# Distinct assembly mechanisms underlie similar biogeographical patterns of rare and abundant bacteria in Tibetan Plateau grassland soils

Mukan Ji[1], Weidong Kong[1,2]*, Kang Zhao[1,2], Linyan Yue[1,2], Chunyan Guo[1,2], Xiaobin Dong[3], James Stegen[4], Don A Cowan[5]

[1]Key Laboratory of Alpine Ecology and Biodiversity, Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing 100101, P.R. China

[2] College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100039, P.R. China

[3]State Key Laboratory of Earth Surface Processes and Resource Ecology, College of Resources Science and Technology, Beijing Normal University, Beijing 100875, China

[4]Pacific Northwest National Laboratory, Richland, Washington 99352, USA

[5]Centre for Microbial Ecology and Genomics, Department of Genetics, University of Pretoria

Corresponding author *:

Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Building 3, Courtyard 16, Lincui Road, Chaoyang District, Beijing 100101, China.

Phone: 8610-84097039; Fax: 8610-84097079; E-mail: wdkong@itpcas.ac.cn

**Running head: Rare biosphere assembly is stochasticity driven**

# Abstract

The rare microbial biosphere, an essential part of biodiversity, plays a key role and acts as a functional reserve in many ecosystems. However, little is known regarding its assembly process. Here, we explore the composition, phylogenetic assembly (inferred by nearest taxon index, NTI) and community turnover (βNTI) of the rare bacterial microbiome in grassland soils on the Tibetan Plateau, China. Our results show that the rare bacterial community assembly is principally driven by stochastic processes at both compositional and phylogenetic levels, and is only weakly influenced by regional factors (mean annual precipitation and spatial distance). In contrast, deterministic processes drive the composition and the phylogenetic assembly processes of the dominant members of the soil community, and these factors are strongly driven by both local (plant diversity, above-ground biomass and soil nutrients) and regional factors. These results indicate that assembly processes affecting the rare bacterial community are distinctively different from the impacting the dominant microbiome fraction in soils, and suggest that the rare biosphere is more sensitive to climate effects such as precipitation changes.

# Introduction

Microbes are major components of natural ecosystems and play essential roles in global biogeochemical cycling[1]. Currently, most studies consider microbes in an ecosystem as a whole[2-4]. However, with improved sampling strategies and next-generation sequencing technologies, the existence of a rare microbial biosphere is now well-accepted[5]. The rare microbial biosphere typically consists of bacterial taxa that are present at low relative abundance[6], but is now known as a crucial part of an ecosystem and to contribute a large proportion of the biodiversity[7]. It has also been reported that the rare and abundant bacteria display distinct phylogeographical patterns, and are driven by different local and regional environmental factors[8]. The rare biosphere was once considered to be redundant and to merely serve as a "seed bank" in an ecosystem, and largely involved in the restoration of ecological function after environmental disturbance[9,10]. However, increasing evidence has shown that rare microbes are actively involved in metabolic turnover, particularly in contributing to nitrogen fixation and sulphur oxidation[11]. Rare bacterial taxa may also contribute indirectly to global warming by providing the energy required for methanogenesis[13].

Our knowledge on functions and diversity of the rare biosphere continues expanding, but their assembly process and driving factors remain unexplored. Two ecological processes (deterministic versus stochastic) have been proposed to drive microbial assembly[14]. The deterministic process is a niche-based process that includes selection imposed by abiotic environmental factors (environmental filtering) and species interaction[15]. In contrast, the stochastic theory conceptualizes from the neutral theory, where the community assembly is a result of unpredictable disturbance, speciation, probabilistic dispersal and random birth-death events[16]. The two assembly processes drive microbial assembly simultaneously, but their relative importance in an ecosystem varies from deterministic[15,17] to stochastic-dominated[2,18] Currently, the assembly process of the rare biosphere in soils remains unexplored, while a few existing studies investigated the ecosystem as a whole[3,4] or focused on less complex ecosystems, such as lakes[19]. It has been further observed that a single set of assembly rules may not fully describe the process of different microbial groups in an ecosystem[4]. Therefore, there is a major gap in the understanding of rare biosphere assembly process.

Several methods have been used to identify the process driving microbial community turnover, and most of which relied on comparing observed community β-diversity to those generated under a null model[4,20]. Webb *et al*. used phylogenetic distance-based index to infer the effect of deterministic and stochastic processes, under the assumption that phylogenetic close-related taxa occupy similar ecological niche and therefore subject to similar environmental selection[15,21,22]. Therefore, by examining the deviations of observed alpha- or beta- phylogenetic diversities from null model expectations, the importance of deterministic and stochastic processes on the community assembly can be interrogated quantitatively.

Here, we used the framework established by Webb *et al*.[21], and applied it on grassland soils on the Tibetan Plateau, China (Supplementary Fig. 1). We aimed to identify the rare biosphere assembly process, and disentangle the contribution of individual biotic and abiotic factor to rare biosphere biogeographical distribution. Here, we hypothesized that the rare sub-community assembly was primarily driven by stochasticity, while the abundant sub-community was assembled predominately by deterministic process.

## Results

## Compositional assembly of rare sub-community and influencing factors

The detailed sequencing results are shown in Supplementary material 1. In brief, a total of 4, 762 operational taxonomic units (OTUs) were grouped in our dataset and 45% were identified as rare OTUs, while abundant OTUs only accounted for 3.4% (Supplementary Fig. 2 and 3). The rare OTUs were classified into 23 bacterial phyla, which was closely approximate to that of the entire community (Supplementary Fig. 4 and 5). In contrast, the abundant OTUs were classified into only 12 phyla with the dominance of Actinobacteria, which was consistent with the entire community (Supplementary Fig. 4 and 5).

The between-sample community compositional variations were estimated using Bray-

Curtis dissimilarity matrix, and its relationship to environmental factors was visualised using distance-based redundancy analysis (Fig. 1). The rare and abundant sub-communities and the entire community all showed a compositional transition along the grassland ecosystems (Fig. 1), and they all exhibited significantly different compositions between desert grassland and steppe soils (one way-ANOVA, P<0.001). In addition, the rare and abundant sub-communities both showed statistically similar compositional variations compared to the entire community (all *P*<0.001), as inferred by Spearman mantel test. However, the rare sub-community showed a much lower similarity with the entire community than the abundant sub-community (Rho=0.442 versus 0.979, all *P*<0.001, Supplementary Fig. 6).

