Global functional spectrum of soil bacterial communities

Gabin Piton (gabinpiton@gmail.com)
French National Institute for Agriculture, Food, and Environment (INRAE) https://orcid.org/0000-0002-6036-5787
Steven Allison
University of California, Irvine https://orcid.org/0000-0003-4629-7842
Mohammad Bahram
University of Tartu
Falk Hildebrand
Quadram Institute Bioscience
Jennifer Martiny
University of California, Irvine https://orcid.org/0000-0002-2415-1247
Kathleen K. Treseder
University of California Irvine
Adam Martiny
University of California, Irvine https://orcid.org/0000-0003-2829-4314

Article

Keywords:

Posted Date: August 2nd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1885110/v1

Global functional spectrum of soil bacterial communities

Gabin Piton[1,2], Steven D Allison[1,3], Mohammad Bahram[4,5], Falk Hildebrand[6,7], Jennifer BH Martiny[3], Kathleen K Treseder[3], Adam C Martiny[1,3]

Author information

Affiliations

1. Department of Earth System Science, University of California, Irvine, California, USA
2. Eco&Sols, INRAE-IRD-CIRAD-SupAgro, University Montpellier, Montpellier, France
3. Department of Ecology and Evolutionary Biology, University of California, Irvine, California, USA
4. Department of Ecology, Swedish University of Agricultural Sciences, Uppsala, Sweden
5. Institute of Ecology and Earth Sciences, University of Tartu, Tartu, Estonia
6. Gut Microbes & Health, Quadram Institute Bioscience, Norwich Research Park, Norwich, Norfolk NR4 7UA, UK
7. Digital Biology, Earlham Institute, Norwich Research Park, Norwich, Norfolk NR4 7UA, UK.

Abstract

Despite the tremendous importance of soil bacterial communities in biogeochemical cycles, the main functional variations developed by soil bacteria communities *in-situ* (as opposed to culture conditions) across the globe, and their environmental drivers has yet to be elucidated. Here we use shotgun metagenomes from all terrestrial biomes to characterize a genomic trait based soil bacteria's functional spectrum, simplifying in two dimensions 53% of the global variation of their functional potential. Using machine learning, we show that soil pH, P content and climate (Precipitation patterns and temperature) predict 78% and 45% of the first and second dimension of the spectrum, respectively. The first dimension captures a continuum from small to large genomes associated with an increase of metabolism and resource acquisition complexity and versatility. Decrease in water stress and increased acidity drives

this shift from small- to large-genome. The second dimension shows a continuum from microbial necromass and nutrient usage specialists to reactive degraders of simple carbons with more CAZy, rRNA and sigma factor genes. This shift to reactive bacteria was observed when climate restrains the growing window (low precipitation with high seasonality and extreme temperature) and pH becomes acidic. Overall, our study showed that metabolism versatility, reactivity and stoichiometry dominate soil bacteria functional spectrum across the globe.

Introduction

Bacteria represent one of the most important biological carbon (C) pools in soil, cycling C and nutrients on a global scale (Fierer 2017). Soil bacterial communities contain enormous genetic diversity which remains mostly functionally uncharacterized (Delgado-Baquerizo et al. 2018, Bahram et al. 2018), hindering progress in our understanding of their response to global changes and their biogeochemical roles (Wieder et al. 2013, Crowther et al. 2019). A promising approach to describe biodiversity is to define functional spectrums representing the dominant associations between traits (Grime 1977, Southwood 1977, Reich et al. 2003, Wright et al. 2004, Diaz et al. 2016) that capture the main organismal tradeoffs underlying adaptations to environments and their effect on ecosystem functioning (Lavorel and Garnier 2002). In plant ecology, functional (phenotypic) traits have been successfully used to define global functional spectrum across plant species (Reich et al. 2003, Wright et al. 2004, Diaz et al. 2016). This has provided a general framework to identify systematic shifts in plant strategies along environmental gradients. One can further extend the spectrum up to the community level (Bruelheide et al. 2018) to understand linkages to ecosystem functioning (eg. fast-slow spectrum role in response to light, water and nutrients and consequences for litter decomposition and primary productivity reviewed in Reich 2014).

Very recently, Westoby et al. (2021) have started to implement similar approach to describe global spectrum of cultured bacteria isolated from various habitats (fresh and marine waters, soils and sediments, animal and plant hosts, and thermal environments) using their genomic attributes and phenotypic traits measured under culture condition (Madin et al. 2020). This global functional spectrum of cultured bacteria showed a first dimension associated with the functional versatility of bacteria (ie.

their capacity to use different resources) that was highly correlated with genome size. The second dimension captured differences in maximum growth rate and was correlated with variation in rRNA gene copy number. However, such culture-based studies might miss part of bacteria functional diversity existing *in situ* in soil environments (Delgado-Baquerizo et al. 2020, Martiny et al. 2019, 2020, Steen et al. 2019) and cannot provide information on the relative importance of each trait under a given environmental condition. Thus, the global functional spectrum across soil bacteria communities (as opposed to spectrum across strains) *in-situ* (as opposed to culture conditions) and its environmental drivers has yet to be elucidated.

Community level patterns of functional variation emerge from a combination of organisms' evolutionary history and community organization in response to biotic and abiotic conditions. Following the filtering theory (Keddy 1992, Weiher et al. 2011, Grime and Pierce, 2012), co-occurring organisms of a given community share primary adaptation needed to colonize, establish and persist in their specific environment. Among co-occurring organisms, dominant taxa are assumed to be best adapted to local conditions (Ackerly 2003, Shipley et al. 2006) and will have the biggest contribution to the community aggregated traits (CAT). Studies based on CAT showed for instance that soil microbial community functional potential responds to experimental nitrogen addition (Fierer et al. 2012) and that genomic traits like bacteria genome size decrease along a steep local temperature gradient (Sorensen et al. 2019), suggesting shift in adaptive strategy dominating these communities. Thus, metagenomic community aggregated traits (Fierer et al. 2014) may offer an in *situ* characterization of functional variation emerging from the main adaptations to environmental gradients with potential impact on ecosystem functioning. However, a global investigation of these functional variations in soil bacteria communities has yet to be conducted.

In this study we tested the hypothesis that the spectrum of soil bacteria community metagenomic profile is dominated by functional dimensions associated with bacteria community metabolism versatility (minimized in small genomes vs. maximized in large genomes) and metabolism reactivity (fast with high number of rRNA and sigma factor genes vs. slow with opposite trait), associated with environmental gradients capturing resource scarcity (favoring maximized versatility) and variability

(favoring fast reactivity). We crossed an exploratory approach mapping the whole metagenomes to the most common functional databases in bacteria biology (KEGG, SEED and eggNOG) with the characterization of genomic traits suggested to link with previously described physiological trait-based spectrums (Malik et al. 2019b, Westoby et al. 2021) and assess the main relationships between these genomic traits and the global metagenomic functional spectrum of soil bacteria communities. Then, we used machine learning to test how the position of bacterial communities along this functional spectrum are linked to environmental conditions related to climate and soil.

Materials and methods

Soil sampling and characteristics

In this study we used a global dataset of 127 soils distributed across continents and latitude (SI figure 1) collected by Bahram et al. (2018). We selected this dataset for our analysis because of its coverage and its use of a highly standardized protocol: 1) to sample top-soils in spatially independent sites across the globe selected to represent all the most important vegetation types; to analyze their chemistry and their metagenomes (Bahram et al. 2018). All samples were processed using similar standardized protocols for their chemistry (Carbon, Nitrogen, Phosphorus content and pH) and metagenome (See Bahram et al. (2018) for protocol details). We checked the global environmental coverage by comparing variation of the main environmental variables (Mean Annual Temperature (MAT), Mean Annual Precipitation (MAP), soil pH and Net Primary productivity (NPP)) in our dataset with global variation from the Atlas of the Biosphere (https://nelson.wisc.edu/sage/data-and-models/atlas/maps.php). This showed an almost complete global coverage, with only extreme Mean Annual Temperature (MAT) of very high latitude (below -11.33°C) and Sahelian Africa (above MAT 27.967°C) as well as very high pH (higher than 7.757) characterizing some part of North Africa, West Asia and Himalaya missing in our dataset (SI Figure 2). As far as we know, when we conducted this analysis, this dataset was the only available with such precise characterization of soil environment done on the same sample as shotgun metagenomic analysis, making this dataset the most robust for our objective to assess environmental drivers of metagenomic profiles. Nevertheless, potential to extend environmental range by adding all natural (Agricultural and contaminated soil excluded) soil metagenomes available (accession date January 28

2021) from the main sequence repositories MG-RAST (Meyer et al. 2008) and IMG:M (Chen et al. 2019) was also tested. This indicated that adding these data would not have extended environmental range and would have greatly decreased precision of soil properties characterization (SI Figure 3).

Metagenomic data

DNA extraction, sequencing, trimming and mapping approaches are detailed in Bahram et al. (2018). In this study, four community aggregated traits databases were built, corresponding to metagenomic reads mapping on different functional annotation systems by Bahram et al. (2018). An additional database was made for this study with genomic traits associated with previously described functional spectrum (See details below).

