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

Huaihai Chen[a,*], Kayan Ma[a], Yu Huang[a], Fu Qi[a], Yingbo Qiu[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

Type of contribution: Original Research Article

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.

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 does 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 shifts 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). The concept of functional redundancy can be "strict redundancy" meaning that microorganisms sharing the exact same set of functions can easily substitute each other, or alternatively "partial redundancy" denoting that

microbes have similarity in certain functions but still harbor difference in other functions, leading to partially dissimilar ecological requirements or environmental preference (Galand et al., 2018). In addition, microbial functional redundancy may be caused by more than just metabolic processes but other mechanical response to environmental disturbance, such as different foraging strategies, particles attachment and biofilm formation, nitrogen source usage, and resistance to antibiotics, which are difficult to be thoroughly evaluated in the current approach mostly focusing on metabolic redundancy (Louca et al., 2018).

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 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). Microbial functional redundancy is inevitable when a high-dimensional trait space is projected to a lower-dimensional function space of interest (Louca et al., 2018). Such apparent contradictory results suggest the degree of functional redundancy may arise from the definition of "redundancy" in different studies, our limitations in measuring the factors controlling niche space, and more importantly depending 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., 2021;Xu et al., 2021). However, a synthetic metagenomic analysis to assess how 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

115 certain groups of soil microbes (Schimel, 1995). The five "broad" functions selected were
116 AAD (Amino Acids and Derivatives), CHO (Carbohydrates), CBS (Clustering-based
117 Subsystems), CVPGP (Cofactors, Vitamins, Prosthetic Groups, Pigments), and Protein
118 (Protein Metabolism), which are the most abundant functional categories in Subsystems
119 level 1, and represent broad-scale functions acquired by a relatively larger group of
120 diverse soil microbes (Balser et al., 2002). We further constructed the pairwise similarity
121 of function and taxonomy based on the relative abundance of functional and taxonomic
122 compositions, respectively, for the five "broad" and the five "narrow" functions. We
123 hypothesized that the taxonomic similarity of soil microbes would be more linearly
124 correlated to the functional similarity for the five "narrow" functions in comparison to the
125 five "broad" functions. Therefore, using these global soil metagenomes, our objective
126 was to test whether the taxonomic compositions of soil microbes that conduct the five
127 "narrow" functions are more dependent on the functional compositions, leading to a
128 lower level of functional redundancy in the "narrow" functions than the "broad"
129 functions.
130
131 **2. Materials and Methods**
132 **2.1. Data collection**
133 To ensure that the quality and completeness of the metagenomes analyzed were of
134 standard, we carefully selected soil metagenomes in MG-RAST server that have been
135 published in peer-reviewed journals. We searched peer-reviewed publications from 2012
136 to 2018 from the Web of Science database using search terms such as "soil
137 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://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., 2021). 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. The genus level was used as the taxonomic classification level across different datasets. Following default setting in MG-RAST, if the species were classified into the higher classification levels than genus but failed to be identified at the genus level, they were classified into "unclassified" groups. Across different studies, there were 2.16 ±0.85 % of sequences belonging to the "unclassified" groups, showing that most taxonomic groups could be classified into the genus level. Total hits of taxonomic compositions of soil microbes conducting each 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 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

184 Orthology (KO) (Kanehisa et al., 2015), Clusters of Orthologous Groups (COG)
185 (Galperin et al., 2014), and Non-supervised Orthologous Groups (NOG) (Huerta-Cepas et
186 al., 2015) databases was that Subsystems had more diverse classification at Level 1,
187 allowing us to conduct direct comparison between "broad" versus "narrow" functions.
188 We chose RefSeq database rather than the traditional ribosomal RNA databases, such as
189 RDP (Ribosomal Database Project) (Cole et al., 2008), Greengenes (DeSantis et al.,
190 2006), or Silva LSU/SSU (Pruesse et al., 2007) databases, because taxonomic hits in the
191 RefSeq database were over 1000-fold higher than the rRNA databases, rendering the
192 resolution comparable to functional hits for comparison between "broad" and "narrow"
193 functions. To increase the coverage of our datasets, soil metagenomes with/without
194 assembly were both included.
195 The geographic coordinates of latitudes (LAT) and longitudes (LONG) of each soil
196 metagenome were directly obtained from publications. Based on LAT and LONG,
197 climate data of mean annual temperature (MAT) and precipitation (MAP) of study sites
198 for each soil metagenome were extracted from the WorldClim dataset (Fick and Hijmans,
199 2017) using the R package 'raster' (Hijmans et al., 2015). To examine how microbial
200 taxonomic diversities of "broad" and "narrow" functions differ globally, soil
201 metagenomic data was classified into seventeen climate zones based on the main
202 classification of Koeppen-Geiger Climatic Zones (Kottek et al., 2006) using the R
203 package 'kgc' (Bryant et al., 2017).
204
205 **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-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 PERMANOVA 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 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 (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 (Guimerà and Nunes Amaral, 2005;Olesen et al., 2006;Guimerà et al., 2007). An overview of data acquisition, transformation, and analysis processes in this study was given in Supplementary Fig. 1.