To disentangle the individual contribution of environmental factors to the compositional variations of rare and abundant sub-communities, distance-based Linear Modelling (DistLM) was employed in the study. The total variation of the rare sub-community compositions that explained by the environmental factors was much lower than that in the abundant sub-community (17.7% versus 48%, Table 1). Compositional variations of both rare and abundant sub-communities could be explained by climate, plant and soil factors, but their dominant environmental factors and explaining proportions differed substantially. The rare bacterial compositional variation was explained by regional factors by 8.3% (mean annual precipitation, MAP, mean annual temperature, MAT and spatial distance), as well as local factors, which explained a further 4.8% by plant (above-ground biomass and Shannon diversity) and 4.6% by soil (total organic carbon, TOC and pH). In contrast, the compositional variation of the abundant sub-community was predominantly explained by local factors by 37.6% (plant and soil) and regional factors by 10.4% (MAP and spatial distance). The dominant individual factors were MAP for the rare bacterial compositional variation (3.6%), and plant above-ground biomass for the abundant (15.6%), respectively.

# Phylogenetic assembly of rare sub-community and influencing factors

Phylogenetic community assembly process of the rare bacterial sub-community was assessed by nearest taxon index (NTI), which integrates bacterial compositions and

their phylogenetic relatedness[21]. For a single community, NTI greater than +2 indicates co-existing taxa being more closely related than expected by chance (phylogenetic cluster), while a NTI less than -2 indicates co-existing taxa being more distantly related (phylogenetic over dispersion)[15].

The mean NTI of rare sub-community was -0.37, and was not significantly different from the randomly generated null communities (mean $P$=0.58, Supplementary Fig. 7). The abundant sub-community NTI was significantly higher than that of the rare sub-communities (unpaired one tail t-test, $P$<0.001), with a mean of 3.1 (Fig. 2). This was true in individual grassland ecosystems (desert grassland and steppe), where the abundant sub-community NTIs were also significant higher than those of the rare sub-community (unpaired one tail t-test, $P$<0.001, Supplementary Fig. 8).

We disentangled the environmental factors that drove the NTI of rare and abundant sub-communities using partial correlation analysis and hierarchical partitioning. The results showed that the rare sub-community NTI did not correlate to any environmental factors (Table 2). In contrast, the abundant sub-community NTI negatively correlated to plant Shannon diversity ($P$<0.01, r=-0.43). Given the strong co-variations among environmental factors (Supplementary Table 2), we used hierarchical partitioning to identify the individual and joint contributions of the environmental factors to the NTI variations. The results demonstrated that the variation of rare sub-community NTI explained by environmental factors was much lower (3.3%) than that of abundant sub-community (32.3%) (Fig. 3). The rare sub-community NTI variation was not significantly explained by any individual factors. In contrast, the abundant bacterial NTI variation was significantly explained by most of the tested individual factors, among which MAP, plant Shannon diversity and TOC played key roles.

## Phylogenetic turnover of rare sub-community and influencing factors

Consistent with the NTI results, 81% of the βNTI for the rare sub-community were between -2 and 2, with an average of -1.41 (Fig. 4). In contrast, over 98% of the βNTI for the abundant sub-community were >2 with a mean of 3.0, and was significantly

higher than that of the rare sub-community. βNTI of the rare sub-community only significantly correlated to regional factors of MAP and spatial distance, but not to any local environmental factors such as plant and soil properties (Table 2). The abundant sub-community βNTI exhibited strong correlations with both regional and local factors, such as climate (MAP and aridity), spatial distance ($P$=0.03, r=0.09), plant (above-ground biomass) and soil nutrients (TOC and ammonium) (all $P \leq 0.05$). Taken together, although the phylogenetic turnover of the rare sub-community was dominated by stochastic process, it was also influenced by regional factors to a certain extent. In contrast, the abundant bacterial turnover was dominated by deterministic process, and driven by both regional and local factors.

## Discussion

Previous studies have predominantly focused soil microbial community assembly as a whole[3,4], which is heavily driven by deterministic process[15]. Here, we demonstrated that the assembly process of soil rare biosphere was driven by stochastic process at the levels of compositions, phylogenetic assembly and turnover, which is distinctly different from the abundant sub-community.

The stochastic assembly processes of rare biosphere were strongly evidenced by the NTI and βNTI, which both fell in the range of -2 to +2 (Fig. 2 and 4). Because NTI/βNTI greater than +2 indicates co-existing taxa being more closely related than expected by chance (phylogenetic clustering), while less than -2 indicates co-existing taxa being more distantly related (phylogenetic overdispersion)[15]. In contrast, except for one sampling site, the NTI and βNTI of the abundant sub-community were all above +2 (Fig. 2 and 4), and were significantly greater than the NTI of the rare sub-community. These results suggest that the abundant OTUs bear phylogenetic conserved traits that are well-adapted to the local environmental factors, and therefore present in greater abundances. This is consistent with previous findings that the abundant sub-community were determined by local environments and driven by deterministic process[15,23].

The correlations between NTI/βNTI and environmental factors further confirmed that environmental filtering has little impact on the rare biosphere phylogenetic assembly

and turnover. The rare sub-community NTI did not correlate to any environmental factors, and their βNTI only correlated to regional factors (MAP and spatial distance) rather than local factors (plant and soil) (Table 2 and Fig. 3). In contrast, the NTI and βNTI of the abundant sub-community significantly correlated to both regional and local factors (Table 2, Fig. 3), therefore indicating that its phylogenetic assembly process was driven by environmental filtering.

The contrasting assembly processes of the rare and abundant sub-communities identified here were in agreement with a previous report in oil-polluted soils, where the abundant bacterial assembly was driven by environmental filtering[24]. The rare taxa were influenced, to a much less extend, by the local environmental filtering[24]. Our findings were distinctively different from the reports in aquatic ecosystems, where both abundant and rare sub-communities were strongly influenced by environmental filtering[8,19,25]. We speculate that the discrepancy observed in soil and aquatic ecosystems was attributed to the environmental homogeneity. The aquatic ecosystem is much more homogenous than soils[26], therefore the environmental filtering driving abundant and rare bacterial species was more even. In contrast, soil is highly heterogeneous and bacterial interactions are much stronger[14], therefore various environmental micro-niches exist that allows the rare bacteria species to be less affected by the overall environments.

