

Biogeography and Landscape-Scale Diversity of the Dominant Crenarchaeota of Soil

David K. Oline¹, Steven K. Schmidt² and Michael C. Grant²

(1) Biology Department, Southern Oregon University, 1250 Siskiyou Blvd., Ashland, OR 97520, USA

(2) Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, Campus Box 334, Boulder, CO 80309, USA

Received: 13 April 2005 / Accepted: 22 April 2005 / Online publication: 15 August 2006

Abstract

We surveyed the diversity of soil Archaea across a large scale elevational gradient of ecosystem types, from foothills forest to alpine tundra in the Front Range of the Rocky Mountains. We used a dilution technique to sequence the single most abundant archaeal 16S rDNA sequence in each of the 40 soil cores distributed across the gradient to compare our results to those of typical 16S clone library studies. We found a greater diversity of sequences than has typically been found in clone library studies from a single site or core, identifying sequences both from the Terrestrial Group and the FFSB Group at several sites. We did not observe any significant environmental correlates with the dominant sequence type, nor was there any relationship between the spatial distance between samples and the phylogenetic similarity of the dominant sequence types. Despite using a very different methodology, our collective results are in remarkably good agreement with other studies of soil Crenarchaeota in terms of the diversity and relative abundance of sequence types identified. We are able to identify two instances of very tightly clustered sequences which we suggest are the results of global selective sweeps—one closely related to SCA1145, an abundant globally distributed group within the Terrestrial Group of Crenarchaeota, and another nested within the more basal FFSB group of sequences. We replicated our sequence results at two levels: first, by repeating the dilution and PCR processes from the same soil core DNA extraction, and second, by performing a replicate DNA extraction from the same homogenized soil core sample. Pairs of sequences produced by the dilution replicates were significantly more similar than the pairs of sequences produced by the extraction replicates, suggest-

ing that soil Crenarchaeota exists in highly localized and discrete clonal populations.

Introduction

The use of molecular methods to investigate uncultivated microbes from natural environments has revolutionized our views of microbial biodiversity and ecology in recent years. Studies based on the extraction of total community DNA from environmental samples followed by polymerase chain reaction (PCR), cloning, and sequencing of 16S rRNA genes have now become commonplace, often comprising one of the first steps in studying the microbiology of an environment of interest.

Most of these studies begin by taking a small sample of the study environment and preparing a 16S rRNA clone library by using universal or bacterial specific PCR primers. Clones from the library are then sequenced by using one of two strategies: (1) sequencing clones at random, to obtain a random sample of all the Bacteria, or 2) screening clones by restriction fragment length polymorphism (RFLP), typing them, and then sequencing a clone of each distinct RFLP type. The latter strategy enhances the diversity of sequences obtained, but does not provide a random, unbiased sample of the community composition. Among the insights gained from the use of this culture-independent molecular approach have been (1) the confirmation of the long-held belief that the vast majority of microbes are not cultivated with currently used techniques, (2) the discovery of numerous novel division-level clades of Bacteria and Archaea that have no cultivated representatives (and whose basic metabolism is thus completely unknown), and (3) the realization that some environments for which we thought we had a basic microbial ecological understanding are

Correspondence to: David K. Oline; E-mail: olined@sou.edu

either dominated by or have a major component consisting of these uncultivated groups.

Although such studies using the standard 16S rRNA gene clone library technique have documented the existence of numerous novel groups of Bacteria and Archaea at every level from the division level to the strain, it is difficult to generalize patterns of distribution and diversity from their results. Because the generation and sequencing of a 16S rRNA gene clone library is both labor-intensive and expensive, researchers typically work with a small number of samples in any given study. For the same reasons, true replication and statistical characterization of clone libraries from an environment, performed in the way that we demand of most other ecological methodologies, is rarely achieved.

In this study, we take a different conceptual and methodological approach; rather than gaining a lot of information from a few samples, we gain some information from a large number of samples. We accomplish this by using a dilution technique; we first perform PCR on a dilution series of a soil DNA extract, which dilutes to extinction all but the most abundant sequences. We then reamplify and sequence the PCR product at the highest dilution for which the product was observed. Because the dilution is performed prior to the PCR, potential issues of PCR bias are minimized. Furthermore, this technique will only recover those sequences that are most numerically abundant (and presumably ecologically important) in the sample; rare sequences are not detected.

We also examine the results produced by this dilution technique by using two levels of replication. The first level of replication, which we call a “dilution replicate,” consists of creating a new dilution series from the original DNA extract, and then repeating the PCR, reamplification, and sequencing to obtain a second 16S rRNA gene sequence. The second level of replication, which we call an “extraction replicate” and in which we draw off a separate 0.5 g of soil from the original homogenized soil core sample, repeats the DNA extraction, PCR, and sequencing to obtain a replicate of the 16S rRNA gene sequence.

To address questions of larger scale patterns of distribution and diversity across a landscape, we have used a multiscale nested sampling approach across the large-scale elevational gradient of the Front Range of the

Rocky Mountains in Northern Colorado. Our sampling scheme uses transects with lengths varying over 4 orders of magnitude along which soil cores are sampled, enabling us to make comparisons of sequences generated by our dilution technique from distances ranging from 1 cm to 22 km. A recent study used a multiscale sampling technique to address questions of larger scale spatial structure of microbial communities using terminal RFLP (T-RFLPs), and found that communities that are closer together do tend to be more similar, even across large spatial scales [27].

We focused our study by examining the biogeography of an understudied group, the soil Crenarchaeota. Previous studies have shown that nonthermophilic Crenarchaeota comprise 1–1.5% of the microbial communities in temperate soils [6, 26]. These previous studies can be divided into those that have identified only members of the terrestrial group [4, 6] or the Finnish forest soil type B (FFSB) group [13, 14] (using the definitions of these groups as presented in [6]), although one recent study has identified members of both groups from grassland soils [21]. We also hope to discover some of the environmental correlates with the local dominance of these two clusters. One of the sites along our gradient is a montane forest, structurally and vegetatively similar to the boreal forest where members of the FFSB cluster have been previously identified as the dominant Crenarchaeota [13, 14].

