environmental microbiology

Environmental Microbiology (2018) **20**(12), 4356–4368 doi:10.1111/1462-2920.14352

Generality of associations between biological richness and the rates of metabolic processes across microbial communities

Deborah Patsch,[1,2] Simon van Vliet,[1,2] Lorenzo Garbani Marcantini[3] and David R. Johnson[2*]

[1]*Department of Environmental Systems Science, ETH Zürich, 8092, Zürich, Switzerland.*

[2]*Department of Environmental Microbiology, Swiss Federal Institute of Aquatic Science and Technology (Eawag), 8600, Dübendorf, Switzerland.*

[3]*Department of Urban Water Management, Swiss Federal Institute of Aquatic Science and Technology (Eawag), 8600, Dübendorf, Switzerland.*

Summary

Biological richness is positively associated with the rates of some metabolic processes performed by microbial communities. It remains unclear, however, whether these positive associations are a general feature of the metabolic processes performed by microbial communities or whether they are specific to certain types of metabolic processes. For example, it was hypothesized that the strength of any particular positive association depends on how many different genotypes within a microbial community perform the metabolic process of interest (i.e. the 'rarity hypothesis'). We tested the generality of these positive associations by measuring the taxonomic richness, functional gene richness and rate constants for 71 different metabolic processes across 30 independent microbial communities. We found that both taxonomic and functional gene richness do indeed tend to positively associate with the rates of metabolic processes. In addition, we found that positive associations occur across a wide range of different environmental conditions. Counter to the 'rarity hypothesis', however, we did not detect a relationship between the strengths of the positive associations and the rarity of each metabolic process. Together, our data provide empirical evidence that positive associations with biological richness may indeed be a general feature of the metabolic processes performed by microbial communities.

Received 23 April, 2018; revised 2 July, 2018; accepted 2 July, 2018. *For correspondence. E-mail david.johnson@eawag.ch; Tel. (+41) 58 765 55 20.

Introduction

Microbial communities perform metabolic processes that provide important services to human society and our environment. The ecological factors that control the rates of particular metabolic processes, however, are not fully understood. Biological richness (e.g. the number of functionally different genotypes within a community) is one factor that may be important (Bell *et al*., 2005; Duffy, 2009; Cardinale, 2011; Gravel *et al*., 2011; Peter *et al*., 2011; Bouvier *et al*., 2012; Cardinale *et al*., 2012; Dell'Anno *et al*., 2012; Hernandez-Raquet *et al*., 2013; Philippot *et al*., 2013; Johnson *et al*., 2015a; Evans *et al*., 2017; Stadler *et al*., 2017). Communities with higher biological richness are more likely to contain genotypes that have direct or indirect positive effects on the rate of a particular metabolic process (Loreau *et al*., 2001; Balvanera *et al*., 2006; Cardinale *et al*., 2006; Cardinale, 2011; Cardinale *et al*., 2012; Tilman *et al*., 2014). These positive effects may emerge as a consequence of several different ecological mechanisms, including complementation, facilitation or sampling effects (Loreau *et al*., 2001; Balvanera *et al*., 2006; Cardinale *et al*., 2006; Cardinale, 2011; Cardinale *et al*., 2012; Tilman *et al*., 2014). Complementation occurs when two or more genotypes perform a particular metabolic process but occupy partially- or non-overlapping niche spaces, where each genotype performs the metabolic process more effectively within its own niche space. For example, one genotype might specialize at performing the metabolic process in a planktonic state while the other might specialize at performing the metabolic process in a biofilm-associated state. The collection of specialized genotypes could consequently attain faster aggregate rates of the metabolic process than would any single genotype. Facilitation occurs when the growth or metabolic activities of genotypes that perform a particular metabolic process are stimulated by the presence of other genotypes (e.g. the other genotype may consume or

detoxify a growth-reducing metabolite). The sampling effect is a probabilistic argument, which states that communities with higher biological richness are more likely to contain genotypes that bestow positive effects on the rate of a particular metabolic process. This may result from the presence of genotypes that perform the metabolic process at faster rates or confer novel complementation or facilitation effects. Counter to complementation and facilitation, however, the sampling effect can also act in the negative direction, where communities with higher biological richness are also more likely to contain genotypes that bestow negative effects on the rate of a particular metabolic process.

While positive associations between biological richness and the rates of particular metabolic processes have been observed (Bell *et al*., 2005; Cardinale, 2011; Gravel *et al*., 2011; Peter *et al*., 2011; Bouvier *et al*., 2012; Cardinale *et al*., 2012; Dell'Anno *et al*., 2012; Hernandez-Raquet *et al*., 2013; Philippot *et al*., 2013; Johnson *et al*., 2015a; Evans *et al*., 2017; Stadler *et al*., 2017), it remains unclear whether these positive associations are a general feature of the metabolic processes performed by microbial communities. For example, while many studies have observed positive associations between biological richness and the rates of some metabolic processes (Bell *et al*., 2005; Cardinale, 2011; Gravel *et al*., 2011; Peter *et al*., 2011; Bouvier *et al*., 2012; Cardinale *et al*., 2012; Dell'Anno *et al*., 2012; Hernandez-Raquet *et al*., 2013; Philippot *et al*., 2013; Johnson *et al*., 2015a; Evans *et al*., 2017; Stadler *et al*., 2017), other studies have not (Salonius, 1981; Griffiths *et al*., 2001; Wertz *et al*., 2006; Szabó *et al*., 2007; Peter *et al*., 2011; Pholchan *et al*., 2013; Graham *et al*., 2014; Roger *et al*., 2016).

One possible explanation for why biological richness may positively associate with the rates of only specific types of metabolic processes is the 'rarity hypothesis' (Levine *et al*., 2011; Johnson *et al*., 2015a). Briefly, if only a few genotypes perform a particular metabolic process (i.e. a rare metabolic process), then differences or changes in biological richness are more likely to have measurable effects on the rate of that metabolic process than if many genotypes perform that metabolic process, thus leading to quantitatively stronger positive associations. This is because the set of possible complementation, facilitation and sampling effects are less likely to be near saturation, and the addition or removal of genotypes is therefore more likely to affect that metabolic process (i.e. functional redundancy is low). In contrast, if many genotypes perform a particular metabolic process (i.e. a common metabolic process), then differences or changes in biological richness are less likely to have measurable effects on the rate of that metabolic process than if only a few genotypes perform that metabolic process, thus leading to quantitatively weaker positive associations. This is because the set of possible complementation, facilitation and sampling effects are more likely to be near saturation, and the addition or removal of genotypes is less likely to affect the rate of that metabolic process (i.e. functional redundancy is high). Importantly, the 'rarity hypothesis' does not require the binary categorization of metabolic processes as either rare or common, but instead considers the magnitude of rarity or commonality (i.e. the number of genotypes that contribute towards a metabolic process or the amount of functional redundancy related to that metabolic process). While the 'rarity hypothesis' provides a plausible explanation, empirical tests of the hypothesis remain limited to only a few types and relatively small sets of metabolic processes (e.g. Levine *et al*., 2011; Johnson *et al*., 2015a).

The goal of this study was twofold. First, we sought to test whether positive associations between biological richness and the rates of metabolic processes are generally observed across a large number of different metabolic processes or whether they are specific to certain types of metabolic processes. Second, we sought to test the 'rarity hypothesis' (Levine *et al*., 2011; Johnson *et al*., 2015a); namely, is the strength of any particular positive association related to how many different genotypes within a microbial community perform that particular metabolic process? To achieve these goals, we obtained 30 independent microbial communities from 30 different wastewater treatment plants (WWTPs). We then measured taxonomic richness (R_t) and functional gene richness (R_f) using non-target DNA sequencing-based approaches for each microbial community. We included functional gene richness (R_f) in our analyses because functional genes are likely responsible for providing many complementation and facilitation effects. In parallel, we measured the rate constants (k) for 71 different metabolic processes for each microbial community using parallelized batch assays. We next tested for associations between taxonomic richness (R_t) or functional gene richness (R_f) and the rate constants (k) for each metabolic process. Finally, we estimated the rarity of each metabolic process and asked whether the strengths of the positive associations are related to the rarity of each metabolic process.

