



HAL
open science

Life history strategies of soil bacterial communities across global terrestrial biomes

Gabin Piton, Steven Allison, Mohammad Bahram, Falk Hildebrand, Jennifer Martiny, Kathleen Treseder, Adam Martiny

► **To cite this version:**

Gabin Piton, Steven Allison, Mohammad Bahram, Falk Hildebrand, Jennifer Martiny, et al.. Life history strategies of soil bacterial communities across global terrestrial biomes. *Nature Microbiology*, 2023, 10.1038/s41564-023-01465-0 . hal-04264306

HAL Id: hal-04264306

<https://hal.inrae.fr/hal-04264306v1>

Submitted on 22 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **Life history strategies of soil bacterial communities across global terrestrial biomes**

2 Gabin Piton^{1,2,*}, Steven D Allison^{1,3}, Mohammad Bahram^{4,5}, Falk Hildebrand^{6,7}, Jennifer BH Martiny³,
3 Kathleen K Treseder³, Adam C Martiny^{1,3}

4 **Author information**

5 *Affiliations*

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

15 * Corresponding author: gabin.piton@inrae.fr

16 **Abstract**

17 The life history strategies of soil microbes determine their metabolic potential and their response to
18 environmental changes. Yet they remain poorly understood. Here we use shotgun metagenomes from
19 terrestrial biomes to characterise overarching covariations of the genomic traits that captures dominant
20 life history strategies in bacterial communities. The emerging patterns show a triangle of life history
21 strategies shaped by two trait dimensions, supporting previous theoretical and isolate-based studies. The
22 first dimension ranges from streamlined genomes with simple metabolisms to larger genomes and

23 expanded metabolic capacities. As metabolic capacities expand, bacteria communities increasingly
24 differentiate along a second dimension that reflects a tradeoff between increasing capacities for
25 environmental responsiveness or nutrient recycling. Random forest analyses shows that soil pH, C:N
26 and precipitation patterns together drive the dominant life history strategy of soil bacteria communities
27 and its biogeographic distribution. Our findings provide a trait-based framework to compare life history
28 strategies of soil bacteria.

29 **Introduction**

30 Bacteria impact carbon (C) and nutrient cycling on a global scale¹. Soil bacterial communities contain
31 enormous, functionally uncharacterized genetic diversity^{2,3} that hinders progress in predicting soil
32 microbial responses to global change^{4,5}. One approach to describe functional biodiversity is to collapse
33 its complexity into one or more dimensions that capture the dominant associations and trade-offs
34 between traits⁶⁻¹⁰. This multivariate trait space - or life history strategy scheme - provides a framework
35 to compare broad organismal strategies^{6,8,10}.

36 While the trait dimensions shaping plant life history strategies is now well established⁶, trait
37 associations for soil microorganisms remain less clear. Initially, studies applied the ‘Competitor’, ‘Stress
38 tolerant’, and ‘Ruderal’ (CSR) strategies proposed for plants⁷ to soil bacteria^{1,11}. This scheme
39 emphasises trade-offs often observed between traits related to maximizing resource capture
40 (Competitor, C), persisting under low resource and stressful condition (Stress tolerant, S), and
41 responding rapidly to exploit growing window between disturbances (Ruderals, R)^{1,7}. Building on the
42 CSR scheme, Malik et al. (2020)¹² emphasised differences between microbial yield (Y), resource
43 acquisition (A) and stress tolerance (S) traits as important for soil carbon cycling¹². While these
44 theoretical papers provide valuable hypotheses on which traits are probably central to soil microbial
45 adaptation, no clear consensus has emerged on the trait dimensions that shape life history strategies of
46 soil bacteria^{1,11,12} (Extended Data table 1). Recently, Westoby and co-workers (2021) analysed bacterial
47 cultures isolated from diverse habitats for genomic and phenotypic traits¹³. This analysis revealed a

48 primary dimension associated with metabolic versatility that was highly correlated with genome size. A
49 secondary dimension separated differences in maximum growth rate and was correlated with variation
50 in ribosomal gene copy number¹⁴. However, there is a lot of variation in how well bacterial cultures
51 represent *in situ* community biodiversity¹⁵⁻¹⁷. Thus, it remains to be tested if the life history strategies
52 of soil bacterial communities matches either the theoretical or culture-based predictions of key trait
53 dimensions.

54 One advantage of studying the traits of microorganisms over those of larger organisms is the
55 ease of which collections of their traits can be measured on the community level. Community aggregated
56 traits (CATs)¹⁸ represent the average functional profile of the community emerging from the
57 combination of organisms' traits and community composition (similar to the idea of community-
58 weighted means of traits proposed for plants)^{19,20}. Hence, it is important to note, that while suggestive,
59 such CAT patterns do not directly inform on the within-organism tradeoffs. Nevertheless, CATs
60 described using metagenomic sequences offer a way to characterize shifts in the organismal strategies
61 dominating bacterial communities *in situ* (eg. ^{21,22}) and thus offer an approach to test theoretical life
62 history strategy schemes to *in situ* microbial communities. In addition, information on the dominant
63 strategy in a bacterial community might be used to predict the response to environmental changes of this
64 key group for global biogeochemical cycles^{4,18}. Elucidating the trait dimensions that shape the dominant
65 life history strategies of soil bacteria would thus provide a framework for comparing soil bacterial
66 communities and developing generic predictions in soil microbial ecology¹⁴.

67

68 In this study, we used a global dataset of soil metagenomic sequences from major biomes to quantify
69 key trait dimensions of soil bacterial communities. We then identified primary environmental factors
70 partitioning the trait dimensions and projected the global biogeography. Finally, we compared the
71 emergent life history strategies with theoretical and culture-based predictions.

72 **Results**

73 *The trait dimensions of soil bacterial communities*

74 Using a multi-table co-inertia analysis (MCOA), we found that two dimensions captured half of the
75 overall variation in metagenomic community aggregated traits (CATs). MCOA1 and MCOA 2 captured
76 29% and 21% of metagenomic trait variation (Figure 1 and Extended Data Figure 1), while MCOA 3
77 and MCOA4 explained 16% and 10% of this variation, respectively (Extended Data Figure 1 and 2).
78 The MCOA revealed the most important associations between traits (Figure 1-2) including traits
79 previously associated with life history strategies (Figure 1-3).

80 Average genome size had the highest contribution to MCOA1 (Figure 1A) with a R^2 of 0.64 for the
81 positive correlation between average genome size and MCOA1 (Extended Data Figure 3A). Mapping
82 coverage decreased along this dimension (Extended Data Figure 4). The lower end of this dimension
83 was characterised by bacterial communities with higher relative abundance of genes for primary
84 metabolism (ie. essential process for survival and growth) and C acquisition machinery (Figure 1). In
85 these communities, carbon acquisition enzymes involved in depolymerization of oligosaccharides were
86 favored over enzymes targeting polysaccharides. This oligosaccharide-degradation enzyme class was
87 dominated by the beta-glucosidases GH1, GH2 and GH3 CAZy families. Finally, chaperones were
88 overrepresented. Thus, the lower end of MCOA1 were defined by communities with a streamlined
89 metabolism (Figure 2).

90 The upper end of MCOA1 defined bacterial communities with a large genome and more complex
91 metabolism and resource acquisition strategies (Figure 1-2). The enriched genes allowed for degradation
92 of complex polysaccharides from fungi, animals and plant lignin. There was also a gene
93 overrepresentation for direct plant pathogenic interactions and negative interactions with other
94 microorganisms. Finally, communities carried a higher proportion of genes encoding for EPS
95 production, Dormancy and Sporulation, membrane, and DNA repair (Figure 1-2). These functions were
96 generally present in lower relative abundance in communities with small genomes at the opposite end

97 of MCOA1. Thus, the first trait dimension captured functional variation associated with genome size
98 and expanded *metabolic capacities* (Figure 2).

99 Bacterial communities differentiated along a second dimension (MCOA2) but only when they increased
100 their *metabolic capacities* along the first trait dimension (MCOA1), shaping a triangle (Figure 2). This
101 distribution indicated that bacteria communities with low *metabolic capacities* and small average
102 genome size are constrained along the second dimension. The MCOA2 separated communities
103 according to genomic traits for *environmental responsiveness* and *nutrient recycling* (Figure 2).
104 Communities associated with the lower end of MCOA2 were enriched in mineral and organic N and P
105 assimilation genes (Figure 1-2). Furthermore, there were also higher relative frequencies of genes
106 encoding for bacterial necromass degradation including peptidoglycan. Communities at the upper end
107 of MCOA2 were defined by an ability to respond to a complex set of environmental cues. This was
108 manifested by an increased presence of genes encoding for activity regulation, resistance to
109 environmental stress, foraging of beneficial conditions, fast growth (*rrn* copy), and building and
110 repairing the cell membrane (Figure 1-2). The communities were also enriched in genes encoding for
111 carbohydrates metabolism of simple substrates like starch, glycogen, and oligosaccharides. Thus, the
112 second trait dimension captured a gradient in the average environmental responsiveness that was
113 positively associated with a specialisation in simple carbon substrate metabolism and negatively with
114 nutrient assimilation and recycling capacities (Figure 2).

115 *Drivers of the trait dimensions*

116 Using random forest analyses, we next found that common soil environmental factors distributed the
117 soil bacterial community along global trait dimensions. Random forest models based on soil pH,
118 precipitation and C:N could predict most of the variation in MCOA1 and MCOA2 with a R^2 of 0.80
119 and 0.58, respectively (Extended Data Figure 5). Mean decrease in mean square error (%MSE) and R
120 squared calculated based on a ten-fold cross-validation of the random forests indicated that soil pH and
121 annual precipitation are the most important predictors for both MCOA1 and MCOA2. However, the two
122 dimensions showed different response patterns to these variables, with MCOA1 decreasing with soil pH

123 but increasing with annual precipitation whereas MCOA2 decreased with both soil pH and annual
124 precipitation, leading to unique position along MCOA1 and MCOA2 depending on the combination of
125 pH and annual precipitation (Figure 3-4). MCOA1 and MCOA2 were also driven by precipitation
126 seasonality whereas soil C:N controls only MCOA1 (Figure 3-4, Extended Data Figure 5). Next, we
127 projected the global variation in the trait dimensions using these random forests (Figure 4 B and D) and
128 global soil and climate databases. It is worth noting that this broad spatial resolution map, using
129 averaged conditions across large spatial units, showed high consistency with values observed locally in
130 our samples (Extended Data Figure 6). Thus, the identified trait dimensions showed a clear global
131 biogeography.

132 The first trait dimension (MCOA1) mainly separated arid, alkaline regions from more acidic and wet
133 ones. More precisely, bacterial communities characterised by a small genome size (i.e., low MCOA1
134 value) were enriched under neutral to alkaline pH, low C:N, low annual precipitation but high
135 precipitation seasonality (Figure 4A). Conversely, communities with larger genome sizes (high MCOA1
136 value), were found in more acidic soils as well as soil with higher C:N and climate with elevated stable
137 precipitation (Figure 4A). Globally, these environmental controls predicted low MCOA1 coordinates ($<$
138 -1) under arid and semi-arid climates at tropical and subtropical latitude as well as in the steppe zones
139 of central Asia and North America (Figure 4B). Conversely, high MCOA1 coordinates (>1) were seen
140 in equatorial forests as well as some temperate zones in northern Europe, Western Canada, New Zealand
141 and south Chile. Steep MCOA1 gradients were estimated to occur in regions separating arid and wet
142 zones and medium coordinates ($-1 < \text{MCOA dimension 1} < 1$) also covered most of temperate and high
143 latitudinal regions (Figure 4B).

