

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

**Mukan Ji**

Institute of Tibetan Plateau Research, Chinese Academy of Sciences

**Weidong Kong** (✉ [wdkong@itpcas.ac.cn](mailto:wdkong@itpcas.ac.cn))

Institute of Tibetan Plateau Research Chinese Academy of Sciences <https://orcid.org/0000-0001-9682-1484>

**James Stegen**

Pacific Northwest National Laboratory

**Linyan Yue**

Institute of Tibetan Plateau Research Chinese Academy of Sciences

**Fei Wang**

Institute of Tibetan Plateau Research Chinese Academy of Sciences

**Xiaobin Dong**

Beijing Normal University

**Don A Cowan**

Department of Genetics, University of Pretoria

**Belinda C Ferrari**

UNSW Sydney

---

## Research

**Keywords:** Soil bacteria, Rare biosphere, Community structure, Assembly processes

**Posted Date:** January 2nd, 2020

**DOI:** <https://doi.org/10.21203/rs.2.19893/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at Environmental Microbiology on April 2nd, 2020. See the published version at <https://doi.org/10.1111/1462-2920.14993>.

# Abstract

**Background:** Bacteria in low abundance represent the majority of Earth's biodiversity and perform vital ecological functions, but little is known about their biogeography nor the ecological processes that drive their community assembly in terrestrial ecosystems. Here, we investigated the community compositions and phylogenies of rare (relative abundance < 0.1%) and abundant (> 1%) soil bacteria along a transect containing three alpine grassland types (meadow, steppe, and desert) on the Tibetan Plateau.

**Results:** Our results revealed similar biogeographical patterns of rare and abundant bacteria, with their community compositions and phylogenies shifting gradually along the transect. The similar patterns, however, were driven by contrasting community assembly processes, with rare subcommunity being more heavily influenced by stochasticity (72%) than abundant (57%). The composition of abundant subcommunity (80%) was better explained by local (including soil and vegetation factors), geospatial and climatic factors than that of rare subcommunity (41%), while the phylogeny of the rare one (36%) was better explained than that of the abundant one (29%). Variation partitioning analysis indicated that pure local factors consistently explained a higher proportion of the community composition than geospatial factors in both rare (12.3% and 8.7%, respectively) and abundant bacteria (18.3% and 14.1%, respectively). In contrast, the phylogeny of rare subcommunity was explained by local and geospatial factors equally (11.5% and 11.9%, respectively), while that of abundant subcommunity was more explained by geospatial (22.1%) than local factors (11.3%). Furthermore, our results revealed a tighter connection between the community phylogeny and composition in rare than in abundant bacteria.

**Conclusions:** Our results revealed consistent biogeographical patterns of rare and abundant bacteria in grassland soils, but their assembly processes were distinct. We further demonstrated that rare subcommunity was less predictable than the abundant subcommunity by environmental and geospatial factors. Rare and abundant bacteria responded differentially to factors, which was attributed to the distinct life strategies. Our study provides novel insights into the assembly processes and biographical patterns of rare and abundant bacteria in terrestrial ecosystems.

## Background

The bacteria in low abundance represent the majority of Earth's biodiversity [1], and these rare bacteria contain an enormous pool of genetic novelty [2]. Accumulating evidence has confirmed the crucial ecological functions of rare bacteria in terrestrial and aquatic ecosystems. For example, rare bacteria are responsible for nitrogen and carbon assimilation in freshwater lakes [3–5], they enhance plant defence against aphids [6], and provide insurance during ecological restoration [7–9]. Therefore, variation in these low abundance, but functional important microorganisms could cause substantial consequences on an ecosystem.

Compared to rare bacteria, abundant bacteria present at much higher relative abundance globally, and potentially exhibit greater tolerance to environmental stresses. Thus, rare bacteria have been identified to

drive bacterial community structure changes both spatially and temporally [10–12]. Several studies have examined the biogeography of rare and abundant bacteria in aquatic ecosystems, such as lake reservoirs [13], subtropical bays [14], the Arctic ocean [15], and the Pacific ocean [16]. These studies have consistently reported that the community compositions of rare and abundant bacteria were influenced by geographical and environmental factors differentially, despite exhibiting similar biogeographical patterns [13, 17]. Specifically, the community composition (typically measured using Bray-Curtis dissimilarity) of rare bacteria was more strongly influenced by local environmental factors, while that of abundant bacteria was predominately explained by geospatial factors [13, 14]. Compared to aquatic ecosystems, the biogeographical patterns of rare and abundant bacteria in terrestrial ecosystems are less understood. Furthermore, as soils are non-fluidic and are much more heterogeneous compared to the aquatic environment [18], the biogeographical pattern of rare bacteria in soil could be different from that of abundant bacteria.

Composition-based metrics have traditionally been used to examine microbial community structure variations, but phylogeny-based community metrics, such as the  $\beta$ -mean nearest taxon distance ( $\beta$ -MNTD) and UniFrac distance, are receiving increased attention [19, 20]. Compared to composition-based community metrics, these phylogeny-based community metrics can be used to infer changes in microbial community function [21–23]. This is based on the assumption that phylogenetically close-related microorganisms have similar habitat associations [20]. Thus, using phylogeny-based community metrics, together with the composition-based community metrics, could provide additional insights into the ecogeographical patterns of rare and abundant subcommunities.

Due to the strong environmental stress tolerance of abundant bacteria and the greater niche differentiation in terrestrial ecosystems, we hypothesize that rare bacteria could exhibit distinct biogeographical pattern compared to abundant bacteria in soils. Furthermore, the relative importance of local environmental factors and geospatial factors on the biogeography of rare and abundant bacteria could be different from that identified in aquatic ecosystems. To test these hypotheses, the bacterial community structure was investigated using DNA-based amplicon sequencing targeting the 16S rRNA gene, and both the community Bray-Curtis compositional dissimilarity and  $\beta$ MNTD phylogenetic distance were calculated. These metrics were used to compare the community compositions and phylogenies of rare and abundant bacteria along a 1200 km grassland transect on the Tibetan Plateau.

## Results

### Taxonomic compositions of rare and abundant bacteria

A total of 7,890 OTUs were identified from the retained 806,358 high-quality sequences at 97% sequence identity. Rare bacteria (relative abundance < 0.1%) comprised 87.7% of the total bacterial richness (6,923 OTUs), but their total relative abundance accounted for only 10.1% of the entire community. In contrast, abundant bacteria (relative abundance > 1%) comprised only 1% of the total richness (114 OTUs), but their relative abundance was 59.4%. Of the retained sequences, 91.3% were classified at the phylum level,

and 34, 13 and 34 bacterial phyla were identified from the rare, abundant and entire communities, respectively (Fig. 1A). Across all samples, rare subcommunity was dominated by Proteobacteria (24.8%), followed by unclassified bacteria (18.1%), Actinobacteria (15.4%), Bacteroidetes (14.9%), and Planctomycetes (7.7%). In contrast, the abundant and entire communities were both dominated by Actinobacteria (57.5% and 47.2%, respectively), and exhibited lower relative abundance of Proteobacteria (12.2% and 18.2%, respectively) than the rare subcommunity.

## Community compositions of rare and abundant bacteria

Both rare and abundant OTUs exhibited positive correlations between their relative abundance and the number of samples they were identified in (Pearson  $r = 0.84$  and  $0.36$ , respectively, both  $P < 0.001$ , Fig. 1B). The abundant OTUs were more widespread geographically than rare OTUs, with over 90% of the abundant OTUs being detected in  $> 50\%$  of samples and 54% of the abundant OTUs being detected from  $> 90\%$  of the samples. In contrast, the majority (97%) of rare OTUs were only detected in  $< 50\%$  of samples analysed.

The community compositions of rare (Fig. 2A), abundant (Fig. 2B) and entire communities (Fig. 2C) were significantly different among the three grassland types (PERMANOVA, all  $P < 0.001$ ). The community compositions of rare, abundant, and entire communities also exhibited gradual shifts along the grassland transect (from meadow to steppe, and then to desert grassland). The measured local, climatic and geospatial factors together explained 41%, 80%, and 60% of the community compositions for rare, abundant, and entire communities, respectively (Figs. 3A, 3B and, 3C, Additional file 1: Table S1). VPA analysis revealed that pure local factors consistently explained higher community composition than pure geospatial factors for both rare and abundant subcommunities, while pure climatic factors explained the lowest community composition (Figs. 3A, 3B and, 3C). Furthermore, soil pH, TOC, and  $\text{NH}_4^+$  explained a higher proportion of the community variation in abundant over rare subcommunities (Additional file 1: Table S1).

