## Supplementary Information for "Bacterial genome size and gene functional diversity negatively correlate with taxonomic diversity along a

pH gradient" by Wang et al.



Supplementary Fig. 1. Bacterial community and functional composition in association with environmental variables. a-b Principal coordinate (PCo) analysis with environmental fitting (envfit) showing significant association of a amplicon-based and b metagenome-based bacterial community composition with soil pH as well as other biotic and abiotic variables. Note a and b are the same as Fig. 1b and Fig. 1c with additional results of envfit. c-e PCo analysis and envfit showing significant association of soil pH as well as other biotic variables with several groups of functional genes: c Kyoto Encyclopedia of Genes and Genomes Ontology (KO), d antibiotic resistance genes (ARG), and e carbohydrate-active enzymes (CAZy). MAP: mean annual precipitation; MAT: mean annual temperature; TC: total carbon; TN: total nitrogen; TP: total phosphorus; ACa: available calcium; AMg: available magnesium; AFe: available iron; AK: available potassium; C\_N: carbon nitrogen

ratio; C\_P: carbon phosphorus ratio; N\_P: nitrogen phosphorus ratio. n = 36 samples. The numerical data is available in Supplementary Data 1. Source data are provided as a Source Data file.



**Supplementary Fig. 2. Genomic traits along a pH gradient.** Bacterial average genome size and protein counts per genome are calculated by annotating 16S rRNA gene sequence to Genome Taxonomy Database (GTDB). **a** Bacterial average genome size decreased as pH changed from acidic to neutral. **b** Bacterial protein counts per genome significantly and positively correlate with average genome size. Linear regression model with two-sided test was used for the statistical analysis, and adjusted R-squared was used. n = 36 samples. The grey area around the smooth line indicates the 95% confidence interval. Source data are provided as a Source Data file.



**Supplementary Fig. 3.** Associations of relative abundance of top 10 bacterial genera with soil pH. The average genome size for each genus is searched from GTDB, NCBI or published literature <sup>49</sup>. Linear regression model with two-sided test was used for the statistical analysis, and adjusted R-squared was used. n = 36 samples. The grey area around the smooth line indicates the 95% confidence interval. Source data are provided as a Source Data file.



Supplementary Fig. 4. Re-analysis of Bahram et al 2018 showing associations of bacterial average genome size with soil pH. a Soil pH negatively correlated with average genome size as detected by shotgun metagenome using MicrobeCensus pipeline. b Soil pH negatively correlated with average genome size as detected by 16S rRNA amplicon referencing Genome Taxonomy Database (GTDB). Linear regression model with two-sided test was used for the statistical analysis, and adjusted R-squared was used. n = 134 samples. The grey area around the smooth line indicates the 95% confidence interval. Source data are provided as a Source Data file.



Supplementary Fig. 5. Relationships between bacterial average genome size and GC content (GC%). a-b Metagenome-based analysis a showed that bacterial average genome size was negatively correlated with GC%, while the 16S rRNA-GTDB method **b** found that bacterial average genome size was not significantly correlated with GC%. Linear regression model with two-sided test was used for the statistical analysis, and adjusted R-squared was used. n = 36 samples. The grey area around the smooth line indicates the 95% confidence interval. Source data are provided as a Source Data file.



**Supplementary Fig. 6. Relationships between bacterial average genome size and GC content (GC%). a-b** Re-analysis on the global data of Bahram et al 2018 showed that **a** metagenome-based bacterial average genome size was negatively correlated with GC%, whereas **b** 16S rRNA-GTDB based bacterial average genome size was significantly positively correlated with GC%. Linear regression model with two-sided test was used for the statistical analysis, and adjusted R-squared was used. n = 134 samples. The grey area around the smooth line indicates the 95% confidence interval. Source data are provided as a Source Data file.



Supplementary Fig. 7. Contrasting distribution patterns for Shannon diversities of bacterial taxonomy and function along pH gradient. Diversities shown are measured by Shannon's index (H'), and diversities measured by richness (S) are provided in Fig. 3. a Bacterial taxonomic diversity (H'.16S) increased as soil pH changed from acid to neutral. Bacterial operational taxonomic units (OTUs) are detected by 16S rRNA gene amplicon metabarcoding sequencing. b Bacterial functional diversity (H'.KO) decreased as pH changed from acid to neutral. Bacterial functional diversity (H'.KO) decreased as pH changed from acid to neutral. Bacterial functions are determined from the shotgun metagenome, as annotated by Kyoto Encyclopedia of Genes and Genomes (KEGG) Ontology (KO). c