Characterization of the global metagenomic functional profiles of bacteria communities

Bahram et al. (2018) mapped reads on three commonly used functional databases (KEGG, eggNOG and CAZy). Data were aggregated at the pathway (KEGG), the functional categories (eggNOG) levels and KEGG annotation were used to produce SEED functional modules. CAZy outputs were read annotations to gene families of Glycolysis Hydrolases (GH) and Auxiliary Activities (AA). All reads mapping was done competitively against both prokaryotic and eukaryotic functional databases and best bit score in the alignment and the taxonomic annotation was used to retrieve only reads annotated as bacteria.

In this study, we used output data from these four annotation processes to provide complementary classification of functional genes (e.g. eggNOG classes include Motility, Cell envelopes and Defense which are not included in SEED whereas SEED classes include Dormancy and Sporulation, Stress response, Virulence, Carbon, Nitrogen and Phosphorus metabolism which are not included in eggNOG). The eggNOG annotation also differed from KEGG and SEED in the construction of orthologous groups with eggNOG using non-supervised construction increasing coverage whereas KEGG used supervised construction increasing annotation robustness. To obtain a more precise picture of C acquisition strategy, the CAZy annotates reads abundance were aggregated on the basis of their targeted substrates (Cellulose, Chitin, Glucan, Lignin, Peptidoglycan, Starch/Glycogen, Xylan, Other Animal Polysaccharides, Other Plant Polysaccharides, Oligosaccharides) using a curated database (SI Table 1)

based on previous works (Berlemont and Martiny 2015, Nguyen et al. 2018, López-Mondéjar et al. 2020). After mapping, for each database, the relative abundance of each gene (or aggregated group of genes) was calculated by dividing for each sample the number of reads annotated to the gene (or functional class) by the total number of bacteria-reads annotated for this sample on the same database. Such normalization corrects for variation between samples in the quantity of annotated reads and avoids biases induced by contamination and sequencing error (Nayfach and Pollard 2016). It involves that the sum of all relative abundances in a sample equal to 1 and thus, the obtained relative abundances for each gene/functional class inform on how much they are important relative to all the other known functions. This perfectly fits with the aim of our analysis, that is to obtain a signal of trade-off between different functions.

Characterization of genomic traits associated with previously described bacteria functional spectrum

Five genomic traits associated with the previously described bacteria spectrum were calculated in this study to relate our observation with previous observational and theoretical works related to soil bacteria spectrums (Fierer 2017, Malik et al. 2019a, Westoby et al. 2021). Average C acquisition enzyme gene copy number was calculated as an indicator of resource acquisition based on the production of extracellular enzymes to depolymerize organic molecules, a central trait in soil bacteria strategy to make resource available (Sinsabaugh and Moorhead 1994, Allison et al. 2010, Malik et al. 2019a). We calculated this genomic trait by normalizing the relative abundance of the sum of GH and AA family (from mapping against CAZy database) by the relative abundance of rpoB gene (COG0085), a well-known single copy core gene (Case et al. 2017). Sigma factor copy number was calculated to capture adaptation to sense and react to intermittent stress (Malik et al. 2019a) as their first role is to regulate gene expression in response to a large array of environmental constraints (osmotic, oxidative, nutritional, temperature) (Paget 2015, Helmann 2002). This genomic trait was calculated by normalizing the the sum of bacteria-COGs encoding for sigma factor (COG0568 for σD, σS and σH COG1191 for σF, σB, COG1508 for σN and COG1595 for extracytoplasmic function (ECF) sigma factors (Chávez et al. 2020) by a single copy core gene (rpoB), as for C-acquisition trait. Normalization to get average gene copy number per genome (as opposed to gene relative abundance) was chosen for C acquisition and Sigma

159 factor traits to make our genomic traits measurement reproducible and our values comparable with future
160 studies. Nevertheless, very similar trends were observed when using both types of normalization (gene
161 copy number per genome vs. relative abundance in the metagenome). For both C acquisition enzyme
162 and sigma factor, we also assessed the correlation between each enzyme class/COG and the final strategy
163 indicator (C acquisition enzyme and sigma factor) to assess their relative importance in the final value
164 of these genomic traits. Three additional CATs (genome size, rRNA copy number and GC content) was
165 calculated with tools using all sequences as input, after a verification that eukaryotic sequences were
166 negligible (less than 2% of annotated reads for all databases verified for all samples) and therefore the
167 samples mostly captured bacteria. GC content was simply calculated using sequence quality check tool
168 and use as an indicator of stress resistance as it has been associated with adaptation to high rate of DNA
169 damage in the form of double-strand breaks under stationary phase induced by stress like soil desiccation
170 (Weissman et al. 2019). Average genome size was measured using MicrobeCensus (Nayfach and Pollard
171 2015) to relate with the versatility dimension described by Westoby et al. (2020). Average rRNA copy
172 number was measured as described in Pereira-Flores et al. (2019) to relate with the fast-slow maximum
173 growth rate dimension described by Westoby et al. (2021).

174 *Statistical analysis*

175 To identify the multivariate axes that best explain the global scale variation in metagenomic community
176 aggregated traits of soil bacteria, we used a multitable co-inertia analysis (MCOA), an exploratory
177 analysis that leverages together the information from our 5 databases (genomic traits, eggNOG
178 categories, SEED modules, KEGG pathway, CAZy types). This method identifies co-relationships
179 between the different databases and uses a covariance optimization criterion to summarize in a common
180 structure the information shared by multiple multivariate (eg. omic) tables (Chessel and Hanafi 1996,
181 Meng et al. 2014, Piton et al. 2020). All variables (CATs) were log transformed (log X +1) before the
182 analysis to improve normality (Meng et al. 2014) and standardized to a mean of zero and a variance of
183 1.

184 Sample coordinates on the first and second dimension of the MCOA were extracted and used as latent
185 variables representing bacterial community positions in the global functional spectrum. First, linear

correlation between bacterial community positions along spectrum dimensions and the relative abundances of bacteria phyla (and classes for Proteobacteria) in the metagenome was assessed. Secondly, random forest models were used to identify predictors of these coordinates among potential environmental drivers (soil properties measured on the same sample as metagenome (see Soil sampling and characteristics) and climatic variables extracted from Worldclim2 (BIO1 = Annual Mean Temperature, BIO4 = Temperature Seasonality (standard deviation), BIO12 = Annual Precipitation and BIO15 = Precipitation Seasonality (standard deviation)) based on sample geographical coordinates. First, we verified that all selected environmental drivers had spearman correlation coefficients lower than 0.7 to mitigate collinearity problems as recommended in Dormann et al. (2013). Second, a variable selection process was carried out using the method implemented in VSURF R package (Genuer et al. 2015). The number of predictors randomly tested at each node of the random forest tree (mtry) was optimized based on randomForest's tuneRF algorithm and the number of trees set to 1000. Third, the random forest models selected following the VSURF selection process were trained using ten-fold cross-validation (100 repetitions) implemented in the caret package (Kuhn 2008) and model performance were then assessed based on Root Mean Square Error (RMSE) and R squared. Finally, random forest predictive models were used to project a broad resolution map of the functional spectrum global biogeography using environmental maps (1600x1200 pixel) as predictors. For this projection, global soil properties maps (https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=569 and data from Yang et al. (2014) were obtained from ORNL DAAC (http://daac.ornl.gov) NASA data center. Worldclim2 (https://www.worldclim.org/) was used for climatic variables. To validate the relevance of this broad resolution map to represent average local values, we tested the correlation between local observations and the predicted value of the cell in which the local observation was done.

Results

The global functional spectrum of soil bacteria communities

A two-dimensional functional spectrum emerged from the co-variations of metagenomic community aggregated traits. Using a multi-table co-inertia analysis (MCOA), the first two dimensions captured 54% of the global soil bacteria genomic variation (Figure 1). The first-dimension (MCOA1, 32%) was

