GENOMIC STUDIES OF UNCULTIVATED ARCHAEA

Christa Schleper, German Jurgens‡ and Melanie Jonuscheit**

Abstract | Archaea represent a considerable fraction of the prokaryotic world in marine and terrestrial ecosystems, indicating that organisms from this domain might have a large impact on global energy cycles. However, many novel archaeal lineages that have been detected by molecular phylogenetic approaches have remained elusive because no laboratory-cultivated strains are available. Environmental genomic analyses have recently provided clues about the potential metabolic strategies of several of the uncultivated and abundant archaeal species, including non-thermophilic terrestrial and marine crenarchaeota and methanotrophic euryarchaeota. These initial studies of natural archaeal populations also revealed an unexpected degree of genomic variation that indicates considerable heterogeneity among archaeal strains. Here, we review genomic studies of uncultivated archaea within a framework of the phylogenetic diversity and ecological distribution of this domain.

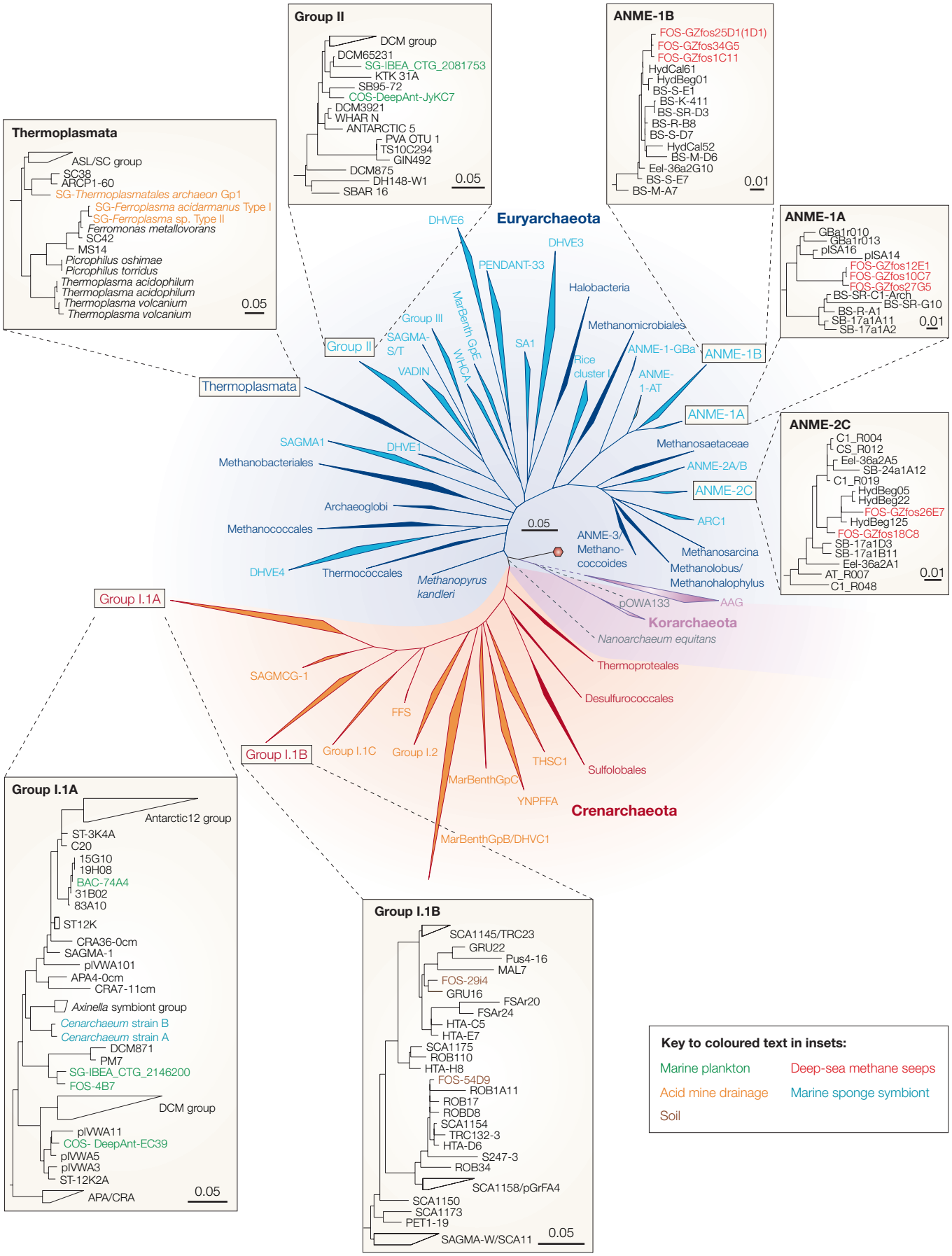
**Department of Biology, University of Bergen, Jahnebakken 5, N-5020 Bergen, Norway.*
‡Department of Applied Chemistry and Microbiology, Viikinkaari 9, FIN-00014 Helsinki University, Finland.
Correspondence to C.S.
e-mail: christa.schleper@bio.uib.no
doi:10.1038/nrmicro1159

Considering the recently accumulated knowledge of genes and genomes of uncultivated archaea, it is time to refine our perception about the ecology and diversity of this third domain of life. Although archaea have been detected in many moderate environments, they are still primarily considered to be extremists, dominating habitats that define the physical limits for biological systems, such as geothermal hot or acidic springs, deep-sea hydrothermal vents, hypersaline ponds or strictly anoxic ecosystems.

From an evolutionary viewpoint, it might be justified to give priority to the extremists: archaea from hot environments are phylogenetically diverse and some of them branch close to the root of the archaeal tree[1–3] — as inferred from 16S ribosomal RNA (rRNA) analyses (FIG. 1) and from phylogeny based on ribosomal or transcriptional proteins[4] — indicating that thermophiles might have arisen first and that some could be progenitors of species that underwent adaptive radiation into moderate habitats. Support for this speculation has been provided by biogeochemical studies and lipid biomarkers[5].

However, from an ecological and phylogenetic perspective, we should invert this picture: Archaea are primarily a diverse and widespread group on Earth, which are found in our gardens and forests[6–11], in the ocean's plankton[12,13] and sediment[14,15], in freshwater lakes[16–19] and deep down in the subsurface[20]. By contrast, those organisms that have been cultivated in the laboratory and studied in detail, such as methanogens, thermophiles and halophiles, represent only a minority of phylotypes and phenotypes.

The broad distribution and abundance of archaea in soils and oceans implies that they contribute to global energy cycles. However, these organisms have been predicted solely from PCR-based surveys and no representatives have been cultivated in the laboratory. Therefore, their specific metabolisms remain elusive. Recent advances in environmental genomic studies show that we now have the necessary technical tools to characterize these organisms in the absence of laboratory cultivation and in the context of indigenous microbial communities[21–23]. This approach involves direct cloning of genomic DNA from the environment and storing this DNA either in small-insert libraries, for large-scale shotgun-sequencing approaches, or in large-insert libraries, for targeted searches for genomic fragments from specific lineages. Hypotheses about the



DEEP SUBSURFACE
Usually 50 m or more below the surface, where the microbiota is not immediately affected by microbial functions or biogeochemical processes of the surface (as opposed to shallow subsurface).

BENTHIC
Living in or on the bottom of a body of water.

specific metabolism of several novel archaeal groups can now be formulated, which supplies the basis for studying the ecological impact of these species in the context of geochemical data and complex microbial communities. The prediction of specific metabolisms might eventually also aid enrichment efforts and enable cultivation of some novel archaeal model organisms.

