

## RESEARCH ARTICLE

# Biogeographical patterns of abundant and rare bacterial biospheres in paddy soils across East Asia

Lu Luan<sup>1</sup> | Guangping Shi<sup>1</sup> | Guofan Zhu<sup>1</sup> | Jie Zheng<sup>1</sup> | Jianbo Fan<sup>1</sup> | Francisco Dini-Andreote<sup>2</sup> | Bo Sun<sup>1</sup> | Yuji Jiang<sup>1</sup>

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

<sup>2</sup>Department of Plant Science & Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, Pennsylvania, USA

## Correspondence

Francisco Dini-Andreote, Department of Plant Science & Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, PA, USA.  
Email: [andreote@psu.edu](mailto:andreote@psu.edu)

Yuji Jiang, State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, China.  
Email: [yjjiang@issas.ac.cn](mailto:yjjiang@issas.ac.cn)

## Funding information

National Natural Science Foundation for Excellent Young Scholars of China, Grant/Award Number: 41922048; National Natural Science Foundation of China, Grant/Award Numbers: 42177298, 42107336; Youth Innovation Promotion Association of CAS, Grant/Award Number: Y2021084; Science and Technology Project of Jiangsu Province, Grant/Award Number: BE2022393; China Postdoctoral Science Foundation, Grant/Award Number: 2021M690155

## Abstract

Soil bacterial communities play fundamental roles in ecosystem functioning and often display a skewed distribution of abundant and rare taxa. So far, relatively little is known about the biogeographical patterns and mechanisms structuring the assembly of abundant and rare biospheres of soil bacterial communities. Here, we studied the geographical distribution of different bacterial sub-communities by examining the relative influence of environmental selection and dispersal limitation on taxa distributions in paddy soils across East Asia. Our results indicated that the geographical patterns of four different bacterial sub-communities consistently displayed significant distance–decay relationships (DDRs). In addition, we found niche breadth and dispersal rates to significantly explain differences in community assembly of abundant and rare taxa, directly affecting the strength of DDRs. While conditionally rare and abundant taxa displayed the strongest DDR due to higher environmental filtering and dispersal limitation, moderate taxa sub-communities had the weakest DDR due to greater environmental tolerance and dispersal rate. Random forest models indicated that soil pH (9.13%–49.78%) and average annual air temperature (16.59%–46.49%) were the most important predictors of the variation in the bacterial community. This study advances our understanding of the intrinsic links between fundamental ecological processes and microbial biogeographical patterns in paddy soils.

## ORIGINALITY—SIGNIFICANCE STATEMENT

Understanding the biogeographical patterns and mechanisms structuring the assembly of abundant and rare bacterial communities is central to predict how soil biological function responds to biotic/abiotic disturbances. Here, we uncovered the distance–decay relationship (DDR) and the corresponding driving mechanism of the rare and abundant soil bacterial communities in paddy soils across East Asia. Our results indicated that four different bacterial sub-communities consistently showed significant DDRs,

with environmental filtering and dispersal processes regulating the strength and direction of DDRs. Taken together, our study indicates an interplay of community assembly mechanisms operating across distinct bacterial sub-communities, directly determining their DDRs and biogeographical distributions.

## INTRODUCTION

The soil microbial diversity underlies major ecosystem processes responsible for ecosystem functioning and maintenance (e.g., carbon, sulfur, and nitrogen cycling)

(Bahram et al., 2018). Advancing our understanding of the processes and mechanisms structuring the distribution and abundance of different microbial taxa (i.e., biogeographical patterns) in soils is essential to better predict their functional outcomes (Bahram et al., 2018; Luan, et al., 2020b; Meyer et al., 2018). The ‘long-tailed’ skewed species abundance distribution of bacterial taxa in soil communities represents a universal pattern. This relates to the existence of a (meta)community containing a large number of low-abundance taxa and a small number of high-abundance taxa (Brown et al., 2014; Hanson et al., 2012; Jousset et al., 2017). Abundant taxa (AT) are generally assumed to be highly active and important for ecosystem functioning (Cottrell & Kirchman, 2003), while rare taxa (RT) are often (albeit not always) assumed to provide complementary functions and/or specific metabolisms (Dimitriu et al., 2010). Recent studies have started to focus on the mechanisms structuring the rare biosphere (Jia et al., 2018), as well as their species diversity and distribution, and co-occurrence patterns (Xue et al., 2018; Zheng et al., 2021).

Distance–decay relationship (DDR) is an approach widely used to study biogeographical patterns of organismal distribution across spatial scales (Hanson et al., 2012; Soininen et al., 2007). This metric quantifies the decrease in compositional similarity ( $\beta$ -diversity) between communities with increasing geographical distance. Thus, DDR is used to demonstrate whether communities in close proximity are more similar to each other than distant ones (Soininen et al., 2007). It can be conceptually devised that ecological communities and DDR are structured by an interplay of four fundamental ecological processes. First, environmental selection filters species out, and it is often spatially distributed in a gradient. Second, ecological drift is associated with randomness in taxa in births and deaths, and it is expected to be more important in RT (Luan, et al., 2020b; Pimm et al., 1988; Villarino et al., 2018). Third, dispersal (that is, the movement of species) tends to counteract compositional differences by homogenizing species distribution in space. And, fourth, diversification affects DDR by increasing local genetic diversity, which lowers the height (not the slope) of the DDR curve and increases its variance (Hedrick, 1999; Slatkin, 1993). Since spatial patterns of community dissimilarity may reflect the outcome and interplay of multiple community assembly processes, it remains a major challenge to elucidate the DDRs among different microbial taxa.

Species abundance and diversity are under the direct influence of distinct community assembly processes, resulting in contrasting patterns of DDRs in abundant and rare taxa (Jia et al., 2018; Jiao & Lu, 2020). For example, AT usually consist of a small number of cosmopolitan species with relatively high

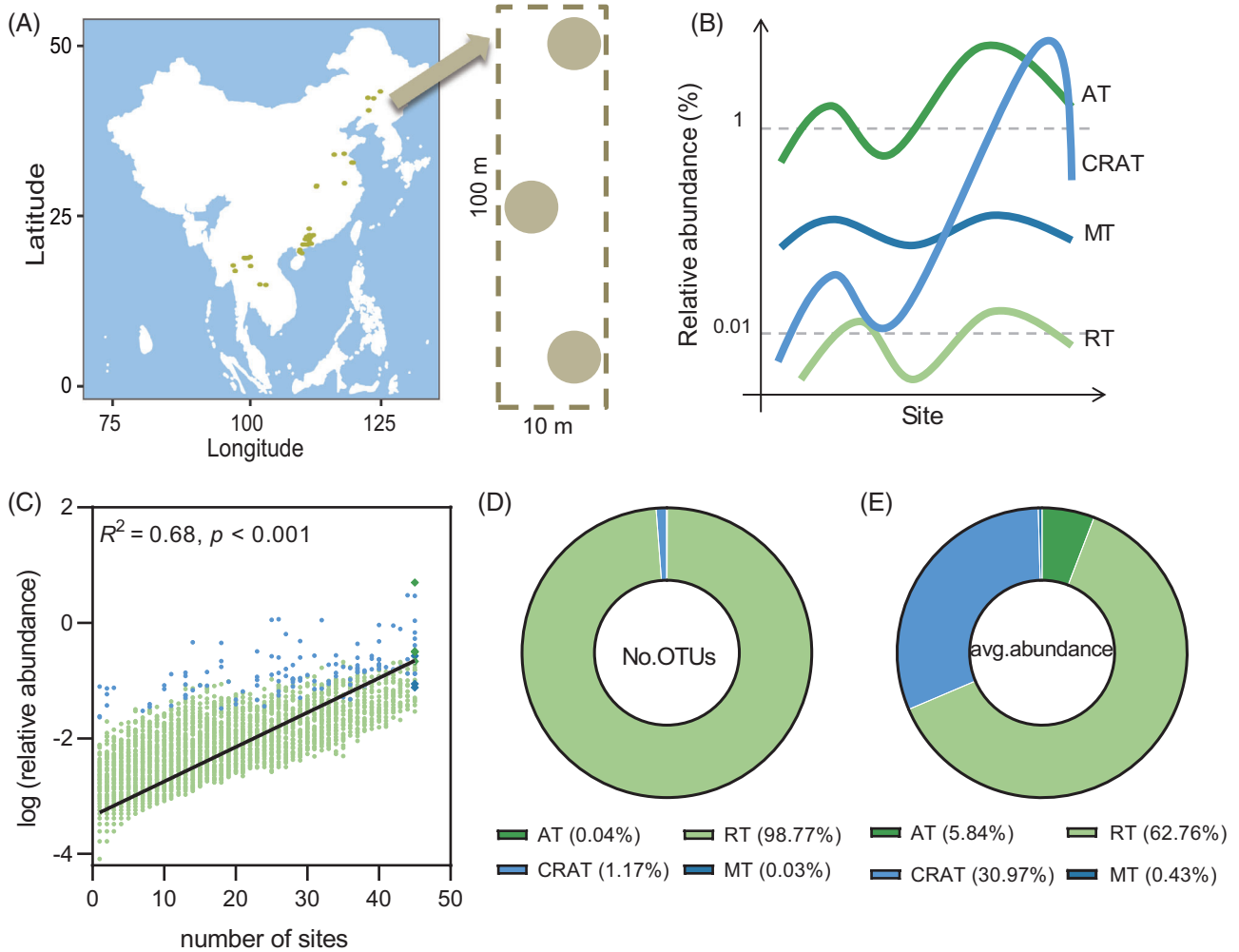
abundance, high dispersal capabilities, and greater environmental tolerance (Jiao & Lu, 2020; Wu et al., 2017). In contrast, RT are oligotrophic and/or endemic species with relatively low abundance, variable dispersal capabilities, and constant environmental filtering. But for conditionally rare and abundant taxa (CRAT), which have a small number of species and relatively low abundance, may face strong dispersal limitations and environmental filtering (Jia et al., 2018; Xue et al., 2018). Given that microbial taxa are potentially shaped by different ecological processes, we hypothesize that the DDRs of microbial taxa vary with the relative influences of distinct ecological processes.

