Supporting Information

Chan et al. 10.1073/pnas.1300643110

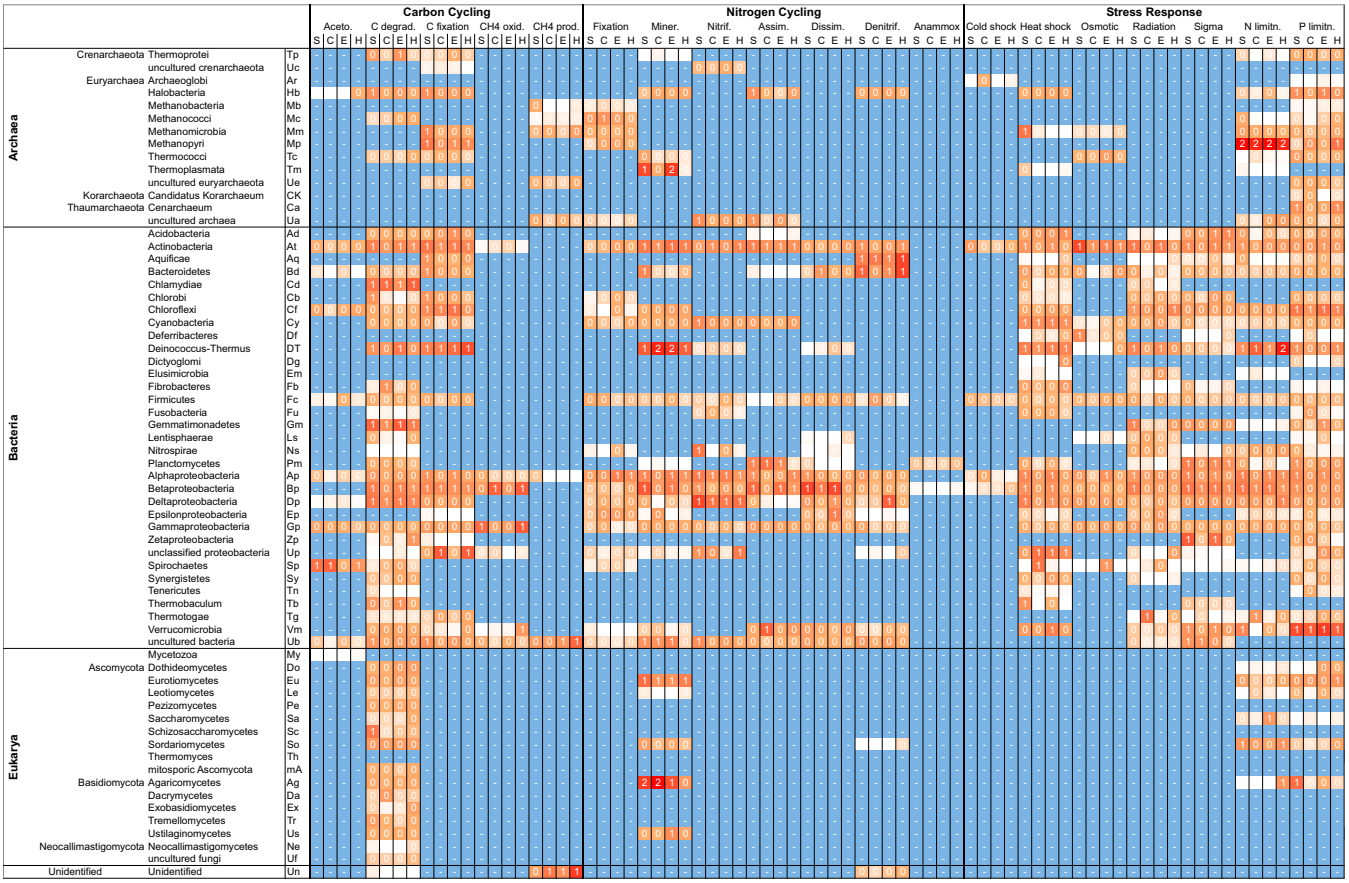

Fig. S1. Heat map of GeoChip data standardized by the taxonomic lineages per gene category in carbon and nitrogen transformations and stress responses. Red shading denotes positive signals, with darker intensity indicating stronger signal. Numerical assignments in each box show mean relative signal intensity rounded to the nearest whole number (i.e., 0 denotes 0–0.49, 1 denotes 0.5–1.49, and 2 denotes 1.5–2). Blue color indicates hybridizations that would not be expected for a given taxon/gene. White boxes represent loci that are not part of the GeoChip microarray. Aceto., acetogenesis; Assim., assimilatory nitrogen reduction; C, chasmoendoliths; Degrad., degradation; Denitrif., denitrification; Dissim., dissimilatory nitrogen reduction; E, cryptoendoliths; H, hypoliths; Miner., mineralization; N limitn., nitrogen limitation; Nitrif., nitrification; Oxid., oxidation; P limitn., phosphorous limitation; Prod., production; S, soils.

