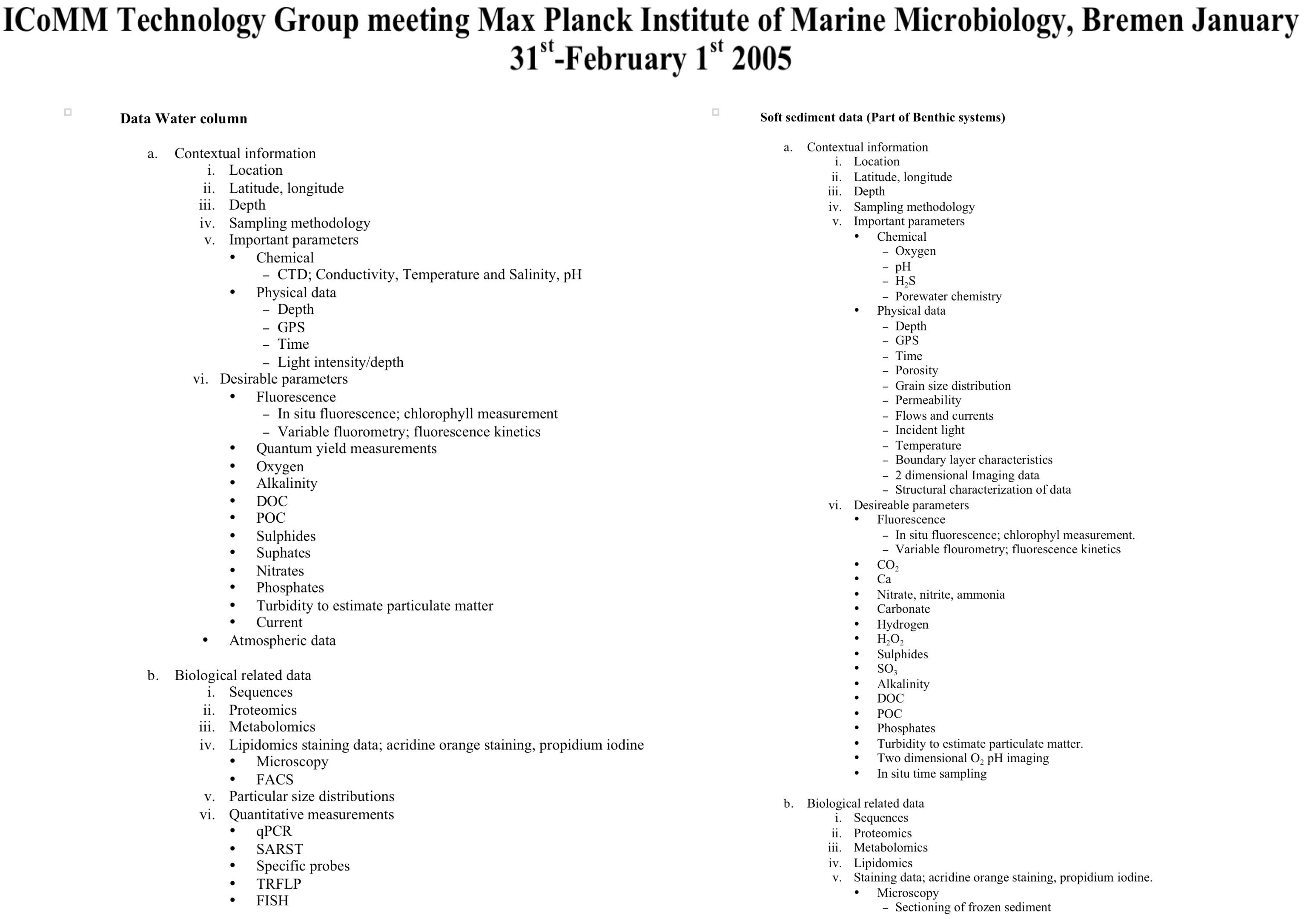

ICoMM Technology Group meeting Max Planck Institute of Marine Microbiology, Bremen January 31st-February 1st 2005

Data Water column

a. Contextual information
   i. Location
   ii. Latitude, longitude
   iii. Depth
   iv. Sampling methodology
   v. Important parameters
      - Chemical
        – CTD; Conductivity, Temperature and Salinity, pH
      - Physical data
        – Depth
        – GPS
        – Time
        – Light intensity/depth
   vi. Desirable parameters
      - Fluorescence
        – In situ fluorescence; chlorophyll measurement
        – Variable fluorometry; fluorescence kinetics
      - Quantum yield measurements
      - Oxygen
      - Alkalinity
      - DOC
      - POC
      - Sulphides
      - Suphates
      - Nitrates
      - Phosphates
      - Turbidity to estimate particulate matter
      - Current
   - Atmospheric data

b. Biological related data
   i. Sequences
   ii. Proteomics
   iii. Metabolomics
   iv. Lipidomics staining data; acridine orange staining, propidium iodine
      - Microscopy
      - FACS
   v. Particular size distributions
   vi. Quantitative measurements
      - qPCR
      - SARST
      - Specific probes
      - TRFLP
      - FISH