Paddy soils represent the third-largest cropland area and the largest anthropogenic wetland on Earth (Jiang et al., 2016; Jiao et al., 2020). These soils are often under waterlogged paddy rice cultivation, resulting in a relatively higher uniformity and physical similarity. Until now, it has remained unclear whether the uniform management resembles an overall homogenization of bacterial communities, for both the abundant and rare biospheres. This might relate to the balance of community assembly processes, in particular, environmental selection and dispersal rates acting on species distributional patterns. Here, we used a continental-scale field sampling of paddy soils across East Asia to explore patterns of soil bacterial biogeography. Specifically, we addressed the following questions: (1) whether different microbial taxa show distinct biogeographical patterns at both taxonomic and phylogenetic levels; and (2) how the community assembly processes determine the DDRs of different bacterial sub-communities. Our results indicated that the geographical patterns of four different bacterial sub-communities consistently displayed significant DDRs, based on both taxa distributions and phylogenetic distances. In addition, we found niche breadth and dispersal rates to significantly explain differences in the community assembly of rare and AT, directly affecting the strength of DDRs.

## EXPERIMENTAL PROCEDURES

### Soil sampling and data collection

We used a paddy soil dataset previously published elsewhere (Jiang et al., 2016). In brief, a total of 45 soil samples were collected in a paddy ecosystem along the north–south transect across East Asia (Figure 1A). For each sampling point, three sub-plots were randomly established in a 10 × 100 m rectangular plot. Within each sub-plot, 20 soil cores (5 cm diameter) of the upper 15 cm were collected randomly and composited into a single bulk sample. The latitude of the sampling area ranged from 15.90 to 44.31°N, the average annual air temperature (MAT) from 2 to 27.5°C, and the average annual precipitation (MAP)



**FIGURE 1** Sampling sites and a detailed description of different bacterial taxa. Soil samples were collected in a paddy ecosystem along the north-south transects across East Asia, and each site was randomly set to three replicates (A). The division of different groups of bacteria (B). The correlation between average relative abundance and frequency (number of occupied sites) of all OTUs (C). Proportion of OTU numbers (D) and accumulative relative abundances (E) of four bacterial sub-communities. AT, abundant taxa; CRAT, conditionally rare and abundant taxa; MT, moderate taxa; RT, rare taxa.

from 550 to 2345 mm. The global positioning system coordinates recorded at each sampling site were used to calculate the MAT and the MAP using the NOAA website (<https://www.noaa.gov>). Soil environmental factors, including soil pH, total potassium (TK), available potassium (AK), total phosphorus (TP), available phosphorus (AP), soil organic carbon (SOC), and total nitrogen (TN), were determined according to standard procedures (Pansu & Gautheyrou, 2006). The soil bacterial community was profiled using high-throughput sequencing as previously described (Luan, et al., 2020a). Alpha-diversity indices (taxon richness, Shannon index, Chao1, Faith's phylogenetic diversity [PD], nearest taxon index [NTI], and net relatedness index [NRI]) of bacterial communities were calculated using the 'vegan' and 'picante' packages after rarifying all samples to the same sequencing depth of 26,994 sequences (Kembel et al., 2010; Oksanen et al., 2019).

## Definition of abundant and rare taxa

We classified all operational taxonomic units (OTUs) into four categories based on the cutoff level of relative abundance (Xue et al., 2018): AT with relative abundances  $\geq 0.01\%$  in all samples and  $\geq 1\%$  in some or all samples; RT with relative abundances  $< 0.01\%$  in some or all samples but never  $\geq 1\%$  in any sample; moderate taxa (MT) with relative abundances between  $0.01\%$  and  $1\%$  in all samples; and CRAT with relative abundances ranging from rare ( $< 0.01\%$ ) to abundant ( $\geq 1\%$ ) (Figure 1B).

## Distance–decay slope and halving-distance

To assess the distance decay of community similarities, we regressed the  $\ln$ -transformed Bray–Curtis

community similarities ( $S$ ) against their geographic distances ( $D$ ). The geographic distances were determined based on the latitude and longitude of the sampling sites. The slopes of the DDRs were used to estimate the species turnover rates (Soininen et al., 2007):

$$\ln S = a + b \ln D, \quad (1)$$

where  $a$  is the intercept and  $b$  is the slope of the DDR.

The halving distance identifies the distance at which community similarity halves, with a smaller halving distance indicating a faster species turnover (Soininen et al., 2007). The halving distance was calculated for each community using a logarithmic decay model:

$$d_H = 10^{\frac{\ln \frac{S_0}{2} - a}{b}}, \quad (2)$$

where  $S_0$  is the initial community similarity at the lowest transit distance (1 km) (Soininen et al., 2007). Compared with other metrics of dispersal scales, the halving distance can be calculated for any type of regression between similarity and distance. As such, it provides a useful and comprehensible metric for the comparison across various habitats and organism types.

## Generalized linear models

To determine the relative importance of environmental factors and spatial factors on bacterial community structure, we modelled the bacterial community composition using multivariate negative binomial generalized linear models (Li et al., 2019a). Environmental factors contained pH, TK, AK, TP, AP, SOC, TN, C/N ratio, MAT, and MAP, while spatial factors contained geographic distance. The number of sequences in each OTU was treated as a taxa abundance. The model fitness of the soil physicochemical properties model and spatial model were compared using OTU-specific akaike information criterion (AIC) scores. A model was considered to support the other model if the difference in AIC ( $\Delta AIC$ ) > 2 (Alzarhani et al., 2019). The total AIC score across all OTUs ( $\Sigma AIC$ ) for each model was then calculated to be compared at the community level.