Although the rare bacterial assembly processes were stochastic, their compositions and phylogenetic turnover were at a certain degree influenced by regional factors and distinct from those of the abundant in grassland soils. This agrees with the findings in aquatic ecosystem, where rare and abundant sub-communities are influenced by different environmental factors[8,19,27]. In rare sub-community, the regional factors (MAP and spatial distance) were consistently identified as key factors influencing their community compositions and phylogenetic turnover. Contrarily, plant Shannon diversity/biomass and soil nutrients were the key factors driving the abundant sub-community assembly processes. MAP has been known to be involved in bacterial dispersion[28], and causes shifts in bacterial community compositions[29], although a large time scale maybe required (in decades) for microbial community variations to be detected[30]. Spatial distance has been observed to limit bacterial dispersal (dispersal

limitation)[31]. Our results showed that the rare sub-community compositional variation and phylogenetic turnover both significantly correlated to the spatial distance (Fig. 1and Table 2), indicating that rare OTUs were dispersal-limited.

The abundant sub-community compositions, phylogenetic assembly and turnover were mainly driven by local rather than regional factors (climate and spatial distance Fig. 1). This was supported by the dominant drive of abundant bacterial composition variation by plant factors (Table 1) and the negative correlations of their phylogenetic assembly and turnover with local factors (plant Shannon diversity/above-ground biomass and soil nutrients, Table 2). Plants promote bacterial diversity in several ways, such as via root exudates[32] and litter deposition[33,34], which all lead to an increased nutrient (carbon) input. Elevated nutrients and diverse resources have been proposed to increase stochasticity by enhancing ecological drift, weakening niche selection and reducing competition[20]. This hypothesis was partly supported by the dominance of Actinobacteria in abundant bacterial sub-community (Supplementary Fig. 4). Actinobacteria are widely regarded as heterotrophs, for their important roles in degrading plant biomaterials such as cellulose and plant litter[35,36]. Therefore the Actinobacteria dominance confirmed the role of plants in selecting abundant bacterial species.

Our results indicated that the soil rare biosphere exhibits contrasting compositional and phylogenetic assembly processes, which were predominately driven by stochasticity, while the abundant taxa were primarily driven by deterministicity or environmental filtering. The rare biosphere was influenced by regional factors (climate and geospatial distance), while the abundant bacteria were more strongly driven by local factors (plant and soil). These findings imply that the rare biosphere is more likely to be affected by large scale global change (such as precipitation change and global warming) than the abundant sub-community. Therefore, the rare biosphere shall be investigated separately from the abundant to capture a more comprehensive understanding to the soil microbial community.

# Materials and Methods

## Study area and sampling

The study area is located in central Tibetan Plateau (TP), and mainly located between 31 and 33°N latitude and 79 and 93°E longitude (east-west sampling area), where the average altitude is above 4400 m above sea level. Soil samples were collected at 11 sampling sites along a 2000 km transect across the TP (Supplementary Fig. 1) in July, 2015. The sampling sites belong to desert grassland (4 sites) and steppe (7 sites) ecosystems from west to east. The plant community was dominated by *Stipa breviflora* and *S. purpurea* in desert grassland and steppe ecosystems, respectivey[37]. At each sampling site, four to five 1 m × 1m quadrats were randomly selected. In each quadrat, five surface soils (0-1 cm) were taken randomly and combined. Soil samples were sieved through 2.0 mm to remove plant material and stones, and were transported to laboratory in coolers with ice bags. Subsamples taken for DNA extraction and physicochemical analyses were stored at -80°C. Root samples were collected using five soil cores (2.5 cm diameter × 10 cm depth) per quadrat from 0-10 cm soil layer, soil cores were collected from the same areas where surface soil was collected. The soil cores were washed by running water using a 0.25 mm sieve to remove soils and stones. Roots were collected carefully into paper bags and oven-dried (65°C for 24 hours) for biomass measurement. The aboveground plants were clipped and stored in paper bags according to species and were also oven-dried for biomass measurement and plant Shannon diversity calculation.

## Physicochemical analysis

Soil pH was measured in a 1:5 soil-to-water suspension using a pH meter (Sartorius PB-10, Germany). Soil nitrate ($NO_3^-$) and ammonium ($NH_4^+$) were extracted with 2 M KCl (1:5) and determined using Smartchem200 Discrete Auto Analyser (Alliance, France). Total organic carbon (TOC) was measured in the solid state using a TOC analyser (TOC-VCPH, Shimadzu, Japan). The aridity index (potential evaporation/precipitation) of each site was obtained by using Global Aridity Index dataset[38], which was available at www.cgiar-csi.org. The mean annual precipitation

(MAP) and the mean annual temperature (MAT) were predicted from the meteorological data from 33 climatic stations (China meteorological Data Sharing Service System; http://cdc.cma.gov.cn/) during period of 2003-2012 using the Kriging interpolation.

## DNA extraction, PCR and high-throughput sequencing

Total genomic DNA was extracted using the MO BIO PowerSoil DNA extraction kit (Mo Bio Laboratories, Carlsbad, CA, USA) according to the manufacturer's instructions. Universal primer 515F (5'-GTGCCAGCMGCCGCGGTAA-3') and 909r (5'-GGACTACHVGGGTWTCTAAT-3') with 12 nt unique barcode, was used to amplify the V4 hyper-variable region of 16S rRNA gene[39]. The PCR mixture (25 μl) contained 1x PCR buffer, 1.5 mM of $MgCl_2$, 0.4 μM each of deoxynucleoside triphosphate, 1.0 μM of each primer, 0.5 U of Ex Taq (TaKaRa, Dalian, China) and 1 μl of DNA template (20 ng). The PCR amplification program included initial denaturation at 94 °C for 3 min, followed by 30 cycles of 94 °C for 20 s, 56 °C for 30 s, and 72 °C for 45 s, and a final extension at 72 °C for 10 min. Triplicate PCR reactions were conducted for each sample, and PCR products were pooled for purification using OMEGA Gel Extraction Kit (Omega Bio-Tek, USA) following electrophoresis. PCR products from different samples were pooled with equal molar amount, and then applied to pair-end sequencing (2x250 bp) using the Illumina MiSeq sequencer at Chengdu Institute of Biology, Chinese Academy of Sciences.

