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

Appendix S1: Supporting information

Mukan Ji[1,2], Weidong Kong[1,2,3]*, James Stegen[4], Linyan Yue[1,2], Fei Wang[1,2], Xiaobin Dong[5], Don A Cowan[6], Belinda C Ferrari[7]

[1]Key Laboratory of Alpine Ecology, Institute of Tibetan Plateau Research, Chinese Academy of Sciences (CAS), Beijing 100101, China

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

[3]CAS Center for Excellence in Tibetan Plateau Earth Sciences, Chinese Academy of Sciences, Beijing 100101, China.

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

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

[6]Centre for Microbial Ecology and Genomics, Department of Genetics, University of Pretoria, Pretoria 0002, South Africa

[7]School of Biotechnology and Biomolecular Sciences, Australian Centre for Astrobiology, UNSW Sydney, Randwick, New South Wales 2052, Australia

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

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

Table S1. Proportion of the community compositional dissimilarity explained by each measured environmental and geospatial factors.

	Rare	Abundant	Entire
Total explained	41%	80%	60%
Soil pH	0.3%	1.0%	1%
TOC	0.5%	1.3%	2%
Annual average temperature	0.2%	0.4%	1%
Annual precipitation	0.3%	0.6%	1%
Aridity	0.0%	0.0%	1%
NH_4^+	1.9%	3.0%	3%
NO_3^-	0.2%	0.3%	2%
Plant above-ground biomass	1.5%	1.3%	1%
Plant Shannon diversity index	0.1%	0.9%	2%
Plant below-ground biomass	0.0%	0.0%	1%
Soil moisture	2.5%	2.6%	3%
Geospatial factors	8.7%	14.1%	12%
Unresolved environmental and geospatial factors	24.9%	54%	28%
Unexplained	59.0%	20%	40%

Analyses are based on DistLM analysis. Rare: rare bacterial subcommunity, Abundant: abundant bacterial subcommunity, Entire: entire bacterial community.

Table S2. Correlation relationships between the community compositional dissimilarity and the measured local, climatic and geospatial factors by Mantel and partial Mantel tests.

		Local		Climatic		Geospatial	
		r	P-value	r	P-value	r	P-value
Rare	Mantel	**0.6617**	< 0.001	**0.4263**	< 0.001	**0.4258**	< 0.001
	Partial Mantel	**0.5551**	< 0.001	0.03025	0.133	**0.2891**	< 0.001
Abundant	Mantel	**0.7028**	< 0.001	**0.4214**	< 0.001	**0.3538**	< 0.001
	Partial Mantel	**0.6196**	< 0.001	0.02329	0.327	**0.1806**	< 0.004
Entire	Mantel	**0.7029**	< 0.001	**0.4174**	< 0.001	**0.3742**	< 0.001
	Partial Mantel	**0.6173**	< 0.001	0.007546	0.422	**0.2105**	0.002

Local factors include soil pH, soil total organic carbon, NH_4^+ and 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. Significant correlations (at $P < 0.05$) are in bold. Rare: rare bacterial subcommunity, Abundant: abundant bacterial subcommunity, Entire: entire bacterial community.

Table S3. Proportion of the community phylogenetic distance explained by the individual measured environmental and geospatial factors.

	Rare	Abundant	Entire
Total explained	36%	29%	44%
Soil pH	0.4%	8.6%	4.5%
TOC	0.4%	3.8%	0.9%
Annual average temperature	-	3.1%	2.5%
Annual precipitation	0.4%	0.2%	3.0%
Aridity	0.0%	4.5%	-
NH_4^+	-	7.0%	1.6%
NO_3^-	0.0%	1.7%	1.4%
Plant above-ground biomass	2.9%	4.9%	1.5%
Plant Shannon diversity index	-	0.0%	2.5%
Plant below-ground biomass	-	0.0%	0.0%
Soil moisture	1.0%	4.7%	1.8%
Geospatial factors	11.9%	22.1%	12.4%
Unresolved environmental and geospatial factors	21.8%	-	11.7%
Unexplained	64%	71%	56%

Analyses are based on DistLM analysis. Rare: rare bacterial subcommunity, Abundant: abundant bacterial subcommunity, Entire: entire bacterial community.

