Running head

Identifying bacterial functional strategies

Title

Data-driven identification of major axes of functional variation in bacteria

List of authors

Geneviève Lajoie*[1,2,3], Steven W. Kembel[1]

* Email for correspondence (G. Lajoie): genevieve.lajoie@umontreal.ca

[1] Département des Sciences Biologiques, Université du Québec à Montréal, 141 Avenue du Président-Kennedy, Montréal, Canada, H2X 1Y4

Current address (G. Lajoie) :

[2] Jardin Botanique de Montréal, 4101 rue Sherbrooke E, Montréal, Canada, H1X 2B2

[3] Institut de Recherche en Biologie Végétale, Université de Montréal, 4101 rue Sherbrooke E, Montréal, Canada, H1X 2B2

Data availability statement

All metagenomic datasets used in this study can be accessed either on the IMG/M (https://img.jgi.doe.gov/cgi-bin/m/main.cgi) or the MG-RAST (https://www.mg-rast.org/) online databases using the sample identification information provided in Appendix S1: Table S2. The functional and phylogenetic genomic datasets can be accessed within the AnnoTree database (http://annotree.uwaterloo.ca/app/downloads.html) using the sample identification information provided in Appendix S1: Table S3. The novel R code developed for processing and analyzing the data is available for review through Figshare (https://figshare.com/s/2d9835e083f160939d66) and will be made public in a GitHub repository (https://github.com/glajoie1/Bact_Fun_Axes) upon publication of the manuscript.

Abstract

The discovery of major axes of correlated functional variation among species and habitats has revealed the fundamental trade-offs structuring both functional and taxonomic diversity in eukaryotes such as plants. Whether such functional axes exist in the bacterial realm and whether they could explain bacterial taxonomic turnover among ecosystems remains unknown. Here we use a data-driven approach to leverage global genomic and metagenomic datasets to reveal the existence of major axes of functional variation explaining both evolutionary differentiation within Bacteria and their ecological sorting across diverse habitats. We show that metagenomic variation among bacterial communities from various ecosystems is structured along a few axes of correlated functional pathways. Similar clusters of traits explained phylogenetic trait variation among >16,000 bacterial genomes, suggesting that functional turnover among bacterial communities from distinct habitats does not only result from the differential filtering of similar functions among communities, but also from phylogenetic correlations among these functions. Concordantly, functional pathways associated with trait clusters that were most important for defining functional turnover among bacterial communities were also those that had the highest phylogenetic signal in the bacterial genomic phylogeny. This study overall underlines the important role of evolutionary history in shaping contemporary distributions of bacteria across ecosystems.

Originality-Significance Statement

In this article, we use a trait screening approach based on genomic and metagenomic data to identify the key functional strategies of bacteria across ecosystems but also across the bacterial tree of life. This novel approach allows us to quantify the role of evolutionary processes in structuring microbial ecological differences among ecosystems. By reducing the high-

dimensionality of trait variation observed among microorganisms around a small number of fundamental axes of trait covariation, we make a significant step towards generalization of the drivers of biological diversity in microbes but also across study systems. This research provide a major advance in our understanding of the origin and maintenance of bacterial biological diversity, expanding on related findings for plants and animals.

Key Words

Functional strategies, Functional traits, Bacteria, Genomics, Metagenomics, Phylogenetics, Trait evolution

Introduction

The identification of major axes of variation in life-history strategies among organisms and habitats has led to advances of both theory and practice in ecology and evolutionary biology (Grime 1977, Westoby et al. 1998, Lajoie and Kembel 2019). For example, in plants, large-scale screening of plant functional traits (Westoby 1998, Díaz et al. 2016) led to the discovery that a few major axes of covarying traits related to leaf economics, seed mass, and plant height explain most of the global variation in plant functional and life-history strategies. The contribution of this work has been twofold; first, it improved our understanding of the major environmental drivers and evolutionary origins of functional variation among organisms (Ackerly and Reich 1999, Wright et al. 2004). Second, the identification of a subset of measurable functional traits that could be used to characterize ecological strategies helped coordinate measurement efforts around a reduced set of interpretable functions, thus facilitating attempts at generalization among study systems worldwide (Díaz et al. 2016, Madani et al. 2018).

Recent conceptual and technological developments in the study of microbes have led to a

69 proliferation of microbial trait measurements and screenings that have opened the door for
70 extending this research agenda to the microbial realm. Still, functional categorization schemes
71 proposed to date in microbial ecology have mostly consisted of direct translations of functional
72 strategy schemes developed for macro-organisms. Classifications focusing on growth responses
73 of organisms, such as the r-K selection spectrum or Grime's CSR (competitive, stress-tolerant,
74 ruderal) scheme (e.g. Fierer et al. 2007, Evans and Wallenstein 2014, Santillan et al. 2019), have
75 drawn particular interest and been useful in improving our understanding of ecosystem
76 functioning in certain habitats (Malik et al. 2020). The relevance of such schemes in explaining
77 global patterns of microbial functional diversity however remains limited by their reliance on in-
78 vitro phenotypic measurements, still unachievable for a vast diversity of microbial strains (Steen
79 et al. 2019).

80 Metagenomic sequencing data represent a pragmatic alternative for the large-scale
81 characterization of microbial functions. Functional predictions from metagenomic sequences
82 have been useful for identifying sets of functional genes that best discriminate microbial samples
83 across ecosystems (Dinsdale et al. 2008, Fierer et al. 2012, Ramírez-Flandes et al. 2019). For
84 example, genes encoding oxidoreductases were shown to be the most useful of enzyme-encoding
85 genes in separating metagenomic samples into biome groups (e.g. terrestrial vs. aquatic)
86 (Ramírez-Flandes et al. 2019). By constraining their analyses along predetermined
87 environmental axes, these studies allow the identification of functional genes that are
88 distinctively enriched in each habitat in comparison to others. They however fall short in
89 determining whether such 'indicator genes' take part in larger strategies that structure functional
90 variation among microbial communities across diverse habitats – as observed for macro-
91 organisms – or whether they tend to explain functional variation among habitat types

idiosyncratically.

Objectives and hypotheses

Here, we evaluate whether functional variation among bacterial metagenomes from different ecosystems is driven by major groups of correlated traits (i.e. functional strategies). We then compare these groups of traits to those explaining the most variation among a set of phylogenetically diverse bacterial genomes to test hypotheses about the ecological or evolutionary origin of these strategies. We predict that if the presence of correlated axes of trait variation across ecosystems is mostly explained by the filtering of groups of traits differentially selected along environmental gradients, then the groups of traits uncovered in genomes vs. metagenomes are unlikely to be the same between datasets (Appendix S1: Table S1). Reversely, if evolutionary selection or constraints on the variation of these groups of traits through time contribute to structuring functional variation among metagenomes (Futuyma 2010, Muir 2015), then the groups of traits should correspond among datasets. We lastly address the contribution of phylogenetic processes to correlated trait variation across metagenomes by testing if traits belonging to groups explaining the most variation among metagenomes also have stronger phylogenetic signal than other traits, and determine the phylogenetic depth at which they are most structured. Overall, this study contributes to a more general understanding of the role of functional strategies in driving bacterial distributions across ecosystems and of their evolutionary origins.

Methods

Metagenomic dataset collection and processing

We searched the IMG/M (Chen et al. 2019) and the MG-RAST (Keegan et al. 2016)

114 databases for shotgun metagenomic sequence datasets from environmental and host-associated
115 sources. We restricted our search to types of habitats for which we could obtain data from >1
116 study and for at least 5 different samples. To limit biases due to sequencing technology (Clooney
117 et al. 2016), only datasets that had been sequenced with an Illumina system were included. To
118 uniformize functional annotations among databases, we performed functional and taxonomic
119 annotations of the IMG/M sequence datasets with the standard MG-RAST annotation pipeline
120 (Keegan et al. 2016) using BLAST annotation identity criteria of a maximum e-value of 1e-10
121 and minimum percent identity of 30%, and a minimum sequence length of 125 nucleotides. To
122 facilitate comparisons with the genomic dataset, metagenomic sequences were filtered to the
123 bacterial domain only, which represented most annotated sequences (~94% of all taxonomic
124 annotations). Past this filtering step, we only kept samples that had at least 100 000 sequences
125 with both functional and taxonomic gene annotations, resulting in a dataset of 114 samples
126 representing ten major bacterial habitats (Appendix S1: Table S2).