144 The second trait dimension (MCOA2) separated regions with high but stable precipitation from places
145 with more seasonal climate and extremely acidic soils. The lower end of MCOA2 covered most high
146 precipitation regions ($>2500\text{mm}$) including equatorial zones of South-America and Asia and wet Europe
147 and North America temperate zones. Medium-high coordinates ($0 < \text{MCOA2} < 1$) covered most of the
148 globe, characterising all tropical-dry, semi-arid and subarctic regions. The projection of this dimension

149 (Figure 4D) predicts very high coordinates (MCOA2 >1) under limited regions of subtropical and high
150 latitudes combining low annual precipitation (<1000mm) and very acidic pH (<4).

151 Finally, we found that trait differences (defined based on euclidian distances along the two first
152 dimensions of the MCOA) were significantly correlated with Unifrac phylogenetic distances ($R^2=0.32$,
153 Extended Data Figure 7). Communities with average genome size below its median values depicted a
154 correlation between trait and phylogenetic distances significantly steeper (slope difference: $p=0.00116$)
155 and tighter ($R^2=0.46$) compared to communities with larger genomes ($R^2=0.15$, Extended Data Figure
156 7).

157 **Discussion**

158 Our study describes two dominant dimensions of community aggregated traits variation across soil
159 bacteria communities (Figure 2-3). In this trait space, communities are constrained in a triangle of three
160 opposing life history strategies: low metabolic capacities; metabolic capacities expanded for
161 environmental responsiveness; metabolic capacities expanded for nutrient recycling. These life history
162 strategies incorporates traits previously identified as CSR strategies^{1,11,12} (Extended Data Table 1).
163 Moreover, it fits into a triangle like the original CSR model^{7,23} (Figure 2-3) which suggests that the
164 constraints on bacterial strategies might scale up to community level. Also consistent with CSR theory,
165 both trait dimensions of our study capture competitor traits that tradeoff with traits of the other strategies.
166 However, while one strategy generally dominates the traits of each end of the trait dimensions, our
167 aggregated profiles often combine traits that had been associated with different strategies. In particular,
168 one or more stress tolerance traits are part of all profiles (Figure 2-3). We hypothesise that these
169 combinations indicate either that the communities are composed of taxa with different strategies or that
170 the majority of bacteria living in soil need stress tolerant traits to survive in this challenging
171 environment.

172 Bacteria with streamlined metabolism dominate the low end of the *metabolic capacity* dimension. The
173 genomic traits of these bacterial communities with small average genome size have only few matches

174 with previous description of stress tolerance strategy (Extended Data Table 1)^{1,11,12}. However, the clear
175 association to arid biomes that we observed suggests that the streamlined bacteria are associated with
176 stress tolerance strategy. This is consistent with recent studies showing that genome streamlining can
177 play a role in adaptation to environmental stressful conditions (eg.^{24,25}). In particular, Liu et al. (2023)
178 used a joint species distribution model to show that soil bacteria with small genomes are selected under
179 arid environments, as seen here. Moreover, these streamlined communities were associated with some
180 low environmental constraints on resource acquisition (low soil C:N and pH near neutrality as observed
181 in ²⁶) that might also reduce fitness benefits for gaining new capabilities²⁷. Thus, genome streamlining
182 and associated change in gene frequency might be central in the soil bacteria stress tolerance, especially
183 in arid biomes.

184 Cells with larger genomes and a more complex metabolism dominate the other end of the *metabolic*
185 *capacity* dimension. The associated variation in the functional gene frequency that we observed is also
186 consistent with previous studies reporting that genome expansion in free-living bacteria is driven by
187 gene additions encoding for new metabolic capabilities or regulation^{14,28}. Large genomes, high catabolic
188 diversity, and antibiotic resistance genes observed for this life history strategy were previously attributed
189 to a competitor strategy (Extended Data Table 1)^{1,11}. This supports the idea that complex substrates
190 acquisition is a key trait of competitors as suggested by Malik et al. (2020). Consistent with competitor
191 traits, these attributes are favoured under stable and wet climates, that reduce the benefits of desiccation
192 stress traits and possibly leading to intense resource competition⁷. We also detected an enrichment in
193 traits associated with sporulation and exopolysaccharides production, two traits often associated with
194 stress tolerance or ruderality (Extended Data Table 1) that might also improve tolerance to antimicrobial
195 compounds or nutritional constraints for such competitor profile^{29,30}. Together, the first trait dimension
196 appears to represent a gradient from stress tolerant communities with small genomes to communities
197 dominated by bacteria with increased *metabolic capacities* associated with other strategies, especially
198 competitors.

199 When average genome size increases, bacteria communities differentiate along the second dimension
200 with opposing profiles of either increased capacities for *environmental responsiveness* or for *nutrient*
201 *recycling*. At the high end of this dimension, communities with high *environmental responsiveness*
202 shared numerous genomic features tied to both the ruderal and stress tolerant strategies (Extended Data
203 Table 1). This includes traits to resist stress, sensing favourable environmental conditions, activate fast
204 growth, and C acquisition. The reduced and fluctuating precipitation patterns associated with this profile
205 are also consistent with original descriptions of these strategies^{1,7}. At the opposite end of this second
206 dimension, bacteria specialised in *nutrient recycling* show a resource acquisition strategy with a high
207 number of transporters and bacterial biomass (Peptidoglycan) recycling and a higher investment towards
208 nitrogen and phosphorus metabolism compared to carbon metabolism. Microbial mineralisation activity
209 and biomass turnover release nutrients and necromass into soil that this profile seems optimised to
210 recycle. Such traits might reflect a strategy that emphasises resource use efficiency and increased
211 competitiveness for nutrients^{11,12}. Further, the environmental parameters associated with this life history
212 strategy (medium-low pH, high precipitation and low seasonality) are the most favourable for resource
213 acquisition³¹, biomass turnover and yield^{32,33}, reinforcing potential selection for competitor traits⁷. In
214 summary, the second trait dimension reflects communities with increased metabolic capacities
215 associated with either a combination of stress tolerance and ruderal traits that maximise their
216 responsiveness or a reinforcement of competitor traits that favour nutrient recycling.

217 Overall, our dimension of *metabolic capacities* matches the versatility dimension described by Westoby
218 et al. (2021) across cultured bacterial taxa, with both studies supporting that genome size plays a central
219 role in differentiating bacteria strategies. Our dimension opposing *environmental responsiveness* and
220 *nutrient recycling* also shows some consistencies with the second trait-dimension described by Westoby
221 et al. (2021) capturing a rate-yield tradeoff, with rrn copy number as principal trait. Indeed, as discussed
222 above, the traits of the *nutrient recycling* profile might favour growth yield, and high *environmental*
223 *responsiveness* is associated with higher rrn copy number. However, these variations of rrn copy
224 numbers have only a limited importance in the second trait dimension of our study, contrasting with the

225 observations of Westoby et al. (2021) for cultured bacteria from diverse habitats. This could be
226 explained by the constraint range of this trait in soil. Indeed, variation in average rrn copy number
227 observed across communities in our study is highly constrained (1 to 1.5 copies, Extended Data Figure
228 3). These observations are consistent with Gao and Wu (2018) reporting that most soil bacteria have
229 less than 2 rrn copies, whereas bacteria from other environments can have up to 15 copies³⁴. Further,
230 variation in the average rrn copy number of whole communities will be more constrained than variation
231 across individual isolates within the community; indeed, some bacteria with more copies might be
232 present in the soil community, with their populations increasing during resource flushes (eg.³⁵). In the
233 oligotrophic environment of soil, our results suggest that increased capacity to recycle resources
234 efficiently, to sense favourable conditions and to survive or escape stressful ones represent more
235 common adaptations for bacteria than growing more rapidly. Investigating the variation of these traits
236 across taxa in soil and their distribution within communities represents a challenging, but fascinating
237 perspective to disentangle how the trait dimensions across taxa scales up to the community level.
238 Overall, life history strategies of soil bacteria that we described using aggregated traits at the community
239 level show some important consistencies with life history strategies described across bacterial taxa from
240 various habitats, but also highlights some specificities and challenges associated with soil environment.

241 Soil bacteria remain poorly characterised with a limited number of reference genomes and gene
242 functional characterization^{36,37}. This reduces annotation coverage of metagenomic data and can limit
243 analysis conclusions. In our study, the proportion of reads annotated (between 5 and 15% depending on
244 the database) were in the range of what is commonly obtained from soil metagenomes³⁸. Our usage of
245 stringent quality filtering criteria in the annotation² also reduced the annotation coverage but increased
246 annotation confidence. Finally, the proportion of unannotated reads is increased by the sequencing error
247 and our usage of short read sequencing technology and read-based profiling (as opposed to assembly
248 based profiling with better annotation but very limited representativity of the community). Our
249 annotation coverage also showed a decrease with genome sizes, as reported across taxa^{36,37}. However,
250 unannotated genes likely belong to accessory genes and not to core metabolism that are well represented

251 in current databases³⁷. Thus, we can expect that increased annotation of large genomes would have
252 accentuated evidence for our conclusion that our first trait dimension captured an increase in metabolic
253 capacities. Overall, our trait dimensions are expected to capture at least the functional variations
254 associated with core metabolism and provide some first elements about functional genes associated with
255 expansion of metabolic capacities.

256 We showed that communities with similar life history strategies tend to be phylogenetically closer,
257 supporting a certain phylogenetic conservatism of the genomic traits shaping life history strategies³⁹.
258 However, this relationship weakens as genome size and metabolic capacities expand (Expanded Data
259 Fig 7). This suggests that metabolic expansion during different evolutionary histories can converge to
260 similar life history strategy⁴⁰. Hence, phylogenetic distance become a poorer predictor of difference in
261 life history strategies for soil bacterial communities with large genomes.

262 The biogeography of dominant life history strategies in soil bacterial communities is mainly driven by
263 the combinations of soil pH and precipitation patterns across the globe. These environmental factors
264 impact stress and competition intensity for soil bacteria, either through direct effect on their physiology
265 and interaction⁴¹⁻⁴³ or indirectly through their modification of abiotic (eg. solubilization of toxic ions
266 Al³⁺) and biotic (eg. plant and fungal communities) characteristics of the ecosystem⁴⁴⁻⁴⁶. The
267 environmental distribution of the life history strategies suggests that bacteria expand their metabolic
268 capacities to deal with conditions associated with increasing soil acidity and annual precipitation until a
269 certain level (Figure 3). Then, expansion of *metabolic capacities* increases either *environmental*
270 *responsiveness* to survive under more extreme pH and fluctuating precipitation or *nutrient recycling* to
271 be competitive under higher precipitation levels. These global effects of pH and precipitation are
272 consistent with previous studies of soil bacteria biogeography^{3,26,47} and provide some new information
273 on the traits associated with these environmental factors.

274 Our global projection (Figure 3B and D) aims at giving a picture of the general biogeographic patterns
275 in the functional profiles of soil bacterial communities. However, it is important to note that

276 transposition of our trait dimensions at local scale will need further investigation. Values predicted for
277 these broad resolution maps can be dissociated from the local situation if its conditions highly differ
278 from the regional mean (Sup Figure 8) and should be used with caution. Despite outstanding issues that
279 remain open, our study demonstrates how metagenomic approaches can provide substantial advance in
280 our understanding of microbial community functioning. Altogether, our results suggest that land use and
281 climate changes impacting soil pH and precipitation gradients at biogeographic scale might be central
282 in shaping future functional potential of soil bacterial communities and thus global biogeochemical
283 cycles.