## Habitat niche breadth

Niche breadth was calculated according to Pandit et al. (2009), using Levins' niche breadth index ( $B$ ):

$$B_j = 1 / \sum_{i=1}^N P_{ij}^2 \quad (3)$$

where  $B_j$  represents the habitat niche breadth of OTU  $j$  in a metacommunity,  $N$  is the total number of

communities in each metacommunity, and  $P_{ij}$  is the proportion of OTU  $j$  in the community  $i$ . A high  $B$  value represents wide habitat niche breadth, indicating that the OTU occurs widely and evenly along with a large range of locations. The niche breadth of bacterial communities was calculated using the 'spaa' package (Zhang & Zhang, 2013).

## Neutral model

We used the Sloan neutral model to estimate the importance of dispersal on community assembly by predicting the relationship between the frequency of each taxon's occurrence in a set of local communities and their abundances across the wider metacommunity (Sloan et al., 2007). In the model, the dispersal rate ( $m$ ) is used to evaluate the probability of a random loss of an individual in a local community being replaced by an immigrant from the metacommunity. The calculation is as follows:

$$\text{Freq}_i = 1 - I(1 \div N | N \times m \times p_i, N \times m \times (1 - p_i)), \quad (4)$$

where  $\text{Freq}_i$  is the occurrence frequency of taxon  $i$  across communities,  $N$  is the number of individuals per community,  $p_i$  is the average relative abundance of taxon  $i$  across communities, and  $I()$  is the probability density function of the beta distribution;  $m$  is the estimated dispersal rate, and higher  $m$  value means a higher dispersal rate. This analysis was performed using nonlinear least-squares fitting with the 'minpack.lm' package (Elzhov et al., 2016). The 95% confidence intervals (CIs) of the model predictions were calculated using the Wilson score interval in the 'Hmisc' package (Harrell, 2013). The overall fit of the model to observed data was carried out by comparing the sum of squares of residuals,  $SS_{\text{err}}$ , with the total sum of squares,  $SS_{\text{total}}$ : model fit =  $1 - SS_{\text{err}}/SS_{\text{total}}$  (generalized  $R$ -squared). The fit of the neutral model and a binomial distribution model were compared to determine whether the model was based on the random sampling of the source metacommunity. Sampling from a binomial distribution represents the case where local communities are random subsets of the metacommunity in the absence of processes of drift and dispersal limitations (Sloan et al., 2007). The AIC and the Bayesian information criterion of each model were calculated based on 1000 bootstraps.

## Statistical analyses

Correlational analyses were based on Pearson ( $p < 0.05$ ). The turnover and nestedness of bacterial communities were calculated based on the Bray–Curtis dissimilarity index using the 'betapart' package (Baselga, 2010; Li, Jiang, et al., 2019b). Random forest



analyses were performed using the ‘randomforest’ package (Liaw & Wiener, 2002). The entire dataset was randomly divided into two parts, with about 2/3 used for the training dataset and the remaining for the ‘out-of-bag’ dataset (Wei et al., 2010). The importance of each factor was evaluated by analysis of the increase in the mean square error between the observed and predicted values seen when the predictor was randomly permuted (Breiman, 2001). The accuracy of the results was measured for each tree and then averaged across the forest containing 500 trees (Cutler et al., 2007).

## RESULTS

### General distribution of bacterial abundant and rare biospheres

After excluding singletons and rarefying sequences to a uniform depth, the dataset contained a total of 26,994 sequences in each sample. These were further subjected to OTU clustering, resulting in a total of 10,379 OTUs. First, we found a significant positive correlation between bacterial relative abundances and frequencies ( $R^2 = 0.68$ ,  $p < 0.001$ ) (Figure 1C). To partition the bacterial abundant and rare biospheres, we classified these OTUs into four categories according to OTU frequency and relative abundance (Figure 1D,E), including AT, RT, MT, and CRAT. The OTUs classified as AT accounted for 0.04% of the total OTUs and 5.84% of the total sequences; the OTUs classified as RT accounted for 98.77% of the total OTUs and 62.76% of the total sequences; the OTUs classified as MT accounted for 0.03% of the total OTUs and 0.43% of the total sequences; the OTUs classified as CRAT accounted for 1.17% of the total OTUs and 30.97% of the total sequences.

### Composition and diversity of bacterial abundant and rare biospheres

We detected a total of 47 phyla/classes, with RT containing 47 phyla/classes, AT containing 2 phyla/classes, CRAT containing 12 phyla/classes, and MT containing 3 phyla/classes (Figure S1). The RT sub-community was mostly composed of *Chloroflexi* (15.61%), *Actinobacteria* (19.88%), *Acidobacteria* (13.15%), *Alphaproteobacteria* (9.20%), *Gammaproteobacteria* (7.20%), *Deltaproteobacteria* (6.40%), *Planctomycetes* (5.13%), *Bacteroidetes* (3.76%), *Gemmatimonadetes* (2.42%), and *Firmicutes* (2.32%). The AT sub-community was composed of *Actinobacteria* (5.84%) and *Alphaproteobacteria* (0.31%). The CRAT sub-community was composed of *Gammaproteobacteria* (10.72%), *Actinobacteria* (9.01%),

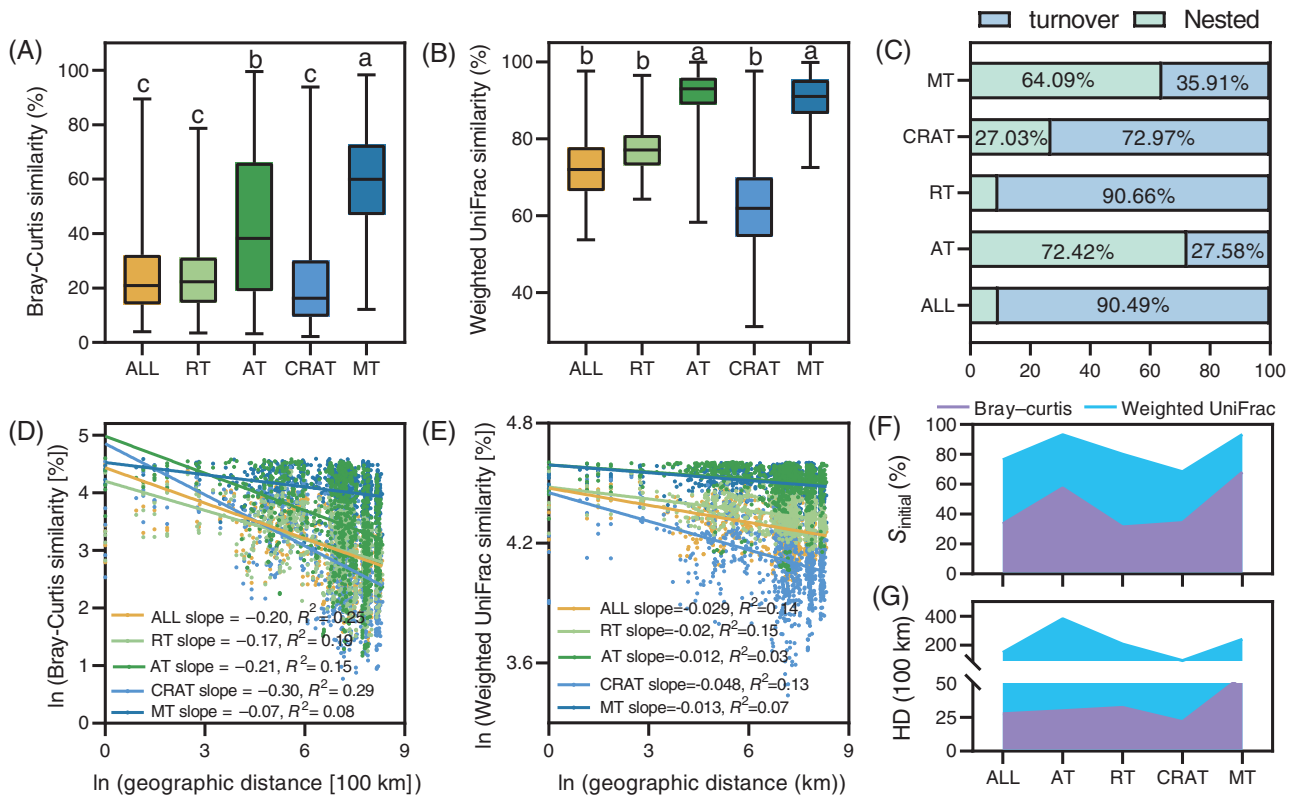
*Chloroflexi* (3.81%), *Bacteroidetes* (1.83%), *Firmicutes* (1.36%), and *Acidobacteria* (1.23%). And, the MT sub-community was composed of *Chloroflexi* (0.27%), *Actinobacteria* (0.09%), and *Acidobacteria* (0.08%).