127 To control for differences in sequencing depth among samples, we randomly picked
128 100 000 sequences from each sample to generate functional and taxonomic community
129 composition tables. Rarefying at that level was sufficient to capture the diversity of functional
130 genes and taxonomic groups present in each sample (Appendix S2: Figure S1). To generate a
131 functional composition table that had more interpretability and lower dimensionality, we
132 summed counts of functional genes for each sample by functional pathway (i.e. Tier 3 KEGG
133 category), the least aggregated of the three levels of the KEGG BRITE gene hierarchy (Kanehisa
134 et al. 2014). Compositional data was transformed into relative abundance data for each sample.
135 The resulting functional composition table spanned 245 functional pathways over 114 samples.
136 A principal coordinates analysis of this dataset constrained by database of origin (i.e. IMG or

137 MG-RAST) confirmed that database did not drive much variation among samples (1.9% of
138 variation explained). The relative abundance of each taxon in each sample was similarly
139 compiled at the phylum, class and order levels using the RefSeq taxonomic classification (Leary
140 et al. 2016). The final taxonomic datasets spanned 28 phyla, 52 classes, 121 orders and 248
141 families of bacteria across all samples (Appendix S2: Figure S2).

142 *Genomic dataset collection and processing*

143 The genomic dataset was retrieved from the AnnoTree server (accessed Dec. 2019)
144 (Mendler et al. 2019). It consists of the functional annotations, taxonomic annotations and
145 phylogenetic relationships of more than 27,000 fully sequenced bacterial genomes. We
146 simplified the dataset by randomly retaining a single genome at each tip of the phylogeny,
147 leading to a final dataset of 15,973 genomes (Appendix S1: Table S3). The taxonomic identity of
148 all genomes retained spanned 111 phyla, 270 classes, 735 orders and 1659 families. The
149 taxonomic composition of bacterial genomes in the genomic dataset encompassed the same
150 major phyla recorded across metagenomes (Appendix S2: Figure S2). Counts of protein-coding
151 genes for each genome were transformed into relative abundances of KEGG functional pathways
152 as with the metagenomic dataset. We overall retrieved 312 pathways across genomes.

153 *Identification of main functional axes in the metagenomic dataset*

154 We used a hierarchical clustering approach to identify groups of correlated functions that
155 could constitute major ecological strategies across metagenomic samples. We first generated a
156 principal coordinate analysis (PCoA) of the metagenomic functional dataset based on the
157 Hellinger distance between samples using R package *ape* (Paradis and Schliep 2018). We then
158 fitted the functional pathways onto the ordination to obtain vectors maximizing the correlation of

159 these pathways with each ordination axis with R package *vegan* (Oksanen et al. 2013). To
160 identify correlated axes of functional variation, we then calculated distances between functional
161 pathways using the coordinates of these vectors across the first four PCoA axes (together
162 explaining 74% of variance in the ordination). We used the resulting distance matrix to perform a
163 hierarchical agglomerative clustering of functional pathways using Ward's minimum variance
164 method (D2) to minimize within-cluster variance. We identified clusters of correlated traits (i.e.
165 trait clusters) by visual examination of the dendrogram based on their contribution to functional
166 variation across the first 4 axes of the PCoA. The number of clusters in the dendrogram was
167 confirmed based on a majority-ruling of 30 indices calculating the optimal number of clusters in
168 a dendrogram using R package *NbClust* (Charrad et al. 2014). We then tested whether the
169 number of pathways from each functional category differed among clusters when compared with
170 their observed prevalence in the dataset using a chi-square test on counts of functional pathways
171 by Tier 2 functional category in the KEGG classification framework, keeping only Tier 2
172 categories for which there were at least 10 observations across clusters.

173 *Drivers of functional variation among metagenomes*

174 We evaluated if the richness of functional pathways within any sample varied among
175 habitats using a type III analysis of variance controlling for unequal sample sizes among factors
176 (Weisberg 2019). We compared habitat types in pairs using least square means comparisons
177 using multivariate-*t* corrections with the R package *emmeans* (Lenth 2022). We then tested
178 whether differences in functional richness among samples explained the functional distances
179 among samples using a PERMANOVA. We also evaluated how much variation in the functional
180 composition among samples could be explained by habitat type, taxonomic composition (class-
181 level) and their interaction using a variation partitioning analysis. These analyses were

182 performed using R package *vegan* (Oksanen et al. 2013).

183 *Phylogenetic structure of functional variation in the genomic dataset*

184 We assessed the phylogenetic covariation of functional traits by performing a
185 phylogenetic principal components analysis (pPCA) on the genomic functional dataset, using R
186 package *adephylo* (Jombart and Dray 2010). This approach finds axes of functional variation that
187 maximize the product of variance of the scores and their phylogenetic autocorrelation, revealing
188 axes of correlated trait variation that are phylogenetically structured (Jombart et al. 2010). We
189 calculated functional distances among pathways using the loadings obtained through this
190 ordination over the first 4 axes of the pPCA, which encompassed ~65% of explained variance.
191 We used these distances as described previously to perform a hierarchical clustering analysis of
192 functional pathways using Ward's minimum variance method.

193 We compared the structure of functional covariation in the metagenomic and genomic
194 dataset with a Procrustes analysis performed on the cophenetic distance matrices describing each
195 of the two hierarchical clusterings. The two dendrograms were pruned beforehand to include the
196 same functional pathways. The significance of the Procrustes statistic, describing the similarity
197 between the two datasets, was assessed by comparing the observed statistic to a distribution of
198 999 statistics generated through permutations of the original data (Oksanen et al. 2013). We
199 generated a tanglegram of the two clusterings for visual comparison of their respective functional
200 groupings. The trees were untangled prior to plotting using R package *dendextend* (Galili 2015).

201 *Phylogenetic signal and depth of ecologically important functions*

202 We next evaluated whether pathways that drove functional differences among ecosystems
203 tended to have high phylogenetic signal and to vary at more recent or ancient scales in the

204 bacterial phylogeny. To test this, we calculated the phylogenetic signal of all functional
205 pathways at increasing depths in the phylogeny using a phylogenetic correlogram. Confidence
206 intervals were estimated through bootstrapping (n=100) of autocorrelation estimates after re-
207 standardizations of phylogenetic weights (Keck et al. 2016). We then tested the effect of cluster
208 affiliation, phylogenetic depth, and their interaction on the phylogenetic signal of traits using a
209 Type III ANOVA for unbalanced designs. Differences among pairs of groups were tested at each
210 depth using contrasts on estimated marginal means, as implemented in R package *emmeans*
211 (Lenth 2022). We also tested whether the loadings of each pathway in the metagenomic PCoA
212 was correlated with the strength of phylogenetic signal (absolute value of the signal) at each of 5
213 different depths using a linear model for each of the first four PCoA dimensions.

214 **Results**

215 *Main strategies driving functional variation among metagenomes*

216 Four major axes of functional variation explained nearly 75% of variance in the
217 composition of functional pathways among metagenomic samples (Figure 1). These axes
218 corresponded to functional transitions among major types of habitats, namely between animal
219 guts, plants and soil, and aquatic habitats across the first 2 axes of variation (Figure 1ab). The
220 third axis was characterized by functional variation between soil and aquatic sediment bacterial
221 communities and saltwater ones, while the transition from (i) free-living and epiphytic habitats to
222 (ii) endophytic ones characterized the fourth axis. Functional pathways could be divided into 9
223 main groups of correlated traits based on their contribution to functional variation across these
224 axes (Figure 2a, Appendix S2: Figure S3, Appendix S1: Table S4). The proportions of functional
225 categories to which the functional pathways belonged also differed significantly among groups
226 (Tier-1 categories: $X^2(32, N=246) = 110.6$, $p<0.001$; Tier-2 categories: $X^2(216, N=246) =$

227 397.62, $p<0.001$) (Figure 2b).