284 **Methods**

285 *Soil sampling and characteristics*

286 We analysed a global dataset of 128 metagenomes each from unique soil samples distributed across
287 continents and latitude (Extended Data Figure 8)². We selected this dataset for our analysis because of
288 its coverage and its use of a highly standardised protocol that: 1) sampled top-soils in spatially
289 independent sites across the globe selected to represent all the most important vegetation types; 2)
290 analysed soil chemistry and metagenomes². All samples were processed using similar standardised
291 protocols for their chemistry (carbon, nitrogen, phosphorus content and pH_{H2O}) and metagenome (See²)
292 for protocol details). We checked the global environmental coverage by comparing variation of the main
293 environmental variables (mean annual temperature (MAT), mean annual precipitation (MAP), soil pH
294 and net primary productivity (NPP)) in our dataset with global variation from the Atlas of the Biosphere
295 (<https://nelson.wisc.edu/sage/data-and-models/atlas/maps.php>). This showed an almost complete global
296 coverage, with only extreme MAT of very high latitude (below -11.33°C) and Sahelian Africa (above
297 MAT 27.97°C) as well as very high pH (higher than 7.76) characterising some parts of North Africa,
298 West Asia and Himalaya missing in our dataset (Extended Data Figure 8). As far as we know, when we
299 conducted this analysis, this dataset was the only available with such precise characterization of soil
300 environment done on the same sample as shotgun metagenomic analysis, making this dataset the most

301 robust for our objective to assess environmental drivers of metagenomic profiles. Nevertheless, potential
302 to extend environmental range by adding all (excluding agricultural and contaminated) soil
303 metagenomes available (accession date January 28 2021) from the main sequence repositories MG-
304 RAST⁴⁸ and IMG:M⁴⁹ was also tested. This indicated that adding these data would not have extended
305 environmental range (excepted a few samples from very cold sites with mean annual temperature lower
306 than -11.5°C available on MG-RAST) and this would have greatly decreased precision of soil properties
307 characterization (Extended Data Figure 9).

308 *Metagenomic and amplicon sequencing data*

309 DNA extraction, sequencing (Illumina with RTA Version 1.18.54 and bcl2fastq v1.8.4), trimming and
310 mapping approaches are detailed in Bahram et al. (2018). In this study, four community aggregated trait
311 databases were built, corresponding to metagenomic reads mapping on different functional annotation
312 systems by Bahram et al. (2018). An additional database was made for this study with genomic traits
313 previously associated with bacterial life history strategies (See details below). Data from 16S rRNA
314 gene amplicon sequencing were also used to characterise phylogenetic distances between bacterial
315 communities using the Unifrac metric⁵⁰

316 *Bacterial community aggregated trait calculation*

317 Bahram et al. (2018) mapped reads to the functional databases (KEGG, eggNOG and CAZy). Data were
318 aggregated at the (1) pathway (KEGG), (2) functional categories (eggNOG) levels, (3) SEED functional
319 modules and (4). Glycolysis Hydrolases (GH) and Auxiliary Activities (AA) gene families from
320 CAZy⁵¹. All read mapping was done competitively against both prokaryotic and eukaryotic functional
321 databases and best bit score in the alignment and the taxonomic annotation was used to retrieve only
322 reads annotated as bacteria.

323 We used output data from these four annotation processes to provide complementary classification of
324 functional genes (e.g. eggNOG categories include Motility, Cell envelopes and Defense which are not
325 included in SEED whereas SEED classes include Dormancy and Sporulation, Stress response,

326 Virulence, Carbon, Nitrogen and Phosphorus metabolism which are not included in eggNOG). The
327 eggNOG annotation also differed from KEGG and SEED in the construction of orthologous groups with
328 eggNOG using non-supervised construction increasing coverage whereas KEGG used supervised
329 construction increasing annotation robustness. To obtain a more precise picture of C acquisition strategy,
330 the CAZy annotated reads abundance were aggregated on the basis of their targeted substrates
331 (Cellulose, Chitin, Glucan, Lignin, Peptidoglycan, Starch/Glycogen, Xylan, Other Animal
332 Polysaccharides, Other Plant Polysaccharides, Oligosaccharides) using a curated database
333 (Supplementary Table 2) based on previous works⁵²⁻⁵⁴. After mapping, the relative abundance of each
334 gene (or aggregated group of genes) was normalised by the total number of bacteria-reads annotated for
335 this sample on the same database. Such normalisation corrects for variation between samples in the
336 quantity of annotated reads and avoids biases induced by contamination and sequencing error⁵⁵. The
337 obtained relative abundances inform on the relative importance of a gene (or gene group) compared to
338 all the other annotated functions.

339 *Life history trait calculation*

340 An additional database was built with genomic traits previously associated with bacteria life history
341 strategies (Extended Data Table 1). For this database, nine life history traits were calculated. Seven traits
342 were calculated by summing the relative abundances of genes associated with Sigma factor⁵⁶,
343 Exopolysaccharides (EPS)⁵⁷, Chaperons^{12,58}, Chemotaxis, and Osmolytes⁵⁹⁻⁶², antibiotic resistance and
344 carbohydrates degradation enzymes (CAZyme). In addition, average genome size was calculated using
345 MicrobeCensus⁶³ and rrn copy number using the method described in ⁶⁴. All sequences were used as
346 input for average genome size and rrn copy number, after a verification that eukaryotic sequences were
347 negligible (less than 2% of annotated reads for all databases verified for all samples) and therefore, that
348 the samples mostly captured bacteria.

349 *Statistical analysis*

350 To identify the multivariate axes that best explain the global scale variation in metagenomic community
351 aggregated traits of soil bacteria, we used a multi-table co-inertia analysis (MCOA), an exploratory
352 analysis that leverages together the information from the 5 databases (genomic traits, eggNOG
353 categories, SEED modules, KEGG pathway, CAZy types). This method identifies co-relationships
354 between the different databases and uses a covariance optimization criterion to summarise in a common
355 structure the information shared by multiple multivariate (eg. omic) tables⁶⁵⁻⁶⁷. All variables (CATs)
356 were log transformed ($\log X + 1$) before the analysis to improve normality⁶⁷ and standardised to a mean
357 of zero and a variance of 1. The R package ade4 was used for the MCOA analysis⁶⁸.

358 Sample coordinates on the first and second dimension of the MCOA were extracted and used as latent
359 variables representing bacterial community positions in the global trait space. Random forest models
360 were then used to identify predictors of these coordinates among potential environmental drivers, which
361 were the soil properties measured on the same sample as metagenome (see Soil sampling and
362 characteristics) and climatic variables extracted from Worldclim2 : BIO1 = Annual Mean Temperature,
363 BIO4 = Temperature Seasonality (standard deviation), BIO12 = Annual Precipitation and BIO15 =
364 Precipitation Seasonality (standard deviation). First, we verified that all selected environmental drivers
365 had spearman correlation coefficients lower than 0.7 to mitigate collinearity problems as recommended
366 in ⁶⁹. Second, a variable selection process was carried out using the method implemented in the VSURF
367 R package⁷⁰. The number of predictors randomly tested at each node of the random forest tree (mtry)
368 was optimised based on randomForest's tuneRF algorithm and the number of trees set to 1000. Third,
369 the random forest models selected following the VSURF selection process were trained using ten-fold
370 cross-validation (100 repetitions) implemented in the caret package⁷¹ and model performance was
371 assessed based on Root Mean Square Error (RMSE) and R squared. Finally, random forest predictive
372 models were used to project a broad resolution map of trait dimension global biogeography using
373 environmental maps (1600x1200 pixel) as predictors. For this projection, we used the the latest map
374 (June 2022) released by ISRIC's World Soil Information Service
375 (https://files.isric.org/soilgrids/latest/data_aggregated/) based on SoilGrids version 2.0⁷². Worldclim2

376 (<https://www.worldclim.org/>) was used for climatic variables. The raster R package was used for the
377 spatial predication and projection. To validate the relevance of this broad resolution map to represent
378 average local values, we tested the correlation between local observations and the predicted value of the
379 cell in which the local observation was done.

380 Finally, we tested the relationship between phylogenetic composition of the bacterial communities and
381 their positions in the MCOA trait space using linear correlation between Euclidean distances along the
382 two first dimensions of the MCOA and Unifrac phylogenetic distance. The influence of average genome
383 size on this relationship was then assessed by comparing the correlation coefficients for communities
384 below and above the median average genome size in the dataset.

385 **Data availability**

386 The five CAT databases used to build the trait dimensions and the associated environmental variables
387 are available on figshare repository : <https://doi.org/10.6084/m9.figshare.22620025> All the original
388 sequences are available in the European Bioinformatics Institute Sequence Read Archive database:
389 soil metagenomes, accession numbers PRJEB18701 (ERP020652), 16S metabarcoding sequences,
390 accession numbers PRJEB19856 (ERP021922).

391 **Code availability**

392 Access to the code used in the analyses done for this research is available by request to the
393 corresponding author.

394 **Acknowledgements**

395 We thank Leho Tedersoo and Peer Bork who conceived and supervised the acquisition of the global
396 dataset used in this study with Mohammad Bahram and Falk Hildebrand. We also wish to thank all their
397 collaborators who contribute to this global data acquisition effort. We also thank Alyse Larkin and Lucas
398 Ustick for their guidance in the bioinformatic analysis conducted in this study. Finally, we thank Kate
399 Buckeridge and two anonymous referees for their insightful reviews. GP, SA, JM, KT and AM were

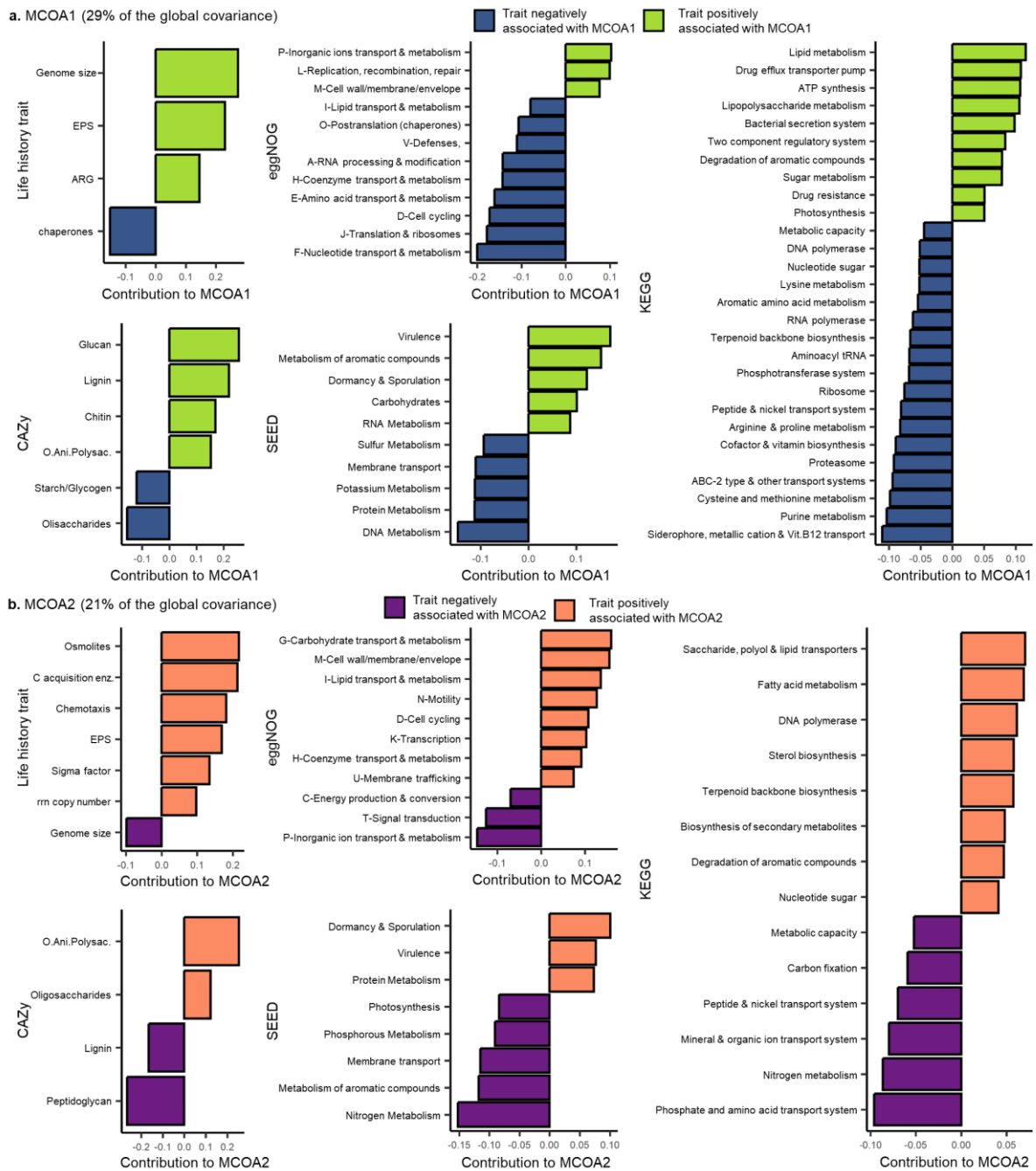
400 supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental
401 Research grants DE-SC0016410 and DE-SC0020382. FH was supported by the European Research
402 Council H2020 StG (erc-stg-948219, EPYC) and the Biotechnology and Biological Sciences research
403 Council (BBSrC) Institute Strategic Program Gut Microbes and Health BB/r012490/1 and its constituent
404 project BBS/e/F/000Pr10355.