Results

Associations between biological richness and the rates of metabolic processes are generally positive in sign

Our first goal was to test whether associations between biological richness and the rates of each metabolic process are generally positive in sign across a large number of different metabolic processes. To achieve this goal, we measured observed taxonomic richness ($R_{t,\ obs}$)

(Supporting Information, Table S1) and observed functional gene richness ($R_{f,\ obs}$) (Supporting Information, Table S1) along with the rate constants (k) for 71 different metabolic processes (Supporting Information, Table S2) across 30 independent microbial communities collected from 30 different WWTPs. We then quantified the Spearman correlation coefficients ($r_{Spearman}$) between observed taxonomic richness ($R_{t,\ obs}$) or observed functional gene richness ($R_{f,\ obs}$) and the rate constants (k) for each metabolic process (Supporting Information, Table S3). We used the non-parametric Spearman rank correlation test because 30 of the 71 metabolic processes have rate constants (k) that significantly deviate from a normal distribution (Shapiro–Wilk test; $P > 0.05$) (Supporting Information, Table S4), thus invalidating a central assumption of parametric correlation tests (e.g. the Pearson correlation test). We note that our goal was not to draw conclusions about any particular metabolic process; the statistical significance of any individual Spearman correlation coefficient ($r_{Spearman}$) is therefore unimportant for our statistical objective. Instead, our goal was to test whether the distribution of Spearman correlation coefficients ($r_{Spearman}$) across a large number of different metabolic processes has a central tendency that significantly deviates from zero towards the positive direction, which would be expected if positive associations were indeed a general feature of the metabolic processes performed by microbial communities.

We found that the rate constants (k) for each metabolic process vary substantially (1.6- to 28-fold for those with all non-zero measurements) across the different microbial communities (Fig. 1), which is an essential prerequisite for our analyses (variation is required to observe an association). We further found that the central tendencies of the Spearman correlation coefficients ($r_{Spearman}$) for observed taxonomic richness ($R_{t,\ obs}$) (Fig. 2A) and observed functional gene richness ($R_{f,\ obs}$) (Fig. 2D) do indeed significantly deviate from zero towards the positive direction (two-sided Wilcoxon rank-sum test; $P < 0.00001$ for observed taxonomic richness [$R_{t,\ obs}$], $P < 0.002$ forobserved functional gene richness [$R_{f,\ obs}$]). For example, observed taxonomic richness ($R_{t,\ obs}$) has positive Spearman correlation coefficients ($r_{Spearman}$) with the rate constants (k) for 55 out of the 71 (e.g. 77%) different metabolic processes (Fig. 2A and Supporting Information, Table S3). This qualitative outcome also occurs when using extrapolated measures (Chao1 or ACE) of taxonomic richness ($R_{t,\ Chao1}$, $R_{t,\ ACE}$) or functional gene richness ($R_{f,\ Chao1}$, $R_{f,\ ACE}$) (two-sided Wilcoxon rank-sum test; $P < 0.00001$ for Chao1 and ACE taxonomic richness [$R_{t,\ Chao1}$, $R_{t,\ ACE}$], $P < 0.001$ for Chao1 and ACE functional gene richness [$R_{f,\ Chao1}$, $R_{f,\ ACE}$]) (Fig. 2). Interestingly, we note that taxonomic richness tends to generate stronger positive associations than functional gene richness. While this outcome has been observed before, this tendency may not be a general phenomenon but instead may be linked to the methodology used to quantify functional richness (Johnson *et al*., 2015a; 2015b). Regardless, our data suggests that positive associations do indeed generally occur across a large number of different metabolic processes.

One potential limitation of our analyses above is that we only tested whether the central tendencies of the Spearman correlation coefficients ($r_{Spearman}$) significantly deviate from a singular value of zero. We therefore performed a randomization test where we randomly re-labelled the rate constants (k) for each metabolic process to a different microbial community and recalculated the Spearman correlation coefficients ($r_{Spearman,\ rand}$) (Supporting Information, Fig. S1). We then tested whether the Spearman correlation coefficients generated from the original data ($r_{Spearman}$) (Fig. 2) significantly deviate from the Spearman correlation coefficients generated from the randomly re-labelled data ($r_{Spearman,\ rand}$) (Supporting Information, Fig. S1). As expected, the central tendencies of the Spearman correlation coefficients generated from the randomly re-labelled data ($r_{Spearman,\ rand}$) do not significantly deviate from zero, regardless of whether we used measures of taxonomic richness (R_t) or functional gene richness (R_f) (two-sided Wilcoxon rank-sum test; $P > 0.3$) (Supporting Information, Fig. S1). Moreover, the central tendencies of the Spearman correlation coefficients generated from the original data ($r_{Spearman}$) are all significantly greater than the central tendencies of the Spearman correlation coefficients generated from the randomly re-labelled data ($r_{Spearman,\ rand}$), again regardless of whether we used measures of taxonomic richness (R_t) or functional gene richness (R_f) (two-sided Wilcoxon rank-sum test; $P < 0.02$) (Fig. 2 and Supporting Information, Fig. S1). This provides further support that positive associations do indeed generally occur across a large number of different metabolic processes.

Positive associations occur across a wide range of different environmental conditions

We next asked whether positive associations between biological richness and the rates of metabolic processes occur across a wide range of different environmental conditions. More specifically, we previously (see above) analysed the rate constants (k) for 71 different metabolic processes under a single standardized environmental condition, while here we analyse the rate constants (k) for a single metabolic process (i.e. the consumption of a single carbon substrate) in the presence of 23 different environmental conditions (different environmental conditions include a different pH, a different concentration of

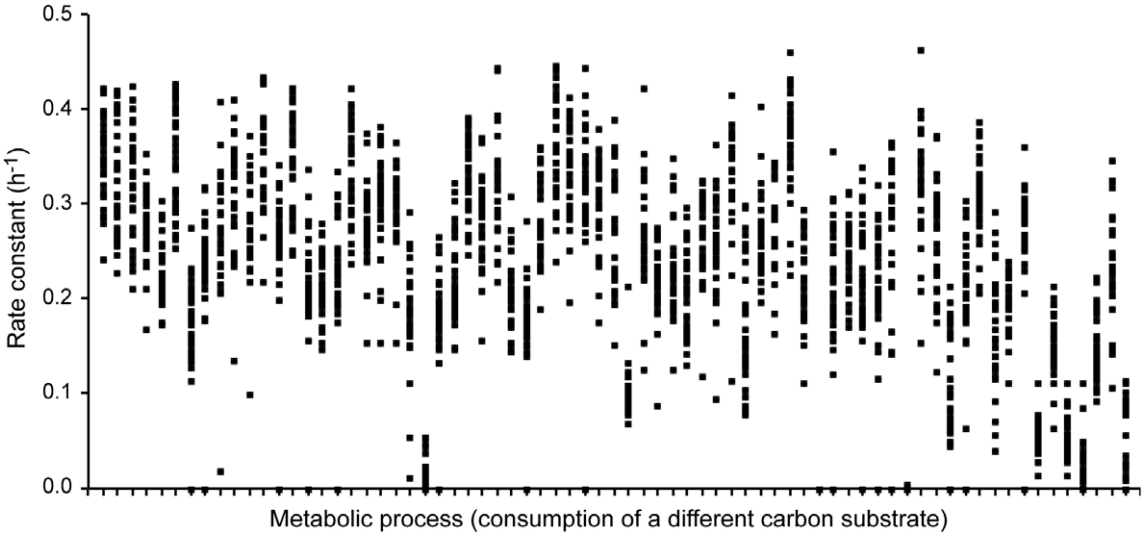

Fig. 1. Rate constants (k) for each metabolic process (n = 71). Symbols are the measurements for each microbial community. Metabolic processes are ordered in columns from left to right in accordance with the ordering in Supporting Information, Table S8. Each metabolic process is the consumption of a different carbon substrate.

sodium chloride or the presence of a different stress-inducing chemical). We found that the central tendencies of the Spearman correlation coefficients ($r_{Spearman}$) between taxonomic richness (R_t) or functional gene richness (R_f) and the rate constants (k) for a single metabolic process across the different environmental conditions again significantly deviate from zero towards the positive direction (two-sided Wilcoxon rank-sum test; $P < 0.05$) (Fig. 3). This supports the argument that positive associations between biological richness and the rates of metabolic processes are generalizable across a wide range of different environmental conditions.