Mantel tests revealed significant linear correlations of the community composition with local, climatic and geospatial factors for the rare, abundant and entire communities (Figs. 3D, 3E, and 3F, Supplementary Fig. 2, and Additional file 1: Table S2). However, partial Mantel tests indicated that the correlations between climatic factors and the community compositional dissimilarity were non-significant. Furthermore, local factors consistently exhibited stronger correlation strength (indicated by Pearson  $r$ ) than geospatial factors.

## Community phylogenies of rare and abundant bacteria

The community phylogenies of rare (Fig. 4A), abundant (Fig. 4B) and entire bacterial communities (Fig. 4C) differed significantly among the three grassland types (PERMANOVA,  $P < 0.001$ ,  $= 0.003$ , and  $< 0.001$ , respectively). DistLM analysis revealed that the measured local, climatic and geospatial factors together explained 36%, 29%, and 44% of the community phylogeny for rare, abundant and entire communities, respectively (Figs. 5A, 5B, 5C, and Additional file 1: Table S3). VPA analysis indicated that

the pure local and geospatial factors contributed equally to the community phylogeny of rare subcommunity (Fig. 5A, 11.5% and 11.9%, respectively). In contrast, the geospatial factors explained a larger proportion of community phylogeny (22.1%) than local factors (11.3%) for abundant subcommunities (Fig. 5B). Consistent with that identified at the community compositional level, climatic factors explained the lowest community phylogeny in rare, abundant, and entire communities (Figs. 5A, 5B, and 5C).

Mantel tests revealed significant correlations of the community phylogeny with the measured local, climatic and geospatial factors in rare subcommunity (all  $P < 0.001$ , Fig. 5D, Additional file 1: Table S4, and Additional file 2: Fig. S3), but these correlations were not significant for abundant or entire communities (Figs. 5E and 5F). Partial Mantel tests further revealed that the community phylogeny of rare bacteria significantly correlated with the local and geospatial factors ( $P < 0.001$  and  $0.002$ , respectively), but not with climatic factors ( $P = 0.133$ ). Furthermore, a significant partial Mantel correlation was detected between the geospatial factors and the community phylogeny of abundant bacteria ( $P = 0.007$ ).

## **Correlation relationships between the community composition and phylogeny**

Significant correlations between community composition and phylogeny were observed in rare, abundant, and entire communities (Fig. 6). The correlations in rare and entire communities were much stronger than that in abundant subcommunity ( $R^2 = 0.65, 0.62$  and  $0.08$ , respectively, all  $P < 0.001$ ). In each case, the observed increase in the community composition was accompanied by a relatively weak change in the community phylogeny until the community compositional dissimilarity further increased to approximately 60%, 40%, and 40% for rare, abundant, and entire communities, respectively. After passing these thresholds, exponential increases in the community phylogenetic distance were then occurred. This observed delay in the community phylogenetic variation was smaller for rare subcommunity than that in abundant subcommunity.

## **Community assembly processes of rare and abundant bacteria**

The ecological processes shaping bacterial community assembly were explored using compositional (modified stochasticity ratio, MST) and phylogenetic ( $\beta$ -Mean Nearest Taxon Index,  $\beta$ NTI) approaches. The MST of rare subcommunity was  $72 \pm 18\%$ , which was higher than that in abundant ( $57 \pm 20\%$ ) and entire communities ( $66 \pm 22\%$ ; Additional file 2: Fig. S4). Furthermore, the within ecosystem community variations were less influenced by stochasticity than those between ecosystems for rare and entire communities.

The  $\beta$ NTI of rare subcommunity varied from  $-8.9$  to  $8.1$ , with 66%, 30%, and 4% of which being  $< -2$ , between  $-2$  and  $2$ , and  $> 2$ , respectively. In contrast, the  $\beta$ NTI of abundant and entire communities varied

from - 8.8 to -4.5 and - 16.5 to 1.7, with 100% and 99% of  $\beta$ NITs being < -2, respectively (Additional file 2: Fig. S5). Significant positive linear correlations of  $\beta$ NIT values with local, climatic and geospatial factors were detected in rare and entire community, whereas the  $\beta$ NIT values of the abundant subcommunity only significantly correlated to local factors (Additional file 2: Fig. S6, Additional file 1: Table S5). Furthermore, partial Mantel tests indicated that the  $\beta$ NIT values of rare ( $P = 0.002$ ), abundant ( $P < 0.001$ ), and entire communities significantly correlated only with local factors ( $P < 0.001$ ), while the correlations with climatic and geospatial factors were not significant.

## Discussion

### Rare and abundant bacteria exhibited similar biogeographical patterns, but were shaped by distinct assembly processes

Rare and abundant bacteria both exhibited similar community compositional shifts along the meadow, steppe, and desert grassland transect (Figs. 2 and 4). The similar biogeographical patterns of rare and abundant subcommunities have also been consistently reported from freshwater lakes [13, 24] and marine environments [13, 14, 16, 24]. Thus, this consistency suggests that the similar biogeographical patterns of rare and abundant bacteria are ubiquitous to both terrestrial and aquatic ecosystems.

The stronger influence of stochasticity on rare subcommunity (Additional file 2: Fig. S4) was consistent with its lower proportion of the community composition explained by environmental factors compared with abundant subcommunity (Figs. 3A and 3B). The varied influence of stochasticity on rare and abundant bacteria may be related to their different life strategies. The abundant subcommunity was comprised of more copiotrophic than oligotrophic bacteria. For example, copiotrophic *Actinobacteria* were presented at higher relative abundances in abundant than rare subcommunities, while oligotrophic bacteria such as *Acidobacteria* and *Planctomycetes* were more abundant in rare subcommunity (Fig. 1A) [25]. Copiotrophic bacteria (such as *Actinobacteria*) are more sensitive to carbon availability than oligotrophic bacteria [26, 27], which is proposed to be due to a lack of carbon and energy regulatory system in copiotrophs [28]. The higher sensitivity of abundant bacteria to soil nutrients was further evidenced by the stronger contribution of soil nutrients (TOC and  $\text{NH}_4^+$ ) in explaining the community composition in abundant than rare subcommunities (Additional file 1: Table S1).

Rare subcommunity exhibited a stronger correlation between the community composition and geospatial factors than abundant bacteria (Fig. 3D and 3E), indicating a stronger influence dispersal limitation [13]. Dispersal limitation is one of the most important stochastic factors influencing bacterial community assembly, and is dependent on cell numbers [29]. Thus, the low relative abundance at less than < 0.1% could be the major limiting factor for rare bacteria being dispersed across large geographic distance [9, 13, 14]. The stronger influence of dispersal limitation on rare over abundant bacteria at the compositional level has also been reported in soil fungal community [30].

Stochasticity has been reported to play less important role than determinism in shaping bacterial community composition in grassland and polar desert soils [31, 32]. Our results further expand this phenomenon, and show that the influence of stochasticity on community assembly could be lineage-specific. The bacterial lineages that present at low abundance, yet potentially perform vital ecological functions [9], could be more strongly influenced by stochasticity than abundant bacteria. Despite the dominance of determinism in shaping the community composition of abundant bacteria, substantial contributions from stochasticity were also observed (Additional file 2: Fig. S4), consistent with previous studies [14, 33].

Despite the similar biogeographical patterns, the community compositions of rare and abundant subcommunities were predominately influenced by stochasticity and determinism, respectively. It is interesting that how two distinct assembly processes lead to consistent biogeographical patterns in rare and abundant subcommunities. We propose that this may be caused by the co-variation of geospatial and environmental heterogeneity. Our results showed that the geospatial and local environmental variations were highly correlated (Additional file 2: Fig. S7). Thus the co-variation of stochastic (geospatial) and deterministic (local environmental) processes could lead to the similar biogeographical patterns of rare and abundant subcommunities observed.

### **Rare bacteria exhibited a greater community phylogenetic variation than abundant bacteria**

Phylogenetically close-related microorganisms have similar habitat associations, thus phylogeny-based community metrics could infer potential community functional change [20]. The community phylogeny of rare bacteria was better separated among the three grassland ecosystems than that of abundant bacteria (Fig. 4A and 4B), potentially suggesting a stronger divergence of ecological functions [34]. Partial Mantel results demonstrated that the community phylogenies of rare and abundant subcommunities both exhibited significant correlations with geospatial factors (Figs. 5D and 5E, Additional file 1: Table S4). This contrasted to the lack of correlation between bacterial community phylogeny and geographical distance reported in Wang *et al.* [35]. The difference could be attributed to the geographical scale, as the sampling distance was less than 10 km in Wang *et al.* (2013), which could be too small for substantial phylogenetic variation to be detected.