## Sequence analysis

Raw sequence data were processed using MOTHUR pipeline (v. 1.34.3)[40]. Paired-end reads were merged and sequences were quality screened with following settings: any sequences with length <300 or >400, more than 1 mismatches at the primer region, average quality <35, ambiguous bases >0 and homopolymer length >9 were removed for further analysis. The remaining sequences were aligned to Silva reference alignment (release 128), which was trimmed to the same region amplified, and those sequences that did not align were removed. Chimera sequences were screened using UCHIME[41]. The sequences were classified using Bayesian classifier against Silva database (release 128), with a minimum confidence score of 80%[42], then all Archaea, Eukaryota,

chloroplasts, mitochondria and unknown sequences were culled. Finally, sequences were classified into operational taxonomic units (OTUs) at 97% identity. The OTUs with only one sequence across the entire dataset were considered as singleton and removed, and then the datasets were sub-sampled to equal depth of 10983, which was smallest sample size across the entire dataset. Community alpha diversity indices (Ace, Chao1, Good's coverage, Shannon, Simpson and species observed) were calculated using Mothur[40].

Currently, there weren't a consistent method of defining abundance and rare OTUs, and current methods are mostly arbitrary[43]. In this study they were defined based on their relative abundance, which was one of the most popular methods been used[8,44,45]. Abundant OTU was defined as an OTU with a relative abundance >1% within any sample, or having an average abundance >0.1% across all samples. Rare OTUs were defined as an OTU with a relative abundance <0.01% in any sample or having an average abundance <0.001% across all samples.

# Data analysis

## Phylogenetic signal

To use phylogenetic information infer ecological processes, it requires phylogenetic signal in OTU's optimal habitat conditions[15] i.e., the phylogenetic related taxa have similar habitat preference. We tested whether phylogenetic signal exist in our dataset and determined the appropriate phylogenetic distance that phylogenetic signal exist. The abundance-weighted environmental niche value of each measured biotic and abiotic factor was calculated for each OTU, and a Euclidean distance matrix was built using 'vegdist' in package 'vegan'. A Mantel correlogram was plotted using 'mantel.correlog' from the same package, which shows the Spearman correlation relationship between the between-OTU niche difference and between-OTU phylogenetic distance (arbitrary value) at various phylogenetic distance bins. This allows identification of phylogenetic distance threshold beyond which niche differences no longer increased with phylogenetic distance[22].

## Phylogenetic community assembly process

A positive relationship (Spearman correlations) was only observed between-OTU niche differences and between-OTU phylogenetic distances across relatively short phylogenetic distances (Supplementary Fig. 8), which is consistent with previous researches[15,22]. Therefore we used mean nearest taxon distance (MNTD) and nearest taxon index (NTI) to infer the assembly process of rare and abundant sub-communities within each sample (unique point in space and time)[15]. MNTD was the averaged minimum phylogenetic distance between each OTU in a sample and the nearest OTUs within the same sample[21]. The MNTD and NTI were calculated using 'mntd' and 'ses.mntd' in package 'picante' with the null model generated by randomized the OTUs and their relative abundances across the tips of phylogeny[15,21]. Only the weighted version of NTI was calculated as Freilich and Connolly[46] suggested that using abundance weighting can substantially increase the power to detect assembly process.

## Community composition and phylogenetic turnover

The community compositional turnover was estimated using Bray-Curtis dissimilarity matrix, which was calculated from square root-transformed sample relative abundance matrix using 'vegdist' in package 'vegan'. The dbRDA plots were generated using 'capscale' in package 'vegan' to visualise the relationship between all measured factors and the community compositional turnover. The turnover in phylogenetic composition through space was quantified using Beta Mean Nearest Taxon Distance (βMNTD) and Beta Nearest Taxon Index (βNTI), which are the between-sample analogue of MNTD and NTI, respectively[15]. Similar to MNTD and NTI, βMNTD quantifies weighted phylogenetic distance among closest taxon in two different communities, βNTI measure the deviations of the observed βMNTD is from the mean of the null distribution. βMNTD and βNTI were calculated using 'comdistnt' in package 'picante'.

## Identify the driving factors of community compositional and phylogenetic turnover

The normalities of the physicochemical variables were checked using Shapiro-Wilk test, and measured biotic and abiotic factors were transformed to reduce skewness. TOC,

Aridity, $NH_4^+$, plant above-ground and below-ground biomass were logarithm transformed; MAP and moisture were square root transformed, and pH and plant Shannon diversity were left untransformed. The spatial distance between each sampling site were calculated based on their longitude and latitude coordinates, and principle coordination analysis were performed using Primer 6 with the PERMANOVA+ package[47] to obtain the principle coordinates of the 1st axis, which were used as spatial factors of each sample. The environmental factors were group into four categories based on their properties, soil factors include pH, TOC, $NH_4^+$, $NO_3^-$ and soil moisture; climate factors include MAT, MAP and aridity; plant factors include plant Shannon diversity, plant above-ground and below-ground biomass; and lastly the spatial factor. The plant and soil factors were further classified as regional factor, while the regional factors contained climate and spatial distance as described previously[8]. To estimate the inter-correlation between measured biotic and abiotic factors, Pearson correlation between all biotic and abiotic factor pairs were calculated using 'rcorr' in package 'Hmisc'.

The contribution of all measured environmental factors to the community compositional variations observed were calculated using distance-based linear modelling (DistLM) using Primer 6 with the PERMANOVA+ package[47]. Due to the strong inter-correlation among measured biotic and abiotic factors, hierarchical partitioning was used to estimate the individual and joint contribution of each factor to NTI, as it has been shown to alleviate multicollinearity[48]. It was calculated using 'hier.part' in package 'hier.part' by calculating the goodness-of-fit with all possible combination of factors, and then the contribution of each factor was estimated based on the increased fit when that particular factor was included in the analysis[49]. The correlation between NTI/βNTI and measured environmental factors were also calculated using partial correlation analysis (with 'pcor.test' in package 'pcor.test') and partial mantel test (with 'mantel.partial' in package 'vegan'), respectively.

