

SCIENTIFIC REPORTS



OPEN

Environmental filtering of bacterial functional diversity along an aridity gradient

Ho-Kyung Song¹, Yu Shi², Teng Yang², Haiyan Chu², Jin-Sheng He³, Hyoki Kim⁴, Piotr Jablonski^{1,5} & Jonathan M. Adams⁶

Studying how metagenome composition and diversity varies along environmental gradients may improve understanding of the general principles of community and ecosystem structuring. We studied soil bacterial metagenomes along a precipitation gradient on the eastern Tibetan Plateau, varying between 500 mm and 60 mm mean annual precipitation (MAP). We found that lower MAP was strongly associated with reduced functional diversity of bacterial genes. It appears that extreme environmental conditions associated with aridity constrain the diversity of functional strategies present in soil biota – analogous to broad scale patterns found in plant functional diversity along environmental gradients. In terms of specific functions, more extreme arid conditions were also associated with increased relative abundance of genes related to dormancy and osmoprotectants. Decreased relative abundance of genes related to antibiotic resistance and virulence in more arid conditions suggests reduced intensity of biotic interaction under extreme physiological conditions. These trends parallel those seen in earlier, more preliminary comparisons of metagenomes across biomes.

Soil metagenomes can potentially reveal patterns of variation in functionality of soil biota – and from these the general ecological patterns of community structure and ecosystem function¹. The search for community-level traits in metagenomes is based upon genes considered to code for these traits, rather than observation of the traits themselves, which would often be difficult to observe and quantify within the soil biota.

Our primary interest in this study was the effect of an environmental stress gradient on the functional diversity of the soil biota. Theory and observation of larger organisms, notably plants, has both predicted and observed broad scale gradients in the diversity of functional traits along climatic stress gradients^{2–4}. The authors of those studies suggested that extreme physiological conditions and reduced energy budgets for organisms constrain the number of different functional strategies which are viable in a particular environment. We were interested in testing whether similar rules also hold true for soil biota, with a more extreme environment being associated with decreased diversity of functional traits.

There have been various studies on the effect of climatic aridity on overall microbial community structure^{5,6}. For example, Maestre *et al.*⁵ studied microbial community structure in a wide range of global drylands. They found that microbial diversity decreased along a gradient of increasing aridity, and that relative abundances of major groups of bacteria and fungi are correlated with the degree of aridity. However, the effects of aridity stress on relative abundance of traits and overall functional trait diversity in soil biota are largely unknown.

In this study, we aimed to use whole metagenome shotgun sequencing of soil DNA to understand how a gradient in aridity - and aridity-associated stress gradients such as salinity - may affect functional diversity of genes in the soil bacterial community, and the relative abundances of genes associated with particular functional traits. Aridity, together with temperature, has long been regarded as one of the ultimate driving factors behind

¹Laboratory of Behavioral Ecology and Evolution, Department of Biological Sciences, Seoul National University, Seoul, 151-742, South Korea. ²State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, East Beijing Road 71, Nanjing, 210008, China. ³Department of Ecology, College of Urban and Environmental Sciences, Peking University, 5 Yiheyuan Road, Beijing, 100871, China. ⁴Celeomics Inc., 19F, Bldg. A, BYC High city, 131, Gasandigital 1-ro, Gwumcheon-gu, Seoul, 153-718, Korea. ⁵Museum and Institute of Zoology, Polish Academy of Sciences, Wilcza 64, Warsaw, Poland. ⁶School of Water, Energy and Environment, Cranfield University, Cranfield, MK43 0AL, UK. Correspondence and requests for materials should be addressed to P.J. (email: snulbee@behecolpiotrsangim.org) or J.M.A. (email: geograph.ecol@gmail.com)

ecosystem variation⁷. We chose the Tibetan Plateau as our study area because it presents a strong aridity gradient under similar mean annual temperature conditions.

We expected functional gene diversity to be more constrained along a gradient of increasing aridity, and reduced functional diversity to be closely associated with reduced OTU (operational taxonomic unit) diversity. Extreme conditions may require a larger proportion of an organism's energy and other resources to be diverted into homeostasis (e.g. against extreme osmotic conditions, extreme pH, lack of water etc.)^{8,9}. This may preclude other functional traits which could otherwise enhance survival (e.g. through additional strategies of nutrient acquisition, mutualism, or interference competition) or allow occupation of a wider range of different types of niches.

We also expected that there would be significant differences in functional gene composition along an aridity gradient, in terms of which gene types are relatively more common in the metagenome. In terms of the prevalence of particular functional traits along an aridity gradient (defined in terms of a spatial climate gradient from higher precipitation to lower precipitation), we hypothesized the following trends based on general ecological principles.

The relative abundance of genes related to stress responses will increase towards more arid conditions. Desert environments are often regarded as 'stressful' because of high disturbance rates through wind erosion, high salinity, low moisture content, low nutrient concentrations, and low primary productivity. We thus expected that bacteria and other micro-organisms will cope with (a) water deficiency by accumulating protectant solutes^{10–12}, and (b) with oxidative stress (due to high radiation) by producing antioxidants¹³. Bacteria will also (c) more frequently possess dormancy and sporulation mechanisms to cope with extreme conditions¹⁴, when their cells cannot sustain basic metabolism.

The relative abundance of genes related to competitive traits will decrease towards more arid conditions, whereas the relative abundance of housekeeping genes will increase. In ecology, it is generally considered that in extreme environments, the physiological costs and functional constraints on organisms leave relatively little of their resources for interference competition^{15,16}. Additionally, the slow rates of growth and population increase in extreme environments - relative to rates of disturbance - are seen as resulting in less intense competition¹⁷. Thus, we anticipated that there would be relatively low abundances of genes relating to competitive traits, for example: (a) production of secondary metabolites and (b) antibiotic resistance^{18,19}. Instead, the core housekeeping genes, for example (c) genes related to cell division, (d) DNA metabolism, or (e) protein metabolism, were expected to play a relatively more important role in maintaining basic metabolism of bacterial cells for survival or reproduction²⁰.

Results

After quality control, the total number of shotgun metagenome sequences in the 29 samples was 48,371,589 ranging from 1,046,994 to 2,274,876 per sample. Among these qualified sequences, 9,686,708 sequences, ranging from 205,826 to 461,729 per sample, were successfully annotated with the *e*-value cutoff of 10^{-5} , minimum % identity of 60%, and minimum alignment length cutoff as 15 bp.