All determined alpha-diversity indices (taxon richness, Chao1 index, and Shannon index) and PD (NRI and NTI) were significantly higher in RT ( $p < 0.05$ ) compared to AT, MT, and CRAT (Figure S2). The AT and MT had the lowest alpha diversity compared with the RT and CRAT. Random forest modelling was performed to assess the interpretability of the diversity of different taxa on the whole bacterial diversity (Figure S3). The models for different diversity indices were consistently significant ( $R^2 = 0.42$ – $0.69$ ,  $p < 0.01$ ). We found that the indices of taxon richness (RT = 16.32% and CRAT = 5.25%), Chao1 (RT = 15.82%), Shannon (RT = 80.12%, CRAT = 77.87%, AT = 62.31%, and MT = 9.23%), PD (RT = 16.75%), NRI (RT = 18.40%, CRAT = 14.34%, and AT = 11.81%), and NTI (RT = 15.92% and AT = 18.94%) of different taxa contributed significantly ( $p < 0.05$ ) to the corresponding indices of the whole bacterial diversity.

### Community dissimilarities and DDR

Both pairwise Bray–Curtis and weighted UniFrac similarity values of MT and AT were significantly greater than those of RT and CRAT ( $p < 0.001$ ) (Figure 2A,B). Nestedness (64.09% and 72.42%) overwhelmed the effect of species turnover for the community variations in MT and AT, and species turnover (72.97% and 90.66%) predicted the community variations in CRAT and RT better than nestedness (Figure 2C). Random forest models for different  $\beta$ -diversity metrics were consistently significant ( $R^2 = 0.43$ – $0.98$ ,  $p < 0.01$ ). We observed that Bray–Curtis (43.92% of RT, 35.75% of CRAT, 22.77% of AT, and 18.92% of MT), weighted UniFrac (47.58% of RT, 70.85% of CRAT, 14.15% of AT, and 18.73% of MT), turnover (71.59% of RT, 31.54% of CRAT, 21.54% of AT, and 22.17% of MT), and nestedness (48.35% of RT, 37.13% of CRAT, 33.45% of AT, and 42.42% of MT) indices of four different sub-communities significantly affected the corresponding  $\beta$ -diversity of the whole bacterial community (Figure S4).

DDRs were determined using both Bray–Curtis and weighted UniFrac distances. The whole bacterial community and four sub-communities consistently had significant DDRs (Figure 2D,E). Overall, the DDRs based on weighted UniFrac distances (slope =  $-0.01$  to  $-0.05$ ,  $R^2 = 0.03$ – $0.15$ ) were weaker than those based on Bray–Curtis (slope =  $-0.07$  to  $-0.30$ ,  $R^2 = 0.08$ – $0.29$ ), for all sub-communities. The DDR slope based on Bray–Curtis was the highest for CRAT, which displayed greater community variability (slope =



**FIGURE 2** General patterns of the bacterial beta-diversity in paddy soils across East Asia. Compositional variation of the whole bacterial community and different sub-communities based on Bray–Curtis (A) and weighted UniFrac (B) distance. Turnover and nestedness of the whole bacterial community and four sub-communities (C). Distance–decay relationship (DDR) regressions are based on In-transformed Bray–Curtis (D) and weighted UniFrac (E) similarities between pairs of communities and In-transformed geographic distances between sampling sites. The initial community similarity (F) and halving distance (G) of the DDRs of the whole bacterial community and different sub-communities. AT, abundant taxa; CRAT, conditionally rare and abundant taxa; MT, moderate taxa; RT, rare taxa.

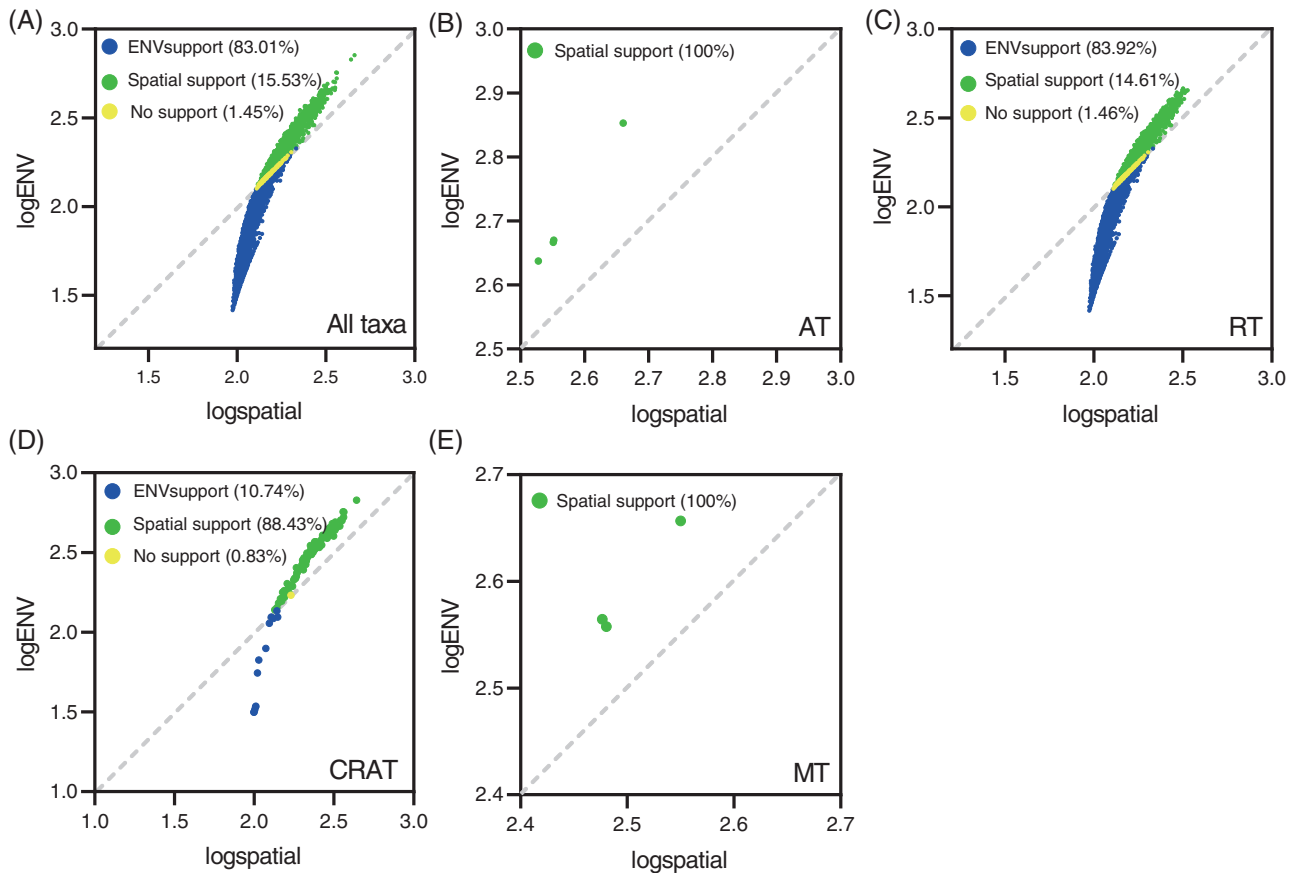
$-0.30$ ,  $p < 0.001$ ), and the lowest values of initial similarity (0.34) and halving distance (2199 km) (Figure 2F,G). In contrast, MT had the weaker DDR with the weakest community variability (slope =  $-0.07$ ,  $p < 0.001$ ) and the largest values of initial similarity (0.68) and halving distance (6394 km). Besides, the DDR slope based on weighted UniFrac distances was the highest for CRAT with the greater community variability (slope =  $-0.048$ ,  $p < 0.001$ ) and the lowest values of initial similarity (0.68) and halving distance (9285 km). Conversely, AT had the lowest DDR with the weakest community variability (slope =  $-0.012$ ,  $p < 0.001$ ), and the largest values of initial similarity (0.93) and halving distance (38,150 km).