# References

1       Torsvik, V., Ovreas, L. & Thingstad, T. F. Prokaryotic diversity--magnitude, dynamics, and controlling factors. *Science* **296**, 1064-1066 (2002).
2       Zhang, X., Johnston, E. R., Liu, W., Li, L. & Han, X. Environmental changes affect the assembly of soil bacterial community primarily by mediating stochastic processes. *Glob. Change Biol.* **22**, 198-207, doi:10.1111/gcb.13080

(2016).

3     Ferrenberg, S. *et al.* Changes in assembly processes in soil bacterial communities following a wildfire disturbance. *ISME J.* **7**, 1102-1111, doi:10.1038/ismej.2013.11 (2013).

4     Caruso, T. *et al.* Stochastic and deterministic processes interact in the assembly of desert microbial communities on a global scale. *ISME J.* **5**, 1406-1413, doi:10.1038/ismej.2011.21 (2011).

5     Sogin, M. L. *et al.* Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proc. Natl. Acad. Sci. U.S.A.* **103**, 12115-12120 (2006).

6     Shade, A. *et al.* Conditionally rare taxa disproportionately contribute to temporal changes in microbial diversity. *mBio* **5** (2014).

7     Naeem, S. & Li, S. Biodiversity enhances ecosystem reliability. *Nature* **390**, 507-509 (1997).

8     Liu, L., Yang, J., Yu, Z. & Wilkinson, D. M. The biogeography of abundant and rare bacterioplankton in the lakes and reservoirs of China. *ISME J.* **9**, 2068-2077, doi:10.1038/ismej.2015.29 (2015).

9     Gibbons, S. M. *et al.* Evidence for a persistent microbial seed bank throughout the global ocean. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 4651-4655 (2013).

10    Mouillot, D. *et al.* Rare species support vulnerable functions in high-diversity ecosystems. *PLOS Biol.* **11**, e1001569, doi:10.1371/journal.pbio.1001569 (2013).

11    Hua, Z. *et al.* Ecological roles of dominant and rare prokaryotes in acid mine drainage revealed by metagenomics and metatranscriptomics. *ISME J.* **9**, 1280-1294, doi:10.1038/ismej.2014.212 (2015).

12    Hol, W. H. G. *et al.* Reduction of rare soil microbes modifies plant–herbivore interactions. *Ecol. Lett.* **13**, 292-301 (2010).

13    Pester, M., Bittner, N., Deevong, P., Wagner, M. & Loy, A. A 'rare biosphere' microorganism contributes to sulfate reduction in a peatland. *ISME J.* **4**, 1591-1602 (2010).

14    Tilman, D. Niche tradeoffs, neutrality, and community structure: a stochastic theory of resource competition, invasion, and community assembly. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 10854-10861 (2004).

15    Stegen, J. C., Lin, X., Konopka, A. E. & Fredrickson, J. K. Stochastic and deterministic assembly processes in subsurface microbial communities. *ISME J.* **6**, 1653-1664, doi:10.1038/ismej.2012.22 (2012).

16    Valverde, A., Makhalanyane, T. P. & Cowan, D. A. Contrasting assembly processes in a bacterial metacommunity along a desiccation gradient. *Front. Microbiol.* **5**, doi:10.3389/fmicb.2014.00668 (2014).

17    Lozupone, C. A. & Knight, R. Global patterns in bacterial diversity. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 11436-11440 (2007).

18    Bahram, M. *et al.* Stochastic distribution of small soil eukaryotes resulting from high dispersal and drift in a local environment. *ISME J.* **10**, 885-896, doi:10.1038/ismej.2015.164 (2016).

19    Liao, J. *et al.* Similar community assembly mechanisms underlie similar biogeography of rare and abundant bacteria in lakes on Yungui Plateau, China. *Limnol. Oceanogr.* **62**, 723-735, doi:10.1002/lno.10455 (2017).

20    Zhou, J. *et al.* Stochasticity, succession, and environmental perturbations in a fluidic ecosystem. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 201324044 (2014).

21    Webb, C. O., Ackerly, D. D. & Kembel, S. W. Phylocom: software for the analysis of phylogenetic community structure and trait evolution.

*Bioinformatics* **24**, 2098-2100 (2008).

22    Stegen, J. C. *et al.* Quantifying community assembly processes and identifying features that impose them. *ISME J.* **7**, 2069-2079, doi:10.1038/ismej.2013.93 (2013).

23    Kembel, S. W. Disentangling niche and neutral influences on community assembly: assessing the performance of community phylogenetic structure tests. *Ecol. Lett.* **12**, 949-960 (2009).

24    Jiao, S. *et al.* Distinct succession patterns of abundant and rare bacteria in temporal microcosms with pollutants. *Environ. Pollut.* **225**, 497-505, doi:10.1016/j.envpol.2017.03.015 (2017).

25    Logares, R. *et al.* Biogeography of bacterial communities exposed to progressive long-term environmental change. *ISME J.* **7**, 937-948, doi:10.1038/ismej.2012.168 (2013).

26    Quesada, E., Ventosa, A., Rodriguezvalera, F., Megias, L. & Ramoscormenzana, A. Numerical taxonomy of moderately halophilic Gram-negative Bacteria from hypersaline soils. *Microbiology* **129**, 2649-2657 (1983).

27    Wu, W., Logares, R., Huang, B. & Hsieh, C.-h. 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, doi:10.1111/1462-2920.13606 (2017).

28    Burrows, S. M., Elbert, W., Lawrence, M. G. & Poschl, U. Bacteria in the global atmosphere - Part 1: review and synthesis of literature data for different ecosystems. *Atmospheric Chem. Phys.* **9**, 9263-9280 (2009).

29    Zhang, K. *et al.* Effects of short-term warming and altered precipitation on soil microbial communities in alpine grassland of the Tibetan Plateau. *Front. Microbiol.* **7**, 1032 (2016).

30    Deangelis, K. M. *et al.* Long-term forest soil warming alters microbial communities in temperate forest soils. *Front. Microbiol.* **6**, 104-104 (2015).

31    Eisenlord, S. D., Zak, D. R. & Upchurch, R. A. Dispersal limitation and the assembly of soil Actinobacteria communities in a long-term chronosequence. *Ecol. Evol.* **2**, 538-549 (2012).

