110417 101844 2
WP 119629307.1 - Methylocaldum marinum (Gammaproteobacteria)
CAE22496.1 - USCg amplicon (forest soil)
OOO02385.1 - USCg Taylor bacterium (Antarctic soil)
JQ838708 USCg amplicon (agricultural soil)
AY654695 JR2 amplicon (grassland soil)
AY654702 JR3 amplicon (grassland soil)
ASD unbinned - 1442377 140861 1
ASD unbinned - 1766486 57811 2
WP 018265987.1 - Methylosinus sp. LW4 (Alphaproteobacteria)
WP 036289770.1 - Methylosinus sp. PW1 (Alphaproteobacteria)
WP 108917565.1 - Methylosinus sporium (Alphaproteobacteria)
DMC unbinned - 7712925 597657 2
WP 085772041.1 - Methylocystis bryophila (Alphaproteobacteria)
WP 014889688.1 - Methylocystis sp. SC2 (Alphaproteobacteria)
WP 016921576.1 - Methylocystis parvus (Alphaproteobacteria)
DMC unbinned - 1690916 172795 1
WP 123175059.1 - Methylocystis hirsuta (Alphaproteobacteria)
DMC unbinned - 2904506 862589 1
WP 085772257.1 - Methylocystis bryophila (Alphaproteobacteria)
WP 003616006.1 - Methylosinus trichosporium (Alphaproteobacteria)
WP 084143866.1 - Methylocapsa acidiphila (Alphaproteobacteria)
WP 084572913.1 - Methylocapsa aurea (Alphaproteobacteria)
WP 091681687.1 - Methylocapsa palsarum (Alphaproteobacteria)
ALO unbinned - 927318 168765 2
AY372363.1 - USCa amplicon (forest soil)
AAD47927.1 - USCa amplicon (forest soil)
FR720141 USCa amplicon (meadow soil)
DMC unbinned - 2387903 210316 2
LEF unbinned - 4501271 1115017 1
WP 107560874.1 - Candidatus Methylomirabilis limnetica (Methylomirabilota)
WP 117279682.1 - Candidatus Methylomirabilis lanthanidiphila (Methylomirabilota)
AWJ65777.1 - TUSC amplicon (agricultural soil)
AWJ65459.1 - TUSC amplicon (agricultural soil)
CBX88146.1 - TUSC amplicon (meadow soil)
AHH01437.1 - TUSC amplicon (wetland soil)
ACD68291.1 - TUSC amplicon (volcanic soil)
ALF41054.1 - TUSC amplicon (forest soil)
DMC bin26 - Candidatus Methylotropicum kingii (Gemmatimonadota / Gemmatimonadales) - k141 4760458
LEF unbinned - 6642991 341173 2
DMC unbinned - 7679776 419945 1
LEF unbinned - 7124126 822426 1
WP 142525986.1 - Methylacidimicrobium cyclopophantes (Verrucomicrobia)
WP 142660696.1 - Methylacidimicrobium tartarophylax (Verrucomicrobia)
WP 142526054.1 - Methylacidimicrobium cyclopophantes (Verrucomicrobia)
WP 039720859.1 - Methylacidiphilum kamchatkense (Verrucomicrobia)
WP 012463843.1 - Methylacidiphilum infernorum (Verrucomicrobia)
WP 048810495.1 - Methylacidiphilum infernorum (Verrucomicrobia)
WP 039720784.1 - Methylacidiphilum kamchatkense (Verrucomicrobia)
WP 045086860.1 - Methylacidiphilum fumariolicum (Verrucomicrobia)
0.1

Figure S13. Metabolic reconstruction of the putative novel methanotroph *Candidatus* Methylotropicum kingii. The core pathways associated with energy conservation and carbon acquisition are shown, with genes detected shown in italics. The bacterium is predicted to use methane, methanol, and acetate as energy and carbon sources. In addition, it can use molecular hydrogen as an electron donor via a group 1f [NiFe]-hydrogenase. The bacterium is predicted to use the electron acceptors oxygen *via* a cytochrome *c* oxidase and nitrous oxide *via* a nitrous oxide reductase. Its particulate methane monooxygenase forms a distinct phylogenetic lineage with amplicons from the Tropical Upland Soil Cluster (TUSC), whereas its methanol dehydrogenase is closely related to those in previously sequenced Gemmatimonadota MAGs inferred to be methylotrophic. The genome encodes key enzymes for the serine cycle for assimilation of one-carbon sources. Abbreviations: H_4F = tetrahydrofolate; Hyd = group 1f [NiFe]; pMMO = particulate methane monooxygenase; MDH = methanol dehydrogenase; PQQ = pyrroloquinoline quinone; I = NADH dehydrogenase (complex I); complex II = succinate dehydrogenase; IV = cytochrome *aa*$_3$ oxidase. Dashed black lines indicate diffusion. Dashed gray lines indicate unknown regulation mechanism/not detected genes.

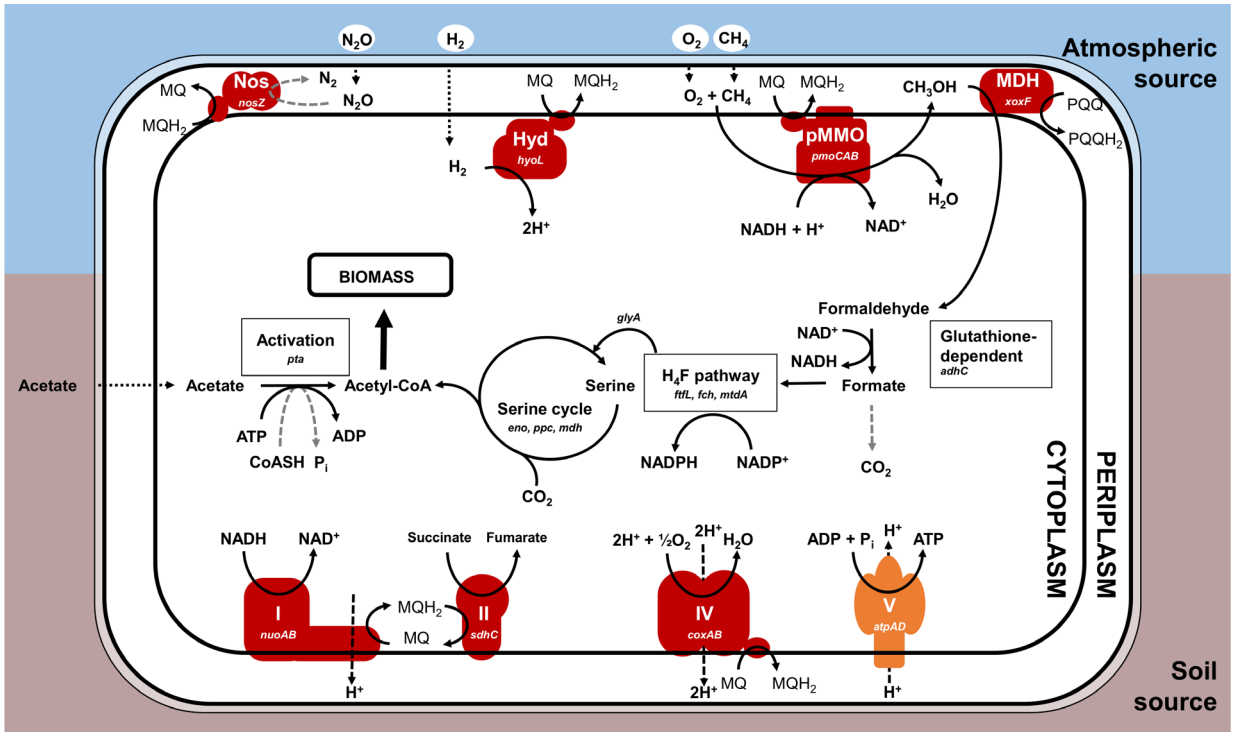


Figure S14. Mean soil-gas profiles normalized to the respective ambient air concentration (dashed line) for each Australian biomes. Note that the different gases were sampled at identical depths, but points are plotted slightly offset on the y-axis for better visibility of error bars.

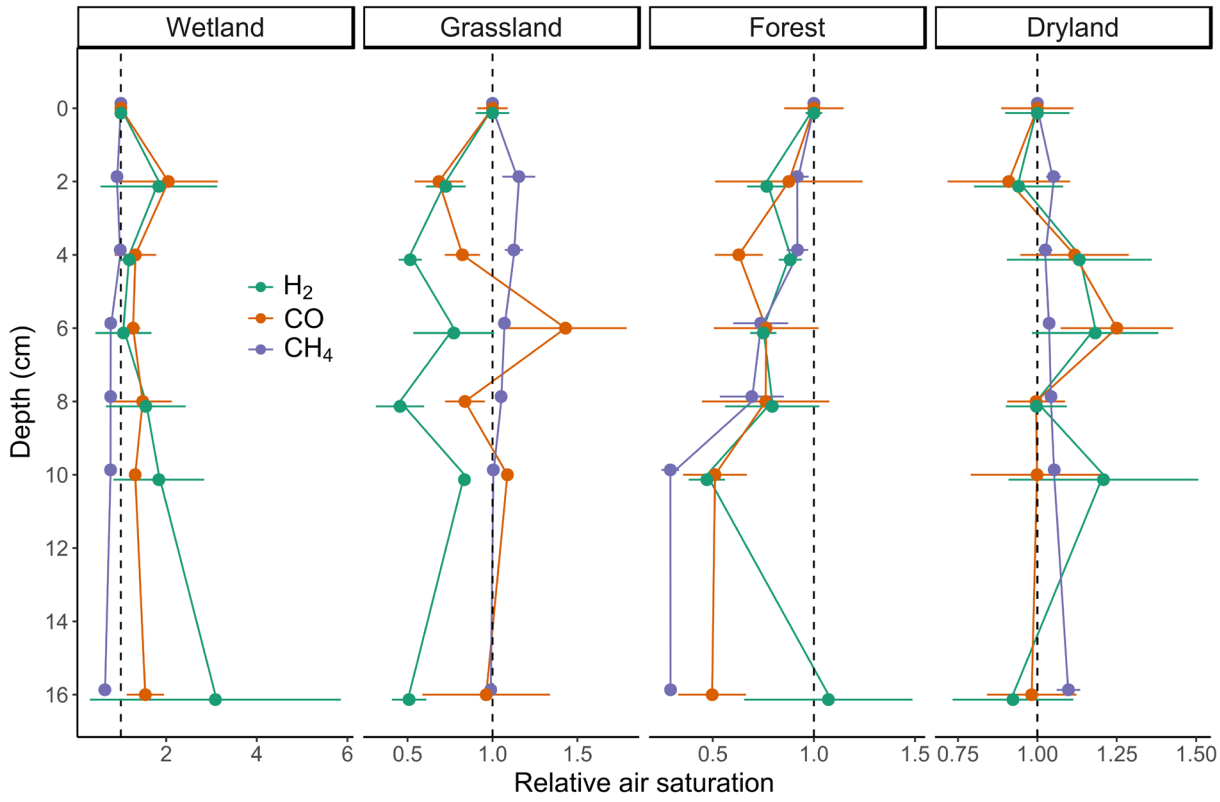


Figure S15. Maximum-likelihood tree of amino acid sequences of group 3 [NiFe]-hydrogenase large subunits, a marker for hydrogen production during fermentation processes. The tree shows sequences from the soil metagenome-assembled genomes (blue) alongside representative reference sequences (black). The subgroup of each reference sequence is denoted according to the HydDB classification scheme [359]. The tree was constructed using the JTT matrix-based model, used all sites, and was bootstrapped with 50 replicates and midpoint-rooted. All sequences shorter than 350 amino acids were omitted.