Also, after quality control, the total number of 16S rRNA amplicon sequences from the 29 samples was 190,058. The number of sequence reads in each sample ranged from 4,036 to 10,400. The most abundant phylum was Actinobacteria (32.92%) followed by Proteobacteria (28.00%), Bacteroidetes (7.84%), Acidobacteria (5.97%), and Firmicutes (2.87%) (Fig. S1) Jing *et al.*²¹. After subsampling into 4,000 reads, the total number of OTUs in the 29 samples was 35,770 ranging from 1,174 to 2,456 per sample.

Functional gene composition along an aridity gradient. The RDA result of Subsystem Level 3 genes showed that mean annual precipitation was the best predictor for functional gene composition, alone explaining 26.7% of total variation (Fig. 1 and Table S1; results of Subsystem Level 4 genes are shown in Fig. S2 and Table S2). A heatmap of standardized Subsystem Level 1 gene relative abundance along the mean annual precipitation gradient displays gene categories variously having significant positive correlation ($p < 0.05$, Pearson's correlation coefficient was automatically larger than 0.52), negative correlation ($p < 0.05$, Pearson's correlation coefficient was automatically lower than -0.44) or no correlation ($p > 0.05$) (Fig. 2). As mean annual precipitation (MAP) increased, the relative abundance of genes related to cell division and cell cycling, dormancy and sporulation, DNA metabolism, RNA metabolism, protein metabolism, nucleotides and nucleosides, decreased. Relative abundance of genes related to nutrient cycling (sulfur metabolism, potassium metabolism, carbohydrates, nitrogen metabolism, iron acquisition and metabolism, phosphorous metabolism, fatty acids) and metabolism of aromatic compounds increased with increasing MAP. The relative abundance of genes related to motility, regulation and cell signaling and secondary metabolism also increased. Under the regulation and cell signaling gene category, only one Subsystem level 2 gene category, "proteolytic pathway", showed a significant positive correlation in relative abundance with MAP (correlation coefficient of 0.535, $p = 0.003$).

Genes related to osmotic stress (under the stress response category) that have significant correlation in their relative abundance with MAP are shown in a more detail in Table S3. Overall the relative abundance of genes related to osmotic stress (Subsystem Level 2) had significant positive correlation with MAP. The relative abundance of genes related to synthesis of osmoregulatory periplasmic glucans increased as MAP increased, whereas the relative abundance of genes related to production of osmoprotectants ("Betaine biosynthesis from glycine", "ectoine biosynthesis and regulation") and osmoprotectant transporters ("osmoprotectant ABC transporter YehZYXW of Enterobacteriales") decreased. The relative abundance of genes related to protein chaperones also decreased significantly with increasing MAP (correlation coefficient of -0.426 , $p = 0.0213$).

Genes related to oxidative stress (under stress response category) that show significant correlation in their relative abundance with MAP are shown in a more detail in Table S4. Overall, genes related to oxidative stress

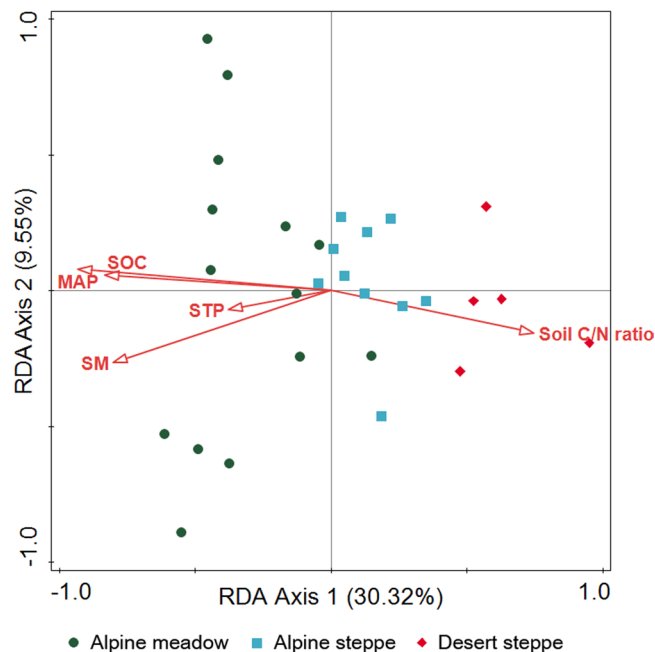


Figure 1. RDA result (Subsystem Level 3) showing sample distribution based on bacterial functional gene composition along the environmental gradient in Tibet. Vegetation types are indicated, with MAP of alpine meadow > alpine steppe > desert steppe. MAP: mean annual precipitation (mm), SM: soil moisture (g/g dried soil), SOC: soil organic carbon (%), STP: soil total phosphorous (%).

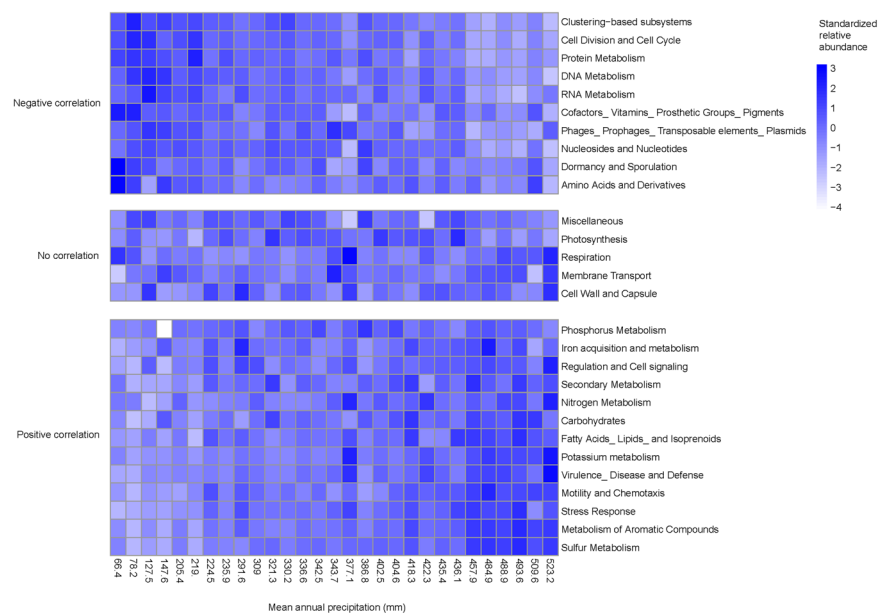


Figure 2. Heatmap of Z-score transformed relative abundance of Subsystem Level 1 genes in samples along the gradient of mean annual precipitation (mm).