228 Habitat type explained an important portion of turnover in functional composition among
229 samples (PERMANOVA: $F(9,104) = 11.917$, $p=0.001$, $R^2 = 51\%$). The functional composition
230 of trait clusters can thus be described in relation to the habitats with which they are most
231 associated. Namely, a cluster (C2) composed of traits pertaining to genetic information
232 processing (i.e. replication and repair; translation; folding sorting and degradation) and
233 nucleotide metabolism explained the functional differentiation of animal gut samples as
234 compared with soil and plants samples (Axis 1). A cluster (C9) characterized by traits related to
235 xenobiotics biodegradation and metabolism, amino acid metabolism as well as cell growth and
236 death explained the variation in the opposite direction. Along the second axis, functional
237 pathways (C6) participating in carbohydrate metabolism, energy metabolism (i.e.
238 photosynthesis) and the metabolism of cofactors and vitamins were together enriched in samples
239 from aquatic communities, while pathways (C1) participating in glycan biosynthesis and
240 metabolism, metabolism of other amino acids and cell motility were enriched in terrestrial ones
241 (both free-living and host-associated).

242 A cluster of traits (C4) involved predominantly in the metabolism of cofactors and
243 vitamins and in environmental information processing (i.e. signal transduction) explained
244 variation along the third axis and were enriched in the saltwater communities as opposed to the
245 soil one. No clear predominance of any functional group emerges for the cluster (C5) correlated
246 with the opposite side of this axis. Compared to other clusters, those (C3, C8) driving variation
247 along the 4th axis both contained pathways involved in prokaryote community sensing and cell-
248 cell interactions. In contrast with cluster C8 associated with endophytic Porifera and plant
249 microbiomes, the cluster (C3) correlated with non-endophytic communities was also enriched in

250 functions related to the biosynthesis of other secondary metabolites.

251 While functional differences among samples could be linked to the turnover in the
252 identity of pathways that characterize these communities, differences in the richness of
253 functional pathways among samples explained 14% of variation in functional distances among
254 samples in a constrained ordination. Indeed, the functional diversity of individual samples varied
255 among bacterial habitats, as revealed for both richness (Type-III ANOVA: $F(9,104) = 7.855$,
256 $p<0.001$) and Shannon diversity (Type-III ANOVA: $F(9,104) = 7.695$, $p<0.001$). Bacterial
257 communities of animal guts (e.g. mammal, bird and insect guts) were generally the least
258 functionally diverse while communities associated with plants (e.g. vascular plants and
259 bryophytes) and soil were the most functionally diverse (Appendix S2: Figure S4). Controlling
260 for the effect of richness on the PCoA however did not affect to a large extent the configuration
261 of samples in the ordination (m12 squared = 0.5071, corr. = 0.7020, sig. = 0.001, # permutations
262 = 999). Most variation in the functional composition of samples among habitats could be
263 explained by turnover in the taxonomic composition of these communities (Figure 1bd,
264 Appendix S2: Figure S5).

265 *Phylogenetic structure of correlated trait variation*

266 The first four axes of the phylogenetic PCA explained 65% of phylogenetic variation in
267 the genomic functional dataset. Clusters of functional pathways that were important in
268 explaining functional variation among ecosystems were also important in explaining
269 phylogenetic variation in the functional composition of genomes (m12 squared = 0.3618, corr. =
270 0.7989, sig. = 0.001, # permutations = 999) (Figure 3). Across all traits, the intensity of
271 phylogenetic signal varied significantly as a function of depth, with higher signal at lower depths
272 and lowest signal at intermediate depths. While the signal was positive at lowest depths, it

273 became negative about mid-way towards the root of the tree. The strength of phylogenetic signal
274 also varied significantly among metagenomic functional clusters and in interaction with
275 phylogenetic depth (Figure 4, Appendix S1: Table S5). Clusters 2 and 9 exhibited more
276 phylogenetic signal than other clusters at low phylogenetic depths, while clusters 1,5 and 7
277 exhibited the least. The latter clusters were also the least phylogenetically structured at deeper
278 depths, while cluster 6 and cluster 2 were the most structured. These differences can be linked
279 with variation in the mean phylogenetic depths of traits belonging to different Tier 1 category,
280 with genetic information processing pathways (enriched in Cluster 2) showing the most signal
281 across phylogenetic depths (Appendix S2: Figure S6). Functional pathways that contributed most
282 to variation among metagenomic samples along the first PCoA axis (characterized by a trade-off
283 between clusters 2 and 9) also consistently displayed the strongest signal in the genomic
284 phylogeny (Appendix S2: Figure S7, Appendix S1: Table S6). This relationship was observed
285 across most depths but was stronger at low phylogenetic depths. It was weaker or null across the
286 other PCoA dimensions.

287 **Discussion**

288 In this study, we show that functional variation among bacterial communities sampled
289 across diverse habitats is structured by clusters of correlated functions defining contrasting life-
290 history strategies among major bacterial habitat types. The structure of these clusters derived
291 from metagenomic annotation data was similar to that obtained from the clustering of functional
292 pathways among bacterial clades using a genomic dataset. This concordance suggests that
293 functional turnover among bacterial communities from distinct habitats does not only result from
294 the differential filtering of similar functions among these communities, but also from
295 phylogenetic correlations among these functions. Functional pathways associated with trait

296 clusters that were most important for defining functional turnover among bacterial communities
297 were also those that were the most phylogenetically structured, underlining the important role of
298 evolutionary history in shaping contemporary distributions of bacteria across ecosystems.

299 *Ecological significance of functional clusters characterizing contrasting habitat types*

300 A few main clusters of correlated functional pathways explained functional variation
301 among bacterial communities in our metagenomic datasets. The prevalence of genetic
302 information processing (i.e. replication and repair; translation; folding sorting and degradation)
303 and nucleotide metabolism in gut bacterial communities can namely be explained by the
304 importance of DNA repair systems for bacterial colonization of intestines (Vogel-Scheel et al.
305 2010, Davies et al. 2011). Reactive nitrogen species and oxidative stress indeed play an
306 important role in host defense against pathogens and have been documented to regulate
307 microbiome assembly in animal hosts (O'Rourke et al. 2003, Fang and Vasquez-Torres 2020).

308 A functional cluster composed of pathways related to xenobiotics degradation and
309 metabolism and to amino acid metabolism were comparatively important in explaining bacterial
310 community composition in soil and plant microbiota. The prevalence of these metabolic
311 functions can be linked to the variety of aromatic compounds (i.e. xenobiotics) and amino acids
312 that plant tissues contain and exude in the soil, which represent a common and varied source of
313 carbon for terrestrial bacteria (Moe 2013, Pérez-Pantoja et al. 2015). The ability to degrade plant
314 aromatic compounds was for example linked to a better colonization and growth of bacteria in
315 rhizosphere habitats (Ledger et al. 2012). Degradation of antibiotic compounds secreted by other
316 microbes and macro-organisms are also important for survival in the dense and diverse biotic
317 networks that interact in the soil (Lucas et al. 2019).

318 Functional pathways linked with the conversion of light to sugar (i.e. photosynthesis and
319 the TCA cycle), and with the metabolism of cofactors and vitamins were predominant in
320 bacterial communities found in aquatic habitats, and especially in saltwater, outlining the
321 importance of autotrophy for bacterial life in aquatic ecosystems. Light-driven proton pumps (i.e.
322 proteorhodopsins) are for example used by a variety of *Flavobacteriia* to generate energy in
323 high-light marine environments (Kumagai et al. 2018). The enrichment of these types of
324 functions in aquatic habitats was contrasted with the enrichment of cell motility pathways in
325 terrestrial habitats, which could be explained by a more general role for bacterial motility in
326 tracking nutrients and carbon sources in the soil (i.e. for heterotrophic lifestyles - Turnbull et al.
327 2001), or in engaging with plants and animals (Akahoshi and Bevins 2022). Finally, the
328 prevalence of cofactors and vitamin metabolic pathways in saltwater bacteria can be linked to the
329 documented depletion of several vitamins in oceanic habitats, which has been proposed as an
330 important filter for planktonic community assembly (Sañudo-Wilhelmy et al. 2012).

331 Our analyses also identified clusters that explained variation between endophytic and
332 bacteria and other host-associated bacteria. These clusters both contained functions linked with
333 community sensing and cell-cell interaction, which likely facilitate establishment on, and
334 interaction with the host. Functional pathways involved in biofilms were more present in non-
335 endophytic bacteria, which can be linked to their role in facilitating colonization of host external
336 and internal surfaces (Bogino et al. 2013, Motta et al. 2021). Quorum sensing functions were
337 comparatively enriched in endophytic communities where they can contribute to bacterial entry
338 in the host tissues (Gosai et al. 2020).