Soft sediment data (Part of Benthic systems)

a. Contextual information
   i. Location
   ii. Latitude, longitude
   iii. Depth
   iv. Sampling methodology
   v. Important parameters
      - Chemical
        – Oxygen
        – pH
        – H_2S
        – Porewater chemistry
      - Physical data
        – Depth
        – GPS
        – Time
        – Porosity
        – Grain size distribution
        – Permeability
        – Flows and currents
        – Incident light
        – Temperature
        – Boundary layer characteristics
        – 2 dimensional Imaging data
        – Structural characterization of data
   vi. Desireable parameters
      - Fluorescence
        – In situ fluorescence; chlorophyl measurement.
        – Variable flourometry; fluorescence kinetics
      - CO_2
      - Ca
      - Nitrate, nitrite, ammonia
      - Carbonate
      - Hydrogen
      - H_2O_2
      - Sulphides
      - SO_3
      - Alkalinity
      - DOC
      - POC
      - Phosphates
      - Turbidity to estimate particulate matter.
      - Two dimensional O_2 pH imaging
      - In situ time sampling

b. Biological related data
   i. Sequences
   ii. Proteomics
   iii. Metabolomics
   iv. Lipidomics
   v. Staining data; acridine orange staining, propidium iodine.
      - Microscopy
        – Sectioning of frozen sediment

Figure S1. Contextual data recommendations for water columns and soft sediment sequence data from the an International Census of Marine Microbes (ICoMM, http://icomm.mbl.edu) technology workshop.

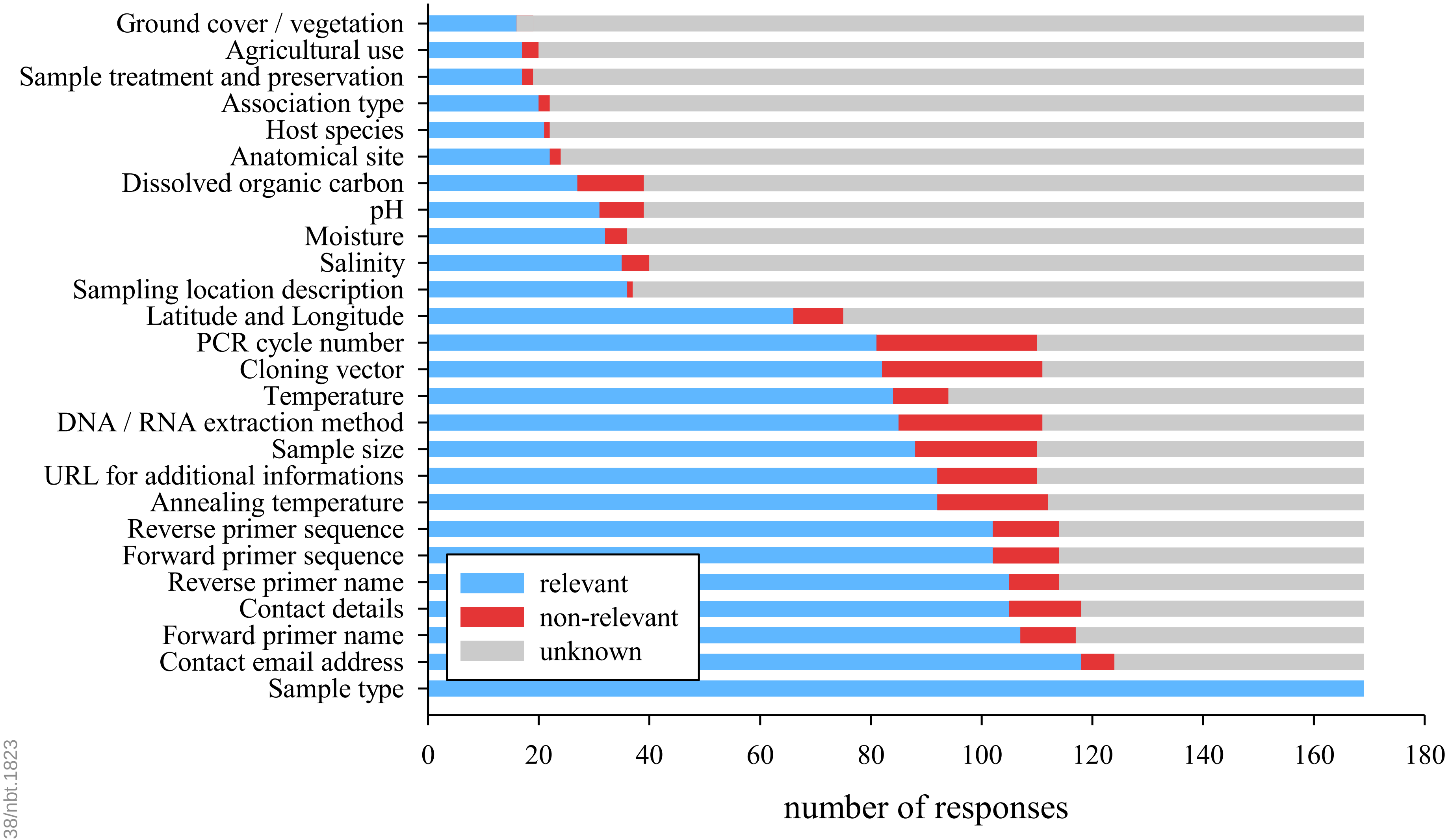

Figure S2. JGI metadata survey results. Bars represent the absolute number of relevant (blue), non-relevant (red) and unknown (gray) responses for a given item. The survey ran between May 2005 and August 2005 and 486 responses were obtained from 78 different locations within six months, with a maximum of 169 responses on one data field. The results shown in the diagram for the selected fields are based on a total number of the 169 responses. "Unknown response" was calculated by subtracting the total number of answers for a given data field from the total number of 169.