(Subsystem Level 2) had a significant positive correlation in relative abundance with MAP. The relative abundance of genes related to antioxidants such as glutathione increased as MAP increased. We searched for taxonomic annotation of those genes that have significant positive correlation in their relative abundance with MAP in alpine meadow samples (where MAP is the highest), and found that *Rhodopseudomonas* and *Bradyrhizobium* were the most abundant genera (Fig. S3). We verified that the relative abundance of these two genera increased along with MAP (*Rhodopseudomonas* had a correlation coefficient of 0.620 and p value of 0.000; *Bradyrhizobium* had a correlation coefficient of 0.638 and p value of 0.000), a trend possibly related to increases in plant root biomass and activity with greater MAP.

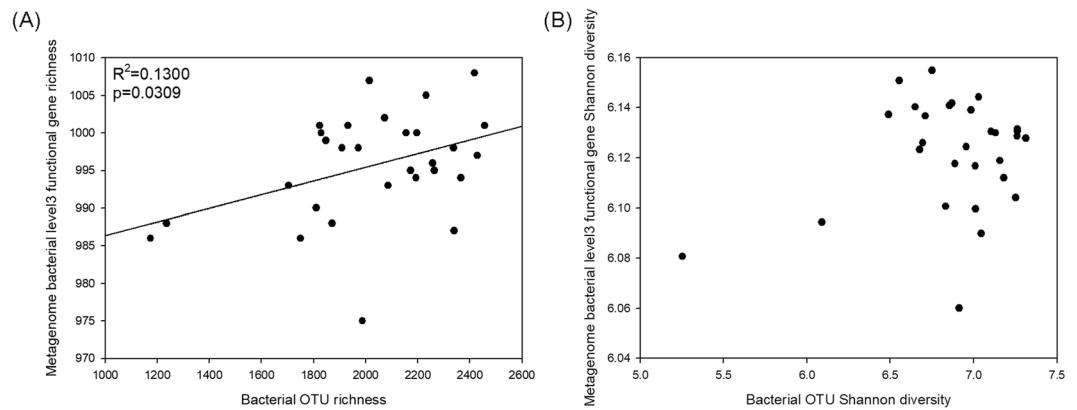


Figure 3. Subsystem Level 3 functional gene richness against bacterial OTU richness. **(A)** Subsystem Level 3 functional gene Shannon diversity against bacterial OTU Shannon diversity. **(B)** Linear regression line was applied only when significant. Presented R-squared value is adjusted R-squared value.

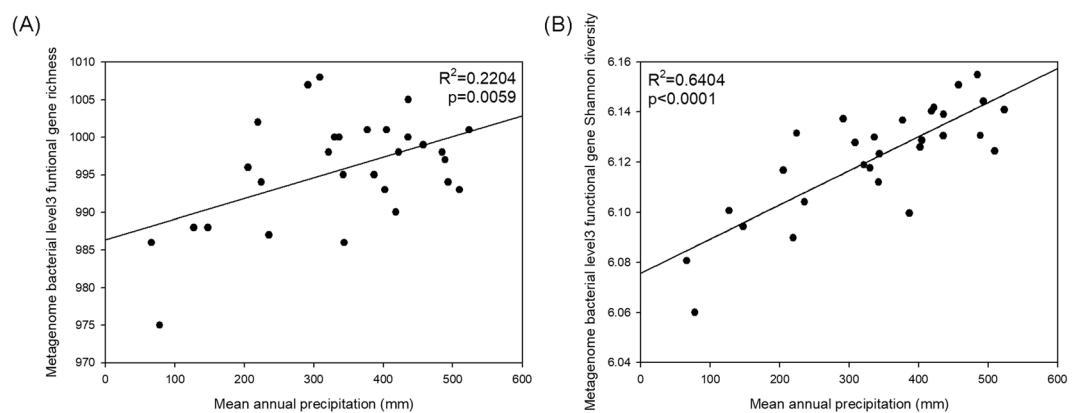


Figure 4. Subsystem Level 3 functional gene richness against mean annual precipitation. **(A)** Subsystem Level 3 functional gene Shannon diversity against mean annual precipitation. **(B)** Linear regression line was applied. Presented R-squared value is adjusted R-squared value.

Since many genes related to antibiotic resistance were grouped together in a very heterogeneous category with genes related to heavy metal resistance under “resistance to antibiotics and toxic compounds (Subsystem Level 2)”, we manually selected genes with annotated labels related to antibiotic resistance and performed correlation analysis. The relative abundance of genes related to antibiotic resistance showed a strong positive correlation with MAP (correlation coefficient = 0.787, $p < 0.0001$)

Functional gene diversity vs OTU diversity. *Alpha diversity.* Subsystem Level 3 functional gene richness increased as bacterial OTU richness increased (Fig. 3-A), but Subsystem Level 3 functional gene Shannon diversity showed no significant correlation with OTU Shannon diversity (Fig. 3-B). Subsystem Level 4 functional gene richness had no significant correlation with bacterial OTU richness (Fig. S4-A) and Subsystem Level 4 functional gene Shannon diversity had no significant correlation with OTU Shannon diversity (Fig. S4-B).

Beta diversity. OTU and functional gene beta diversity between samples calculated based on Bray-Curtis dissimilarity between sample points showed a significant correlation. (Mantel test of 999 permutations, Pearson correlation, for Subsystem Level 3 genes, Mantel static $r = 0.8376$, $p < 0.001$; for Subsystem Level 4 genes, Mantel static $r = 0.8549$, $p < 0.001$).

Functional gene diversity along aridity gradients. *Alpha diversity.* Both richness and Shannon diversity of functional genes decreased towards more arid conditions (Fig. 4; Subsystem Level 4 gene result is shown in Fig. S5). The correlation was stronger than that between OTU richness/Shannon diversity and aridity (Fig. S6). When both OTU richness and MAP were included as explanatory variables in a linear regression model predicting functional gene richness, only MAP was picked as a significant variable (p value for OTU richness: 0.129, p value for MAP: 0.0239).

