

# Northumbria Research Link

Citation: Zhang, Jianwei, Dolfing, Jan, Liu, Wenjing, Chen, Ruirui, Zhang, Jiabao, Lin, Xiangui and Feng, Youzhi (2022) Beyond the snapshot: identification of the timeless, enduring indicator microbiome informing soil fertility and crop production in alkaline soils. *Environmental Microbiomes*, 17 (1). p. 25. ISSN 2524-6372

Published by: BMC

URL: <https://doi.org/10.1186/s40793-022-00420-6> <<https://doi.org/10.1186/s40793-022-00420-6>>

This version was downloaded from Northumbria Research Link:  
<http://nrl.northumbria.ac.uk/id/eprint/49304/>

Northumbria University has developed Northumbria Research Link (NRL) to enable users to access the University's research output. Copyright © and moral rights for items on NRL are retained by the individual author(s) and/or other copyright owners. Single copies of full items can be reproduced, displayed or performed, and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided the authors, title and full bibliographic details are given, as well as a hyperlink and/or URL to the original metadata page. The content must not be changed in any way. Full items must not be sold commercially in any format or medium without formal permission of the copyright holder. The full policy is available online: <http://nrl.northumbria.ac.uk/policies.html>

This document may differ from the final, published version of the research and has been made available online in accordance with publisher policies. To read and/or cite from the published version of the research, please visit the publisher's website (a subscription may be required.)

RESEARCH ARTICLE

Open Access



# Beyond the snapshot: identification of the timeless, enduring indicator microbiome informing soil fertility and crop production in alkaline soils

Jianwei Zhang<sup>1,2</sup>, Jan Dolfing<sup>3</sup>, Wenjing Liu<sup>1,2</sup>, Ruirui Chen<sup>1</sup>, Jiabao Zhang<sup>1</sup>, Xiangui Lin<sup>1</sup> and Youzhi Feng<sup>1\*</sup> 

## Abstract

**Background:** Microorganisms are known to be important drivers of biogeochemical cycling in soil and hence could act as a proxy informing on soil conditions in ecosystems. Identifying microbiomes indicative for soil fertility and crop production is important for the development of the next generation of sustainable agriculture. Earlier researches based on one-time sampling have revealed various indicator microbiomes for distinct agroecosystems and agricultural practices as well as their importance in supporting sustainable productivity. However, these microbiomes were based on a mere snapshot of a dynamic microbial community which is subject to significant changes over time. Currently true indicator microbiomes based on long-term, multi-annual monitoring are not available.

**Results:** Here, using samples from a continuous 20-year field study encompassing seven fertilization strategies, we identified the indicator microbiomes ecophysiologically informing on soil fertility and crop production in the main agricultural production base in China. Among a total of 29,184 phylotypes in 588 samples, we retrieved a streamlined consortium including 2% of phylotypes that were ubiquitously present in alkaline soils while contributing up to half of the whole community; many of them were associated with carbon and nutrient cycling. Furthermore, these phylotypes formed two opposite microbiomes. One indicator microbiome dominated by *Bacillus asahii*, characterized by specific functional traits related to organic matter decomposition, was mainly observed in organic farming and closely associated with higher soil fertility and crop production. The counter microbiome, characterized by known nitrifiers (e.g., *Nitrosospira multiformis*) as well as plant pathogens (e.g., *Bacillus anthracis*) was observed in nutrient-deficit chemical fertilizations. Both microbiomes are expected to be valuable indicators in informing crop yield and soil fertility, regulated by agricultural management.

**Conclusions:** Our findings based on this more than 2-decade long field study demonstrate the exciting potential of employing microorganisms and maximizing their functions in future agroecosystems. Our results report a “most-wanted” or “most-unwanted” list of microbial phylotypes that are ready candidates to guide the development of sustainable agriculture in alkaline soils.

**Keywords:** Soil microorganisms, Soil fertility, Sustainable agroecosystem, Archived soils, Fertilization

\*Correspondence: yzfeng@issas.ac.cn

<sup>1</sup> State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, People's Republic of China

Full list of author information is available at the end of the article

## Background

Soil microorganisms with their incredible genetic and ecophysiological diversity play essential roles in regulating multiple ecosystem services, including nutrient cycling, organic matter decomposition, plant



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

productivity, and pathogen control [1–3]. However, the immense diversity of soil bacterial communities has stymied efforts to characterize functional taxa and/or consortia [4]. Recent research suggests alternative strategies to focus on a relatively small subset of phylotypes instead of exhaustively screening that overwhelming diversity [5–7]. These refined phylotypes are indispensably present in communities with a high frequency and they self-assemble into consortia that could characterize divergent environmental conditions (hereafter termed indicator microbiomes) [4, 8]. Such streamlined but ubiquitous microbiomes will be fruitful targets for genomic and cultivation-based efforts to improve our understanding of soil microbes and their contributions to ecosystem functioning [4, 6–8].

Agroecosystems are shaped by entangled webs of interactions among diversified microorganisms. Species composition and interactions in microbial communities determine the ecological functions and services they provide and hence they play a pivotal role in maintaining ecosystem sustainability [9]. Harnessing microorganisms and maximizing their functions in agroecosystems offers one of the most promising long-term solutions to meet the integral demand for simultaneously promoting crop production and environmental sustainability [6, 10]. Phylogenetically distinct soil microbes inherently exhibit divergent niche preferences, which in turn mirror different environmental conditions [4, 8], e.g., a high fertility soil with high crop production, and vice versa.

Fertilization is one of the most impactful human manipulations of agroecosystems [11], where divergent farming systems supported distinct microbiomes depending on and indicative of the conditions in the agroecosystem concerned [12, 13]. Divergent fertilization practices provide us with the opportunity to unravel the indicator microbiome associated with nutrient cycling, for example, nitrogen addition management shifting the microbiome related to N cycling [12] and organic matter amendment linking the microbiome associated with the decomposition of external organic matter [13, 14]. In this context, a systematic long-term field experiment including the frequent application of multiple fertilization types could give us a general indicator microbiome associated with soil nutrient cycling and hence crop production [15–17].

Without strong external perturbation, soil microbial community composition can vary on scales from days to years due to stochastic drift and speciation in combination with ongoing dispersal processes [18–20]. Given this situation, microbial temporal patterns cannot necessarily be generalized even if soils are exposed to the same climatic conditions [21], which questions the existence of consistent indicator microbiomes that could resist

temporal variability [18, 21]. On the other hand, the situation is different when deterministic influences from strong external perturbations exceed the inherent stochastic temporal variability in the microbial community [21, 22]. For example, fertilization is a strong environmental filter that significantly shapes the environment, and hence shifts the microbial community therein [6, 23, 24]. In this respect, we hypothesized that specific indicator microbiomes consistently emerge during long-term distinct fertilization strategies, which can be indicative of soil fertility and crop yield. Indeed, such a comprehensive temporal analysis of indicator microbiomes would transcend noisy information from typical snapshot studies [7, 21] and is essential for refinement of the identification and application of indicator microbiomes that are critical for crop yield and soil fertility.

Here, we performed a retrospective analysis of a 20-year long fertilization experiment encompassing 588 samples from seven contrasting fertilizer treatments, representing trajectories of farming systems from conventional to organic and from nutrient-deficiency to nutrient-balance. These unique samples enable us to investigate the existence, composition and structure of indicator microbiomes across a decadal scale that are closely related to soil fertility and crop production under various management practices. This long-term field experiment was conducted on the Northern China Plain, where winter wheat-summer maize rotation is the dominant cropping system and crop production is primarily constrained by low indigenous fertility of the alkaline soil [14]. Approximately 30% of arable land in the world suffers from alkaline stress limiting crop production [25], and this challenge is compounded by global aridification [26, 27]. Wheat and maize are among the most widely grown commodity crops around the world, accounting for a considerable amount of cropland arable area, yielding non-meat calories and animal feed [28]. In this respect, the identified indicator microbiome for soil fertility and crop yield from this study, although generated from a specific sampling site, could provide critical information for dryland ecosystems on a global scale that suffer from soil alkalinity.

