Community-level signatures of ecological succession in natural bacterial communities

Alberto Pascual-García[1, 2,*] and Thomas Bell[1]

November 29, 2019

(1) Department of Life Sciences. Silwood Park Campus. Imperial College London, Ascot, United Kingdom
(2) Current address: Institute of Integrative Biology. ETH-Zürich, Zürich, Switzerland
(*) Correspondence: alberto.pascual.garcia@gmail.com; palberto@env.ethz.ch

Abstract

A central goal in microbial ecology is to simplify the extraordinary biodiversity that inhabits natural environments into ecologically coherent units. We present an integrative top-down analysis of over 700 bacterial communities sampled from water-filled beech tree-holes in the United Kingdom at distances between <5m to >100km, combining an analyses of community composition (16S rRNA sequencing) with assays of community functional capacity (exo-enzymatic activities, ATP production, CO_2 dissipation and yield). The communities were grown in laboratory conditions in a complex growth medium, what allowed us to investigate the relationship between composition and function, excluding confounding environmental factors. We found a distance-decay trend in the similarity of the communities, and simulated data allowed us to reject the hypothesis that stochastic processes dominated the assembly of the communities, suggesting that niche effects prevailed. Consistent with this idea, clustering of communities lead us to identify six distinct community classes encompassing samples collected at often distant locations. Using structural equation modelling, we explored how functions were interrelated, demonstrating that a representative functional signature can be associated with each community class. We obtained a more mechanistic understanding of the classes using metagenomic predictions with PiCRUST. Notably, this approach allowed us to show that these classes contain distinct genetic repertoires reflecting community-level ecological strategies. We finally formulated an over-arching ecological hypothesis about how local conditions drive succession in these habitats. The ecological strategies resemble the classical distinction between r- and K-strategists and could be extrapolated to other systems, suggesting that a coarse grained picture of microbial ecological succession may be explained by relatively simple ecological mechanisms.

1 Introduction

The microbial communities inhabiting natural environments are unmanageably complex. It is therefore difficult to establish causal relationships between community composition, environmental conditions and ecosystem functions (such as rates of biogeochemical cycles) because of the large number of factors influencing these relationships. There is great interest in developing methods that reduce this complexity in order to understand whether there are predictable changes in community composition across space and time, and whether those differences alter microbe-associated ecosystem functioning. The most common approach has been to search for physical (e.g. disturbance) and chemical (e.g. pH) features that correlate with community structure and function. This approach has often been successful in identifying some major major differences among bacterial communities associated with different habitats [1] and some of the edaphic correlates [2]. However, even if significant correlations between environmental variables and microbial functioning are found, we are still far from understanding the underlying biological mechanisms explaining these relationships. For instance, adding variables such as biomass or diversity to models in which environmental variables are good predictors of function do not strongly improve model predictions [3], suggesting that there is a need for variables that increase the accuracy of biological processes [4].

The development of a more mechanistic picture is hindered for several reasons, such as difficulties in identifying the relative role of stochastic and deterministic processes in shaping microbial communities [5, 6, 7], the pervasiveness of functional redundancy [8, 9], and of priority effects [10]. In addition, it is often difficult to identify which functions to assess. Microbes inhabitating a host sometimes have a substantial impact on host performance, for example turning a "healthy" into a "diseased" host [11]. Such extreme impacts of individual taxa make it relatively simple to infer a direct link between community composition and function. In open, natural environments (e.g. soil, lakes, oceans), the impact of individual taxa on ecosystem functions is often minor, and generalisations may depend on subjective choices of which functions to measure.

An important step forward comes from manipulative experiments in natural environments, that have identified variables such as pH [12], salinity [8], sources of energy [13], the number of species [14], and environmental complexity [4] as key players in the relationship between bacterial community structure and functioning. Improved control can be obtained by "domesticating" communities surveyed from natural environments by growing them in a synthetic (albeit complex) environment, and quantifying their functioning under such controlled conditions [15, 16]. With these experiments it becomes possible to directly test the hypothesis that more similar communities have more similar functions without the confounding influence of extrinsic environmental conditions.

Community similarity can be assessed using a rich array of analytic tools that identify $\beta-$diversity clusters within multivariate datasets, such as the detection of communities in species co-occurrences networks [17] or the reduction of the dimensionality of $\beta-$diversity similarities [18]. These approaches have been pervasive in the medical microbiome literature, for example in the search for "enterotypes"– i.e. whether individuals are characterised by diagnostic sets of species representing alternative community states [18, 19] which, in this paper, we call "classes" of communities. The existence of classes in communities sampled from different locations may be due to variable environmental conditions that select for different taxa, or may be explained more parsimoniously by stochastic processes together with strong dispersal limitation [20]. Deciphering the likelihood of different ecological mechanisms can be assessed by adopting a suitable null model, see e.g. [21]. Community classes arising from environmental selection would also be functionally different, whereas we would not expect functioning to differ among community classes created by stochastic processes.

Once classes and functional differences have been identified, it is possible to step down into key biological processes focusing on the genetic repertoires of the constituent taxa [22]. Investigating the dominant genes present in the different community classes allows explanations of functional differences and the determination of ecological strategies. For example, community classes that differ in genes related with environmental sensing, degradation of extracellular substrates, or metabolic preferences, could be used as hypotheses of the molecular mechanisms responsible for functional differences. Therefore, the last step aims to explain how the functional and genetic differences arise from the prevailing environmental conditions [23], and could point to the specific environmental parameters that could be measured. This approach solves the problem of measuring many environmental parameters in the hope that some will be significantly associated with community structure or ecosystem functioning. Lack of any clear functional differentiation among community classes is also informative, and would indicate alternative community states with redundant functions [24, 25, 26]. Such redundancy could arise in the absence of environmental variability, which could also help explain the lack of a dominant environmental axis that explains variation in composition.

In this work we followed the above pipeline investigating a large dataset consisting of more than 700 samples

of rainwater-filled puddles (phytotelmata) that can form at the base of beech trees. The bacterial communities present in the tree-holes are key players in the decomposition of leaf litter, and therefore of great interests more broadly for understanding decomposition in forest soils and riparian zones. This is an ideal system to follow this procedure given the relatively similar conditions found across different locations, making it unique in terms of replicability of a natural aquatic environment [27, 28], and its relatively low diversity. Indeed, while effects caused by environmental variation on phytotelmata ecosystems has been investigated in meio- and macrofaunal communities [28], the influence in microbial communities is largely unknown. Moreover, emphasis was made for understanding bottom-up drivers of tree-holes diversity like nutrients [29, 30, 31], but top-down approaches that may help us to understand other drivers of microbial composition like stochastic dispersion or interactions have received less attention [28].

Previous work using this dataset showed that larger (smaller) taxa abundances influenced broad (narrow) community-level functions [32]. Here we aim to illuminate the mechanistic basis of this relationship. The large dataset allowed us to study natural variation in bacterial community composition through the top-down categorisation of communities into classes. We then linked the classes with bacterial functioning, analysing a set of community-level functional profiles obtained from laboratory assays of the same communities [32], and investigating whether the classes differed in their functional capacity. Instead of focusing on each function individually, we investigated how the functional profiles varied across the community classes. We then used metagenomic information to understand whether similar compositions and functions were translated into different classes of genetic repertoires.

We found significant differences in the genetic repertoires and functional measurements among classes, that we interpreted in the context of changing environmental conditions. In deciduous forests, bacterial life on the forest floor is characterised by seasonal and daily changes to temperature and resource availability. We therefore address whether differences in the communities are due to the historical processes at the different geographic locations, or if they are rather more influenced by contingent local conditions. These factors are often difficult to resolve [26] but may both be important due to the high temporal variability in these systems, as observed in compost ecosystems [33]. Interestingly, interpreting the signatures found in the functional measurements and in the genetic repertoires led us to hypothesize the existence of community-level ecological strategies reflecting an ecological succession driven by local environmental dynamics of the tree-holes. These ecological strategies ressemble the classical distinction between r− and K−strategists described for single species [34].

Results

Microbial communities classes are determined by local conditions

We analysed 753 bacterial communities sampled from water-filled beech tree-holes in the South West of UK [32] (see Suppl. Table 1). Communities were grown in a medium made of beech leaves as substrate for seven days and then their composition interrogated through 16S rRNA sequencing (see Methods). We analysed the $\beta-$diversity of these communities according to two different metrics: the Jensen-Shannon divergence (D_{JSD}) [35], and a transformation of the SparCC metric (D_{SparCC}, see Methods [36]). Spatial distances and the two $\beta-$diversity distances were significantly correlated (Mantel test: $r = 0.21; p < 10^{-3}$ for D_{SparCC} and $r = 0.19; p < 10^{-3}$ for D_{JSD}). To test if this trend was maintained across the different scales, we clustered samples that were closer in space, and retrieved the classifications found at 10 distance thresholds spanning 5 orders of magnitude (from <5 m to >100km). We used three statistics (ANOSIM, MRPP and PERMANOVA [37, 38], see Methods) to test whether the $\beta-$diversity distances within clusters are significantly smaller than those between clusters for the 10 classifications. In all cases, the three tests supported the hypothesis that communities within locations were significantly more similar than between locations (permutation tests, $p < 10^{-3}$, see Suppl. Fig. 1). This spatial autocorrelation may be indicative of an important role for stochastic assemblage and dispersal limitation [20].