Archaeal diversity based on molecular surveys

Among the first discoveries of PCR-based molecular ecological surveys was the detection of 16S rRNA of members of the Crenarchaeota in the marine plankton[12,13]. Since then, novel archaeal 'phylotypes' (bacterial phylogenetic types) have been detected in most environmental surveys that have targeted archaeal sequences. Three years ago, Philip Hugenholtz dissected 18 different archaeal phyla (10 of which contained no cultivated representatives) as opposed to 35 phyla of bacteria (13 without cultivated representatives), based on comparative analyses of 16S rRNA gene sequences[24]. Meanwhile, almost 8,000 archaeal 16S rRNA gene sequences from environmental studies have been deposited in public databases, extending the known groups and increasing the number of novel lineages (FIG. 1). Many novel archaeal groups seem (so far) to be confined to specific geographical locations or to ecosystems that have similar geochemistry, but other groups seem to be widely distributed. For example, the two crenarchaeal lineages, which are mostly defined from sequences of marine plankton (group I.1A, FIG. 1) or soils (group I.1B, FIG. 1), are both also found in freshwater samples and DEEP SUBSURFACES[8,20].

It is surprising that most of the newly discovered lineages seem to expand the two major phyla — Euryarchaeota and Crenarchaeota — that were defined as early as 1986 based on only a few cultured archaeal species[25] (FIG. 1). However, the discovery of a few more distant and deeply branching lineages through molecular surveys (Korarchaeota and AAG (ancient archaeal group) in FIG. 1, REFS 1,3) or cultivation (Nanoarchaeota)[2], indicates that greater archaeal diversity is to be expected and that more lineages might be recovered through improved molecular ecological searches, more sophisticated cultivation techniques[26] and perhaps by metagenomic approaches.

◀ Figure 1 | **The domain Archaea — from diversity to function.** A 16S rRNA tree of the Archaea is shown, with groups of uncultivated species that have been targeted in genomic studies emphasized as boxes. Genomic fosmids (prefix FOS), cosmids (COS), bacterial artificial chromosomes (BACs) or sequences that have been assembled from shotgun sequencing projects (SG) are shown in expanded boxes. Triangles in light colours represent branches with exclusively uncultivated species, dark triangles show branches with cultivated species. The size of the triangle is proportional to the number of sequences analysed. The red hexagon in the centre indicates rooting to species in the domain Bacteria. The phylogenetic tree is based on comparisons of 16S rRNA sequences from the Euryarcheota, Crenarcheota and Korarchaeota phyla together with *Nanoarchaeum equitans* (dotted lines indicate uncertain phylogenetic positions), clone pOWA133 and the ancient archaeal group (AAG), which are not classified within these phyla. In total, 1,344 16S rRNA sequences were included (mostly full-length). Sequence data were analysed with the ARB software package[82]. The backbone tree was calculated by using maximum likelihood in combination with filters excluding highly variable positions with 55 full-length sequences (1,329 positions in length). Partial sequences were inserted into the reconstructed tree by using parsimony criteria without allowing changes to the overall tree topology. The scale bar for the whole tree represents 0.05 changes per nucleotide. The complete dataset is available from the authors on request.

Evidence for living archaeal populations

Several molecular ecological studies indicate that many of the novel archaeal organisms that are predicted by PCR-based studies do in fact represent metabolically active populations. For example, marine planktonic archaea[27], a sponge-associated crenarchaeote[28], crenarchaeota on plant roots[10] and euryarchaeota in sediments[14] as well as in sulphurous marsh water enriched on polyethylene nets[26] have all been visualized with fluorescence *in situ* hybridization (FIG. 2), indicating a distinct morphology and high rRNA content, as expected for living cells. Furthermore, the abundance, distribution and dynamics of several groups shows patterns that are characteristic of active microbial populations. For example, marine planktonic archaea, which are found in many different oceanic provinces and represent approximately 20% of the total microbial planktonic population[29], vary in relative abundance with respect to water depth and season[27,30]. Similarly, marine BENTHIC crenarchaeota and euryarchaeaota show depth-related variability in deep-sea sediments[31]. A specific phylotype of cold-water archaea, the crenarchaeote *Cenarchaeum symbiosum*, was found in association with a marine sponge[28] and related phylotypes were recovered from other sponges in different oceanic regions, indicating specific metazoan–archaeal associations[32–34]. Euryarchaeota species from the three different anaerobic methane oxidation (ANME) lineages are found within the microbial mats of cold seeps, sometimes at high abundance (up to 50% for ANME-1), and have various distributions according to depth and geographical location[35]. Furthermore, these methanotrophic archaea form specific aggregates with their syntrophic sulphate-reducing bacterial partners[14]. Crenarchaeota representatives have been found in diverse soil samples, including sandy ecosystems, pristine forest soil, agricultural fields, contaminated soil and the rhizosphere[6–10,36], and represent a considerable fraction (up to 5%) of the total prokaryotic community[7,8]. They show spatial heterogeneity and changes in abundance and community structure dependent on succession, land-management strategies, heavy-metal contamination or rhizosphere type[9,37–39]. These findings indicate dynamic and active archaeal populations that react according to changing environmental parameters.

Despite the accumulated knowledge of several novel archaeal groups and their apparent abundance, none of these archaea have been obtained in pure laboratory cultures. The preliminary evidence for their specific physiologies mostly stems from environmental genomic studies.

Genome analyses of cultivated archaea

The availability of complete genome sequences from cultivated archaeal species — 45 genomes completed or near completion (listed in REF. 40) — has not only stimulated new research areas that reach far beyond the archaeal community but has served as an important framework for tracking genomes of uncultivated archaea in complex communities.

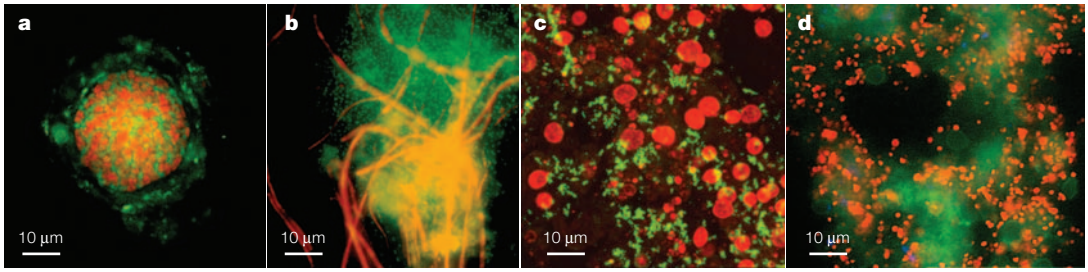


Figure 2 | **Visualization of uncultivated archaea in various environments by fluorescence *in situ* hybridization (FISH).** **a** | ANME-2C euryarchaeota (red) in association with sulphate-reducing bacteria (green) in sediments above methane hydrates. Image courtesy of T. Lösekann and K. Knittel, Max-Planck Institute for Marine Microbiology, Bremen, Germany. **b** | The freshwater euryarchaeon SM1 (green; belonging to group Pendant-33 in FIG. 1) in association with *Thiotrix* (red). Image courtesy of C. Moissl and R. Huber, Univeristy of Regensburg, Germany. **c** | *Cenarchaeum symbiosum* (green; belonging to group I.1A in FIG. 1) from the sponge *Axinella mexicana* (nuclei in red). Image courtesy of C. Preston, Monterey Bay Aquarium Research Institute, Moss Landing, USA. **d** | Crenarchaeota (red; belonging to group I.1B in FIG. 1) of tomato roots, enriched (associated bacteria in green). Image courtesy of H. Simon, Oregon Health & Science University, Beaverton, USA.