Table S4. Correlation relationships between the community phylogenetic distance and the measured local, climatic and geospatial factors by Mantel and partial Mantel tests.

		Local		Climatic		Geospatial	
		r	P-value	r	P-value	r	P-value
Rare	Mantel	**0.5754**	0.001	**0.3249**	0.001	**0.3122**	0.001
	Partial Mantel	**0.4841**	0.001	-0.02751	0.1333	**0.1612**	0.002
Abundant	Mantel	-0.4572	1	-0.449	1	-0.03027	0.713
	Partial Mantel	-0.3922	1	-0.3041	1	**0.1722**	0.007
Entire	Mantel	0.004308	0.455	0.055	0.174	-0.00655	0.54
	Partial Mantel	-0.0092	0.534	0.06811	0.18	-0.01237	0.577

Local factors include soil pH, soil total organic carbon, NH_4^+ and 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. Significant correlations (at $P < 0.05$) are in bold. Rare: rare bacterial subcommunity, Abundant: abundant bacterial subcommunity, Entire: entire bacterial community.
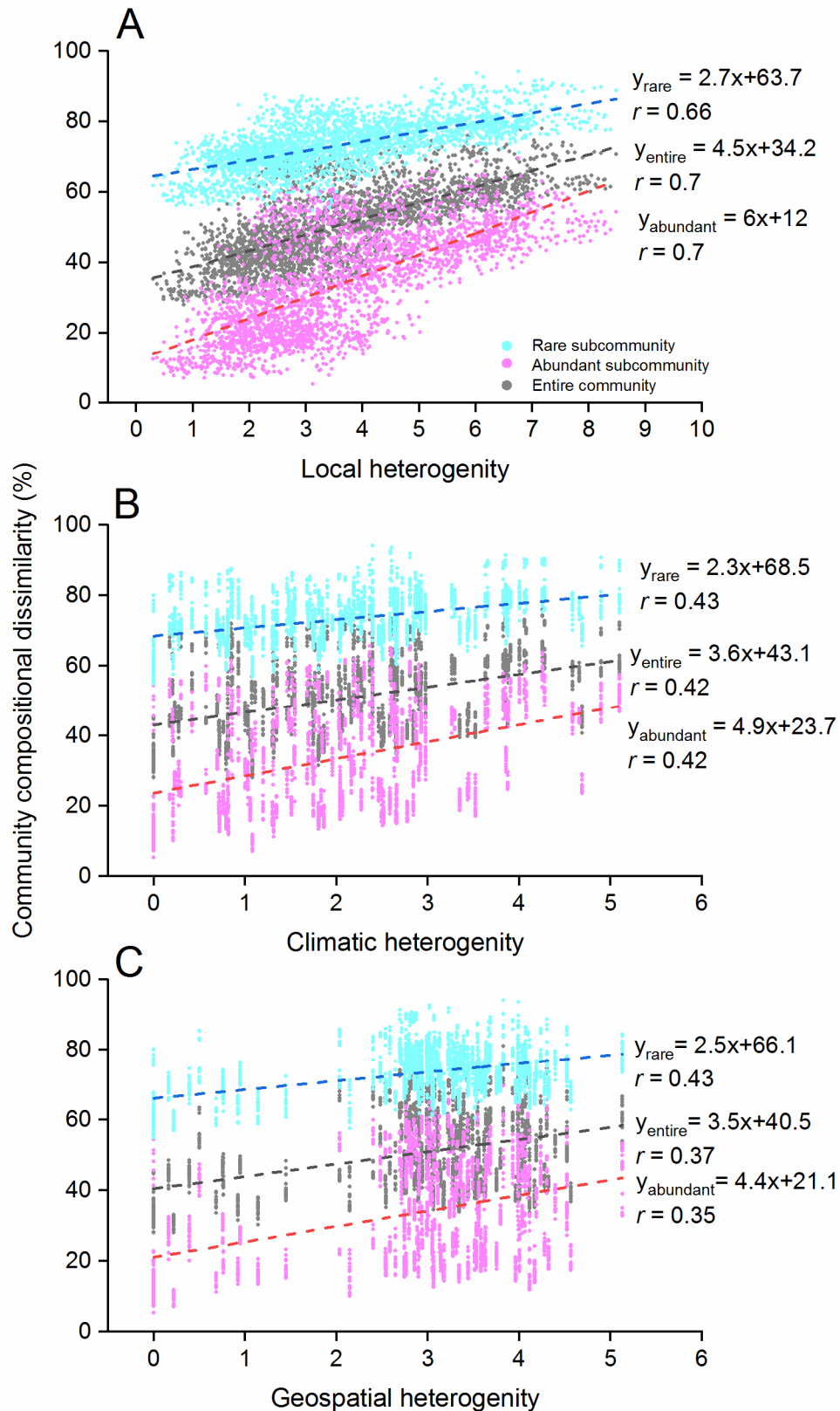
Table S5. Correlation relationships between the □NTI index and the measured local, climatic and geospatial factors by Mantel and partial Mantel tests.