### Community assembly of bacterial abundant and rare biospheres and GLMs analysis

Multivariate negative binomial generalized linear models (GLMs) analysis was used to disentangle the relative influence of environmental factors and spatial factors on the composition of the abundant and rare

biospheres compared with the whole community (Figure 3). All AT, MT, and most of CRAT (88.43%) were better explained by spatial variables than environmental ones, while most RT (83.92%) followed the opposite trend.

We determined the dispersal rate (indicated by the  $m$  parameter) of each OTU by fitting the frequency distribution to the Sloan neutral model (Figure 4A). All AT and MT, as well as 8.26% of CRAT and 25% of RT, were selectively enriched (Figure 4B). There were 9.5% of RT and 85.12% of CRAT OTUs neutrally assembled. Our results showed that CRAT ( $m = 0.007$ ,  $R^2 = 0.16$ ) and RT ( $m = 0.09$ ,  $R^2 = 0.50$ ) had the lowest and highest  $m$  values, respectively (Figure 4C, Table S1). The niche breadth was calculated to evaluate the influence of selection on community assembly, with a wider niche breadth ( $B_{\text{com}}$ ) reflecting lower selection (Figure 4D). We observed that RT and MT had the lowest and highest  $B_{\text{com}}$  values (7.67 and 29.72), respectively. Random forest models indicated that soil factors (pH = 9.13%–49.78% and TK = 10.12%–42.37%), climate factors (MAT = 16.59%–46.49% and MAP = 10.86%–37.56%), and



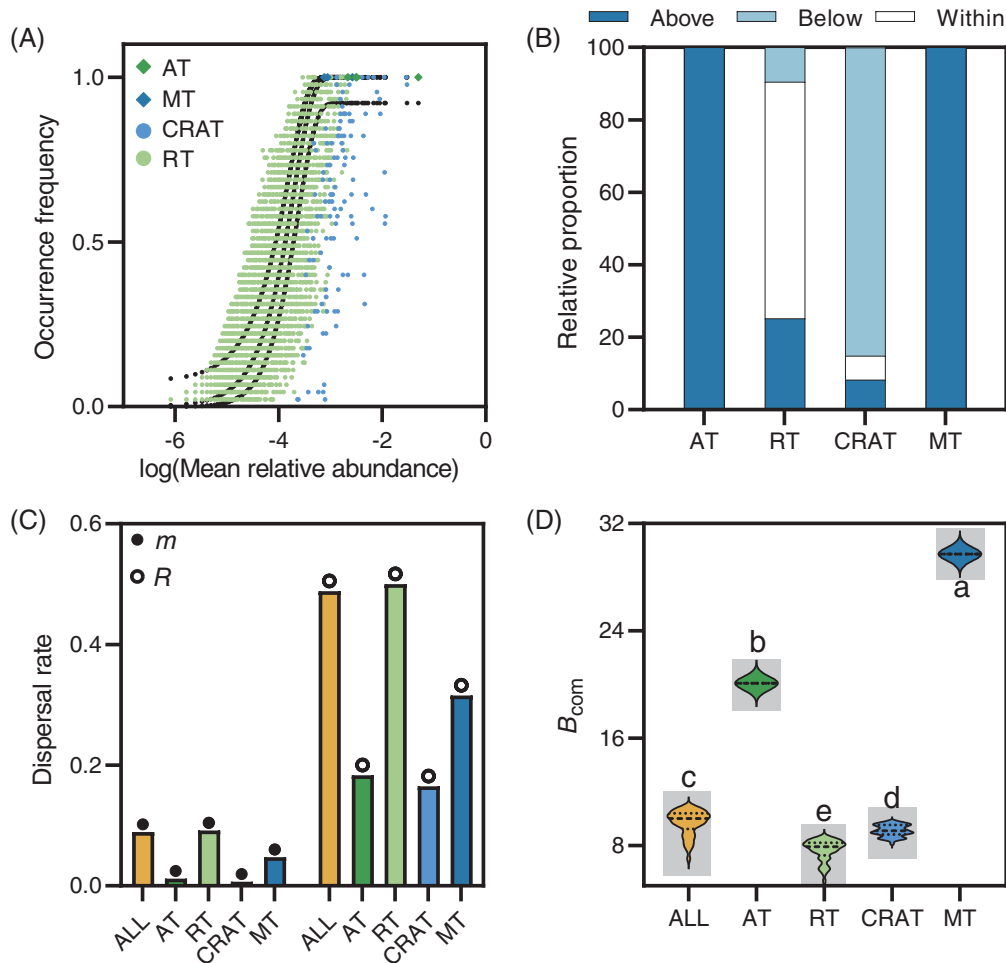
**FIGURE 3** Influence factors predicting the composition of the whole bacterial community and different sub-communities. The effects of environmental factors and geographic distance on the composition of the whole bacterial community (A) and different sub-communities (B)–(E) are examined via multivariate negative binomial generalized linear models. A lower AIC score represents a superior fit relative to the number of variables in each model. Grey dotted lines indicate equal AIC scores for the two models. An OTU in green indicates that the OTU is more affected by spatial factors; an OTU in blue indicates that the OTU is more affected by environmental factors; and an OTU in yellow indicates that the OTU is influenced by neither spatial factors nor environmental factors. AT, abundant taxa; CRAT, conditionally rare and abundant taxa; MT, moderate taxa; RT, rare taxa.

biological factors (Shannon = 7.16%–102.68%, NTI = 9.49%–60.81%, and NRI = 5.99%–61.78%) were the most important predictors of the variation in a bacterial community (Figure 5).

## DISCUSSION

In this study, we used a continental paddy soil dataset to investigate the biogeographical patterns and ecological mechanisms underpinning the bacterial abundant and rare biosphere assemblies. We determined the dynamic interplay of ecological processes (environmental filtering and dispersal limitation) structuring DDR patterns for the whole bacterial community, as well as abundant and rare sub-communities. Overall, the obtained results corroborate our initial hypothesis that four sub-communities had significantly distinct DDR patterns, with environmental filtering and dispersal limitations differentially affecting the intensities of their respective DDRs.

We found significant DDRs for all tested bacterial sub-communities, indicating that bacteria were not ubiquitously distributed across space. The DDRs calculated based on the taxonomic index (Bray–Curtis similarity) were significantly stronger than the DDRs based on phylogenetic distances (weighted UniFrac). This suggests that overall phylogenetic metrics tend to cluster communities and/or operate at a lower resolution. The literatures suggest that the complex microbial communities—in some cases—appear to be generally similar, albeit their differences in composition always exist at finer taxonomic resolutions (e.g., species or strain level) (Bryant et al., 2008; Clark et al., 2021). Our results indicated that weighted UniFrac similarity and initial similarity were approximately three and two times higher than the Bray–Curtis similarity and initial similarity, respectively (Figure 2). It has also been reported that communities with less common species can exhibit high phylogenetic similarity when the species share many branches in the phylogenetic tree, thereby reducing the decay of similarity over geographical distance