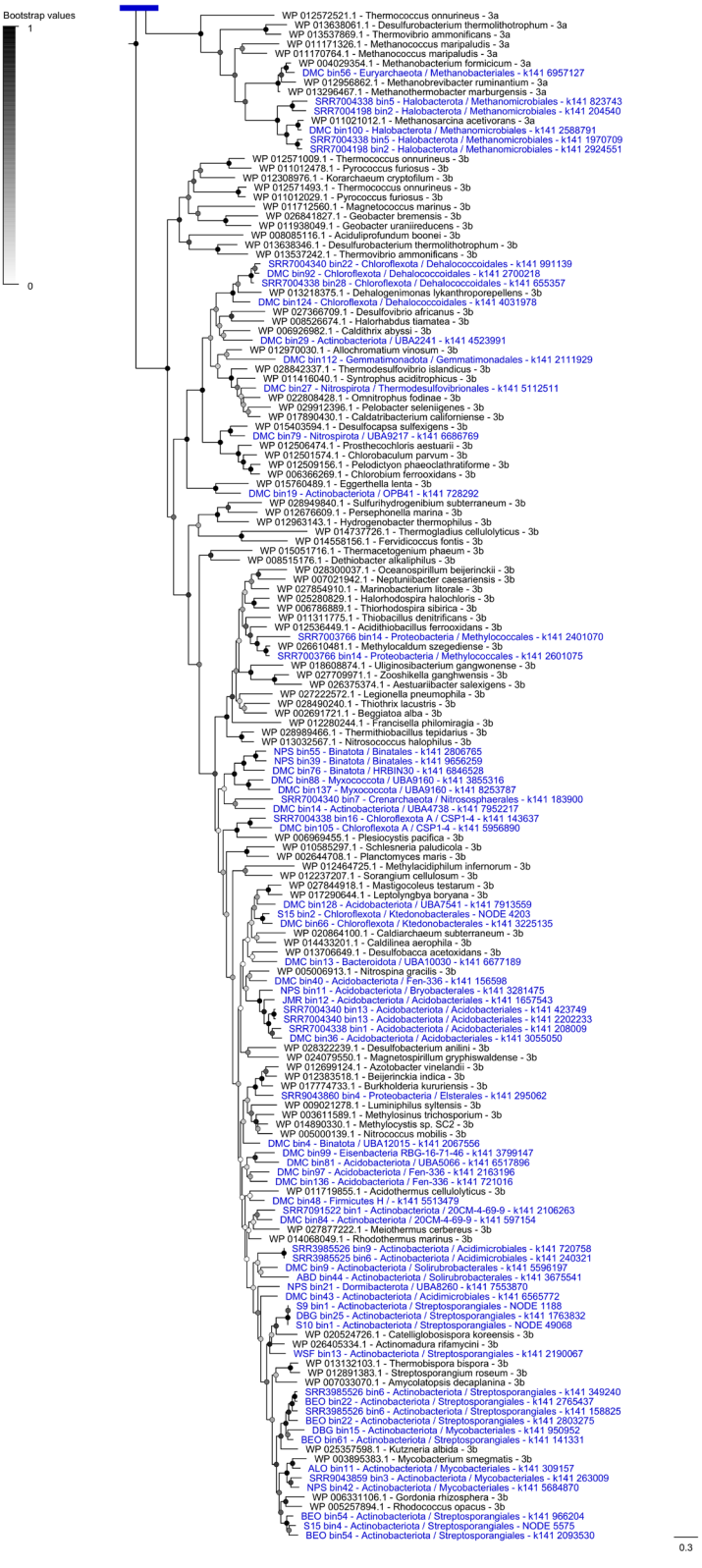

Bootstrap values
1
0
WP 012572521.1 - Thermococcus onnurineus - 3a
WP 013638061.1 - Desulfurobacterium thermolithotrophum - 3a
WP 013537869.1 - Thermovibrio ammonificans - 3a
WP 011171326.1 - Methanococcus maripaludis - 3a
WP 011170764.1 - Methanococcus maripaludis - 3a
WP 004029354.1 - Methanobacterium formicicum - 3a
DMC bin56 - Euryarchaeota / Methanobacteriales - k141 6957127
WP 012956862.1 - Methanobrevibacter ruminantium - 3a
WP 013296467.1 - Methanothermobacter marburgensis - 3a
SRR7004338 bin5 - Halobacterota / Methanomicrobiales - k141 823743
SRR7004198 bin2 - Halobacterota / Methanomicrobiales - k141 204540
WP 011021012.1 - Methanosarcina acetivorans - 3a
DMC bin100 - Halobacterota / Methanomicrobiales - k141 2586791
SRR7004338 bin5 - Halobacterota / Methanomicrobiales - k141 1970709
SRR7004198 bin2 - Halobacterota / Methanomicrobiales - k141 2924551
WP 012571009.1 - Thermococcus onnurineus - 3b
WP 011012478.1 - Pyrococcus furiosus - 3b
WP 012308976.1 - Korarchaeum cryptofilum - 3b
WP 012571493.1 - Thermococcus onnurineus - 3b
WP 011012029.1 - Pyrococcus furiosus - 3b
WP 011712560.1 - Magnetococcus marinus - 3b
WP 026841827.1 - Geobacter bremensis - 3b
WP 011938049.1 - Geobacter uraniireducens - 3b
WP 008085116.1 - Acidulprofundum boonei - 3b
WP 013638346.1 - Desulfurobacterium thermolithotrophum - 3b
WP 013537242.1 - Thermovibrio ammonificans - 3b
SRR7004340 bin22 - Chloroflexota / Dehalococcoidales - k141 991139
DMC bin92 - Chloroflexota / Dehalococcoidales - k141 2700218
SRR7004338 bin28 - Chloroflexota / Dehalococcoidales - k141 655357
WP 013218375.1 - Dehalogenimonas lykanthroporepellens - 3b
DMC bin124 - Chloroflexota / Dehalococcoidales - k141 4031978
WP 027366709.1 - Desulfovibrio africanus - 3b
WP 008526674.1 - Halorhabdus tiamatea - 3b
WP 006926982.1 - Caldithrix abyssi - 3b
DMC bin29 - Actinobacteriota / UBA2241 - k141 4523991
WP 012970030.1 - Allochromatium vinosum - 3b
DMC bin112 - Gemmatimonadota / Gemmatimonadales - k141 2111929
WP 028842337.1 - Thermodesulfovibrio islandicus - 3b
WP 011416040.1 - Syntrophus aciditrophicus - 3b
DMC bin27 - Nitrospirota / Thermodesulfovibrionales - k141 5112511
WP 022808428.1 - Omnitrophus fodinae - 3b
WP 029912396.1 - Pelobacter seleniigenes - 3b
WP 017890430.1 - Caldatribacterium californiense - 3b
WP 015403594.1 - Desulfocapsa sulfexigens - 3b
DMC bin79 - Nitrospirota / UBA9217 - k141 6686769
WP 012506474.1 - Prosthecochloris aestuarii - 3b
WP 012501574.1 - Chlorobaculum parvum - 3b
WP 012509156.1 - Pelodictyon phaeoclathratiforme - 3b
WP 006366269.1 - Chlorobium ferrooxidans - 3b
WP 015760489.1 - Eggerthella lenta - 3b
DMC bin19 - Actinobacteriota / OPB41 - k141 728292
WP 028949840.1 - Sulfurihydrogenibium subterraneum - 3b
WP 012676609.1 - Persephonella marina - 3b
WP 012963143.1 - Hydrogenobacter thermophilus - 3b
WP 014737726.1 - Thermogladius cellulolyticus - 3b
WP 014558156.1 - Fervidicoccus fontis - 3b
WP 015051716.1 - Thermotogenium phaeum - 3b
WP 008515176.1 - Dethiobacter alkaliphilus - 3b
WP 028300037.1 - Oceanospirillum beijerinckii - 3b
WP 007021942.1 - Neptuniibacter caesariensis - 3b
WP 027854910.1 - Marinobacterium litorale - 3b
WP 025280829.1 - Halorhodospira halochloris - 3b
WP 006788889.1 - Thiorhodospira sibirica - 3b
WP 011311775.1 - Thiobacillus denitrificans - 3b
WP 012536449.1 - Acidithiobacillus ferrooxidans - 3b
SRR7003766 bin14 - Proteobacteria / Methylococcales - k141 2401070
WP 026610481.1 - Methylocaldum szegediense - 3b
SRR7003766 bin14 - Proteobacteria / Methylococcales - k141 2601075
WP 018608874.1 - Uliginosibacterium gangwonense - 3b
WP 027709971.1 - Zoogloea caeni - 3b
WP 026375374.1 - Aestuariibacter salexigens - 3b
WP 027222572.1 - Legionella pneumophila - 3b
WP 028490240.1 - Thiothrix lacustris - 3b
WP 002691721.1 - Beggiatoa alba - 3b
WP 012280244.1 - Francisella philomiragia - 3b
WP 028989466.1 - Thermithiobacillus tepidarius - 3b
WP 013032567.1 - Nitrosococcus halophilus - 3b
NPS bin55 - Binatota / Binatales - k141 2806765
NPS bin39 - Binatota / Binatales - k141 9656259
DMC bin76 - Binatota / HRBIN30 - k141 6846528
DMC bin88 - Myxococcota / UBA9160 - k141 3855316
DMC bin137 - Myxococcota / UBA9160 - k141 8253787
SRR7004340 bin7 - Crenarchaeota / Nitrososphaerales - k141 183900
DMC bin14 - Actinobacteriota / UBA4738 - k141 7952217
SRR7004338 bin16 - Chloroflexota A / CSP1-4 - k141 143637
DMC bin105 - Chloroflexota A / CSP1-4 - k141 5956890
WP 006969455.1 - Plesiocystis pacifica - 3b
WP 010585297.1 - Schlesneria paludicola - 3b
WP 002644708.1 - Planctomyces maris - 3b
WP 012464725.1 - Methylacidiphilum infernorum - 3b
WP 012237207.1 - Sorangium cellulosum - 3b
WP 027844918.1 - Mastigocoleus testarum - 3b
WP 017290644.1 - Leptolyngbya boryana - 3b
DMC bin128 - Acidobacteriota / UBA7541 - k141 7913559
S15 bin2 - Chloroflexota / Ktedonobacterales - NODE 4203
DMC bin66 - Chloroflexota / Ktedonobacterales - k141 3225135
WP 020864100.1 - Caldiarchaeum subterraneum - 3b
WP 014433201.1 - Caldilinea aerophila - 3b
WP 013706649.1 - Desulfobacca acetoxidans - 3b
DMC bin13 - Bacteroidota / UBA10030 - k141 6677189
WP 005006913.1 - Nitrospina gracilis - 3b
DMC bin40 - Acidobacteriota / Fen-336 - k141 156598
NPS bin11 - Acidobacteriota / Bryobacterales - k141 3281475
JMR bin12 - Acidobacteriota / Acidobacteriales - k141 1657543
SRR7004340 bin13 - Acidobacteriota / Acidobacteriales - k141 423749
SRR7004340 bin13 - Acidobacteriota / Acidobacteriales - k141 2202233
SRR7004338 bin1 - Acidobacteriota / Acidobacteriales - k141 208009
DMC bin36 - Acidobacteriota / Acidobacteriales - k141 3055050
WP 028322239.1 - Desulfobacterium anilini - 3b
WP 024079550.1 - Magnetospirillum gryphiswaldense - 3b
WP 012699124.1 - Azotobacter vinelandii - 3b
WP 012383518.1 - Beijerinckia indica - 3b
WP 017774733.1 - Burkholderia ruricensis - 3b
SRR9043860 bin4 - Proteobacteria / Elsterales - k141 295062
WP 009021278.1 - Luminiphilus syltensis - 3b
WP 003611589.1 - Methylosinus trichosporium - 3b
WP 014890330.1 - Methylocystis sp. SC2 - 3b
WP 005000139.1 - Nitrococcus mobilis - 3b
DMC bin4 - Binatota / UBA12015 - k141 2067656
DMC bin99 - Eisenbacteria RBG-16-71-46 - k141 3799147
DMC bin81 - Acidobacteriota / UBA5066 - k141 6517896
DMC bin97 - Acidobacteriota / Fen-336 - k141 2163196
DMC bin136 - Acidobacteriota / Fen-336 - k141 7210116
WP 011719855.1 - Acidothermus cellulolyticus - 3b
DMC bin48 - Firmicutes H / - k141 5513479
SRR7091522 bin1 - Actinobacteriota / 20CM-4-69-9 - k141 2106263
DMC bin84 - Actinobacteriota / 20CM-4-69-9 - k141 597154
WP 027877222.1 - Meiothermus cerbereus - 3b
WP 014068049.1 - Rhodothermus marinus - 3b
SRR3985526 bin9 - Actinobacteriota / Acidimicrobiales - k141 720758
SRR3985525 bin6 - Actinobacteriota / Acidimicrobiales - k141 240321
DMC bin9 - Actinobacteriota / Solirubrobacterales - k141 5596197
ABD bin44 - Actinobacteriota / Solirubrobacterales - k141 3675541
NPS bin21 - Dormibacterota / UBA8260 - k141 7553870
DMC bin43 - Actinobacteriota / Acidimicrobiales - k141 6565772
S9 bin1 - Actinobacteriota / Streptosporangiales - NODE 1188
DBG bin25 - Actinobacteriota / Streptosporangiales - k141 1763832
S10 bin1 - Actinobacteriota / Streptosporangiales - NODE 49068
WP 020524726.1 - Catelliglobosispora koreensis - 3b
WP 026405334.1 - Actinomadura rifamycini - 3b
WSF bin13 - Actinobacteriota / Streptosporangiales - k141 2190067
WP 013132103.1 - Thermobispora bispora - 3b
WP 012891383.1 - Streptosporangium roseum - 3b
WP 007033070.1 - Amycolatopsis decaplanina - 3b
SRR3985526 bin6 - Actinobacteriota / Streptosporangiales - k141 349240
BEO bin22 - Actinobacteriota / Streptosporangiales - k141 2765437
SRR3985526 bin6 - Actinobacteriota / Streptosporangiales - k141 158825
BEO bin22 - Actinobacteriota / Streptosporangiales - k141 2803275
DBG bin15 - Actinobacteriota / Mycobacteriales - k141 950952
BEO bin61 - Actinobacteriota / Streptosporangiales - k141 141331
WP 025357598.1 - Kutzneria albida - 3b
WP 003895383.1 - Mycobacterium smegmatis - 3b
ALO bin11 - Actinobacteriota / Mycobacteriales - k141 309157
SRR9043859 bin3 - Actinobacteriota / Mycobacteriales - k141 263009
NPS bin42 - Actinobacteriota / Mycobacteriales - k141 5684870
WP 006331106.1 - Gordonia rhizosphera - 3b
WP 005257894.1 - Rhodococcus opacus - 3b
BEO bin54 - Actinobacteriota / Streptosporangiales - k141 966204
S15 bin4 - Actinobacteriota / Streptosporangiales - NODE 5575
BEO bin54 - Actinobacteriota / Streptosporangiales - k141 2093530
S10 bin6 - Acidobacteriota / Acidobacteriales - NODE 176
DBG bin5 - Acidobacteriota / Acidobacteriales - k141 158236
DBG bin14 - Acidobacteriota / Acidobacteriales - k141 1173837
DBG bin3 - Acidobacteriota / Acidobacteriales - k141 10915
KBS bin143 - Acidobacteriota / Acidobacteriales - k141 12040566
JMR bin5 - Acidobacteriota / Acidobacteriales - k141 1525134
DMC bin36 - Acidobacteriota / Acidobacteriales - k141 8499729
WP 011686721.1 - Solibacter usitatus - 3d
DMC bin66 - Chloroflexota / Ktedonobacterales - k141 8375351
SRR7091524 bin1 - Verrucomicrobiota / Chthoniobacterales - k141 2508790
NPS bin46 - Verrucomicrobiota / Chthoniobacterales - k141 9095622
WP 014808238.1 - Desulfomonile tiedjei - 3d
DMC bin96 - Acidobacteriota / Bryobacterales - k141 1534264
SRR3985625 bin19 - Actinobacteriota / Acidothermales - k141 162518
SRR3985625 bin18 - Actinobacteriota / Acidothermales - k141 413600
WP 020504162.1 - Lamprocystis purpurea - 3d
WP 012971066.1 - Allochromatium vinosum - 3d
SRR3985525 bin9 - Acidobacteriota / Acidobacteriales - k141 105197
S13 bin2 - Acidobacteriota / Acidobacteriales - NODE 2262
ALO bin43 - Acidobacteriota / Acidobacteriales - k141 4191670
WP 015724758.1 - Desulfobulbus propionicus - 3d
DMC bin97 - Acidobacteriota / Fen-336 - k141 3235530
KBS bin124 - Acidobacteriota / Bryobacterales - k141 10898952
WP 014551946.1 - Geobacter sulfurreducens - 3d
WP 027388897.1 - Chromohalobacter arsenatis - 3d
WP 013506608.1 - Desulfurispirillum indicum - 3d
WP 014434113.1 - Caldilinea aerophila - 3d
WP 020876264.1 - Desulfococcus multivorans - 3d
WP 011699561.1 - Syntrophobacter fumaroxidans - 3d
DMC bin97 - Acidobacteriota / Fen-336 - k141 591371
WP 015941206.1 - Chloroflexus aggregans - 3d
WP 006562473.1 - Oscillochloris trichoides - 3d
WP 012119893.1 - Roseiflexus castenholzii - 3d
WP 016514321.1 - Microcystis aeruginosa - 3d
WP 015221229.1 - Cyanobacterium aponinum - 3d
WP 011378516.1 - Synechococcus elongatus - 3d
KBS bin155 - Verrucomicrobiota / Pedosphaerales - k141 9273821
DMC bin11 - Verrucomicrobiota / Pedosphaerales - k141 2328337
WP 012167698.1 - Acaryochloris marina - 3d
WP 016860165.1 - Fischerella muscicola - 3d
WP 011321324.1 - Anabaena variabilis - 3d
WP 006276675.1 - Cylindrospermopsis raciborskii - 3d
DMC bin13 - Bacteroidota / UBA10030 - k141 2177153
LEF bin58 - Bacteroidota / Chitinophagales - k141 879073
DMC bin124 - Chloroflexota / Dehalococcoidales - k141 2885310
WP 024821667.1 - Aminobacterium mobile - 3d
WP 014163745.1 - Thermovirga lienii - 3d
WP 012548375.1 - Dictyoglomus thermophilum - 3d
WP 013559391.1 - Anaerolinea thermophila - 3d
DMC bin121 - Desulfobacterota RBG-13-43-22 - k141 251258
WP 008526412.1 - Halorhabdus tiamatea - 3d
WP 027856531.1 - Marinobacterium jannaschii - 3d
WP 005505390.1 - Grimontia hollisae - 3d
WP 011516268.1 - Cupriavidus metallidurans - 3d
WP 011154013.1 - Ralstonia eutropha - 3d
WP 014208491.1 - Mycobacterium rhodesiae - 3d
WP 011394103.1 - Hahella chejuensis - 3d
WP 006931831.1 - Labrenzia aggregata - 3d
NPS bin7 - Proteobacteria / Reyranellales - k141 1416627
NPS bin26 - Proteobacteria / Acetobacterales - k141 6833001
WP 012591738.1 - Methylocella silvestris - 3d
WP 028864775.1 - Psychromonas aquimarina - 3d
ABD bin103 - Proteobacteria / Methylococcales - k141 4327626
WP 020393771.1 - Thiothrix disciformis - 3d
WP 009849435.1 - Mariprofundus ferrooxydans - 3d
WP 010959484.1 - Methylococcus capsulatus - 3d
WP 029132296.1 - Sedimenticola selenatireducens - 3d
WP 020504783.1 - Lamprocystis purpurea - 3d
NPS bin31 - Proteobacteria / Rhizobiales - k141 3708132
NPS bin22 - Proteobacteria / Rhizobiales - k141 646431
DMC bin22 - Proteobacteria / Rhizobiales - k141 4502769
WP 028994701.1 - Azonexus hydrophilus - 3d
WP 020162845.1 - Methyloversatilis universalis - 3d
WP 011466117.1 - Rhodoferax ferrireducens - 3d
WP 012346489.1 - Leptothrix cholodnii - 3d
WP 011493687.1 - Burkholderia xenovorans - 3d
WP 011385756.1 - Magnetospirillum magneticum - 3d
WP 027457859.1 - Dechloromonas agilata - 3d
WP 026687097.1 - Azovibrio restrictus - 3d
WP 013293541.1 - Gallionella capsiferriformans - 3d
DMC bin21 - Proteobacteria / Burkholderiales - k141 7575049
WP 013029742.1 - Sideroxydans lithotrophicus - 3d
WP 011380994.1 - Nitrosospira multiformis - 3d
WP 009205622.1 - Sulfuricella denitrificans - 3d
SRR7004340 bin11 - Halobacterota / Methanomicrobiales - k141 998908
DMC bin100 - Halobacterota / Methanomicrobiales - k141 6816299
WP 011171638.1 - Methanococcus maripaludis - 3c
WP 013296307.1 - Methanothermobacter marburgensis - 3c
WP 012956704.1 - Methanobrevibacter ruminantium - 3c
WP 011170767.1 - Methanococcus maripaludis - 3c
DMC bin56 - Euryarchaeota / Methanobacteriales - k141 7231595
DMC bin114 - Nitrospirota / UBA9217 - k141 8782499
WP 015724435.1 - Desulfobulbus propionicus - 3c
WP 011862256.1 - Desulfotalea psychrophila - 3c
WP 011700738.1 - Syntrophobacter fumaroxidans - 3c
WP 015903075.1 - Desulfobacterium autotrophicum - 3c
WP 027368703.1 - Desulfovibrio africanus - 3c
DMC bin61 - Desulfobacterota / BSN033 - k141 2103073
WP 013218178.1 - Dehalogenimonas lykanthroporepellens - 3c
WP 010936390.1 - Dehalococcoides mccartyi - 3c
WP 015233716.1 - Natronobacterium gregoryi - 3c
WP 015722755.1 - Desulfobulbus propionicus - 3c
WP 013756665.1 - Thermodesulfobium narugense - 3c
WP 013908379.1 - Thermodesulfatator indicus - 3c
WP 013637988.1 - Desulfurobacterium thermolithotrophum - 3c
WP 012965352.1 - Ferroglobus placidus - 3c
WP 010878869.1 - Archaeoglobus fulgidus - 3c
SRR7004338 bin6 - Thermoplasmatota / Methanomassiliicoccales - k141 3030368
DMC bin82 - Thermoplasmatota / Methanomassiliicoccales - k141 2759615
SRR7004338 bin1 - Acidobacteriota / Acidobacteriales - k141 469912
DMC bin142 - Gemmatimonadota / Gemmatimonadales - k141 4819574
WP 012376486.1 - Opitutus terrae - 3c
WP 006930005.1 - Caldithrix abyssi - 3c
WP 012309762.1 - Korarchaeum cryptofilum - 3c
WP 015935331.1 - Anaeromyxobacter dehalogenans - 3c
DMC bin97 - Acidobacteriota / Fen-336 - k141 1037952
WP 022775381.1 - Symbiobacterium thiodismis - 3c
DMC bin77 - Chloroflexota A / CSP1-4 - k141 8259662
SRR7004338 bin16 - Chloroflexota A / CSP1-4 - k141 445989
DMC bin105 - Chloroflexota A / CSP1-4 - k141 2521235
DMC bin61 - Desulfobacterota / BSN033 - k141 3786115
DMC bin92 - Chloroflexota / Dehalococcoidales - k141 88237
SRR7004340 bin22 - Chloroflexota / Dehalococcoidales - k141 1543770
SRR7004338 bin28 - Chloroflexota / Dehalococcoidales - k141 2172240
SRR7004338 bin31 - Actinobacteriota / UBA2241 - k141 1939820
DMC bin28 - Actinobacteriota / UBA2241 - k141 3226005
WP 020448391.1 - Methanomassiliicoccus intestinalis - 3c
WP 019177457.1 - Methanomassiliicoccus luminyensis - 3c
SRR7004338 bin6 - Thermoplasmatota / Methanomassiliicoccales - k141 2237998
DMC bin82 - Thermoplasmatota / Methanomassiliicoccales - k141 768055
DMC bin93 - Desulfobacterota / BSN033 - k141 2754227
DMC bin61 - Desulfobacterota / BSN033 - k141 5191897
WP 015617925.1 - Desulfomonas gibsoniae - 3c
WP 015739967.1 - Ammonifex degensii - 3c
WP 012302733.1 - Desulforudis audaxviator - 3c
0.3

Figure S16. Rates of *ex situ* trace gas consumption of soils. Depicted are the oxidation of **(a)** atmospheric H_2, **(b)** atmospheric CO, and **(c)** atmospheric CH_4 by soils at each depth compared to heat-killed controls. For **(d)**, oxidation of the three gases was also measured in the dryland soils following hydration. Error bars represent standard deviations of four biological replicates.