We studied how the different statistics changed across the 10 classifications. We observed an increase in the mean distances within clusters (quantified with the MRPP statistics) and a decay in the ANOSIM R statistics (Fig. 1 c and d), while PERMANOVA remains roughly constant across scales (Suppl. Fig. 2). To interpret these trends we analysed the behaviour of these metrics with synthetic data in which artificial $\beta-$diversity distances matrices were generated under different assumptions on the mean and variances within and between locations across scales (Suppl. Figs. 3—6). We found that the most informative observation is the ANOSIM-R decay when the variance of the simulated distances were large (Suppl. Figs. 6 middle column, bottom) a pattern consistent with the one observed for the real data (Fig. 1d). To give a sense of the implications of this finding, in Suppl.

Fig. 7 we show that for such variance in the generated distances, approximately 3% of the $\beta-$diversity distances between samples located at more than 100 km. away should be as high as those found in communities samples within 5 m. of distance.

This result implies high levels of dispersal that are difficult to explain for tree-hole bacteria. We explored the alternative hypothesis: that similar communities found at distant locations result from similar underlying environmental conditions. We performed unsupervised clusterings with D_{JSD} and D_{SparCC}, revealing in both cases six distinct community classes (Fig. 2A and B, Suppl. Fig. 8 and Suppl. Table 2 for global characteristic metrics such as diversity). The whole set of communities are dominated by *Proteobacteria*, and the community classes were distinguished at the genus level (95% sequence similarity), including a higher presence of the genus *Klebsiella* and *Pantoea* (classes 1, red; and class 3, pink); *Paenibacillus* and *Sphingobioum* (class 2, green); *Serratia* (class 5 blue); *Sphingomonas, Streptomyces* and *Pseudomonas* (classes 4, yellow) and low abundant genera like *Brevundimonas* and *Herbaspirillum* and, again, *Pseudomonas* for class 6 (grey). In the following, we refer to class 1 (red) as the reference class because it encompassed the largest number of communities (Suppl. Table 2). We refer to the remaining communities by their most distinctive taxon as *Paenibacillus* (class 2), *Klebsiella* (class 3), *Streptomyces* (class 4), *Serratia* (class 5). For class 6 we observed that, although the *Pseudomonas* genus was also high in other communities, classes 4 and 6 were dominated specifically by *Pseudomonas putida* (Suppl. Fig. 9), which we selected as representative of class 6. In Ref. [39] we use a network approach to identify modules of coocurring species that confirm the key role of the taxa selected as representatives.

We illustrate how these classes are distributed in space by representing the class identity of each community in a coloured bar, alongside the Site and Date in which the community was sampled (Fig. 2C). As expected from the previous analysis, some communities belong to the same class even if they were distant in space (see examples in Suppl. Fig. 10). For the latter, in Fig. 2C we highlight with dotted rectangles some of the cases in which there is a better correspondence with the date of sampling than with the site. To test this observation, we showed that a classification based on the date of sampling is consistently more similar to the $\beta-$diversity classes than the site (Suppl. Table. 3). Moreover, computing the ANOSIM statistics when tree-holes are clustered according to sampling location (Site values in Fig. 2e) or according to the sampling date (Day and Month values) consistently show that the specific date (Day) is more informative than the site. The Day was also more informative than the Month, suggesting that seasonal environmental conditions were not the main drivers of the similarities, but were rather due to daily variation in local conditions. Notably, the value we obtain for the ANOSIM statistics when the classification considered are the community classes found, reaches the same value than the one found at 50 m (Fig. 1c). In summary, the classification successfully grouped communities into just six groups, with communties often separated by far more than 50 m. In addition, the date of sampling is more informative than the sites. Taken together, the results suggest that the classes capture similarities in local environmental conditions even in tree-holes that were spatially separated by considerable distances.

Community classes reflect different functional performances

If environmental conditions determine compositional differences in the communities, we expect that these differences are translated into different community functional capacities, a question that we investigate analysing data that quantify the function performance of these communities [32]. The sampled communities were cryo-preserved after sequencing, and later revived in a medium made of beech leaves as substrate. Cells were grown for seven days while monitoring CO_2 dissipation and, after this period the following measurements were taken: community cell counts, community metabolic capacity (measured as ATP concentration), and community capacity to secrete four ecologically relevant exoenzymes [40] related with i) uptake of carbon: xylosidase (X) and $\beta-$glucosidase (G); ii) carbon and nitrogen: $\beta-$chitinase (N); and iii) phosphate: phosphatase (P) .

