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

Mukan Ji1,2, Weidong Kong1,2,3*, James Stegen4, Linyan Yue1,2, Fei Wang1,2, Xiaobin Dong5, Don A Cowan6, Belinda C Ferrari7

1Key Laboratory of Alpine Ecology, Institute of Tibetan Plateau Research, Chinese Academy of Sciences (CAS), Beijing 100101, China
2College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100039, China
3CAS Center for Excellence in Tibetan Plateau Earth Sciences, Chinese Academy of Sciences, Beijing 100101, China.
4Pacific Northwest National Laboratory, Richland, Washington 99352, USA
5State Key Laboratory of Earth Surface Processes and Resource Ecology, College of Resources Science and Technology, Beijing Normal University, Beijing 100875, China
6Centre for Microbial Ecology and Genomics, Department of Genetics, University of Pretoria, Pretoria 0002, South Africa
7School of Biotechnology and Biomolecular Sciences, Australian Centre for Astrobiology, UNSW Sydney, Randwick, New South Wales 2052, Australia

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/1462-2920.14993
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: wd Kong@itpcas.ac.cn

Running title: Assembly patterns of rare and abundant bacteria

**Originality-Significance Statement**

We revealed the similar biogeographical patterns of rare and abundant bacteria at both compositional and phylogenetic levels across a grassland transect of over 1200 km. We also showed that the similar patterns were driven by distinct assembly processes, with rare and abundant being influenced by stochasticity and determinism respectively. Our work substantially enhanced the understanding of bacterial biogeography and assembly in terrestrial ecosystems.

**Summary**

Rare biosphere represents the majority of Earth’s biodiversity and perform vital ecological functions, yet little is known about their biogeographical patterns and community assembly processes in terrestrial ecosystems. Herein we investigated the community composition and phylogeny of rare (relative abundance <0.1%) and abundant (>1%) bacteria in dryland grassland soils on the Tibetan Plateau. Results revealed similar biogeographical patterns of rare and abundant bacteria at both compositional and phylogenetic levels, but rare subcommunity was more heavily...
influenced by stochasticity (72%) than the abundant (57%). The compositional variation of rare bacteria was less explained by environmental factors (41%) than that of the abundant (80%), while the phylogeny of rare bacteria (36%) was more explained than that of the abundant (29%). The phylogeny of rare bacteria was equally explained by local factors (soil and vegetation) and geospatial distance (11.5% and 11.9%, respectively), while that of the abundant was more explained by geospatial distance (22.1%) than local factors (11.3%). Furthermore, a substantially tighter connection between the community phylogeny and composition was observed in rare ($R^2=0.65$) than in abundant bacteria ($R^2=0.08$). Our study provides novel insights into the assembly processes and biographical patterns of rare and abundant bacteria in dryland soils.

**Keywords:** Soil bacteria; Rare biosphere; Community structure; Assembly processes.
Introduction

The bacteria in low abundance represent the majority of Earth’s biodiversity (Naeem and Li, 1997), and these rare bacteria contain an enormous pool of genetic novelty (Vigneron et al., 2018). 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 (Montoya et al., 2004; Hua et al., 2015; Hausmann et al., 2016), they enhance plant defence against aphids (Hol et al., 2010), and provide insurance during ecological restoration (Gibbons et al., 2013; Mouillot et al., 2013; Jousset et al., 2017). 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 (Delgado-Baquerizo et al., 2018). Thus, rare bacteria have been identified to drive bacterial community structure changes both spatially and temporally (Gobet et al., 2012; Hugoni et al., 2013; Alonso-Saez et al., 2015). Several studies have examined the biogeography of rare and abundant bacteria in aquatic ecosystems, such as lake reservoirs (Liu et al., 2015), subtropical bays (Mo et al., 2018), the Arctic ocean (Galand et al., 2009), and the Pacific ocean (Wu et al., 2017). 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 (Galand et al., 2009; Liu et al., 2015). 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 (Liu et al., 2015; Mo et al., 2018). 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 (Zhou et al., 2014), the biogeographical pattern of rare bacteria in soil could be different from that of abundant bacteria.