No observed relationship between association strength and rarity

Our second goal was to test the 'rarity hypothesis' (Levine *et al*., 2011; Johnson *et al*., 2015a); namely, the strengths of the positive associations between biological richness and the rates of metabolic processes should be greater for rare metabolic processes than for common metabolic processes. We first tested this hypothesis using a dilution-to-extinction approach to estimate the rarity of each metabolic process, where we assumed that the extent of dilution required for a metabolic process to no longer be measurable provides a proxy estimate of its rarity (Garland and Lehman, 1999). We report the dilutions at which each of the 71 different metabolic processes were no longer measurable in Supporting Information, Table S5 (values are for one randomly selected microbial community). When we tested the specific hypothesis that the extents of dilution negatively associate with the magnitudes of the Spearman correlation coefficients ($r_{Spearman}$) between taxonomic richness (R_t) or functional gene richness (R_f) and the rate constants (k) for each metabolic process (smaller extents of dilution indicate greater rarity), we did not detect a statistically significant negative association (Spearman rank correlation test; $P > 0.2$) (Fig. 4). Thus, counter to the 'rarity hypothesis' (Levine *et al*., 2011; Johnson *et al*., 2015a), we have no statistical evidence for a relationship between the rarity of a particular metabolic process and the strength of its positive association between biological richness and the rates of that metabolic process.

We next tested the 'rarity hypothesis' using an alternative estimate of rarity: the lag-time before each metabolic process is measurable. Here we assumed that the lag-time provides an alternative proxy estimate of its rarity. We report the lag-times before each of the 71 different metabolic processes became measurable in Supporting Information, Table S6. We found that the lag-times are significantly negatively associated with the dilutions at which each ecosystem function was no longer measurable (Spearman rank correlation test; $P < 0.001$) (Fig. 5), which would be expected if both approaches do indeed contain substantial information about the rarity of each metabolic process. However, when we tested the specific hypothesis that the lag-times positively associate with the magnitudes of the Spearman correlation coefficients ($r_{Spearman}$) between taxonomic richness (R_t) or functional gene richness (R_f) and the rate constants (k) for each metabolic process (longer lag-times indicate greater rarity), we again did not detect a statistically significant positive association (Spearman rank correlation test; $P > 0.99$) (Fig. 6). Thus, similar to the outcome from the dilution-to-extinction approach, we have no statistical

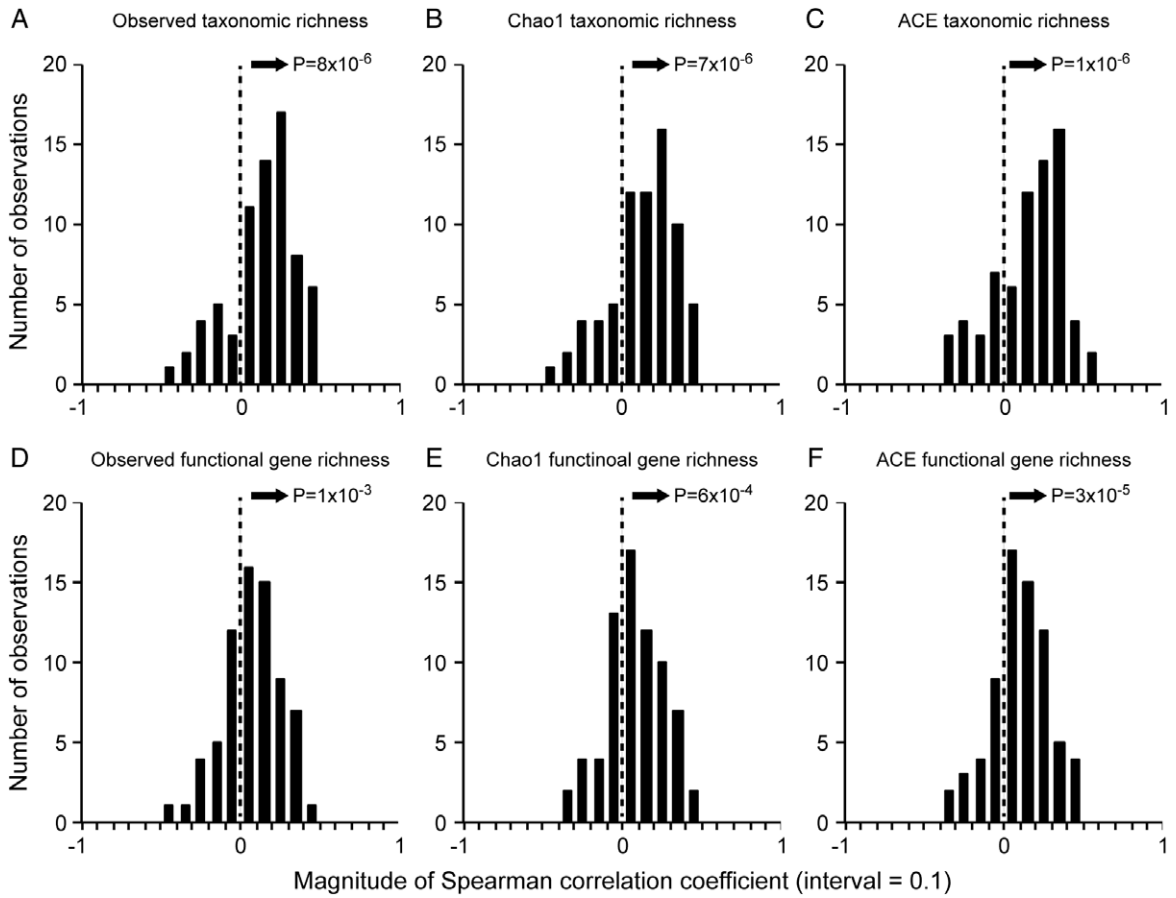

Fig. 2. Spearman correlation coefficients ($r_{Spearman}$) for the associations between taxonomic richness (R_t) or functional gene richness (R_f) and the rate constants (k) of each metabolic process (n = 71). Each metabolic process is the consumption of a different carbon substrate. Richness measurements include A., observed taxonomic richness ($R_{t,\ obs}$); B., Chao1 taxonomic richness ($R_{t,\ Chao1}$); C., ACE taxonomic richness ($R_{t,\ ACE}$); D., observed functional richness ($R_{f,\ obs}$); E., Chao1 functional richness ($R_{f,\ Chao1}$); or F., ACE functional richness ($R_{f,\ ACE}$). The dashed vertical lines indicate a Spearman correlation coefficient ($r_{Spearman}$) of zero. The P values indicate that the central tendencies of the Spearman correlation coefficients ($r_{Spearman}$) significantly deviate from a value of zero. The horizontal arrows indicate the direction of the deviation of the central tendencies from a value of zero.

evidence for a relationship between the rarity of a particular metabolic process and the strength of its positive association between biological richness and the rates of that metabolic process.

Discussion

We found that across a large number of different metabolic processes, biological richness tends to positively associate with the rates of those metabolic processes (Fig. 2). This was true regardless of whether we used measures of taxonomic richness (R_t) or functional gene richness (R_f) or whether we used observed or extrapolated (Chao1 or ACE) measures of biological richness (Fig. 2). Moreover, we also observed these positive associations for a single metabolic process across a wide range of different environmental conditions (Fig. 3). Thus, our data lend credence to the suggestion that positive associations with biological richness may indeed be a general feature of the metabolic processes performed by microbial communities. Additionally, the pervasiveness of positive associations suggests that functional redundancy within WWTP microbial communities in particular may be lower than sometimes assumed, as excessive functional redundancy should indicate near-saturation of biodiversity effects and consequently obscure these positive associations (Johnson *et al*., 2015a). Together, our data suggest that promoting and maintaining biological richness may be particularly important for optimizing the rates of metabolic processes performed by WWTP microbial communities. Moreover, assessing differences or changes in biological richness may be an effective strategy to predict the functional performance of WWTPs.