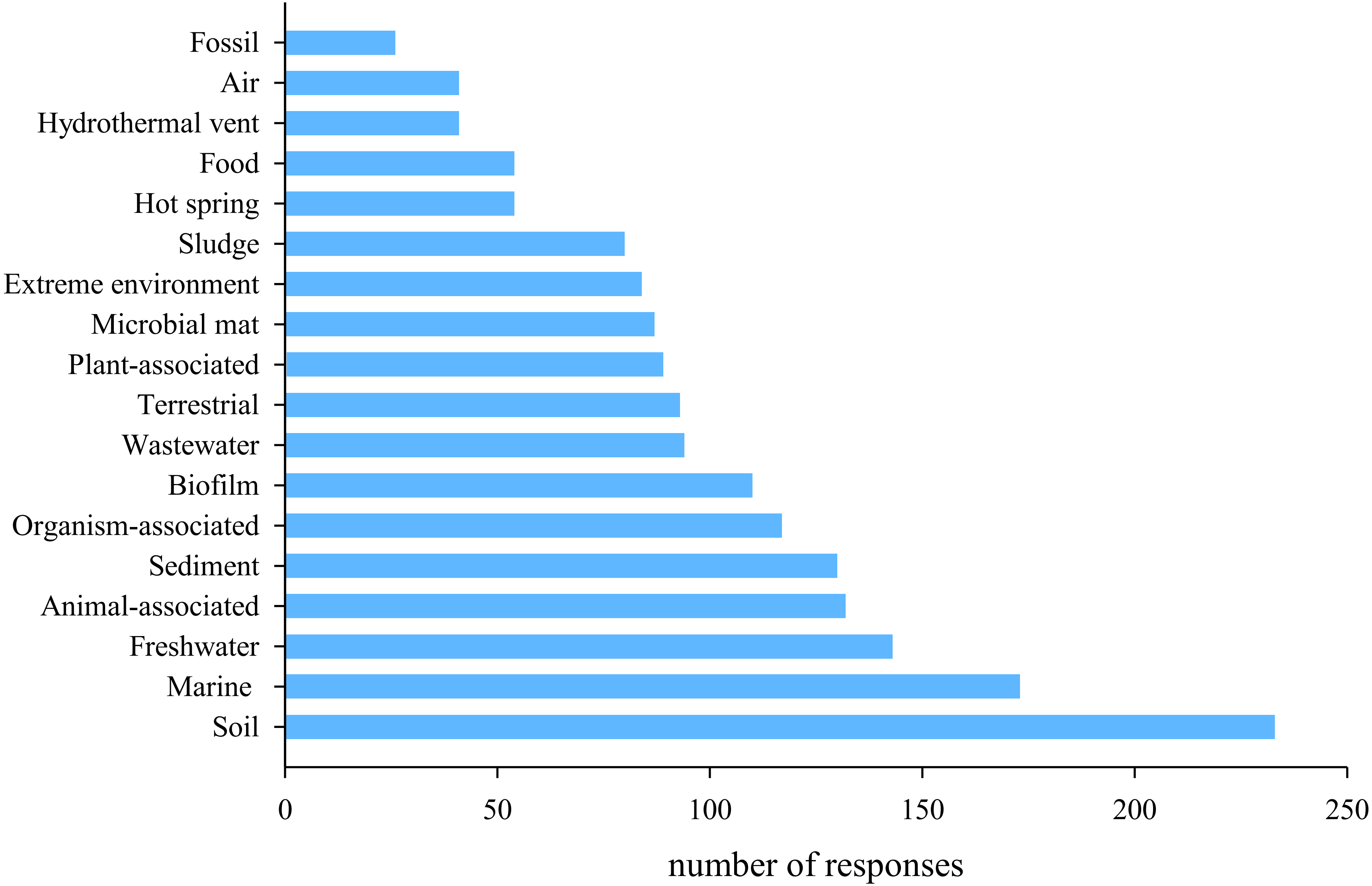

Figure S3. RDP user survey #5 for habitat descriptors results. Bars represent the absolute number of responses for a given item.