Phylogeny-based community metrics, such as the \( \beta \)-mean nearest taxon distance (\( \beta \)-MNTD) and UniFrac distance, are receiving increased attention (Lozupone et al., 2011; Stegen et al., 2012). Compared to composition-based community metrics, these phylogeny-based community metrics can be used to infer changes in microbial community function (Stegen et al., 2013; Dini-Andreote et al., 2015; Moroenyane et al., 2016). This is based on the assumption that phylogenetically close-related microorganisms have similar habitat associations (Stegen et al., 2012). Thus, using
phylogeny-based community metrics, together with the composition-based community metrics, could provide additional insights into the geographical 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 of 16S rRNA gene, and both the community Bray-Curtis compositional dissimilarity and βMNTD 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, 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, total organic carbon (TOC), and ammonium (\( \text{NH}_4^+ \)) explained a higher proportion of the community variation in abundant over rare subcommunities (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 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 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, Table S4, and 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 (β-Mean Nearest Taxon Index, βNTI) approaches. The MST of rare subcommunity was 72±18%, which was higher than that in abundant (57±20%) and entire communities (66±22%; Fig. S4). Furthermore, the within ecosystem community variations were less influenced by stochasticity than those between ecosystems for rare and entire communities.

The βNTI 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 βNTI of abundant and entire communities varied from -8.8 to -4.5 and -16.5 to 1.7, with 100% and 99% of βNTIs being < -2, respectively (Fig. S5). Significant positive linear correlations of βNTI values with local, climatic and geospatial factors were detected in rare and entire community, whereas the βNTI values of the abundant subcommunity only significantly correlated to local factors (Fig. S6, Table S5).

Furthermore, partial Mantel tests indicated that the βNTI 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 (Liu et al., 2015; Liao et al., 2017) and marine environments (Liu et al., 2015; Liao et al., 2017; Wu et al., 2017; Mo et al., 2018). 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 was consistent with its lower proportion of the community composition explained by environmental factors compared with abundant subcommunity (Figs. S4, 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) (Ho et al., 2017). Copiotrophic bacteria (such as Actinobacteria) are more sensitive to carbon availability than oligotrophic bacteria (Eilers et al., 2010; Leff et al., 2015), which is proposed to be due to a lack of carbon
and energy regulatory system in copiotrophs (Koch, 2001). The higher sensitivity of abundant bacteria to soil nutrients was further evidenced by the stronger contribution of soil nutrients (TOC and NH$_4^+$) in explaining the community composition in abundant than rare subcommunities (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 (Liu et al., 2015). Dispersal limitation is one of the most important stochastic factors influencing bacterial community assembly, and is dependent on cell numbers (Taylor and Buckling, 2010). 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 (Liu et al., 2015; Jousset et al., 2017; Mo et al., 2018). The stronger influence of dispersal limitation on rare over abundant bacteria at the compositional level has also been reported in soil fungal community (Oono et al., 2017).

Stochasticity has been reported to play less important role than determinism in shaping bacterial community composition in grassland and polar desert soils (Cao et al., 2016; Ferrari et al., 2016). 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 (Jousset et al., 2017), 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 (Fig. S4), consistent with previous studies (Evans et al., 2017; Mo et al., 2018).

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 (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 (Stegen et al., 2012). 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 (Cadotte, 2015). Partial Mantel results demonstrated that the community phylogenies of rare and abundant subcommunities both exhibited significant correlations with geospatial factors (Figs. 5D and 5E, Table S4). This contrasted to the lack of correlation between bacterial community phylogeny and geographical distance reported in Wang et al. (2013). 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 (Wu et al., 2017). 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 (Bissett et al., 2010; Delgado-Baquerizo et al., 2018). 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 (Fig. S5). Thus, dispersal limitation could be the main factor constraining the closeness of community phylogeny in abundant subcommunity across the transect (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 (Bissett et al., 2010). Rare bacteria generally exhibit lower environmental adaptability and smaller niche breadth (Kurm et al., 2017) 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 (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 (Wu et al., 2017).

