

# Soil pH and temperature regulate assembly processes of abundant and rare bacterial communities in agricultural ecosystems

Shuo Jiao  and Yahai Lu <sup>\*</sup>

College of Urban and Environmental Sciences, Peking University, Beijing 100871, P. R. China.

## Summary

**The factors determining stochastic and deterministic processes that drive microbial community structure, specifically the balance of abundant and rare bacterial taxa, remain underexplored. Here we examined biogeographic patterns of abundant and rare bacterial taxa and explored environmental factors influencing their community assembly processes in agricultural fields across eastern China. More phylogenetic turnover correlating with spatial distance was observed in abundant than rare sub-communities. Homogeneous selection was the main assembly process for both the abundant and rare sub-communities; however, the abundant sub-community was more tightly clustered phylogenetically and was more sensitive to dispersal limitations than the rare sub-community. Rare sub-community of rice fields and abundant sub-community of maize fields were more governed by stochastic assembly processes, which showed higher operational taxonomic unit richness. We propose a conceptual paradigm wherein soil pH and mean annual temperature mediate the assembly of the abundant and rare sub-communities respectively. A higher soil pH leads to deterministic assembly of the abundant sub-community. For the rare sub-community, the dominance of stochasticity in low-temperature regions indicates weaker niche-based exclusion and the arrival of more evolutionary lineages. These findings suggest that the community assembly processes for abundant and rare bacterial taxa are dependent on distinct environmental variables in agro-ecosystems.**

## Introduction

Microbial communities, which are highly diverse at both taxonomic and metabolic levels, are a crucial part of global biogeochemical processes (Prosser *et al.*, 2007; Schloss and Handelsman, 2007; Falkowski *et al.*, 2008). Biogeography describes the distribution of taxa over space and time (Meyer *et al.*, 2018). Understanding microbial biogeography has led to fundamental insights into the maintaining and generating species diversity. It can also help determine the links between community stability and ecosystem functions (Hanson *et al.*, 2012; Meyer *et al.*, 2018). Microbial communities commonly exhibit unbalanced distributions, whereby they contain a large number of low-abundance taxa and a small number of high-abundance taxa (Hanson *et al.*, 2012; Brown *et al.*, 2014; Jousset *et al.*, 2017). These low-abundance taxa are often referred to as the 'rare biosphere' (Sogin *et al.*, 2006; Pedrós-Alió, 2012). The highly abundant microbes that account for the majority of microbial biomass are thought to be most important in carbon cycling (Pedrós-Alió, 2012; Wu *et al.*, 2017). In oil-contaminated soils, abundant taxa had closer relationships with and a greater influence on the co-occurrence of other taxa than rare taxa did (Jiao *et al.*, 2017a). However, recent studies have increasingly emphasized the importance of rare taxa (Lynch and Neufeld, 2015). For example, rare bacterial species have been shown to play fundamental roles in sulphate reduction (Pester *et al.*, 2010) and phenanthrene degradation (Sauret *et al.*, 2014), as well as mediating ecosystem stability and function (Elshahed *et al.*, 2008; Reid and Buckley, 2011). Currently, little is known about the fundamental processes underlying the microbial biogeography of abundant and rare taxa, as well as the mechanisms controlling rare taxa (Jia *et al.*, 2018).

Microbial community ecology is progressing from the description of patterns towards a mechanistic understanding the underlying processes in microbial community assembly (Ovaskainen *et al.*, 2017). There is an ongoing debate regarding the relative contributions of deterministic (niche-based theory) and stochastic (neutral theory) processes in the assembly of microbial communities (Stegen *et al.*, 2012; Wang *et al.*, 2013; Dini-

Received 15 July, 2019; revised 27 September, 2019; accepted 30 September, 2019. <sup>\*</sup>For correspondence. E-mail luyh@pku.edu.cn; Tel. +86 29 87091175; Fax +86 29 87091175.

Andreote *et al.*, 2015). Over the last decade, high-throughput sequencing technologies have allowed researchers to explore microbial community assembly in the rare biosphere (Logares *et al.*, 2012). Indeed, studies have been conducted in a wide range of environments, such as marine water (Logares *et al.*, 2014; Wu *et al.*, 2017; Mo *et al.*, 2018), inland freshwater (Liu *et al.*, 2015) and oil-contaminated soil (Jiao *et al.*, 2017a). In oil-contaminated soil, the rare subcommunity is primarily governed by stochastic processes; meanwhile, deterministic processes are the primary contributor to the structure of the abundant sub-community (Jiao *et al.*, 2017a). However, the opposite results were observed in inland freshwater (Liu *et al.*, 2015). These outcomes indicate that microbial assembly processes can differ between terrestrial and aquatic ecosystems. Agricultural fields are human-managed ecosystems that provide global food security for a growing world population (Mondiale, 2008). Due to contrasting water management practices, agricultural soils can be divided into unsaturated (e.g. corn) and saturated (e.g. paddy) fields that have distinct microbial diversity patterns. Nevertheless, studies assessing the mechanisms that influence the community assembly of abundant and rare bacterial taxa in unsaturated and saturated agro-ecosystems remain scarce.

To advance mechanistic understanding of community assembly processes, the factors that mediate the balance between stochastic and deterministic processes must be elucidated (Feng *et al.*, 2018). A study showed that soil pH influences the balance between stochastic and deterministic processes in shaping soil bacterial community assembly during the paedogenic processes (Tripathi *et al.*, 2018). In addition, it has been observed that increases in soil organic matter (OM) lead to a transition from homogeneous selection to weak selection and stochasticity, and then to variable selection during primary ecosystem succession (Dini-Andreote *et al.*, 2015). The initial amount of and the changes in soil OM due to nutrient addition influence the relative contributions of different assembly processes (Feng *et al.*, 2018). However, it is still unclear what underlying environmental factors lead to the dominance of particular assembly processes for abundant and rare bacterial taxa in agro-ecosystems.

In the present study, we aim to (i) explore the biogeographic patterns of abundant and rare bacterial taxa in agro-ecosystems and (ii) uncover the underlying environmental factors regulating assembly processes of abundant and rare bacterial taxa. These questions could be well addressed based on the existing high-throughput sequencing data sets obtained from adjacent pairs of maize (water unsaturated) and rice (water saturated) cultivated fields across eastern China (Jiao *et al.*, 2019). Our results highlight the importance of considering the different linkages between environmental conditions and

ecological processes in governing the assembly of abundant and rare taxa in agro-ecosystems.

## Results

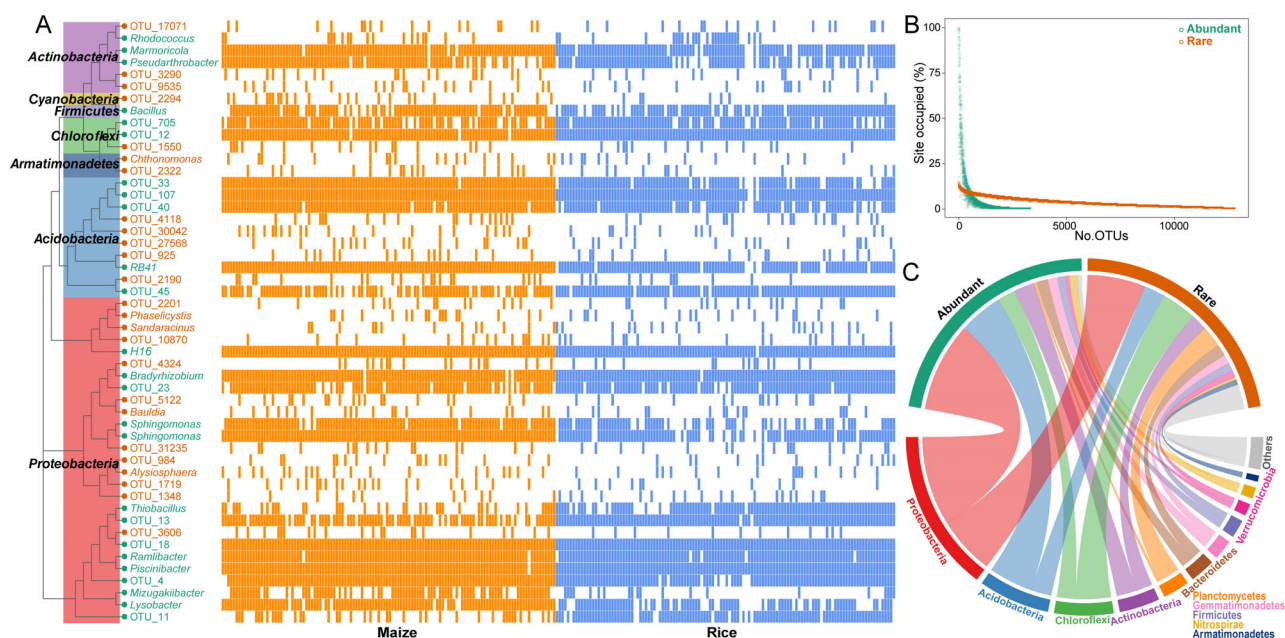
### General patterns of abundant and rare taxa

There was large phylogenetic diversity in both the rare and abundant sub-communities (Fig. 1A and C). The abundant taxa were primarily comprised of *Proteobacteria*, *Acidobacteria* and *Actinobacteria*; meanwhile, the rare taxa were primarily comprised of *Chloroflexi*, *Planctomycetes* and *Firmicutes*. The rare taxa contained more taxonomic groups than the abundant taxa. In addition, the abundant taxa had greater presence across soil samples than the rare taxa (Fig. 1A and B). Some abundant taxa were present in 100% of samples; meanwhile, none of the rare taxa were present in > 20% of sites. To explore the distributions of rare and abundant taxa across soil samples we calculated the per cent changes of rare or abundant taxa in each sample to determine whether these taxa were abundant, intermediate, rare or not detected in other samples (Fig. S1). Of the taxa that were locally abundant, 35.3% were also abundant and only 2.4% were rare in other samples. Of the taxa that were locally rare, 77.7% were not detected, 5.2% were still rare and almost none (0.9%) were abundant in the other samples. These results indicate distinct distribution patterns of abundant and rare taxa in agricultural fields.

