

SCIENTIFIC REPORTS



OPEN

Distribution of Bathyarchaeota Communities Across Different Terrestrial Settings and Their Potential Ecological Functions

Received: 17 October 2016

Accepted: 17 February 2017

Published: 21 March 2017

Xing Xiang^{1,*}, Ruicheng Wang^{1,*}, Hongmei Wang^{1,2}, Linfeng Gong³, Baiying Man¹ & Ying Xu¹

High abundance and widespread distribution of the archaeal phylum Bathyarchaeota in marine environment have been recognized recently, but knowledge about Bathyarchaeota in terrestrial settings and their correlation with environmental parameters is fairly limited. Here we reported the abundance of Bathyarchaeota members across different ecosystems and their correlation with environmental factors by constructing 16S rRNA clone libraries of peat from the Dajiuhu Peatland, coupling with bioinformatics analysis of 16S rRNA data available to date in NCBI database. In total, 1456 Bathyarchaeota sequences from 28 sites were subjected to UniFrac analysis based on phylogenetic distance and multivariate regression tree analysis of taxonomy. Both phylogenetic and taxon-based approaches showed that salinity, total organic carbon and temperature significantly influenced the distribution of Bathyarchaeota across different terrestrial habitats. By applying the ecological concept of 'indicator species', we identify 9 indicator groups among the 6 habitats with the most in the estuary sediments. Network analysis showed that members of Bathyarchaeota formed the "backbone" of archaeal community and often co-occurred with Methanomicrobia. These results suggest that Bathyarchaeota may play an important ecological role within archaeal communities via a potential symbiotic association with Methanomicrobia. Our results shed light on understanding of the biogeography, potential functions of Bathyarchaeota and environment conditions that influence Bathyarchaea distribution in terrestrial settings.

Many uncultured archaeal groups have been unveiled by culture-independent methods, such as the Marine Benthic Group D (MBGD), Miscellaneous Crenarchaeota Group (MCG, recently assigned into a novel archaeal phylum Bathyarchaeota) and Terrestrial Miscellaneous Euryarchaeota Group (TMEG). Since the first observation of MCG via DNA sequencing in 2003¹, MCG have been detected in different marine^{2–4} and terrestrial^{5–7} environments and dominated in various sediments^{8,9} and soils^{10,11} as indicated by 16S rRNA clone libraries.

Although MCG has been detected in various ecosystems, it is still difficult to elucidate their physiological functions and ecological roles due to their slow growth rate and lack of pure cultures. The detection of archaeal Bacteriochlorophyll synthase gene suggested that the member of MCG may be able to fix CO₂ via photosynthesis¹². However, evidence about their heterotrophic life is emerging recently. For instance, MCG has been shown to be able to assimilate ¹³C-labeled acetate¹³ and sedimentary organic carbon¹⁴ as indicated by DNA-SIP and bulk isotope of intact membrane lipids, respectively. Another line of evidence of its heterotrophic lifestyle is that detection of the gene encoding extracellular protein-degrading enzyme by single-cell sequencing¹⁵. Besides, some assembled genomes of MCG showed that they may be capable of hydrolyzing extracellular plant-derived carbohydrates¹⁶, utilizing various methylated compounds for methylotrophic methanogenesis¹⁷ and fermenting organic substrates for acetogenesis¹⁸. Furthermore, the expression of the gene involved in degradation of aromatic compounds in MCG has been demonstrated with protocatechuate as the substrate¹⁹.

¹State Key Laboratory of Biogeology and Environmental Geology, China University of Geosciences, Wuhan, 430074, China. ²Laboratory of Basin Hydrology and Wetland Eco-restoration, China University of Geosciences, Wuhan, 430074, China. ³State Key Laboratory Breeding Base of Marine Genetic Resources, Key Laboratory of Marine Genetic Resources, Third Institute of Oceanography, SOA, Key Laboratory of Marine Genetic Resources of Fujian Province, Xiamen, 361005, China. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to H.W. (email: wanghmei04@163.com)

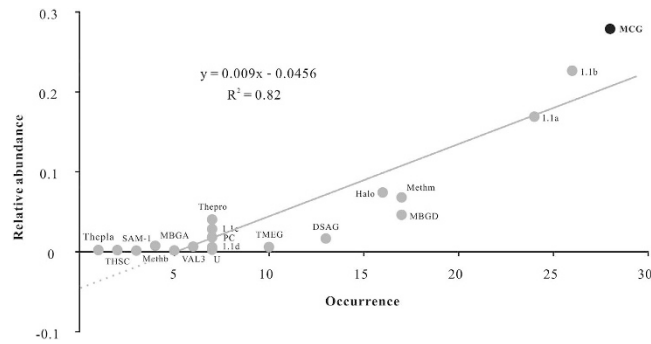


Figure 1. Species abundance distribution (SAD) pattern of archaeal lineages in the libraries analyzed.

Abbreviations: 1.1a: Thaumarchaeota 1.1a; 1.1b: Thaumarchaeota 1.1b; 1.1c: Thaumarchaeota 1.1c; 1.1d: Thaumarchaeota 1.1d; DSAG: Deep Sea Archaeal Group; Halo: Halobacteria; MBGA: Marine Benthic Group A; MBGD: Marine Benthic Group D; MCG: Miscellaneous Crenarchaeota Group; Methb: Methanobacteria; Methm: Methanomicrobia; PC: Peat Crenarchaeota; SAM-1: South African Gold Mine Euryarchaeota Group; Thpfla: Thermoplasmata; Thepro: Thermoprotei; THSC: Terrestrial Hot Spring Crenarchaeota group; TMEG: Terrestrial Miscellaneous Euryarchaeota Group; VAL3: Val Kotinen lake clade III group; U: unclassified archaeal group.

Phylogenetically, MCG is rather divergent. Totally 21 subgroups were identified and the similarity between the most distant sequences is 75%²⁰. Indicator value analysis indicates that some MCG subgroups appear to adapt well to specific environments, such as subgroups 1, 3 and 8 to saline sediments, subgroups 5a, 5b, 7, 9 and 11 to freshwater sediments²⁰. Moreover, salinity is demonstrated to be the main factor controlling the adaptive evolution of subgroups as indicated by ancestor state reconstruction analysis²⁰.

Despite these advances, the geographic distribution pattern of MCG subgroups among terrestrial habitats remains largely unknown. One reason is that, systematic and comprehensive evaluation about abundance and diversity of MCG subgroups in terrestrial environments is still sparse. On the other hand, multiple factors may affect the structures of MCG communities which make it difficult to clearly picture the general distribution pattern of MCG on large scale. For example, abundant and diverse MCG lineage is detected in acidic and neutral coastal sediments²¹, contrarily only a small amount of MCG is observed in alkaline lagoon sediments⁷. Temperature also impacted the abundance of MCG as indicated by mesocosm experiments with high arctic peat soils which showed an increase of temperature will decrease the relative abundance of MCG members²².