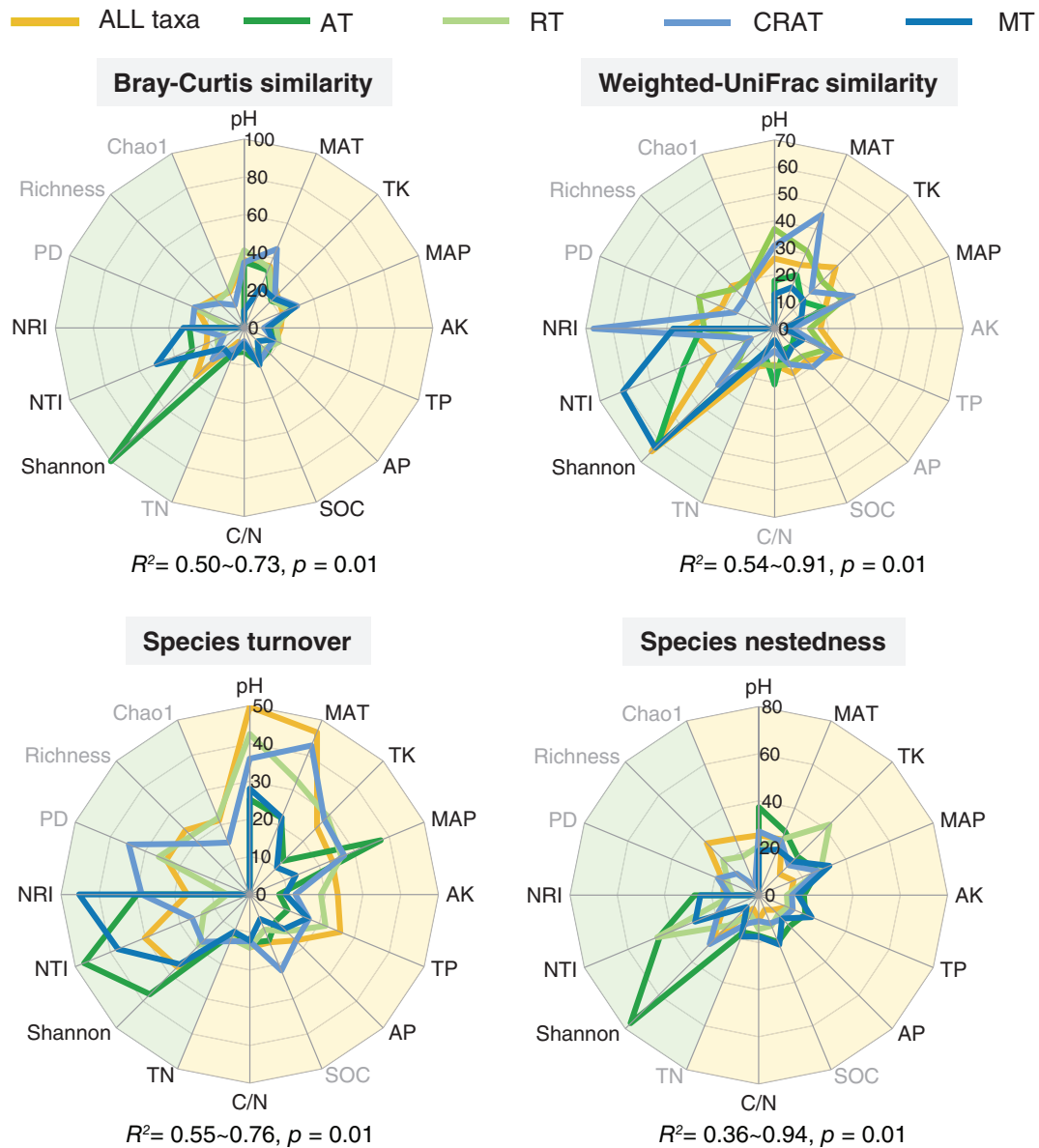
**FIGURE 4** The ecological processes acting upon different bacterial taxa. The fit of the neutral model for bacterial taxa (A). The predicted occurrence frequency is shown as a solid black line, and points with different colours represent the distribution of different bacterial sub-communities in the model. OTUs of different bacterial sub-communities that occur more or less frequently than predicted by the neutral model (B). The fit of the neutral model for different bacterial sub-communities (C). The niche breadths of different bacterial sub-communities (D). Different lower cases indicate significant differences between different sub-communities based on Tukey's HSD test ( $p < 0.05$ ). AT, abundant taxa; CRAT, conditionally rare and abundant taxa; MT, moderate taxa; RT, rare taxa.

(Bryant et al., 2008; Clark et al., 2021). Within this general pattern, CRAT ( $R^2 = 0.29$ ) and RT ( $R^2 = 0.19$ ) had stronger DDRs than AT ( $R^2 = 0.15$ ) and MT ( $R^2 = 0.07$ ), indicating that the community turnover was higher in CRAT and RT sub-communities than in AT and MT sub-communities. Besides, such results might likely be structured by the dynamic interplay of ecological processes operating in these sub-communities.

We revealed that RT had a higher dispersal rate ( $m = 0.092$ ) than MT ( $m = 0.048$ ), AT ( $m = 0.012$ ), and CRAT ( $m = 0.007$ ), implying that RT had a greater dispersal potential. The effects of dispersal limitation on rare and abundant taxa are not always uniform and may depend largely on ecosystem type and geographic scale (Liu et al., 2015; Shi et al., 2018; Wang et al., 2013). Dispersal limitation significantly influenced the AT sub-community rather than the RT sub-community, possibly because the continental scale in our study

spanned a long geographic distance of 4000 km. It is widely known that large geographic distances make it difficult to detect highly ubiquitous species worldwide (Martiny et al., 2006). We observed that the AT sub-community (4 OTUs) represented 5.84% of the total sequences (the average abundance of each OTU was 1.46%), while the RT sub-community (10,251 OTUs) accounted for 62.76% of the total sequences (the average abundance of each OTU was 0.006%). As such, significantly greater species numbers and population densities can enhance the likelihood of passive dispersal of RT via passive mechanisms (e.g., wind, water, and/or hitchhiking on macro-organisms) (Villarino et al., 2018; Woo et al., 2018). We note that the CRAT sub-community with a large relative abundance was mostly influenced by dispersal limitation. The CRAT sub-community was more sensitive to environmental changes during the dispersal process, potentially resulting in dispersal limitation (Jia



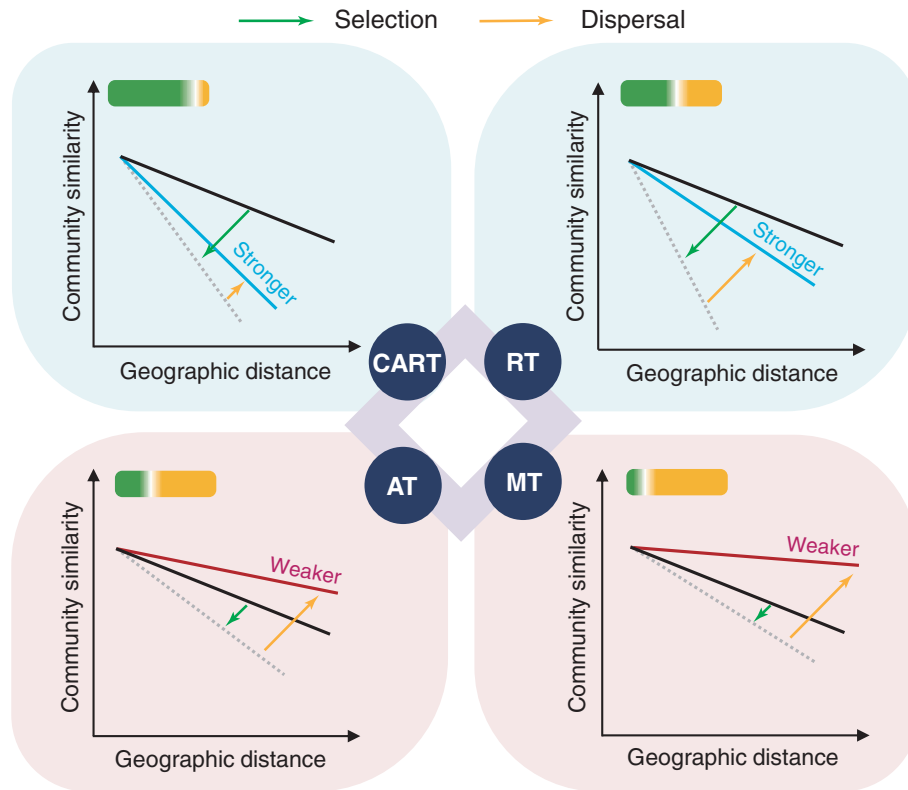


**FIGURE 5** Mean predictor importance (percentage increased mean square error) of different bacterial sub-community variations. Environmental factors include the average annual air temperature (MAT), the average annual precipitation (MAP), soil pH, total potassium (TK), available potassium (AK), total phosphorus (TP), available phosphorus (AP), soil organic carbon (SOC), and total nitrogen (TN). Biological factors include Shannon, phylogenetic diversity (PD), richness, nearest taxon index (NRI), Chao1, and net relatedness index (NTI). Significance levels of predictors are indicated as follows: \*\*\* $p < 0.001$ ; \*\* $p < 0.01$ ; \* $p < 0.05$ .