While we observed positive associations between biological richness and the rates of metabolic processes, our data do not provide statistical support for the 'rarity hypothesis' (Figs. 4 and 6). We emphasize, however, that the absence of evidence supporting a hypothesis is not

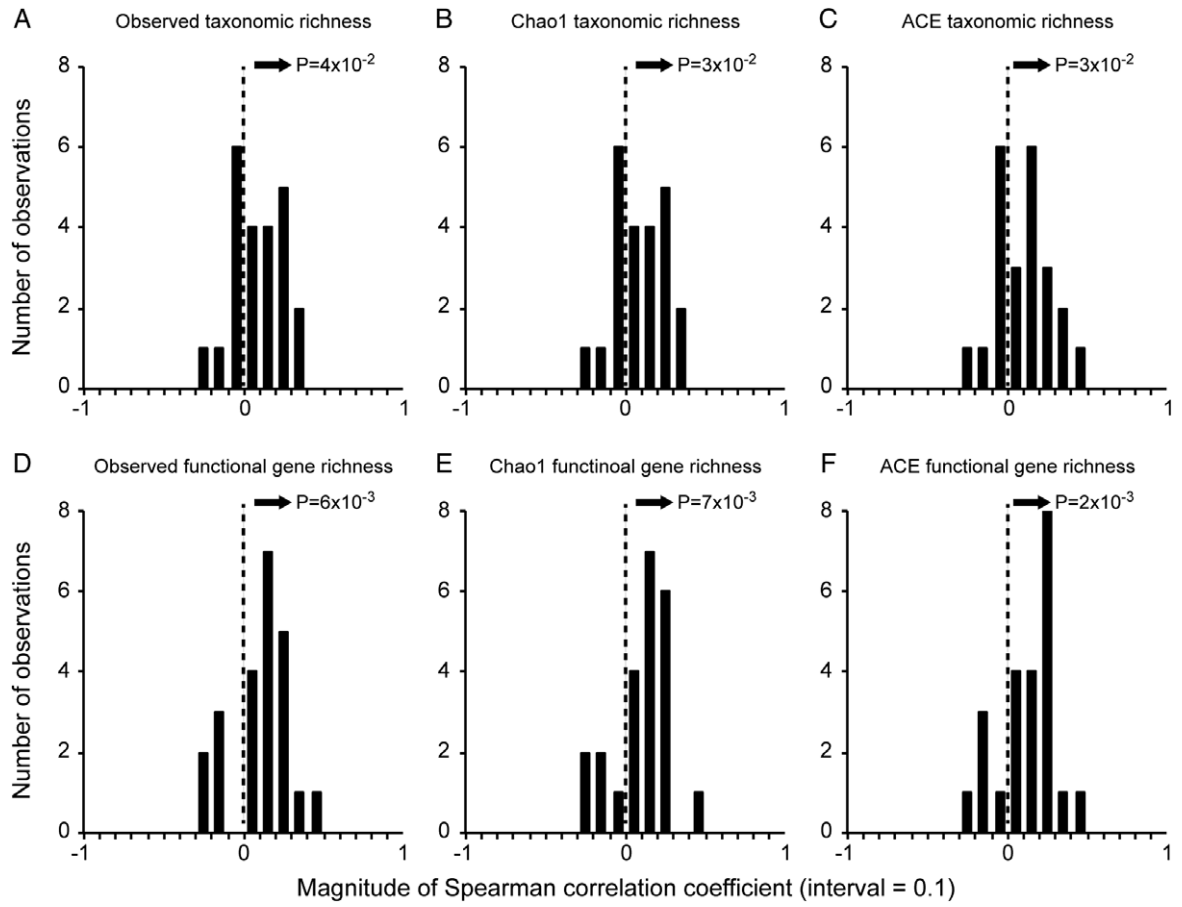

Fig. 3. Spearman correlation coefficients ($r_{Spearman}$) for the associations between taxonomic richness (R_t) or functional richness (R_f) and the rate constants (k) of a single metabolic process in the presence of different environmental conditions (n = 23). Richness measurements include A., observed taxonomic richness ($R_{t,\ obs}$); B., Chao1 taxonomic richness ($R_{t,\ Chao1}$); C., ACE taxonomic richness ($R_{t,\ ACE}$); D., observed functional richness ($R_{f,\ obs}$); E., Chao1 functional richness ($R_{f,\ Chao1}$); or F., ACE functional richness ($R_{f,\ ACE}$). The dashed vertical lines indicate a Spearman correlation coefficient ($r_{Spearman}$) of zero. The P values indicate that the central tendencies of the Spearman correlation coefficients ($r_{Spearman}$) significantly deviate from a value of zero. The horizontal arrows indicate the direction of the deviation of the central tendencies from a value of zero.

the same as evidence against the hypothesis. One aspect of our experimental design that could obscure the predicted effects of rarity is that growth and substrate consumption were coupled during our analyses. We diluted the microbial communities prior to measuring the rates of each metabolic process, which provided opportunity for the preferential growth of certain genotypes over the time-scale of the experimental analyses. If preferential growth were strong, then a few genotypes with especially large metabolic rates may have rapidly increased in frequency. This could modify the native genotype frequencies and consequently disrupt the native complementation and facilitation effects, thus obscuring the expected relationship between rarity and the strengths of the positive associations. Indeed, in our previous study where we did observe the expected relationship between rarity and the strengths of the positive associations (Johnson *et al*., 2015a), we used very low substrate concentrations and types of substrates that were unlikely to support substantial growth. This resulted in the approximate uncoupling of growth from substrate consumption, which likely helped preserve the native genotype frequencies and maintain more of the native complementation and facilitation effects present within those microbial communities. An alternative explanation is that functional redundancy is generally low within our microbial communities, which could mask the effect of rarity on the associations between biological richness and the rates of metabolic processes.

We believe that our results are potentially generalizable beyond WWTP microbial communities. It is plausible that positive relationships between biological richness and the rates of metabolic processes may be particularly weak in environments with high immigration rates, as high immigration rates may introduce new genotypes that increase biological richness but have no positive effects on the rates of particular metabolic processes. Importantly, WWTPs have this feature (high immigration rates),