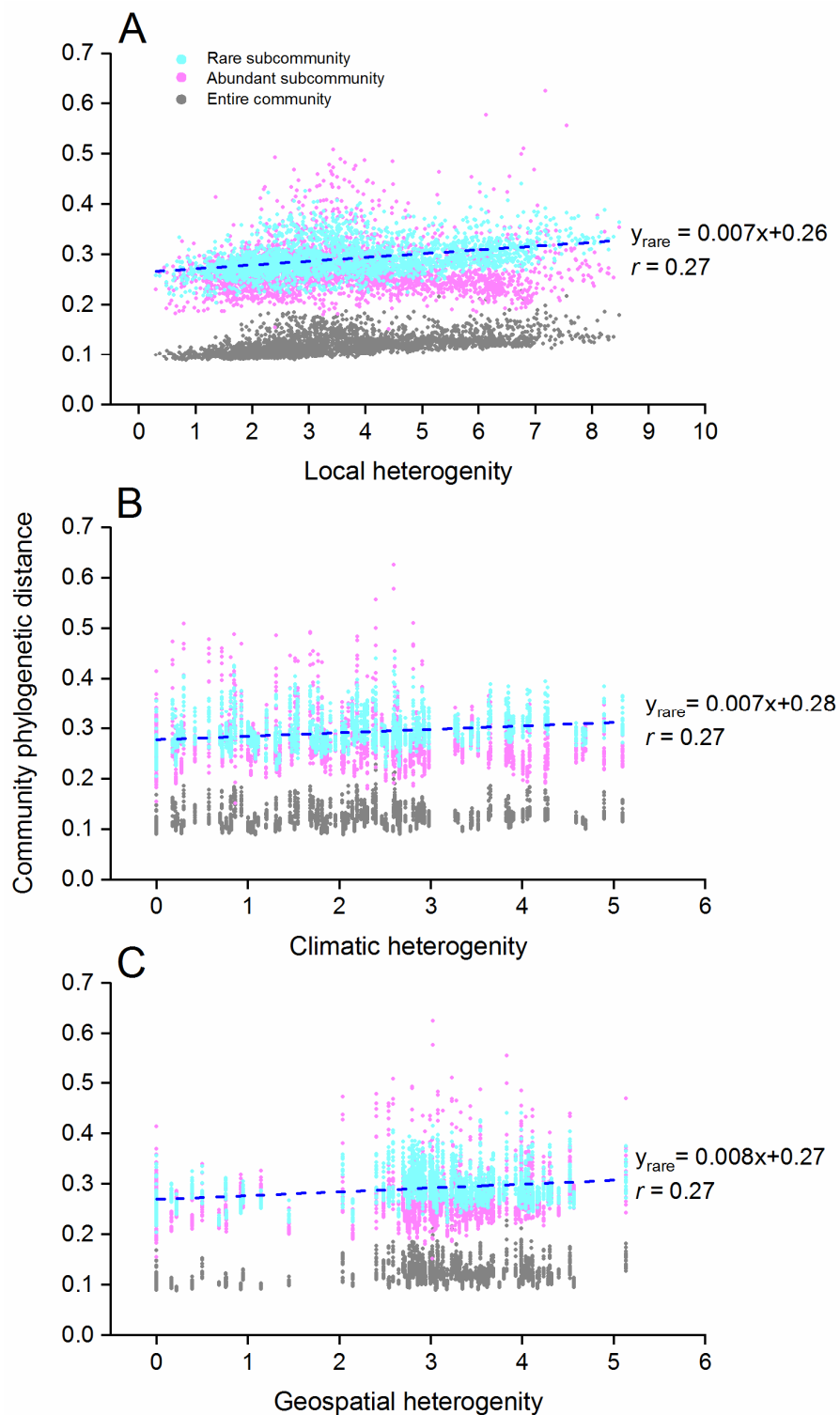|		Local		Climatic		Geospatial	
		r	P-value	*r*	P-value	*r*	P-value
Rare	Mantel	**0.6025**	<0.001	**0.3049**	<0.001	**0.3354**	<0.001
	Partial Mantel	**0.517**	<0.001	-0.1442	0.988	-0.1917	1
Abundant	Mantel	**0.2358**	0.004	0.02147	0.358	0.06624	0.14
	Partial Mantel	**0.2388**	0.002	-0.1461	0.967	-0.1449	0.976
Entire	Mantel	**0.6487**	<0.001	**0.2867**	<0.001	**0.3479**	<0.001
	Partial Mantel	**0.5747**	<0.001	-0.2118	1	-0.2365	1