Visual inspection indicated substantial differences in the functional capacities among the community classes. In some cases, communities belonging to different classes clearly separated, as shown in the histograms in Suppl. Fig. 12. Therefore, we explored if these differences among the community classes were significant using structural equation models (SEM) [41]. A model that made no distinction among classes yielded an excellent fit (RMSEA$<10^{-3}$, CI=[0–0.023], AIC=7493 see Methods and Suppl. Fig. 14). From this structural model we next considered models with different parameters for each community class (up to six parameters per pathway, see Methods). We then investigated which model better explained the data according to 3 scenarios: i) a model in which all the parameters were constrained to be the same for all the classes; ii) a model in which each class had a different parameter for each pathway; iii) an intermediate model, in which some parameters were constrained for some classes. Accounting for penalizations for models with more degrees of freedom (RMSEA$<10^{-3}$, CI=[0–0.035], AIC=6658, see Methods and Suppl. Material), the best model belonged to the scenario (iii) (see Fig. 3
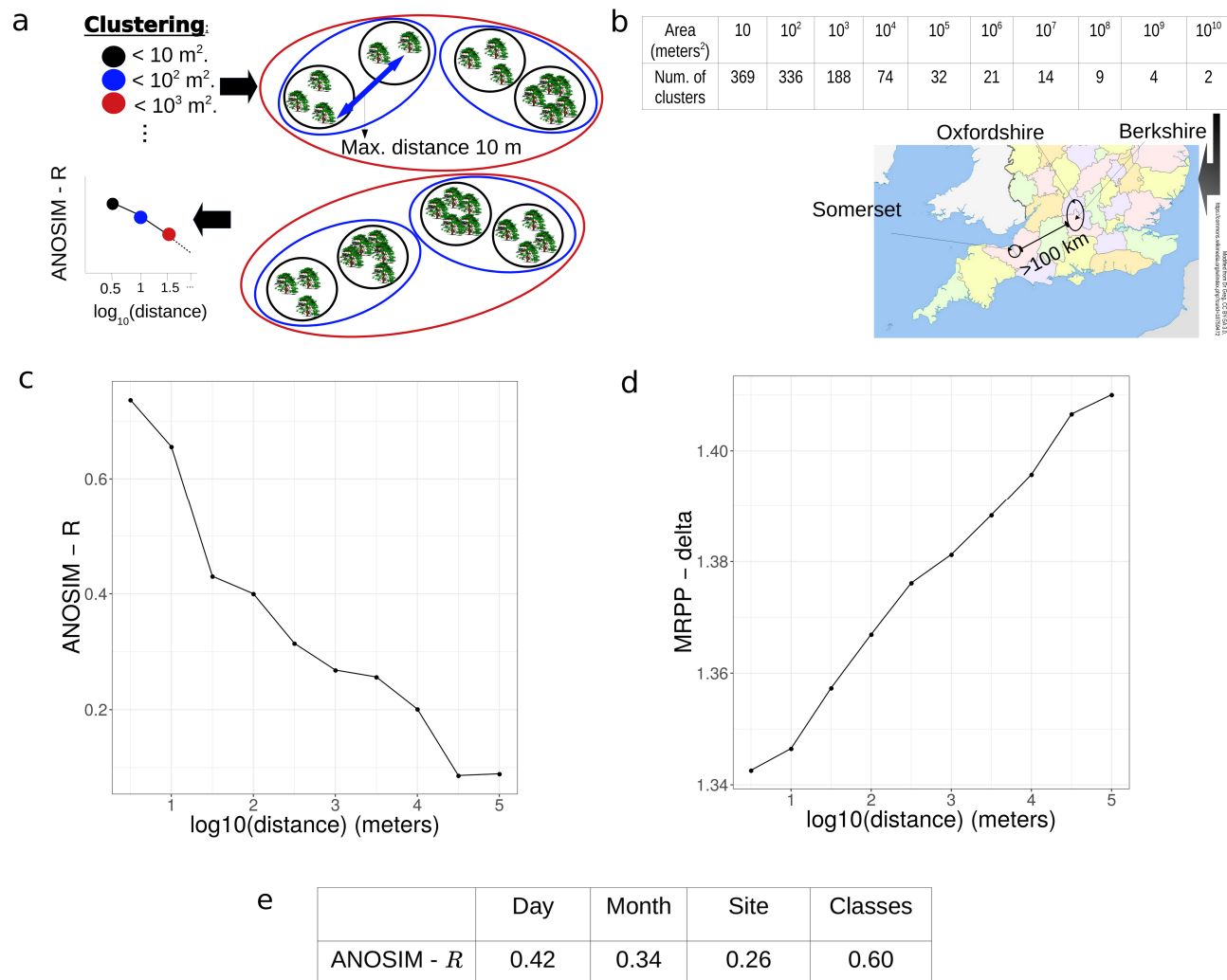
Figure 1: **Analysis of decay of communities' similarity with distance.** (A) Illustration of the procedure. Trees are clustered within areas A of increasing sizes, leading to one classification every order of magnitude (ellipses of different colours). We approximate these areas, stopping the clustering at spatial distances thresholds equal to $\sqrt{A}$, e.g. the classification for areas of 100 m^2 (blue ellipses) is obtained stopping at $d = 10m$. For each classification, an ANOSIM and mrpp value was computed to test if the similarity of the communities within the clusters is larger than between clusters, and the statistics were plotted against the spatial threshold. (B) Number of clusters at each spatial threshold. The three counties from which the communities were sampled are shown in the map. The classification at the last threshold consists of two clusters separated by more than 100km. (C and D) Observed ANOSIM and mrpp statistics for each of the distance thresholds, reflecting a decay in the similarity of the communities (see Main Text for details). (E) ANOSIM values for communities clustered according to the day and month in which were sampled, the site from which they were collected (optimal classification of spatial distances), and the six community classes (optimal classification of $\beta-$diversity distances).
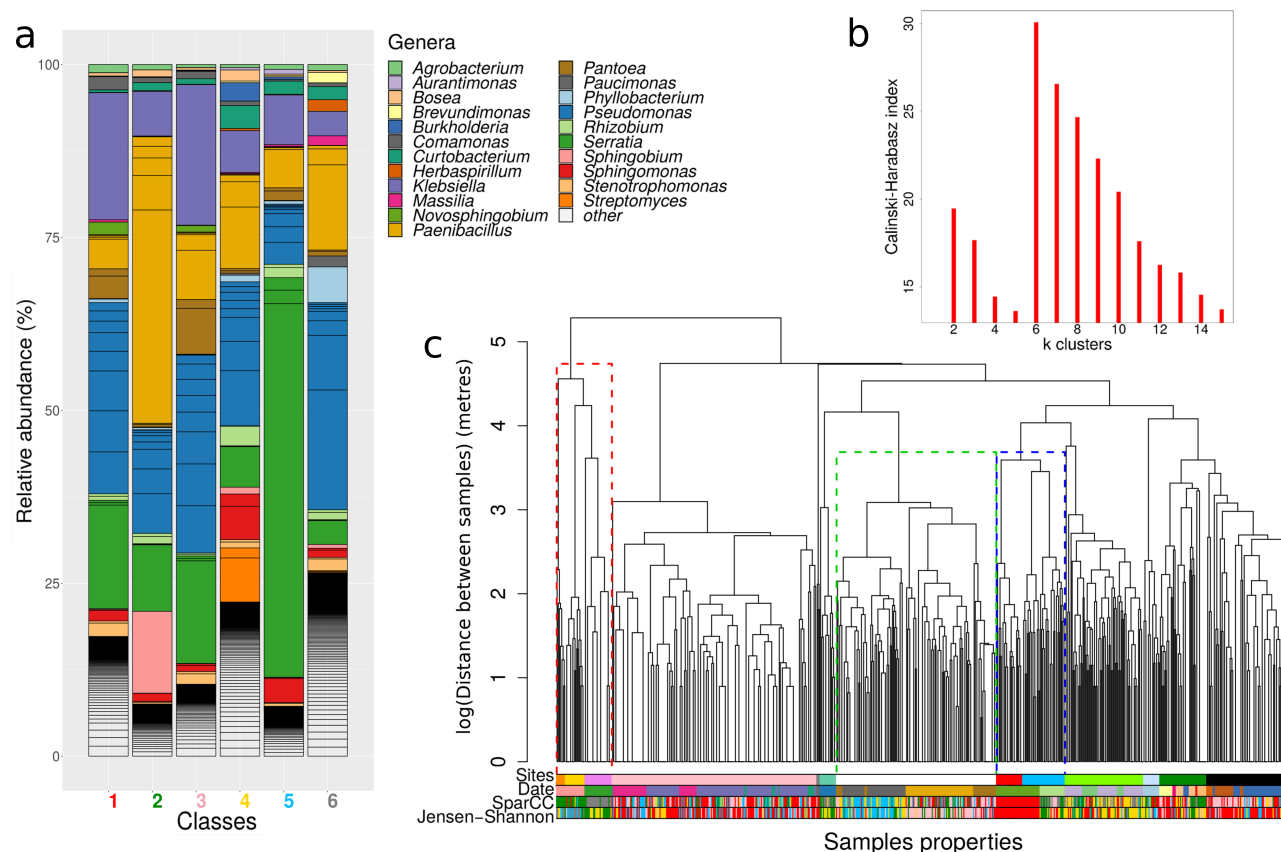
Figure 2: **Communities classes description and relation with sampling sites.** (A) Genus frequencies in each of the six community classes found after unsupervised clustering of the communities using D_{SparCC}. Only the 15 most abundant genera in each community class are shown, with the remainder classified as Others. (B) Calinski-Harabasz index versus the number of final clusters in the classification, showing a maximum at six classes. (C) The dendrogram represents the clustering of the samples (leafs in the tree) according to their spatial distances. The coloured bars represent membership of each sample to different classifications (from top to bottom): the sites, the day of sampling collection (Date), and the classes found using either D_{SparCC} and D_{JSD}. Dotted rectangles in the dendrogram indicate examples of: (green) locations where one site was sampled on different days; (blue) two sites sampled on the same day: (red) a mix of both situations. It is apparent that the date is a better match to the community classes than the site, which is confirmed with a comparison between the classifications (Suppl. Table 3).

161 and Suppl. Table 6). This result supports the hypothesis that the classes had differentiated functional capacities
162 (Suppl. Table 7 and Suppl. Fig. 15).
163   The model showed that measurements related with uptake of nutrients were all exogenous, including ATP
164 production, cell yields and CO2 production (Fig. 3). In addition, ATP production influenced yield, which in turn
165 influenced CO2. Among exoenzyme variables, N influenced ATP and, notably, only X affected yield, while G and
166 P influenced both ATP and CO2. Analysis and clustering of the standarized partial regression coefficients (Suppl.
167 Fig. 11) established a classification of the classes in three groups according to their function, with *Paenibacillus*
168 and the reference classes being the most similar, then *Streptomyces* and *P. putida* classes, and *Klebsiella* and
169 *Serratia*.
170   We then explored if distinctive pathways for each class could be determined. Given the complexity of the
171 SEM models, we first ruled out the possibility that differences in pathway coefficients were due to the influence
172 of other (confounding) variables. To control for this possibility, for each pair of endogenous-exogenous variables,
173 we searched for its set of confounding variables with dagitty [42]. Next, for each pair of variables involved in a
174 pathway, we performed a linear regression including its adjustment set of confounding factors, and an interaction
175 term with a factor coding for the different classes. Coefficients should be interpreted as deviations with respect to
176 the reference class (see Methods). The significant interaction terms (Fig. 3, panel D) show how the relationships
177 among the functional variables differed among the community classes. For example, the analysis revealed that
178 cell yield was negatively influenced by $\beta-$chitinase activity for the *Paenibacillus* class, for ATP production for
179 the *Serratia* class, while being positive related with $\beta-$glucosidase for the classes of *Klebsiella* and *P. putida*. We
180 therefore concluded that the community classes had significantly different functional capacities, thus producing
181 different relationships in the models.