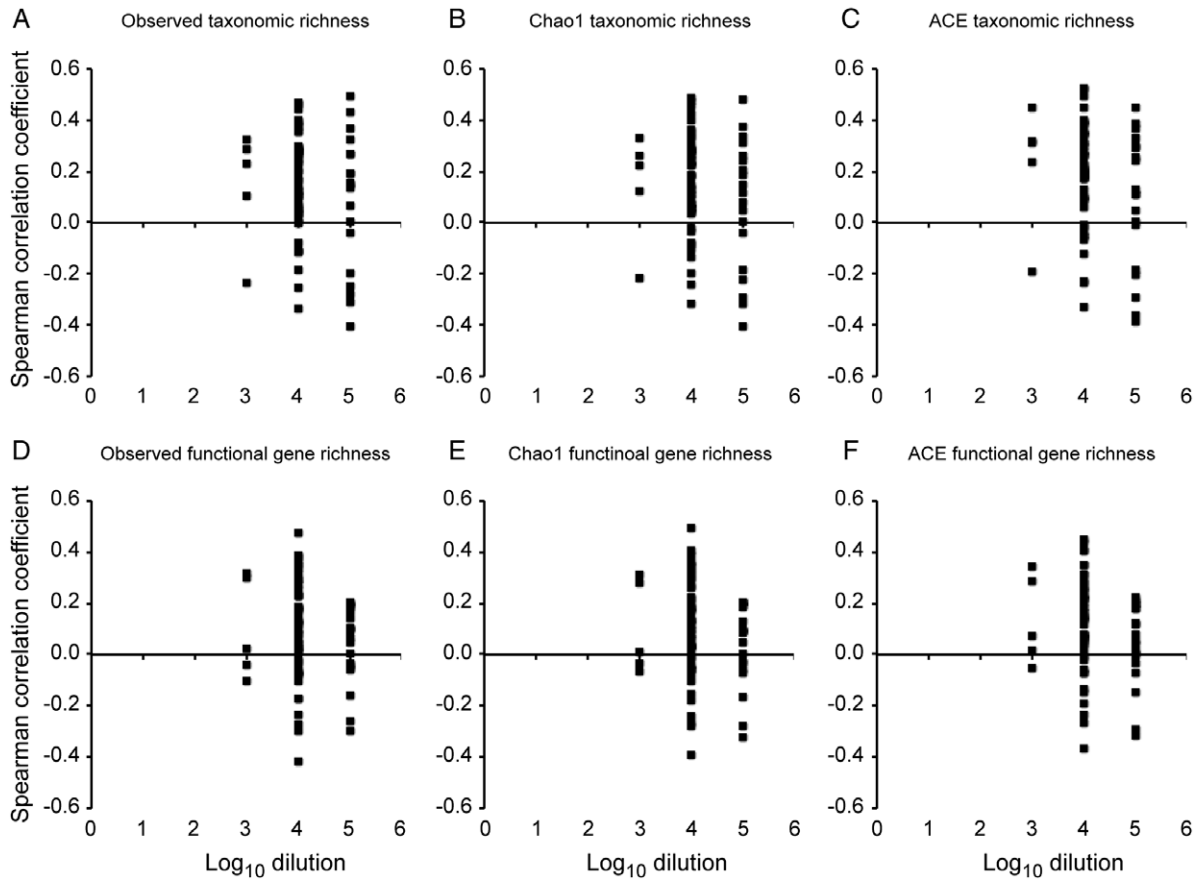

Fig. 4. Spearman correlation coefficients ($r_{Spearman}$) plotted against the extent of dilution required for each metabolic process to no longer be measurable. The Spearman correlation coefficients ($r_{Spearman}$) are for the associations between taxonomic richness (R_t) or functional richness (R_f) and the rate constants (k) of each metabolic process, where each metabolic process is the consumption of a different carbon substrate. Richness measurements include A., observed taxonomic richness ($R_{t,\ obs}$); B., Chao1 taxonomic richness ($R_{t,\ Chao1}$); C., ACE taxonomic richness ($R_{t,\ ACE}$); D., observed functional richness ($R_{f,\ obs}$); E., Chao1 functional richness ($R_{f,\ Chao1}$); or F., ACE functional richness ($R_{f,\ ACE}$).

yet we still observed statistically significant positive associations for a large number of metabolic processes (Fig. 2) and across a wide range of different environmental conditions (Fig. 3). Thus, we believe that our conclusions are conservative with regards to their generalizability to other types of microbial communities and environmental conditions. In conclusion, our data suggest that biological richness may indeed be an important determinant of the rates of many different types of metabolic processes performed by microbial communities.

Experimental procedures

Microbial communities

We obtained 30 independent microbial communities, each of which was collected from a different WWTP located in Switzerland (29 WWTPs) or Austria (1 WWTP) (Supporting Information, Table S7). We collected the microbial communities using a consistent protocol described elsewhere (Helbling *et al*., 2012). Briefly, we collected a single 1 l liquid sample directly from the aeration basin of each WWTP in a 2 l glass bottle. We then transported the sample back to the laboratories at the Swiss Federal Institute of Aquatic Science and Technology (Eawag) and began preparing and analysing the sample (see below) ~2.5 h after collection from the WWTP.

Sample preparation

Prior to measuring biological richness or the rates of metabolic processes, we first homogenized each microbial community to minimize any artefacts that may arise from sampling small volumes from spatially structured microbial communities, such as the microbial communities residing within WWTPs. To achieve this, we used a method described previously for WWTP microbial communities (Victorio *et al*., 1996). Briefly, we transferred 2 × 40 ml aliquots of each microbial community from the 2 l glass bottle into 2 × 50 ml screw-top plastic tubes. We next mixed each aliquot with a laboratory homogenizer for 10 s, centrifuged the aliquots at 2800g, and discarded

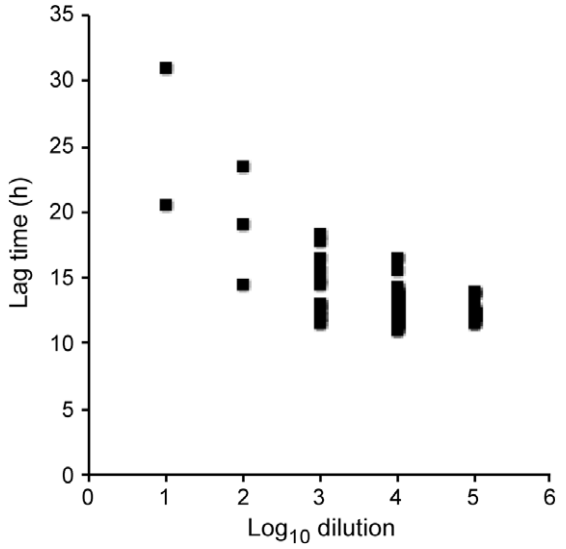

Fig. 5. Relationship between the extent of dilution required for each metabolic process to no longer be measurable and the lag time (median value among all 30 microbial communities) before each metabolic process was measurable. The extent of dilution is significantly and negatively associated with the lag time (Spearman rank correlation test; $P < 0.001$).

the supernatants. We then suspended the remaining cells in 40 ml of 0.85% NaCl and repeated the centrifugation/suspension for a total of three times to remove residual nutrients. Finally, we transferred the cell suspensions to new 50 ml screw-top plastic tubes and measured the optical density at 590 nm [OD_{590}] using a spectrophotometer. For each microbial community, we used one 40 ml cell aliquot for measuring taxonomic richness and functional gene richness (described below) and the other 40 ml cell aliquot for measuring the rates of each metabolic process (described below).

Isolation of genomic DNA for biodiversity measurements

We centrifuged one of the 40 ml cell aliquots for each microbial community at 4700g for 5 min, removed the supernatant and suspended the remaining cells in 1 ml of IF-A solution (Biolog, Hayward, CA). IF-A solution is a cell suspension solution that contains no carbon substrate but does contain the tetrazolium dye used to measure the rates of respiration coupled to each metabolic process via the GEN III MicroPlate™ (Biolog, Hayward, CA) (described below). This creates the same conditions that we used to measure the rates of each metabolic process; hence, if some genotypes were disturbed by the IF-A solution when we measured the rates of each metabolic process, they would also be disturbed prior to measuring taxonomic richness and functional gene richness. After 30 min of incubation in IF-A solution, we centrifuged the cells again at 4700g for 5 min, discarded the supernatant and stored the samples at –20°C until further processing.

We used a conventional phenol-chloroform-based method to isolate genomic DNA from the cell samples as described elsewhere (Johnson *et al*., 2015b). We quantified the masses of isolated genomic DNA using a Qubit Fluorometer with the Qubit dsDNA High Sensitivity Assay Kit (Thermo Fisher Scientific, Waltham, MA). We then sent the isolated genomic DNA on dry ice to the Quantitative Genomics Facility (QGF) at ETH Zürich in Basel, Switzerland for sequencing.

Non-target genomic DNA sequencing

At the QGF, genomic DNA was fragmented with an E220 focused ultrasonicator (Covaris, Woburn, MA) and libraries with an insert size of 400–700 base pairs were prepared using the KAPA Hyper Prep Kit (Applied Biosystems, Waltham, MA). Sequencing was performed with an Illumina HiSeq 2500 system (Illumina, San Diego, CA) with paired-end sequencing (PE125). The sequencing effort was designed to generate ~10 million sequence reads per sample.