Bacterial diversity of antibiotic resistance genes (H'.ARG) decreased as pH changed from acid to neutral. Bacterial antibiotic resistance genes are detected based on shotgun metagenome annotated by the Resfam database. d Bacterial diversity of carbohydrate-active enzymes (H'.Cazy) genes decreased as pH changed from acid to neutral. Bacterial carbohydrate-active enzymes genes are detected based on shotgun metagenome annotated by database of CAZy. e Bacterial taxonomic diversity (H'.16S) negatively correlated with functional gene diversity (H'.KO). f-h Bacterial average genome size (AGS) positively correlated with functional diversities as measured by g H'.ARG and h H'.Cazy, and not f H'.KO. Linear regression model with two-sided test was used for the statistical analysis, and adjusted R-squared was used. n = 36 samples. The grey area around the smooth line indicates the 95% confidence interval. Source data are provided as a Source Data file.



Supplementary Fig. 8. Bacterial taxonomic and functional compositions in association with biotic and abiotic variables. The heatmap in leftbottom triangle showing the intercorrelations among microbial genomic traits, microbial diversity indices, plant, soil and geographical

variables as detected by Spearman's correlation analysis, with a color gradient (red to blue) and box size denoting Spearman's correlation coefficients (rho). The significance of Spearman's correlation is denoted by \* P < 0.05, \*\* P < 0.01, \*\*\* P < 0.001. The curved lines in the top-right triangle show the association of bacterial taxonomic (the right-top dot) and functional (the right-bottom dot) compositions with biotic and abiotic variables, as detected by Mantel tests. The width of curved lines corresponds to the r statistic for Mantel test, and the color of curved lines denotes the statistical significance. AGS: average genome size; ACN: average 16S rRNA gene copy number; Doubling time: minimal doubling time; MAP: mean annual precipitation; MAT: mean annual temperature; TC: total carbon; TN: total nitrogen; TP: total phosphorus; ACa: available calcium; AMg: available magnesium; AFe: available iron; AK: available potassium; C\_N: carbon nitrogen ratio; C\_P: carbon phosphorus ratio; N\_P: nitrogen phosphorus ratio. Linear regression model with two-sided test was used for the statistical analysis, and adjusted R-squared was used. n = 36 samples. Source data are provided as a Source Data file.



**Supplementary Fig. 9.** Bacterial taxonomic diversity based on metagenomics data, as measured by **a** richness (S) and **b** Shannon's index (H'), did not change significantly as soil pH changed from acidic to neutral. The bacterial taxonomic diversity was detected from metagenome using Kaiju annotation. Linear regression model with two-sided test was used for the statistical analysis, and adjusted R-squared was used. n = 36 samples. Source data are provided as a Source Data file.



**Supplementary Fig. 10.** Heatmap showing significant, positive (blue), and negative (red) Spearman's correlations (rho) between environmental factors (left) and indices of bacterial diversities (bottom), as measured by Shannon's index (H') and richness (S). Diversities of bacterial community were both detected by 16S rRNA amplicon (16S) and shotgun metagenome (mgm) annotated by Kaiju method. Diversities of bacterial functions were detected by shotgun metagenome annotated by Kyoto Encyclopedia of Genes and Genomes Ontology (KO), antibiotic resistance genes (ARG), and carbohydrate-active enzymes (CAZy). The values of significant Spearman's correlations (rho) are labeled on the box. Non-significant correlations are colored white. n = 36 samples. Source data are provided as a Source Data file.



**Supplementary Fig. 11.** Bacterial average genome size negatively correlated with **a** richness (S.16S) and **b** Shannon diversity (H'.16S) of bacterial community. Linear regression model with two-sided test was used for the statistical analysis, and adjusted R-squared was used. n = 36 samples. The grey area around the smooth line indicates the 95% confidence interval. Source data are provided as a Source Data file.



**Supplementary Fig. 12.** Heatmap showing significant, positive (blue), and negative (red) Spearman's correlations (rho) between environmental variables (left) and metagenome annotated by **a** Clusters of Orthologous Genes (COGs) and **b** Kyoto Encyclopedia of Genes and Genomes Ontology (KO). Annotation information of the 11,065 KOs in the horizontal axis is provided in the source data. Non-significant correlations are colored white. n = 36 samples. Source data are provided as a Source Data file.