405 **Authors contributions**

406 Data collection was designed and supervised by M.B. Initial bioinformatics analysis to obtain
407 functional genes abundance tables (eggNOG, KEGG, SEED, CAZy) was designed and performed by
408 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
409 quantification of genomic traits, Unifrac and data analyses were performed by G.P. First draft and
410 following editing was conducted by G.P. with inputs from all co-authors.

411 **Competing interests**

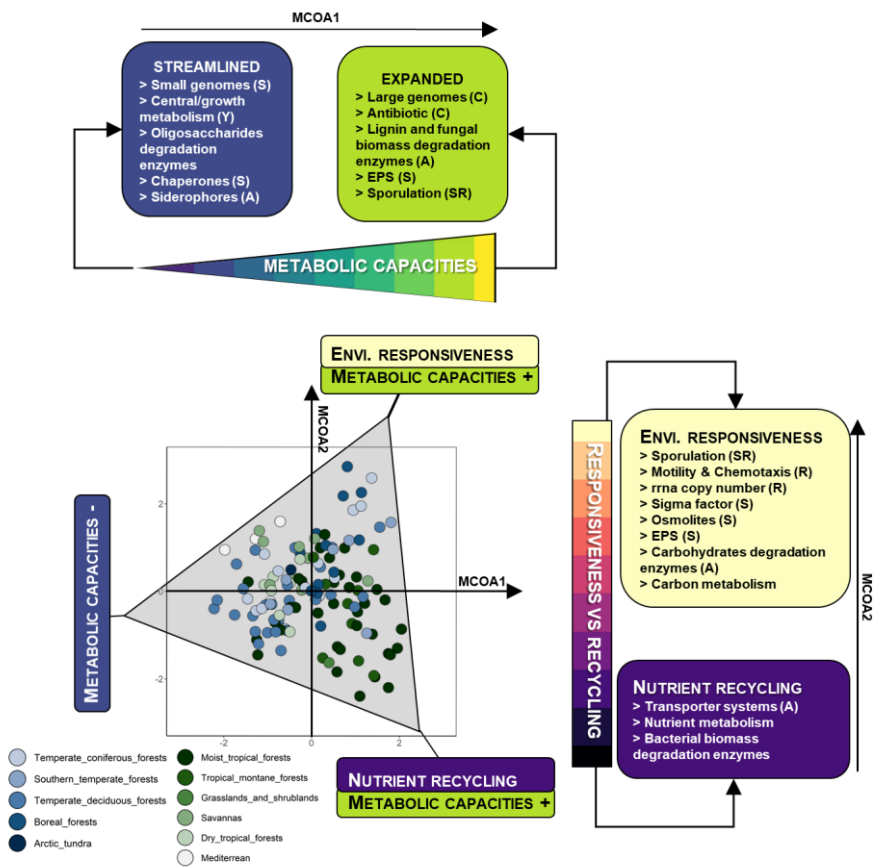
412 The authors declare no competing interests.



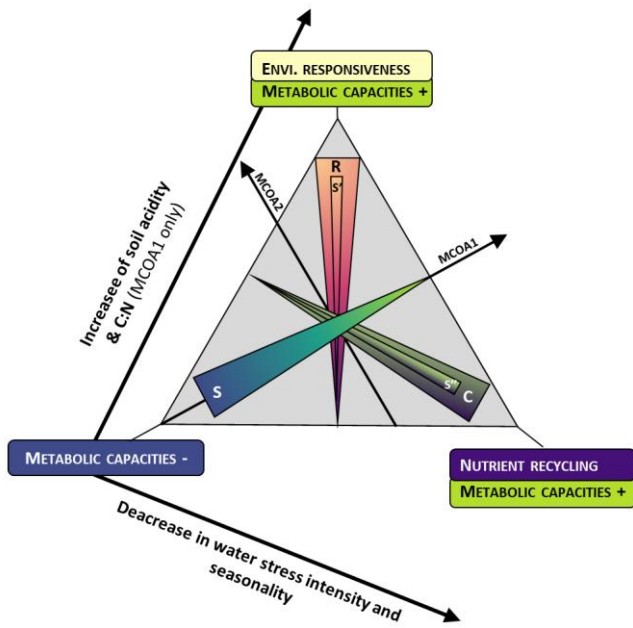
413

414 *Figure 1. Global trait dimensions of soil bacteria metagenomes. Variable contributions to the multiple*
 415 *co-inertia analysis (MCOA) summarising in a common structure (MCOA dimensions 1 and 2) the*
 416 *information shared by 5 community aggregated trait (CAT) databases (Life history trait, CAZY,*
 417 *eggNOG, SEED and KEGG). Only the most important variables with significant correlation ($p < 0.001$)*
 418 *with each dimension are reported in this figure. a and b panels present variable contributions to MCOA*

419 Dimension 1 and 2 respectively. Bar colours indicate the direction of the associations between the
 420 variable and the MCOA dimensions.

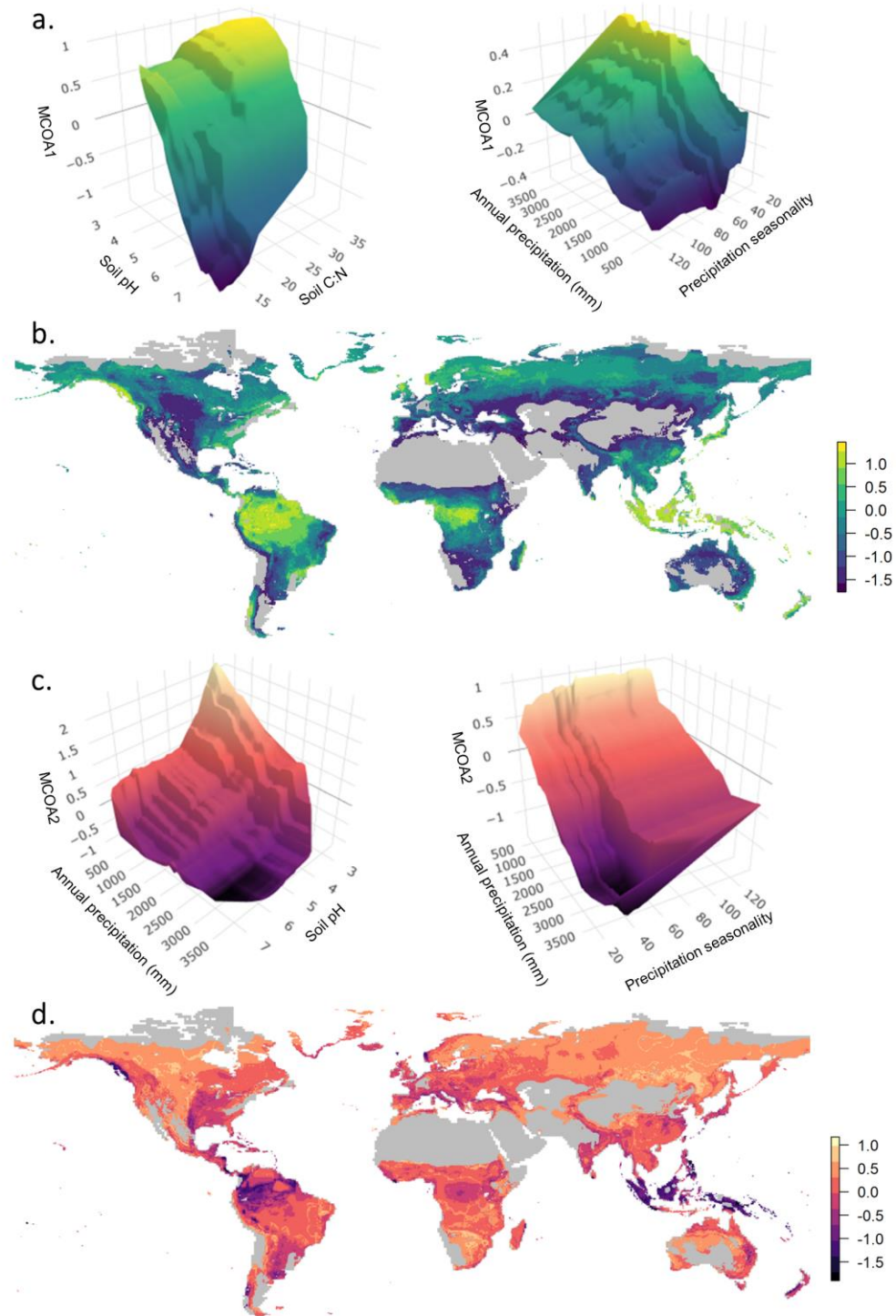


421
 422 *Figure 2. The global life history strategies of soil bacteria communities. Two-dimensional trait space*
 423 *from a MCOA depicting trait associations across soil bacteria communities, with traits inferred from*
 424 *enriched genes in bacteria metagenomes. Dots represent the positions of the 128 bacterial communities*
 425 *used in this study along these two dimensions. In the trait lists, letters in brackets represent how CSR*
 426 *(Competitors, Stress tolerant, Ruderal) and YAS (High Yield, resource Acquisition, Stress tolerance)*
 427 *strategies have been associated with these traits in previous theoretical works (Extended Data Table1).*



428

429 *Figure 3. Hypothesised role of competitor (C), Ruderal (R) and Stress tolerant (S) traits in shaping the*
 430 *life history strategy observed at the community level and associated environmental gradients. Details*
 431 *of CSR traits association are provided in Figure 1 and 2. S, S' and S'', represent the different S traits*
 432 *associated with each dimension and detailed in Figure 1 and 2.*



433

434 *Figure 4. Environmental control and global scale projection of bacterial communities' coordinates*
 435 *along MCOA dimension 1 and 2. a and c, random forest partial dependence plots describing*
 436 *relationships between bacterial communities' coordinates along MCOA dimension 1 (a) and 2 (c) and*
 437 *their most significant environmental predictors (Extended Data Figure 5). b and d, random forest*

438 predictions for MCOA dimension 1 (**b**) and 2 (**d**) projected across the globe using broad resolution map
439 of mean soil and climate conditions (1600x1200 pixel), with land out of the dataset range in grey. Colour
440 bars represent the predicted coordinates along MCOA dimension 1 (**b**) and MCOA dimension 2 (**d**).
441 SoilGrids version 2.0 was used for soil properties and Worldclim2 for climate variables. Accuracy of
442 the prediction was verified by ten-fold cross-validation of the random forest (Extended Data Figure 5)
443 and by comparing the predicted values of the broad resolution projection with local observations
444 (Extended Data Figure 6).