et al., 2018). In fact, the dispersal rates of all sub-communities were lower than 0.1, indicating that all taxa were subject to some extent to dispersal limitation. In addition to the large geographic distance, reduced fluidity may have negative consequences for the dispersal potentials of natural bacterial communities in terrestrial ecosystems (Jiao et al., 2020; Luan et al., 2020a).

Our results further showed that environmental factors (especially soil pH and MAT) exhibited a more pronouncedly influence on RT and CRAT communities than AT and MT communities. This was further confirmed by the results of niche breadth, which showed

that RT (avg.  $B_{com} = 7.67$ ) and CRAT (avg.  $B_{com} = 9.10$ ) had narrower niche breadth compared to AT (avg.  $B_{com} = 20.09$ ) and MT (avg.  $B_{com} = 29.72$ ). This indicates that RT and CRAT were under greater influence from environmental filtering than AT and MT (Figure 4). The narrower niche breadth may result from the rare species' particular life-history strategies and ecological trade-offs. For example, oligotrophic organisms can successfully persist under low-nutrient conditions at low abundances, mostly due to intrinsic metabolic constraints (Fierer et al., 2007). Alternatively, the persistent low abundance of some RT may be a consequence of selective exclusion by shifting environmental conditions



**FIGURE 6** Conceptual frameworks showing the distance–decay relationships (DDRs) of different bacterial sub-communities. Environmental filtering and dispersal limitations jointly shape the geographic DDR of conditionally rare and abundant taxa (CRAT), leading to a strong distance–decay relationship (A). Although the rare taxa (RT) sub-community is more influenced by strong environmental filtering, their large dispersal rate flattens the DDR pattern (B). moderate taxa (MT) and abundant taxa (AT) sub-communities were predominantly influenced by dispersal limitation and thus produced weaker DDR relationships (C, D). MT had the weakest DDR relationship due to its wider niche breadth (D). The relative contributions of environmental filtering and dispersal rate to the distance–decay relationship of different sub-communities were coloured green and orange, respectively.

(Jia et al., 2018). In the case of bacteria, some taxa can enter a reversible state of decreased metabolic activity by engaging in dormancy and can persist at low densities for long periods to contend with unfavourable or harsh local environmental conditions (Lennon & Jones, 2011). Conversely, the higher niche breadth of abundant species may be dependent upon a combination of intrinsic species characteristics and environmental selection. Abundant bacterial taxa have been proposed as habitat generalists able to utilize a wide spectrum of resources available in the environment (Hambricht et al., 2015). The unique microhabitats in paddy soils could selectively enrich specific bacterial populations to be widespread (Figure 4B). However, CRAT is composed of many opportunistic species with narrow niche breadth and is therefore highly sensitive to climatic–environmental changes (Jia et al., 2018). When environmental conditions are unfavourable, CRAT exists at low abundances or in a dormant state, or grows slowly. Once they encounter a suitable environment, CRAT will gradually recover their population densities and become abundant members of the community (Evans et al., 2014; Jia et al., 2018).

Last, we propose a conceptual framework describing the relative influence of environmental selection and dispersal rate on DDRs in paddy soils. In brief, environmental filtering and dispersal limitation collectively mediate the geographic distribution of the CRAT sub-community, leading to a strong DDR pattern (Figure 6A). The RT sub-community is strongly affected by environmental filtering, but greater dispersal rates flatten the intensity of the DDR (Figure 6B). The MT and AT sub-communities are less influenced by environmental filtering and dispersal limitation, thus resulting in weak DDRs. The MT sub-community has the weaker DDR due to the wider niche breadth compared to AT sub-community (Figure 6C,D). Corroborating previous studies, we also suggest that MT and AT were not strongly structured by environmental filtering, probably because of their wide range of inherent growth rates and/or their significant advantages in adapting to unique environments. However, we do not rule out the importance of unmeasured abiotic and biotic variables potentially accounting for the observed differences in the MT and RT sub-communities (Luan et al., 2020a).

## CONCLUSION

Our results demonstrated that both the entire bacterial community and four sub-communities had significant DDRs at taxonomic and phylogenetic levels in paddy soils. In particular, we revealed that the CRAT sub-community had the strongest DDRs due to greater environmental filtering and dispersal limitations. On the contrary, the MT sub-community had the weakest DDRs owing to the lowest intensity of environmental filtering and greater dispersal. Taken together, our study highlights an interplay of ecological processes dynamically structuring the bacterial abundant and rare biospheres, with direct implications for DDR patterns in paddy soils.

## AUTHOR CONTRIBUTIONS

Yuji Jiang, Lu Luan, and Bo Sun designed the experiments. Lu Luan and Yuji Jiang performed soil sampling, field and laboratory experiments, and contributed to the data analysis. All the authors wrote the manuscript, discussed the results, and commented on the manuscript.

## ACKNOWLEDGEMENTS

This study was supported by the National Natural Science Foundation for Excellent Young Scholars of China (41922048), the National Natural Science Foundation of China (42177298, 42107336), the Youth Innovation Promotion Association of CAS (Y2021084), the Science and Technology Project of Jiangsu Province (BE2022393), and the China Postdoctoral Science Foundation (2021M690155).

## CONFLICT OF INTEREST

The authors have declared that no competing interests exist.

## DATA AVAILABILITY STATEMENT

Data used in this work are available from the corresponding authors upon request. The sequences of the 16 S rRNA gene have been deposited in the Sequence Read Archive (SRA) at the National Center for Biotechnology Information (NCBI) with the accession number PRJNA607877. The environmental data and geographical location information of soil samples have been deposited in the Figshare database (<https://doi.org/10.6084/m9.figshare.12622829>).

## REFERENCES

Alzarhani, A.K., Clark, D.R., Underwood, G.J.C., Ford, H., Cotton, T. E.A. & Dumbrell, A.J. (2019) Are drivers of root-associated fungal community structure context specific? *The ISME Journal*, 13, 1330–1344.

Bahram, M., Hildebrand, F., Forslund, S.K., Anderson, J.L., Soudzilovskaia, N.A., Bodegom, P.M. et al. (2018) Structure and function of the global topsoil microbiome. *Nature*, 560, 233–237.

Baselga, A. (2010) Partitioning the turnover and nestedness components of beta diversity. *Global Ecology and Biogeography*, 19, 134–143.

Breiman, L. (2001) Random forests. *Machine Learning*, 45, 5–32.

Brown, M.V., Ostrowski, M., Grzymalski, J.J. & Lauro, F.M. (2014) A trait based perspective on the biogeography of common and abundant marine bacterioplankton clades. *Marine Genomics*, 15, 17–28.

Bryant, J.A., Lamanna, C., Morlon, H., Kerkhoff, A.J., Enquist, B.J. & Green, J.L. (2008) Microbes on mountainsides: contrasting elevational patterns of bacterial and plant diversity. *Proceedings of the National Academy of Sciences of the USA*, 105, 11505–11511.

Clark, D.R., Underwood, G.J.C., Mcgenity, T.J. & Dumbrell, A.J. (2021) What drives study-dependent differences in distance–decay relationships of microbial communities? *Global Ecology and Biogeography*, 30(4), 811–825.