Community classes depict different genetic repertoires

183 To get a more mechanistic understanding of the above results, we analysed the genetic repertoire of each com-
184 munity class by performing metagenomic predictions with PiCRUST [43], and further statistical analysis with
185 STAMP [44]. The Nearest Sequence Taxon Index is 0.059, reflecting a high quality prediction [43]. This is because
186 most of the dominant genera in this system are found in gut microbiomes (e.g. Fig. 5 in Ref. [45]).
187   The fraction of exo-enzymatic genes belonging to *Paenibacillus*, *Streptomyces* and *P. putida* classes was sig-
188 nificantly larger than the fractions found for the *Klebsiella*, *Serratia* classes and the reference class, suggesting
189 that the former classes are specialized in degrading a wider array of substrates (Fig. 4).
190   Clustering the KO annotations into KEGG pathways (see Methods) showed that the 6 community classes
191 differed in their genetic repertoires. Furthermore, these divergent genetic repertoires suggested different ecological
192 adaptations, which are summarized in Fig. 5. Consistent with PCA analysis of the KEGG pathways (Suppl.
193 Figs. 18–21), we divided the classes in two groups: the reference, *Klebsiella* and *Serratia* classes carried the
194 genetic machinery for fast growth, while *Paenibacillus*, *Streptomyces* and *P. putida* classes carried the genetic
195 machinery for autonomous amino-acid biosynthesis. Evidence for fast growth in the reference, *Klebsiella* and
196 *Serratia* classes comes from the large fraction of genes related with genetic information processing (Suppl. Fig.
197 24), mostly related with DNA replication such as DNA replication proteins genes, transcription factors, mismatch
198 repair, homologous recombination genes or ribosome biogenesis —the latter being a good genetic predictor of fast
199 growth [46]. Secondly, communities from these classes also carried a larger fraction of genes related with intake
200 of readily-available extracellular compounds (Suppl. Fig. 25), including ABC transporters, phosphotransferase
201 system, or peptidases, and environmental adaptations including motility proteins, synthesis of siderophores, and
202 the two-component systems. Rapid replication often requires a more accurate control of protein folding and
203 trafficking. Consistent with this hypothesis, we found a significantly inflated fraction of genes involved in folding
204 stability, sorting and degradation, including chaperones and genes involved in the phosphorelay system (Suppl.
205 Fig. 26).
206   A second series of evidences pointing towards orthogonal ecological strategies came from differences in the
207 metabolic pathways associated with the community classes. *Serratia*-dominated class (5) had an inflated fraction
208 of genes related to carbohydrate degradation, including genes involved in glycolysis and in the trycaborxylic
209 acid (TCA) cycle (Suppl. Fig. 27). In contrast, the *Paenibacillus, Streptomyces* and *P. putida* classes were
210 associated with genes involved in alternative pathways like nitrogen/methane metabolism, and in secondary
211 metabolic pathways related with degradation of xenobiotics/chlorophyl metabolism. Notably, the genes involved
212 in the exoenzymes that were experimentally assayed were higher in these classes, suggesting that they were adapted
213 to environments with more recalcitrant nutrients (Fig. 5). In addition, *Paenibacillus, Streptomyces* and *P. putida*
214 classes had a remarkable repertoire of genes for amino acids biosynthesis –possibly at odds with the reference class
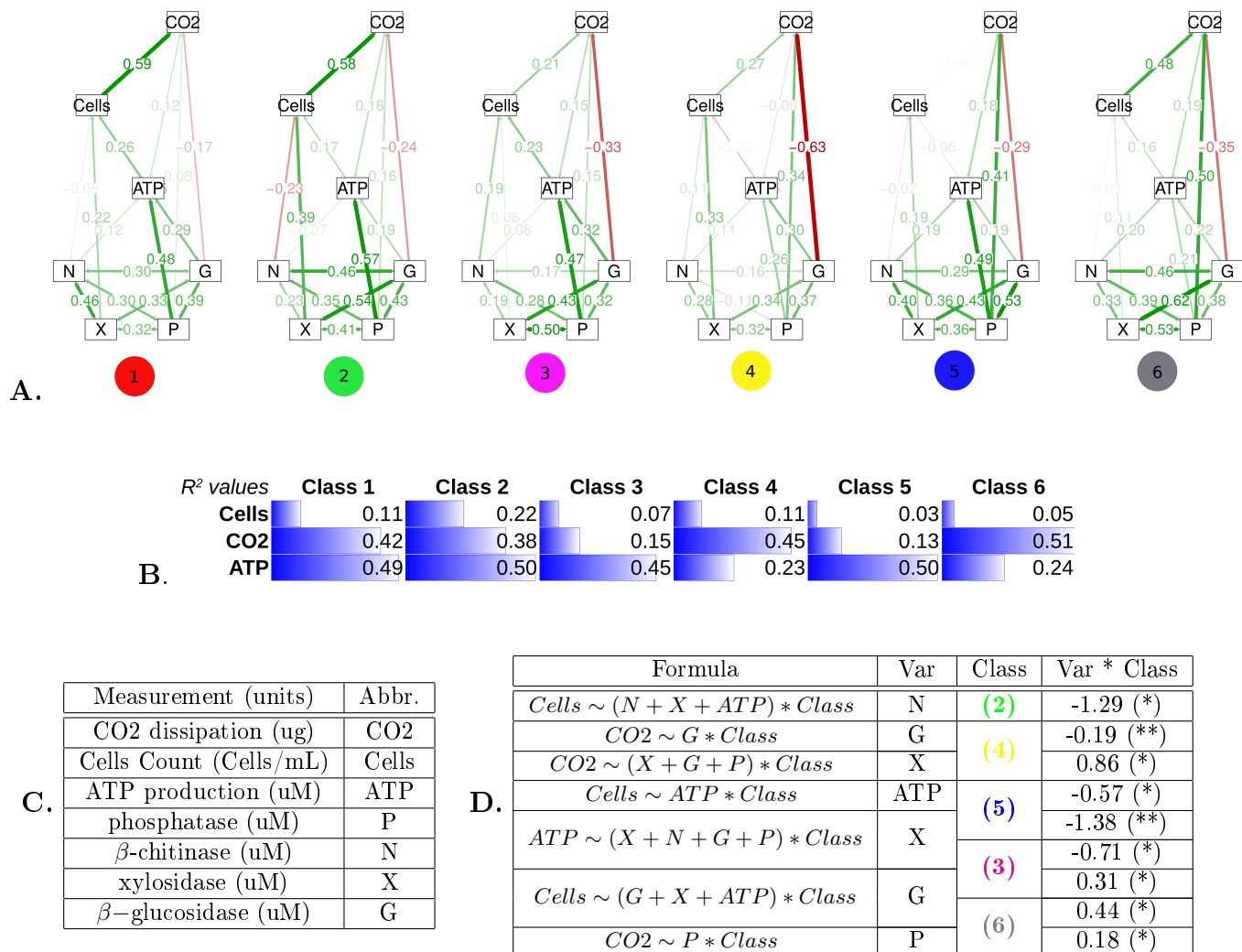
7

Figure 3: **Functional capacity of the the communities classes.** (A) Diagrams of the final SEMs. The standarized coefficients for each of the six classes are shown on the corresponding pathway. (B) R^2 values for the three endogenous variables in the SEM model. (C) Experimental measurements used in this study and their abbreviations. (D) Causal analysis of the influence of the exogenous variable (Var) for each SEM pathway (determined by the response variable and Var in Formula) when the identity of the classes is included as a factor. Confounding factors involved in the pathway are also included (remainder variables in Formula). The column Class reflects the class identity of any significant coefficient found with its value shown in the column Var * Class (significance code: * = p-val<0.01, ** = pval<0.001).
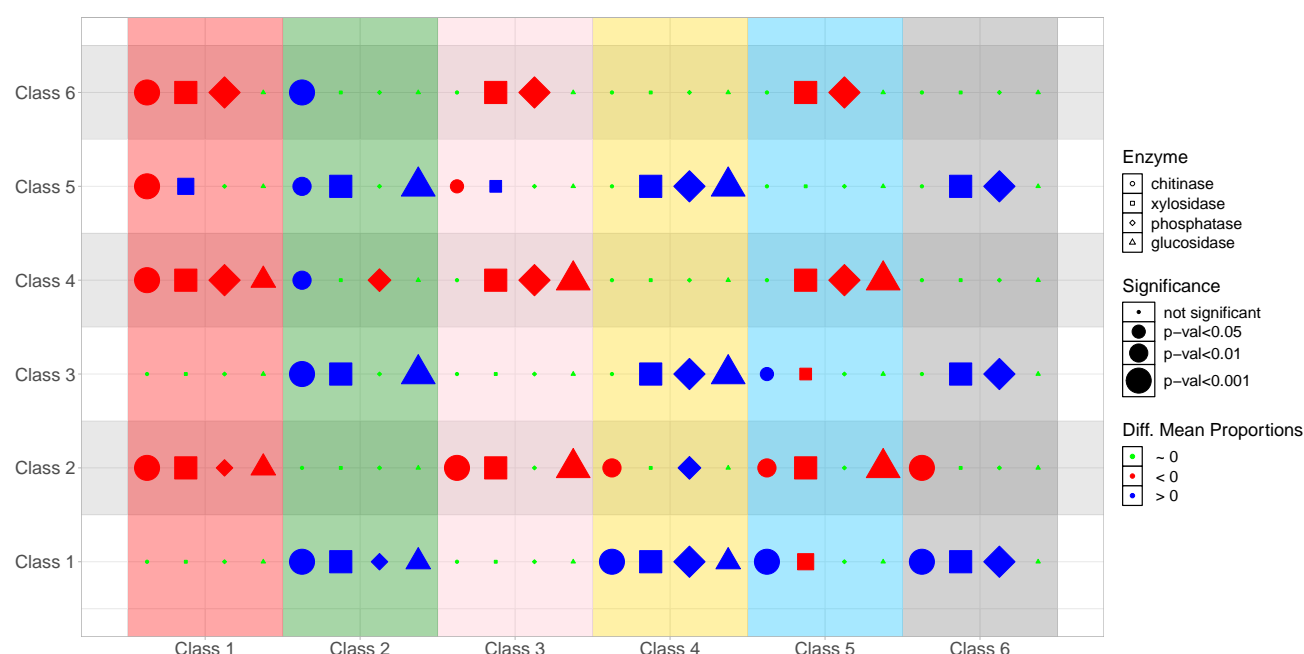
Figure 4: **Summary of post-hoc analysis of exo-enzymatic genes predictions.** The significance of differences in mean proportions of chitinase (circles), β-xylosidase (squares), β-glucosidase (triangles) and phosphatase (diamonds) were tested across all pairwise combinations of community classes. Red (blue) symbols denote that the class in the correspondent column has a significantly lower (higher) mean proportion of the enzyme for the class shown in the row. The size of the symbol is larger for more significant differences. All tests are provided in Suppl. Fig. 4.

and *Klebsiella* and *Serratia* classes, which invested in proteases for amino acid uptake (Suppl. Fig. 28 and 29). The apparently low glycolytic capabilities of these communities could result in pyruvate deficiencies, which would hindered the production of sufficient acetyl-CoA and oxaloacetate required to activate the TCA cycle. Consistent with this observation, we observed that these communities exhibited a significantly larger proportion of genes related with glyoxylate metabolism and degradation of benzoate, that may be used as alternatives to glycolysis (Suppl. Fig. 30). Finally, we observed that communities in the reference class and Klebsiella and *Serratia* classes had a significantly larger repertoire of genes needed to synthesise amino acids requiring pyruvate (valine, leucine and isoleucine), and which, according to our interpretation, they would generate through glycolysis (Suppl. Fig. 27). By contrast, *Paenibacillus, Streptomyces* and *P. putida* classes had a significantly larger proportion of genes used to degrade these amino acids (Suppl. Fig. 28) and hence, either they take these essential amino acids from the environment or they generate them from other pathways. Consistent with this observation, genes in these classes were enriched for glycine, serine and threonine metabolism (Suppl. Fig. 28), through which it is possible to obtain valine, leucine and isoleucine, and which could provide an alternative source of acetyl-CoA (Fig. 5).