With the accumulation of DNA sequencing data from various terrestrial environments, it is highly possible to further understand the distribution pattern of MCG on large scale via bioinformatics analysis. Thus to fully address the distribution pattern of MCG on large ecological scale, archaeal clone libraries were firstly constructed with peat sediments from the acidic Dajiuhe Peatland, central China. Subsequently phylogenetic analysis was conducted by pooling sequences from other terrestrial locations that were available on July 2015. Our results provide new insights into overview on the distribution pattern of MCG across terrestrial ecosystems, their correlation with environmental conditions and their ecological functions.

Results

Widespread distribution of MCG lineages. A total of 890 archaeal OTUs (with 97% cutoff) from 51 sites (Supplementary Table S1) around the world were classified into 19 lineages (Supplementary Fig. S1), whose mean relative abundance was plotted against their occurrence (number of sites in which given lineage was found) (Fig. 1). A significant positive correlation ($R^2 = 0.82$) between mean relative abundance and occurrence (Fig. 1) suggested that broadly-occurred archaeal lineages are more abundant than those with restricted occurrence. MCG was the most frequently observed and abundant archaeal group in terrestrial environments. It was detected in 28 sites and accounted for 27.8% of the total archaeal sequences (Fig. 1) with a relative abundance of 0.7–82.1% among those sites.

Diversity of MCG lineage. The numbers of clone libraries and MCG 16S rRNA sequences varied from one site to another (Table 1). Specifically the number of clone libraries in lagoon was the highest, whilst the number of MCG 16S rRNA sequences was highest in peat sediments. Both the numbers of clone libraries and sequences were the lowest in samples from hot springs (Table 1). However, little difference was observed in Shannon-Wiener indices of MCG among peat sediments, mangrove soils and estuary sediments. The lowest PD index was observed within soils and the highest was in mangrove soils (Table 1).

Different from the biodiversity indexes of Shannon-Wiener and PD, PSV quantifies the phylogenetic relatedness of all species in a community. Observed PSV (Table 1) was remarkably different among habitats. The highest value of PSV in soils indicated a distant phylogenetic relatedness which means a phylogenetic overdispersion. Low PSV values in hot springs, estuary sediments and mangroves suggested a close phylogenetic relatedness which means a phylogenetic clustering. Overall, the mean observed PSV value (0.56) was significantly lower than those given by two null hypothesis models Null 1 (0.64, $p < 0.05$) and Null 2 (0.61, $p < 0.05$). These results indicated that coexisting taxa within habitats were more closely related (phylogenetic clustering) than expected by chance.

Habitat	Nib	N	Shannon-Wiener	PD	PSV
Peat sediment	4	472	3.42 ± 0.61	1.76 ± 0.47	0.56 ± 0.13
Soil ^a	2	130	1.12	0.21	0.86
Hot spring	2	7	2.69	0.30	0.36
Lagoon	13	217	1.87 ± 1.19	0.59 ± 0.47	0.61 ± 0.21
Mangrove soil	4	433	3.68 ± 0.77	2.96 ± 1.08	0.48 ± 0.02
Estuary sediment	3	197	3.56 ± 0.23	2.45 ± 0.40	0.47 ± 0.05

Table 1. Diversity index of MCG lineages in different habitats. Nib: number of libraries or studying sites. N: number of clones. ^aLibraries containing only one OTU were excluded from the calculation of the mean biodiversity index.

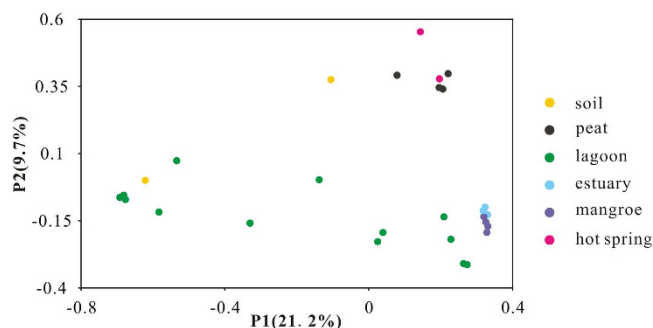


Figure 2. Principal coordinate analysis (PCoA) of 28 MCG clone libraries based on UniFrac-weighted distance matrix.

Beta diversity of these clone libraries were analyzed with weighted PCoA according to phylogenetic community similarity. Principle coordination 1 (P1) and 2 (P2) explained 21.2% and 9.7% of the total variance, respectively (Fig. 2). MCG communities showed a high specificity to habitats, samples from a specific habitat clustered together well except those from lagoons with a scattering distribution. The high specificity was further strongly supported by permutation manova test ($R^2 = 0.53$, $P < 0.01$). It is interesting to see that phylogenetic community composition of estuary sediments and mangrove soils, peat sediments and hot springs were close related to each other, respectively (Fig. 2).

Phylogenetic analysis and community compositions of MCG lineages. Totally 1456 MCG sequences (353 OTUs, at the threshold of 97% similarity) were used to construct MCG phylogenetic tree (Fig. 3). The MCG phylogenetic tree was classified into 18 subgroups with intergroup identity ranging from 82% to 96%. Notably, new cluster peat MCG (pMCG) displayed high stability with a bootstrap value of 100%. However, about 1.2% of MCG sequences (17 sequences) still remained unclassified (uMCG).

As indicated by phylogenetic analysis (Fig. 3), MCG 6 and MCG 5b dominated the MCG lineages with a relative abundance of 38.3% and 20.0%, respectively, followed by MCG 8, MCG 17 and MCG 15. The remaining 13 subgroups contributed <3% of the total sequences (Supplementary Table S2). Overall, MCG communities showed high similarities both in group number and the dominant groups in a specific habitat (Fig. 4). For example, subgroup number ranged from 4 to 6 and the dominant subgroup was MCG 5b in all peat samples (Fig. 4). However, both the number of MCG subgroups and community composition varied significantly in estuaries. Totally 16 subgroups have been detected and MCG 8 and MCG 17 dominated estuarine MCG communities with a relative abundance of 25.6% and 23.1%, respectively. Fourteen MCG subgroups were observed in mangrove soils. Among the MCG subgroups, subgroup 8 and 15 were detected from various environments such as lagoons, estuary sediments and mangroves. In addition, MCG subgroup 6 was the sole subgroup detected in soil samples.

Indicator value (IndVal) function analysis identified 9 indicator groups from all habitats investigated (Fig. 4). Estuary sediments had 4 indicator groups, followed by peat sediments (2), mangrove soils (1), hot springs (1) and soils (1). No indicator group was observed in lagoon samples (Fig. 4). Furthermore, MCG subgroup 3, 5a, 5b, 13 and peat MCG with high IndVal values (0.83–1.0) suggested a high specificity to their habitats. The other indicator MCG subgroups showed a moderate specificity to their habitats with IndVal values of 0.43–0.67. IndVal value of MCG 6 was rather low (0.43) due to its extensive distribution in various environments. In contrast, peat MCG lineage was exclusively detected in peat habitat with a relative abundance of 2.5% with the highest IndVal value of 1.0 (Fig. 4).

