Supporting Information

Gubry-Rangin et al. 10.1073/pnas.1109000108

SI Materials and Methods

Sequence Database and Alignment. Sequences were retrieved from the National Center for Biotechnology Information (NCBI) GenBank database with the Entrez search terms "amoA + archaea" together with contextual data (April 2010). A database was constructed from an initial collection of >10,000 sequences, comprising 3,619 soil-derived sequences and five reference metagenomes or sequences from cultivated organisms (Tables S1 and S2). All sequences were aligned within ARB software (1) using ClustalW and edited manually. Sequences were referenced against GenBank data or associated publications, and those recovered from soils of unknown pH, those with fewer than 586 nucleotide positions spanning primer positions 23F/616R (2), and those that did not translate perfectly were excluded from further analysis. Sequences of ambiguous phylogenetic placement identified from partial treeing analysis were removed as well, resulting in an alignment of 606 sequences on 586 nucleotide positions. Other datasets were constructed for samples with information on country of origin and ecosystem type, but analysis of other environmental characteristics was precluded by a lack of robust and comprehensive data across all studies.

Phylogenetic Analysis and Affiliation of Database Sequences to Clusters. Potentially distinct phylogenetic clusters associated with specific pH ranges were defined using a series of analyses on unambiguously aligned DNA sequences (606 taxa, 586 nucleotide positions) using distance and maximum likelihood (ML) analyses (Fig. 1*A*). Given the large dataset, bootstrap support for individual clusters was calculated by distance analysis in MEGA 4.02 (3) using general time-reversible correction, ML-estimated variable sites only, and gamma-distributed site variation calculated using PhyML (4). A wider range of phylogenetic analyses were performed on a restricted selection of inferred amino acid sequences (three sequences per cluster) (Fig. 1*B*). ProtTest (5) was used to identify suitable models of correction before performing ML and distance analyses with PhyML and MEGA, respectively. Maximum parsimony analysis was performed using MEGA, and Bayesian posterior probabilities were calculated with MrBayes (6). Bootstrap support for distance, ML, and parsimony methods was calculated with 1,000 replicates each, and Bayesian analysis was performed with 1 million iterations with an SD <0.04.

The minimum level of divergence giving high bootstrap support for all clusters was ~20% (~0.2 change per nucleotide position, using estimated variable sites only), with the maximum divergence of 83% between two sequences within any group. Initial phylogenies indicated the possible presence of specific pH-adapted groups within these clusters. Thus, phylogenetic clusters were formed as close to 0.2 change per nucleotide position as possible (i.e., 0.1 change in one direction along the scale bar) and only where bootstrap support was high (90% or more if possible). This resulted in the formation of 18 clusters of soil sequences, plus 1 cluster containing *N. yellowstonii*-like sequences. UCLUST (7) was then used to assign novel 454 sequence data to each of the 18 predetermined clusters. Because this software does not use a phylogenetically defined method of comparison (but simply % pairwise similarity), phylogenetic analyses also were performed to verify the placement of UCLUST-assigned 454 sequences into the 18 predetermined clusters. No discrepancies among the methods was observed.

Soil Collection, Characterization, and DNA Extraction. Forty soil cores were selected from the UK Countryside Survey (www.countrysidesurvey.org.uk/), a survey of more than 1,000 soil samples. Each sample represents a soil core (5 cm diameter, 15 cm depth) sampled in the United Kingdom between May and November 2007. Neighboring cores were sampled simultaneously to measure pH, % carbon (C), % nitrogen (N), C:N ratio; % organic matter (loss on ignition), soil moisture content (% moisture), and phosphorous (Olsen P mg kg^{-1}) (see ref. 8 for a full description). In addition, seven soil samples were collected across a pH gradient of 4.5–7.5 (Scottish Agricultural College, Aberdeen, Scotland; grid reference NJ872104) that has been maintained for 50 y. Soil cores were homogenized, and total nucleic acids were extracted from 0.25 g of soil as described previously (9), with a supplementary 30-min hexadecyltrimethylammonium bromide freeze-thaw, soft lysis stage.

