https://doi.org/10.5194/soil-2021-148
Preprint. Discussion started: 13 January 2022

Lower functional redundancy in "narrow" than "broad" functions in global soil metagenomics

Huaihai Chen[a,*], Kayan Ma[a], Yu Huang[a], Jiajiang Lin[b], Christopher W. Schadt[c,d], Hao Chen[a]

[a]State Key Laboratory of Biocontrol, School of Ecology, Sun Yat-sen University, Shenzhen, 518107, China

[b]Fujian Key Laboratory of Pollution Control and Resource Reuse, College of Environmental Science and Engineering, Fujian Normal University, Fuzhou, 350007, China

[c]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, 37831, USA

[d]Department of Microbiology, University of Tennessee, Knoxville, TN, 37996, USA

[*]Corresponding author:
Huaihai Chen, State Key Laboratory of Biocontrol, School of Ecology, Sun Yat-sen University, Guangzhou, 510006, China; Email: chenhh68@mail.sysu.edu.cn

Jiajiang Lin, Fujian Key Laboratory of Pollution Control and Resource Reuse, College of Environmental Science and Engineering, Fujian Normal University, Fuzhou, 350007, China; Email: jjlin@fjnu.edu.cn

https://doi.org/10.5194/soil-2021-148
Preprint. Discussion started: 13 January 2022

Abstract

Understanding the relationship between soil microbial taxonomic compositions and functional profiles is essential for predicting ecosystem functions under various environmental disturbances. However, even though microbial communities are sensitive to disturbance, ecosystem functions remain relatively stable, as soil microbes are likely to be functionally redundant. Microbial functional redundancy may be more associated with "broad" functions carried out by a wide range of microbes, than with "narrow" functions specialized by specific microorganisms. Thus, a comprehensive study to evaluate how microbial taxonomic compositions correlate with "broad" and "narrow" functional profiles is necessary. Here, we evaluated soil metagenomes worldwide to assess whether functional and taxonomic diversities differ significantly between the five "broad" and the five "narrow" functions that we chose. Our results revealed that compared with the five "broad" functions, soil microbes capable of performing the five "narrow" functions were more taxonomically diverse, and thus their functional diversity was more dependent on taxonomic diversity, implying lower levels of functional redundancy in "narrow" functions. Co-occurrence networks indicated that microorganisms conducting "broad" functions were positively related, but microbes specializing "narrow" functions were interacting mostly negatively. Our study provides strong evidence to support our hypothesis that functional redundancy is significantly different between "broad" and "narrow" functions in soil microbes, as the association of functional diversity with taxonomy were greater in the five "narrow" rather than the five "broad" functions.

https://doi.org/10.5194/soil-2021-148
Preprint. Discussion started: 13 January 2022

Keywords Functional redundancy, Soil metagenomics, Functional traits, Taxonomic compositions,

1. Introduction

Microbial communities often exhibit incredible taxonomic diversity, with one gram of soil harboring millions of microbial species (Gans et al., 2005). However, how such diversity governs microbial functional potential and ecosystem processes is largely unknown. Though microbial taxonomic composition is generally sensitive to disturbance and often do not rapidly recover (Allison and Martiny, 2008), it is unclear how changes in microbial community composition would regulate ecosystem functioning. Mechanistic understanding of microbial systems, including microbial taxonomic compositions and functional potential, is essential for predicting ecosystem functioning under various environmental disturbances (Torsvik and Øvreås, 2002;Wellington et al., 2003;McGill et al., 2006).

Though microbial community composition usually shift in response to disturbance, ecosystem functions could remain relatively stable due to functional redundancy (Allison and Martiny, 2008). Microbial functional redundancy is an inevitable emergent property of microbial systems (Louca et al., 2018), as some metabolic functions can be performed by multiple species, which may thus be substitutable in certain ecosystem processes (Rosenfeld, 2002), implying that microbial taxonomy and function can be decoupled (Louca et al., 2016;Louca et al., 2017). Microbial functional redundancy has been mainly observed in “broad” ecosystem processes (Yin et al., 2000;Rousk et al., 2009;Banerjee et al., 2016), but is perhaps less significant in “narrow” functions specialized by certain

https://doi.org/10.5194/soil-2021-148
Preprint. Discussion started: 13 January 2022

microorganisms (Schimel, 1995;Balser et al., 2002). However, some studies simulating microbial diversity reduction and physiological processes challenged the hypothesis of microbial redundancy in soil microbes (Peter et al., 2011;Philippot et al., 2013;Delgado-Baquerizo et al., 2016). Such apparent contradictory results suggest the degree of functional redundancy may depend on the function of interest. Microbes conducting "broad" metabolic functions, such as carbon decomposition, are likely to distribute across most taxa (Crowther et al., 2019) and associate with high level of functional redundancy (Beier et al., 2017;Rivett and Bell, 2018). "Narrow" functions, such as nitrification or methanogenesis, may be restricted to a few phylogenetic clades (Schimel and Gulledge, 1998), and are hypothesized to exhibit less redundancy than "broad" functions (Schimel, 1995;Rocca et al., 2015). Today, multifunctionality (Hector and Bagchi, 2007) has to be accounted for to avoid overestimating functional redundancy (Gamfeldt et al., 2008). By assessing multiple functions, the relationship between microbial diversity and ecosystem function can be better quantified in the soil (Bastida et al., 2016;Delgado-Baquerizo et al., 2016).

Nowadays, metagenomics have been increasingly used as a promising comparative tool (Tringe et al., 2005) to study the relationship between functional and taxonomic diversities (Fierer et al., 2012a;Fierer et al., 2012b;Fierer et al., 2013;Pan et al., 2014;Leff et al., 2015;Souza et al., 2015). The growing wealth of soil metagenome data thus poised well to aid in the generalization of global patterns of microbial attributes and standardizing frameworks for consistent representation of microbial community (Chen et al., 2021a;Chen et al., 2021b). However, a synthetic metagenomic analysis to assess how

https://doi.org/10.5194/soil-2021-148
Preprint. Discussion started: 13 January 2022

general microbial taxonomic and functional diversities differ between "broad" and "narrow" functions across the globe is still lacking.

Here, we constructed soil metagenomic datasets of taxonomic and functional diversities of five "broad" and five "narrow" functions across seventeen climate zones. We typically chose SEED Subsystems database (Overbeek et al., 2013) that has diverse classification at level 1, allowing us to conduct comparison between "broad" versus "narrow" functions. We selected five "narrow" functions, namely N (Nitrogen Metabolism), P (Phosphorous Metabolism), K (Potassium Metabolism), S (Sulfur Metabolism), and Fe (Iron Acquisition and Metabolism). These are typical functional categories of specific nutrient cycling in Subsystems Level 1 and are only performed by certain groups of soil microbes (Schimel, 1995). The five "broad" functions selected were AAD (Amino Acids and Derivatives), CHO (Carbohydrates), CBS (Clustering-based Subsystems), CVPGP (Cofactors, Vitamins, Prosthetic Groups, Pigments), and Protein (Protein Metabolism), which are the most abundant functional categories in Subsystems level 1, and represent broad-scale functions acquired by a relatively larger group of diverse soil microbes (Balser et al., 2002). We further contructed the pairswise similarity of function and taxonomy based on the relative abundance of functional and taxonomic compositions, respectively, for the five "broad" and the five "narrow" functions. We hypothesized that the taxonomic similarity of soil microbes would be more linearly correlated to the functional similarity for the five "narrow" functions in comparison to the five "broad" functions. Therefore, using these global soil metagenomes, our objective was to test whether the taxonomic compositions of soil microbes that conduct the five "narrow" functions are more dependent on the functional compositions, leading to a

https://doi.org/10.5194/soil-2021-148
Preprint. Discussion started: 13 January 2022

lower level of functional redundancy in the "narrow" functions than the "broad" functions.