Supplementary Fig. 13. Genes of glycan biosynthesis and metabolism and carbohydrate-active enzymes (CAZy) associated with biotic and abiotic variables. a Heatmap showing positive (blue), and negative (red) Spearman's correlations (rho) between environmental variables (left) and relative abundance of genes of glycan biosynthesis and metabolism. b Heatmap showing positive (blue), and negative (red) Spearman's correlations (rho) between environmental variables (left) and relative abundance of genes of carbohydrate-active enzymes. Non-significant correlations are colored white. n = 36 samples. Source data are provided as a Source Data file.



Supplementary Fig. 14. Genes of cell motility, signal transduction and cellular community associated with biotic and abiotic variables. a Heatmap showing positive (blue), and negative (red) Spearman's correlations (rho) between environmental variables (left) and relative abundance of genes of cell motility. **b** Heatmap showing positive (blue), and negative (red) Spearman's correlations (rho) between environmental variables (left) and relative abundance of genes of signal transduction (tow component system). **c** Heatmap showing positive (blue), and negative (red) Spearman's correlations (rho) between environmental variables (left) and relative abundance of genes of signal transduction (tow component system). **c** Heatmap showing positive (blue), and negative (red) Spearman's correlations (rho) between environmental variables (left) and relative abundance of genes of cellular community. n = 36 samples. Source data are provided as a Source Data file.



Supplementary Fig. 15. Genes of metabolism of terpenoids and polyketides, xenobiotics biodegradation and metabolism and antibiotic resistance genes associated with biotic and abiotic variables. a Heatmap showing positive (blue), and negative (red) Spearman's correlations (rho) between environmental variables (left) and relative abundance of genes of metabolism of terpenoids and polyketides. b Heatmap showing positive (blue), and negative (red) Spearman's correlations (rho) between environmental variables (left) and relative abundance of genes of xenobiotics biodegradation and metabolism. c Heatmap showing positive (blue), and negative (red) Spearman's correlations (rho) between environmental variables (left) and relative abundance of genes of antibiotic resistance. Non-significant correlations are colored white. n = 36 samples. Source data are provided as a Source Data file.



Supplementary Fig. 16. Prevailing of positive associations between pH and genes of energy metabolism and membrane transport. a

Heatmap showing positive (blue), and negative (red) Spearman's correlations (rho) between environmental variables (left) and relative

abundance of genes of energy metabolism. **b** Heatmap showing positive (blue), and negative (red) Spearman's correlations (rho) between environmental variables (left) and relative abundance of genes of membrane transport. n = 36 samples. Source data are provided as a Source Data file.



**Supplementary Fig. 17.** Heatmap showing positive (blue), and negative (red) Spearman's correlations (rho) between environmental variables (left) and relative abundance of genes involved in translation function. Non-significant correlations are colored white. n = 36 samples. Source data are provided as a Source Data file.



**Supplementary Fig. 18. Genes involving biogeochemical cycle associated with environmental variables.** a Six sulfur cycle genes positively correlate with soil pH, as compared to that two sulfur cycle genes negatively correlate with soil pH. **b** Five Fe cycle genes positively correlate with soil pH, and five iron transport genes negatively correlate with soil pH. Source data are provided as a Source Data file.

Reference	Sampling	Scale	pH range	Curve	Method	Note
1	98 samples; Cross biome	North and South America	3.5 -8.5	Peak in neutral	16S-tRFLP	
2	88 sites (subset of 1)	North and South America	3.5-8.5	Peak in neutral	16S-454seq	
3	13 samples; Wetland	Carolina coastal plain	3.5 - 8.5	Peak in neutral	16S-clone-Seq	
4	7 samples	Qiantang River	6.0-8.0	Increase	16S-454seq	
5	20 samples	Tibetan forest	4.5-6.1	Increase	18S-454seq	Fungi
6	110 sites	China forests		?		No pH ~ diversity
7	17 sediments	15 Tibetan alkaline lake	7-10.5	Decrease	16S-454seq	
8	16 samples?	200m scale	4-8	Increase	16S-454seq	
9	36 samples	Soils from a pH gradient with	4-6	Increase	16S-Illumina	Simpson
		lime addition in agricultural				
		region of southern Brazil				
10	30 samples	Altitude gradient (10 sites)	3.5-7	Increase	18S-Illumina	Fungi
11	18 samples	Tibetan alkaline soils	7.3-9.3	Increase	16S-Illumina	
12			3.5-9.5		16S-Illumina	Stochasticity
13						Stochasticity
14						no pH ~ diversity
15	6 chronosequences (high	Meta-analysis	3.5-9	Peak in neutral	16S-Illumina	
	productivity system)					
16	Dilution					Stochasticity
17	Global, 237 locations				16S-Illumina,	Network, pH cluster
					Shotgun?	
18	179 sites, six ecosystem types,	Scotland	-	Increase	16S-454seq	Stoichiometry, TC, N:P, C:N,

