Assessment of spike-AMP and qPCR-AMP in soil microbiota quantitative research

Meiling Zhang, Liyu Zhang, Shuyu Huang, Wentao Li, Wei Zhou, Laurent Philippot, Chao Ai

Types of paper: Research Paper

Date of preparation: August, 2020

Number of text pages: 37

Number of tables and figures: 1 table and 7 figures

Supplementary Information: 6 methods, 15 figures and 8 tables

Title: Assessment of spike-AMP and qPCR-AMP in soil microbiota quantitative research

Authors: Meiling Zhang [ID a, 1], Liyu Zhang [a], Shuyu Huang [a], Wentao Li [b], Wei Zhou [a], Laurent Philippot [ID c, *], Chao Ai [ID a, c, 1, *]

[a] Ministry of Agriculture Key Laboratory of Plant Nutrition and Fertilizer, Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing, 100081, PR China.

[b] Jiangsu Coastal Area Institute of Agricultural Sciences, Yancheng, 224002, PR China.

[c] Université Bourgogne Franche-Comté, INRAE, AgroSup Dijon, Agroécologie, 21000 Dijon, France.

[1] Meiling Zhang and Chao Ai contributed equally to this work.

*** Corresponding authors:**

Chao Ai, Tel: +86 10 82108669; E-mail: aichao@caas.cn.

Laurent Philippot, Tel: +86 10 82108671; E-mail: laurent.philippot@inrae.fr.

Abstract

Relative microbiome profiling (RMP) using new sequencing approaches has limited capacity to detect shifts in microbial abundances. The growing need for absolute abundances has led to advances in absolute microbiome profiling (AMP). However, the performance and universal applicability of these various AMP methods remain unclear. Here, the two most popular AMP methods, spike-in method (spike-AMP) and quantitative PCR combined with high-throughput sequencing (qPCR-AMP), were evaluated in soil microbiota research. Our results showed that the quantitative results based on spike-AMP were inconsistent with expected trends. The spike-derived absolute abundance was indeterminate and highly dependent on the amount of spike added. Furthermore, no good correlation was found between the addition of spike copies and output of spike reads, especially at low spike levels, contradicting the theoretical assumption of the spike-in method. Spike addition consumed substantial sequencing resources, and more importantly, it altered the original microbial community structure, explaining 16.1%–36.2% of structural variation. In contrast, the more common qPCR-AMP method provided valuable insights into the understanding of soil microbial dynamics in response to straw addition. Our results showed that the straw-induced variations in some dominant phyla such as *Proteobacteria*, *Actinobacteriota* and *Ascomycota* could only be detected by absolute rather than relative microbial profiling. We inferred microbial networks based on absolute and relative data matrices, respectively, and observed that the choice of data type essentially impacted the patterns of co-occurrence networks and the recognition of module hubs. The keystones and enriched phyla only detected by absolute microbial profiling were confirmed to be involved in straw decomposition by a stable isotope probing experiment. Overall, AMP can provide valuable insights into the understanding

of soil microbial dynamics in response to environmental fluctuations. Given its stability and technical feasibility, qPCR-AMP may be broadly applicable to soil microbiota quantitative research.

Keywords: absolute microbiome profiling, spike-in, absolute abundance, qPCR, soil microbial community

1. Introduction

Microorganisms are diverse forms of life and thrive in almost all environments. Their composition and function have substantial impacts on human health (Fan and Pedersen, 2021), global element cycling (Crowther et al., 2019), crop production (Charpentier and Oldroyd, 2010) and plant disease resistance (Kwak et al., 2018). Advances in high-throughput sequencing technologies have contributed to the surge of microbial sequencing data (White et al., 2016), but similar to previous fingerprinting approaches, such as denaturing gradient gel electrophoresis or terminal restriction fragment length polymorphism, relative microbiome profiling (RMP) obtained from sequencing data overlooks absolute microbial abundance. However, without absolute quantification, it is challenging to build a more comprehensive understanding of how dynamics of microbiome abundance vary across space, time, and in response to environmental fluctuations (Vandeputte et al., 2017; Zhang et al., 2017; Tkacz et al., 2018; Guo et al., 2019; Boshier et al., 2020).

Currently, absolute microbiome profiling (AMP) has been developed to overcome the innate limitation of high-throughput sequencing (Tourlousse et al., 2017; Vandeputte et al., 2017). There are three main types of AMP: (i) spike-in method (spike-AMP) (Fig. 1a–d, f); (ii) quantitative PCR combined with sequencing data (qPCR-AMP) (Fig. 1e, f); and (iii) flow cytometry (FCM) combined with sequencing data (FCM-AMP). Spike-AMP, similar to strategies adopted for RNA-seq (Jiang et al., 2011), has been used to extrapolate absolute abundances of microbial communities by adding a known amount of spike to environmental samples. According to the form of the spike, spike-AMP can be further divided into three different workflows (Fig. 1b–d). First, a known amount of single spike was added into environmental samples (Smets et al., 2016) or into DNA extracts (Guo et al., 2019), and the absolute abundance of a specific taxon can be obtained based on the input of spike copies and output of

sequencing reads (i.e., method #1 in Fig. 1b, f). However, the challenge of the single spike-in method was the selection of a suitable spike concentration because low levels of spikes can easily be retained in environmental samples (e.g., soils) and high levels of spikes might be oversaturated (Tkacz et al., 2018). In general, preliminary tests that a single spike with different gradient concentrations was added into environmental samples (Tkacz et al., 2018) or DNA extracts (Guo et al., 2019) were conducted to determine the appropriate amount of spike addition (i.e., method #2 in Fig. 1c). In an exploratory study, Tkacz et al. (2018) found that the optimum spike levels of 16S rRNA gene, 18S rRNA gene and ITS region should account for 20%–80% of total sequencing reads. If the absolute microbial abundances between samples vary greatly, the optimal spike concentration of each sample needs to be determined. Because of the laboriousness and complexities of method #2, an alternative method was to use one spike mixture to replace method #2 (Tourlousse et al., 2017; Jiang et al., 2019; Gao and Sun, 2020; Mou et al., 2020) (i.e., method #3 in Fig. 1d). The absolute abundance of specific taxa can be determined based on the generated spike linear relationship (Fig. 1d, f). However, the spike mixture might account for a larger proportion of sequencing data when compared with the addition of a single spike. For example, the spike mixture with nine different spikes could consume 41.7% of total reads in a given amplicon library (Jiang et al., 2019). In addition to spike-AMP, integrating qPCR-based gene copies (Zhang et al., 2017; Lou et al., 2018; Boshier et al., 2020) (Fig. 1e, f) or FCM-based cell counts (Vandeputte et al., 2017; Zhang et al., 2017; Vieira-Silva et al., 2019) into sequencing workflow have been applied to calculate the absolute abundance of a specific taxon in human gut, soil and vagina microbiomes. In addition to using spike-AMP or qPCR/FCM-AMP alone, Zemb et al. (2020) have proposed a framework in which spike-ins and qPCR data were integrated to calculate the absolute microbial abundances.

AMP represented a significant advance in microbial quantification, enriching the interpretation of microbial sequencing data from human gut (Stammler et al., 2016; Vandeputte et al., 2017; Vieira-Silva et al., 2019; Rao et al., 2021), vagina (Boshier et al., 2020), soil (Smets et al., 2016; Tourlousse et al., 2017; Zhang et al., 2017; Tkacz et al., 2018; Yang et al., 2018; Jiang et al., 2019), plant (Guo et al., 2019) and ocean water (Gao and Sun, 2020). However, the universal applicability of these methods has not yet been demonstrated in complex and diverse soil samples. Here, we first designed a set of bacterial and fungal spikes with different gradient concentrations and then added these spikes to distinct soil DNA extracts to evaluate the universal feasibility of spike-AMP (i.e., method #2 in Fig. 1c). After finding that spike-AMP did not exhibit any advantages in accuracy and reliability, we further applied qPCR-AMP to determine the absolute profiling of soil bacterial and fungal communities in response to the addition of organic matter (i.e., maize straw). The superiority of AMP was further confirmed by a stable isotope probing (SIP) experiment. The objectives of this investigation were to (i) evaluate the universal feasibility of spike-AMP and the effect of spike application on the original microbial community structure; (ii) verify whether AMP has significant advantages in determining key species that respond to environmental fluctuations (i.e., the addition of maize straw).

2. Materials and Methods

2.1. Soil samples collection

Soil cores were collected from two locations distant from each other: Hailun, North China (47.86° N, 127.01° E) (HL) and Sanya, South China (18.34° N, 109.65° E) (SY) (Fig. S1a). Three soil cores as biological replicates were collected from both upland and paddy soils at each location, thus resulting in 12 soil samples that were used for the further incubation experiment. The physico-chemical properties, microbial community structure and taxonomic distribution for these samples are shown in Fig. S2 and

Table S1. Next, SIP microcosms were prepared by adding 0.1 g ground ^{13}C-labeled maize straw (95.05 atom % ^{13}C) to the ~23 g fresh soil (equivalent to 20 g on a dry weight) (Fig. S1b). The soil microcosms without straw addition were performed as pairwise comparison for each SIP microcosms (i.e., control microcosms). Then, a 60-day incubation study was conducted on the 12 SIP microcosms and 12 control microcosms. Detailed processing steps about the ^{13}C labeling and incubation experiment are outlined in Methods 1 and 2, respectively (see Supplementary Information). During the incubation period, approximately 5 g of incubated soil was successively collected from the same set of 24 microcosms at four time points (days 0, 7, 30 and 60). Finally, we collected a total of 96 incubated soil samples.

Based on the collected incubated soil samples, we implemented the following two sections in this study (Fig. S3). In section I, considering the representativeness of samples, we selected a set of straw-amended soil samples at three time points (one of the three biological replicates at days 0, 7 and 30) from two soil types (HL paddy soils (HLP) and SY upland soils (SYU)), thus resulting in 6 soil samples to evaluate the feasibility of the spike-AMP method. In section II, qPCR-AMP were performed to detect the absolute microbial profiling of 72 incubated samples (including control and straw-amended soils at days 7, 30 and 60 from two field types (upland and paddy) from two sites (HL and SY)). In contrast to RMP, the superiority of AMP was verified by the SIP experiment. Detailed information of the qPCR-AMP and SIP experiment can be found in Methods 2 and 3 (see Supplementary Information).

133 *2.2. Overview of spike-AMP*

134 *2.2.1. Determination of the total copy numbers of 16S rRNA gene and ITS region*

135 To estimate the background abundances of 16S rRNA gene and ITS region in the six soil samples
136 used in section I, we performed qPCR by using 338F/806R (338F: 5'-ACT CCT ACG GGA GGC
137 AGC A-3'; 806R: 5'-GGA CTA CHV GGG TWT CTA AT-3') and ITS1F/ITS2 (ITS1F: 5'-CTT GGT
138 CAT TTA GAG GAA GTA A-3'; ITS2: 5'-GCT GCG TTC TTC ATC GAT GC-3') primer sets,
139 respectively. In particular, to improve the confidence of quantitative results, the 16S rRNA gene and
140 ITS region abundances of each sample were reproduced in two independent qPCR experiments (i.e.,
141 Experiments 1 and 2; Fig. 2). The qPCR processing steps including DNA samples preparation, qPCR
142 amplification and data analysis, are described in Method 4 (see Supplementary Information). The
143 qPCR reaction efficiency for targeted 16S rRNA gene and ITS region ranged from 87.59% to 92.53%
144 and 90.24% to 104.03%, respectively. The presence of PCR inhibitors in the soil DNA extracts was
145 verified by mixing a known amount of plasmid DNA (pMD^{TM}18-T vector) either with DNA extracts or
146 sterilized water; no inhibition was observed in the assays in this study.

147 *2.2.2. Design of synthetic spike*

148 The synthetic spike sequences included two regions: 1) conserved primer binding sites from
149 338F/806R and ITS1F/ITS2 for amplification of 16S rRNA gene and ITS region, respectively; 2)
150 artificial variable regions that lack identity to nucleotide sequences in public databases (Tourlousse et
151 al., 2017; Tkacz et al., 2018) (Fig. 1a). Design of the artificial variable sequence was conducted in
152 accordance with a previous publication (Tourlousse et al., 2017); the amplicon lengths were 469 and
153 421 bp for bacterial and fungal spikes, respectively (Table S2). Detailed steps on how the final spikes

were obtained are outlined in Method 5 (see Supplementary Information). Spike concentrations were measured by Quant-iT™ PicoGreen™ dsDNA Assay (Thermo Fisher Scientific, Eugene, OR, USA). Finally, spike copies were calculated according to the equation described in a previous study (Lee et al., 2006):

$$Spike\ copies\ number\ (copies\ \mu L^{-1}) = \frac{6.02\times10^{23}(copies\ mol^{-1})\times Spike\ concentration\ (ng\ \mu L^{-1})\times10^{-9}}{Spike\ length\ (bp)\times660\ (dalton\ bp^{-1})} \quad (1)$$

2.2.3. Spiking into DNA samples and high-throughput sequencing

On the basis of the background abundances of 16S rRNA gene and ITS region in the six soil samples, we designed eight bacterial and nine fungal spike gradient concentrations ranging from 6.94×10^{2} to 2.81×10^{8} and 4.91×10^{1} to 1.99×10^{7} copies μL^{-1}, respectively (Table 1). To ensure that the spike dilution level was correct, we tested the spike gradient concentration using qPCR. The standard curves of spike dilutions showed that both bacterial and fungal spikes exhibited excellent linearity between spike concentration and Ct values with $R^2 \geq 0.999$. Then, bacterial and fungal spikes were added separately to DNA samples extracted from the six tested samples. For each DNA samples, 114 samples were prepared corresponding to the bacterial/fungal control, eight bacterial and nine fungal spike levels performed in six technical replicates (Table 1). The DNA samples and synthetic spikes were co-amplified in duplicate using the bacterial 338F/806R or fungal ITS1F/ITS2 primer pairs. An 8 bp barcode sequence located in front of the forward primer was used for multiplexing of samples during sequencing. The description of amplicon library preparation for Illumina NovaSeq 6000-PE250 sequencing can be found in Method 6 (see Supplementary Information).

