FEMS Microbiology Ecology, 91, 2015, fiv029

doi: 10.1093/femsec/fiv029
Advance Access Publication Date: 20 March 2015
Research Article

RESEARCH ARTICLE

Physiological and evolutionary potential of microorganisms from the Canterbury Basin subseafloor, a metagenomic approach

Frédéric Gaboyer[1], Gaëtan Burgaud[2] and Karine Alain[3,4,5,*]

[1]CNRS, Centre de Biophysique Moléculaire, Imagerie Spectroscopie et Chimie du Vivant rue Charles Sadron Orléans, FR 45071 Cedex2, France, [2]Université de Brest, UEB, Laboratoire Universitaire de Biodiversité et d'Écologie Microbienne EA 3882n IFR 148 SFR ScInBioS, ESIAB, Plouzané, France, [3]CNRS, IUEM – UMR 6197, Laboratoire de Microbiologie des Environnements Extrêmes (LMEE), rue Dumont d'Urville, F-29280 Plouzané, France, [4]Ifremer, UMR6197, Laboratoire de Microbiologie des Environnements Extrêmes (LMEE), Technopôle Pointe du diable, F-29280 Plouzané, France and [5]Université de Bretagne Occidentale (UBO, UEB), Institut Universitaire Européen de la Mer (IUEM) – UMR 6197, Laboratoire de Microbiologie des Environnements Extrêmes (LMEE), rue Dumont d'Urville, F-29280 Plouzané, France

*Corresponding author: CNRS, IUEM UMR 6197, Laboratoire de Microbiologie des Environnements Extrêmes (LMEE), rue Dumont d'Urville, F-29280 Plouzané, France. Tel: +33 2 98 49 88 53; E-mail: karine.alain@univ-brest.fr

One sentence summary: We report two subseafloor sediment metagenomes (one of them representing the deepest subseafloor metagenome ever obtained) and discuss about microbial ecology (diversity, physiology, energy production and evolution) in subsurface sediments.

Editor: Gary King

ABSTRACT

Subseafloor sediments represent a large reservoir of organic matter and are inhabited by microbial groups of the three domains of life. Besides impacting the planetary geochemical cycles, the subsurface biosphere remains poorly understood, notably questions related to possible metabolic pathways and selective advantages that may be deployed by buried microorganisms (sporulation, response to stress, dormancy). In order to better understand physiological potentials and possible lifestyles of subseafloor microbial communities, we analyzed two metagenomes from subseafloor sediments collected at 31 mbsf (meters below the sea floor) and 136 mbsf in the Canterbury Basin. Metagenomic phylogenetic and functional diversities were very similar. Phylogenetic diversity was mostly represented by *Chloroflexi*, *Firmicutes* and *Proteobacteria* for *Bacteria* and by *Thaumarchaeota* and *Euryarchaeota* for *Archaea*. Predicted anaerobic metabolisms encompassed fermentation, methanogenesis and utilization of fatty acids, aromatic and halogenated substrates. Potential processes that may confer selective advantages for subsurface microorganisms included sporulation, detoxication equipment or osmolyte accumulation. Annotation of genomic fragments described the metabolic versatility of *Chloroflexi*, Miscellaneous Crenarchaeotic Group and *Euryarchaeota* and showed frequent recombination events within subsurface taxa. This study confirmed that the subseafloor habitat is unique compared to other habitats at the (meta)-genomic level and described physiological potential of still uncultured groups.

Keywords: metagenomics; subsurface microbiology; microbial diversity; microbial physiology; selective advantages

Received: 24 October 2014; **Accepted:** 12 March 2015

INTRODUCTION

Subseafloor environments cover about 70% of the Earth and contain a large part of the planetary organic carbon pool, possibly more than 90% of earth's carbon (Javoy 1997; Schrenk, Huber and Edwards 2010). Estimates of microbial abundance in marine sediments reported a biomass representing 0.6% of Earth's total biomass (Kallmeyer *et al.* 2012), and a cell number as high as 5.39×10^{29} (Parkes *et al.* 2014). In areas where the energy flow is considerably low, turnover rates of marine subsurface microbial communities are thought to be extremely low, in the range of hundreds to thousands of years, impacting geochemical cycles over geological times (Lomstein *et al.* 2012; Roy *et al.* 2012; Hoehler and Jørgensen 2013). Diverse approaches have described viable and active microorganisms inhabiting deep marine sediments, including cultivation (Batzke *et al.* 2007), CARD-FISH (Schippers *et al.* 2005), stable isotope tracer incubation followed by NanoSIMS analyses (Morono *et al.* 2011), geochemical profiles and activity measurements (D'Hondt, Rutherford and Spivack 2002) and metatranscriptomics (Orsi *et al.* 2013). Microbial diversity of buried sediments encompasses the three domains of life, i.e. *Archaea*, *Bacteria* and *Eukarya*, and also comprises viruses and endospores (Lomstein *et al.* 2012; Engelhardt *et al.* 2014). This diversity remains mainly resistant to cultivation. Metabolism of the few isolates from deep subseafloor sediments is based on aerobic and anaerobic respirations (of nitrate, manganese oxides or sulfate), fermentation and methanogenesis (Batzke *et al.* 2007; Fry *et al.* 2008). DNA-based studies also suggested a potential for the utilization of halogenated compounds and aromatic substrates produced during diagenesis (Futagami *et al.* 2009; Wasmund *et al.* 2013). Microbial studies of subsurface samples remain relatively scarce due to obvious technical challenges. Only two metagenomic studies (Biddle *et al.* 2008, 2011), three single cell genomic (SCG) analyses (Lloyd *et al.* 2013; Wasmund *et al.* 2013; Kaster *et al.* 2014) and one metatranscriptomic survey (Orsi *et al.* 2013) have focused on the subseafloor biosphere. Metagenomics was shown to be an informative approach to understand putative microbial lifestyles in the deep biosphere (Wrighton 2012; Castelle *et al.* 2013). Adaptive strategies deployed by buried microorganisms to sustain stressful conditions are largely unknown, albeit being of significant ecological and evolutionary importance. Previous metagenomic studies focusing on the well-studied sites of Peru Margin and Gulf of Mexico showed a distinct phylogenetic and functional landscape at the genomic level for buried habitats from 1 to 50 mbsf, meters below the seafloor (Biddle *et al.* 2008, 2011).

The IODP 317 Expedition was conducted in the Canterbury Basin (CB) (New-Zealand) and drilled sediments up to 1922 mbsf. Microbiological characterization of these sediments revealed changes in the community structure with depth, and extended the presence of surviving *Bacteria* up to 1922 mbsf and of *Eukarya* molecular signatures up to 1740 mbsf (Ciobanu *et al.* 2014). This survey about depth limit of microbial evidence raised questions about putative microbial metabolisms that may be based on fermentation, although genes for sulfate reduction were also amplified.

To obtain an integrated view of the CB microbiome and highlight the physiological and evolutionary traits of occurring microorganisms, total DNA extracted from sediments samples at 31 mbsf (4H5) and 136 mbsf (15H4) depths were sequenced and analyzed. This study reports on two subseafloor sediment metagenomes, one of them representing the deepest subseafloor metagenome ever obtained. Putative microbial lifestyles in subsurface sediments revealed by metagenomics data are presented and discussed herein.

MATERIALS AND METHODS

Sediment coring

Sediments of the CB were collected in 2009 during the IODP 317 cruise at site U1352 (44°56′26.62″S; 172°1′36.30″E), at 344 m water depth (Fulthorpe *et al.* 2011). A sediment core reaching a total depth of 1927.5 m long was obtained, using an ACP/XCB assembly for unconsolidated sediments and with a RCB assembly for sedimentary rock. Properties of the whole sediment core are described elsewhere (Ciobanu *et al.* 2014) and summarized in Table S1 (Supporting Information). Contamination during drilling was monitored with fluorescent microspheres. Only the center parts of the sediments were used for metagenomics after exposure of the samples to ultraviolet radiation. Samples used for molecular analyses were immediately stored at −80°C.

DNA extraction, amplification and sequencing

DNA was extracted with a protocol optimized for low biomass clay sediments. This protocol aimed at inhibiting DNA adsorption on clay using a casein solution (Sigma Aldrich, 80 mg mL^{-1}, in 300 mM NaH_2PO_4) added before cellular lysis. DNA was then extracted with the UltraCLean DNA Isolation Kit (MOBIO Laboratories) under strict nucleic acid free conditions (in a PCR cabinet dedicated to low biomass samples and with material and solutions filtered with 0.2 μm filters, UV treated and autoclaved when possible). Due to biomass limitations, only two independent extractions were pooled together, totalizing about 7 g for each sedimentary horizon (4H5 and 15H4).

Multiple displacement amplification (MDA) was performed with Illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare) following the manufacturer's instructions. To lower biases related to MDA amplification, three independent amplifications were performed and DNA products were subsequently pooled for sequencing. Illumina technology was used to paired-end sequence DNA with a 500X coverage (Fasteris, Plan-les-Ouattes, Switzerland). To prevent DNA contamination, an extraction control (all extracting steps without added sediments) was also sequenced. Sequences were filtered according to their sequencing quality and 86.7% (4H5) and 88.8% (15H4) of sequences were thus retained (with a score threshold sets at 30). Contig filtration was then performed (i) with a BLAST analysis to remove all 4H5 and 15H4 contigs aligned with the 'control' sequences (E-value threshold 10^{-5}) and (ii) by removing all contigs with length <500 bp.

DNA assembly and analysis

Metagenomic reads for the two samples were separately assembled in contigs with MetaVelvet (Namiki *et al.* 2012). Assemblies with the highest N50 (the contig size where all contigs of equal or larger size add up to half the assembly size) were kept for metagenomic analysis (N50 = 5019 bp for 4H5 with 63% mapped reads and N50 = 3410 bp for 15H4 with 52% mapped reads).