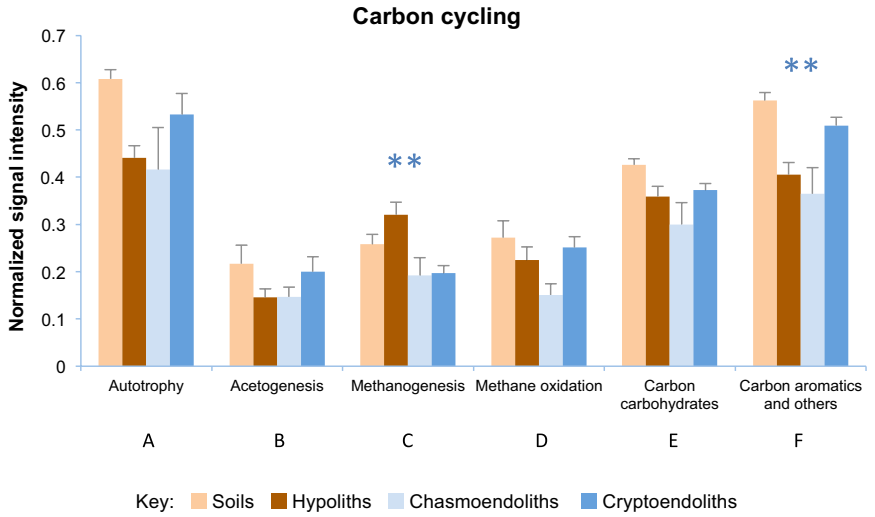

Fig. S2. Relative abundance of various carbon-cycling genes grouped by the major functional processes: (*A*) autotrophy, (*B*) acetogenesis, (*C*) methanogenesis, (*D*) methane oxidation, (*E*) carbon carbohydrates degradation, and (*F*) carbon aromatics and others degradation. The signal intensity was normalized by the number of probes for each category and the relative abundance was determined by dividing the total signal of all carbon-cycling genes. Mean values of samples from respective niches (soils, hypoliths, chasmoendoliths, and cryptoendoliths) were plotted with the SEs ($n = 3$ for rocks and $n = 5$ for soils). The following pathway-critical enzymes/loci were selected to represent the functional processes: (*A*) autotrophy was depicted by the key enzymes ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCo) in Calvin–Benson–Bassham cycle, the propionyl-CoA/acetyl-CoA carboxylase (*pcc*) in the 3-hydroxypropionate/malyl-CoA cycle, the ATP citrate lyase (*acl*B), and the carbon-monoxide dehydrogenase (CODH) in the reductive acetyl-CoA pathway; (*B*) acetogenesis was represented by the Formyltetrahydrofolate synthetase (FTHFS) in the Wood–Ljungdahl pathway; (*C*) methanogenesis was represented by the key enzyme methyl coenzyme M reductase encoded by mcrA; (*D*) methane oxidation was detected by the methane monooxygenase (mmoX) and particulate methane monooxygenase (pmoA); (*E*) degradations of various carbon carbohydrates were shown by (*i*) alpha-amylase (amyA), amylopullulanase (amyX, apu), cyclo-maltodextrin dextrin-hydrolase (cda), glucoamylase, isopullulanase, neopullulanase II (nplT), and pullulanase (pulA) for starch; (*ii*) cellobiose dehydrogenase, cellobiase (bgl), endoglucanase (egl), and exoglucanase for cellulose; (*iii*) bacterial arabinofuranosidase, fungal arabinofuranosidase, mannanase, xylose isomerase (*xyl*A), and xylanase for hemicellulose; (*iv*) acetylglucosaminidase, endochitinase, and exochitinase for chitin; and (*v*) pectinase for pectin; (*F*) degradations/metabolisms of a wide range of other carbon compounds, aromatics, and organic contaminants were detected by 188 different enzymes/gene loci. The variations among niches were tested with one-way ANOVA and labeled with ***$P < 0.01$, **$P < 0.05$, and *$P < 0.10$.