Quantitative PCR. The abundance of archaeal *amoA* genes was determined as described previously (10) using primers CrenamoA23f (5′-ATGGTCTGGCTWAGACG-3′) and CrenamoA616r (5′-GCCATCCATCTGTATGTCCA-3′) (11) and an *amoA* gene-containing a 1,934-bp PCR product derived from fosmid 54d9 as the standard template (12). Efficiency of the amplification was 95%, and the r^2 value was 1.

Sequencing and Classification of Archaeal *amoA* Genes. Bar-coded amplicons for multiplexing were prepared with the primers CrenamoA23f and CrenamoA616r extended as amplicon fusion primers with respective primer A and B adapters, key sequence, and one of 24 multiplex identifiers as recommended by Roche. The use of 24 different barcode decamers enabled bidirectional amplicon sequencing of 12 different samples in a single plate region, with analysis of 47 samples within four quarters of a 454 plate. Triplicate PCR amplifications were performed on each soil DNA template and pooled. Primer dimers were removed by electrophoresis of PCR products on an agarose gel, excision, and purification using a NucleoSpin Extract II Kit (Fisher). Amplicons were further purified with AMpure beads (Beckman Coulter) and pooled in an equimolar ratio as specified by Roche. Emulsion PCR, emulsion breaking of DNA-enriched beads, and sequencing runs of the amplicon pools were performed on a second-generation pyrosequencer (454 GS FLX Titanium; Roche) using titanium reagents and titanium procedures as recommended by the manufacturer following protocols for bidirectional amplicon sequencing.

Quality filtering of the pyrosequencing reads was performed using the automatic amplicon pipeline of the GS Run Processor (Roche) to remove failed and low-quality reads from raw data. Reads were truncated at the first ambiguous base and removed if that occurred in the first 50% of a flowgram (13). Reads were truncated to 400 bp, because the last portion of titanium reads has high levels of noise (14, 15). Barcodes were removed, and reads were dereplicated to unique sequences and their frequencies calculated. Unique reverse reads were reversed and complemented. All unique reads were translated to amino acids, and any with a stop codon were removed. Forward and reverse DNA sequences were then assembled, requiring an exact match over at least 100 bp with a custom C program using exact pairwise Needleman–Wunsch alignments. Multiple matches were resolved by only assembling each forward sequence with a single reverse sequence, with forward sequences matched in decreasing order of abundance and each forward sequence being matched

with the most abundant reverse sequence not already matched. Total abundance was the sum of forward and reverse sequences. All assemblies with length other than 629 bp were removed.

Each 454 sequence was compared with the 606 sequence dataset obtained from the NCBI database, searching for a match exceeding a 90% identity threshold using UCLUST and allowing for the detection of new clusters not identified in the initial analysis.

The assembly of forward and reverse sequences was based on a 100% minimum overlap of 100 nt and complete identity over forward and reverse strands. Thus, there is potential for the creation of chimeric sequences during the assembly procedure. This potential was examined using collections of cloned sequences from individual soil samples deposited in GenBank (Fig. S2). Sequences were grouped together on the basis of having 100% identity over three different 100-nt regions in the middle of the 629-bp fragment amplified in this study. The amount of sequence divergence was then calculated between all sequences over the full *amoA* amplicon. This analysis indicated that the potential for chimera formation (1% error from true sequences) is much lower than that used to assign sequences within specific clusters (≥83% identity).

Statistical Analysis. Alpha diversity was estimated by the species richness and Shannon diversity index, both calculated using the vegan package for R (16). Coverage by 454 analysis was analyzed by rarefaction analysis, calculated with EstimateS (17) using different thresholds, and plotted with SigmaPlot (Systat Software).