Relationship between MCG community structure and environmental factors. We tested the relationship between MCG phylogenetic communities and environmental factors including pH, salinity, temperature, TOC and TN using permutational analysis of variation. Results indicated that salinity was the strongest environmental force ($R^2 = 0.39$, $P < 0.01$), which explains 39% of variation of MCG assemblages. Temperature

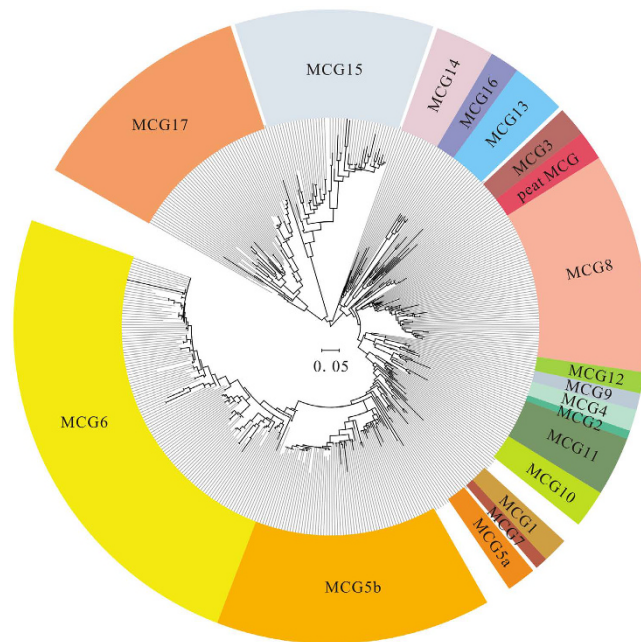


Figure 3. Phylogenetic tree of MCG 16S rRNA genes from different environments was constructed by Maximum Likelihood method under the optimal GTR + G mode. The scale bar represented 5% estimated sequence divergence. The colored leaves presented different MCG subgroups, and uncolored leaves showed unclassified sequences.

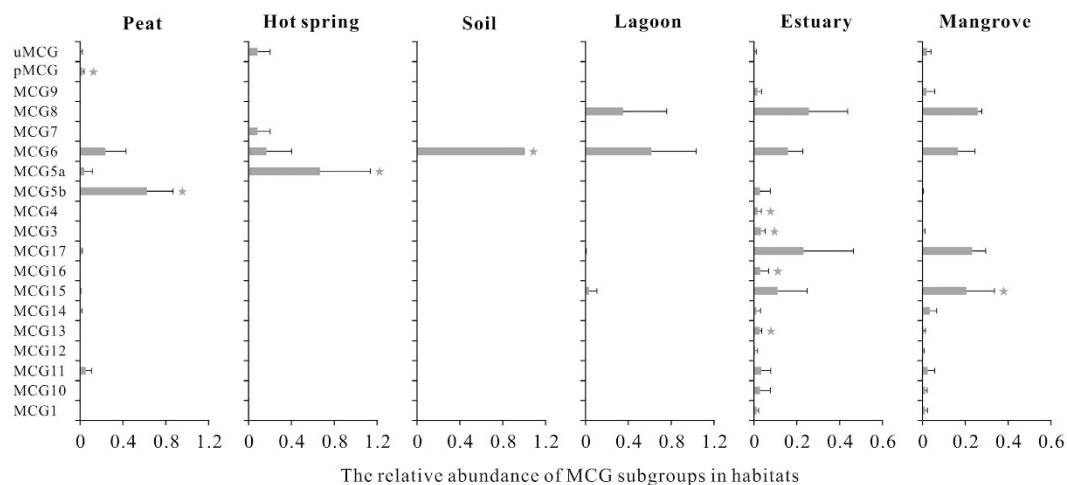


Figure 4. Mean relative abundance of MCG subgroups within different habitats. The error bar represented variance, and asterisks showed indicator groups with a significant threshold of $p = 0.05$.

($R^2 = 0.10$, $P = 0.02$) and TOC ($R^2 = 0.08$, $P = 0.02$) also had significant influence on MCG community structure, accounting for 10% and 8% of the variation, respectively. The remaining environmental factors did not show significant impact on MCG groups and contributed to only 6% total variance of UniFrac matrix.

The effects of the environment factors on the community structure of MCG subgroups were determined by multivariate regression tree (MRT) analysis. The variance was shown in a 5-leaf tree mainly based on TOC content, followed by salinity and temperature, which can explain 77% of the variance in total (Fig. 5). Samples collected from peat sediments were grouped together by high TOC contents in the sediments, and estuary sediment samples were clustered together by their similar salinity. However, samples from lagoons were split into two groups as large variation of salinity among these sites (Fig. 5 and Supplementary Table S1). In addition, samples collected from soils and hot springs were clustered with samples from low salinity lagoon and mangrove, respectively, which were affected by temperature (Fig. 5).

The pie charts in MRT showed the relative abundance of MCG subgroups in each leaf. Each leaf was specified by indicator groups identified by IndVal index, which were mainly responsible for separation of the topology of the MRT. For example, MCG subgroup 5b and pMCG lead to the separation of MCG in peat sediments and in