In general, the rare taxa accounted for a high proportion of the operational taxonomic units (OTUs) (mean = 63.2%); however, they only accounted for an average of 11.5% of the relative abundance in each sample (Fig. S2A). The abundant taxa accounted for a very low proportion of OTUs (mean = 4.7%); however, they accounted for 48.6% of the average relative abundance in each sample. The abundant sub-community had a higher OTU richness and relative abundance in maize soils than rice soils; the rare sub-community exhibited the opposite trend (i.e. lower OTU richness and relative abundance in maize soils than rice soils; Fig. S2B). Modelling using kriging interpolation showed that both the abundant and rare sub-communities had higher OTU richness in northern fields than southern fields (Fig. S2C and D).

### Turnover of abundant and rare taxa

The mean values of standardized effect size measure of the mean nearest taxon distance (SES.MNTD) were significantly higher for abundant than rare taxa, which were all less than zero (Wilcoxon rank-sum test,  $p < 0.001$ , Fig. S3). Significant distance-decay relationships (DDRs) were observed for the abundant and rare sub-communities ( $p < 0.05$ ; Fig. S4A). The slopes of the DDRs for abundant



**Fig. 1.** Taxonomic and occupied distribution of the abundant and rare bacterial sub-communities in agricultural soils.

A. Phylogenetic distribution of the top 25 most dominant taxa in rare and abundant sub-communities and their local occupancies in maize and rice soils. The phylogenetic tree constructed using the neighbour-joining method. Taxa that could be assigned to genus level were shown as genus, otherwise were shown as OTU id.

B. The proportion of sites occupied for each abundant and rare taxa.

C. Taxonomic distribution of abundant and rare taxa at phylum level shown as chord plot. The thickness of each ribbon represents the number of abundant and rare taxa assigned to different phyla. [Color figure can be viewed at [wileyonlinelibrary.com](http://onlinelibrary.wiley.com)]

taxa were steeper than those for rare taxa; this indicates that the phylogenetic  $\beta$ -diversity variation is greater in the abundant sub-community and increases with geographic distance. In addition, the abundant sub-community exhibited a significantly higher phylogenetic  $\beta$ -diversity than the rare sub-community (Fig. S4B).

We performed a distance-based linear model and forward selection procedure to identify the major environmental variables responsible for shaping the abundant and rare sub-communities. A constrained analysis of principal coordinates (CAP) revealed that soil pH and mean annual temperature (MAT) were the most important variables for the community assembly of abundant and rare taxa respectively (Fig. 2, Tables S1 and S2). Similar trends were observed separately in maize and rice soils (Fig. S5, Tables S3–S6).

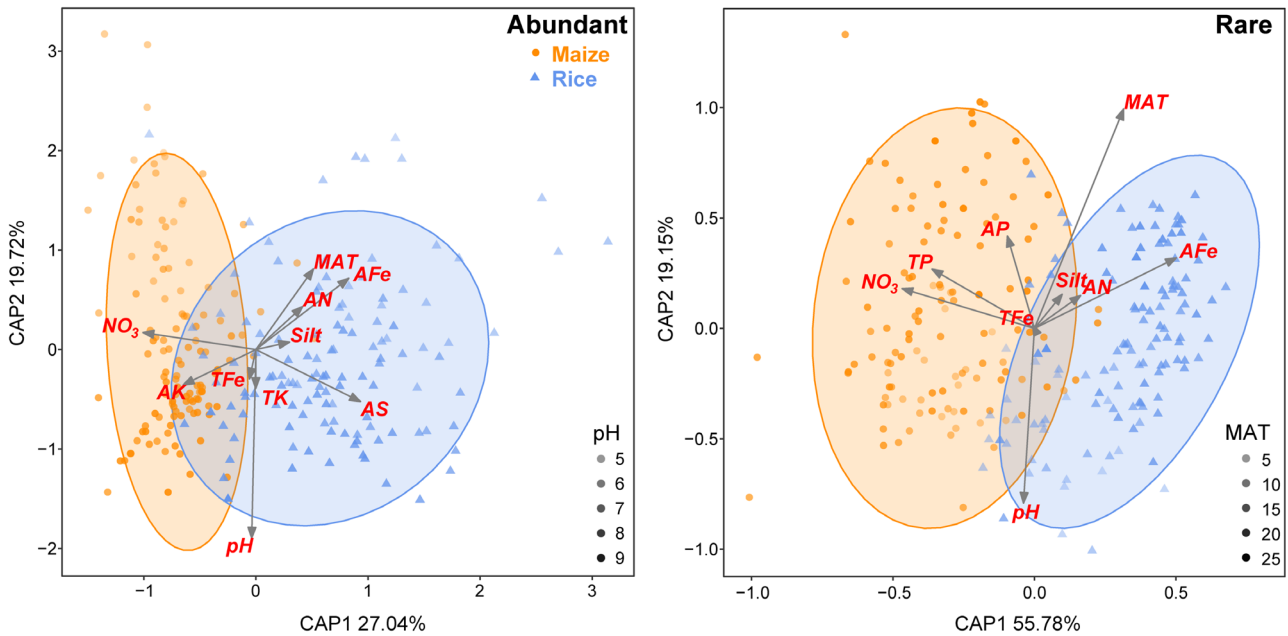
The strength of the relationship between phylogeny and environmental response was examined to determine whether environmental responses could be predictive of phylogenetic levels. Spearman's correlations between relative abundances of bacterial taxa and environmental variables (soil pH or MAT) across all of the soil samples were used to identify 'alkaline-responsive' or 'acid-responsive' indicator taxa, as well as 'warm-responsive' or 'cold-responsive' indicator taxa. In the abundant sub-community, 426, 518, 380 and 311 OTUs were identified as 'alkaline-responsive', 'acid-responsive', 'warm-responsive' and 'cold-

responsive' indicator taxa, respectively; in the rare sub-community, 506, 571, 564 and 626 OTUs were identified as 'alkaline-responsive', 'acid-responsive', 'warm-responsive' and 'cold-responsive' indicator taxa respectively. Indicator taxa mainly belonged to the *Proteobacteria*, *Chloroflexi*, *Acidobacteria* and *Actinobacteria* phyla (Fig. S6).

We assessed trait conservatism in response to environmental changes (Table 1). The abundant sub-community exhibited a stronger and significant phylogenetic signal for soil pH ( $K = 0.091$ ;  $p < 0.001$ ) than for MAT ( $K = 0.078$ ;  $p < 0.001$ ). This indicates that within abundant taxa, closely related species had more similar responses to soil pH than MAT. By contrast, the phylogenetic signals of rare taxa were stronger for MAT ( $K = 0.068$ ;  $p < 0.001$ ) than for soil pH ( $K = 0.067$ ;  $p < 0.001$ ). Phylum-level taxonomy confirmed these observations (Table 1). For instance, most phyla in the abundant sub-community had a stronger phylogenetic signal for soil pH than for MAT; meanwhile, most rare phyla, including *Proteobacteria*, *Gemmatimonadetes* and *Verrucomicrobia*, had a stronger phylogenetic signal for MAT than for pH.

#### Relative influence of assembly processes in rare and abundant sub-communities

The relationships between  $\beta$ -nearest taxon index ( $\beta$ NTI) and major environmental variables were used to infer



**Fig. 2.** Major environmental variables in shaping the abundant and rare phylogenetic sub-communities based on MNTD metric ( $\beta$ MNTD) identified by CAP. The abbreviations of environmental factors according to "Methods" section. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

**Table 1.** Phylogenetic signal showing the level of trait conservatism for environmental response of all members ('All') and major phyla in abundant and rare bacterial communities using Blomberg's *K*.

	Abundant		Rare	
	pH	MAT	pH	MAT
All	<b>0.091***</b>	<b>0.078***</b>	<b>0.067**</b>	<b>0.068***</b>
<i>Proteobacteria</i>	<b>0.104***</b>	0.074	0.078	<b>0.081*</b>
<i>Chloroflexi</i>	<b>0.103***</b>	<b>0.090*</b>	<b>0.073**</b>	0.069
<i>Acidobacteria</i>	<b>0.174***</b>	<b>0.133***</b>	0.084	0.089
<i>Actinobacteria</i>	<b>0.124*</b>	<b>0.130**</b>	0.125	0.092
<i>Bacteroidetes</i>	0.077	0.047	0.052	0.066
<i>Gemmatimonadetes</i>	0.105	0.140*	0.085	<b>0.099*</b>
<i>Planctomycetes</i>	<b>0.105**</b>	0.080	<b>0.074***</b>	0.064
<i>Firmicutes</i>	0.102	<b>0.118*</b>	0.069	0.089
<i>Verrucomicrobia</i>	0.091	0.080	0.099	<b>0.127***</b>

\* $p \leq 0.05$ ; \*\* $p \leq 0.01$ ; \*\*\* $p \leq 0.001$ .