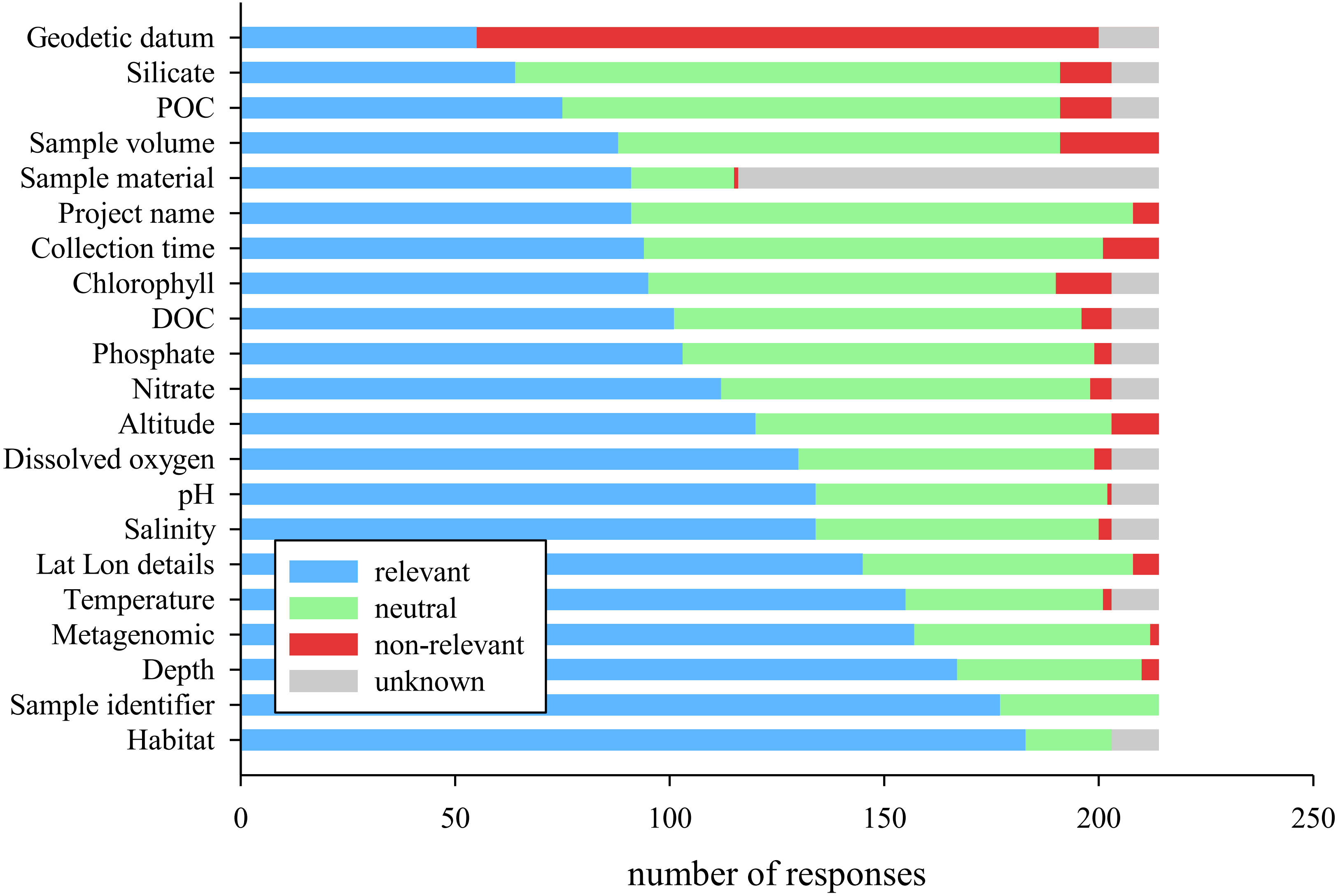

Figure S4. SILVA contextual data survey results. Bars represent the absolute number of relevant (blue), neutral (green), non-relevant (red) and unknown (gray) responses for a given item. Number of relevant responses are indicated on the bars. The survey was from May to October 2008. 214 responses from 26 different countries were acquired within a six month period with 182 complete responses. The responses are classified as "relevant", "neutral", "non-relevant" and "unknown response" for each suggested field. "Unknown response" was added as a result of questions that were not answered.