## Material and methods

### Study sites and sample collection

Soil samples were collected from a long-term nutrient fertilization experiment at the Fengqiu Ecological Experiment Station affiliated to Institute of Soil Science, Chinese Academy of Sciences (Kaifeng City, Henan Province, 35° 00' N, 114° 24' E). The soil is classified as Aquic Inceptisol (a calcareous fluvoaquic soil), a typical alkaline soil in the North China Plain derived from alluvial

sediments of the Yellow River. The soil pH is approximately 8.65, with total organic matter at 5.83 g/kg, total N at 0.45 g/kg, total P at 0.5 g/kg, and total K at 18.5 g/kg at the beginning of the experiment [14]. The crop succession was winter wheat (*Triticum aestivum* L.) and summer maize (*Zea mays* L.). This experiment was established in 1989 and consisted of seven fertilization regimes with four replicates in randomized plots as follows [14]: (1) OM: organic manure (supplemented with P and K as chemical fertilizers for the same amount of nutrients as in the other treatments); (2) NPKM: half organic manure plus chemical fertilizer NPK; (3) NPK: balanced chemical fertilizer NPK; (4) NP: chemical fertilizer NP; (5) PK: chemical fertilizer PK; (6) NK: chemical fertilizer NK; and (7) Control: no fertilization. The detailed experimental design and fertilization regimes have been documented in previous work [29, 30], and could be referred in Additional file 2: Table S1. Briefly, each plot measured 9.5\*5 m<sup>2</sup>. All phosphorus (calcium superphosphate), potassium (potassium sulfate) and organic manure fertilizers (mixture of wheat straw, oil rapeseed cake, and cottonseed cake in a ratio of 100:40:45) were applied as basal fertilizers, whereas N (urea) was added in two applications as both the basal (60 kg ha<sup>-1</sup> N for maize and 90 kg ha<sup>-1</sup> N for wheat, respectively) and supplementary fertilizer (90 kg ha<sup>-1</sup> N for maize and 60 kg ha<sup>-1</sup> N for wheat, respectively). Crop yields were recorded annually and were generally increased under fertilizations except for those under N- and P-deficient fertilizers (PK and NK, respectively) and unfertilized Control (Additional file 1: Fig. S1). Soil samples were collected annually after the harvest of summer maize (typically in October), air-dried and archived in stoppered glass bottles. In 2019, we retrieved archived soil samples from 1989 to 2009, yielding 588 samples (21 years × 7 fertilization regimes × 4 replicates) in total. Soil C and N were measured with a LECO analyzer, and soil P and K were measured by colorimetry.

#### DNA sequencing and bacterial diversity analysis

Genomic DNA was extracted from archived soils using the FastDNA Spin Kit for Soil (MP Biochemicals, Solon, OH, USA), with some modifications to the manufacturer protocol. Briefly, soils of around 500 mg (as dry weight) were weighed into prepared centrifuge tubes (prechilled to 4 °C) containing buffer and extraction glass beads. The samples were incubated for 30 min for rehydration, and then cells were lysed in a FastPrep-24 Homogenizer (6.0 setting, 40 s; MP Biochemicals, USA). The samples were then incubated at 70 °C for 10 min to aid the lysis of Gram-positive bacteria, and briefly (~ 10 s) re-shaken in the homogenizer. The extracted soil DNA was dissolved in 50 µl of TE buffer, quantified by a spectrophotometer

and stored at -20 °C (long-term storage in -80 °C). Despite air-drying and long-term storage, high-quality DNA was recovered from the soils archived for decades (260/280~1.8 ([31])). We have been previously published based on DGGE analyses of these samples [14], which confirmed the systematic difference between samples from divergent farming systems and reported that *Bacillus asahii* bloomed from 2 to 4 years onward under organic fertilization. In this study, Illumina high-throughput sequencing platform was used to dissect more sophisticated information on the bacterial community composition. For each soil sample, the primer set 519F/907R [32] was used to amplify approximately 400 bp of V4-V5 region of bacterial 16S rRNA gene fragments. The 5-bp length oligonucleotide sequences were fused to the forward primer to distinguish individual samples. PCR was carried out in 50 µl mixtures with the following components (final concentrations): 200 µM dNTPs; 0.6 µM forward and reverse primers; 2 units/50 µl Taq DNA polymerase (TaKaRa, Japan); around 1 ng/µl template DNA; add sterile water to 50 µl. Twenty-five cycles (95 °C for 45 s, 56 °C for 45 s, and 72 °C for 60 s) were performed with a final extension at 72 °C for 7 min on a Bio-Rad C1000 thermal cycler. Triplicate PCR were performed per sample and then were pooled, purified using the QIAquick PCR Purification kit (QIAGEN), and quantified using a NanoDrop ND-1000 (Thermo Scientific, USA). The bar-coded PCR products from all samples were normalized in equimolar amounts to construct the library using a MiSeq Reagent kit v3 (PE 2 × 300), and then were sequenced on an Illumina MiSeq platform in our institute (Illumina Inc., CA, USA).

Bacterial 16S rRNA gene sequence data were processed using the USEARCH pipeline. Raw paired-end (forward and reverse) sequences were merged with the “fastq\_mergepairs” function and low-quality sequences (length < 200 or quality score < 20) were filtered with the “fastq\_filter” function in USEARCH software [33]. We used the “fastx\_uniques” function to perform the dereplication which could greatly reduce the computational memory demand. The functions “cluster\_otus” and “unoise3” were then used to cluster unique sequences into operational taxonomic units (OTUs), and to perform error-correction on sequences (denoising into zero-radius OTUs, zOTUs), respectively. Singletons were discarded as they potentially result from erroneous sequencing or prediction. Phylotype identification was achieved using RDP classifier with a ≥ 80% probability threshold [34]. OTU table was constructed using the “otutab” function, and then combined with phylotype identification in R. We removed phylotypes that were not assigned as Bacteria and detected in less than 10 times across the data set. Finally, validated data were rarefied

to 8200 (for 97% clustering) and 9500 sequences (for denoising) per sample to balance the sampling efforts, respectively. Beta diversity plots were generated with the rarefied data using the QIIME `beta_diversity_through_plots.py` script based on Bray–Curtis metrics [35]. The two approaches produced considerably overlapping communities (21,112 phylotypes for 97% clustering and 29,184 phylotypes for denoising, respectively), with alpha and beta diversity metrics being highly positively correlated ( $r > 0.73$  for observed richness, and  $r > 0.98$  for Bray–Curtis dissimilarity, Additional file 1: Fig. S2). The denoising approach only increased (by 1.4 times) the resolution of bacterial diversity, in line with recent systematic investigations that demonstrated the limited effect of the OTU approach on the resulting microbial community structure [8]. Although the 97% similarity threshold is widely accepted to define molecular OTUs from the bacterial 16S rRNA region [4], the use of the denoised zOTUs has the important advantage at a much higher resolution of allowing for more precise identification down to the species level and even potentially beyond [8]. Therefore, given the substantial overlap between the two approaches, we decided to use the denoised zOTUs for all the downstream analyses. The datasets generated and analyzed during the current study are available in the National Center for Biotechnology Information (NCBI) repository under the accession number PRJNA726588.

#### Validation for retrospective microbial analysis based on archived soil samples

We first estimated the fidelity of the archived soil samples for retrospect analysis according to their soil chemical characteristics (i.e., SOC, TN, TP, TK) and bacterial community composition via the Random Forest algorithm. This method has previously been shown to outperform other modeling approaches when used for environmental bacterial datasets [13, 36]. In these classification models (number of trees = 500, permutations = 999), fertilization strategy (a vector including seven factors that correspond to the seven fertilization treatments used) was set as response variable while soil chemical characteristics or community composition was set as predictor variable, in which fertilization time (in years from 0 to 20) were set as strata (i.e. a factor variable that is used for stratified sampling) to take into consideration the temporal dynamics in field conditions. The results showed that our archived soil samples were well stored to the extent that we could use the soil chemical properties (accuracy of 91.5% estimated via the out of bag error rate in Random Forest models) and bacterial community (accuracy of 86.73%) to accurately identify the fertilizer treatment the sample originated from (Additional file 1: Fig. S3). This high match accuracy could be comparable to those based on

fresh samples [37], and provides high confidence in our results. Random Forest analysis was performed using the R package `rfPermute` (with default parameters, v2.1.81 [38]).