Our goal is to use the dilution PCR technique to ask what are the abundant Crenarchaeota at our sampling sites, as well as specific questions of which groups may be associated with particular soil types and whether or not sequence similarity among the dominant groups is correlated with distance over a wide range of spatial scales.

Materials and Methods

Sampling Sites. For our large-scale environmental gradient, we used the elevational gradient of the Front Range of the Rocky Mountains in Northern Colorado. We chose four sites, each of which represents one of the major ecosystem types of the Front Range [17]: (1) foothills forest, dominated by ponderosa pine (*Pinus ponderosa*); (2) montane forest, dominated by ponderosa

Table 1. Locations and characteristics of sampling sites, including elevation, dominant vegetation, soil type, and annual precipitation

Site	Location	Elevation (m)	Vegetation	Soil type	Annual precipitation (cm)	
A	40.01N, 105.38W	1800	Foothills forest	Sandy loam	51	<i>Pinus ponderosa</i> <i>Carex heliophila</i>
B	40.02N, 105.43W	2200	Montane forest	Sandy loam	53	<i>Pinus ponderosa</i> <i>Psuedotsuga menziessi</i>
C	40.04N, 105.55W	2600	Subalpine forest	Sandy loam	69	<i>Pinus contorta</i> <i>Vaccinium myrtillus</i>
D	40.06N, 105.62W	3400	Alpine tundra	Silt loam	93	<i>Kobresia myosuroides</i> <i>Acomastylus rossi</i>

pine and Douglas fir (*Psuedotsuga menziesii*); (3) subalpine forest, dominated by lodgepole pine (*Pinus contorta*); and (4) alpine tundra, dominated by the sedge *Kobresia myosuroides* and alpine avens (*Acomastylis rossii*). Details and locations of the collection sites are shown in Table 1. The alpine tundra site is the Niwot Ridge Long-Term Ecological Research (LTER) site, which has been the site of numerous studies of vegetation, soils, and microbial processes. Together, the four sites comprise an elevational gradient ranging from 1800 m at the foothills site to 3400 m at Niwot Ridge. A straight-line distance of 22 km separates these sites, which lie roughly along a line extending west from the University of Colorado campus perpendicular to the continental divide, which trends in a north–south direction.

Numerous east–west canyons incise the eastern slope of the Front Range, and the four study plots are situated along the crest of a broad east–west ridge between two such canyons. Because marked differences between the microclimate and plant communities on north-facing versus south-facing ridges are typical of this portion of the Front Range [17], we selected the ridge top study plots to eliminate additional variance in soil properties that might be associated with these differences in solar exposure.

Sampling Scheme. At each of the four sites, we established a 100-m transect in a generally east–west direction, beginning from a randomly chosen starting point. We took our samples in June and July of 1999, beginning at the lower elevation A site and then proceeding to the higher elevation B, C, and D sites in successive weeks. This strategy attempts to minimize potential temporal variation in community structure as our sampling took place at a similar time relative to local snowmelt at each of the sites.

Within our 100-m transects, we took soil cores every 10 m. We then randomly chose one of the 10-m intervals and took cores every 1 m within that interval. We then randomly chose one of the 1-m intervals and took soil cores every 10 cm within that interval. Finally, we randomly chose one of the 10-cm intervals and established it as a 10-cm minitranssect, and took soil cores every 1 cm within this minitranssect, effectively removing a 1×10 cm slice from the soil. We repeated this sampling design at each of our four sampling sites along the elevational gradient. The four study sites along the elevational gradient are separated by 22 km, and are roughly evenly spaced with a 7-km average distance between adjacent sites; together, these form an effective 22-km transect. We performed measurements of soil water content, organic matter content, pH, and microbial biomass on all samples from all transects, the results of which have been previously reported [22]. For the

analysis of the crenarchaeotal communities, we chose a subset of 10 samples from each site to use for DNA extraction and sequence identification.

We took soil cores to a depth of 10 cm with a 1 cm diameter metal soil corer, individual sections of which were autoclaved prior to collection and brought into the field in individually sterilized and sealed plastic bags. After resealing each core section in its individual bag, we placed the cores on ice and returned them to the laboratory the same day.

Sample processing. After extruding the soil cores, we rapidly sieved the < 2 mm fraction of soil directly into a sterilized mortar in which it was ground with a sterilized pestle, before collecting the homogenized soil into sterile 50-mL tubes, which we then stored at -80°C . Although larger root sections were removed by the 2-mm sieve, we did not specifically remove roots from our samples as we include the rhizosphere within the total soil habitat.

DNA Extraction. We used a modification of the method described by Moré *et al.* [20], with additional purification steps that we found necessary in order to perform reliable PCR given the diversity of soils from which our samples were taken. The procedure consists of bead beating in a lysis buffer followed by ammonium acetate precipitation, phenol and chloroform extractions, and a final spin column purification, and is consistent with the recommendations of Miller *et al.* [19]. For the bead beating, we used 2-mL screw-cap tubes that were filled with 0.3 g of 1-mm-diameter glass beads, 0.3 g of 0.5-mm zirconia–silica beads, and 0.3 g of 0.1-mm zirconia–silica beads (BioSpec Products). We autoclaved the tubes containing the beads before adding soil equivalent to 0.5 g dry weight. We then added 0.5 mL of 100 mM sodium phosphate buffer (pH 8.0) to each tube, followed by 0.5 mL of lysis buffer (100 mM Tris–HCl, 100 mM NaCl, 10% SDS, pH 8.0). We then placed the samples in a Mini Bead Beater (BioSpec Products) for 2 min at a rate of 4200/min; when finished, the samples were immediately placed on ice.