Since pH dominates many aspects of soil ecology^{1,22}, we also tested whether there is any correlation between pH and OTU/functional gene diversity that might better explain the variation seen along the aridity gradient. The

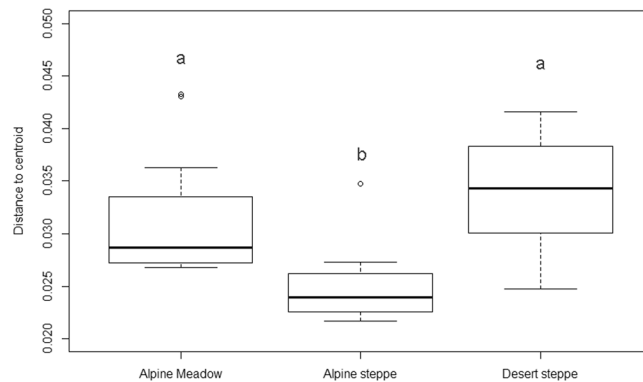


Figure 5. Functional beta diversity of bacterial functional genes, amongst samples from each vegetation type calculated based on Bray-Curtis dissimilarity of Subsystem Level 3 genes from group centroid. Alphabet denotes posthoc test result of Tukey's HSD.

regression analysis of bacterial OTU diversity against pH showed a non-significant result (Fig. S7), but pH was able to explain 22% of variation in functional gene diversity (Fig. S8). Compared to MAP, which explained 64% of variation in functional gene diversity, far less variation was explained by pH. When we put both MAP and pH as explanatory variables in a linear regression model, only MAP was selected as a significant variable (p value of MAP was 8.08×10^{-6} whereas p value of pH was 0.688).

Beta diversity. We grouped samples by vegetation type (which is broadly representative of sections of the aridity gradient) and compared beta diversity calculated based on Bray-Curtis dissimilarity from group centroid of functional gene composition (Fig. 5). In contrast to alpha diversity, although the geographical distance from each group centroid does not differ much between groups (Fig. S9), averaged beta diversity was greatest in desert steppe samples, which had lowest average mean annual precipitation (128 mm) (Fig. 5; Subsystem Level 4 gene result is shown in Fig. S10).

Discussion

In this study, we show in a field based observational study how total functional diversity and the relative abundance of gene categories varies with a broad scale aridity trend. Previously, such differences have been studied mostly by culturing individual bacterial groups isolated from desert environments²³. Our findings mostly parallel those seen in earlier studies of arid environments such as Fierer *et al.*²⁴ - which demonstrated functional differences of microbial community in desert biome compared to non-desert biomes in a worldwide scale - or Tripathi *et al.*²², which compared the functional profile of microbes in arid, semi-arid, Mediterranean, and humid Mediterranean environments in Israel. However, our study is distinguished from these two other studies in scale, geographical location, and sampling methods (e.g. sampling at points scattered along a gradient instead of discrete sampling clusters).

Functional gene diversity against MAP or OTU diversity. As hypothesized, bacterial functional gene diversity decreased along the gradient of increasing aridity. It appears that more arid conditions may indeed constrain the number of traits that are viable and hence retained in the community, as discussed by Kleidon *et al.*³, and Kleidon and Mooney² for plants. Similarly, Mullen-Cooper *et al.*²⁵ found in their study on biocrust that aridity had strong and significantly negative effects on functional diversity, suggesting that a reduced number of functional guilds of microbes are able to survive in arid environments. Our findings appear to strengthen the general principle, applicable to both microorganisms and macroorganisms, that extreme environments constrain functional diversity.

We found that the relationship between bacterial functional gene diversity and aridity (Fig. 4) was much stronger than the relationship between OTU diversity and aridity (Fig. S6). It seems that the bacterial species found towards the moister end of the gradient in alpine meadow or alpine steppe samples may be more specialized ecologically (for example, in having genes related to degradation of complex nutrient sources) and that there may be lower functional redundancy in these samples compared to desert steppe samples. It is widely known that reduced taxonomic diversity does not necessarily correlate with reduced functional diversity due to functional redundancy^{26,27}. It seems that functional redundancy, which indicates that the same ecological function can be carried out by multiple bacterial species²⁶, should be considered as separate from the relationship between taxonomic diversity and functional diversity.

Intriguingly, functional *beta* diversity was greatest in the desert samples. For macro-organisms, there is some evidence that the final community structure in extreme conditions is spatially more uniform than in mesic, productive conditions^{28,29}. This may not be applicable to microbes, since microbes exist in much smaller microsites than macro-organisms. Greater beta-diversity of soil microbes in more arid environments in our study might be derived by a habitat heterogeneity effect, due to stronger influence of substrate rock in an arid environment where weathering has exerted a weaker effect on soil chemistry³⁰. It might also be due to an 'island of fertility' effect whereby locally more resource-rich environments could produce greater patchiness in bacterial community traits^{31,32}.

Functional gene relative abundances along aridity gradient. *The relative abundance of genes related to stress responses will increase along a gradient of increasing aridity.* The overall assemblage of genes related to stress responses showed an unexpected pattern in Fierer *et al.*²⁴ and Tripathi *et al.*³³ because their relative abundance was lower in arid conditions, compared to mesic conditions. We also found in our study that overall relative abundance of genes related to stress responses decreased along a gradient of increasing aridity.

However, when we looked in more detail within these categories down to Subsystem Level 3, we were able to find some reasons for these unexpected patterns. In our study, the relative abundance of the general category of genes related to bacterial stress response (Subsystem Level 1) and genes related to bacterial osmotic stress (Subsystem Level 2) decreased in relative abundance as aridity increased – but within the overall category of genes related to osmotic stress response, this trend was dominated by bacterial genes relating to “Synthesis of osmoregulated periplasmic glucans”, which are produced when there is excessive water³⁴. This might then explain why moister climates were associated with greater relative abundance of these genes, due to increased average soil water content, and the frequency of bursts of very high soil water content. However, the relative abundance of the subcategories of bacterial genes related to production of osmoprotectants such as betain or ectoine increased with increasing aridity as would be expected^{10,11}. Hence, investigating this hypothesis seems to be partly a case of studying the correct subset of stress related genes in relation to aridity gradients. Also, the relative abundance of bacterial genes relating to “protein chaperones” increased with increasing aridity, as expected. In these senses, then, the hypothesis was supported by the results.

However, the relative abundance of genes related to oxidative stress did not show the expected pattern even when we investigated Subsystem Level 3 genes. The relative abundance of core genes related to oxidative stress response (e.g. production of glutathione) was higher in the samples with increased MAP. This might be due to greater abundance of *Bradyrhizobium* and *Rhodopseudomonas* in these samples. These two genera belong to the family Bradyrhizobiae which play a significant role in global nitrogen cycles³⁵. Both genera possess nitrogenase which is a key enzyme in nitrogen fixation and is extremely oxygen sensitive³⁶. There are numerous papers reporting the importance of oxidative stress response of these taxa^{37,38}. This may relate to greater plant root biomass and activity in alpine meadows, as plant roots may support diazotrophs in the rhizosphere^{39,40}.

Also, we found the relative abundance of genes related to dormancy to be higher in more arid conditions in all three published studies that have studied metagenomes in arid and non-arid environments (our study, Fierer *et al.*²⁴, and Tripathi *et al.*³³). Dormancy is a well-known mechanism for microbes to cope with stressful conditions, helping bacterial species to conserve energy until a brief window of time in which to grow, feed and reproduce comes¹⁴. High relative abundances of these genes in the arid conditions found in our study and others place emphasis on the role of dormancy as a survival strategy of bacterial species in extreme conditions.