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.

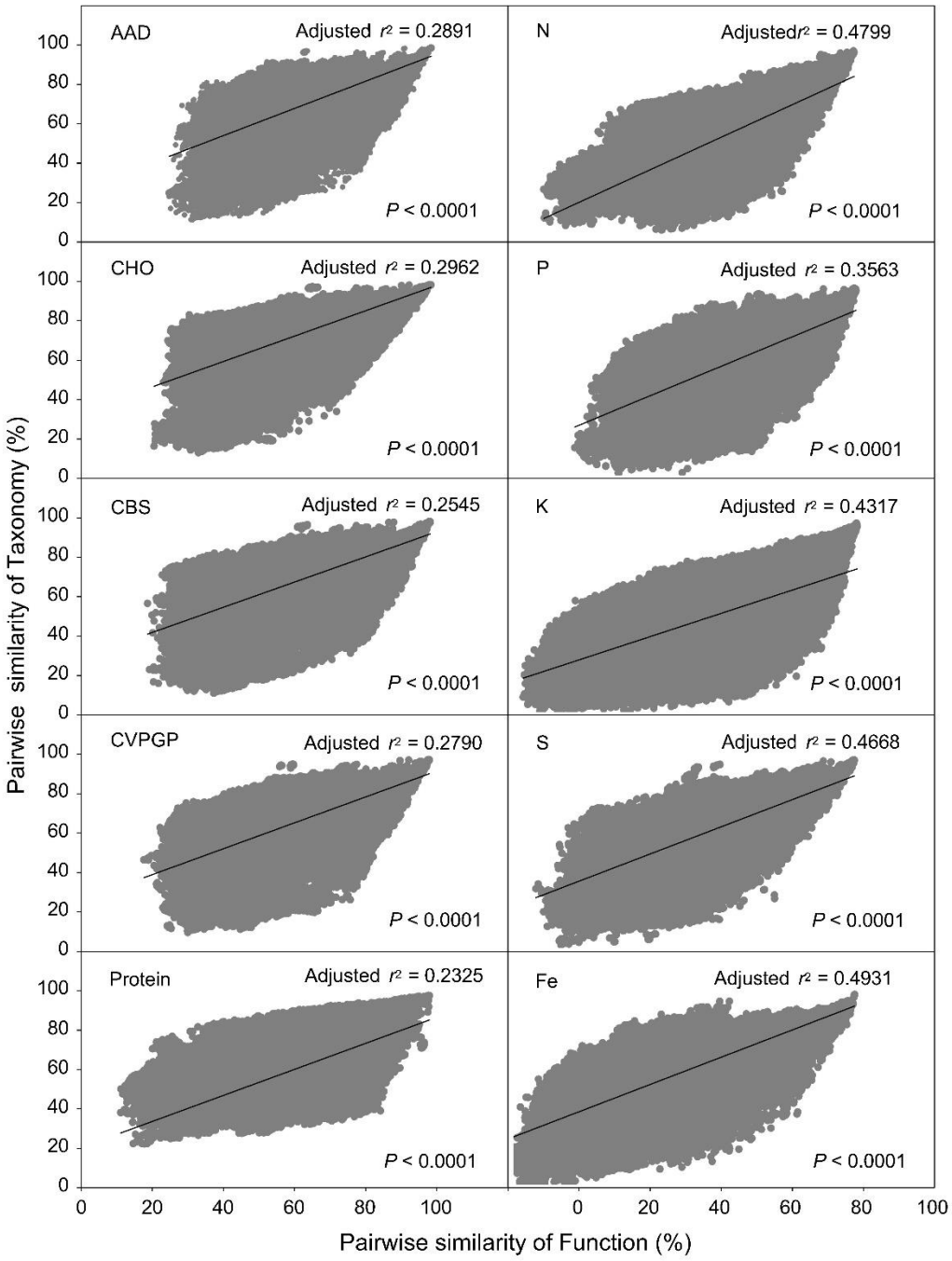

Fig. 1. Relations between functional and taxonomic beta-diversities for "broad" and "narrow" functions. Pearson's correlations between pairwise Bray-curtis similarity of microbial taxonomic and functional compositions for "broad" and "narrow" functions annotated using Subsystems at function level (Function) 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-

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.

To compare similarity ranges of these two compositions related to the five "broad" functions versus the five "narrow" functions, the boxplots were constructed based on the pairwise similarity of function and taxonomy. 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

difference, in which the five “narrow” functions were more distinct from each other. These evidences together suggest that the taxonomic composition of soil microbes conducting the five “broad” functions were more conserved in taxonomy than those conducting the five “narrow” functions. However, it should be noted that the current analysis had some limitations as the metagenomics datasets consisted of sequencing data that are phylogenetically classified and assigned based on certain taxonomic and functional databases. Thus, our results may to some extent depend on the databases chosen, of which the classification and assignment may contain potential bias. Future studies should continue to test this hypothesis using regional samples and individual datasets.