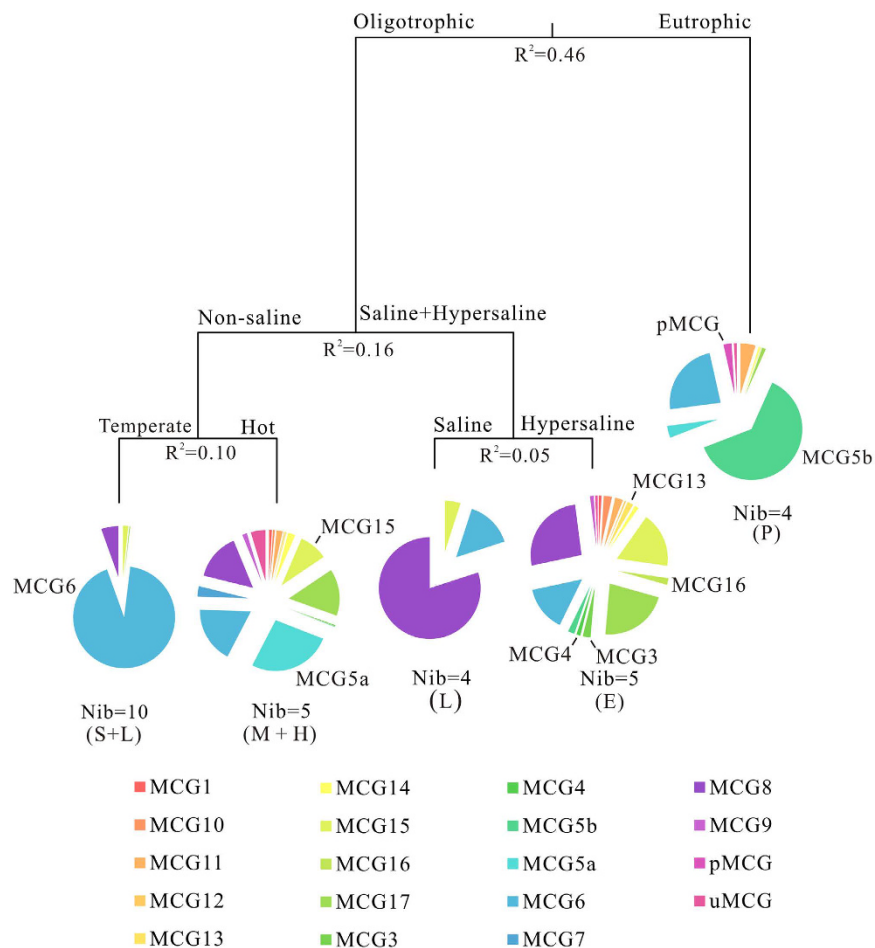


Figure 5. Multivariate regression tree analysis of correlation between MCG subgroups and environmental factors. The pie charts indicated composition and relative abundance of each MCG subgroup in each leaf. Letters in parenthesis displayed habitat types for each leaf: P: peatland; E: estuary; L: lagoon; S: soil; M: mangrove; H: hot spring; Nib: number of libraries or studying sites.

other samples. Similarly, MCG subgroups 3, 4, 13 and 16 resulted in the splitting of MCG in estuary sediments and samples in lagoons with high salinity.

Interrelation between MCG and other archaeal lineages. A total of 99 archaeal OTUs from 51 sites with a cutoff of 90% were subjected to the co-occurrence network interface analysis to explore potential commensalism that MCG may involve in. The network of terrestrial archaea consisted of 88 nodes and 269 edges among which 40% (35 out of 88) nodes and 68% (184 out of 269) edges were composed by MCG (Fig. 6a). The network topology of terrestrial archaea was characterized by a diameter of 5, average path length of 1.62 and 437 shortest paths. The individual node with top 10 network indices were affiliated with MCG, Halobacteria, 1.1b, peat Crenarchaeota and Methanomicrobia, respectively (Table 2). Furthermore, these network indices of MCG nodes were much higher than those of others archaeal groups, particular for centrality and degree. These results suggested that MCG lineages possess pivotal ecological function among archaeal groups.

The network showed modular structure with a modular index 0.67. The network was classified into 11 modules (Fig. 6b). Modules could be considered as ecological niches (Fisher's test, $P < 0.001$) in which OTUs were closely connected to each other and connectivity was significantly higher than those between modules. For example, modules 2, 5 and 8 represented sub-networks of peat sediments, hot springs and saline lakes, respectively. MCG was observed in 5 out of the 11 modules (Fig. 6b). Co-occurrence association was observed between MCG and other archaeal groups such as DSAG, MBGD, peat Crenarchaeota, Thermoprotei, and VAL3 (Fig. 6c). It should be noted that, the co-occurrence of MCG and Methanomicrobia were detected in 3 modules out of 5 containing MCG modules indicating a rather close ecological association (Fig. 6c). Members of Methanomicrobia which co-occurred with MCG were affiliated with *Methanosaeta*, *Methanoregula*, *Methanocella*, *Methanoperedens* and uncultured Methanomicrobiaceae.

Discussion

MCG is a ubiquitous archaeal group in terrestrial environments. It occurred in 28 of 51 sites investigated and accounted for 27.8% of total archaeal sequences analyzed. MCG widely distributed across various terrestrial