Supplementary Table 1 A list of	publications on the relat	tionship between pH and	microbial diversity
---------------------------------	---------------------------	-------------------------	---------------------

	Scotland					C:P
19	Global	metanalysis	3.1-10.4			
20	France, 2173 soils		3.7-9.2			No pH ~ diversity
21	16 samples across biome	Desert, tundra, prairie, boreal,	4.12-9.95	Increase	16S-Shotgun	
		temperate and tropical forests				
22	Guangdong, 18 samples	Three subtropical forests with	3.6-4.5	Negative	16S- Illumina	H' negative pH
		two soil layers				
23	Zhejiang, China, 95 samples	Bulk soil and rizhospheric soil	3.9-6.5	Increase	16S- Illumina	
		in forests with different age				
24					16s	pH with different phyla
25	China	Meta-analysis, 47 forests and	4.19-8.74	Increase	165	
		105 samples				
26	China, 343 soil samples	Five natural mountain forests	3.8-8.5	Increase, peak	16S-Illumina	Assembly process
		across eastern China		at pH =7		
27	Ningbo, China, 30 samples	pH gradient by adding CaCO3	4.62-7.46	Increase	16S-Illumina	
28	Tibet, China, 72 samples	Upper and deeper soil in	5-9	Negative	16S-Illumina	
		wetland				
29	Australia, 16 samples	Four 78-years old plantations	4.49-6.01	Increase	16S-Illumina	
		with different tree species				
30	29 samples	Agricultural soils	4.26-8.43	Increase	16S-Illumina	pH with different phyla
31	40 samples	Grassland soil with nitrogen	4.0-7.5	Increase	16S-454seq	
		addition				
32	38 samples	Soils from oil refineries	7.5-9.5	Increase	16S- Illumina	
33	115 mineral soil samples	Typical forests across north	4.0-7.5	Increase	16S- Illumina	pH with different phyla

		and south China				
		with the latitudes ranging				
		from 18.70°N to 51.53°N				
34	15 samples	Temperate wheat land with	4.64-6.36	Increase (ns)	16S- Illumina	
		fertilizations				
35	7 samples	Sediment of freshwater river	6.0-8.25	Increase	16S-454seq	
36	26 samples	Soils across the black soil zone	4.56-6.57	Increase	16S-454seq	
		of northeast China				
37	14 samples	soil samples from a defined	4.5-7.5	Increase	16S- Illumina	
		agricultural soil pH gradient in				
		Craibstone, Scotland				
38		Water samples from river		Increase	16S- Illumina	Simpson
39	24 samples	Temperate forest soils with	3.8-6.5	Increase	16S-454seq	pH with different phyla
		altitudinal gradient from 530				
		to 2200 m in Changbai				
		Mountain				
40	1010 samples	Soil samples across the UK	3-9	Increase	16S-TRFLP- BigDye	Simpson
					v3.1 chemistry	
41	Local		3.6-7	Increase	165	
42	Local	12 soil samples with				Negative correlation between
		contrasting land cover				pH and Acidobacteria
43	Global scale	615 composite topsoil samples	3.49-9.54	Increase	165	
		from 151 locations from all	pH=2.99 for			
		continents and 23 countries	one sample			
L		1	1	1		

44	Global scale	30 studies and 1,998 samples				Negative correlation between
		from 21 countries				pH and Acidobacteria
45	Global scale	3986 samples mostly collected	4-12	Peak near	165	16S rRNA gene copy number
		from soil and fresh water	pH=2.0 for one	neutral pH		
			sample	(around 7)		
46	Global scale	Samples of top soils from	4.00-9.60	Increase	165	
		global drylands, Americans,				
		Australia, China, New South				
		Wales				
47		15 soil samples with	3.8-7.75	Increase	165	
		contrasting land use				
48	Global scale	237 sites globally distributed	4.00-9.00	Arid and	165	
		across arid, continental and		Temperate:		
		temperate climates		increase		
				Global: peak at		
				around pH=7		