The community phylogeny of abundant bacteria was more strongly influenced by dispersal limitation than that of rare bacteria (Figs. 5A). This finding contradicts the observations at the community compositional level (Figs. 3A and 3B), but is consistent with the previous phylogeny-based investigation at the northwest Pacific Ocean [36]. This suggests the factors influence the community composition and phylogeny could be different. The stronger influence of geospatial than environmental factors in shaping the community phylogeny of abundant bacteria could be explained by their environmental adaptability. Abundant bacteria have been reported to exhibit strong tolerance to alkaline and saline conditions [37, 38]. This is also evidenced by the wide distribution of abundant bacteria across the transect in the present study (Fig. 1B) and the dominance of homogenous selection in shaping their community phylogeny (Additional file 2: Fig. S5). Thus, dispersal limitation could be the main factor constraining the

closeness of community phylogeny in abundant subcommunity across the transect (Additional file 1: Table S4).

The equal importance of environmental and geospatial factors in shaping the community phylogeny of rare subcommunity can be attributed to their life strategies [37]. Rare bacteria generally exhibit lower environmental adaptability and smaller niche breadth [39] than abundant bacteria, which hinder their colonization in new habitats after being dispersed. This resulted in their community phylogeny being gradual shifted across environmental gradient (Fig. 4A) and the strong correlation between community phylogeny and environmental heterogeneity (Additional file 2: Fig. S3), which was not detected in the abundant subcommunity. The critical contribution of environmental filtering on the community phylogeny of rare bacteria is consistent with that observed in the marine environment [36]. However, unlike aquatic environment, the non-fluidic soil [18] further constrain the chance of dispersal, thus geospatial factors exhibited greater influence on the community phylogeny of rare subcommunity in soil than in marine ecosystems [14, 36].

### **Decoupled community compositional and phylogenetic variations in abundant bacteria**

A tighter connection (Fig. 6, larger  $R^2$  value) between the community composition and phylogeny was observed in rare than abundant subcommunities. This is mainly due to the phylogeny of abundant subcommunity being less sensitive to environmental changes (Figs. 4 and 5). The ability to maintain a community's phylogeny could reflect its ecological niche preservation capacity under changed environmental conditions [20]. Thus, the decoupling between community composition and phylogeny indicates that abundant bacteria could be better in preserving ecological niches than rare bacteria. This could explain the easier loss of narrow niche functions (such as the degradation of toxic compounds, which is typically performed by rare bacteria) than broad niche functions (such as the degradation of organic compounds in general, which are performed by all bacteria) [8, 40].

## **Conclusions**

Our study provides novel insights to explain the bacterial distribution patterns of rare and abundant bacteria in alpine grassland soils. The similar biogeographical patterns of rare and abundant subcommunities were consistent with that observed in aquatic ecosystems, highlighting their global consistency. Our results also revealed that the similar biogeographical patterns of rare and abundant subcommunities could be predominately influenced by stochastic and deterministic processes, respectively. We propose that this could be due to the co-variation of stochastic (geospatial) and deterministic (environmental) factors. This shall be further investigated to disentangle the contributions of stochasticity and determinism in other ecosystems with inconsistent changing patterns of geospatial and environmental factors.

## **Materials And Methods**

## Site description and soil sampling

The study area is located on the Tibetan Plateau, between 31 and 33°N latitude and 79 and 93°E longitude, with an average altitude above 4400 m a.s.l. Soil samples were collected along a 1200 km transect (Additional file 2: Fig. S1) in July 2015. The 1200 km transect contained ecosystems classified as alpine desert grassland (4 sites), alpine steppe (11 sites), and alpine meadow (4 sites) from west to east, thereby forming a natural gradient with broad environmental conditions. At each sampling site, four 1 m × 1 m quadrats were randomly selected. In each quadrat, approximately 200 g of bulk soils (0–10 cm) were collected after plant material being removed by sieving (2.0 mm), and were transported on ice to laboratories at Beijing, then stored at -80 °C until processed. The plant Shannon diversity, plant above-ground biomass, and plant below-ground biomass were calculated as described previously [41].

## Soil physiochemical analysis

Soil pH was measured in a 1:5 soil-to-water suspension using a pH meter (Sartorius, Goettingen, Germany). Soil nitrate ( $\text{NO}_3^-$ ) and ammonium ( $\text{NH}_4^+$ ) were extracted with 2 M KCl (1:5) and their concentrations were determined using a Smartchem200 Discrete Auto Analyzer (Alliance, Frépillon, France). Total organic carbon (TOC) was measured in the solid-state using a TOC analyzer (Shimadzu, Kyoto, Japan). The aridity index (potential evaporation/precipitation) of each site was obtained from the Global Aridity Index dataset ([www.cgiar-csi.org](http://www.cgiar-csi.org)) [42]. The mean annual precipitation (MAP) and mean annual temperature (MAT) were predicted using the meteorological data between 2003 and 2012 from 33 climatic stations (China meteorological Data Sharing Service System; <http://cdc.cma.gov.cn/>) using the Kriging interpolation.

## DNA extraction, PCR and high-throughput 16S rRNA gene sequencing

Total DNA was extracted using the MO BIO Power Soil DNA extraction kit (Mo Bio Laboratories, Carlsbad, CA, USA) according to manufacturer's instructions. Universal primers 515F (5'-GTGCCAGCMGCCGCGGTAA-3') and 909r (5'-CCCCGYCAATTCMTTTRAGT-3') with 12 nt unique barcodes were used to amplify the V4 and V5 hyper-variable regions of the 16S rRNA gene [43]. The PCR mixture (25  $\mu$ l) contained 1x PCR buffer, 1.5 mM of  $\text{MgCl}_2$ , 0.4  $\mu$ M each of deoxynucleoside triphosphate bases, 1.0  $\mu$ M of each primer, 0.5 U of Ex Taq (TaKaRa, Dalian, China) and 20 ng of DNA template. The PCR amplification program included an 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. PCR products were purified using a OMEGA Gel Extraction Kit (Omega Bio-Tek, Georgia, USA) following electrophoresis. PCR products from different samples were pooled in equimolar amounts and then used for pair-end sequencing (2 × 250 bp) on the Illumina MiSeq platform at the Chengdu Institute of Biology, Chinese Academy of Sciences.

## Sequence data processing

Three samples were not sequenced successfully, generating zero reads. These samples were not re-sequenced but removed from downstream analysis to avoid artificial community structure differences caused by batch-to-batch sequencing variation. Raw sequence data of the remaining 73 samples were processed using the MOTHUR pipeline (v. 1.34.3) [44]. Paired-end reads were merged and quality screened by removing any reads with > 1 mismatch in the primer region, average quality < 30, ambiguous bases > 0, and homopolymer length > 9 from downstream analysis. The remaining sequences were aligned against the Silva reference alignment (release 128), which was trimmed to include only the same region amplified, and those sequences that did not align were removed. Chimeric sequences were identified using UCHIME [45] and removed. The remaining sequences were classified using the Bayesian classifier against the Silva database (release 128), with a minimum confidence score of 80% [46]. All archaeal and unknown sequences were removed using the `remove.lineage` command. The archaeal sequences were removed due to a recent report showing that the currently used prokaryotic primers may have poor coverage on the archaeal domain [47]. Finally, sequences were classified into operational taxonomic units (OTUs) at 97% identity and singletons were removed. To correct for possible uneven sequencing coverage between samples, samples were randomly sub-sampled without replacement to an equal depth of 11,046, which was the smallest sample size in the dataset. This ended with a total of 806,358 high-quality bacterial sequences, with an average sequence length of 329 bp. Bacterial richness was calculated using the `summary.single` command in Mothur [44].

## Defining rare and abundant OTUs

Various arbitrary thresholds have been used to define rare and abundant OTUs in literatures. For example, abundant OTUs have been defined as OTUs with relative abundance > 1% or 0.1%, whereas rare OTUs have been defined as OTUs with relative abundance < 0.1%, 0.001% or 0.006% [13, 24, 48–50]. Here, rare OTUs were defined as the OTUs with relative abundance < 0.1% in all samples analysed, whereas abundant OTUs were defined as the OTUs with a relative abundance > 1% in one or more samples. The latter contained both OTUs that were abundant ubiquitously and occasionally across the transect (Shade et al., 2014).

## Quantifying community structure variation and assembly process

The community structure variations of the rare, abundant and entire communities were calculated separately. Bray-Curtis dissimilarity matrices were calculated using the square-root transformed community abundance matrices, and were used to indicate community compositional dissimilarity between samples. Weighted  $\beta$ -MNTD distance matrices were calculated to indicate community phylogenetic distance between samples. The Bray-Curtis dissimilarity was calculated using Primer V6 [51], whereas the  $\beta$ -MNTD distance was calculated using the function '`comdistnt`' in the package '`Picante`' in R (<http://www.r-project.org>), with the `exclude.conspecific` parameter being set to TRUE.