339 The functional classification of bacterial communities portrayed here among habitats is
340 similar in some ways to those suggested in other studies of microbial functional variation.

341 Dinsdale et al. (2008) identified functions associated with certain habitats, but focused on broad,
342 high-level functional categories. Other studies (e.g. Ramírez-Flandes et al. 2019, Malik et al.
343 2020) found that different genes allowed the functional classification of microbial communities
344 into different habitat types but focused on a more limited range of habitat types (for example, on
345 soil microbial communities). By expanding analyses of bacterial functions to identify functional
346 strategies across metagenomes and genomes without an a priori selection of subsets of functions
347 or habitats to examine, our analyses revealed that major axes of bacterial trait variation are
348 structured by clusters of correlated traits that correspond to coherent functional categories across
349 diverse habitats. It remains an open question whether ecological strategies identified for
350 macroorganisms such as plants are relevant for understanding the strategies of microorganisms.
351 In this study, we did not find clear mappings between major groups of correlated traits
352 explaining variation among bacterial genomes and habitats and the competitor-stress tolerator-
353 ruderal (CSR) scheme of Grime (1977), whose application to microbial communities has
354 recently been proposed (Fierer 2017). Our results overall emphasize the role of data-driven
355 approaches in providing novel insights into the ecological and evolutionary factors that drive
356 microbial trait variation.

357 *Phylogenetic structure of functional pathways and clusters*

358 Phylogenetically constrained functional clusters based on genomic data were generally
359 correlated with those obtained from metagenomic data. Concordantly, clusters of functions that
360 were associated with the first axis of the PCoA and thus explained the most functional variation
361 among habitats (C2, C9) were also the most phylogenetically 'conserved', having the strongest
362 signal in the genomic phylogeny across most depths. These results suggest that selection or
363 constraint (e.g. genetic correlations (Futuyma 2010) or functional barriers to recombination

364 (Hendrickson et al. 2018)) have been important in shaping the evolution of clusters of correlated
365 traits in bacteria in a consistent fashion across both clades and habitats, leading to predictable
366 variation in the ecological strategies of bacteria associated with particular clades. The fact that
367 we found such clusters despite the likely presence of many transient organisms in environmental
368 metagenomes provides additional confidence in these results.

369 Despite identifying three main axes of functional covariation among bacterial traits, other
370 drivers of functional trait correlations appeared to play a role in structuring trait variation among
371 ecosystems (Appendix S1: Table S1). First, some metagenomically-derived trait clusters (e.g.
372 C4, C5, C8) were not observed in the genomic dataset, suggesting that they might result from
373 environmental filtering on pre-existing trait variation or lateral gene transfer among host-
374 associated clades rather than a fundamental constraint or trade-off in genome architecture and
375 evolution. These clusters still showed some level of phylogenetic signal in the genomic
376 phylogeny, but more so at shallower phylogenetic depths, consistent with filtering of genes
377 transferred horizontally among related organisms living in a shared habitat (Shapiro et al. 2012).
378 Second, traits belonging to the cluster C7 did not explain much variation across the axes and
379 were not enriched in any particular type of high-level functional categories, suggesting they do
380 not represent part of a functional strategy that has arisen in response to the major axes of
381 environmental variation encompassed by this dataset. Their lack of correlations in the genomic
382 dataset as well suggests that they are likely to have been evolving idiosyncratically or to be
383 subject to more extensive horizontal gene transfer. It is possible that other types of functional
384 strategies than those identified above are also important for driving within-community niche
385 partitioning among bacteria, but at the broad scales of the bacterial tree of life and among
386 metagenomes from distinct habitat types, these types of trade-offs are not important enough to be

consistently identified as clusters of covarying traits. Testing how important the main functional axes identified in this study are in explaining microbial fitness at smaller spatial and environmental scales will represent a valuable next research direction for the field.

Conclusion

In conclusion, we used a data-driven functional trait screening approach to identify the major axes of functional trait covariation in bacterial metagenomes and genomes based on gene functional annotations. By comparing functional trait clusters in both datasets, we show that the major strategies driving functional differentiation between metagenomes of co-occurring bacteria in ecological communities tend to be phylogenetically structured among bacterial genomes. We also point to the limited overlap between the bacterial ecological strategy axes described here and the trait classification schemes developed for plants and animals. By reducing the high-dimensionality of trait variation observed among microorganisms around a small number of fundamental axes of trait covariation, we make an important step towards generalization of microbial ecology and of the drivers of biological diversity and bacterial biogeography across study systems.

Acknowledgments

We are grateful to Jesse Shapiro and Mario Muscarella for comments on a version of this manuscript. We acknowledge funding from the Natural Sciences and Engineering Research Council of Canada (GL, SWK) and the Canada Research Chairs (SWK).

Literature cited

Ackerly, D. D., and P. B. Reich. 1999. Convergence and correlations among leaf size and function in seed plants: comparative tests using independent contrast. American Journal of

409 Botany 86:1272–1281.
410 Akahoshi, D. T., and C. L. Bevins. 2022. Flagella at the host-microbe interface : Key functions
411 intersect with redundant responses. Frontiers in Immunology 13:828758.
412 Bogino, P. C., M. de la Mercedes Oliva, F. Sorroche, and W. Giordano. 2013. The role of
413 bacterial biofilms and surface components in plant-bacterial associations. International Journal
414 of Molecular Sciences 14:15838–15859.
415 Charrad, M., N. Ghazzali, V. Boiteau, and A. Niknafs. 2014. NbClust: An R Package for
416 determining the relevant number of clusters in a data set. Journal of Statistical Software 61:1–
417 36.
418 Chen, I. A., K. Chu, K. Palaniappan, M. Pillay, A. Ratner, J. Huang, M. Huntemann, N.
419 Varghese, J. R. White, R. Seshadri, T. Smirnova, E. Kirton, S. P. Jungbluth, T. Woyke, N. N.
420 Ivanova, and N. C. Kyrpides. 2019. IMG/M v.5.0 : an integrated data management and
421 comparative analysis system for microbial genomes and microbiomes. Nucleic Acids Research
422 47:666–677.
423 Clooney, A. G., F. Fouhy, R. D. Sleator, and A. O. Driscoll. 2016. Comparing apples and
424 oranges ?: Next generation sequencing and its impact on microbiome analysis. PLoS ONE
425 11:e0148028.
426 Davies, B. W., R. W. Bogard, N. M. Dupes, T. A. I. Gerstenfeld, L. A. Simmons, and J. John.
427 2011. DNA damage and reactive nitrogen species are barriers to *Vibrio cholerae* colonization
428 of the infant mouse intestine. PLoS Pathogens 7:e1001295.
429 Díaz, S., J. Kattge, J. H. C. Cornelissen, I. J. Wright, S. Lavorel, S. Dray, B. Reu, M. Kleyer, C.
430 Wirth, I. C. Prentice, E. Garnier, G. Bönisch, M. Westoby, H. Poorter, P. B. Reich, A. T.
431 Moles, J. Dickie, A. N. Gillison, A. E. Zanne, J. Chave, S. J. Wright, S. N. Sheremet'ev, H.

Jactel, B. Christopher, B. Cerabolini, S. Pierce, B. Shipley, D. Kirkup, F. Casanoves, J. S. Joswig, A. Günther, V. Falczuk, N. Rüger, M. D. Mahecha, and L. D. Gorné. 2016. The global spectrum of plant form and function. Nature 529:167–171.

Dinsdale, E. A., R. A. Edwards, D. Hall, F. Angly, M. Breitbart, J. M. Brulc, M. Furlan, C. Desnues, M. Haynes, L. Li, L. McDaniel, M. A. Moran, K. E. Nelson, C. Nilsson, R. Olson, J. Paul, B. R. Brito, Y. Ruan, B. K. Swan, R. Stevens, D. L. Valentine, R. V. Thurber, L. Wegley, B. A. White, and F. Rohwer. 2008. Functional metagenomic profiling of nine biomes. Nature 452:629–632.

Evans, S. E., and M. D. Wallenstein. 2014. Climate change alters ecological strategies of soil bacteria. Ecology Letters 17:155–164.

Fang, F. C., and A. Vasquez-Torres. 2020. Reactive nitrogen species in host-bacterial interactions. Current Opinion in Immunology 60:96–102.