Local factors include soil pH, soil total organic carbon, NH_4^+ and 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. Significant correlations (at $P < 0.05$) are in bold. Rare: rare bacterial subcommunity, Abundant: abundant bacterial subcommunity, Entire: entire bacterial community.
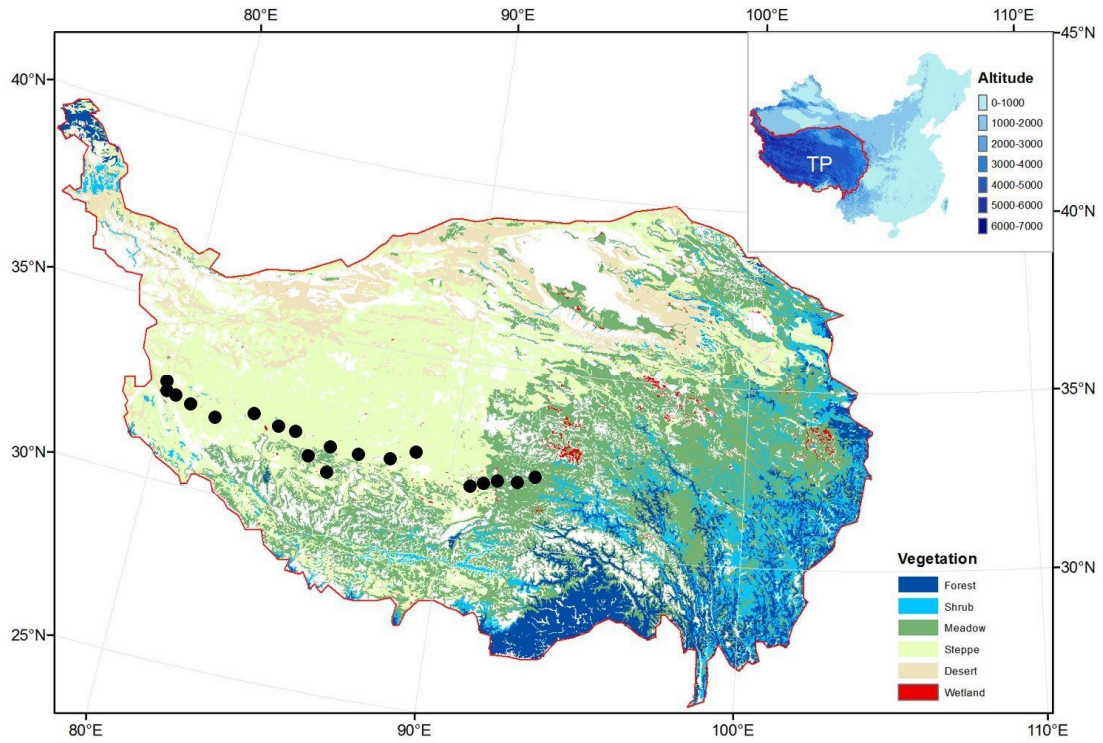
Fig. S1. Soil sampling sites across the Tibetan Plateau. Sampling sites are marked with black dots. Map is modified from the China vegetation map (National specimen information infrastructure, http://www.nsii.org.cn/2017/tdb.php?id=chinavegetaion). TP stands for the Tibetan Plateau.