After 10 min of centrifugation at $14,000 \times g$, we collected 500 μL of the supernatant into autoclaved 1.5-mL microcentrifuge tubes. We then added 200 μL of 7.5 M ammonium acetate and precipitated on ice for 5 min. After a 5-min centrifugation at $14,000 \times g$, we poured off the supernatant into new 1.5-mL tubes. We then added 500 μL equilibrated phenol (pH 8.0), inverted the tubes several times, and centrifuged at $14,000 \times g$ for 5 min. We collected 400 μL of the supernatant into fresh 1.5-mL tubes and added 500 μL of 24:1 chloroform/isoamyl alcohol, inverted the tubes several times, and centrifuged at $14,000 \times g$ for 5 min. We collected 300 μL of the supernatant into new 1.5-mL tubes, added 200 μL isopropanol, and precipitated at -20°C overnight. We pelleted the

Table 2. Nonredundant sequences obtained from the four sampling sites, including the GenBank accession no., length (nt), the dilution factor of the soil DNA extract at which the sequence was obtained, and the major clade affinity of the sequence

Sequence ID	Accession no.	Length	Dilution	Affinity
FRA0	AY016469	420	10 ⁻⁴	SCA1170
FRA1	AY016470	800	10 ⁻⁴	
FRA27	AY016471	836	10 ⁻⁵	FFSB
FRA27x2	AY016472	805	10 ⁻⁵	SCA1145
FRA31B	AY016473	804	10 ⁻³	SCA1145
FRA32	AY016474	805	10 ⁻³	SCA1145
FRA33	AY016475	792	10 ⁻⁴	SCA1145
FRA9	AY016476	806	10 ⁻⁵	
FRB1	AY016577	805	10 ⁻⁶	
FRB15	AY016478	805	10 ⁻⁴	SCA1145
FRB25	AY016479	805	10 ⁻⁵	
FRB27	AY016480	805	10 ⁻⁴	
FRB31	AY016481	804	10 ⁻⁴	SCA1145
FRB32B	AY016482	805	10 ⁻⁴	SCA1145
FRB32x2	AY016483	805	10 ⁻³	FRD
FRB33	AY016484	797	10 ⁻⁴	SCA1170
FRB38	AY016485	805	10 ⁻⁴	SCA1145
FRB9A	AY016486	805	10 ⁻⁴	SCA1145
FRC0	AY016487	469	10 ⁻⁶	FFSB1 cluster
FRC15	AY016488	803	10 ⁻⁷	SCA1145
FRC1B	AY016489	805	10 ⁻⁵	SCA1145
FRC1x2	AY016490	664	10 ⁻⁵	FFSB1 cluster
FRC27	AY016491	805	10 ⁻⁶	SCA1145
FRC32	AY016492	805	10 ⁻⁵	SCA1145
FRC33A	AY016493	664	10 ⁻⁴	FFSB1 cluster
FRC33B	AY016494	835	10 ⁻⁴	FFSB
FRC38	AY016495	805	10 ⁻⁵	SCA1145
FRC9	AY016496	805	10 ⁻⁴	FRD
FRD0	AY016497	426	10 ⁻⁴	SCA1170
FRD15	AY016498	805	10 ⁻⁴	
FRD25B	AY016499	805	10 ⁻⁶	FRD
FRD25x2	AY016500	460	10 ⁻⁶	FRD
FRD31	AY016501	805	10 ⁻⁴	FRD
FRD32	AY016502	805	10 ⁻⁴	FRD
FRD33	AY016503	805	10 ⁻⁵	SCA1145
FRD38	AY016504	835	10 ⁻³	FFSB1 cluster
FRD9	AY016505	805	10 ⁻⁵	
FRD9x2	AY016506	453	10 ⁻⁵	SCA1145 cluster

Dilution is provided as a reference, and is not intended for quantitative comparisons.

DNA by centrifuging for 10 min at 14,000 × g, poured off the isopropanol, and dried down the samples with a spin-vac. We then resuspended the DNA in 200 μL of water, and further purified the DNA solution with Glass-milk spin columns (Bio 101), eluting in water to a final volume of 50 μL. We performed replicate extractions from the same homogenized soil cores for a subsample of 10 soil cores.

PCR Amplification. For each of the samples, we made a tenfold dilution series from 10⁻¹ to 10⁻⁸. We then performed PCR (as described below) by using the 10⁻¹ to 10⁻⁸ dilutions as templates. The highest dilution that still produced a visible PCR product was then used as the template for reamplification. The reamplification step was necessary to produce sufficient quantities of DNA for sequencing, as the higher dilutions typically produced

very faint bands. By using this dilution approach, conceptually similar to the MPN technique, the highest dilution product will consist of only the most numerically abundant lineages in our soil sample.

We performed the initial PCR with primers CREN24F (5'-GGACCCGACYGCTATC-3') and ARCH927R (5'-CCCGCCAATTCCTTTAAGTTTC-3'), which together amplify the first ~900 bases of the 16S ribosomal RNA gene of the Crenarchaeota, which contains the V1, V2, and V3 variable regions. We used a 20 μL total reaction volume with a 60 mM Tris-HCl/15 mM (NH₄)₂SO₄ buffer, 3.0 mM MgCl₂, 0.4 mM dNTPs, 0.4 μM of each primer, and 1 U of *Taq* polymerase (Promega). After an initial denaturation step of 2 min at 94°C, we ran 40 cycles of 30 s at 94°C, 45 s at 52°C, and 90 s at 72°C, with a final extension step of 5 min at 72°C. We deliberately kept the PCR conditions at moderately low stringency to enhance