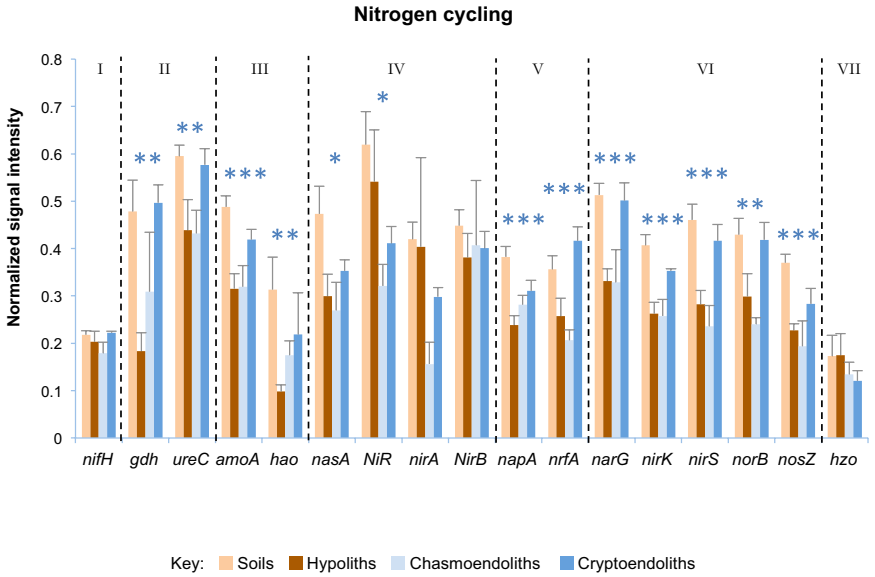

Fig. S3. Relative abundance of pathway-critical loci for various nitrogen transformation pathways: (*I*) Nitrogen fixation, nitrogenase reductase (nifH); (*II*) mineralization, glutamate dehydrogenase (gdh) and urea amidohydrolase (ureC); (*III*) nitrification, ammonia monooxygenase (amoA) and hydroxylamine oxidase (hao); (*IV*) assimilatory nitrogen reduction, assimilatory nitrate reductase (nasA) and nitrite reductase (nir, nirA, nirB); (*V*) dissimilatory nitrogen reduction, nitrate reductase (napA) and c-type cytochrome nitrite reductase (nrfA); (*VI*) denitrification, nitrate reductase (narG), Cu-nitrite reductase (nirK), cytochrome-nitrite reductase (nirS), nitric oxide reductase (norB), and nitrous oxide reductase (nosZ); and (*VII*) anaerobic ammonium oxidation (anammox), hydrazine oxidoreductase (hzo). The signal intensity was normalized by the number of probes for each gene and the relative abundance was determined by dividing the total signal of all nitrogen-cycling genes. Mean values of samples from respective niches (soils, hypoliths, chasmoendoliths, and cryptoendoliths) were plotted with the SEs. The variations among niches were tested with one-way ANOVA and labeled with ***$P < 0.01$, **$P < 0.05$, and *$P < 0.10$.