However, unlike aquatic environment, the non-fluidic soil (Zhou et al., 2014) 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 (Wu et al., 2017; Mo et al., 2018).

Decoupled community compositional and phylogenetic variations in abundant
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 (Stegen et al., 2012). 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) (Girvan et al., 2005; Mouillot et al., 2013).

**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.

**Experimental Procedures**

*Soil sampling and soil physiochemical analysis*

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 (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. The sample collection procedures, vegetation characteristics (plant Shannon diversity, plant above-ground biomass, plant below-ground biomass), soil physiochemical properties (soil pH, nitrate, ammonium, total organic carbon), and the collection of climatic factors (aridity index, mean annual precipitation, and mean annual temperature) have been described previously (Zhao et al., 2018).

*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,
Carlsbad, CA, USA). The V4 and V5 hyper-variable regions of the 16S rRNA gene was amplified using primers 515F and 909r (Tuan et al., 2014), and sequenced using an Illumina MiSeq PE250 × 2 platform at the Chengdu Institute of Biology, Chinese Academy of Sciences.

Sequence data processing
Three samples were not sequenced successfully, generating zero reads, and thus removed from downstream analysis. Raw sequence data of the remaining 73 samples were processed using the MOTHUR pipeline (v. 1.34.3) (Schloss et al., 2009). In brief, paired-end reads were merged and quality screened, then aligned against the Silva reference alignment (release 128). Chimeric sequences were identified and removed using UCHIME (Edgar et al., 2011). The remaining sequences were classified using the Bayesian classifier against the Silva database (release 128), and archaeal and unknown sequences were removed. Finally, sequences were classified into operational taxonomic units (OTUs) at 97% identity and singletons were removed. Samples were randomly sub-sampled without replacement to an equal depth of 11,046. The raw sequencing reads generated have been deposited in the NCBI Sequence Read Archive under project ID PRJNA419993.

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 (Zhang et al., in press). 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 composition and phylogenetic variations of the rare, abundant and entire communities were calculated based on the Bray-Curtis dissimilarity matrices and the weighted β-MNTD distance matrices were calculated to indicate community phylogenetic distance between samples. The Bray-Curtis dissimilarity was calculated using Primer V6 (Clarke and Warwick, 2006), whereas the β-MNTD distance was calculated using the function ‘comdistnt’ in the package ‘Picante’ in R (http://www.r-project.org).

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 (Ning et al., 2019). The MST index ranges from 0% to 100%, a 0% indicates zero contribution of stochasticity, whereas 100% indicates the community assembly being completely stochasticity-driven. We also calculated the Nearest Taxon index (NTI) to quantify the relative contributions of stochastic and deterministic processes in shaping community phylogenetic assembly (Stegen et al., 2012) using the script provided in
Swenson (Swenson, 2014). βNTI values $< -2$ indicate the between-sample phylogenetic distance being significantly lower than expected by chance (homogeneous selection); βNTI values $> 2$ indicate the phylogenetic distance being significantly higher than expected (variable selection). βNTI 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.

**Correlation and multivariate analyses**

Measured environmental factors were divided into locally measured (local) and climatic factors. The local factors included soil pH, soil total organic carbon (TOC), ammonium (NH$_4^+$) and nitrate (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, and was converted into a set of geospatial factors using the principal coordinates of neighbour matrices analysis (PCNM) (Dixon, 2003; Mo et al., 2018). The environmental factors (both local and climatic) and the derived geospatial factors were transformed and normalized to reduce data skewness improve data normality for multivariate statistical analysis.
Permutational analysis of variance (PERMANOVA) was used to test the significance of community compositional and phylogenetic differences among the three grassland types (Legendre and Anderson, 1999). 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) (Legendre and Anderson, 1999). Then, distance-based redundancy analysis (dbRDA) ordination plots were used to visualize the associations between bacterial communities and the measured environmental/geospatial factors (Legendre and Anderson, 1999). 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 (Legendre and Legendre, 2012). The PERMANOVA, DistLM and dbRDA ordination analyses were performed using Primer V6 (Clarke and Warwick, 2006), whereas the Mantel and partial Mantel tests were performed using the VEGAN package (Dixon, 2003) in R (http://www.r-project.org).