The Bold values represent significant ( $p < 0.05$ ) Blomberg's *K*.

changes in the relative influences of deterministic and stochastic assembly processes. Partial mantel test showed that soil pH and MAT were, respectively, the best predictor of assembly processes of the abundant and the rare sub-communities, and the relationships remained significant after controlling for spatial distance and other measured environmental variables (Table 2). Pairwise comparisons of  $\beta$ NTI values for the abundant and rare sub-communities were significantly and positively correlated with differences in soil pH and MAT respectively (Fig. 3A). This indicated that an increasing divergence of soil pH and MAT, respectively, led to a shift from homogeneous selection to stochasticity in abundant and rare bacterial community assembly.

Samples were separated into sub-groups based on soil pH and MAT and correlated with  $\beta$ NTI. In the abundant sub-community, the relative influence of homogeneous selection increased with soil pH. In the rare sub-community, the relative influence of stochastic assembly decreased with MAT (Fig. 3B). Similar results were observed in both maize and rice soils (Figs S7 and S8, Table 2).

#### Quantitative analysis of assembly processes in abundant and rare sub-communities

Homogeneous selection contributed the largest fraction to the assembly of both abundant (45.0%) and rare (65.8%) sub-communities, with higher contributions for rare sub-communities (Fig. 4). Dispersal limitations (31.0%) and variable selection (9.3%) influenced the abundant sub-community more than the rare sub-community (dispersal limitations: 3.1%, variable selection: 0.2%). The neutral community model explained a larger fraction of variation in the abundant sub-community than the rare sub-community, which were observed in both maize and rice soils (Table 3). Then, we compared the different assembly processes of maize versus rice (Figs 5). For abundant taxa, deterministic processes contributed a larger fraction to the community assembly in rice (64.0%) than maize (61.7%) fields. While stochastic assembly processes influenced the rare sub-community in rice fields (35.0%) fivefold more than that in maize fields (7.1%).

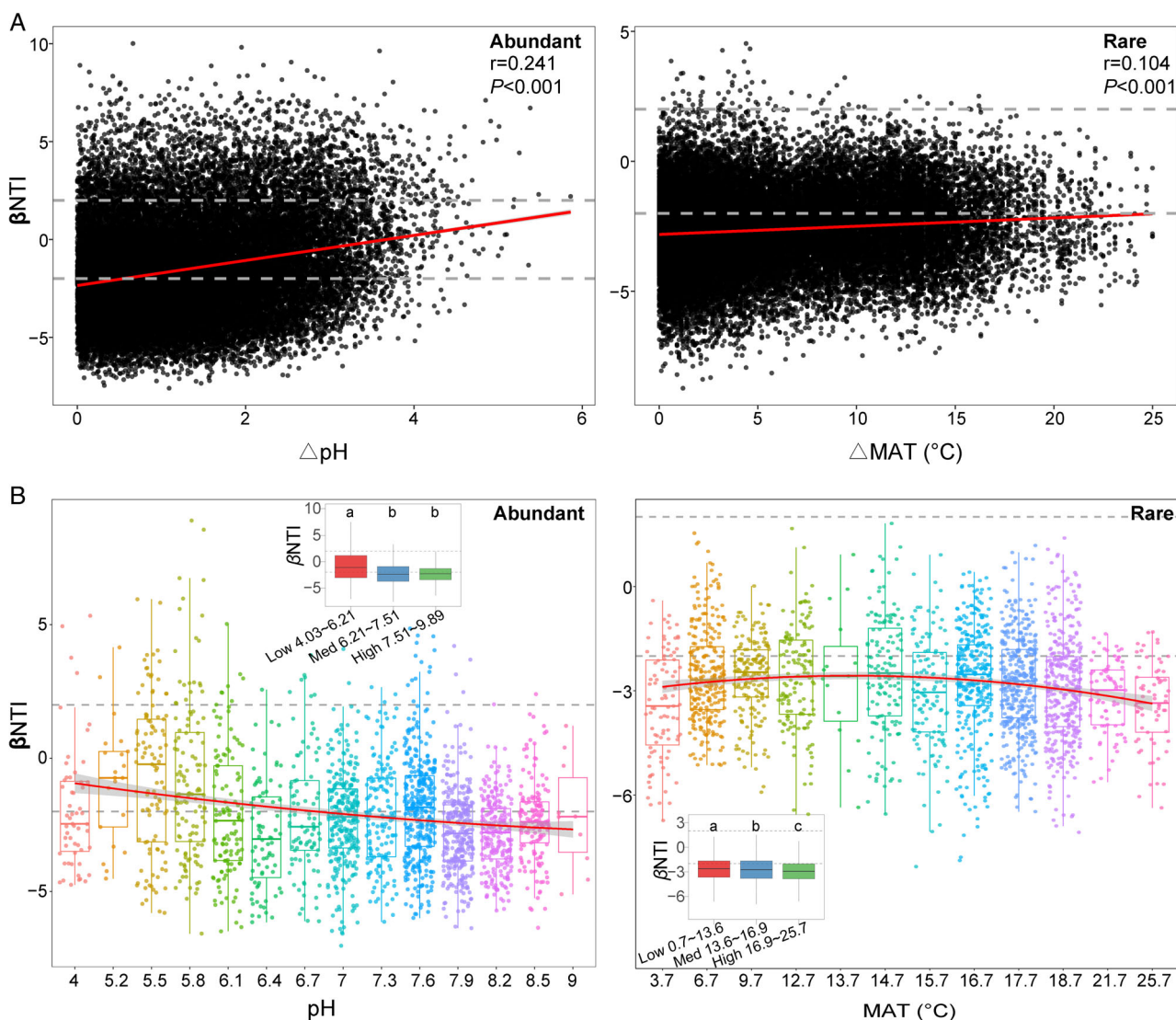


**Table 2.** Partial mantel test results showing relationship between phylogenetic turnover ( $\beta$ NTI) versus soil pH/MAT and spatial and environmental distances for abundant/rare subcommunities after controlling for the other two explanatory distance constant in all ('Whole'), maize ('Maize') and rice ('Rice') soil samples.

	Effect of selected variables on $\beta$ NTI	Controlling for	Whole	Maize	Rice
Abundant	Soil pH	Spatial + environmental (excluding pH)	<b>0.232***</b>	<b>0.300***</b>	<b>0.246***</b>
	Spatial	Environmental (excluding pH) + soil pH	<b>0.049*</b>	<b>0.083*</b>	<b>0.119**</b>
	Environmental (excluding pH)	Spatial + soil pH	<b>0.069*</b>	0.007	0.077
Rare	MAT	Spatial + environmental (excluding MAT)	<b>0.034*</b>	<b>0.089*</b>	<b>0.199***</b>
	Spatial	Environmental (excluding MAT) + MAT	0.028	−0.007	0.001
	Environmental (excluding MAT)	Spatial + MAT	0.012	−0.045	0.005

\* $p \leq 0.05$ ; \*\* $p \leq 0.01$ ; \*\*\* $p \leq 0.001$ .

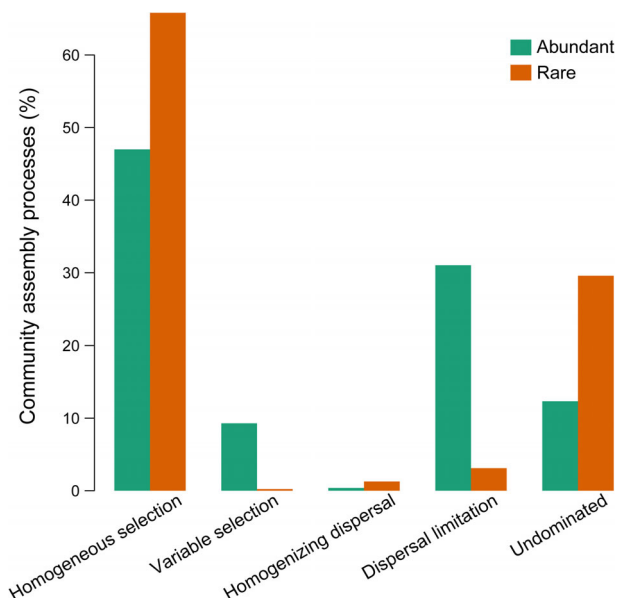
The Bold values represent significant ( $p < 0.05$ ) Blomberg's K.



**Fig. 3.** Relative influence of deterministic and stochastic assembly processes in agricultural soils.

A. The relationships between  $\beta$ NTI and differences in soil pH for abundant bacterial sub-community and in MAT for rare sub-community. Linear regressions models (shown as red lines) and associated correlation coefficients are provided on each panel. Horizontal dashed lines indicate the  $\beta$ NTI significance thresholds of +2 and -2.