Fig. S2. Linear correlation relationships between community compositional dissimilarity and the measured local (A), climatic (B) and geospatial factors (C). Local, climatic, and geospatial heterogeneity are based on the corresponding Euclidean distance between samples.
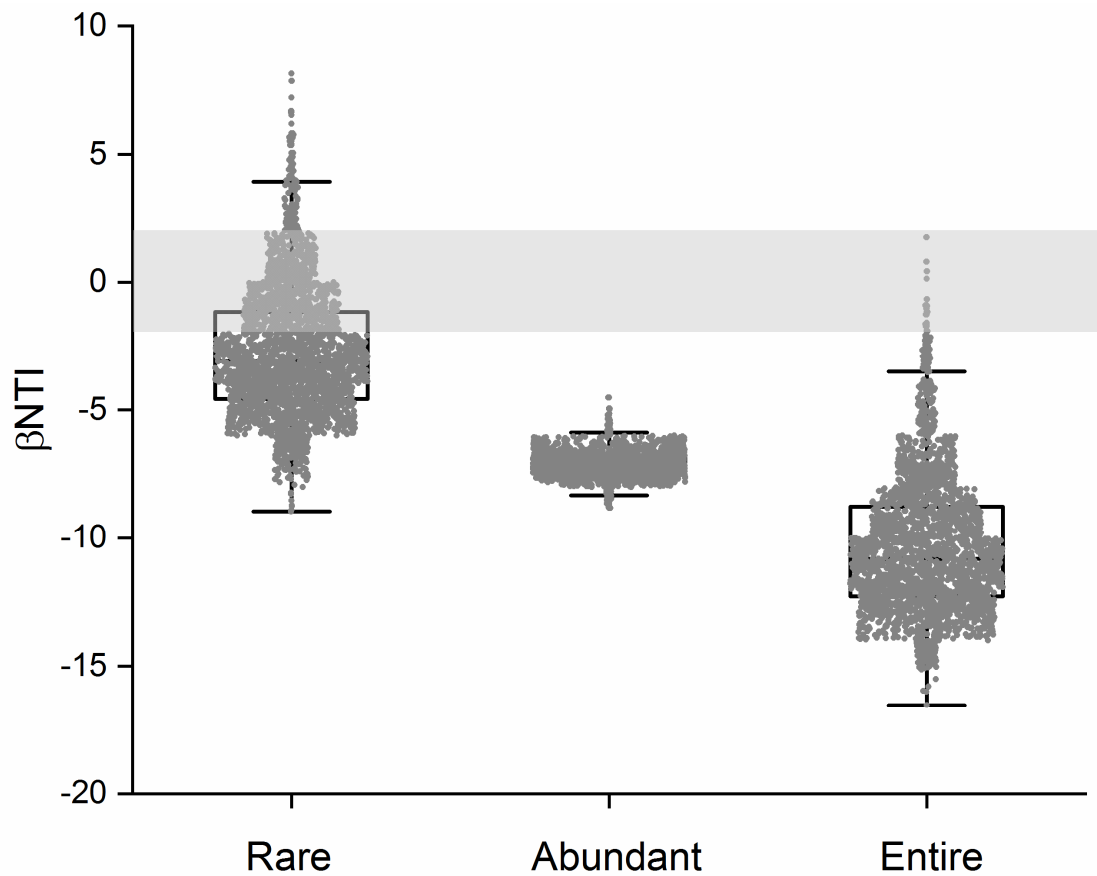
Fig. S3. Linear correlation relationships between the community phylogenetic distance and the measured local (A), climatic (B) and geospatial factors (C). Local, climatic, and geospatial heterogeneity are based on Euclidean distance between samples. Local, climatic, and geospatial heterogeneity