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

Fierer, N., M. A. Bradford, and R. B. Jackson. 2007. Toward an ecological classification of soil bacteria. Ecology 88:1354–1364.

Fierer, N., J. W. Leff, B. J. Adams, U. N. Nielsen, S. Thomas, C. L. Lauber, S. Owens, J. A. Gilbert, D. H. Wall, and J. G. Caporaso. 2012. Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. PNAS 109:21390–21395.

Futuyma, D. J. 2010. Evolutionary constraint and ecological consequences. Evolution 64:1865–1884.

Galili, T. 2015. dendextend: an R package for visualizing, adjusting, and comparing trees of hierarchical clustering. Bioinformatics 31:3718–3720.

455 Genz, A., F. Bretz, T. Miwa, X. Mi, F. Leisch, F. Scheipl, and T. Hothorn. 2021. mvtnorm:
456 Multivariate normal and t distributions. R Package version 1.1-3.
457 Gosai, J., S. Anandhan, A. Bhattacharjee, and G. Archana. 2020. Elucidation of quorum sensing
458 components and their role in regulation of symbiotically important traits in Ensifer nodulating
459 pigeon pea. Microbiological Research 231:126354.
460 Grime, J. P. 1977. Evidence for the existence of three primary strategies in plants and its
461 relevance to ecological and evolutionary theory. The American Naturalist 111:1169–1194.
462 Hendrickson, H. L., D. Barbeau, R. Ceschin, and J. G. Lawrence. 2018. Chromosome
463 architecture constrains horizontal gene transfer in bacteria. PLoS Genetics 14:e1007421.
464 Jombart, T., and S. Dray. 2010. adephylo: exploratory analyses for the phylogenetic comparative
465 method. Bioinformatics 26:1907–1909.
466 Jombart, T., S. Pavoine, S. Devillard, and D. Pontier. 2010. Putting phylogeny into the analysis
467 of biological traits : a methodological approach. Journal of Theoretical Biology 264:693–701.
468 Kanehisa, M., S. Goto, Y. Sato, M. Kawashima, M. Furumichi, and M. Tanabe. 2014. Data,
469 information, knowledge and principle: Back to metabolism in KEGG. Nucleic Acids Research
470 42:199–205.
471 Keck, F., F. Rimet, A. Bouchez, and A. Franc. 2016. phylosignal: an R package to measure, test,
472 and explore the phylogenetic signal. Ecology and Evolution 6:2774–2780.
473 Keegan, K., E. Glass, and F. Meyer. 2016. MG-RAST, a metagenomics service for analysis of
474 microbial community structure and function. Methods in Molecular Biology 1399:207–233.
475 Kumagai, Y., S. Yoshizawa, Y. Nakajima, M. Watanabe, T. Fukunaga, Y. Ogura, T. Hayashi, K.
476 Oshima, M. Hattori, and M. Ikeuchi. 2018. Solar-panel and parasol strategies shape the
477 proteorhodopsin distribution pattern in marine *Flavobacteriia*. The ISME Journal 12:1329–

478 1343.

479 Lajoie, G., and S. W. Kembel. 2019. Making the most of trait-based approaches for microbial
480 ecology. Trends in Microbiology 27:814–823.

481 Leary, N. A. O., M. W. Wright, J. R. Brister, S. Ciufo, D. Haddad, R. Mcveigh, B. Rajput, B.
482 Robbertse, B. Smith-white, D. Ako-adjei, A. Astashyn, A. Badretdin, Y. Bao, O. Blinkova, V.
483 Brover, V. Chetvernin, J. Choi, E. Cox, O. Ermolaeva, C. M. Farrell, T. Goldfarb, T. Gupta, D.
484 Haft, E. Hatcher, W. Hlavina, S. Joardar, V. K. Kodali, W. Li, D. Maglott, P. Masterson, M.
485 Mcgarvey, M. R. Murphy, K. O. Neill, S. Pujar, S. H. Rangwala, D. Rausch, L. D. Riddick, C.
486 Schoch, A. Shkeda, S. S. Storz, H. Sun, F. Thibaud-nissen, I. Tolstoy, R. E. Tully, R. Vatsan,
487 C. Wallin, D. Webb, W. Wu, M. J. Landrum, A. Kimchi, T. Tatusova, M. Dicuccio, P. Kitts, T.
488 D. Murphy, and K. D. Pruitt. 2016. Reference sequence (RefSeq) database at NCBI : current
489 status, taxonomic expansion, and functional annotation. Nucleic Acids Research 44:733–745.

490 Ledger, T., A. Zuniga, T. Kraiser, P. Dasencich, R. Donoso, D. Pérez-Pantoja, and B. Gonzalez.
491 2012. Aromatic compounds degradation plays a role in colonization of *Arabidopsis thaliana*
492 and *Acacia caven* by *Cupriavidus pinatubonensis* JMP134. Antonie van Leeuwenhoek
493 101:713–723.

494 Lenth, R. 2022. Package 'Estimated Marginal Means, aka Least-Squares Means'.

495 Lucas, J. M., E. Gora, A. Salzberg, M. Kaspari, and J. M. Lucas. 2019. Antibiotics as chemical
496 warfare across multiple taxonomic domains and trophic levels in brown food webs.
497 Proceedings of the Royal Society B: Biological Sciences.

498 Madani, N., J. S. Kimball, A. P. Ballantyne, D. L. R. Affleck, P. M. Van Bodegom, P. B. Reich,
499 J. Kattge, A. Sala, M. Nazeri, M. O. Jones, M. Zhao, and S. W. Running. 2018. Future global
500 productivity will be affected by plant trait response to climate. Scientific Reports 8:2870.

501 Malik, A. A., J. B. H. Martiny, E. L. Brodie, A. C. Martiny, K. K. Treseder, and S. D. Allison.
502 2020. Defining trait-based microbial strategies with consequences for soil carbon cycling under
503 climate change. The ISME Journal 14:1–9.
504 Mendler, K., H. Chen, D. H. Parks, B. Lobb, L. A. Hug, and A. C. Doxey. 2019. AnnoTree :
505 visualization and exploration of a functionally annotated microbial tree of life. Nucleic Acids
506 Research 47:4442–4448.
507 Moe, L. 2013. Amino acids in the rhizosphere: from plants to microbes. American Journal of
508 Botany 100:1692–1705.
509 Motta, J., J. Wallace, A. Buret, C. Deraison, and N. Vergnolle. 2021. Gastrointestinal biofilms in
510 health and disease. Nature Reviews Gastroenterology & Hepatology 18:314–334.
511 Muir, C. D. 2015. Making pore choices: repeated regime shifts in stomatal ratio. Proceedings of
512 the Royal Society B: Biological Sciences 282:20151498.
513 O'Rourke, E. J., C. Chevalier, A. V. Pinto, J. M. Thiberge, L. Ielpi, and J. P. Radicella. 2003.
514 Pathogen DNA as target for host-generated oxidative stress : Role for repair of bacterial DNA
515 damage in *Helicobacter pylori* colonization. PNAS. 100:2789–2794.
516 Oksanen, A. J., F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. Mcglinn, P. R. Minchin,
517 R. B. O. Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, and E. Szoecs. 2013. Package '
518 vegan'
519 Paradis, E., and K. Schliep. 2018. ape 5.0: an environment for modern phylogenetics and
520 evolutionary analyses in R. Bioinformatics 35:526–528.
521 Pérez-Pantoja, D., P. Leiva-Novoa, R. A. Donoso, C. Little, M. Godoy, D. H. Pieper, and B.
522 Gonzàlez. 2015. Hierarchy of carbon source utilization in soil bacteria : Hegemonic preference
523 for benzoate in complex aromatic compound mixtures degraded by *Cupriavidus pinatubonensis*

strain JMP134. Applied and Environmental Microbiology 81:3914–3924.

Ramírez-Flandes, S., B. González, and O. Ulloa. 2019. Redox traits characterize the organization of global microbial communities. PNAS 116:3630–3635.

Santillan, E., H. Seshan, F. Constancias, and S. Wuertz. 2019. Trait-based life-history strategies explain succession scenario for complex bacterial communities under varying disturbance. Environmental Microbiology 21:3751–3764.