Discussion

Our analysis of a large set of tree-hole bacterial communities found a strong distance-decay in the similarity of the communities across several orders of magnitude. The existence of spatial autocorrelation has previously reported in soil environments [47, 12], but this study extends the findings to scales above the short distances (<10m) previously reported [47]. The high ANOSIM statistics require unrealistic levels of dispersal for the pattern to be explained by stochastic processes alone (Suppl. Figs. 28), and therefore points towards a hypothesis that similar environmental conditions occur at distant locations.

We observed that the communities could be arranged into classes, and that the classes corresponded to the site and the date of collection, which are tightly entangled. There was a geographic influence that was particularly apparent at larger distances, which suggests the influence of broad environmental conditions and perhaps historical processes [26]. The finding is consistent with the idea that environmental conditions on a particular day strongly
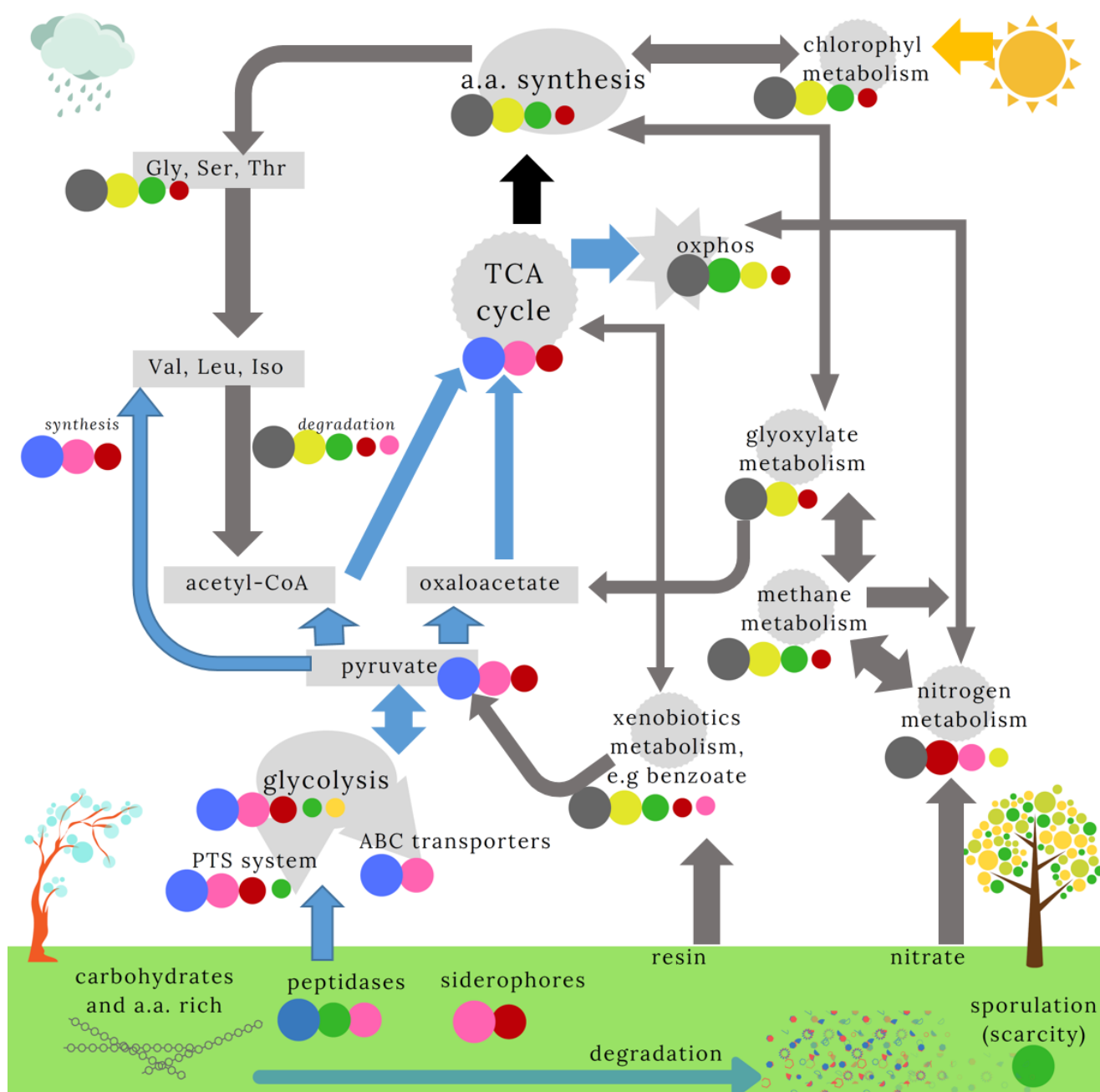
9

Figure 5: **Scheme describing the genetic repertoires of the six community classes.** Pathways from the KEGG database that were most relevant for describing the classes are shown. We ranked the mean proportion of genes in each pathway (see Suppl. Material), indicated by the size of the circles (Fig. 4). Only classes with significant pairwise post-hoc comparisons are shown. *P. putida* (gray) and *Serratia* (blue) classes appear to dominate orthogonal pathways. We therefore indicated how the pathways were influenced by the dominant community (indicated by the arrow colour). The link between the TCA cycle and amino acid synthesis (black arrow) is unclear. We further illustrate the substrates and hypothetical environmental conditions expected for *Serratia* (rain/cold) and *P. putida* (dry/hot) communities. We suggest the other communities are intermediates between these two classes. The *Paenibacillus* class, with a large number of sporulation and germination genes, may reflect conditions of very low nutrient availability.

influenced species composition, consistent with previous findings on macro-invertebrate tree-holes communities [48]. Moreover, particular classes were found in different seasons suggesting that factors like temperature were of secondary importance, despite results highlighting their importance in similar systems [25]. Communities collected within the same day were more similar if they were closer in space, demonstrating significant spatial autocorrelation, which has previously been reported in soil environments [47, 12].

Laboratory experiments confirmed that these classes were associated with different functional capacities, which we believe strongly implies that the classes are ecologically meaningful subgroups. The result was compatible with a scenario of ecological succession in which there was a transition from communities dominated by r-strategists to K-strategists [49]. We suggest that early successional stages were characterized by the *Serratia* class. This class had a negative relationship between ATP and cell yield, indicating low resource use efficiency. In addition, investing in xylosidase had a much lower transfer into ATP production than for the reference class, implying a preference for labile substrates like sugar monomers. Analysis of the metagenome revealed pathways responsible for extracellular degradation and uptake of nutrients, and metabolic processes associated with glycolysis. The class also had many genes associated with environmental processing, fast replication and accurate molecular control of protein folding and trafficking. The mean Shannon diversity of communities belonging to this class was almost the lowest (Suppl. Table 2) which might be expected in a rich environment dominated by a few well adapted fast growers, consistent with the notion of r-strategists.

The next communities in the succession were the reference and *Klebsiella* classes. Although still sharing some of the features of the *Serratia* class, they had distinctive features such as a higher conversion of ATP into yield. Later successional stages were characterized by the *P. putida* and *Streptomyces* classes, exhibiting high respiration values. These classes contained an inflated fraction of genes related to oxidative phosphorylation and were able to synthesise most amino acids. They were also associated with secondary metabolic pathways that may be valuable in environments in which resources are low but where it is possible to scavenge the metabolic by-products of former inhabitants. This is particularly apparent for the *P. putida* class, which also had a higher Shannon diversity including many rare species, consistent with communities dominated by K-strategists competing for rare and heterogeneous resources.

Finally, the *Paenibacillus* class contained many of the metagenomic characteristics of the *P. putida* and *Streptomyces* classes. It was the class with lowest Shannon diversity, and also a large fraction of sporulation and germination genes (Suppl. Fig. 31). These results imply that these communities lived particularly unproductive environments. The laboratory results are consistent with this hypothesis: this community had the largest conversion of chitinase activity into yield, which may reflect its ability to take advantage of the remaining nutrients such as dead arthropod exoskeletons or fungi. Water volume is among the main driver of fungi sporulation in this system [50], which would match our interpretation. Taken together, the results imply this class is the last stage of the succession, where nutrients have been depleted to low levels.

There are several environmental conditions that might be driving succession. First, succession may be due to nutrients dynamics in the tree-holes. A main source of carbon is beech leaf litter, supports meio- and macrofaunal communities [51, 52]. Degradation of leaf litter would be compatible with the succession described. Following leaf fall, any imple sugars would rapidly be used over days to weeks, while starch and cellulose degrade much more slowly [53]. If this is the main driver of succession in tree-holes, we would expect a strong seasonal signal, with a class dominating in autumn. Our data do not support these observations because the month of the year was a relatively poor classifier of the samples, and members of the classes we identified were often from different times of year.

Second, succession may be due to patterns of rainfall. Rainwater can bring nitrogen, sulfate, and other ions into the tree-hole, but the pathway followed by the water (stemflow or throughfall) will influence the final chemical compositions [29, 31]. For example, flushing after heavy rain can reduce phosphate levels to a minimum [30], and labile orthophosphate is expected to increase at later successional stages [31]. In addition, a progressive acidification in tree-holes that do not receive water inputs for long periods is also expected due to nitrification [29, 31]. Rain pulses can therefore have rapid impacts on tree-hole conditions and may explain the similarity of some samples collected at the same date even at distant locations, while other properties of the tree-holes like size, litter content, and the modes of water collection may preclude complete synchronization.