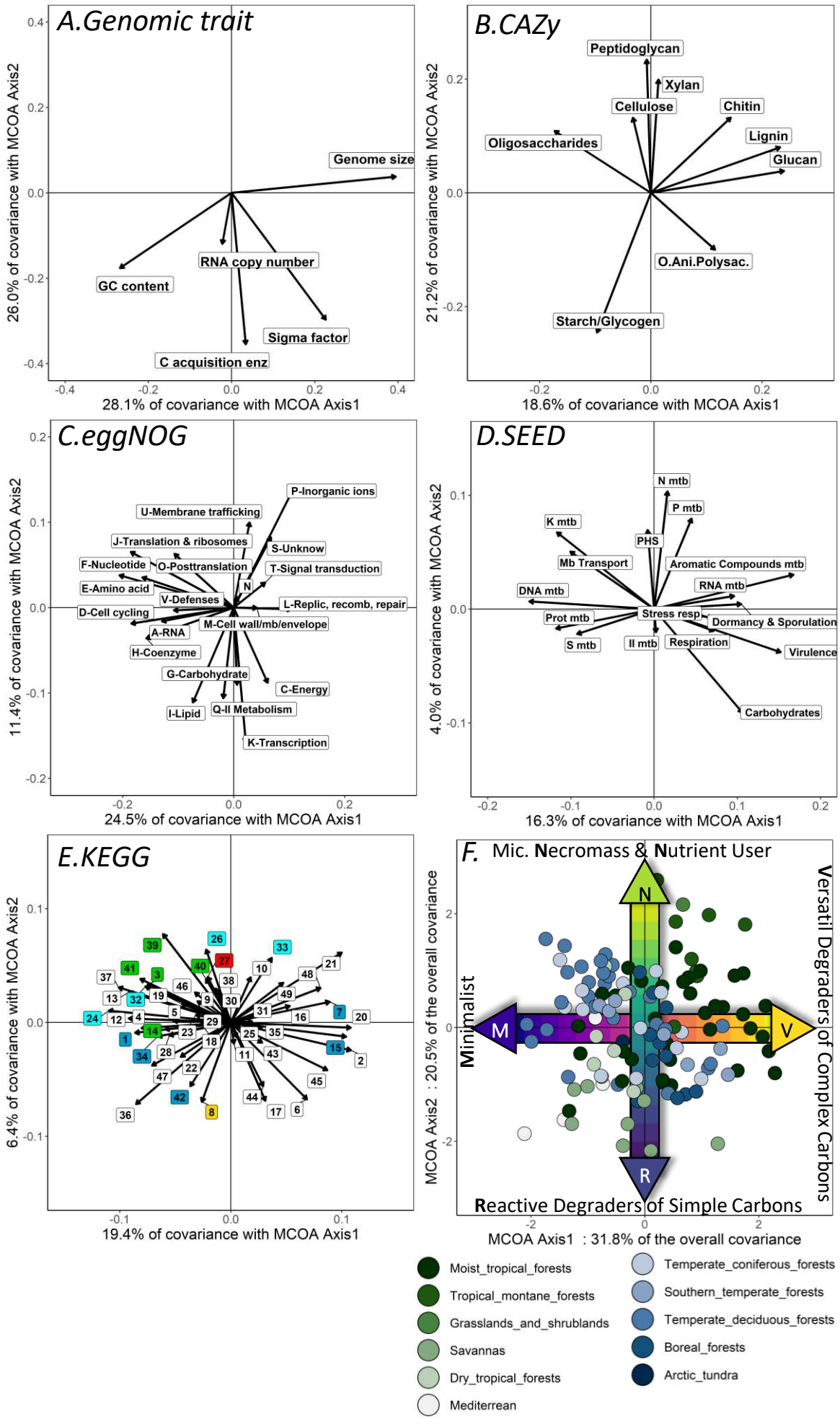
associated with genome size (Figure 1A, Supplementary Figure 4, R^2 of 0.75 for the correlation of MCOA1 with genome size). At one extreme, bacterial communities had small genome encoding for basic growth functions and C acquisition machinery, and were defined as the minimalist (M) profile. On the other extreme, bacterial communities had large genomes with more complex metabolism and resource acquisition strategy, and were defined as the versatile degraders of complex carbons (V) profile. More precisely, metagenomic mapping on the functional databases (eggNOG, KEGG, SEED, CAZy) showed that bacteria community with smaller genomes (M profile) had a higher proportion of gene categories associated with growth metabolism in their annotated genomes (Figure 1, eggNOG: F-Nucleotide, H-Coenzyme and E-Amino acid transport and metabolism, D-Cell cycling, J-ribosomes and translation. SEED: DNA and protein metabolisms. KEEG: 4-Arginine & Proline, 13-Cysteine & Methionine, 37-Purine, 41-Ribosome, 12-cofactors and vitamins metabolisms, 36-Proteosome). In these communities, carbon acquisition enzymes involved in last step of organic carbon depolymerization directly providing assimilable carbon from oligosaccharides (eg. beta-glucosidases encoded in GH1, GH2 and GH3 CAZy family dominates oligosaccharides degradation enzyme class) were favored over enzymes targeting polysaccharides. On the opposite end of this first dimension, the more complex metabolism and resource acquisition strategy of the large genome C was characterized by an enrichment in accessory genes like genes associated with lignin degrading enzymes (CAZy class), metabolism of derived aromatic compounds (SEED: Metabolism of Aromatic Compounds, KEGG: 6-Aromatics degradation), fungal cell wall degrading enzymes (Chitin and Glucan), as well as virulence genes. These communities were also characterized by higher proportion of genes associated with resistance to antimicrobial compounds or other stresses like membrane and DNA repair related genes (eggNOG: L-Replication, recombination & repairs, KEGG: 20-Lipid and 21-lipopolysaccharide metabolism) and genes to sense environment and regulate metabolism (sigma factor copy number, KEGG : 48-two component regulatory system). Finally, a large genome was also associated with higher relative abundance of energy production related genes (KEGG: 2-ATP synthesis). Thus, covariation patterns along the MCOA1 captured functional variation associated with metabolism and resource acquisition complexity tightly associated with genome size.

The second dimension (MCOA2) of the bacterial functional spectrum (Figure 1, 21% of the total variation) separated communities according to their metabolism stoichiometry (C- vs. Nutrient-oriented) and their C acquisition enzymes, sigma factors, and rRNA gene copy number (Figure 1, Supplementary Figure 4). On one side of this dimension, bacteria communities showed a specialization of the genomic profile for acquiring and metabolizing energy rich C and for regulating activity, and were defined as the reactive degraders of simple carbons (R) side of the spectrum. On the other side of MCOA2, bacteria showed a nutrient oriented metabolism relying on transporters and microbial biomass recycling enzymes, and was defined as the Necromass N-oriented usage side of the spectrum (N). More precisely, C acquisition enzymes gene abundance increased at the R side of the spectrum (Figure 1, Supplementary Figure 4), mostly driven by an increase of CAZy genes encoding for simple substrate targeting Starch/Glycogen and Oligosaccharides classes representing from 57% to 73% of the C acquisition genes. In parallel, sigma factor gene abundance increased (Figure 1, Supplementary Figure 4), firstly driven by extracytoplasmic function (ECF) sigma factors representing from 42 to 61% of sigma factor genes. Associated with higher C acquisition and sigma gene relative abundances, the R side of the second dimension showed a specialization to C metabolism (Figure 1, eggNOG: G-Carbohydrates, SEED: Carbohydrate, KEGG: 45-Sugar metabolism) to process energy (eggNOG : C-Energy production and conservation) as well as higher relative abundance of gene categories associated with capacity to regulate activity and build/repair membrane in front of environmental fluctuations (eggNOG: K-Transcription and I-Lipid, KEEG: 17-Fatty acid metabolism, 44-Sterol biosynthesis) and secondary metabolism (eggNOG : Q-Secondary metabolites biosynthesis, transport and catabolism, KEGG : 8-Biosynthesis of secondary metabolites). At the opposite end of MCOA2, the N profile showed clear specialization for metabolism and transport of mineral and organic nutrients (eggNOG: P-Inorganic ion transport and metabolism, U-Membrane trafficking, SEED: Membrane Transport, Nitrogen and Phosphorus Metabolism. KEGG: transport systems of 26-Mineral & organic ion, 33-Phosphate and amino acid, 32-Peptide and nickel) and their CAZy profile indicated a specialization for acquisition of bacterial necromass compounds (Peptidoglycan) but also to a lower extent fungal biomass (Chitin), cellulose and Xylan. Thus, the second dimension of the spectrum captured variation of metabolism orientation (stoichiometry) and regulation independent of genome size.

While genome size, C acquisition gene and RNA copy number were tightly linked with only one on of the spectrum, it is noteworthy that GC content and sigma factor showed association with the two dimensions (Supplementary Figure 4). Genome GC content showed a negative association with both dimensions indicating a more GC rich genome when bacteria communities get closer to the small genome Minimalist (M) end of the spectrum (R^2=0.36) or to a lower extent to the Reactive C-degraders (R) end (R^2=0.16). Sigma factor gene abundance was also significantly associated with the two dimensions (Supplementary Figure 4), especially with the Reactive C-degraders end of the MCOA2 (R^2=0.44) but also with the large genome Versatile (V) end of MCOA1 (R^2=0.26). Thus, these genomic traits are able to capture functional variation along one or two dimensions of the spectrum.

Relationships between bacteria community composition and the global functional spectrum

We next assessed the relationships between the two MCOA dimensions and phyla (or class for Proteobacteria) relative abundances in the metagenome. Taxa positively correlated with the small genome Minimalist end of the first dimension were Bacteroidetes (R^2=0.35), Actinobacteria (R^2=0.32) and Chloroflexi (R^2=0.08) whereas taxa positively associated with large genome Versatile end were Acidobacteria (R^2=0.45), Planctomycetes (R^2=O.21) and Cyanobacteria (R^2=0.16) and the Alpha-Proteobacteria (R^2=0.72) and Gamma-Proteobacteria (R^2=0.38) classes. Taxa positively associated with the Reactive C-degraders end of the second dimension were Actinobacteria (R^2=0.40), Firmicutes (R^2=0.09) whereas phyla associated with the microbial necromass and nutrient user (N) end were Bacteroidetes (R^2=0.20) and the Beta- (R^2=0.05) and Delta-Proteobacteria (R^2=0.08) classes. Thus, variation of the functional profile covaried with shift in bacteria community composition at the phylum level.