Sequence analysis

We used Kraken (version 0.10.5) with the provided MiniKraken database (version 141208) (Wood and Salzberg, 2014; Lindgreen *et al*., 2016) to assign the DNA sequence reads obtained from the genomic DNA sequencing effort to taxonomic identities. We only considered exact matches to the Kraken-Genbank database using fully default parameters (Wood and Salzberg, 2014; Lindgreen *et al*., 2016). We considered all DNA sequence reads for taxonomic identification rather than restricting the analyses to 16S-rDNA sequence reads. We used the EBI Metagenomics (EMG) tool (https://www.ebi.ac.uk/metagenomics/) (Mitchell *et al*., 2016) and the associated InterProScan tool (Jones *et al*., 2014) to assign the protein-encoding DNA sequence reads obtained from the genomic DNA sequencing effort to putative functional groups using fully default parameters. Putative functional annotations are obtained based on matches to the InterPro database (Finn *et al*., 2017), which returns a single functional annotation based on identified putative protein domains. We define a functional group as the single functional annotation returned by InterProScan (Jones *et al.,* 2014), and we refer to these functional groups as putative because they are bioinformatics hypotheses. We also used the EMG tool to merge paired ends, perform quality control (including trimming of adapter sequences and masking of non-coding sequences) and predict coding sequences (Mitchell *et al*., 2016). All metadata, raw genomic DNA sequence reads and MD5checksums are publicly

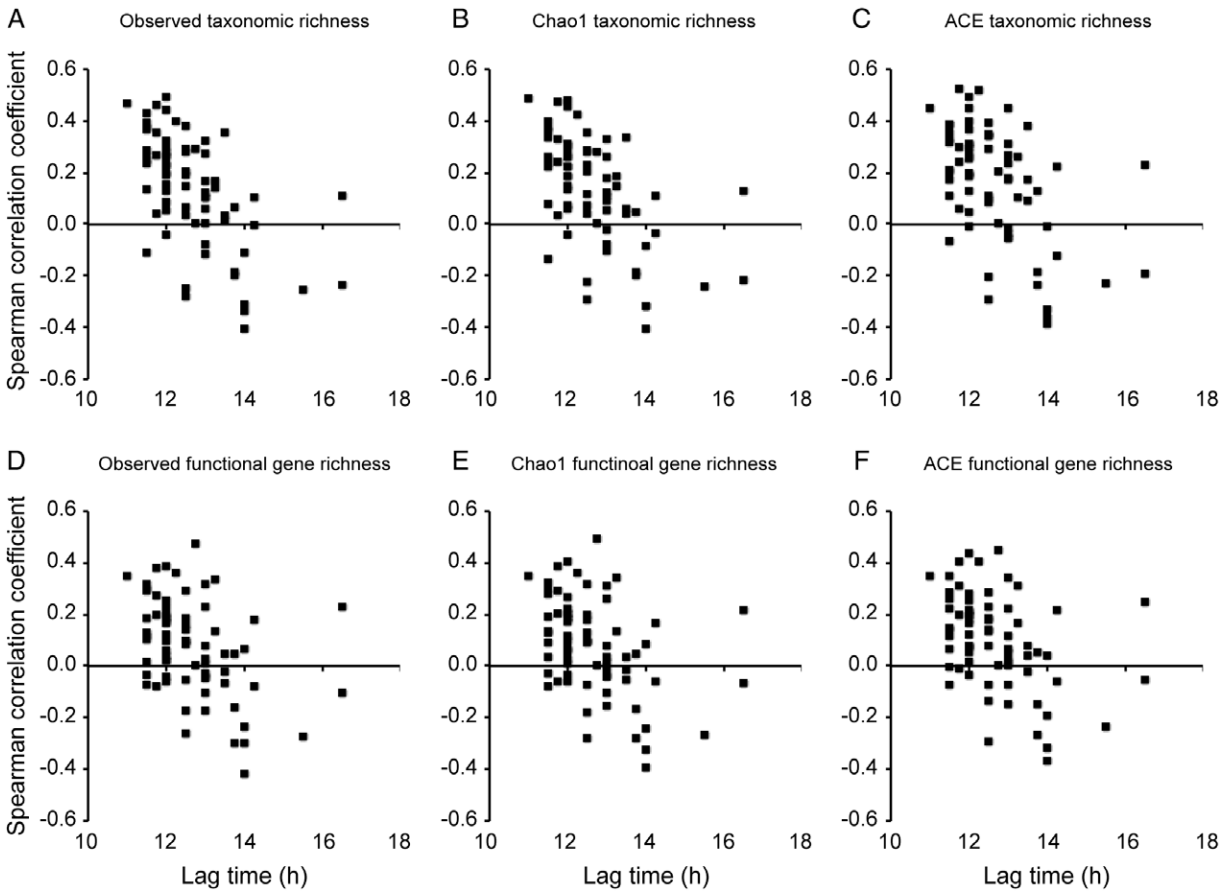

Fig. 6. Spearman correlation coefficients ($r_{Spearman}$) plotted against the lag times (median values among all 30 microbial communities) before each metabolic process was measurable. The Spearman correlation coefficients ($r_{Spearman}$) are for the associations between taxonomic richness (R_t) or functional richness (R_f) and the rate constants (k) of each metabolic process, where each metabolic process is the consumption of a different carbon substrate. Richness measurements include A., observed taxonomic richness ($R_{t,\ obs}$); B., Chao1 taxonomic richness ($R_{t,\ Chao1}$); C., ACE taxonomic richness ($R_{t,\ ACE}$); D., observed functional richness ($R_{f,\ obs}$); E., Chao1 functional richness ($R_{f,\ Chao1}$); or F., ACE functional richness ($R_{f,\ ACE}$).

available at the EBI metagenomics database under accession number PRJEB13232.

Quantifying taxonomic richness and functional gene richness

We quantified taxonomic richness (R_t) from the DNA sequence reads that were assigned to taxonomic identities in Kraken (Wood and Salzberg, 2014; Lindgreen *et al*., 2016) (i.e. we discarded all DNA sequence reads that did not match perfectly to the Kraken-Genbank database). We quantified functional gene richness (R_f) from the protein-encoding DNA sequence reads that were assigned a putative functional annotation in EMG (Mitchell *et al*., 2016) (i.e. we discarded all protein-encoding DNA sequence reads that did not match to the InterPro database when using default parameters). We rarefied all sequence datasets to the single sequence dataset containing the smallest number of taxonomically identified DNA sequence reads or functional group assigned protein-encoding sequence reads. We quantified observed taxonomic richness ($R_{t,\ obs}$) as the observed number of unique taxonomic identities per rarified sequence dataset. We quantified observed functional gene richness ($R_{f,\ obs}$) as the observed number of unique functional groups per rarified sequence dataset as described elsewhere (Johnson *et al*., 2015a,b). We additionally quantified Chao1 extrapolated richness ($R_{t,\ Chao1}$, $R_{f,\ Chao1}$) and ACE extrapolated richness ($R_{t,\ ACE}$, $R_{f,\ ACE}$) (Chao, 1987; Chao *et al*., 1993; Hughes *et al*., 2001; Gotelli and Ellison, 2004) from the rarified sequence datasets. All the reported taxonomic richness (R_t) and functional gene richness (R_f) measurements are the mean values from 100 independent rarefactions. We performed rarefaction and calculated taxonomic richness (R_t) and functional gene richness (R_f) in the R environment (R Development Core Team, 2011) using functions from the vegan (Oksanen *et al*., 2013) and fossil (Vavrek, 2011) packages. We report all the taxonomic richness (R_t) and functional gene richness (R_f) measurements for each microbial community in Supporting Information, Table S1.

Quantifying the rates of each metabolic process

For this study, we define the rate of a metabolic process as the rate of respiration coupled to the consumption of a particular carbon substrate. In total, we quantified the rates of respiration coupled to the consumption of 71 different carbon substrates in parallel using the GEN III MicroPlate™ (Biolog, Hayward, CA) (hereafter referred to as a plate). These plates consist of 96 wells, each of which contains a different growth-limiting carbon substrate (n = 71), a single growth-limiting carbon substrate in the presence of a different environmental condition (n = 23; different environmental conditions include a different pH, a different concentration of sodium chloride or the presence of a different stress-inducing chemical), or a negative or positive control (n = 2) (summarized in Supporting Information, Table S8). The wells also contain a tetrazolium dye, which is oxidized during respiration and can be used to measure the rate of respiration coupled to the consumption of each carbon substrate as per the manufacturer's recommendations (Biolog, Hayward, CA).