32    Chaparro, J. M., Badri, D. V. & Vivanco, J. M. Rhizosphere microbiome assemblage is affected by plant development. *ISME J.* **8**, 790-803, doi:10.1038/ismej.2013.196 (2014).

33    Prescott, C. E. & Grayston, S. J. Tree species influence on microbial communities in litter and soil: current knowledge and research needs. *For. Ecol. Manag.* **309**, 19-27 (2013).

34    Baldrian, P. *et al.* Active and total microbial communities in forest soil are largely different and highly stratified during decomposition. *ISME J.* **6**, 248-258, doi:10.1038/ismej.2011.95 (2012).

35    Větrovský, T., Steffen, K. T. & Baldrian, P. Potential of cometabolic transformation of polysaccharides and lignin in lignocellulose by soil Actinobacteria. *PLoS One* **9** (2014).

36    Das, M., Royer, T. V. & Leff, L. G. Diversity of Fungi, Bacteria, and Actinomycetes on leaves decomposing in a stream. *Appl. Environ. Microbiol.* **73**, 756-767 (2007).

37    Jin, D. *et al.* Legumes in Chinese natural grasslands: species, biomass, and distribution. *Rangeland. Ecol. Manag.* **66**, 648-656 (2013).

38    Zomer, R. J., Trabucco, A., Bossio, D. A. & Verchot, L. Climate change mitigation: a spatial analysis of global land suitability for clean development

mechanism afforestation and reforestation. *Agric. Ecosyst. Environ.* **126**, 67-80 (2008).

39    Caporaso, J. G. *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* **6**, 1621-1624, doi:10.1038/ismej.2012.8 (2012).

40    Schloss, P. D. *et al.* Introducing mothur: open-source, platform-Independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**, 7537-7541 (2009).

41    Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**, 2194-2200 (2011).

42    Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73**, 5261-5267 (2007).

43    Lynch, M. D. J. & Neufeld, J. D. Ecology and exploration of the rare biosphere. *Nat. Rev. Microbiol.* **13**, 217-229 (2015).

44    Logares, R. *et al.* Biogeography of bacterial communities exposed to progressive long-term environmental change. *ISME J.* **7**, 937-948 (2013).

45    Galand, P. E., Casamayor, E. O., Kirchman, D. L. & Lovejoy, C. Ecology of the rare microbial biosphere of the Arctic Ocean. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 22427-22432 (2009).

46    Freilich, M. A. & Connolly, S. R. Phylogenetic community structure when competition and environmental filtering determine abundances. *Global. Ecol. Biogeogr.* **24**, 1390-1400, doi:10.1111/geb.12367 (2015).

47    Clarke, K. R. Non-parametric multivariate analyses of changes in community structure. *Austral. Ecol.* **18**, 117-143 (1993).

48    Poyry, J., Luoto, M., Heikkinen, R. K., Kuussaari, M. & Saarinen, K. Species traits explain recent range shifts of Finnish butterflies. *Glob. Change Biol.* **15**, 732-743 (2009).

49    Nally, R. M. Hierarchical partitioning as an interpretative tool in multivariate inference. *Austral. Ecol.* **21**, 224-228 (1996).

## Acknowledgements

## Author contributions

W.K. conceived the study and developed the idea with M.J. K.Z., L.Y. and W.K. sampled the field soils and surveyed plants. K.Z. performed soil physiochemical analysis, DNA extraction and collected climate and plant data. M.J. conducted the data

statistical analysis. M.J. and W.K. wrote the first draft of the manuscript and all authors commented on the manuscript drafts and contributed to writing.

**Additional information**

**Supplementary Information**: The manuscript contains 2 supplementary tables and 8 supplementary figures.

**Competing interests**: The authors declare no conflict of interest.

**Data availability**

The raw sequencing reads generated during the current study have been deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive under accession numbers SRR6326306-SRR6326301.

**Tables**

Table 1 Percentage of variations explained by abiotic and biotic factors using distance-based linear modelling (DistLM).


Table 2 Correlations of all environmental factors and Nearest Taxon Index (NTI) and Beta Nearest Taxon Index (□NTI) for the rare and abundant sub-communities.

**Figure Legends**

**Fig. 1 Distance based redundancy analysis (dbRDA) performed on Bray-Curtis dissimilarity matrices for (A) rare sub-community, (B) abundant sub-community and (C) the entire community and environmental factors.** Samples from desert grassland (○) and steppe (●) were marked. Arrows indicate correlation between environmental factors and microbial structure. The shown environmental factors were chosen based on distance-based linear modelling results. MAP: mean annual precipitation; MAT: mean annual temperature; BGB: plant below-ground biomass; AGB: plant above-ground biomass; TOC: total organic carbon.

**Fig. 2 NTI of the rare and abundant sub-communities.** Steppe samples were coloured red, while the desert grassland samples were coloured black. The dashed lines indicate the thresholds of deterministic (|NTI|>2) or stochastic (-2<NTI<2) dominated assembly processes. NTIs of the rare sub-community were mostly between -2 and +2, while that of the abundant sub-community were predominately >2.

**Fig. 3 Independent and joint contributions of biotic and abiotic factors in relation to the NTI variations observed according to a hierarchical partitioning analysis.** A: Rare sub-community; B: abundant sub-communities. Significant (*P*<0.05) independent effects are indicated by asterisks (*). MAP: mean annual precipitation; MAT: mean annual temperature; BGB: plant below-ground biomass; AGB: plant above-ground biomass; TOC: total organic carbon.

**Fig. 4 βNTI distributions of the rare and abundant sub-communities.** The dashed lines indicate the thresholds of deterministic (|NTI|>2) or stochastic (-2<NTI<2) dominated assembly processes. βNTIs of the rare sub-community were mostly between -2 and +2, while that of the abundant sub-community were predominately >2.