B. Patterns of  $\beta$ NTI across different categories in soil pH for abundant sub-community and MAT for rare sub-community. Boxplots within the panels showed the variation in  $\beta$ NTI across low, medium ('Med') and high categories of soil pH and MAT. Boxplots that do not share a letter are significantly different ( $p < 0.05$ ; multiple comparisons with Kruskal–Wallis). [Color figure can be viewed at wileyonlinelibrary.com]



**Fig. 4.** The per cent of turnover in abundant and rare bacterial community assembly governed primarily by various deterministic, including homogeneous and variable selection, and stochastic processes, including dispersal limitations and homogenizing dispersal, as well as the fraction that was not dominated by any single process ('Undominated') in agricultural soils. [Color figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

**Table 3.** Fit of the neutral model in abundant and rare bacterial subcommunities for all ('Whole'), maize ('Maize') and rice ('Rice') soil samples.

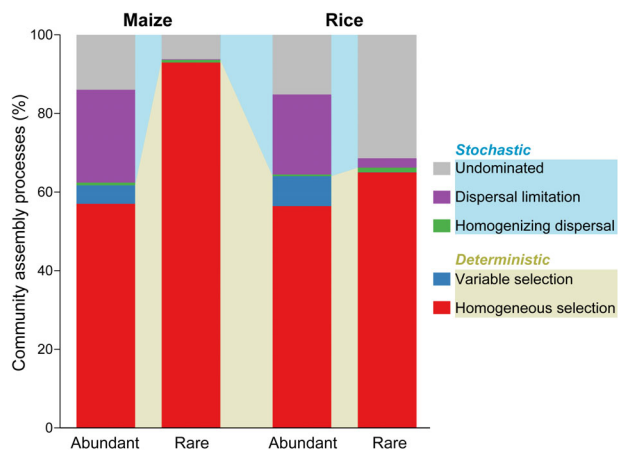
	Abundant		Rare	
	<i>m</i>	<i>R</i> <sup>2</sup>	<i>m</i>	<i>R</i> <sup>2</sup>
Whole	0.0036	0.9264	0.4626	−0.5150
Maize	0.0037	0.9290	0.4692	−0.3843
Rice	0.0042	0.9320	0.4655	−0.5118

*R*<sup>2</sup> and *m* values indicate the fit to the neutral model and the estimated migration rate respectively.

## Discussion

Knowledge of the assembly processes driving ecological communities is essential for ecosystem management (Hanson *et al.*, 2012; Pagaling *et al.*, 2014; Meyer *et al.*, 2018). Determining the biogeographic patterns and community assembly of abundant and rare microbial taxa is useful for understanding the generation and maintenance of ecosystem diversity (Pedrós-Alió, 2012; Lynch and Neufeld, 2015; Jia *et al.*, 2018). Here, we show that (i) abundant and rare bacterial diversity exhibits distinct biogeographic and ecological patterns that are driven by different assembly processes, and (ii) soil pH and MAT predominately mediate the assembly of abundant and rare sub-communities, respectively, in agricultural fields.

Novel sequencing technologies have provided increasing resolutions to shed light on the role of the rare



**Fig. 5.** The per cent of turnover in abundant and rare bacterial community assembly governed primarily by various deterministic, including homogeneous and variable selection, and stochastic processes, including dispersal limitations and homogenizing dispersal, as well as the fraction that was not dominated by any single process ('Undominated') in maize and rice soils. [Color figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

biosphere, which is currently one of the new frontiers in microbial ecology (Jousset *et al.*, 2017). In the present study, abundant taxa were more ubiquitous than rare taxa, suggesting that rare taxa are mediated by the presence of distinct ecological niches in agricultural ecosystems. This conclusion is supported by the notion that abundant species occupy a diverse niche, competitively utilize an array of resources, and effectively adapt to the environment (Jiao *et al.*, 2017a). Rare taxa, meanwhile, are restricted by habitat specificity (Barberan *et al.*, 2014; Jousset *et al.*, 2017). A study suggested that rare taxa may act as a 'seed bank' for taxa that can become dominant under the proper conditions (Lennon and Jones, 2011). However, in our study only a few taxa shifted from rare to abundant, and most rare taxa (77.7%) were non-existent or below the limits of detection in other samples. This implies that rare species might not be waiting for favourable conditions based on the homogenized habitats in maize or rice fields. Rice soils possessed more rare taxa than maize soils, which could be due to the presence of both aerobic and anaerobic taxa due to saturation–unsaturation cycles in rice fields (Jiao *et al.*, 2019). Rare taxa, which are more abundant in rice fields, can increase functional redundancy and resiliency to environmental disturbances (Yachi and Loreau, 1999; Pedrós-Alió, 2012; Lynch and Neufeld, 2015).

The globally biogeographic microbial assemblages could have potential to be efficient biological indicators for the environmental health of the complex natural systems (Astudillo-Garcia *et al.*, 2019). Therefore, revealing microbial biogeography and mechanisms underlying community assembly is a central issue in microbial ecology (Nemergut *et al.*, 2013). In our agricultural soils, both

abundant and rare sub-communities had robust phylogenetic DDRs, which indicate the spatial structure for abundant and rare bacteria; this is consistent with previous findings (Jiao *et al.*, 2017a; Mo *et al.*, 2018). Soil pH had the strongest influence on the structure of the abundant sub-community among all environmental factors tested; by contrast, the rare sub-community was most highly correlated with MAT. Studies have demonstrated that different controlling factors constrain the structures of the rare and abundant sub-communities (Liu *et al.*, 2015; Jiao *et al.*, 2017a). Additionally, the significantly lower mean values of SES.MNTD for rare taxa indicate that these taxa are more closely phylogenetically clustered than the abundant sub-community (Fan *et al.*, 2017). The negative SES.MNTD values indicate that both abundant and rare taxa were phylogenetically clustered more closely than expected by chance. Abundant and rare taxa also exhibited more phylogenetic niche conservation and ecological similarity than soil pH and MAT respectively. This implies that the sub-communities have distinct phylogenetic adaptations to different environmental conditions. Although the phylogenetic signals were significant, the *K* values were all below 1 and close to 0. This indicates that closely related species were actually less correlated with the traits (e.g. soil pH, temperature) than expected under the Brownian motion model of trait evolution, which mimics a random or convergent pattern of evolution (Blomberg *et al.*, 2003).

A central challenge in biogeography is quantifying the contributions of deterministic and stochastic processes to microbial community assembly (Wang *et al.*, 2015). The neutral and niche-based mechanisms could lead to microbial communities with low-stability properties, for example, resistance and recovery (Liu *et al.*, 2019b). Analysis of phylogenetic and taxonomic  $\beta$ -diversity using a null-model approach provides a systematic framework to classifying community pairs based on the underlying forces driving community assembly (Stegen *et al.*, 2013). We can also make ecological inferences using phylogenetic turnover (Cavender-Bares *et al.*, 2009; Tripathi *et al.*, 2018). Microbial diversity of extreme ecosystems, like Arctic soils, could be substantially different, while other non-extreme environments, for example, forest and agricultural fields, may harbour similar communities (Malard and Pearce, 2018). Long-term crop cultivation (e.g. maize, rice) and irrigation (e.g. unsaturated, saturated) results in homogenized habitats in agricultural fields. As expected, homogeneous selection was found to be the main process driving the assembly of both abundant and rare sub-communities irrespective of crop type (i.e. maize or rice). Recent studies have demonstrated that deterministic processes drive community assembly in wheat (Shi *et al.*, 2018) and soybean (Zhang *et al.*, 2018) fields. Homogenizing selection, estimated by

null model analysis, was predicted to have played a major role in the assembly of soil microbial communities, while dispersal limitation was a dominant factor structuring mineral-associated communities in a rhizosphere microcosm (Whitman *et al.*, 2018). Moreover, Ogonowski and colleagues (2018) provided experimental evidences for the significant effect of selection process in the structuring of biofilm microbial community assemblages and demonstrated that differences in biofilm community structure were strongly related to the variation in substrate hydrophobicity.

In our study, dispersal significantly influenced the abundant sub-community but not the rare sub-community. This result is supported by a study showing that abundant taxa were more limited by dispersal than rare taxa in the surface layer of the north-western Pacific Ocean (Wu *et al.*, 2017). Other studies have shown that rare taxa are more limited by dispersal than abundant taxa in oil-contaminated soils (Jiao *et al.*, 2017a) and subtropical bays (Mo *et al.*, 2018). These discrepancies may be due to the difference in habitats (Wang *et al.*, 2013, 2017) and geography (Shi *et al.*, 2018). Langenheder and Lindstrom (2019) proposed a conceptual overview to predict changes in the relative importance of community assembly processes in dependence of spatial scale, and demonstrated the important roles of spatial scale in the assembly processes. In addition, different definitions of rare and abundant taxa could partly result in the different results (Langenheder and Lindstrom, 2019). The high fraction of non-dominant processes (29.6%) that contributed to the assembly of the rare sub-community suggests that more complex assembly mechanisms may shape the rare sub-community (Mo *et al.*, 2018). Neutral-based processes are another force that shapes the microbial community (Hubbell, 2001; Dumbrell *et al.*, 2010; Langenheder and Szekely, 2011). Neutral mechanisms were found to influence the assembly of microbial communities with dominated strength in steady-state insular environments (Liu *et al.*, 2019b). Our neutral model explained a large fraction ( $R^2 > 0.9$ ) of the variability in occurrence frequency of the abundant sub-community, indicating that neutral processes have a strong role in shaping the abundant sub-community but not the rare sub-community. Multiple analyses support the notion that different assembly processes drive the structure of the abundant and rare sub-communities, which leads to their distinct biogeographic distributions. However, this is inconsistent with other studies exploring the rare microbial biosphere of the Arctic Ocean (Galand *et al.*, 2009) and subtropical bays (Mo *et al.*, 2018), which showed that rare taxa had similar patterns to those of abundant taxa. It is possible that dispersal is limited by diminished fluidity in terrestrial ecosystems and that selection strength may be higher in

terrestrial ecosystems than in aquatic ecosystems, especially for abundant taxa (Jiao *et al.*, 2017a).