are based on the corresponding Euclidean distance between samples. Only significant correlations (Mantel test at $P < 0.05$) are shown.
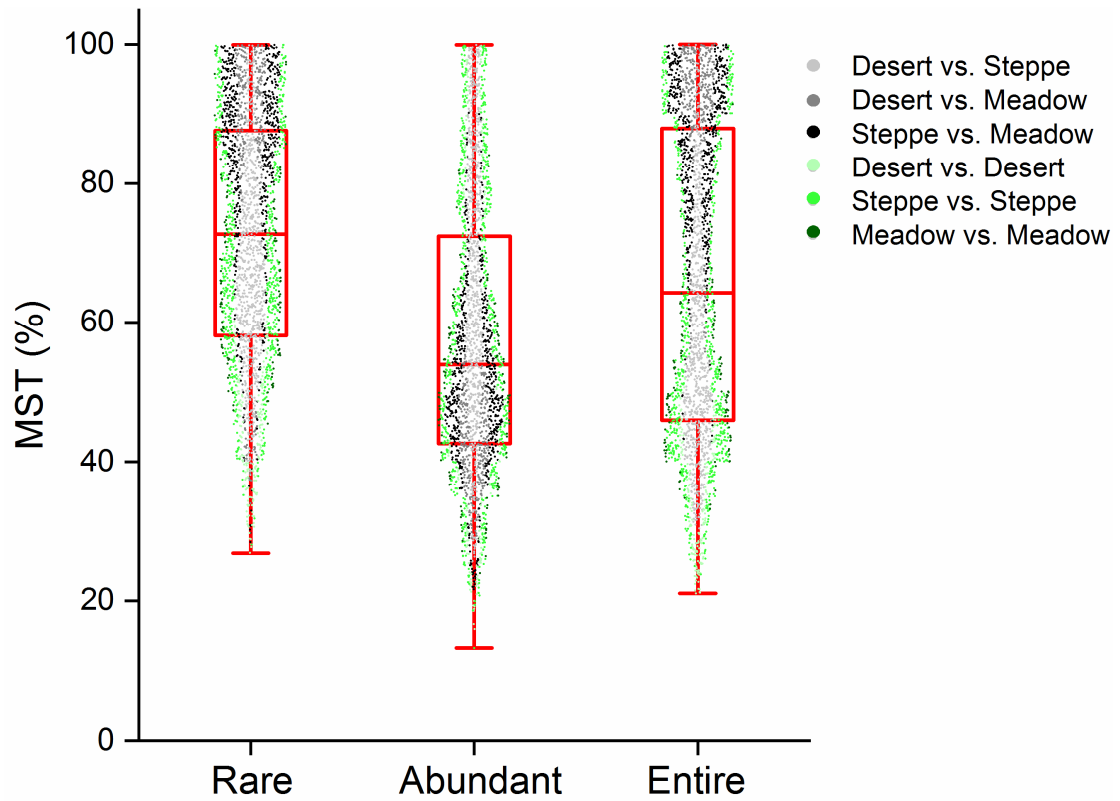
Fig. S4. Modified stochasticity ratio (MST) of rare, abundant and entire bacterial communities (n = 2628). The box signifies the upper and lower quartiles, and the median is represented by a short line within the box. Desert vs. Steppe, Desert vs. Steppe, and Steppe vs. Meadow represent between ecosystem variations, whereas Desert vs. Desert, Steppe vs. Steppe, and Meadow vs. Meadow represent within ecosystem variations.