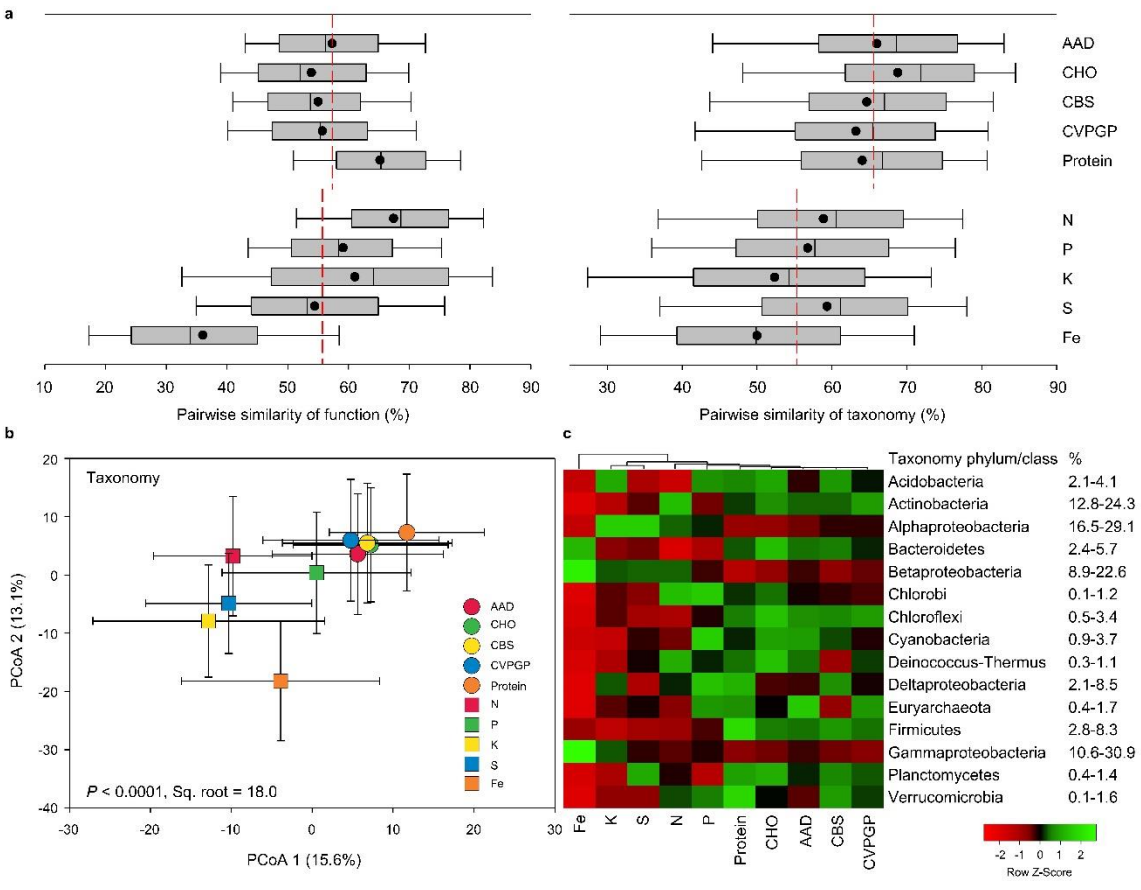

Fig. 2. Functional and taxonomic diversities for "broad" versus "narrow" functions. **a**, Box plots and mean values of pairwise Bray-curtis similarity of microbial functional and taxonomic diversities for "broad" versus "narrow" functions. **b**, PCoA (Principal coordinates analysis) showing beta-diveristy 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. 2). 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

microbial taxonomic and functional profiles is still inevitable when a low-dimensional functional space is projected to a high-dimensional taxonomic space (Louca et al., 2018), especially for "broad" functions. Moreover, certain environmental factors may have significant effects on the coupling of taxonomy and function due to their already existent selective pressure, such as the extreme environment of ice cap, and thus future research can focus on comparison of relationship between function and taxonomy among terrestrial ecosystems of different selective pressure levels.