Uncovering which assembly processes in structuring microbial communities are important under different habitats could provide us with an overall better mechanistic understanding of the maintenance and generation of species diversity (Langenheder and Lindstrom, 2019). Bacterial  $\beta$ -diversity showed strong habitat-specific patterns (Wang *et al.*, 2017). In the present study, we suggested that rare taxa in the rice fields, more governed by stochastic assembly processes were less constrained by environment than that in the maize fields. On the other hand, weaker deterministic processes for abundant sub-community were observed in maize fields compared with rice soils. A potential explanation is that paddy soils could be unique habitats due to constant flooding management activities and long-term regular dry-wet cycles (Hu *et al.*, 2013), resulting in less environment filtering. This effort might be stronger for rare taxa, because they are typically restricted by habitat specificity (Barberan *et al.*, 2014; Jousset *et al.*, 2017). When facing less environment filtering, rare taxa might exhibit stronger stochastic dispersal than abundant ones, which could utilize an array of resources and effectively adapt to the environment (Barberan *et al.*, 2014; Jousset *et al.*, 2017). Stochastic processes may overwhelm deterministic processes under less environmental variation (Wang *et al.*, 2013). Interestingly, the abundant sub-community in maize fields and rare sub-community in rice fields were more governed by stochastic assembly processes, corresponding to their higher OTU richness. This could be supported by a recent study, demonstrating that stochastic assembly processes were dominant in high  $\alpha$ -diversity communities (Xun *et al.*, 2019).

Environmental filtering is a key determinant of community assembly (Kraft *et al.*, 2015) and has been shown to cause phylogenetic clustering in bacterial communities (Horner-Devine and Bohannan, 2006). Many studies have shown that soil pH, OM and salinity can influence bacterial community assembly processes in various environments (Dini-Andreote *et al.*, 2015; Feng *et al.*, 2018; Tripathi *et al.*, 2018; Zhang *et al.*, 2019). Here, we showed that soil pH and MAT mediated the balance between stochastic and deterministic assembly for abundant and rare sub-communities, respectively, in agro-ecosystems. It is not surprising that soil pH influences the assembly of abundant sub-communities because pH is known to affect bacterial community composition and diversity at multiple scales (Fierer and Jackson, 2006; Rousk *et al.*, 2010; Jiao *et al.*, 2019). In the present study, the increased role of stochasticity in acidic environments may mean that adapted lineages of abundant taxa accumulate in acidic soils. However, another study showed that stochasticity played a larger role in neutral pH soils (Tripathi *et al.*, 2018). The differences between studies might be due to the divergences between human-

managed agricultural fields and natural ecosystems. MAT drove the assembly of the rare sub-community, which was dominated by homogeneous selection in high-MAT regions and stochasticity in low MAT regions. Temperature is known to influence soil microbial community structure (Davidson and Janssens, 2006; Leininger *et al.*, 2006; Jurburg *et al.*, 2017). Climate warming accelerates temporal turnover and divergent succession of soil microbial communities (Guo *et al.*, 2018, 2019). Species diversity, metabolic activity and population growth rates increase exponentially with temperature (Zhou *et al.*, 2016). Because rare taxa are important contributors to microbial diversity (Lynch and Neufeld, 2015), changing environmental temperatures might have a larger effect on rare taxa. The dominance of homogeneous selection in high MAT regions suggests that rare taxa are more sensitive to hot environments; by contrast, the dominance of stochastic assembly in low MAT regions implies that weaker niche-based exclusion and increased arrival of rare taxa lineages may dominate in colder regions.

## Conclusions

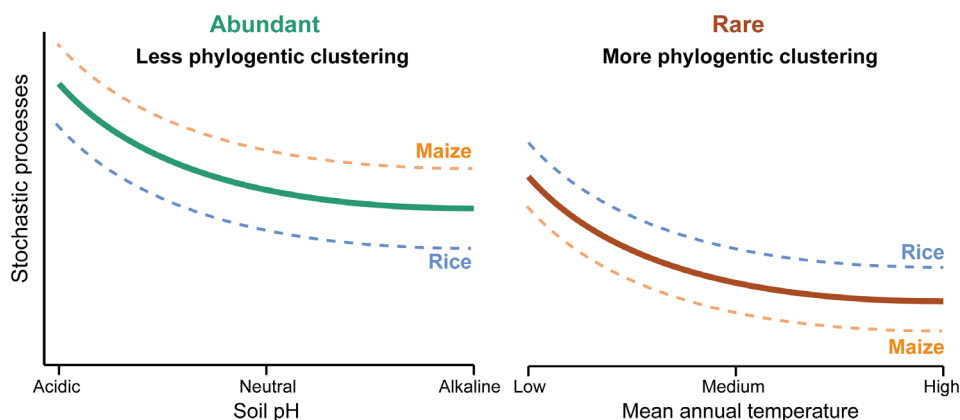
Together, we propose a conceptual paradigm describing the assembly of abundant and rare bacterial sub-communities based on environmental factors in agro-ecosystems (Fig. 6). Phylogenetic clustering of bacterial communities is different in rare and abundant sub-communities. The rare sub-community is more phylogenetically closely clustered than the abundant sub-community, which were less governed by stochastic assembly. With regard to different habitats, stochastic processes contributed more to the abundant sub-community assembly in maize fields than rice fields, while the rare sub-community exhibited the opposite trend. Soil pH and MAT mediate the assembly of the abundant and rare sub-communities respectively. In the abundant sub-community, higher pH leads to an increase in deterministic assembly, whereas in the rare sub-community, increased MAT leads to a decrease in stochastic assembly. This model helps to explain the forces responsible for the generation and maintenance of bacterial diversity, which is essential for ecosystem function (Jing *et al.*, 2015; Delgado-Baquerizo *et al.*, 2016). These factors should be considered when determining ecosystem management practices. Our study was limited to two agricultural ecosystems, and a broader examination of various ecosystems is warranted.

## Methods

### Data sets

We used high-throughput sequencing data from studies of soil bacterial diversity in fields cultivated with maize





**Fig. 6.** Conceptual paradigm showing the stochastic processes for abundant and rare bacterial community assembly, respectively, changing with soil pH and MAT in agricultural ecosystems. [Color figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

and rice (Jiao *et al.*, 2019). The sampling sites extend from 18.30° N to 48.35° N and 87.61° E to 99.91° E across eastern China. Briefly, 251 soils samples were collected from 133 agricultural fields under long-term cultivation with maize and rice; there were 118 paired sites (i.e. < 5 km apart), 8 maize-only sites and 7 rice-only sites. Edaphic variables, including pH, cation exchange capacity (CEC), OM, dissolved organic carbon (DOC), total nitrogen (TN), available nitrogen (AN), nitrate–nitrogen (NO<sub>3</sub>), ammonium–nitrogen (NH<sub>4</sub>), total phosphorus (TP), available phosphorus (AP), total potassium (TK), available potassium (AK), microbial biomass carbon (MBC), microbial biomass nitrogen (MBN), total iron (TFe), available iron (AF<sub>e</sub>), total sulphur (TS), available sulphur (AS), total manganese (TMn), available manganese (AMn), total copper (TCu), available copper (ACu), total zinc (TZn), available zinc (AZn) and soil particle size distribution (clay, silt and sand), were measured using standard soil testing procedures (Ma *et al.*, 2016; Shi *et al.*, 2018). MATs and mean annual precipitation values were obtained from the WorldClim database ([www.worldclim.org](http://www.worldclim.org)). Microbial diversity was analysed using high-throughput sequencing of the V4–V5 regions of the 16S *rRNA* gene on a HiSeq2500 platform (Illumina Inc., San Diego, CA). Sequences were filtered for quality control, split into OTUs using a 97% similarity cut-off and assigned to taxonomic groups (Jiao *et al.*, 2019).

#### Data analyses

All statistical analyses were performed in R (v3.5.1; <http://www.r-project.org/>). OTUs that contained less than 20 reads were discarded. Sequences were rarefied at 26,547 sequences from each sample. Give the rarefaction and locations proximity, 242 samples (120 maize and 122 rice soils) were further analysed. After quality filtering and the removal of chimeric sequences, 6 588 444 high-quality sequences were clustered into 13,149 OTUs. We defined ‘abundant’ OTUs as those having relative

abundances above 0.1% of total sequences, ‘rare’ OTUs as those having relative abundances below 0.01% and ‘intermediate’ OTUs as those having relative abundances between 0.01% and 0.1% (Campbell *et al.*, 2011; Alonso-Saez *et al.*, 2015; Jiao *et al.*, 2017b). Given the aims of this study, we mainly focused on the abundant and rare taxa, rather than intermediate taxa. Recently, a new method has been proposed for the definition of abundant and rare taxa (Mo *et al.*, 2018; Liu *et al.*, 2019a). While, our definition might be more suitable for the present work, which concentrate more on the local level (i.e. in one sample). A kriging interpolation method in the ‘automap’ package of R was used to model spatial distributions of OTU richness (Hiemstra *et al.*, 2009). Maps were cross-validated using ‘autoKrigcv’. Predicted richness values were extracted for selected soil samples and correlated with the observed values from corresponding sites using a Pearson correlation analysis.