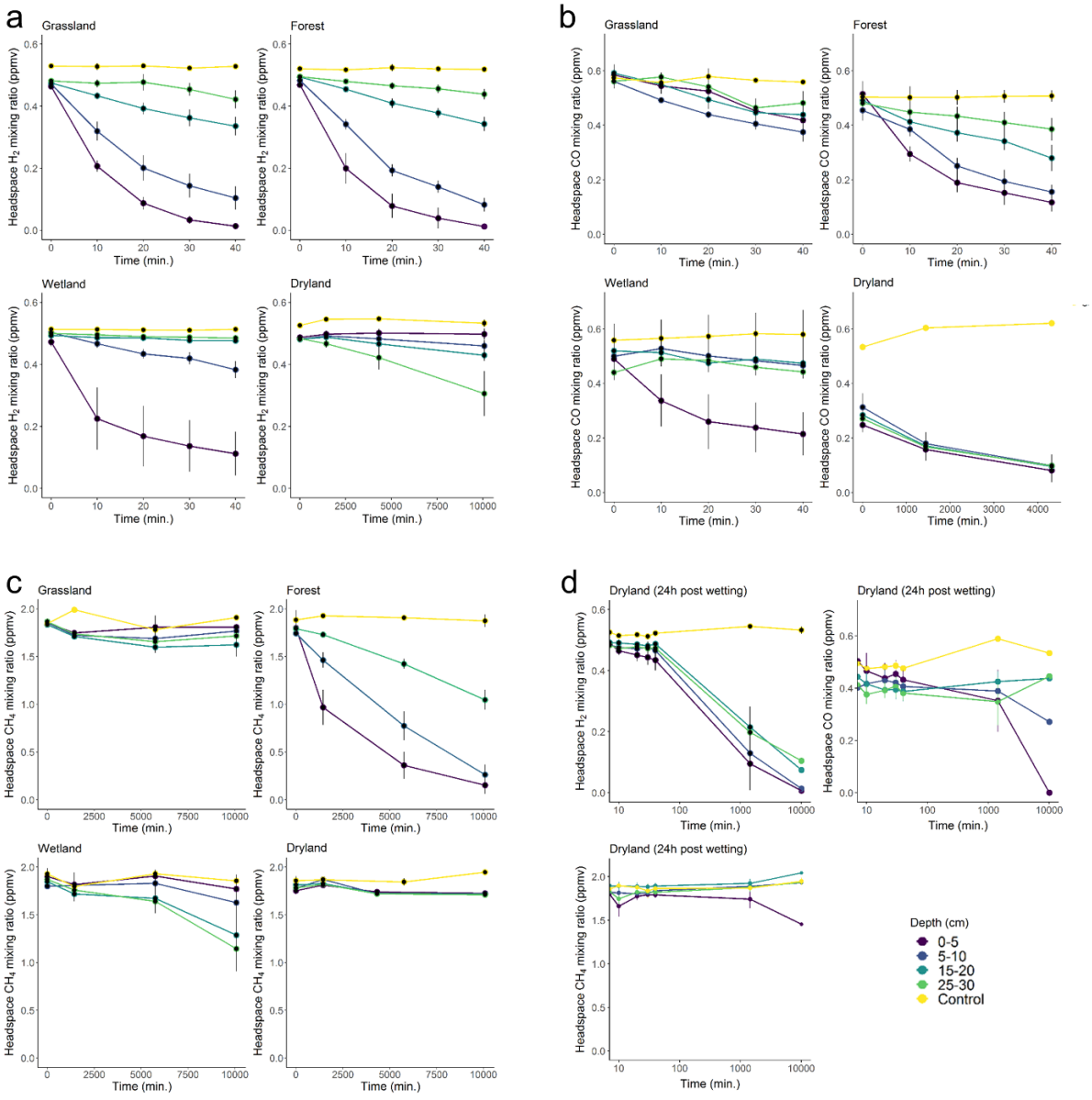


Figure S17. Copy number of the 16S rRNA gene per gram of soil dry weight in the soil samples at each depth.

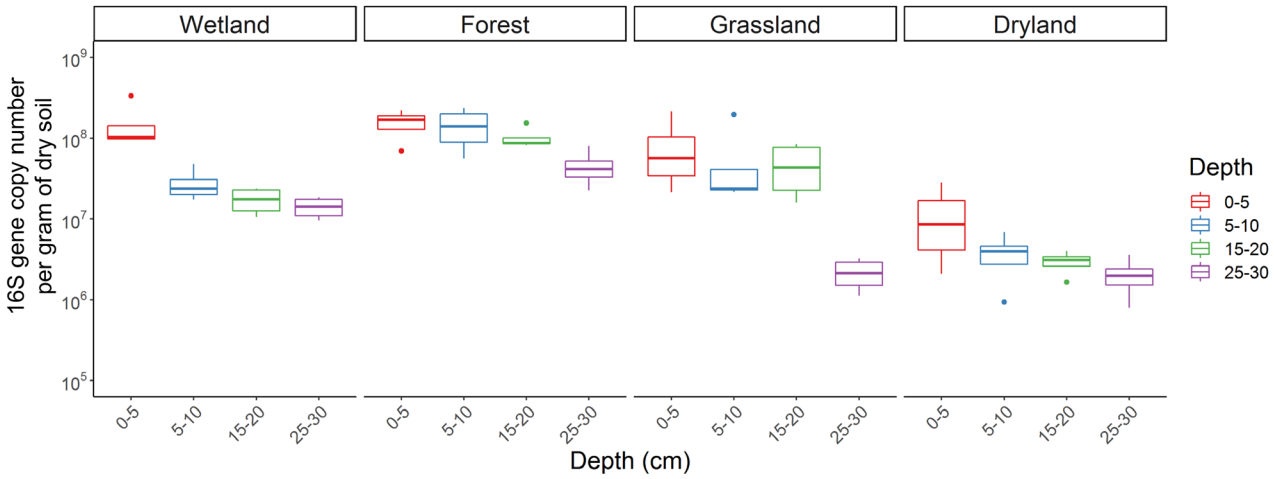
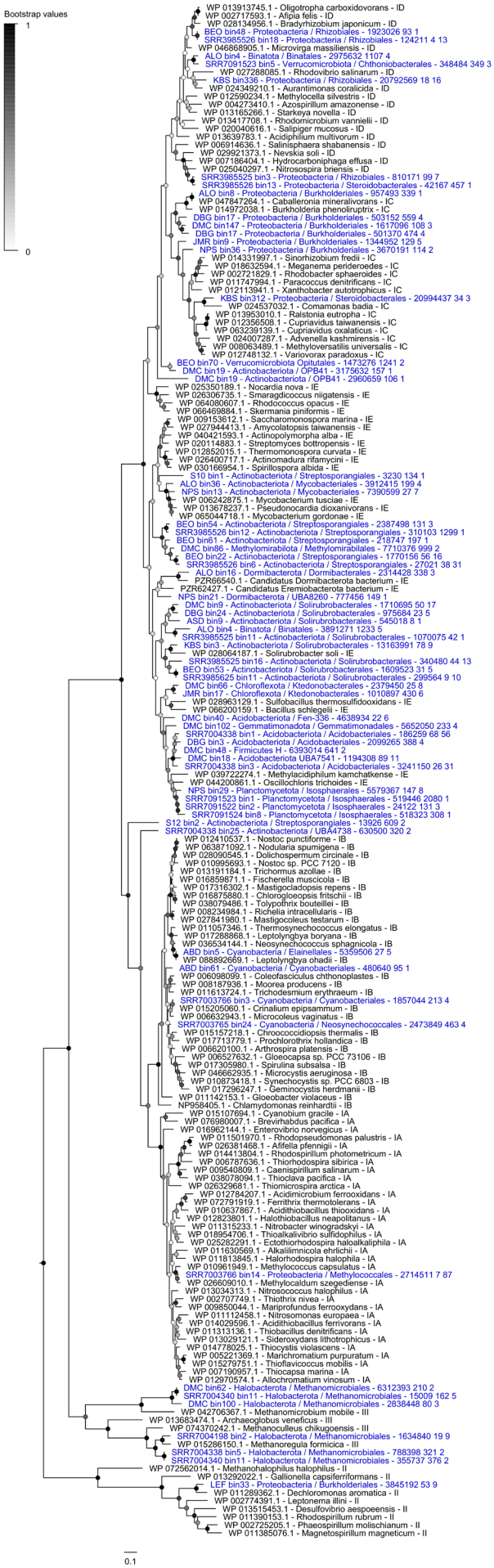


Figure S18. Maximum-likelihood tree of amino acid sequences of ribulose 1,5-bisphosphate carboxylase/oxygenase (RbcL), a marker for carbon fixation through the Calvin-Benson cycle. The tree shows sequences from soil metagenome-assembled genomes (blue) alongside representative reference sequences (black). The subtype of each reference sequence is denoted. The tree was constructed using the JTT matrix-based model, used all sites, and was bootstrapped with 50 replicates and midpoint-rooted.


Bootstrap values
1
0
WP 013913745.1 - Oligotropha carboxidovorans - ID
WP 002717593.1 - Afipia felis - ID
WP 028134956.1 - Bradyrhizobium japonicum - ID
BEO bin48 - Proteobacteria / Rhizobiales - 1923026 93 1
SRR3985526 bin18 - Proteobacteria / Rhizobiales - 124211 4 13
WP 046868905.1 - Microvirga massiliensis - ID
ALO bin4 - Binatota / Binatales - 2975632 1107 4
SRR7091523 bin5 - Verrucomicrobiota / Chthoniobacterales - 348484 349 3
WP 027288085.1 - Rhodovibrio salinarum - ID
KBS bin336 - Proteobacteria / Rhizobiales - 20792569 18 16
WP 024349210.1 - Aurantimonas coralicida - ID
WP 012590234.1 - Methylocella silvestris - ID
WP 004273410.1 - Azospirillum amazonense - ID
WP 013165266.1 - Starkeya novella - ID
WP 013417708.1 - Rhodomicrobium vannielii - ID
WP 020040616.1 - Salipiger mucosus - ID
WP 013639783.1 - Acidiphilium multivorum - ID
WP 006914636.1 - Salinisphaera shabanensis - ID
WP 029921373.1 - Nevskia soli - ID
WP 007186404.1 - Hydrocarboniphaga effusa - ID
WP 025040297.1 - Nitrosospira briensis - ID
SRR3985525 bin3 - Proteobacteria / Rhizobiales - 810171 99 7
SRR3985526 bin13 - Proteobacteria / Steroidobacterales - 42167 457 1
ALO bin8 - Proteobacteria / Burkholderiales - 957493 339 1
WP 047847264.1 - Caballeronia mineralivorans - IC
WP 014972038.1 - Burkholderia phenoliruptrix - IC
DBG bin17 - Proteobacteria / Burkholderiales - 503152 559 4
DMC bin147 - Proteobacteria / Burkholderiales - 1617096 108 3
DBG bin17 - Proteobacteria / Burkholderiales - 501370 474 4
JMR bin9 - Proteobacteria / Burkholderiales - 1344952 129 5
NPS bin36 - Proteobacteria / Burkholderiales - 3670191 114 2
WP 014331997.1 - Sinorhizobium fredii - IC
WP 018632594.1 - Meganema perideroedes - IC
WP 002721829.1 - Rhodobacter sphaeroides - IC
WP 011747994.1 - Paracoccus denitrificans - IC
WP 012113941.1 - Xanthobacter autotrophicus - IC
KBS bin312 - Proteobacteria / Steroidobacterales - 20994437 34 3
WP 024537032.1 - Comamonas badia - IC
WP 013953010.1 - Ralstonia eutropha - IC
WP 012356508.1 - Cupriavidus taiwanensis - IC
WP 063239139.1 - Cupriavidus oxalaticus - IC
WP 024007287.1 - Advenella kashmirensis - IC
WP 008063489.1 - Methyloversatilis universalis - IC
WP 012748132.1 - Variovorax paradoxus - IC
BEO bin70 - Verrucomicrobiota Opitutales - 1473276 1241 2
DMC bin19 - Actinobacteriota / OPB41 - 3175632 157 1
DMC bin19 - Actinobacteriota / OPB41 - 2960659 106 1
WP 025350189.1 - Nocardia nova - IE
WP 026306735.1 - Smaragdicoccus niigatensis - IE
WP 064080607.1 - Rhodococcus opacus - IE
WP 066469884.1 - Skermania piniformis - IE
WP 009153612.1 - Saccharomonospora marina - IE
WP 027944413.1 - Amycolatopsis taiwanensis - IE
WP 040421593.1 - Actinopolymorpha alba - IE
WP 020114883.1 - Streptomyces bottropensis - IE
WP 012852015.1 - Thermomonospora curvata - IE
WP 026400717.1 - Actinomadura rifamycini - IE
WP 030166954.1 - Spirillospora albida - IE
S10 bin1 - Actinobacteriota / Streptosporangiales - 3230 134 1
ALO bin36 - Actinobacteriota / Mycobacteriales - 3912415 199 4
NPS bin13 - Actinobacteriota / Mycobacteriales - 7390599 27 7
WP 006242875.1 - Mycobacterium tusciae - IE
WP 013678237.1 - Pseudonocardia dioxanivorans - IE
WP 065044718.1 - Mycobacterium gordonae - IE
BEO bin54 - Actinobacteriota / Streptosporangiales - 2387498 131 3
SRR3985526 bin12 - Actinobacteriota / Streptosporangiales - 310103 1299 1
BEO bin61 - Actinobacteriota / Streptosporangiales - 218747 197 1
DMC bin86 - Methylomirabilota / Methylomirabilales - 7710376 999 2
BEO bin22 - Actinobacteriota / Streptosporangiales - 1770156 56 16
SRR3985526 bin6 - Actinobacteriota / Streptosporangiales - 27021 38 31
ALO bin16 - Dormibacterota / Dormibacterales - 2314428 338 3
PZR66540.1 - Candidatus Dormibacterota bacterium - IE
PZR62427.1 - Candidatus Eremiobacterota bacterium - IE
NPS bin21 - Dormibacterota / UBA8260 - 777456 149 1
DMC bin9 - Actinobacteriota / Solirubrobacterales - 1710695 50 17
DBG bin24 - Actinobacteriota / Solirubrobacterales - 975684 23 5
ASD bin9 - Actinobacteriota / Solirubrobacterales - 545018 8 1
ALO bin4 - Binatota / Binatales - 3891271 1233 5
SRR3985525 bin11 - Actinobacteriota / Solirubrobacterales - 1070075 42 1
KBS bin3 - Actinobacteriota / Solirubrobacterales - 13163991 78 9
WP 028064187.1 - Solirubrobacter soli - IE
SRR3985525 bin16 - Actinobacteriota / Solirubrobacterales - 340480 44 13
BEO bin53 - Actinobacteriota / Solirubrobacterales - 1609523 31 5
SRR3985625 bin11 - Actinobacteriota / Solirubrobacterales - 299564 9 10
DMC bin66 - Chloroflexota / Ktedonobacterales - 2379450 25 8
JMR bin17 - Chloroflexota / Ktedonobacterales - 1010897 430 6
WP 028963129.1 - Sulfobacillus thermosulfidooxidans - IE
WP 066200159.1 - Bacillus schlegelii - IE
DMC bin40 - Acidobacteriota / Fen-336 - 4638934 22 6
DMC bin102 - Gemmatimonadota / Gemmatimonadales - 5652050 233 4
SRR7004338 bin1 - Acidobacteriota / Acidobacteriales - 186259 68 56
DBG bin3 - Acidobacteriota / Acidobacteriales - 2099265 388 4
DMC bin48 - Firmicutes H - 6393014 641 2
DMC bin18 - Acidobacteriota UBA7541 - 1194308 89 11
SRR7004338 bin3 - Acidobacteriota / Acidobacteriales - 3241150 26 31
WP 039722274.1 - Methylacidiphilum kamchatkense - IE
WP 044200861.1 - Oscillochloris trichoides - IE
NPS bin29 - Planctomycetota / Isosphaerales - 5579367 147 8
SRR7091523 bin1 - Planctomycetota / Isosphaerales - 519446 2080 1
SRR7091522 bin2 - Planctomycetota / Isosphaerales - 24122 131 3
SRR7091524 bin8 - Planctomycetota / Isosphaerales - 518323 308 1
S12 bin2 - Actinobacteriota / Streptosporangiales - 13926 609 2
SRR7004338 bin25 - Actinobacteriota / UBA4738 - 630500 320 2
WP 012410537.1 - Nostoc punctiforme - IB
WP 063871092.1 - Nodularia spumigena - IB
WP 028090545.1 - Dolichospermum circinale - IB
WP 010995693.1 - Nostoc sp. PCC 7120 - IB
WP 013191184.1 - Trichormus azollae - IB
WP 016859871.1 - Fischerella muscicola - IB
WP 017316302.1 - Mastigocladopsis repens - IB
WP 016875880.1 - Chlorogloeopsis fritschii - IB
WP 038079486.1 - Tolypothrix bouteillei - IB
WP 008234984.1 - Richelia intracellularis - IB
WP 027841980.1 - Mastigocoleus testarum - IB
WP 011057346.1 - Thermosynechococcus elongatus - IB
WP 017288868.1 - Leptolyngbya boryana - IB
WP 036534144.1 - Neosynechococcus sphagnicola - IB
ABD bin5 - Cyanobacteria / Elainellales - 5359506 27 5
WP 088892669.1 - Leptolyngbya chadii - IB
ABD bin61 - Cyanobacteria / Cyanobacteriales - 480640 95 1
WP 006098099.1 - Coleofasciculus chthonoplastes - IB
WP 008187936.1 - Moorea producens - IB
WP 011613724.1 - Trichodesmium erythraeum - IB
SRR7003766 bin3 - Cyanobacteria / Cyanobacteriales - 1857044 213 4
WP 015205060.1 - Crinalium epipsammum - IB
WP 006632943.1 - Microcoleus vaginatus - IB
SRR7003765 bin24 - Cyanobacteria / Neosynechococcales - 2473849 463 4
WP 015157218.1 - Chroococcidiopsis thermalis - IB
WP 017713779.1 - Prochlorothrix hollandica - IB
WP 006620100.1 - Arthrospira platensis - IB
WP 006527632.1 - Gloeocapsa sp. PCC 73106 - IB
WP 017305980.1 - Spirulina subsalsa - IB
WP 046662935.1 - Microcystis aeruginosa - IB
WP 010873418.1 - Synechocystis sp. PCC 6803 - IB
WP 017296247.1 - Geminocystis herdmanii - IB
WP 011142153.1 - Gloeobacter violaceus - IB
NP958405.1 - Chlamydomonas reinhardtii - IB
WP 015107694.1 - Cyanobium gracile - IA
WP 076980007.1 - Brevirhabdus pacifica - IA
WP 016962144.1 - Enterovibrio norvegicus - IA
WP 011501970.1 - Rhodopseudomonas palustris - IA
WP 026381468.1 - Afifella pfennigii - IA
WP 014413804.1 - Rhodospirillum photometricum - IA
WP 006787636.1 - Thiorhodospira sibirica - IA
WP 009540809.1 - Caenispirillum salinarum - IA
WP 038078094.1 - Thioclava pacifica - IA
WP 026329681.1 - Thiomicrospira arctica - IA
WP 012784207.1 - Acidimicrobium ferrooxidans - IA
WP 072791919.1 - Ferrithrix thermotolerans - IA
WP 010637867.1 - Acidithiobacillus thiooxidans - IA
WP 012823801.1 - Halothiobacillus neapolitanus - IA
WP 011315233.1 - Nitrobacter winogradskyi - IA
WP 018954706.1 - Thioalkalivibrio sulfidiphilus - IA
WP 025282291.1 - Ectothiorhodospira haloalkaliphila - IA
WP 011630569.1 - Alkalilimnicola ehrlichii - IA
WP 011913845.1 - Halorhodospira halophila - IA
WP 010961949.1 - Methylococcus capsulatus - IA
SRR7003766 bin14 - Proteobacteria / Methylococcales - 2714511 7 87
WP 026609010.1 - Methylocaldum szegediense - IA
WP 013034313.1 - Nitrosococcus halophilus - IA
WP 002707749.1 - Thiothrix nivea - IA
WP 009850044.1 - Mariprofundus ferrooxydans - IA
WP 011112458.1 - Nitrosomonas europaea - IA
WP 014029596.1 - Acidithiobacillus ferrivorans - IA
WP 011313136.1 - Thiobacillus denitrificans - IA
WP 013029121.1 - Sideroxydans lithotrophicus - IA
WP 014778025.1 - Thiocystis violascens - IA
WP 005221369.1 - Marichromatium purpuratum - IA
WP 015279751.1 - Thioflavicoccus mobilis - IA
WP 007190957.1 - Thiocapsa marina - IA
WP 012970574.1 - Allochromatium vinosum - IA
DMC bin62 - Halobacterota / Methanomicrobiales - 6312393 210 2
SRR7004340 bin11 - Halobacterota / Methanomicrobiales - 15009 162 5
DMC bin100 - Halobacterota / Methanomicrobiales - 2838448 80 3
WP 042706367.1 - Methanomicrobium mobile - III
WP 013683474.1 - Archaeoglobus veneficus - III
WP 074370242.1 - Methanoculleus chikugoensis - III
SRR7004198 bin2 - Halobacterota / Methanomicrobiales - 1634840 19 9
WP 015286150.1 - Methanoregula formicica - III
SRR7004338 bin5 - Halobacterota / Methanomicrobiales - 788398 321 2
SRR7004340 bin11 - Halobacterota / Methanomicrobiales - 355737 376 2
WP 072562014.1 - Methanohalophilus halophilus - II
WP 013292022.1 - Gallionella capsiferriformans - II
LEF bin33 - Proteobacteria / Burkholderiales - 3845192 53 9
WP 011289362.1 - Dechloromonas aromatica - II
WP 002774391.1 - Leptonema illini - II
WP 013515453.1 - Desulfovibrio aespoeensis - II
WP 011390153.1 - Rhodospirillum rubrum - II
WP 002725205.1 - Phaeospirillum molischianum - II
WP 011385076.1 - Magnetospirillum magneticum - II
0.1