The relative abundance of genes related to competitive traits will be decrease with increasing aridity whereas the relative abundance of housekeeping genes will increase with increasing aridity. The relative abundance of genes related to competitive traits - for example genes classified as related to “virulence, disease, and defense” (including genes related to antibiotic resistance), and genes related to secondary metabolism or proteolytic pathways - was also found to be higher in the least arid environments in all the three studies. A similar pattern has also been found by Goberna *et al.* (2014)⁴¹. They compared vegetated patches with open unvegetated patches (considered as indicative of stressful conditions) in a semi-arid environment. They found competitive traits to be consistently more abundant in the less stressful, plant-covered patches. Le *et al.*⁴² also found that the types of defense related genes in their hypolith metagenomes clustered more closely with the desert samples of Fierer *et al.*²⁴ rather than their non-desert samples, supporting their expectations. As aridity increases, cell density decreases⁴³, and this may also be expected to decrease selection for competitive traits, by analogy with the stress-selected strategy of Grime for plants¹⁵.

Also, we found the relative abundance of housekeeping genes (genes related to DNA metabolism, RNA metabolism, and protein metabolism, cell division and cell cycle) increased with increasing aridity – though this may simply reflect the lack of the other ‘extra’ genes for competitive traits rather than any particularly strong selection for housekeeping genes themselves. This pattern was also found in Fierer *et al.*²⁴ or Tripathi *et al.*³³. Koo *et al.*²⁰ also reported high abundances of genes related to DNA replication, protein folding, protein biosynthesis, and transcription termination protein in extreme cold conditions. Though it was not discussed in detail by Fierer *et al.*²⁴ or Tripathi *et al.*³³, it seems these genes are needed for bacterial species to maintain steady-state of cellular metabolism, growth, or cell division.

We were also able to find additional interesting trends that had not been hypothesized. For example, that the relative abundance of genes related to nutrient cycling decreased with increasing aridity. High concentrations of nutrients and higher plant biomass in non-arid environments may have selected for a variety of types of nutrient cycling related genes. For instance, aromatic compounds are naturally produced by many macro-organisms, an abundant example being the lignin of plants⁴⁴, and these may have selected for bacterial species carrying genes which can allow them to utilize aromatics as their nutrient source.

Also, we found the relative abundance of genes related to motility decreased with increasing aridity. In plant ecology, the average spatial dispersal distance of seeds in the plant community decreases towards more arid conditions, due to differences in investment in dispersal traits by plants^{16,45}. Favorable growing conditions are ephemeral and rare in the desert landscape, and as a result plant seeds in deserts have evolved to stay close to the mother plant and in a dormant state until rain occurs, rather than disperse across the landscape to find favorable conditions⁴⁵. For bacteria too, a more conservative strategy of remaining in place in favorable microsites may be ecologically selected in an arid environment. The genes involved in motility which varied in their relative abundance along the aridity gradient were mostly those related to flagellae: motility in the bacterial world most often depends upon flagellae, which represent an active form of movement. It is also necessary to bear in mind that as flagellae require water surrounding the cells, these might be of less ecological benefit within a soil that is dry most of the time.

Conclusions

Overall, our findings in this study suggest certain predictable functional patterns in the microbial world, in response to environmental gradients. A striking trend towards decreased functional gene diversity resembles the trends in functional trait diversity seen in macro-organisms such as plants. The trends in relative abundance of particular categories of genes – for example decreased relative abundance of genes associated with competitive traits along aridity gradient – also imply certain basic principles in common between the world of macro-organisms and the world of micro-organisms. It is interesting that most of the trends we found in this study parallel those found by other recent studies on arid-land microbial communities, reinforcing the view that these may be general trends^{24,33}.

As a basis for future work, it is important to recognize that the bacterial functional diversity as recognized here is based only on gene diversity, which is not necessarily the same as the phenotypic functional diversity of the whole bacterial community. For instance, some bacterial OTUs, while present, may be relatively more active⁴⁶ such that their abundances do not reflect their functional importance in the community. Furthermore, some gene functional types may be misassigned by MG-RAST (metagenomics Rapid Annotation using Subsystem Technology)⁴⁷, which is an ongoing project to categorize the huge diversity of naturally occurring genes, and is limited by deducing function from homologues which have already been studied. It would be informative to directly study soil catabolic diversity⁴⁸ and other measures of soil activity, as an additional and more direct way of understanding the full functional trait diversity of these soils and how it varies.

It is also possible that the functional gene profile of these soils is significantly altered by a legacy of ‘old’ DNA from dead or lysed cells⁴⁹. Whether this old DNA negates the relevance of what is seen is a moot point, as it represents a time-integrated view of total community gene diversity rather than a snapshot in time. In any case, it is important to realize that old DNA would only be expected to add to the gene diversity of the desert soils, where DNA may be preserved for longer due to slower turnover rates – yet the more arid soils in our gradient still have lower functional gene diversity. Perhaps, without the signal of old DNA, their gene diversity would be even lower compared to that of moister climate soil. Additionally, since relative abundance of genes in samples was compared, an observed pattern might be due to selection against other genes rather than selection for the particular gene functions that become more abundant in one part of the gradient.

It would also be interesting to study whether other apparent ‘stress’ gradients (e.g. gradients towards in salinity, heat, and concentrations of chemical poisons) are associated with reduced gene functional diversity towards the more extreme end of the gradient. Additionally, a systematic global comparison of metagenomes along broader aridity gradients – expanding on the fairly small number of samples in the preliminary scale study by Fierer *et al.*²⁴ would be rewarding in understanding the true extent and scale of these trends, particularly whether these trends actually differ as a result of idiosyncratic factors from one geographical region to another.

Materials and Methods

Soil sampling and DNA extraction. We collected soil sample from 29 grassland sites across in Tibetan plateau (Fig. S11). We sampled during the growing season in July–August. The survey area encompassed a wide range of climatic and edaphic conditions (mean annual precipitation range from 66 mm to 523 mm, pH range from 6.0 to 9.1, soil total carbon range from 1.12% to 15.22%). Detailed information is provided in Supplementary Tables S5 and S6. Find Supplementary Information provided as a separate document for methodological details used for measurement of each environmental variable. In each site, seven randomly located 0–5 cm depth soil cores (each 5 cm diameter) were collected across a single 1 square meter quadrat (m²). The seven individual soil core samples were placed in one bag and thoroughly mixed to homogenize them – making one sample from which the DNA for a single metagenome was obtained. DNA was extracted from 0.5 g of soil in each sample using a FastDNA Spin kit (Bio 101, Carlsbad, CA, USA), following the manufacturer’s instructions and stored at –40 °C.