Nature Biotechnology: doi:10.1038/nbt.1823

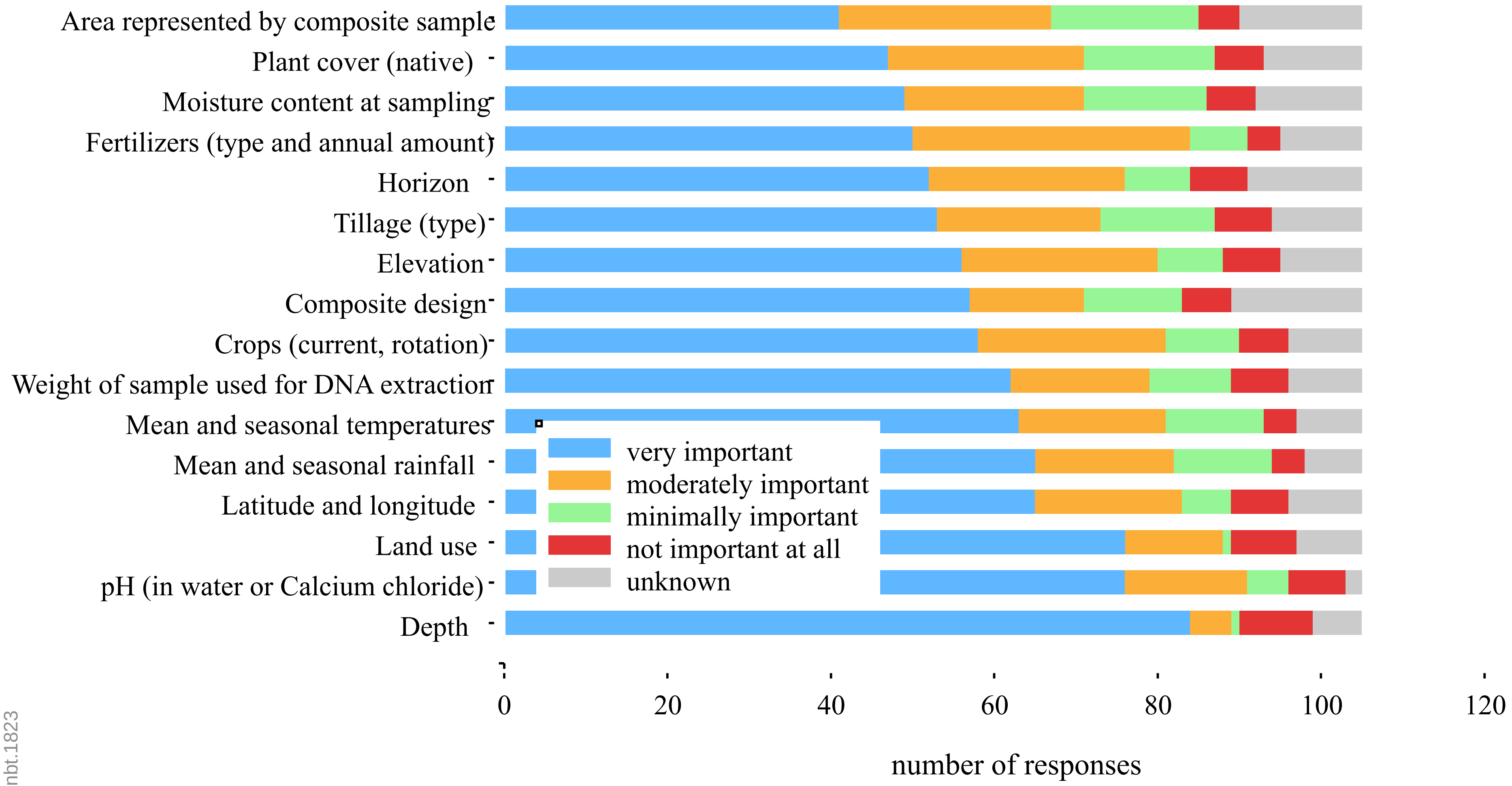

Figure S5. Soil metadata for microbial studies-importance survey results. Bars represent the absolute number of very important (blue), moderately important (orange), minimally important (green), not important at all (red) and unknown (gray) responses for a given item. The survey was ran between March 2009 and July 2009, and 105 unique IP responses were obtained. The survey items were grouped into seven categories as chemical, biological, geographical, physical, soil/geological, management and sampling. For each item, the respondent could rank the importance of each item and the difficulty of obtaining the particular field. For convenience the results displayed below are selected from items which ranked as being the most important and easiest to obtain.