Bray-Curtis dissimilarity-based modified stochasticity ratio (MST) index was calculated using the '`NST`' package in R (<http://www.r-project.org>) to represent the contribution of stochasticity to community

compositional assembly [52]. The MST index ranges from 0–100%, a 0% indicates zero contribution of stochasticity, whereas 100% indicates the community assembly being completely stochasticity-driven. We also calculated the  $\beta$ -Nearest Taxon index ( $\beta$ NTI) to quantify the relative contributions of stochastic and deterministic processes in shaping community phylogenetic assembly [20] using the script provided in Swenson [53].  $\beta$ NTI values  $< -2$  indicate the between-sample phylogenetic distance being significantly lower than expected by chance (homogeneous selection);  $\beta$ NTI values  $> 2$  indicate the phylogenetic distance being significantly higher than expected (variable selection) [23, 54].  $\beta$ NTI values between  $-2$  and  $2$  indicate that the observed phylogenetic distance does not significantly deviate from the null model and the assembly process is not the result of deterministic selection, but due to stochastic processes.

## Transformation and normalization of environmental and geospatial factors

Measured environmental factors were divided into locally measured (local) and climatic factors. The local factors included soil pH, soil total organic carbon (TOC), ammonium ( $\text{NH}_4^+$ ) and nitrate ( $\text{NO}_3^-$ ) concentrations, soil moisture, plant Shannon diversity, plant above-ground biomass, and plant below-ground biomass. The climatic factors included mean annual temperature (MAT), mean annual precipitation (MAP), and aridity index. The geographical distances between samples were calculated from GPS coordinates. This geographical distance matrix was converted into a set of geospatial factors using the principal coordinates of neighbour matrices analysis (PCNM) [14, 55].

The environmental factors (both local and climatic) and the derived geospatial factors were combined, and then transformed to reduce data skewness. Specifically, TOC,  $\text{NH}_4^+$ , plant above-ground biomass, plant below-ground biomass, and aridity indices were log-transformed, whereas MAP,  $\text{NO}_3^-$ , and soil moisture were square-root transformed. The resulting environmental and geospatial factors were further normalized ( $(x-\text{mean})/\text{standard deviation}$ ) to improve data normality for multivariate statistical analysis.

## Correlation and multivariate analyses

Permutational analysis of variance (PERMANOVA) was used to test the significance of community compositional and phylogenetic differences among the three grassland types [56]. Variation partitioning analysis (VPA) based on distance-based linear model (DistLM) was used to partition the relative contributions of local, climatic and geospatial factors, as well as the contribution of each individual environmental factor (the pure effect) [56]. Then, distance-based redundancy analysis (dbRDA) ordination plots were used to visualize the associations between bacterial communities and the measured environmental/geospatial factors [56]. In addition to VPA, Mantel and partial Mantel tests were conducted to explore the correlations between bacterial community composition dissimilarity (or phylogenetic distance) and local /climatic/geospatial factors [57]. The PERMANOVA, DistLM and dbRDA ordination analyses were performed using Primer V6 [51], whereas the Mantel and partial Mantel tests were performed using the VEGAN package [55] in R (<http://www.r-project.org>).

# Declarations

## Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

## Consent for publication

Not applicable.

## Availability of data and material

The raw sequencing reads generated in the present study supporting the conclusions of this article are available in the National Center for Biotechnology Information (NCBI) Sequence Read Archive under project ID PRJNA419993, available at <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA419993>.

## Conflict interests

The authors declare no conflict of interest.

## Funding

This project was financially supported by Chinese Academy of Sciences (XDA19070304, QYZDB-SSW-DQC033 and XDA20050101), and National Natural Science Foundation of China (41771303). J.S. was supported by the US Department of Energy (DOE), Office of Biological and Environmental Research, as part of the Subsurface Biogeochemical Research Program's Scientific Focus Area at the Pacific Northwest National Laboratory (PNNL). PNNL is operated for DOE by Battelle Memorial Institute under contract DE-AC06-76RLO 1830.

## Authors' contributions

WK conceived the study and developed the idea with MJ, LY, XD, and WK and LY sampled the field soils and surveyed plants. MJ conducted the data statistical analysis. MJ and WK wrote the first draft of the manuscript, JS, DC, and BF revised the manuscript substantially. All authors read and approved the final manuscript.

## Acknowledgments

The authors greatly thank Dr. Xiagnzhen Li in the Chengdu Institute of Biology, Chinese Academy of Sciences (CAS) for thoughtful discussions. We thank Dr. Jiangshan Lai in the Institute of Botany, CAS for the statistical analysis assistance, and Dr. Yongqin Liu in the Institute of Tibetan Plateau Research, CAS for the computational platform in partial data processing.

## Corresponding authors

## References

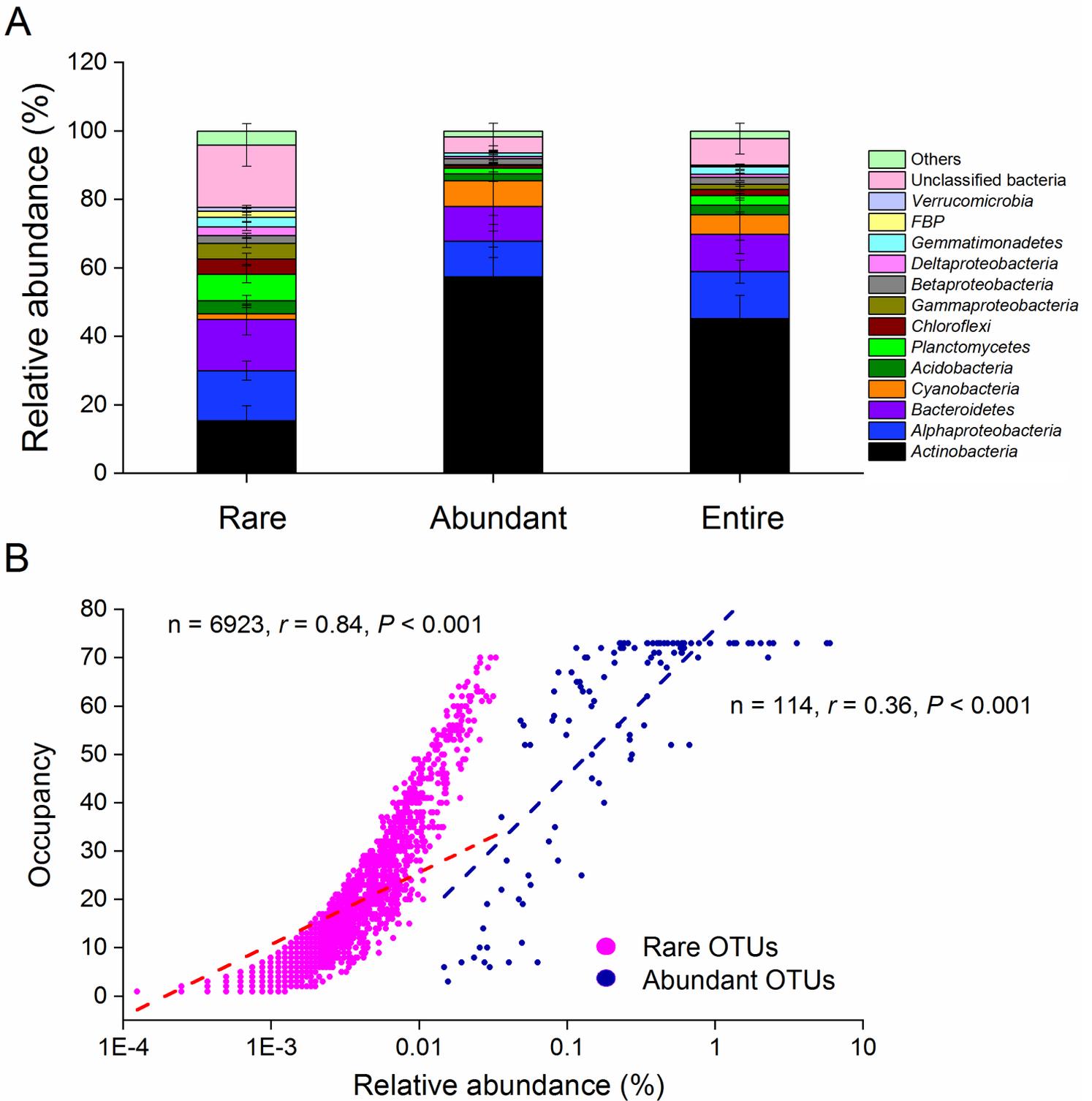
1. Naeem S, Li S. Biodiversity enhances ecosystem reliability. *Nature*. 1997;390:507-09.
2. Vigneron A, Cruaud P, Alsop E, de Rezende JR, Head IM, Tsesmetzis N. Beyond the tip of the iceberg: a new view of the diversity of sulfite- and sulfate-reducing microorganisms. *ISME J*. 2018;12:2096-99.
3. Hua Z, Han Y, Chen L, Liu J, Hu M, Li S, et al. Ecological roles of dominant and rare prokaryotes in acid mine drainage revealed by metagenomics and metatranscriptomics. *ISME J*. 2015;9:1280-94.
4. Montoya JP, Holl CM, Zehr JP, Hansen A, Villareal TA, Capone DG. High rates of N<sub>2</sub> fixation by unicellular diazotrophs in the oligotrophic Pacific Ocean. *Nature*. 2004;430:1027-31.
5. Hausmann B, Knorr K-H, Schreck K, Tringe SG, del Rio TG, Loy A, et al. Consortia of low-abundance bacteria drive sulfate reduction-dependent degradation of fermentation products in peat soil microcosms. *ISME J*. 2016;10:2365-75.
6. Hol WHG, De Boer W, Termorshuizen AJ, Meyer KM, Schneider JHM, Van Dam NM, et al. Reduction of rare soil microbes modifies plant–herbivore interactions. *Ecol Lett*. 2010;13:292-301.
7. Gibbons SM, Caporaso JG, Pirrung M, Field D, Knight R, Gilbert JA. Evidence for a persistent microbial seed bank throughout the global ocean. *Proc Natl Acad Sci U S A*. 2013;110:4651-55.
8. Mouillot D, Bellwood DR, Baraloto C, Chave J, Galzin R, Harmelinvivié M, et al. Rare species support vulnerable functions in high-diversity ecosystems. *PLOS Biol*. 2013;11:e1001569.
9. Jousset A, Bienhold C, Chatzinotas A, Gallien L, Gobet A, Kurm V, et al. Where less may be more: how the rare biosphere pulls ecosystems strings. *ISME J*. 2017;11:853-62.
10. Alonso-Saez L, Diaz-Perez L, Moran XAG. The hidden seasonality of the rare biosphere in coastal marine bacterioplankton. *Appl Environ Microbiol*. 2015;17:3766-80.
11. Gobet A, Boer SI, Huse SM, van Beusekom JEE, Quince C, Sogin ML, et al. Diversity and dynamics of rare and of resident bacterial populations in coastal sands. *ISME J*. 2012;6:542-53.
12. Hugoni M, Taib N, Debroas D, Domaizon I, Dufournel IJ, Bronner G, et al. Structure of the rare archaeal biosphere and seasonal dynamics of active ecotypes in surface coastal waters. *Proc Natl Acad Sci U S A*. 2013;110:6004-09.
13. Liu L, Yang J, Yu Z, Wilkinson DM. The biogeography of abundant and rare bacterioplankton in the lakes and reservoirs of China. *ISME J*. 2015;9:2068-77.
14. Mo Y, Zhang W, Yang J, Lin Y, Yu Z, Lin S. Biogeographic patterns of abundant and rare bacterioplankton in three subtropical bays resulting from selective and neutral processes. *ISME J*. 2018;12:2198-210.
15. Galand PE, Casamayor EO, Kirchman DL, Lovejoy C. Ecology of the rare microbial biosphere of the Arctic Ocean. *Proc Natl Acad Sci USA*. 2009;106:22427-32.