173 *2.2.4. Bioinformatics analysis of sequence data*

174 The sequence data were processed using QIIME2-2020.8 (Bolyen et al., 2019), QIIME v.1.9.1
175 (Caporaso et al., 2010), USEARCH v.11.0 (Edgar, 2010), VSEARCH v.2.12.0 (Rognes et al., 2016)
176 and mothur v.1.40.4 (Schloss et al., 2009). The quality of paired-end sequencing data was confirmed by
177 FastQC v.0.10.1 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Subsequently,
178 sequencing data were processed using VSEARCH and QIIME commands as follows: join paired-end (-
179 fastq_mergepairs), extract barcodes (extract_barcodes.py) and demultiplex paired-end fastq (demux).
180 The representative sequences were obtained by USEARCH and VSEARCH pipelines based on the
181 merged sequences using the following commands: remove primers (-fastx_filter), find non-redundancy
182 reads (-derep_fulllength), cluster unique reads (-cluster_size) and remove chimeric sequences (-
183 uchime3_denovo). All sequences were clustered at 97% nucleotide similarity to obtain operational
184 taxonomic units (OTUs) (-usearch_global). OTUs were aligned against the databases of bacterial
185 SILVA 138 (Quast et al., 2012) or fungal UNITE (Abarenkov et al., 2010). The OTUs defined as
186 unknown, chloroplast, mitochondria, eukaryote, cyanophyta, cyanobacteria, cercozoa and protista were
187 removed. In the spike-AMP experiment, OTU1 was mapped into synthetic spike sequences with a
188 perfect match (-usearch_global, id=1). Abnormal samples with spike reads more than 2.5 times the
189 mean values of the other technical replicates were removed, and thus five fungal samples were
190 removed. The OTUs table were rarefied at 21,073 and 32,734 sequences per sample for subsequent
191 analysis of bacterial and fungal spike-AMP, respectively.

192 *2.2.5. Calculation in the spike-AMP method*

193 The absolute microbial abundance can be calculated using a single spike (Fig. 1b) or using a spike
194 linear relationship (i.e., the blue dashed box in Fig. 1c).

In theory, there is a correlation between copy number and sequencing reads number as follows:

$$\frac{C_s}{C_m} = \frac{R_s}{R_m} \quad \rightarrow \quad C_m = C_s \times \frac{R_m}{R_s} \qquad (2)$$

where C_s and C_m are the spike copies and microbial copies (i.e., the absolute abundance of 16S rRNA gene or ITS region) of the sample, respectively. The corresponding spike sequencing reads and microbial sequencing reads are denoted as R_s and R_m, respectively. Therefore, the absolute microbial abundance (i.e., C_m) can be calculated by Eq. (2) according to the amount of spike added (i.e., C_s).

The alternative method is based on a linear relationship as follows:

$$\frac{C_s}{C_m} = \frac{R_s}{R_m} \quad \rightarrow \quad \frac{R_s}{R_m} = \frac{1}{C_m} \times C_s \quad \rightarrow \quad y = a \times x \quad \rightarrow \quad C_m = \frac{1}{a} \qquad (3)$$

where we can define $\frac{R_s}{R_m}$ as the dependent variable, y, and C_s as the independent variable, x. In theory, C_m is invariable in a given sample, thus we can define $\frac{1}{C_m}$ as a constant, a. This linear relationship ($y = a \times x$) can be obtained by the gradient concentration of spike addition, where a is the slope of the linear model. In this case, the absolute microbial abundance can be calculated via $C_m = \frac{1}{a}$.

If we present the above equation on a $\log_{10}$ scale, the linear relationship still exists as follows:

$$Log_{10}\frac{R_s}{R_m} = Log_{10}\left(\frac{1}{C_m} \times C_s\right) \rightarrow Log_{10}\left(\frac{R_s}{R_m}\right) = Log_{10}\left(\frac{1}{C_m}\right) + Log_{10}(C_s)$$

$$\rightarrow \quad Y = b + X \quad \rightarrow \quad C_m = \frac{1}{10^b} \qquad (4)$$

where we can define $Log_{10}\left(\frac{R_s}{R_m}\right)$ as the dependent variable, Y, $Log_{10}(C_s)$ as the independent variable, X, and $Log_{10}\left(\frac{1}{C_m}\right)$ as a constant, b, where b is the intercept of the linear model. The C_m can be calculated via $C_m = \frac{1}{10^b}$.

The theoretical assumption of the spike-based method is $\frac{C_s}{C_m} = \frac{R_s}{R_m}$ (Eq. (2)). To evaluate the consistency between $\frac{C_s}{C_m}$ and $\frac{R_s}{R_m}$ in actual sequencing data, we defined the bias degree as follows:

$$Bias\ degree = Log_{10}\frac{R_s}{R_{non\text{-}spike\ total}} - Log_{10}\frac{C_s}{C_{non\text{-}spike\ total}} \qquad (5)$$

where $C_{non\text{-}spike\ total}$ and $R_{non\text{-}spike\ total}$ represent the total non-spike microbial copies estimated by qPCR and the total non-spike microbial sequencing reads, respectively. C_s and R_s represent the spike copies added and the corresponding spike sequencing reads, respectively. If bias degree tends to zero (i.e., the absolute abundance calculated by the single spike is equivalent to the qPCR data), it means that the spike-based calculation of absolute abundance is theoretically feasible; on the contrary, it means that the spike-based method has bias in quantifying absolute microbial abundance.

To show the difference in OTUs relative abundance (RA) between spike-added samples and control, we calculated OTU abundance error (OA error) as follows:

$$OA\ error = Log_{10}\ RA\ (spike\text{-}added\ samples) - Log_{10}\ RA\ (control) \quad (6)$$

where RA (spike-added samples) and RA (control) represent the OTUs relative abundance in spike-added samples and no-spike samples, respectively. The OTUs relative abundance in spike-added samples was calculated after removing spike sequences. Only OTUs with relative abundance above 0.01% were selected for this comparison. To keep all values finite when working with a $\log_{10}$ scale, the zero relative abundance was mapped to 1/ (sequencing depth).

2.3. Statistical analysis

SPSS Statistics 23 (https://www.ibm.com/products/spss-statistics) was employed to perform statistical tests including one-way ANOVA, univariate analysis of variance and nonparametric test. Significance was set for $P < 0.05$. All pairs of comparisons between samples were assessed by post hoc Duncon's test. Corrections for multiple testing were performed using p.adjust function where applicable (Benjamini and Hochberg, 1995).

Spike-AMP. Differences in quantitative results between spike levels were assessed using univariate analysis of variance. One-way ANOVA was employed to analyze the differences in absolute gene abundance between the six samples. After deleting spike sequences from all samples, we performed permutational multivariate analysis of variance (PERMANOVA) based on Bray-Curtis distance matrices with 999 permutations to assess the effect of adding spikes on microbial β-diversity (Anderson, 2001). Canonical analysis of principal coordinates (CAP) was calculated using the capscale function in R (Anderson and Willis, 2003), by constraining for the variable of spike level. Bar plots were generated using GraphPad Prism 8. Box plots and heatmaps were generated using the ggplot2 and pheatmap packages in R, respectively. To evaluate the quantitative performance of a single spike at the OTUs level, we conducted linear regression analysis in GraphPad Prism 8 to determine the consistency of results provided by the single-spike-based and qPCR-based methods. To reduce rare OTUs in the data set, we only chose OTUs with mean absolute abundance above 10^2 copies g^{-1} soil when performing the linear regression analysis. To keep all values finite when working with a $\log_{10}$ scale, the zero absolute abundance was mapped to 1.

qPCR-AMP. The OTUs tables of both qPCR-AMP and SIP study were rarefied at 10000 sequences per sample for subsequent analysis according to the minimum reads number of samples. The sequencing data at three sampling times were merged given that the incubation time had no significant impact on both overall bacterial and fungal community structures according to ADONIS analysis (Fig. S4). Differences in top 10 phyla between control and straw-added soils were assessed using univariate analysis of variance. To assess the influence of AMP on the outcomes of differential OTU abundance analysis, we investigated the differentially abundant OTUs between control and straw-added soils by

using the edgeR package in R (Robinson et al., 2010). Co-occurrence networks based on the relative abundance (RMP networks) or absolute abundance (AMP networks) were reconstructed by performing OTUs Pearson correlation in the Molecular Ecological Network Analyses (MENA) pipeline (http://ieg4.rccc.ou.edu/mena/). The networks were graphed using Gephi (https://gephi.org/). To reduce rare OTUs in the data set, only OTUs with relative abundance above 0.01% that were detected in 75% of all soil samples were selected for network construction. To depict the topology of the AMP and RMP networks, a set of indexes including total nodes, total edges, average degree (avgK), betweenness centrality and modularity were characterized according to a previous study (Deng et al., 2012). Nonparametric tests (Kruskal-Wallis tests) were performed to evaluate the differences in avgK and betweenness centrality between control and straw-added soil networks. Module hubs were defined as those nodes with the degree value >10 in a network, and module hubs detected as unclassified genera were not displayed in heatmaps. All the sequence data in the present study have been deposited in the NCBI Sequence Read Archive (SRA) database under accession numbers SAMN19600335-SAMN19601257. Rmarkdown code to reproduce the results described in this paper is available at https://github.com/PlantNutrition/ZhangSBB.

3. Results

3.1. Section I: spike-AMP vs. qPCR-AMP

3.1.1. Determination of the total copies of 16S rRNA gene and ITS region

Considering the complexity of soil samples, two independent qPCR experiments were employed to quantify total abundances of bacterial 16S rRNA gene and fungal ITS region in each soil samples (i.e., Experiments 1 and 2). Overall, two independent qPCR experiments yielded the same trends in

quantitative results of 16S rRNA gene and ITS region. For example, the abundance of both genes increased with time in SYU soils (Fig. 2a, b). Further, qPCR exhibited good reproducibility between two independent experiments, with mean variation of 1.18- and 1.19-fold in 16S rRNA gene and ITS region abundances, respectively (Table S3). After combining the two independent qPCR results, 16S rRNA gene and ITS region abundances (per g of soil) among six samples broadly ranged from 1.98×10^{9} to 1.61×10^{10} and 3.03×10^{7} to 2.02×10^{9}, respectively (Table S3). The 16S rRNA gene abundances were 4.86–65.31 times higher than the ITS region abundances in six soils. On the basis of these results, the total copies of 16S rRNA gene and ITS region determined by qPCR were defined as the background values of six soils in further analysis.

3.1.2. Quantitative performance of spike-AMP

On the basis of the 16S rRNA gene and ITS region abundances estimated by qPCR, we designed eight bacterial spike levels and nine fungal spike levels, with the mean proportion of spike copies to total copies per PCR reaction ranging from approx. 0.01% to 97% (i.e., C_s/C_{total}) (Table 1). After sequencing, we calculated the ratio of spike reads to total sequencing reads (i.e., R_s/R_{total}) (Fig. 3a, b). There was approximately 10-fold gradient variation in C_s/C_{total} at spike levels 1–4 for both bacteria and fungi; however, the R_s/R_{total} did not show a consistent increase (Tables S4 and 5). For example, the C_s/C_{total} of HLP_Day7 at bacterial spike levels 2 and 3 was 0.04% and 0.44% respectively, but the R_s/R_{total} at these two spike levels were both 0.27% (Table S4). At the higher spike levels such as bacterial levels 6 and 7, the R_s/R_{total} (6.33% and 44.33%) was 5.85- and 1.84-fold lower than the C_s/C_{total} (37.06% and 81.73%), respectively (Table S4). These results indicated that there was no good correlation between the addition of spike copies and the output of spike reads.