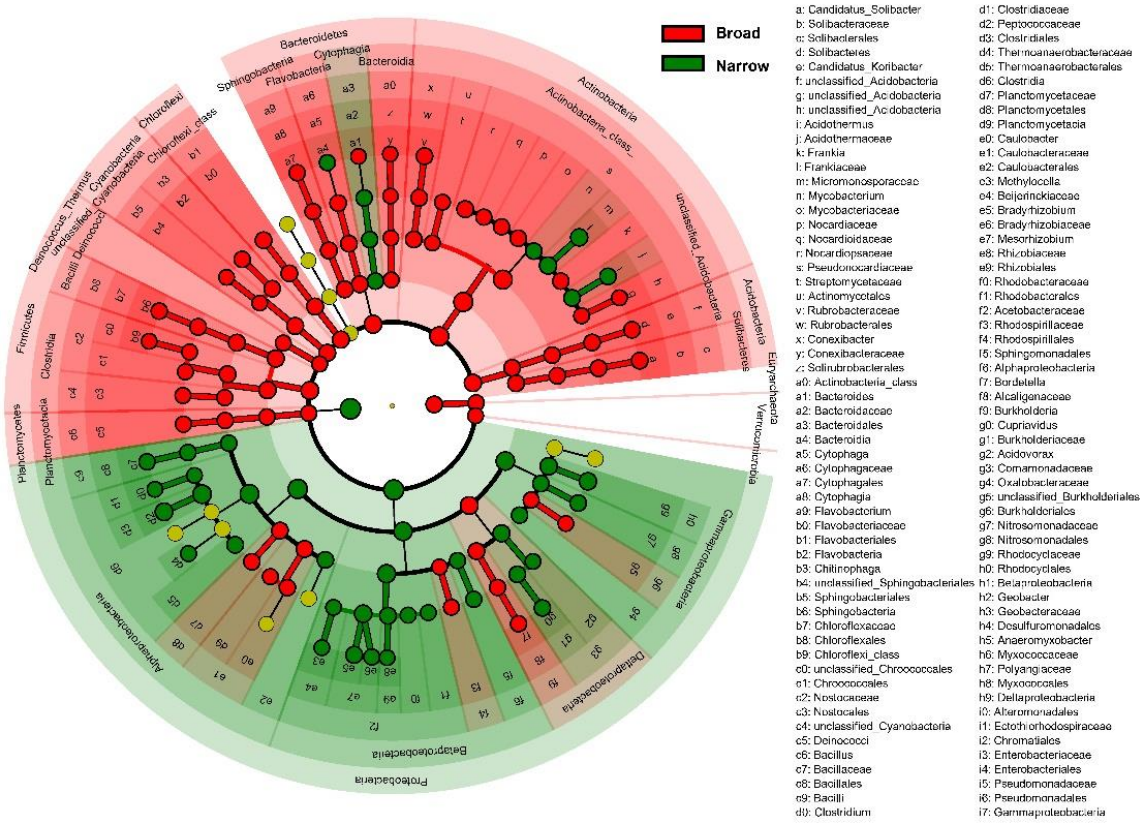

Fig. 3. Difference of taxonomic compositions between "broad" and "narrow functions". LEfSe (linear discriminant analysis effect size) results showing the significant differences in the relative abundance of dominant microbial taxonomic groups (mean > 0.5%) between "broad" (red) versus "narrow" (green) functions annotated using RefSeq (Taxonomy). From the center outward, each circle represents the level of domain,

401 phylum, class, order, family, and genus, respectively. The taxonomic groups with
402 significant differences are labeled by colors.
403
404 The taxonomic compositions of microbes conducting the five "broad" functions were
405 more abundant in most major phyla, such as Acidobacteria, Actinobacteria,
406 Bacteroidetes, and Firmicutes, while the relative abundance of the taxonomic
407 composition of microbes conducting the five "narrow" functions were greater in
408 Proteobacteria, especially Alphaproteobacteria and Betaproteobacteria (Fig. 2c). Other
409 studies also found that some bacteria conducting N cycling, such as ammonia-oxidizers
410 and rhizobia for N fixation, mainly belong to Alphaproteobacteria or Betaproteobacteria
411 (Stephen et al., 1996;Moulin et al., 2001).
412 To find out the dominant microbial groups that were statistically different between
413 the five "broad" and the five "narrow" functions, LEfSe analysis was conducted based on
414 the relative abundances at the taxonomic levels of domain, phylum, class, order, family,
415 and genus. In particular, among the Proteobacteria conducting the five "narrow"
416 functions, *Bacillaceae* from Bacilli, *Clostridium*, *Peptococcaceae*, and
417 *Thermoanaerobacteraceae* from Clostridia, *Methylocella*, *Bradyrhizobium*,
418 *Bradyrhizobiaceae*, and *Rhizobiaceae* from Rhodospirillaceae, and *Cupriavidus* from
419 Comamonadaceae had higher relative abundance than the others (Fig. 3). The Venn's
420 diagrams indicated that the taxonomic compositions of soil microbes performing the
421 "broad" functions shared 68% dominant genera among the five functional categories,
422 while the proportion was reduced to only 41% for the five "narrow" functions (Fig. 4).
423 However, it should be stated that all the analyses performed in our study were based on
424 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 linkage framework to account for the microbial contributions to major biogeochemical cycles.

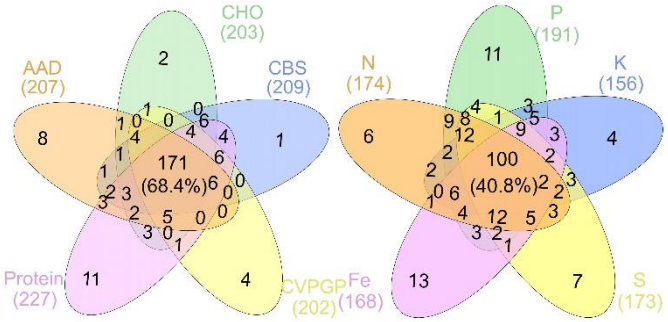

Fig. 4. Taxonomic compositions shared among "broad" and "narrow" functions. Venn's diagrams showing dominant microbial taxonomic groups (mean > 0.1%) annotated using RefSeq at genus levels (Taxonomy) shared among "broad" and "narrow" functions.