16. Wu WX, Logares R, Huang BQ, Hsieh CH. Abundant and rare picoeukaryotic sub-communities present contrasting patterns in the epipelagic waters of marginal seas in the northwestern Pacific Ocean. *Environ Microbiol.* 2017;19:287-300.
17. Galand PE, Casamayor EO, Kirchman DL, Lovejoy C. Ecology of the rare microbial biosphere of the Arctic Ocean. *Proc Natl Acad Sci U S A.* 2009;106:22427-32.
18. Zhou J, Deng Y, Zhang P, Xue K, Liang Y, Van Nostrand JD, et al. Stochasticity, succession, and environmental perturbations in a fluidic ecosystem. *Proc Natl Acad Sci U S A.* 2014;111:E836-E45.
19. Lozupone C, Lladser ME, Knights D, Stombaugh J, Knight R. UniFrac: an effective distance metric for microbial community comparison. *Isme Journal.* 2011;5:169-72.
20. Stegen JC, Lin X, Konopka AE, Fredrickson JK. Stochastic and deterministic assembly processes in subsurface microbial communities. *ISME J.* 2012;6:1653-64.
21. Stegen JC, Lin X, Fredrickson JK, Chen X, Kennedy DW, Murray CJ, et al. Quantifying community assembly processes and identifying features that impose them. *ISME J.* 2013;7:2069-79.
22. Moroenyane I, Dong K, Singh D, Chimphango SBM, Adams JM. Deterministic processes dominate nematode community structure in the Fynbos Mediterranean heathland of South Africa. *Ecol Evol.* 2016;30:685-701.
23. Dini-Andreote F, Stegen JC, van Elsas JD, Salles JF. Disentangling mechanisms that mediate the balance between stochastic and deterministic processes in microbial succession. *Proc Natl Acad Sci U S A.* 2015;112:E1326-E32.
24. Liao J, Cao X, Wang J, Zhao L, Sun J, Jiang D, et al. Similar community assembly mechanisms underlie similar biogeography of rare and abundant bacteria in lakes on Yungui Plateau, China. *Limnol Oceanogr.* 2017;62:723-35.
25. Ho A, Di Lonardo DP, Bodelier PLE. Revisiting life strategy concepts in environmental microbial ecology. *FEMS Microbiol Ecol.* 2017;93:fix006.
26. Leff JW, Jones SE, Prober SM, Barberán A, Borer ET, Firn JL, et al. Consistent responses of soil microbial communities to elevated nutrient inputs in grasslands across the globe. *Proc Natl Acad Sci U S A* 2015;112:10967-72.
27. Eilers KG, Lauber CL, Knight R, Fierer N. Shifts in bacterial community structure associated with inputs of low molecular weight carbon compounds to soil. *Soil Biol Biochem.* 2010;42:896-903.
28. Koch AL. Oligotrophs versus copiotrophs. *Bioessays.* 2001;23:657-61.
29. Taylor TB, Buckling A. Competition and Dispersal in *Pseudomonas aeruginosa*. *Am Nat.* 2010;176:83-89.
30. Oono R, Rasmussen A, Lefevre E. Distance decay relationships in foliar fungal endophytes are driven by rare taxa. *Environ Microbiol.* 2017;19:2794-805.
31. Cao P, Wang J, Hu H, Zheng Y, Ge Y, Shen J, et al. Environmental filtering process has more important roles than dispersal limitation in shaping large-scale prokaryotic beta diversity patterns of grassland soils. *Microb Ecol.* 2016;72:221-30.

32. Ferrari BC, Bissett A, Snape I, Van Dorst J, Palmer AS, Ji M, et al. Geological connectivity drives microbial community structure and connectivity in polar, terrestrial ecosystems. *Environ Microbiol.* 2016;18:1834-49.
33. Evans S, Martiny JBH, Allison SD. Effects of dispersal and selection on stochastic assembly in microbial communities. *ISME J.* 2016;11:176-85.
34. Cadotte MW. Phylogenetic diversity-ecosystem function relationships are insensitive to phylogenetic edge lengths. *Funct Ecol.* 2015;29:718-23.
35. Wang J, Shen J, Wu Y, Tu C, Soininen J, Stegen JC, et al. Phylogenetic beta diversity in bacterial assemblages across ecosystems: deterministic versus stochastic processes. *ISME J.* 2013;7:1310-21.
36. Wu W, Logares R, Huang B, Hsieh CH. Abundant and rare picoeukaryotic sub-communities present contrasting patterns in the epipelagic waters of marginal seas in the northwestern Pacific Ocean. *Environ Microbiol.* 2017;19:287-300.
37. Bissett A, Richardson AE, Baker G, Wakelin S, Thrall PH. Life history determines biogeographical patterns of soil bacterial communities over multiple spatial scales. *Mol Ecol.* 2010;19:4315-27.
38. Delgado-Baquerizo M, Oliverio AM, Brewer TE, Benavent-Gonzalez A, Eldridge DJ, Bardgett RD, et al. A global atlas of the dominant bacteria found in soil. *Science.* 2018;359:320-25.
39. Kurm V, van der Putten WH, de Boer W, Naus-Wiezer S, Hol WHG. Low abundant soil bacteria can be metabolically versatile and fast growing. *Ecology.* 2017;98:555-64.
40. Girvan MS, Campbell CD, Killham K, Prosser JI, Glover LA. Bacterial diversity promotes community stability and functional resilience after perturbation. *Environ Microbiol.* 2005;7:301-13.
41. Zhao K, Kong W, Wang F, Long X-E, Guo C, Yue L, et al. Desert and steppe soils exhibit lower autotrophic microbial abundance but higher atmospheric CO<sub>2</sub> fixation capacity than meadow soils. *Soil Biol Biochem.* 2018;127:230-38.
42. Zomer RJ, Trabucco A, Bossio DA, Verchot L. Climate change mitigation: a spatial analysis of global land suitability for clean development mechanism afforestation and reforestation. *Agric Ecosyst Environ.* 2008;126:67-80.
43. Tuan NN, Chang Y-C, Yu C-P, Huang S-L. Multiple approaches to characterize the microbial community in a thermophilic anaerobic digester running on swine manure: A case study. *Microbiol Res.* 2014;169:717-24.
44. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol.* 2009;75:7537-41.
45. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics.* 2011;27:2194-200.
46. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol.* 2007;73:5261-67.