299 A linear relationship between C_s and $\frac{R_s}{R_m}$ (Eq. (4)) is essential for spike-AMP, as shown by the
300 dashed line in Fig. 3c, d. However, our results revealed that the theoretical linear relationship only
301 partly occurred at bacterial spike levels 6–8 and fungal spike levels 4–6 (i.e., the blue rectangle in Fig.
302 3c, d), with the corresponding mean R_s/R_{total} of 6.33%–83.57% and 13.49%–54.13%, respectively
303 (Tables S4 and 5). Similar patterns also were confirmed in our preliminary survey (Fig. S5 and Table
304 S6). Furthermore, we found that the gene abundance estimated by the single spike method (Eq. (2))
305 was strongly correlated with spike level, even within the partial linear relationship between C_s and $\frac{R_s}{R_m}$
306 (Fig. 3e, f). For example, the estimated 16S rRNA gene abundances showed a notable decrease from
307 spike levels 6 to 8 (univariate analysis, P <0.0001; Fig. 3e and Table S7). In contrast, there was no
308 significant difference in ITS region abundances estimated by fungal spike levels 4, 5 and 6 (univariate
309 analysis, $P \geq 0.19$; Fig. 3f). These results demonstrated that the single-spike-based quantification was
310 strongly dependent on the amount of spike added and the objective gene abundance.
311 The gene abundance inferred by single spike inside the partial linear relationship (Eq. (2)) was
312 theoretically equal to that detected by linear relationship (Eq. (3)). However, we found that the 16S
313 rRNA gene abundances calculated by spike levels 6 and 7 were significantly higher than those
314 computed by linear relationship (univariate analysis, P <0.0001; Fig. 3e and Table S7). In comparison,
315 there were no noteworthy discrepancies in ITS region abundances estimated by spike levels 4–6 and by
316 linear relationship (univariate analysis, $P \geq 0.37$; Fig. 3f). Additionally, the variation trends between
317 samples calculated by the two spike-based methods (i.e., single-spike-based or linear-relationship-
318 based methods) were obviously different from those calculated by qPCR (Fig. 3e, f). For instance, both
319 spike-based methods revealed that 16S rRNA gene abundances in HLP_Day0 was remarkably lower
320 than that in HLP_Day30, but the qPCR results showed no significant differences between them (Fig.

3e). Collectively, these results indicated that even though a linear relationship between C_s and $\frac{R_s}{R_m}$ was found, the estimated gene abundance may not be reliable.

3.1.3. Comparison of OTUs absolute abundances calculated by spike-AMP and qPCR-AMP

By definition, when bias degree was closest to zero, the corresponding spike level was the best matching spike level (Eq. (5)). However, the best matching spike level was variable and dependent on the absolute abundances of original samples. In general, as the original abundance of 16S rRNA gene or ITS region increased, the best matching spike level increased (Fig. 4a, b). Focusing on the best matching spike level, we further compared the differences in OTUs absolute abundances estimated by spike-AMP or by qPCR-AMP (Fig. 4c, e). For example, these two methods produced consistent results when bacterial and fungal OTUs abundance were higher than 10^7 and 10^5 copies g^{-1} soil, respectively, in HLP_Day30 sample (Fig. 4d, f). However, this consistency remarkably decreased with declining OTUs abundance, especially when bacterial and fungal OTUs abundance were lower than 10^6 and 10^4 copies g^{-1} soil, respectively. Similar results were also found in the other five samples (Fig. S6–10). Therefore, there was potential error in the estimation of low-abundant OTUs, even when using the best matching spike level.

3.1.4. Effects of spike addition on the OTUs relative abundances and microbial β-diversity

To assess the effects of spike addition on the OTUs relative abundances, we compared the variation in OTUs relative abundances between control (i.e., no spike) and spike-added samples (i.e., the OA error defined by Eq. (6)). In both bacterial and fungal communities, 91.30%–99.66% of >0.5 OA error occurred in OTUs group with relative abundances below 0.1% (Fig. 5a, b), indicating that low-abundant OTUs were more susceptible to interference from spike addition. Specially, the larger error (e.g., OA error >2) was mainly detected in higher spike level samples, as indicated by the blue

343 "tail" in the scatter plots (Fig. 5a, b). Furthermore, CAP revealed a substantial effect of spike addition
344 on bacterial and fungal microbial community structure, explaining 16.1%–18.9% and 20.2%–36.2% of
345 overall structural variation, respectively (P = 0.001; Fig. 5c, e and Fig. S11). Saliently, the best
346 matching spike level in Fig. 4a and b also significantly changed the structures of bacterial and fungal
347 communities (i.e., the red border) (Fig. 5d, f).

348 *3.2. Section II: The utilization of qPCR-AMP in assessing soil microbial dynamics*

349 With the finding that spike-AMP method did not exhibit any advantages in accuracy, stability or
350 labor intensity, we then used the more common qPCR-AMP method to detect the absolute profiling of
351 soil bacterial and fungal communities in response to the addition of organic matter (i.e., maize straw).

352 *3.2.1. Differences between RMP and AMP in revealing key microbiota involved in straw decomposition*

353 We applied qPCR to a sequencing dataset from a 60-day incubation study. On average, the total
354 copies of 16S rRNA gene and ITS region in HL soils were 1.40 and 1.36 times higher than those in SY
355 soils, respectively (univariate analysis, P <0.0001; Fig. 6a, b). The 16S rRNA gene and ITS region
356 copies in SYU soils increased 1.03- and 4.38-fold, respectively, with straw addition (univariate analysis,
357 P <0.0001; Fig. 6a, b). Furthermore, we compared the differences in the top 10 phyla between straw-
358 added and control soils using RMP and AMP analysis. These two quantification methods produced
359 substantially different results, especially in SYU soils (Fig. 6c–f). Among these discrepancies, we
360 highlighted that *Actinobacteriota* was notably increased by straw addition in SYU soils when using
361 AMP, whereas this straw-induced increase was not observed when using RMP (univariate analysis, P
362 <0.0001; Fig. 6c, d). In addition, RMP seriously underestimated the changes in bacterial
363 *Proteobacteria* and fungal *Ascomycota* abundances. For example, the abundances of *Ascomycota*, a

dominant phylum in the fungal community, were significantly increased by straw addition in SYU soils, with average increases of 4.42-fold based on AMP compared with 0.11-fold based on RMP (Fig. 6e, f). Similar results were also found in paddy soil (Fig. S12). To validate whether the enriched phyla detected by AMP play crucial roles in straw decomposition, we performed a SIP experiment on the same samples to explore the ^{13}C-labeled communities. The results showed that *Proteobacteria*, *Actinobacteriota* and *Ascomycota* were dominant in ^{13}C-enriched phyla, accounting for 40.54%, 31.58% and 82.30% of total relative abundance in ^{13}C-labeled communities, respectively (Fig. 6d, f). These results suggested that the enriched phyla detected by qPCR-AMP were strongly associated with straw decomposition.

The compositions of straw-associated communities based on RMP and AMP analyses was obviously different over time. A far larger number of significantly enriched OTUs were detected by AMP (Fig. 6g and Fig. S12). When arranging these enriched OTUs according to their taxonomic information, we observed that the compositions of the enriched OTUs detected by AMP and RMP could not be overlapped (Fig. 6h, i). For example, the enriched bacterial OTUs belonging to *Planctomycetota*, *Myxococcota*, *Gemmatimonadota*, *Elusimicrobiota*, *Desulfobacterota*, *Chloroflexi*, *Armatimonadota* and *Acidobacteriota* were only detected by AMP, implying that key microbiota dynamics related to straw decomposition may be masked by RMP results.

3.2.2. Different patterns of co-occurrence network based on RMP and AMP

To assess the impact of quantitative information on the OTUs co-occurrence pattern, we reconstructed networks using both RMP and AMP data matrices. Multiple network topological indexes consistently showed that the OTU–OTU interactions were remarkably different between AMP-based and RMP-based networks (Table S8). In paddy soils (i.e., HLP and SYP), straw application resulted in

386 a simpler network in comparison with the control when using RMP data, whereas an inverse pattern
387 occurred when applying AMP data (Fig. 7a, b and Fig. S13). Taking the HLP networks as an example,
388 the degree and betweenness were averagely 1.24- and 9.01-fold lower in straw-added soils than in the
389 control, respectively, when using RMP data (nonparametric tests, P <0.0001), while this prominent
390 difference could not be observed when applying AMP data (Fig. 7c). Focusing on the straw-networks,
391 we observed that more connections were detected in AMP-based networks than in the RMP-based
392 networks (954 vs. 573) and there were only 237 common connections between these two networks,
393 indicating that the correlations detected by the two methods were quite different (Fig. 7d). Notably,
394 most of the unique links in AMP-based networks belonged to *Ascomycota*, *Proteobacteria* and
395 *Actinobacteriota*, all of which were associated with straw decomposition (Fig. 6d, f). Moreover, a set
396 of module hubs only detected by AMP in the straw-added network were the dominant ^{13}C-enriched
397 groups (i.e., the black box in Fig. 7e), accounting for 19.29% of ^{13}C-labeled communities. These results
398 indicated that these module hubs were involved in straw decomposition (Fig. 7e). Similar results were
399 obtained for the other soils tested (Fig. S13–15). Taken together, the data type determined by RMP or
400 AMP had a considerable impact on the co-occurrence network patterns.

401 **4. Discussion**

402 AMP is crucial in deciphering the variation in absolute microbial abundance between samples or
403 over time. For example, the absolute abundance of *Proteobacteria* increased remarkably in fertilized
404 soil when compared with the unfertilized soil, while this growth could not be identified by RMP (Jiang
405 et al., 2019). The growing interest in absolute abundance has led to methodological and technological
406 advances, such as spike-AMP (Tourlousse et al., 2017; Tkacz et al., 2018), qPCR-AMP (Lou et al.,
407 2018; Boshier et al., 2020) and FCM-AMP (Vandeputte et al., 2017; Vieira-Silva et al., 2019). The

spike-AMP, as an emerging quantification method, potentially offered some advantages over the traditional qPCR/FCM-AMP in some special studies. For example, spike-AMP may be an indispensable strategy for host-associated microbiome quantification, such as plant root microbiome (Guo et al., 2019). The host genome (chloroplasts and mitochondria) typically accounted for >80% of the 16S rRNA gene sequences in root microbiome samples (Bulgarelli et al., 2012; Lundberg et al., 2013). Universal primers, such as 799F/1193R for the 16S rRNA gene, cannot distinguish the sequences from root endophytes or plant genome, restricting accurate estimation of the absolute gene abundance by commonly used qPCR (Guo et al., 2019). Such issues can be solved by use of the spike-AMP approach. In addition, the operational procedures of spike-AMP are simple for a sample where optimal spike level was predetermined. Currently, the wide-ranging applicability of spike-AMP and qPCR-AMP having not yet been fully demonstrated for complex and diverse soil samples. In this study, we used a set of soil samples with dramatic differences in total gene abundances to evaluate the universal applicability of both spike-AMP and qPCR-AMP.

4.1. The evaluation of spike-AMP in soil microbiota quantitative research

In section I, we first evaluated the universal applicability of spike-AMP by applying a set of synthetic spikes with a gradient of concentrations to a set of soil samples with dramatic differences in 16S rRNA gene and ITS region (Table 1 and Fig. 2). We found that there was no good correlation between the input of spike copies and the output of spike sequencing reads (Tables S4 and 5), violating the theoretical assumption of the spike-in method (Jiang et al., 2011; Guo et al., 2019). This phenomenon was especially obvious at low spike levels, such as levels 1–4, which might be explained by the fact that low spike levels might remain undetected because of poor amplification (Reid and

Heathfield, 2020) or low sequencing depth. Because of this inconsistent input-output ratio, the 16S rRNA gene and ITS region abundance calculated by single spike seemed to be indeterminate and strongly correlated with spike addition (Fig. 3e, f). For example, the maximum differences in 16S rRNA gene and ITS region abundance estimated by single spike were up to 840- and 269-fold, respectively. Furthermore, the perfect input-output ratio of spike was extremely dependent on the targeted gene abundance in the original sample (Fig. 4a, b). However, large spatio-temporal variations in microbial population abundance have been reported (Hallam and McCutcheon, 2015; Leach et al., 2017), which means that the optimal spike level for different samples may be variable. Therefore, the single-spike-based method needs a preliminary test to determine the optimal spike level for each given environmental sample, which substantially increases the time and workload when using this method.

A spike mixture (e.g., spike-in method #3 in Fig. 1d) (Tourlousse et al., 2017; Jiang et al., 2019; Mou et al., 2020) was used to circumvent the flaws of the single-spike-based method by constructing a linear regression between input of spike amount and output of sequencing reads. We found that the linear relationship was partially obtained, with corresponding bacterial and fungal spike reads accounting for 6.33%–83.57% and 13.49%–54.13% of total sequencing reads, respectively (Fig. 3c, d). In a soil microbiota research, Tkacz et al. (2018) showed that the optimum spike amount for 16S rRNA gene, 18S rRNA gene and ITS region should account for 20%–80% of total sequencing reads. In a plant study, Guo et al. (2019) provided evidence that the coverage of spike concentration should be 10%–60% in the amplicon library. These discrepancies could be because the linear interval may be related to the nature of complex environmental samples, such as microbial population size. Furthermore, the quantitative results calculated even by linear-relationship-based method were also

questionable. As shown in HLP_Day30, the 16S rRNA gene abundance calculated by linear interval (levels 6–8) was 336.05% higher than the qPCR abundance (Fig. 3e). This deviation from the expected result may be partly attributed to the skew proportions of spike reads in sequencing data, which may occur during multiple steps of the spike-AMP process, such as mixing inaccuracies, PCR bias or sequencing error (Tourlousse et al., 2017). Meanwhile, spike addition consumed substantial sequencing resources. For example, the bacterial spike at level 8 accounted for 77.02%–98.02% of total sequencing reads in each soil (Fig. 3a), which makes the process infeasible in an actual study. More importantly, spike addition significantly changed the original community structure (Fig. 5d, f). The relative abundance of OTUs changed from the original values, especially for low-abundant OTUs (Fig. 5a, b). For instance, deviations of more than 3.16-fold (i.e., OA error=0.5) mostly occurred in groups with OTUs relative abundances below 0.1%. Thus, our results suggested that spike-AMP may not be suitable for quantifying absolute abundance of soil microbiota, at least for those samples with distinct microbial characteristics.