Gene prediction and annotation were performed using CAMERA (Community Cyber infrastructure for Advanced Microbial Ecology Research and Analysis), using Metagene to predict open reading frame (ORF) and with a 10^{-5} threshold for BLAST in databases. The MG-RAST (Meyer *et al.* 2008) platform, based on the SEED classification (a classification used

for gene annotation and structured in subsystems and functional roles; http://www.theseed.org), was also used. In addition to annotations proposed by MG-RAST and CAMERA, ORFs were downloaded and used for additional manual analyses. The P2CS (Barakat, Ortet and Whitworth 2010) and P2TF (http://www.p2tf.org/) databases were used to predict respectively two-component systems and transcription factors (similarity and *E*-value thresholds of 40% and 10^{-5}, respectively). PrediSi was used to find peptide signal, retaining predictions with scores above 0.5 (http://www.predisi.de/). Phylogenetic trees were performed by collecting sequences from the GenBank database, with SeaView4 (Gouy, Guindon and Gascuel 2010) and using Muscle to align sequences and PhyML to construct tree topologies, with 100 bootstrap reiterations.

To predict SSU and LSU rRNA genes, CAMERA rRNA genes identifications were manually controlled to remove false positives. The default parameters of MEGAN (Huson *et al.* 2011) were used to investigate microbial diversity. For control, genomes of *Methanosaeta thermophila*, *Thermosediminibacter oceani* and *Anaerolinea thermophila* were truncated in 3000 bp sequences (the average contig length) with a local Perl script and then submitted to BLASTX and MEGAN analysis. To get insight in the overall representation of specific groups in our metagenomes, we did metagenomic recruitments with microbial genomes downloaded from GenBank against 4H5 and 15H4 metagenomes and also against metagenomes of the Peru Margin (1 mbsf amplified and 50 mbsf), metagenome of Brazos-Trinity (8 mbsf) and metagenome of the Ionian Sea (3000m KM3), downloaded from MG-RAST. To generate metagenome clustering, we used the R software and the HeatMap script after metagenome size normalization. To detect hypothetical viral genes, metagenomes were submitted to the METAVIR server (http://metavir-meb.univ-bpclermont.fr), and results were imported and locally filtered (*E*-value and identity percent thresholds 10^{-10} and 30%). Since automatic annotation of genes requires a manual validation, we used GenomeView (http://genomeview.org/) to manually annotate contigs >15kb and to manually define their ORF coordinates. For this, we did BLASTX against the nr NCBI database to define the best Blast hits (BBHs) for each ORF. Recombination analysis of these large contigs was done with Easyfig (http://easyfig.sourceforge.net/) using information contained in separated GenBank files generated by RAST.

All text and tables data were manipulated with Access and the Galaxy GenOuest (Le Bras *et al.* 2013) and the Orione Galaxy platforms (http://orione.crs4.it).

Assembled scaffolds have been deposited in GenBank under accession number JRER00000000 (4H5) and accession number JRER00000000 (15H4).

RESULTS AND DISCUSSION

General metagenomic features

Two sedimentary horizons (samples 4H5 and 15H4 from 31 and 136 mbsf, respectively) and one negative control were subjected to DNA extraction for further DNA sequencing. After assembly and contig filtration, two metagenomes were obtained (named 4H5 and 15H4 hereafter), whose main properties are summarized in Table 1.

Metagenome 4H5 was represented by a 16.9 Mb assembly, distributed in 6765 contigs, with an average contig size of 2498 bp and metagenome 15H4 had similar features, presenting a 18.3 Mb assembly, distributed in 6175 contigs with an average contig size of 2963 bp. Considering a mean prokaryotic genome size of 2–3 Mb (Bentley and Parkhill 2004), metagenomic data are equivalent to 5–8 genomes. An occurrence list of 19 single-copy genes (Ciccarelli 2006) supports such results with means of 5.5 ± 2.5 genomes for 4H5 and of 5.1 ± 2.5 genomes for 15H4 (see Table S2, Supporting Information).

A total of 20 942 and 21 205 ORFs were predicted in metagenomes 4H5 and 15H4. About half of these genes had significant hits (*E*-value < 10^{-5}) in the Pfam and in the COG databases and 64% (4H5) to 66% (15H4) had a significant Blast hit in GenBank. ORF distributions in SEED and COG groups were very similar between metagenomes, as detailed in Fig. S1 (Supporting Information).

Two populations of contigs were observed based on the GC content, a major population with a GC content of 41–43% and a minor population characterized by a GC content of 28–30% (Fig. 1A).

True biological replicates are crucial in metagenomics, notably to demonstrate that functions and diversity of a metagenome are truly characteristic of an environmental sample (Knight *et al.* 2012), as it was for example described with permafrost samples (Mackelprang *et al.* 2011). Subsurface

Table 1. General features of metagenomes 4H5 and 15H4.

		4H5 (31 mbsf)	15H4 (136 mbsf)
Assembly length	Initial assembly	19.9 Mbp	21 Mbp
	After quality filtration	16.9 Mbp	18.3 Mbp
Number of contigs	Initial assembly	17 112	15 999
	After quality filtration	6777	6177
Maximum contig length		96 664 bp	52 727 bp
Average contig length		2486 bp	2958 bp
N50		4542 bp	6085 bp
Mean GC percent		40.7	41.2
Median GC percent		42.1	42.8
Predicted number of ORFs		20 942	21 205
Number of ORFs per kbp		1.24	1.16
Mean protein length		194 aa	208 aa
Longest ORF size		2776 bp	2427 bp
% ORFs affiliated with COG database		54.1	55.2
% ORFs affiliated with Pfam database		48.5	50.8

Abbreviations: mbsf, meters below the seafloor; ORFs, open reading frames; aa, aminoacids; COG, cluster of orthologous groups; Pfam, protein families.

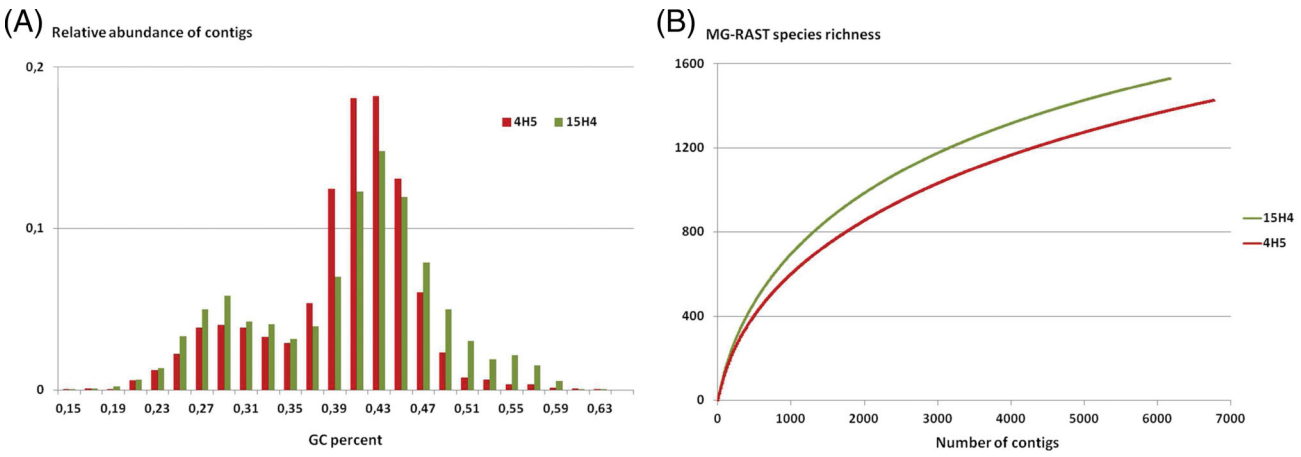

Figure 1. (**A**) Distribution of contigs based on the GC content distribution within metagenomes. (**B**) MG-Rast rarefaction curves of metagenomes 4H5 and 15H4.

sediments are challenging to obtain and the limited quantity of CB sediments for our study did not enable us to obtain such replicates. However, independent DNA extractions and MDA amplifications were done and pooled for each depth. A comparative analysis of 4H5 and 15H4 metagenomes is enabled by the fact that both metagenomes were strictly identically handled, which considerably limits biases related to sample storage (Mills, Reese and St Peter 2012) or to DNA extraction (Alain *et al.* 2011).

16S/23S rRNA gene diversity

In order to determine the taxonomic composition of the metagenomes, we first searched for SSU (small subunit: 16S) and LSU (large subunit: 23S) rRNA genes within contigs (see Table S3, Supporting Information).

A total of 26 and 19 partial 16S/23S rRNA genes were found in metagenomes 4H5 and 15H4, respectively, belonging to only four phyla (*Chloroflexi*, *Firmicutes*, *Proteobacteria* and *Thaumarchaeota*) with a dominance of *Chloroflexi* (up to 47% of all rRNA genes in 15H4) and *Thaumarchaeota* (up to 60% in 4H5). In comparison to the specific richness of other public metagenomes contained in MG-RAST, the specific richness of 4H5 and 15H4 metagenomes is low, in agreement with previous description of low microbial diversity in the CB (Ciobanu *et al.* 2014) and consistent with rarefaction curves of metagenomes (Fig. 1B). However, we cannot exclude the possibility of MDA amplification biases (Binga, Lasken and Neufeld 2008; Rodrigue *et al.* 2009; Direito *et al.* 2014) and incomplete DNA recovery during DNA extraction.

The SSU rRNA genes were more similar to environmental DNA sequences (99–82% sequence identity) than to sequences belonging to cultured representatives (91–75%).

Recently, Yarza *et al.* (2014) proposed new thresholds based on similarity of SSU rRNA sequences to define distinct taxonomic groups. Since five of our SSU rRNA genes had similarities below 94.5% (the threshold proposed by Yarza *et al.* 2014 to delineate two genera), including two SSU rRNA genes below 86.5% (the threshold to distinguish families), these genes might belong to representatives of novel genera or families. However, sequence conservation differs between regions of the SSU rRNA genes, so the reliability of taxonomic assignments depends on the sequence length (Yarza *et al.* 2014). Since most of our 16S/23S sequences are partial, it is thus difficult to conclude if they really belong to new genera or families.