#### Identifying the indicator microbiome and co-occurrence network construction

We first selected dominant taxa by retaining ubiquitous zOTUs that were present in more than 80% of all soil samples ( $n = 544$ ), which yielded a subset community consisting of 604 phylotypes. We did not apply abundance criteria in this analysis because microbial keystone taxa might exert their influence (either individually or as a guild) on ecosystem functioning irrespective of their abundance across space and time (such as those involved in nitrogen fixation or ammonia oxidation) [5]. Ecological co-occurrence networks provide critical information on the potential associations among soil microorganisms and can be used to identify soil indicator phylotypes and/or consortia associated with the high levels of soil functioning and crop production [5, 23]. All possible Spearman rank correlations between OTUs across samples and corresponding  $P$  values were calculated. We corrected the false discovery rate according to Benjamini–Hochberg [39]. We considered a valid co-occurrence event to be robust if the Spearman correlation coefficient  $|\rho| > 0.6$  and statistically significant at  $P < 0.01$  [40, 41]. Pairwise Spearman correlation was calculated using the function `'rcorr'` in the R package `Hmisc` (v4.4–0) [42], and the co-occurrence network was constructed using the function `'graph.adjacency'` in the R package `picante` (v1.8.1). We also detected potential clusters/modules within co-occurrence network using the function `'fastgreedy.community'` in the R `picante` package, as ecological clusters represent important ecological units that provide the opportunity to identify highly connected and indicator taxa [43]. An interactive platform “Gephi” (default parameters set, v0.9.1 [44]) was used to visualize the co-occurrence network. In the network graph, the nodes represent the individual phylotypes, with the colors denoting the ecological modules to which they belong, and the sizes of the nodes are proportional to their relative abundances in the community. Edges represent significant pairwise correlations with their width proportional to the coefficient. We calculated the relative abundance of each cluster by summarizing the relative abundances of all phylotypes therein. We are aware of the compositional nature of microbiome data [45] and of the fact that Spearman correlation is not a compositionally aware correlation metric [46]. When we tested with compositionally aware correlation metric, that is SparCC [47], despite the difference in network layout and scales, the outcomes and conclusions were essentially the same (Additional file 1: Fig. S4). In addition,

Spearman approach generated a bigger-scale network whereby providing more potential candidates taxa that are fruitful target in management practices [4, 41]. Therefore, we decided to choose Spearman approach in downstream analysis.

#### Identification of accurate taxonomy and microbial functional traits

Given that the denoised zOTUs allow for precise identification down to the species level [8], we further validated taxonomy information by blasting our phylotypes against the 16S rRNA database in NCBI. Result showed that, 163 out of the 604 selected phylotypes had 16S rRNA sequences that could be assigned with >98.5% identity (similarity threshold for differentiating two species based on 16S rRNA gene sequence [48]) to a type specimen with a sequenced genome in the NCBI database. Then, the corresponding type specimen annotations were taken as the taxonomy of these phylotypes. Accounting for duplicate matches, these 163 phylotypes accounted for 140 separate type organisms. We matched those phylotypes at the species level with publicly available functional microorganisms related to nitrogen (nitrogen fixation/M00175, nitrification/M00528, denitrification/M00529, dissimilatory nitrate reduction/M00530, assimilatory nitrate reduction/M00531, complete nitrification/comammox/M00804), phosphorus (phosphonate and phosphinate metabolism/ko00440, alkaline phosphatase/M00126), and sulfur (assimilatory sulfate reduction/M00176, dissimilatory sulfate reduction/M00596, thio-sulfate oxidation by SOX complex/M00595) metabolism from the KEGG database (last accessed on April 25, 2021). We calculated the module functional proportion by summarizing the proportion of successfully matched phylotypes within each module.

#### Statistical analysis

The importance of these dominant taxa was estimated by calculating their count ratio as well as their relative abundance ratio. Pearson correlation between phylotype/module proportion and soil chemical characteristics was calculated using the function “rcorr” in the R package Hmisc (v4.4–0) [42]. Ordinary least square method was used to estimate the relationship between community (alpha and beta) diversities generated by 97% clustering and denoising approaches as well as the dissimilarity between the whole community and the subset community consisting of selected dominant phylotypes. ANOVA was used to test the significant difference in the proportion of each module among fertilizer treatments. We used the “aov” function in the R package Stats (v3.6.3) to conduct this analysis with the formula

$aov(\text{module} \% \sim \text{fertilizer} + \text{Error}(\text{duration}/\text{fertilizer}))$ , in which fertilizer is the fixed variable, but experiment duration (years) is random and fertilizer is nested within duration to take into account of temporal variation in field conditions. The circular phylogenetic tree was visualized using the online tool iTOL [49]. All statistical analyses in this study were based on R version v3.6.3 [42].

## Results

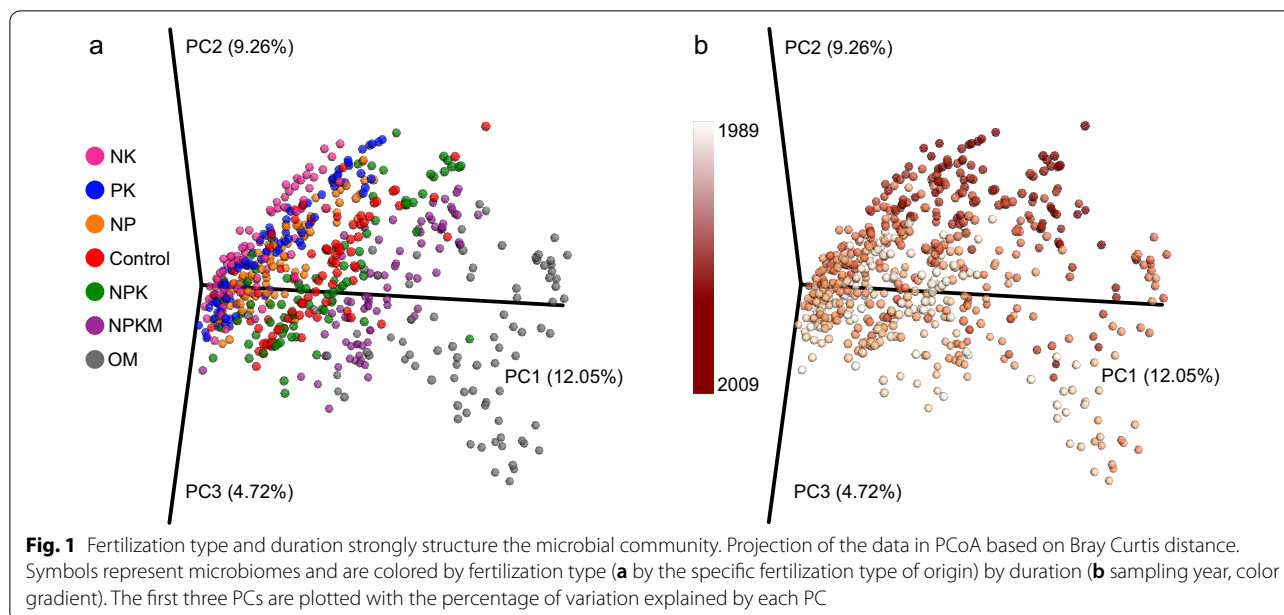
### Fidelity of the archived soil samples for the retrospective analysis

We first checked the chemical fidelity of the archived soil samples. Based on their chemical signature we could accurately assign a certain sample to a certain treatment with >91% confidence (Additional file 1: Fig. S3a). Then we performed the same approach based on the microbial community signatures of the samples, and the result was again very positive: we could identify with 86.73% certainty from what treatment series the (stored) sample was derived (Additional file 1: Fig. S3b).

Principal coordinate analysis (PCoA) based on Bray–Curtis distances showed that soil bacterial community exhibited distinct patterning according to the fertilization type (Fig. 1a), and to fertilization duration within each type (Fig. 1b). Strikingly, the samples formed four clusters based on fertilization types: two types of organic fertilization (OM and NPKM) were distinct from the chemical fertilizers and formed two independent clusters, whereas the three types of element-deficit fertilization (NP, NK, and PK) clustered together, while balanced chemical fertilizer (NPK) was more closely related to the no fertilized Control (Fig. 1a).

### Identification of the indicator phylotypes

As microbes that are consistently present with high frequency are likely to provide critical ecological functions, we searched for microbial taxa that are detected in more than 80% of samples. This streamlined microbial consortium consisted of 604 OTUs from a small set of bacterial phyla (i.e., phyla Proteobacteria, Firmicutes, Actinobacteria, Chloroflexi and Acidobacteria), which only represented ~2.1% of all the 29,184 sampled bacterial phylotypes, while contributing to nearly half of the total reads identified as Bacteria (Fig. 2 and Additional file 2: Table S2). The variation in Bray–Curtis dissimilarity index of this streamlined consortium was positively correlated with the diversity of the whole bacterial community ( $r=0.95$ ,  $P<0.001$ , Additional file 1: Fig. S5a). Individual phylotype abundance and frequency were positively correlated, with relatively few phylotypes being both highly frequent and abundant, and the majority of



the retrieved bacterial taxa restricted in their abundance (Additional file 1: Fig. S5b).