Mostly through comparative genomic studies, it has become evident that the core components of archaeal information processing systems — replication, transcription, translation and DNA repair — show striking similarities to those of eukaryotes (reviewed in REFS 4,40,41). This finding has supported the status of the Archaea as a unique and separate domain, as predicted by Carl Woese more than 25 years ago[42]. The study of information processing in the simpler, more streamlined systems of the Archaea has become increasingly relevant to understanding cellular evolution and the complex interactions that occur in the eukaryal nucleus[43–46]. Long before the genomic era, the intriguing relationship with eukaryotes was revealed by Zillig and colleagues, who realized that the archaeal DNA-dependent RNA polymerase and the basal promoter sequences have striking similarities to those of eukaryotes[47,48]. Nevertheless, archaeal genomes are typically prokaryotic in terms of their small size and organization of genes and operon structures. Archaeal genome diversity reflects the physiological versatility of the Archaea and the wide range of growth conditions, as well as frequent gene acquisition by horizontal gene transfer[41]. Based on the unique archaeal information-processing machineries, conserved archaeal 'core' genes[41] and archaeal transcription signals, we can differentiate archaeal and bacterial genomes, and can even differentiate euryarchaeotes and crenarchaeotes. Even in complex environmental DNA libraries, it is often possible to distinguish large archaeal genome fragments from bacterial fragments, based on gene content and phylogenetic marker genes. However, assigning as-yet-uncultivated archaea to distinct lineages is not easy. As often nothing but the 16S rRNA gene sequence of a novel lineage is known, this information has initially been used as a phylogenetic anchor to retrieve specific genome fragments from complex environmental libraries. More comprehensive environmental genomic studies of specific lineages became possible with large-scale sequence analyses from samples with some highly enriched organisms that allow the subsequent assembly of longer genomic contigs or even almost complete genomes (see below).

Genomic studies of marine planktonic archaea

The new field of cultivation-independent genomic studies of microorganisms was initiated when Edward DeLong and collaborators attempted to characterize genome fragments of marine planktonic archaea[49]. Inspired by the rapid advances in genomic techniques applied to cultivated microorganisms, they used a bacterial artificial chromosome (BAC)-derived fosmid vector to prepare a large-insert library from marine plankton of the North-Eastern Pacific. A 38.5-kb genomic fragment of an uncultivated mesophilic crenarchaeote was identified within 3,552 clones, using archaea-specific 16S rDNA probes (FOS-4B7, Group I.1A in FIG. 1). Even snapshot sequencing of this clone confirmed its archaeal origin[49]. Further genome fragments of marine archaea have been isolated from BAC, fosmid or cosmid libraries of surface[50,51] and deep waters[52,53] of the Antarctic and the North Pacific. Conservation of gene order around the 16S rRNA gene confirmed the close relationship of the planktonic crenarchaeota, even in strains from different oceanic regions[50,52] (FIG. 3). By contrast, it is noticeable that considerable genomic variation can be dissected, including microheterogeneity in protein-encoding regions and intergenic spacers, when genome fragments with otherwise identical or almost identical 16S RNA genes were compared from the same DNA library[50].

The planktonic archaeal clones share several genomic features with their hyperthermophilic relatives, including the estimated low GC content (~32–36%) as well as the gene repertoire and the structure of the rRNA operon. However, some genes that are so far unique to planktonic archaea have also been identified, such as a putative RNA-binding protein that shares features with the bacterial cold-shock family and a novel zinc-finger protein that was previously only found in eukaryotes[50].

Given the abundance and ubiquity of marine planktonic archaea, it is plausible that large numbers of archaeal genes would be detected in a random sequencing survey of DNA obtained from filtered surface waters. This is in fact the case. The large dataset

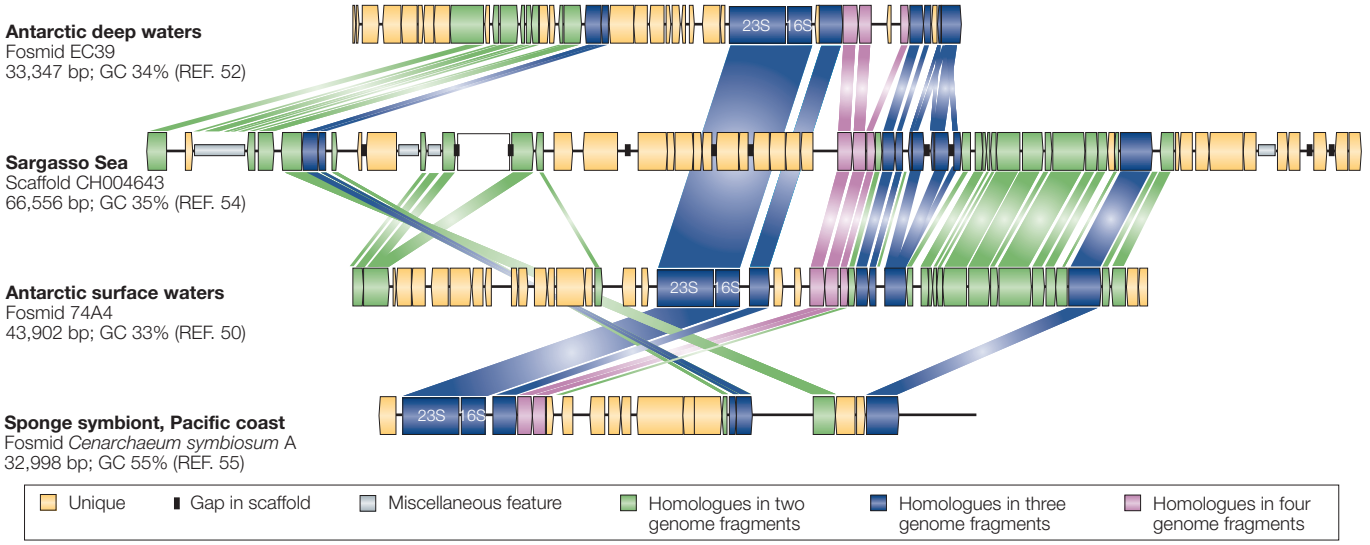


Figure 3 | **Conservation among genomic regions of uncultivated marine crenarchaeota.** This figure shows a comparison of genomic regions from marine crenarchaeota that were isolated from different environments (surface, deep water or metazoan tissue) in four different ocean locations, and reveals high genomic conservation with respect to gene similarities and order. Orthologous genes with a score of more than 50 using blastp searches (comparing an amino acid query sequence against a protein sequence database) were identified (coloured boxes) using the Artemis comparison tool ACT[83]. Three genome fragments that were cloned directly from the environment and one assembled scaffold (Sargasso Sea) are shown. Note that the Sargasso Sea scaffold is interrupted by regions of unknown sequence composition (black bars) and by ribosomal RNA sequences of *Shewanella* that have apparently been misassembled (white box).

of roughly 1 billion basepairs with 1.2 million novel genes obtained by shotgun sequencing samples from the Sargasso Sea[54] contains considerable amounts of archaeal sequences. Complete and partial 16S rRNA genes (FIG. 1) and other phylogenetic markers indicate the presence of crenarchaeota and euryarchaeota. In many cases, the genomic fragments cannot be clearly ascribed to either branch. Furthermore, the assembly of larger scaffolds in the dataset should be re-examined as misassemblies are present (see also the article by E.F. DeLong in this issue). But with more sophisticated annotation tools or simply with the help of large genomic fragments of marine archaeal species isolated earlier from BAC or fosmid libraries, a considerable number of archaeal genes and scaffolds can clearly be assigned (REF. 54 and C.S., unpublished data). Scaffold CH004643 from the Sargasso Sea dataset, for example, links three crenarchaeal fragments that were previously isolated from three different oceanic locations (FIG. 3).