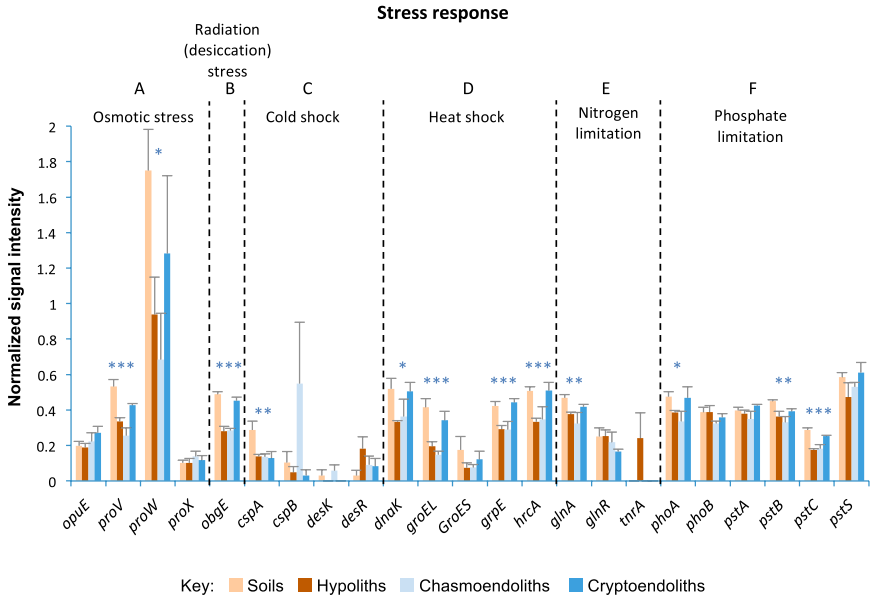

Fig. S4. Relative abundance of stress response genes. These include (*A*) osmotic stress response genes that encode the osmoregulated proline symporter (*opu*E) and the proU operon that encodes the glycine betaine/proline ABC transporter system (proVWX) that facilitates the uptake of osmoprotectants; (*B*) the radiation (desiccation) stress response gene that encodes the GTP-binding protein (obgE) for DNA repair; (*C*) cold-shock genes that encode cold-shock proteins (cspA and cspB) and the sensor/repressor (desK and desR); (*D*) heat-shock genes that encode molecular chaperone hsp (heat-shock proteins) (dnaK, groES, and grpE) and the hsp transcriptional repressor (hrcA); (*E*) nitrogen limitation-responsive genes such as glnA that encodes glutamine synthetase (GS), which is the key enzyme in nitrogen metabolism in virtually all living cells, and the GS transcription regulators encoded by glnR and tnrA; and (*F*) phosphate limitation-responsive genes including the Alkaline phosphatase (phoA), the response regulator (phoB), and the pst operon that encodes the high-affinity phosphate-specific ABC (ATP-binding cassette) transport system (pstSCAB). The signal intensity was normalized by the number of probes for each gene and the relative abundance was determined by dividing the total signal of all stress reponse genes. Mean values of samples from respective niches (soils, hypoliths, chasmoendoliths, and cryptoendoliths) were plotted with the SEs. The variations among niches were tested with one-way ANOVA and labeled with $***P < 0.01$, $**P < 0.05$, and $*P < 0.10$.

Table S1. Summary of functional loci detected and relative signal intensity, using the GeoChip functional microarray

	Soils					Chasmoendoliths			Cryptoendoliths			Hypoliths			
	B1	B2	B3	B4	B5	C1	C2	C3	E1	E2	E3	H1	H2	H3	Total, average
No. probes,	18,638	17,146	15,621	17,094	19,526	15,818	10,857	18,362	14,714	15,770	16,042	14,988	17,172	14,815	35,858
%	52.0	47.8	43.6	47.7	54.5	44.1	30.3	51.2	41.0	44.0	44.7	41.8	47.9	41.3	45.1
No. genes,	374	372	364	370	376	373	353	377	361	367	369	365	369	366	401
%	93.3	92.8	90.8	92.3	93.8	93.0	88.0	94.0	90.0	91.5	92.0	91.0	92.0	91.3	91.8

Table S2. Number of detected gene overlap between pairwise samples (%)

	Soils					Chasmoendoliths			Cryptoendoliths			Hypoliths			
	B1	B2	B3	B4	B5	C1	C2	C3	E1	E2	E3	H1	H2	H3	
Soils	2 (0.53)	366 (98.4)	360 (98.9)	363 (98.1)	368 (98.4)	362 (97.1)	348 (98.6)	367 (98.1)	357 (98.9)	362 (98.6)	364 (98.6)	362 (99.2)	360 (97.6)	362 (98.9)	B1
		1 (0.27)	358 (98.4)	359 (97.0)	364 (97.8)	359 (96.5)	345 (97.7)	365 (98.1)	355 (98.3)	359 (97.8)	362 (98.1)	360 (98.6)	358 (97.0)	359 (98.1)	B2
			1 (0.27)	357 (98.1)	360 (98.9)	357 (98.1)	344 (97.5)	361 (99.2)	354 (98.1)	356 (97.8)	356 (96.5)	355 (97.5)	354 (97.3)	353 (97.0)	B3
				4 (1.1)	363 (98.1)	358 (96.8)	345 (97.5)	364 (98.4)	355 (98.3)	358 (97.5)	358 (97.0)	356 (97.5)	357 (96.7)	355 (97.0)	B4
					1 (0.27)	366 (98.1)	349 (98.9)	370 (98.4)	357 (98.9)	364 (99.2)	365 (98.9)	362 (99.2)	362 (98.1)	361 (98.6)	B5
Chasmoendoliths						0	347 (98.3)	365 (97.9)	357 (98.9)	362 (98.6)	359 (89.5)	360 (98.6)	363 (98.4)	359 (98.1)	C1
							0	347 (98.3)	345 (97.7)	345 (97.7)	348 (86.8)	346 (98.0)	344 (97.5)	344 (97.5)	C2
								0	359 (99.4)	365 (99.5)	364 (90.8)	363 (99.5)	366 (99.2)	362 (98.9)	C3
Cryptoendoliths									0	356 (98.6)	355 (88.5)	356 (98.6)	354 (98.1)	353 (97.8)	E1
										0	360 (89.8)	358 (98.1)	359 (97.8)	356 (97.3)	E2
											0	360 (98.6)	358 (97.0)	359 (98.1)	E3
Hypoliths												0	358 (98.1)	360 (98.6)	H1
													1 (0.27)	358 (97.8)	H2
														0	H3