The relationship between pH and the relative abundance of the six most abundant clusters in the 47 soils was analyzed using generalized additive modeling (gam) with the mgcv package (18). The same approach was used to analyze the relationships between pH and the relative abundances of the three *amoA* lineages A, B, and C, associated with 16S rRNA-defined groups 1.1a, 1.1b, and 1.1a-associated, respectively. A heat map graphical representation of relative abundance data distributions was created for each environmental variable using the R package gplots (19), including a dendrogram of archaeal ammonia oxidizer community similarity. To estimate pairwise similarity in archaeal communities, the square root-transformed proportional abundance was used to generate a Bray–Curtis (BC) dissimilarity matrix using R. The BC matrix was used for ordination of soil samples by nonmetric multidimensional scaling using the vegan package. The BC matrix was compared with the environmental characteristic pairwise distance matrix (i.e., normalized Euclidean distances) using a canonical correspondence analysis. This model was tested with Monte Carlo permutation tests (199 randomized runs) using the vegan package to determine significance, and each environmental parameter was tested by stepwise analysis to detect the significant predictors.

1. Ludwig W, et al. (2004) ARB: A software environment for sequence data. *Nucleic Acids Res* 32:1363–1371.
2. Tourna M, Freitag TE, Nicol GW, Prosser JI (2008) Growth, activity and temperature responses of ammonia-oxidizing archaea and bacteria in soil microcosms. *Environ Microbiol* 10:1357–1364.
3. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596–1599.
4. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52:696–704.
5. Abascal F, Zardoya R, Posada D (2005) ProtTest: Selection of best-fit models of protein evolution. *Bioinformatics* 21:2104–2105.
6. Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574.
7. Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26:2460–2461.
8. Emmett BA, et al. (2008) CS Technical Report 3/07: Soils manual v1.0. Available at: www.countrysidesurvey.org.uk/pdf/reports2007/CS_UK_2007_TR3.pdf. Accessed November 11, 2011.
9. Griffiths RI, Whiteley AS, O'Donnell AG, Bailey MJ (2000) Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA- and rRNA-based microbial community composition. *Appl Environ Microbiol* 66: 5488–5491.
10. Gubry-Rangin C, Nicol GW, Prosser JI (2010) Archaea rather than bacteria control nitrification in two agricultural acidic soils. *FEMS Microbiol Ecol* 74:566–574.
11. Tourna M, Freitag TE, Nicol GW, Prosser JI (2008) Growth, activity and temperature responses of ammonia-oxidizing archaea and bacteria in soil microcosms. *Environ Microbiol* 10:1357–1364.
12. Treusch AH, et al. (2005) Novel genes for nitrite reductase and Amo-related proteins indicate a role of uncultivated mesophilic crenarchaeota in nitrogen cycling. *Environ Microbiol* 7:1985–1995.
13. Huse SM, Huber JA, Morrison HG, Sogin ML, Welch DM (2007) Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol* 8:R143.
14. Balzer S, Malde K, Jonassen I (2011) Systematic exploration of error sources in pyrosequencing flowgram data. *Bioinformatics* 27:i304–i309.
15. Gilles A, et al. (2011) Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing. *BMC Genomics* 12:245.
16. Oksanen J, et al. (2009) Vegan: Community ecology package. R package version 1.16-33. Available at: http://cran.r-project.org/web/packages/vegan/. Accessed August 11, 2011.