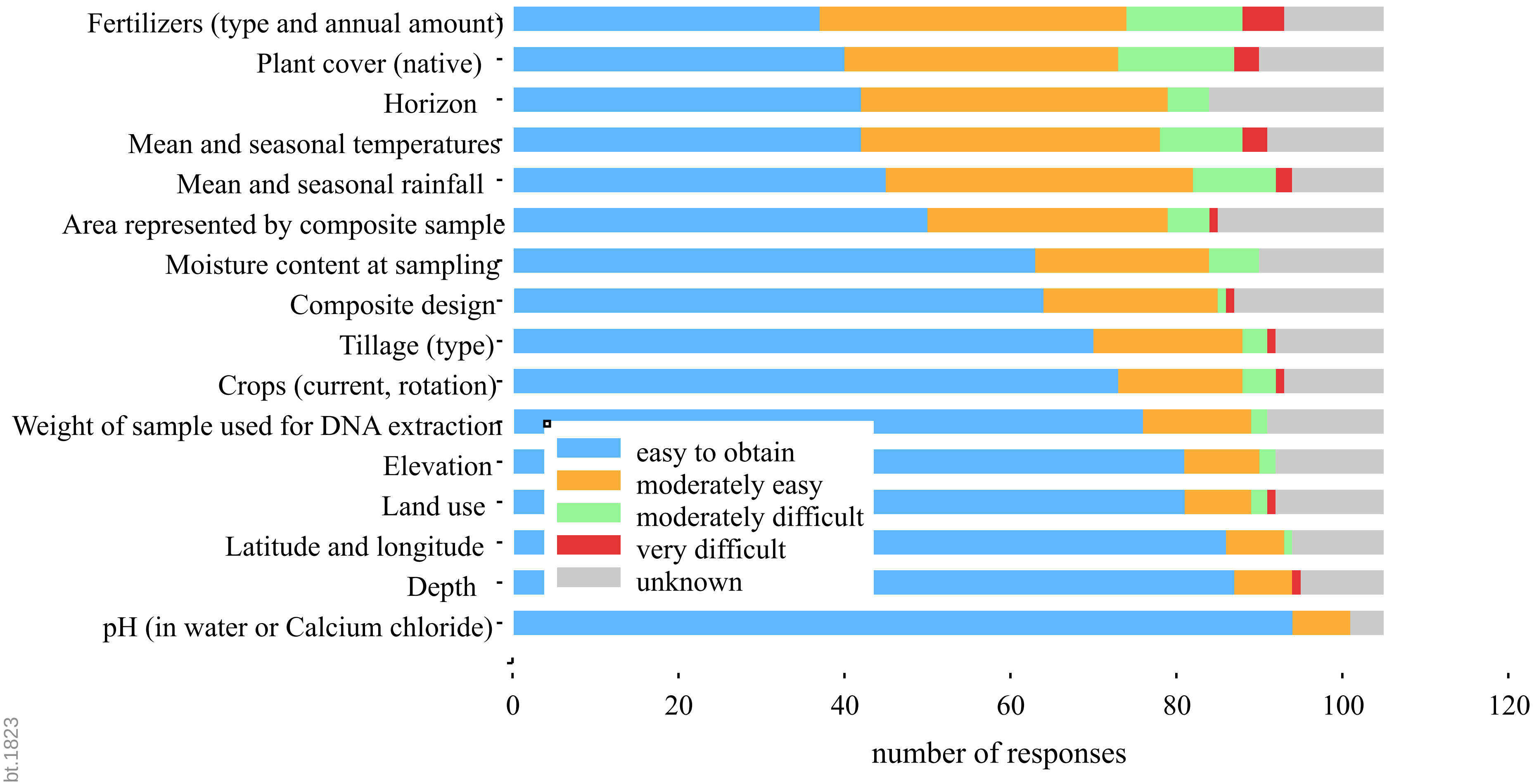

Figure S6. Soil metadata for microbial studies-difficulty survey results. Bars represent the absolute number of easy to obtain (blue), moderately easy (orange), moderately difficult (green), very difficult to obtain (red) and unknown (gray) responses for a given item. The survey was ran between March 2009 and July 2009, and 105 unique IP responses were obtained. The survey items were grouped into seven categories as chemical, biological, geographical, physical, soil/geological, management and sampling. For each item, the respondent could rank the importance of each item and the difficulty of obtaining the particular field. For convenience the results displayed below are selected from items which ranked as being the most important and easiest to obtain.

Nature Biotechnology: doi:10.1038/nbt.1823

Table S1. List of publications, the respective digital object identifiers (DOI), high-level habitat term associated with the study, the number of sequences that have been submitted to the INSDC databases and the metadata fields reported in each publication. The publications were selected on the basis of the number of submitted sequences, but the selection was extended to cover studies from different habitats uniformly. ‘[ ]’ denotes concentration. The table is continued in the following pages.

Title	DOI	Study Type	Number of sequences	Parameters
Pyrosequencing enumerates and contrasts soil microbial diversity	doi: 10.1038/ismej.2007.53	soil	149159	depth below surface elevation pH soil type
Evolution of mammals and their gut Microbes	doi: 10.1126/science.1155725	organism-associated (gut)	26160	age host species diet country
Assessment of bias associated with incomplete extraction of microbial DNA from soil	doi: 10.1128/AEM.00120-09	soil	21471	depth below surface pH total organic carbon [ ] particle classification (silt, clay %)
Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases	doi: 10.1073/pnas.0706625104	organism-associated (gut)	15172	sample size (cm^2) disease status sample preservation
Global diversity in the human salivary microbiome	doi: 10.1101/gr.084616.108	organism-associated (oral)	14115	age country sample size sample preservation

Towards the definition of a core of microorganisms involved in anaerobic digestion of sludge	doi: 10.1038/ismej.2009.2	sludge	12057	reactor type sewage type temperature sludge retention time efficiency industrial effluent % 1º treatment 2º treatment
Diversity of the human intestinal microbial flora	doi: 10.1126/science.1110591	organism-associated(gut)	11831	age medication anatomical site sample preservation sampling procedure-biopsy
A core gut microbiome in obese and lean twins	doi: 10.1038/nature07540	organism-associated(gut)	9773	age race/ethnicity BMI medication
A diversity profile of the human skin microbiota	doi: 10.1101/gr.075549.107	organism-associated (skin)	6213	age sex ethnicity disease status sample size lab host strain
Changes in microbial community structure in the wake of hurricanes Katrina and Rita	doi: 10.1021/es801904z	lake/lacunial	5804	sampling device filter pore size sample preservation
Reciprocal gut microbiota transplants from zebrafish and mice to germ-free recipients reveal host habitat selection	doi: 10.1016/j.cell.2006.08.043	organism-associated (gut)	5534	host strain/species diet host growth conditions