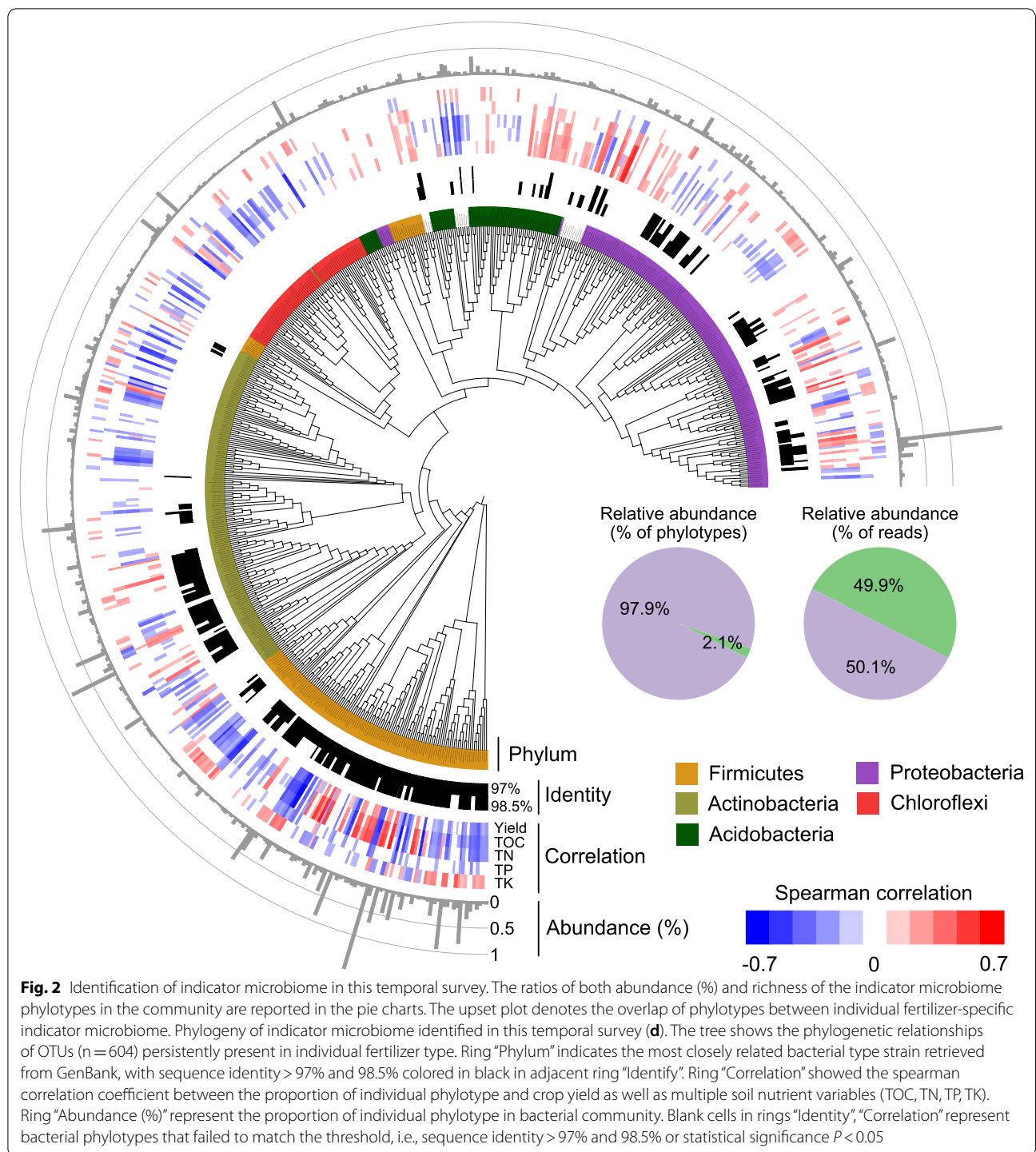
When compared against the GenBank database, 163 out of the 604 phylotypes had >98.5% sequence identity with sequences obtained from publicly available bacterial type strains (Fig. 2 and Additional file 2: Table S2). The most abundant taxa included *Bacillus asahii* (2.05%), *Methylobacterium radiotolerans* (1.99%), *Priestia filamentosa* (0.94%) and *Blastococcus colisei* (0.93%). Importantly, majority members with their abundance were closely associated with multiple soil nutrient conditions as well as maize yield (Fig. 2), suggesting that this refined microbial consortium could well act as the indicator microbiome of soil fertility and crop production. Specifically, 40% of phylotypes affiliated with the phylum Firmicutes were negatively ( $P < 0.05$ ) correlated with crop production, while more than one quarter (27%) of phylotypes in the phylum Proteobacteria (especially those in Alphaproteobacteria and in Gammaproteobacteria) were positively ( $P < 0.05$ ) correlated with higher production (Additional file 2: Table S2). Nevertheless, several abundant phylotypes in the class Bacilli of phylum Firmicutes, such as *B. asahii* ( $r = 0.3$ ), *Lysinibacillus sphaericus* ( $r = 0.31$ ) were also positively correlated with higher production.

#### Identification of the indicator microbiome

We used co-occurrence network analysis to identify phylotypes in the indicator microbiome that were highly connected with each other, and therefore, potentially shared environmental preferences and functional potentials (i.e., the effect on maize production in this study), which could

go beyond predictions based on their taxonomic affiliations alone. This approach was feasible because the contrasting fertilization types supported different levels of maize production (Additional file 1: Fig. S1) and tended to give rise to distinct microbial communities (Fig. 1a).

We detected and identified several ecological clusters from the refined microbiome in the co-occurrence network. Many of the members (66 out of 165 phylotypes) of these clusters could be accurately identified at the species level based on the similarity threshold 98.5% (Fig. 2 and Additional file 2: Table S2). Specifically, *Cytobacillus firmus* (abundance = 0.88%), *Neobacillus mesonae* (0.81%) and *Bacillus anthracis* (0.56%) in phylum Firmicutes were the main members in module #1; module #2 was mainly comprised of *Nitrospira japonica* (0.27%) in phylum Nitrospirae and many other unclassified taxa in phyla Proteobacteria and Acidobacteria; module #3 was dominated by Actinobacteria, such as *Blastococcus colisei* (0.93%) and *Micrococcus terreus* (0.63%); and phylotypes in genus *Bacillus* of phylum Firmicutes dominated the module #4, in which *B. asahii* (1.98%) and *L. sphaericus* (0.31%) were the most abundant members (Fig. 3a). Members within each module were consistently and positively correlated with each other. Modules #1 and #2 were the main components in the network, and were connected through several phylotypes with negative correlations. Furthermore, while some members (e.g., *N. mesonae* and *B. anthracis* in module #1, *B. colisei* in module #3 and *B. asahii* in module #4) are important phylotypes that dominate the community and construct the module, dominant taxa are not necessarily guaranteed to be the keystone taxa in a community (e.g., *Priestia flexa*