A.Genomic trait
Genome size
RNA copy number
GC content
Sigma factor
C acquisition enz
26.0% of covariance with MCOA Axis2
28.1% of covariance with MCOA Axis1
B.CAZy
Peptidoglycan
Xylan
Cellulose
Chitin
Lignin
Glucan
Oligosaccharides
O.Ani.Polysac.
Starch/Glycogen
21.2% of covariance with MCOA Axis2
18.6% of covariance with MCOA Axis1
C.eggNOG
P-Inorganic ions
U-Membrane trafficking
J-Translation & ribosomes
S-Unknow
F-Nucleotide
O-Posttranslation
T-Signal transduction
E-Amino acid
N
V-Defenses
L-Replic, recomb, repair
D-Cell cycling
M-Cell wall/mb/envelope
A-RNA
H-Coenzyme
G-Carbohydrate
C-Energy
I-Lipid
Q-II Metabolism
K-Transcription
11.4% of covariance with MCOA Axis2
24.5% of covariance with MCOA Axis1
D.SEED
N mtb
P mtb
K mtb
PHS
Mb Transport
Aromatic Compounds mtb
RNA mtb
DNA mtb
Prot mtb
Stress resp
Dormancy & Sporulation
S mtb
II mtb
Respiration
Virulence
Carbohydrates
4.0% of covariance with MCOA Axis2
16.3% of covariance with MCOA Axis1
E.KEGG
6.4% of covariance with MCOA Axis2
19.4% of covariance with MCOA Axis1
F. Mic. Necromass & Nutrient User
N
M
V
R
Minimalist
Versatil Degraders of Complex Carbons
Reactive Degraders of Simple Carbons
MCOA Axis2 : 20.5% of the overall covariance
MCOA Axis1 : 31.8% of the overall covariance
Moist_tropical_forests
Tropical_montane_forests
Grasslands_and_shrublands
Savannas
Dry_tropical_forests
Mediterrean
Temperate_coniferous_forests
Southern_temperate_forests
Temperate_deciduous_forests
Boreal_forests
Arctic_tundra

290 *Figure 1. Global functional spectrum of soil bacteria metagenomes represented by a multiple co-inertia*
291 *analysis (MCOA) summarizing in a common structure (MCOA Axes 1 and 2) the information shared by*
292 *5 community aggregated trait (CAT) databases. Variable contribution (Arrow-length) to the common*
293 *structure for the 5 CAT databases: genomic trait (A), CAZy (B), eggNOG (C), SEED (D), and KEGG*
294 *(E, with light blue labels for nutrient transport, dark blue for other transport KEGG categories, green*
295 *label for categories associated with translation and ribosomes, red label for nitrogen metabolism,*
296 *yellow label for secondary metabolism, white label for other metabolic categories, see SI Table3 and*
297 *text for KEGG pathways corresponding to numbers on the plot). F plot: sample coordinates (small dots)*
298 *on the common structure and double arrows representing the color scale used in Figure 2 to represent*
299 *bacterial community position along the first and second dimension of the MCOA. Letters at each extreme*
300 *of these dimensions represent extreme profiles (see text for details).*

301

302 *Environmental drivers and geographic distribution of the global functional spectrum*

303 We next assessed how environmental gradients separated bacterial communities across the global
304 functional spectrum. We used metagenome coordinates along the dimension 1 and 2 of the MCOA to
305 represent variation across this spectrum using a random forest model for each dimension. Random
306 forests based on environmental factors were able to predict 78% and 45% of the variation along the
307 dimension 1 and 2 respectively (Figure 2A). These models were used to project a global map describing
308 predicted position of the bacteria communities along dimension 1 and 2 based on broad resolution maps
309 of soil and climate variables (Figure 2. C). Predicted values in these broad cells, representing averaged
310 conditions across a large surface, showed high consistency with values observed locally in our samples
311 ($p<0.001$), with a R^2 of 0.40 and 0.45 for dimensions 1 and 2 respectively (Supplementary Figure 5).
312 Thus, the global functional spectrum was tightly associated with environmental factors.

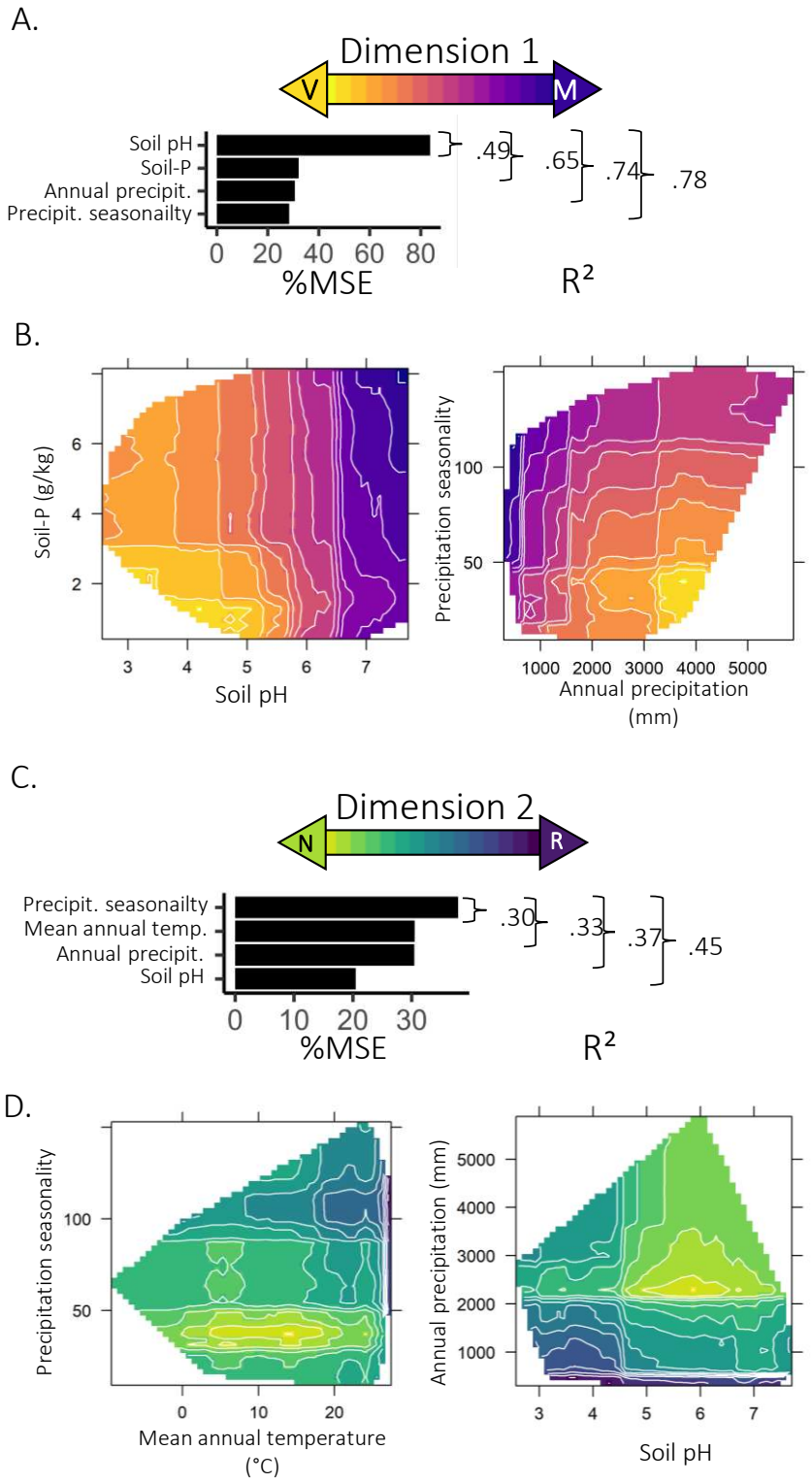
313 More precisely, bacterial community functional profiles (coordinates along the two dimensions of the
314 functional spectrum) shifts along environmental gradients were mainly progressive along the two
315 dimensions, while we observed a few threshold effects indicating clear shifts from one extreme profile
316 to another. Bacterial communities shifted toward small genome Minimalist functional profile (low

317 coordinate on Dimension 1, Figure 1), under neutral to alkaline pH, high soil-P, low annual precipitation
318 but high precipitation seasonality (Figure 2 B). Conversely, the large genome Versatile functional profile
319 (high coordinate on Dimension 1, Figure 1), was favored when pH ran below 5.6, especially when soil-
320 P content was low and when annual precipitations were high and stable (Figure 2B).