2. Materials and Methods

2.1. Data collection

To ensure that the quality and completeness of the metagenomes analyzed were of standard, we carefully selected soil metagenomes in MG-RAST server that have been published in peer-reviewed journals. We searched peer-reviewed publications from 2012 to 2018 from the Web of Science database using search terms such as "soil metagenome", "shotgun sequencing", and "MG-RAST" to source the metagenomic data used in this study to their publications. We included soil metagenomes publicly available in the MG-RAST database that are generated using shotgun sequencing without amplification or that were directly deposited by peer-reviewed studies into the MG-RAST database. We then extracted data matrix of taxonomic and functional compositions of soil metagenomes from MG-RAST public server (https://www.mg-rast.org/) based on the Study ID and/or MG-RAST ID reported in the publications. Details of each soil metagenome extracted from publications and MG-RAST database was given in Supplementary Table S1.

The functional database that we used in this study, SEED Subsystems, is a categorization system which organizes gene functional categories into a hierarchy with three levels of resolution (Level 3, 2 and 1) (Overbeek et al., 2013). To download the taxonomic compositions to soil microbes to conduct "broad" and "narrow" functions, for each soil metagenome, in the 'Analysis' function of the MG-RAST server

https://doi.org/10.5194/soil-2021-148
Preprint. Discussion started: 13 January 2022

(https://www.mg-rast.org/mgmain.html?mgpage=analysis), we loaded both SEED Subsystems (Level 3, 2 and 1) as functional profiles and RefSeq (Tatusova et al., 2013) databases (genus, family, order, class, and phylum levels) as taxonomic compositions (Chen et al., 2021b). The detailed protocols of MG-RAST server were followed to analyze the metagenomic functions (Meyer et al., 2008;Wilke et al., 2017). To obtain the taxonomic compositions of soil microbes that conduct the selected "broad" and "narrow" functions, we chose 'RefSeq' as source and 'genus' as level, and in 'function filter' we added the functional categories in Subsystems Level 1 that we are interested in, including five "broad" functions of AAD (Amino Acids and Derivatives), CHO (Carbohydrates), CBS (Clustering-based Subsystems), CVPGP (Cofactors, Vitamins, Prosthetic Groups, Pigments), and Protein (Protein Metabolism), of which the relative abundance was 5-13%. The functions of AAD, CHO, CBS, CVPGP, and Protein were the most abundant functional categories in Subsystems Level 1, which were used to represent broad-scale functions acquired by a large group of diverse soil microbes. Correspondingly, five "narrow" functions were chosen, namely N (Nitrogen Metabolism), P (Phosphorous Metabolism), K (Potassium Metabolism), S (Sulfur Metabolism), and Fe (Iron Acquisition and Metabolism), of which the relative abundance was 0.8-1.4%, as these are typical functional categories of specific nutrient cycling in Subsystems Level 1 and are only performed by certain groups of soil microbes. Total hits of taxonomic compositions of soil microbes conductingeach function at Subsystems Level 1 were calculated as the sums of hits in different taxonomic categories at RefSeq genus level.

The comparative metagenomic analyses were performed using default settings (maximum e-value cutoff = $1e^{-5}$, minimum identity cutoff = 60%, and minimum

https://doi.org/10.5194/soil-2021-148
Preprint. Discussion started: 13 January 2022

alignment length = 50) (Meyer et al., 2008). We then merged the taxonomic compositions of data matrix of each functions extracted from different studies together to generate new datasets of microbial taxonomic compositions annotated by the RefSeq database. The reason why we chose the Subsystems database for functional grouping rather than KEGG Orthology (KO) (Kanehisa et al., 2015), Clusters of Orthologous Groups (COG) (Galperin et al., 2014), and Non-supervised Orthologous Groups (NOG) (Huerta-Cepas et al., 2015) databases was that Subsystems had more diverse classification at Level 1, allowing us to conduct direct comparison between "broad" versus "narrow" functions. We chose RefSeq database rather than the traditional ribosomal RNA databases, such as RDP (Ribosomal Database Project) (Cole et al., 2008), Greengenes (DeSantis et al., 2006), or Silva LSU/SSU (Pruesse et al., 2007) databases, because taxonomic hits in the RefSeq database were over 1000-fold higher than the rRNA databases, rendering the resolution comparable to functional hits for comparison between "broad" and "narrow" functions. To increase the coverage of our datasets, soil metagenomes with/without assembly were both included.

The geographic coordinates of latitudes (LAT) and longitudes (LONG) of each soil metagenome were directly obtained from publications. Based on LAT and LONG, climate data of mean annual temperature (MAT) and precipitation (MAP) of study sites for each soil metagenome were extracted from the WorldClim dataset (Fick and Hijmans, 2017) using the R package 'raster' (Hijmans et al., 2015). To examine how microbial taxonomic diversities of "broad" and "narrow" functions differ globally, soil metagenomic data was classified into seventeen climate zones based on the main

https://doi.org/10.5194/soil-2021-148
Preprint. Discussion started: 13 January 2022

classification of Koeppen-Geiger Climatic Zones (Kottek et al., 2006) using the R package 'kgc' (Bryant et al., 2017).

2.2. Statistical Analyses

To minimize bias caused by different sequencing depths and read lengths among studies, we standardize the hits of each taxonomic (or functional) category in each data to relative abundance by dividing them by the total number of hits. To calculate the pairwise similarity of taxonomy based on the relative taxonomic abundance at genus level of microbes conducting the five "broad" and five "narrow" functions, we calculated Bray-Curtis similarity following log transformation of the compositional taxonomic data by constructing pairwise Bray-Curtis similarity matrix between each pair of two samples for each functional categories at Subsystems database at Level 1, which were further transformed to lists of pairwise Bray-Curtis similarities ordered by sample pair names in PRIMER 7 (Plymouth Routines in Multivariate Ecological Research Statistical Software, v7.0.13, PRIMER-E Ltd, UK) (Clarke and Gorley, 2015). To calculate the pairwise similarity of function, based on the functional abundance at function gene level within each of the five "broad" and five "narrow" functions, we calculated Bray-Curtis similarity following log transformation of the compositional functional data by constructing pairwise Bray-Curtis similarity matrix between each pair of two samples for each functional categories at Subsystems database at Level 1, which were further transformed to lists of pairwise Bray-Curtis similarities ordered by sample pair names in PRIMER 7. To examine the relationship between functional and taxonomic diversities, Pearson's correlations were constructed between the transformed lists of pairwise Bray-

https://doi.org/10.5194/soil-2021-148
Preprint. Discussion started: 13 January 2022

Curtis similarity of soil metagenomes annotated using Subsystems database at Level 3 (Function) and the RefSeq database at genus level (Taxonomy). The approaches for processing the relative abundance of compositional data follow the requirements (Gloor et al., 2017). To analyze the taxonomic composition structures of soil metagenomes annotated using the RefSeq database at genus level (Taxonomy) of the five "broad" and five "narrow" functions, PCoA (principal coordinates analysis) and PERMANOVA (Permutational multivariate analysis of variance) were conducted using the pairwise Bray-Curtis similarity matrix in PRIMER 7.