Nature Biotechnology: doi:10.1038/nbt.1823

Short-term temporal variability in airborne bacterial and fungal populations	doi: 10.1128/AEM.01467-07	air	5196	altitude filter pore size temperature windspeed wind direction humidity barometric pressure solar irradiance
Extensive phylogenetic analysis of a soil bacterial community illustrates extreme taxon evenness and the effects of amplicon length, degree of coverage, and DNA fractionation on classification and ecological parameters	doi: 10.1128/AEM.01757-08	soil	4891	dominant vegetation soil type sample processing
Relative impacts of land-use, management intensity and fertilization upon soil microbial community structure in agricultural systems	doi: 10.1016/j.soilbio.2008.07.030	soil	3706	soil type dominant vegetation sampling device sample size conductivity total carbon [ ] total nitrogen [ ] particulate organic carbon [ ] particulate organic nitrogen [ ]
Novelty and uniqueness patterns of rare members of the soil biosphere	doi: 10.1128/AEM.00410-08	soil	3663	dominant vegetation soil type pH NO_3 [ ] SO_3 [ ] HCO_3^- [ ] Cl [ ] Boron [ ] Na [ ] Ca [ ] total organic matter [ ] total organic nitrogen [ ]

Bacterial and fungal community structure in arctic tundra tussock and shrub soils	doi: 10.1111/j.1574-6941.2006.00260.x	soil	3338	dominant vegetation depth below surface NH_4 [ ] soil type NO_3 [ ] climate
Temporal and spatial variability in nearshore bacterioplankton communities of lake michigan	doi: 10.1111/j.1574-6941.2008.00639.x	lake/lacunial	2717	temperature conductivity lake type filter pore size
Microbes on mountainsides: contrasting elevational patterns of bacterial and plant diversity	doi: 10.1073/pnas.0801920105	soil	2196	altitude soil horizon temperature total carbon [ ] total nitrogen [ ] pH moisture
Effects of host plant environment and ustilago maydis infection on the fungal endophyte community of maize (zea mays)	doi: 10.1111/j.1469-8137.2007.02350.x	organism-associated(plant)	1576	sampling host strain anatomical site
Culturable aerobic bacteria from the upstream region of a karst water rivulet	doi: 10.2436/20.1501.01.48	lake/lacunial	944	pH Ca [ ] CO_2 [ ] partial pressure
Novel root fungal consortium associated with a dominant desert grass	doi: 10.1128/AEM.02769-07	organism-associated(plant)	657	sampling host anatomical site

Biogeochemistry and community composition of iron- and sulfur-precipitating microbial mats at the chefren mud volcano (nile deep sea fan, eastern mediterranean)	doi: 10.1128/ AEM. 01751-07	microbial mat/mud volcano	520	CH_4 [ ] SO_4 [ ] organic carbon [ ] Na, Cl [ ] $Fe^{+2/+3}$ [ ] H_2S [ ]
Spatial and temporal patterns in the microbial diversity of a meromictic soda lake in washington state	doi: 10.1128/ AEM. 00455-08	lake/lacunial	508	lake type salinity pH depth below surface dissolved O_2 [ ] dissolved organic carbon [ ] total nitrogen [ ] NO_3 [ ] NO_2 [ ] PO_4 [ ] total organic carbon [ ] H_2S [ ] total phosphorus [ ]
Active root-inhabiting microbes identified by rapid incorporation of plant-derived carbon into RNA	doi: 10.1073/ pnas. 0705902104	organism-associated (plant)	467	anatomical site dominant vegetation
Changes in bacterial and archaeal community structure and functional diversity along a geochemically variable soil profile	doi: 10.1128/ AEM. 01787-07	soil	460	soil type soil horizon pH particle size moisture organic carbon [ ] total nitrogen [ ] total carbon [ ]

Species richness and phylogenetic diversity comparisons of soil microbial communities affected by nickel-mining and revegetation efforts in new Caledonia	doi: 10.1016/j.ejsobi.2006.12.001	soil	435	dominant vegetation soil horizon depth below surface sieve pore size total carbon [ ] total nitrogen [ ]
Microbial community succession in an unvegetated, recently deglaciated soil	doi: 10.1007/s00248-006-9144-7	soil	411	elevation soil type rock type dominant vegetation depth below surface sieve pore size total nitrogen [ ] total carbon [ ] pH labile phosphorus content
The infuence of ultramafic rocks on microbial communities at the Logatchev hydrothermal field, located 15N on the Mid-Atlantic ridge	doi: 10.1111/j.1574-6941.2007.00325.x	hydrothermal vent	354	sample material filter pore size temperature pH H_2S [ ] CH_4 [ ] dissolved H_2 [ ]
An oligarchic microbial assemblage in the anoxic bottom waters of a volcanic subglacial lake	doi: 10.1038/ismej.2008.124	lake/lacunial	334	temperature pH trophic status particulate organic carbon [ ] total dissolved solids [ ] sample size depth below surface dissolved O_2[ ] SO_4[ ] H_2S [ ]