This study (Wang et al)		Ramoneda et al 2023	Malik et al 2018		
Function (KEGG Level3 or gene)	pH association	Function (pfam)	pH association	Function	pH association
Carbon metabolism	(+)	Sugar metabolism	(+)	Carbon metabolism	(+)
Carbon fixation	(+)				
Glyoxylate and dicarboxylate metabolism	(+)				
Citrate cycle	(+)	Citrate transporter	(+)		
Methane metabolism	(+)				
ABC transporters	(+)	Transmembrane cation transporters/ Transmembrane anion transporter/Na+/H+ antiporters	(+)	ABC transporters	(+)
Metabolism of other amino acids	(+)	Methionine metabolism	(+)	Biosynthesis of amino acids	(+)
Oxidative phosphorylation	(+)			Oxidative phosphorylation	(+)
Ribosome	(+)			Ribosome	(+)
Flagellar assembly	(-)	Motility	(-)		
Metabolism of terpenoids polyketides	(-)				
Xenobiotics biodegradation metabolism	(-)	Phenol degradation	(-)		
Glycan biosynthesis metabolism	(-)				
Bacterial secretion system	(-)	Type IV secretion system	(-)		
Bacterial chemotaxis	(-)				
Two component system	(-)	Transmembrane proteins	(-)		
Porphyrin metabolism	(-)				
Biosynthesis of siderophore	(-)				
Lipopolysaccharide biosynthesis	(-)				

## Supplementary Table 2 Associations of certain microbial functions in this study\*, in Ramoneda et al 2023 and in Malik et al 2018

MTHFD; methylenetetrahydrofolate dehydrogenase	(+)	Folate metabolism	(+)	
ATPase family AAA domain-containing protein 3A/B	(+)	ATPases/AAA_25	(+)	
uvrA/B, excinuclease ABC subunit B	(+)	UV damage repair endonuclease	(+)	
ACH1; acetyl-CoA hydrolase [EC:3.1.2.1]	(-)	Acetyl-CoA hydrolase/transferase	(-)	
pufC; photosynthetic reaction center cytochrome c subunit	(-)	Cytochrome C photosynthetic reaction center	(-)	
dexA; dextranase	(-)	Hydrolases of dextrans Sugar metabolism	(-)	
hipA/B	(-)	Kinase that inhibits tRNA synthase (antibiosis)	(-)	
hypF; hydrogenase maturation protein HypF	(-)	Hydrogenase maturation	(-)	
csxA; exo-1,4-beta-D-glucosaminidase [EC:3.2.1.165]	(-)	Chasseminides		
NAGLU; alpha-N-acetylglucosaminidase [EC:3.2.1.50]	(-)	Glucosaminidase	(-)	
kdpA/B/C	(-)	K <sup>+</sup> transporter	(-)	
rhtB; homoserine/homoserine lactone efflux protein	(-)	Acylated homoserine lactone metabolism	(-)	
scrY; sucrose porin	(-)	Carbohydrate porin	(-)	
TC.FEV.OM3, tbpA, hemR, lbpA, hpuB, bhuR, hugA, hmbR; hemoglobin/transferrin/lactoferrin receptor protein	(-)	Transferrin dimerization domain	(-)	
plc; phospholipase C [EC:3.1.4.3]	(-)	Phospholipase C	(-)	
K14645; serine protease [EC:3.4.21]	(+)	Serine protease	(-)	
ccdA; cytochrome c-type biogenesis protein	(-)	Matif part of C tupa sutashroma		
napC; cytochrome c-type protein NapC	(+)		(-)	

ccmF; cytochrome c-type biogenesis protein CcmF	(+)				
nrfE; cytochrome c-type biogenesis protein NrfE	(+)				
perR; Fur family transcriptional regulator, peroxide stress response regulator	(-)				
furA; Fur family transcriptional regulator, stress- responsive regulator	(-)	Stress response	(-)		
universal stress protein A/G/F/E	(+)				
InuA_C_D_E, lin; lincosamide nucleotidyltransferase A/C/D/E	(+)	Nucleatidulturanterran			
pnp, PNPT1; polyribonucleotide nucleotidyltransferase [EC:2.7.7.8]	(+)	Nucleotidyltransferase	(-)		
				Purine metabolism	(+)
				RNA degradation	(-)
		Photolyase	(+)		
		Endonuclease/Exonuclease/phosphatase family	(+)		
		Lyase of methionine metabolism	(+)		
		Methyltransferases	(+)		
		Kinase that does AMPylation to proteins	(+)		
		Ca-dependent nuclease	(+)		
		Heme binding proteins	(+)		
		Gluc to Fruc 6-phosphate	(+)		
		Kinase that does AMPylation to proteins	(+)		
		Ca-dependent nuclease	(+)		
		Heme binding proteins	(+)		
		Ig like domains	(-)		
		Phosphatases in polysaccharide synthesis	(-)		

	Cytidylate kinase	(-)	
	Dehydrokinase	(-)	
	Helix-turn-helix domain	(-)	
	Lipid hydratase	(-)	
	Quinolones	(-)	
	Thyamin pirophsophate binding domain	(-)	
	Fe receptors	(-)	
	Lipid hydratase	(-)	

\*Note a list of all KOs significantly associated with soil pH of our dataset as detected by threshold indicator analysis (TITAN) is provided in Supplementary Data 2.