To compare microbial taxonomic compositions among the five "broad" and the five "narrow" functions, one-factor PEMANOVA was conducted using the main test and pair-wise test in PRIMER 7 with *P* values and Sq. root reported. Pearson's correlations were constructed to assess the relationships between functional and taxonomic diversities in the "broad" and "narrow" functions with adjusted P-Square reported. A RELATE analysis was also performed to evaluate the relatedness among "broad" and "narrow" functions by calculating a Spearman's Rho correlation coefficient in PRIMER 7. To examine the relative abundance of dominant microbial at phylum and class level (mean > 1%) among the five "broad" and five "narrow" functions, heatmaps were constructed using HeatMapper (Babicki et al., 2016). One-way analysis of variance (ANOVA) with *P* values adjusted by Bonferroni-correction for multiple comparisons was conducted using SPSS 22.0 software (Chicago, IL, USA) to evaluate the differences in the relative abundance of dominant taxonomic compositions (mean > 1%) among climate zones after the normality of residues and homogeneity of variance were checked using Shapiro-Wilk and Levene test, respectively. The significance level was set at α=0.05 unless otherwise

https://doi.org/10.5194/soil-2021-148
Preprint. Discussion started: 13 January 2022

stated. To calculate the statistical difference between the relative abundance of dominant microbial taxonomic groups (mean > 1%) in the "broad" and "narrow" functions, LEfSe (linear discriminant analysis effect size) method was used (http://huttenhower.sph.harvard.edu/lefse/) (Segata et al., 2011). Venn's diagrams were constructed to visualize the amount of dominant microbial taxonomic groups at genus levels or network nodes shared between the five "broad" and the five "narrow" functions using InteractiVenn (Heberle et al., 2015).

To find out potential interactions of microbial taxonomic compositions between "broad" and "narrow" functions across the globe, co-occurrence network analysis was performed using the Molecular Ecological Network Analyses Pipeline (http://ieg4.rccc.ou.edu/MENA/) (Zhou et al., 2011;Deng et al., 2012). To make the minimum observed value close to but no less than 1 as required by the pipeline, the data of relative abundance were multiplied by 10^6, which would not change the correlation coefficients. The data matrix of transformed data matrix was uploaded to construct the network with default settings, including (1) keeping only the species present in more than a half of all samples; (2) only filling with 0.01 in blanks with paired valid values; (3) taking logarithm with recommended similarity matrix of Pearson's correlation coefficient; and (4) calculation ordered to decrease the cutoff from top using regress poisson distribution only. A default cutoff value (similarity threshold, S_t) for the similarity matrix was used to assign a link between the pair of species. After that, the global network properties, the individual nodes' centrality, and the module separation and modularity were analyzed based on default settings using greedy modularity optimization. Network files were exported and visualized using Cytoscape software

https://doi.org/10.5194/soil-2021-148
Preprint. Discussion started: 13 January 2022

(Shannon et al., 2003). The scatter plots of within-module connectivity (zi) and among-module connectivity (Pi) were constructed to show the network node distribution of module-based topological roles of taxonomic compositions for the "broad" and "narrow" functions. The threshold values of Zi and Pi for categorizing were 2.5 and 0.62 respectively.

3. Results and Discussion

3.1. Microbial taxonomic compositions

This study included 845 soil metagenomes across seventeen climate zones around the world extracted from 56 MG-RAST studies published in 51 peer-reviewed papers. They resulted in 356090 pairwise comparisons of Bray-curtis similarity in functional (Subsystems L3) and taxonomic (RefSeq genus) diversities for the five "broad" and five "narrow" functions, which were analyzed to find out whether the correlations of function and taxonomy were greater in the five "narrow" functions. Overall, for the five "narrow" functions, the positive correlations of the pairwise similarity of taxonomy and function between either two samples (r^2 = 0.36-0.49) were greater than those for the five "broad" functions (r^2 = 0.23-0.29) (Fig. 1). This suggests that rare phylotypes could be more associated with narrow ecosystem processes than broad-scale functions, supporting the notion that the abundance of particular specialists could influence narrow functional measures (Peter et al., 2011;Rivett and Bell, 2018), leading to a lower degree of functional redundancy associated with "narrow" functions, such as the nutrient cycling examined in this study.

https://doi.org/10.5194/soil-2021-148
Preprint. Discussion started: 13 January 2022

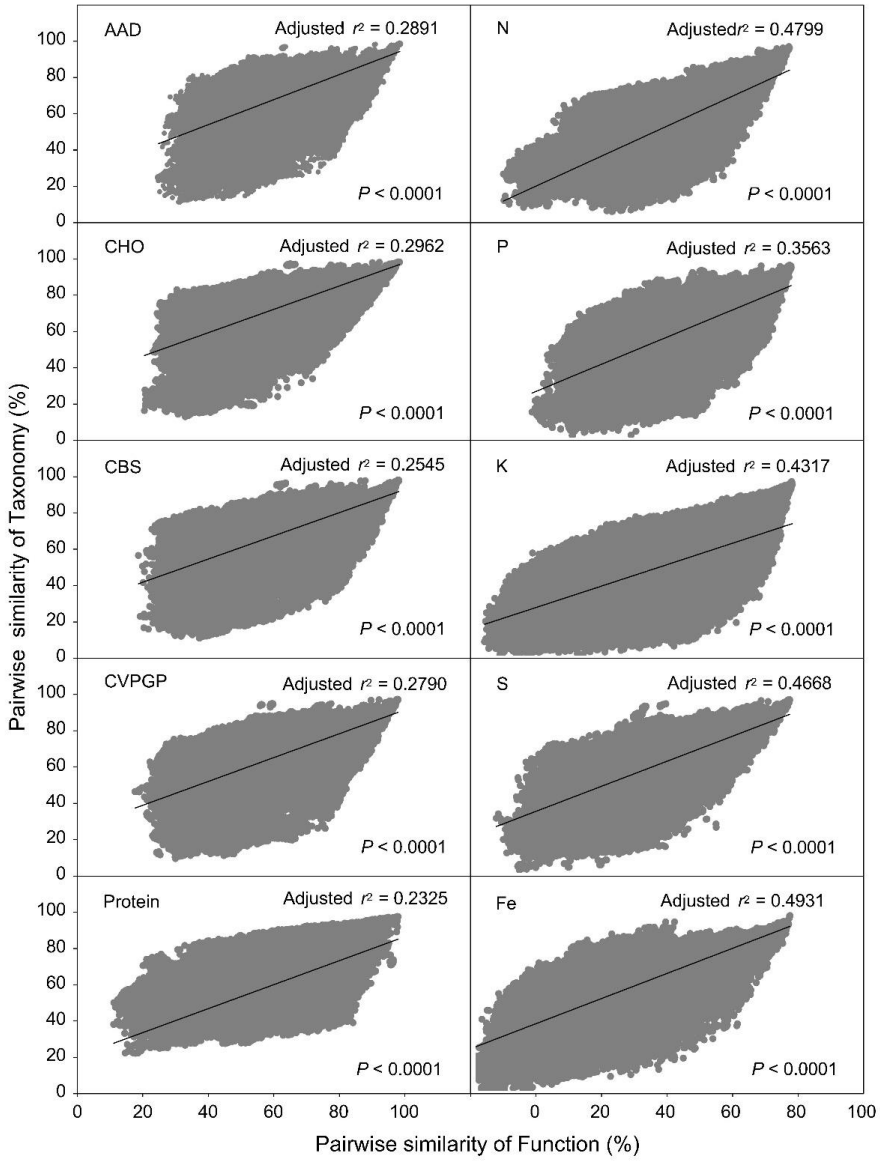


Fig. 1. Relations between functional and taxonomic beta-diversities for "broad" and "narrow" functions. Pearson's correlations between pairwise Bray-curtis similarity of microbial taxonomic and functional compositions for "broad" and "narrow" functions annotated using Subsystems at function level (Function) and RefSeq at genus level (Taxonomy). Correlation adjursted *r*-squared and *P* values are given. "Broad" functions include AAD (Amino Acids and Derivatives), CHO (Carbohydrates), CBS (Clustering-

https://doi.org/10.5194/soil-2021-148
Preprint. Discussion started: 13 January 2022

based Subsystems), CVPGP (Cofactors, Vitamins, Prosthetic Groups, Pigments), and Protein (Protein Metabolism). "Narrow" functions include N (Nitrogen Metabolism), P (Phosphorous Metabolism), K (Potassium Metabolism), S (Sulfur Metabolism), and Fe (Iron Acquisition and Metabolism).