6.3.2. Supplementary Tables

Table S1 – S10 are available on the Google Drive folder under **Chapter 4:**
https://drive.google.com/open?id=16UnlZtajRSBZ_vSvJzw3Pl_XraDWUNV2

Table S1 (xlsx). Physicochemical properties of the Australian soil sampled. The particle size and nutrient content is shown for soil samples from each depth.

Table S2 (xlsx). Metadata and sequencing statistics on the Australian soil metagenomes sequenced and global soil metagenomes analyzed.

Table S3 (xlsx). Quality statistics and taxonomy information of the 757 metagenome-assembled genomes from the Australian and global metagenomes.

Table S4 (xlsx). Statistical testing of differences in community composition of the samples. PERMANOVA and BETADISPERSION were used to test significant differences of beta diversity between ecosystem types and soil depths.

Table S5 (xlsx). Relative abundance of metabolic marker genes in metagenomic short reads expressed as percentage of total community (%) and reads per kilobase million (RPKM).

Table S6 (xlsx). Summary of metabolic marker genes detected in the metagenome-assembled genomes.

Table S7 (xlsx). Amino acid sequences of hydrogenase, CoxL, PmoA, AmoA, Sqr, FCC, DsrA, SoxB, and RbcL in the binned and unbinned contigs of the metagenomes.

Table S8 (xlsx). Soil gas concentrations, *in situ* gas fluxes, and *ex situ* oxidation rates measured for the four Australian soils.

Table S9 (xlsx). Significance testing of differences in *ex situ* oxidation rates between ecosystem type and soil depth.

Table S10 (xlsx). Thermodynamic modelling of power obtained per cell based on *in situ* and *ex situ* rates of trace gas oxidation.

7. References

1. Horn, R., Taubner, H., Wuttke, M. & Baumgartl, T. Soil physical properties related to soil structure. *Soil Tillage Res.* **30**, 187–216 (1994).
2. Lal, R., Kimble, J. & Follett, R. F. Pedospheric processes and the carbon cycle. in *Soil processes and the carbon cycle* (CRC Press, 2018).
3. Jenny, H. *Factors of soil formation: a system of quantitative pedology*. (Courier Corporation, 1994).
4. Lehmann, J. & Kleber, M. The contentious nature of soil organic matter. *Nature* **528**, 60–68 (2015).
5. Allison, F. E. *Soil organic matter and its role in crop production*. (Elsevier, 1973).
6. Badawy, S. H., Helal, M. I. D., Chaudri, A. M., Lawlor, K. & McGrath, S. P. Soil solid-phase controls lead activity in soil solution. *J. Environ. Qual.* **31**, 162–167 (2002).
7. Kemper, W. D., Rosenau, R. & Nelson, S. Gas displacement and aggregate stability of soils. *Soil Sci. Soc. Am. J.* **49**, 25–28 (1985).
8. Rhoades, J. D., Raats, P. A. C. & Prather, R. J. Effects of liquid-phase electrical conductivity, water content, and surface conductivity on bulk soil electrical conductivity 1. *Soil Sci. Soc. Am. J.* **40**, 651–655 (1976).
9. Cary, S. C., McDonald, I. R., Barrett, J. E. & Cowan, D. A. On the rocks: The microbiology of Antarctic Dry Valley soils. *Nat. Rev. Microbiol.* **8**, 129–138 (2010).
10. Bahram, M. *et al.* Structure and function of the global topsoil microbiome. *Nature* **560**, 233–237 (2018).
11. Tedersoo, L. *et al.* Global diversity and geography of soil fungi. *Science* **346**, 1256688 (2014).
12. Lladó, S., López-Mondéjar, R. & Baldrian, P. Forest Soil Bacteria: Diversity, Involvement in Ecosystem Processes, and Response to Global Change. *Microbiol. Mol. Biol. Rev.* **81**, e00063-16 (2017).
13. Bardgett, R. D. & Van Der Putten, W. H. Belowground biodiversity and ecosystem functioning. *Nature* **515**, 505–511 (2014).
14. Cavicchioli, R. *et al.* Scientists' warning to humanity: microorganisms and climate change. *Nature Reviews Microbiology* **17**, 569–586 (2019).
15. Zhou, J. *et al.* Microbial mediation of carbon-cycle feedbacks to climate warming. *Nat. Clim. Chang.* **2**, 106–110 (2012).
16. Van Der Heijden, M. G. A., Bardgett, R. D. & Van Straalen, N. M. The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. *Ecol. Lett.* **11**, 296–310 (2008).
17. Hayat, R., Ali, S., Amara, U., Khalid, R. & Ahmed, I. Soil beneficial bacteria and their role in plant growth promotion: a review. *Ann. Microbiol.* **60**, 579–598 (2010).

18. Štursová, M., Žifčáková, L., Leigh, M. B., Burgess, R. & Baldrian, P. Cellulose utilization in forest litter and soil: identification of bacterial and fungal decomposers. *FEMS Microbiol. Ecol.* **80**, 735–746 (2012).
19. Greening, C. *et al.* Atmospheric hydrogen scavenging: From enzymes to ecosystems. *Appl. Environ. Microbiol.* **81**, 1190–1199 (2015).
20. Conrad, R. Soil Microorganisms as Controllers of Atmospheric Trace Gases (H2 , CO , CH4 , OCS , N2O , and NO). *Microbiol. Rev.* **60**, 609–640 (1996).
21. Giardina, C. P., Litton, C. M., Crow, S. E. & Asner, G. P. Warming related increases in soil CO2 efflux are explained by increased below-ground carbon flux. *Nat. Clim. Chang.* **4**, 822–827 (2014).
22. Weller, D. M. Biological control of soilborne plant pathogens in the rhizosphere with bacteria. *Annu. Rev. Phytopathol.* **26**, 379–407 (1988).
23. Liu, Q., Tang, J., Gao, K., Gurav, R. & Giesy, J. P. Aerobic degradation of crude oil by microorganisms in soils from four geographic regions of China. *Sci. Rep.* **7**, 1–12 (2017).
24. Saccá, M. L., Caracciolo, A. B., Lenola, M. Di & Grenni, P. *Soil Biological Communities and Ecosystem Resilience*. *Springer International Publishing* (2017).
25. Barrios, E. Soil biota, ecosystem services and land productivity. *Ecol. Econ.* **64**, 269–285 (2007).
26. Bryant, D. A. & Frigaard, N.-U. Prokaryotic photosynthesis and phototrophy illuminated. *Trends Microbiol.* **14**, 488–496 (2006).
27. Daniel, R. The soil metagenome–a rich resource for the discovery of novel natural products. *Curr. Opin. Biotechnol.* **15**, 199–204 (2004).
28. Ji, M. *et al.* Atmospheric trace gases support primary production in Antarctic desert surface soil. *Nature* **552**, 400–403 (2017).
29. Handa, I. T. *et al.* Consequences of biodiversity loss for litter decomposition across biomes. *Nature* **509**, 218–221 (2014).
30. Wang, J. T., Egidi, E., Li, J. & Singh, B. K. Linking microbial diversity with ecosystem functioning through a trait framework. *J. Biosci.* **44**, 1–3 (2019).
31. Delgado-Baquerizo, M. *et al.* Microbial diversity drives multifunctionality in terrestrial ecosystems. *Nat. Commun.* **7**, 1–8 (2016).
32. Laforest-Lapointe, I., Paquette, A., Messier, C. & Kembel, S. W. Leaf bacterial diversity mediates plant diversity and ecosystem function relationships. *Nature* **546**, 145–147 (2017).
33. Bell, T. Experimental tests of the bacterial distance–decay relationship. *ISME J.* **4**, 1357–1365 (2010).
34. Hug, L. A. *et al.* A new view of the tree of life. *Nat. Microbiol.* **1**, 1–6 (2016).
35. Delgado-Baquerizo, M. *et al.* A global atlas of the dominant bacteria found in soil. *Science* **359**, 320–325 (2018).
36. Janssen, P. H. Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S

rRNA genes. *Applied and Environmental Microbiology* **72**, 1719–1728 (2006).

37. Karimi, B. *et al.* Biogeography of soil bacteria and archaea across France. *Sci. Adv.* **4**, (2018).
38. Bull, A. T., Asenjo, J. A., Goodfellow, M. & Gómez-Silva, B. The Atacama Desert: Technical Resources and the Growing Importance of Novel Microbial Diversity. *Annu. Rev. Microbiol.* **70**, 215–234 (2016).
39. Smith, J. J., Tow, L. A., Stafford, W., Cary, C. & Cowan, D. A. Bacterial diversity in three different antarctic cold desert mineral soils. *Microb. Ecol.* **51**, 413–421 (2006).
40. Barka, E. A. *et al.* Taxonomy, physiology, and natural products of Actinobacteria. *Microbiol. Mol. Biol. Rev.* **80**, 1–43 (2016).
41. Makhalanyane, T. P. *et al.* Microbial ecology of hot desert edaphic systems. *FEMS Microbiol. Rev.* **39**, 203–221 (2015).
42. Tourna, M. *et al.* Nitrososphaera viennensis, an ammonia oxidizing archaeon from soil. *Proc. Natl. Acad. Sci.* **108**, 8420–8425 (2011).
43. Leininger, S. *et al.* Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature* **442**, 806–809 (2006).
44. Bar-On, Y. M., Phillips, R. & Milo, R. The biomass distribution on Earth. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 6506–6511 (2018).
45. Griffiths, R. I. *et al.* Mapping and validating predictions of soil bacterial biodiversity using European and national scale datasets. *Appl. Soil Ecol.* **97**, 61–68 (2016).
46. Jousset, A. *et al.* Where less may be more: how the rare biosphere pulls ecosystems strings. *ISME J.* **11**, 853–862 (2017).
47. Liu, L., Yang, J., Yu, Z. & Wilkinson, D. M. The biogeography of abundant and rare bacterioplankton in the lakes and reservoirs of China. *ISME J.* **9**, 2068–2077 (2015).
48. Lynch, M. D. J. & Neufeld, J. D. Ecology and exploration of the rare biosphere. *Nat. Rev. Microbiol.* **13**, 217–229 (2015).
49. Schuster, S. C. Next-generation sequencing transforms today's biology. *Nat. Methods* **5**, 16–18 (2008).
50. Petti, C. A., Polage, C. R. & Schreckenberger, P. The role of 16S rRNA gene sequencing in identification of microorganisms misidentified by conventional methods. *J. Clin. Microbiol.* **43**, 6123–6125 (2005).
51. Janda, J. M. & Abbott, S. L. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *J. Clin. Microbiol.* **45**, 2761–2764 (2007).
52. Böttger, E. C. Rapid determination of bacterial ribosomal RNA sequences by direct sequencing of enzymatically amplified DNA. *FEMS Microbiol. Lett.* **65**, 171–176 (1989).
53. Nearing, J. T., Douglas, G. M., Comeau, A. M. & Langille, M. G. I. Denoising the denoisers: an independent evaluation of microbiome sequence error-correction approaches. *PeerJ* **6**, e5364 (2018).

54. Edgar, R. C. Updating the 97% identity threshold for 16S ribosomal RNA OTUs. *Bioinformatics* **34**, 2371–2375 (2018).
55. Amir, A. *et al.* Deblur rapidly resolves single-nucleotide community sequence patterns. *Am. Soc. Microbiol.* **2**, 1–7 (2017).
56. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
57. Philip, H., Tyson, G. W., Hugenholtz, P. & Tyson, G. W. Metagenomics. *Nature* **455**, 481–483 (2008).
58. Tyson, G. W. *et al.* Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**, 37–43 (2004).
59. Venter, J. C. *et al.* Environmental Genome Shotgun Sequencing of the Sargasso Sea. *Science* **304**, 66–74 (2004).
60. Daniel, R. The metagenomics of soil. *Nat. Rev. Microbiol.* **3**, 470–478 (2005).
61. Shi, Y., Tyson, G. W. & DeLong, E. F. Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column. *Nature* **459**, 266–269 (2009).
62. Greening, C., Grinter, R. & Chiri, E. Uncovering the Metabolic Strategies of the Dormant Microbial Majority: towards Integrative Approaches. *mSystems* **4**, e00107-19 (2019).
63. Nielsen, E. S. The use of radio-active carbon (C^) for measuring organic production in the sea. *J. Cons.* **18**, 117–140 (1952).
64. Morris, I., Yentsch, C. M. & Yentsch, C. S. Relationship Between Light Carbon Dioxide Fixation and Dark Carbon Dioxide Fixation By Marine Algae. *Limnol. Oceanogr.* **16**, 854–858 (1971).
65. Yuan, H., Ge, T., Chen, C., O'Donnell, A. G. & Wu, J. Significant role for microbial autotrophy in the sequestration of soil carbon. *Appl. Environ. Microbiol.* **78**, 2328–2336 (2012).
66. Kanai, R., Miyachi, S. & Takamiya, A. Knall-gas reaction-linked fixation of labelled carbon dioxide in an autotrophic streptomyces. *Nature* **188**, 873–875 (1960).
67. Ji, M. *et al.* Atmospheric trace gases support primary production in Antarctic desert surface soil. *Nature* **552**, 400–403 (2017).
68. Constant, P., Chowdhury, S. P., Hesse, L. & Conrad, R. Co-localization of atmospheric H2 oxidation activity and high affinity H2-oxidizing bacteria in non-axenic soil and sterile soil amended with Streptomyces sp. PCB7. *Soil Biol. Biochem.* **43**, 1888–1893 (2011).
69. Greening, C., Villas-Bôas, S. G., Robson, J. R., Berney, M. & Cook, G. M. The growth and survival of Mycobacterium smegmatis is enhanced by co-metabolism of atmospheric H2. *PLoS One* **9**, e103034 (2014).
70. Greening, C. *et al.* Persistence of the dominant soil phylum *Acidobacteria* by trace gas scavenging. *Proc. Natl. Acad. Sci.* **112**, 10497–10502 (2015).
71. Cordero, P. R. F. *et al.* Atmospheric carbon monoxide oxidation is a widespread mechanism