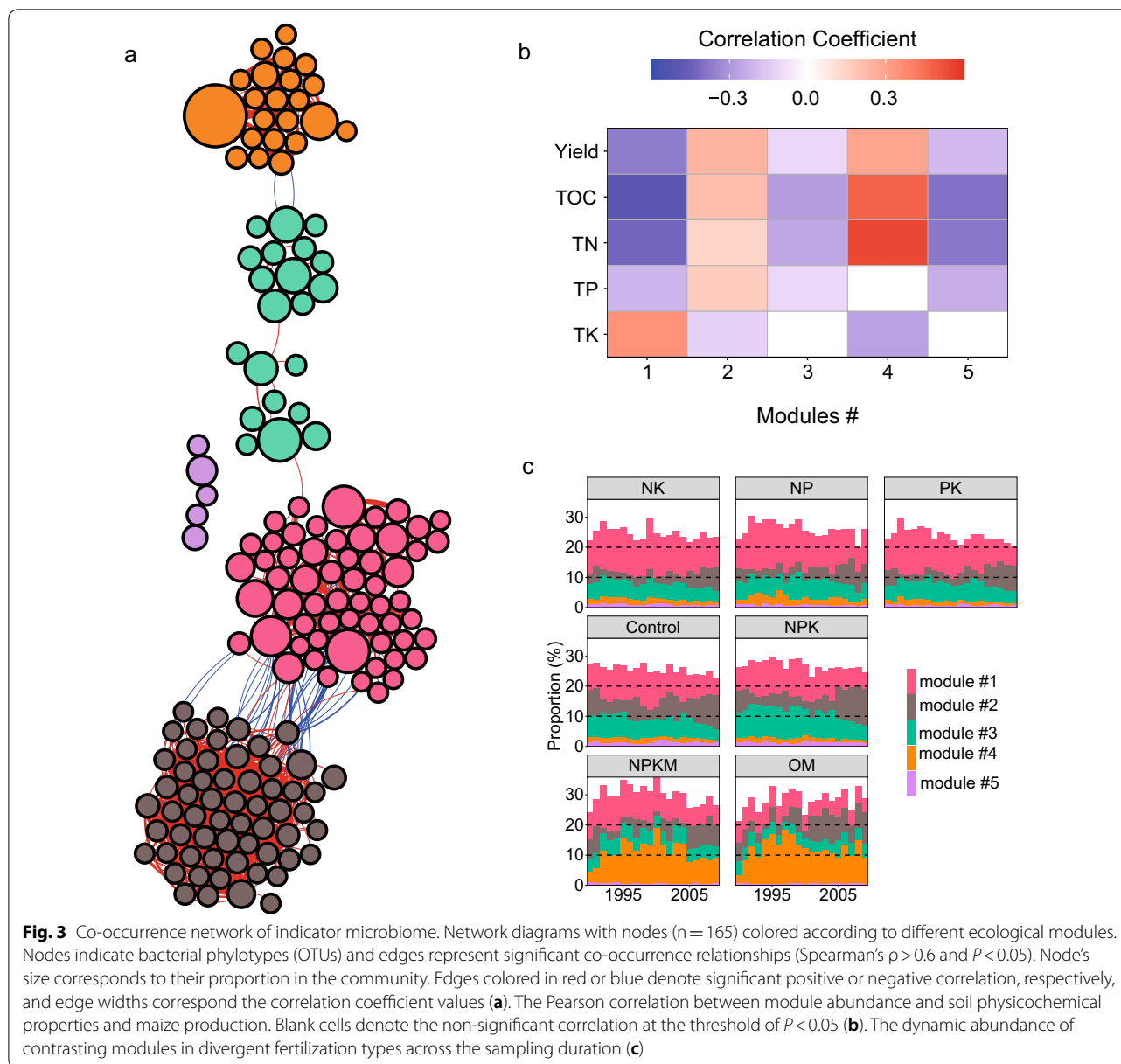


in module #1, *N. japonica* in module #2 and *L. sphaericus* in module #4) (Additional file 2: Table S2).

Network analysis revealed that the combined phylotype abundances of modules #2 and #4 were significantly and positively correlated with maize production and multiple soil nutrient contents (e.g., SOC and TN, TP), whereas

those of modules #1, #3 and #5 exhibited opposite patterns (Fig. 3b). Furthermore, these two contrasting groups of modules also showed different distributional patterns across divergent fertilization types. Compared to the unfertilized control and chemical fertilizers, organic fertilization (e.g., OM and NPKM) consistently



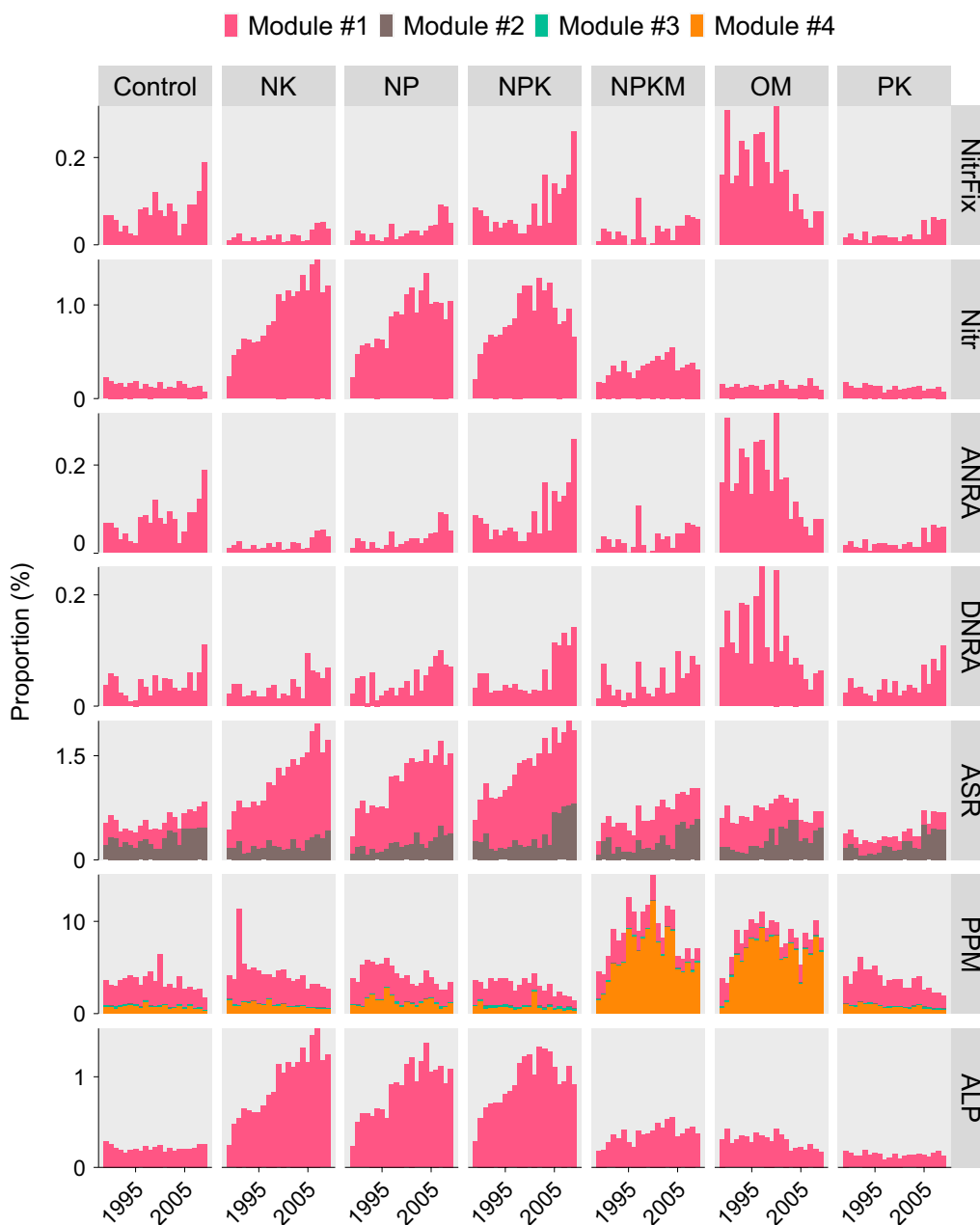


harbored a higher proportion of phylotypes in module #4 ( $F = 4.54$ ,  $P < 0.001$ ), while phylotypes in modules #1 ( $F = 2.78$ ,  $P = 0.011$ ) and #3 ( $F = 5.3$ ,  $P < 0.001$ ) were persistently rare in OM fertilization compared to chemical fertilization types (Fig. 3c).

#### Multifunctionality of the indicator microbiome

Bacterial phylotypes in the indicator microbiome exhibited various metabolic functions related to nitrogen, phosphorus and sulfate cycling, with most of the members capable of decomposing organic phosphorus via phosphonate and phosphonite metabolism (Fig. 4). Bacterial phylotypes in module #2 (*N. japonica*) was

associated with assimilatory sulfate reduction, and those in module #4 (i.e., *L. sphaericus*, and *B. asahii*) were involved in phosphonate/phosphinate metabolism. Members in module #1 were suggested to play important roles in multiple functions: *B. anthracis*, *P. filamentosa* and *N. mesonae* combined play key roles in phosphonate and phosphinate metabolism, while *Nitrosospira multififormis* is involved in nitrification, assimilatory sulfate reduction and alkaline phosphatase production, and was enriched under chemical nitrogen fertilization, *Rhodopseudomonas palustris* is active in both nitrogen fixation and assimilatory nitrate reduction to ammonium (Fig. 4). Nitrogen fertilizer treatment exhibited a



**Fig. 4** Multifunctionality of indicator microbiomes. We matched our bacterial phylotypes against the KEGG module table related to nitrogen, phosphorus and sulfate metabolism. *NitrFix* nitrogen fixation, *Nitr* nitrification, *DNRA/ARNA* dissimilatory/assimilatory nitrate reduction to ammonium, *ASR* assimilatory sulfate reduction, *PPM* phosphonate and phosphinite metabolism, *ALP* alkaline phosphatase

higher proportion of phylotypes in module #1 that were involved in nitrification ( $F=47.59, P<0.001$ ) and alkaline phosphatase ( $F=54.37, P<0.001$ ) as well as assimilatory sulfate reduction ( $F=55.71, P<0.001$ ), while organic fertilizer samples showed higher potential in nitrogen fixation ( $F=14, P<0.001$ ), nitrate reduction to ammonium ( $F=14$  for ANRA and  $F=11.20$  for DNRA,  $P<0.001$ ) in

module #1 and phosphonate and phosphonite metabolism ( $F=5.58, P<0.001$ ) in module #4.