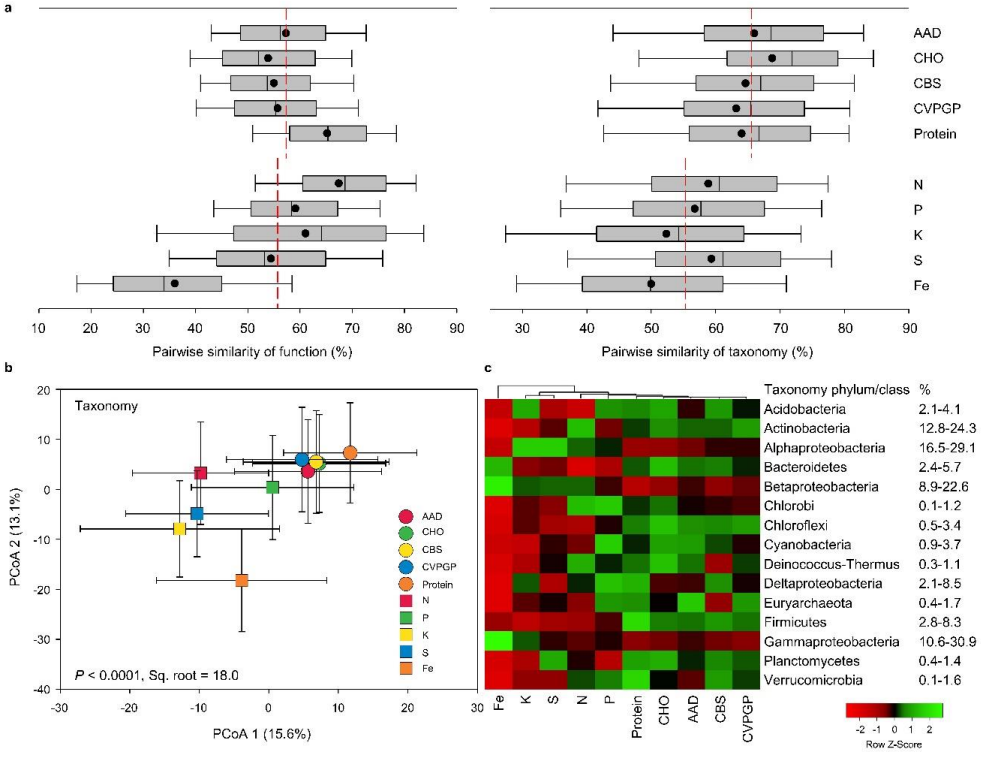
Several soil metagenomic studies have reported a linear relationship between functional and taxonomic diversities (Fierer et al., 2012b;Fierer et al., 2013;Leff et al., 2015), indicating a somewhat dependency of microbial functional profiles on taxonomic compositions. This dependency, however, does not necessarily imply an absence of microbial functional redundancy. In fact, those studies all showed lower variation of beta-diversity of metagenomic functions than taxonomy (Fierer et al., 2012b;Fierer et al., 2013;Pan et al., 2014;Souza et al., 2015) or higher similarity in composition of functional profiles than taxonomic composition (Leff et al., 2015). Those findings support that microbial functions are relatively more stable than taxonomy responding to ecological and environmental perturbations. In this study, the five "broad" and the five "narrow" functions had relative abundance of 5-13% and 0.8-1.4%, respectively. Thus, the five "broad" functions are more abundant than the five "narrow" functions. In addition, the numbers of genes within the categories of the five "broad" functions were also greater than those of the "narrow" functions. As the diversities of the microbes conducting the five "broad" functions were also greater than those conducting the "narrow" functions, we calculated the relationship between the diversities of taxonomy and of function, and compared these relationships between the five "broad" and the five "narrow" functions. Our study further evidenced a lower extent of functional redundancy in the five "narrow" functions compared to the five "broad" functions despite the linear correlations found in our study.

https://doi.org/10.5194/soil-2021-148
Preprint. Discussion started: 13 January 2022

The boxplots were constructed based on the pairwise similarity of function and taxonomy to compare similarity ranges of these two compositions related to the five "broad" functions versus the five "narrow" functions. For the functional compositions at specific function gene levels, the average similarity of the five "broad" functional diversity (58%) was comparable to that of the five "narrow" functions (56%) (Fig. 2a). However, the pairwise similarity of the five "narrow" functions had larger variation, in which Fe function had the lowest similarity of 36% and N function had the highest similarity of 69%. On the contrary, the taxonomic similarity of the five "broad" functions were consistently greater (63-69%) than those of the five "narrow" functions (50-59%). The PERMANOVA pairwise test was conducted to find out the difference between taxonomic similarity of microbes conducting the five "broad" and the five "narrow" functions based on the relative abundance. Our results indicated that the microbial taxonomic compositions of the five "broad" functions were more phylogenetically different from those of the five "narrow" functions (13-22%) than from each other (8-13%) (Supplementary Table 2). The RELATE test was also conducted to evaluate the relationship of the taxonomic compositions of microbes conducting the five "broad" and the five "narrow" functions. Our results confirmed that the microbial taxonomic compositions of the five "broad" functions were more correlated with each other (0.97-0.99) than those of the five "narrow" functions (0.77-0.94) (Supplementary Table 3). When the microbial taxonomic compositions of the ten functional categories were combined in PCoA analysis, the resulting scatter plot showed that the five "broad" functions were grouped closely together and separated from the five "narrow" functions (Fig. 2b). Grouping of the ten functions generally explain up to 18.0% of the community

https://doi.org/10.5194/soil-2021-148
Preprint. Discussion started: 13 January 2022

328 difference, in which the five "narrow" functions were more distinct from each other.
329 These evidences together suggest that the taxonomic composition of soil microbes
330 conducting the five "broad" functions were more conserved in taxonomy than those
331 conducting the five "narrow" functions. However, it should be noted that the current
332 analysis had some limitations as the metagenomics datasets consisted of sequencing data
333 that are phylogenetically classified and assigned based on certain the taxonomic and
334 functional databases. Thus, our results may to some extent depend on the databases
335 chosen, of which the classification and assignment may not contain potential bias. Future
336 studies should continue to test this hypothesis using regional samples and individual
337 datasets.



338

https://doi.org/10.5194/soil-2021-148
Preprint. Discussion started: 13 January 2022