Sañudo-Wilhelmy, S. A., L. S. Cutter, R. Durazo, E. A. Smail, L. Gómez-Consarnau, E. Webb, M. Prokopenko, W. Berelson, and D. Karl. 2012. Multiple B-vitamin depletion in large areas of the coastal ocean. PNAS 109:14041–14045.

Shapiro, B., J. Friedman, O. Cordero, S. Preheim, S. Timberlake, G. Szabó, M. F. Polz, and E. J. Alm. 2012. Population genomics of early events in the ecological differentiation of bacteria. Science 336:48–51.

Steen, A. D., A. C. P. Carini, K. G. Lloyd, J. C. Thrash, K. M. Deangelis, and N. Fierer. 2019. High proportions of bacteria and archaea across most biomes remain uncultured. The ISME Journal:3126–3130.

Turnbull, G. A., J. A. W. Morgan, J. M. Whipps, and J. R. Saunders. 2001. The role of bacterial motility in the survival and spread of *Pseudomonas fluorescens* in soil and in the attachment and colonisation of wheat roots. FEMS Microbiology Ecology 36:21–31.

Vogel-Scheel, J., C. Alpert, W. Engst, G. Loh, M. Blaut, and H. N. Potsdam-rehbru. 2010. Requirement of purine and pyrimidine synthesis for colonization of the mouse intestine by Escherichia coli. Applied and Environmental Microbiology 76:5181–5187.

Weisberg, F. 2019. An R Companion to Applied Regression. Sage, Thousand Oaks, California.

Westoby, M. 1998. A leaf-height-seed (LHS) plant ecology strategy scheme. Plant and Soil

199:213–227.

Wright, I. J., P. B. Reich, M. Westoby, D. D. Ackerly, Z. Baruch, F. Bongers, J. Cavender-bares, T. Chapin, J. H. C. Cornelissen, M. Diemer, J. Flexas, E. Garnier, P. K. Groom, and J. Gulias. 2004. The worldwide leaf economics spectrum. Nature 428:821–827.

553 **Figure captions**

554 **Figure 1. Principal coordinates analysis (PCoa) of bacterial functional pathways across**
555 **habitats.** Communities are represented by alphanumeric symbols where letters designate the
556 habitat (A=Aves, B=Bryophyte, FW=Freshwater, I=Insecta, M=Mammalia, P=Porifera, S=Soil,
557 SW=Saltwater, SWS=Saltwater sediment, VP=Vascular plants) and numbers designate the
558 sample ID within that habitat (see Appendix S1: Table S2). Ellipses are drawn for each habitat
559 and correspond to a 0.95 confidence interval from the group centroid. Functional pathways were
560 fitted onto dimensions 1 and 2 (panel A), and dimensions 3 and 4 (panel C) of the ordination.
561 Only the 20 functional pathways that were the most correlated with each set of dimensions are
562 represented. They are colored by functional cluster (see Figure 2). Taxonomic classes composing
563 these same communities were also fitted onto dimensions 1 and 2 (panel B) and dimensions 3
564 and 4 (panel D). All taxonomic classes whose relative abundances across samples were
565 significantly correlated with each set of dimensions ($\alpha<0.001$) are represented on the panels.

566 **Figure 2. Hierarchical clustering of bacterial functional pathways across metagenomic**
567 **samples (panel A) and contribution of functional pathways to the composition of each**
568 **cluster, per Tier 1 and Tier 2 KEGG category (panel B).** Panel A: Each dot at the end of
569 branch represents a functional pathway, for which color codes for Tier 1 KEGG functional
570 category (see legend). Each cluster (labelled C1 to C9) is color-coded to facilitate visual
571 identification in other figures. The standardized explanatory power of each functional pathway
572 across each of the first four dimensions of the metagenomic PCoA is indicated in a heatmap
573 below the clustering. In the heatmap, blue is positively correlated with the dimension, and red is
574 negatively correlated. The relative abundance of each pathway across all samples is also
575 represented. Panel B: Dot size is proportional to the residuals of a chi-square test evaluating how

576 different from normal is the distribution of functional pathways of a given functional category
577 among clusters. The more functional pathways of a given category there is in a cluster compared
578 to the expectation, the bigger and the bluer the dot (see legend). The less functional pathways of
579 a given category there is in a cluster compared to the expectation, the bigger and the redder the
580 dot. Color-coding of Tier-2 functional category follows that of panel A.

581 **Figure 3. Tanglegram comparing functional clusters calculated from correlations between**
582 **bacterial functional pathways across (left) metagenomic samples from different habitats**
583 **and across (right) bacterial genomes.** Hierarchical clustering of the genomic data is
584 phylogenetically constrained (see Methods). Dots at the end of each branch are color-coded by
585 the Tier-1 KEGG functional category to which the functional pathway belongs. Lines connect
586 the same functional pathways among datasets and are color-coded by metagenomic cluster.

587 **Figure 4. Phylogenetic signal (Moran's *I*) of functional pathways among clusters and across**
588 **phylogenetic depths.** Moran's *I* was calculated per pathway using phylogenetic correlograms
589 across the bacterial genomic phylogeny, from the tip (depth = 0) towards the root of the tree
590 (depth = 0.0444). Black dots represent the value of estimated marginal means of Moran's I
591 across the functional pathways of each cluster at each depth. The blue bars are confidence
592 intervals. The absence of overlap among red arrows of two clusters at a given depth indicates
593 that their mean phylogenetic signals are significantly different. These pairwise comparisons were
594 adjusted for multiple testing using the multivariate *t* distribution method (Genz et al. 2021).