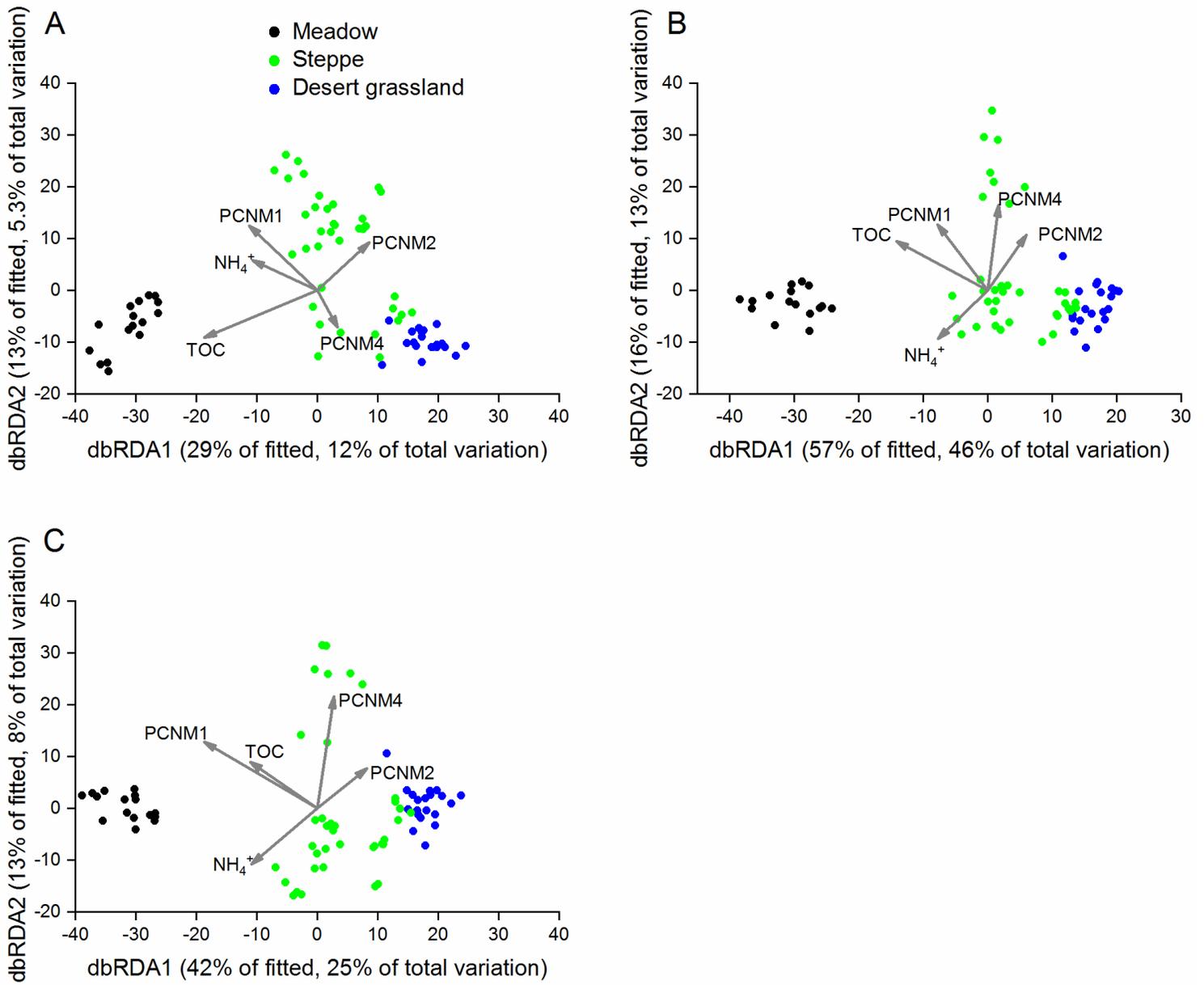
47. Bahram M, Anslan S, Hildebrand F, Bork P, Tedersoo L. Newly designed 16S rRNA metabarcoding primers amplify diverse and novel archaeal taxa from the environment. *Environ Microbiol Rep.* 2019;11:487-94.
48. Jiao S, Chen W, Wei G. Biogeography and ecological diversity patterns of rare and abundant bacteria in oil-contaminated soils. *Mol Ecol.* 2017;26:5305-17.
49. Jiao S, Luo Y, Lu M, Xiao X, Lin Y, Chen W, et al. Distinct succession patterns of abundant and rare bacteria in temporal microcosms with pollutants. *Environ Pollut.* 2017;225:497-505.
50. Yan WK, Ma HM, Shi GT, Li YS, Sun B, Xiao X, et al. Independent shifts of abundant and rare bacterial populations across East Antarctica glacial foreland. *Front Microbiol.* 2017;8:1534.
51. Clarke KR, Warwick RM. PRIMER v6: user manual/tutorial. 2 ed: PRIMER-E, Plymouth; 2006.
52. Ning DL, Deng Y, Tiedje JM, Zhou JZ. A general framework for quantitatively assessing ecological stochasticity. *Proc Natl Acad Sci U S A.* 2019;116:16892-98.
53. Swenson NG. Functional and phylogenetic ecology in R (Use R! series). Gentleman R, Hornik K, Parmigiani G, editors. New York: Springer Science+Business Media; 2014. 144-45.
54. Stegen JC, Lin X, Fredrickson JK, Konopka AE. Estimating and mapping ecological processes influencing microbial community assembly. *Frontiers in Microbiology.* 2015;6:370.
55. Dixon P. VEGAN, a package of R functions for community ecology. *J Veg Sci.* 2003;14:927-30.
56. Legendre P, Anderson MJ. Distance-based redundancy analysis: Testing multispecies responses in multifactorial ecological experiments. *Ecol Monogr.* 1999;69:1-24.
57. Legendre P, Legendre L. Numerical ecology. 3rd ed. Amsterdam: Elsevier Science Ltd; 2012.