**Discussion**

**Functional traits of indicator phylotypes for soil fertility and crop yield**

Harnessing microbiomes offers an invaluable approach to achieve agroecosystem sustainability with joint

consideration of food security and environmental stress [6, 10]. In this study, we addressed this issue by characterizing the identity, fertilization-dependency and ecology of indicator microbiomes identified in a long-term (>20 years) fertilization experiment conducted in the North China Plain that encompassed seven frequently applied fertilization types. To confidently detect and characterize the indicator microbiota, we selected those phylotypes that appeared at high frequency over 20 years. This stringent criterion generated a set of streamlined indicator phylotypes (~2.1%) from a few bacterial phyla that accounted for almost half (~50%) of the community population (Fig. 2a). These indicator phylotypes encompassed both dominant taxa with considerable abundance affecting broad processes (e.g., organic matter decomposition) and rare microbes exerting disproportional influences on specific well-defined processes (e.g., nitrogen fixation and ammonia oxidation) (Figs. 3a, 4). Many members of these indicator phylotypes are on record as indicator or keystone microbes in various agroecosystems, such as the most abundant taxa identified in this study like *B. asahii* and *M. radiotolerans* (Fig. 2d and Additional file 2: Table S2). As most bacteria generally do not take up macromolecules, *Bacillus* spp. naturally secrete a wide range of hydrolytic enzymes to decompose complex organic matter and to release mineral nutrients (such as phosphonate) into their environment, making them keystone taxa in the soil food-web [14, 50], and pioneers when exposed to organic matter amendment in alkaline soils [14, 51]. Indeed, *B. asahii* was found to dominate the bacterial community from 2 to 4 years onward under organic fertilization in our previous investigation [14]. In comparison, *M. radiotolerans* was more abundant under nutrient-deficit chemical fertilization types (Additional file 2: Table S2), helping the crop resist nutrient deficiencies by forming a symbiotic relationship with the plants in which *M. radiotolerans* consumes plant waste as a carbon source and in turn generates substances promoting plant growth [52].

With respect to the indicators of rare phylotypes, for example, *R. palustris* is a soil-borne phototroph that has been shown to exert positive effects on plant growth by stimulating nitrogen uptake and elevating auxin levels in expanding leaves [53]. *R. palustris* was enriched under OM treatment in this study (Additional file 2: Table S2), suggesting that organic fertilization supported a stronger soil nitrogen fixation ability. *N. multiformis* is an important nitrifier in upland soil and was enriched under chemical N fertilization (Additional file 2: Table S2), which might limit N use efficiency in upland soil by accelerating nitrification and denitrification [54]. Evidence of the consistent presence of such rare taxa with a potentially disproportional influence raises the possibility that

members of the rare biosphere are nevertheless indispensable in agroecosystem functioning [5]. Overall, the close association between limited indicator phylotypes and specific fertilization-induced nutrient cycling across the two-decadal duration suggests the possibility and/or feasibility of employing microbes to achieve the desired results by altering farming practices.

#### **Multifunctionality of the indicator microbiome for soil fertility and crop yield**

Ecological co-occurrence networks have been widely used to identify co-occurring microorganisms in the community and could potentially provide critical information on the multifunctionality of microbiomes [5, 23]. In this study, we identified several ecological modules comprised of the identified indicator phylotypes, in which members in each module positively correlated with each other (Fig. 3a). Importantly, these modules exhibited distinct ecosystem functions related to soil nutrients and carbon cycling (Fig. 4) and hence presented multifunctionality with soil fertility and crop yield (Fig. 3b, c). For example, module #1 was closely involved in biogeochemical processes such as nitrogen fixation, nitrification and organic phosphate decomposition (Fig. 4). N and P are the most limiting nutrients in terrestrial ecosystems; however, their availability for plants and soil biota largely depends on the depolymerization of soil organic matter and net mineralization by soil microorganisms [55]. Specifically, we found that OM treatment was associated with a higher abundance of phylotypes involved in nitrogen fixation (e.g., *R. palustris*) and nitrate reduction to ammonium, while chemical nitrogen fertilizer types enriched phylotypes active in nitrification (e.g., *N. multiformis*) and alkaline phosphatase (Fig. 4), suggesting competition of soil nutrients between soil microbes and plants under conventional chemical fertilization types [13]. *B. anthracis* in module #1 is a widely distributed soil-borne plant pathogen [56], and was found to be more enriched under nutrient-deficit fertilization (Additional file 2: Table S2), which might also account for the lower crop production in those farming systems. Accordingly, the microorganisms in module #1 were distributed in soils based on fertilization types, with their combined abundance negatively associated with soil fertility and crop yield (Fig. 3b).

In contrast, module #4 in the network exhibited solely and powerful functions in solubilizing phosphate by decomposing external organic matter (especially *B. asahii*), making it the most distinctive microbial signature for the sampling site in the North China Plain, where agricultural production is hindered by lower P availability [14]. In this context, amendment with external organic matter is necessary for promoting native microorganisms

to increase the soil P availability, termed the coupled soil C and P cycling [51]. In addition, the external C-induced higher soil P availability paved the way for some beneficial phylotypes (i.e., *L. sphaericus* [57]) that could promote plant growth or control insects and pests. As a result, we found that microorganisms in module #4 were most abundant in OM treatments (Fig. 3c), and were closely associated with higher soil fertility and crop yield (Fig. 3b). Similarly, it has been reported that OM farming systems are characterized by specific microbial guilds known to be involved in the degradation of complex organic compounds, whereas systems receiving chemical fertilizer presumably harbor oligotrophic organisms adapted to nutrient-limited environments [13]. Importantly, long-term chemical fertilization could potentially suppress inherent soil nitrogen fixation by inhabiting nitrogen fixers [58]. Deploying microorganisms to increase plant nutrient uptake and resistance to biotic and abiotic stress has been postulated as one of the most promising long-term solutions to the integral challenges of achieving food security while supporting a healthy environment [6, 10]. Our results revealed that OM farming systems were more likely to accomplish this goal compared to conventional systems receiving chemical fertilizers in alkaline soils.

#### Fidelity of the archived soil samples for the retrospective analysis

This study was based on archived soil samples collected from a long-term field experiment. We are aware that retrospective analysis of a microbial community based on archived soil samples poses a series of methodological and technical challenges. While archived samples are widely accepted for retrospective analysis of their chemical properties [59, 60], their application in retrospective microbiological studies remains controversial [61, 62]. In this study, we showed that the soil microbial communities were very well preserved to the extent that we could accurately assign, with high confidence (accuracy of 86.7%, Additional file 1: Fig. S3b) comparable to the confidence associated with fresh samples [37], a certain sample to a certain treatment based on their microbial signature. The microbial community showed a strong patterning of samples by fertilization type, with an obvious configuration according to the duration for each fertilization type (Fig. 1). Interestingly, across the duration of the field experiment, microbial communities under OM treatment were consistently separated from those that had developed under chemical fertilization and unfertilized control, confirming the frequently recorded distinct microbial diversity under conventional and OM farming systems at the time-series scale [13, 31, 63]. Strikingly, three types of unbalanced

chemical fertilization with different element deficiencies harbored a similar microbial community composition, which was distinct from their balanced counterparts. The results based on soil chemical properties were even better (accuracy of 91.5%, Additional file 1: Fig. S3a). This high fidelity gives confidence that the microbial community in archived samples reflects the original community—while possibly not a 100% accurate reflection—to the extent that they can be used to identify indicator microbiome for soil fertility and crop yield in the sampling site on the Northern China Plain.

#### Conclusions

In this study, we leveraged a retrospective analysis to investigate indicator microbiomes for soil fertility and crop yield in fields exposed to contrasting fertilization management practices. Using a stringent criterion, we identified an indicator microbiome consisting of limited but ubiquitous and abundant phylotypes, the majority of which were strongly associated with soil fertility and crop production. Furthermore, we identified two small subsets of coexisting members that exhibited opposite relationships with soil fertility and crop production in alkaline soils. One group consisted of members enriched under organic fertilization, which showed specific metabolic functions related to organic matter decomposition and was strongly associated with higher soil fertility and crop production at this type of sampling site. The other group encompassed microbial phylotypes that potentially limiting plant health and nutrient uptake and was found to be enriched under nutrient-deficit chemical fertilization practices. In summary, this study reports a “most-wanted” or “most-unwanted” list of microbial phylotypes that are feasible candidates for human manipulation in supporting crop production with explicit consideration of environmental sustainability.

#### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40793-022-00420-6>.

**Additional file 1 Supplementary information including yearly crop yield and additional statistical support for alpha and beta diversity measures:** **Figure S1:** Maize production of different fertilizer types across the duration; **Figure S2:** Comparison of community  $\alpha$ (a) and  $\beta$ (b) diversities for the 97% clustering and sequence denoising approaches; **Figure S3:** Fidelity of archived samples in current study for retrospect analysis; **Figure S4:** Co-occurrence network of indicator microbiome based on SparCC correlation. **Figure S5:** The diversity of indicator microbial phylotypes.