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.

**Conflict interests**

The authors declare no conflict of interest.

**References**

Alonso-Saez, L., Diaz-Perez, L., and Moran, X.A.G. (2015) The hidden seasonality of the rare biosphere in coastal marine bacterioplankton. *Appl Environ Microbiol* 17: 3766-3780.

Bissett, A., Richardson, A.E., Baker, G., Wakelin, S., and Thrall, P.H. (2010) Life history determines biogeographical patterns of soil bacterial communities over multiple spatial scales. *Mol Ecol* 19: 4315-4327.

Cadotte, M.W. (2015) Phylogenetic diversity-ecosystem function relationships are insensitive to phylogenetic edge lengths. *Funct Ecol* 29: 718-723.

Cao, P., Wang, J., Hu, H., Zheng, Y., Ge, Y., Shen, J., and He, J. (2016) Environmental filtering process has more important roles than dispersal limitation in shaping large-scale prokaryotic beta diversity patterns of grassland soils. *Microb Ecol* 72: 221-230.

This article is protected by copyright. All rights reserved.
Clarke, K.R., and Warwick, R.M. (2006) *PRIMER v6: user manual/tutorial*: PRIMER-E, Plymouth.

Delgado-Baquerizo, M., Oliverio, A.M., Brewer, T.E., Benavent-Gonzalez, A., Eldridge, D.J., Bardgett, R.D. *et al.* (2018) A global atlas of the dominant bacteria found in soil. *Science* **359**: 320-325.

Dini-Andreote, F., Stegen, J.C., van Elsas, J.D., and Salles, J.F. (2015) Disentangling mechanisms that mediate the balance between stochastic and deterministic processes in microbial succession. *Proc Natl Acad Sci U S A* **112**: E1326-E1332.

Dixon, P. (2003) VEGAN, a package of R functions for community ecology. *J Veg Sci* **14**: 927-930.

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

Eilers, K.G., Lauber, C.L., Knight, R., and Fierer, N. (2010) Shifts in bacterial community structure associated with inputs of low molecular weight carbon compounds to soil. *Soil Biol Biochem* **42**: 896-903.

Evans, S., Martiny, J.B.H., and Allison, S.D. (2017) Effects of dispersal and selection on stochastic assembly in microbial communities. *ISME J* **11**: 176-185.

Ferrari, B.C., Bissett, A., Snape, I., Van Dorst, J., Palmer, A.S., Ji, M. *et al.* (2016) Geological connectivity drives microbial community structure and connectivity in
polar, terrestrial ecosystems. *Environ Microbiol* **18**: 1834-1849.

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

Gibbons, S.M., Caporaso, J.G., Pirrung, M., Field, D., Knight, R., and Gilbert, J.A. (2013) Evidence for a persistent microbial seed bank throughout the global ocean. *Proc Natl Acad Sci U S A* **110**: 4651-4655.

Girvan, M.S., Campbell, C.D., Killham, K., Prosser, J.I., and Glover, L.A. (2005) Bacterial diversity promotes community stability and functional resilience after perturbation. *Environ Microbiol* **7**: 301-313.

Gobet, A., Boeer, S.I., Huse, S.M., van Beusekom, J.E.E., Quince, C., Sogin, M.L. *et al.* (2012) Diversity and dynamics of rare and of resident bacterial populations in coastal sands. *ISME J* **6**: 542-553.