We envisage a scenario in which rain events were the primary drivers of bacterial composition, illustrated in bottom-left corner of Fig. 5, which would be modified by tree-hole features (e.g. volume, leaf inputs). Rain would generate pulsed resources of different type and frequency [54], and tree-holes features would determine the rate of resource attenuation [55]. For instance, large tree-holes or those with large leaf contents would have a slower rate of succession since resources are depleted less rapidly. This hypothesis would explain why, on some dates, all

the tree-holes had similar compositions (recent rain or long standing drought conditions) while, beyond that, the classes are distributed across different dates and sites (due to the differential tempos of succession in tree-holes with different features).

Dissolved oxygen may be a third environmental component that influences community composition. The *P. Putida* class were associated with genes involved in aerobic respiration and high levels of phosphate. We observed an increase in abundances of strict aerobes, including *Brevundimonas*, *Paucimonas* and *Phyllobacterium*. There was also an increase in genes related with metabolism of nitrate, methane, degradation of benzoate (likely associated with the presence of resines), or chlorophyl (which indicates an increase in photo-heterotrophs). This class might also be able to run the TCA cycle generating acetyl-CoA from acetate, and from the degradation of valine, leucine and isoleucine, further complemented with glyoxylate metabolism and the degradation of benzoate to generate oxaloacetate. Finally, the class was found in summer and winter, and clustered in specific areas. This makes it less likely that temperature is an important variable, and points towards the amount of water and oxygen as key variables. This observation could also hold for the *Paenibacillus* class, for which long drought periods could lead to lack of water regardless of other tree-holes features (Suppl. Fig. 10).

We cannot rule out other site-based conditions like the type of forest management. A study analysing this factor did not find substantial differences in enzymatic activities despite different community compositions [56], perhaps because the low number of samples did not bring sufficient resolution. Another possible local influence for the composition are trophic ecological interactions, like the prevalence of invertebrates in certain areas (e.g. mosquito larvae) [57]. Insects with flying stages may also influence dispersal among tree-holes, which might contribute to microbial community similarity within a site [58], resulting in a metacommunity structure [48].

The approach taken here provides detailed insights into the community ecology of the bacterial communities inhabiting rainwater pools. By identifying community classes a priori, we were able to piece together the natural history of this environment from the perspective of the bacterial taxa. The spatial and temporal distribution of these classes, combined with the inferred metagenomes, indicate how environmental conditions reflect the metabolic specialisations of the dominant members. In this way, we were able to identify classes resembling r- vs. K-strategists [49] inhabiting tree-holes that were at different successional stages, a distinction also apparent in gut's microbiomes [?]. Although this is no doubt an oversimplification, in general we find this conceptual framework is useful for microbes [34], since this ecological dichotomy may well be supported by thermodynamic [59] and protein-allocation trade-offs [60], which might also underlie other observed life history tradeoffs in microbes (e.g. olitgotrophic vs. copiotrophic strategies, [61]). We believe this approach therefore holds great promise for reducing the complexity of microbial community datasets [62], particularly in systems where the microbial communities have not yet been well characterised. In these systems, the approach we have used would generate hypotheses that could become the focus of future experiments or more detailed sampling strategies, therefore forming the basis of a bottom-up synthetic ecology that can be predictive in the wild.

Methods

Dataset

We analyzed 753 bacterial communities sampled in from rainwater-filled beech tree-holes (Fagus spp.) from different locations in the South West of United Kingdom, see Suppl. Table 1. 95% of the samples were collected between 28 of August and 03 of December 2013, being the remaining 5% collected in April 2014. Spatial distances between samples spanned five orders of magnitude (from <1m to >100km). Sampled communities were grown in standard laboratory conditions using a tea of beech leaves as a substrate. After 7 days of growth, communities were characterized by sequencing 16S rRNA amplicon libraries from Ref. [32]. We considered only samples with more than 10K reads, and species with fewer than 100 reads across all samples were removed. This led to a final dataset comprising 680 samples and 618 Operational Taxonomic Units (OTUs) at the 97% of 16 rRNA sequence similarity. In previous work [32], four replicates of each of these communities were revived and regrown in the same media, further supplemented with low quantities of 4 substrates labelled with 4-methylumbelliferon. After 7 days, the experiments quantified the capacity of the communities to degrade xylosidase (abreviated X in the text, cleaves the labile substrate xylose, a monomer prevalent in hemicellulose), of β-chitinase (N, breaks down chitin, which is the major component of arthropod exoskeletons and fungal cell walls), β-glucosidase (G, break down cellulose, the structural component of plants), and phosphatase (P, breaks down organic monoesters for the mineralisation and acquisition of phosphorus). Cells were also counted at the end of the experiment and CO_2 dissipation quantified as a single accumulative measure along the seven days of experiment. Full experimental

346 details can be found in [32].

Determination of classes

348 We computed all-against-all communities dissimilarities with Jensen-Shannon divergence [35], D_{JSD}, and a trans-
349 formation of the SparCC metric [36], D_{SparCC} (see Suppl. Material), and then clustered the samples following
350 a similar approach to the one proposed in Ref. [18] to identify enterotypes. In the text, we call these clusters
351 community classes. The method consists of a Partition Around Medoids (PAM) clustering for both metrics, with
352 the function PAM implemented in the R package CLUSTER [63]. This clustering requires as input the number of
353 output clusters desired k. We performed the clustering considering a wide range of k values and also computing
354 the Calinski-Harabasz index (CH) that quantifies the quality of the classification, and selecting as optimal clas-
355 sification $k_{\mathrm{opt}} = \arg\max_k(CH)$, shown in Fig. 2b. Processing of data and taxa summaries provided as Suppl.
356 Material were generated with QIIME [64] and Phyloseq [65].

Relation between community similarity and the sampling date and location

358 To investigate the relationship between the sampling location, the sampling date and the similarity in composition
359 of bacterial communities, we performed analysis of the similarities of the communities grouping them with different
360 criteria and testing if the similarities within groups were significantly different than the similarities between
361 groups, using both D_{JSD} and D_{SparCC}. We considered as grouping units one automatic spatial classification and
362 two temporal classifications in which samples are joined in clusters depending on whether they were collected in
363 the same day, or in the same month. Details for the spatial automatic classification and results for two other
364 definitions of sampling sites (see Suppl. Material). We clustered the communities in spatial areas A of increasing
365 sizes every order of magnitude, from 10 m^2. to 100 km^2, which we approximate considering spatial distances'
366 cut-offs of $\sqrt{A}$ meters. We then computed the ANOSIM, MRPP and PERMANOVA tests (see Refs. [37, 38]) for
367 each of the resultant classifications, using the R functions ANOSIM, MRPP and ADONIS2, respectively (available in
368 the R package VEGAN [66]) and assessing the significance with permutation tests (10^3 permutations). To interpret
369 the observed trends of these metrics we created synthetic distance matrices following different criteria, available
370 in Suppl. Material.

Structural equation modelling

372 Structural equation models [41] were built and analysed with LAVAAN (version 0.523) and visualized with SEMPLOT
373 R package [67, 68]. The modelling procedure was split into different stages detailed in Suppl. Material. First, a
374 global model considering all data was investigated following several theoretical assumptions about the relationship
375 between the functions, until a final model was achieved. Then, we looked for a second series of models in which it
376 was possible to fit a different coefficient for each of the parameters in the global model, constraining the data into
377 subsets corresponding to the community classes (i.e., six possible coefficients for each SEM pathway). Minor re-
378 specification of the model was performed (see Suppl. Material). We investigated whether altering the constraints
379 on the models provided better fits, and penalized the models according to the number of degrees of freedom.
380 The main criterion to accept a change was that the AIC of the modified model was smaller than the original
381 model [69]. We verified that a several estimators were improved after any modification, including the RMSEA,
382 the Comparative Fit Index and the Tucker-Lewis Index [70, 71].
383   Investigating causal relationships between endogenous and exogenous variables within the final specified model
384 required controlling for confounding factors. For each pathway in the regression in the SEM model, we identified
385 its adjustment set with dagitty [42]. We then performed a linear regression of each pathway adjusted by the
386 confounding factors, adding a factor coding for the different classes. The coefficients obtained from the regression
387 were estimated with respect to the reference class. Finally, we identified significant interaction terms between
388 classes and the exogenous variable under investigation in the pathway. A significant interaction coefficient involv-
389 ing a given class was interpreted as a different performance of that class with respect to the reference class, and
390 was therefore used to identify distinctive functional features of each class.

Metagenomic analysis

392 Metagenomics predictions were performed using PiCRUST v1.1.2 [43] and quality controls computed (Suppl.
393 Table. 8). A subset of genes appearing at intermediate frequencies was selected (Suppl. Fig. 17) and aggregated
394 into KEGG pathways [72]. The mean proportion of genes assigned to a specific pathway was computed across

395   communities belonging to the same class. Then we tested if the differences in mean proportions between classes
396   were statistically significant using post-hoc tests with STAMP [73] (see Suppl. Material). To create Fig. 5 we
397   visually inspected each post-hoc test and ranked the classes according with the number of pairwise tests in which
398   they appeared significantly inflated (Suppl. Figs. 22–32). We qualitatively represent this ranking with circles of
399   different sizes. Classes that do not appear inflated in any pairwise test in the pathway are not represented.