17. Colwell RK (1994) EstimateS: Statistical estimation of species richness and shared species from samples. Available at:http://purl.oclc.org/estimates. Accessed August 11, 2011.
18. Wood SN (2008) Fast stable direct fitting and smoothness selection for generalized additive models. *J R Stat Soc B* 70:495–518.
19. Warnes GR (2010) Gplots:Various R programming tools for plotting data. R package version 2.8.0. Available at: http://cran.r-project.org/web/packages/gplots. Accessed November 11, 2011.
20. Leininger S, et al. (2006) Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature* 442:806–809.
21. He JZ, et al. (2007) Quantitative analyses of the abundance and composition of ammonia-oxidizing bacteria and ammonia-oxidizing archaea of a Chinese upland red soil under long-term fertilization practices. *Environ Microbiol* 9:2364–2374.
22. Tourna M, Freitag TE, Nicol GW, Prosser JI (2008) Growth, activity and temperature responses of ammonia-oxidizing archaea and bacteria in soil microcosms. *Environ Microbiol* 10:1357–1364.
23. Jia Z, Conrad R (2009) Bacteria rather than Archaea dominate microbial ammonia oxidation in an agricultural soil. *Environ Microbiol* 11:1658–1671.
24. Shen J-P, Zhang L-M, Zhu YG, Zhang J-B, He J-Z (2008) Abundance and composition of ammonia-oxidizing bacteria and ammonia-oxidizing archaea communities of an alkaline sandy loam. *Environ Microbiol* 10:1601–1611.
25. Chen X-P, Zhu Y-G, Xia Y, Shen J-P, He JZ (2008) Ammonia-oxidizing archaea: Important players in paddy rhizosphere soil? *Environ Microbiol* 10:1978–1987.
26. Hansel CM, Fendorf S, Jardine PM, Francis CA (2008) Changes in bacterial and archaeal community structure and functional diversity along a geochemically variable soil profile. *Appl Environ Microbiol* 74:1620–1633.
27. Onodera Y, Nakagawa T, Takahashi R, Tokuyama T (2010) Seasonal change in vertical distribution of ammonia-oxidizing archaea and bacteria and their nitrification in temperate forest soil. *Microbes Environ* 25:28–35.
28. He J-H, Zhang L, Wang M, Prosser JI, Zheng Y-M (2009) Altitude ammonia-oxidizing bacteria and archaea in soils of Mount Everest. *FEMS Microbiol Ecol* 70:208–217.
29. Offre P, Prosser JI, Nicol GW (2009) Growth of ammonia-oxidizing archaea in soil microcosms is inhibited by acetylene. *FEMS Microbiol Ecol* 70:99–108.
30. Nakaya A, et al. (2009) Analysis of ammonia monooxygenase and archaeal 16S rRNA gene fragments in nitrifying acid-sulfate soil microcosms. *Microbes Environ* 24: 168–174.
31. Ying J-Y, Zhang L-M, He J-Z (2010) Putative ammonia-oxidizing bacteria and archaea in an acidic red soil with different land utilization patterns. *Environ Microbiol Rep* 2: 304–312.
32. Höfferle Š, et al. (2010) Ammonium supply rate influences archaeal and bacterial ammonia oxidizers in a wetland soil vertical profile. *FEMS Microbiol Ecol* 74:302–315.
33. de la Torre JR, Walker CB, Ingalls AE, Könneke M, Stahl DA (2008) Cultivation of a thermophilic ammonia-oxidizing archaeon synthesizing crenarchaeol. *Environ Microbiol* 10:810–818.
34. Treusch AH, et al. (2005) Novel genes for nitrite reductase and Amo-related proteins indicate a role of uncultivated mesophilic crenarchaeota in nitrogen cycling. *Environ Microbiol* 7:1985–1995.
35. Hatzenpichler R, et al. (2008) A moderately thermophilic ammonia-oxidizing crenarchaeote from a hot spring. *Proc Natl Acad Sci USA* 105:2134–2139.
36. Hallam SJ, et al. (2006) Pathways of carbon assimilation and ammonia oxidation suggested by environmental genomic analyses of marine Crenarchaeota. *PLoS Biol* 4: e95.
37. Könneke M, et al. (2005) Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* 437:543–546.