4.2. The verification of AMP in soil microbiota quantitative research

We also applied qPCR-AMP to soil samples from a 60-day incubation study to analyze dynamic changes in microbial profiles. We found that microbial abundance in SYU soils showed a drastic response to straw addition (mean 1.03-fold and 4.38-fold in 16S rRNA gene and ITS region, respectively), implying that the variation in total absolute abundance possibly represents a key feature of microbiota in response to environmental disturbance (Vandeputte et al., 2017; Guo et al., 2019; Jiang et al., 2019). For instance, the genuine increases in abundance of bacterial phyla *Proteobacteria* and *Actinobacteriota*, and fungal phyla *Ascomycota* were only detected by qPCR-AMP in straw-added soil, and these three phyla were further shown to be highly correlated with straw decomposition in the

SIP experiment (Fig. 6c–f and Fig. S6c–f). These findings were in line with previous studies that many sub-groups belonging to these three phyla were involved in assimilation of carbon from plant residues (Lee et al., 2011; Fan et al., 2014; Zhao et al., 2019). Therefore, AMP allowed us to identify the dominant species involved in straw decomposition. Without the information on absolute abundance, the underlying physiology and ecological responses of specific phyla to organic matter addition may be masked by relative abundance.

Inappropriate data types and statistical methods can lead to spurious results or hide useful information when applied to compositional analysis of sequencing data (Vandeputte et al., 2017; Carr et al., 2019). We found that the overlap of OTU–OTU links between RMP-based and AMP-based networks was only 13.7% in this study. More unique connections detected by AMP belonged to the ^{13}C-dominant phyla (Fig. 7c), indicating that AMP can better reflect the impacts of straw addition on microbial communities. Previous studies suggested that the module hubs may play a critical role in maintaining the structure and function of ecological communities (Jiao et al., 2016; Shi et al., 2016). Indeed, most of the network module hubs in AMP-based straw-networks were related to straw decomposition (Fig. 7e). For example, the ^{13}C-enriched module hubs (genus level), such as *Intrasporangium*, *Arenimonas*, *Cellvibrio* and *Gibberella*, were only detected by AMP-based networks and their organic matter degradation ability has been observed for lignocellulose and plant residues (DeBoy et al., 2008; Cai et al., 2018; Song et al., 2018; Zhan et al., 2021). Additionally, some of the module hubs, such as *Pseudarthrobacter* and *Micromonospora*, were shown to be important participants during straw decomposition in this study, although they have not previously been reported to be responsible for straw degradation. However, these details of microbial dynamic changes and microbiota interactions may be ignored or misinterpreted when using RMP-based network analysis.

5. Conclusion

In this study, we highlight several prominent issues of spike-AMP including reliability, stability and labor intensity, all of which stymie the universal feasibility of spike-AMP in soil microbiota quantitative research. This is because neither the potentially optimal spike level is determined, nor can the straight line between C_s and $\frac{R_s}{R_m}$ be fixed. In contrast to spike-AMP, the commonly used qPCR-AMP provides a straightforward and high-throughput tool for quantifying absolute profiling of soil microbiota. Therefore, in the absence of a gold-standard quantitative approach, qPCR-AMP may be the preferred method in soil microbial research. However, in special ecological niches, such as root endophytes, the existence of plant plastids (mitochondria and chloroplasts) prevents the accurate detection of total microbial load (Guo et al., 2019). Further research is required to explore the implementation of AMP in these environments.

6. Acknowledgements

This research was financially supported by the National Natural Science Foundation of China (31601829), the Earmarked Fund for China Agriculture Research System (No. CARS-01-24), and Young Elite Scientists Sponsorship Program by CAST (2017QNRC001). We thank Catherine Dandie, PhD for editing the English text of a draft of this manuscript. We are grateful to two anonymous reviewers for their constructive comments and suggestions on this paper.

References

Abarenkov, K., Henrik Nilsson, R., Larsson, K.H., Alexander, I.J., Eberhardt, U., Erland, S., Høiland, K., Kjøller, R., Larsson, E., Pennanen, T., 2010. The UNITE database for molecular identification of fungi–recent updates and future perspectives. New Phytologist 186, 281-285.

515 Anderson, M.J., 2001. A new method for non-parametric multivariate analysis of variance. Austral
516 Ecology 26, 32-46.
517 Anderson, M.J., Willis, T.J., 2003. Canonical analysis of principal coordinates: A useful method of
518 constrained ordination for ecology. Ecology 84, 511-525.
519 Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful
520 approach to multiple testing. Journal of the Royal statistical society: series B (Methodological) 57, 289-
521 300.
522 Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C., Al-Ghalith, G.A., Alexander, H.,
523 Alm, E.J., Arumugam, M., Asnicar, F., 2019. Reproducible, interactive, scalable and extensible
524 microbiome data science using QIIME 2. Nature Biotechnology 37, 852-857.
525 Boshier, F.A.T., Srinivasan, S., Lopez, A., Hoffman, N.G., Proll, S., Fredricks, D.N., Schiffer, J.T.,
526 Caporaso, J.G., 2020. Complementing 16S rRNA gene amplicon sequencing with total bacterial load to
527 infer absolute species concentrations in the vaginal microbiome. mSystems 5, e00777-00719.
528 Bulgarelli, D., Rott, M., Schlaeppi, K., Ver Loren van Themaat, E., Ahmadinejad, N., Assenza, F.,
529 Rauf, P., Huettel, B., Reinhardt, R., Schmelzer, E., Peplies, J., Gloeckner, F.O., Amann, R., Eickhorst,
530 T., Schulze-Lefert, P., 2012. Revealing structure and assembly cues for Arabidopsis root-inhabiting
531 bacterial microbiota. Nature 488, 91-95.
532 Cai, L., Chen, T.B., Zheng, S.W., Liu, H.T., Zheng, G.D., 2018. Decomposition of lignocellulose and
533 readily degradable carbohydrates during sewage sludge biodrying, insights of the potential role of
534 microorganisms from a metagenomic analysis. Chemosphere 201, 127-136.

535 Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N.,
536 Pena, A.G., Goodrich, J.K., Gordon, J.I., 2010. QIIME allows analysis of high-throughput community
537 sequencing data. Nature Methods 7, 335.
538 Carr, A., Diener, C., Baliga, N.S., Gibbons, S.M., 2019. Use and abuse of correlation analyses in
539 microbial ecology. The ISME Journal 13, 2647-2655.
540 Charpentier, M., Oldroyd, G., 2010. How close are we to nitrogen-fixing cereals? Current Opinion in
541 Plant Biology 13, 556-564.
542 Crowther, T.W., van den Hoogen, J., Wan, J., Mayes, M.A., Keiser, A.D., Mo, L., Averill, C., Maynard,
543 D.S., 2019. The global soil community and its influence on biogeochemistry. Science 365, eaav0550.
544 DeBoy, R.T., Mongodin, E.F., Fouts, D.E., Tailford, L.E., Khouri, H., Emerson, J.B., Mohamoud, Y.,
545 Watkins, K., Henrissat, B., Gilbert, H.J., Nelson, K.E., 2008. Insights into plant cell wall degradation
546 from the genome sequence of the soil bacterium Cellvibrio japonicus. Journal of Bacteriology 190,
547 5455-5463.
548 Deng, Y., Jiang, Y.-H., Yang, Y., He, Z., Luo, F., Zhou, J., 2012. Molecular ecological network
549 analyses. BMC Bioinformatics 13, 113.
550 Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26,
551 2460-2461.
552 Fan, F., Yin, C., Tang, Y., Li, Z., Song, A., Wakelin, S.A., Zou, J., Liang, Y., 2014. Probing potential
553 microbial coupling of carbon and nitrogen cycling during decomposition of maize residue by 13C-
554 DNA-SIP. Soil Biology and Biochemistry 70, 12-21.
555 Fan, Y., Pedersen, O., 2021. Gut microbiota in human metabolic health and disease. Nature Reviews
556 Microbiology 19, 55-71.

557 Gao, R., Sun, C., 2020. A marine bacterial community that degrades poly(ethylene terephthalate) and
558 polyethylene. bioRxiv.
559 Guo, X., Zhang, X., Qin, Y., Liu, Y.-X., Zhang, J., Zhang, N., Wu, K., Qu, B., He, Z., Wang, X.,
560 Zhang, X., Hacquard, S., Fu, X., Bai, Y., 2019. Host-associated quantitative abundance profiling
561 reveals the microbial load variation of root microbiome. Plant Communications 1, 100003.
562 Hallam, S.J., McCutcheon, J.P., 2015. Microbes don't play solitaire: how cooperation trumps isolation
563 in the microbial world. Environmental Microbiology Reports 7, 26-28.
564 Jiang, L., Schlesinger, F., Davis, C.A., Zhang, Y., Li, R., Salit, M., Gingeras, T.R., Oliver, B., 2011.
565 Synthetic spike-in standards for RNA-seq experiments. Genome Research 21, 1543-1551.
566 Jiang, S.Q., Yu, Y.N., Gao, R.W., Wang, H., Zhang, J., Li, R., Long, X.H., Shen, Q.R., Chen, W., Cai,
567 F., 2019. High-throughput absolute quantification sequencing reveals the effect of different fertilizer
568 applications on bacterial community in a tomato cultivated coastal saline soil. Science of the Total
569 Environment 687, 601-609.
570 Jiao, S., Liu, Z., Lin, Y., Yang, J., Chen, W., Wei, G., 2016. Bacterial communities in oil contaminated
571 soils: Biogeography and co-occurrence patterns. Soil Biology and Biochemistry 98, 64-73.
572 Kwak, M.J., Kong, H.G., Choi, K., Kwon, S.K., Song, J.Y., Lee, J., Lee, P.A., Choi, S.Y., Seo, M., Lee,
573 H.J., Jung, E.J., Park, H., Roy, N., Kim, H., Lee, M.M., Rubin, E.M., Lee, S.W., Kim, J.F., 2018.
574 Rhizosphere microbiome structure alters to enable wilt resistance in tomato. Nature Biotechnology 36,
575 1100-1109.
576 Leach, J.E., Triplett, L.R., Argueso, C.T., Trivedi, P., 2017. Communication in the Phytobiome. Cell
577 169, 587-596.

578 Lee, C., Kim, J., Shin, S.G., Hwang, S., 2006. Absolute and relative QPCR quantification of plasmid
579 copy number in Escherichia coli. Journal of Biotechnology 123, 273-280.
580 Lee, C.G., Watanabe, T., Sato, Y., Murase, J., Asakawa, S., Kimura, M., 2011. Bacterial populations
581 assimilating carbon from 13C-labeled plant residue in soil: Analysis by a DNA-SIP approach. Soil
582 Biology and Biochemistry 43, 814-822.
583 Lou, J., Yang, L., Wang, H., Wu, L., Xu, J., 2018. Assessing soil bacterial community and dynamics by
584 integrated high-throughput absolute abundance quantification. PeerJ 6, 4514.
585 Lundberg, D.S., Yourstone, S., Mieczkowski, P., Jones, C.D., Dangl, J.L., 2013. Practical innovations
586 for high-throughput amplicon sequencing. Nature Methods 10, 999-1002.
587 Mou, J., Li, Q., Shi, W., Qi, X., Song, W., Yang, J., 2020. Chain conformation, physicochemical
588 properties of fucosylated chondroitin sulfate from sea cucumber Stichopus chloronotus and its in vitro
589 fermentation by human gut microbiota. Carbohydrate Polymers 228, 115359.
590 Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O., 2012.
591 The SILVA ribosomal RNA gene database project: improved data processing and web-based tools.
592 Nucleic Acids Research 41, D590-D596.
593 Rao, C., Coyte, K.Z., Bainter, W., Geha, R.S., Martin, C.R., Rakoff-Nahoum, S., 2021. Multi-kingdom
594 ecological drivers of microbiota assembly in preterm infants. Nature 591, 633-638.
595 Reid, K.M., Heathfield, L.J., 2020. Evaluation of direct PCR for routine DNA profiling of non-
596 decomposed deceased individuals. Science & Justice 60, 567-572.
597 Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR: a Bioconductor package for differential
598 expression analysis of digital gene expression data. Bioinformatics 26, 139-140.