Nevertheless, phylogenetic analyses performed with the most complete SSU rRNA genes (four genes with alignment lengths of 1395–1467 bp) showed that contig 3051 may belong to a novel *Chloroflexi* genus (91% identity shared with a cultured *Dehalococcoidia*) and that *Thaumarchaeota* genes belonged to the MCG. This group is commonly detected in deep sedimentary environments (Fry *et al.* 2008) but does not possess any cultured representative. The diversity of MCG in our metagenomes includes representatives of the MCG-A, MCG-B and MCG-G groups (see Fig. S2, Supporting Information).

Among other phyla, two *Firmicutes* SSU rRNA genes differed greatly from known *Firmicutes* sequences, and phylogenetic analyses showed that they clustered with a sequence sampled from a mud volcano (96%), in a strongly divergent deep subsurface *Firmicutes*-related group (see Fig. S2, Supporting Information).

Three LSU rRNA genes of *Proteobacteria*-related sequences were also found in the metagenomic libraries, distantly related to *Deltaproteobacteria*.

Whole-metagenomes phylogenetic diversity

Description of the microbial diversity implies methodological biases leading sometimes to conflicting results, as already observed with marine subsurface metagenomes (Biddle *et al.* 2008). We thus investigated our metagenomic diversity using two approaches: the Best Hit Classification predicted by MG-RAST platform and MEGAN affiliation of contigs submitted to BLASTX against GenBank (Fig. 2).

The reliability of MEGAN affiliations was first tested by submitting the truncated genomes of *M. thermophila* PT, *T. oceani* JW1228P and *A. thermophila* UNI-1 and the truncated fosmids of six MCG (Meng *et al.* 2009, 2014; Li *et al.* 2012) to BLASTX and MEGAN analyses. The three genomes were correctly affiliated but MCG sequences were not, probably due to an assignment of MCG fosmids as *Crenarchaeota* in the NCBI classification. It is thus likely that MCG genes could be incorrectly classified as crenarchaeal sequences.

Despite differences in taxonomic composition, several trends were common to both approaches. *Bacteria* and *Archaea* represented more than 99% of the assigned sequences and proportions of viral and eukaryotic assigned sequences were merely insignificant (<0.01%). *Archaea* occurred at a lower level in the 15H4 metagenome compared to the 4H5 one. Within the domains *Bacteria* and *Archaea*, diversity was very similar between metagenomes: *Chloroflexi*, *Firmicutes* and *Proteobacteria* dominated *Bacteria* and *Thaumarchaeota*, *Euryarchaeota* and

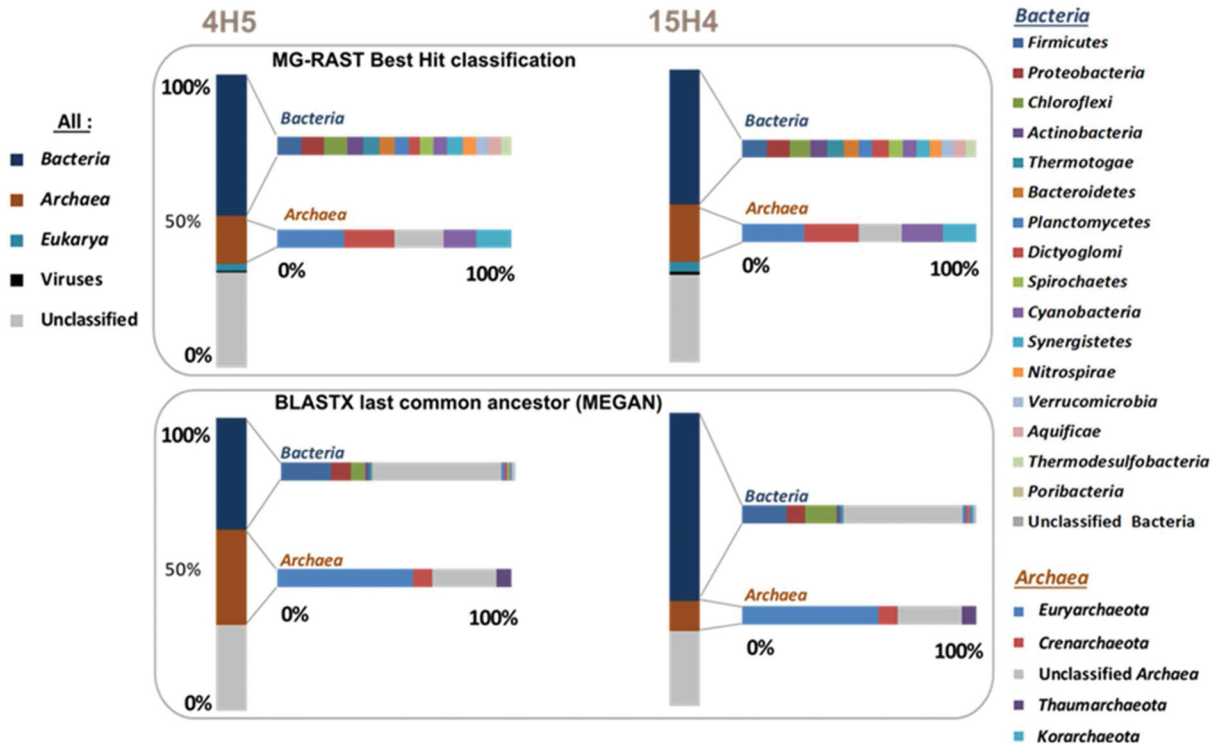

Figure 2. Phylogenetic diversity of 4H5 and 15H4 metagenomes, predicted by MG-RAST and MEGAN.

Crenarchaeota were always present within *Archaea*. The proportion of unclassified sequences reached 31% (MEGAN with 4H5 metagenome) indicating that databases lack subsurface DNA sequences.

In agreement with the SSU rRNA gene data, *Chloroflexi* were mostly affiliated (50–60% of *Chloroflexi*) with the *Dehalococcoidetes* class, in which they were annotated as 'Unclassified *Dehalococcoidetes*'. Other classes encompassed *Chloroflexi* (20–23%), *Thermomicrobia* (10–14%) and *Anaerolinea* (5–8%). *Firmicutes* sequences were similar to many subsurface representatives. The *Clostridia* class (80–82% of *Firmicutes*) was for instance mostly represented by the *Thermoanaerobacteraceae* and *Peptococcaceae* families, gathering subsurface sulfate-reducers like *Desulfotomaculum* species or *candidatus Desulforudis audaxviator* (Chivian *et al.* 2008). The *Bacilli* class represented 15–16% of *Firmicutes* sequences.

Proteobacteria were dominated by *Deltaproteobacteria* (59–63% of *Proteobacteria*), mostly by *Desulforomonadales*, *Synthrophobacteriales* and *Desulfobacteriales* orders. *Gammaproteobacteria* (14–17%) and *Alphaproteobacteria* (12–13%) were also represented but without any distinguishable dominant group.

Archaeal diversity was often dominated by sequences with BBHs with methanogenic orders (*Methanomicrobiales*, *Methanosarcinales* and *Methanococcales*) and *Thermococci*. *Crenarchaeota* were associated with the *Desulfurococcales* and the *Thermoproteales* orders. *Thaumarchaeota* were classified as 'Unclassified *Thaumarchaeota*' or as *Nitrosopumilaceae*. Due to a lack of MCG sequences in the databases, crenarchaeal hits could actually correspond to thaumarchaeal genes.

If this diversity was broadly similar to the one described by pyrosequencing (Ciobanu *et al.* 2014), groups like the SAGMEG (South African Gold Mine Euryarchaeota Group) and the MBG-B (Marine Benthic Group-B) were not found. This might be explained by a classification of SAGMEG and MBG-B as representatives of methanogens (because of the lack of reference sequences) or by MDA amplification biases.

To deeper investigate the dominant groups within the metagenomes, we performed recruitment analysis with genomes thought to be the most represented according to MG-RAST predictions: *Thermincola potens* (*Firmicutes*), *Methanosarcina acetivorans* (*Euryarcheaota*), a single amplified genome (SAG) of a subsurface *Chloroflexi* (*Dehalococcoidia* DscP2) (Kaster *et al.* 2014) and a SAG of MCG (MCG-E09) (Lloyd *et al.* 2013) (See Fig. S3, Supporting Information). Results were very similar for both metagenomes. They showed that *Dehalococcoidia* DscP2 genome was the most recruited (up to >1450 hits representing 50% of the genome) and at a second level, *Th. potens* (up to 547 hits corresponding to 6% of the genome). Due to a shorter size, fewer hits (up to 96) were found with the partial genome of strain MCG E-09, totalizing only 9% of this SAG. *Methanosarcina acetivorans* genome showed 226 hits but, when normalized to its size (5751 492 bp), this corresponded to only ~1% of its total length. The same analysis performed with three subsurface sediment metagenomes (from 1 and 50 mbsf at the Peru margin and 8 mbsf in the Gulf of Mexico) and 1 bathypelagic metagenome (KM3 at 3000m depth in the Ionian Sea (Martín-Cuadrado *et al.* 2007) also showed that the genome of *Dehalococcoidia* was well represented in subsurface sediments and, to a lesser extent the one of *Th. potens* (see Fig. S4, Supporting Information).

General metabolic information

Cell abundances at 4H5 and 15H4 sedimentary horizons were estimated to be in the range of 5×10^5 cells g^{-1} (Ciobanu *et al.* 2014). Although one cannot exclude that a part of these cells may be dead biological entities preserved on clay, it is also possible that they represent (in)active, but viable, biological forms. The presence of genes related to stress resistance and sporulation in

our metagenomes, as discussed below, is in agreement with this second hypothesis.