Table 1 Percentage of variations explained by abiotic and biotic factors using distance-based linear modelling (DistLM).

|  | Rare sub-community | Abundant sub-community |
|---|---|---|
| Total variations explained | 17.7% | 48.0% |
| Climate |  |  |
| MAP | 3.6% | 4.3% |
| MAT | 2.3% | n.d. |
| Aridity | n.d. | n.d. |
| Spatial distance | 2.4% | 6.1% |
| Soil |  |  |
| pH | 2.1% | 3.7% |
| TOC | 2.5% | 9.4% |
| $NH_4^+$ | n.d. | 3.1% |
| $NO_3^-$ | n.d. | n.d. |
| Plants |  |  |
| AGB | 2.4% | 15.6% |
| PSD | 2.4% | 5.8% |

Summary of distance-based linear modelling testing the correlations between community Bray-Curtis dissimilarity matrix and biotic and abiotic factors: mean annual temperature (MAT), mean annual precipitation (MAP), aridity, pH, total organic carbon (TOC), $NH_4^+$, $NO_3^-$, above-ground biomass (AGB) and plant Shannon diversity (PSD). n.d., the proportion of variations explained was not significant.

Table 2 Correlations of all environmental factors and Nearest Taxon Index (NTI) and Beta Nearest Taxon Index (βNTI) for the rare and abundant sub-communities.

| | NTI | | | | βNTI | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | Rare | | Abundant | | Rare | | Abundant | |
| | *P*-value | r | *P*-value | r | *P*-value | r | *P*-value | r |
| Climate | n.a. | n.a. | n.a. | n.a. | 0.08 | -0.10 | **0.03** | **0.12** |
| MAT | 0.41 | -0.17 | 0.68 | 0.07 | 0.36 | 0.03 | 0.31 | -0.05 |
| MAP | 0.21 | 0.23 | 0.87 | -0.03 | **<0.01** | **-0.14** | **<0.01** | **0.12** |
| Aridity | 0.50 | -0.10 | 0.94 | 0.01 | 0.15 | -0.08 | **<0.01** | **0.25** |
| Soil | n.a. | n.a. | n.a. | n.a. | 0.14 | -0.10 | **<0.01** | **-0.29** |
| pH | 0.44 | 0.12 | 0.88 | -0.03 | 0.47 | 0.01 | 0.26 | -0.05 |
| Moisture | 0.82 | 0.04 | 0.99 | -0.01 | 0.05 | -0.10 | 0.10 | -0.08 |
| TOC | 0.63 | -0.08 | 0.12 | -0.25 | 0.09 | -0.08 | **0.01** | **-0.18** |
| $NH_4^+$ | 0.22 | -0.20 | 0.58 | -0.09 | 0.06 | -0.13 | **0.04** | **-0.16** |
| $NO_3^-$ | 0.35 | -0.15 | 0.20 | -0.21 | 0.14 | 0.09 | 0.10 | -0.11 |
| Plants | n.a. | n.a. | n.a. | n.a. | 0.18 | -0.06 | **0.04** | **-0.11** |
| Above ground biomass | 0.28 | 0.18 | 0.39 | 0.14 | 0.06 | -0.11 | **0.01** | **-0.17** |
| Plant Shannon diversity | 0.50 | 0.07 | **<0.01** | **-0.43** | 0.13 | -0.08 | 0.23 | -0.05 |
| Root biomass | 0.34 | -0.15 | 0.52 | 0.11 | 0.25 | 0.04 | 0.43 | 0.01 |
| Geospatial | 0.14 | 0.24 | 0.86 | -0.03 | **0.02** | **-0.10** | **0.03** | **0.09** |

The Pearson correlations between the NTI and environmental factors were calculated using partial correlation analysis, while the correlations for the βNTI were calculated using partial Mantel analysis. For each tested environmental factor and category, the controlling matrix was defined to contain all other factors excluding the ones been tested. Significant correlations are in bold