599 Rognes, T., Flouri, T., Nichols, B., Quince, C., Mahé, F., 2016. VSEARCH: a versatile open source
600 tool for metagenomics. PeerJ 4, e2584.
601 Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A.,
602 Oakley, B.B., Parks, D.H., Robinson, C.J., Sahl, J.W., Stres, B., Thallinger, G.G., Van Horn, D.J.,
603 Weber, C.F., 2009. Introducing mothur: open-source, platform-independent, community-supported
604 software for describing and comparing microbial communities. Applied and Environmental
605 Microbiology 75, 7537-7541.
606 Shi, S., Nuccio, E.E., Shi, Z.J., He, Z., Zhou, J., Firestone, M.K., Johnson, N., 2016. The
607 interconnected rhizosphere: High network complexity dominates rhizosphere assemblages. Ecology
608 Letters 19, 926-936.
609 Smets, W., Leff, J.W., Bradford, M.A., McCulley, R.L., Lebeer, S., Fierer, N., 2016. A method for
610 simultaneous measurement of soil bacterial abundances and community composition via 16S rRNA
611 gene sequencing. Soil Biology and Biochemistry 96, 145-151.
612 Song, Y., Kong, Y., Wang, J., Ruan, Y., Huang, Q., Ling, N., Shen, Q., 2018. Identification of the
613 produced volatile organic compounds and the involved soil bacteria during decomposition of
614 watermelon plant residues in a Fusarium-infested soil. Geoderma 315, 178-187.
615 Stammler, F., Glasner, J., Hiergeist, A., Holler, E., Weber, D., Oefner, P.J., Gessner, A., Spang, R.,
616 2016. Adjusting microbiome profiles for differences in microbial load by spike-in bacteria.
617 Microbiome 4, 28.
618 Tkacz, A., Hortala, M., Poole, P.S., 2018. Absolute quantitation of microbiota abundance in
619 environmental samples. Microbiome 6, 110.

620 Tourlousse, D.M., Yoshiike, S., Ohashi, A., Matsukura, S., Noda, N., Sekiguchi, Y., 2017. Synthetic
621 spike-in standards for high-throughput 16S rRNA gene amplicon sequencing. Nucleic Acids Research
622 45, e23.
623 Vandeputte, D., Kathagen, G., D'Hoe, K., Vieira-Silva, S., Valles-Colomer, M., Sabino, J., Wang, J.,
624 Tito, R.Y., De Commer, L., Darzi, Y., Vermeire, S., Falony, G., Raes, J., 2017. Quantitative
625 microbiome profiling links gut community variation to microbial load. Nature 551, 507-511.
626 Vieira-Silva, S., Sabino, J., Valles-Colomer, M., Falony, G., Kathagen, G., Caenepeel, C., Cleynen, I.,
627 van der Merwe, S., Vermeire, S., Raes, J., 2019. Quantitative microbiome profiling disentangles
628 inflammation- and bile duct obstruction-associated microbiota alterations across PSC/IBD diagnoses.
629 Nature Microbiology 4, 1826-1831.
630 White, R.A., Callister, S.J., Moore, R.J., Baker, E.S., Jansson, J.K., 2016. The past, present and future
631 of microbiome analyses. Nature Protocols 11, 2049-2053.
632 Yang, L., Lou, J., Wang, H., Wu, L., Xu, J., 2018. Use of an improved high-throughput absolute
633 abundance quantification method to characterize soil bacterial community and dynamics. Science of
634 the Total Environment 633, 360-371.
635 Zemb, O., Achard, C.S., Hamelin, J., De Almeida, M.L., Gabinaud, B., Cauquil, L., Verschuren,
636 L.M.G., Godon, J.J., 2020. Absolute quantitation of microbes using 16S rRNA gene metabarcoding: A
637 rapid normalization of relative abundances by quantitative PCR targeting a 16S rRNA gene spike-in
638 standard. Microbiologyopen 9, e977.
639 Zhan, P., Liu, Y., Wang, H., Wang, C., Xia, M., Wang, N., Cui, W., Xiao, D., Wang, H., 2021. Plant
640 litter decomposition in wetlands is closely associated with phyllospheric fungi as revealed by microbial
641 community dynamics and co-occurrence network. Science of the Total Environment 753, 142194.

642 Zhang, Z., Qu, Y., Li, S., Feng, K., Wang, S., Cai, W., Liang, Y., Li, H., Xu, M., Yin, H., Deng, Y.,
643 2017. Soil bacterial quantification approaches coupling with relative abundances reflecting the changes
644 of taxa. Scientific Reports 7, 4837.
645 Zhao, S., Qiu, S., Xu, X., Ciampitti, I.A., Zhang, S., He, P., 2019. Change in straw decomposition rate
646 and soil microbial community composition after straw addition in different long-term fertilization soils.
647 Applied Soil Ecology 138, 123-133.

648

649 **Figure legends**

650 **Fig. 1 A review of absolute microbiome profiling (AMP). a–f** Two main types of AMP: the spike-in
651 method (spike-AMP; **a–d**, **f**) and quantitative PCR (qPCR) combined with high-throughput sequencing
652 (qPCR-AMP; **e–f**). Spike-AMP includes several key steps: (1) The design of a synthetic spike
653 containing synthetic sequences with negligible identity to known nucleotide sequences in public
654 databases and primer binding sites (e.g., 515F/806R targeting V4 variable region of the bacterial 16S
655 rRNA gene, ITS1F/ITS2 targeting variable region of the fungal ITS region or a combination of primer
656 regions) (**a**). (2) A known amount of synthetic spike is added to environmental samples or DNA
657 extracts in the form of a single spike with certain concentration (spike-in method #1, **b**), a single spike
658 with different gradient concentrations (spike-in method #2, **c**), or a spike mixture with different spikes
659 and concentrations (spike-in method #3, **d**). (3) Environmental samples and synthetic spike are co-
660 amplified and co-sequenced. (4–5) The absolute abundance of each taxon can be calculated based on
661 the relationship between input of spike copies and output of sequence reads (equations in red box) (**a–d**,
662 **f**). In general, the optimal spike concentration is determined by preliminary tests using spike-in method

#2 (**c**), and then spike-in method #1 (**b**) is used to calculate the absolute abundance of each taxon. The equations in the "blue dashed box" were constructed in this study (**c**). For detailed description, please see the main text.

Fig. 2 The total copies of 16S rRNA gene and ITS region detected by quantitative PCR (qPCR) in six soil samples. The test soil samples included three time points (days 0, 7 and 30) in two soil types (Hailun paddy (HLP) and Sanya upland (SYU)). **a–b** The copy numbers of bacterial 16S rRNA gene (**a**) and fungal ITS region (**b**) were quantified by qPCR. The gene abundance of each sample was reproduced in two independent qPCR experiments (i.e., Experiments 1 and 2). Each dot represents a technical PCR replicate and error bar represents the standard deviation. Different uppercase and lowercase letters indicate significant difference between the six soil samples in Experiments 1 and 2, respectively.

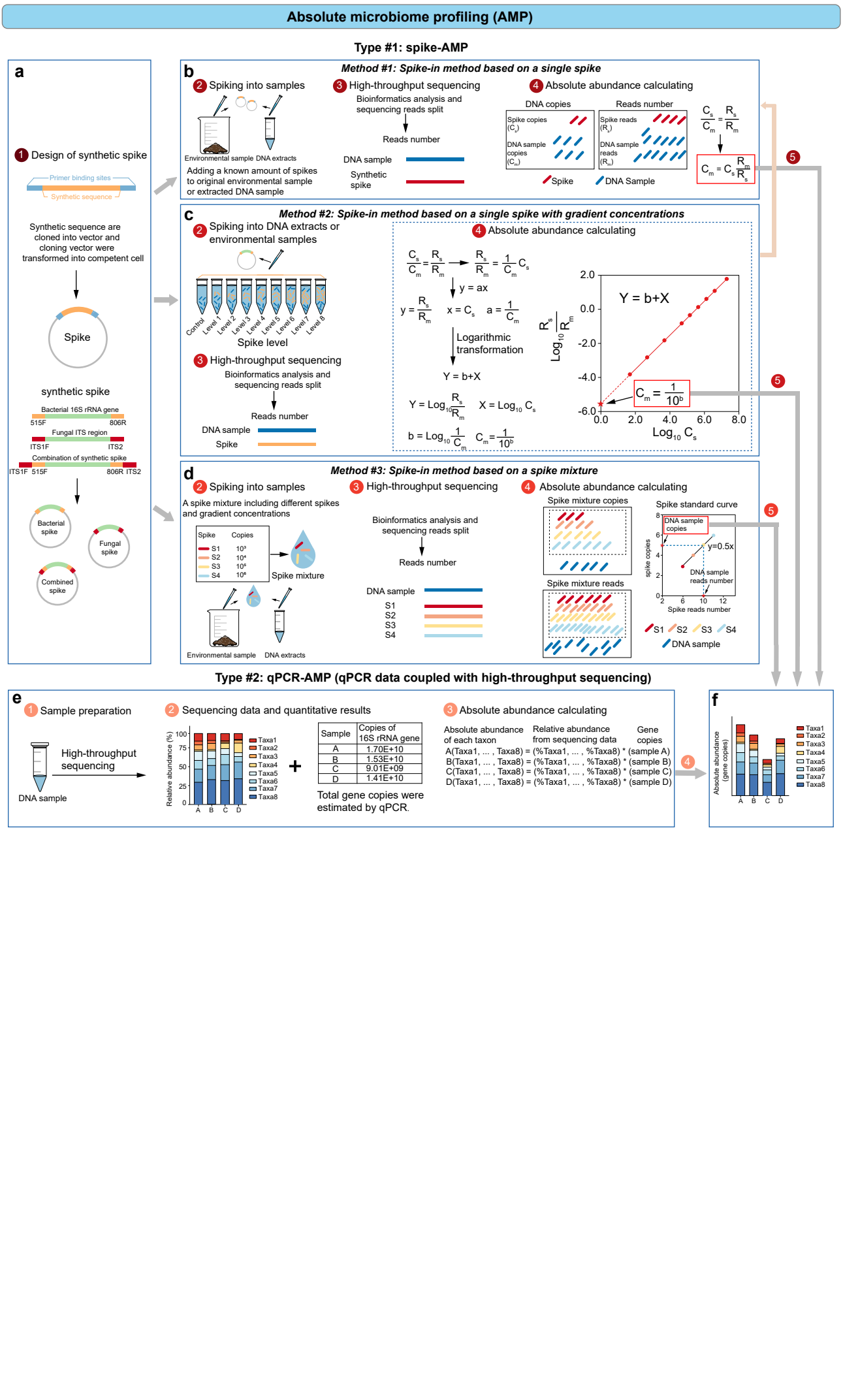
Fig. 3 Application of spike-in method #2 based on a gradient of bacterial and fungal single spike. **a–b** The proportion of spike reads in total sequencing reads at each spike level (i.e., R_s/R_{total}). **c–d** The relationship between the spike copies (C_s) and the ratio of spike reads to DNA sample reads ($\frac{R_s}{R_m}$), as described in Eqs. (3–4). In theory, there is a straight line between C_s and $\frac{R_s}{R_m}$, which is shown by a dashed line. For measured data, the range of spike levels marked by the blue rectangle represents a partial linear response of $\frac{R_s}{R_m}$ to C_s. Data in **c** and **d** are pooled from two independent experiments (i.e., Preliminary test and spike-AMP). **e–f** Bar plots showing estimated absolute abundance of bacterial 16S rRNA gene (**e**) and fungal ITS region (**f**) by using single spike (Eq. (2), $C_m=C_s\ \frac{R_m}{R_s}$), spike linear relationship (Eq. (3), $C_m=\ \frac{1}{a}$) and qPCR results (as shown in Fig. 2), respectively. The range of spike