To perform these analyses, we used the second 40 ml cell aliquot from each microbial community. We first adjusted the OD_{590} of the cell suspension to 0.08 by diluting the sample in 0.85% NaCl solution as proposed elsewhere (Yang *et al*., 2011). We then added 100 μl of cell suspension to each well of a plate and measured the OD_{590} over time using an Eon plate reader (BioTek, Luzern, Switzerland) with the following protocol: continuous linear shaking, constant temperature at 25 °C and OD_{590} measurements every 0.5 h for 36 h for a total of 72 measurements. In parallel, we measured the background signal by filling three plates with 0.85% NaCl solution that did not contain any cells. We next calculated the mean value of these background measurements for each well and subtracted the mean values from all the measurements that contained cells. We then calculated the respiration rate constant (h^{-1}) coupled to the consumption of each carbon substrate in the MATLAB environment (MathWorks, Natick, MA) using the following equation:

$$OD_{590}(t) = OD_{590}(0)e^{kt}$$

where k is the respiration rate constant (h^{-1}), $OD_{590}(0)$ is the OD_{590} measurement at time zero and $OD_{590}(t)$ is the OD_{590} measurement at time t (h). We note here that the cumulative amount of tetrazolium dye that is oxidized via respiration is measured at OD_{590} as per the manufacturer's recommendations (Biolog, Hayward, CA). This measure is therefore not an exact measure of cumulative respiration activity because it is, to some extent, confounded with cell density. However, the presence of this confounding factor should not affect the qualitative outcome or main conclusions of this study, as both cumulative respiration activity and cell density are coupled to the metabolic process of substrate consumption.

We estimated the respiration rate constant (k) coupled to the consumption of each carbon substrate using a sliding time-window analysis. Briefly, we performed a linear regression between the natural log-transformed OD_{590} measurements across six consecutive time-points, calculated the slope and Pearson correlation coefficient ($r_{Pearson}$), and repeated the analysis at increments of one time-point. We then identified the maximum slope that also had a statistically significant ($P < 0.05$) Pearson correlation coefficient ($r_{Pearson}$) and used this value as a measure of the respiration rate constant (k). We report the measured rate constants (k) for each metabolic process and for each microbial community in Supporting Information, Table S2.

Estimating the rarity of each metabolic process

We first estimated the rarity of each metabolic process using a dilution-to-extinction approach as described elsewhere (Garland and Lehman, 1999). The main assumption here is that rare metabolic processes should require less extensive dilution of a microbial community to no longer be measurable than common metabolic processes (Garland and Lehman, 1999). The extent of dilution at which respiration activity is no longer measurable therefore provides a proxy estimate of the rarity of each metabolic process. To perform these measurements, we selected a single microbial community at random for analysis and serially diluted the community with 85% NaCl solution at 10-fold increments up to 10^{-8}. We selected one microbial community at random to maintain objectivity in the analysis and adhere to the underlying assumptions of our statistical approaches. We next added each dilution of the microbial community to a separate plate as described above, incubated the plates for 48 h at 25°C, and measured the OD_{590} at 48 h. Finally, we identified the extent of dilution required for each metabolic process to no longer be measurable after 48 h of incubation (i.e. an OD_{590} measurement that is not statistically larger than the background signal). We report the extent of dilution required for each metabolic process to no longer be measureable in Supporting Information, Table S5.

We additionally estimated the rarity of each metabolic process based on the lag-time between inoculation of the plates and the onset of measurable respiration activity. The main assumption here is that rare metabolic processes should have longer lag-times than common metabolic processes. This is because fewer genotypes perform rare metabolic processes, and they are therefore more likely to be at lower initial abundances. Thus, the lag-time before each metabolic process is measurable

provides an alternative proxy estimate of the rarity of each metabolic process. We calculated the lag-time in the MATLAB environment (MathWorks, Natick, MA) using the following equation (Swinnen *et al*., 2004):

$$t_{lag} = t(OD_{590} > \text{threshold})$$

where t_{lag} is the lag-time and the threshold value was set to 0.2 based on visual inspection of the OD_{590} profiles. More specifically, we first calculated the mean of the first five OD_{590} measurements for each metabolic process and subtracted the background signal (see above). We next repeated the analysis at increments of one time point and estimated t_{lag} as the time point when the background-subtracted OD_{590} measurement was statistically larger than the threshold value. We report the lag-times for each metabolic process in Supporting Information, Table S6.

Association tests

We used the Shapiro–Wilk test to test whether the rate constants (k) for each metabolic process across the 30 microbial communities significantly deviate from a normal distribution. This was indeed the case for a large proportion of the metabolic processes (Shapiro–Wilk test; $P < 0.05$) (Supporting Information, Table S4). We therefore used non-parametric statistical tests for all of our analyses. While non-parametric statistical tests are valid for distributions that deviate from normality, they can also have lower sensitivity and higher false negative rates than their parametric analogues. We therefore took caution when interpreting statistical outcomes that do not significantly deviate from null hypotheses. We used the Spearman rank correlation test to test for associations between taxonomic richness (R_t) or functional gene richness (R_f) and the rate constants (k) for metabolic processes. We used the Wilcoxon rank-sum test to test whether the central tendencies of the Spearman correlation coefficients ($r_{Spearman}$) significantly deviate from zero or from the central tendencies of randomly re-labelled datasets (see the main text for details regarding random re-labelling). We performed all statistical tests in the R environment using core functions (R Development Core Team, 2011).

Acknowledgements

We thank Teresa Colangelo Failla for assistance with collecting and processing the samples and Martin Ackermann for helpful discussions regarding the design, analysis and interpretation of the work. Portions of the data produced and analysed in this manuscript were generated in collaboration with the Genetic Diversity Centre (GDC), ETH Zürich. This work was supported by MERMAID, an initial training network funded by the People Programme – Maria Sklowdowska–Curie Actions – of the European Union's Seventh Framework Programme FP7/2007-2013 under REA grant agreement number 607492.

References

Balvanera, P., Pfisterer, A. B., Buchmann, N., He, J. S., Nakashizuka, T., Raffaelli, D., and Schmid, B. (2006) Quantifying the evidence for biodiversity effects on ecosystem functioning and services. *Ecol Lett* **9**: 1146–1156.

Bell, T., Newmann, J. A., Silverman, B. W., Turner, S. L., and Lilley, A. K. (2005) The contribution of species richness and composition to bacterial services. *Nature* **436**: 1157–1160.

Bouvier, T., Venail, P., Pommier, T., Bouvier, C., Barbera, C., and Nouquet, N. (2012) Contrasted effects of diversity and immigration on ecological insurance in marine bacterioplankton communities. *PLoS ONE* **7**: e37620.

Cardinale, B. J. (2011) Biodiversity improves water quality through niche partitioning. *Nature* **472**: 86–89.

Cardinale, B. J., Srivastava, D. S., Duffy, J. E., Wright, J. P., Downing, A. L., Sankaran, M., *et al*. (2006) Effects of biodiversity on the functioning of trophic groups and ecosystems. *Nature* **443**: 989–992.

Cardinale, B. J., Duffy, J. E., Gonzales, A., Hooper, D. U., Perrings, C., Venail, P., *et al*. (2012) Biodiversity loss and its impact on humanity. *Nature* **486**: 59–67.

Chao, A. (1987) Estimating the population size for capture-recapture data with unequal catchability. *Biometrics* **43**: 783–791.

Chao, A., Ma, M. C., and Yang, M. C. K. (1993) Stopping rules and estimation for recapture debugging with unequal failure rates. *Biometrika* **80**: 193–201.

Dell'Anno, A., Beolchini, F., Rocchetti, L., Luna, G. M., and Danovaro, R. (2012) High bacterial biodiversity increases degradation performance of hydrocarbons during bioremediation of contaminated harbor marine sediments. *Environ Pollut* **167**: 85–92.

Duffy, J. E. (2009) Why biodiversity is important for the functioning of real-world ecosystems. *Front Ecol Environ* **7**: 437–444.

Evans, R., Alessi, A. M., Bird, S., McQueen-Mason, S. J., Bruce, N. C., and Brockhurst, M. A. (2017) Defining the functional traits that drive bacterial decomposer community productivity. *ISME J* **11**: 1680–1687.

Finn, R. D., Attwood, T. K., Babbitt, P. C., Bateman, A., Bork, P., Bridge, A. J., *et al*. (2017) InterPro in 2017 – beyond protein family and domain annotations. *Nucleic Acids Res* **45**: D190–D199.