Fig. S5. $\square$NTI distribution of rare, abundant and entire bacterial communities. βNTI values < −2 indicate significantly less than expected phylogenetic dissimilarity (homogeneous selection); βNTI values > 2 indicate significantly more than expected phylogenetic dissimilarity (variable selection); βNTI values between -2 and 2 indicate the influence of stochastic processes. Grey box indicates the influence of stochasticity.
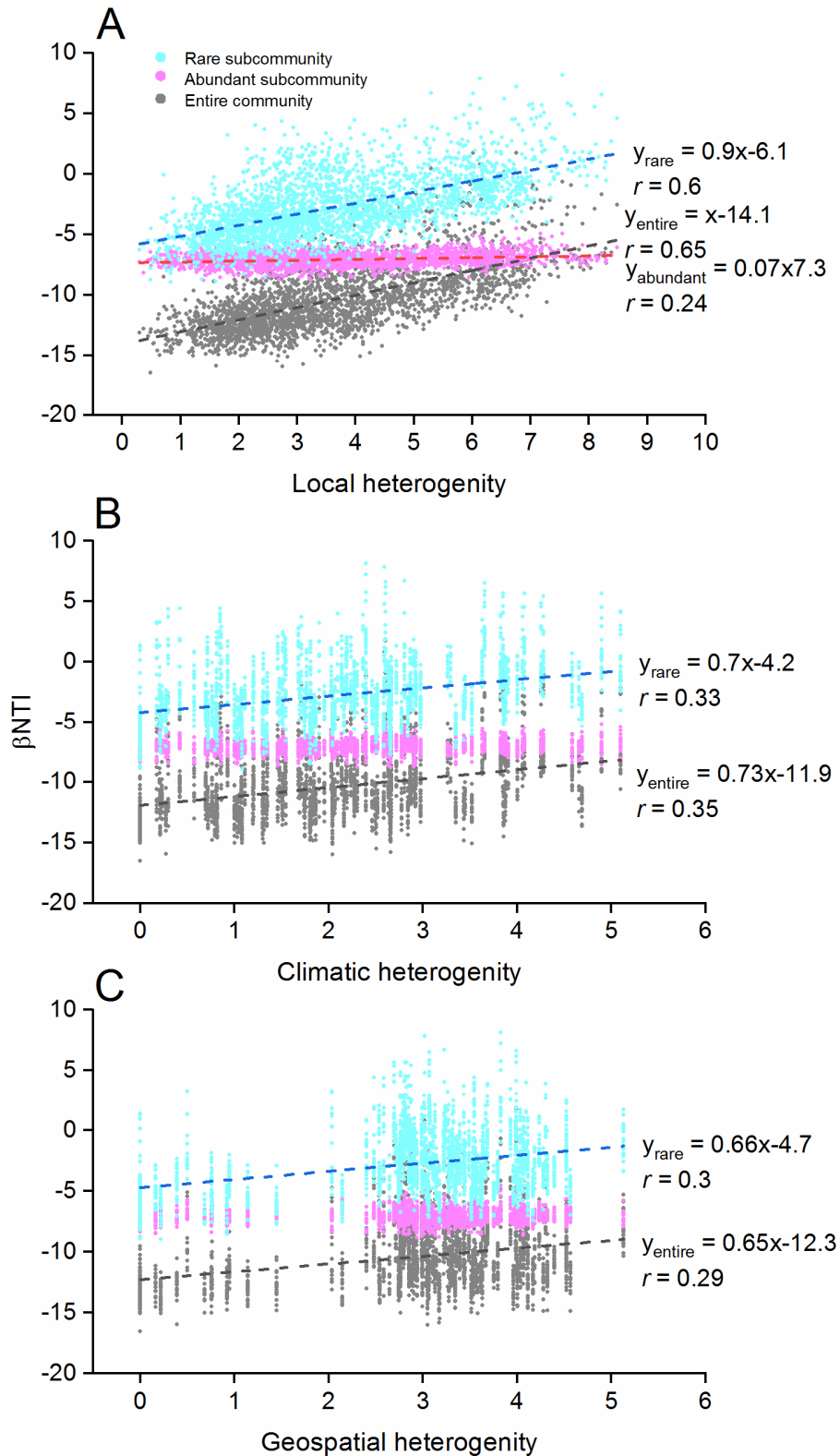
Fig. S6. Linear correlation relationships between the community □NTI and the measured