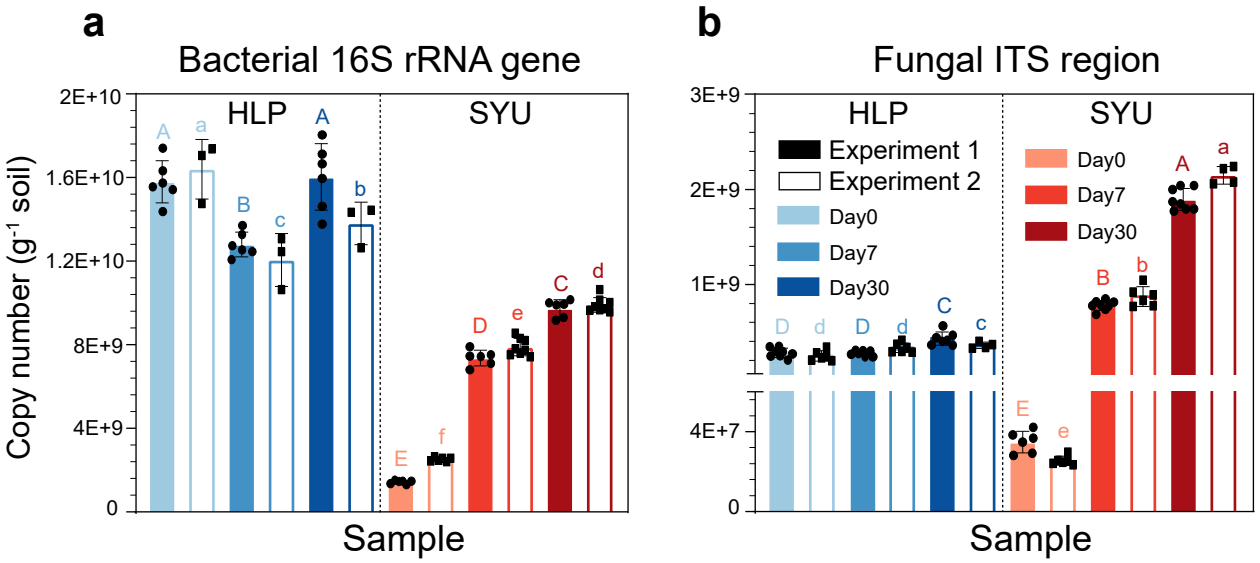
685 levels marked by the blue background represents a partial linear response of $\frac{R_s}{R_m}$ to C_s as shown in **c–d**.
686 Differences in absolute abundances between the six tested soil were analyzed using one-way ANOVA
687 followed by post hoc Duncon's multiple comparisons test. Different letters indicate significant
688 difference among six soil samples. Each dot represents a technical replicate and error bars represent
689 standard deviation (In spike-AMP, n=6 technical replicates for each bar). Note that the results from
690 preliminary tests are shown in Fig. S5 and Table S6.
691
692 **Fig. 4 Comparison of operational taxonomic units (OTUs) absolute abundances (AAs) calculated**
693 **by qPCR and single spike. a–b** Heatmaps displaying the bias degree (Eq. (5)) and bar plots showing
694 the copy numbers of 16S rRNA gene and ITS region estimated by qPCR in the tested samples (as
695 shown in Fig. 2). The bias degree with absolute values >0.5 is marked in white. Black borders around
696 cells indicate that the bias degree is lowest in each sample (i.e., the best matching spike level). **c**, **e**
697 Taking HLP_Day30 as an example, scatter plots showing bacterial (**c**) and fungal (**e**) OTUs AAs
698 calculated by qPCR data (x-axis) and each single spike data (y-axis). Color scale indicates spike level
699 and each point represents an OTU. Bacterial OTUs and fungal OTUs with AAs < 1×10^2 copies g^{-1} soil
700 are not displayed. The solid black lines (i.e., the y=x line) suggest that the AAs calculated by single
701 spike data are equivalent to those by qPCR data. **d**, **f** In accordance with qPCR data, the OTU AAs are
702 divided into three different abundance levels (i.e., AA $>10^7$, $10^6\leq$ AA $\leq10^7$ and AA $<10^6$ for bacteria;
703 AA $>10^5$, $10^4\leq$ AA $\leq10^5$ and AA $<10^4$ for fungi). For the best matching spike level (i.e., Level 3 for
704 bacteria (**d**) and Level 6 for fungi (**f**)), the difference between OTU AAs calculated by qPCR and single
705 spike is further shown at the three different abundance levels. Gray dotted lines represent the fitted
706 linear regressions. Note that all results from remaining soil samples are shown in Fig. S6–10.

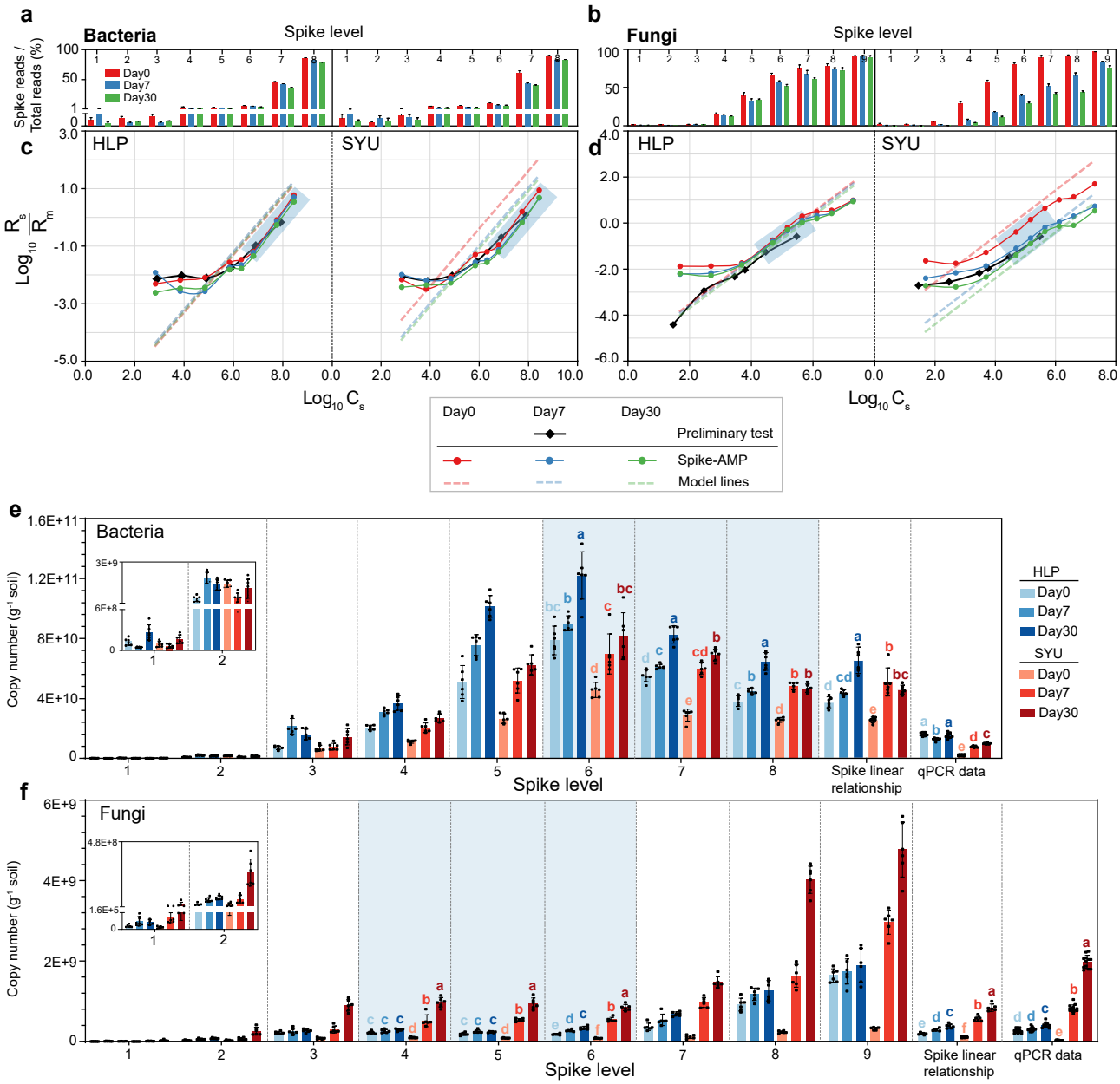
Fig. 5 Effects of spike addition on the operational taxonomic units (OTUs) relative abundances and microbial β-diversity. a–b The OTUs relative abundances (RAs) of bacteria (**a**) and fungi (**b**) in control soils (i.e., no spike) are plotted against the OTU abundance error (OA error), as described in Eq. (6). Color scale represents spike levels and each point corresponds to an OTU. OTUs with OA error >0.5 are indicated with solid red lines (i.e., the variations in OTU RAs between spike-added samples and control are 3.16-fold). All spike sequences have been deleted in the calculation of relative abundance in spike-added samples. The dotted black line divides OTU RAs into three groups (i.e., RA >1%, 0.1%≤ RA ≤1% and RA <0.1%), and the occurrence of >0.5 OA error in each RA group is marked with percentages. **c**, **e** Canonical analysis of principal coordinates (CAP) was performed based on bacterial (**c**) and fungal (**e**) Bray-Curtis distance matrices by constraining for the variable of spike level. Each point corresponds to a different sample colored by spike level. **d**, **f** The effects of adding spikes on β-diversity were assessed by permutational multivariate analysis of variance (PERMANOVA). Note that all spike sequences have been deleted when performing CAP and PERMANOVA. White cells indicate that there are no significant changes in community structure between spike-added samples and control, whereas light (ADONIS, P <0.05) and dark (ADONIS, P <0.01) blue cells represent that adding spike caused a significant change in bacterial (**d**) or fungal (**f**) community structure. Red borders around cells indicate that the absolute abundance calculated by single spike was closest to those calculated by qPCR (i.e., the best matching spike level), as shown in Fig. 4a, b. The CAP plots of SYU soils are shown in Fig. S11.

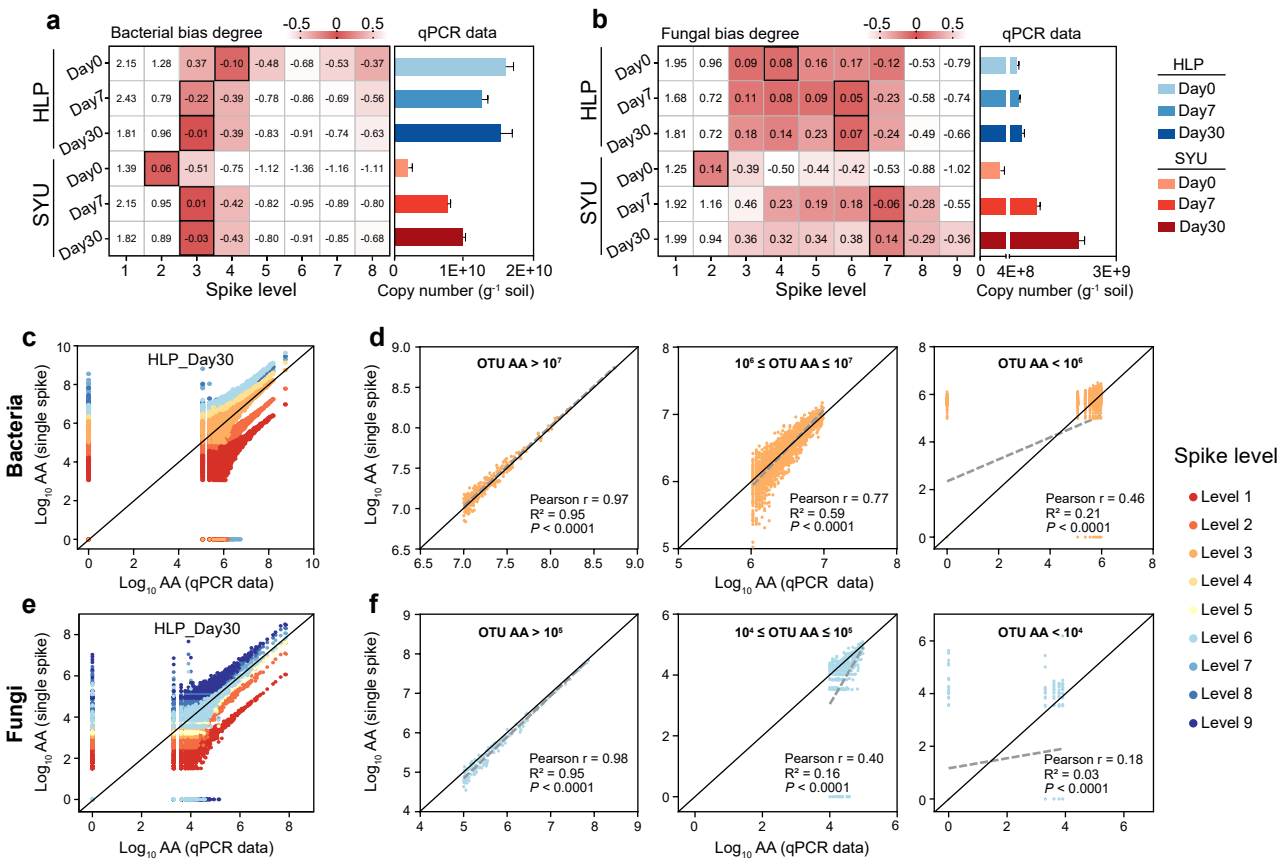
728 **Fig. 6 Influence of straw addition on soil microbial community structure.** Taking upland soils
729 (Hailun upland (HLU) and Sanya upland (SYU)) as an example, the impacts of straw addition on
730 bacterial and fungal community structure were assessed by using both relative microbiome profiling
731 (RMP) and qPCR-based absolute microbiome profiling (qPCR-AMP). We merged sequencing data
732 across all three time points and mainly focused on the comparison between control and straw-added
733 soils in **a–f**. **a–b** Box plots showing the copies of 16S rRNA gene (**a**) and ITS region (**b**) in Hailun (HL)
734 and Sanya (SY) soils. The color scales of green represent time points. The horizontal bold lines within
735 boxes represent medians. The top and bottoms of boxes indicate the 75th and 25th percentiles,
736 respectively. Univariate analysis of variance was performed to evaluate the differences in gene
737 abundance, ns, no significant difference, **** indicates P <0.0001. **c**, **e** Bacterial (**c**) and fungal (**e**)
738 phylum-level community composition determined by RMP and qPCR**-**AMP. **d**, **f** The significant
739 differences in phyla abundance between control (i.e., no straw addition) and straw-added soils were
740 assessed using univariate analysis of variance followed by post hoc Duncon's multiple comparisons
741 test (left panel). The bar plots (right panel) show the relative abundance of bacterial (**d**) and fungal (**f**)
742 phyla in ^{13}C-labeled communities (stable isotope probing (SIP) experiment). **g** A comparison of
743 numbers of enriched OTUs in straw-added soils between RMP and qPCR-AMP at each time point. **h–i**
744 Taxonomic distribution of straw-enriched OTUs at bacterial phylum level (**h**) and fungal class level (**i**).
745 The red and blue circle sizes represent the relative abundance and absolute abundance (on a $\log_{10}$ scale),
746 respectively. All results from paddy soils are shown in Fig. S12.
747
748 **Fig. 7 Co-occurrence networks of operational taxonomic units (OTUs) in control and straw-**
749 **added soils. a–b** Taking Hailun paddy soil (HLP) as an example, visualization of bacterial and fungal