**Additional file 2 Supplementary information including detailed experiment setup and indicator phylotypes:** **Table S1:** Detailed fertilization regimes in this study; **Table S2:** Detailed information of indicator phylotypes (n = 604).

### Acknowledgements

The authors would like to thank Dr. Zhiying Guo for his assistance in data analysis, and Dr. Yanan Huang for her constructive comments in constructing this manuscript.

### Author contributions

Y.Z.F., J.W.Z. and J.D. developed the ideas in this article. J.B.Z., X.G.L., Y.Z.F. and R.R.C. led the soil sampling. J.W.Z., W.J.L., Y.Z.F., J.D. and R.R.C. analyzed the data. The final manuscript was written by J.W.Z., J.D. and Y.Z.F. with contributions from all the co-authors. All authors read and approved the final manuscript.

### Funding

This work is supported by the National Natural Science Foundation of China (Project No. 41771294), National Key R&D Program (2019YFC1520700) and the Youth Innovation Promotion Association, CAS (Member No. 2014271).

### Availability of data and materials

The datasets supported current study are available in the NCBI repository under the accession number PRJNA726588.

### Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, People's Republic of China. <sup>2</sup>University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China. <sup>3</sup>Faculty of Engineering and Environment, Northumbria University, Newcastle Upon Tyne, UK.

Received: 5 January 2022 Accepted: 30 April 2022

Published online: 12 May 2022

### References

- Falkowski PG, Fenchel T, DeLong EF. The microbial engines that drive Earth's biogeochemical cycles. *Science*. 2008;320(5879):1034–9. <https://doi.org/10.1126/science.1153213>.
- Delgado-Baquerizo M, Maestre FT, Reich PB, Jeffries TC, Gaitan JJ, Encinar D, et al. Microbial diversity drives multifunctionality in terrestrial ecosystems. *Nat Commun*. 2016. <https://doi.org/10.1038/ncomms10541>.
- Delgado-Baquerizo M, Reich PB, Trivedi C, Eldridge DJ, Abades S, Alfaro FD, et al. Multiple elements of soil biodiversity drive ecosystem functions across biomes. *Nat Ecol Evol*. 2020;4(2):210–20. <https://doi.org/10.1038/s41559-019-1084-y>.
- Delgado-Baquerizo M, Oliverio AM, Brewer TE, Benavent-González A, Eldridge DJ, Bardgett RD, et al. A global atlas of the dominant bacteria found in soil. *Science*. 2018;359(6373):320–5. <https://doi.org/10.1126/science.aap9516>.
- Banerjee S, Schlaeppi K, Van Der Heijden MGA. Keystone taxa as drivers of microbiome structure and functioning. *Nat Rev Microbiol*. 2018;16(9):567–76. <https://doi.org/10.1038/s41579-018-0024-1>.
- Toju H, Peay KG, Yamamichi M, Narisawa K, Hiruma K, Naito K, et al. Core microbiomes for sustainable agroecosystems. *Nat Plants*. 2018;4(5):247–57. <https://doi.org/10.1038/s41477-018-0139-4>.
- Walters WA, Jin Z, Youngblut N, Wallace JG, Sutter J, Zhang W, et al. Large-scale replicated field study of maize rhizosphere identifies heritable microbes. *Proc Natl Acad Sci U S A*. 2018;115(28):7368–73. <https://doi.org/10.1073/pnas.1800918115>.
- Egidi E, Delgado-Baquerizo M, Plett JM, Wang J, Eldridge DJ, Bardgett RD, et al. A few Ascomycota taxa dominate soil fungal communities worldwide. *Nat Commun*. 2019;10(1):2369. <https://doi.org/10.1038/s41467-019-10373-z>.
- Stegen JC, Botts EM, Jansson JK. A unified conceptual framework for prediction and control of microbiomes. *Curr Opin Microbiol*. 2018;44:20–7. <https://doi.org/10.1016/j.mib.2018.06.002>.
- Singh BK, Trivedi P, Egidi E, Macdonald CA, Delgado-Baquerizo M. Crop microbiome and sustainable agriculture. *Nat Rev Microbiol*. 2020;18(11):601–2. <https://doi.org/10.1038/s41579-020-00446-y>.
- Tilman D, Balzer C, Hill J, Befort BL. Global food demand and the sustainable intensification of agriculture. *Proc Natl Acad Sci U S A*. 2011;108(50):20260–4. <https://doi.org/10.1073/pnas.1116437108>.
- Dai ZM, Su WQ, Chen HH, Barberan A, Zhao HC, Yu MJ, et al. Long-term nitrogen fertilization decreases bacterial diversity and favors the growth of Actinobacteria and Proteobacteria in agro-ecosystems across the globe. *Glob Change Biol*. 2018;24(8):3452–61. <https://doi.org/10.1111/gcb.14163>.
- Hartmann M, Frey B, Mayer J, Mader P, Widmer F. Distinct soil microbial diversity under long-term organic and conventional farming. *ISME J*. 2015;9(5):1177–94. <https://doi.org/10.1038/ismej.2014.210>.
- Feng YZ, Chen RR, Hu JL, Zhao F, Wang JH, Chu HY, et al. Bacillus asahii comes to the fore in organic manure fertilized alkaline soils. *Soil Biol Biochem*. 2015;81:186–94. <https://doi.org/10.1016/j.soilbio.2014.11.021>.
- Fierer N, Wood SA, Bueno De Mesquita CP. How microbes can, and cannot, be used to assess soil health. *Soil Biol Biochem*. 2021. <https://doi.org/10.1016/j.soilbio.2020.108111>.
- Guerra CA, Bardgett RD, Caon L, Crowther TW, Delgado-Baquerizo M, Montanarella L, et al. Tracking, targeting, and conserving soil biodiversity. *Science*. 2021;371(6526):239–41. <https://doi.org/10.1126/science.abd7926>.
- Shade A, Handelsman J. Beyond the Venn diagram: the hunt for a core microbiome. *Environ Microbiol*. 2012;14(1):4–12. <https://doi.org/10.1111/j.1462-2920.2011.02585.x>.
- Chernov TI, Zhelezova AD. The dynamics of soil microbial communities on different timescales: a review. *Eurasian Soil Sci*. 2020;53(5):643–52. <https://doi.org/10.1134/S106422932005004x>.
- Liu W, Graham EB, Zhong L, Zhang J, Li S, Lin X, et al. Long-term stochasticity combines with short-term variability in assembly processes to underlie rice paddy sustainability. *Front Microbiol*. 2020;11:873. <https://doi.org/10.3389/fmicb.2020.00873>.
- Rasche F, Knapp D, Kaiser C, Koranda M, Kitzler B, Zechmeister-Boltenstern S, et al. Seasonality and resource availability control bacterial and archaeal communities in soils of a temperate beech forest. *ISME J*. 2011;5(3):389–402. <https://doi.org/10.1038/ismej.2010.138>.
- Lauber CL, Ramirez KS, Aanderud Z, Lennon J, Fierer N. Temporal variability in soil microbial communities across land-use types. *ISME J*. 2013;7(8):1641–50. <https://doi.org/10.1038/ismej.2013.50>.
- Fierer N, Jackson RB. The diversity and biogeography of soil bacterial communities. *Proc Natl Acad Sci U S A*. 2006;103(3):626–31. <https://doi.org/10.1073/pnas.0507535103>.
- Fan KK, Delgado-Baquerizo M, Guo XS, Wang DZ, Zhu YG, Chu HY. Biodiversity of key-stone phylotypes determines crop production in a 4-decade fertilization experiment. *ISME J*. 2021;15(2):550–61. <https://doi.org/10.1038/s41396-020-00796-8>.
- Jiao S, Xu Y, Zhang J, Hao X, Lu Y. Core microbiota in agricultural soils and their potential associations with nutrient cycling. *mSystems*. 2019. <https://doi.org/10.1128/mSystems.00313-18>.
- Takahashi M, Nakanishi H, Kawasaki S, Nishizawa NK, Mori S. Enhanced tolerance of rice to low iron availability in alkaline soils using barley nicotianamine aminotransferase genes. *Nat Biotechnol*. 2001;19(5):466–9. <https://doi.org/10.1038/88143>.
- Berdugo M, Delgado-Baquerizo M, Soliveres S, Hernandez-Clemente R, Zhao YC, Gaitan JJ, et al. Global ecosystem thresholds driven by aridity. *Science*. 2020;367(6479):787–90. <https://doi.org/10.1126/science.aay5958>.
- Slessarev EW, Lin Y, Bingham NL, Johnson JE, Dai Y, Schimel JP, et al. Water balance creates a threshold in soil pH at the global scale. *Nature*. 2016;540(7634):567–9. <https://doi.org/10.1038/nature20139>.
- Loell DB, Field CB. Global scale climate—crop yield relationships and the impacts of recent warming. *Environ Res Lett*. 2007. <https://doi.org/10.1088/1748-9326/2/1/014002>.