Phylogenetic analyses were performed in the ‘picante’ package in R (Kembel *et al.*, 2010). To evaluate the phylogenetic community assembly, we calculated the SES. MNTD using the null model ‘taxa.labels’ (999 randomization) in the ‘ses.mntd’ function. Pairwise phylogenetic turnover between communities was calculated as the MNTD metric ( $\beta$ MNTD) using the ‘comdistnt’ function (abundance.weighted = TRUE). The phylogenetic DDR was calculated as the slope of an ordinary least-squares regression between geographic distance and phylogenetic community similarity ( $1 - \beta$ MNTD metric). The impact of environmental variables on  $\beta$ -diversity was assessed using a distance-based linear model and forward selection procedure based on the  $\beta$ MNTD distance matrix by estimating the proportion of variance explained ( $R^2$ ). These results were displayed by CAP.

We estimated the phylogenetic depths of environmental responses for the abundant and rare bacterial sub-communities and their major phyla using Blomberg’s *K*, calculated with the ‘multiPhylosignal’ function. These *K* values indicate how well the species is correlated with

phylogeny as expected by a Brownian motion-based metric of the strength of the phylogenetic signal (Blomberg *et al.*, 2003); higher  $K$  values indicate better correlation. To test for significant phylogenetic signal,  $p$  values for each  $K$  value were calculated based on the variance of the phylogenetically independent comparisons relative to a null model that randomly reshuffles trait values (999 simulations).

A null model analysis was used to evaluate assembly processes of abundant and rare taxa (Stegen *et al.*, 2013). The variations in phylogenetic and taxonomic diversity were measured using null model-based phylogenetic and taxonomic  $\beta$ -diversity metrics ( $\beta$ NTI and Bray–Curtis-based Raup–Crick,  $RC_{Bray}$ ). The  $\beta$ NTI quantifies the magnitude and direction of deviation between an observed  $\beta$ MNTD value and the null  $\beta$ MNTD distribution. It also estimates the phylogenetic turnover given stochastic and deterministic ecological processes (Stegen *et al.*, 2012; Wang *et al.*, 2013; Dini-Andreote *et al.*, 2015). A significant deviation (i.e.  $|\beta$ NTI  $> 2$ ) indicates the dominance of deterministic processes.  $\beta$ NTI values  $< -2$  indicate significantly less phylogenetic turnover than expected (i.e. homogeneous selection); meanwhile,  $\beta$ NTI values  $> +2$  indicate significantly more phylogenetic turnover than expected (i.e. variable selection). The taxonomic  $\beta$ -diversity metric  $RC_{Bray}$  was used to analyse pairwise comparisons that were not assigned to deterministic processes (i.e.  $|\beta$ NTI  $< 2$ ). The relative influence of homogeneous dispersal was quantified as the fraction of pairwise comparisons with  $|\beta$ NTI  $< 2$  and  $RC_{Bray} < -0.95$ . The relative influence of dispersal limitations was quantified as the fraction of pairwise comparisons with  $|\beta$ NTI  $< 2$  and  $RC_{Bray} > 0.95$ . The fraction of all pairwise comparisons with  $|\beta$ NTI  $< 2$  and  $|RC_{Bray}| < 0.95$  was used to estimate the influence of the ‘non-dominant’ fraction (Zhou and Ning, 2017; Tripathi *et al.*, 2018).

To assess the relative influence of assembly processes, we compared all possible pairwise comparisons of  $\beta$ NTI values with each major environmental variable. The variation in community assembly processes along the gradients of the derived environmental variables (soil pH for the abundant sub-community and MAT for the rare sub-community) was assessed using regression analysis comparing  $\beta$ NTI values with the Euclidean distance matrices of these variables. The statistical significance of these comparisons was determined by Mantel tests with 999 permutations. The same method was used to assess the relationship between phylogenetic turnover and the derived environmental variables after controlling for spatial or other environmental variables distance. These analyses were performed using the ‘mantel’ function of the ‘ecodist’ package in R (Goslee and Urban, 2007). A neutral community model was used to determine contribution of neutral processes to bacterial community assembly by

predicting the relationship between the frequency of taxonomic occurrence in a set of local communities and the abundance of those taxa across the wider metacommunity (Sloan *et al.*, 2006).

## Acknowledgements

This work was supported by the National Science Foundation of China (41807030), the National Key Research and Development Program of China (2016YFD0200306), the National Postdoctoral Program for Innovative Talents (BX201700005), and the China Postdoctoral Science Foundation (2018M630041).

## Availability of data and materials

The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive (Genomics, Proteomics & Bioinformatics 2017) in Beijing Institute of Genomics (BIG) Data Center (Nucleic Acids Research 2018), BIG, Chinese Academy of Sciences, under accession numbers PRJCA001121 that are publicly accessible at <http://bigd.big.ac.cn/gsa>

## Author contributions

S.J. conceived and designed the experiments with the help of Y.L.; S.J. performed the experiments and analysed the data. S.J. and Y.L. wrote the manuscript.

## References

- Alonso-Saez, L., Diaz-Perez, L., and Moran, X.A.G. (2015) The hidden seasonality of the rare biosphere in coastal marine bacterioplankton. *Environ Microbiol* **17**: 3766–3780.
- Astudillo-Garcia, C., Hermans, S.M., Stevenson, B., Buckley, H.L., and Lear, G. (2019) Microbial assemblages and bioindicators as proxies for ecosystem health status: potential and limitations. *Appl Microbiol Biot* **103**: 6407–6421.
- Barberan, A., Ramirez, K.S., Leff, J.W., Bradford, M.A., Wall, D.H., and Fierer, N. (2014) Why are some microbes more ubiquitous than others? Predicting the habitat breadth of soil bacteria. *Ecol Lett* **17**: 794–802.
- Blomberg, S.P., Garland, T., and Ives, A.R. (2003) Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* **57**: 717–745.
- Brown, M.V., Ostrowski, M., Grzymalski, J.J., and Lauro, F.M. (2014) A trait based perspective on the biogeography of common and abundant marine bacterioplankton clades. *Marine genomics* **15**: 17–28.
- Campbell, B.J., Yu, L., Heidelberg, J.F., and Kirchman, D.L. (2011) Activity of abundant and rare bacteria in a coastal ocean. *Proc Natl Acad Sci USA* **108**: 12776–12781.
- Cavender-Bares, J., Kozak, K.H., Fine, P.V.A., and Kembel, S.W. (2009) The merging of community ecology and phylogenetic biology. *Ecol Lett* **12**: 693–715.