supporting microbial survival. *ISME J.* **13**, 2868–2881 (2019).
72. Islam, Z. F. *et al.* Two Chloroflexi classes independently evolved the ability to persist on atmospheric hydrogen and carbon monoxide. *ISME J.* **13**, 1801–1813 (2019).
73. Conrad, R. & Seiler, W. Decomposition of atmospheric hydrogen by soil microorganisms and soil enzymes. *Soil Biol. Biochem.* **13**, 43–49 (1981).
74. Conrad, R. & Seiler, W. Contribution of hydrogen production by biological nitrogen fixation to the global hydrogen budget. *J. Geophys. Res. Ocean.* **85**, 5493–5498 (1980).
75. Whalen, S. C. & Reeburgh, W. S. A methane flux time series for tundra environments. *Global Biogeochem. Cycles* **2**, 399–409 (1988).
76. Tveit, A. T. *et al.* Widespread soil bacterium that oxidizes atmospheric methane. *Proc. Natl. Acad. Sci.* **116**, 8515–8524 (2019).
77. Kits, K. D. *et al.* Kinetic analysis of a complete nitrifier reveals an oligotrophic lifestyle. *Nature* **549**, 269–272 (2017).
78. Imachi, H. *et al.* Isolation of an archaeon at the prokaryote–eukaryote interface. *Nature* **577**, 519–525 (2020).
79. MacArthur, R. H. & Wilson, E. O. *Island biogeography*. (Princeton, 1967).
80. Tobler, W. R. A Computer Movie Simulating Urban Growth in the Detroit Region. *Econ. Geogr.* **46**, 234 (1970).
81. Hanson, C. A., Fuhrman, J. A., Horner-Devine, M. C. & Martiny, J. B. H. Beyond biogeographic patterns: Processes shaping the microbial landscape. *Nat. Rev. Microbiol.* **10**, 497–506 (2012).
82. Woodcock, S., Curtis, T. P., Head, I. M., Lunn, M. & Sloan, W. T. Taxa-area relationships for microbes: the unsampled and the unseen. *Ecol. Lett.* **9**, 805–812 (2006).
83. Scheiner, S. M. Six types of species-area curves. *Glob. Ecol. Biogeogr.* **12**, 441–447 (2003).
84. Ramirez, K. S. *et al.* Detecting macroecological patterns in bacterial communities across independent studies of global soils. *Nat. Microbiol.* **3**, 189–196 (2018).
85. van der Gast, C. J. Microbial biogeography: The end of the ubiquitous dispersal hypothesis? *Environmental Microbiology* **17**, 544–546 (2015).
86. Zhou, J. & Ning, D. Stochastic community assembly: does it matter in microbial ecology? *Microbiol. Mol. Biol. Rev.* **81**, e00002-17 (2017).
87. Green, J. L., Bohannan, B. J. M. & Whitaker, R. J. Microbial biogeography: From taxonomy to traits. *Science* **320**, 1039–1043 (2008).
88. Martiny, J. B. H. *et al.* Microbial biogeography: Putting microorganisms on the map. *Nat. Rev. Microbiol.* **4**, 102–112 (2006).
89. Andam, C. P. *et al.* A latitudinal diversity gradient in terrestrial bacteria of the genus Streptomyces. *MBio* **7**, 1–9 (2016).
90. Lawton, J. H. Are There General Laws in Ecology? *Oikos* **84**, 177 (1999).
91. Hairston, N. G., Smith, F. E. & Slobodkin, L. B. Community structure, population control, and

competition. *Am. Nat.* **94**, 421–425 (1960).
92. Hutchinson, G. E. Cold spring harbor symposium on quantitative biology. in *Concluding remarks* **22**, 415–427 (1957).
93. Vandermeer, J. H. Niche theory. *Annu. Rev. Ecol. Syst.* **3**, 107–132 (1972).
94. Peterson, A. T. Predicting species' geographic distributions based on ecological niche modeling. *Condor* **103**, 599–605 (2001).
95. Vellend, M. Conceptual Synthesis in Community Ecology. *Q. Rev. Biol.* **85**, 183–206 (2010).
96. Hubbell, S. P. *The unified neutral theory of biodiversity and biogeography (MPB-32)*. (Princeton University Press, 2001).
97. Matthews, T. J. & Whittaker, R. J. Neutral theory and the species abundance distribution: Recent developments and prospects for unifying niche and neutral perspectives. *Ecol. Evol.* **4**, 2263–2277 (2014).
98. Stegen, J. C., Lin, X., Konopka, A. E. & Fredrickson, J. K. Stochastic and deterministic assembly processes in subsurface microbial communities. *ISME J.* **6**, 1653–1664 (2012).
99. Dumbrell, A. J., Nelson, M., Helgason, T., Dytham, C. & Fitter, A. H. Relative roles of niche and neutral processes in structuring a soil microbial community. *ISME J.* **4**, 337–345 (2010).
100. Dini-Andreote, F., Stegen, J. C., van Elsas, J. D. & Salles, J. F. Disentangling mechanisms that mediate the balance between stochastic and deterministic processes in microbial succession. *Proc. Natl. Acad. Sci.* **112**, E1326–E1332 (2015).
101. Chase, J. M. Stochastic community assembly causes higher biodiversity in more productive environments. *Science* **328**, 1388–1391 (2010).
102. Caruso, T. *et al.* Stochastic and deterministic processes interact in the assembly of desert microbial communities on a global scale. *ISME J.* **5**, 1406–1413 (2011).
103. Powell, J. R. *et al.* Deterministic processes vary during community assembly for ecologically dissimilar taxa. *Nat. Commun.* **6**, 1–10 (2015).
104. Nemergut, D. R. *et al.* Patterns and processes of microbial community assembly. *Microbiol. Mol. Biol. Rev.* **77**, 342–356 (2013).
105. Stegen, J. C., Lin, X., Fredrickson, J. K. & Konopka, A. E. Estimating and mapping ecological processes influencing microbial community assembly. *Front. Microbiol.* **6**, 1–15 (2015).
106. Fierer, N. Embracing the unknown: Disentangling the complexities of the soil microbiome. *Nature Reviews Microbiology* **15**, 579–590 (2017).
107. Neilson, J. W. *et al.* Significant impacts of increasing aridity on the arid soil microbiome. *mSystems* **2**, e00195-16 (2017).
108. Andrew, D. R. *et al.* Abiotic factors shape microbial diversity in Sonoran desert soils. *Appl. Environ. Microbiol.* **78**, 7527–7537 (2012).
109. Van Horn, D. J. *et al.* Factors Controlling Soil Microbial Biomass and Bacterial Diversity and Community Composition in a Cold Desert Ecosystem: Role of Geographic Scale. *PLoS One*

8, (2013).

110. Fierer, N., Strickland, M. S., Liptzin, D., Bradford, M. A. & Cleveland, C. C. Global patterns in belowground communities. *Ecol. Lett.* **12**, 1238–1249 (2009).
111. Jiao, S. & Lu, Y. Soil pH and temperature regulate assembly processes of abundant and rare bacterial communities in agricultural ecosystems. *Environ. Microbiol.* **22**, 1052–1065 (2019).
112. Chase, J. M. & Myers, J. A. Disentangling the importance of ecological niches from stochastic processes across scales. *Philos. Trans. R. Soc. B Biol. Sci.* **366**, 2351–2363 (2011).
113. Zhou, J. *et al.* Stochasticity, succession, and environmental perturbations in a fluidic ecosystem. *Proc. Natl. Acad. Sci.* **111**, E836–E845 (2014).
114. Zhou, J. *et al.* Stochastic assembly leads to alternative communities with distinct functions in a bioreactor microbial community. *MBio* **4**, e00584-12 (2013).
115. Jia, X., Dini-Andreote, F. & Salles, J. F. Community assembly processes of the microbial rare biosphere. *Trends Microbiol.* **26**, 738–747 (2018).
116. Hamilton, W. D. & May, R. M. Dispersal in stable habitats. *Nature* **269**, 578–581 (1977).
117. Dechesne, A., Wang, G., Gülez, G., Or, D. & Smets, B. F. Hydration-controlled bacterial motility and dispersal on surfaces. *Proc. Natl. Acad. Sci.* **107**, 14369–14372 (2010).
118. Herbold, C. W., Lee, C. K., McDonald, I. R. & Cary, S. C. Evidence of global-scale aeolian dispersal and endemism in isolated geothermal microbial communities of Antarctica. *Nat. Commun.* **5**, 1–10 (2014).
119. Grossart, H.-P., Dziallas, C., Leunert, F. & Tang, K. W. Bacteria dispersal by hitchhiking on zooplankton. *Proc. Natl. Acad. Sci.* **107**, 11959–11964 (2010).
120. Urban, M. C. & De Meester, L. Community monopolization: local adaptation enhances priority effects in an evolving metacommunity. *Proc. R. Soc. B Biol. Sci.* **276**, 4129–4138 (2009).
121. Chanal, A. *et al.* The desert of Tataouine: an extreme environment that hosts a wide diversity of microorganisms and radiotolerant bacteria. *Environ. Microbiol.* **8**, 514–525 (2006).
122. Lowe, W. H. & McPeek, M. A. Is dispersal neutral? *Trends Ecol. Evol.* **29**, 444–450 (2014).
123. Morlon, H. Phylogenetic approaches for studying diversification. *Ecol. Lett.* **17**, 508–525 (2014).
124. Louca, S. *et al.* Bacterial diversification through geological time. *Nat. Ecol. Evol.* **2**, 1458–1467 (2018).
125. Chubukov, V., Gerosa, L., Kochanowski, K. & Sauer, U. Coordination of microbial metabolism. *Nat. Rev. Microbiol.* **12**, 327–340 (2014).
126. Madigan, M. T., Martinko, J. M. & Parker, J. *Brock biology of microorganisms*. **11**, (Prentice hall, 1997).

127. Nicholls, D. G. *Bioenergetics*. (Academic Press, 2013).
128. Galston, A. W. Photosynthesis as a basis for life support on Earth and in space. *Bioscience* **42**, 490–493 (1992).
129. Green, B. R. & Durnford, D. G. The chlorophyll-carotenoid proteins of oxygenic photosynthesis. *Annu. Rev. Plant Biol.* **47**, 685–714 (1996).
130. Blankenship, R. E. & Hartman, H. The origin and evolution of oxygenic photosynthesis. *Trends Biochem. Sci.* **23**, 94–97 (1998).
131. Summons, R. E., Jahnke, L. L., Hope, J. M. & Logan, G. A. 2-Methylhopanoids as biomarkers for cyanobacterial oxygenic photosynthesis. *Nature* **400**, 554–557 (1999).
132. Fischer, W. W., Hemp, J. & Valentine, J. S. How did life survive Earth's great oxygenation? *Curr. Opin. Chem. Biol.* **31**, 166–178 (2016).
133. McFadden, G. I. Endosymbiosis and evolution of the plant cell. *Curr. Opin. Plant Biol.* **2**, 513–519 (1999).
134. Vikram, S. *et al.* Metagenomic analysis provides insights into functional capacity in a hyperarid desert soil niche community. *Environ. Microbiol.* **18**, 1875-1888. (2016).
135. Niederberger, T. D. *et al.* Carbon-fixation rates and associated microbial communities residing in arid and ephemerally wet antarctic dry valley soils. *Front. Microbiol.* **6**, 1–9 (2015).
136. Lee, K. C. *et al.* Niche filtering of bacteria in soil and rock habitats of the Colorado Plateau Desert, Utah, USA. *Front. Microbiol.* **7**, 1–7 (2016).
137. Rasuk, M. C. *et al.* Bacterial Diversity in Microbial Mats and Sediments from the Atacama Desert. *Microb. Ecol.* **71**, 44–56 (2016).
138. Valverde, A., Makhalanyane, T. P., Seely, M. & Cowan, D. A. Cyanobacteria drive community composition and functionality in rock-soil interface communities. *Mol. Ecol.* **24**, 812–821 (2015).
139. Bryant, D. A. *et al.* Comparative and functional genomics of anoxygenic green bacteria from the taxa Chlorobi, Chloroflexi, and Acidobacteria. in *Functional genomics and evolution of photosynthetic systems* 47–102 (Springer, 2012).
140. Imhoff, J. F., Rahn, T., Künzel, S. & Neulinger, S. C. Photosynthesis is widely distributed among Proteobacteria as demonstrated by the phylogeny of PufLM reaction center proteins. *Front. Microbiol.* **8**, 2679 (2018).
141. Lanyi, J. K. Bacteriorhodopsin. *Annu. Rev. Physiol.* **66**, 665–688 (2004).
142. Gómez-Consarnau, L. *et al.* Proteorhodopsin phototrophy promotes survival of marine bacteria during starvation. *PLoS Biol.* **8**, e1000358 (2010).
143. Steindler, L., Schwalbach, M. S., Smith, D. P., Chan, F. & Giovannoni, S. J. Energy starved Candidatus Pelagibacter ubique substitutes light-mediated ATP production for endogenous carbon respiration. *PLoS One* **6**, e19725 (2011).
144. Finkel, O. M., Béja, O. & Belkin, S. Global abundance of microbial rhodopsins. *ISME J.* **7**,

448–451 (2013).
145. Guerrero, L. D., Vikram, S., Makhalanyane, T. P. & Cowan, D. A. Evidence of microbial rhodopsins in A ntarctic D ry V alley edaphic systems. *Environ. Microbiol.* **19**, 3755–3767 (2017).
146. Kelly, D. P. Autotrophy: concepts of lithotrophic bacteria and their organic metabolism. *Annu. Rev. Microbiol.* **25**, 177–210 (1971).
147. Starr, M. P., Stolp, H., Trüper, H. G., Balows, A. & Schlegel, H. G. *The prokaryotes: a handbook on habitats, isolation and identification of bacteria*. (Springer Science & Business Media, 2013).
148. Shively, J. M., van Keulen, G. & Meijer, W. G. SOMETHING FROM ALMOST NOTHING: Carbon Dioxide Fixation in Chemoautotrophs. *Annu. Rev. Microbiol.* **52**, 191–230 (1998).
149. Tolli, J. & King, G. M. Diversity and structure of bacterial chemolithotrophic communities in pine forest and agroecosystem soils. *Appl. Environ. Microbiol.* **71**, 8411–8418 (2005).
150. Uzgent, G., Nanba, K., King, G. M. & Dunfield, K. Analysis of Facultative Lithotroph Distribution and Diversity on Volcanic Deposits by Use of the Large Subunit of Analysis of Facultative Lithotroph Distribution and Diversity on Volcanic Deposits by Use of the Large Subunit of Ribulose. *Appl. Environ. Microbiol.* **70**, 2245–2253 (2004).
151. Ellis, R. J. The most abundant protein in the world. *Trends Biochem. Sci.* **4**, 241–244 (1979).
152. Purkhold, U. *et al.* Phylogeny of all recognized species of ammonia oxidizers based on comparative 16S rRNA and amoA sequence analysis: implications for molecular diversity surveys. *Appl. Environ. Microbiol.* **66**, 5368–5382 (2000).
153. Purkhold, U., Wagner, M., Timmermann, G., Pommerening-Röser, A. & Koops, H.-P. 16S rRNA and amoA-based phylogeny of 12 novel betaproteobacterial ammonia-oxidizing isolates: extension of the dataset and proposal of a new lineage within the nitrosomonads. *Int. J. Syst. Evol. Microbiol.* **53**, 1485–1494 (2003).
154. Arp, D. J., Sayavedra-Soto, L. A. & Hommes, N. G. Molecular biology and biochemistry of ammonia oxidation by Nitrosomonas europaea. *Arch. Microbiol.* **178**, 250–255 (2002).
155. Suzuki, I., Dular, U. & Kwok, S. C. Ammonia or ammonium ion as substrate for oxidation by Nitrosomonas europaea cells and extracts. *J. Bacteriol.* **120**, 556–558 (1974).
156. Chain, P. *et al.* Complete genome sequence of the ammonia-oxidizing bacterium and obligate chemolithoautotroph Nitrosomonas europaea. *J. Bacteriol.* **185**, 2759–2773 (2003).
157. Rawlings, D. E. The molecular genetics of Thiobacillus ferrooxidans and other mesophilic, acidophilic, chemolithotrophic, iron-or sulfur-oxidizing bacteria. *Hydrometallurgy* **59**, 187–201 (2001).
158. Wakao, N., Hanada, K., Takahashi, A., Sakurai, Y. & Shiota, H. Morphological, Physiological, and chemotaxonomical characteristics of iron-and sulfur-oxidizing bacteria isolated from acid mine drainage waters. *J. Gen. Appl. Microbiol.* **37**, 35–48 (1991).
159. Barreto, M., Jedlicki, E. & Holmes, D. S. Identification of a gene cluster for the formation of

extracellular polysaccharide precursors in the chemolithoautotroph Acidithiobacillus ferrooxidans. *Appl. Environ. Microbiol.* **71**, 2902–2909 (2005).

160. Schwartz, E., Fritsch, J. & Friedrich, B. *H2-metabolizing prokaryotes*. (Springer Berlin Heidelberg, 2013).
161. Cordero, P. R. F. *et al.* Atmospheric carbon monoxide oxidation is a widespread mechanism supporting microbial survival. *ISME J.* **13**, 2868–2881 (2019).
162. Greening, C. *et al.* Genomic and metagenomic surveys of hydrogenase distribution indicate H_2 is a widely utilised energy source for microbial growth and survival. *ISME J.* **10**, 761–777 (2016).
163. Greening, C., Berney, M., Hards, K., Cook, G. M. & Conrad, R. A soil actinobacterium scavenges atmospheric H2 using two membrane-associated, oxygen-dependent [NiFe] hydrogenases. *Proc. Natl. Acad. Sci.* **111**, 4257–4261 (2014).
164. Novelli, P. C. Molecular hydrogen in the troposphere: Global distribution and budget. *J. Geophys. Res. Atmos.* **104**, 30427–30444 (1999).
165. Novelli, P. C., Masarie, K. A., Tans, P. P. & Lang, P. M. Recent changes in atmospheric carbon monoxide. *Science* **263**, 1587–1590 (1994).
166. Conrad, R. The global methane cycle: recent advances in understanding the microbial processes involved. *Environ. Microbiol. Rep.* **1**, 285–292 (2009).
167. Schlegel, H. G. Production, modification, and consumption of atmospheric trace gases by microorganisms. *Tellus* **26**, 11–20 (1974).
168. Ehhalt, D. H. & Rohrer, F. The tropospheric cycle of H2: A critical review. *Tellus, Series B: Chemical and Physical Meteorology* **61**, 500–535 (2009).
169. Prinn, R. G. The cleansing capacity of the atmosphere. *Annu. Rev. Environ. Resour.* **28**, 29–57 (2003).
170. Conrad, R. & Seiler, W. Role of microorganisms in the consumption and production of atmospheric carbon monoxide by soil. *Appl. Environ. Microbiol.* **40**, 437–445 (1980).
171. Conrad, R., Meyer, O. & Seiler, W. Role of carboxydobacteria in consumption of atmospheric carbon monoxide by soil. *Appl. Environ. Microbiol.* **42**, 211–215 (1981).
172. Peters, J. W. *et al.* [FeFe]- and [NiFe]-hydrogenase diversity, mechanism, and maturation. *Biochim. Biophys. Acta - Mol. Cell Res.* **1853**, 1350–1369 (2015).
173. Conrad, R. Soil microorganisms oxidizing atmospheric trace gases (CH4, CO, H2, NO). *Indian Journal of Microbiology* **39**, 193–203 (1999).
174. Conrad, R., Aragno, M. & Seiler, W. The inability of hydrogen bacteria to utilize atmospheric hydrogen is due to threshold and affinity for hydrogen. *FEMS Microbiol. Lett.* **18**, 207–210 (1983).
175. Constant, P., Poissant, L. & Villemur, R. Isolation of Streptomyces sp. PCB7, the first microorganism demonstrating high-affinity uptake of tropospheric H2. *ISME J.* **2**, 1066–1076 (2008).