Hausmann, B., Knorr, K.H., Schreck, K., Tringe, S.G., del Rio, T.G., Loy, A., and Pester, M. (2016) Consortia of low-abundance bacteria drive sulfate reduction-dependent degradation of fermentation products in peat soil microcosms. *ISME J* **10**: 2365-2375.

Ho, A., Di Lonardo, D.P., and Bodelier, P.L.E. (2017) Revisiting life strategy concepts in environmental microbial ecology. *FEMS Microbiol Ecol* **93**: fix006.
Hol, W.H.G., De Boer, W., Termorshuizen, A.J., Meyer, K.M., Schneider, J.H.M., Van Dam, N.M. et al. (2010) Reduction of rare soil microbes modifies plant–herbivore interactions. *Ecol Lett* 13: 292-301.

Hua, Z., Han, Y., Chen, L., Liu, J., Hu, M., Li, S. et al. (2015) Ecological roles of dominant and rare prokaryotes in acid mine drainage revealed by metagenomics and metatranscriptomics. *ISME J* 9: 1280-1294.

Hugoni, M., Taib, N., Debroas, D., Domaizon, I., Dufournel, I.J., Bronner, G. et al. (2013) Structure of the rare archaeal biosphere and seasonal dynamics of active ecotypes in surface coastal waters. *Proc Natl Acad Sci U S A* 110: 6004-6009.

Jousset, A., Bienhold, C., Chatzinotas, A., Gallien, L., Gobet, A., Kurm, V. et al. (2017) Where less may be more: how the rare biosphere pulls ecosystems strings. *ISME J* 11: 853-862.

Koch, A.L. (2001) Oligotrophs versus copiotrophs. *Bioessays* 23: 657-661.

Kurm, V., van der Putten, W.H., de Boer, W., Naus-Wiezer, S., and Hol, W.H.G. (2017) Low abundant soil bacteria can be metabolically versatile and fast growing. *Ecology* 98: 555-564.

Leff, J.W., Jones, S.E., Prober, S.M., Barberán, A., Borer, E.T., Firn, J.L. et al. (2015) Consistent responses of soil microbial communities to elevated nutrient inputs in grasslands across the globe. *Proc Natl Acad Sci U S A* 112: 10967-10972.
Legendre, P., and Anderson, M.J. (1999) Distance-based redundancy analysis: Testing multisspecies responses in multifactorial ecological experiments. *Ecol Monogr* **69**: 1-24.

Legendre, P., and Legendre, L. (2012) *Numerical ecology*. Amsterdam: Elsevier Science Ltd.

Liao, J., Cao, X., Wang, J., Zhao, L., Sun, J., Jiang, D., and Huang, Y. (2017) Similar community assembly mechanisms underlie similar biogeography of rare and abundant bacteria in lakes on Yungui Plateau, China. *Limnol Oceanogr* **62**: 723-735.

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

Lozupone, C., Lladser, M.E., Knights, D., Stombaugh, J., and Knight, R. (2011) UniFrac: an effective distance metric for microbial community comparison. *ISME J* **5**: 169-172.

Mo, Y., Zhang, W., Yang, J., Lin, Y., Yu, Z., and Lin, S. (2018) Biogeographic patterns of abundant and rare bacterioplankton in three subtropical bays resulting from selective and neutral processes. *ISME J* **12**: 2198-2210.

Montoya, J.P., Holl, C.M., Zehr, J.P., Hansen, A., Villareal, T.A., and Capone, D.G. (2004) High rates of N2 fixation by unicellular diazotrophs in the oligotrophic Pacific Ocean. *Nature* **430**: 1027-1031.
Moroenyane, I., Dong, K., Singh, D., Chimphango, S.B.M., and Adams, J.M. (2016) Deterministic processes dominate nematode community structure in the Fynbos Mediterranean heathland of South Africa. *Ecol Evol* **30**: 685-701.