## Figures



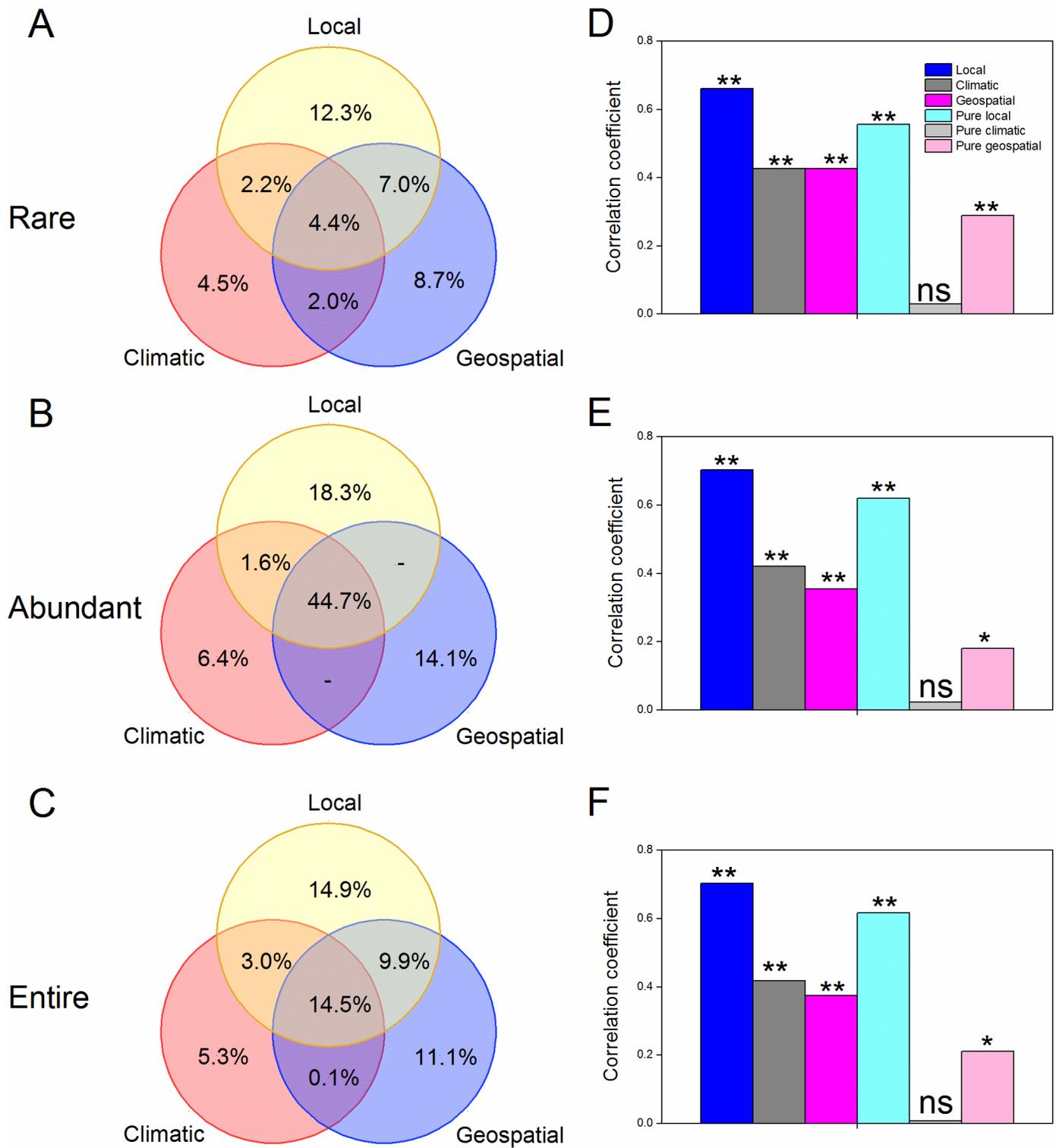
**Figure 1**

Taxonomic compositions and abundance-occupancy relationships. The taxonomic compositions of rare, abundant and entire bacterial communities are shown in (A), and the relative abundance-occupancy relationships of rare and abundant OTUs are shown in (B). Occupancy is the number of samples an OTU being detected from.



**Figure 2**

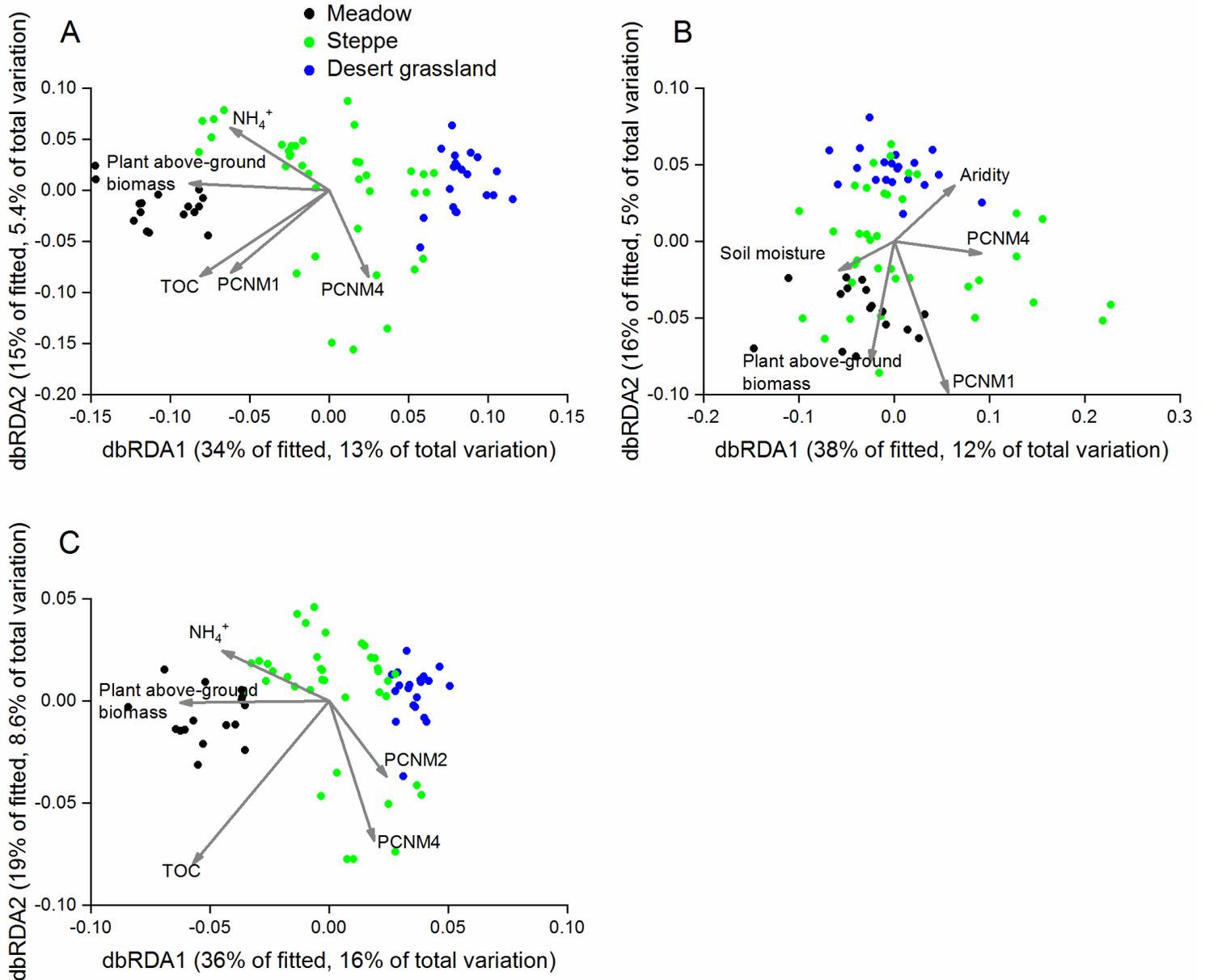
Community composition comparison for rare (A), abundant (B), and entire (C) bacterial communities. The community compositional variation is quantified using Bray-Curtis dissimilarity. The proportion of community variation explained by the measured local, climatic and geospatial factors are calculated using DistLM analysis, and visualized on dbRDA ordination plots (only the top five environmental and geospatial factors are displayed). TOC, total organic carbon; PCNM1-5 are geographical distance derived geospatial factors.