## **Supplementary References**

- 1 Fierer, N. & Jackson, R. B. The diversity and biogeography of soil bacterial communities. *Proc. Natl. Acad. Sci. USA.* **103**, 626-631 (2006). https://doi.org:10.1073/pnas.0507535103
- Lauber, C. L., Hamady, M., Knight, R. & Fierer, N. Pyrosequencing-Based Assessment of Soil pH as a Predictor of Soil Bacterial
  Community Structure at the Continental Scale. *Appl. Environ. Microb.* 75, 5111-5120 (2009). <a href="https://doi.org:doi:10.1128/AEM.00335-09">https://doi.org:doi:10.1128/AEM.00335-09</a>
- 3 Hartman, W., Richardson, C., Vilgalys, R. & Bruland, G. Environmental and anthropogenic control of bacterial communities in wetland soils. *Proc. Natl. Acad. Sci. USA.* **105**, 17842-17847 (2008). <u>https://doi.org:10.1073/pnas.0808254105</u>
- 4 Hao, K. *et al.* pH levels drive bacterial community structure in the Qiantang River as determined by 454 pyrosequencing. *Front. Microbiol.* **6**, 285 (2015). <u>https://doi.org:10.3389/fmicb.2015.00285</u>
- 5 Wang, J.-T. *et al.* Soil pH determines the alpha diversity but not beta diversity of soil fungal community along altitude in a typical Tibetan forest ecosystem. *J. Soil. Sediment.* **15**, 1224-1232 (2015). <u>https://doi.org:10.1007/s11368-015-1070-1</u>
- 6 Ma, B. *et al.* Distinct biogeographic patterns for archaea, bacteria, and fungi along the vegetation gradient at the continental scale in eastern China. *mSystems* **2**, e00174-00116 (2017). <u>https://doi.org:10.1128/mSystems.00174-16</u>
- Xiong, J. *et al.* Geographic distance and pH drive bacterial distribution in alkaline lake sediments across Tibetan Plateau. *Environ. Microbiol.* 14, 2457-2466 (2012). <u>https://doi.org/10.1111/j.1462-2920.2012.02799.x</u>
- 8 Rousk, J. *et al.* Soil bacterial and fungal communities across a pH gradient in an arable soil. *ISME J* **4**, 1340-1351 (2010).
- 9 Lammel, D. R. *et al.* Direct and indirect effects of a pH gradient bring insights into the mechanisms driving prokaryotic community structures. *Microbiome* 6, 106 (2018). <u>https://doi.org:10.1186/s40168-018-0482-8</u>
- 10 Liu, D., Liu, G.-h., li, C., Wang, J. & Zhang, L. Soil pH determines fungal diversity along an elevation gradient in Southwestern China. *Sci. China Life Sci.* **61** (2018). <u>https://doi.org:10.1007/s11427-017-9200-1</u>
- 11 Shen, C. *et al.* Soil pH dominates elevational diversity pattern for bacteria in high elevation alkaline soils on the Tibetan Plateau. *FEMS Microbiol. Ecol.* **95** (2018). <u>https://doi.org:10.1093/femsec/fiz003</u>
- 12 Tripathi, B. M. *et al.* Soil pH mediates the balance between stochastic and deterministic assembly of bacteria. *ISME J.* **12**, 1072-1083 (2018). <u>https://doi.org:10.1038/s41396-018-0082-4</u>
- 13 Jiao, S. & Lu, Y. Soil pH and temperature regulate assembly processes of abundant and rare bacterial communities in agricultural ecosystems. *Environ. Microbiol.* **22**, 1052-1065 (2020). <u>https://doi.org/10.1111/1462-2920.14815</u>
- 14 Malik, A. A. *et al.* Land use driven change in soil pH affects microbial carbon cycling processes. *Nat. Commun.* **9**, 3591 (2018). https://doi.org:10.1038/s41467-018-05980-1
- 15 Delgado-Baquerizo, M. *et al.* Changes in belowground biodiversity during ecosystem development. *Proc. Natl. Acad. Sci. USA.* **116**, 6891-6896 (2019). <u>https://doi.org:doi:10.1073/pnas.1818400116</u>