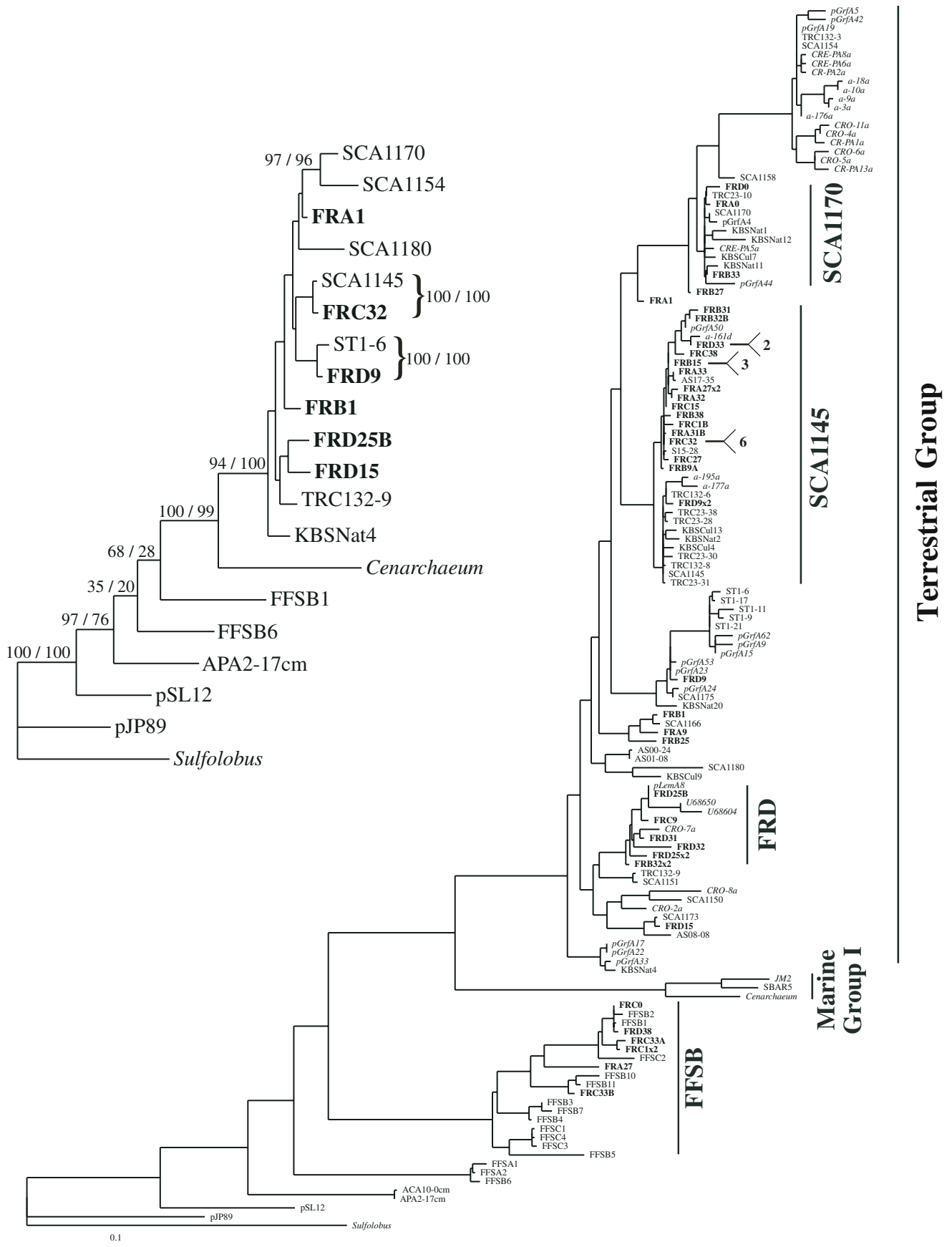


Table 3. Numbers (and percentages) of clones belonging to the SCA1145 and SCA1170 clades from this and previously published unbiased studies, in which random 16S rDNA clones from soil environments were sequenced

Study	SCA1145	SCA1170	Other clades
This study	22 (55%)	3 (8%)	15 (37%)
[4]	16 (47%)	2 (6%)	16 (47%)
[28]	7 (70%)	1 (10%)	2 (20%)
[26]	3 (38%)	0 (0%)	5 (62%)
[6]	3 (25%)	4 (33%)	5 (42%)

Despite low sample sizes in several studies, there is good agreement of the relative numbers despite the varying environments sampled and different techniques used.

the sensitivity to the low template concentrations in the dilution series. We visualized the PCR products on 1% agarose gels stained in ethidium bromide under UV transillumination. For the reamplification, we used 1 μ L of a 10^{-3} dilution of the initial PCR product and identical PCR conditions except for using a 60 μ L reaction volume, 2.5 mM MgCl₂, and primers 26F (5'-CCCGACTGCTATCRGA-3') and 900R (5'-CTNCCCAGGCGGC-3'). These produced a product of typically 805 bases containing the V1, V2, and V3 variable regions of the 16S ribosomal RNA gene, which we then sequenced.

Sequencing. We prepared the PCR products from the reamplifications for sequencing by purifying with a Qia-Quick PCR cleanup column (Qiagen Corp). The sequencing was performed in the Department of Molecular, Cellular, and Developmental Biology at the University of Colorado, Boulder, by using an ABI 377 sequencer. We used the internal sequencing primers SQ1 (5'-CTACGGATGCTTTAGGCC-3') and SQ2 (5'-TGGGCAATGCTGGTGTCA-3') to sequence the full ~805-bp product generated by the reamplification in two reactions. These two primers go in opposite directions from the midregion of the ~805-bp fragment with sufficient overlap to align the two fragments into a single sequence. On several occasions we observed sequencing chromatograms which indicated that the reamplified product did not consist of a single sequence type; these sequences were discarded and we went back

to the original DNA extraction, prepared a new dilution series, and repeated the PCR and sequencing procedures to obtain a pure sequence.

Sequence and Phylogenetic Analysis. We used known 16S rRNA secondary structure diagrams for *Sulfolobus solfataricus* and several of the FFSB group of nonthermophilic Crenarchaeota [7] as templates to create a structure-based alignment including our new sequences and additional crenarchaeotal sequences from GenBank. Our goal was to include sequences from all major studies of uncultivated Archaea that fell within the terrestrial group of the nonthermophilic Crenarchaeota, as defined in [6], in our phylogenetic analysis. Because not all of the GenBank sequences fully overlapped the ~805-bp region that we sequenced, we first performed a maximum likelihood analysis with fastDNAm1 [11, 23] using sequences containing the full ~805-bp region. Using this tree, we then grafted on the additional partially overlapping sequences by using the restart option in fastDNAm1 and disallowing local rearrangements. This technique allowed us to use a large number of sequences obtained from studies of diverse environments to identify major clades of terrestrial Crenarchaeota that have been found in multiple studies. Because of the large number of taxa and the incomplete overlap of all the sequences, we selected representative taxa containing the full ~805-bp region from major clades and performed bootstrap analyses on this subset of the data to estimate the level of support for the relationships of the major clades.