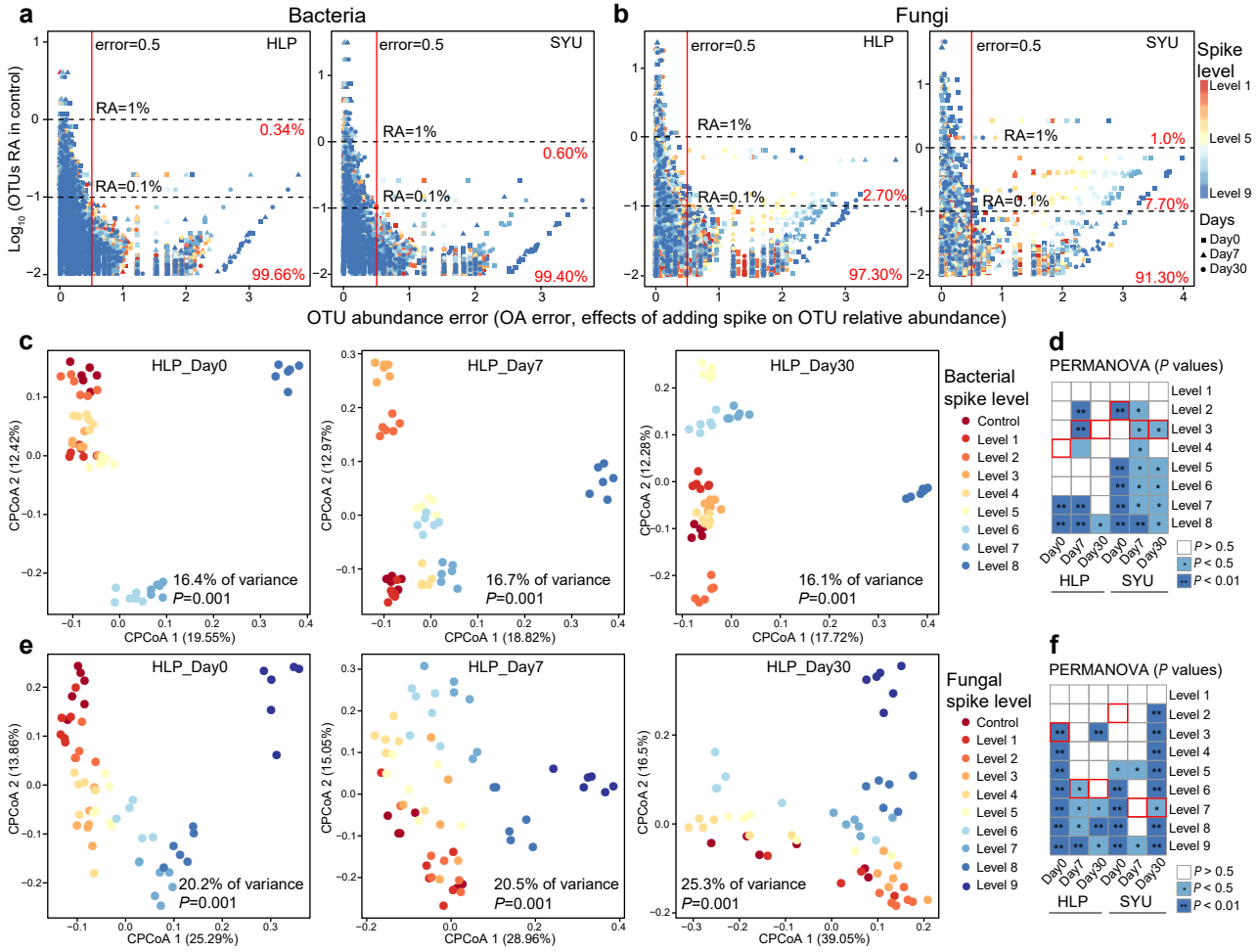
OTUs interactions by using relative microbiome profiling (RMP) (**a**) and qPCR-based absolute microbiome profiling (qPCR-AMP) (**b**). Larger modules with nodes >10 are labeled with different colors, and smaller modules are shown in gray. The nodes represent individual OTUs and node size corresponds to their abundance. Topological features of each network are listed in Table S8. **c** Degree and betweenness centrality (on a $\log_{10}$ scale) of nodes in control and straw-added soil networks detected by RMP and AMP. Nonparametric tests (Kruskal-Wallis tests) were performed to evaluate the differences in the two topological indexes between control and straw-added soil networks. *** indicates P <0.001, ns, no significant difference. **d** Venn plot showing the difference between links in the straw-added soil network between RMP and AMP. Bar plot showing the taxonomic distribution of nodes that belong to the unique links detected by RMP and AMP, respectively. **e** Heatmap showing the taxonomic distribution of module hubs detected in **a** and **b** networks. The barplot (right panel) showing the relative abundances of module hubs in ^{13}C-labeled communities (stable isotope probing (SIP) experiment). The genera in the black boxes represent module hubs that were only found in the straw-added network based on AMP. Network plots of the other three soil types are shown in Fig. S13–15.

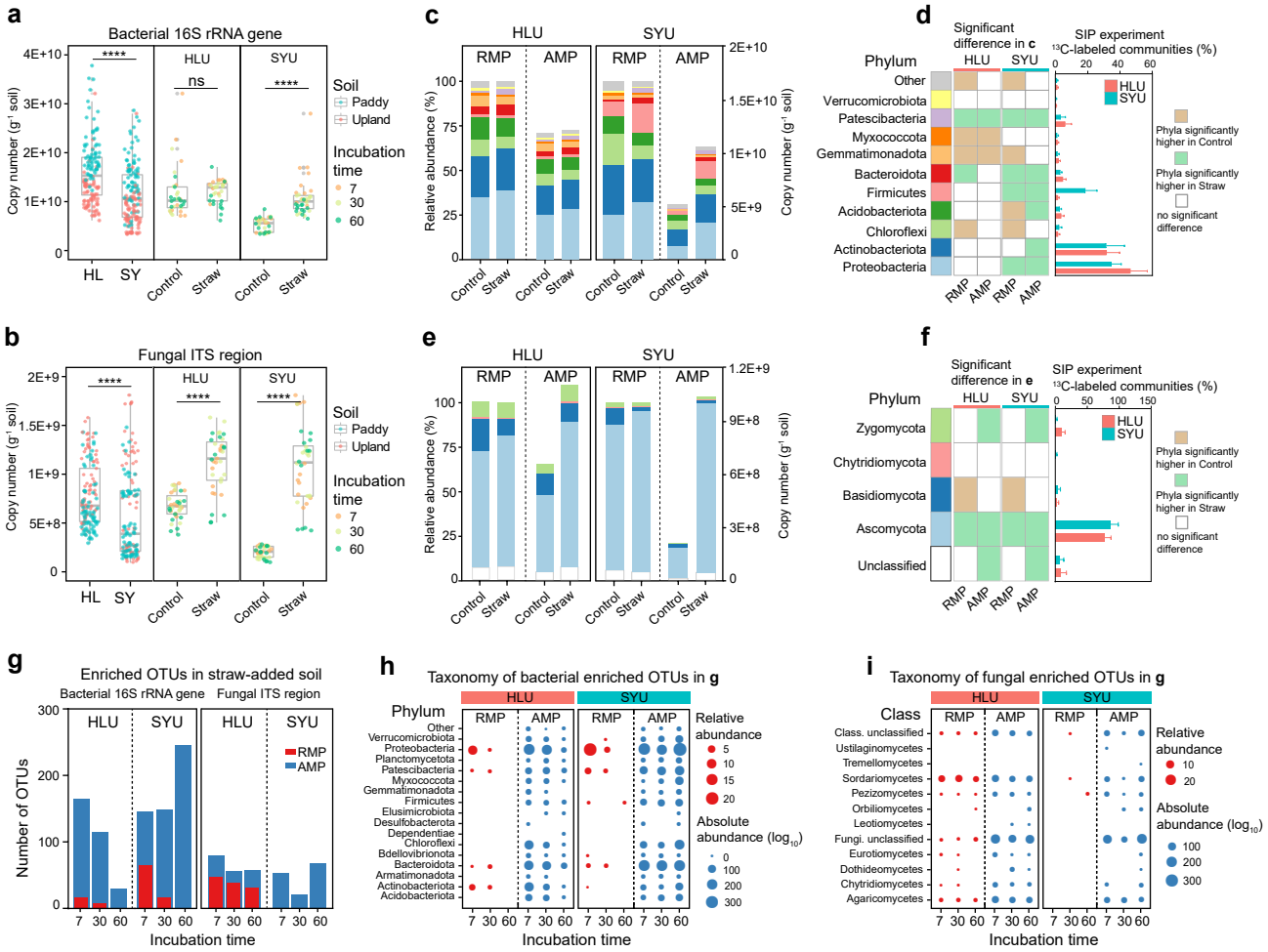

Absolute microbiome profiling (AMP)
Type #1: spike-AMP
a
1 Design of synthetic spike
Primer binding sites
Synthetic sequence
Synthetic sequence are cloned into vector and cloning vector were transformed into competent cell
Spike
synthetic spike
Bacterial 16S rRNA gene
515F
806R
Fungal ITS region
ITS1F
ITS2
Combination of synthetic spike
ITS1F 515F
806R ITS2
Bacterial spike
Fungal spike
Combined spike
b
Method #1: Spike-in method based on a single spike
2 Spiking into samples
Environmental sample DNA extracts
Adding a known amount of spikes to original environmental sample or extracted DNA sample
3 High-throughput sequencing
Bioinformatics analysis and sequencing reads split
Reads number
DNA sample
Synthetic spike
4 Absolute abundance calculating
DNA copies
Reads number
Spike copies (C_s)
DNA sample copies (C_m)
Spike reads (R_s)
DNA sample reads (R_m)
Spike
DNA Sample
$\frac{C_s}{C_m} = \frac{R_s}{R_m}$
$C_m = C_s \frac{R_m}{R_s}$
5
c
Method #2: Spike-in method based on a single spike with gradient concentrations
2 Spiking into DNA extracts or environmental samples
Control Level 1 Level 2 Level 3 Level 4 Level 5 Level 6 Level 7 Level 8
Spike level
3 High-throughput sequencing
Bioinformatics analysis and sequencing reads split
Reads number
DNA sample
Spike
4 Absolute abundance calculating
$\frac{C_s}{C_m} = \frac{R_s}{R_m} \rightarrow \frac{R_s}{R_m} = \frac{1}{C_m} C_s$
$y = ax$
$y = \frac{R_s}{R_m}$ $x = C_s$ $a = \frac{1}{C_m}$
Logarithmic transformation
$Y = b+X$
$Y = \mathrm{Log}_{10} \frac{R_s}{R_m}$ $X = \mathrm{Log}_{10} C_s$
$b = \mathrm{Log}_{10} \frac{1}{C_m}$ $C_m = \frac{1}{10^b}$
$\mathrm{Log}_{10} \frac{R_s}{R_m}$
Y = b+X
$C_m = \frac{1}{10^b}$
$\mathrm{Log}_{10} C_s$
5
d
Method #3: Spike-in method based on a spike mixture
2 Spiking into samples
A spike mixture including different spikes and gradient concentrations
Spike Copies
S1 10^3
S2 10^4
S3 10^5
S4 10^6
Spike mixture
Environmental sample DNA extracts
3 High-throughput sequencing
Bioinformatics analysis and sequencing reads split
Reads number
DNA sample
S1
S2
S3
S4
4 Absolute abundance calculating
Spike mixture copies
Spike mixture reads
Spike standard curve
DNA sample copies
y=0.5x
DNA sample reads number
spike copies
Spike reads number
S1 S2 S3 S4
DNA sample
5
Type #2: qPCR-AMP (qPCR data coupled with high-throughput sequencing)
e
1 Sample preparation
High-throughput sequencing
DNA sample
2 Sequencing data and quantitative results
Relative abundance (%)
Taxa1 Taxa2 Taxa3 Taxa4 Taxa5 Taxa6 Taxa7 Taxa8
A B C D
+
Sample | Copies of 16S rRNA gene
A | 1.70E+10
B | 1.53E+10
C | 9.01E+09
D | 1.41E+10
Total gene copies were estimated by qPCR.
3 Absolute abundance calculating
Absolute abundance of each taxon
Relative abundance from sequencing data
Gene copies
A(Taxa1, ... , Taxa8) = (%Taxa1, ... , %Taxa8) * (sample A)
B(Taxa1, ... , Taxa8) = (%Taxa1, ... , %Taxa8) * (sample B)
C(Taxa1, ... , Taxa8) = (%Taxa1, ... , %Taxa8) * (sample C)
D(Taxa1, ... , Taxa8) = (%Taxa1, ... , %Taxa8) * (sample D)
4
f
Absolute abundance (gene copies)
Taxa1 Taxa2 Taxa3 Taxa4 Taxa5 Taxa6 Taxa7 Taxa8
A B C D