- Davidson, E.A., and Janssens, I.A. (2006) Temperature sensitivity of soil carbon decomposition and feedbacks to climate change. *Nature* **440**: 165–173.
- Delgado-Baquerizo, M., Maestre, F.T., Reich, P.B., Jeffries, T.C., Gaitan, J.J., Encinar, D., *et al.* (2016) Microbial diversity drives multifunctionality in terrestrial ecosystems. *Nat Commun* **7**: 10541.
- Dini-Andreote, F., Stegen, J.C., van Elsas, J.D., and Salles, J.F. (2015) Disentangling mechanisms that mediate the balance between stochastic and deterministic processes in microbial succession. *Proc Natl Acad Sci USA* **112**: E1326–E1332.
- Dumbrell, A.J., Nelson, M., Helgason, T., Dytham, C., and Fitter, A.H. (2010) Relative roles of niche and neutral processes in structuring a soil microbial community. *ISME J* **4**: 337–345.
- Elshahed, M.S., Youssef, N.H., Spain, A.M., Sheik, C., Najjar, F.Z., Sukharnikov, L.O., *et al.* (2008) Novelty and uniqueness patterns of rare members of the soil biosphere. *Appl Environ Microb* **74**: 5422–5428.
- Falkowski, P.G., Fenchel, T., and Delong, E.F. (2008) The microbial engines that drive Earth's biogeochemical cycles. *Science* **320**: 1034–1039.
- Fan, K.K., Cardona, C., Li, Y.T., Shi, Y., Xiang, X.J., Shen, C.C., *et al.* (2017) Rhizosphere-associated bacterial network structure and spatial distribution differ significantly from bulk soil in wheat crop fields. *Soil Biol Biochem* **113**: 275–284.
- Feng, Y., Chen, R., Stegen, J.C., Guo, Z., Zhang, J., Li, Z., and Lin, X. (2018) Two key features influencing community assembly processes at regional scale: initial state and degree of change in environmental conditions. *Mol Ecol* **27**: 5238–5251.
- Fierer, N., and Jackson, R.B. (2006) The diversity and biogeography of soil bacterial communities. *Proc Natl Acad Sci USA* **103**: 626–631.
- Galand, P.E., Casamayor, E.O., Kirchman, D.L., and Lovejoy, C. (2009) Ecology of the rare microbial biosphere of the Arctic Ocean. *Proc Natl Acad Sci USA* **106**: 22427–22432.
- Goslee, S.C., and Urban, D.L. (2007) The ecodist package for dissimilarity-based analysis of ecological data. *J Stat Softw* **22**: 1–19.
- Guo, X., Feng, J.J., Shi, Z., Zhou, X.S., Yuan, M.T., Tao, X. Y., *et al.* (2018) Climate warming leads to divergent succession of grassland microbial communities. *Nat Clim Change* **8**: 813–818.
- Guo, X., Zhou, X., Hale, L., Yuan, M., Ning, D., Feng, J., *et al.* (2019) Climate warming accelerates temporal scaling of grassland soil microbial biodiversity. *Nat Ecol Evol* **3**: 612–619.
- Hanson, C.A., Fuhrman, J.A., Horner-Devine, M.C., and Martiny, J.B. (2012) Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat Rev Microbiol* **10**: 497–506.
- Hiemstra, P.H., Pebesma, E.J., Twenhöfel, C.J.W., and Heuvelink, G.B.M. (2009) Real-time automatic interpolation of ambient gamma dose rates from the Dutch radioactivity monitoring network. *Comput Geosci UK* **35**: 1711–1721.
- Horner-Devine, M.C., and Bohannan, B.J.M. (2006) Phylogenetic clustering and overdispersion in bacterial communities. *Ecology* **87**: S100–S108.
- Hu, H.W., Zhang, L.M., Yuan, C.L., and He, J.Z. (2013) Contrasting Euryarchaeota communities between upland and paddy soils exhibited similar pH-impacted biogeographic patterns. *Soil Biol Biochem* **64**: 18–27.
- Hubbell, S.P. (2001) *The Unified Neutral Theory of Biodiversity and Biogeography (MPB-32)*: New Jersey, USA: Princeton University Press.
- Jia, X., Dini-Andreote, F., and Salles, J.F. (2018) Community assembly processes of the microbial rare biosphere. *Trends Microbiol* **26**: 738–747.
- Jiao, S., Chen, W.M., and Wei, G.H. (2017a) Biogeography and ecological diversity patterns of rare and abundant bacteria in oil-contaminated soils. *Mol Ecol* **26**: 5305–5317.
- Jiao, S., Luo, Y.T., Lu, M.M., Xiao, X., Lin, Y.B., Chen, W.M., and Wei, G.H. (2017b) Distinct succession patterns of abundant and rare bacteria in temporal microcosms with pollutants. *Environ Pollut* **225**: 497–505.
- Jiao, S., Xu, Y., Zhang, J., Hao, X., and Lu, Y. (2019) Core microbiota in agricultural soils and their potential associations with nutrient cycling. *mSystems* **4**: e00313–e00318.
- Jing, X., Sanders, N.J., Shi, Y., Chu, H.Y., Classen, A.T., Zhao, K., *et al.* (2015) The links between ecosystem multifunctionality and above- and belowground biodiversity are mediated by climate. *Nat Commun* **6**: 8159.
- Jousset, A., Bienhold, C., Chatzinotas, A., Gallien, L., Gobet, A., Kurm, V., *et al.* (2017) Where less may be more: how the rare biosphere pulls ecosystems strings. *ISME J* **11**: 853–862.
- Jurburg, S.D., Nunes, I., Brejnrod, A., Jacquiod, S., Prieme, A., Sorensen, S.J., *et al.* (2017) Legacy effects on the recovery of soil bacterial communities from extreme temperature perturbation. *Front Microbiol* **8**: 1832.
- Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D., *et al.* (2010) Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* **26**: 1463–1464.
- Kraft, N.J.B., Adler, P.B., Godoy, O., James, E.C., Fuller, S., and Levine, J.M. (2015) Community assembly, coexistence and the environmental filtering metaphor. *Funct Ecol* **29**: 592–599.
- Langenheder, S., and Lindstrom, E.S. (2019) Factors influencing aquatic and terrestrial bacterial community assembly. *Environ Microbiol Rep* **11**: 306–315.
- Langenheder, S., and Szekely, A.J. (2011) Species sorting and neutral processes are both important during the initial assembly of bacterial communities. *ISME J* **5**: 1086–1094.
- Leininger, S., Urich, T., Schlöter, M., Schwark, L., Qi, J., Nicol, G.W., *et al.* (2006) Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature* **442**: 806–809.
- Lennon, J.T., and Jones, S.E. (2011) Microbial seed banks: the ecological and evolutionary implications of dormancy. *Nat Rev Microbiol* **9**: 119–130.
- Liu, L., Yang, J., Yu, Z., and Wilkinson, D.M. (2015) The biogeography of abundant and rare bacterioplankton in the lakes and reservoirs of China. *ISME J* **9**: 2068–2077.

- Liu, M., Xue, Y., and Yang, J. (2019a) Rare plankton sub-communities are far more affected by DNA extraction kits than abundant plankton. *Front Microbiol* **10**: 454.
- Liu, Z., Cichocki, N., Hübschmann, T., Süring, C., Ofitt, I.D., Sloan, W.T., *et al.* (2019b) Neutral mechanisms and niche differentiation in steady-state insular microbial communities revealed by single cell analysis. *Environ Microbiol* **21**: 164–181.
- Logares, R., Audic, S., Bass, D., Bittner, L., Boute, C., Christen, R., *et al.* (2014) Patterns of rare and abundant marine microbial eukaryotes. *Curr Biol* **24**: 813–821.
- Logares, R., Haverkamp, T.H., Kumar, S., Lanzén, A., Nederbragt, A.J., Quince, C., and Kausrud, H. (2012) Environmental microbiology through the lens of high-throughput DNA sequencing: synopsis of current platforms and bioinformatics approaches. *J Microbiol Method* **91**: 106–113.
- Lynch, M.D., and Neufeld, J.D. (2015) Ecology and exploration of the rare biosphere. *Nat Rev Microbiol* **13**: 217–229.
- Ma, B., Wang, H., Dsouza, M., Lou, J., He, Y., Dai, Z., *et al.* (2016) Geographic patterns of co-occurrence network topological features for soil microbiota at continental scale in eastern China. *ISME J* **10**: 1891–1901.
- Malard, L.A., and Pearce, D.A. (2018) Microbial diversity and biogeography in Arctic soils. *Environ Microbiol Rep* **10**: 611–625.
- Meyer, K.M., Memiaghe, H., Korte, L., Kenfack, D., Alonso, A., and Bohannan, B.J.M. (2018) Why do microbes exhibit weak biogeographic patterns? *ISME J* **12**: 1404–1413.
- Mo, Y., Zhang, W., Yang, J., Lin, Y., Yu, Z., and Lin, S. (2018) Biogeographic patterns of abundant and rare bacterioplankton in three subtropical bays resulting from selective and neutral processes. *ISME J* **12**: 2198–2210.
- Mondiale, B. (2008) *World development report: agriculture for development*. Washington, DC: The World Bank.
- Nemergut, D.R., Schmidt, S.K., Fukami, T., O'Neill, S.P., Bilinski, T.M., Stanish, L.F., *et al.* (2013) Patterns and processes of microbial community assembly. *Microbiol Mol Biol R* **77**: 342–356.
- Ogonowski, M., Motiei, A., Ininbergs, K., Hell, E., Gerdes, Z., Udekwu, K.I., *et al.* (2018) Evidence for selective bacterial community structuring on microplastics. *Environ Microbiol* **20**: 2796–2808.
- Ovaskainen, O., Tikhonov, G., Norberg, A., Blanchet, F.G., Duan, L., Dunson, D., *et al.* (2017) How to make more out of community data? A conceptual framework and its implementation as models and software. *Ecol Lett* **20**: 561–576.
- Pagaling, E., Strathdee, F., Spears, B.M., Cates, M.E., Allen, R.J., and Free, A. (2014) Community history affects the predictability of microbial ecosystem development. *ISME J* **8**: 19–30.
- Pedros-Álió, C. (2012) The rare bacterial biosphere. *Ann Rev Mar Sci* **4**: 449–466.
- Pester, M., Bittner, N., Deevong, P., Wagner, M., and Loy, A. (2010) A 'rare biosphere' microorganism contributes to sulfate reduction in a peatland. *ISME J* **4**: 1591–1602.
- Prosser, J.I., Bohannan, B.J., Curtis, T.P., Ellis, R.J., Firestone, M.K., Freckleton, R.P., *et al.* (2007) The role of ecological theory in microbial ecology. *Nat Rev Microbiol* **5**: 384–392.
- Reid, A., and Buckley, M. (2011) *The Rare Biosphere: A Report from the American Academy of Microbiology*. Washington, DC: American Academy of Microbiology, p. 436.
- Rousk, J., Baath, E., Brookes, P.C., Lauber, C.L., Lozupone, C., Caporaso, J.G., *et al.* (2010) Soil bacterial and fungal communities across a pH gradient in an arable soil. *ISME J* **4**: 1340–1351.
- Sauret, C., Séverin, T., Vétion, G., Guigue, C., Goutx, M., Pujo-Pay, M., *et al.* (2014) 'Rare biosphere' bacteria as key phenanthrene degraders in coastal seawaters. *Environ Pollut* **194**: 246–253.
- Schloss, P.D., and Handelsman, J. (2007) The last word: books as a statistical metaphor for microbial communities. *Annu Rev Microbiol* **61**: 23–34.
- Shi, Y., Li, Y., Xiang, X., Sun, R., Yang, T., He, D., *et al.* (2018) Spatial scale affects the relative role of stochasticity versus determinism in soil bacterial communities in wheat fields across the North China plain. *Microbiome* **6**: 27.
- Sloan, W.T., Lunn, M., Woodcock, S., Head, I.M., Nee, S., and Curtis, T.P. (2006) Quantifying the roles of immigration and chance in shaping prokaryote community structure. *Environ Microbiol* **8**: 732–740.
- Sogin, M.L., Morrison, H.G., Huber, J.A., Welch, D.M., Huse, S.M., Neal, P.R., *et al.* (2006) Microbial diversity in the deep sea and the underexplored 'rare biosphere'. *Proc Natl Acad Sci USA* **103**: 12115–12120.
- Stegen, J.C., Lin, X.J., Fredrickson, J.K., Chen, X.Y., Kennedy, D.W., Murray, C.J., *et al.* (2013) Quantifying community assembly processes and identifying features that impose them. *ISME J* **7**: 2069–2079.
- Stegen, J.C., Lin, X.J., Konopka, A.E., and Fredrickson, J.K. (2012) Stochastic and deterministic assembly processes in subsurface microbial communities. *ISME J* **6**: 1653–1664.
- Tripathi, B.M., Stegen, J.C., Kim, M., Dong, K., Adams, J.M., and Lee, Y.K. (2018) Soil pH mediates the balance between stochastic and deterministic assembly of bacteria. *ISME J* **12**: 1072–1083.
- Wang, J., Shen, J., Wu, Y., Tu, C., Soininen, J., Stegen, J.C., *et al.* (2013) Phylogenetic beta diversity in bacterial assemblages across ecosystems: deterministic versus stochastic processes. *ISME J* **7**: 1310–1321.
- Wang, K., Ye, X., Chen, H., Zhao, Q., Hu, C., He, J., *et al.* (2015) Bacterial biogeography in the coastal waters of northern Zhejiang, East China Sea is highly controlled by spatially structured environmental gradients. *Environ Microbiol* **17**: 3898–3913.
- Wang, X.B., Lu, X.T., Yao, J., Wang, Z.W., Deng, Y., Cheng, W.X., *et al.* (2017) Habitat-specific patterns and drivers of bacterial beta-diversity in China's drylands. *ISME J* **11**: 1345–1358.
- Whitman, T., Neurath, R., Perera, A., Chu-Jacoby, I., Ning, D., Zhou, J., *et al.* (2018) Microbial community assembly differs across minerals in a rhizosphere microcosm. *Environ Microbiol* **20**: 4444–4460.