321 These environmental controls predicted Minimalist M profile especially under arid climate concentrated
322 under dry tropical and subtropical latitude as well as in central Asia, Mongolia, and northwest China
323 (Figure 3A). Conversely, the Versatile profile was clearly representative of the bacteria communities
324 from equatorial forests (Figures 3A) but was also predicted in some temperate zones in northern Europe,
325 western Canada, New Zealand and south Chile combining high precipitation and low pH. Intermediate
326 coordinates indicating a profile with medium genome size between extreme M and V profile were
327 predicted for most of the other pedoclimatic zones (subequatorial, temperate and high latitude zones)
328 covering most of the global surface.

329 Along the second functional dimension, a clear shift to the Reactive C-degraders profile (low coordinate
330 on Dimension 2, Figure 1) was observed when conditions ran below a threshold of pH 4.6 (Figure 2D)
331 as well as when mean annual temperature (MAT) approached its low and high extremes and when annual
332 precipitation was low. Above the threshold of pH 4.6, bacteria communities showed an intermediate
333 position between the two extremes of this dimension but got closer to the R profile when precipitation
334 seasonality increased (large coefficient of variation). On the other hand, bacteria communities shift to
335 a more N profile (high coordinate on Dimension 2, Figure 1) when soil pH is above 4.6 and climate
336 conditions are stable and not extreme (Figure 2D).

337 Extreme R profiles were observed under the most aridic subtropical zone of Africa, Australia and North
338 Mexico (Figure 3B) whereas metagenomes shift toward extreme N profile were only observed under
339 cold temperate zone of Europe, Asia and America but also under some equatorial zones (Southeast Asia
340 and Central America), where precipitation is stable across the year (low seasonality) and soil pH not too
341 acid (> 4.6). As for the first dimension of the global bacterial spectrum, it is noteworthy that most of the
342 global surface is characterized by pedoclimatic conditions predicted to favor an intermediate profile
343 between R and N.



344

345 *Figure 2. Variable importance (mean decrease in mean square error (%MSE) and R squared estimated*

346 *using ten-fold cross validation) from most parsimonious random forest models (A). Partial dependence*

347 *plots from most parsimonious random forest (B).*

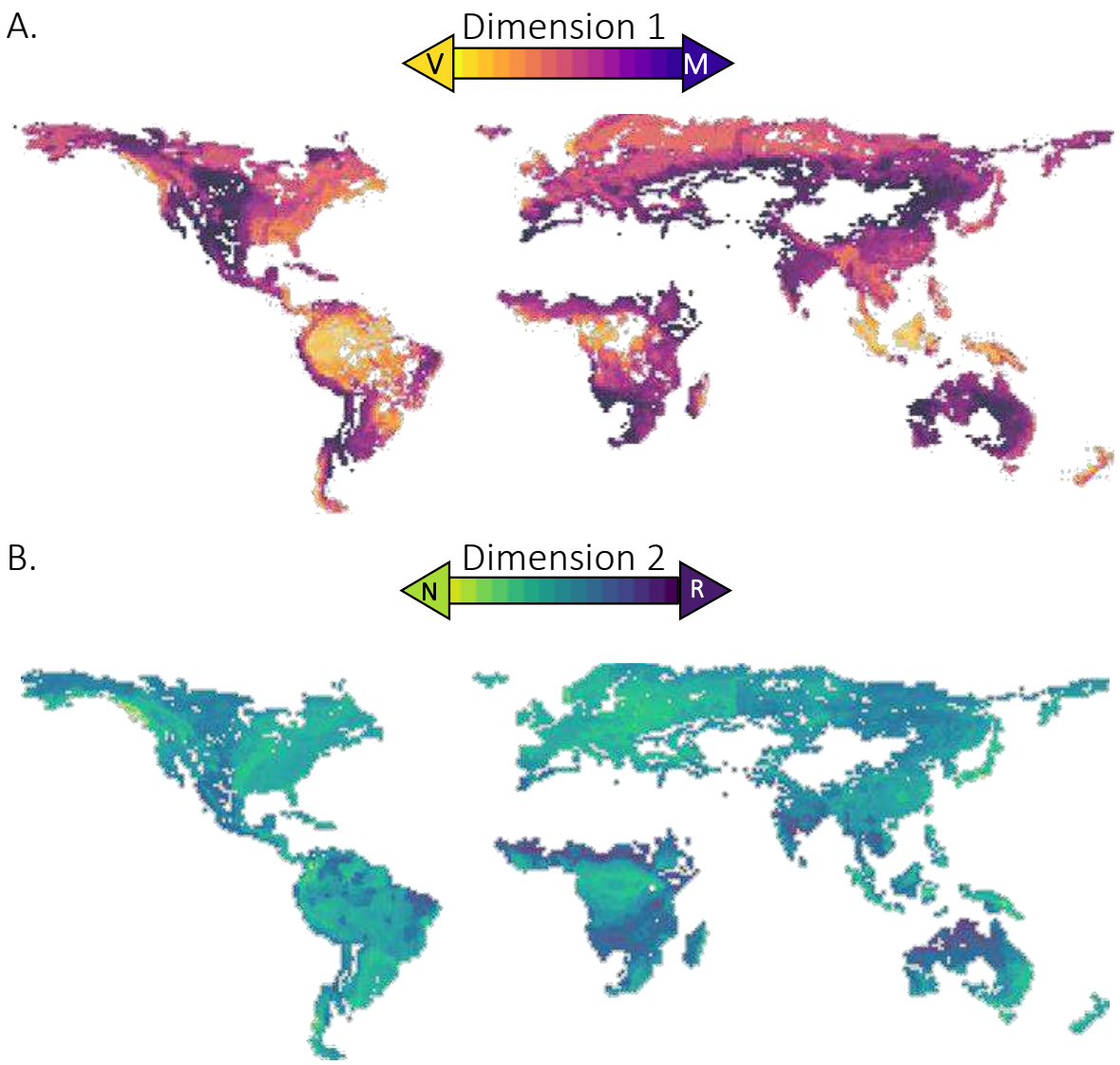


Figure 3. Predicted global distribution of the bacteria spectrum dimension 1 (A) and 2 (B) based on broad resolution of mean soil and climate conditions across the globe, with area out of the dataset range excluded.

Discussion

Our results support our hypothesis that the spectrum of soil bacteria community metagenomic profile is dominated by a first functional dimension associated with bacteria community metabolism versatility and a second dimension capturing metabolism reactivity. Our community aggregated trait spectrum is

partially consistent with the spectrum across cultured bacterial taxa described by Westoby et al. (2021) describing one dimension associated with versatility (captured by genome size) and one dimension associated with a slow vs fast growth rate (captured by rRNA copies). This suggests that the functional tradeoffs in bacterial tree of life, captured by genomic traits, play a central role in bacteria adaptation and community organization along environmental gradients and scale up to the level of the bacterial community functional profiles. Indeed, ranges of genome size (median=6.8Mb, minimum=5.2Mb and maximum=10.3Mb) observed in our community level spectrum matched with patterns reported for soil bacteria isolates by Westoby et al. (2021). This equivalent range and the fact that both studies indicate genome size to capture the first functional dimension suggest a tight connection between taxa and community level spectrums. However, on another hand, rRNA copy variation observed in our study is highly constrained between 1 and 1.5 copies, a range consistent with the study of Gao and Wu (2018), recently reporting that the large majority of soil bacteria have less than 2 rRNA copies, whereas bacteria from other environments can have up to 15 copies. Hence, global dominance of bacteria carrying few copies in soil drive our small range of community level aggregated values, but this does not exclude that some bacteria with more copies might be present in the soil community (eg. Li et al. 2019). This constraint range of rRNA copies for community aggregated traits might explain that Westoby et al. (2021) showed rRNA copy tightly related with the second slow-fast dimension of their taxa level spectrum, whereas our community level study showed other genomic traits (eg. C acquisition enzymes and sigma factor) better capturing a second dimension associated with reactivity rate. Indeed, under the oligotrophic condition of soil, the reactivity of bacteria to new conditions might be more associated with their ability to sense environmental stimuli and regulate their activity and to produce resource acquisition enzymes than just to grow rapidly. Overall, our results describe a two dimensional spectrum at the community levels, consistent with the idea that bacteria versatility and reactivity dominate their bacterial functional spectrum in soil.