Figure 1

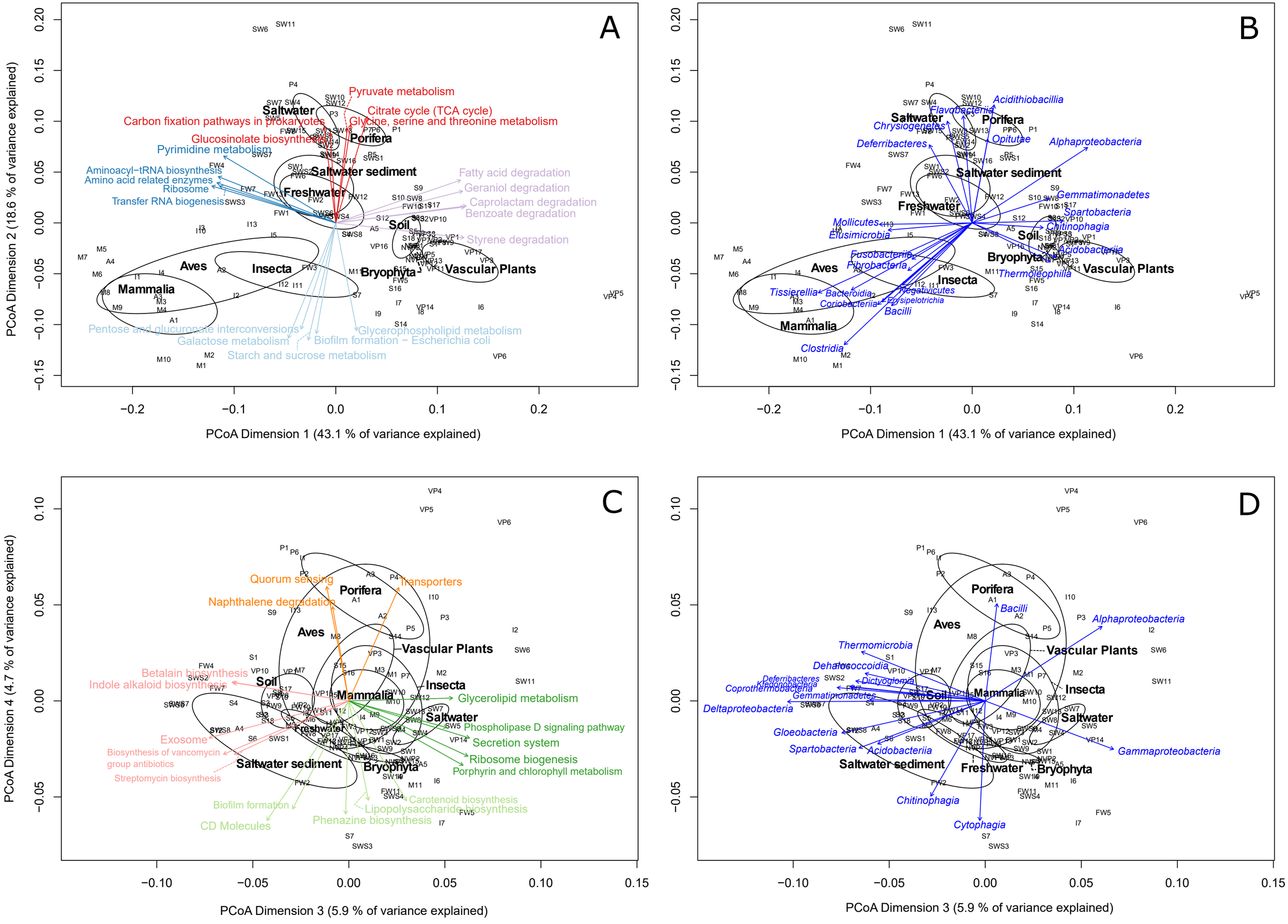

A
B
C
D
PCoA Dimension 1 (43.1 % of variance explained)
PCoA Dimension 2 (18.6 % of variance explained)
PCoA Dimension 3 (5.9 % of variance explained)
PCoA Dimension 4 (4.7 % of variance explained)
Pyruvate metabolism
Citrate cycle (TCA cycle)
Carbon fixation pathways in prokaryotes
Glycine, serine and threonine metabolism
Glucosinolate biosynthesis
Pyrimidine metabolism
Aminoacyl−tRNA biosynthesis
Amino acid related enzymes
Ribosome
Transfer RNA biogenesis
Fatty acid degradation
Geraniol degradation
Caprolactam degradation
Benzoate degradation
Styrene degradation
Pentose and glucuronate interconversions
Galactose metabolism
Starch and sucrose metabolism
Biofilm formation − Escherichia coli
Glycerophospholipid metabolism
Saltwater
Porifera
Saltwater sediment
Freshwater
Soil
Bryophyta
Vascular Plants
Aves
Insecta
Mammalia
Acidithiobacillia
Flavobacteriia
Chrysiogenetes
Opitutae
Deferribacteres
Alphaproteobacteria
Gemmatimonadetes
Spartobacteria
Chitinophagia
Mollicutes
Elusimicrobia
Acidobacteriia
Fusobacteriia
Fibrobacteria
Thermoleophilia
Tissierellia
Bacteroidia
Negativicutes
Coriobacteriia
Erysipelotrichia
Bacilli
Clostridia
Quorum sensing
Transporters
Naphthalene degradation
Betalain biosynthesis
Indole alkaloid biosynthesis
Glycerolipid metabolism
Phospholipase D signaling pathway
Secretion system
Ribosome biogenesis
Porphyrin and chlorophyll metabolism
Exosome
Biosynthesis of vancomycin group antibiotics
Streptomycin biosynthesis
Biofilm formation
CD Molecules
Phenazine biosynthesis
Lipopolysaccharide biosynthesis
Carotenoid biosynthesis
Thermomicrobia
Dehalococcoidia
Dictyoglomia
Deferribacteres
Ktedonobacteria
Coprothermobacteria
Gemmatimonadetes
Deltaproteobacteria
Gloeobacteria
Spartobacteria
Acidobacteriia
Gammaproteobacteria
Chitinophagia
Cytophagia

Figure 2

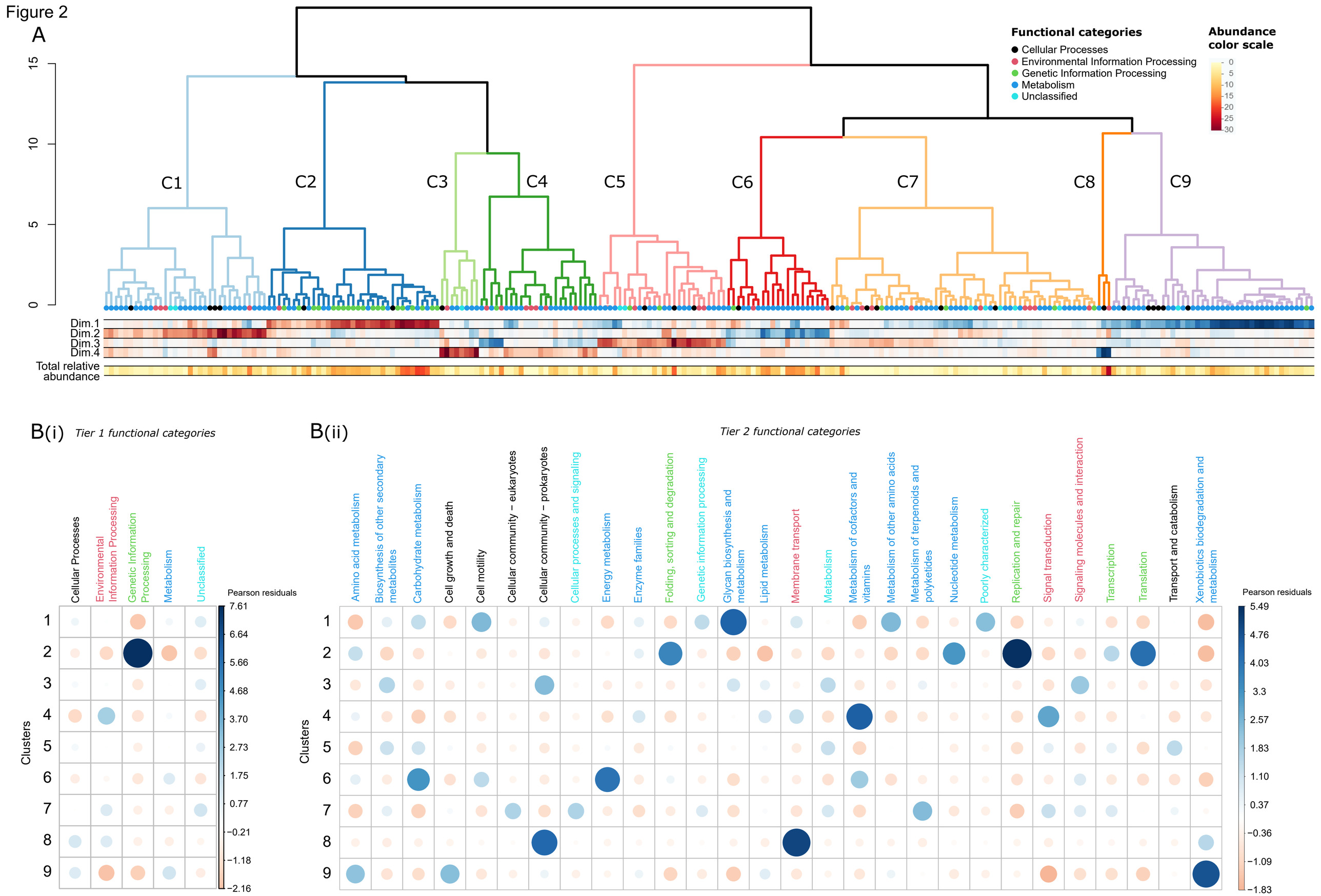

A
Functional categories
Cellular Processes
Environmental Information Processing
Genetic Information Processing
Metabolism
Unclassified
Abundance color scale
C1
C2
C3
C4
C5
C6
C7
C8
C9
Dim.1
Dim.2
Dim.3
Dim.4
Total relative abundance
B(i) Tier 1 functional categories
Cellular Processes
Environmental Information Processing
Genetic Information Processing
Metabolism
Unclassified
Clusters
Pearson residuals
B(ii) Tier 2 functional categories
Amino acid metabolism
Biosynthesis of other secondary metabolites
Carbohydrate metabolism
Cell growth and death
Cell motility
Cellular community – eukaryotes
Cellular community – prokaryotes
Cellular processes and signaling
Energy metabolism
Enzyme families
Folding, sorting and degradation
Genetic information processing
Glycan biosynthesis and metabolism
Lipid metabolism
Membrane transport
Metabolism
Metabolism of cofactors and vitamins
Metabolism of other amino acids
Metabolism of terpenoids and polyketides
Nucleotide metabolism
Poorly characterized
Replication and repair
Signal transduction
Signaling molecules and interaction
Transcription
Translation
Transport and catabolism
Xenobiotics biodegradation and metabolism
Clusters
Pearson residuals