16S rRNA gene analysis using amplicon sequencing and sequence processing. Bacterial community composition data were obtained from Jing *et al.*²¹ where the same sample cores we analysed here by whole genome sequencing were used for amplicon sequencing. For assessment of bacterial community composition, V4–V5 regions of bacterial 16S rRNA genes were amplified using universal bacterial primer F515 (5′-GTGCCAGCMGCCGCGG-3′)⁵⁰ and R907 (5′-CCGTC AATTCMTTTRAGTTT-3′)⁵¹. PCR process included 30 cycles of denaturation (94 °C, 30 s), annealing (55 °C, 30 s), and extension (72 °C, 30 s), followed by final extension (72 °C, 10 min). PCR product was purified with agarose gel DNA purification kit (TaKaRa) and sequenced using a Roche FLX454 pyrosequencing machine (Roche Diagnostics Corp., Branford, CT, USA).

Sequences were processed following the Quantitative Insights into Microbial Ecology pipeline (<http://qiime.sourceforge.net/>). Low quality sequences (average quality score less than 25) and sequences with ambiguous characters were removed. Sequences with more than 200-bp length were selected for further analysis. Sequences were aligned against the Greengenes database using the NAST algorithm. Sequences were assigned into OTUs with 97% similarity and classified using the Greengenes database. More details can be found in Jing *et al.*²¹ on bacterial 16S rRNA sequence analysis.

Shotgun metagenome sequencing and sequence processing. Genomic DNA was sheared into 150-bp to 250-bp size fragments and processed for a shot-gun sequencing run. The process included the following steps; end-repair, dA-tailing, adapter ligation and pre-PCR for indexed NGS library. The PCR product were sequenced in Celemics Inc. (Seoul, Korea) with the HiSeq. 2500 platform (Illumina, USA) using 2 × 150 bp paired-end runs. Sequences were uploaded to MG-RAST (metagenomics Rapid Annotation using Subsystem Technology) server for quality control and for annotation⁴⁷ (project accession number: mgp20476). Sequences that had more than 5 low quality (with 15 phred score cutoff) bases were removed. Sequences were annotated

using BLAT (the BLAST-like alignment tool) algorithm⁵² against M5NR database⁵³, which provides nonredundant integration of many databases. Among the integrated databases, we chose Refseq database⁵⁴ for taxonomic annotation of functional genes and SEED subsystem hierarchy⁵⁵ for functional gene annotation. Sequences were annotated using default settings (maximum e-value cutoff of 10^{-5} , minimum % identity cutoff of 60%, and minimum alignment length cutoff as 15 bp). Sequences that failed to be annotated by MG-RAST were discarded for further analysis. About 97% of annotated functional gene sequences on average comprised of bacterial sequences, so we focused only on bacterial functional genes and removed sequences derived from other organisms (archaea, viruses, and eukaryotes).

Statistical analysis. To calculate the relative abundance of bacterial phyla in each sample, we grouped OTUs belonging to the same phylum and summed the number of sequence reads. The number of sequence reads for each phylum was divided by the total number of bacterial sequence reads in each sample. The relative abundance of functional genes was calculated similarly. To calculate the relative abundance of functional gene categories in each sample, we divided the number of sequence reads assigned to a specific functional gene category with the total number of bacterial functional gene sequence reads in each sample that has been annotated by MG-RAST. We used non-subsampled data for calculating relative abundances.

To find functional gene categories which have significant correlation with precipitation gradients, we performed correlation analysis using the relative abundance of each gene. To provide an intuitive visualization of the relationship between MAP and each gene, we drew a heatmap using “pheatmap” package in R⁵⁶ with z-score transformed relative abundance of functional gene categories.

We subsampled our data for whole the other analysis described from now, including RDA and diversity calculations. 16S rRNA sequence reads were randomly subsampled to 4,000 reads per sample. Shot-gun metagenome sequence reads were randomly subsampled to 205,826 reads per sample.

To find the environmental variables that best explained functional gene composition, we performed redundancy analysis (RDA) using Canonco ver. 5 (Biometrics, Wageningen, The Netherlands). We applied forward selection of environmental variables and the Monte Carlo permutation test with 999 random permutations was used for significance test. All of the environmental variables listed in Table S5 were included in the RDA analysis as explanatory variables. Subsampled sequence reads of Subsystem level 3 functional genes were square-root transformed and used as a response variable in the RDA analysis.

To calculate alpha and beta diversity of functional genes we used subsampled sequence reads of Subsystem level 3 functional genes. We focused on Subsystem Level 3 genes to calculate functional diversity, since Level 1 and Level 2 gene categories are more likely to be too broad to be used as a representative indicator of diversity. Also, the finest level (one level below Level 3, which has been called Subsystem Level 4 in this paper) showed very similar results to Level 3, so we only presented our data based on Level 3 functions in our main text and included Level 4 functions as supplementary files.

To calculate beta diversity of OTUs and functional genes, subsampled sequence reads were square root transformed and the Bray-Curtis dissimilarity between each sample point was calculated. To compare beta diversity of OTUs and functional genes, a Mantel test was performed by using “mantel.test” function in R package “vegan”⁵⁷. To compare beta diversity of functional genes in different vegetation types (alpine meadow, alpine steppe, and desert steppe), subsampled Subsystem level 3 functional gene data were square root transformed and Bray-Curtis dissimilarity from group centroid was calculated using “betadisper” function in R package “vegan”⁵⁷. Beta diversity between groups was compared using analysis of variance (ANOVA).