176. Morita, R. Y. Is H_2 the universal energy source for long-term survival? *Microb. Ecol.* **38**, 307–320 (1999).
177. Schuler, S. & Conrad, R. Soils contain two different activities for oxidation of hydrogen. *FEMS Microbiol. Ecol.* **6**, 77–83 (1990).
178. Häring, V. & Conrad, R. Demonstration of two different H 2-oxidizing activities in soil using an H 2 consumption and a tritium exchange assay. *Biol. Fertil. soils* **17**, 125–128 (1994).
179. Fritsch, J. *et al.* The crystal structure of an oxygen-tolerant hydrogenase uncovers a novel iron-sulphur centre. *Nature* **479**, 249–52 (2011).
180. Myers, M. R. & King, G. M. Isolation and characterization of Acidobacterium ailaaui sp. Nov., a novel member of Acidobacteria subdivision 1, from a geothermally heated Hawaiian microbial mat. *Int. J. Syst. Evol. Microbiol.* **66**, 5328–5335 (2016).
181. Sondergaard, D., Pedersen, C. N. S. & Greening, C. HydDB: A web tool for hydrogenase classification and analysis. *Sci. Rep.* **6**, 1–8 (2016).
182. Greening, C. *et al.* Genomic and metagenomic surveys of hydrogenase distribution indicate H 2 is a widely utilised energy source for microbial growth and survival. *ISME J.* **10**, 761–777 (2016).
183. Berney, M., Greening, C., Hards, K., Collins, D. & Cook, G. M. Three different [NiFe] hydrogenases confer metabolic flexibility in the obligate aerobe M ycobacterium smegmatis. *Environ. Microbiol.* **16**, 318–330 (2014).
184. Heinrich, D., Raberg, M. & Steinbüchel, A. Studies on the aerobic utilization of synthesis gas (syngas) by wild type and recombinant strains of Ralstonia eutropha H16. *Microb. Biotechnol.* **11**, 647–656 (2018).
185. Schäfer, C. *et al.* Structure of an Actinobacterial-Type [NiFe]-Hydrogenase Reveals Insight into O2-Tolerant H2 Oxidation. *Structure* **24**, 285–292 (2016).
186. Sato, Y. *et al.* Occurrence of hydrogen-oxidizing Ralstonia species as primary microorganisms in the Mt. Pinatubo volcanic mudflow deposits. *Soil Sci. plant Nutr.* **50**, 855–861 (2004).
187. King, G. M. Characteristics and significance of atmospheric carbon monoxide consumption by soils. *Chemosph. - Glob. Chang. Sci.* **1**, 53–63 (1999).
188. Meyer, O. & Schlegel, H. G. Biology of aerobic carbon monoxide oxidising bacteria. 277–310 (1983).
189. Hille, R. Molybdenum-containing hydroxylases. *Arch. Biochem. Biophys.* **433**, 107–116 (2005).
190. King Gary, M. Uptake of carbon monoxide and hydrogen at environmentally relevant concentrations by mycobacteria. *Appl. Environ. Microbiol.* **69**, 7266–7272 (2003).
191. King Gary, M. Molecular and culture-based analyses of aerobic carbon monoxide oxidizer diversity. *Appl. Environ. Microbiol.* **69**, 7257–7265 (2003).
192. Dobbek, H., Gremer, L., Kiefersauer, R., Huber, R. & Meyer, O. Catalysis at a dinuclear

[CuSMo (O) OH] cluster in a CO dehydrogenase resolved at 1.1-Å resolution. *Proc. Natl. Acad. Sci.* **99**, 15971–15976 (2002).

193. Dobbek, H., Gremer, L., Meyer, O. & Huber, R. Crystal structure and mechanism of CO dehydrogenase, a molybdo iron-sulfur flavoprotein containing S-selanylcysteine. *Proc. Natl. Acad. Sci.* **96**, 8884–8889 (1999).
194. Kraut, M. & Meyer, O. Plasmids in carboxydotrophic bacteria: physical and restriction analysis. *Arch. Microbiol.* **149**, 540–546 (1988).
195. Gadkari, D., Schricker, K., Acker, G., Kroppenstedt, R. M. & Meyer, O. Streptomyces thermoautotrophicus sp. nov., a thermophilic CO and H2 oxidizing obligate chemolithoautotroph. *Appl. Environ. Microbiol.* **56**, 3727–3734 (1990).
196. King, C. E. & King, G. M. Description of Thermogemmatispora carboxidivorans sp. nov., a carbon-monoxide-oxidizing member of the class Ktedonobacteria isolated from a geothermally heated biofilm, and analysis of carbon monoxide oxidation by members of the class Ktedonobacteria. *Int. J. Syst. Evol. Microbiol.* **64**, 1244–1251 (2014).
197. Weber, C. F. & King, G. M. Volcanic soils as sources of novel CO-oxidizing Paraburkholderia and Burkholderia: Paraburkholderia hiiakae sp. nov., Paraburkholderia metrosideri sp. nov., Paraburkholderia paradisi sp. nov., Paraburkholderia peleae sp. nov., and Burkholderia alpina sp. . *Front. Microbiol.* **8**, 207 (2017).
198. McDuff, S., King, G. M., Neupane, S. & Myers, M. R. Isolation and characterization of extremely halophilic CO-oxidizing Euryarchaeota from hypersaline cinders, sediments and soils and description of a novel CO oxidizer, Haloferax namakaokahaiae Mke2. 3T, sp. nov. *FEMS Microbiol. Ecol.* **92**, fiw028 (2016).
199. Knief, C. Diversity and habitat preferences of cultivated and uncultivated aerobic methanotrophic bacteria evaluated based on pmoA as molecular marker. *Front. Microbiol.* **6**, 1346 (2015).
200. Murrell, J. C., McDonald, I. R. & Bourne, D. G. Molecular methods for the study of methanotroph ecology. *FEMS Microbiol. Ecol.* **27**, 103–114 (1998).
201. van Teeseling, M. C. F. *et al.* Expanding the verrucomicrobial methanotrophic world: description of three novel species of Methylacidimicrobium gen. nov. *Appl. Environ. Microbiol.* **80**, 6782–6791 (2014).
202. Schmitz, R. A. *et al.* The thermoacidophilic methanotroph Methylacidiphilum fumariolicum SolV oxidizes subatmospheric H 2 with a high-affinity, membrane-associated [NiFe] hydrogenase. *ISME J.* 1–10 (2020).
203. Bowman, J. The methanotrophs—the families Methylococcaceae and Methylocystaceae. *Prokaryotes Prokaryotic Physiol. Biochem.* 266–289 (2006).
204. Ettwig, K. F. *et al.* Nitrite-driven anaerobic methane oxidation by oxygenic bacteria. *Nature* **464**, 543–548 (2010).
205. Ross, M. O. & Rosenzweig, A. C. A tale of two methane monooxygenases. *JBIC J. Biol.*

Inorg. Chem. **22**, 307–319 (2017).
206. Heyer, J., Galchenko, V. F. & Dunfield, P. F. Molecular phylogeny of type II methane-oxidizing bacteria isolated from various environments. *Microbiology* **148**, 2831–2846 (2002).
207. McGenity, T. J., Crombie, A. T. & Murrell, J. C. Microbial cycling of isoprene, the most abundantly produced biological volatile organic compound on Earth. *ISME J.* **12**, 931–941 (2018).
208. Crombie, A. T. & Murrell, J. C. Trace-gas metabolic versatility of the facultative methanotroph Methylocella silvestris. *Nature* **510**, 148–151 (2014).
209. Blagodatskaya, E. & Kuzyakov, Y. Active microorganisms in soil: Critical review of estimation criteria and approaches. *Soil Biol. Biochem.* **67**, 192–211 (2013).
210. Lennon, J. T. & Jones, S. E. Microbial seed banks: The ecological and evolutionary implications of dormancy. *Nat. Rev. Microbiol.* **9**, 119–130 (2011).
211. Jones, S. E. & Lennon, J. T. Dormancy contributes to the maintenance of microbial diversity. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 5881–5886 (2010).
212. Liu, G., Chater, K. F., Chandra, G., Niu, G. & Tan, H. Molecular regulation of antibiotic biosynthesis in Streptomyces. *Microbiol. Mol. Biol. Rev.* **77**, 112–143 (2013).
213. Shoemaker, W. R. & Lennon, J. T. Evolution with a seed bank: The population genetic consequences of microbial dormancy. *Evol. Appl.* **11**, 60-75. (2018).
214. Levin, D. A. The seed bank as a source of genetic novelty in plants. *Am. Nat.* **135**, 563–572 (1990).
215. Epstein, S. S. Microbial awakenings. *Nature* **457**, 1083 (2009).
216. Dworkin, J. & Shah, I. M. Exit from dormancy in microbial organisms. *Nat. Rev. Microbiol.* **8**, 890–896 (2010).
217. Thomas, D. S. G. Arid Environments: Their Nature and Extent. in *Arid Zone Geomorphology: Process, Form and Change in Drylands* 1–16 (John Wiley & Sons, Ltd, 2011).
218. Warren-Rhodes, K. A. *et al.* Hypolithic cyanobacteria, dry limit of photosynthesis, and microbial ecology in the hyperarid Atacama Desert. *Microb. Ecol.* **52**, 389–398 (2006).
219. Wong, F. K. Y. *et al.* Hypolithic Microbial Community of Quartz Pavement in the High-Altitude Tundra of Central Tibet. *Microb. Ecol.* **60**, 730–739 (2010).
220. Crits-Christoph, A. *et al.* Phylogenetic and functional substrate specificity for endolithic microbial communities in hyper-arid environments. *Front. Microbiol.* **7**, 1–15 (2016).
221. Martínez, I. *et al.* Small-scale patterns of abundance of mosses and lichens forming biological soil crusts in two semi-arid gypsum environments. *Aust. J. Bot.* **54**, 339 (2006).
222. Pointing, S. B. & Belnap, J. Microbial colonization and controls in dryland systems. *Nature Reviews Microbiology* **10**, 551–562 (2012).
223. Watson, G. M. F. & Tabita, F. R. Regulation, unique gene organization, and unusual primary structure of carbon fixation genes from a marine phycoerythrin-containing cyanobacterium.

Plant Mol. Biol. **32**, 1103–1115 (1996).

224. Badger, M. R. & Bek, E. J. Multiple Rubisco forms in proteobacteria: Their functional significance in relation to CO2 acquisition by the CBB cycle. in *Journal of Experimental Botany* **59**, 1525–1541 (2008).
225. Drees, K. P. *et al.* Bacterial community structure in the hyperarid core of the Atacama Desert, Chile. *Appl. Environ. Microbiol.* **72**, 7902–7908 (2006).
226. Stomeo, F. *et al.* Hypolithic and soil microbial community assembly along an aridity gradient in the Namib Desert. *Extremophiles* **17**, 329–337 (2013).
227. Tebo, B. M. *et al.* Microbial communities in dark oligotrophic volcanic ice cave ecosystems of Mt. Erebus, Antarctica. *Front. Microbiol.* **6**, 1–14 (2015).
228. Grostern, A. & Alvarez-Cohen, L. RubisCO-based CO2 fixation and C1 metabolism in the actinobacterium Pseudonocardia dioxanivorans CB1190. *Environ. Microbiol.* **15**, 3040–3053 (2013).
229. Bay, S., Ferrari, B. & Greening, C. Life without water: how do bacteria generate biomass in desert ecosystems? *Microbiol. Aust.* **39**, 38–32 (2018).
230. Cameron, E. K. *et al.* Global gaps in soil biodiversity data. *Nat. Ecol. Evol.* **2**, 1042–1043 (2018).
231. Coyle, D. R. *et al.* Soil fauna responses to natural disturbances, invasive species, and global climate change: Current state of the science and a call to action. *Soil Biol. Biochem.* **110**, 116–133 (2017).
232. Terrat, S. *et al.* Mapping and predictive variations of soil bacterial richness across France. *PLoS One* **12**, e0186766 (2017).
233. Serna-Chavez, H. M., Fierer, N. & Van Bodegom, P. M. Global drivers and patterns of microbial abundance in soil. *Glob. Ecol. Biogeogr.* **22**, 1162–1172 (2013).
234. Shade, A. *et al.* Macroecology to unite all life, large and small. *Trends Ecol. Evol.* **33**, 731–744 (2018).
235. Fierer, N. & Jackson, R. B. The diversity and biogeography of soil bacterial communities. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 626–631 (2006).
236. Rousk, J. *et al.* Soil bacterial and fungal communities across a pH gradient in an arable soil. *ISME J.* **4**, 1340 (2010).
237. Horner-Devine, M. C., Lage, M., Hughes, J. B. & Bohannan, B. J. M. A taxa – area relationship for bacteria. *Nature* **432**, 750–754 (2004).
238. Wang, J. *et al.* Phylogenetic beta diversity in bacterial assemblages across ecosystems: deterministic versus stochastic processes. *ISME J.* **7**, 1310–1321 (2013).
239. Tripathi, B. M. *et al.* Soil pH mediates the balance between stochastic and deterministic assembly of bacteria. *ISME J.* **12**, 1072–1083 (2018).
240. Wang, X. B. *et al.* Habitat-specific patterns and drivers of bacterial β-diversity in China's drylands. *ISME J.* **11**, 1345–1358 (2017).

241. Evans, S., Martiny, J. B. H. & Allison, S. D. Effects of dispersal and selection on stochastic assembly in microbial communities. *ISME J.* **11**, 176–185 (2017).
242. Albright, M. B. N. & Martiny, J. B. H. Dispersal alters bacterial diversity and composition in a natural community. *ISME J.* **12**, 296–299 (2018).
243. Ferrenberg, S. *et al.* Changes in assembly processes in soil bacterial communities following a wildfire disturbance. *ISME J.* **7**, 1102–1111 (2013).
244. Dini-Andreote, F., Stegen, J. C., van Elsas, J. D. & Salles, J. F. Disentangling mechanisms that mediate the balance between stochastic and deterministic processes in microbial succession. *Proc. Natl. Acad. Sci.* **112**, 1326–1332 (2015).
245. Delgado-Baquerizo, M. *et al.* Changes in belowground biodiversity during ecosystem development. *Proc. Natl. Acad. Sci.* **116**, 6891–6896 (2019).
246. Legendre, P. Spatial autocorrelation: trouble or new paradigm? *Ecology* **74**, 1659–1673 (1993).
247. Baker, K. L. *et al.* Environmental and spatial characterisation of bacterial community composition in soil to inform sampling strategies. *Soil Biol. Biochem.* **41**, 2292–2298 (2009).
248. Diniz-Filho, J. A. F., Bini, L. M. & Hawkins, B. A. Spatial autocorrelation and red herrings in geographical ecology. *Glob. Ecol. Biogeogr.* **12**, 53–64 (2003).
249. Meyer, K. M. *et al.* Why do microbes exhibit weak biogeographic patterns? *ISME J.* **12**, 1404–1413 (2018).
250. Bell, T. *et al.* Larger Islands House More Bacterial Taxa. *Science* **308**, 1884 (2005).
251. Zhou, J., Kang, S., Schadt, C. W. & Garten, C. T. Spatial scaling of functional gene diversity across various microbial taxa. *Proc. Natl. Acad. Sci.* **105**, 7768–7773 (2008).
252. Ranjard, L. *et al.* Turnover of soil bacterial diversity driven by wide-scale environmental heterogeneity. *Nat. Commun.* **4**, 1434 (2013).
253. Fierer, N. & Jackson, R. B. The diversity and biogeography of soil bacterial communities. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 626–631 (2006).
254. Barreto, D. P., Conrad, R., Klose, M., Claus, P. & Enrich-Prast, A. Distance-decay and taxa-area relationships for bacteria, archaea and methanogenic archaea in a tropical lake sediment. *PLoS One* **9**, e110128 (2014).
255. Wang, X.-B. *et al.* Habitat-specific patterns and drivers of bacterial β-diversity in China's drylands. *ISME J.* **11**, 1345–1358 (2017).
256. Jiang, Y. *et al.* Crop rotations alter bacterial and fungal diversity in paddy soils across East Asia. *Soil Biol. Biochem.* **95**, 250–261 (2016).
257. Noguez, A. M. *et al.* Microbial macroecology: Highly structured prokaryotic soil assemblages in a tropical deciduous forest. *Glob. Ecol. Biogeogr.* **14**, 241–248 (2005).
258. Zinger, L., Boetius, A. & Ramette, A. Bacterial taxa-area and distance-decay relationships in marine environments. *Mol. Ecol.* **23**, 954–964 (2014).
259. Reche, I., Pulido-Villena, E., Morales-Baquero, R. & Casamayor, E. O. Does ecosystem size

determine aquatic bacterial richness? *Ecology* **86**, 1715–1722 (2005).
260. Jones, S. E. & Lennon, J. T. Dormancy contributes to the maintenance of microbial diversity. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 5881–5886 (2010).
261. Locey, K. *et al.* Dormancy dampens the microbial distance-decay relationship. *bioRxiv* (2019).
262. Martiny, J. B. H., Eisen, J. A., Penn, K., Allison, S. D. & Horner-Devine, M. C. Drivers of bacterial β-diversity depend on spatial scale. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 7850–7854 (2011).
263. O'Brien, S. L. *et al.* Spatial scale drives patterns in soil bacterial diversity. *Environ. Microbiol.* **6**, 2039–2051 (2016).
264. Hermans, S. M., Buckley, H. L. & Lear, G. Perspectives on the Impact of Sampling Design and Intensity on Soil Microbial Diversity Estimates. *Front. Microbiol.* **10**, 1820 (2019).
265. Castle, S. C. *et al.* Impacts of Sampling Design on Estimates of Microbial Community Diversity and Composition in Agricultural Soils. *Microb. Ecol.* 1–11 (2019).
266. Yeh, C., Soininen, J., Teittinen, A. & Wang, J. Elevational patterns and hierarchical determinants of biodiversity across microbial taxonomic scales. *Mol. Ecol.* **28**, 86–99 (2019).
267. Terrat, S. *et al.* Improving soil bacterial taxa-area relationships assessment using DNA meta-barcoding. *Heredity (Edinb).* **114**, 468–475 (2015).
268. Storch, D. & Šizling, A. L. The concept of taxon invariance in ecology: Do diversity patterns vary with changes in taxonomic resolution? *Folia Geobot.* **43**, 329–344 (2008).
269. Bent, S. J. & Forney, L. J. The tragedy of the uncommon: understanding limitations in the analysis of microbial diversity. *ISME J.* **2**, 689–695 (2008).
270. McMurdie, P. J. & Holmes, S. Waste not, want not: why rarefying microbiome data is inadmissible. *PLoS Comput. Biol.* **10**, (2014).
271. De Vos, W. M. A perspective on 16S rRNA operational taxonomic unit. *npj Biofilms Microbiomes* **57**, 10–13 (2002).
272. Beck, J., Holloway, J. D. & Schwanghart, W. Undersampling and the measurement of beta diversity. *Methods Ecol. Evol.* **4**, 370–382 (2013).
273. Hui, C. & McGeoch, M. A. Zeta diversity as a concept and metric that unifies incidence-based biodiversity patterns. *Am. Nat.* **184**, 684–694 (2014).
274. Crist, T. O., Veech, J. A., Gering, J. C. & Summerville, K. S. Partitioning species diversity across landscapes and regions: a hierarchical analysis of alpha, beta, and gamma diversity. *Am. Nat.* **162**, 734–743 (2003).
275. Wagner, H. H., Wildi, O. & Ewald, K. C. Additive partitioning of plant species diversity in an agricultural mosaic landscape. *Landsc. Ecol.* **15**, 219–227 (2000).
276. Amir, A. *et al.* Deblur rapidly resolves single-nucleotide community sequence patterns. *mSystems* **2**, e00191-16 (2017).
277. Latombe, G., Hui, C. & McGeoch, M. A. Multi-site generalised dissimilarity modelling: using

zeta diversity to differentiate drivers of turnover in rare and widespread species. *Methods Ecol. Evol.* **8**, 431–442 (2017).