In order to get insights into the possible types of metabolisms deployed in 4H5 and 15H4 sedimentary horizons, we looked for genes involved in autotrophic and heterotrophic pathways, thought to be present in buried sediments, i.e. sulfate reduction, fermentation, dehalogenation, methanogenesis and CO_2 fixation.

A total of 356 (4H5) and 342 (15H4) genes involved in these pathways were found, with a dominance of genes related to heterotrophy (61% of genes in 4H5 and 74% in 15H4).

All genes of glycolysis were predicted in 4H5 and 15H4 metagenomes and only a single gene of the tricarboxylic acid cycle was missing (α-ketoglutarate dehydrogenase).

Surprisingly, while *Deltaproteobacteria* are well represented in our metagenomes, no more than two genes involved in sulfate reduction were found (polysulfide reductase and sulfite reductase subunit B) suggesting that sulfate reduction may be rare at these horizons, in agreement with a sulfate methane transition zone (SMTZ) located at ~15.5 mbsf (see Table S1, Supporting Information).

Metabolic potential of the microbial communities from the CB

Many genes involved in fermentation were predicted in both metagenomes (27–36% of genes of Table 2). Microorganisms typically use phospho-transacetylase and acetate kinase to generate ATP by acetyl-CoA conversion to acetate. No phosphotransacetylase was predicted in metagenomes and only three acetate kinases were found. However, 29 genes coding for acetyl-CoA synthases were predicted and would enable acetate and ATP production from acetyl-CoA. Genes coding for the four enzymes responsible for butyrate production from acetyl-CoA were also detected (acetyl-CoA acetyltransferase, 3-hydroxybutyryl-CoA dehydrogenase, 3-hydroxybutyryl-CoA dehydratase and the electron transfer flavoprotein complex). The presence of 30 genes for pyruvate-formate lyase-activating proteins provides another example of a genetic potential for fermentation in the CB sediments. The possibility for fermentation in the CB is supported by the presence of organic matter in our samples (see Table S1, Supporting Information), as well as by previous descriptions of fermentation in the subseafloor with different methods, either in the CB (Ciobanu *et al.* 2014) or at other sites (Mikucki *et al.* 2003; Orsi *et al.* 2013).

SCG applied to other subsurface sediments previously showed the potential for extracellular degradation of detrital matter by subseafloor *Archaea* (Lloyd *et al.* 2013). To investigate the possibility of such extracellular degradation in the CB, we looked for secreted peptidases. We found that among the 4440 genes harboring putative signal peptides, 126 encoded for peptidases and could initiate the first steps of extracellular organic matter degradation. Only four genes involved in polysaccharide degradation were found (two putative xylanases and two glycogen-debranching enzymes). This might indicate a predominance of proteinaceous matter utilization over complex carbohydrate degradation in CB sediments. Interestingly, genes encoding for clostripain and gingipain cysteine proteases, already described in partial genomes of MCG and MBG-D (Lloyd *et al.* 2013), were present in our metagenomes. Phylogenetic reconstructions showed that the cysteine proteases from the CB clustered together with *Chloroflexi* genes (see Fig. S6, Supporting Information).

The import of fermentation substrates would be enabled by genes of ABC transporters for peptides (515 genes) and sugars (427 genes) and genes of the tripartite ATP-independent periplasmic transport system for organic acid (75 genes).

Heterotrophy in deep sediments can also be supplied by diverse lipid and aromatic substrates, partially generated during diagenesis. In agreement with this hypothesis, we found genes involved in benzoate and propanoate metabolisms in both metagenomes, like α-, β-, γ-, and δ-subunits of benzoyl-CoA reductases or o-succinylbenzoate-CoA ligase. Fatty acid oxidation and decarboxylation to acetyl-CoA is done via β-oxidation of fatty acids, a process for which four different enzymes have been predicted, distributed into 83 genes. Since 4H5 and 15H4 sediments harbor volatile hydrocarbons and kerogens (see Table S1, Supporting Information), these compounds may also supply microbial metabolisms in the CB.

Natural organohalides are produced biotically and abiotically in marine sediments and can serve as electron acceptors for microbial metabolisms (Gribble 2010). The key gene of organohalide respiration, reductive dehalogenase (*rdh*A), was previously detected in distant sedimentary drilling sites, down to 350 mbsf (Futagami *et al.* 2009), and its expression in sediments has also been described (Futagami *et al.* 2013). Six genes coding for putative *rdh*A were found in the CB metagenomes, and phylogenetic analyses showed that they display various evolutionary histories and may belong to *Chloroflexi* (see Fig. S7, Supporting Information). A total of 19 *had* genes (haloacid dehalogenases) were also found and may enable microorganisms to use halogenated compounds as carbon sources.

Due to the environmental conditions of buried sediments, methanogenesis was early thought to be the last metabolism in the remineralization of organic matter in marine sediments, below the SMTZ. Cultural (Mikucki *et al.* 2003; Kendall 2006; Imachi *et al.* 2011) and metatranscriptomics (Orsi *et al.* 2013) approaches confirmed the occurrence of methanogenesis in the subseafloor as deep as 247 mbsf. Nevertheless, only few sequences of methanogens were generally detected using molecular tools in the deep subseafloor (Fry *et al.* 2008).

Metagenomes 4H5 and 15H4 are located deeper than the SMTZ, and are characterized by the presence of abiotic methane (Table S2, Supporting Information). A total of 231 genes encoding for enzymes of the methanogenesis were predicted, representing 37 different enzymes and 35 different COG groups. However, no gene encoding for the methyl-CoM reductase (*mcr*), the key gene of the final step of the methanogenesis, was predicted. This absence might be explained by amplification biases or by a high degree of divergence of subsurface sequences with sequences in the databases, preventing the identification of unusual *mcrA*-related genes. Thus, metabolism(s) of methane, if not true methanogenesis, might occur in CB subseafloor.

Some of these genes were related to methylated compounds, which have different sources in marine sediments: production by benthic invertebrates (methylamine) and macroalgae (trimethylamine-N-oxide), anaerobic decomposition of dead fishes, plants and algae (production of methanol from pectin) and releasing of compatible solutes by members of the three domains of life that can be used directly by methanogens or after hydrolysis (Watkins *et al.* 2014). Several genes required for mono/di/tri-methylamine conversion to methyl-CoM were predicted, like a corrinoid-monomethyltransferase or a dimethylamine-corrinoid protein. Conversion of trimethylamine to (di)methylamine is suggested by trimethylamine transferase genes. In the CB basin, formate-dependent methanogenesis might occur since genes for the conversion of

Table 2. Number of genes involved in different pathways found in the 4H5 and 15H4 metagenomes of the CB subseafloor.

		4H5	15H4
Nitrogen fixation	NADH dehydrogenase/NAD(P)H nitroreductase	2	2
	Nitrate reductase	1	3
	Nitrogenase reductase-like protein	2	0
	Nitrogen-fixing NifU like protein	9	6
	Nitroreductase	29	29
	Total	43	40
Fermentation	3-Hydroxyacyl-CoA dehydrogenase	14	12
	3-Hydroxybutyryl-CoA dehydratase	3	3
	Acetyl-CoA acetyltransferase	26	18
	Alcohol dehydrogenase	23	48
	Butyryl-CoA dehydrogenase	5	5
	Pyruvate formate-lyase	12	18
	Pyruvate-formate lyase-activating enzyme	1	0
	Total	84	104
Fatty acid β-oxidation	Enoyl-CoA hydratase/isomerase	7	11
	Butyryl-CoA dehydrogenase	12	14
	Propionyl-CoA carboxylase subunits	2	4
	Methylmalonyl-CoA mutase subunits	21	12
	Total	42	41
Benzoate utilization	3-Octaprenyl-4-hydroxybenzoate carboxy-lyase	2	2
	3-Oxoadipate CoA-transferase subunits	1	0
	4-Hydroxybenzoate decarboxylase	1	0
	4-Hydroxybenzoate prenyltransferase	4	0
	4-Hydroxybenzoyl-CoA reductase, alpha subunit	0	1
	4-Hydroxybenzoyl-CoA thioesterase	1	0
	4-OH-benzoyl-CoA reductase cofactor biosynthesis protein	0	1
	Benzoquinol methylase	0	1
	Benzoyl CoA reductase subunits	7	3
	Benzylsuccinyl-CoA dehydrogenase BbsG	1	0
	Nitrobenzoate nitroreductase	0	1
	o-succinylbenzoate-CoA ligase	2	2
	Para-benzoquinone reductase	1	0
	Polyferredoxin benzoyl-CoA reductase electron transfer	1	0
	Total	21	11
Propanoate utilization	Acetyl/propionyl-CoA carboxylase subunits	2	4
Dehalogenation	Haloacid dehalogenase	13	6
	Reductive dehalogenase	3	3
	Tetrachloroethene reductive dehalogenase PceA	0	1
	Total	18	14
Methanogenesis	Corrinoid-monomethylaminemethyltransferase	1	0
	Corrinoid-trimethylaminemethyltransferase	2	5
	Dimethylamine-corrinoid protein	1	0
	Trimethylamine methyltransferase	1	11
	Polyamine-transporting ATPase	1	0
	Acetyl-CoA decarbonylase/synthase complex subunits	9	13
	Formate dehydrogenase subunits	8	10
	Formate dehydrogenase accessory protein	1	1
	Formate dehydrogenase family accessory protein FdhD	1	1
	NAD-dependent formate dehydrogenase catalytic subunit	0	2
	5,10-Methylenetetrahydrofolate reductase	1	0
	5,10-Methylenetetrahydromethanopterin reductase	0	1
	5–10-Methylenetetrahydrofolate reductase	1	2
	Coenzyme F420 hydrogenase subunits	31	18
	Formylmethanofuran dehydrogenase subunits	28	4
	Formylmethanofuran/THM N-formyltransferase	13	5
	Methenyl tetrahydrofolate cyclohydrolase	4	5
	Methylene tetrahydromethanopterin dehydrogenase	8	4
	Methylenetetrahydrofolate reductase	6	10
	N(5),N(10)-methenyl-THM cyclohydrolase	6	3
	THM-S-methyltransferase subunit A	8	5
	Total	131	100
Reductive acetyl-CoA	Acetyl-CoA decarbonylase/synthase complex	13	19
	CO dehydrogenase/acetyl-CoA synthase complex	15	19
	Total	28	38