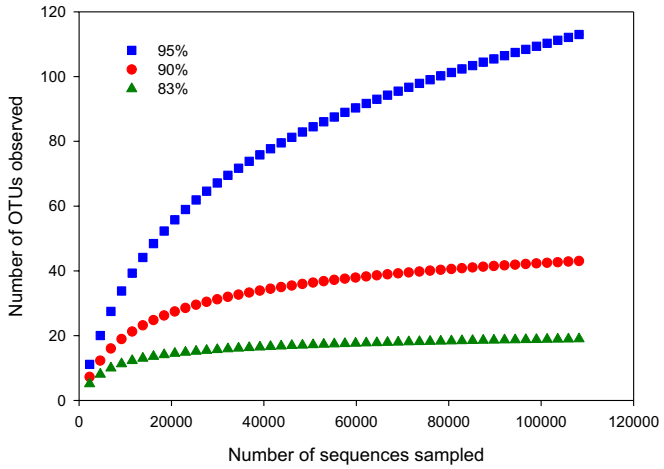
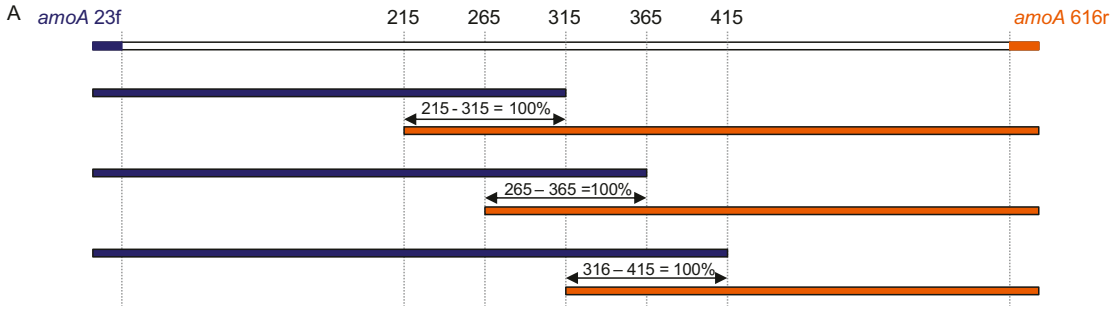


Fig. S1. Rarefaction curves of *amoA* gene sequence diversity amplified in a regional analysis of 47 soil samples. Operational taxonomic units were grouped at 95%, 90%, and 83% similarity.



B

Sequence (accession no.)	Reference	No of sequences 100% identical over 100 bp derived from same soil sample			Full amplicon comparison	
		215-315	265-365	315-415	Actual length compared	Minimum identity (%)
AOAB_sH03 (EU339453)	1	1	14	14	588	99.1
AOAC-u_sH07 (EU339457)	1	18	14	14	588	98.9
CSamoA_31_4.9 (EU856208)	2	4	4	4	506	99.6
CSamoA_63_7.5 (EU856239)	2	7	7	6	506	99.4
R60-70_284 (DQ534888)	3	10	10	9	604	99.6
OKR-C-19 (DQ148879)	4	4	4	4	606	99.6

Fig. S2. Comparison of sequence variability across 629-bp amplicons grouped on the basis of having identical ≥100-nt sections in defined regions of cloned PCR products to determine the potential for technical chimera formation during assembly of forward and reverse 454-sequenced amplicons. (*A*) Schematic diagram illustrating where 100-bp sections where dissected *in silico* from the center of cloned amplicons at three regions relative to 23f/616r *amoA* gene PCR products (positions 215–315, 265–365, and 315–415). (*B*) Comparison of six individual *amoA* gene sequences amplified from soil with other sequences amplified from the same soil DNA sample. Between 4 and 18 other sequences were found to have identical 100-bp sections to each selected clone, and the minimum overall sequence identity was ≥98.9% over the entire length of the amplicon sequences based on groupings at any of the three 100-nt positions. These results indicate that chimeras may be produced but will not result in the misassignment of individual sequences into one of 19 defined clusters containing sequences with ≥83% identity.