Cenarchaeum symbiosum — a model archaeon

Among the complex prokaryotic community harboured in the tissues of the marine sponge *Axinella mexicana,* a sole archaeal phylotype, *Cenarchaeum symbiosum*, was detected by 16S-RNA-based PCR surveys[28]. It can be maintained over many months in stable association with its host, under controlled laboratory conditions at temperatures 40–50°C below the optimum of its closest cultivated relatives, the hyperthermophilic crenarchaeota. This crenarchaeal–metazoan association therefore provides a tractable system for the study of non-thermophilic marine crenarchaeota and gives access to relatively large amounts of biomass (and DNA) from this species. Although *C. symbiosum* has not been cultivated or completely physically separated from the tissues of its host and the coexisting bacteria, cell fractions that are enriched for the archaeon have allowed the construction of large-insert genomic libraries[55], facilitating the isolation of genome fragments and leading to a genome sequencing project (E.F. DeLong, personal communication). Many features, including a homologue of a family B DNA polymerase that is encoded on an isolated fosmid, show the close relationship of *C. symbiosum* to cultivated hyperthermophiles[56]. The biochemical characterization of this protein, however, substantiated the prediction that the organism is adapted to low temperatures[56]. Unexpectedly, the analysis of genomic contigs from *C. symbiosum* revealed the presence of two closely related variants that were found in most sponge individuals analysed[55]. The two genomic entities have less than 0.7% deviation in the 16S rRNA gene and identical gene order, but vary by up to 20% in the protein-encoding regions and up to 30% in intergenic regions. This study was among the first to reveal genomic microheterogeneity at the species level based on environmental genomic studies, and indicates considerable functional diversity and perhaps adaptation to special niches in populations of coexisting, closely related prokaryotic strains. It also reveals the primary complications inherent to this approach (that is, the risk of genome misassembly of closely related species) when genomes of natural populations are being analysed instead of genomes from pure cultures of laboratory strains. This observation was confirmed by the study of closely affiliated archaeal planktonic fosmids[50] and was largely extended in the shotgun-sequencing projects of the Sargasso Sea and of an acidic mine drainage during the assembly process of huge datasets[54,57].

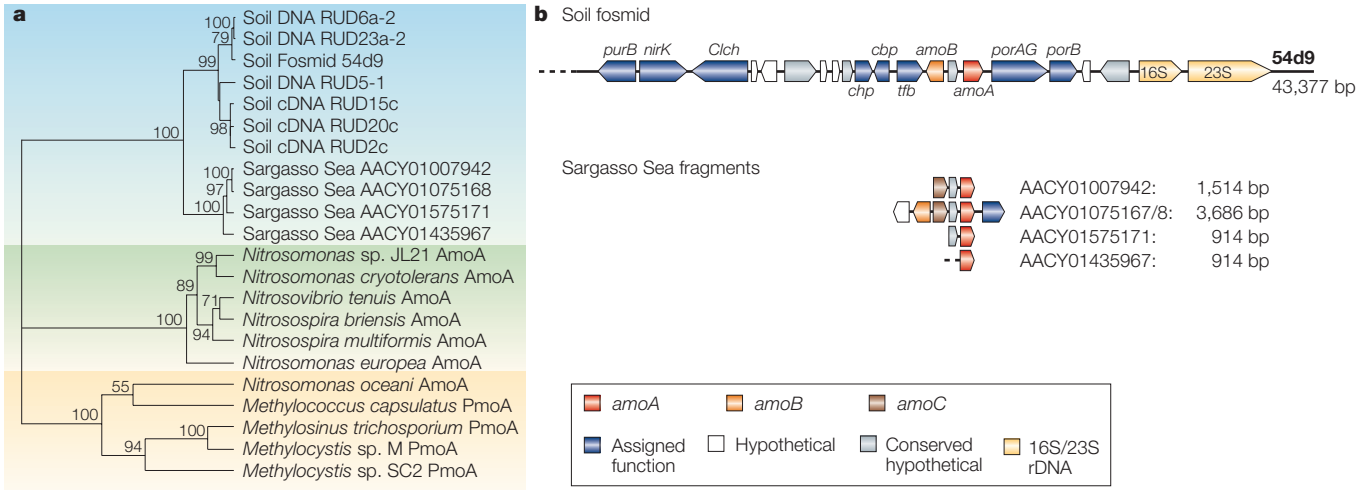


Figure 4 | **Expansion of the Amo/Pmo protein family by environmental genomic sequencing — hints of ammonia oxidation in moderate *Crenarchaeota*. a** | An *amoA/pmoA* phylogeny that was reconstructed using 200 aligned amino-acid positions of AmoA (ammonia monooxygenases) and PmoA (particulate methane monooxygenases subunit A) proteins, as well as homologous proteins that were predicted from environmental genome fragments or PCR products. DNA sequences were directly amplified from nucleic acids isolated from soil using PCR (soil DNA) or reverse transcription-PCR (soil cDNA) or were obtained from genomic fosmid 54d9 (AJ627422) and the Sargasso Sea data set (REF. 66). Bootstrap values shown are the maximum values for the respective groups and were obtained from three independent reconstructions using 1,000 rounds of PROTPARS (protein sequence parsimony method) and FITCH (distance matrix) each, as well as 100 rounds of PROML (protein maximum likelihood program). All phylogenetic programs were from PHYLIP (phylogeny inference package). **b** | A comparison between fosmid clone 54d9 from soil with sequences from the Sargasso Sea database, indicating the similarity of *amoA/B* and the linked genes. The similarities of *amoA/B* genes from the Sargasso Sea to the homologues that are located on clone 54d9 were more than 70% at the nucleotide level. Clone 54d9, which contains ribosomal RNA genes (shown in yellow) therefore clearly links the *amoA/B* genes identified in soil and marine plankton to the non-thermophilic crenarchaeota.

The *C. symbiosum* genome has a considerably higher GC content[55] (>55%) than its relatives in the marine plankton (~34%) (FIG. 3), which might reflect adaptation to the lifestyle in the metazoan host instead of a large evolutionary distance. Despite this difference, however, a considerable number of genomic scaffolds with homologous, SYNTENIC regions can be identified using *C. symbiosum* genome fragments to query the environmental database from the Sargasso Sea (FIG. 3 and C.S., unpublished data). This indicates a close affiliation to planktonic relatives, and the potential of using *C. symbiosum* as a model system to study free-living marine crenarchaeota.