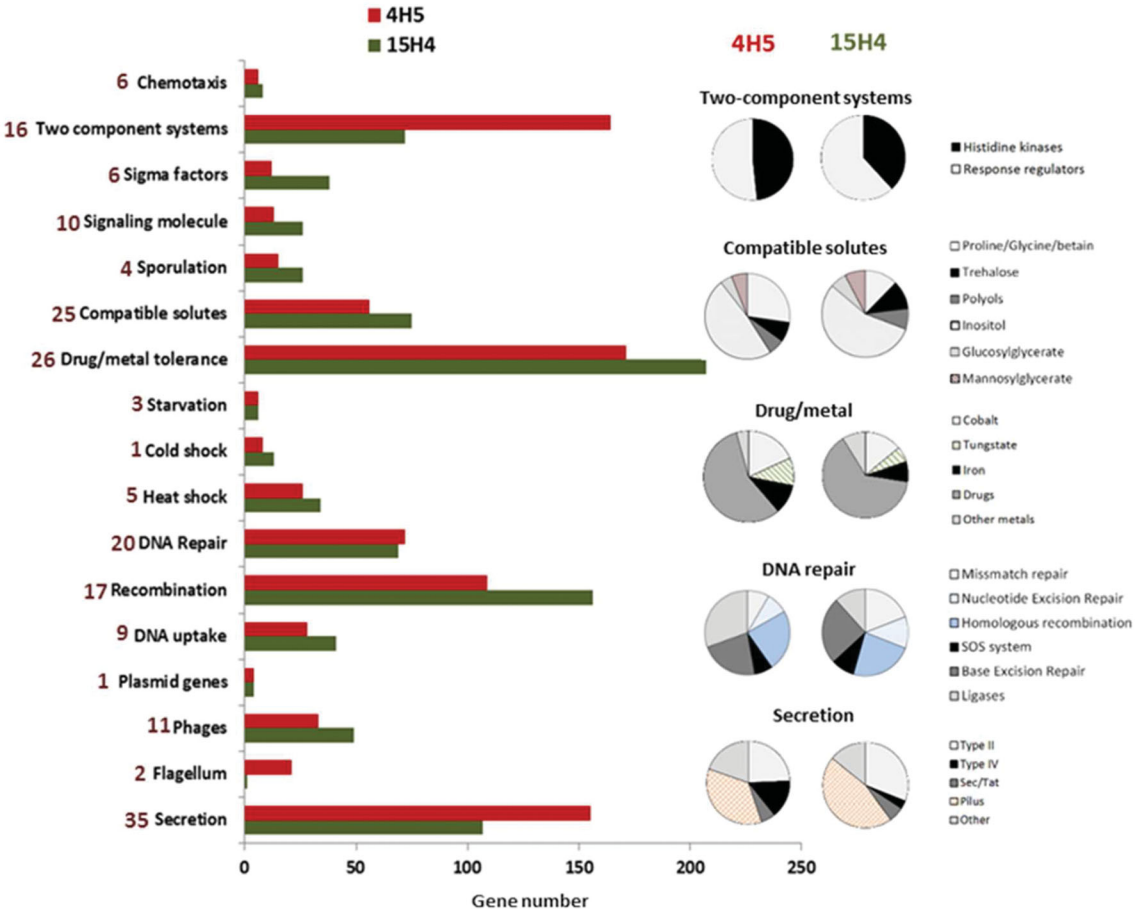

Figure 3. Number and distribution of genes involved in different physiological processes found in the metagenomes 4H5 and 15H4 from the CB subseafloor. The numbers in red represent the total count of COG families associated with these processes.

formate to CO_2 (formate dehydrogenases) and for the synthesis of formylmethanofuran (formylmethanofuran dehydrogenases) were detected.

Other genes related to the hydrogenotrophic (165 genes) and acetoclastic (27 genes for acetyl-CoA decarbonylases/synthetases) methanogenesis were also found.

A potential for acetogenesis is suggested by the presence of genes involved in the reductive acetyl-CoA (Wood-Ljungdah) pathway, which enables the utilization of CO_2 and other C1 compounds. This pathway was described in anaerobic microorganisms, like *Firmicutes*, methanogens and subsurface *Chloroflexi* (Wasmund *et al.* 2013). A total of 20 putative carbon monoxide dehydrogenases were found, mostly similar to *Euryarchaeota*, *Deltaproteobacteria*, *Firmicutes* and *Chloroflexi* sequences.

Hypothetical microbial lifestyles in the subseafloor of the CB

To better understand the possible lifestyles and interactions of *Bacteria* and *Archaea* that coexist in 4H5 and 15H4 sediments, we looked for genes of ecological importance in both metagenomes (Fig. 3, see Table S5, Supporting Information).

Due to energetic and porosity confinement, microbial motility might be limited at great depths. We only found 22 genes involved in archaeal flagellum assembly, although porosity at the 4H5 and 15H4 horizons (40–45%) would enable microbial motility and ORFs involved in different types of motility have been recently described down to 159 mbsf (Orsi *et al.* 2013). Consistently with this, we found only 13 *che* genes and a single methyl-accepting chemotaxis protein, all involved in chemotaxis.

An attached lifestyle is suggested by the presence of 109 genes related to pilus assembly and secretion systems (*sec*, *tat*, type II and type IV secretion systems) (Fig. 3).

Low activity of water, frequent in buried sediments, leads to cellular volume changes and is amplified by lithospheric and hydrostatic pressures. Osmoadaptive strategies represented in metagenomes comprised accumulation and synthesis of compatible solutes (131 genes) related to various osmolytes, in agreement with previous description of osmoadaptation potential in subsurface *Chloroflexi* (Wasmund *et al.* 2013).

A striking feature of our metagenomes regarding stress tolerance is the high number of genes related to drug and metal tolerance (378 genes) (Fig. 3).

DNA degradation, a major limitation to long-term dormancy, rends DNA repair necessary to maintain microorganisms in the subseafloor over geological times. Expression of DNA repair systems in the subseafloor has been demonstrated (Orsi *et al.* 2013). The potential for DNA repair in our metagenomes is suggested by numerous genes encoding for proteins involved in various DNA repair systems (Fig. 3), like the Mismatch Repair, the Base Excision Repair or the Nucleotide Excision Repair systems.

Oligotrophic conditions of the deep biosphere make important for microorganisms to lower their activity in response to starvation. It appears that in the CB, this could be done via a

stringent response, since genes involved in (p)ppGpp alarmone synthesis (GTP pyrophosphokinases) were detected. By contrast, few genes related to the carbon starvation (Cst) system were found.

The importance of sporulation in the subseafloor was notably demonstrated by the description of spore-forming cells as abundant as vegetative forms (Lomstein *et al.* 2012). A total of 41 genes involved in sporulation were predicted, mostly affiliated to the *Firmicutes*. This includes for instance genes for the sporulation transcriptional regulators *spoVT*/*abrB* or for the septation stage V *spoVG* gene. It is likely that a fraction of the prokaryotic cells of the CB can sporulate under certain stressful conditions and wait for more favorable conditions.

The ability for microorganisms to adapt their cellular activity to environmental conditions is a crucial selective advantage in natural environments. To regulate gene expression, we predicted 1263 transcriptional regulators distributed in 58 families, mostly represented by the AsnC and the Xre families. Forty-four sigma factors of the σ^{24}, σ^{54} and σ^{70} families and six anti-sigma factors were also detected, as well as adenylate cyclases and diguanylate cyclases, indicating that cyclic AMP and cyclic-di GMP may regulate gene expression by binding to their targets. This is in agreement with previous description of an adenylate cyclase gene in a MCG fosmid (Meng *et al.* 2009).

Even though subseafloor sediments represent a relatively stable habitat, environmental sensing may be crucial for microbial survival, especially with respect to starvation. Predicted two-component systems were represented by 82 histidine kinases (HKs) and 117 response regulators (RRs). The RR/HK ratio (1 to 2) may indicate that several RRs can be associated to a single HK. Unfortunately, it is difficult to predict the nature of signals activating HKs. A similarity analysis performed between these HKs and the well-studied HKs of *Escherichia coli* showed that the detected HKs were mostly similar to the EvgS/EvgA and the BarA/UvrY HKs families, involved in carbon storage and drug resistance (Chavez *et al.* 2010).

Evolutionary potential

The study of microbial evolution in the deep biosphere is still in its early stages. The stability of this ecosystem, altogether with low metabolic rates and long generation times, might make difficult mutation fixation in microbial populations. We investigated the possibility for genome rearrangements in the CB subseafloor.

We found 82 genes sharing significant similarities to known phages, and five CRISPR spacers regions (see Table S4, Supporting Information). To confirm the presence of virus-related genes, we used the METAVIR server and found that 215 kb (4H5) and 230 kb (15H4) were significantly recruited in the RefSeqVir database, representing ~1.3% of both metagenomes. Several viral phylogenetic markers (DNA polymerase B *polB2*, Terminase *terL*) were also detected, distantly related to *Bacillus* or to *Thermus* phage genes (data not shown). Interesting BBHs of these viral sequences notably involved hits to *Desulfotomaculum* species or to a metagenome obtained from subsurface aquifer sediments (Wrighton 2012). The presence of virus-like particles in subsurface sediments, as already described (Engelhardt *et al.* 2011, 2014), is important in terms of ecology since cellular lysis induced by viral cycles may significantly impact the pool of organic matter in buried sediments.