**Figure 3**

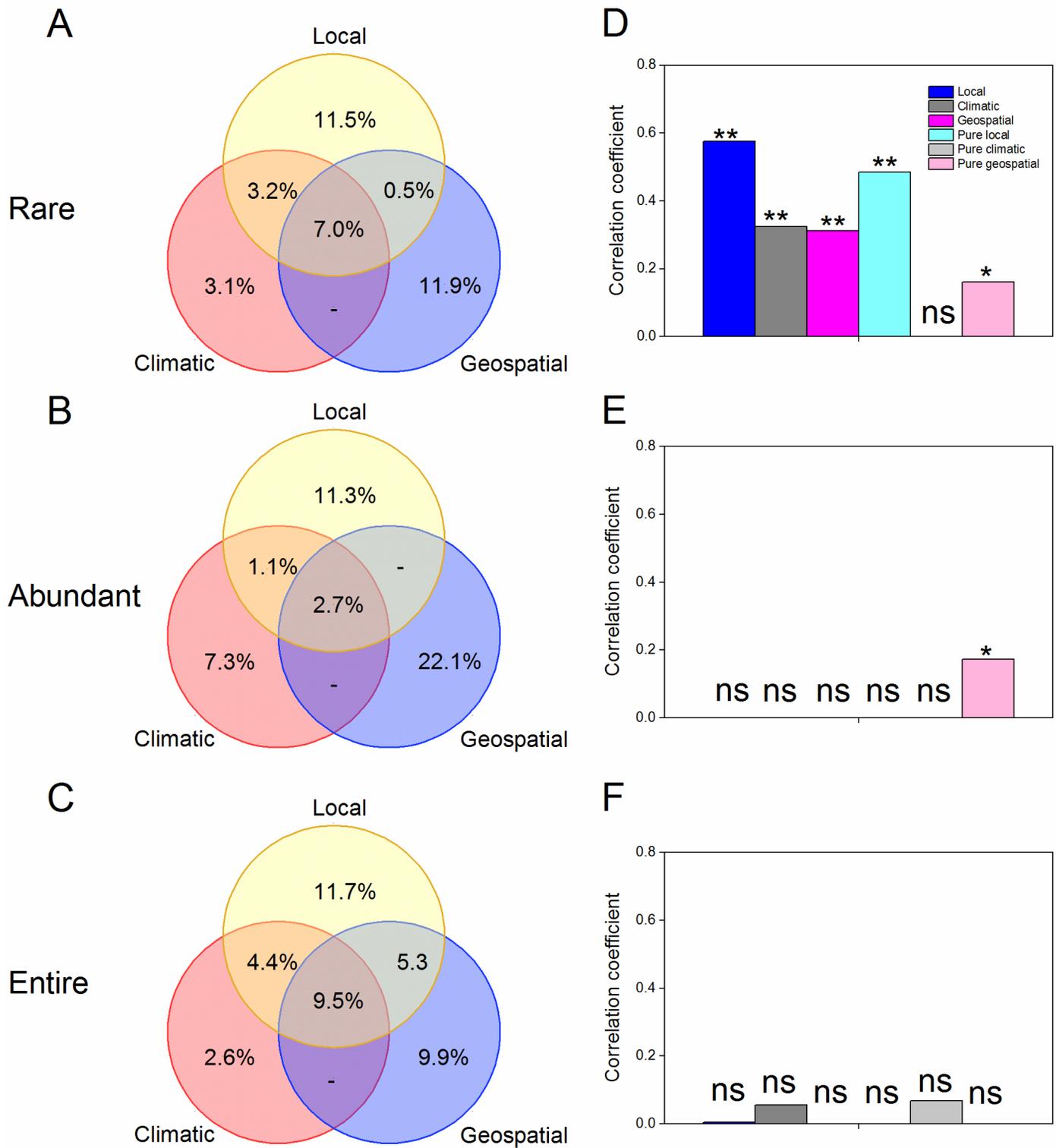
The variation of bacterial community composition explained by local, climatic and geospatial factors. Variation partitioning analysis for rare (A), abundant (B), and entire (C) bacterial communities. The variation explained by pure local, climatic and geospatial factors corresponds to the proportion of community composition explained without the effect of the others by the DistLM analysis. Mantel and partial Mantel tests for the correlations between community composition and the measured local,

climatic and geospatial factors using Pearson's coefficients for rare (D), abundant (E), and entire (F) communities. \*:  $P < 0.01$ , \*\*:  $P < 0.001$ , ns: not significant. Local factors include soil pH, soil total organic carbon,  $\text{NH}_4^+$  and  $\text{NO}_3^-$  concentrations, soil moisture, plant Shannon diversity, plant above-ground biomass, and plant below-ground biomass; the climatic factors include mean annual temperature, mean annual precipitation, and aridity index. – indicates negative or zero proportion of the community composition explained.



**Figure 4**

Community phylogeny comparison for rare (A), abundant (B), and entire (C) bacterial communities. The community phylogeny variation is quantified using the  $\bar{\lambda}$ -mean nearest taxon distance ( $\bar{\lambda}$ -MNTD). The proportion of community phylogeny explained by the measured local, climatic and geospatial factors are calculated using DistLM analysis, and visualized on dbRDA ordination plots (only the top five environmental and geospatial factors are displayed). TOC, total organic carbon; PCNM1-5 are geographical distance derived geospatial factors.



**Figure 5**

The variation of bacterial community phylogeny explained by local, climatic and geospatial factors. Variation partitioning analysis for rare (A), abundant (B), and entire (C) communities. The variation explained by pure local, climatic and geospatial factors corresponds to the proportion of community phylogeny explained without the effect of the others by the DistLM analysis. Mantel and partial Mantel tests for the correlation between community phylogenetic distance and the measured local, climatic and

geospatial factors using Pearson's coefficients for rare (A), abundant (B), and entire (C) communities. \*:  $P < 0.01$ , \*\*:  $P < 0.001$ , ns: not significant. Local factors included soil pH, soil total organic carbon,  $\text{NH}_4^+$  and  $\text{NO}_3^-$  concentrations, soil moisture, plant Shannon diversity, plant above-ground biomass, and plant below-ground biomass; the climatic factors included average annual temperature, annual precipitation, and aridity index. – indicates negative or zero proportion of the community phylogeny explained.

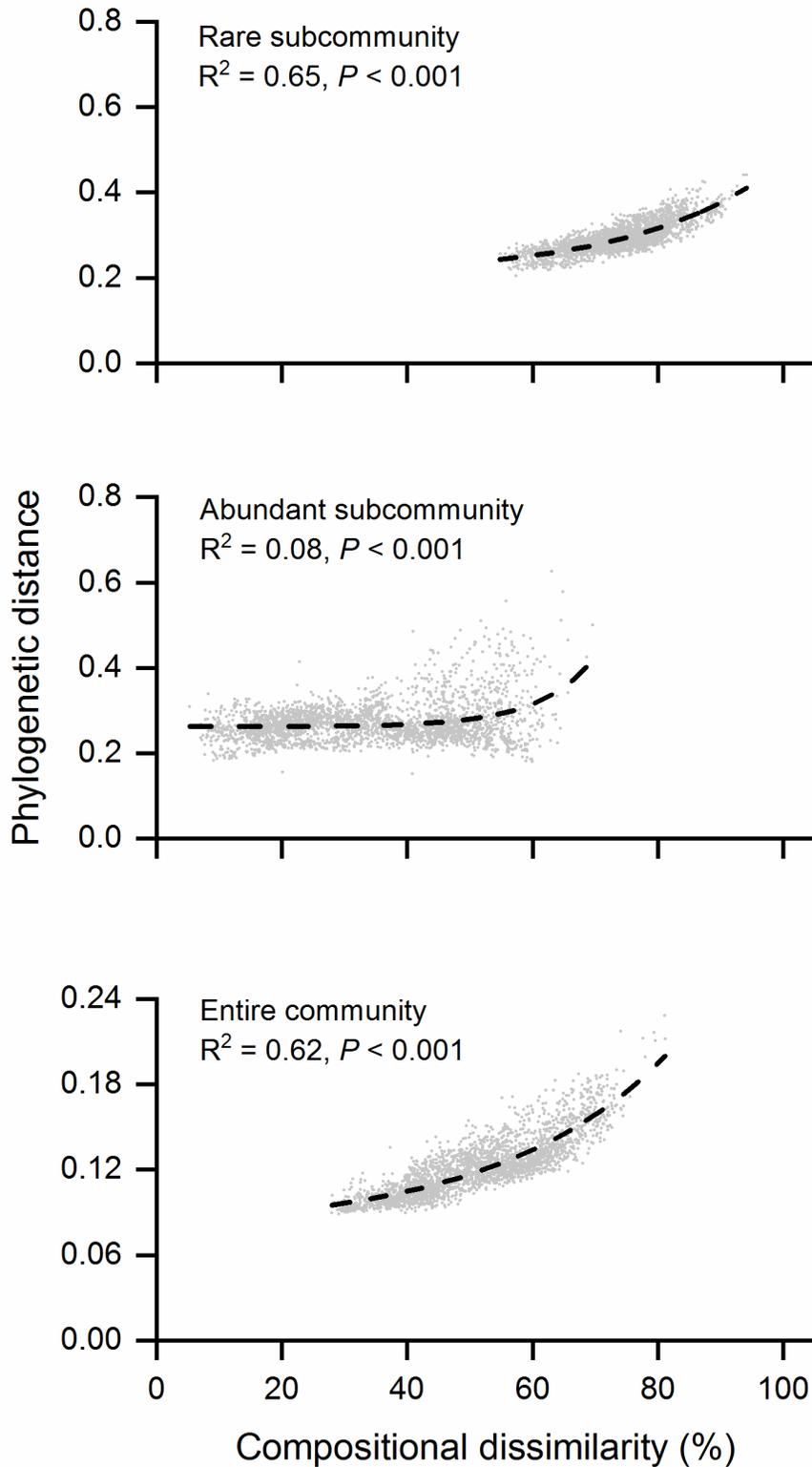


Figure 6

Correlations between the community composition and phylogeny for rare, abundant, and entire communities. The compositional dissimilarity is measured using Bray-Curtis dissimilarity, whereas the phylogenetic distance is measured using  $\bar{r}$ -mean nearest taxon distance. The correlations are fitted by power law correlations.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1.docx](#)
- [Additionalfile2.docx](#)