Crenarchaeota in soil might be nitrifiers

16S rRNA sequences of crenarchaeota from the upper layer of soils can be recovered on all continents from almost any terrestrial ecosystem. Similar to the marine crenarchaeota, most of these species seem to be closely related (group I.1B in FIG. 1), but not all (for example, group FFS from forest soils[58] in FIG. 1). The characterization of soil microorganisms by genomic techniques is particularly challenging. Soils are the most diverse ecosystems on the planet, with an estimated 12,000 to 18,000 different dominant species in one small sample[59] and a species richness that is 20-fold higher than that of seawater[60]. Therefore, huge DNA libraries are needed to capture a reasonable fraction of the genomic diversity. Furthermore, the use of soil samples is particularly challenging, because DNA preparations are heavily contaminated with polyphenolic compounds (humic and fulvic acids) and need to be specifically purified before cloning[61,62]. Several large genome fragments from crenarchaeota that were identified by 16S RNA or archaeal-specific core genes have been isolated from complex soil libraries[61,63,64]. Ribosomal RNA operon structure and phylogenetic analyses, as well as the low GC content and predicted proteins, confirm the crenarchaeal origin of these fragments. Several genes provide first clues about the energy metabolism of these archaea. In particular, the identification of two genes encoding proteins related to subunits of ammonia monooxygenases (AmoAB), the central enzymes of ammonia oxidation in bacterial nitrifiers[65], led to the speculation that non-thermophilic crenarchaeota use ammonia as their primary energy source[66] (FIG. 4). Proteins from the family of ammonia monooxygenases and particulate methane monooxygenases (of methanotrophs) have so far only been detected in the Proteobacteria. The novel, quite distant homologues from archaea expand this well known family to a third group (FIG. 4a). Interestingly, homologues of the *amo*-like genes from the soil crenarchaeote were found in the dataset from the Sargasso Sea, indicating that marine crenarchaeota could also be capable of ammonia oxidation (FIG. 4b). As autotrophic growth of marine crenarchaeota has been inferred by radiocarbon studies[67], it is tempting to speculate that at least some groups of non-thermophilic crenarchaeota could be chemolithoautotrophic ammonia oxidizers (nitrifiers). As bacterial nitrifiers are not easily

SYNTENIC
Relationship between chromosomal regions of different species where homologous genes occur in the same order.

recovered in pure laboratory cultures, this might even explain why non-thermophilic archaea, although cosmopolitan, have not yet been cultured.

Interestingly, a homologue of a copper-containing nitrite reductase (NirK), the key enzyme of dissimilatory nitrate reduction[68], was also identified in the archaeal soil clone (FIG. 4) as well as in the Sargasso Sea dataset (not shown). As this protein is also found in ammonia-oxidizing bacteria (nitrifiers) and might even have a key role in their metabolism[69], it lends further support to the hypothesis that the primary energy metabolism of mesophilic terrestrial and marine crenarchaeota is based on ammonia oxidation.

Several, but not all, archaeal contigs that were recovered from soil libraries show considerable genomic overlap (that is, syntenic regions of high similarity) with scaffolds from the Sargasso Sea dataset (M. J. and C. S., unpublished observations), indicating that some, but not all, soil groups share metabolic capacities with marine planktonic crenarchaeota.

Methane-oxidizing archaea in the deep sea

Different lineages of Euryarchaeota that are affiliated with Methanosarcinales (ANME-1, ANME-2 and ANME-3 in FIG. 1) have been found in the upper subsurface sediments that lie above the large reservoirs of methane deep below the ocean floors[14,15,70,71]. The anaerobic oxidation of methane in these sediments is an important microbial process, as it prevents the flux of considerable amounts of the greenhouse gas into the hydrosphere. *In situ* analyses based on lipid and isotope signatures, as well as 16S-rRNA-based surveys of the microbial communities feeding on methane seeps, have indicated that archaea belonging to the ANME lineages should be methanotrophs, which convert methane into CO_2 and reduce the by-products in a process that is coupled to sulphate reduction by closely associated bacteria[14,15,70,71]. Genes for methyl coenzyme M reductase (MCR), which carries out the terminal step in methanogenesis, were found in association with ANME lineages[72], and a novel nickel compound, a variant of the MCR cofactor F430 of methanogenic archaea associated with an MCR-like protein, was characterized in a combined biochemical and environmental genomic study[73]. The hypothesis that ANME archaea carry out a 'reverse methanogenesis' was recently supported by Hallam *et al.*[74], who have isolated nearly all the genes typically associated with methane production in an approach that combined shotgun-sequencing efforts and large-insert libraries made from highly enriched samples of methane seep sediments (see also the article by E.F. DeLong in this issue).

Reconstructing a genome of *Ferroplasma*

A composite genome of an extremophilic archaeon, a *Ferroplasma* species, has been reconstructed from a dataset obtained by random shotgun sequencing from a natural biofilm of an acidic mine drainage comprising only very few microbial species[57]. In addition to the full genome of a *Leptospirillum* species (*Nitrospira* class), an almost complete *Ferroplasma* genome of 1.82 Mb was reconstructed within the 76 Mb dataset of environmental sequences.

The genome contained a 16S RNA with 99% identity to that of *Ferroplasma acidarmanus* (isolated earlier from the same location[75]), but there was a sequence divergence of 22% on the nucleotide level when comparing the two largely syntenic genomes. Most interestingly, the reconstructed genome of the *Ferroplasma* type II strain revealed an extensive degree of nucleotide polymorphism (different in its pattern from that observed with the assembled *Leptospirillum* genome) that was most probably explained by frequent recombination events among closely related *Ferroplasma* type II strains. Apart from cryptic prophages and putative mobile genetic (retro)-elements, no evidence was found for the mechanism underlying this frequent gene exchange. Genetic exchange among laboratory archaeal strains of euryarchaeota[76] and crenarchaeota[77,78] was shown earlier. Furthermore, frequent events of homologous recombination[52], even leading to a degree of linkage equilibrium resembling that of a sexual population[79], have recently been suggested for other naturally occuring archaeal assemblages, indicating that extensive genetic exchange could be a more general feature within the archaea.

Conclusions

Even the few initial environmental genomic studies on archaea (TABLE 1) have shed light on the potential of this novel research field. The approaches range from the identification of potential key metabolic traits based in specifically identified large genome fragments (as for soil crenarchaeota), the use of model organisms within well-circumscribed systems (such as *C. symbiosum*) and the modelling of complex biochemical pathways from more comprehensive datasets (as exemplified for methanotrophs) to the collection of large datasets through random shotgun sequencing (such as those of the Sargasso Sea and an acid mine drainage). A compilation of these initial environmental studies shows that the different approaches are useful individually but are also complementary. Furthermore, it is evident that studying the genomics of natural microbial communities can provide new insights into population structures, genome dynamics and speciation that cannot be obtained solely through the analysis of laboratory strains.

This novel approach marks the beginning of a second dimension in microbial genomics, as it takes into account the true microbial diversity. Extrapolating from the impact of these first datasets, one can speculate that we might soon be able to comprehensively study the capacity and dynamics of microorganisms in complex natural assemblies within a framework of changing environmental parameters. We will also be able to compare the genomic diversity and activity of different microbial consortia.

However, several challenges must be met in the future. Even more than with genome analyses of cultivated organisms, the full potential of environmental sequences can only be exploited if the hypotheses based on *in silico* data can be supported by functional studies.

Table 1 | **Published studies involving genomic analyses of uncultivated archaea**

Environment	Archaeal group	Library type	Comments	References
North Pacific, marine plankton	Group I.1A, marine planktonic crenarchaeota	Fosmids	First described genomic fragment from an uncultured prokaryote	49,50
Californian pacific coast, marine sponge *Axinella mexicana*	Group I.1A, crenarchaeote: *Cenarchaeum symbiosum*	Fosmids	Archaeal extracellular symbiont, stable association in laboratory aquaria, libraries enriched for archaeon, ongoing genome project	55,56
Californian coast, surface waters	Group II, marine planktonic euryarchaeota	BACs	60-kb clone, identified through 23S rRNA gene	51
Antarctic, coastal waters	Group I.1A, planktonic crenarchaeota	Fosmids	Comparative analysis of highly related fragments	50
Calcerous grassland soil, aerobic layer	Group I.1B, soil crenarchaeota	Fosmids	Identification of archaeal fosmids through 16S RNA genes or random end sequencing	61,64,66
Eel River Basin, Monterey Canyon, microbial mats associated with deep methane seeps	ANME-1 and -2	Fosmids	Methanotrophs in syntrophy with sulphate-reducing bacteria, genes for reverse methanogenesis	72,74
Northwestern Black Sea shelf, microbial mats from methane seeps	ANME-1	Fosmids	Biochemical isolation of a novel Ni-cofactor of methyl-coenzyme M reductase from the same sample	73
Antarctic polar front, waters of 500 m depth	Group I.1A crenarchaeota, group II, euryarchaeote	Cosmids	Study of horizontal gene transfer in crenarchaeote	52,53
Acid mine drainage biofilm	*Ferroplasma*	Plasmids (small inserts)	*Ferroplasma* type II (complete genome almost reconstructed), also A-type, G-type fragments	57
Sargasso Sea surface waters	Planktonic crenarchaeota and euryarchaeota	Plasmids (small inserts)	>1 Gb of environmental DNA sequenced	54
Methanogenic consortium	Rice cluster I methanogens	Fosmids	Stably maintained, 3-year old bacterial/archaeal consortium, anaerobic, 50°C	84

BACs, bacterial artificial chromosomes.