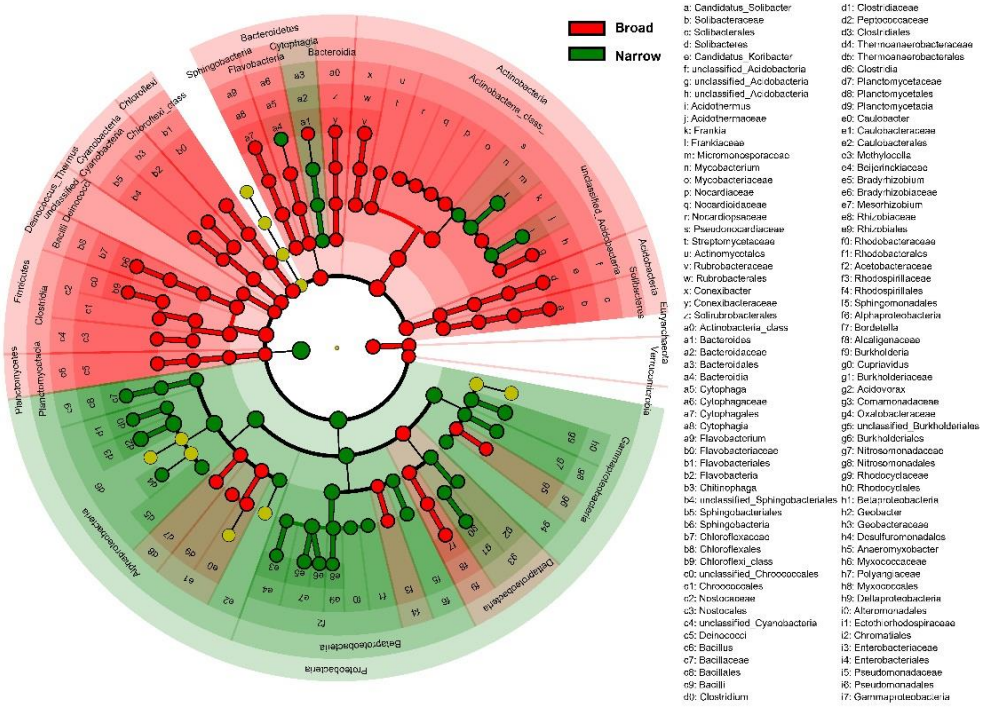
Fig. 2. Functional and taxonomic diversities for "broad" versus "narrow" functions. **a**, Box plots and mean values of pairwise Bray-curtis similarity of microbial functional and taxonomic diversities for "broad" versus "narrow" functions. **b**, PCoA (Principal coordinates analysis) showing beta-diveristy of microbial taxonomic diversity for "broad" and "narrow" functions annotated using RefSeq at genus level (Taxonomy). The error bars represent the standard deviation of data ranges. Variations (by percentage) explained by the two principal coordinate dimensions aare given in parentheses. *P* values and sq. root of PERMANOVA are also given. **c**, Heatmaps showing relative abundance of dominant microbial taxonomic composition (mean > 0.5%) for "broad" and "narrow" functions annotated using RefSeq at phylum/class levels (Taxonomy). "Broad" functions include AAD (Amino Acids and Derivatives), CHO (Carbohydrates), CBS (Clustering-based Subsystems), CVPGP (Cofactors, Vitamins, Prosthetic Groups, Pigments), and Protein (Protein Metabolism); "Narrow" functions include N (Nitrogen Metabolism), P (Phosphorous Metabolism), K (Potassium Metabolism), S (Sulfur Metabolism), and Fe (Iron Acquisition and Metabolism).

To investigate how microbial taxonomic diversities differ globally, the taxonomic compositions of soil microbes conducting the five "broad" and the five "narrow" functions were analyzed among the seventeen climate zones based on the PCoA analysis. Across climate zones, microbial taxonomic compositions of the five "narrow" functions (sq. root = 15.2-18.8) were more distinct than the five "broad" functions (sq. root = 13.4-15.1) based on the PERMANOVA anaysis (Supplementary Fig. 1). This suggests that microorganisms relating to "broad" functions were similar to each other in taxonomy, because "broad" functions are more broadly distributed across most taxa, but soil microbes performing "narrow" functions were more phylogenetically diverse due to the specialty of "narrow" functions. Thus, though microbial metabolic functions can be strongly coupled to elemental cycles and certain environmental factors, the decoupling of

https://doi.org/10.5194/soil-2021-148
Preprint. Discussion started: 13 January 2022

366 microbial taxonomic and functional profiles is still inevitable when a low-dimensional
367 functional space is projected to a high-dimensional taxonomic space (Louca et al., 2018),
368 especially for “broad” functions.

369 The taxonomic compositions of microbes conducting the five “broad” functions were
370 more abundant in most major phyla, such as Acidobacteria, Actinobacteria,
371 Bacteroidetes, and Firmicutes, while the relative abundance of the taxonomic
372 composition of microbes conducting the five “narrow” functions were greater in
373 Proteobacteria, especially Alphaproteobacteria and Betaproteobacteria (Fig. 2c). Other
374 studies also found that some bacteria conducting N cycling, such as ammonia-oxidizers
375 and rhizobia for N fixation, mainly belong to Alphaproteobacteria or Betaproteobacteria
376 (Stephen et al., 1996;Moulin et al., 2001).



377

https://doi.org/10.5194/soil-2021-148
Preprint. Discussion started: 13 January 2022

Fig. 3. Taxonomic compositions shared among "broad" and "narrow" functions. Venn's diagrams showing dominant microbial taxonomic groups (mean > 0.1%) annotated using RefSeq at genus levels (Taxonomy) shared among "broad" and "narrow" functions.

LEfSe analysis was used to show the dominant microbial groups at the taxonomic levels of domain, phylum, class, order, family, and genus that were statistically different between the five "broad" and the five "narrow" functions based on the their relative abundances. In particular, among the Proteobacteria conducting the five "narrow" functions, *Bacillaceae* from Bacilli, *Clostridium*, *Peptococcaceae*, and *Thermoanaerobacteraceae* from Clostridia, *Methylocella*, *Bradyrhizobium*, *Bradyrhizobiaceae*, and *Rhizobiaceae* from Rhodospirillaceae, and *Cupriavidus* from Comamonadaceae had higher relative abundance than the others (Fig. 3). The Venn's diagrams indicated that the taxonomic compositions of soil microbes performing the "broad" functions shared 68% dominant genera among the five functional categories, while the proportion was reduced to only 41% for the five "narrow" functions (Fig. 4). However, it should be stated that all the analyses performed in our study were based on relative abundance data that is compositional, so it is difficult to directly compare taxonomic diversities among samples and/or datasets. Despite the differences in the identification protocol and quantification of soil metagenomes, we deem the effects of these differences to be trivial for our analyses as we intended to understand the general patterns of microbial taxonomic and functional linkages, rather than simply compare soil community structures across samples. By uncovering universal patterns of these relationships within the microbial community, we can then further establish a potential

https://doi.org/10.5194/soil-2021-148
Preprint. Discussion started: 13 January 2022

linkage framework to account for the microbial contributions to major biogeochemical cycles.

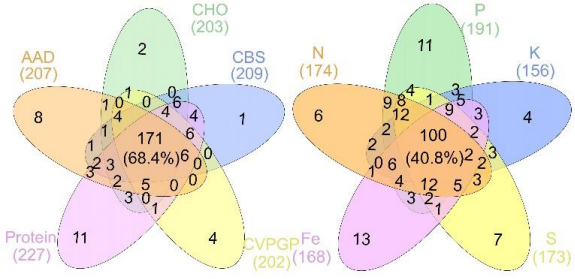


Fig. 4. Difference of taxonomic compositions between "broad" and "narrow functions". LEfSe (linear discriminant analysis effect size) results showing the significant differences in the relative abundance of dominant microbial taxonomic groups (mean > 0.5%) between "broad" (red) versus "narrow" (green) functions annotated using RefSeq (Taxonomy). From the center outward, each circle represents the level of domain, phylum, class, order, family, and genus, respectively. The taxonomic groups with significant differences are labeled by colors.

Because of functional redundancy of soil microbes, understanding what types of functions that have more significant association with microbial taxonomy can be critical for accurate prediction of microbial metabolic activity and flexibility across space and time. As microbial taxonomic composition and diversity plays critical role in maintaining ecosystem function (Allison and Martiny, 2008), our results suggest that taxonomic information alone provides limited utility in predicting basic metabolic capabilities, but may be capable of forecasting biogeochemical transformations or changes in the rate of biogeochemical process at ecosystem level (Hall et al., 2018). Investigating functional redundancy with respect to functions associated with elemental cycles provides useful information for guiding the development of explicit microbial biogeochemical prediction,

https://doi.org/10.5194/soil-2021-148
Preprint. Discussion started: 13 January 2022

and further delving into major pathways of C and N cycles will be a fruitful approach for scrutinizing microbes' functional potentials.

Table 1. Summary of key properties of co-occurrence networks for the five "broad" and the five "narrow" functions.