- 16 Xun, W. *et al.* Diversity-triggered deterministic bacterial assembly constrains community functions. *Nat. Commun.* **10**, 3833 (2019). https://doi.org:10.1038/s41467-019-11787-5
- 17 Delgado-Baquerizo, M. *et al.* A global atlas of the dominant bacteria found in soil. *Science* **359**, 320-325 (2018). https://doi.org:10.1126/science.aap9516
- 18 Delgado-Baquerizo, M. *et al.* It is elemental: soil nutrient stoichiometry drives bacterial diversity. *Environ. Microbiol.* **19**, 1176-1188 (2017). <u>https://doi.org/10.1111/1462-2920.13642</u>
- 19 Delgado-Baquerizo, M. *et al.* Carbon content and climate variability drive global soil bacterial diversity patterns. *Ecol. Monogr.* **86**, 373-390 (2016). <u>https://doi.org/10.1002/ecm.1216</u>
- 20 Karimi, B. *et al.* Biogeography of soil bacteria and archaea across France. *Sci. Adv.* **4**, eaat1808 (2018). https://doi.org:doi:10.1126/sciadv.aat1808
- 21 Fierer, N. *et al.* Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proc. Natl. Acad. Sci.* USA. **109**, 21390-21395 (2012). <u>https://doi.org:10.1073/pnas.1215210110</u>
- 22 Wang, Y.-F. *et al.* Key factors shaping prokaryotic communities in subtropical forest soils. *Appl. Soil Ecol.* **169** (2022). https://doi.org:10.1016/j.apsoil.2021.104162
- 23 Wang, B. *et al.* Factors driving the assembly of prokaryotic communities in bulk soil and rhizosphere of Torreya grandis along a 900-year age gradient. *Sci. Total Environ.* **837**, 155573 (2022). <u>https://doi.org:10.1016/j.scitotenv.2022.155573</u>
- 24 Muneer, M. A. *et al.* Soil pH: a key edaphic factor regulating distribution and functions of bacterial community along vertical soil profiles in red soil of pomelo orchard. *BMC Microbiol.* **22**, 38 (2022). <u>https://doi.org:10.1186/s12866-022-02452-x</u>
- 25 Yan, K., Dong, Y., Gong, Y., Zhu, Q. & Wang, Y. Climatic and edaphic factors affecting soil bacterial community biodiversity in different forests of China. *Catena* **207** (2021). <u>https://doi.org:10.1016/j.catena.2021.105675</u>
- 26 Ni, Y. *et al.* Soil pH determines bacterial distribution and assembly processes in natural mountain forests of eastern China. *Global Ecol. Biogeogr.* **30**, 2164-2177 (2021). <u>https://doi.org:10.1111/geb.13373</u>
- Li, H. Q. *et al.* Soil pH has a stronger effect than arsenic content on shaping plastisphere bacterial communities in soil. *Environ. Pollut.* 287, 117339 (2021). <u>https://doi.org:10.1016/j.envpol.2021.117339</u>
- 28 Kang, E. *et al.* Soil pH and nutrients shape the vertical distribution of microbial communities in an alpine wetland. *Sci. Total Environ.* **774** (2021). <u>https://doi.org:10.1016/j.scitotenv.2021.145780</u>
- 29 Zhou, X., Guo, Z., Chen, C. & Jia, Z. Soil microbial community structure and diversity are largely influenced by soil pH and nutrient quality in 78-year-old tree plantations. *Biogeosciences* **14**, 2101-2111 (2017). <u>https://doi.org:10.5194/bg-14-2101-2017</u>
- 30 Wu, Y., Zeng, J., Zhu, Q., Zhang, Z. & Lin, X. pH is the primary determinant of the bacterial community structure in agricultural soils impacted by polycyclic aromatic hydrocarbon pollution. *Sci. Rep.* **7**, 40093 (2017). <u>https://doi.org:10.1038/srep40093</u>