This will involve the integration of biological disciplines on a large scale. In the context of environmental genomics, expertise from systems biology, transcriptomics and biogeochemistry will be needed. Ten years ago, when the first environmental or 'metagenomic' libraries were constructed, the assembly of a complete or near complete genome of an uncultivated microorganism was a huge challenge. With the first large-scale sequencing projects, this proof of principle was confirmed. But it has also become clear that even larger efforts than originally anticipated, involving novel and cheaper sequencing technologies and more sophisticated bioinformatic tools, will be needed in the future if complex microbial communities are to be characterized on a genomic scale. In the future, however, genomes of specific microbial species in complex communities might be isolated with less effort, as new technologies are being developed that allow an *a priori* enrichment before the construction of environmental libraries. These more focused second-generation, metagenomic libraries could be constructed based on metabolic predictions and parameters obtained in the first round of environmental studies. Similarly, libraries enriched for DNA from metabolically active populations can be made by isolating heavy DNA after incubation of environmental samples with stable-isotope-labelled substrates — SIP-enabled metagenomics (see also the article by M.G. Dumont and J.C. Murrell in this issue).

From the perspective of a researcher of Archaea, other inspirations have emerged from this first wave of population genomics. After the astonishing finding that archaea are so widespread on Earth, we now have the tools to recognize, or at least approach, their ecological impact and interaction with other life forms. It is also possible that population genomics will lead to the identification of archaeal pathogens[80] or to novel, divergent archaeal lineages that have eluded PCR-based molecular surveys. As the extent of prokaryotic diversity and horizontal gene transfer is far from completely understood[81], this could mean that the phylogenetic trees will again be radically altered.

Pioneering groundwork in the archaeal research field has been the isolation and cultivation of hyperthermophilic organisms and other extremophiles. This has not only led to the discovery of novel metabolisms and special adaptations of archaea, but also to a more fundamental understanding of the features that unify the organisms of this third domain of life. It will be as exciting and important to isolate species of those archaeal groups that have so far solely been studied by molecular techniques. In particular, some of the organisms that are commonly found in moderate, aerobic environments should eventually be brought into culture, perhaps assisted by predictions made in metagenomic studies. Although the study of model organisms remains crucial, it has become clear over the past years of archaeal research that cultivation-independent techniques, including population genomics, will be indispensable if we want to fully understand the diversity and ecological impact of archaea.