Other processes of DNA exchange likely to occur in the CB subseafloor include (i) the transformation of competent cells by extracellular DNA, since 69 genes involved in DNA uptake were predicted, or (ii) the conjugation by means of conjugative pili, for which we also found many genes. A capability for genome rearrangements is also suggested by the presence of 88 putative integrases and 120 putative transposases. Genes involved in homologous recombination were also detected (*recA*, *reB*, *recF*, *recO* and *recR* genes), totalizing 33 ORFs. The occurrence of plasmid genes (plasmid maintenance protein, toxin/antitoxin systems) raises the questions of plasmid maintenance and exchange in the subseafloor. Hence, genome rearrangements is likely to occur in the CB subseafloor.

To confirm previous analyses that showed that subseafloor has its own genetic space (Biddle *et al.* 2011), we performed hierarchical clustering analysis with SEED categories of various subsurface and surface habitats (Fig. 4). Results showed that subseafloor metagenomes clustered together while metagenomes from oceans and surface sediments (1 mbsf) grouped into another cluster. The higher relative abundance of SEED groups like capsular and cell-wall components, aromatic compounds/methylamine utilization, protein translocation or resistance to antibiotics within subsurface metagenomes could reveal the biological importance of these processes in the subseafloor (Fig. 4). The peculiarity of the subseafloor compared to other habitats was also confirmed by clustering metagenomes according to their taxonomic composition, with higher abundance of *Firmicutes*, *Chloroflexi* and *Euryarchaeota* in the subseafloor (See Fig. S5, Supporting Information).

Manual annotation of genomic fragments

Due to the low microbial diversity of 4H5 and 15H4 metagenomes and high sequencing coverage (~500X), the assembly led to several contigs of significant length, allowing the analysis of 'genomic fragments'. To correlate functional information and phylogeny of microorganisms, we searched and manually annotated large contigs (>15kb) harboring phylogenetic markers. A total of 18 contigs was selected (Table 3), belonging to *Chloroflexi*, *Euryarchaeota* and *Thaumarchaeota* and to the OP8 and OD1 candidate divisions. Phylogenetic trees of to these identity markers are detailed in Fig. S8 (Supporting Information).

Chloroflexi: five contigs of *Chloroflexi* displayed high coding densities (88–96%) and G+C contents from 0.44 to 0.57, which is in the same range than the genome of *Dehalogenimonas lykanthroporepellens* (50%) and two SAGs from a subsurface *Chloroflexi* (46–47%) (Wasmund *et al.* 2013; Meng *et al.* 2014). Genes encoded by these contigs suggest a heterotrophic metabolism based on amino acids and fatty acids, and strengthens the hypothesis of a fermentative metabolism within sedimentary *Chloroflexi* (Wasmund *et al.* 2013; Meng *et al.* 2014). For example, genes encoding for the aspartate-semialdehyde dehydrogenase, the dihydrodipicolinate reductase and the dihydrodipicolinate synthase formed a hypothetical operon in contig 1078.

Euryarchaeota

Four contigs appeared to be of euryarchaeal origin, ranging from 15.8 to 54.6 kb. These euryarchaeal sequences were distantly related to other *Euryarchaeota* (see Fig. S8, Supporting Information) and could belong to the SAGMEG group, since their DNA was amplified from the samples (Ciobanu *et al.* 2014). Genes involved in carbon monoxide fixation (CO dehydrogenase accessory protein), sugar metabolism (phosphoglycerate kinase, xylose isomerase), amino acid metabolism (ornithine carbamoyltransferase, tryptophanase) or β-oxidation (isopentenyl phosphate kinase, dihydroxy-butanone-phosphate synthase) were predicted in these contigs. The co-occurrence of

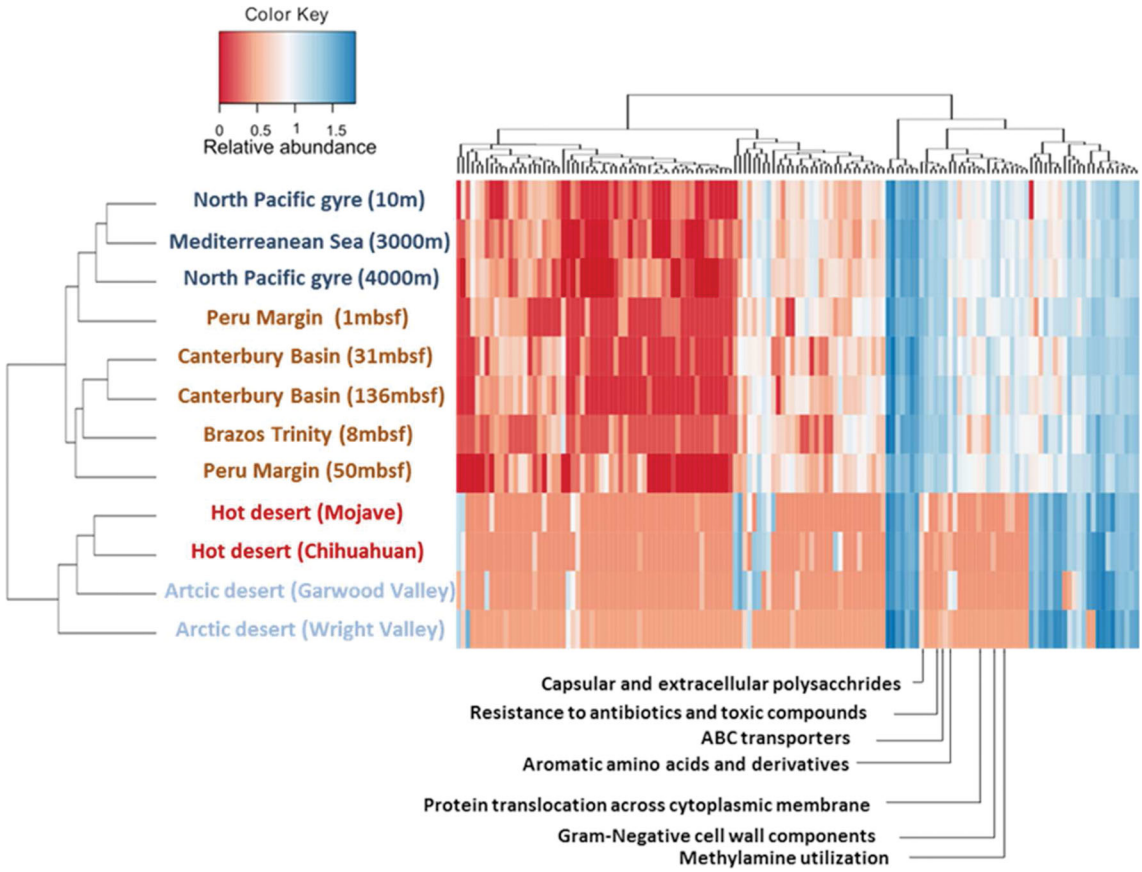

Figure 4. Comparison of functional diversity of the two metagenomes from the CB subseafloor with other habitats by metagenomic clustering of functional diversity predicted by MG-RAST. The clustering was done with the SEED database data, at the level 2 of functional information. Groups of habitats are colored: in red for hot deserts; in light blue for cold deserts; in dark blue for oceans; in brown for subsurface sediments.

genes of methanogenesis (formylmethanofuran dehydrogenase, *hypABCDE* genes coding for hydrogenase nickel incorporation-associated proteins), of fermentation (acetyl-CoA synthetase or pyruvate formate-lyase) and of β-oxidation in contig 308 was unexpected and raises the question of a possible metabolic flexibility for these *Euryarchaeota*. A potential for secretion and attachment is also suggested by the presence of secretion (*ftsY*) and pili (*pilT*) genes.

MCG

Seven manually annotated contigs were predicted to belong to MCG, ranging from 15.1 to 87.7 kb in size and with 0.42 to 0.46 G+C contents, which is less variable than the 6 MCG fosmids already described (0.38 to 0.57). Overall, these MCG genes reflected a potential for metabolic versatility, secretion and transport processes. This is illustrated by contig 613, encoding genes for benzoate and butanoate metabolisms, aminoacid utilization, or a maltodextrin ABC transport system (*malK*, *malF*, *malG* genes). Genes involved in stress tolerance comprised an alkyl hydroperoxide reductase, a cobalt-zinc-cadmium resistance *czcD* gene or a glyoxylase. Glyoxylases are enzymes protecting cells from the toxic methylglyoxal produced during glycolysis and have been already described in MCG fosmids (Li *et al.* 2012). A haloacid dehalogenase-like (*had*) gene was also present in a contig of MCG and was most similar to the *had* gene of an uncultured *Thermoplasmata* (47% identity and E-value 2.10^{-63}). Interestingly, contig 1311 harbored a large region (15.5 kb) coding for 24 hypothetical proteins, a phage tail protein and a phage-like integrase. This region could thus represent the first prophage signature found in MCG genomes.

Candidate divisions

Many hypothetical proteins were encoded in OP8 and OD1 contigs, making any physiological prediction difficult. An OD1 candidate division contig encoded for *pilX*, *pilM* genes, *pulF* and *pulE* genes (secretory pathway) and proteases, suggesting attachment to surfaces. Predicted genes in the OP8 contig 465 are involved in transport, exopolysaccharide secretion and sugar or amino acid metabolisms.

Structure and recombination of genomic fragments

In a last step, we investigated recombination events between these large contigs and genomes of (un)cultured representatives. Genome recombination in subseafloor microorganisms has already been reported within *Chloroflexi* and MCG (Meng *et al.* 2009, 2014; Wasmund *et al.* 2013) and is also supported by our metagenomic data (Fig. 5 and S9, Supporting Information).

For example, the comparative analysis of MCG contig 5051 with the MCG 26-B6 fosmid sequence (Meng *et al.* 2014) showed that ancient recombination events occurred since the divergence of the two corresponding species (Fig. 5).