Acknowledgements

References

[1] L. R. Thompson, J. G. Sanders, D. McDonald, A. Amir, J. Ladau, K. J. Locey, R. J. Prill, A. Tripathi, S. M. Gibbons, G. Ackermann, *et al.*, "A communal catalogue reveals Earth's multiscale microbial diversity," *Nature*, vol. 551, no. 7681, p. 457, 2017.

[2] R. I. Griffiths, B. C. Thomson, P. James, T. Bell, M. Bailey, and A. S. Whiteley, "The bacterial biogeography of british soils," *Environmental microbiology*, vol. 13, no. 6, pp. 1642–1654, 2011.

[3] E. B. Graham, J. E. Knelman, A. Schindlbacher, S. Siciliano, M. Breulmann, A. Yannarell, J. Beman, G. Abell, L. Philippot, J. Prosser, *et al.*, "Microbes as engines of ecosystem function: when does community structure enhance predictions of ecosystem processes?," *Frontiers in microbiology*, vol. 7, p. 214, 2016.

[4] S. Langenheder, M. T. Bulling, M. Solan, and J. I. Prosser, "Bacterial biodiversity-ecosystem functioning relations are modified by environmental complexity," *PloS One*, vol. 5, no. 5, p. e10834, 2010.

[5] G. Bell, "Neutral macroecology," *Science*, vol. 293, no. 5539, pp. 2413–2418, 2001.

[6] A. J. Dumbrell, M. Nelson, T. Helgason, C. Dytham, and A. H. Fitter, "Relative roles of niche and neutral processes in structuring a soil microbial community," *The ISME Journal*, vol. 4, pp. 337–345, Nov. 2009.

[7] T. Caruso, Y. Chan, D. C. Lacap, M. C. Y. Lau, C. P. McKay, and S. B. Pointing, "Stochastic and deterministic processes interact in the assembly of desert microbial communities on a global scale," *The ISME Journal*, vol. 5, pp. 1406–1413, Sept. 2011.

[8] M. Berga, Y. Zha, A. J. Székely, and S. Langenheder, "Functional and compositional stability of bacterial metacommunities in response to salinity changes," *Frontiers in microbiology*, vol. 8, p. 948, 2017.

[9] E. Schnyder, P. L. Bodelier, M. Hartmann, R. Henneberger, and P. A. Niklaus, "Positive diversity-functioning relationships in model communities of methanotrophic bacteria," *Ecology*, vol. 99, no. 3, pp. 714–723, 2018.

[10] T. Fukami, I. A. Dickie, J. Paula Wilkie, B. C. Paulus, D. Park, A. Roberts, P. K. Buchanan, and R. B. Allen, "Assembly history dictates ecosystem functioning: evidence from wood decomposer communities," *Ecology letters*, vol. 13, no. 6, pp. 675–684, 2010.

[11] V. K. Ridaura, J. J. Faith, F. E. Rey, J. Cheng, A. E. Duncan, A. L. Kau, N. W. Griffin, V. Lombard, B. Henrissat, J. R. Bain, M. J. Muehlbauer, O. Ilkayeva, C. F. Semenkovich, K. Funai, D. K. Hayashi, B. J. Lyle, M. C. Martini, L. K. Ursell, J. C. Clemente, W. V. Treuren, W. A. Walters, R. Knight, C. B. Newgard, A. C. Heath, and J. I. Gordon, "Gut Microbiota from Twins Discordant for Obesity Modulate Metabolism in Mice," *Science*, vol. 341, p. 1241214, June 2013.

[12] D. Bru, A. Ramette, N. P. A. Saby, S. Dequiedt, L. Ranjard, C. Jolivet, D. Arrouays, and L. Philippot, "Determinants of the distribution of nitrogen-cycling microbial communities at the landscape scale," *The ISME Journal*, vol. 5, pp. 532–542, Mar. 2011.

[13] B. Jenkins, R. Kitching, and S. Pimm, "Productivity, disturbance and food web structure at a local spatial scale in experimental container habitats," *Oikos*, pp. 249–255, 1992.

[14] D. Gravel, T. Bell, C. Barbera, T. Bouvier, T. Pommier, P. Venail, and N. Mouquet, "Experimental niche evolution alters the strength of the diversity–productivity relationship," *Nature*, vol. 469, no. 7328, p. 89, 2011.

[15] D. W. Rivett and T. Bell, "Abundance determines the functional role of bacterial phylotypes in complex communities," *Nature Microbiology*, vol. 3, no. 7, p. 767, 2018.

[16] T. N. Enke, M. S. Datta, J. Schwartzman, N. Cermak, D. Schmitz, J. Barrere, A. Pascual-García, and O. X. Cordero, "Modular assembly of polysaccharide-degrading marine microbial communities," *Current Biology*, vol. 29, no. 9, pp. 1528–1535, 2019.

[17] K. Faust, J. F. Sathirapongsasuti, J. Izard, N. Segata, D. Gevers, J. Raes, and C. Huttenhower, "Microbial co-occurrence relationships in the human microbiome," *PLoS computational biology*, vol. 8, no. 7, p. e1002606, 2012.

16

[18] M. Arumugam, J. Raes, E. Pelletier, D. Le Paslier, T. Yamada, D. R. Mende, G. R. Fernandes, J. Tap, T. Bruls, J.-M. Batto, *et al.*, "Enterotypes of the human gut microbiome," *Nature*, vol. 473, no. 7346, pp. 174–180, 2011.

[19] O. Koren, D. Knights, A. Gonzalez, L. Waldron, N. Segata, R. Knight, C. Huttenhower, and R. E. Ley, "A guide to enterotypes across the human body: Meta-analysis of microbial community structures in human microbiome datasets," *PLOS Computational Biology*, vol. 9, p. e1002863, Jan. 2013.

[20] J. M. Chase and J. A. Myers, "Disentangling the importance of ecological niches from stochastic processes across scales," *Philosophical transactions of the Royal Society B: Biological sciences*, vol. 366, no. 1576, pp. 2351–2363, 2011.

[21] F. Dini-Andreote, J. C. Stegen, J. D. van Elsas, and J. F. Salles, "Disentangling mechanisms that mediate the balance between stochastic and deterministic processes in microbial succession," *Proceedings of the National Academy of Sciences*, vol. 112, no. 11, pp. E1326–E1332, 2015.

[22] J. E. Koenig, A. Spor, N. Scalfone, A. D. Fricker, J. Stombaugh, R. Knight, L. T. Angenent, and R. E. Ley, "Succession of microbial consortia in the developing infant gut microbiome," *Proceedings of the National Academy of Sciences*, vol. 108, no. Supplement 1, pp. 4578–4585, 2011.

[23] S. Louca, L. W. Parfrey, and M. Doebeli, "Decoupling function and taxonomy in the global ocean microbiome," *Science*, vol. 353, no. 6305, pp. 1272–1277, 2016.

[24] J. M. Bhatnagar, K. G. Peay, and K. K. Treseder, "Litter chemistry influences decomposition through activity of specific microbial functional guilds," *Ecological Monographs*, vol. 88, pp. 429–444, Aug. 2018.

[25] X. Zhang, E. R. Johnston, A. Barberán, Y. Ren, X. Lü, and X. Han, "Decreased plant productivity resulting from plant group removal experiment constrains soil microbial functional diversity," *Global Change Biology*, vol. 23, pp. 4318–4332, Oct. 2017.

[26] Y. Shi, P. Grogan, H. Sun, J. Xiong, Y. Yang, J. Zhou, and H. Chu, "Multi-scale variability analysis reveals the importance of spatial distance in shaping Arctic soil microbial functional communities," *Soil Biology and Biochemistry*, vol. 86, pp. 126–134, July 2015.

[27] K. A. R. Nishadh, A. DAS, and K. Sakthidas, "Tree-hole aquatic habitats: inhabitants, processes and experiments. A review.," *International Journal of Conservation Science*, vol. 5, no. 2, 2014.

[28] R. Kitching, "Food webs in phytotelmata: bottom-up and top-down explanations for community structure," *Annual Review of Entomology*, vol. 46, no. 1, pp. 729–760, 2001.

[29] S. R. Carpenter, "Stemflow chemistry: effects on population dynamics of detritivorous mosquitoes in tree-hole ecosystems," *Oecologia*, vol. 53, no. 1, pp. 1–6, 1982.

[30] E. D. Walker, D. L. Lawson, R. W. Merritt, W. T. Morgan, and M. J. Klug, "Nutrient dynamics, bacterial populations, and mosquito productivity in tree hole ecosystems and microcosms," *Ecology*, vol. 72, pp. 1529–1546, Oct. 1991.

[31] R. C. Verdonschot, C. M. Febria, and D. D. Williams, "Fluxes of dissolved organic carbon, other nutrients and microbial communities in a water-filled treehole ecosystem," *Hydrobiologia*, vol. 596, no. 1, pp. 17–30, 2008.

[32] D. W. Rivett and T. Bell, "Abundance determines the functional role of bacterial phylotypes in complex communities," *Nature microbiology*, p. 1, 2018.

[33] K. Nakasaki, K. Nag, and S. Karita, "Microbial succession associated with organic matter decomposition during thermophilic composting of organic waste," *Waste Management & Research*, vol. 23, pp. 48–56, Feb. 2005.

[34] J. H. Andrews and R. F. Harris, "r-and k-selection and microbial ecology," in *Advances in microbial ecology*, pp. 99–147, Springer, 1986.