We selected two major clades for further analyses, the SCA1145 clade and the FFSB clade, to assess the fine structure of sequence relationships within them. For these analyses, we used only sequences containing the full ~805-bp region and performed both maximum likelihood (using the SEQBOOT and DNAML programs from the PHYLIP package) and parsimony (using PAUP version 4.0b4a) bootstrapping.

Replication. For the dilution replicates, we selected 10 DNA extracts at random and then made two dilution series from each extract. On each of these dilution series, we performed initial PCR, reamplification, and sequenc-

Figure 1. Maximum likelihood phylogenetic tree of Crenarchaeota based on 16S rRNA genes including the sequences identified from this study (in bold type), which are denoted by the prefix FR- with the third letter representing the sampling site. Sequences with the prefix SCA- are from [4], SBAR5 is from [10], JM2 is from [18], pJP89 and pSL12 are from [3], the prefixes pGrf- and pLem- are from [12], U68650 and U68604 are from [5], prefixes FFS- are from [13] and [14], prefixes KBS- are from [6], prefixes CR- are from [9], prefixes ACA- and APA- are from [30], prefixes AS- are from [15], prefixes TRC- are from [26], prefixes S15- and ST1- are from [8], and prefixes a- are from [26]. Sequences in italics are those that do not fully overlap with the ~800-bp region used for the full analysis (except for *Cenarchaeum* and *Sulfolobus*, which are italicized by convention); they were added using fastDNAm1 without allowing local rearrangements (see *Materials and Methods*). Identical sequences that we recovered multiple times from different soil samples are indicated with the number of different samples from which that sequence was recovered. The *inset* shows the bootstrap values (maximum likelihood/parsimony) for a subset of the sequences to show the confidence levels of the relationships between the major clades.

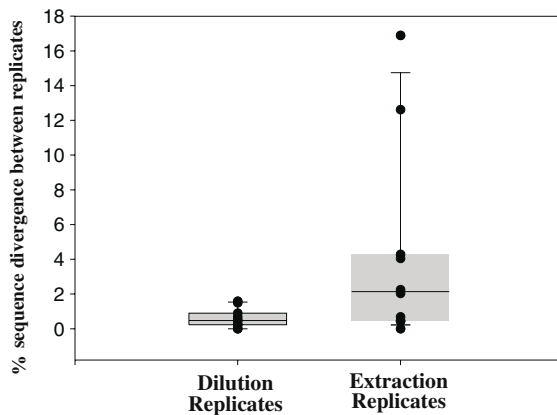


Figure 2. Box plot of the distribution of sequence difference values between pairs of sequences generated from dilution and extraction replicates, respectively. The median and the 10th, 25th, 75th, and 90th percentiles are indicated.

ing as described above. For each of the 10 pairs of dilution replicates, we then calculated the percentage of nonidentical bases for that pair. For the extraction replicates, we selected 10 soil cores at random and then withdrew two separate 0.5-g soil samples upon which we performed the DNA extraction, dilution series PCR, reamplification, and then sequencing as described above. For each of the 10 pairs of extraction replicates, we then calculated the percentage of nonidentical bases for that pair.

Nucleotide sequence accession numbers. The nonredundant sequences from this study have been deposited in GenBank and have accession numbers AY016469 through AY016506, and AF102692.

Results

We initially obtained 10 sequences from each of the four sampling sites, each of which is likely the most numerically dominant sequence in its particular soil sample (Table 2). Of these 40 sequences, a single sequence type, identical over ~805 bp, was identified as the most dominant in seven different soil cores. Three different sequence types were dominant in two soil cores each, and the remaining 27 soil cores produced dominant sequences that were unique.

A maximum likelihood phylogenetic hypothesis including all of the nonredundant sequences identified from this study and representative sequences from a range of other published studies of environmental Crenarchaeota is presented in Fig. 1. The Terrestrial Group, Marine Group I, and the FFSB group are as defined in [6]. At the base of the Terrestrial Group, there is poor resolution among the major lineages, although a number of well-defined clades are indicated. We define additional fine structure within the Terrestrial Group by identifying

several of these clades, which have consistently high bootstrap support: the SCA1145 group, the SCA1170 group (which are named for the first published sequences from these groups [4]), and the FRD group (named for its abundance at our alpine tundra site). We define these clades as consisting of a cluster of closely related sequences that have been identified from multiple studies in diverse environments.

Having defined these clades, we then used them to classify our 40 sequences. Twenty sequences (55%) belonged to the SCA1145 clade, 3 (7.5%) belonged to the SCA1170 clade, 4 (10%) belonged to the FRD clade, and 4 (10%) belonged to the FFSB clade; 7 (17.5%) were not readily classifiable, belonging to other clades within the Terrestrial Group (Tables 2 and 3). There were no significant associations between particular sites and particular clades, nor was there a significant relationship