278. McGeoch, M. A. *et al.* Measuring continuous compositional change using decline and decay in zeta diversity. *Ecology* **100**, e02832 (2019).
279. Hui, C., Vermeulen, W. & Durrheim, G. Quantifying multiple-site compositional turnover in an Afrotemperate forest, using zeta diversity. *For. Ecosyst.* **5**, 15 (2018).
280. Angel, R., Soares, M. I. M., Ungar, E. D. & Gillor, O. Biogeography of soil archaea and bacteria along a steep precipitation gradient. *ISME J.* **4**, 553–563 (2010).
281. Štovíček, A., Kim, M., Or, D. & Gillor, O. Microbial community response to hydration-desiccation cycles in desert soil. *Sci. Rep.* **7**, 1–9 (2017).
282. Rayment, G. E. & Lyons, D. J. *Soil chemical methods: Australasia*. **3**, (CSIRO publishing, 2011).
283. Trabucco, A. and Zomer, R. J. Global aridity index (global-aridity) and global potential evapo-transpiration (global-PET) geospatial database. *CGIAR Consortium for Spatial Information* (2009). Available at: https://cgiarcsi.community/data/global-aridity-and-pet-database/.
284. Antonio, T. & Zomer, R. J. Global aridity index (global-aridity) and global potential evapo-transpiration (global-PET) geospatial database. *CGIAR Consort. Spat. Inf.* (2009).
285. Paulin, M. M. *et al.* Improving Griffith's protocol for co-extraction of microbial DNA and RNA in adsorptive soils. *Soil Biol. Biochem.* **63**, 37–49 (2013).
286. Marotz, C. *et al.* Triplicate PCR reactions for 16S rRNA gene amplicon sequencing are unnecessary. *Biotechniques* **67**, 1–4 (2019).
287. Gilbert, J. A., Jansson, J. K. & Knight, R. The earth microbiome project: successes and aspirations. *BMC Biology* **12**, 69 (2014).
288. Bolyen, E. *et al.* Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* **37**, 852–857 (2019).
289. Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**, 2584 (2016).
290. Parks, D. H. *et al.* A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* **36**, 996–1004 (2018).
291. Pedregosa Fabian *et al.* Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
292. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
293. Johnson, J. S. *et al.* Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat. Commun.* **10**, 1–11 (2019).
294. Větrovský, T. & Baldrian, P. The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. *PLoS One* **8**, e57923 (2013).

295. Woodcroft, B. J. SingleM. (2019). Available at: https://github.com/wwood/singlem.
296. Boyd, J. A., Woodcroft, B. J. & Tyson, G. W. GraftM: a tool for scalable, phylogenetically informed classification of genes within metagenomes. *Nucleic Acids Res.* **46**, e59–e59 (2018).
297. Lan, Y., Rosen, G. & Hershberg, R. Marker genes that are less conserved in their sequences are useful for predicting genome-wide similarity levels between closely related prokaryotic strains. *Microbiome* **4**, 18 (2016).
298. Wickham, H. Ggplot2 elegant graphics for data analysis. *Springer* (2016).
299. McMurdie, P. J. & Holmes, S. Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* **8**, e61217 (2013).
300. Oksanen, J. *et al.* Vegan: community ecology package. *R Package Version 2. 4-6* (2018).
301. McGeoch, M. A. & Gaston, K. J. Occupancy frequency distributions: patterns, artefacts and mechanisms. *Biol. Rev. Camb. Philos. Soc.* **77**, 311–331 (2002).
302. Scheiner, S. M. *et al.* The underpinnings of the relationship of species richness with space and time. *Ecol. Monogr.* **81**, 195–213 (2011).
303. Drakare, S., Lennon, J. J. & Hillebrand, H. The imprint of the geographical, evolutionary and ecological context on species-area relationships. *Ecol. Lett.* **9**, 215–227 (2006).
304. Wang, Y., Naumann, U., Wright, S. T. & Warton, D. I. Mvabund- an R package for model-based analysis of multivariate abundance data. *Methods Ecol. Evol.* **3**, 471–474 (2012).
305. Warton, D. I., Thibaut, L. & Wang, Y. A. The PIT-trap — A “model-free” bootstrap procedure for inference about regression models with discrete, multivariate responses. *PLoS One* **12**, e0181790 (2017).
306. Morton, J. T. *et al.* Uncovering the Horseshoe Effect in Microbial Analyses. *mSystems* **2**, e00166-16 (2017).
307. Johnson, R. M., Ramond, J., Gunnigle, E., Seely, M. & Cowan, D. A. Namib Desert edaphic bacterial, fungal and archaeal communities assemble through deterministic processes but are influenced by different abiotic parameters. *Extremophiles* **21**, 381–392 (2017).
308. Ramette, A. & Tiedje, J. M. Multiscale responses of microbial life to spatial distance and environmental heterogeneity in a patchy ecosystem. *Proc. Natl. Acad. Sci.* **104**, 2761–2766 (2007).
309. Yergeau, E. *et al.* Influences of space, soil, nematodes and plants on microbial community composition of chalk grassland soils. *Environ. Microbiol.* **12**, 2096–2106 (2010).
310. Bell, T. *et al.* Ecology: Larger islands house more bacterial taxa. *Science* **308**, 1884 (2005).
311. Bankevich, A. & Pevzner, P. A. Joint analysis of long and short reads enables accurate estimates of microbiome complexity. *Cell Syst.* **7**, 192–200 (2018).
312. Zinger, L., Boetius, A. & Ramette, A. Bacterial taxa–area and distance–decay relationships in marine environments. *Mol. Ecol.* **23**, 954–964 (2014).
313. Oono, R., Rasmussen, A. & Lefèvre, E. Distance decay relationships in foliar fungal

endophytes are driven by rare taxa. *Environ. Microbiol.* **19**, 2794–2805 (2017).

314. Nekola, J. C. & White, P. S. The distance decay of similarity in biogeography and ecology. *J. Biogeogr.* **26**, 867–878 (1999).
315. Logares, R. *et al.* Biogeography of bacterial communities exposed to progressive long-term environmental change. *ISME J.* **7**, 937–948 (2013).
316. Székely, A. J. & Langenheder, S. The importance of species sorting differs between habitat generalists and specialists in bacterial communities. *FEMS Microbiol. Ecol.* **87**, 102–112 (2014).
317. Chen, Y.-J. *et al.* Metabolic flexibility allows generalist bacteria to become dominant in a frequently disturbed ecosystem. *bioRxiv* 2020.02.12.945220 (2020).
318. Latombe, G., Richardson, D. M., Pyšek, P., Kučera, T. & Hui, C. Drivers of species turnover vary with species commonness for native and alien plants with different residence times. *Ecology* **99**, 2763–2775 (2018).
319. del Barrio, G., Sanjuán, M. E., Ruiz, A. & Martínez Valderrama, J. Puigdefábregas, J. Case study: Land condition surveillance using geospatial data (Iberian Peninsula and Maghreb). in *World Atlas of Desertification* (EU Joint Research Centre, 2018).
320. Yan-Gui, S., Xin-Rong, L., Ying-Wu, C., Zhi-Shan, Z. & Yan, L. Carbon fixation of cyanobacterial-algal crusts after desert fixation and its implication to soil organic carbon accumulation in desert. *L. Degrad. Dev.* **24**, 342–349 (2013).
321. Housman, D. C., Powers, H. H., Collins, A. D. & Belnap, J. Carbon and nitrogen fixation differ between successional stages of biological soil crusts in the Colorado Plateau and Chihuahuan Desert. *J. Arid Environ.* **66**, 620–634 (2006).
322. Van Goethem, M. W., Makhalanyane, T. P., Cowan, D. A. & Valverde, A. Cyanobacteria and Alphaproteobacteria may facilitate cooperative interactions in niche communities. *Front. Microbiol.* **8**, 2099 (2017).
323. Ramond, J. B., Woodborne, S., Hall, G., Seely, M. & Cowan, D. A. Namib Desert primary productivity is driven by cryptic microbial community N-fixation. *Sci. Rep.* **8**, 6921 (2018).
324. Fierer, N. *et al.* Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proc. Natl. Acad. Sci.* **109**, 21390–21395 (2012).
325. Yergeau, E., Newsham, K. K., Pearce, D. A. & Kowalchuk, G. A. Patterns of bacterial diversity across a range of Antarctic terrestrial habitats. *Environ. Microbiol.* **9**, 2670–2682 (2007).
326. Leung, P. M. *et al.* Energetic Basis of Microbial Growth and Persistence in Desert Ecosystems. *mSystems* **5**, e00495-19 (2020).
327. Locey, K. J., Fisk, M. C. & Lennon, J. T. Microscale insight into microbial seed banks. *Front. Microbiol.* **7**, 2040 (2017).
328. Kieft, T. L., Soroker, E. & Firestone, M. K. Microbial Biomass Response to a Rapid Increase in Water Potential When Dry Soil Is Wetted. *Soil Biol. Biochem.* **19**, 119–126 (1987).

329. Sponseller, R. A. Precipitation pulses and soil CO2 flux in a Sonoran Desert ecosystem. *Glob Chang Biol* **13**, 426–436 (2007).
330. Constant, P., Chowdhury, S. P., Pratscher, J. & Conrad, R. Streptomycetes contributing to atmospheric molecular hydrogen soil uptake are widespread and encode a putative high-affinity [NiFe]-hydrogenase. *Environ. Microbiol.* **12**, 821–829 (2010).
331. Greening, C., Berney, M., Hards, K., Cook, G. M. & Conrad, R. A soil actinobacterium scavenges atmospheric H_2 using two membrane-associated, oxygen-dependent [NiFe] hydrogenases. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 4257–4261 (2014).
332. Greening, C., Villas-Bôas, S. G., Robson, J. R., Berney, M. & Cook, G. M. The growth and survival of Mycobacterium smegmatis is enhanced by co-metabolism of atmospheric H2. *PLoS One* **9**, e103034 (2014).
333. Berney, M. & Cook, G. M. Unique flexibility in energy metabolism allows mycobacteria to combat starvation and hypoxia. *PLoS One* **5**, e8614 (2010).
334. Liot, Q. & Constant, P. Breathing air to save energy - new insights into the ecophysiological role of high-affinity [NiFe]-hydrogenase in Streptomyces avermitilis. *Microbiologyopen* **6**, 47–49 (2016).
335. Berney, M., Greening, C., Conrad, R., Jacobs, W. R. & Cook, G. M. An obligately aerobic soil bacterium activates fermentative hydrogen production to survive reductive stress during hypoxia. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 11479–11484 (2014).
336. Conrad, R. Soil Microorganisms as Controllers of Atmospheric Trace Gases (H2 , CO, CH4 , OCS, N2O, and NO). *Microbiol. Rev.* **60**, 609–640 (1996).
337. Cordero, P. R. F. *et al.* Two uptake hydrogenases differentially interact with the aerobic respiratory chain during mycobacterial growth and persistence. *J. Biol. Chem.* **294**, 18980–18991 (2019).
338. Shomura, Y., Yoon, K.-S., Nishihara, H. & Higuchi, Y. Structural basis for a [4Fe-3S] cluster in the oxygen-tolerant membrane-bound [NiFe]-hydrogenase. *Nature* **479**, 253–256 (2011).
339. Ortiz, E. M. *et al.* Bacterial communities across an Antarctic ecotone use a novel hydrogenase lineage to scavenge atmospheric H_2. In preparation (2020).
340. Constant, P., Chowdhury, S. P., Hesse, L. & Conrad, R. Co-localization of atmospheric H_2 oxidation activity and high affinity H_2-oxidizing bacteria in non-axenic soil and sterile soil amended with *Streptomyces* sp. PCB7. *Soil Biol. Biochem.* **43**, 1888–1893 (2011).
341. Carere, C. R. *et al.* Mixotrophy drives niche expansion of verrucomicrobial methanotrophs. *ISME J.* **11**, 2599–2610 (2017).
342. Bay, S. K. *et al.* Soil bacterial communities exhibit strong biogeographic patterns at fine taxonomic resolution. *mSystems* In review (2020).
343. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. MetaSPAdes: A new versatile metagenomic assembler. *Genome Res.* **27**, 824–834 (2017).
344. Kang, D. D. *et al.* MetaBAT 2: An adaptive binning algorithm for robust and efficient genome

reconstruction from metagenome assemblies. *PeerJ* **2019**, e7359 (2019).
345. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
346. Bowers, R. M. *et al.* Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat. Biotechnol.* **35**, 725–731 (2017).
347. Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* In press (2020).
348. Parks, D. H. *et al.* A proposal for a standardized bacterial taxonomy based on genome phylogeny. *Nat. Biotechnol.* In press (2018).
349. Dixon, P. VEGAN, a package of R functions for community ecology. *J. Veg. Sci.* **14**, 927–930 (2003).
350. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
351. Anantharaman, K. *et al.* Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat. Commun.* **7**, 13219 (2016).
352. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
353. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* msw054 (2016).
354. Wickham, H. 'The tidyverse.' R package ver.1.1 1. (2017).
355. Lehtovirta-Morley, L. E. *et al.* Isolation of 'Candidatus Nitrosocosmicus franklandus', a novel ureolytic soil archaeal ammonia oxidiser with tolerance to high ammonia concentration. *FEMS Microbiol. Ecol.* **92**, fiw057 (2016).
356. Rao, B. *et al.* Influence of dew on biomass and photosystem II activity of cyanobacterial crusts in the Hopq Desert, northwest China. *Soil Biol. Biochem.* **41**, 2387–2393 (2009).
357. Soo, R. M., Hemp, J., Parks, D. H., Fischer, W. W. & Hugenholtz, P. On the origins of oxygenic photosynthesis and aerobic respiration in Cyanobacteria. *Science* **355**, 1436–1440 (2017).
358. Carnevali, P. B. M. *et al.* Hydrogen-based metabolism as an ancestral trait in lineages sibling to the Cyanobacteria. *Nat. Commun.* **10**, 1–15 (2019).
359. Søndergaard, D., Pedersen, C. N. S. & Greening, C. HydDB: a web tool for hydrogenase classification and analysis. *Sci. Rep.* **6**, 34212 (2016).
360. Moxley, J. M. & Smith, K. A. Factors affecting utilisation of atmospheric CO by soils. *Soil Biol. Biochem.* **30**, 65–79 (1998).
361. Quiza, L., Lalonde, I., Guertin, C. & Constant, P. Land-use influences the distribution and activity of high affinity CO-oxidizing bacteria associated to type I-coxL genotype in soil.

Front. Microbiol. **5**, 271 (2014).

362. Hagemann, M. *et al.* Cyanobacterial diversity in biological soil crusts along a precipitation gradient, Northwest Negev Desert, Israel. *Microb. Ecol.* **70**, 219–230 (2015).
363. Friedmann, I., Lipkin, Y. & Ocampo-Paus, R. Desert algae of the Negev (Israel). *Phycologia* **6**, 185–200 (1967).
364. Herrnstadt, I., Heyn, C. C. & Crosby, M. R. A checklist of the mosses of Israel. *Bryologist* 168–178 (1991).
365. Grostern, A. & Alvarez-Cohen, L. RubisCO-based CO_2 fixation and C1 metabolism in the actinobacterium *Pseudonocardia dioxanivorans* CB1190. *Environ. Microbiol.* **15**, 3040–3053 (2013).
366. Belnap, J. & Lange, O. L. *Biological soil crusts: structure, function, and management*. **150**, (Springer Science & Business Media, 2013).
367. Belnap, J. The potential roles of biological soil crusts in dryland hydrologic cycles. *Hydrol. Process. An Int. J.* **20**, 3159–3178 (2006).
368. Hill, A. J., Dawson, T. E., Shelef, O. & Rachmilevitch, S. The role of dew in Negev Desert plants. *Oecologia* **178**, 317–327 (2015).
369. Zangvil, A. Six years of dew observations in the Negev Desert, Israel. *J. Arid Environ.* **32**, 361–371 (1996).
370. Kidron, G. J., Herrnstadt, I. & Barzilay, E. The role of dew as a moisture source for sand microbiotic crusts in the Negev Desert, Israel. *J. Arid Environ.* **52**, 517–533 (2002).
371. Kuske, C. R., Yeager, C. M., Johnson, S., Ticknor, L. O. & Belnap, J. Response and resilience of soil biocrust bacterial communities to chronic physical disturbance in arid shrublands. *ISME J.* **6**, 886–897 (2012).
372. Steven, B., Gallegos-Graves, L. V., Belnap, J. & Kuske, C. R. Dryland soil microbial communities display spatial biogeographic patterns associated with soil depth and soil parent material. *FEMS Microbiol. Ecol.* **86**, 101–113 (2013).
373. Steven, B., Gallegos-Graves, L. V., Starkenburg, S. R., Chain, P. S. & Kuske, C. R. Targeted and shotgun metagenomic approaches provide different descriptions of dryland soil microbial communities in a manipulated field study. *Environ. Microbiol. Rep.* **4**, 248–256 (2012).
374. Swenson, T. L., Karaoz, U., Swenson, J. M., Bowen, B. P. & Northen, T. R. Linking soil biology and chemistry in biological soil crust using isolate exometabolomics. *Nat. Commun.* **9**, 1–10 (2018).
375. Li, J.-Y. *et al.* Comparative metagenomics of two distinct biological soil crusts in the Tengger Desert, China. *Soil Biol. Biochem.* **140**, 107637 (2020).
376. Blay, E. S. *et al.* Variation in biological soil crust bacterial abundance and diversity as a function of climate in cold steppe ecosystems in the Intermountain West, USA. *Microb. Ecol.* **74**, 691–700 (2017).