Comparative analysis of *Chloroflexi* contigs 1078 and 5624 with the genome of *D. lykanthroporepellens* showed that synteny was locally highly conserved (see Fig. S9, Supporting Information).

Table 3. General features of large contigs (>15kb) manually annotated and harboring phylogenetic a marker.

Phylum	Contig	Metagenome	Length (bp)	Phylogenetic marker*	Genes	Hypothetical proteins	GC content	Coding density
Chloroflexi	Contig 1078	4H5	17 393	t-IF-2	19	3	0.46	91%
	Contig 1324	15H4	21 239	t-IF-2	23	5	0.51	96%
	Contig 5624	4H5	18 239	S2 SSU protein	20	3	0.44	93%
	Contig 5148	15H4	16 953	Phe-/Thr-tRNA Sase	20	7	0.44	90%
	Contig 535	15H4	16 587	Phe-tRNA Sase	20	6	0.57	88%
Euryarchaeota	Contig 295	4H5	24596	L11 LSU protein	24	6	0.46	83%
	Contig 308	4H5	54553	S9 SSU protein	56	11	0.47	88%
	Contig 557	15H4	15 794	O-P-Ser-tRNA Sase	15	5	0.54	94%
	Contig 841	4H5	24 881	O-P-Ser-tRNA Sase	25	9	0.46	80%
Thaumarchaeota	Contig 1311	4H5	42 488	L5 LSU protein	49	27	0.43	86%
	Contig 164	15H4	17 504	t-aIF-2	21	6	0.43	83%
	Contig 21	4H5	87 704	t-aIF-2	89	27	0.46	83%
	Contig 2033	15H4	15 121	L5 LSU protein	23	4	0.43	75%
	Contig 5051	4H5	20 223	L2 LSU protein	25	6	0.42	80%
	Contig 3565	4H5	25 078	L5 LSU protein	35	3	0.46	84%
	Contig 613	4H5	43 444	S7P SSU protein	39	2	0.46	86%
OP8 division	Contig 465	4H5	19 374	Pro-tRNA Sase	22	8	0.40	88%
OD1 division	Contig 1729	15H4	18 746	t-IF-2	20	8	0.37	92%

*The following abbreviations were used: t-IF-2, Translation Initiation Factor-2; t-aIF2: Archaeal Translational Initiation Factor-2; Pro/Phe/Ser/Thr-tRNA Sase, Prolyl/Phenylalanyl/Seryl/Threonyl-tRNA Synthetase; O-P-Ser-tRNA Sase, O-phospho-Seryl-tRNA Synthetase; LSU, Large subunit; SSU, Small subunit.

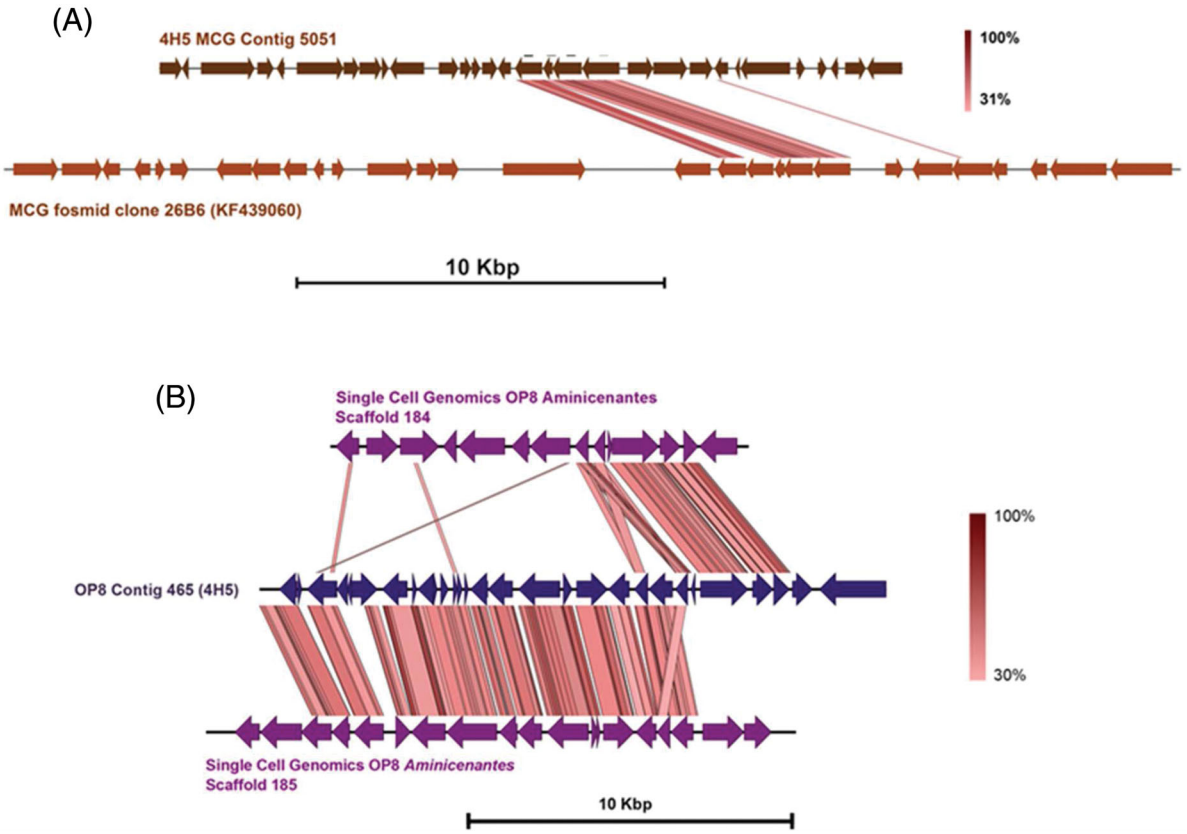

Figure 5. Examples of genomic recombinations in metagenomes. The figure shows comparisons of (**A**) a MCG contig from the CB with an MCG fosmid previously described from estuarine sediments (Meng *et al.* 2014), and (**B**) an OP8 contig from the CB with two scaffolds of an OP8 SAG obtained from a freshwater lake (Rinke *et al.* 2013).

Analysis of the OD1 contig 1729 showed that several ORFs had BBHs with an OD1 SAG sampled in freshwater (Rinke *et al.* 2013). Comparative analysis of these sequences showed a synteny, even though traces of ancient recombination events are also present.

Interestingly, contig 465 of the candidate division OP8 was found to be homologous to two different scaffolds of an OP8 SAG (Rinke *et al.* 2013): scaffold 185 matched the 5′ part of our contig and scaffold 184 matched its 3′ part. To our best knowledge, this is the first description of genome rearrangements

in OP8 and OD1 candidate divisions (see Fig. S9, Supporting Information).

CONCLUSION

Metagenomic studies applied to subseafloor sediments collide with numerous technical problems mostly related to DNA extraction and to incomplete databases (Binga, Lasken and Neufeld 2008; Rodrigue *et al.* 2009; Direito *et al.* 2014). The MDA amplification performed on DNA extracted from 4H5 and 15H4 sedimentary horizons do not enable to draw quantitative conclusions. However, if it is impossible to know what was missed by MDA amplification, it is however possible to conclude about the subseafloor physiological potential revealed by whole metagenome sequencing. It is thus necessary to underline that the phylogenetic and functional diversities described in our metagenomes represent only one fraction of the diversities present in 4H5 and 15H4 sediments.

This study described the metagenomes from two sedimentary horizons of the CB, encompassing typical subseafloor archaeal and bacterial groups, and that were clearly distinct from surface environment metagenomes. The main selective advantages of these groups might be related to their metabolic capacities and energy conversion properties, as well as to their physiological versatility. Description of genes involved in fermentation, methanogenesis and utilization of aromatic or halogenated compounds indicates that these pathways are part of the whole metabolisms occurring in the CB subseafloor. The geochemical settings of these sediments also support this. The presence of viral genes is also of major importance since the organic matter produced by cellular lysis could be used as substrates in anaerobic pathways.

Selective advantages that may be deployed in response to subsurface conditions were predicted, including sporulation, osmolyte accumulation, chemical detoxication or stringent response. Metagenomic data also suggested that microbial lifestyles in the subseafloor could involve surface attachment and secretion, possibly highlighting the importance of microbial interactions with the sedimentary environment.

It is difficult to draw clear-cut conclusions from a genetic potential, especially when we know that the picture that we have is incomplete and that we cannot exclude the possibility of the persistence of viable but dormant biological entities. Nevertheless, several lines of evidence, like the importance of anaerobic metabolisms based on amino acids, lipids and methylated compounds highly represented in our metagenomes and expressed in other sites (Orsi *et al.* 2013), hint that subseafloor sediments are maybe not only based on the use of buried refractory and aromatic compounds. Life/death cycle ('cannibalism') and persistence of some microbial cells on dead microbial remains, as previously demonstrated on extended stationary phase of bacterial cultures (Hoehler and Jørgensen 2013), appeared as other processes to maintain microbial activities.

The descriptions of genome rearrangements within various taxa (MCG, *Chloroflexi*, OD1 and OP8 candidate divisions) provide evidence for complex evolutionary histories of the subseafloor microbial communities. Recombination may involve gene transfers and suggests that the subseafloor biosphere represents a huge reservoir a genetic innovation that remains to be explored.

SUPPLEMENTARY DATA

Supplementary data is available at FEMSEC online.

ACKNOWLEDGEMENTS

We thank Grégoire Michoud for his help in bioinformatics, the Fasteris team for their help in metagenomic assembly and two anonymous referees for their constructive comments.

FUNDING

This work was performed with funding from the InEE-CNRS program APEGE-2013 and by the EU program MaCuMBA to Karine Alain. F.G. was supported by a grant from the Conseil Régional de Bretagne and the Université de Bretagne Occidentale (UBO).

Conflict of interest. None declared.