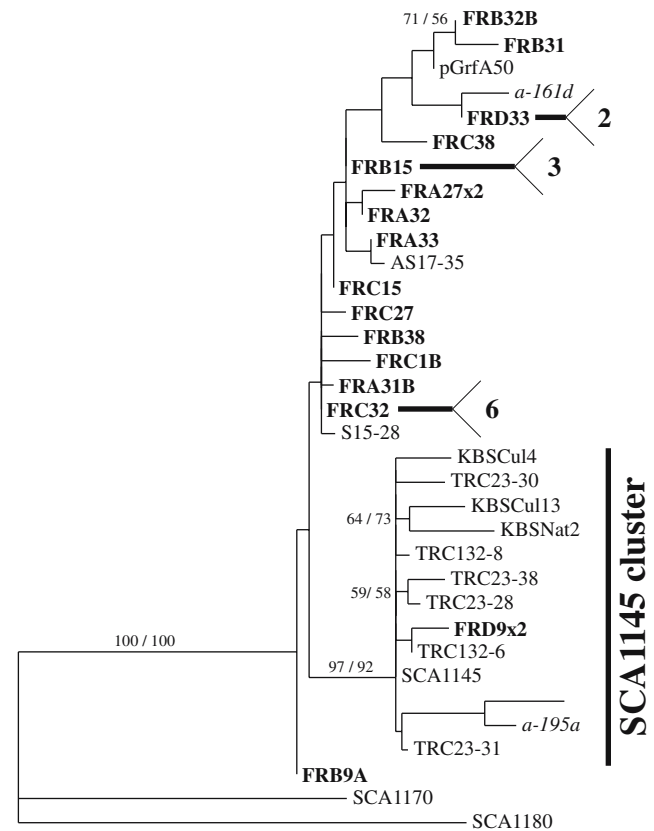


Figure 3. Maximum likelihood phylogenetic tree of the SCA1145 clade of sequences. Values on specific nodes are bootstrap support using maximum likelihood/parsimony criteria, respectively. Most sequences within this group are very closely related and thus phylogenetic resolution is very poor; however, support for a distinct cluster containing primarily SCA-, TRC-, and KBS-sequences is very strong. Sequences recovered from multiple different samples are indicated with the number of different samples from which that sequence was recovered. Sequences in italics do not fully overlap with the ~800-bp region of the FR-sequences and were not included in the bootstrap analysis.

reliably identify to which of the major clades (as defined if Fig. 1) the most numerically abundant lineage within the sample belongs. Furthermore, as the greatest divergence between our dilution replicate pairs was 98.4%, this suggests that this technique may be able to identify the dominant lineage to a higher resolution within these major clades.

The lack of exact sequence identity among the dilution replicate pairs suggests that either there is a great deal of fine-scale genetic diversity among the most abundant sequences within a sample, or that such sequence microheterogeneity may be generated by errors in PCR from a mixed template sample. The observation of sequence microheterogeneity is ubiquitous among unbiased studies (those which sequence clones at random and do not screen by RFLP type first) of environmental clone libraries. In this study, we have used a completely different dilution technique on 40 different samples and observe similar sequence microheterogeneity. The effect of using a highly diluted template in such mixed template PCR reactions has not been explored, but is likely to reduce the possibility of amplifying errors or artifacts, or producing chimeras [25]. Furthermore, we have identified a single sequence (FRA25), identical over 800 bp, from seven different soil samples at three different sampling sites, and have three other cases where identical sequences were found in two or three soil samples each. Analyses of the differences between closely related sequences show that many of these base differences between closely related sequences are phylogenetically informative as well as nonrandomly distributed, being most prevalent in the V2 and V3 variable regions of the 16S rRNA gene. Basic population genetic theory predicts that the most abundant species, which our dilution technique identifies, contain the most genetic variation [2], which is consistent with our results.

Our extraction replicates produced a much higher average (4.4%) and maximum (16.9%) divergence between sequences identified from two separate DNA extractions of the same homogenized soil sample (Fig. 2), and in four out of 10 cases, the replicate sequences were in different clades (Fig. 1). The two extraordinarily high values of sequence divergence (12.8% and 16.9%; Fig. 2), represent cases in which one sequence was in the FFSB clade and the other in the Terrestrial Group. Different 0.5-g subsamples of the soil core contained different dominant sequence types almost as often as not. This suggests that populations of soil Crenarchaeota, which comprise a minor fraction of the soil microbial community (from 1.1% to 1.4% of total prokaryotic cell numbers [6, 26]), exist in discrete clumps localized within the soil environment. We hypothesize that the withdrawal of a 0.5-g subsample effectively randomly samples a small number of discrete localized populations of Crenarchaeota by the chance inclusion of a discrete

localized population. If the number of these localized populations in 0.5 g of our soils is small, than random sampling becomes important and can explain why two different subsamples can be numerically dominated by different groups. Recent observations have shown that members of the Terrestrial Group are components of the rhizosphere [28, 29]. If the closely related soil Crenarchaeota we have identified from our soils are similarly associated with the rhizosphere, this could explain their discrete, clumped distribution because of the presence or absence of particular root fragments in different 0.5-g subsamples.

In spite of sampling across a steep environmental gradient of diverse soil types, we did not find any statistically significant associations of particular groups of Crenarchaeota with any of our four sampling sites (data not shown), although our sample size was fairly small. Members of the SCA1145 clade were found at all four sites, whereas members of the SCA1170, FRD, and FFSB groups were found at three sites. Although not a significant association, members of the FFSB clade were most dominant (three out of 11 samples) at the montane forest site, which is most ecologically similar to the boreal forest sites where members of this group have been exclusively detected in surveys of archaeal diversity [13, 14]. We speculate that low soil pH may be an important factor for the FFSB, as our soil sample that produced both extraction replicates from the FFSB group had a pH value of 4.35, one of the lowest of all our samples and similar to the values reported in [13] and [14] at their boreal forest sites. The alpine tundra site, despite being markedly different from the three forested sites in site and soil characteristics (Table 1), did not produce a markedly different group of sequences—sequences from all four phylogenetic groups were identified from this site.