[35] D. M. Endres and J. E. Schindelin, "A new metric for probability distributions," *IEEE Transactions on Information theory*, vol. 49, no. 7, pp. 1858–1860, 2003.

[36] J. Friedman and E. J. Alm, "Inferring correlation networks from genomic survey data," *PLoS computational biology*, vol. 8, no. 9, p. e1002687, 2012.

[37] K. R. Clarke, "Non-parametric multivariate analyses of changes in community structure," *Austral Ecology*, vol. 18, no. 1, pp. 117–143, 1993.

[38] D. I. Warton, S. T. Wright, and Y. Wang, "Distance-based multivariate analyses confound location and dispersion effects," *Methods in Ecology and Evolution*, vol. 3, no. 1, pp. 89–101, 2012.

[39] A. Pascual-García and T. Bell, "functionink: An efficient method to detect functional groups in multidimensional networks reveals the hidden structure of ecological communities," *bioRxiv*, p. 656504, 2019.

[40] R. L. Sinsabaugh, B. H. Hill, and J. J. F. Shah, "Ecoenzymatic stoichiometry of microbial organic nutrient acquisition in soil and sediment," *Nature*, vol. 462, no. 7274, p. 795, 2009.

[41] J. B. Grace, *Structural equation modeling and natural systems*. Cambridge University Press, 2006.

[42] J. Textor, B. van der Zander, M. S. Gilthorpe, M. Liśkiewicz, and G. T. Ellison, "Robust causal inference using directed acyclic graphs: The R package dagitty," *International journal of epidemiology*, vol. 45, no. 6, pp. 1887–1894, 2016.

[43] M. G. Langille, J. Zaneveld, J. G. Caporaso, D. McDonald, D. Knights, J. A. Reyes, J. C. Clemente, D. E. Burkepile, R. L. V. Thurber, R. Knight, *et al.*, "Predictive functional profiling of microbial communities using 16s rRNA marker gene sequences," *Nature biotechnology*, vol. 31, no. 9, pp. 814–821, 2013.

[44] D. H. Parks, G. W. Tyson, P. Hugenholtz, and R. G. Beiko, "Stamp: statistical analysis of taxonomic and functional profiles," *Bioinformatics*, vol. 30, no. 21, pp. 3123–3124, 2014.

[45] A. Pascual-García, J. Tamames, and U. Bastolla, "Bacteria dialog with santa rosalia: Are aggregations of cosmopolitan bacteria mainly explained by habitat filtering or by ecological interactions?," *BMC Microbiology*, vol. 14, no. 1, p. 284, 2014.

[46] B. R. Roller, S. F. Stoddard, and T. M. Schmidt, "Exploiting rRNA operon copy number to investigate bacterial reproductive strategies," *Nature microbiology*, vol. 1, p. 16160, 2016.

[47] R. B. Franklin and A. L. Mills, "Multi-scale variation in spatial heterogeneity for microbial community structure in an eastern Virginia agricultural field," *FEMS Microbiology Ecology*, vol. 44, pp. 335–346, June 2003.

[48] C. J. Paradise, J. D. Blue, J. Q. Burkhart, J. Goldberg, L. Harshaw, K. D. Hawkins, B. Kegan, T. Krentz, L. Smith, and S. Villalpando, "Local and regional factors influence the structure of treehole metacommunities," *BMC ecology*, vol. 8, no. 1, p. 22, 2008.

[49] M. Gadgil and O. T. Solbrig, "The concept of r-and k-selection: evidence from wild flowers and some theoretical considerations," *The American Naturalist*, vol. 106, no. 947, pp. 14–31, 1972.

[50] J. Gönczöl and Á. Révay, "Treehole fungal communities: aquatic, aero-aquatic and dematiaceous hyphomycetes," *Fungal Diversity*, vol. 12, pp. 19–34, 2003.

[51] C. Ptatscheck and W. Traunspurger, "Meio-and macrofaunal communities in artificial water-filled tree holes: Effects of seasonality, physical and chemical parameters, and availability of food resources," *PloS one*, vol. 10, no. 8, p. e0133447, 2015.

[52] C. Ptatscheck and W. Traunspurger, "The meiofauna of artificial water-filled tree holes: colonization and bottom-up effects," *Aquatic ecology*, vol. 48, no. 3, pp. 285–295, 2014.

[53] M. K. Aneja, S. Sharma, F. Fleischmann, S. Stich, W. Heller, G. Bahnweg, J. C. Munch, and M. Schloter, "Microbial colonization of beech and spruce litter–influence of decomposition site and plant litter species on the diversity of microbial community," *Microbial Ecology*, vol. 52, pp. 127–135, July 2006.

[54] D. A. Yee and S. A. Juliano, "Concurrent effects of resource pulse amount, type, and frequency on community and population properties of consumers in detritus-based systems," *Oecologia*, vol. 169, no. 2, pp. 511–522, 2012.

[55] D. W. Rivett, T. Scheuerl, C. T. Culbert, S. B. Mombrikotb, E. Johnstone, T. G. Barraclough, and T. Bell, "Resource-dependent attenuation of species interactions during bacterial succession," *The ISME journal*, vol. 10, no. 9, p. 2259, 2016.

[56] W. Purahong, D. Kapturska, M. J. Pecyna, E. Schulz, M. Schloter, F. Buscot, M. Hofrichter, and D. Krüger, "Influence of different forest system management practices on leaf litter decomposition rates, nutrient dynamics and the activity of ligninolytic enzymes: A case study from Central European forests," *PLoS ONE*, vol. 9, p. e93700, Apr. 2014.

[57] Y. Xu, S. Chen, M. G. Kaufman, S. Maknojia, M. Bagdasarian, and E. D. Walker, "Bacterial community structure in tree hole habitats of ochlerotatus triseriatus: Influences of larval feeding," *Journal of the American Mosquito Control Association*, vol. 24, pp. 219–227, June 2008.

[58] B. Maguire Jr, "The passive dispersal of small aquatic organisms and their colonization of isolated bodies of water," *Ecological Monographs*, vol. 33, no. 2, pp. 161–185, 1963.

[59] T. Pfeiffer, S. Schuster, and S. Bonhoeffer, "Cooperation and competition in the evolution of ATP-producing pathways," *Science*, vol. 292, pp. 504–507, Apr. 2001.

[60] D. W. Erickson, S. J. Schink, V. Patsalo, J. R. Williamson, U. Gerland, and T. Hwa, "A global resource allocation strategy governs growth transition kinetics of escherichia coli," *Nature*, vol. 551, no. 7678, p. 119, 2017.

[61] F. M. Lauro, D. McDougald, T. Thomas, T. J. Williams, S. Egan, S. Rice, M. Z. DeMaere, L. Ting, H. Ertan, J. Johnson, *et al.*, "The genomic basis of trophic strategy in marine bacteria," *Proceedings of the National Academy of Sciences*, vol. 106, no. 37, pp. 15527–15533, 2009.

[62] A. Pascual-García, S. Bonhoeffer, and T. Bell, "Microbial metabolically cohesive consortia and ecosystem functioning," *Under review*, 2019.

[63] A. P. Reynolds, G. Richards, B. de la Iglesia, and V. J. Rayward-Smith, "Clustering rules: a comparison of partitioning and hierarchical clustering algorithms," *Journal of Mathematical Modelling and Algorithms*, vol. 5, no. 4, pp. 475–504, 2006.

[64] J. G. Caporaso, J. Kuczynski, J. Stombaugh, K. Bittinger, F. D. Bushman, E. K. Costello, N. Fierer, A. G. Peña, J. K. Goodrich, J. I. Gordon, *et al.*, "Qiime allows analysis of high-throughput community sequencing data," *Nature methods*, vol. 7, no. 5, pp. 335–336, 2010.

[65] P. J. McMurdie and S. Holmes, "phyloseq: an r package for reproducible interactive analysis and graphics of microbiome census data," *PloS One*, vol. 8, no. 4, p. e61217, 2013.

[66] J. Oksanen, F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, E. Szoecs, and H. Wagner, *vegan: Community Ecology Package*, 2019. R package version 2.5-5.

[67] Y. Rosseel, "lavaan: An R package for structural equation modeling," *Journal of Statistical Software*, vol. 48, no. 2, pp. 1–36, 2012.

[68] S. Epskamp, "semplot: Unified visualizations of structural equation models," *Structural Equation Modeling*, vol. 22, no. 3, pp. 474–483, 2015.

[69] L.-C. Lin, P.-H. Huang, and L.-J. Weng, "Selecting path models in sem: A comparison of model selection criteria," *Structural Equation Modeling: A Multidisciplinary Journal*, vol. 24, no. 6, pp. 855–869, 2017.

[70] P. Barrett, "Structural equation modelling: Adjudging model fit," *Personality and Individual differences*, vol. 42, no. 5, pp. 815–824, 2007.

585 [71] J. C. Loehlin, *Latent variable models: An introduction to factor, path, and structural analysis.* Lawrence
586   Erlbaum Associates Publishers, 1998.

587 [72] M. Kanehisa and S. Goto, "Kegg: kyoto encyclopedia of genes and genomes," *Nucleic acids research*, vol. 28,
588   no. 1, pp. 27–30, 2000.

589 [73] J. R. White, N. Nagarajan, and M. Pop, "Statistical methods for detecting differentially abundant features
590   in clinical metagenomic samples," *PLoS computational biology*, vol. 5, no. 4, p. e1000352, 2009.