Cottrell, M.T. & Kirchman, D.L. (2003) Contribution of major bacterial groups to bacterial biomass production (thymidine and leucine incorporation) in the Delaware estuary. *Limnology and Oceanography*, 48, 168–178.

Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J. et al. (2007) Random forests for classification in ecology. *Ecology*, 88, 2783–2792.

Dimitriu, A., David, P., Lee, J. & Grayston, S. (2010) An evaluation of the functional significance of peat microorganisms using a reciprocal transplant approach. *Soil Biology and Biochemistry*, 42, 65–71.

Elzhov, T.V., Mullen, K.M., Spiess, A.-N., and Bolker, B. (2016) Minpack. Im: R Interface to the levenberg–marquardt nonlinear least-squares algorithm found in MINPACK, plus support for bounds (CRAN repository).

Evans, S.E., Wallenstein, M.D. & Fierer, N. (2014) Climate change alters ecological strategies of soil bacteria. *Ecology Letters*, 17(2), 155–164.

Fierer, N., Bradford, M.A. & Jackson, R.B. (2007) Toward an ecological classification of soil bacteria. *Ecology*, 88, 1354–1364.

Hambright, K.D., Beyer, J.E., Easton, J.D., Zamor, R.M., Easton, A. C. & Halliday, T.C. (2015) The niche of an invasive marine microbe in a subtropical freshwater impoundment. *The ISME Journal*, 9, 256–264.

Hanson, C.A., Fuhrman, J.A., Horner-Devine, M.C. & Martiny, J.B. (2012) Beyond biogeographic patterns: processes shaping the microbial landscape. *Nature Reviews Microbiology*, 10, 497–506.

Harrell, F. E. (2013) Jr. Hmisc: Harrell miscellaneous. R. package version 3.0–12.

Hedrick, P.W. (1999) Perspective: highly variable loci and their interpretation in evolution and conservation. *Evolution*, 53, 313–318.

Jia, X., Dini-Andreote, F. & Salles, J.F. (2018) Community assembly processes of the microbial rare biosphere. *Trends in Microbiology*, 26(9), 738–747.

Jiang, Y.J., Liang, Y.T., Li, C.M., Wang, F., Sui, Y.Y., Suvannang, N. et al. (2016) Crop rotations alter bacterial and fungal diversity in paddy soils across East Asia. *Soil Biology and Biochemistry*, 95, 250–261.

Jiao, S. & Lu, Y. (2020) Soil pH and temperature regulate assembly processes of abundant and rare bacterial communities in agricultural ecosystems. *Environmental Microbiology*, 22(3), 1052–1065.

Jiao, S., Yang, Y., Xu, Y., Zhang, J. & Lu, Y. (2020) Balance between community assembly processes mediates species coexistence in agricultural soil microbiomes across eastern China. *The ISME Journal*, 14, 202–216.

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. *The ISME Journal*, 11, 853–862.

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.

- Lennon, J.T. & Jones, S.E. (2011) Microbial seed banks: the ecological and evolutionary implications of dormancy. *Nature Reviews Microbiology*, 9, 119–130.
- Li, P.F., Liu, J., Jiang, C.Y., Wu, M., Liu, M. & Li, Z.P. (2019a) Distinct successions of common and rare bacteria in soil under humic acid amendment—a microcosm study. *Frontiers in Microbiology*, 10, 2271.
- Li, P.F., Liu, M., Ma, X.Y., Wu, M., Jiang, C.Y., Liu, K. et al. (2019b) Responses of microbial communities to a gradient of pig manure amendment in red paddy soils. *Science of the Total Environment*, 705, 135884.
- Liaw, A. & Wiener, M. (2002) Classification and regression by random forest. *R News*, 2, 18–22.
- Liu, L., Yang, J., Yu, Z. & Wilkinson, D.M. (2015) The biogeography of abundant and rare bacterioplankton in the lakes and reservoirs of China. *The ISME Journal*, 9, 2068–2077.
- Luan, L., Jiang, Y.J., Cheng, M.H., Dini-Andreote, F., Sui, Y.Y., Xu, Q.S. et al. (2020a) Organism body size structures the soil microbial and nematode community assembly at a continental and global scale. *Nature Communications*, 11, 6406.
- Luan, L., Liang, C., Chen, L.J., Wang, H.T., Xu, Q.S., Jiang, Y.J. et al. (2020b) Coupling bacterial community assembly to microbial metabolism across soil profiles. *mSystems*, 5, e00298-20.
- Martiny, J.B.H., Bohannan, B.J.M., Brown, J.H., Colwell, R.K., Fuhrman, J.A., Green, J.L. et al. (2006) Microbial biogeography: putting microorganisms on the map. *Nature Reviews Microbiology*, 4, 102–112.
- Meyer, K.M., Memiaghe, H., Korte, L., Kenfack, D., Alonso, A. & Bohannan, B.J.M. (2018) Why do microbes exhibit weak biogeographic patterns? *The ISME Journal*, 12, 1404–1413.
- Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGinn, D. et al. (2019) Vegan: community ecology package. R. package version 2.5–4.
- Pandit, S.N., Kolasa, J. & Cottenie, K. (2009) Contrasts between habitat generalists and specialists: an empirical extension to the basic metacommunity framework. *Ecology*, 90, 2253–2262.
- Pansu, M. & Gautheryrou, J. (2006) *Handbook of soil analysis: mineralogical, organic, and inorganic methods*. Berlin: Springer.
- Pimm, S.L., Jones, H.L. & Diamond, J. (1988) On the risk of extinction. *The American Naturalist*, 132, 757–785.
- 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.
- Slatkin, M. (1993) Isolation by distance in equilibrium and non-equilibrium populations. *Evolution*, 47, 264–279.
- Sloan, W.T., Woodcock, S., Lunn, M., Head, I.M. & Curtis, T.P. (2007) Modeling taxa-abundance distributions in microbial communities using environmental sequence data. *Microbial Ecology*, 53, 443–455.
- Soininen, J., McDonald, R. & Hillebrand, H. (2007) The distance decay of similarity in ecological communities. *Ecography*, 30, 3–12.
- Villarino, E., Watson, J.R., Jonsson, B., Gasol, J.M., Salazar, G., Acinas, S.G. et al. (2018) Large-scale ocean connectivity and planktonic body size. *Nature Communications*, 9, 142.
- 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. *The ISME Journal*, 7, 1310–1321.
- Wei, C.L., Rowe, G.T., Escobar-Briones, E., Boetius, A., Soltwedel, T., Caley, M.J. et al. (2010) Global patterns and predictions of seafloor biomass using random forests. *PLoS One*, 5(12), e15323.
- Woo, C., An, C., Xu, S., Yi, S.M. & Yamamoto, N. (2018) Taxonomic diversity of fungi deposited from the atmosphere. *The ISME Journal*, 12, 2051–2060.
- Wu, W.X., Logares, R., Huang, B.Q. & 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. *Environmental Microbiology*, 19, 287–300.
- Xue, Y., Chen, H., Yang, J.R., Liu, M., Huang, B. & Yang, J. (2018) Distinct patterns and processes of abundant and rare eukaryotic plankton communities following a reservoir cyanobacterial bloom. *The ISME Journal*, 12, 2263–2277.
- Zhang, J., and Zhang, M.J. (2013) Spaa: species association analysis. R package version 0.2.1.
- Zheng, W., Zhao, Z.Y., Lv, F.L., Wang, R.Z., Wang, Z.H., Zhao, Z.Y. et al. (2021) Assembly of abundant and rare bacterial and fungal sub-communities in different soil aggregate sizes in an apple orchard treated with cover crop and fertilizer. *Soil Biology and Biochemistry*, 156, 108222.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Luan, L., Shi, G., Zhu, G., Zheng, J., Fan, J., Dini-Andreote, F. et al. (2022) Biogeographical patterns of abundant and rare bacterial biospheres in paddy soils across East Asia. *Environmental Microbiology*, 1–12. Available from: <https://doi.org/10.1111/1462-2920.16281>