We next detailed the functional profiles of bacteria communities across this versatility and reactivity dimensions using metagenome mapping on general functional databases and assess environmental conditions driving these different profiles. At the small genome end of the first dimension, bacteria

showed a genome streamlined to the basic growth and resource acquisition functions (DNA and protein metabolism, oligosaccharide acquisition), consistent with patterns previously described across free-living bacteria genomes (Konstantinidis and Tiedje 2003). Such a minimal versatility profile suggests a high streamlining pressure and/or low fitness benefit to gain new capability (Guieysse and Wuertz 2012). pH near neutrality and high soil P content combined with challenging precipitation patterns (low precipitation and high seasonality) select this small genome profile. We suggest that survival constraint of their environment due to precipitation pattern might amplify genome streamlining for economic reasons while low abiotic constraint on physiology (neutral pH and high P content) might limit fitness gain associated with gene addition. Consistent with adaptation to stress, this small genome profile showed higher GC content, as observed by Chuckran et al. (2021), which might represent adaptation to minimize damage during stressing stationary phases (Weissman et al. 2019) and C cost (Hellweger et al. 2018). Moving along the first dimension, bacteria communities run to more complex, larger genomes, carrying the machinery in the genome to access resources with low return on investment (eg. Complex carbons like lignin and other plant polysaccharides). Selection of such energetically costly strategy suggest that high return on investment resource (eg. oligosaccharides) are not directly available because of low provisioning or intense competition. Increased versatility associated with this large genome size (Westoby et al. 2021), combined with metabolism regulation capacity (sigma factor genes) might also be beneficial for competitiveness by enabling bacteria to shift between different resources when resources become scarce. Consistent with an intense competition, this profile is selected under stable and wet climate conditions that involve low constraint on survival traits, making competition to access nutritional resource the main selecting force (Grime et al. 1977). This transition from small to large genome profile is also clearly linked with low soil pH, with a steep transition to large genomes observed especially between pH 6.6 and 5.6. This tight association might be directly linked with the difference in pH optimum of soil bacteria that are known to be narrow (Jones et al. 2021), with adaptation to low pH involving a metabolic specialization requiring gene addition. This association might also be linked with the numerous soil characteristics directly or indirectly linked with pH (eg. nutrient availability, aluminum solubility, plant communities and organic carbon inputs). Consistent with this idea, the threshold of 5.6 below which V profile clearly dominates is consistent with the breaking point at which

toxic Al3+ ion becomes soluble (5.5) with demonstrated consequences for microbial physiology (Jones et al. 2019). Indeed, such acidic conditions are challenging for bacteria who require molecular costly adaptation to maintain cytoplasm homeostasis and cope with Aluminum toxicity (Fernández-Calviño et al. 2011, Auger et al. 2013, Jones et al. 2019). The enrichment of their genome with genes encoding for energy production, DNA and membrane repair and ion transporters that we observed might represent an adaptation to these acidic conditions. Overall, our results indicate that genome size and associated metabolic complexity (basic growth function vs. more complex metabolism and resource acquisition increasing versatility) are central components of global functional variation of soil bacteria communities and are mainly associated with soil pH and precipitation patterns.

The second dimension of our spectrum depicts functional variation associated with metabolism reactivity independent of genome size and tightly linked with different resource acquisition strategy and metabolism stoichiometry (ie. C- vs. Nutrient-oriented). It runs from a C-oriented and regulated R profile to nutrient-oriented bacteria relying on transporters and microbial necromass recycling (N profile). More precisely, when pH is low (below a threshold of 4.6 for which we don't have explanation) and when climatic patterns restrict growing window (extreme temperatures and low precipitations with high seasonality), R bacteria combine capacity to sense condition and to grow and acquire C rapidly during favorable periods (eg. sigma factor, energy rich C usage, and high rRNA copies associated with faster growth after resource pulse (Li et al. 2019)). Running to the other side of this dimension, bacteria profiles N depict a transporter based resource acquisition profile (Malik et al. 2019a) specialized in acquiring nutrients and in recycling microbial Nitrogen-rich necromass (ie. peptidoglycan and chitin), two attributes expected to optimize competitive resource use efficiency (Malik et al. 2019a, Krause et al. 2014, Malik et al. 2019a). Their environment (neutral pH, medium temperature and precipitation and low seasonality) are favorable for resource acquisition (Sinsabaugh and Shah 2012), biomass turnover and yield (Zheng et al. 2019, Buckeridge et al. 2020) likely making of this transporter based resource acquisition and biomass recycling an efficient strategy. Overall, these results show that metabolism regulation to react rapidly, resource acquisition (depolymerization vs transporter and recycling based) and metabolism stoichiometry (Carbon vs nutrient oriented) are structuring a second dimension in global

soil bacterial communities' functional spectrum along a gradient from low to high pH and climatic constraints.

This community level spectrum across environmental gradients can contain a signal of the strategy variation across organisms dominating these communities (eg. for plants in Bruelheide et al. 2018). Here we discuss how our community level spectrum might capture such signals for soil bacteria relying on the Competitive-Stress Tolerant-Ruderal frameworks (Grime 1977, Krause 2014, Fierer 2017). Functional attributes of the Versatile degraders of complex carbons described in this study match with several features previously proposed for competitor strategies like large genome and high catabolic diversity (Fierer et al. 2017) and their stable environment is theoretically expected to select for competitiveness (Grime et al. 1977) suggesting that dominance of such strategy might have imprinted the community V profile. Minimalist bacteria community profile does not especially match with traits previously proposed to represent stress tolerant strategies (Westoby et al. 2021, Fierer et al. 2017) but their environment is likely constrain their physiology (water stress and C-limitation), inviting for more investigation on the role of genome simplification in stress tolerance (eg. Weissman et al. 2019, Simonsen et al. (2022)). Reactive degraders of simple carbons (ie. R profile) described here shared several genomic features proposed for ruderal strategy like responsiveness and rRNA copy number (Krause et al. 2014, Fierer et al. 2017) and also environmental properties expected to select this strategy (high precipitation seasonality). Finally, the microbial necromass and nutrient user N profile showed genomic features that have been proposed to be associated with the transporter based resource acquisition strategy of Malik et al. (2019), which can be seen as an alternative competitor strategy different from depolymerization and versatility based competitor profile. The microbial necromass recycling genomic signature of this N profile also invites more investigation of this function in bacteria strategy. Overall, our community spectrum brings some new elements in genomic features potentially playing a key role in bacteria adaptation and community organization across global soil gradients.

This study provides for the first time a global picture of the soil bacteria community functional spectrum but also highlights three challenges. First, the genomic spectrum relies on sequences successfully assigned on different databases and we observed that mapping coverage on eggNOG, KEGG and SEED

databases decreases along the first dimension of our spectrum running from small simple genome to more complex large genome (Sup Figure 6). This highlights that genes associated with large and more complex genomes, especially characteristics of equatorial forest, are less represented in databases and need more investigation. We can expect that increasing representation of these complex genomes in databases might accentuate even more the trend from basic to complex metabolism captured in our spectrum as genes associated with basic metabolism are likely the best represented in current databases. Second, it is noteworthy that environmental factors identified here as the main drivers of the global spectrum (ie. pH and precipitation pattern) also affect plant community functioning globally (eg. Le Bagousse-Pinguet et al. 2017, Bruelheide et al. 2018). The central role of plant compounds degradation enzymes in structuring our global spectrum calls for deeper investigation of the links with plant community properties to fully elucidate how bacterial community functional variations are shaped worldwide. Third, it is important to note that our global projection (Figure 3) aims at giving a picture of the general biogeographic patterns in bacteria community profiles but values predicted for these broad cells can be dissociated from the local situation of soils far from average regional (cell) condition (Sup Figure 5). Indeed, soil pH and moisture not only vary globally but can also greatly vary at local scale because of bedrock or land use history variations (eg. Malik al. 2018). Thus, transposition of our global scale spectrum at local scale using our global projection needs to be considered with caution. Despite these challenges, our study demonstrates how metagenomic approaches can provide significant advance in our understanding of microbial communities functioning.

Acknowledgements

We thank Leho Tedersoo and Peer Bork who conceived and supervised the acquisition of the global dataset used in this study with Mohammad Bahram and Falk Hildebrand. We also wish to thank all their collaborators who contribute to this global data acquisition effort. We also thank Alyse Larkin and Lucas Ustick for their guidance in the bioinformatic analysis conducted in this study. This work was supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research grants DE-SC0016410 and DE-SC0020382. FH was supported by the European Research Council H2020 StG (erc-stg-948219, EPYC) and the Biotechnology and Biological Sciences research Council

(BBSrC) Institute Strategic Program Gut Microbes and Health BB/r012490/1 and its constituent project BBS/e/F/000Pr10355.

Authors contributions

Data collection was designed and supervised by M.B. Initial bioinformatics analysis to obtain functional genes abundance tables (eggNOG, KEGG, SEED, CAZy) was designed and performed by F.H. Idea of this new analysis was conceived by G.P. with inputs from A.M., S.A, J.M. and K.T. New quantification of genomic traits and data analysis was performed by G.P. First draft and following editing was conducted by G.P. with inputs from all co-authors.

References

Ackerly, D. D. 2003. Community assembly, niche conservatism, and adaptive evolution in changing environments. International Journal of Plant Sciences 164:S165–S184.

Auger, C., S. Han, V. P. Appanna, S. C. Thomas, G. Ulibarri, and V. D. Appanna. 2013. Metabolic reengineering invoked by microbial systems to decontaminate aluminum: implications for bioremediation technologies. Biotechnology advances 31:266–273.

Bahram, M., F. Hildebrand, S. K. Forslund, J. L. Anderson, N. A. Soudzilovskaia, P. M. Bodegom, J. Bengtsson-Palme, S. Anslan, L. P. Coelho, H. Harend, and others. 2018. Structure and function of the global topsoil microbiome. Nature 560:233–237.

Berlemont, R., and A. C. Martiny. 2015. Genomic potential for polysaccharide deconstruction in bacteria. Appl. Environ. Microbiol. 81:1513–1519.