1. Hansel CM, Fendorf S, Jardine PM, Francis CA (2008) Changes in bacterial and archaeal community structure and functional diversity along a geochemically variable soil profile. *Appl Environ Microbiol* 74:1620–1633.
2. Nicol GW, Leininger S, Schleper C, Prosser JI (2008) The influence of soil pH on the diversity, abundance and transcriptional activity of ammonia oxidizing archaea and bacteria. *Environ Microbiol* 10:2966–2978.
3. Leininger S, et al. (2006) Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature* 442:806–809.
4. Francis CA, Roberts KJ, Beman JM, Santoro AE, Oakley BB (2005) Ubiquity and diversity of ammonia-oxidizing archaea in water columns and sediments of the ocean. *Proc Natl Acad Sci USA* 102:14683–14688.

Table S1. Details of soil studies from which *amoA* gene sequences deposited in the NCBI database were used to analyze globally distributed thaumarchaeal *amoA* gene sequences

Study	pH	Soil habitat	Country	Number of sites	Number of sequences	Reference
1	6.0	Grassland	United States	1	24	Unpublished; accession nos. DQ312267–DQ312293
2	7.1	Grassland	Germany	1	72	20
3	5.8	Agricultural	China	1	15	21
4	7.0	Agricultural	Scotland	1	22	22
5	6.9	Agricultural	Germany	1	45	23
6	8.5	Agricultural	China	1	14	24
7	6.8	Paddy	China	1	17	25
8	4.5–6.9	Forest	United States	3	89	26
9	4.2	Grassland	Austria	1	5	Unpublished; accession nos. DQ534697–DQ534701
10	5.3	Forest	Austria	1	30	Unpublished; accession nos. EU770834–EU770880
11	5.8–6.6	Forest	Japan	1	61	27
12	8.8	Grassland	Tibet	1	66	28
13	7.0	Agricultural	Scotland	1	6	29
14	3.5	Paddy	Thailand	1	5	30
15	4.1	Forest	China	1	30	31
16	4.3	Forest	Slovenia	1	87	32
17	3.7	Forest	Scotland	1	13	Unpublished; accession nos. JQ014678–JQ014690

Table S2. Additional reference sequences from both soil and nonsoil habitats

Metagenomic/genome/cloned amplicon sequences	Source	Reference
Nitrosocaldus yellowstonii, HL3.H08, NyCr.F07	Terrestrial hot spring	33
Fosmid 54d9	Meadow soil	34
Nitrososphaera gargensis	Terrestrial hot spring	35
Cenarchaeum symbiosum	*Axinella mexicana* symbiont	36
Nitrosopumilus maritimus	Public aquarium	37

Table S3. Characteristics of 47 UK soils used in analysis of the regional distribution of thaumarchaeal *amoA* gene sequences