377. Wilson, J. M. & Griffin, D. M. Water potential and the respiration of microorganisms in the soil. *Soil Biol. Biochem.* **7**, 199–204 (1975).
378. Iovieno, P. & Bååth, E. Effect of drying and rewetting on bacterial growth rates in soil. *FEMS Microbiol. Ecol.* **65**, 400–407 (2008).
379. Bell, C., McIntyre, N., Cox, S., Tissue, D. & Zak, J. Soil microbial responses to temporal variations of moisture and temperature in a Chihuahuan Desert grassland. *Microb. Ecol.* **56**, 153–167 (2008).
380. Bell, C. W. *et al.* Soil microbial and nutrient responses to 7 years of seasonally altered precipitation in a Chihuahuan Desert grassland. *Glob. Chang. Biol.* **20**, 1657–1673 (2014).
381. Brockett, B. F. T., Prescott, C. E. & Grayston, S. J. Soil moisture is the major factor influencing microbial community structure and enzyme activities across seven biogeoclimatic zones in western Canada. *Soil Biol. Biochem.* **44**, 9–20 (2012).
382. Stark, J. M. & Firestone, M. K. Mechanisms for soil moisture effects on activity of nitrifying bacteria. *Appl. Environ. Microbiol.* **61**, 218–221 (1995).
383. Jarvis, P. *et al.* Drying and wetting of Mediterranean soils stimulates decomposition and carbon dioxide emission: The 'Birch effect'. *Tree Physiol.* **27**, 929–940 (2007).
384. Meredith, L. K. *et al.* Consumption of atmospheric hydrogen during the life cycle of soil-dwelling actinobacteria. *Environ. Microbiol. Rep.* **6**, 226–38 (2014).
385. Greening, C. *et al.* Persistence of the dominant soil phylum Acidobacteria by trace gas scavenging. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 10497–10502 (2015).
386. Khdhiri, M. Soil Carbon Content and Relative Abundance of High Affinity H 2 -Oxidizing Bacteria Predict Atmospheric H 2 Soil Uptake Activity Better than Soil Microbial Community Composition. *Soil Biol. Biochem.* **85**, 1–9 (2015).
387. Khdhiri, M., Piché-Choquette, S., Tremblay, J., Tringe, S. G. & Constant, P. Meta-omics survey of [NiFe]-hydrogenase genes fails to capture drastic variations in H2-oxidation activity measured in three soils exposed to H2. *Soil Biol. Biochem.* **125**, 239–243 (2018).
388. Lynch, R. C., Darcy, J. L., Kane, N. C., Nemergut, D. R. & Schmidt, S. K. Metagenomic evidence for metabolism of trace atmospheric gases by high-elevation desert actinobacteria. *Front. Microbiol.* **5**, 1–13 (2014).
389. Dunfield, P. F. *et al.* Methane oxidation by an extremely acidophilic bacterium of the phylum Verrucomicrobia. *Nature* **450**, 879–882 (2007).
390. Thomas Bell Bernard W. Silverman, Sarah L. Turner and Andrew K. Lilley, J. A. N. The contribution of species richness and composition to bacterial services. *Nature* **436**, 1157 (2005).
391. Joseph, S. J., Hugenholtz, P., Sangwan, P., Osborne, C. A. & Janssen, P. H. Laboratory cultivation of widespread and previously uncultured soil bacteria. *Appl. Environ. Microbiol.* **69**, 7210–7215 (2003).
392. Beeckman, F., Motte, H. & Beeckman, T. Nitrification in agricultural soils: impact, actors and

mitigation. *Curr. Opin. Biotechnol.* **50**, 166–173 (2018).
393. Conrad, R. Soil microorganisms as controllers of atmospheric trace gases (H_2, CO, CH_4, OCS, N_2O, and NO). *Microbiol. Mol. Biol. Rev.* **60**, 609–640 (1996).
394. Novelli, P. C., Masarie, K. A. & Lang, P. M. Distributions and recent changes of carbon monoxide in the lower troposphere. *J. Geophys. Res. Atmos.* **103**, 19015–19033 (1998).
395. Ehhalt, D. H. & Rohrer, F. The tropospheric cycle of H_2: a critical review. *Tellus B* **61**, 500–535 (2009).
396. Constant, P., Poissant, L. & Villemur, R. Tropospheric H2 budget and the response of its soil uptake under the changing environment. *Sci. Total Environ.* **407**, 1809–1823 (2009).
397. Tarr, M. A., Miller, W. L. & Zepp, R. G. Direct carbon monoxide photoproduction from plant matter. *J. Geophys. Res.* **100**, 403–414 (1995).
398. Mörsdorf, G., Frunzke, K., Gadkari, D. & Meyer, O. Microbial growth on carbon monoxide. *Biodegradation* **3**, 61–82 (1992).
399. King, G. M. Contributions of atmospheric CO and hydrogen uptake to microbial dynamics on recent Hawaiian volcanic deposits. *Appl. Environ. Microbiol.* **69**, 4067–4075 (2003).
400. Chen, Q., Popa, M. E., Batenburg, A. M. & Röckmann, T. Isotopic signatures of production and uptake of H_2 by soil. *Atmos. Chem. Phys.* **15**, 13003–13021 (2015).
401. Hüppi, R. *et al.* Restricting the nonlinearity parameter in soil greenhouse gas flux calculation for more reliable flux estimates. *PLoS One* **13**, (2018).
402. Caporaso, J. G. *et al.* Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci.* **108**, 4516–4522 (2011).
403. Uritskiy, G. V., DiRuggiero, J. & Taylor, J. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* **6**, 158 (2018).
404. Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676 (2015).
405. Kang, D. D., Froula, J., Egan, R. & Wang, Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **3**, e1165 (2015).
406. Wu, Y.-W., Simmons, B. A. & Singer, S. W. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**, 605–607 (2015).
407. Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J.* **11**, 2864–2868 (2017).
408. Markowitz, V. M. *et al.* IMG: the integrated microbial genomes database and comparative analysis system. *Nucleic Acids Research* **40**, D115-22 (2012).
409. Leinonen, R., Sugawara, H., Shumway, M. & Collaboration, I. N. S. D. The sequence read archive. *Nucleic Acids Res.* **39**, D19–D21 (2010).
410. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND.

Nat. Methods **12**, 59–60 (2015).

411. Helgeson, H., Kirkham, D. & Flowers, G. Theoretical prediction of the thermodynamic behavior of aqueous electrolytes by high pressures and temperatures; IV, Calculation of activity coefficients, osmotic coefficients, and apparent molal and standard and relative partial molal properties to 600 d. *Am. J. Sci.* **281**, 1249–1516 (1981).
412. Tanger, J. C. & Helgeson, H. C. Calculation of the thermodynamic and transport properties of aqueous species at high pressures and temperatures; revised equations of state for the standard partial molal properties of ions and electrolytes. *Am. J. Sci.* **288**, 19–98 (1988).
413. Shock, E. L. (University of C. at B., Oelkers, E. H., Johnson, J. W., Sverjensky, D. A. & Helgeson, H. C. Calculation of the thermodynamic properties of aqueous Species at high pressures and temperatures. *J. Chem. Soc. Faraday Trans.* **88**, 803–826 (1992).
414. Johnson, J. W., Oelkers, E. H. & Helgeson, H. C. A software package for calculating the standard molal thermodynamic properties of minerals, gases, aqueous species and reactions from 1 to 5000 bars and 0° to 1000° C. *Comput. Geosci.* **18**, 899–947 (1992).
415. Shock, E. L. & Helgeson, H. C. Calculation of the thermodynamic and transport properties of aqueous species at high pressures and temperatures: Correlation algorithms for ionic species and equation of state predictions to 5 kb and 1000°C. *Geochim. Cosmochim. Acta* **52**, 2009–2036 (1988).
416. Shock, E. L., Helgeson, H. C. & Sverjensky, D. A. Calculation of the thermodynamic and transport properties of aqueous species at high pressures and temperatures: Standard partial molal properties of inorganic neutral species. *Geochim. Cosmochim. Acta* **53**, 2157–2183 (1989).
417. Shock, E. L. & Helgeson, H. C. Calculation of the thermodynamic and transport properties of aqueous species at high pressures and temperatures: Standard partial molal properties of organic species. *Geochim. Cosmochim. Acta* **54**, 915–945 (1990).
418. Sverjensky, D. A., Shock, E. L. & Helgeson, H. C. Prediction of the thermodynamic properties of aqueous metal complexes to 1000°C and 5 kb. *Geochim. Cosmochim. Acta* **61**, 1359–1412 (1997).
419. Schulte, M. D., Shock, E. L. & Wood, R. H. The temperature dependence of the standard-state thermodynamic properties of aqueous nonelectrolytes. *Geochim. Cosmochim. Acta* **65**, 3919–3930 (2001).
420. Amend, J. P. & LaRowe, D. E. Minireview: demystifying microbial reaction energetics. *Environ. Microbiol.* **21**, 3539–3547 (2019).
421. Constant, P., Chowdhury, S. P., Hesse, L., Pratscher, J. & Conrad, R. Genome data mining and soil survey for the novel group 5 [NiFe]-hydrogenase to explore the diversity and ecological importance of presumptive high-affinity H 2-oxidizing bacteria. *Appl. Environ. Microbiol.* **77**, 6027–6035 (2011).
422. Pandelia, M. E., Lubitz, W. & Nitschke, W. Evolution and diversification of Group 1 [NiFe]

hydrogenases. Is there a phylogenetic marker for O2-tolerance? *Biochim. Biophys. Acta - Bioenerg.* **1817**, 1565–1575 (2012).

423. van Kessel, M. A. H. J. *et al.* Complete nitrification by a single microorganism. *Nature* **528**, 555 (2015).
424. Daims, H. *et al.* Complete nitrification by *Nitrospira* bacteria. *Nature* **528**, 504 (2015).
425. Kirschke, S. *et al.* Three decades of global methane sources and sinks. *Nat. Geosci* **6**, 813–823 (2013).
426. Dutaur, L. & Verchot, L. V. A global inventory of the soil CH4 sink. *Global Biogeochem. Cycles* **21**, GB4013 (2007).
427. Ehhalt, D. H. & Rohrer, F. The dependence of soil H2 uptake on temperature and moisture: A reanalysis of laboratory data. *Tellus, Ser. B Chem. Phys. Meteorol.* **63**, 1040–1051 (2011).
428. Smith-Downey, N. V, Randerson, J. T. & Eiler, J. M. Molecular hydrogen uptake by soils in forest, desert, and marsh ecosystems in California. *J. Geophys. Res* **113**, 3037 (2008).
429. Tijhuis, L., Van Loosdrecht, M. C. & Heijnen, J. J. A thermodynamically based correlation for maintenance gibbs energy requirements in aerobic and anaerobic chemotrophic growth. *Biotechnol. Bioeng.* **42**, 509–519 (1993).
430. Marschall, E., Jogler, M., Henßge, U. & Overmann, J. Large-scale distribution and activity patterns of an extremely low-light-adapted population of green sulfur bacteria in the Black Sea. *Environ. Microbiol.* **12**, 1348–1362 (2010).
431. LaRowe, D. E. & Amend, J. P. Power limits for microbial life. *Front. Microbiol.* **6**, 718 (2015).
432. Ekblad, A. & Nordgren, A. Is growth of soil microorganisms in boreal forests limited by carbon or nitrogen availability? *Plant Soil* **242**, 115–122 (2002).
433. Demoling, F., Figueroa, D. & Bååth, E. Comparison of factors limiting bacterial growth in different soils. *Soil Biol. Biochem.* **39**, 2485–2495 (2007).
434. Schultz, M. G., Diehl, T., Brasseur, G. P. & Zittel, W. Air pollution and climate-forcing impacts of a global hydrogen economy. *Science* **302**, 624–627 (2003).
435. Chase, J. M. Drought mediates the importance of stochastic community assembly. *Proc. Natl. Acad. Sci.* **104**, 17430–17434 (2007).
436. Jia, X. & Dini-Andreote, F. Community Assembly Processes of the Microbial Rare Biosphere. *Artic. Trends Microbiol.* **8**, 1–5 (2018).
437. Finlay, B. J. *et al.* Global dispersal of free-living microbial eukaryote species. *Science* **296**, 1061–3 (2002).
438. Finlay, B. J. & Clarke, K. J. Ubiquitous dispersal of microbial species. *Nature* **400**, 828 (1999).
439. Bar, M., Hardenberg, J., Meron, E. & Provenzale, A. Modelling the survival of bacteria in drylands: the advantage of being dormant. *Proc. R. Soc. B Biol. Sci.* **269**, 937–942 (2002).
440. Cáceres, C. E. Temporal variation, dormancy, and coexistence: a field test of the storage

effect. *Proc. Natl. Acad. Sci.* **94**, 9171–9175 (1997).

441. Locey, K. J. & Lennon, J. T. Scaling laws predict global microbial diversity. *Proc. Natl. Acad. Sci.* **113**, 5970–5975 (2016).
442. Park, S. W. *et al.* Presence of duplicate genes encoding a phylogenetically new subgroup of form I ribulose 1,5-bisphosphate carboxylase/oxygenase in Mycobacterium sp. strain JC1 DSM 3803. *Res. Microbiol.* **160**, 159–165 (2009).
443. Meyer, C., Weigelt, P. & Kreft, H. Multidimensional biases, gaps and uncertainties in global plant occurrence information. *Ecol. Lett.* **19**, 992–1006 (2016).
444. Platts, P. J., McClean, C. J., Lovett, J. C. & Marchant, R. Predicting tree distributions in an East African biodiversity hotspot: model selection, data bias and envelope uncertainty. *Ecol. Modell.* **218**, 121–134 (2008).
445. Ready, R. C., Champ, P. A. & Lawton, J. L. Using respondent uncertainty to mitigate hypothetical bias in a stated choice experiment. *Land Econ.* **86**, 363–381 (2010).
446. Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G. A. C. & Baird, D. J. Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. *PLoS One* **6**, e17497 (2011).
447. Harris, R. M. B. *et al.* Biological responses to the press and pulse of climate trends and extreme events. *Nat. Clim. Chang.* **8**, 579 (2018).
448. Conrad, R. & Eiler, W. Influence of Temperature, Moisture, and Organic Carbon on the Flux of and CO Between Soil and Atmosphere' Field Studies in Subtropical Regions. *Exch. Organ. Behav. Teach. J.* **90**, 5699–5709 (1985).
449. Van Asperen, H. *et al.* The role of photo- And thermal degradation for CO2 and CO fluxes in an arid ecosystem. *Biogeosciences* **12**, 4161–4174 (2015).
450. Neilson, J. W. *et al.* Significant Impacts of Increasing Aridity on the Arid Soil Microbiome. *mSystems* **2**, 5–16 (2017).
451. Scola, V. *et al.* Namib Desert Soil Microbial Community Diversity, Assembly, and Function Along a Natural Xeric Gradient. *Microb. Ecol.* **75**, 193–203 (2018).
452. Tait, A. W., Gagen, E. J., Wilson, S. A., Tomkins, A. G. & Southam, G. Microbial populations of stony meteorites: substrate controls on first colonizers. *Front. Microbiol.* **8**, 1227 (2017).
453. Chiri, E., Nauer, P. A., Henneberger, R., Zeyer, J. & Schroth, M. H. Soil–methane sink increases with soil age in forefields of Alpine glaciers. *Soil Biol. Biochem.* **84**, 83–95 (2015).
454. Lucas-Borja, M. E. *et al.* Immediate fire-induced changes in soil microbial community composition in an outdoor experimental controlled system. *Sci. Total Environ.* **696**, 134033 (2019).
455. Harris, J. Measurements of the soil microbial community for estimating the success of restoration. *Eur. J. Soil Sci.* **54**, 801–808 (2003).
456. Liao, M. & Xie, X. M. Effect of heavy metals on substrate utilization pattern, biomass, and activity of microbial communities in a reclaimed mining wasteland of red soil area.

Ecotoxicol. Environ. Saf. **66**, 217–223 (2007).

457. Palmer, A. N. Origin and morphology of limestone caves. *Geol. Soc. Am. Bull.* **103**, 1–21 (1991).
458. Greeley, R. & Hyde, J. H. Lava tubes of the cave basalt, Mount St. Helens, Washington. *Geol. Soc. Am. Bull.* **83**, 2397–2418 (1972).
459. Ortiz, M. *et al.* Profiling bacterial diversity and taxonomic composition on speleothem surfaces in Kartchner Caverns, AZ. *Microb. Ecol.* **65**, 371–383 (2013).
460. Ortiz, M. *et al.* Making a living while starving in the dark: metagenomic insights into the energy dynamics of a carbonate cave. *ISME J.* **8**, 478–491 (2014).
461. Webster, K. D., Mirza, A., Deli, J. M., Sauer, P. E. & Schimmelmann, A. Consumption of atmospheric methane in a limestone cave in Indiana, USA. *Chem. Geol.* **443**, 1–9 (2016).
462. Portillo, M. C., Leff, J. W., Lauber, C. L. & Fierer, N. Cell size distributions of soil bacterial and archaeal taxa. *Appl. Environ. Microbiol.* **79**, 7610–7617 (2013).
463. Sait, M., Hugenholtz, P. & Janssen, P. H. Cultivation of globally distributed soil bacteria from phylogenetic lineages previously only detected in cultivation-independent surveys. *Environ. Microbiol.* **4**, 654–666 (2002).