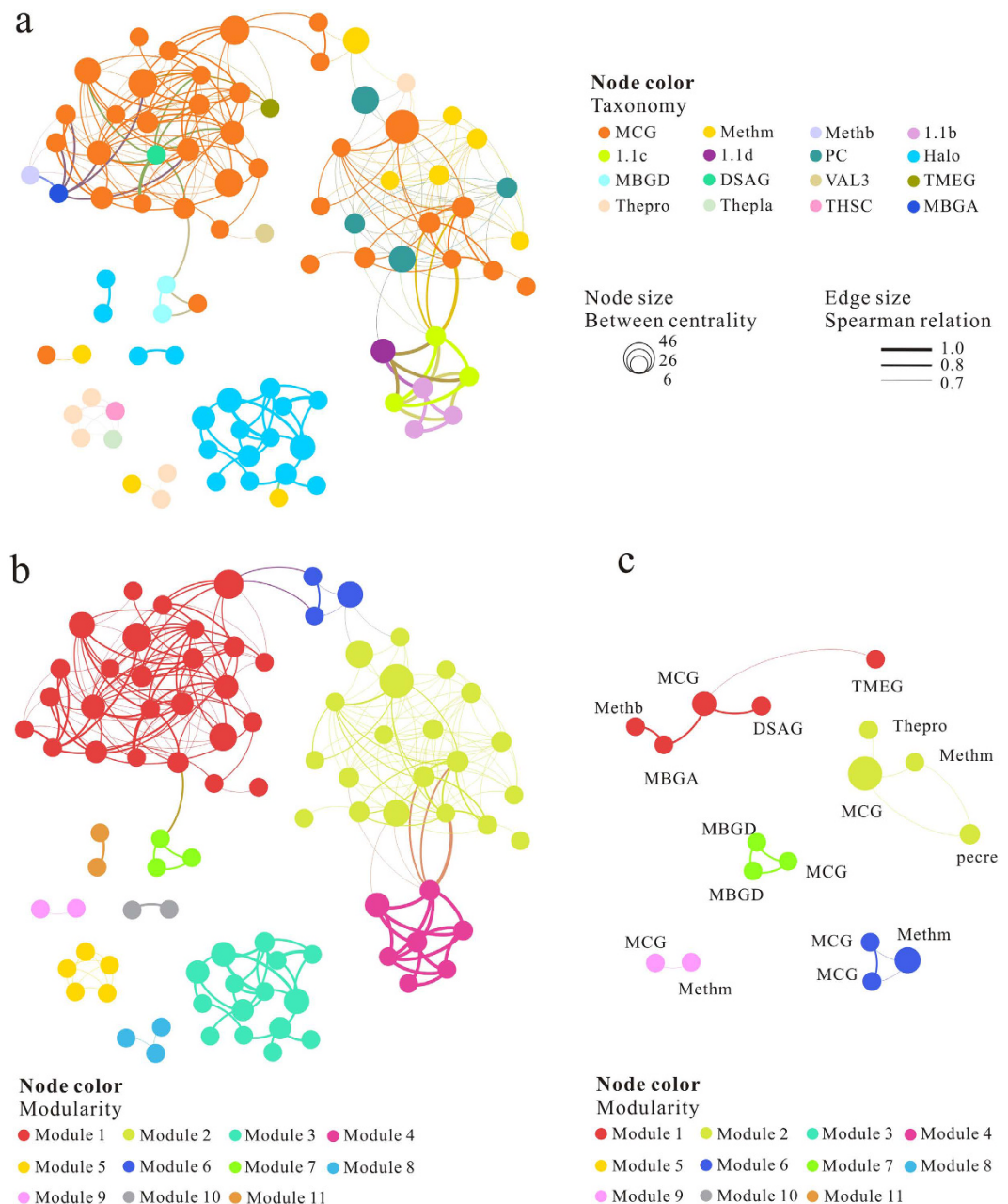


Figure 6. Network analysis of archaeal OTUs in terrestrial settings based on Spearman correlation. Nodes represented archaeal OTUs (90% cutoff) and the node size is proportional to between centrality. Edges between nodes showed significant ($P < 0.01$) Spearman correlation. (a) Nodes were colored by archaeal taxonomy, (b) Same network as a, but the node was dyed according to modularity. (c) Sub-network of OTU clustering by modularity, only sub-network related with MCG were presented.

habitats including hot springs, soils and lagoons with the relative abundance of MCG ranging from 0.7% to 80.0%. Furthermore MCG dominated the archaeal communities in peat sediments, mangroves and estuary sediments. Compared with previous studies, which found that MCG accounted for 35–80% of archaeal communities in peat sediments⁸, 10–100% in high mountain lake²³, and 43% in southern red soil¹¹, our result provided an overall view of terrestrial distribution of MCG group on a large scale. Wide distribution and relatively high abundance strongly suggest that MCG is an important archaeal group in terrestrial ecological process.

Despite of ubiquitous distribution of MCG in terrestrial environment, some MCG subgroups show well-adapted and high specificity to certain habitats, indicating a close relationship between MCG and environmental conditions. This made them suitable to serve as ecological indicator for specific habitats. In this study, nine indicator subgroups were identified based on indicator value function analysis, and at least one indicator subgroup was observed in most habitats investigated except lagoons.

Lineages	Closeness centrality	Lineages	Betweenness centrality	Lineages	Degree
Halo	3.50	MCG	46.0	MCG	20
MCG	3.27	MCG	32.0	MCG	18
MCG	3.18	MCG	30.0	MCG	14
Halo	2.86	MCG	27.0	MCG	14
1.1b	2.73	PC	27.0	MCG	14
Halo	2.60	PC	24.2	MCG	14
MCG	2.50	MCG	22.2	MCG	14
1.1c	2.50	Methm	22.0	Methm	14
Halo	2.50	Halo	20.5	PC	13
MCG	2.43	Halo	18.0	MCG	13

Table 2. Network indices of the top 10 nodes in the network. Abbreviations are the same as listed for Fig. 1.

Previously, salinity has been shown to be a driving force that shapes archaeal communities both within a specific estuary habitat²⁴ and on large-scale aquatic habitats²⁵. The phylogenetic communities of MCG also showed a close relationship with salinity which significantly architects the composition and abundance of MCG as indicated by MRT. In addition to that, salinity was also found to be an evolutionary barrier that leads to the diversification of MCG between marine and freshwater sediments²⁰. This is highly consistent with our observation that salinity was the most significant factor affecting MCG phylogenetic structure.

It is demonstrated that TOC greatly constrains the distribution and diversity of heterotrophs due to their high dependence on TOC for carbon source^{26,27}. It should be noted that in our research TOC also significantly shaped the distribution of MCG on a large scale. Even through the knowledge about MCG life style is still limited, several lines of evidence support the heterotrophic lifestyle of MCG recently^{13,15,18}. Our observation of the significant impact of TOC on MCG distribution may mount another line of heterotrophic lifestyle of MCG²⁸.

Temperature is also an important factor controlling microbial communities. For example, it has been found to be a dominant factor that drives microbial diversity and composition in geothermal springs²⁹. Importantly, increase of temperature will decrease the relative abundance of MCG as indicated by both laboratory experiment²² and field investigation⁷. Our study also demonstrated the impact of temperature on MCG communities to certain extent.

Compared with restricted distribution of Thermoprotei in hyperthermal environments, the wide distribution of MCG across terrestrial environments indicates that MCG have adapted to various environments and potentially have versatile metabolic pathways for survival. Co-occurrence network analysis indicates the dominance of MCG in nodes, edges and modules (Fig. 6b). Moreover, individual nodes with top 10 highest values of network indices are mainly affiliated with MCG. All of these suggest that MCG is the most abundant and widespread archaeal group and plays an important role in construction of archaeal network. Likely, MCG serve as “keystone species” which functions to maintain the stability and adaptability of archaeal community³⁰.

Furthermore, the co-occurrence pattern also reveals some of interesting ecological patterns among archaeal divisions which have not been well revealed previously. For example, MCG co-occurred with the members of 1.1a, 1.1b, DSAG, MBGD, peat Crenarchaeota, Thermoprotei, and VAL3. These co-occurrences suggest a preference of similar environmental conditions of these archaeal groups or a commensalism relationship between them. However, the currently available data do not permit us to distinguish these two possibilities at this point.

Notably, members of MCG co-occur more often with the Methanomicrobia (including *Methanosaeta*, *Methanoregula*, *Methanocella*, *Methanoperedens* and Methanomicrobiaceae) than with other archaeal groups, which indicates a high possibility of a syntrophic association between these two groups. We proposed that acetate may be response for this close association between those two groups. Acetate has been demonstrated to be important organic substrate for growth and methane production of acetoclastic *Methanosaeta*³¹. Meanwhile, acetate is also a necessary growth factor for hydrogenotrophic *Methanoregula*³², *Methanocella*³³ and Methanomicrobiaceae³⁴. Importantly, some MCG members were demonstrated to possess a gene encoding acetate kinase¹⁸, which is involved in acetate production. This may explain the mechanism of symbiosis between MCG and Methanomicrobia. In fact the close physical association between the members of MCG and Methanomicrobia (*Methanosaeta*) has been demonstrated by fluorescence *in situ* hybridization (FISH) in active sludge³⁵.