REFERENCES

Alain K, Callac N, Ciobanu M-C, *et al.* DNA extractions from deep subseafloor sediments: Novel cryogenic-mill-based procedure and comparison to existing protocols. *J Microbiol Methods* 2011;**87**:355–62.

Barakat M, Ortet P, Whitworth DE. P2CS: a database of prokaryotic two-component systems. *Nucleic Acids Res* 2010;**39**:D771–6.

Batzke A, Engelen B, Sass H, *et al.* Phylogenetic and physiological diversity of cultured deep-biosphere bacteria from equatorial Pacific ocean and Peru margin sediments. *Geomicrobiol J* 2007;**24**:261–73.

Bentley SD, Parkhill J. Comparative genomic structure of prokaryotes. *Ann Rev Gen* 2004;**38**:771–91.

Biddle JF, Fitz-Gibbon S, Schuster SC, *et al.* Metagenomic signatures of the Peru Margin subseafloor biosphere show a genetically distinct environment. *P Natl Acad Sci USA* 2008;**105**:10583–8.

Biddle JF, White JR, Teske AP, *et al.* Metagenomics of the subsurface Brazos-Trinity Basin (IODP site 1320): comparison with other sediment and pyrosequenced metagenomes. *ISME J* 2011;**5**:1038–47.

Binga EK, Lasken RS, Neufeld JD. Something from (almost) nothing: the impact of multiple displacement amplification on microbial ecology. *ISME J* 2008;**2**:233–41.

Castelle CJ, Hug LA, Wrighton KC, *et al.* Extraordinary phylogenetic diversity and metabolic versatility in aquifer sediment. *Nat Com* 2013;**4**:2120.

Chavez RG, Alvarez AF, Romeo T, *et al.* The physiological Stimulus for the BarA sensor kinase. *J Bacteriol* 2010;**192**:2009–12.

Chivian D, Brodie EL, Alm EJ, *et al.* Environmental genomics reveals a single-species ecosystem deep within earth. *Science* 2008;**322**:275–8.

Ciccarelli FD. Toward automatic reconstruction of a highly resolved tree of life. *Science* 2006;**311**:1283–7.

Ciobanu M-C, Burgaud G, Dufresne A, *et al.* Microorganisms persist at record depths in the subseafloor of the Canterbury Basin. *ISME J* 2014;**8**:1370–80.

D'Hondt S, Rutherford S, Spivack A. Metabolic activity of subsurface life in deep-sea sediments. *Science* 2002;**295**:2067–70.

Direito S, Zaura E, Little M, *et al.* Systematic evaluation of bias in microbial community profiles induced by whole genome amplification. *Environ Microbiol* 2014;**16**:643–57.

Engelhardt T, Kallmeyer J, Cypionka H, *et al.* High virus-to-cell ratios indicate ongoing production of viruses in deep subsurface sediments. *ISME J* 2014;**8**:1503–9.

Engelhardt T, Sahlberg M, Cypionka H, *et al.* Induction of prophages from deep-subseafloor bacteria. *Environ Microbiol Rep* 2011;**3**:459–65.

Fry JC, Parkes RJ, Cragg BA, *et al.* Prokaryotic biodiversity and activity in the deep subseafloor biosphere. *FEMS Microbiol Ecol* 2008;**66**:181–96.

Fulthorpe CS, Hoyanagi K, Blum P, eds. *Proceedings of the IODP Integrated Ocean Drilling Program* Management International, Inc., for the Integrated Ocean Drilling Program, Vol. 317, Washington D.C., USA, 2011. http://publications.iodp.org/proceedings/317/317title.htm (26 August 2013, date last accessed).

Futagami T, Morono Y, Terada T, *et al.* Dehalogenation activities and distribution of reductive dehalogenase homologous genes in marine subsurface sediments. *Appl Environ Microb* 2009;**75**:6905–9.

Futagami T, Morono Y, Terada T, *et al.* Distribution of dehalogenation activity in subseafloor sediments of the Nankai Trough subduction zone. *Philos T R Soc B* 2013;**368**:20120249.

Gouy M, Guindon S, Gascuel O. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* 2010;**27**:221–4.

Gribble G. *Naturally Occurring Organohalogen Compounds—A Comprehensive Update*. New York: Springer, 2010.

Hoehler TM, Jørgensen BB. Microbial life under extreme energy limitation. *Nat Rev Microbiol* 2013;**11**:83–94.

Huson DH, Mitra S, Ruscheweyh H-J, *et al.* Integrative analysis of environmental sequences using MEGAN4. *Genome Res* 2011;**21**:1552–60.

Imachi H, Aoi K, Tasumi E, *et al.* Cultivation of methanogenic community from subseafloor sediments using a continuous-flow bioreactor. *ISME J* 2011;**5**:1913–25.

Javoy M. The major volatile elements of the Earth: their origin, behavior, and fate. *Geophys Res Lett* 1997;**24**:177–80.

Kallmeyer J, Pockalny R, Adhikari RR, *et al.* Global distribution of microbial abundance and biomass in subseafloor sediment. *P Natl Acad Sci USA* 2012;**109**:16213–6.

Kaster A-K, Mayer-Blackwell K, Pasarelli B, *et al.* Single cell genomic study of *Dehalococcoidetes* species from deep-sea sediments of the Peruvian Margin. *ISME J* 2014;**8**:1831–42.

Kendall MM. *Methanococcus aeolicus* sp. nov., a mesophilic, methanogenic archaeon from shallow and deep marine sediments. *Int J Syst Evol Micr* 2006;**56**:1525–9.

Knight R, Jansson J, Field D, *et al.* Unlocking the potential of metagenomics through replicated experimental design. *Nat Biotechnol* 2012;**30**:513–20.

Le Bras Y, Roult A, Monjeaud C, *et al.* *JOBIM 2013: Towards a Life Sciences Virtual Research Environment*, eBGO: Rennes, France, 2013. https://www.e-biogenouest.org/resources/129/download/jobim˙YLeBras˙2013.pdf (20 April 2015, date last accessed).

Li P-Y, Xie B-B, Zhang X-Y, *et al.* Genetic structure of three fosmid-fragments encoding 16S rRNA genes of the Miscellaneous Crenarchaeotic Group (MCG): implications for physiology and evolution of marine sedimentary archaea: MCG crenarchaea. *Environ Microbiol* 2012;**14**:467–79.

Lloyd KG, Schreiber L, Petersen DG, *et al.* Predominant archaea in marine sediments degrade detrital proteins. *Nature* 2013;**496**:215–8.

Lomstein BA, Langerhuus AT, D'Hondt S, *et al.* Endospore abundance, microbial growth and necromass turnover in deep sub-seafloor sediment. *Nature* 2012;**484**:101–4.

Mackelprang R, Waldrop MP, DeAngelis KM, *et al.* Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. *Nature* 2011;**480**:368–71.

Martín-Cuadrado A-B, López-García P, Alba J-C, *et al.*, Metagenomics of the deep Mediterranean, a warm bathypelagic habitat. *PLoS One* 2007;**2**:e914.

Meng J, Wang F, Wang F, *et al.* An uncultivated crenarchaeota contains functional bacteriochlorophyll a synthase. *ISME J* 2009;**3**:106–16.

Meng J, Xu J, Qin D, *et al.* Genetic and functional properties of uncultivated MCG archaea assessed by metagenome and gene expression analyses. *ISME J* 2014;**8**:650–9.

Meyer F, Paarmann D, D'Souza M, *et al.* The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 2008;**9**:386.

Mikucki JA, Liu Y, Delwiche M, *et al.* Isolation of a methanogen from deep marine sediments that contain methane hydrates, and description of *Methanoculleus submarinus* sp. nov. *Appl Environ Microb* 2003;**69**:3311–6.

Mills HJ, Reese BK, St Peter C. Characterization of microbial population shifts during sample storage. *Front Microbiol* 2012;**3**:49.

Morono Y, Terada T, Nishizawa M, *et al.* Carbon and nitrogen assimilation in deep subseafloor microbial cells. *P Natl Acad Sci USA* 2011;**108**:18295–300.

Namiki T, Hachiya T, Tanaka H, *et al.* MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. *Nucleic Acids Res* 2012;**40**.

Orsi WD, Edgcomb VP, Christman GD, *et al.* Gene expression in the deep biosphere. *Nature* 2013;**499**:205–8.

Parkes RJ, Cragg B, Roussel E, *et al.* A review of prokaryotic populations and processes in sub-seafloor sediments, including biosphere: geosphere interactions. *Marine Geol* 2014;**352**:409–25.

Rinke C, Schwientek P, Sczyrba A, *et al.* Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 2013;**499**:431–7.

Rodrigue S, Malmstrom RR, Berlin AM, *et al.* Whole genome amplification and de novo assembly of single bacterial cells Ahmed N, ed. *PLoS One* 2009;**4**:e6864.

Roy H, Kallmeyer J, Adhikari RR, *et al.* Aerobic microbial respiration in 86-million-year-old deep-sea red clay. *Science* 2012;**336**:922–5.

Schippers A, Neretin L, Kallmeyer J, *et al.* Prokaryotic cells of the deep sub-seafloor biosphere identified as living bacteria. *Nature* 2005;**433**:861–4.

Schrenk MO, Huber JA, Edwards KJ. Microbial provinces in the subseafloor. *Ann Rev Mar Sci* 2010;**2**:279–304.

Wasmund K, Schreiber L, Lloyd KG, *et al.* Genome sequencing of a single cell of the widely distributed marine subsurface *Dehalococcoidia*, phylum *Chloroflexi*. *ISME J* 2013;**8**:383–97.

Watkins AJ, Roussel EG, Parkes RJ, *et al.* Glycine betaine as a direct substrate for methanogens (*Methanococcoides* spp.). *Appl Environ Microb* 2014;**80**:289–93.

Wrighton KC. Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla (vol 337, pg 1661, 2012). *Sciences* 2012;**338**:742.

Yarza P, Yilmaz P, Pruesse E, *et al.* Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat Rev Microbiol* 2014;**12**:635–45.