Of our 40 samples, 22 (55%) were dominated by the SCA1145 clade and 3 (7.5%) were dominated by the SCA1170 clade. These percentages are remarkably similar to those reported in [4]: out of 34 clones, 16 (47%) were closely related to SCA1145, and 2 (6%) were closely related to SCA1170 (Table 3). This degree of similarity of relative abundance exists in spite of the fact that the sequences in [4] were identified from only a single sample of an agricultural soil, whereas we have used 40 different samples across an environmental gradient including diverse forest and tundra soil types. Although the sample sizes used by Bintrim *et al.* [4], Sandaa *et al.* [26], and Simon *et al.* [28] are small, there is still good agreement among the relative numbers of each clade despite the diversity of soils sampled and the techniques used. This agreement is suggestive that the microhabitats to which these different clades are adapted occur in similar relative abundances regardless of scale. Additionally, they suggest that the SCA1145 clade is perhaps the most

important group within the Terrestrial Group of non-thermophilic Crenarchaeota, being numerically dominant and cosmopolitan in a wide range of soil environments.

Our nested multiscale sampling scheme allowed us to compare the genetic similarity among the dominant 16S rRNA genes from different soil samples across a range of distances from 1 cm to 22 km; however, we found no significant association of sequence similarity with distance (data not shown). Not only did we identify the exact FRC32 sequence at three of our sampling sites, but we also identified multiple pairs of adjacent soil cores (1 cm apart) in which one was dominated by a member of the FFSB group and the other dominated by a member of the Terrestrial Group (approximately 16% sequence divergence). Additionally, the results of our extraction replicates suggest that the changeover between dominant groups may shift even within a single 1 cm diameter soil core.

Our more detailed analysis of the sequences within the SCA1145 clade shows that there exists within the larger clade a well-supported cluster of nearly identical sequences arising from within a poorly resolved basal group of sequences. A close analysis of the sequences identifies specific signature bases that are diagnostic for this SCA1145 cluster (Table 4). Recently published results [21] presented four sequences (SUPA1, SUPA2, SUPA3, and SUPA4) all of which share these signature sequences and so belong to the SCA1145 cluster. A recently deposited, but as yet unpublished, sequence from the waters of a South African gold mine (AB050221) also contains all of the signature sequences shown in Table 4 to identify it as a member of this cluster. This cluster of sequences, all of which share 99% sequence identity, is thus truly globally distributed, in addition to being locally abundant at some sites. Furthermore, because all of these sequences share the signature bases shown in Table 4, they can be distinguished as a distinct lineage within the larger, similarly globally distributed SCA1145 clade.

A similar analysis of sequences within the FFSB clade revealed a number of signature sequences that define the well-supported FFSB1 cluster within that clade (Table 3). This group contains sequences from two different studies of boreal forest soil and from three of our sampling sites, indicating that this group can be locally abundant in a diversity of soil types. In addition, Jurgens and Saano [14] reported that 82 out of 138 (57%) of their forest soil clones had RFLP types that would presumably place them in this group (by virtue of matching either the FFSB1, FFSB2, or FFSC2 RFLP pattern; see Table 1 in [14]). Although there are relatively few known sequences belonging to the FFSB group, the FFSB1 cluster of very closely related sequences appears to be dominant in multiple studies. The SUPA7 sequence recently reported by Nicol *et al.* [21] shares all of these FFSB1 signature

sequences over the range of overlap with the sequences reported by this study and those of Jurgens *et al.* [13] and Jurgens and Saano [14]. Our use of the dilution PCR technique across a landscape scale environmental gradient has detected a wider diversity of sequence types than has been identified from most studies, which have a small number of samples from just one or two field sampling sites. In spite of this, our collection of the most abundant archaeal types is in remarkably good agreement with the relative frequencies of these same groups in studies from even single samples.

We have identified four well-defined major clades within the Terrestrial Group that have been identified from diverse soil types and studies on multiple continents. We suggest, based on their abundances in this and other studies, that these groups represent the major groups of globally distributed Terrestrial Crenarchaeota. Furthermore, within two of these major groups we have identified two very narrowly defined clusters of closely related sequences as defined by sets of unambiguous signature bases that are both globally distributed. The identification of this pattern within two groups of the nonthermophilic Archaea suggests that this may be a general pattern of evolution and diversification within these groups, and perhaps natural microbial communities in general. These data are consistent with the theoretical predictions of Palys *et al.* [24] and Majewski and Cohan [16], who predicted that ecological and evolutionary processes acting on clonal bacterial lineages should produce patterns in which large fractions of the total community would fall into distinct clusters. As 79% of our sequences fall within four groups of 99% relatedness, our data are similar to those of Acinas *et al.* [1], who showed that over 50% of the bacterial sequences from a large clone library fell into 99% relatedness groups. Importantly, we suggest that these patterns, and perhaps other intriguing patterns, can only be identified by using a scale of taxonomic resolution that is much finer than that obtained by using universal or bacterial specific 16S rRNA gene primers. Surveys of the entire bacterial community are an important first step in studying a particular environment, but the macroecological properties are so vast in comparison to the scale of sampling, typically a few hundred clones, that only a very coarse resolution view of the community can be obtained in this way. Many of the results of ecological and evolutionary processes that we are interested in, such as the globally distributed SCA1145 and FFSB1 clades we have identified, can only be observed by looking with an appropriately fine taxonomic resolution.

Acknowledgments

The authors wish to thank J. Mitton, B. Kreiser, R. Ramey, A. Martin, T. Seastedt, and B. Bowman. We also

thank the Arapahoe-Roosevelt National Forest, the University of Colorado Mountain Research Station, and the Niwot Ridge LTER site for their support. Thanks to R. Gutell for secondary structure diagrams of rRNA sequences, R. Sandaa for providing sequence data, and H. Simon for providing a preprint of her manuscript. Special thanks to samiam.colorado.edu for an outstanding effort devoted to the cause of molecular phylogeny.

Funding for this research was provided by the Colorado Mountain Club, the David Paddon Memorial Scholarship of the Indian Peaks Working Group, the Beverley Sears Deans Small Grants, Sigma Xi, the Department of EPO Biology at the University of Colorado, the National Science Foundation Dissertation Improvement Grants (DEB-9972713), and the Microbial Observatories program (DEB-0084243).