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, 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

To identify potential interaction patterns of microbial groups that conduct the five "broad" and the five "narrow" functions, the co-occurrence networks of taxonomic compositions were generated based on the taxonomic composition at the genus level across the globe. Network graphs with submodule structures indicated distinct topology of taxonomic networks between the "broad" and "narrow" functions (Table 1, Supplementary Fig. 3 and Supplementary Fig. 4). 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. These significant difference of network properties between "broad" and "narrow" functions suggests that taxonomic composition of "narrow" functions had both facilitative and inhibitive interactions, while taxonomic compositions of the "broad" function are all cooperative (Faust and Raes, 2012). Thus, soil microbes with "broad" functions tended to respond to the environment in a similar way, indicating functional sharing and association, while distinct microorganisms to conduct "narrow" functions competitively interact with each other, reflecting regulatory or suppression relationships (Delgado-Baquerizo et al., 2018).

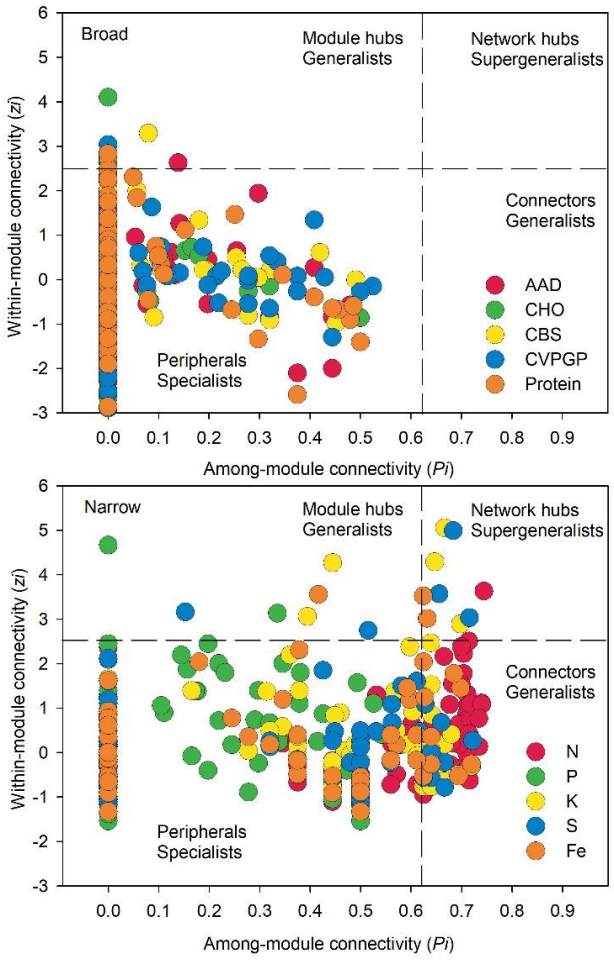

Fig. 5. Network information of taxonomic compositions for "broad" and "narrow" functions. Node distribution of module-based topological roles of taxonomic compositions for "broad" and "narrow" functions determined by the scatter plot of within-module connectivity (*zi*) and among-module connectivity (*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 "broad" function networks had links mainly within modules of similar species. In the co-occurrence network of taxonomic composition of the "narrow" functions, 13% of the nodes were identified as connectors linking several modules (high *Pi*) connectors, while 3% were identified as module hubs that connected other nodes within their own modules (high *Zi*), indicated by the *Zi-Pi* plot (Olesen et al., 2007;Deng et al., 2012). Thus, significantly less nodes were identified as module hubs in the co-occurrence network of the taxonomic composition of the "broad" functions, indicting less correlations found among different modules. This is expected given that module was comprised of genera that were mainly from the same phylogenetic groups. This difference was consistent with the Venn's diagrams showing significantly more nodes (54%) shared among the five functional categories representing the "broad" functions, while only 5% of the nodes were overlaid among the five "narrow" function networks (Fig. 6). Environmental conditions likely determine the microbial taxonomic composition, and microbial phylotypes sharing similar habitat preferences tend to co-occur (Delgado-Baquerizo et al., 2018;Ram ́ırez-Flandes et al., 2019). We emphasize that this analysis is a combination

of snapshots of microbial communities compared across space, thus environmental conditions (at the same geographic location) may vary, and the levels of functional redundancy may change depending on the mechanisms selecting specific functions and the phylogenetic distribution of those functions (Louca et al., 2018).

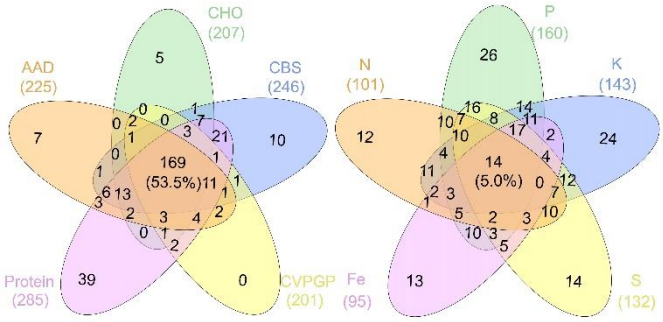

Fig. 6. Taxonomic network nodes shared among "broad" and "narrow" functions. Venn's diagrams showing the microbial taxonomic network nodes shared among "broad" and "narrow" functions.