Diversity and ubiquity of thermophilic methanogenic archaea in temperate anoxic soils	doi: 10.1111/j.1462-2920.2005.00904.x	soil	323	soil type dominant vegetation country depth below surface
Comparative analysis of bacterial diversity in freshwater sediment of a shallow eutrophic lake by molecular and improved cultivation-based techniques	doi: 10.1128/AEM.71.4.2162-2169.2005	sediment (lake/lacunial)	322	temperature pH dissolved O_2 [ ] turbidity trophic status depth below surface water content pH total carbon [ ] total nitrogen [ ] particle classification (silt, clay %)
Bacterial diversity in a subseafloor habitat following a deep-sea volcanic eruption	doi: 10.1111/j.1574-6941.2003.tb01080.x	sediment (marine/hydrothermal vent)	287	filter pore size temperature pH H_2S [ ] Mg [ ] Cl [ ] Si [ ] $Fe^{+2/+3}$ [ ] salinity

Nature Biotechnology: doi:10.1038/nbt.1823

Comparison of deep-sea sediment microbial communities in the Eastern Mediterranean	doi: 10.1111/j.1574-6941.2008.00463.x	sediment (marine)	200	core size depth below surface water content density total nitrogen [ ] total carbon [ ] pH total organic carbon [ ] NH_4 [ ] NO_2 [ ] NO_3 [ ] SO_4 [ ] S^{-2} [ ] PO_4 [ ] Br [ ] Cl [ ]
Comparison of microbial communities associated with phase-separation-induced hydrothermal fluids at the Yonaguni Knoll IV hydrothermal field, the Southern Okinawa Trough	doi: 10.1111/j.1574-6941.2008.00636.x	hydrothermal field	140	sample material temperature H_2 [ ] CH_4 [ ] CO_2 [ ] Cl [ ]
Variability in microbial community and venting chemistry in a sediment-hosted backarc hydrothermal system: Impacts of subseafloor phase-separation	doi: 10.1016/j.femsec.2005.03.007	hydrothermal vent	112	temperature pH H_2S [ ] Mg[ ] Cl [ ] Si [ ] K [ ] Ca [ ]
Temporal and spatial archaeal colonization of hydrothermal vent deposits	doi: 10.1111/j.1462-2920.2007.01505.x	hydrothermal vent	89	sample material temperature

Table S2. Metadata field statistics from SILVA Parc small subunit (ssu) rRNA databases. The numbers are absolute for each item, and "-" denotes that the item has not been parsed for a release. The parsed metadata is obtained from source feature key qualifiers and other feature keys from EMBL with each release. The total number of sequences in the parc ssu databases are 422987, 461823, 504295, 566047, 606879, 682303, 756668, 868390, and 995747 from release 90 through 100, respectively. The table is continued in the following pages.

	alternative name	biomaterial	clone	clone library	collection date	collector	country	culture collection
silva_r90_ssu	885	-	273685	-	18303	7379	118711	-
silva_r91_ssu	926	-	303016	-	23234	8333	135901	-
silva_r92_ssu	966	-	337132	-	33907	14142	160048	-
silva_r93_ssu	1090	0	382364	-	42192	21065	187480	-
silva_r94_ssu	1152	0	413190	-	51514	25948	211498	-
silva_r95_ssu	1356	0	477631	20474	57329	28932	245783	-
silva_r96_ssu	1368	0	539395	22977	81066	35033	282935	-
silva_r98_ssu	1613	0	629885	26227	94024	39208	326341	-
silva_r100_ssu	2169	0	693317	34091	111891	57760	486501	1651

	description	haplotype	habitat	identified by	isolate	isolation source	lab host	latlong
silva_r90_ssu	422987	-	-	-	45870	280919	-	15835
silva_r91_ssu	461823	-	-	-	49446	314661	-	17954
silva_r92_ssu	504295	-	-	-	52022	351415	-	27456
silva_r93_ssu	566047	-	-	-	54976	404797	-	39446
silva_r94_ssu	606879	-	-	-	58472	440066	-	46139
silva_r95_ssu	682303	-	-	-	61407	511139	-	50191
silva_r96_ssu	756668	-	-	-	65576	579739	-	63577
silva_r98_ssu	868390	-	-	-	69612	676049	-	72084
silva_r100_ssu	995747	448	8699	11386	72320	936127	995	94337

	molecule type	organelle	organism name	pcr primers	plasmid name	specific host	strain	subspecies
silva_r90_ssu	422987	5336	422987	-	-	47009	83107	-
silva_r91_ssu	461823	5594	461823	-	-	53129	87690	-
silva_r92_ssu	504295	5613	504295	-	-	57339	93891	-
silva_r93_ssu	566047	5975	566047	0	-	76534	103499	-
silva_r94_ssu	606879	6010	606879	0	-	80904	109149	-
silva_r95_ssu	682303	6338	682303	0	-	110662	116464	-
silva_r96_ssu	756668	6580	756668	0	-	136667	124204	-
silva_r98_ssu	868390	8846	868390	0	-	-	137399	-
silva_r100_ssu	995747	10380	995747	0	75	235618	161578	5478