Bruelheide, H., J. Dengler, O. Purschke, J. Lenoir, B. Jiménez-Alfaro, S. M. Hennekens, Z. Botta-Dukát, M. Chytr, R. Field, F. Jansen, and others. 2018. Global trait–environment relationships of plant communities. Nature Ecology & Evolution 2:1906–1917.

Buckeridge, K. M., Mason, K. E., McNamara, N. P., Ostle, N., Puissant, J., Goodall, T., ... & Whitaker, J. (2020). Environmental and microbial controls on microbial necromass recycling, an important precursor for soil carbon stabilization. *Communications Earth & Environment*, *1*(1), 1-9.

518 Chávez, J., D. P. Devos, and E. Merino. 2020. Complementary tendencies in the use of regulatory
519 elements (transcription factors, sigma factors, and riboswitches) in bacteria and archaea. Journal of
520 bacteriology 203:413–20.

521 Chen, I.-M. A., K. Chu, K. Palaniappan, M. Pillay, A. Ratner, J. Huang, M. Huntemann, N. Varghese,
522 J. R. White, R. Seshadri, and others. 2019. IMG/M v. 5.0: an integrated data management and
523 comparative analysis system for microbial genomes and microbiomes. Nucleic acids research 47:D666–
524 D677.

525 Chessel, D., and M. Hanafi. 1996. Analyses de la co-inertie de K nuages de points. Revue de statistique
526 appliquée 44:35–60.

527 Chuckran, P. F., B. A. Hungate, E. Schwartz, and P. Dijkstra. 2021. Soil, ocean, hot spring, and host-
528 associated environments reveal unique selection pressures on genomic features of bacteria in microbial
529 communities. bioRxiv.

530 Cotrufo, M. F., M. D. Wallenstein, C. M. Boot, K. Denef, and E. Paul. 2013. The Microbial Efficiency-
531 Matrix Stabilization (MEMS) framework integrates plant litter decomposition with soil organic matter
532 stabilization: do labile plant inputs form stable soil organic matter? Global Change Biology 19:988–
533 995.

534 Crowther, T. W., J. Van den Hoogen, J. Wan, M. A. Mayes, A. Keiser, L. Mo, C. Averill, and D. S.
535 Maynard. 2019. The global soil community and its influence on biogeochemistry. Science
536 365:eaav0550.

537 Delgado-Baquerizo, M., A. M. Oliverio, T. E. Brewer, A. Benavent-González, D. J. Eldridge, R. D.
538 Bardgett, F. T. Maestre, B. K. Singh, and N. Fierer. 2018. A global atlas of the dominant bacteria found
539 in soil. Science 359:320–325.

540 Diaz, S., J. Kattge, J. H. Cornelissen, I. J. Wright, S. Lavorel, S. Dray, B. Reu, M. Kleyer, C. Wirth, I.
541 C. Prentice, and others. 2016. The global spectrum of plant form and function. Nature 529:167.

542 Diaz, S., S. Lavorel, F. de Bello, F. Quétier, K. Grigulis, and T. M. Robson. 2007. Incorporating plant
543 functional diversity effects in ecosystem service assessments. Proceedings of the National Academy of
544 Sciences 104:20684–20689.

545 Dormann, C. F., J. Elith, S. Bacher, C. Buchmann, G. Carl, G. Carré, J. R. G. Marquéz, B. Gruber, B.
546 Lafourcade, P. J. Leitao, and others. 2013. Collinearity: a review of methods to deal with it and a
547 simulation study evaluating their performance. Ecography 36:27–46.

548 Escalas, A., L. Hale, J. W. Voordeckers, Y. Yang, M. K. Firestone, L. Alvarez-Cohen, and J. Zhou.
549 2019. Microbial functional diversity: From concepts to applications. Ecology and evolution 9:12000–
550 12016.

551 Fernández-Calviño, D., J. Rousk, P. C. Brookes, and E. Bååth. 2011. Bacterial pH-optima for growth
552 track soil pH, but are higher than expected at low pH. Soil Biology and Biochemistry 43:1569–1575.

553 Fierer, N. 2017. Embracing the unknown: disentangling the complexities of the soil microbiome. Nature
554 Reviews Microbiology 15:579–590.

555 Fierer, N., A. Barberán, and D. C. Laughlin. 2014. Seeing the forest for the genes: using metagenomics
556 to infer the aggregated traits of microbial communities. Frontiers in microbiology 5.

557 Fierer, N., Lauber, C. L., Ramirez, K. S., Zaneveld, J., Bradford, M. A., & Knight, R. 2012. Comparative
558 metagenomic, phylogenetic and physiological analyses of soil microbial communities across nitrogen
559 gradients. *The ISME journal*, *6*(5), 1007-1017.

560 Fierer, N., M. A. Bradford, and R. B. Jackson. 2007. Toward an ecological classification of soil bacteria.
561 Ecology 88:1354–1364.
562
563 Gao, Y., and M. Wu. 2018. Free-living bacterial communities are mostly dominated by oligotrophs.
564 bioRxiv:350348.

565 Garnier, E., J. Cortez, G. Billès, M.-L. Navas, C. Roumet, M. Debussche, G. Laurent, A. Blanchard, D.
566 Aubry, A. Bellmann, and others. 2004. Plant functional markers capture ecosystem properties during
567 secondary succession. Ecology 85:2630–2637.

568 Genuer, R., J.-M. Poggi, and C. Tuleau-Malot. 2015. VSURF: an R package for variable selection using
569 random forests. The R Journal 7:19–33.

570 Green, J. L., B. J. Bohannan, and R. J. Whitaker. 2008. Microbial biogeography: from taxonomy to
571 traits. science 320:1039–1043.

572 Grime, J. P. 1977. Evidence for the existence of three primary strategies in plants and its relevance to
573 ecological and evolutionary theory. The American Naturalist 111:1169–1194.

574 Guieysse, B., & Wuertz, S. (2012). Metabolically versatile large-genome prokaryotes. *Current Opinion*
575 *in Biotechnology*, *23*(3), 467-473.

576 Hellweger, F. L., Y. Huang, and H. Luo. 2018. Carbon limitation drives GC content evolution of a
577 marine bacterium in an individual-based genome-scale model. The ISME journal 12:1180–1187.

578 Helmann, J. D. 2002. The extracytoplasmic function (ECF) sigma factors.

579 Jones, D. L., E. C. Cooledge, F. C. Hoyle, R. I. Griffiths, and D. V. Murphy. 2019. pH and exchangeable
580 aluminum are major regulators of microbial energy flow and carbon use efficiency in soil microbial
581 communities. Soil Biology and Biochemistry:107584.

582 Jones, B., Goodall, T., George, P. B., Gweon, H. S., Puissant, J., Read, D. S., ... & Griffiths, R. I. (2021).
583 Beyond Taxonomic Identification: Integration of ecological responses to a soil bacterial 16S rRNA gene
584 database. *Frontiers in microbiology*, *12*.

585 Konstantinidis, K. T., & Tiedje, J. M. (2004). Trends between gene content and genome size in
586 prokaryotic species with larger genomes. *Proceedings of the National Academy of Sciences*, *101*(9),
587 3160-3165.

588 Krause, S., X. Le Roux, P. A. Niklaus, P. M. Van Bodegom, J. T. Lennon, S. Bertilsson, H.-P. Grossart,
589 L. Philippot, and P. L. Bodelier. 2014. Trait-based approaches for understanding microbial biodiversity
590 and ecosystem functioning. Frontiers in Microbiology 5.

591 Kuhn, M. 2008. Building predictive models in R using the caret package. Journal of statistical software
592 28:1–26.

593 Lajoie, G., and S. W. Kembel. 2019. Making the most of trait-based approaches for microbial ecology.
594 Trends in microbiology.

595 Lauber, C.L., Hamady, M., Knight, R., Fierer, N., 2009. Pyrosequencing-based assessment of soil pH
596 as a predictor of soil bacterial community structure at the continental scale. Applied and Environmental
597 Microbiology 75, 5111–5120.

598 Laughlin, D.C., Abella, S.R., Covington, W.W., Grace, J.B., 2007. Species

599 Lavorel, S., and E. Garnier. 2002. Predicting changes in community composition and ecosystem
600 functioning from plant traits: revisiting the Holy Grail. Functional ecology 16:545–556.

601 Le Bagousse-Pinguet, Y., Gross, N., Maestre, F. T., Maire, V., de Bello, F., Fonseca, C. R., ... &
602 Liancourt, P. (2017). Testing the environmental filtering concept in global drylands. Journal of Ecology,
603 105(4), 1058-1069.

604 Li, J., R. L. Mau, P. Dijkstra, B. J. Koch, E. Schwartz, X.-J. A. Liu, E. M. Morrissey, S. J. Blazewicz,
605 J. Pett-Ridge, B. W. Stone, and others. 2019. Predictive genomic traits for bacterial growth in culture
606 versus actual growth in soil. The ISME journal 13:2162–2172.