Due to the high specificity to a given environment, indicator MCG subgroups may play important roles in biogeochemical processes under specific conditions. For example, MCG subgroup 15 (the indicator group of mangrove soils) showed a high abundance in coast and marine environments and their genomic data indicated that they are highly capable of extracellular protein degradation, acetate-producing and dissimilatory nitrite reduction¹⁶. MCG subgroup 6 (indicator group for soils) may also capable of degrading plant-derived carbohydrate under the anaerobic condition besides the functions proposed for MCG subgroup 15¹⁶. As indicator groups of estuary sediments, MCG subgroup 3 is proposed to be methanogens and contributes to global methane cycling¹⁷. Subgroup 16 (indicator group for estuary) could utilize a variety of organic substrate via fermentation to produce acetate¹⁸ which could be subsequently utilized by various microbial groups³⁶.

Lack of available physiologic and genomic information renders it difficult to elucidate the ecological function of the remaining 5 indicator MCG subgroups (MCG 4, 13, 5a, 5b and peat MCG). Based on their phylogenetic distance with the groups whose genomic data are available, the functions of some of these subgroups are

speculated. For MCG subgroup 13, we hypothesize that they play a significant role in degrading various organic compounds such as protein, starch, and persistent aromatic compounds according to their close phylogenetic relationship with MCG 16 (Fig. 3). Because they are phylogenetically related to MCG subgroup 6 and 7 (Fig. 3), MCG subgroup 5a and 5b might degrade carbohydrate, protein debris and reduce nitrite to ammonia. As they show no close relationship with other MCG subgroups, the ecological functions of MCG subgroup 4 and peat MCG cannot be speculated.

Materials and Methods

Site description and sampling. The Dajihu Peatland (31°28′50″N, 110°00′90″E) locates in northwest Hubei province and is characterized by typically sub alpine monsoon climate. The annual temperature and precipitation were 7.2 °C and 1,560 mm, respectively. The organic carbon was abundant due to the low temperature condition and high ground water level³⁷.

Four peat sediment samples were collected with 50 ml sterilized centrifuge tubes along a profile in Sep, 2012 and transported to the geomicrobiology lab in China University of Geosciences (Wuhan) on dry ice immediately. Samples for molecular analysis were stored at −80 °C. Samples for physicochemical analysis were freezing dried and grinded into powder.

Physicochemical analysis. Homogeneous suspension was obtained by mixing the grinded peat sediments with ultra-pure water with a ratio of 1:2.5 (w/v) via stirring with a glass rod for 10 min. After sitting quietly for 30 min, the pH of supernatant was measured three times for each sample and the mean value of three measurements was reported. The total organic carbon (TOC) and total nitrogen (TN) in sediment samples were analyzed with Elemental Analyzer (Elementar Vario EL III) with a precision of ±0.4%³⁸.

Clone library construction of archaeal 16S rRNA. Genomic DNA was extracted with Fast DNA spin kit for soil (MP, USA) according to manufacturer's instructions. Extracted DNA was stored at −80 °C after the concentration was measured with Nanodrop ND-2000 spectrophotometer (ThermoFisher, Germany). Archaeal 16S rRNA gene fragments were amplified with the primer set A21F and A958R³⁹ by the following steps: 94 °C for 5 min, 36 cycles of 94 °C for 30 s, 58 °C for 45 s, 72 °C for 1 min, and final extension at 72 °C for 10 min. PCR mixture (20 µl) contained 2.0 µl 10 × PCR buffer, 2.0 µl Mg²⁺ (25 mM), 2.0 µl dNTP (2.5 mM), 0.5 µl of each primer (20 mM), 0.25 µl of Taq DNA polymerase (5 U/µl), 1 µl of BSA (20 µg/µl) (TaKaRa, Dalian, China), 2 µl of template DNA (20 ng/µl) and 9.75 µl of PCR water. All the PCR reactions were run on a Master Cycler 5331 Gradient PCR (Eppendorf, Hamberg, Germany).

PCR product was visualized with 1% agarose gel. Target bands were excised and recovered with QIAquick gel extraction kit (Qiagen, Germany) according to the manufacturer's instructions. Purified PCR product fragments were ligated into pMD18-T plasmid vector (TaKaRa, Dalian, China) and transformed into the *Escherichia coli* JM109 cell (TaKaRa, Dalian, China). Transformants were identified based on blue-white screening on LB media containing 100 µg/ml of ampicillin. The presence of the insert was screened directly with M13F and M13R vector primers and positive clones were sequenced with A21F primer by Genscript Company (Nanjing, China). Possible vector contamination was checked using NCBI Vecscreen (<http://www.ncbi.nlm.nih.gov/tools/vecscreen/>). Totally 216 valid sequences were acquired from Dajihu peat sediments which are deposited in GenBank under accession numbers of KM514090-KM514305. All the sequences were pooled together as one clone library to increase the representativeness.

Selection criteria for archaeal 16S rRNA sequence through NCBI nr database. To discern the distribution pattern of MCG communities across terrestrial ecosystems and the association between MCG communities and environmental parameters, we collected archaeal 16S rRNA sequences available in Genbank database. Only the sequences of 1) that were amplified with the universal primer set (A21F and A958R); 2) that contained no nucleotide ambiguities and whose lengths were >700 bp; 3) whose sampling site were well defined were used.

Construction of phylogenetic tree. ClustalW program⁴⁰ was used to align the 5226 archaeal 16S rRNA sequences from 51 investigated sites and Mothur software was used to calculate distance matrix and cluster sequences⁴¹. Those sequences were clustered into Operation Taxonomy Unit (OTU) at the identity level of 97% (890 OTUs) for diversity analysis and 90% (237 OTUs) for network analysis. The selected 890 OTUs (97% similarity) were used to construct archaeal taxonomy tree based on Maximum likelihood method by GTR + G + T mode with Mega 6.0 software⁴². Among 890 archaeal OTUs, 353 MCG OTUs were selected to construct MCG phylogenetic tree. The best phylogenetic tree was constructed with GTR + G2 mode with the bootstrap 1000 and was drawn by iTOL⁴³.

Statistical analysis. Phylogenetic diversity (PD)⁴⁴ and Shannon-Wiener index⁴⁵ were used to characterize microbial diversity. PD is the sum of phylogenetic branch length within a habitat⁴⁴. Mean PD value of 1000 random subsamples were used to correct unequal sequence numbers between habitats. The Shannon-Wiener index indicates the number and abundance of species within a habitat which positively correlates with richness and evenness, and is sensitive to the abundance variation of rare species⁴⁶.

UniFrac was conducted to analyze beta diversity which quantifies community similarity based on phylogenetic relatedness⁴⁷. Principle coordinate analysis (PCoA) was used to visualize beta diversity obtained from UniFrac distance matrix. Permutation manova based on 1000 permutations with R (version 3.2.2)⁴⁸ package *vagan* (version 2.4-1)⁴⁹ was used to explain phylogenetic variance from UniFrac distance.

Species abundance distribution (SAD) pattern and ecological importance of MCG lineage were estimated according to the relative abundance of archaeal lineages. A multivariate regression tree (MRT) was carried out by R package mvpart (version 1.6-2)⁵⁰ to determine the relationship between MCG subgroups and environmental factors.