445

446 **References**

- 447 1. Fierer, N. Embracing the unknown: disentangling the complexities of the soil microbiome.
448 *Nature Reviews Microbiology* **15**, 579–590 (2017).
- 449 2. Bahram, M. *et al.* Structure and function of the global topsoil microbiome. *Nature* **560**, 233–237
450 (2018).
- 451 3. Delgado-Baquerizo, M. *et al.* A global atlas of the dominant bacteria found in soil. *Science* **359**,
452 320–325 (2018).
- 453 4. Crowther, T. W. *et al.* The global soil community and its influence on biogeochemistry. *Science*
454 **365**, eaav0550 (2019).
- 455 5. Wieder, W. R., Bonan, G. B. & Allison, S. D. Global soil carbon projections are improved by
456 modelling microbial processes. *Nature Climate Change* **3**, 909–912 (2013).
- 457 6. Diaz, S. *et al.* The global spectrum of plant form and function. *Nature* **529**, 167 (2016).
- 458 7. Grime, J. P. Evidence for the existence of three primary strategies in plants and its relevance to
459 ecological and evolutionary theory. *The American Naturalist* **111**, 1169–1194 (1977).
- 460 8. Wright, I. J. *et al.* The worldwide leaf economics spectrum. *Nature* **428**, 821 (2004).
- 461 9. Southwood, T. R. Habitat, the templet for ecological strategies? *Journal of animal ecology* **46**,
462 337–365 (1977).
- 463 10. Reich, P. B. *et al.* The evolution of plant functional variation: traits, spectra, and strategies.
464 *International Journal of Plant Sciences* **164**, S143–S164 (2003).
- 465 11. Krause, S. *et al.* Trait-based approaches for understanding microbial biodiversity and ecosystem
466 functioning. *Frontiers in Microbiology* **5**, (2014).
- 467 12. Malik, A. A. *et al.* Defining trait-based microbial strategies with consequences for soil carbon
468 cycling under climate change. *The ISME Journal* **14**, 1–9 (2019).
- 469 13. Madin, J. S. *et al.* A synthesis of bacterial and archaeal phenotypic trait data. *Scientific Data* **7**, 1–
470 8 (2020).
- 471 14. Westoby, M. *et al.* Trait dimensions in bacteria and archaea compared to vascular plants. *Ecology*
472 *Letters* (2021).
- 473 15. Steen, A. D. *et al.* High proportions of bacteria and archaea across most biomes remain
474 uncultured. *The ISME journal* **13**, 3126–3130 (2019).
- 475 16. Martiny, A. C. High proportions of bacteria are culturable across major biomes. *The ISME*
476 *Journal* **13**, 2125–2128 (2019).
- 477 17. Martiny, A. C. The “1% culturability paradigm” needs to be carefully defined. *The ISME journal*
478 **14**, 10–11 (2020).

- 479 18. Fierer, N., Barberán, A. & Laughlin, D. C. Seeing the forest for the genes: using metagenomics to
480 infer the aggregated traits of microbial communities. *Frontiers in microbiology* **5**, (2014).
- 481 19. Garnier, E. *et al.* Plant functional markers capture ecosystem properties during secondary
482 succession. *Ecology* **85**, 2630–2637 (2004).
- 483 20. Violle, C. *et al.* Let the concept of trait be functional! *Oikos* **116**, 882–892 (2007).
- 484 21. Fierer, N. *et al.* Comparative metagenomic, phylogenetic and physiological analyses of soil
485 microbial communities across nitrogen gradients. *The ISME journal* **6**, 1007–1017 (2012).
- 486 22. Sorensen, J. W., Dunivin, T. K., Tobin, T. C. & Shade, A. Ecological selection for small
487 microbial genomes along a temperate-to-thermal soil gradient. *Nature microbiology* **4**, 55–61
488 (2019).
- 489 23. Grime, J. P. & Pierce, S. *The evolutionary strategies that shape ecosystems*. (John Wiley & Sons:
490 2012).
- 491 24. Liu, H. *et al.* Warmer and drier ecosystems select for smaller bacterial genomes in global soils.
492 *iMeta* e70 (2023).
- 493 25. Simonsen, A. K. Environmental stress leads to genome streamlining in a widely distributed
494 species of soil bacteria. *The ISME Journal* **16**, 423–434 (2021).
- 495 26. Chuckran, P. F. *et al.* Edaphic controls on genome size and GC content of bacteria in soil
496 microbial communities. *Soil Biology and Biochemistry* **178**, 108935 (2023).
- 497 27. Guieysse, B. & Wuertz, S. Metabolically versatile large-genome prokaryotes. *Current Opinion in*
498 *Biotechnology* **23**, 467–473 (2012).
- 499 28. Konstantinidis, K. T. & Tiedje, J. M. Trends between gene content and genome size in
500 prokaryotic species with larger genomes. *Proceedings of the National Academy of Sciences* **101**,
501 3160–3165 (2004).
- 502 29. Paul, C. *et al.* Bacterial spores, from ecology to biotechnology. *Advances in applied microbiology*
503 **106**, 79–111 (2019).
- 504 30. Singh, S., Datta, S., Narayanan, K. B. & Rajnish, K. N. Bacterial exo-polysaccharides in
505 biofilms: role in antimicrobial resistance and treatments. *Journal of Genetic Engineering and*
506 *Biotechnology* **19**, 1–19 (2021).
- 507 31. Sinsabaugh, R. L. & Follstad Shah, J. J. Ecoenzymatic stoichiometry and ecological theory.
508 *Annual Review of Ecology, Evolution, and Systematics* **43**, 313–343 (2012).
- 509 32. Buckeridge, K. M. *et al.* Environmental and microbial controls on microbial necromass recycling,
510 an important precursor for soil carbon stabilization. *Communications Earth &mathsemicolon*
511 *Environment* **1**, (2020).
- 512 33. Zheng, Q. *et al.* Growth explains microbial carbon use efficiency across soils differing in land use
513 and geology. *Soil Biology and Biochemistry* **128**, 45–55 (2019).
- 514 34. Gao, Y. & Wu, M. Free-living bacterial communities are mostly dominated by oligotrophs.
515 *bioRxiv* 350348 (2018).
- 516 35. Li, J. *et al.* Predictive genomic traits for bacterial growth in culture versus actual growth in soil.
517 *The ISME journal* **13**, 2162–2172 (2019).
- 518 36. Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J. & Segata, N. Shotgun metagenomics,
519 from sampling to analysis. *Nature biotechnology* **35**, 833–844 (2017).
- 520 37. Lobb, B., Tremblay, B. J.-M., Moreno-Hagelsieb, G. & Doxey, A. C. An assessment of genome
521 annotation coverage across the bacterial tree of life. *Microbial Genomics* **6**, (2020).
- 522 38. Coelho, L. P. *et al.* Towards the biogeography of prokaryotic genes. *Nature* **601**, 252–256 (2022).
- 523 39. Martiny, J. B., Jones, S. E., Lennon, J. T. & Martiny, A. C. Microbiomes in light of traits: a
524 phylogenetic perspective. *Science* **350**, aac9323 (2015).
- 525 40. Allison, S. D. & Martiny, J. B. Resistance, resilience, and redundancy in microbial communities.
526 *Proceedings of the National Academy of Sciences* **105**, 11512–11519 (2008).
- 527 41. Jones, D. L., Cooledge, E. C., Hoyle, F. C., Griffiths, R. I. & Murphy, D. V. pH and
528 exchangeable aluminum are major regulators of microbial energy flow and carbon use efficiency
529 in soil microbial communities. *Soil Biology and Biochemistry* 107584 (2019).

- 530 42. Fernández-Calviño, D. & Bååth, E. Growth response of the bacterial community to pH in soils
531 differing in pH. *FEMS microbiology ecology* **73**, 149–156 (2010).
- 532 43. Auger, C. *et al.* Metabolic reengineering invoked by microbial systems to decontaminate
533 aluminum: implications for bioremediation technologies. *Biotechnology advances* **31**, 266–273
534 (2013).
- 535 44. Bruelheide, H. *et al.* Global trait–environment relationships of plant communities. *Nature*
536 *Ecology & Evolution* **2**, 1906–1917 (2018).
- 537 45. Tedersoo, L. *et al.* Regional-scale in-depth analysis of soil fungal diversity reveals strong pH and
538 plant species effects in Northern Europe. *Frontiers in Microbiology* **11**, 1953 (2020).
- 539 46. Bagousse-Pinguet, Y. L. *et al.* Testing the environmental filtering concept in global drylands.
540 *Journal of Ecology* **105**, 1058–1069 (2017).
- 541 47. Lauber, C. L., Hamady, M., Knight, R. & Fierer, N. Pyrosequencing-based assessment of soil pH
542 as a predictor of soil bacterial community structure at the continental scale. *Appl. Environ.*
543 *Microbiol.* **75**, 5111–5120 (2009).
- 544 48. Meyer, F. *et al.* The metagenomics RAST server—a public resource for the automatic
545 phylogenetic and functional analysis of metagenomes. *BMC bioinformatics* **9**, 1–8 (2008).
- 546 49. Chen, I.-M. A. *et al.* IMG/M v. 5.0: an integrated data management and comparative analysis
547 system for microbial genomes and microbiomes. *Nucleic acids research* **47**, D666–D677 (2019).
- 548 50. Lozupone, C. & Knight, R. UniFrac: a new phylogenetic method for comparing microbial
549 communities. *Applied and environmental microbiology* **71**, 8228–8235 (2005).
- 550 51. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The
551 carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic acids research* **42**, D490–D495
552 (2014).
- 553 52. Nguyen, L. T. *et al.* Responses of the soil microbial community to nitrogen fertilizer regimes and
554 historical exposure to extreme weather events: Flooding or prolonged-drought. *Soil Biology and*
555 *Biochemistry* **118**, 227–236 (2018).
- 556 53. Berlemont, R. & Martiny, A. C. Genomic potential for polysaccharide deconstruction in bacteria.
557 *Appl. Environ. Microbiol.* **81**, 1513–1519 (2015).
- 558 54. López-Mondéjar, R. *et al.* Metagenomics and stable isotope probing reveal the complementary
559 contribution of fungal and bacterial communities in the recycling of dead biomass in forest soil.
560 *Soil Biology and Biochemistry* **148**, 107875 (2020).
- 561 55. Nayfach, S. & Pollard, K. S. Toward accurate and quantitative comparative metagenomics. *Cell*
562 **166**, 1103–1116 (2016).
- 563 56. Chávez, J., Devos, D. P. & Merino, E. Complementary tendencies in the use of regulatory
564 elements (transcription factors, sigma factors, and riboswitches) in bacteria and archaea. *Journal*
565 *of bacteriology* **203**, 413–20 (2020).
- 566 57. Cania, B. *et al.* Site-specific conditions change the response of bacterial producers of soil
567 structure-stabilizing agents such as exopolysaccharides and lipopolysaccharides to tillage
568 intensity. *Frontiers in microbiology* **11**, 568 (2020).
- 569 58. Finn, D., Yu, J. & Penton, C. R. Soil quality shapes the composition of microbial community
570 stress response and core cell metabolism functional genes. *Applied Soil Ecology* **148**, 103483
571 (2020).
- 572 59. Sharma, M. P. *et al.* Deciphering the role of trehalose in tripartite symbiosis among rhizobia,
573 arbuscular mycorrhizal fungi, and legumes for enhancing abiotic stress tolerance in crop plants.
574 *Frontiers in microbiology* **11**, 509919 (2020).
- 575 60. Yaakop, A. S. *et al.* Characterization of the mechanism of prolonged adaptation to osmotic stress
576 of *Jeotgalibacillus malaysiensis* via genome and transcriptome sequencing analyses. *Scientific*
577 *reports* **6**, 1–14 (2016).
- 578 61. Boch, J., Kempf, B., Schmid, R. & Bremer, E. Synthesis of the osmoprotectant glycine betaine in
579 *Bacillus subtilis*: characterization of the gbsAB genes. *Journal of Bacteriology* **178**, 5121–5129
580 (1996).