Numbers of unique genes in each sample are shaded in peach.

Table S3. Biodiversity estimates using phylogenetic marker gyrB from GeoChip and comparison with 16S rRNA gene-based estimates from clone libraries

	Soils		Hypoliths		Chasmoendoliths		Cryptoendoliths	
Biodiversity estimates	rRNA	gyrB	rRNA	gyrB	rRNA	gyrB	rRNA	gyrB
Alpha Diversity Indexes								
Shannon's diversity	3.3	5.2	2.3	5.2	2.8	5.0	2.8	5.1
Simpson diversity index	1	1.0	0.8	1.0	0.9	1.0	0.9	1.0
Pielou's evenness	0.9	0.9	0.7	0.9	0.7	0.9	0.7	0.9
Relative abundance, %								
Archaea								
Euryarchaeota	—	3.833	—	5.225	—	4.746	—	5.506
Bacteria								
Uncultured bacterium	12.778	1.720	0	1.704	7.848	2.404	1.998	1.866
Thermobaculum	0	1.872	0	5.904	0	3.576	0	3.605
Acidobacteria	31.111	0.583	1.998	0.660	1.962	1.035	4.995	0.281
Actinobacteria	32.778	6.892	1.998	5.990	4.905	4.319	3.996	4.914
Aquificae	0	0.857	0	0.392	0	0.000	0	1.560
Bacteroidetes	2.222	0.768	0	1.205	6.867	0.954	15.984	1.065
Chlamydiae	0	1.686	0	0.128	0	0.067	0	0.389
Chlorobi	0	2.281	0	1.809	0	3.160	0	2.432
Chloroflexi	2.778	0.480	0	0.205	0	1.560	0	0.047
Cyanobacteria	0	2.980	94.905	5.018	63.765	5.177	55.944	5.184
Deinococcus-Thermus	5.556	0.873	0	0.000	0	0.251	1.998	1.463
Firmicutes	0	3.564	0	3.368	0	2.616	0	2.343
Gemmatimonadetes	8.3333	0.000	0	0.000	0	0.000	0	0.000
Lentisphaerae	0	4.968	0	1.165	0	1.423	0	0.927
Planctomycetes	0.555	5.089	0	9.506	0	6.024	1.998	8.408
Alphaproteobacteria	0	7.051	0.999	6.575	3.924	5.370	3.996	5.835
Betaproteobacteria	3.889	5.479	0	5.019	0	3.068	0	5.610
Deltaproteobacteria	0	8.500	0	8.266	0	4.526	0	6.144
Epsilonproteobacteria	0	8.890	0	7.852	0	4.730	0	5.219
Gammaproteobacteria	0	2.344	0	2.214	7.848	1.728	3.996	1.926
Unclassified Proteobacteria	0	21.435	0	19.136	0	39.591	0	29.987
Spirochaete	0	1.140	0	1.524	0	0.686	0	0.567
Tenericutes	0	1.470	0	2.214	0	1.286	0	1.021
Thermotogae	0	3.252	0	3.363	0	0.039	0	0.944
Verrucomicrobia	0	1.992	0	1.559	0	1.662	0	2.757
Eukarya								
Ascomycota	—	—	0	—	0.323	—	0.011	—
Basidiomycota	—	—	0	—	0.076	—	0.013	—
Chlorophyta	—	—	0.1	—	2.482	—	4.072	—

In each library, $n = 250$ approximately. Relative abundance of taxa was estimated and standardized at the phylum level for all domains. —, data not available.