Network Indexes	Total nodes	Total links (positive%)	Average connectivity	Average clustering coefficient	Average geodesic distance	Modularity (modules numbers)
AAD	225	1472 (100%)	13.084	0.663	2.873	0.695 (11)
CHO	207	1155 (99%)	11.159	0.615	3.805	0.672 (10)
CBS	246	1622 (99%)	13.187	0.663	2.859	0.671 (11)
CVPGP	201	1293 (99%)	12.866	0.65	3.303	0.697 (9)
Protein	285	1651 (99%)	11.586	0.638	2.992	0.749 (14)
N	101	519 (12%)	10.277	0.349	1.903	0.184 (5)
P	160	449 (4%)	5.612	0.299	3.298	0.615 (10)
K	143	364 (67%)	5.091	0.08	2.676	0.429 (6)
S	132	264 (15%)	4	0.09	2.563	0.486 (12)
Fe	95	215 (11%)	4.526	0.071	2.601	0.435 (6)

3.2. Microbial taxonomic co-occurrence networks

Co-occurrence networks of taxonomic compositions were generated to identify potential interaction patterns of microbial groups that conduct the five "broad" and the five "narrow" functions across the globe. Network graphs with submodule structures indicated distinct topology of taxonomic networks between the "broad" and "narrow" functions (Table 1, Supplementary Fig. 2 and Supplementary Fig. 3). Compared to the "narrow" functions, the "broad" functions harbored larger and more complex networks with more nodes (201-285 vs. 95-160) and links (1293-1651 vs. 215-519), with higher average connectivity (11.2-13.2 vs. 4.0-10.3) and average clustering coefficient (0.64-0.66 vs. 0.07-0.35). The "broad" function network had 99-100% positive links, while the "narrow" function had 33-96% negative links.

https://doi.org/10.5194/soil-2021-148
Preprint. Discussion started: 13 January 2022

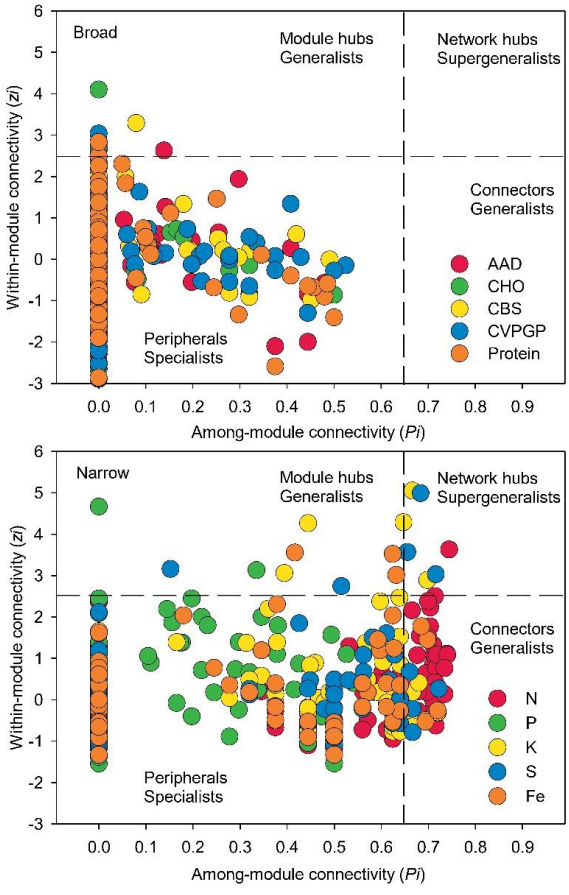
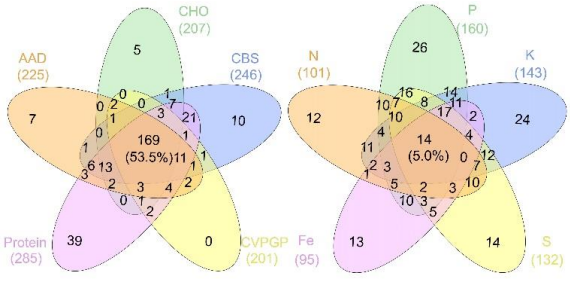


Fig. 5. Network information of taxonomic compositions for "broad" and "narrow" functions. Node distribution of module-based topological roles of taxonomic compositions for "broad" and "narrow" functions determined by the scatter plot of within-module connectivity (*zi*) and among-module connectivity (*Pi*). The threshold values of Zi and Pi for categorizing were 2.5 and 0.62 respectively.

In addition, network modularity was greater in the "broad" functions, indicating that significant correlations between taxonomic compositions of microbes that conduct the five "broad" functions are mainly within similar taxonomic groups. No node could be classfied as connectors in the five "broad" function networks (Fig. 5), reaffirming that the

https://doi.org/10.5194/soil-2021-148
Preprint. Discussion started: 13 January 2022

450 "broad" function networks had links mainly within modules of similar species. In the co-
451 occurrence network of taxonomic composition of the "narrow" functions, 13% of the
452 nodes were identified as connectors linking several modules (high *Pi*) connectors, while
453 3% were identified as module hubs that connected other nodes within their own modules
454 (high *Zi*), indicated by the *Zi-Pi* plot (Olesen et al., 2007;Deng et al., 2012). Thus,
455 significantly less nodes were identified as module hubs in the co-occurrence network of
456 the taxonomic composition of the "broad" functions, indicting less correlations found
457 among different modules. This is expected given that module was comprised of genera
458 that were mainly from the same phylogenetic groups. This difference was consistent with
459 the Venn's diagrams showing significantly more nodes (54%) shared among the five
460 functional categories representing the "broad" functions, while only 5% of the nodes
461 were overlaid among the five "narrow" function networks (Fig. 6). Environmental
462 conditions likely determine the microbial taxonomic composition, and microbial
463 phylotypes sharing similar habitat preferences tend to co-occur (Delgado-Baquerizo et
464 al., 2018;Ram ́ırez-Flandes et al., 2019). We emphasize that this analysis is a combination
465 of snapshots of microbial communities compared across space, thus environmental
466 conditions (at the same geographic location) may vary, and the levels of functional
467 redundancy may change depending on the mechanisms selecting specific functions and
468 the phylogenetic distribution of those functions (Louca et al., 2018).



469

https://doi.org/10.5194/soil-2021-148
Preprint. Discussion started: 13 January 2022

Fig. 6. Taxonomic network nodes shared among "broad" and "narrow" functions. Venn's diagrams showing the microbial taxonomic network nodes shared among "broad" and "narrow" functions.

3.3. Conclusion

By analyzing and generalizing microbial taxonomic and functional profiles, we provide strong evidence that the degree of soil microbial functional redundancy differ significantly between "broad" and "narrow" functions across the global. The level of functional redundancy varies depending on the functions of interest. Here, by contrasting the five "broad" metabolic functions and the five "narrow" functions that are important for elemental cycles, we found lower levels of functional redundancy associated with the five "narrow" functions of biogeochemical cycling, despite the fact that even for the five "narrow" functions, there is still a high level of functional redundancy in the soil communities. Although there is a caveat concerning direct comparison of metagenomic data, the present study demonstrated the use of comparative metagenome and co-occurrence network analysis in generalizing patterns of microbial characteristics regulating biogeochemical cycling of major elements. With the increasing advancement of sequencing techniques and data coverage, future sequencing efforts will likely increase our confidence in comparative metagenomes and provide time-series information to further identify to what extent microbial functional redundancy regulates dynamic ecological fluxes across space and time.