- Wu, W.X., Logares, R., Huang, B.Q., and Hsieh, C.H. (2017) Abundant and rare picoeukaryotic sub-communities present contrasting patterns in the epipelagic waters of marginal seas in the northwestern Pacific Ocean. *Environ Microbiol* **19**: 287–300.
- Xun, W., Li, W., Xiong, W., Ren, Y., Liu, Y., Miao, Y., et al. (2019) Diversity-triggered deterministic bacterial assembly constrains community functions. *Nat Commun* **10**: 3833.
- Yachi, S., and Loreau, M. (1999) Biodiversity and ecosystem productivity in a fluctuating environment: the insurance hypothesis. *Proc Natl Acad Sci USA* **96**: 1463–1468.
- Zhang, B.G., Zhang, J., Liu, Y., Guo, Y.Q., Shi, P., and Wei, G.H. (2018) Biogeography and ecological processes affecting root-associated bacterial communities in soybean fields across China. *Sci Total Environ* **627**: 20–27.
- Zhang, K.P., Shi, Y., Cui, X.Q., Yue, P., Li, K.H., Liu, X.J., et al. (2019) Salinity is a key determinant for soil microbial communities in a desert ecosystem. *mSystems* **4**: e00225–e00218.
- Zhou, J., and Ning, D. (2017) Stochastic community assembly: does it matter in microbial ecology? *Microbiol Mol Biol R* **81**: e00002–e00017.
- Zhou, J.Z., Deng, Y., Shen, L.N., Wen, C.Q., Yan, Q.Y., Ning, D.L., et al. (2016) Temperature mediates continental-scale diversity of microbes in forest soils. *Nat Commun* **7**: 12083.

## Supporting Information

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

**Table S1** ANOVA of environmental factors correlated with soil bacterial  $\beta$ -diversity of abundant sub-communities in agricultural fields of eastern China. The abbreviations of environmental factors accorded to the *Methods*.

**Table S2** ANOVA of environmental factors correlated with soil bacterial  $\beta$ -diversity of rare sub-communities in agricultural fields of eastern China. The abbreviations of environmental factors accorded to the *Methods*.

**Table S3** ANOVA of environmental factors correlated with soil bacterial  $\beta$ -diversity of abundant sub-communities in maize fields of eastern China. The abbreviations of environmental factors accorded to the *Methods*.

**Table S4** ANOVA of environmental factors correlated with soil bacterial  $\beta$ -diversity of rare sub-communities in maize fields of eastern China. The abbreviations of environmental factors accorded to the *Methods*.

**Table S5** ANOVA of environmental factors correlated with soil bacterial  $\beta$ -diversity of abundant sub-communities in rice fields of eastern China. The abbreviations of environmental factors accorded to the *Methods*.

**Table S6** ANOVA of environmental factors correlated with soil bacterial  $\beta$ -diversity of rare sub-communities in rice fields of eastern China. The abbreviations of environmental factors accorded to the *Methods*.

**Fig. S1** Distributions of rare and abundant bacteria in agricultural soils. The boxplots display the percent changes for abundant or rare taxa in one sample to abundant, intermediate, rare, or not detected in other samples. For example, the

percentages of rare taxa in Sample 1 that become abundant, intermediate, remained rare, or were not detected in Samples 2, through all samples are displayed.

**Fig. S2** General distribution patterns of abundant and rare bacteria in agricultural soils. **(A)** The proportion of the OTUs richness ('Richness') and relative abundance of abundant and rare taxa, compared to the whole bacterial community in each sample. **(B)** Differences in OTUs richness and relative abundance of abundant and rare taxa between maize and rice soils (\*\*\*,  $P < 0.001$ ; Wilcoxon rank-sum test). Spatial distributions of OTU richness of abundant **(C)** and rare **(D)** taxa were mapped using a kriging interpolation method. Cross-validation ('CV') of the maps was based on Pearson correlation between the predicted and observed values in each sampling site. The Pearson correlation coefficient and  $P$ -value were shown in the maps.

**Fig. S3** Boxplot showing difference in SES.MNTD values between abundant and rare sub-communities (\*\*\*,  $P < 0.001$ ; Wilcoxon rank-sum test).

**Fig. S4** General beta-diversity patterns of abundant and rare bacteria in agricultural soils. **(A)** Phylogenetic distance-decay curves showing community similarity ( $1 - \beta$ MNTD metric) against geographic distances between sampling sites in all ('Whole'), maize ('Maize') and rice ('Rice') soil samples. Solid lines denote the ordinary least squares linear regression. Asterisks represent significance of correlation (\*,  $P < 0.05$ ; \*\*\*,  $P < 0.001$ ). **(B)** Boxplot showing difference in phylogenetic  $\beta$ -diversity between abundant and rare sub-communities (\*\*\*,  $P < 0.001$ ; Wilcoxon rank-sum test).

**Fig. S5** Major environmental variables in shaping the abundant and rare phylogenetic sub-communities based on mean nearest taxon distance metric ( $\beta$ MNTD) in maize and rice soils, identified by constrained analysis of principal coordinates (CAP). The shade of the points represented the values of soil pH or mean annual temperature (MAT). The colour of the points represented low, medium ('Med') and high categories of soil pH or MAT. The abbreviations of environmental factors accorded to the *Methods*.

**Fig. S6** Taxonomic distribution of 'alkaline-responsive', 'acid-responsive', 'warm-responsive' and 'cold-responsive' bio-indicator taxa in abundant and rare sub-communities at phylum level shown as circos plot. The thickness of each ribbon represents the number of each categories assigned to different phyla.

**Fig. S7** Relative influence of deterministic and stochastic assembly processes in maize soils. **(A)** The relationships between  $\beta$ NTI and differences in soil pH for abundant bacterial sub-community and in mean annual temperature (MAT) for rare sub-community. Linear regressions models (shown as red lines) and associated correlation coefficients are provided on each panel. Horizontal dashed lines indicate the  $\beta$ NTI significance thresholds of +2 and - 2. **(B)** Patterns of  $\beta$ NTI across different categories in soil pH for abundant sub-community and MAT for rare sub-community. Boxplots showed the variation in  $\beta$ NTI across low, medium ('Med') and high categories of soil pH and MAT. Boxplots that do not share a letter are significantly different ( $P < 0.05$ ; multiple comparisons with Kruskal-Wallis).

**Fig. S8** Relative influence of deterministic and stochastic assembly processes in rice soils. **(A)** The relationships



between  $\beta$ NTI and differences in soil pH for abundant bacterial sub-community and in mean annual temperature (MAT) for rare sub-community. Linear regressions models (shown as red lines) and associated correlation coefficients are provided on each panel. Horizontal dashed lines indicate the  $\beta$ NTI significance thresholds of +2 and - 2. **(B)** Patterns of

$\beta$ NTI across different categories in soil pH for abundant sub-community and MAT for rare sub-community. Boxplots showed the variation in  $\beta$ NTI across low, medium ('Med') and high categories of soil pH and MAT. Boxplots that do not share a letter are significantly different ( $P < 0.05$ ; multiple comparisons with Kruskal-Wallis).