Table S4. Abiotic data for soil and rocks

	Soils					Hypoliths			Chasmoendoliths			Cryptoendoliths		
Abiotic measures	B1	B2	B3	B4	B5	H1	H2	H3	C1	C2	C3	E1	E2	E3
Moisture content	0.6	0.6	0.6	0.6	0.7	ND	ND	ND	ND	ND	ND	ND	ND	ND
Total organic content	0.5	0.5	0.5	0.5	0.5	ND	ND	ND	ND	ND	ND	ND	ND	ND
pH	8.6	9	8.6	8.8	8.1	ND	ND	ND	ND	ND	ND	ND	ND	ND
Porosity	NA	NA	NA	NA	NA	ND	ND	ND	0.89	0.3	1.35	1.61	0.56	1.71
Soluble salts	<0.05	0.15	<0.05	<0.05	0.19	ND	ND	ND	ND	ND	ND	ND	ND	ND
Total carbon	0.082	<0.05	<0.05	<0.05	0.068	12.92	13.7	13.8	6.06	0	0	36.53	14.91	11.51
Total nitrogen	<0.05	<0.05	<0.05	<0.05	<0.05	ND	ND	ND	ND	ND	ND	ND	ND	ND
Al	ND	ND	ND	ND	ND	0.91	0.75	1.71	ND	ND	ND	0.78	1.14	0.48
Ca	7,200	6,700	7,100	6,800	13,000	0.81	1.01	3.73	1.29	0	0.94	ND	ND	ND
Cl	ND	ND	ND	ND	ND	0.09	0.03	0.2	ND	ND	ND	ND	ND	ND
Fe	11,000	14,000	13,000	11,000	19,000	ND	ND	1.06	ND	ND	ND	ND	ND	ND
K	520	620	580	500	1,000	0.19	0.07	0.34	ND	ND	ND	ND	ND	ND
Mg	2,900	3,800	3,900	3,000	5,600	0.24	0.3	2.17	ND	1.39	ND	0.89	1.36	0.13
Na	1,400	2,200	1,200	1,500	2,000	ND	0.12	0.36	ND	ND	ND	0.34	0.87	ND
O	ND	ND	ND	ND	ND	63.56	64.8	60.33	72.61	74.01	74.85	48.97	62.8	65.43
P	240	370	310	250	350	21.29	19.23	16.26	20.04	23.66	25.15	ND	ND	ND

Moisture content, total organic content, and porosity are expressed as percentage wt/wt. Soluble salts, total carbon, and total nitrogen are expressed in g/100 g dry mass. Elemental analysis is expressed as mg/kg dry mass for soil and as percentage of relative abundance for rocks. NA, not applicable; ND, not detectable.

Table S5. Two-character codes for the taxa in Figs. 2–4

Code	Taxa
Ad	Acidobacteria
Ap	Alphaproteobacteria
Aq	Aquificae
Ar	Archaeoglobi
As	Ascomycota
At	Actinobacteria
Ba	Basidiomycota
Bd	Bacteroidetes
Bp	Betaproteobacteria
Cb	Chlorobi
Cd	Chlamydiae
Cf	Chloroflexi
CK	Candidatus Korarchaeum
Cr	Crenarchaeota
Cy	Cyanobacteria
Df	Deferribacteres
Dg	Dictyoglomi
Dp	Deltaproteobacteria
DT	Deinococcus-Thermus
Em	Elusimicrobia
Ep	Epsilonproteobacteria
Ey	Euryarchaea
Fb	Fibrobacteres
Fc	Firmicutes
Fu	Fusobacteria
Gm	Gemmatimonadetes
Gp	Gammaproteobacteria
Hb	Halobacteria
Ko	Korarchaeota
Ls	Lentisphaerae
Mb	Methanobacteria
Mc	Methanococci
Mm	Methanomicrobia
Mp	Methanopyri
My	Mycetozoa
Na	Neocallimastigomycota
Ns	Nitrospirae
Pm	Planctomycetes
Sc	Schizosaccharomycetes
Sp	Spirochaetes
Sy	Synergistetes
Ta	Thaumarchaeota
Tb	Thermobaculum
Tc	Thermococci
Tg	Thermotogae
Tm	Thermoplasmata
Tn	Tenericutes
Tp	Thermoprotei
Ua	Uncultured archaea
Ub	Uncultured bacteria
Uc	Uncultured crenarchaeote
Ue	Uncultured euryarchaeote
Uf	Uncultured fungi
Un	Unidentified
Up	Unclassified proteobacteria
Vm	Verrucomicrobia
Zp	Zetaproteobacteria
Zy	Zygomycota