Garland, J. L., and Lehman, R. M. (1999) Dilution/extinction of community phenotypic characters to estimate relative structural diversity in mixed communities. *FEMS Microbiol Ecol* **30**: 333–343.

Gotelli, N. J., and Ellison, A. M. (2004) *A Primer of Ecological Statistics*. Sunderland, MA, USA: Sinauer Associates Publishers.

Graham, E. B., Wieder, W. R., Leff, J. W., Weintraub, S. R., Townsend, A. R., Cleveland, C. C., *et al*. (2014) Do we

need to understand microbial communities to predict ecosystem function? A comparison of statistical models of nitrogen cycling processes. *Soil Biol Biochem* **68**: 279–282.

Gravel, D., Bell, T., Barbera, C., Bouvier, T., Pommier, T., Venail, P., and Mouquet, N. (2011) Experimental niche evolution alters the strength of the diversity-productivity relationship. *Nature* **469**: 89–U1601.

Griffiths, B. S., Ritz, K., Wheatley, R., Kuan, H. L., Boag, B., Christensen, S., *et al*. (2001) An examination of the biodiversity-ecosystem function relationship in arable soil microbial communities. *Soil Biol Biochem* **33**: 1713–1722.

Helbling, D. E., Johnson, D. R., Honti, M., and Fenner, K. (2012) Micropollutant biotransformation kinetics associate with WWTP process parameters and microbial community characteristics. *Environ Sci Technol* **46**: 10579–10588.

Hernandez-Raquet, G., Durand, E., Braun, F., Cravo-Laureau, C., and Godon, J. J. (2013) Impact of microbial diversity depletion on xenobiotic degradation by sewage-activated sludge. *Environ Microbiol Rep* **5**: 588–594.

Hughes, J. B., Hellmann, J. J., Ricketts, T. H., and Bohannan, B. J. (2001) Counting the uncountable: statistical approaches to estimating microbial diversity. *Appl Environ Microbiol* **67**: 4399–4406.

Johnson, D. R., Helbling, D. E., Lee, T. K., Park, J., Fenner, K., Kohler, H. P., *et al*. (2015a) Association of biodiversity with the rates of micropollutant biotransformations among full-scale wastewater treatment plant communities. *Appl Environ Microbiol* **81**: 666–675.

Johnson, D. R., Lee, T. K., Park, J., Fenner, K., and Helbling, D. E. (2015b) The functional and taxonomic richness of wastewater treatment plant microbial communities are associated with each other and with ambient nitrogen and carbon availability. *Environ Microbiol* **17**: 4851–4860.

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., *et al*. (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**: 1236–1240.

Levine, U. Y., Teal, T. K., Robertson, G. P., and Schmidt, T. M. (2011) Agriculture's impact on microbial diversity and associated fluxes of carbon dioxide and methane. *ISME J* **5**: 1683–1691.

Lindgreen, S., Adair, K. L., and Gardner, P. P. (2016) An evaluation of the accuracy and speed of metagenome analysis tools. *Sci Rep* **6**: 19233.

Loreau, M., Naeem, S., Inchausti, F., Bengtsson, J., Grime, J. F., Hector, A., *et al*. (2001) Biodiversity and ecosystem functioning: current knowledge and future challenges. *Science* **294**: 804–808.

Mitchell, A., Bucchini, F., Cochrane, G., Denise, H., ten Hoopen, P., Fraser, M., *et al*. (2016) EBI metagenomics in 2016--an expanding and evolving resource for the analysis and archiving of metagenomic data. *Nucleic Acids Res* **44**: D595–D603.

Oksanen, A. J., Blanchet, F. G., Kindt, R., Minchin, P. R., O'Hara, R. B. O., Simpson, G. L., *et al*. (2013) *Vegan: community ecology package. R Package Version 2.5*. Vienna, Austria: R Foundation for Statistical Computing.

Peter, H., Beier, S., Bertilsson, S., Lindström, E. S., Langenheder, S., and Tranvik, L. J. (2011) Function-specific response to depletion of microbial diversity. *ISME J* **5**: 351–361.

Philippot, L., Spor, A., Hénault, C., Bru, D., Bizouard, F., Jones, C. M., *et al*. (2013) Loss in microbial diversity affects nitrogen cycling in soil. *ISME J* **7**: 1609–1619.

Pholchan, M. K., Batistia, J. C., Davenport, R. J., Sloan, W. T., and Curtis, T. P. (2013) Microbial community assembly, theory and rare functions. *Front Microbiol* **4**: 68.

R Development Core Team. (2011) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Roger, F., Bertilsson, S., Langenheder, S., Osman, O. A., and Gamfeldt, L. (2016) Effects of multiple dimensions of bacterial diversity on functioning, stability and multifunctionality. *Ecology* **97**: 2716–2728.

Salonius, P. O. (1981) Metabolic capabilities of forest soil microbial populations with reduced species diversity. *Soil Biol Biochem* **13**: 1–10.

Stadler, L. B., Vela, J. D., Jain, S., Dick, G. J., and Love, N. G. (2017) Elucidating the impact of microbial community biodiversity on pharmaceutical biotransformation during wastewater treatment. *Microbiol Biotechnol* [In press] doi: 10.1111/1751-7915.12870

Swinnen, I. A. M., Bernaerts, K., Dens, E. J. J., Geeraerd, A. H., and Van Impe, J. F. (2004) Predictive modelling of the microbial lag phase: a review. *Int J Food Microbiol* **94**: 137–159.

Szabó, K. E., Itor, P., Bertilsson, S., Tranvik, L., and Eiler, A. (2007) Importance of rare and abundant populations for the structure and functional potential of freshwater bacterial communities. *Aquat Microb Ecol* **47**: 1–10.

Tilman, D., Isbell, F., and Cowles, J. M. (2014) Biodiversity and ecosystem functioning. *Annu Rev Ecol Evol Syst* **45**: 471–493.

Vavrek, M. J. (2011) Fossil: palaeoecological and palaeogeographical analysis tools. *Palaeontol Electron* **14**: 1T.

Victorio, L., Gilbride, K. A., Grant Allen, D., and Liss, S. N. (1996) Phenotypic fingerprinting of microbial communities in wastewater treatment systems. *Water Res* **30**: 1077–1086.

Wertz, S., Degrange, V., Prosser, J. I., Poly, F., Commeaux, C., Freitag, T., *et al*. (2006) Maintenance of soil functioning following erosion of microbial diversity. *Environ Microbiol* **8**: 2162–2169.

Wood, D. E., and Salzberg, S. L. (2014) Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol* **15**: R46.

Yang, C., Zhang, W., Liu, R., Li, Q., Li, B., Wang, S., *et al*. (2011) Phylogenetic diversity and metabolic potential of activated sludge microbial communities in full-scale wastewater treatment plants. *Environ Sci Technol* **45**: 7408–7415.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Figure S1. Spearman correlation coefficients (r) for the associations between taxonomic or functional richness and the randomly re-labelled rates of each metabolic process (n = 71). Each metabolic process is the

consumption of a different carbon substrate, and the rate constants (k) for each metabolic process were randomly re-labelled to a different microbial community. Richness measurements include A., observed taxonomic richness; B., Chao1 taxonomic richness; C., ACE taxonomic richness; D., observed functional richness; E., Chao1 functional richness; or F., ACE functional richness. The dashed vertical lines indicate a Spearman correlation coefficient of zero.

Table S1. Measured taxonomic richness (Rt) or functional gene richness (Rf) for each microbial community.

Table S2. Measured rate constants (h – 1) (k) for each metabolic process or environmental condition.

Table S3. Spearman correlation coefficients ($r_{Spearman}$) and P values between taxonomic richness (R_t) or functional gene richness (R_f) and the rate constants (k) for each metabolic process or environmental condition.

Table S4. Shapiro–Wilk tests (P values) for the rate constants (k) for each metabolic process or environmental condition.

Table S5. Dilutions at which each metabolic process was no longer measurable.

Table S6. Lag time (h) before each metabolic process became measurable.

Table S7. Locations of each WWTP.

Table S8. Summary of the GEN III MicroPlate™.