References

1. Fierer, N. Embracing the unknown: disentangling the complexities of the soil microbiome. *Nat Rev Micro* **15**, 579–590, <https://doi.org/10.1038/nrmicro.2017.87> (2017).
2. Kleidon, A. & Mooney, H. A. A global distribution of biodiversity inferred from climatic constraints: results from a process-based modelling study. *Global Change Biology* **6**, 507–523, <https://doi.org/10.1046/j.1365-2486.2000.00332.x> (2000).
3. Kleidon, A., Adams, J., Pavlick, R. & Reu, B. Simulated geographic variations of plant species richness, evenness and abundance using climatic constraints on plant functional diversity. *Environmental Research Letters* **4**, 014007 (2009).
4. Swenson, N. G. *et al.* The biogeography and filtering of woody plant functional diversity in North and South America. *Global Ecology and Biogeography* **21**, 798–808, <https://doi.org/10.1111/j.1466-8238.2011.00727.x> (2012).
5. Maestre, F. T. *et al.* Increasing aridity reduces soil microbial diversity and abundance in global drylands. *Proceedings of the National Academy of Sciences* **112**, 15684–15689, <https://doi.org/10.1073/pnas.1516684112> (2015).
6. Wang, D. *et al.* Precipitation drives the biogeographic distribution of soil fungal community in Inner Mongolian temperate grasslands. *Journal of Soils and Sediments* **18**, 222–228, <https://doi.org/10.1007/s11368-017-1727-z> (2018).
7. Jiang, M., Felzer, B. S. & Nielsen, U. N. Biome-specific climatic space defined by temperature and precipitation predictability. *Global Ecology and Biogeography* **26**, 1270–1282, <https://doi.org/10.1111/geb.12635> (2017).
8. Lauber, C. L., Hamady, M., Knight, R. & Fierer, N. Soil pH as a predictor of soil bacterial community structure at the continental scale: a pyrosequencing-based assessment. *Applied and Environmental Microbiology*, <https://doi.org/10.1128/aem.00335-09> (2009).
9. Nunes, A. *et al.* Which plant traits respond to aridity? A critical step to assess functional diversity in Mediterranean drylands. *Agricultural and Forest Meteorology* **239**, 176–184, <https://doi.org/10.1016/j.agrformet.2017.03.007> (2017).
10. Csonka, L. N. Physiological and genetic responses of bacteria to osmotic stress. *Microbiological Reviews* **53**, 121–147 (1989).
11. Schimel, J., Balsler, T. C. & Wallenstein, M. Microbial stress-response physiology and its implications for ecosystem function. *Ecology* **88**, 1386–1394, <https://doi.org/10.1890/06-0219> (2007).
12. Warren, C. R. Response of osmolytes in soil to drying and rewetting. *Soil Biology and Biochemistry* **70**, 22–32, <https://doi.org/10.1016/j.soilbio.2013.12.008> (2014).
13. Mittler, R. Oxidative stress, antioxidants and stress tolerance. *Trends in Plant Science* **7**, 405–410, [https://doi.org/10.1016/S1360-1385\(02\)02312-9](https://doi.org/10.1016/S1360-1385(02)02312-9) (2002).
14. Schimel, J. P. Life in Dry Soils: Effects of Drought on Soil Microbial Communities and Processes. *Annual Review of Ecology, Evolution, and Systematics* **49**, null, <https://doi.org/10.1146/annurev-ecolsys-110617-062614> (2018).
15. Grime, J. P. *Plant strategies, vegetation processes, and ecosystem properties.* (John Wiley & Sons, 2006).