- 31 Zeng, J. *et al.* Nitrogen fertilization directly affects soil bacterial diversity and indirectly affects bacterial community composition. *Soil Biol. Biochem.* **92**, 41-49 (2016). <u>https://doi.org:10.1016/j.soilbio.2015.09.018</u>
- Jiao, S. *et al.* Bacterial communities in oil contaminated soils: Biogeography and co-occurrence patterns. *Soil Biol. Biochem.* **98**, 64-73 (2016). <u>https://doi.org:10.1016/j.soilbio.2016.04.005</u>
- 33 Xia, Z. *et al.* Biogeographic distribution patterns of bacteria in typical Chinese forest soils. *Front. Microbiol.* **7**, 1106 (2016). https://doi.org:10.3389/fmicb.2016.01106
- 34 Zhou, J. *et al.* Influence of 34-years of fertilization on bacterial communities in an intensively cultivated black soil in northeast China. *Soil Biol. Biochem.* **90**, 42-51 (2015). <u>https://doi.org:10.1016/j.soilbio.2015.07.005</u>
- Liu, S. *et al.* pH levels drive bacterial community structure in sediments of the Qiantang River as determined by 454 pyrosequencing. *Front. Microbiol.* **6**, 285 (2015). <u>https://doi.org:10.3389/fmicb.2015.00285</u>
- Liu, J. *et al.* High throughput sequencing analysis of biogeographical distribution of bacterial communities in the black soils of northeast China. *Soil Biol. Biochem.* **70**, 113-122 (2014). <u>https://doi.org:10.1016/j.soilbio.2013.12.014</u>
- 37 Bartram, A. K. *et al.* Exploring links between pH and bacterial community composition in soils from the Craibstone Experimental Farm. *FEMS Microbiol. Ecol.* **87**, 403-415 (2014). <u>https://doi.org:10.1111/1574-6941.12231</u>
- 38 Staley, C. *et al.* Application of Illumina next-generation sequencing to characterize the bacterial community of the Upper Mississippi River. *J. Appl. Microbiol.* **115**, 1147-1158 (2013). <u>https://doi.org:10.1111/jam.12323</u>
- Shen, C. *et al.* Soil pH drives the spatial distribution of bacterial communities along elevation on Changbai Mountain. *Soil Biol. Biochem.* 57, 204-211 (2013). <u>https://doi.org:10.1016/j.soilbio.2012.07.013</u>
- 40 Griffiths, R. I. *et al.* The bacterial biogeography of British soils. *Environ. Microbiol.* **13**, 1642-1654 (2011). <u>https://doi.org:10.1111/j.1462-2920.2011.02480.x</u>
- 41 Zhalnina, K. *et al.* Soil pH Determines Microbial Diversity and Composition in the Park Grass Experiment. *Microb. Ecol.* **69**, 395-406 (2015). <u>https://doi.org:10.1007/s00248-014-0530-2</u>
- 42 Lauber, C. L., Strickland, M. S., Bradford, M. A. & Fierer, N. The influence of soil properties on the structure of bacterial and fungal communities across land-use types. *Soil Biol. Biochem.* **40**, 2407-2415 (2008). <u>https://doi.org:10.1016/j.soilbio.2008.05.021</u>
- 43 Guerra, C. A. *et al.* Global hotspots for soil nature conservation. *Nature* **610**, 693-698 (2022). <u>https://doi.org:10.1038/s41586-022-05292-x</u>
- 44 Ramirez, K. S. *et al.* Detecting macroecological patterns in bacterial communities across independent studies of global soils. *Nat. Microbiol.* **3**, 189-196 (2018). <u>https://doi.org:10.1038/s41564-017-0062-x</u>
- 45 Thompson, L. R. *et al.* A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* **551**, 457-463 (2017). https://doi.org:10.1038/nature24621

- 46 Delgado-Baquerizo, M. *et al.* Palaeoclimate explains a unique proportion of the global variation in soil bacterial communities. *Nat. Ecol. Evol.* **1**, 1339-1347 (2017). <u>https://doi.org:10.1038/s41559-017-0259-7</u>
- 47 Terrat, S. *et al.* Meta-barcoded evaluation of the ISO standard 11063 DNA extraction procedure to characterize soil bacterial and fungal community diversity and composition. *Microb. Biotechnol.* **8**, 131-142 (2015). <u>https://doi.org:10.1111/1751-7915.12162</u>
- 48 Delgado-Baquerizo, M. & Eldridge, D. J. Cross-Biome Drivers of Soil Bacterial Alpha Diversity on a Worldwide Scale. *Ecosystems* **22**, 1220-1231 (2019). <u>https://doi.org:10.1007/s10021-018-0333-2</u>
- 49 Brewer, T.E., Handley, K.M., Carini, P., et al. (2017). Genome reduction in an abundant and ubiquitous soil bacterium '*Candidatus Udaeobacter copiosus*'. *Nat. Microbiol.* 2, 16198, 10.1038/nmicrobiol.2016.198.