1. Barns, S. M., Delwiche, C. F., Palmer, J. D. & Pace, N. R. Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. *Proc. Natl Acad. Sci. USA* **93**, 9188–9193 (1996).
2. Huber, H. *et al.* A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature* **417**, 63–67 (2002).
3. Takai, K. & Horikoshi, K. Genetic diversity of archaea in deep-sea hydrothermal vent environments. *Genetics* **152**, 1285–1297 (1999).
4. Forterre, P., Brochier, C. & Philippe, H. Evolution of the Archaea. *Theor. Popul. Biol.* **61**, 409–422 (2002).
5. Kuypers, M. M. *et al.* Massive expansion of marine archaea during a mid-Cretaceous oceanic anoxic event. *Science* **293**, 92–95 (2001).
6. Bintrim, S. B., Donohue, T. J., Handelsman, J., Roberts, G. P. & Goodman, R. M. Molecular phylogeny of Archaea from soil. *Proc. Natl Acad. Sci. USA* **94**, 277–282 (1997).
7. Buckley, D. H., Graber, J. R. & Schmidt, T. M. Phylogenetic analysis of nonthermophilic members of the kingdom crenarchaeota and their diversity and abundance in soils. *Appl. Environ. Microbiol.* **64**, 4333–4339 (1998).
8. Ochsenreiter, T., Selezi, D., Quaiser, A., Bonch-Osmolovskaya, L. & Schleper, C. Diversity and abundance of Crenarchaeota in terrestrial habitats studied by 16S RNA surveys and real time PCR. *Environ. Microbiol.* **5**, 787–797 (2003).
9. Sandaa, R. A., Enger, O. & Torsvik, V. Abundance and diversity of Archaea in heavy-metal-contaminated soils. *Appl. Environ. Microbiol.* **65**, 3293–3297 (1999).
10. Simon, H. M., Dodsworth, J. A. & Goodman, R. M. Crenarchaeota colonize terrestrial plant roots. *Environ. Microbiol.* **2**, 495–505 (2000).
11. Jurgens, G., Lindstrom, K. & Saano A. Novel group within the kingdom Crenarchaeota from boreal forest soil. *Appl. Environ. Microbiol.* **63**, 803–805 (1997).
12. DeLong, E. F. Archaea in coastal marine environments. *Proc. Natl Acad. Sci. USA* **89**, 5685–5689 (1992).
13. Fuhrman, J. A., McCallum, K. & Davis, A. A. Novel major archaebacterial group from marine plankton. *Nature* **356**, 148–149 (1992).
14. Boetius, A. *et al.* A marine microbial consortium apparently mediating anaerobic oxidation of methane. *Nature* **407**, 623–626 (2000).
15. Orphan, V. J., House, C. H., Hinrichs, K. U., McKeegan, K. D. & DeLong, E. F. Multiple archaeal groups mediate methane oxidation in anoxic cold seep sediments. *Proc. Natl Acad. Sci. USA* **99**, 7663–7668 (2002).
16. Keough, B. P., Schmidt, T. M. & Hicks, R. E. Archaeal nucleic acids in picoplankton from great lakes on three continents. *Microb. Ecol.* **46**, 238–248 (2003).
17. MacGregor, B. J., Moser, D. P., Alm, E. W., Nealson, K. H. & Stahl, D. A. Crenarchaeota in Lake Michigan sediment. *Appl. Environ. Microbiol.* **63**, 1178–1181 (1997).
18. Schleper, C., Holben, W. & Klenk, H. P. Recovery of crenarchaeotal ribosomal DNA sequences from freshwater-lake sediments. *Appl. Environ. Microbiol.* **63**, 321–323 (1997).
19. Jurgens, G. *et al.* Identification of novel Archaea in bacterioplankton of a boreal forest lake by phylogenetic analysis and fluorescent *in situ* hybridization. *FEMS Microbiol. Ecol.* **34**, 45–56 (2000).
20. Takai, K., Moser, D. P., DeFlaun, M., Onstott, T. C. & Fredrickson, J. K. Archaeal diversity in waters from deep South African gold mines. *Appl. Environ. Microbiol.* **67**, 5750–5760 (2001).
21. DeLong, E. F. Microbial population genomics and ecology. *Curr. Opin. Microbiol.* **5**, 520–524 (2002).
22. Handelsman, J. Metagenomics: application of genomics to uncultured microorganisms. *Microbiol. Mol. Biol. Rev.* **68**, 669–685 (2004).
23. Treusch, A.H., Schleper, C. Environmental genomics: a novel tool to study uncultivated microorganisms. In *Handbook of Genome Research* (ed. Sensen, C. W.) in the press (Wiley–VCH, Weinheim, 2005).
24. Hugenholtz, P. Exploring prokaryotic diversity in the genomic era. *Genome Biol.* **3**, REVIEWS0003 (2002).
25. Woese, C. R. & Olsen, G. J. Archaebacterial phylogeny: perspectives on the urkingdoms. *Syst. Appl. Microbiol.* **7**, 161–177 (1986).
26. Moissl, C., Rudolph, C., Rachel, R., Koch, M. & Huber, R. *In situ* growth of the novel SM1 euryarchaeon from a string-of-pearls-like microbial community in its cold biotope, its physical separation and insights into its structure and physiology. *Arch. Microbiol.* **180**, 211–217 (2003).
Describes an 'in nature' enrichment technique for a novel and interesting group of euryarchaeota in freshwater habitats.
27. DeLong, E. F., Taylor, L. T., Marsh, T. L. & Preston, C. M. Visualization and enumeration of marine planktonic archaea and bacteria by using polyribonucleotide probes and fluorescent *in situ* hybridization. *Appl. Environ. Microbiol.* **65**, 5554–5563 (1999).
28. Preston, C. M., Wu, K. Y., Molinski, T. F. & DeLong, E. F. A psychrophilic crenarchaeon inhabits a marine sponge: *Cenarchaeum symbiosum* gen. nov., sp. nov. *Proc. Natl Acad. Sci. USA* **93**, 6241–6246 (1996).
29. Karner, M. B., DeLong, E. F. & Karl, D. M. Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature* **409**, 507–510 (2001).
30. Massana, R., Murray, A. E., Preston, C. M. & DeLong, E. F. Vertical distribution and phylogenetic characterization of marine planktonic Archaea in the Santa Barbara Channel. *Appl. Environ. Microbiol.* **63**, 50–56 (1997).
31. Vetriani, C., Jannasch, H. W., MacGregor, B. J., Stahl, D. A. & Reysenbach, A. L. Population structure and phylogenetic characterization of marine benthic Archaea in deep-sea sediments. *Appl. Environ. Microbiol.* **65**, 4375–4384 (1999).
32. Lee, E. Y., Lee, H. K., Lee, Y. K., Sim, C. J. & Lee, J. H. Diversity of symbiotic archaeal communities in marine sponges from Korea. *Biomol. Eng.* **20**, 299–304 (2003).
33. Margot, H., Acebal, C., Toril, E., Amils, R. & Fernandez, P. J. L. Consistent association of crenarchaeal Archaea with sponges of the genus *Axinella*. *Marine Biology* **140**, 739–745 (2002).
34. Webster, N. S., Watts, J. E. & Hill, R. T. Detection and phylogenetic analysis of novel crenarchaeote and euryarchaeote 16S ribosomal RNA gene sequences from a Great Barrier Reef sponge. *Mar. Biotechnol. (NY)* **3**, 600–608 (2001).
35. Knittel, K., Losekann, T., Boetius, A., Kort, R. & Amann, R. Diversity and distribution of methanotrophic archaea at cold seeps. *Appl. Environ. Microbiol.* **71**, 467–479 (2005).
36. Nicol, G. W., Glover, L. A. & Prosser, J. I. Spatial analysis of archaeal community structure in grassland soil. *Appl. Environ. Microbiol.* **69**, 7420–7429 (2003).
37. Nicol, G. W., Glover, L. A. & Prosser, J. I. The impact of grassland management on archaeal community structure in upland pasture rhizosphere soil. *Environ. Microbiol.* **5**, 152–162 (2003).
38. Nicol, G. W., Webster, G., Glover, L. A. & Prosser, J. I. Differential response of archaeal and bacterial communities to nitrogen inputs and pH changes in upland pasture rhizosphere soil. *Environ. Microbiol.* **6**, 861–867 (2004).
39. Sliwinski, M. K. & Goodman, R. M. Spatial heterogeneity of crenarchaeal assemblages within mesophilic soil ecosystems as revealed by PCR-single-stranded conformation polymorphism profiling. *Appl. Environ. Microbiol.* **70**, 1811–1820 (2004).
40. Allers, T. & Mevarech, M. Archaeal genetics — the third way. *Nature Rev. Genet.* **6**, 58–73 (2005).
41. Makarova, K. S. & Koonin, E. V. Comparative genomics of Archaea: how much have we learned in six years, and what's next? *Genome Biol.* **4**, 115 (2003).
42. Woese, C. R., Magrum, L. J. & Fox, G. E. Archaebacteria. *J. Mol. Evol.* **11**, 245–251 (1978).
43. Bell, S. D. & Jackson, S. P. Transcription and translation in Archaea: a mosaic of eukaryal and bacterial features. *Trends Microbiol.* **6**, 222–228 (1998).
44. Roberts, J. A., Bell, S. D. & White, M. F. An archaeal XPF repair endonuclease dependent on a heterotrimeric PCNA. *Mol. Microbiol.* **48**, 361–371 (2003).
45. Robinson, N. P. *et al.* Identification of two origins of replication in the single chromosome of the archaeon *Sulfolobus solfataricus*. *Cell* **116**, 25–38 (2004).
46. White, M. F. & Bell, S. D. Holding it together: chromatin in the Archaea. *Trends Genet.* **18**, 621–626 (2002).
47. Huet, J., Schnabel, R., Sentenac, A. & Zillig, W. Archaebacteria and eukaryotes possess DNA-dependent RNA polymerases of a common type. *EMBO J.* **2**, 1291–1294 (1983).
48. Reiter, W. D., Palm, P. & Zillig, W. Analysis of transcription in the archaebacterium *Sulfolobus* indicates that archaebacterial promoters are homologous to eukaryotic pol II promoters. *Nucleic Acids Res.* **16**, 1–19 (1988).
49. Stein, J. L., Marsh, T. L., Wu, K. Y., Shizuya, H. & DeLong, E. F. Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. *J. Bacteriol.* **178**, 591–599 (1996).
50. Beja, O. *et al.* Comparative genomic analysis of archaeal genotypic variants in a single population and in two different oceanic provinces. *Appl. Environ. Microbiol.* **68**, 335–345 (2002).
51. Beja, O. *et al.* Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. *Environ. Microbiol.* **2**, 516–529 (2000).
This hallmark environmental genomics paper describes the cloning of large DNA fragments from marine samples into a BAC vector. The libraries have average insert sizes of 60–80 kb.
52. Lopez-Garcia, P., Brochier, C., Moreira, D. & Rodriguez-Valera, F. Comparative analysis of a genome fragment of an uncultivated mesopelagic crenarchaeote reveals multiple horizontal gene transfers. *Environ. Microbiol.* **6**, 19–34 (2004).
53. Moreira, D., Rodriguez-Valera, F. & Lopez-Garcia, P. Analysis of a genome fragment of a deep-sea uncultivated Group II euryarchaeote containing 16S rDNA, a spectinomycin-like operon and several energy metabolism genes. *Environ. Microbiol.* **6**, 959–969 (2004).
54. Venter, J. C. *et al.* Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**, 66–74 (2004).
Large-scale shotgun-sequencing project which shows the advantages and problems of environmental genomics and delivered a huge dataset of novel genes.
55. Schleper, C. *et al.* Genomic analysis reveals chromosomal variation in natural populations of the uncultured psychrophilic archaeon *Cenarchaeum symbiosum*. *J. Bacteriol.* **180**, 5003–5009 (1998).
56. Schleper, C., Swanson, R. V., Mathur, E. J. & DeLong, E. F. Characterization of a DNA polymerase from the uncultivated psychrophilic archaeon *Cenarchaeum symbiosum*. *J. Bacteriol.* **179**, 7803–7811 (1997).
57. Tyson, G. W. *et al.* Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**, 37–43 (2004).
First study that describes the assembly of (almost) complete microbial genomes from an environmental sample.
58. Jurgens, G., Lindstrom, K. & Saano, A. Novel group within the kingdom Crenarchaeota from boreal forest soil. *Appl. Environ. Microbiol.* **63**, 803–805 (1997).
59. Torsvik, V., Ovreas, L. & Thingstad, T. F. Prokaryotic diversity — magnitude, dynamics, and controlling factors. *Science* **296**, 1064–1066 (2002).
Summary on the extent and controlling factors of natural microbial biodiversity — an important background for environmental genomics.
60. Curtis, T.P., Sloan, W.T. & Scannell, J.W. Estimating prokaryotic diversity and its limits. *Proc. Natl Acad. Sci. USA* **99**,10494–10499 (2002).
61. Quaiser, A. *et al.* First insight into the genome of an uncultivated crenarchaeote from soil. *Environ. Microbiol.* **4**, 603–611 (2002).
One of the few described methods suitable for the isolation of high-molecular-weight DNA from soils or sediments.
62. Rondon, M. R., Goodman, R. M. & Handelsman, J. The Earth's bounty: assessing and accessing soil microbial diversity. *Trends Biotechnol.* **17**, 403–409 (1999).
63. Quaiser, A. *et al.* Acidobacteria form a coherent but highly diverse group within the bacterial domain: evidence from environmental genomics. *Mol. Microbiol.* **50**, 563–575 (2003).
64. Treusch, A. H. *et al.* Characterization of large-insert DNA libraries from soil for environmental genomic studies of Archaea. *Environ. Microbiol.* **6**, 970–980 (2004).
65. Arp, D. J., Sayavedra-Soto, L. A. & Hommes, N. G. Molecular biology and biochemistry of ammonia oxidation by *Nitrosomonas europaea*. *Arch. Microbiol.* **178**, 250–255 (2002).
66. Schleper, C. Population Genomics of Soil Microbial communities. *104th General meeting of the American Society for Microbiology, New Orleans, Louisiana* (2004).
67. Wuchter, C., Schouten, S., Boschker, H. T. & Sinninghe Damste, J. S. Bicarbonate uptake by marine Crenarchaeota. *FEMS Microbiol. Lett.* **219**, 203–207 (2003).
68. Zumft, W. G. Cell biology and molecular basis of denitrification. *Microbiol. Mol. Biol. Rev.* **61**, 533–616 (1997).
69. Schmidt, I., van Spanning, R. J. & Jetten, M. S. Denitrification and ammonia oxidation by *Nitrosomonas europaea* wild-type, and NirK- and NorB-deficient mutants. *Microbiology* **150**, 4107–4104 (2004).