607 López-Mondéjar, R., V. Tláskal, T. Vetrovsk, M. Štursová, R. Toscan, U. N. da Rocha, and P. Baldrian.
608 2020. Metagenomics and stable isotope probing reveal the complementary contribution of fungal and
609 bacterial communities in the recycling of dead biomass in forest soil. Soil Biology and Biochemistry
610 148:107875.

611 Madin, J. S., D. A. Nielsen, M. Brbic, R. Corkrey, D. Danko, K. Edwards, M. K. Engqvist, N. Fierer, J.
612 L. Geoghegan, M. Gillings, and others. 2020. A synthesis of bacterial and archaeal phenotypic trait data.
613 Scientific Data 7:1–8.

614 Malik, A. A., J. B. H. Martiny, E. L. Brodie, A. C. Martiny, K. K. Treseder, and S. D. Allison. 2019a.
615 Defining trait-based microbial strategies with consequences for soil carbon cycling under climate
616 change. The ISME Journal 14:1–9.

617 Malik, A. A., J. Puissant, K. M. Buckeridge, T. Goodall, N. Jehmlich, S. Chowdhury, H. S. Gweon, J.
618 M. Peyton, K. E. Mason, M. van Agtmaal, and others. 2018. Land use driven change in soil pH affects
619 microbial carbon cycling processes. Nature communications 9:3591.

620 Malik, A. A., T. Swenson, C. Weihe, E. Morrison, J. B. Martiny, E. L. Brodie, T. R. Northen, and S. D.
621 Allison. 2019b. Physiological adaptations of leaf litter microbial communities to long-term drought.
622 BioRxiv:631077.

623 Martiny, A. C. 2019. High proportions of bacteria are culturable across major biomes. The ISME Journal
624 13:2125–2128.
625
626 Martiny, A. C. 2020. The “1% culturability paradigm”needs to be carefully defined. The ISME journal
627 14:10–11.
628
629 Meng, C., B. Kuster, A. C. Culhane, and A. M. Gholami. 2014. A multivariate approach to the
630 integration of multi-omics datasets. BMC bioinformatics 15:1–13.

631 Meyer, F., D. Paarmann, M. D’Souza, R. Olson, E. M. Glass, M. Kubal, T. Paczian, A. Rodriguez, R.
632 Stevens, A. Wilke, and others. 2008. The metagenomics RAST server–a public resource for the
633 automatic phylogenetic and functional analysis of metagenomes. BMC bioinformatics 9:1–8.

634 Nayfach, S., and K. S. Pollard. 2015. Average genome size estimation improves comparative
635 metagenomics and sheds light on the functional ecology of the human microbiome. Genome biology
636 16:1–18.

637 Nayfach, S., and K. S. Pollard. 2016. Toward accurate and quantitative comparative metagenomics. Cell
638 166:1103–1116.

639 Nguyen, S. T., H. L. Freund, J. Kasanjian, and R. Berlemont. 2018. Function, distribution, and
640 annotation of characterized cellulases, xylanases, and chitinases from CAZy. Applied microbiology and
641 biotechnology 102:1629–1637.

642 Okie, J. G., A. T. Poret-Peterson, Z. M. Lee, A. Richter, L. D. Alcaraz, L. E. Eguiarte, J. L. Siefert, V.
643 Souza, C. L. Dupont, and J. J. Elser. 2020. Genomic adaptations in information processing underpin
644 trophic strategy in a whole-ecosystem nutrient enrichment experiment. Elife 9:e49816.

645 Paget, M. S. 2015. Bacterial sigma factors and anti-sigma factors: structure, function and distribution.
646 Biomolecules 5:1245–1265.

647 Pereira-Flores, E., F. O. Glöckner, and A. Fernandez-Guerra. 2019. Fast and accurate average genome
648 size and 16S rRNA gene average copy number computation in metagenomic data. BMC bioinformatics
649 20:1–13.

650 Pérez-Ramos, I. M., C. Roumet, P. Cruz, A. Blanchard, P. Autran, and E. Garnier. 2012. Evidence for
651 a “plant community economics spectrum”driven by nutrient and water limitations in a Mediterranean
652 rangeland of southern France. Journal of Ecology 100:1315–1327.

653 Piton, G., N. Legay, C. Arnoldi, S. Lavorel, J. C. Clément, and A. Foulquier. 2020. Using proxies of
654 microbial community-weighted means traits to explain the cascading effect of management intensity,
655 soil and plant traits on ecosystem resilience in mountain grasslands. Journal of Ecology 108:876–893.

656 Pold, G., L. A. Domeignoz-Horta, E. W. Morrison, S. D. Frey, S. A. Sistla, and K. M. DeAngelis. 2020.
657 Carbon Use Efficiency and Its Temperature Sensitivity Covary in Soil Bacteria. mBio 11.

658 Reich, P. B., I. Wright, J. Cavender-Bares, J. Craine, J. Oleksyn, M. Westoby, and M. Walters. 2003.
659 The evolution of plant functional variation: traits, spectra, and strategies. International Journal of Plant
660 Sciences 164:S143–S164.

661 Shipley, B., D. Vile, and É. Garnier. 2006. From plant traits to plant communities: a statistical
662 mechanistic approach to biodiversity. science 314:812–814.

663 Simonsen, A. K. (2022). Environmental stress leads to genome streamlining in a widely distributed
664 species of soil bacteria. *The ISME journal*, *16*(2), 423-434.

665 Sinsabaugh, R. L., & Shah, J. J. F. (2012). Ecoenzymatic stoichiometry and ecological theory. *Annual*
666 *Review of Ecology, Evolution and Systematics*, *43*(313), 2012.

667 Slonczewski, J. L., M. Fujisawa, M. Dopson, and T. A. Krulwich. 2009. Cytoplasmic pH measurement
668 and homeostasis in bacteria and archaea. Advances in microbial physiology 55:1–317.

669 Sorensen, J. W., Dunivin, T. K., Tobin, T. C., & Shade, A. (2019). Ecological selection for small
670 microbial genomes along a temperate-to-thermal soil gradient. *Nature microbiology*, *4*(1), 55-61.

671 Southwood, T. R. 1977. Habitat, the templet for ecological strategies? Journal of animal ecology
672 46:337–365.

673 Starke, R., D. Morais, T. Vetrovsk, R. Lopez Mondejar, P. Baldrian, and V. Brabcová. 2020. Feeding
674 on fungi: genomic and proteomic analysis of the enzymatic machinery of bacteria decomposing fungal
675 biomass. Environmental Microbiology 22:4604–4619.

676 Steen, A. D., A. Crits-Christoph, P. Carini, K. M. DeAngelis, N. Fierer, K. G. Lloyd, and J. C. Thrash.
677 2019. High proportions of bacteria and archaea across most biomes remain uncultured. The ISME
678 journal 13:3126–3130.

679

680 Violle, C., M.-L. Navas, D. Vile, E. Kazakou, C. Fortunel, I. Hummel, and E. Garnier. 2007. Let the
681 concept of trait be functional! Oikos 116:882–892.

682 Wallenstein, M. D., and E. K. Hall. 2012. A trait-based framework for predicting when and where
683 microbial adaptation to climate change will affect ecosystem functioning. Biogeochemistry 109:35–47.

684 Weissman, J. L., W. F. Fagan, and P. L. Johnson. 2019. Linking high GC content to the repair of double
685 strand breaks in prokaryotic genomes. PLoS genetics 15:e1008493.

686 Westoby, M., M. R. Gillings, J. S. Madin, D. A. Nielsen, I. T. Paulsen, and S. G. Tetu. 2021. Trait
687 dimensions in bacteria and archaea compared to vascular plants. Ecology Letters.

688 Wieder, W. R., G. B. Bonan, and S. D. Allison. 2013. Global soil carbon projections are improved by
689 modelling microbial processes. Nature Climate Change 3:909–912.

690 Wright, I. J., P. B. Reich, M. Westoby, D. D. Ackerly, Z. Baruch, F. Bongers, J. Cavender-Bares, T.
691 Chapin, J. H. Cornelissen, M. Diemer, and others. 2004. The worldwide leaf economics spectrum.
692 Nature 428:821.

693 Xu, X., J. P. Schimel, P. E. Thornton, X. Song, F. Yuan, and S. Goswami. 2014. Substrate and
694 environmental controls on microbial assimilation of soil organic carbon: a framework for Earth system
695 models. Ecology letters 17:547–555.

696 Yang, X., W.M. Post, P.E. Thornton, and A. Jain. 2014. Global Gridded Soil Phosphorus Distribution
697 Maps at 0.5-degree Resolution. Data set. Available on-line [http://daac.ornl.gov] from Oak Ridge
698 National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, USA.
699 http://dx.doi.org/10.3334/ORNLDAAC/1223

700 Zheng, Q., Hu, Y., Zhang, S., Noll, L., Böckle, T., Richter, A., & Wanek, W. (2019). Growth explains
701 microbial carbon use efficiency across soils differing in land use and geology. *Soil Biology and*
702 *Biochemistry*, *128*, 45-55.

703

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformationGlobalfunctionalspectrumofsoilbacteria.pdf