a
Bacterial 16S rRNA gene
HLP
SYU
Copy number (g-1 soil)
2E+10
1.6E+10
1.2E+10
8E+9
4E+9
0
A
a
B
c
A
b
E
f
D
e
C
d
Sample
b
Fungal ITS region
HLP
SYU
3E+9
2E+9
1E+9
4E+7
0
Experiment 1
Experiment 2
Day0
Day7
Day30
D
d
D
d
C
c
E
e
B
b
A
a
Sample


a
Bacteria
Spike level
Spike reads / Total reads (%)
Day0
Day7
Day30
b
Fungi
Spike level
c
HLP
SYU
R_s/R_m
Log_{10}
$Log_{10} C_s$
d
HLP
SYU
$Log_{10} C_s$
Day0
Day7
Day30
Preliminary test
Spike-AMP
Model lines
e
Bacteria
Copy number (g^{-1} soil)
Spike level
Spike linear relationship
qPCR data
HLP
Day0
Day7
Day30
SYU
Day0
Day7
Day30
f
Fungi
Copy number (g^{-1} soil)
Spike level
Spike linear relationship
qPCR data


a
Bacterial bias degree
-0.5 0 0.5
qPCR data
HLP
SYU
Day0
Day7
Day30
Spike level
Copy number (g-1 soil)
b
Fungal bias degree
HLP
Day0
Day7
Day30
SYU
Day0
Day7
Day30
c
Bacteria
HLP_Day30
Log10 AA (single spike)
Log10 AA (qPCR data)
d
OTU AA > 10^7
Pearson r = 0.97
R² = 0.95
P < 0.0001
10^6 ≤ OTU AA ≤ 10^7
Pearson r = 0.77
R² = 0.59
P < 0.0001
OTU AA < 10^6
Pearson r = 0.46
R² = 0.21
P < 0.0001
Log10 AA (qPCR data)
Spike level
Level 1
Level 2
Level 3
Level 4
Level 5
Level 6
Level 7
Level 8
Level 9
e
Fungi
HLP_Day30
Log10 AA (single spike)
Log10 AA (qPCR data)
f
OTU AA > 10^5
Pearson r = 0.98
R² = 0.95
P < 0.0001
10^4 ≤ OTU AA ≤ 10^5
Pearson r = 0.40
R² = 0.16
P < 0.0001
OTU AA < 10^4
Pearson r = 0.18
R² = 0.03
P < 0.0001
Log10 AA (qPCR data)


a
Bacteria
b
Fungi
error=0.5
HLP
SYU
RA=1%
RA=0.1%
0.34%
99.66%
0.60%
99.40%
2.70%
97.30%
1.0%
7.70%
91.30%
Log10 (OTUs RA in control)
Spike level
Level 1
Level 5
Level 9
Days
Day0
Day7
Day30
OTU abundance error (OA error, effects of adding spike on OTU relative abundance)
c
HLP_Day0
HLP_Day7
HLP_Day30
CPCoA 2 (12.42%)
CPCoA 1 (19.55%)
16.4% of variance
P=0.001
CPCoA 2 (12.97%)
CPCoA 1 (18.82%)
16.7% of variance
P=0.001
CPCoA 2 (12.28%)
CPCoA 1 (17.72%)
16.1% of variance
P=0.001
Bacterial spike level
Control
Level 1
Level 2
Level 3
Level 4
Level 5
Level 6
Level 7
Level 8
d
PERMANOVA (P values)
Level 1
Level 2
Level 3
Level 4
Level 5
Level 6
Level 7
Level 8
Day0
Day7
Day30
HLP
SYU
P > 0.5
P < 0.5
P < 0.01
e
HLP_Day0
HLP_Day7
HLP_Day30
CPCoA 2 (13.86%)
CPCoA 1 (25.29%)
20.2% of variance
P=0.001
CPCoA 2 (15.05%)
CPCoA 1 (28.96%)
20.5% of variance
P=0.001
CPCoA 2 (16.5%)
CPCoA 1 (39.05%)
25.3% of variance
P=0.001
Fungal spike level
Control
Level 1
Level 2
Level 3
Level 4
Level 5
Level 6
Level 7
Level 8
Level 9
f
PERMANOVA (P values)
Level 1
Level 2
Level 3
Level 4
Level 5
Level 6
Level 7
Level 8
Level 9
Day0
Day7
Day30
HLP
SYU
P > 0.5
P < 0.5
P < 0.01


a
Bacterial 16S rRNA gene
Copy number (g-1 soil)
HL
SY
HLU
SYU
ns
****
Control
Straw
Soil
Paddy
Upland
Incubation time
7
30
60
b
Fungal ITS region
c
HLU
SYU
RMP
AMP
Relative abundance (%)
Copy number (g-1 soil)
d
Significant difference in c
Phylum
Other
Verrucomicrobiota
Patescibacteria
Myxococcota
Gemmatimonadota
Bacteroidota
Firmicutes
Acidobacteriota
Chloroflexi
Actinobacteriota
Proteobacteria
SIP experiment
13C-labeled communities (%)
Phyla significantly higher in Control
Phyla significantly higher in Straw
no significant difference
e
f
Significant difference in e
Zygomycota
Chytridiomycota
Basidiomycota
Ascomycota
Unclassified
g
Enriched OTUs in straw-added soil
Bacterial 16S rRNA gene
Fungal ITS region
Number of OTUs
Incubation time
h
Taxonomy of bacterial enriched OTUs in g
Relative abundance
Absolute abundance (log10)
i
Taxonomy of fungal enriched OTUs in g
Class
Class. unclassified
Ustilaginomycetes
Tremellomycetes
Sordariomycetes
Pezizomycetes
Orbiliomycetes
Leotiomycetes
Fungi. unclassified
Eurotiomycetes
Dothideomycetes
Chytridiomycetes
Agaricomycetes

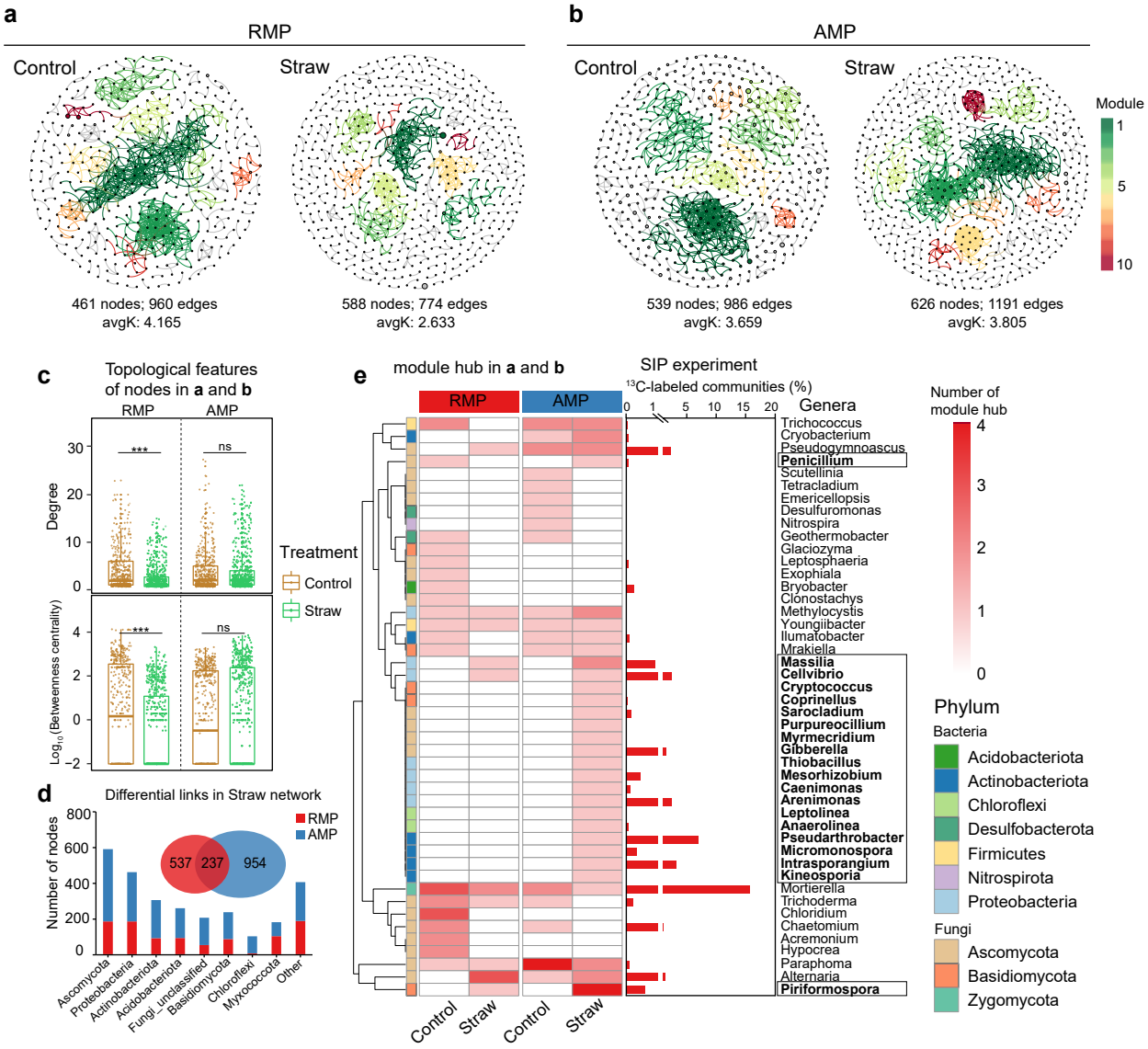

a
RMP
Control
Straw
461 nodes; 960 edges
avgK: 4.165
588 nodes; 774 edges
avgK: 2.633
b
AMP
Control
Straw
539 nodes; 986 edges
avgK: 3.659
626 nodes; 1191 edges
avgK: 3.805
Module
1
5
10
c
Topological features
of nodes in a and b
RMP
AMP
Degree
***
ns
Log10 (Betweenness centrality)
Treatment
Control
Straw
d
Differential links in Straw network
Number of nodes
RMP
AMP
537
237
954
Ascomycota
Proteobacteria
Actinobacteriota
Acidobacteriota
Fungi_unclassified
Basidiomycota
Chloroflexi
Myxococcota
Other
e
module hub in a and b
SIP experiment
13C-labeled communities (%)
RMP
AMP
Genera
Trichococcus
Cryobacterium
Pseudogymnoascus
Penicillium
Scutellinia
Tetracladium
Emericellopsis
Desulfuromonas
Nitrospira
Geothermobacter
Glaciozyma
Leptosphaeria
Exophiala
Bryobacter
Clonostachys
Methylocystis
Youngiibacter
Ilumatobacter
Mrakiella
Massilia
Cellvibrio
Cryptococcus
Coprinellus
Sarocladium
Purpureocillium
Myrmecridium
Gibberella
Thiobacillus
Mesorhizobium
Caenimonas
Arenimonas
Leptolinea
Anaerolinea
Pseudarthrobacter
Micromonospora
Intrasporangium
Kineosporia
Mortierella
Trichoderma
Chloridium
Chaetomium
Acremonium
Hypocrea
Paraphoma
Alternaria
Piriformospora
Control
Straw
Control
Straw
Number of
module hub
Phylum
Bacteria
Acidobacteriota
Actinobacteriota
Chloroflexi
Desulfobacterota
Firmicutes
Nitrospirota
Proteobacteria
Fungi
Ascomycota
Basidiomycota
Zygomycota

Table 1 Amount of synthetic spike added into DNA sample.

	Bacterial synthetic spike		Fungal synthetic spike	
Synthetic spike level	Spike copies added per PCR reaction	Spike copies / total gene copies (%) [a]	Spike copies added per PCR reaction	Spike copies / total gene copies (%)
Control	0	0	0	0
Level 1	6.94E+02	0.01	4.91E+01	0.03
Level 2	6.94E+03	0.09	4.91E+02	0.29
Level 3	6.94E+04	0.86	4.91E+03	2.65
Level 4	6.94E+05	7.54	4.91E+04	16.05
Level 5	2.08E+06	18.22	1.47E+05	29.86
Level 6	6.24E+06	37.06	4.42E+05	49.17
Level 7	5.62E+07	81.73	1.33E+06	69.73
Level 8	2.81E+08	95.59	3.98E+06	85.35
Level 9	-	-	1.99E+07	96.28

[a] The total gene copies is the sum of synthetic spike copies and DNA copies per PCR reaction.