Author Contributions

https://doi.org/10.5194/soil-2021-148
Preprint. Discussion started: 13 January 2022

493 Huaihai Chen conceived the study, performed the data analysis, interpreted the results,
494 and drafted the manuscript. CL, YY, CWS, and Hao Chen secured the research funding.
495 KM, YH, YY, and Hao Chen critically assessed and interpreted the findings. All authors
496 discussed results, commented on, edited, revised, and approved the manuscript.
497
498 **Funding**
499 This study is supported by the National Natural Science Foundation of China
500 (31872691), Basic and Applied Basic Research Foundation of Guangdong Province.
501
502 **Acknowledgements**
503 We thank the authors of the publications included in our study, without which this global
504 metagenomic analysis would not be possible.
505
506 **Data Availability Statement**
507 The data that support the findings of this study are available from the corresponding
508 author upon request. All metagenomic data used in this study are publicly assessable in
509 the MG-RAST server with study and MG-RAST ID reported in supplementary files.
510
511 **Competing Interests**
512 The authors declare no competing interests.
513
514 **References**

https://doi.org/10.5194/soil-2021-148
Preprint. Discussion started: 13 January 2022

Allison, S.D., and Martiny, J.B. (2008). Resistance, resilience, and redundancy in microbial communities. *Proceedings of the National Academy of Sciences* 105**,** 11512-11519.

Babicki, S., Arndt, D., Marcu, A., Liang, Y., Grant, J.R., Maciejewski, A., and Wishart, D.S. (2016). Heatmapper: web-enabled heat mapping for all. *Nucleic Acids Research* 44**,** W147-W153.

Balser, T.C., Kinzig, A.P., and Firestone, M.K. (2002). Linking soil microbial communities and ecosystem functioning. *The functional consequences of biodiversity: empirical progress and theoretical extensions***,** 265-293.

Banerjee, S., Kirkby, C.A., Schmutter, D., Bissett, A., Kirkegaard, J.A., and Richardson, A.E. (2016). Network analysis reveals functional redundancy and keystone taxa amongst bacterial and fungal communities during organic matter decomposition in an arable soil. *Soil Biology and Biochemistry* 97**,** 188-198.

Bastida, F., Torres, I.F., Moreno, J.L., Baldrian, P., Ondoño, S., Ruiz-Navarro, A., Hernández, T., Richnow, H.H., Starke, R., and García, C. (2016). The active microbial diversity drives ecosystem multifunctionality and is physiologically related to carbon availability in Mediterranean semi-arid soils. *Molecular ecology* 25**,** 4660-4673.

Beier, S., Shen, D., Schott, T., and Jürgens, K. (2017). Metatranscriptomic data reveal the effect of different community properties on multifunctional redundancy. *Molecular ecology* 26**,** 6813-6826.

Bryant, C., Wheeler, N., Rubel, F., and French, R. (2017). "kgc: Koeppen–Geiger climatic zones. R package version 1.0. 0.2".).

https://doi.org/10.5194/soil-2021-148
Preprint. Discussion started: 13 January 2022

538 Chen, H., Ma, K., Huang, Y., Yang, Y., Ma, Z., and Chu, C. (2021a). Salinity drives
539 functional and taxonomic diversities in global water metagenomes. *Frontiers in*
540 *Microbiology* 12.
541 Chen, H., Ma, K., Huang, Y., Yao, Z., and Chu, C. (2021b). Stable soil microbial
542 functional structure responding to biodiversity loss based on metagenomic
543 evidences. *Frontiers in Microbiology* 12.
544 Clarke, K., and Gorley, R. (2015). Getting started with PRIMER v7. *PRIMER-E:*
545 *Plymouth, Plymouth Marine Laboratory*.
546 Cole, J.R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R.J., Kulam-Syed-
547 Mohideen, A., Mcgarrell, D.M., Marsh, T., and Garrity, G.M. (2008). The
548 Ribosomal Database Project: improved alignments and new tools for rRNA
549 analysis. *Nucleic acids research* 37**,** D141-D145.
550 Crowther, T.W., Van Den Hoogen, J., Wan, J., Mayes, M.A., Keiser, A., Mo, L., Averill,
551 C., and Maynard, D.S. (2019). The global soil community and its influence on
552 biogeochemistry. *Science* 365**,** eaav0550.
553 Delgado-Baquerizo, M., Maestre, F.T., Reich, P.B., Jeffries, T.C., Gaitan, J.J., Encinar,
554 D., Berdugo, M., Campbell, C.D., and Singh, B.K. (2016). Microbial diversity
555 drives multifunctionality in terrestrial ecosystems. *Nature communications* 7**,**
556 10541.
557 Delgado-Baquerizo, M., Oliverio, A.M., Brewer, T.E., Benavent-Gonz áez, A., Eldridge,
558 D.J., Bardgett, R.D., Maestre, F.T., Singh, B.K., and Fierer, N. (2018). A global
559 atlas of the dominant bacteria found in soil. *Science* 359**,** 320-325.

https://doi.org/10.5194/soil-2021-148
Preprint. Discussion started: 13 January 2022

Delgado-Baquerizo, M., Giaramida, L., Reich, P.B., Khachane, A.N., Hamonts, K., Edwards, C., Lawton, L.A., and Singh, B.K. (2016). Lack of functional redundancy in the relationship between microbial diversity and ecosystem functioning. *Journal of Ecology* 104**,** 936-946.

Deng, Y., Jiang, Y.-H., Yang, Y., He, Z., Luo, F., and Zhou, J. (2012). Molecular ecological network analyses. *BMC bioinformatics* 13**,** 113.

Desantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P., and Andersen, G.L. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* 72**,** 5069-5072.

Fick, S.E., and Hijmans, R.J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International journal of climatology* 37**,** 4302-4315.

Fierer, N., Ladau, J., Clemente, J.C., Leff, J.W., Owens, S.M., Pollard, K.S., Knight, R., Gilbert, J.A., and Mcculley, R.L. (2013). Reconstructing the microbial diversity and function of pre-agricultural tallgrass prairie soils in the United States. *Science* 342**,** 621-624.

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

Fierer, N., Leff, J.W., Adams, B.J., Nielsen, U.N., Bates, S.T., Lauber, C.L., Owens, S., Gilbert, J.A., Wall, D.H., and Caporaso, J.G. (2012b). Cross-biome metagenomic

https://doi.org/10.5194/soil-2021-148
Preprint. Discussion started: 13 January 2022

analyses of soil microbial communities and their functional attributes. *Proceedings of the National Academy of Sciences* 109**,** 21390-21395.

Galperin, M.Y., Makarova, K.S., Wolf, Y.I., and Koonin, E.V. (2014). Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic acids research* 43**,** D261-D269.

Gamfeldt, L., Hillebrand, H., and Jonsson, P.R. (2008). Multiple functions increase the importance of biodiversity for overall ecosystem functioning. *Ecology* 89**,** 1223-1231.

Gans, J., Wolinsky, M., and Dunbar, J. (2005). Computational improvements reveal great bacterial diversity and high metal toxicity in soil. *Science* 309**,** 1387-1390.

Gloor, G.B., Macklaim, J.M., Pawlowsky-Glahn, V., and Egozcue, J.J. (2017). Microbiome datasets are compositional: and this is not optional. *Frontiers in microbiology* 8**,** 2224.

Hall, E.K., Bernhardt, E.S., Bier, R.L., Bradford, M.A., Boot, C.M., Cotner, J.B., Del Giorgio, P.A., Evans, S.E., Graham, E.B., and Jones, S.E. (2018). Understanding how microbiomes influence the systems they inhabit. *Nature microbiology* 3**,** 977-982.

Heberle, H., Meirelles, G.V., Da Silva, F.R., Telles, G.P., and Minghim, R. (2015). InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC bioinformatics* 16**,** 169.

Hector, A., and Bagchi, R. (2007). Biodiversity and ecosystem multifunctionality. *Nature* 448**,** 188.

https://doi.org/10.5194/soil-2021-148
Preprint. Discussion started: 13 January 2022

Hijmans, R.J., Van Etten, J., Cheng, J., Mattiuzzi, M., Sumner, M., Greenberg, J.A., Lamigueiro, O.P., Bevan, A., Racine, E.B., and Shortridge, A. (2015). Package 'raster'. *R package* 734.

Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M.C., Rattei, T., Mende, D.R., Sunagawa, S., and Kuhn, M. (2015). eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic acids research* 44**,** D286-D293.

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2015). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research* 44**,** D457-D462.

Kottek, M., Grieser, J., Beck, C., Rudolf, B., and Rubel, F. (2006). World map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift* 15**,** 259-263.

Leff, J.W., Jones, S.E., Prober, S.M., Barberán, A., Borer, E.T., Firn, J.L., Harpole, W.S., Hobbie, S.E., Hofmockel, K.S., and Knops, J.M. (2015). Consistent responses of soil microbial communities to elevated nutrient inputs in grasslands across the globe. *Proceedings of the National Academy of Sciences* 112**,** 10967-10972.

Louca, S., Jacques, S.M., Pires, A.P., Leal, J.S., Srivastava, D.S., Parfrey, L.W., Farjalla, V.F., and Doebeli, M. (2017). High taxonomic variability despite stable functional structure across microbial communities. *Nature ecology & evolution* 1**,** 0015.

Louca, S., Parfrey, L.W., and Doebeli, M. (2016). Decoupling function and taxonomy in the global ocean microbiome. *Science* 353**,** 1272-1277.

https://doi.org/10.5194/soil-2021-148
Preprint. Discussion started: 13 January 2022

Louca, S., Polz, M.F., Mazel, F., Albright, M.B., Huber, J.A., O'connor, M.I., Ackermann, M., Hahn, A.S., Srivastava, D.S., and Crowe, S.A. (2018). Function and functional redundancy in microbial systems. *Nature ecology & evolution* 2**,** 936.

Mcgill, B.J., Enquist, B.J., Weiher, E., and Westoby, M. (2006). Rebuilding community ecology from functional traits. *Trends in ecology & evolution* 21**,** 178-185.

Meyer, F., Paarmann, D., D'souza, M., Olson, R., Glass, E.M., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., and Wilke, A. (2008). The metagenomics RAST server–a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC bioinformatics* 9**,** 386.

Moulin, L., Munive, A., Dreyfus, B., and Boivin-Masson, C. (2001). Nodulation of legumes by members of the β-subclass of Proteobacteria. *Nature* 411**,** 948.

Olesen, J.M., Bascompte, J., Dupont, Y.L., and Jordano, P. (2007). The modularity of pollination networks. *Proceedings of the National Academy of Sciences* 104**,** 19891-19896.

Overbeek, R., Olson, R., Pusch, G.D., Olsen, G.J., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Parrello, B., and Shukla, M. (2013). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic acids research* 42**,** D206-D214.

Pan, Y., Cassman, N., De Hollander, M., Mendes, L.W., Korevaar, H., Geerts, R.H., Van Veen, J.A., and Kuramae, E.E. (2014). Impact of long-term N, P, K, and NPK fertilization on the composition and potential functions of the bacterial community in grassland soil. *FEMS microbiology ecology* 90**,** 195-205.

https://doi.org/10.5194/soil-2021-148
Preprint. Discussion started: 13 January 2022

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

Philippot, L., Spor, A., Hénault, C., Bru, D., Bizouard, F., Jones, C.M., Sarr, A., and Maron, P.-A. (2013). Loss in microbial diversity affects nitrogen cycling in soil. *The ISME journal* 7**,** 1609.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., and Glöckner, F.O. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic acids research* 35**,** 7188-7196.

Ramírez-Flandes, S., González, B., and Ulloa, O. (2019). Redox traits characterize the organization of global microbial communities. *Proceedings of the National Academy of Sciences* 116**,** 3630-3635.

Rivett, D.W., and Bell, T. (2018). Abundance determines the functional role of bacterial phylotypes in complex communities. *Nature microbiology* 3**,** 767.

Rocca, J.D., Hall, E.K., Lennon, J.T., Evans, S.E., Waldrop, M.P., Cotner, J.B., Nemergut, D.R., Graham, E.B., and Wallenstein, M.D. (2015). Relationships between protein-encoding gene abundance and corresponding process are commonly assumed yet rarely observed. *The ISME journal* 9**,** 1693.

Rosenfeld, J.S. (2002). Functional redundancy in ecology and conservation. *Oikos* 98**,** 156-162.

https://doi.org/10.5194/soil-2021-148
Preprint. Discussion started: 13 January 2022

671 Rousk, J., Brookes, P.C., and Bååth, E. (2009). Contrasting soil pH effects on fungal and
672 bacterial growth suggest functional redundancy in carbon mineralization. *Appl.*
673 *Environ. Microbiol.* 75**,** 1589-1596.
674 Schimel, J. (1995). "Ecosystem consequences of microbial diversity and community
675 structure," in *Arctic and alpine biodiversity: patterns, causes and ecosystem*
676 *consequences*. Springer), 239-254.
677 Schimel, J.P., and Gulledge, J. (1998). Microbial community structure and global trace
678 gases. *Global change biology* 4**,** 745-758.
679 Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W.S., and
680 Huttenhower, C. (2011). Metagenomic biomarker discovery and explanation.
681 *Genome biology* 12**,** R60.
682 Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N.,
683 Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for
684 integrated models of biomolecular interaction networks. *Genome research* 13**,**
685 2498-2504.
686 Souza, R.C., Hungria, M., Cantão, M.E., Vasconcelos, A.T.R., Nogueira, M.A., and
687 Vicente, V.A. (2015). Metagenomic analysis reveals microbial functional
688 redundancies and specificities in a soil under different tillage and crop-
689 management regimes. *Applied Soil Ecology* 86**,** 106-112.
690 Stephen, J.R., Mccaig, A.E., Smith, Z., Prosser, J.I., and Embley, T.M. (1996). Molecular
691 diversity of soil and marine 16S rRNA gene sequences related to beta-subgroup
692 ammonia-oxidizing bacteria. *Appl. Environ. Microbiol.* 62**,** 4147-4154.

https://doi.org/10.5194/soil-2021-148
Preprint. Discussion started: 13 January 2022

693 Tatusova, T., Ciufo, S., Fedorov, B., O'neill, K., and Tolstoy, I. (2013). RefSeq microbial
694 genomes database: new representation and annotation strategy. *Nucleic acids*
695 *research* 42**,** D553-D559.
696 Torsvik, V., and Øvreås, L. (2002). Microbial diversity and function in soil: from genes
697 to ecosystems. *Current opinion in microbiology* 5**,** 240-245.
698 Tringe, S.G., Von Mering, C., Kobayashi, A., Salamov, A.A., Chen, K., Chang, H.W.,
699 Podar, M., Short, J.M., Mathur, E.J., and Detter, J.C. (2005). Comparative
700 metagenomics of microbial communities. *Science* 308**,** 554-557.
701 Wellington, E.M., Berry, A., and Krsek, M. (2003). Resolving functional diversity in
702 relation to microbial community structure in soil: exploiting genomics and stable
703 isotope probing. *Current opinion in microbiology* 6**,** 295-301.
704 Wilke, A., Gerlach, W., Harrison, T., Paczian, T., Trimble, W.L., and Meyer, F. (2017).
705 MG-RAST manual for version 4, revision 3. *Lemont, IL: Argonne National*
706 *Laboratory*.
707 Yin, B., Crowley, D., Sparovek, G., De Melo, W.J., and Borneman, J. (2000). Bacterial
708 functional redundancy along a soil reclamation gradient. *Appl. Environ.*
709 *Microbiol.* 66**,** 4361-4365.
710 Zhou, J., Deng, Y., Luo, F., He, Z., and Yang, Y. (2011). Phylogenetic molecular
711 ecological network of soil microbial communities in response to elevated CO2.
712 *MBio* 2**,** e00122-00111.

713