Site	pH	*amoA* gene abundance (ng^{-1}DNA)	C (%)	N (%)	C:N ratio	H_2O (%)	Organic matter (%)	Simplified aggregate vegetation class
1	4.3	9.8E+05	2.2	0.9	2.4	NA	3.8	Grassland
2	5.0	2.5E+06	1.6	0.2	10.7	NA	2.8	Grassland
3	3.9	4.7E+04	NA	NA	NA	NA	66.9	Grassland
4	4.8	2.8E+05	2.1	0.2	9.5	22.8	4.9	Grassland
5	7.6	2.5E+06	3.2	0.3	9.4	30.4	6.3	Agricultural
6	8.3	6.7E+05	19.0	1.3	14.8	47.5	24.1	Grassland
7	8.5	8.1E+05	11.0	0.7	15.7	16.4	12.1	Grassland
8	6.9	1.4E+06	2.0	0.2	12.5	24.1	4.1	Grassland
9	4.2	6.7E+04	3.3	0.2	19.4	33.3	5.7	Forest
10	6.7	2.0E+06	1.4	0.1	10.8	15.4	2.9	Agricultural
11	8.1	1.4E+06	2.9	0.3	9.4	23.3	5.8	Agricultural
12	5.7	9.0E+04	13.4	1.1	12.1	69.8	24.0	Moorland
13	8.4	4.5E+06	4.1	0.3	12.4	24.5	6.4	Grassland
14	8.7	5.8E+03	1.3	0.1	10.0	15.8	2.7	Agricultural
15	4.3	1.7E+05	3.1	0.3	11.1	26.1	6.8	Forest
16	4.3	4.3E+04	47.9	2.6	18.8	91.9	85.1	Moorland
17	4.0	5.2E+04	49.8	2.3	21.5	83.2	92.0	Moorland
18	3.7	1.9E+04	8.0	0.4	22.2	39.9	15.0	Forest
19	6.2	7.4E+04	1.8	0.2	11.3	17.3	3.8	Grassland
20	6.4	1.4E+05	2.0	0.2	11.1	17.1	3.7	Agricultural
21	8.2	4.2E+05	1.7	0.2	9.4	19.9	3.7	Grassland
22	3.9	5.7E+04	11.8	0.5	26.2	39.8	19.9	Moorland
23	6.2	6.2E+05	3.1	0.3	10.0	29.2	5.8	Grassland
24	6.8	1.4E+05	13.0	0.7	18.6	55.9	20.1	Forest
25	6.8	1.0E+05	7.5	0.5	14.7	35.6	12.3	Grassland
26	8.1	1.5E+05	2.5	0.2	10.9	20.6	5.1	Agricultural
27	8.0	1.0E+06	2.4	0.2	10.9	20.5	5.1	Grassland
28	5.5	6.3E+05	2.7	0.2	12.9	20.2	4.7	Agricultural
29	3.9	5.8E+04	30.9	1.6	19.6	79.9	59.0	Moorland
30	6.8	3.1E+05	2.5	0.3	10.0	29.9	4.8	Grassland
31	7.2	7.9E+05	2.5	0.2	14.7	15.8	4.5	Agricultural
32	3.5	2.1E+05	22.6	1.1	20.9	75.2	40.6	Moorland
33	5.2	4.3E+05	8.3	0.4	21.8	49.9	15.9	Forest
34	6.0	2.6E+04	6.8	0.5	15.1	52.6	11.8	Moorland
35	6.9	2.4E+05	1.7	0.2	10.6	15.7	3.2	Grassland
36	5.6	3.9E+05	8.0	0.6	14.0	46.3	16.0	Grassland
37	8.5	1.6E+06	4.5	0.3	14.5	33.0	8.9	Grassland
38	6.7	9.4E+03	4.7	0.4	11.5	38.4	10.0	Agricultural
39	4.7	8.2E+04	7.3	0.4	20.3	28.3	12.5	Forest
40	4.5	4.6E+05	7.0	0.4	18.6	29.7	12.1	Agricultural
41	5.0	6.4E+05	6.6	0.4	18.0	30.9	11.3	Agricultural
42	5.5	1.1E+06	7.4	0.4	18.5	31.0	12.8	Agricultural
43	6.0	5.5E+05	6.6	0.3	21.4	30.2	11.3	Agricultural
44	6.5	8.5E+05	6.4	0.3	22.0	29.4	11.0	Agricultural
45	7.0	5.0E+05	8.0	0.4	22.1	31.0	13.7	Agricultural
46	7.5	3.7E+05	7.1	0.3	20.8	29.1	12.3	Agricultural
47	5.0	6.6E+05	7.3	0.4	20.3	24.6	12.4	Forest

Table S4. Canonical correspondence analysis of the relative abundances of archaeal lineages and measured physicochemical characteristics in 47 soils used for regional and local analysis of archaeal *amoA* gene sequence distribution

	Degrees of freedom	χ^2	*F* value	Number of permutations	Pr(>*F*)
pH	1	0.7178	11.8625	99	0.01
C:N	1	0.1004	1.6588	99	0.19
OM (LOI)	1	0.0619	1.0222	99	0.39
Moisture	1	0.0910	1.5047	99	0.13
Vegetation	3	0.2555	1.4076	99	0.14
Residual	36	2.1783			