Phylogenetic structure was estimated with phylogenetic species variability (PSV)⁵¹ which based on species presence/absence and their phylogenetic relatedness. This metrics is between 0 and 1, and asymptotically approach 0 as relatedness increases. Comparison between observed PSV and null values from different randomized models were conducted by R package picante (version 1.6-2)⁵² to statistically test whether MCG community composition of different habitats is more closely related or distantly related than expected by chance.

Correlation-based method⁵³ as well as probabilistic graphic model method⁵⁴ is a common approach to explore underlying mechanism of microbial ecological process. Despite of the deficiency in the information of control and dependency, correlation-based method shows direct relatedness between (environmental and biological) variables and is widely used to study potential interactions between microorganisms^{20,55}. In this study we analyzed correlation between OTUs affiliated with MCG and other archaeal lineages using correlation-based method based on Spearman coefficient⁵⁶. We calculated all possible Spearman rank correlation between those OTUs (at the threshold of 90% identity) with more than 5 sequences (99 OTUs). We removed those poor representative OTUs and reduced network complexity, facilitating the determination of core archaeal community. Valid co-occurrence event with rank correlation coefficients above 0.6 and statistical significant below 0.01⁵⁷ were considered to be a robust correlation. The network was characterized by topology indices (diameter, average path length, modularity, number of the shortest path) and indices of an individual node (degree, closeness centrality, between centrality). Specially, modularity was calculated using Louvain algorithm⁵⁸, which include two phases of modularity optimization and community aggravation. Aggravation is repeated until modularity no longer can be increased. All these analysis were performed by R package psych (version 1.6.9)⁵⁹ and Gephi (version 0.8.2 beta) software⁶⁰. Network images were visualized by Gephi software. The correlation between modules and habitats were statistically tested by the Fisher's test under software R.

References

- Inagaki, F. *et al.* Microbial communities associated with geological horizons in coastal seafloor sediments from the sea of Okhotsk. *Appl. Environ. Microbiol.* **69**, 7224–7235 (2003).
- Dang, H., Luan, X., Zhao, J. & Li, J. Diverse and novel nifH and nifH-like gene sequences in the deep-sea methane seep sediments of the Okhotsk Sea. *Appl. Environ. Microbiol.* **75**, 2238–2245 (2009).
- Sørensen, K. B. & Teske, A. Stratified communities of active archaea in deep marine subsurface sediments. *Appl. Environ. Microbiol.* **72**, 4596–4603 (2006).
- Webster, G. *et al.* Enrichment and cultivation of prokaryotes associated with the sulphate–methane transition zone of diffusion-controlled sediments of Aarhus Bay, Denmark, under heterotrophic conditions. *FEMS Microbiol. Ecol.* **77**, 248–263 (2011).
- Sayeh, R. *et al.* Microbial diversity in Tunisian geothermal springs as detected by molecular and culture-based approaches. *Extremophiles* **14**, 501–514 (2010).
- Dong, H. *et al.* Microbial diversity in sediments of saline Qinghai Lake, China: linking geochemical controls to microbial ecology. *Microbiol. Ecol.* **51**, 65–82 (2006).
- Llirós, M., Trias, R., Borrego, C. & Bañeras, L. Specific Archaeal communities are selected on the root surfaces of *Ruppia* spp. and *Phragmites australis*. *Wetlands* **34**, 403–411 (2014).
- Etto, R. M. *et al.* Prokaryotic communities of acidic peatlands from the southern Brazilian Atlantic Forest. *Braz. J. Microbiol.* **43**, 661–674 (2012).
- Jiang, L., Zheng, Y., Chen, J., Xiao, X. & Wang, F. Stratification of archaeal communities in shallow sediments of the Pearl River Estuary, Southern China. *Antonie Van Leeuwenhoek* **99**, 739–751 (2011).
- Wei, S. *et al.* Diversity and distribution of archaea community along a stratigraphic permafrost profile from Qinghai-Tibetan Plateau, China. *Archaea* **2014**, 240817 (2014).
- Cao, P. *et al.* Distribution and diversity of archaeal communities in selected Chinese soils. *FEMS Microbiol. Ecol.* **80**, 146–158 (2012).
- Meng, J. *et al.* An uncultivated crenarchaeota contains functional bacteriochlorophyll a synthase. *ISME J.* **3**, 106–116 (2009).
- Na, H., Lever, M. A., Kjeldsen, K. U., Schulz, F. & Jørgensen, B. B. Uncultured desulfobacteraceae and crenarchaeotal group C3 incorporate ¹³C-acetate in coastal marine sediment. *Env. Microbiol. Rep.* **7**, 614–622 (2015).
- Biddle, J. F. *et al.* Heterotrophic archaea dominate sedimentary subsurface ecosystems off Peru. *Proc. Natl. Acad. Sci. USA* **103**, 3846–3851 (2006).
- Lloyd, K. G. *et al.* Predominant archaea in marine sediments degrade detrital proteins. *Nature* **496**, 215–218 (2013).
- Lazar, C. S. *et al.* Genomic evidence for distinct carbon substrate preferences and ecological niches of Bathyarchaeota in estuarine sediments. *Environ. Microbiol.* **18**, 1200–1211 (2016).
- Evans, P. N. *et al.* Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric metagenomics. *Science* **350**, 434–438 (2015).
- He, Y. *et al.* Genomic and enzymatic evidence for acetogenesis among multiple lineages of the archaeal phylum Bathyarchaeota widespread in marine sediments. *Nat. Microbiol.* **1**, 16035 (2016).
- Meng, J. *et al.* Genetic and functional properties of uncultivated MCG archaea assessed by metagenome and gene expression analyses. *ISME J.* **8**, 650–659 (2014).
- Fillol, M., Auguet, J.-C., Casamayor, E. O. & Borrego, C. M. Insights in the ecology and evolutionary history of the Miscellaneous Crenarchaeotic Group lineage. *ISME J.* **10**, 665–677 (2015).
- Ren, J., Yan, B. & Hong, K. Comparison of bacterial and archaeal community of mangrove soil under different vegetation in Dongzhaigang, Hainan Island. *Acta Microbiologica Sinica* **52**, 736–743 (2012).
- Høj, L., Olsen, R. A. & Torsvik, V. L. Effects of temperature on the diversity and community structure of known methanogenic groups and other archaea in high Arctic peat. *ISME J.* **2**, 37–48 (2007).
- Auguet, J. C. & Casamayor, E. O. A hotspot for cold crenarchaeota in the neuston of high mountain lakes. *Environ. Microbiol.* **10**, 1080–1086 (2008).
- Xie, W., Zhang, C., Zhou, X. & Wang, P. Salinity-dominated change in community structure and ecological function of Archaea from the lower Pearl River to coastal South China Sea. *Appl. Microbiol. Biot.* **98**, 7971–7982 (2014).
- Auguet, J.-C., Barberan, A. & Casamayor, E. O. Global ecological patterns in uncultured Archaea. *ISME J.* **4**, 182–190 (2010).
- Zhou, J. *et al.* Spatial and resource factors influencing high microbial diversity in soil. *Appl. Environ. Microbiol.* **68**, 326–334 (2002).
- Drenovsky, R. E., Vo, D., Graham, K. J. & Scow, K. M. Soil water content and organic carbon availability are major determinants of soil microbial community composition. *Microbiol. Ecol.* **48**, 424–430 (2004).