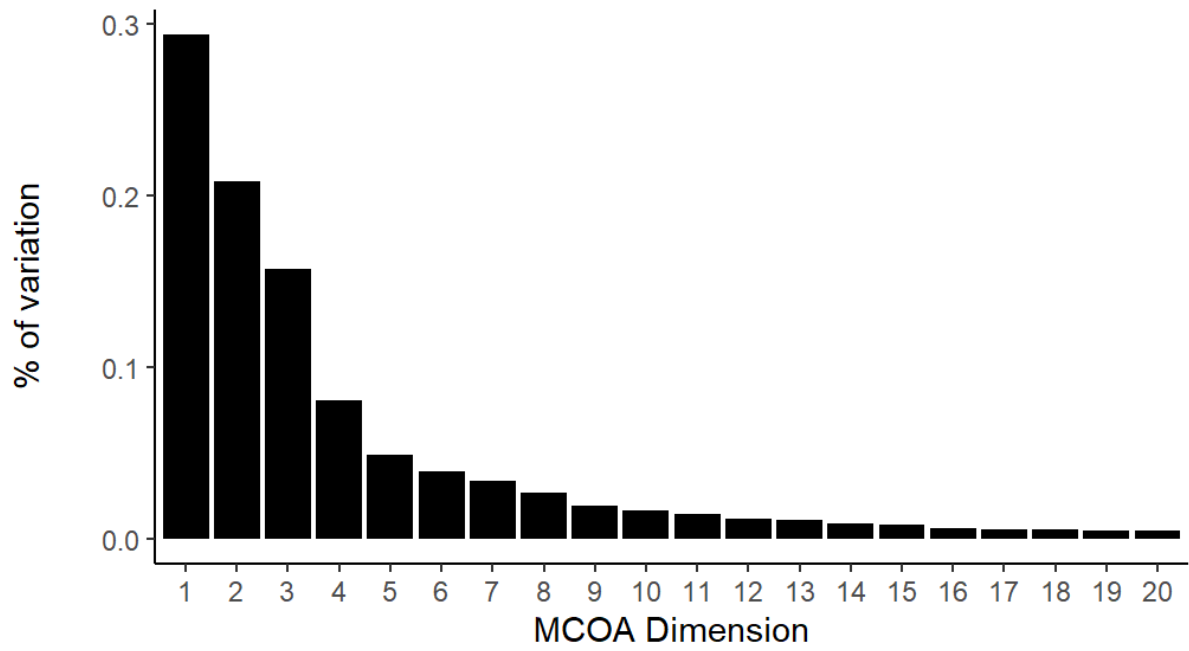
- 581 62. Wargo, M. J. Homeostasis and catabolism of choline and glycine betaine: lessons from
582 *Pseudomonas aeruginosa*. *Applied and environmental microbiology* **79**, 2112–2120 (2013).
- 583 63. Nayfach, S. & Pollard, K. S. Average genome size estimation improves comparative
584 metagenomics and sheds light on the functional ecology of the human microbiome. *Genome*
585 *biology* **16**, 1–18 (2015).
- 586 64. Pereira-Flores, E., Glöckner, F. O. & Fernandez-Guerra, A. Fast and accurate average genome
587 size and 16S rRNA gene average copy number computation in metagenomic data. *BMC*
588 *bioinformatics* **20**, 1–13 (2019).
- 589 65. Chessel, D. & Hanafi, M. Analyses de la co-inertie de K nuages de points. *Revue de statistique*
590 *appliquée* **44**, 35–60 (1996).
- 591 66. Piton, G. *et al.* Using proxies of microbial community-weighted means traits to explain the
592 cascading effect of management intensity, soil and plant traits on ecosystem resilience in
593 mountain grasslands. *Journal of Ecology* **108**, 876–893 (2020).
- 594 67. Meng, C., Kuster, B., Culhane, A. C. & Gholami, A. M. A multivariate approach to the
595 integration of multi-omics datasets. *BMC bioinformatics* **15**, 1–13 (2014).
- 596 68. Dray, S., Dufour, A. B. & Chessel, D. The ade4 package-II: Two-table and K-table methods. *R*
597 *news* **7**, 47–52 (2007).
- 598 69. Dormann, C. F. *et al.* Collinearity: a review of methods to deal with it and a simulation study
599 evaluating their performance. *Ecography* **36**, 27–46 (2013).
- 600 70. Genuer, R., Poggi, J.-M. & Tuleau-Malot, C. VSURF: an R package for variable selection using
601 random forests. *The R Journal* **7**, 19–33 (2015).
- 602 71. Kuhn, M. Building predictive models in R using the caret package. *Journal of statistical software*
603 **28**, 1–26 (2008).
- 604 72. Poggio, L. *et al.* SoilGrids 2.0: producing soil information for the globe with quantified spatial
605 uncertainty. *Soil* **7**, 217–240 (2021).
- 606
- 607

608 **Extended Data**

609 *Extended Data Table 1. Life history traits used in this study. Traits were selected based on their previous*
610 *association with CSR ('Competitor', 'Stress tolerant', and 'Ruderal') strategies by Fierer (2017) [1] or*
611 *Krause et al. (2014) [2] or YAS strategies ("Yield", "Resource acquisition", and "Stress tolerant") by*
612 *Malik et al. (2020) [3]. Cells associated with CSR and YAS have been greyed based on the strategy to*
613 *facilitate comparisons between references. Same gray has been used for C and A, and for R and Y*
614 *strategies as they have some important theoretical linkages (Malik et al. 2020).*

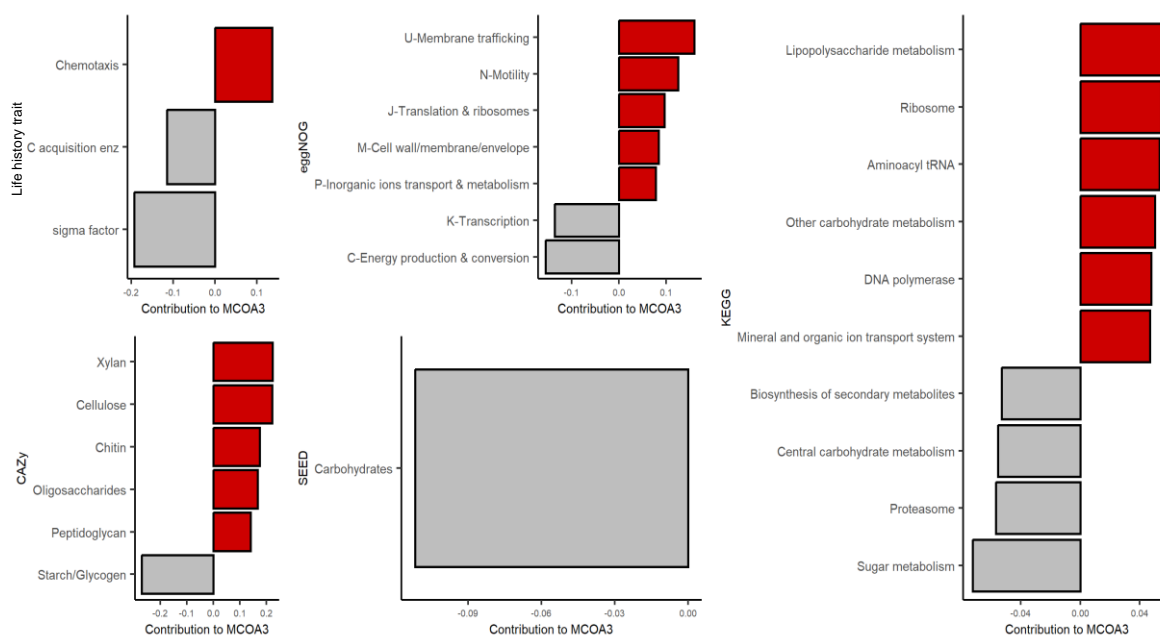
Life history traits	Associated metagenomic community aggregated traits used in this study	CSR [1]	CSR [2]	YAS [3]
Amino acid, fatty acid and nucleotide synthesis [3]	> eggNOG category, KEGG pathway and SEED modules associated with amino acid, lipid and nucleotide metabolism			Y
Chaperons [3]	> Chaperons genes : <i>GroEL</i> (COG0459), <i>dnaK</i> (COG0443) and <i>dnaJ</i> (COG0484) (Malik et al. 2020, Finn et al. 2020)			S
Siderophores [1,3]	> KEGG pathway "Metallic cation iron siderophore and vitamin B12 transport system "	C		A
Oligosaccharides degradation enzymes	> Genes associated with Oligosaccharides degradation among other GH and AA genes			
Carbohydrate central metabolism [3]	> KEGG pathway "Central_carbohydrate_metabolism"			Y
Primary metabolism	> eggNOG categories : F-Nucleotide transport and metabolism, J-Translation and ribosomes, D-Cell cycling, E-Amino acid transport and metabolism, H-Coenzyme transport and metabolism and A-RNA processing and modification, SEED modules : DNA and protein metabolisms. KEGG pathways: Purine, Cysteine, Methionine, Arginine, Proline and Lysine metabolism, Proteasome, cofactors and vitamins metabolisms and Ribosome, Aminoacyl tRNA, RNA and DNA polymerase and Nucleotide sugars			
Genome size [1,2]	> Average genome size (Nayfach and Pollard 2015)	C	R	
Complex polymers degradation enzymes [3]	> Genes associated with Lignin degradation among other GH and AA genes			A
Fungal biomass degradation enzymes	> Genes associated with Chitin and Glucan degradation among other GH and AA genes			
Antibiotic [1,2]	> Antibiotic Resistance Genes	C	C	
Pathogenic interactions with plants	> SEED module : Virulence			
Sporulation [1,2]	> SEED module "Dormancy_and_Sporulation"	R	S	
EPS [1,2,3]	> EPS genes : <i>WcaB</i> (COG1596), <i>WcaF</i> (COG0110), <i>Wza</i> (COG1596), <i>KpsE</i> and <i>RkpR</i> (COG3524) and <i>wcaK</i> (COG2327) (Cania et al. 2020)	S	S	S
Membrane synthesis and repair [3]	> eggNOG categories : L-Replication, recombination & repairs, M-Cell wall, membrane and envelope, KEGG pathways : Lipid and lipopolysaccharide metabolism			S
rRNA gene copies [1,2]	> Average rRNA copy number (Pereira-Flores et al. 2019)	R	C	
Motility [2,3]	> eggNOG category : "N-Motility"		R	A
Chemotaxis [2,3]	> Genes associated with chemotaxis : <i>CheA</i> (COG0643), <i>CheY</i> (COG0784), <i>CheW</i> (COG0835), <i>CheB</i> (COG2201), <i>CheX</i> (COG1406), <i>CheD</i> (COG1871), Methyl-accepting chemotaxis proteins (COG0840, COG1352)		R	A
Sigma factor [3]	> σ factor genes : σD , σS and σH (COG0568), σF and σB (COG1191), σN (COG1508) and extracytoplasmic function σ factors (COG1595) (Chávez et al. 2020)			S
Osmolytes [3]	> Genes associated with Trehalose and glycine betaine (Malik et al. 2020, Sharma et al. 2020, Suriaty Yaakop et al. 2016, Bochet al. 1996, Wargo et al. 2013)			S
Exoenzymes (All) [2,3]	> GH and AA genes in global metabolism		S	A
Bacterial biomass degradation enzyme	> Genes associated with Peptidoglycan degradation			
Uptake system [2,3]	> KEGG pathway and SEED modules associated with transport systems		S	A