Figure 1 Distance based Redundancy analysis (dbRDA) performed on Bray-Curtis dissimilarity matrices for (A) rare sub-community, (B) abundant sub-community and (C) the entire community and environmental factors. Samples from desert grassland (○) and steppe (●) were marked. Arrows indicate correlation between environmental factors and microbial structure. The shown environmental factors were chosen based on distance-based linear modelling results. MAP: mean annual precipitation; MAT: mean annual temperature; BGB: plant below-ground biomass; AGB: plant above-ground biomass; TOC: total organic carbon.
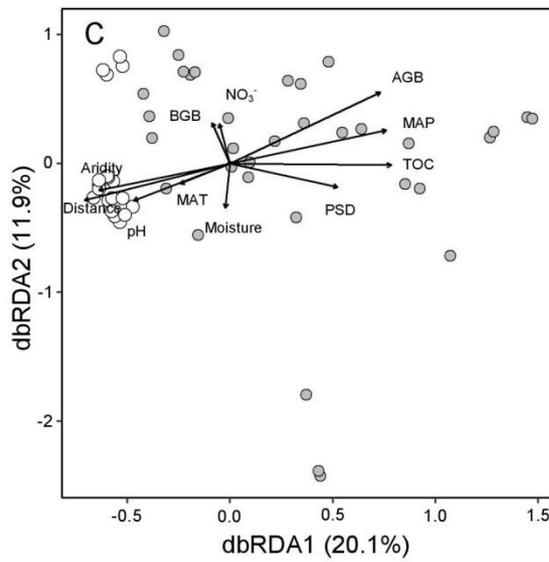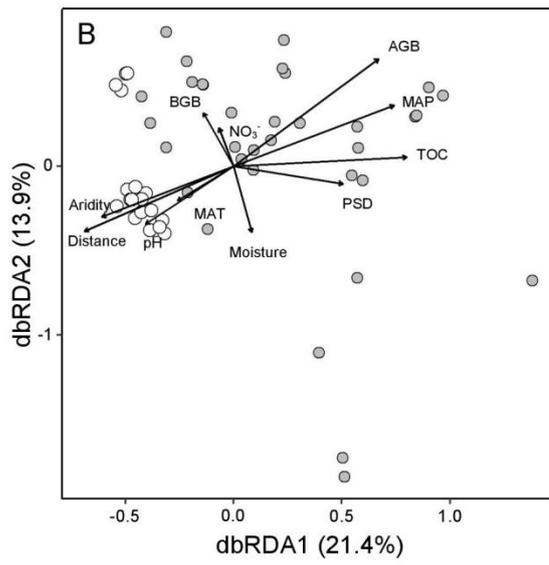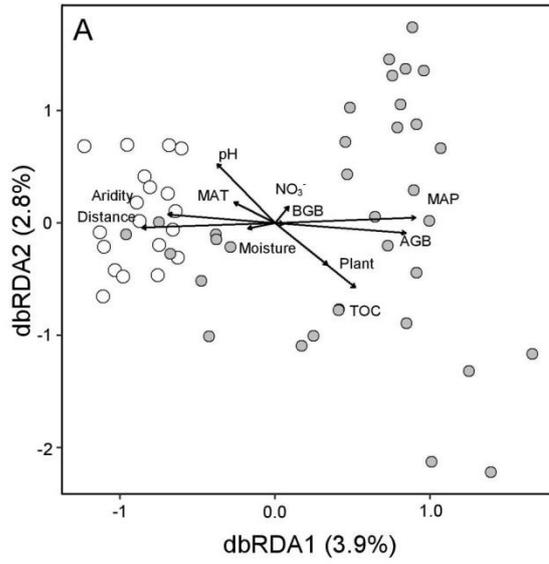
Figure 2 NTI of the rare and abundant sub-communities. Steppe samples were coloured red, while the desert grassland samples were coloured black. The dashed lines indicate the thresholds of deterministic (|NTI|>2) or stochastic (-2<NTI<2) dominated assembly processes. NTIs of the rare sub-community were mostly between -2 and +2, while that of the abundant sub-community were predominately >2.
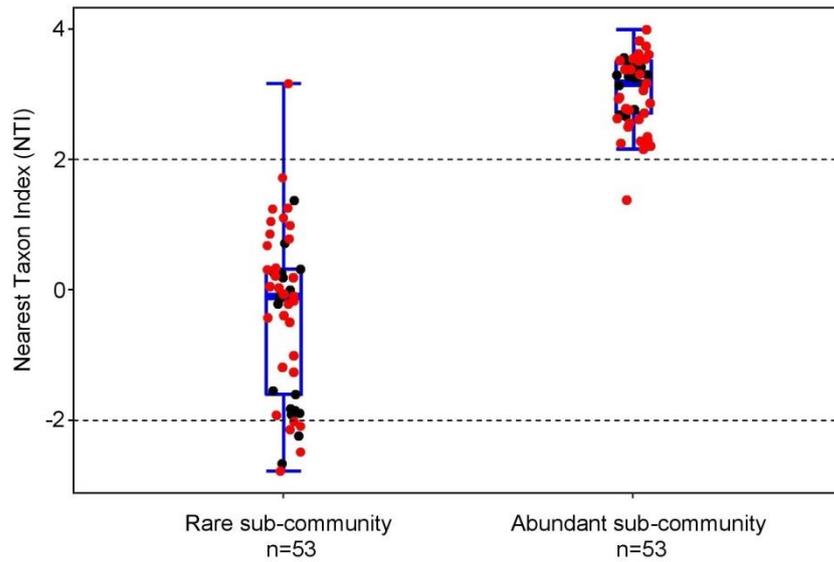
Figure 3 Independent and joint contributions of biotic and abiotic factors in relation to the NTI variations observed according to a hierarchical partitioning analysis. A: Rare sub-community; B: abundant sub-communities. Significant (*P*<0.05) independent effects are indicated by asterisks (*).MAP: mean annual precipitation; MAT: mean annual temperature; BGB: plant below-ground biomass; AGB: plant above-ground biomass; TOC: total organic carbon.
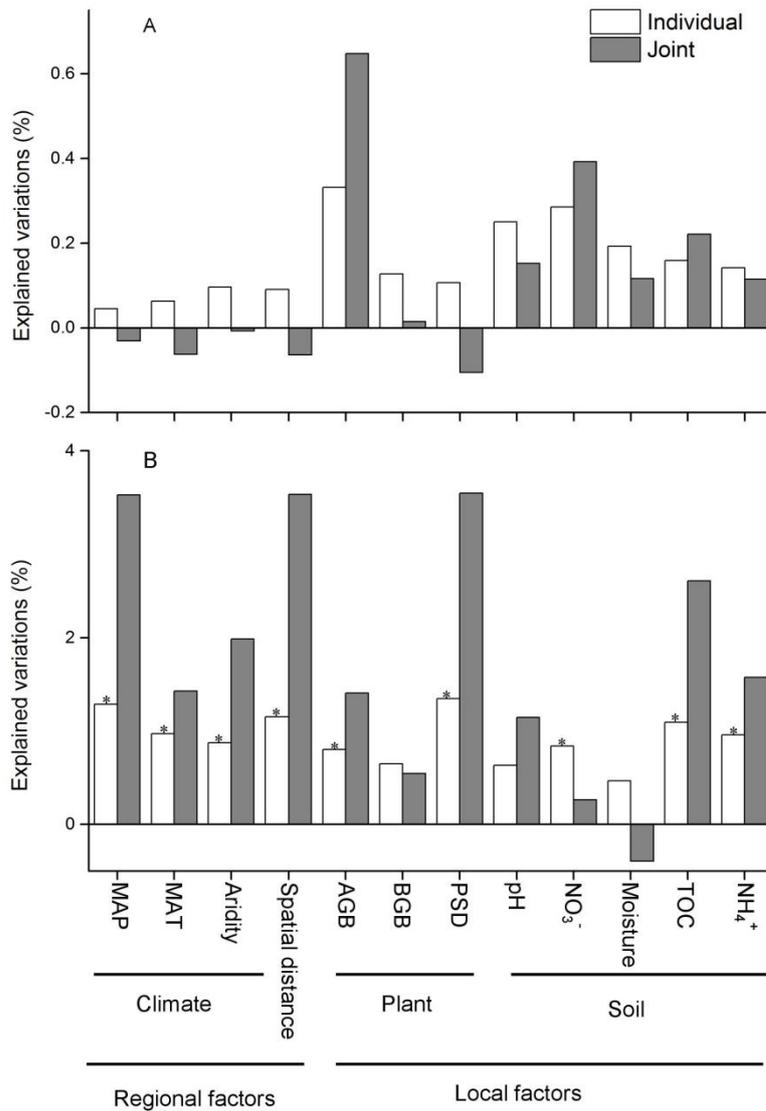
Figure 4 βNTI distributions of the rare and abundant sub-communities. The dashed lines indicate the thresholds of deterministic ($|NTI|>2$) or stochastic ($-2<NTI<2$) dominated assembly processes. βNTIs of the rare sub-community were mostly between -2 and +2, while that of the abundant sub-community were predominately >2.