3.3. Conclusion

By analyzing and generalizing microbial taxonomic and functional profiles, we provide strong evidence that the degree of soil microbial functional redundancy differs 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

Huaihai Chen conceived the study, performed the data analysis, interpreted the results, and drafted the manuscript. JL, CWS, and Hao Chen secured the research funding. KM, YH, QF, YQ, and Hao Chen critically assessed and interpreted the findings. All authors discussed results, commented on, edited, revised, and approved the manuscript.

Funding

This study is supported by the National Natural Science Foundation of China (No. 41977106, No. 31872691), Basic and Applied Basic Research Foundation of Guangdong Province.

Acknowledgements

We thank the authors of the publications included in our study, without which this global metagenomic analysis would not be possible.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon request. All metagenomic data used in this study are publicly assessable in the MG-RAST server with study and MG-RAST ID reported in supplementary files.

Competing Interests

The authors declare no competing interests.

References

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

565 microbial diversity drives ecosystem multifunctionality and is physiologically
566 related to carbon availability in Mediterranean semi-arid soils. *Molecular ecology*
567 25**,** 4660-4673.
568 Beier, S., Shen, D., Schott, T., and Jürgens, K. (2017). Metatranscriptomic data reveal the
569 effect of different community properties on multifunctional redundancy. *Molecular*
570 *ecology* 26**,** 6813-6826.
571 Bryant, C., Wheeler, N., Rubel, F., and French, R. (2017). "kgc: Koeppen–Geiger
572 climatic zones. R package version 1.0. 0.2".).
573 Chen, H., Ma, K., Huang, Y., Yao, Z., and Chu, C. (2021). Stable soil microbial
574 functional structure responding to biodiversity loss based on metagenomic
575 evidences. *Frontiers in Microbiology* 12.
576 Clarke, K., and Gorley, R. (2015). Getting started with PRIMER v7. *PRIMER-E:*
577 *Plymouth, Plymouth Marine Laboratory*.
578 Cole, J.R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R.J., Kulam-Syed-
579 Mohideen, A., Mcgarrell, D.M., Marsh, T., and Garrity, G.M. (2008). The
580 Ribosomal Database Project: improved alignments and new tools for rRNA analysis.
581 *Nucleic acids research* 37**,** D141-D145.
582 Crowther, T.W., Van Den Hoogen, J., Wan, J., Mayes, M.A., Keiser, A., Mo, L., Averill,
583 C., and Maynard, D.S. (2019). The global soil community and its influence on
584 biogeochemistry. *Science* 365**,** eaav0550.
585 Delgado-Baquerizo, M., Maestre, F.T., Reich, P.B., Jeffries, T.C., Gaitan, J.J., Encinar,
586 D., Berdugo, M., Campbell, C.D., and Singh, B.K. (2016). Microbial diversity
587 drives multifunctionality in terrestrial ecosystems. *Nature communications* 7**,** 10541.

588 Delgado-Baquerizo, M., Oliverio, A.M., Brewer, T.E., Benavent-González, A., Eldridge,
589 D.J., Bardgett, R.D., Maestre, F.T., Singh, B.K., and Fierer, N. (2018). A global
590 atlas of the dominant bacteria found in soil. *Science* 359**,** 320-325.
591 Delgado-Baquerizo, M., Giaramida, L., Reich, P.B., Khachane, A.N., Hamonts, K.,
592 Edwards, C., Lawton, L.A., and Singh, B.K. (2016). Lack of functional redundancy
593 in the relationship between microbial diversity and ecosystem functioning. *Journal*
594 *of Ecology* 104**,** 936-946.
595 Deng, Y., Jiang, Y.-H., Yang, Y., He, Z., Luo, F., and Zhou, J. (2012). Molecular
596 ecological network analyses. *BMC bioinformatics* 13**,** 113.
597 Desantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T.,
598 Dalevi, D., Hu, P., and Andersen, G.L. (2006). Greengenes, a chimera-checked 16S
599 rRNA gene database and workbench compatible with ARB. *Appl. Environ.*
600 *Microbiol.* 72**,** 5069-5072.
601 Faust, K., and Raes, J. (2012). Microbial interactions: from networks to models. *Nature*
602 *Reviews Microbiology* 10**,** 538.
603 Fick, S.E., and Hijmans, R.J. (2017). WorldClim 2: new 1-km spatial resolution climate
604 surfaces for global land areas. *International journal of climatology* 37**,** 4302-4315.
605 Fierer, N., Ladau, J., Clemente, J.C., Leff, J.W., Owens, S.M., Pollard, K.S., Knight, R.,
606 Gilbert, J.A., and Mcculley, R.L. (2013). Reconstructing the microbial diversity and
607 function of pre-agricultural tallgrass prairie soils in the United States. *Science* 342**,**
608 621-624.