616



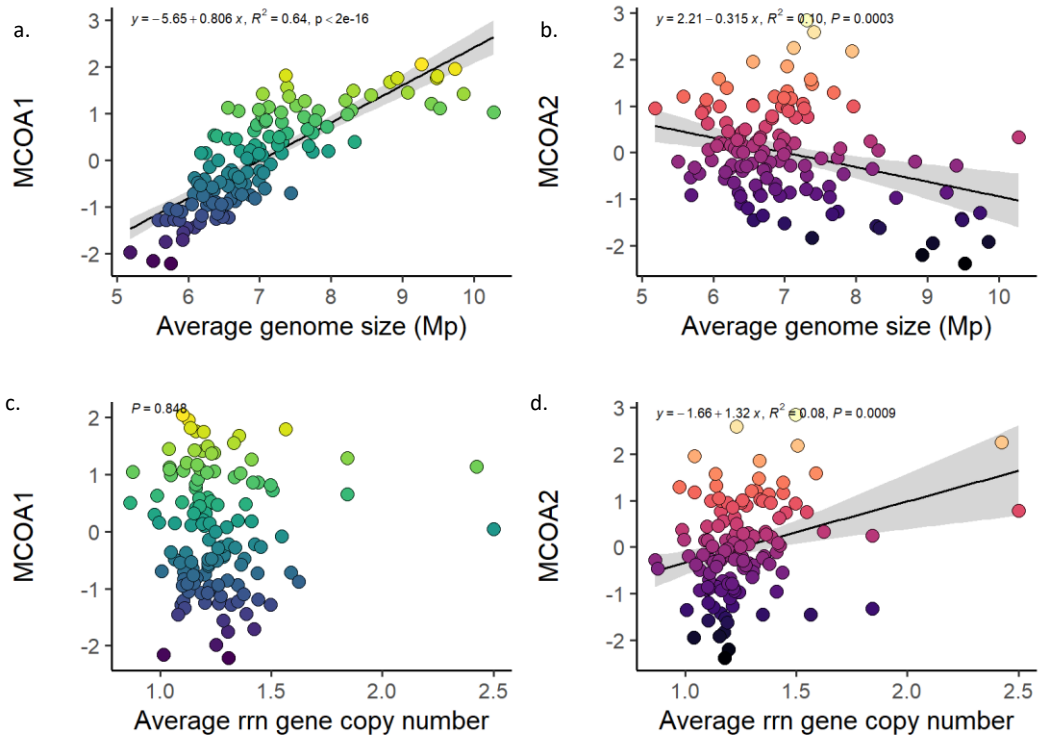
617

618 *Extended Data Figure 1. Stress plot representing the % of variation of the global dataset captured by*
619 *each dimension of the MCOA*



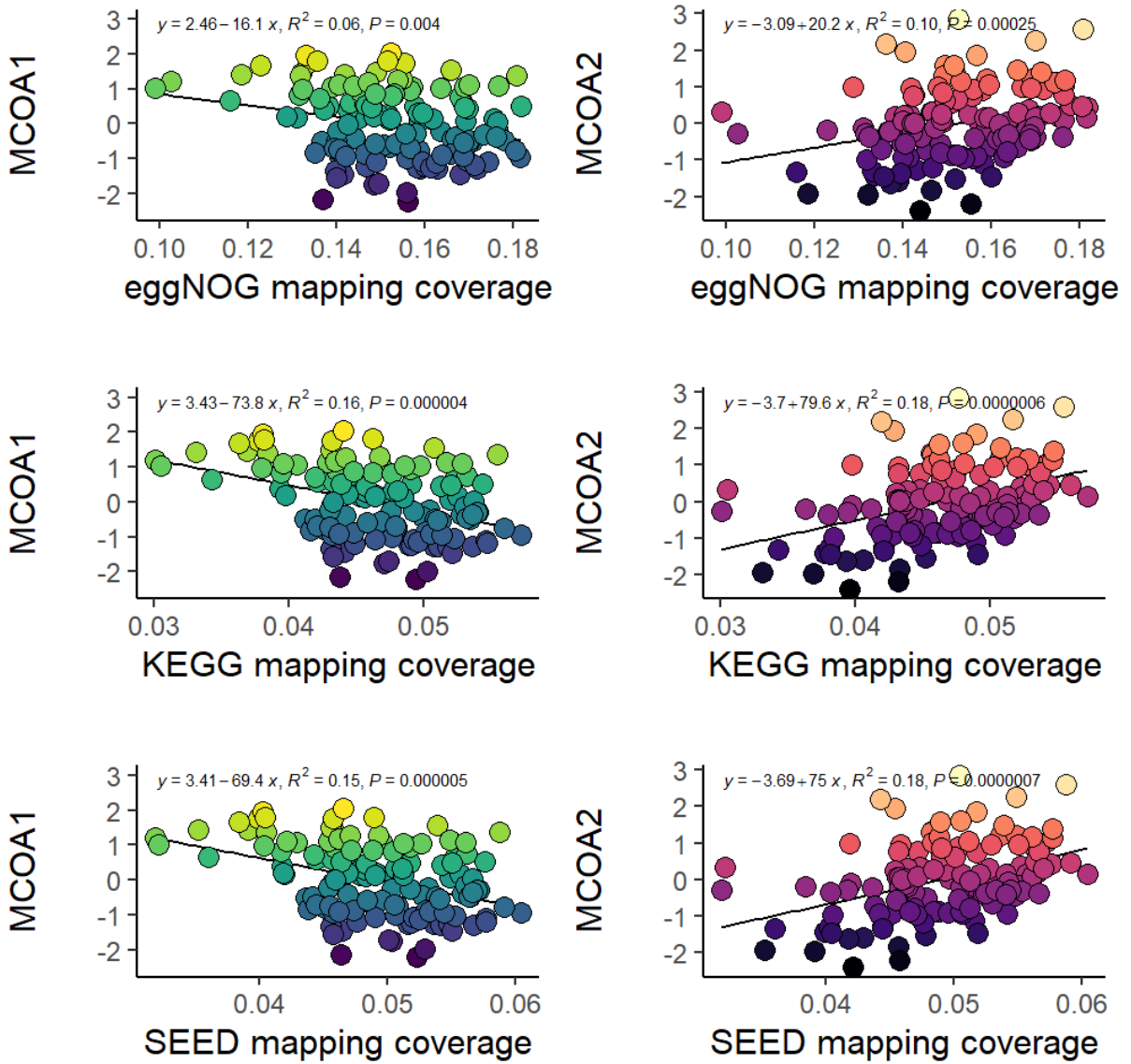
620

621 *Extended Data Figure 2. Variable contributions to the third trait dimension of the multiple co-inertia*
 622 *analysis (MCOA). The MCOA summarizes in a common structure the information shared by 5*
 623 *community aggregated trait (CAT) databases (Genomic trait, CAZy, eggNOG, SEED and KEGG). Only*
 624 *the most important variables with significant correlation ($p < 0.001$) with each dimension are reported*
 625 *in this figure.*



626

627 *Extended Data Figure 3. Correlations between genomic traits and coordinates along dimensions 1 and*
 628 *2 of the MCOA. The P value indicates the significance of the regression slope obtained using a t-test.*
 629 *Shade represents the estimated 95% confidence interval. Color gradients follow MCOA dimensions and*
 630 *match with figure 1 and 3 in the main text.*



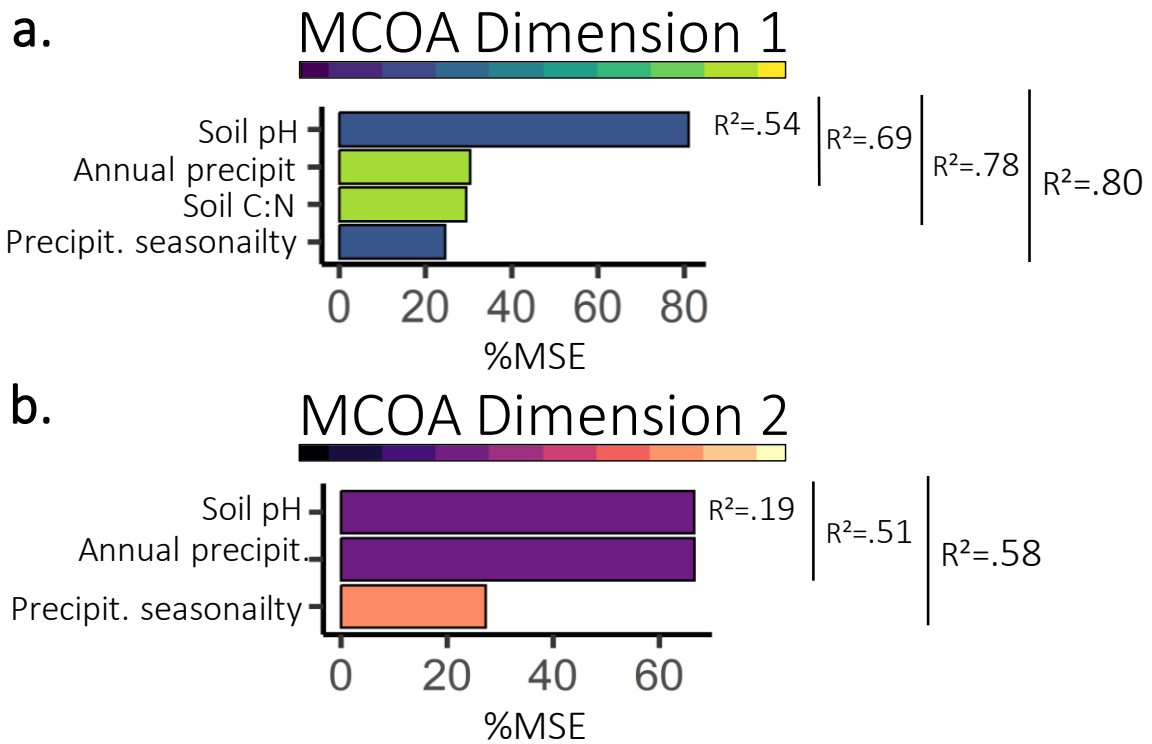
631

632 *Extended Data Figure 4. Correlations between MCOA dimensions (MCOA1 and MCOA2) and mapping*

633 *coverages on the 3 general databases (eggNOG, KEGG, SEED) used in this study. The P value indicates*

634 *the significance of the regression slope obtained using a t-test.*

635



636

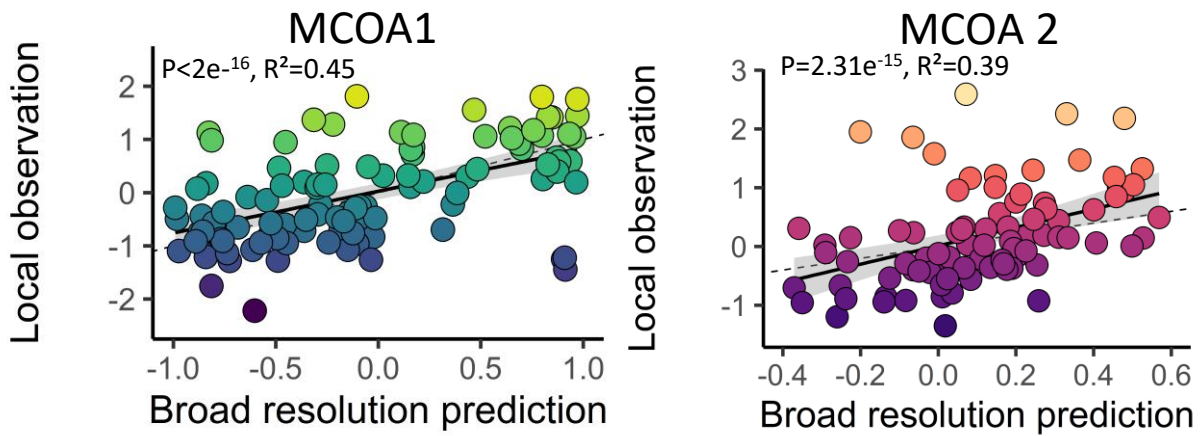
637 *Extended Data Figure 5. Environmental drivers of the bacterial community trait dimensions.*

638 *Environmental variable importances are represented as the mean decrease in mean square error*

639 *(%MSE) and R squared in random forest models predicting MCOA Dimension 1 (a) and 2 (b). Bar*

640 *colours indicate which end of the dimension (Figure 1 and 3) is positively correlated with the variable.*

641



642

643 *Extended Data Figure 6. Correlations between local trait dimension observations and global spatial*

644 *prediction. Correlations between local observations of bacterial community positions along the first*