Figure 3

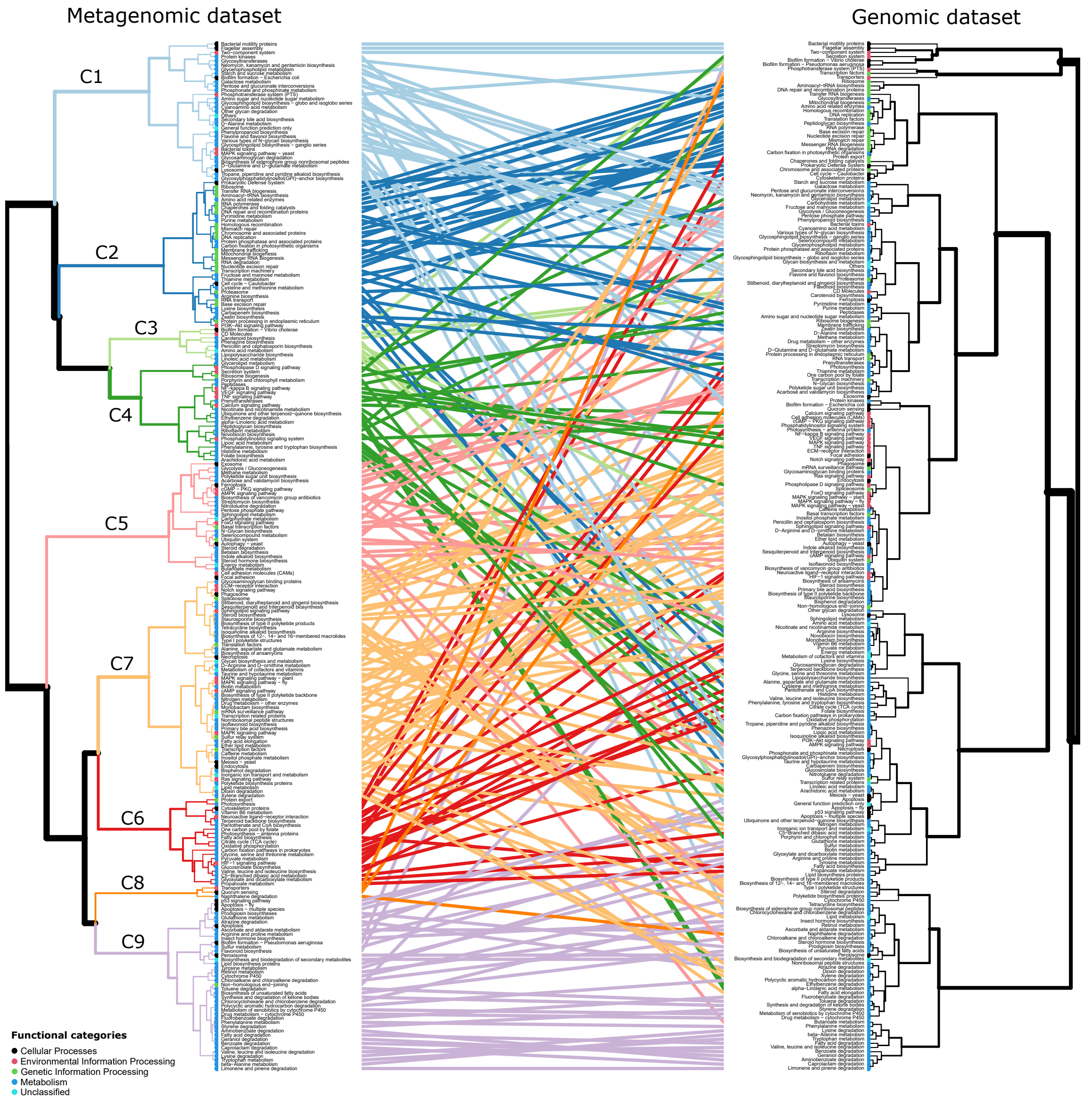
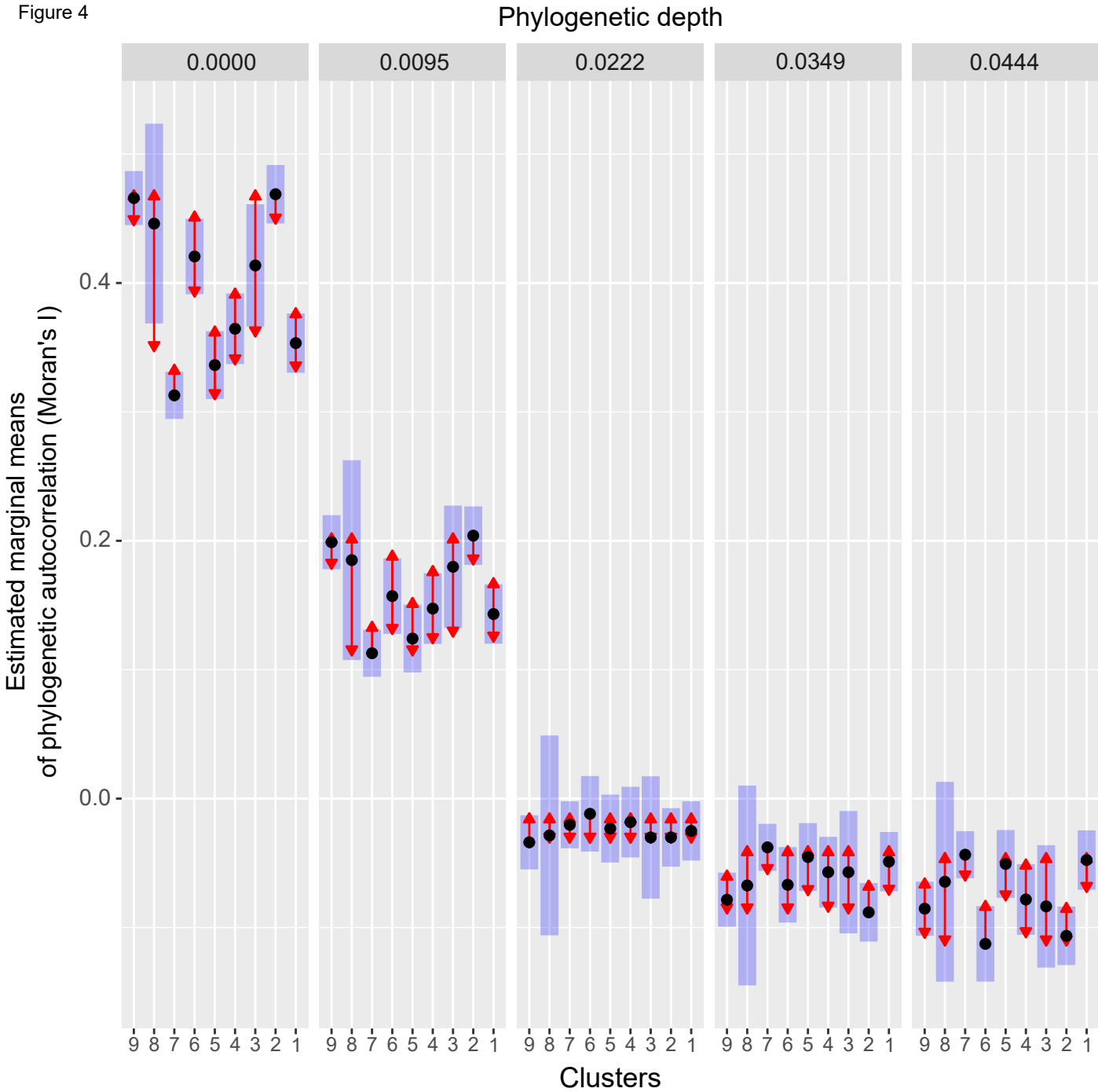

Metagenomic dataset
Genomic dataset
C1
C2
C3
C4
C5
C6
C7
C8
C9
Functional categories
Cellular Processes
Environmental Information Processing
Genetic Information Processing
Metabolism
Unclassified
15
10
5
0
0
5
10
15
20

Figure 4


Phylogenetic depth
0.0000
0.0095
0.0222
0.0349
0.0444
Estimated marginal means
of phylogenetic autocorrelation (Moran's I)
0.4
0.2
0.0
9 8 7 6 5 4 3 2 1
Clusters