70. Hinrichs, K. U., Hayes, J. M., Sylva, S. P., Brewer, P. G. & DeLong, E. F. Methane-consuming archaebacteria in marine sediments. *Nature* **398**, 802–805 (1999).
71. Orphan, V. J., House, C. H., Hinrichs, K. U., McKeegan, K. D. & DeLong, E. F. Methane-consuming archaea revealed by directly coupled isotopic and phylogenetic analysis. *Science* **293**, 484–487 (2001).
72. Hallam, S. J., Girguis, P. R., Preston, C. M., Richardson, P. M. & DeLong, E. F. Identification of methyl coenzyme M reductase A (*mcrA*) genes associated with methane-oxidizing archaea. *Appl. Environ. Microbiol.* **69**, 5483–5491 (2003).
73. Kruger, M. *et al.* A conspicuous nickel protein in microbial mats that oxidize methane anaerobically. *Nature* **426**, 878–881 (2003).
Excellent paper on biochemical and environmental genomic analyses combined on a natural sample highly enriched for methanotrophic euryarchaeota.
74. Hallam, S. J. *et al.* Reverse methanogenesis: testing the hypothesis with environmental genomics. *Science* **305**, 1457–1462 (2004).
Reconstruction of an (almost) complete metabolic pathway based on genomics of uncultured microorganisms.
75. Dopson, M., Baker-Austin, C., Hind, A., Bowman, J. P. & Bond, P. L. Characterization of *Ferroplasma* isolates and *Ferroplasma acidarmanus* sp. nov., extreme acidophiles from acid mine drainage and industrial bioleaching environments. *Appl. Environ. Microbiol.* **70**, 2079–2088 (2004).
76. Rosenshine, I., Tchelet, R. & Mevarech, M. The mechanism of DNA transfer in the mating system of an archaebacterium. *Science* **245**, 1387–1389 (1989).
77. Grogan, D. W. Exchange of genetic markers at extremely high temperatures in the archaeon *Sulfolobus acidocaldarius*. *J. Bacteriol.* **178**, 3207–3211 (1996).
78. Schleper, C., Holz, I., Janekovic, D., Murphy, J. & Zillig, W. A multicopy plasmid of the extremely thermophilic archaeon *Sulfolobus* effects its transfer to recipients by mating. *J. Bacteriol.* **177**, 4417–4426 (1995).
79. Papke, R. T., Koenig, J. E., Rodriguez-Valera, F. & Doolittle, W. F. Frequent recombination in a saltern population of *Halorubrum*. *Science* **306**, 1928–1929 (2004).
80. Cavicchioli, R., Curmi, P. M., Saunders, N. & Thomas, T. Pathogenic archaea: do they exist? *Bioessays* **25**, 1119–1128 (2003).
81. Doolittle, W. F. Phylogenetic classification and the universal tree. *Science* **284**, 2124–2129 (1999).
82. Ludwig, W. *et al.* ARB: a software environment for sequence data. *Nucleic Acids Res.* **32**, 1363–1371 (2004).
83. Rutherford, K. *et al.* Artemis: sequence visualization and annotation. *Bioinformatics* **16**, 944–945 (2000).
84. Erkel, C. *et al.* Retrieval of first genome data for rice cluster I methanogens by a combination of cultivation and molecular techniques. *FEMS Microbiol. Ecol.* 2005 (doi:10.1016).

Acknowledgements
We dedicate this article to Wolfram Zillig, a pioneer in archaeal research and good friend, who died April 23, 2005 in Munich. Thanks to H.-P. Klenk for the *amoA* tree, to A. Treusch, S. Leininger and S. Schuster for work on soil clone 54d9, to A. Kletzin and S. Norland for bioinformatic support and to V. Torsvik and colleagues (University of Bergen) for fruitful discussions. Special thanks from C.S. to E.F. DeLong for continuous encouragement since the first days of environmental genomics. The work in C.S.'s laboratory is supported by the Bundesministerium für Bildung und Forschung, the EMBO Young Investigator Programme and the University of Bergen.

Competing interests statement
The authors declare no competing financial interests.

Online links

DATABASES
The following terms in this article are linked online to:
Entrez: http://www.ncbi.nlm.nih.gov/Entrez
Ferroplasma acidarmanus

FURTHER INFORMATION
Christa Schleper's homepage: http://www.bio.uib.no/schleper
The ARB Project: http://www.arb-home.de
EMBO Young Investigator Programme: http://www.embo.org/projects/yip
Access to this interactive links box is free online.