local (A), climatic (B) and geospatial factors (C). Local, climatic, and geospatial heterogeneity are based on the corresponding Euclidean distance between samples. Only significant correlations (Mantel test at $P < 0.05$) are shown.
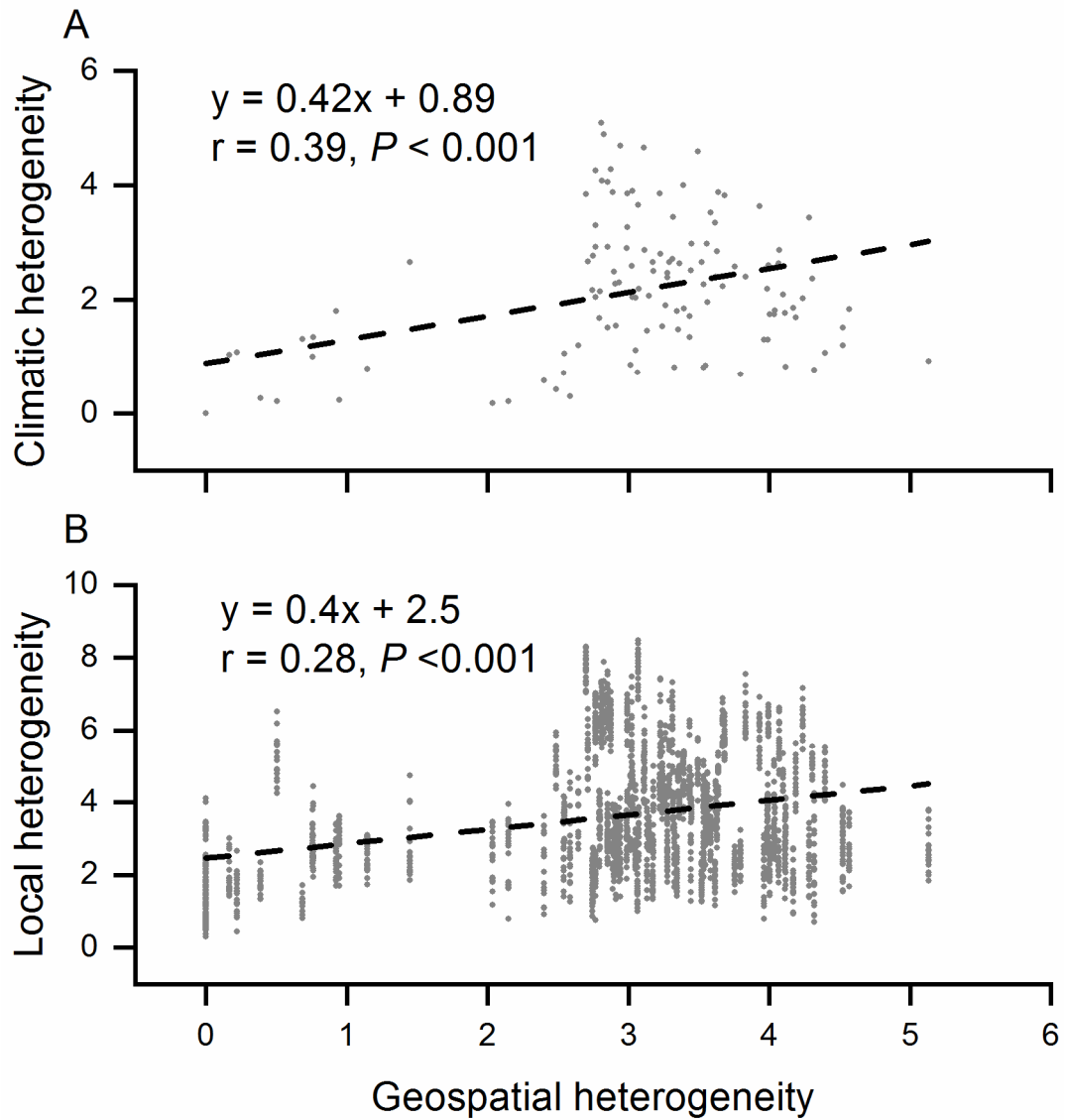
Fig. S7. Linear correlation relationships between geospatial heterogeneity and the measured climatic (A) and local (B) factors. Local, climatic, and geospatial heterogeneity are based on the corresponding Euclidean distance between samples. The significance was tested with Mantel tests.