Mouillot, D., Bellwood, D.R., Baraloto, C., Chave, J., Galzin, R., Harmelin-vivien, M. *et al.* (2013) Rare species support vulnerable functions in high-diversity ecosystems. *PLOS Biol* **11**: e1001569.

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

Ning, D.L., Deng, Y., Tiedje, J.M., and Zhou, J.Z. (2019) A general framework for quantitatively assessing ecological stochasticity. *Proc Natl Acad Sci U S A* **116**: 16892-16898.

Oono, R., Rasmussen, A., and Lefevre, E. (2017) Distance decay relationships in foliar fungal endophytes are driven by rare taxa. *Environ Microbiol* **19**: 2794-2805.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B. *et al.* (2009) Introducing mothur: open-source, platform-Independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**: 7537-7541.

Shade, A., Jones, S.E., Caporaso, J.G., Handelsman, J., Knight, R., Fierer, N., and Gilbert, J.A. (2014) Conditionally rare taxa disproportionately contribute to temporal
changes in microbial diversity. *mBio* 5: doi:10.1128/mBio.01371-14.

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

Stegen, J.C., Lin, X., Fredrickson, J.K., Chen, X., Kennedy, D.W., Murray, C.J. *et al.* (2013) Quantifying community assembly processes and identifying features that impose them. *ISME J* 7: 2069-2079.

Swenson, N.G. (2014) *Functional and phylogenetic ecology in R (Use R! series).* New York: Springer Science+Business Media.

Taylor, T.B., and Buckling, A. (2010) Competition and Dispersal in Pseudomonas aeruginosa. *Am Nat* 176: 83-89.

Tuan, N.N., Chang, Y.C., Yu, C.P., and Huang, S.L. (2014) Multiple approaches to characterize the microbial community in a thermophilic anaerobic digester running on swine manure: A case study. *Microbiol Res* 169: 717-724.

Vigneron, A., Cruaud, P., Alsop, E., de Rezende, J.R., Head, I.M., and Tsesmetzis, N. (2018) Beyond the tip of the iceberg: a new view of the diversity of sulfite- and sulfate-reducing microorganisms. *ISME J* 12: 2096-2099.

Wang, J., Shen, J., Wu, Y., Tu, C., Soininen, J., Stegen, J.C. *et al.* (2013) Phylogenetic beta diversity in bacterial assemblages across ecosystems: deterministic versus
stochastic processes. *ISME J* 7: 1310-1321.

Wu, W., Logares, R., Huang, B., and Hsieh, C. (2017) Abundant and rare picoeukaryotic sub-communities present contrasting patterns in the epipelagic waters of marginal seas in the northwestern Pacific Ocean. *Environ Microbiol* 19: 287-300.

Zhang, H.J., Hou, F.R., Xie, W.J., Wang, K., Zhou, X.Y., Zhang, D.M., and Zhu, X.Y. (in press) Interaction and assembly processes of abundant and rare microbial communities during a diatom bloom process. *Environ Microbiol*. doi: 10.1111/1462-2920.14820.

Zhao, K., Kong, W., Wang, F., Long, X.-E., Guo, C., Yue, L. *et al.* (2018) Desert and steppe soils exhibit lower autotrophic microbial abundance but higher atmospheric CO2 fixation capacity than meadow soils. *Soil Biol Biochem* 127: 230-238.

Zhou, J., Deng, Y., Zhang, P., Xue, K., Liang, Y., Van Nostrand, J.D. *et al.* (2014) Stochasticity, succession, and environmental perturbations in a fluidic ecosystem. *Proc Natl Acad Sci USA* 111: E836-E845.
**Figure legends**

**Fig. 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.

**Fig. 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.

**Fig. 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.

**Fig. 4.** Community phylogeny comparison for rare (A), abundant (B), and entire (C) bacterial communities. The community phylogeny variation is quantified using the $\beta$-mean nearest taxon distance ($\beta$-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.

**Fig. 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.

**Fig. 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 $\beta$-mean nearest taxon distance. The correlations are fitted by power law correlations.