609 Fierer, N., Lauber, C.L., Ramirez, K.S., Zaneveld, J., Bradford, M.A., and Knight, R.
610 (2012a). Comparative metagenomic, phylogenetic and physiological analyses of soil
611 microbial communities across nitrogen gradients. *The ISME journal* 6**,** 1007.
612 Fierer, N., Leff, J.W., Adams, B.J., Nielsen, U.N., Bates, S.T., Lauber, C.L., Owens, S.,
613 Gilbert, J.A., Wall, D.H., and Caporaso, J.G. (2012b). Cross-biome metagenomic
614 analyses of soil microbial communities and their functional attributes. *Proceedings*
615 *of the National Academy of Sciences* 109**,** 21390-21395.
616 Galand, P.E., Pereira, O., Hochart, C., Auguet, J.C., and Debroas, D. (2018). A strong
617 link between marine microbial community composition and function challenges the
618 idea of functional redundancy. *The ISME Journal* 12**,** 2470-2478.
619 Galperin, M.Y., Makarova, K.S., Wolf, Y.I., and Koonin, E.V. (2014). Expanded
620 microbial genome coverage and improved protein family annotation in the COG
621 database. *Nucleic acids research* 43**,** D261-D269.
622 Gamfeldt, L., Hillebrand, H., and Jonsson, P.R. (2008). Multiple functions increase the
623 importance of biodiversity for overall ecosystem functioning. *Ecology* 89**,** 1223-
624 1231.
625 Gans, J., Wolinsky, M., and Dunbar, J. (2005). Computational improvements reveal great
626 bacterial diversity and high metal toxicity in soil. *Science* 309**,** 1387-1390.
627 Gloor, G.B., Macklaim, J.M., Pawlowsky-Glahn, V., and Egozcue, J.J. (2017).
628 Microbiome datasets are compositional: and this is not optional. *Frontiers in*
629 *microbiology* 8**,** 2224.
630 Guimerà, R., and Nunes Amaral, L.A. (2005). Functional cartography of complex
631 metabolic networks. *Nature* 433**,** 895-900.

Guimerà, R., Sales-Pardo, M., and Amaral, L.a.N. (2007). Classes of complex networks defined by role-to-role connectivity profiles. *Nature Physics* 3**,** 63-69.

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.

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.

653 Kottek, M., Grieser, J., Beck, C., Rudolf, B., and Rubel, F. (2006). World map of the
654 Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift* 15**,** 259-
655 263.
656 Leff, J.W., Jones, S.E., Prober, S.M., Barberán, A., Borer, E.T., Firn, J.L., Harpole, W.S.,
657 Hobbie, S.E., Hofmockel, K.S., and Knops, J.M. (2015). Consistent responses of
658 soil microbial communities to elevated nutrient inputs in grasslands across the globe.
659 *Proceedings of the National Academy of Sciences* 112**,** 10967-10972.
660 Louca, S., Jacques, S.M., Pires, A.P., Leal, J.S., Srivastava, D.S., Parfrey, L.W., Farjalla,
661 V.F., and Doebeli, M. (2017). High taxonomic variability despite stable functional
662 structure across microbial communities. *Nature ecology & evolution* 1**,** 0015.
663 Louca, S., Parfrey, L.W., and Doebeli, M. (2016). Decoupling function and taxonomy in
664 the global ocean microbiome. *Science* 353**,** 1272-1277.
665 Louca, S., Polz, M.F., Mazel, F., Albright, M.B., Huber, J.A., O'connor, M.I.,
666 Ackermann, M., Hahn, A.S., Srivastava, D.S., and Crowe, S.A. (2018). Function and
667 functional redundancy in microbial systems. *Nature Ecology & Evolution* 2**,** 936.
668 Mcgill, B.J., Enquist, B.J., Weiher, E., and Westoby, M. (2006). Rebuilding community
669 ecology from functional traits. *Trends in ecology & evolution* 21**,** 178-185.
670 Meyer, F., Paarmann, D., D'souza, M., Olson, R., Glass, E.M., Kubal, M., Paczian, T.,
671 Rodriguez, A., Stevens, R., and Wilke, A. (2008). The metagenomics RAST server–
672 a public resource for the automatic phylogenetic and functional analysis of
673 metagenomes. *BMC bioinformatics* 9**,** 386.
674 Moulin, L., Munive, A., Dreyfus, B., and Boivin-Masson, C. (2001). Nodulation of
675 legumes by members of the β-subclass of Proteobacteria. *Nature* 411**,** 948.