References

- Acinas, SG, Klepac-Ceraj, V, Hunt, DE, Pharino, C, Ceraj, I, Distel, DL, Polz, MF (2004) Fine-scale phylogenetic architecture of a complex bacterial community. *Nature* 430: 551–554
- Avise, JC (1994) *Molecular Markers, Natural History, and Evolution*. Chapman and Hall, New York
- Barns, SM, Delwiche, CF, Palmer, JD, Pace, NR (1996) Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. *Proc Natl Acad Sci USA* 93: 9188–9193
- Bintrim, SB, Donohue, TJ, Handelsman, J, Roberts, GP, Goodman, RM (1997) Molecular phylogeny of Archaea from soil. *Proc Natl Acad Sci USA* 94: 277–282
- Borneman, J, Triplett, EW (1997) Molecular microbial diversity in soils from eastern Amazonia: evidence for unusual microorganisms and microbial population shifts associated with deforestation. *Appl Environ Microbiol* 63: 2647–2653
- Buckley, DH, Graber, JR, Schmidt, TM (1998) Phylogenetic analysis of nonthermophilic members of the Kingdom Crenarchaeota and their diversity and abundance in soils. *Appl Environ Microbiol* 64: 4333–4339
- Cannone, JJ, Subramanian, S, Schnare, MN, Collett, JR, D'Souza, LM, Du, Y, Feng, B, Lin, N, Madabusi, LV, Muller, KM, Pande, N, Shang, Z, Yu, N, Gutell, RR (2002) The Comparative RNA Web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* 3: 2 [Correction: *BioMed Central Bioinformatics*. 3:15]
- Chin, KJ, Lukow, T, Conrad, R (1999) Effect of temperature on structure and function of the methanogenic archaeal community in an anoxic rice field soil. *Appl Environ Microbiol* 65: 2341–2349
- Crump, BC, Baross, JA (2000) Archaeoplankton in the Columbia river, its estuary, and the adjacent coastal ocean, USA. *FEMS Microbiol Ecol* 31: 231–239
- DeLong, EF (1992) Archaea in coastal marine environments. *Proc Natl Acad Sci USA* 89: 5685–5689
- Felsenstein, J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol* 17: 368–376
- Hershberger, KL, Barns, SM, Reysenbach, A, Dawson, SC, Pace, NR (1996) Wide diversity of Crenarchaeota. *Nature* 384: 420
- Jurgens, G, Lindstrom, K, Saano, A (1997) Novel group within the Kingdom Crenarchaeota from boreal forest soil. *Appl Environ Microbiol* 63: 803–805
- Jurgens, G, Saano, A (1999) Diversity of soil Archaea in boreal forest before, and after clear-cutting and prescribed burning. *FEMS Microb Ecol* 29: 205–213
- Lueders, T, Friedrich, MW (2000) Archaeal population dynamics during sequential reduction processes in rice field soil. *Appl Environ Microbiol* 66: 2732–2742
- Majewski, J, Cohan, FM (1999) Adapt globally, act locally: the effect of selective sweeps on bacterial sequence diversity. *Genetics* 152: 1459–1474
- Marr, JW (1961) *Ecosystems of the east slope of the Front Range in Colorado*. Univ Colo Ser Biol, vol. 8. University of Colorado Press, Boulder, pp 1–134
- McInerney, JO, Wilkinson, M, Patching, JW, Embley, TM, Powell, R (1995) Recovery and phylogenetic analysis of novel archaeal rRNA sequences from a deep-sea deposit feeder. *Appl Environ Microbiol* 61: 1646–1648
- Miller, DN, Bryant, JE, Madsen, EL, Ghiorse, WC (1999) Evaluation and optimization of DNA extraction and purification procedures for soil and sediment samples. *Appl Environ Microbiol* 65: 4715–4724
- Moré, MI, Herrick, JB, Silva, MC, Ghiorse, WC, Madsen, EL (1994) Quantitative cell lysis of indigenous microorganisms and rapid extraction of microbial DNA from sediment. *Appl Environ Microbiol* 60: 1572–1580
- Nicol, GW, Glover, LA, Prosser, JL (2003) The impact of grassland management on archaeal community structure in upland pasture rhizosphere soil. *Environ Microbiol* 5: 152–162
- Oline, DK, Grant, MC (2001) Scaling patterns of biomass and soil properties: an empirical analysis. *Landscape Ecol* 17: 13–26
- Olsen, GJ, Matsuda, H, Hagstrom, R, Overbeek, R (1994) fast DNAm1: A tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. *Comput Appl Biosci* 10: 41–48
- Palys, T, Nakamura, LK, Cohan, FM (1997) Discovery and classification of ecological diversity in the bacterial world: the role of DNA sequence data. *Int J Syst Bacteriol* 47: 1145–1156
- Qiu, X, Wu, L, Huang, H, McDonel, PE, Palumbo, AV, Tiedje, JM, Zhou, J (2001) Evaluation of PCR-generated chimeras, mutations, and heteroduplexes with 16S rRNA gene-based cloning. *Appl Environ Microbiol* 67(2): 880–887
- Sandaa, RA, Enger, O, Tørsvik, V (1999) Abundance and diversity of Archaea in heavy-metal-contaminated soils. *Appl Environ Microbiol* 65: 3293–3297
- Scala, DJ, Kerkhof, LJ (2000) Horizontal heterogeneity of denitrifying bacterial communities in marine sediments by terminal restriction fragment length polymorphism analysis. *Appl Environ Microbiol* 66: 1980–1986
- Simon, HM, Dodsworth, JA, Goodman, RM (2000) Crenarchaeota colonize terrestrial plant roots. *Environ Microbiol* 2: 495–505
- Sliwinski, MK, Goodman, RM (2004) Comparison of Crenarchaeal consortia inhabiting the rhizosphere of diverse terrestrial plants with those in bulk soil in native environments. *Appl Environ Microbiol* 70: 1821–1826
- Vetriani, C, Jannasch, HW, MacGregor, BJ, Stahl, DA, Reysenbach, AL (1999) Population structure and phylogenetic characterization of marine benthic Archaea in deep-sea sediments. *Appl Environ Microbiol* 65: 4375–4384