29. Meng L, Ding W, Cai Z. Long-term application of organic manure and nitrogen fertilizer on N<sub>2</sub>O emissions, soil quality and crop production in a sandy loam soil. *Soil Biol Biochem*. 2005;37(11):2037–45. <https://doi.org/10.1016/j.soilbio.2005.03.007>.
30. Gong W, Yan X-Y, Wang J-Y, Hu T-X, Gong Y-B. Long-term manuring and fertilization effects on soil organic carbon pools under a wheat–maize cropping system in North China Plain. *Plant Soil*. 2008;314(1–2):67–76. <https://doi.org/10.1007/s11104-008-9705-2>.
31. Liang YT, Ning DL, Lu ZM, Zhang N, Hale L, Wu LY, et al. Century long fertilization reduces stochasticity controlling grassland microbial community succession. *Soil Biol Biochem*. 2020. <https://doi.org/10.1016/j.soilbio.2020.108023>.
32. Schwieger F, Tebbe CC. A new approach to utilize PCR-single-strand-conformation polymorphism for 16S rRNA gene-based microbial community analysis. *Appl Environ Microbiol*. 1998;64(12):4870–6. <https://doi.org/10.1128/AEM.64.12.4870-4876.1998>.
33. Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods*. 2013;10(10):996–8. <https://doi.org/10.1038/Nmeth.2604>.
34. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol*. 2007;73(16):5261–7. <https://doi.org/10.1128/Aem.00062-07>.
35. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*. 2010;7(5):335–6. <https://doi.org/10.1038/nmeth.f.303>.
36. Smith MB, Rocha AM, Smillie CS, Olesen SW, Paradis C, Wu LY, et al. Natural bacterial communities serve as quantitative geochemical biosensors. *MBio*. 2015. <https://doi.org/10.1128/mBio.00326-15>.
37. Hermans SM, Buckley HL, Case BS, Curran-Cournane F, Taylor M, Lear G. Using soil bacterial communities to predict physico-chemical variables and soil quality. *Microbiome*. 2020. <https://doi.org/10.1186/s40168-020-00858-1>.
38. Archer E. rFPermute: estimate permutation p-values for random forest importance metrics. 2020.
39. Benjamini Y, Hochberg Y. Controlling the false discovery rate—a practical and powerful approach to multiple testing. *J R Stat Soc B*. 1995;57(1):289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
40. Widder S, Besemer K, Singer GA, Ceola S, Bertuzzo E, Quince C, et al. Fluvial network organization imprints on microbial co-occurrence networks. *Proc Natl Acad Sci U S A*. 2014;111(35):12799–804. <https://doi.org/10.1073/pnas.1411723111>.
41. Barberan A, Bates ST, Casamayor EO, Fierer N. Using network analysis to explore co-occurrence patterns in soil microbial communities. *ISME J*. 2012;6(2):343–51. <https://doi.org/10.1038/ismej.2011.119>.
42. R Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2013. <http://www.R-project.org/>.
43. De Menezes AB, Prendergast-Miller MT, Richardson AE, Toscas P, Farrell M, Macdonald LM, et al. Network analysis reveals that bacteria and fungi form modules that correlate independently with soil parameters. *Environ Microbiol*. 2015;17(8):2677–89. <https://doi.org/10.1111/1462-2920.12559>.
44. Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. In: Proceedings of the international AAAI conference on web and social media; 2009. <https://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154>.
45. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome datasets are compositional: and this is not optional. *Front Microbiol*. 2017;8:2224. <https://doi.org/10.3389/fmicb.2017.02224>.
46. Peschel S, Müller CL, Von Mutius E, Boulesteix A-L, Depner M. NetCoMi: network construction and comparison for microbiome data in R. *Brief Bioinform*. 2021;22(4):bbaa290. <https://doi.org/10.1093/bib/bbaa290>.
47. Kurtz ZD, Müller CL, Miraldi ER, Littman DR, Blaser MJ, Bonneau RA. Sparse and compositionally robust inference of microbial ecological networks. *PLoS Comput Biol*. 2015;11(5): e1004226. <https://doi.org/10.1371/journal.pcbi.1004226>.
48. Kim M, Oh HS, Park SC, Chun J. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int J Syst Evol Microbiol*. 2014;64:346–51. <https://doi.org/10.1099/ijs.0.059774-0>.
49. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res*. 2019;47(W1):W256–9. <https://doi.org/10.1093/nar/gkz239>.
50. Pohl S, Harwood CR. Heterologous protein secretion by *Bacillus* species: from the cradle to the grave. In: Laskin AI, Sariaslani S, Gadd GM, editors. *Advances in applied microbiology*. London: Academic Press; 2010. p. 1–25.
51. Jing ZW, Chen RR, Wei SP, Feng YZ, Zhang JB, Lin XG. Response and feedback of C mineralization to P availability driven by soil microorganisms. *Soil Biol Biochem*. 2017;105:111–20. <https://doi.org/10.1016/j.soilbio.2016.11.014>.
52. Dourado MN, Neves AAC, Santos DS, Araujo WL. Biotechnological and agronomic potential of endophytic pink-pigmented methylotrophic *Methylobacterium* spp. *Biomed Res Int*. 2015. <https://doi.org/10.1155/2015/909016>.
53. Hsu SH, Shen MW, Chen JC, Lur HS, Liu CT. The photosynthetic bacterium *Rhodospseudomonas palustris* strain PS3 exerts plant growth-promoting effects by stimulating nitrogen uptake and elevating auxin levels in expanding leaves. *Front Plant Sci*. 2021. <https://doi.org/10.3389/fpls.2021.573634>.
54. Alonso-Ayuso M, Gabriel JL, Quemada M. Nitrogen use efficiency and residual effect of fertilizers with nitrification inhibitors. *Eur J Agron*. 2016;80:1–8. <https://doi.org/10.1016/j.eja.2016.06.008>.
55. Gao DC, Bai E, Li MH, Zhao CH, Yu KL, Hagedorn F. Responses of soil nitrogen and phosphorus cycling to drying and rewetting cycles: a meta-analysis. *Soil Biol Biochem*. 2020. <https://doi.org/10.1016/j.soilbio.2020.107896>.
56. Pohl S, Harwood CR. Heterologous protein secretion by *Bacillus* species: from the cradle to the grave. *Adv Appl Microbiol*. 2010;73:1–25. [https://doi.org/10.1016/S0065-2164\(10\)73001-X](https://doi.org/10.1016/S0065-2164(10)73001-X).
57. Berry C. The bacterium, *Lysinibacillus sphaericus*, as an insect pathogen. *J Invertebr Pathol*. 2012;109(1):1–10. <https://doi.org/10.1016/j.jip.2011.11.008>.
58. Fan KK, Delgado-Baquerizo M, Guo XS, Wang DZ, Wu YY, Zhu M, et al. Suppressed N fixation and diazotrophs after four decades of fertilization. *Microbiome*. 2019. <https://doi.org/10.1186/s40168-019-0757-8>.
59. Blake L, Goulding KWT, Mott CJB, Johnston AE. Changes in soil chemistry accompanying acidification over more than 100 years under woodland and grass at Rothamsted Experimental Station, UK. *Eur J Soil Sci*. 1999;50(3):401–12. <https://doi.org/10.1046/j.1365-2389.1999.00253.x>.
60. Blake L, Goulding KWT, Mott CJB, Poulton PR. Temporal changes in chemical properties of air-dried stored soils and their interpretation for long-term experiments. *Eur J Soil Sci*. 2000;51(2):345–53. <https://doi.org/10.1046/j.1365-2389.2000.00307.x>.
61. Benucci GMN, Rennick B, Bonito G. Patient propagules: do soil archives preserve the legacy of fungal and prokaryotic communities? *PLoS ONE*. 2020. <https://doi.org/10.1371/journal.pone.0237368>.
62. Tzenea VA, Salles JF, Naumova N, De Vos WA, Kuikman PJ, Dolfing J, et al. Effect of soil sample preservation, compared to the effect of other environmental variables, on bacterial and eukaryotic diversity. *Res Microbiol*. 2009;160(2):89–98. <https://doi.org/10.1016/j.resmic.2008.12.001>.
63. Schmidt JE, Kent AD, Brisson VL, Gaudin ACM. Agricultural management and plant selection interactively affect rhizosphere microbial community structure and nitrogen cycling. *Microbiome*. 2019. <https://doi.org/10.1186/s40168-019-0756-9>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.