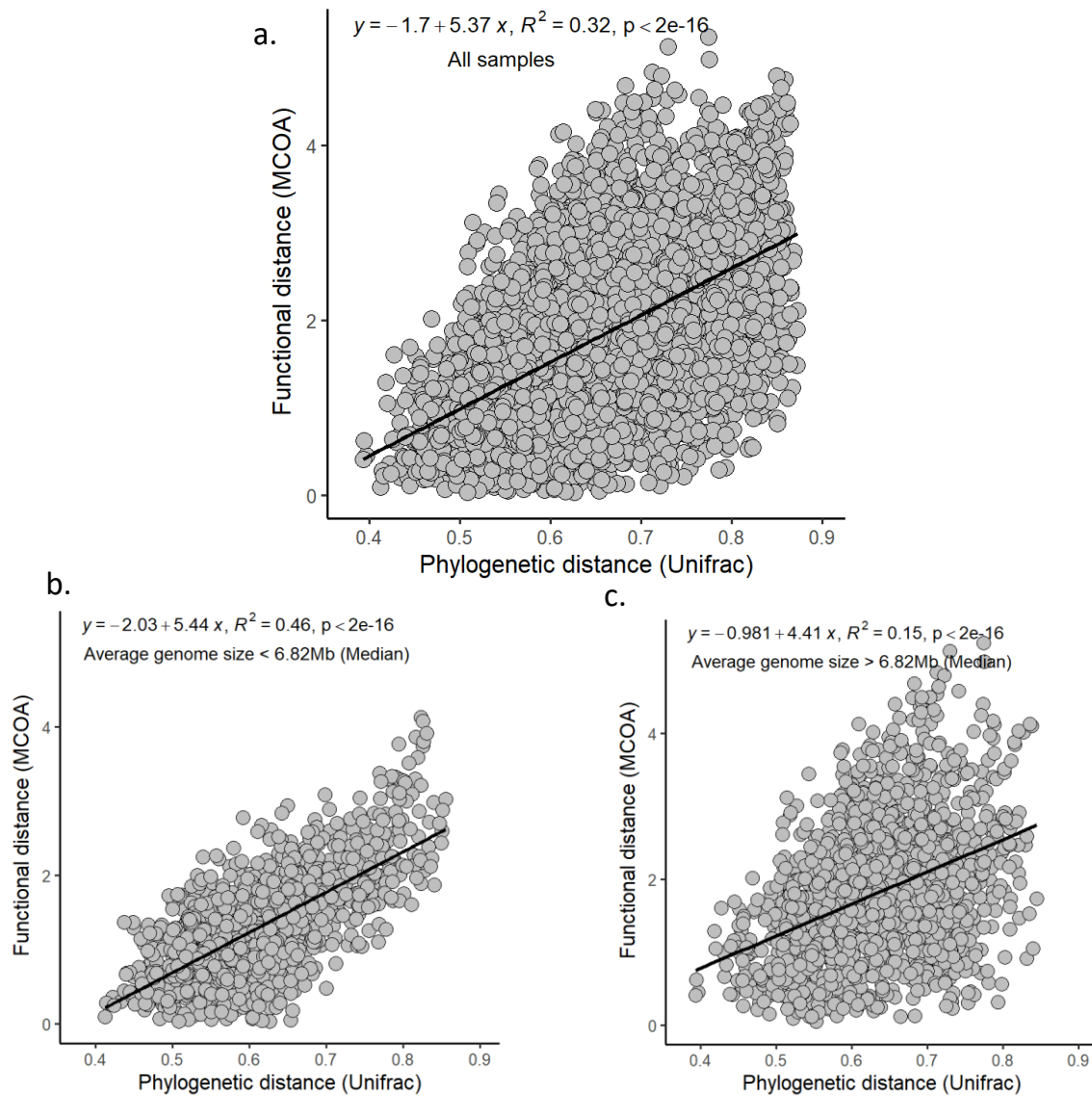
645 *and second trait dimensions from the MCOA (Figure 1-2) and the predicted value of the global map cell*

646 *(Figure 4) corresponding to where the local observations have been done. Dashed line represents a 1:1*

647 *correlation. The P value indicates the significance of the regression slope obtained using a t-test. Shade*

648 *represents the estimated 95% confidence interval. Color gradients follow MCOA dimension and match*

649 *with figure 1,2 and 4 in the main text.*

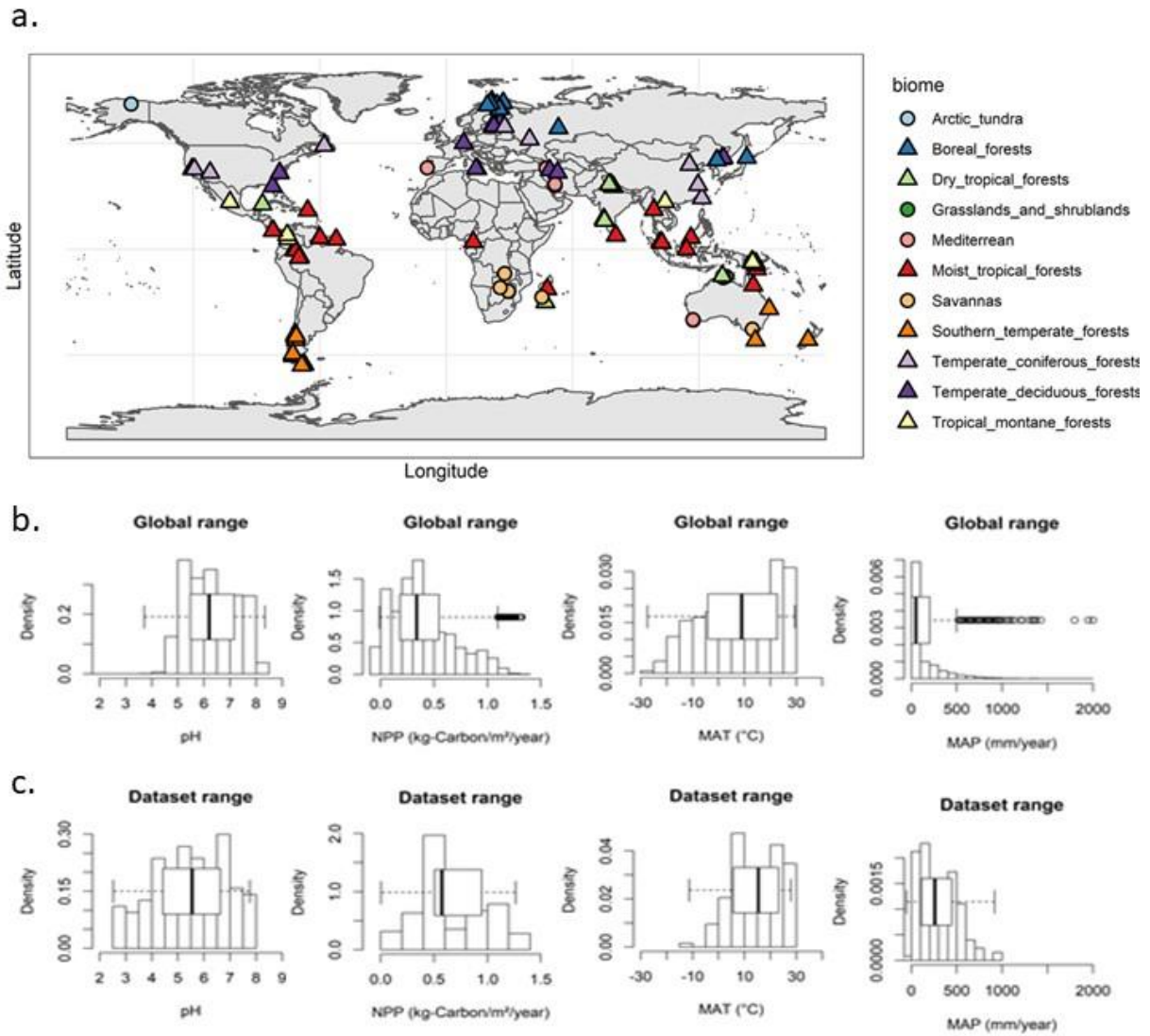


650

651

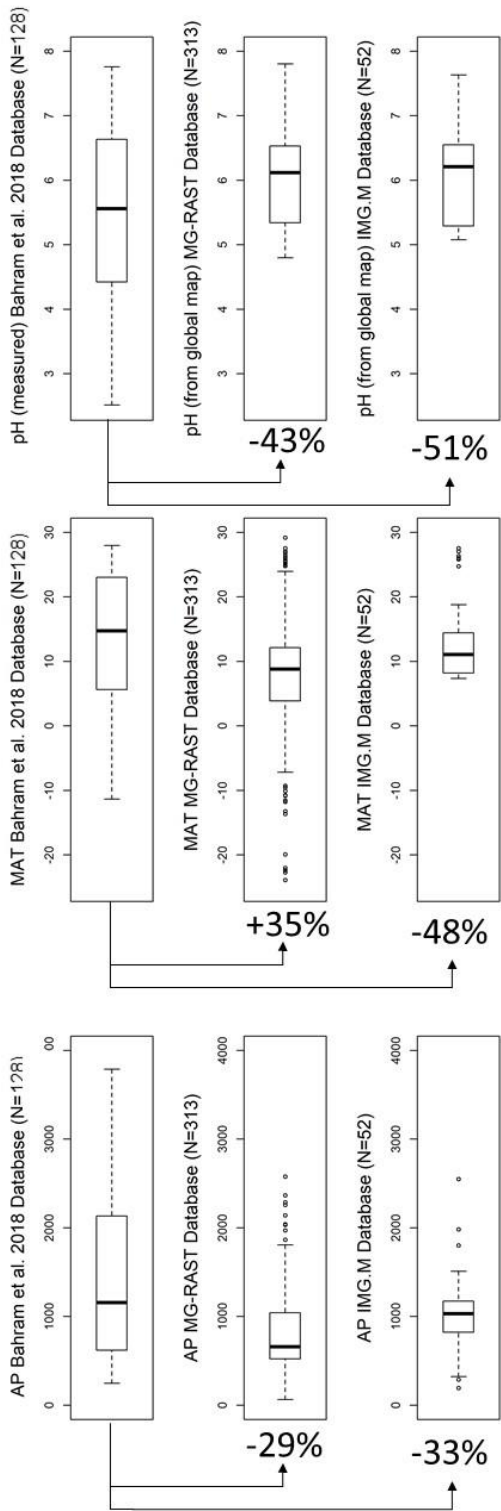
652 *Extended Data Figure 7. Correlation between phylogenetic distance (Unifrac metric) and functional*
 653 *distance (Euclidian distance in MCOA space using coordinates of the two principal dimensions).*
 654 *Correlation for all samples (a) and restricted to samples with average genome size below (b) and*
 655 *above (c) its median value in the dataset. The P value indicates the significance of the regression slope*
 656 *obtained using a t-test.*

657



658

659 *Extended Data Figure 8. Dataset distribution and environmental coverage. a. Sample localisations and*
 660 *associated biomes b-c. Comparison between global range of environmental variables from the Atlas of*
 661 *the Biosphere (b) and the environmental coverage of dataset (n=128) used in this study (c). Boxplot*
 662 *elements: Center line=median; box limits=upper and lower quartiles; whiskers=1.5x interquartile*
 663 *range; points=outliers. World map was done with rnaturalearth R package*
 664 *(<https://github.com/ropensci/rnaturalearth>).*



665

666 *Extended Data Figure 9. Environmental coverage comparison between the database used in this study*

667 *from Bahram et al. (2018) and databases from the main metagenomes repositories (MG-RAST and*

668 *IMG:M). N corresponds to the number of metagenomes available in each database. MAT=Mean Annual*
669 *Temperature, AP=Annual Precipitation. Boxplot elements: Center line=median; box limits=upper and*
670 *lower quartiles; whiskers=1.5x interquartile range; points=outliers.*