28. Oni, O. E. *et al.* Microbial communities and organic matter composition in surface and subsurface sediments of the Helgoland Mud Area, North Sea. *Front. Microbiol.* **6**, 1290 (2015).
29. Cole, J. K. *et al.* Sediment microbial communities in Great Boiling Spring are controlled by temperature and distinct from water communities. *ISME J.* **7**, 718–729 (2013).
30. Bäckhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A. & Gordon, J. I. Host-bacterial mutualism in the human intestine. *Science* **307**, 1915–1920 (2005).
31. Ma, K., Liu, X. & Dong, X. *Methanosaeta harundinacea* sp. nov., a novel acetate-scavenging methanogen isolated from a UASB reactor. *Int. J. Syst. Evol. Microbiol.* **56**, 127–131 (2006).
32. Bräuer, S. L., Cadilloquiroz, H., Ward, R. J., Yavitt, J. B. & Zinder, S. H. *Methanoregula boonei* gen. nov., sp. nov., an acidiphilic methanogen isolated from an acidic peat bog. *Int. J. Syst. Evol. Microbiol.* **61**, 45–52 (2010).
33. Sakai, S., Conrad, R., Liesack, W. & Imachi, H. *Methanocella arvoryzae* sp. nov., a hydrogenotrophic methanogen isolated from rice field soil. *Int. J. Syst. Evol. Microbiol.* **60**, 2918–2923 (2010).
34. Oren, A. The Family Methanomicrobiaceae in *The prokaryotes* (eds Rosenberg, E. *et al.*) 231–146 (Springer, 2014).
35. Collins, G. *et al.* Distribution, localization, and phylogeny of abundant populations of Crenarchaeota in anaerobic granular sludge. *Appl. Environ. Microbiol.* **71**, 7523–7527 (2005).
36. Vandieken, V. & Thamdrup, B. Identification of acetate-oxidizing bacteria in a coastal marine surface sediment by RNA-stable isotope probing in anoxic slurries and intact cores. *FEMS Microbiol. Ecol.* **84**, 373–386 (2013).
37. Huang, X. *et al.* Occurrence of diplotene in moss species from the Dajiu Peatland in southern China. *Org. Geochem.* **41**, 321–324 (2010).
38. Yang, H. *et al.* Soil pH impact on microbial tetraether lipids and terrestrial input index (BIT) in China. *Sci. China Earth Sci.* **55**, 236–245 (2012).
39. DeLong, E. F. Archaea in coastal marine environments. *Proc. Natl. Acad. Sci. USA* **89**, 5685–5689 (1992).
40. Li, K.-B. ClustalW-MPI: ClustalW analysis using distributed and parallel computing. *Bioinformatics* **19**, 1585–1586 (2003).
41. Schloss, P. D. *et al.* Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**, 7537–7541 (2009).
42. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729 (2013).
43. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**, 127–128 (2007).
44. Faith, D. P. Conservation evaluation and phylogenetic diversity. *Biol. Conserv.* **61**, 1–10 (1992).
45. Spellerberg, I. F. & Fedor, P. J. A tribute to Claude Shannon (1916–2001) and a plea for more rigorous use of species richness, species diversity and the Shannon–Wiener Index. *Global Ecol. Biogeog.* **12**, 177–179 (2003).
46. Hughes, J. B., Hellmann, J. J., Ricketts, T. H. & Bohannan, B. J. Counting the uncountable: statistical approaches to estimating microbial diversity. *Appl. Environ. Microbiol.* **67**, 4399–4406 (2001).
47. Lozupone, C., Hamady, M. & Knight, R. UniFrac—an online tool for comparing microbial community diversity in a phylogenetic context. *BMC bioinformatics* **7**, 1 (2006).
48. R Core Team R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/> (2013).
49. Oksanen, J. *et al.* The vegan package. *Community ecology package* **10** (2007).
50. De'Ath, G. Multivariate regression trees: a new technique for modeling species–environment relationships. *Ecology* **83**, 1105–1117 (2002).
51. Helmus, M. R., Bland, T. J., Williams, C. K. & Ives, A. R. Phylogenetic measures of biodiversity. *Am. Nat.* **169**, E68–E83 (2007).
52. Kembel, S. W. *et al.* Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* **26**, 1463–1464 (2010).
53. Fierer, N. & Jackson, R. B. The diversity and biogeography of soil bacterial communities. *Proc. Natl. Acad. Sci. USA* **103**, 626–631 (2006).
54. Metcalf, J. L. *et al.* Microbial community assembly and metabolic function during mammalian corpse decomposition. *Science* **351**, 158–162 (2015).
55. Ma, B. *et al.* Geographic patterns of co-occurrence network topological features for soil microbiota at continental scale in eastern China. *ISME J.* **10**, 1891–1901 (2016).
56. Hollander, M., Wolfe, D. A. & Chicken, E. Nonparametric statistical methods in *Probability and statistics*. (eds Balding, David J. *et al.*) 115–149 (Wiley, 2013).
57. Barberán, A., Bates, S. T., Casamayor, E. O. & Fierer, N. Using network analysis to explore co-occurrence patterns in soil microbial communities. *ISME J.* **6**, 343–351 (2012).
58. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *Mech-theory. E.* **2008**, P10008 (2008).
59. Revelle, W. psych: Procedures for psychological, psychometric, and personality research. R package version 1.6.9. <http://personality-project.org/r/psych.manual.pdf> (2012).
60. Bastian, M., Heymann, S. & Jacomy, M. Gephi: an open source software for exploring and manipulating networks. *ICWSM* **8**, 361–362 (2009).

Acknowledgements

This work was supported by National Natural Science Foundation of China (Grant Nos 41572325 and 41130207).

Author Contributions

X.X., H.W. conceived and designed this experiment, X.X. and R.W. analyzed those data. Author X.X. and R.W. contributed equally. All of authors including X.X., R.W., H.W., L.G., B.M. and Y.X. assisted in writing this manuscript, discussed the results and commented on the manuscript.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing Interests: The authors declare no competing financial interests.

How to cite this article: Xiang, X. *et al.* Distribution of Bathyarchaeota Communities Across Different Terrestrial Settings and Their Potential Ecological Functions. *Sci. Rep.* **7**, 45028; doi: 10.1038/srep45028 (2017).

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



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017