16. Pueyo, Y., Kefi, S., Alados, C. L. & Rietkerk, M. Dispersal strategies and spatial organization of vegetation in arid ecosystems. *Oikos* **117**, 1522–1532, <https://doi.org/10.1111/j.0030-1299.2008.16735.x> (2008).
17. Huston, M. A General Hypothesis of Species Diversity. *The American Naturalist* **113**, 81–101, <https://doi.org/10.1086/283366> (1979).
18. Brook, I. Bacterial Interference. *Critical Reviews in Microbiology* **25**, 155–172, <https://doi.org/10.1080/10408419991299211> (1999).
19. Hibbing, M. E., Fuqua, C., Parsek, M. R. & Peterson, S. B. Bacterial competition: surviving and thriving in the microbial jungle. *Nature reviews. Microbiology* **8**, 15–25, <https://doi.org/10.1038/nrmicro2259> (2010).
20. Koo, H. *et al.* Metagenomic Analysis of Microbial Community Compositions and Cold-Responsive Stress Genes in Selected Antarctic Lacustrine and Soil Ecosystems. *Life* **8**, 29 (2018).
21. Jing, X. *et al.* The links between ecosystem multifunctionality and above- and belowground biodiversity are mediated by climate. **6**, 8159, <https://doi.org/10.1038/ncomms9159> <https://www.nature.com/articles/ncomms9159#supplementary-information> (2015).
22. Tripathi, B. M. *et al.* Tropical Soil Bacterial Communities in Malaysia: pH Dominates in the Equatorial Tropics Too. *Microbial Ecology* **64**, 474–484, <https://doi.org/10.1007/s00248-012-0028-8> (2012).
23. Heulin, T. *et al.* In *Adaption of Microbial Life to Environmental Extremes: Novel Research Results and Application* (eds Helga Stan-Lotter & Sergiu Fendrihan) 69–85 (Springer Vienna, 2012).
24. Fierer, N. *et al.* Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proceedings of the National Academy of Sciences* **109**, 21390–21395, <https://doi.org/10.1073/pnas.1215210110> (2012).
25. Mallen-Cooper, M., Eldridge, D. J. & Delgado-Baquerizo, M. Livestock grazing and aridity reduce the functional diversity of biocrusts. *Plant and Soil* **429**, 175–185, <https://doi.org/10.1007/s11104-017-3388-5> (2018).
26. Mayfield, M. M. *et al.* What does species richness tell us about functional trait diversity? Predictions and evidence for responses of species and functional trait diversity to land-use change. *Global Ecology and Biogeography* **19**, 423–431, <https://doi.org/10.1111/j.1466-8238.2010.00532.x> (2010).
27. Cadotte, M. W., Carscadden, K. & Mirotchnick, N. Beyond species: functional diversity and the maintenance of ecological processes and services. *Journal of Applied Ecology* **48**, 1079–1087, <https://doi.org/10.1111/j.1365-2664.2011.02048.x> (2011).
28. Chase, J. M. Drought mediates the importance of stochastic community assembly. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 17430–17434, <https://doi.org/10.1073/pnas.0704350104> (2007).
29. Pagaling, E. *et al.* Community history affects the predictability of microbial ecosystem development. *The ISME Journal* **8**, 19–30, <https://doi.org/10.1038/ismej.2013.150> (2014).
30. Shachak, M., Gosz, J., Pickett, S. T. A. & Perevolotsky, A. *Biodiversity in Drylands: Towards a Unified Framework*. (Oxford University Press, 2005).
31. Bolling, J. D. & Walker, L. R. Fertile island development around perennial shrubs across a Mojave desert chronosequence. *Western North American Naturalist* **62**, 88–100 (2002).
32. Gonzalez-Polo, M. & Austin, A. T. Spatial heterogeneity provides organic matter refuges for soil microbial activity in the Patagonian steppe, Argentina. *Soil Biology and Biochemistry* **41**, 1348–1351, <https://doi.org/10.1016/j.soilbio.2009.03.008> (2009).
33. Tripathi, B. M. *et al.* Trends in Taxonomic and Functional Composition of Soil Microbiome Along a Precipitation Gradient in Israel. *Microbial Ecology* **74**, 168–176, <https://doi.org/10.1007/s00248-017-0931-0> (2017).
34. Jean-Pierre, B. & Jean-Marie, L. In *The Periplasm* (ed Ehrmann M) (American Society of Microbiology, 2007).
35. Marcondes de Souza, J. A., Carareto Alves, L. M., de Mello Varani, A. & de Macedo Lemos, E. G. In *The Prokaryotes: Alphaproteobacteria and Betaproteobacteria* (eds Eugene Rosenberg *et al.*) 135–154 (Springer Berlin Heidelberg, 2014).
36. Gallon, J. R. Reconciling the incompatible: N₂ fixation and O₂. *New Phytologist* **122**, 571–609, <https://doi.org/10.1111/j.1469-8137.1992.tb00087.x> (1992).
37. Oda, Y. *et al.* Functional Genomic Analysis of Three Nitrogenase Isozymes in the Photosynthetic Bacterium *Rhodospseudomonas palustris*. *Journal of Bacteriology* **187**, 7784–7794, <https://doi.org/10.1128/jb.187.22.7784-7794.2005> (2005).
38. Masloboeva, N. *et al.* Reactive Oxygen Species-Inducible ECF σ Factors of *Bradyrhizobium japonicum*. *PLoS ONE* **7**, e43421, <https://doi.org/10.1371/journal.pone.0043421> (2012).
39. Santi, C., Bogusz, D. & Franche, C. Biological nitrogen fixation in non-legume plants. *Annals of Botany* **111**, 743–767, <https://doi.org/10.1093/aob/mct048> (2013).
40. Cocking, E. C. Endophytic colonization of plant roots by nitrogen-fixing bacteria. *Plant and Soil* **252**, 169–175, <https://doi.org/10.1023/a:1024106605806> (2003).
41. Goberna, M., Navarro-Cano, J. A., Valiente-Banuet, A., García, C. & Verdú, M. Abiotic stress tolerance and competition-related traits underlie phylogenetic clustering in soil bacterial communities. *Ecology Letters* **17**, 1191–1201, <https://doi.org/10.1111/ele.12341> (2014).
42. Le, P. T. *et al.* Comparative Metagenomic Analysis Reveals Mechanisms for Stress Response in Hypoliths from Extreme Hyperarid Deserts. *Genome Biology and Evolution* **8**, 2737–2747, <https://doi.org/10.1093/gbe/evw189> (2016).
43. Clark, J. S., Campbell, J. H., Grizzle, H., Acosta-Martinez, V. & Zak, J. C. Soil Microbial Community Response to Drought and Precipitation Variability in the Chihuahuan Desert. *Microbial Ecology* **57**, 248–260, <https://doi.org/10.1007/s00248-008-9475-7> (2009).
44. Masai, E., Katayama, Y. & Fukuda, M. Genetic and biochemical investigations on bacterial catabolic pathways for lignin-derived aromatic compounds. *Bioscience, biotechnology, and biochemistry* **71**, 1–15 (2007).
45. Venable, D. L., Flores-Martinez, A., Muller-Landau, H. C., Barron-Gafford, G. & Becerra, J. X. Seed dispersal of desert annuals. *Ecology* **89**, 2218–2227, <https://doi.org/10.1890/07-0386.1> (2008).
46. Blazewicz, S. J., Barnard, R. L., Daly, R. A. & Firestone, M. K. Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses. *The ISME Journal* **7**, 2061–2068, <https://doi.org/10.1038/ismej.2013.102> (2013).
47. Meyer, F. *et al.* The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9**, 386, <https://doi.org/10.1186/1471-2105-9-386> (2008).
48. Degens, B. P., Schipper, L. A., Sparling, G. P. & Duncan, L. C. Is the microbial community in a soil with reduced catabolic diversity less resistant to stress or disturbance? *Soil Biology and Biochemistry* **33**, 1143–1153, [https://doi.org/10.1016/S0038-0717\(01\)00018-9](https://doi.org/10.1016/S0038-0717(01)00018-9) (2001).
49. Carini, P. *et al.* Relic DNA is abundant in soil and obscures estimates of soil microbial diversity. **2**, 16242, <https://doi.org/10.1038/nmicrobiol.2016.242> <https://www.nature.com/articles/nmicrobiol2016242#supplementary-information> (2016).
50. Caporaso, J. G. *et al.* Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences* **108**, 4516–4522, <https://doi.org/10.1073/pnas.1000080107> (2011).
51. Muyzer, G., Teske, A., Wirsén, C. O. & Jannasch, H. W. Phylogenetic relationships of *Thiomicrospira* species and their identification in deep-sea hydrothermal vent samples by denaturing gradient gel electrophoresis of 16S rDNA fragments. *Archives of Microbiology* **164**, 165–172, <https://doi.org/10.1007/bf02529967> (1995).
52. Kent, W. J. BLAT—the BLAST-like alignment tool. *Genome research* **12**, 656–664 (2002).
53. Wilke, A. *et al.* The M5nr: a novel non-redundant database containing protein sequences and annotations from multiple sources and associated tools. *BMC Bioinformatics* **13**, 141, <https://doi.org/10.1186/1471-2105-13-141> (2012).
54. Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research* **35**, D61–D65, <https://doi.org/10.1093/nar/gkl842> (2007).

55. Overbeek, R. *et al.* The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Research* **33**, 5691–5702, <https://doi.org/10.1093/nar/gki866> (2005).
56. Kolde, R. & Kolde, M. R. Package 'pheatmap'. (2018).
57. Oksanen, J. *et al.* Package 'vegan'. *Community ecology package, version 2* (2013).

Acknowledgements

We thank Xin Jing, Litong Chen, Ke Zhao, Xiaoxia Yang, Kaoping Zhang, Congcong Shen, Huaibo Sun and Xingjia Xiang for assistance in soil sampling and lab analyses. This work was supported by the National Program on Key Basic Research Project [No. 2014CB954002] and the National Research Foundation of Korea (NRF) grant [No. NRF-0409-20160032].

Author Contributions

Initial concept of project: J.M.A. Study design: J.-S.H., H.C., J.M.A. Soil surveys, field sampling, and DNA extraction: Y.S., T.Y. Data analysis: H.-K.S. Study monitoring: H.-K.S., P.J., H.C., J.M.A. Metagenome sequencing: H.K. Manuscript preparation: H.-K.S., J.M.A.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-37565-9>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019