676 Olesen, J.M., Bascompte, J., Dupont, Y.L., and Jordano, P. (2006). The smallest of all
677 worlds: Pollination networks. *Journal of Theoretical Biology* 240**,** 270-276.
678 Olesen, J.M., Bascompte, J., Dupont, Y.L., and Jordano, P. (2007). The modularity of
679 pollination networks. *Proceedings of the National Academy of Sciences* 104**,** 19891-
680 19896.
681 Overbeek, R., Olson, R., Pusch, G.D., Olsen, G.J., Davis, J.J., Disz, T., Edwards, R.A.,
682 Gerdes, S., Parrello, B., and Shukla, M. (2013). The SEED and the Rapid
683 Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic*
684 *acids research* 42**,** D206-D214.
685 Pan, Y., Cassman, N., De Hollander, M., Mendes, L.W., Korevaar, H., Geerts, R.H., Van
686 Veen, J.A., and Kuramae, E.E. (2014). Impact of long-term N, P, K, and NPK
687 fertilization on the composition and potential functions of the bacterial community
688 in grassland soil. *FEMS microbiology ecology* 90**,** 195-205.
689 Peter, H., Beier, S., Bertilsson, S., Lindström, E.S., Langenheder, S., and Tranvik, L.J.
690 (2011). Function-specific response to depletion of microbial diversity. *The ISME*
691 *journal* 5**,** 351.
692 Philippot, L., Spor, A., Hénault, C., Bru, D., Bizouard, F., Jones, C.M., Sarr, A., and
693 Maron, P.-A. (2013). Loss in microbial diversity affects nitrogen cycling in soil. *The*
694 *ISME journal* 7**,** 1609.
695 Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., and Glöckner,
696 F.O. (2007). SILVA: a comprehensive online resource for quality checked and
697 aligned ribosomal RNA sequence data compatible with ARB. *Nucleic acids*
698 *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.

Rousk, J., Brookes, P.C., and Bååth, E. (2009). Contrasting soil pH effects on fungal and bacterial growth suggest functional redundancy in carbon mineralization. *Appl. Environ. Microbiol.* 75**,** 1589-1596.

Schimel, J. (1995). "Ecosystem consequences of microbial diversity and community structure," in *Arctic and alpine biodiversity: patterns, causes and ecosystem consequences*. Springer), 239-254.

Schimel, J.P., and Gulledge, J. (1998). Microbial community structure and global trace gases. *Global change biology* 4**,** 745-758.

Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W.S., and Huttenhower, C. (2011). Metagenomic biomarker discovery and explanation. *Genome biology* 12**,** R60.

721 Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N.,
722 Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for
723 integrated models of biomolecular interaction networks. *Genome research* 13**,** 2498-
724 2504.
725 Souza, R.C., Hungria, M., Cantão, M.E., Vasconcelos, A.T.R., Nogueira, M.A., and
726 Vicente, V.A. (2015). Metagenomic analysis reveals microbial functional
727 redundancies and specificities in a soil under different tillage and crop-management
728 regimes. *Applied Soil Ecology* 86**,** 106-112.
729 Stephen, J.R., Mccaig, A.E., Smith, Z., Prosser, J.I., and Embley, T.M. (1996). Molecular
730 diversity of soil and marine 16S rRNA gene sequences related to beta-subgroup
731 ammonia-oxidizing bacteria. *Appl. Environ. Microbiol.* 62**,** 4147-4154.
732 Tatusova, T., Ciufo, S., Fedorov, B., O'neill, K., and Tolstoy, I. (2013). RefSeq microbial
733 genomes database: new representation and annotation strategy. *Nucleic acids*
734 *research* 42**,** D553-D559.
735 Torsvik, V., and Øvreås, L. (2002). Microbial diversity and function in soil: from genes
736 to ecosystems. *Current opinion in microbiology* 5**,** 240-245.
737 Tringe, S.G., Von Mering, C., Kobayashi, A., Salamov, A.A., Chen, K., Chang, H.W.,
738 Podar, M., Short, J.M., Mathur, E.J., and Detter, J.C. (2005). Comparative
739 metagenomics of microbial communities. *Science* 308**,** 554-557.
740 Wellington, E.M., Berry, A., and Krsek, M. (2003). Resolving functional diversity in
741 relation to microbial community structure in soil: exploiting genomics and stable
742 isotope probing. *Current opinion in microbiology* 6**,** 295-301.

743 Wilke, A., Gerlach, W., Harrison, T., Paczian, T., Trimble, W.L., and Meyer, F. (2017).
744 MG-RAST manual for version 4, revision 3. *Lemont, IL: Argonne National*
745 *Laboratory*.
746 Xu, X., Qiu, Y., Zhang, K., Yang, F., Chen, M., Luo, X., Yan, X., Wang, P., Zhang, Y.,
747 and Chen, H. (2021). Climate warming promotes deterministic assembly of
748 arbuscular mycorrhizal fungal communities. *Global Change Biology*.
749 Yin, B., Crowley, D., Sparovek, G., De Melo, W.J., and Borneman, J. (2000). Bacterial
750 functional redundancy along a soil reclamation gradient. *Appl. Environ. Microbiol.*
751 66**,** 4361-4365.
752 Zhou, J., Deng, Y., Luo, F., He, Z., and Yang, Y. (2011). Phylogenetic molecular
753 ecological network of soil microbial communities in response to elevated CO2.
754 *MBio* 2**,** e00122-00111.
755