FEMS Microbiology Ecology, 92, 2016, fiv138

doi: 10.1093/femsec/fiv138
Advance Access Publication Date: 5 November 2015
Research Article

RESEARCH ARTICLE

A minimalistic microbial food web in an excavated deep subsurface clay rock

Alexandre Bagnoud[1,#], Ino de Bruijn[2], Anders F. Andersson[3], Nikitas Diomidis[4], Olivier X. Leupin[4], Bernhard Schwyn[4] and Rizlan Bernier-Latmani[1,*]

[1]Environmental Microbiology Laboratory, EPFL, Building CH, CH A1 375, Station 6, Lausanne, 1015, Switzerland, [2]Science for Life Laboratory, Bioinformatics Infrastructure for Life Sciences, Tomtebodavägen 23, Stockholm, 171 65, Sweden, [3]Science for Life Laboratory, KTH Royal Institute of Technology, Tomtebodavägen 23, Stockholm, 171 65, Sweden and [4]NAGRA, National Cooperative for the Disposal of Radioactive Waste, Hardstrasse 73, Wettingen, 5430, Switzerland

*Corresponding author: EPFL ENAC IIE EML, CH A1 375 (Bâtiment CH), Station 6, CH-1015 Lausanne, Switzerland. Tel: +41-21-693-5001; Fax: +41-21-693-62-05; E-mail: rizlan.bernier-latmani@epfl.ch
#Present address: Stream Biofilm and Ecology Research Laboratory, EPFL, Station 2, Lausanne, 1015, Switzerland.
One sentence summary: A simple heterotrophic microbial community thrives in borehole water in a deep subsurface clay rock.
Editor: Tillmann Lueders

ABSTRACT

Clay rocks are being considered for radioactive waste disposal, but relatively little is known about the impact of microbes on the long-term safety of geological repositories. Thus, a more complete understanding of microbial community structure and function in these environments would provide further detail for the evaluation of the safety of geological disposal of radioactive waste in clay rocks. It would also provide a unique glimpse into a poorly studied deep subsurface microbial ecosystem. Previous studies concluded that microorganisms were present in pristine Opalinus Clay, but inactive. In this work, we describe the microbial community and assess the metabolic activities taking place within borehole water. Metagenomic sequencing and genome-binning of a porewater sample containing suspended clay particles revealed a remarkably simple heterotrophic microbial community, fueled by sedimentary organic carbon, mainly composed of two organisms: a *Pseudomonas* sp. fermenting bacterium growing on organic macromolecules and releasing organic acids and H_2, and a sulfate-reducing Peptococcaceae able to oxidize organic molecules to CO_2. In Opalinus Clay, this microbial system likely thrives where pore space allows it. In a repository, this may occur where the clay rock has been locally damaged by excavation or in engineered backfills.

Keywords: deep subsurface; Opalinus Clay; microbial activity; sulfate reduction; metagenomics

INTRODUCTION

The existence of microbial life in the deep terrestrial subsurface has been documented (Pedersen 2000; Colwell and Smith 2004) but little is known about the diversity and functioning of the microbial communities in these systems. We depend on the subsurface for several critical services such as groundwater supply, resource recovery, gas storage and nuclear waste disposal. Hence, understanding the distribution and diversity of microbial

Received: 15 July 2015; **Accepted:** 3 November 2015

communities as well as their metabolic potential and activity in the subsurface is an active area of research. Unfortunately, it is hampered by the logistical difficulties associated with limited access to this environment (Edwards, Becker and Colwell 2012).

Clay formations such as Opalinus, Boom, or Callovo-Oxfordian Clay are low-permeability rocks that are being considered in several European countries for the disposal of radioactive waste. Such disposal sites, named geological repositories, represent an international consensus on the favored option to contain the waste until radionuclides decay to an acceptable level of radioactivity. Repository layouts and engineered barrier systems are specific to each country depending on national regulations and local geology (IAEA 2003). Extensive study of the physical and chemical characteristics of potential host clay rocks has been conducted over the past 30 years but the extant microbial communities and their associated metabolic potential remain poorly described. The rationale for investigating the metabolic potential of repository rock microbial communities is to contribute to the scientific basis for the evaluation of their possible impact on the safety of the repository. There are a number of microbial processes that could impact the repository barriers either positively or negatively. For instance, gas production could lead to pressure build-up in the vaults and enhancement of the weathering of the barriers around the waste (Stroes-Gascoyne *et al.* 2007). Positive effects include the consumption of gas produced by steel corrosion in order to avoid critical pressure build-up in repositories, and the reduction of radionuclide mobility when microorganisms precipitate and/or assimilate these compounds. Delineating which of the above metabolic processes are significant in these environments is the first step of a more complete microbial characterization of repository rocks and a more exhaustive safety case.

In Switzerland, one of the preferred host rocks for geological repository is Opalinus Clay, a sedimentary rock that was formed about 170 million years ago. Salient features of this rock formation are its very small pore size, with a mean of 10–20 nm, a very low hydraulic conductivity ($<10^{-13}$ m/s) and a strongly reducing environment, which is due to the presence of minerals such as siderite and pyrite (Mazurek 1999; Thury and Bossart 1999). The Opalinus Clay also contains up to 1.5% (w/w) organic matter that is mainly associated with the solid phase (Pearson *et al.* 2003; Courdouan *et al.* 2007). The slow diffusion of water occurring in this formation results in the porewater remaining in chemical equilibrium with the rock, and thus retaining the marine signature of its depositional environment, which explains sulfate concentrations between 10 and 20 mM (Pearson *et al.* 2003). The pore size distribution, favoring very small pores, also results in limited microbial activity in the rock, simply due to the lack of physical space. This was clearly shown by previous studies demonstrating that living microorganisms can be recovered from Opalinus Clay samples (Stroes-Gascoyne *et al.* 2007; Poulain *et al.* 2008), but their activity and distribution were below the detection limits and no DNA could be extracted from this substrate. The limiting factor for growth was shown to be lack of space, since autochtonous bacteria can grow when Opalinus Clay rock is placed in anoxic water (Stroes-Gascoyne *et al.* 2007). Thus, it is expected that during the construction of a repository, a network of micro- and macro-scale fractures will develop as a result of excavation, leading to the formation of an excavation-disturbed zone (EDZ) around the galleries that is characterized by slightly higher hydraulic conductivity (Bossart *et al.* 2002), and that represents a potential site of enhanced microbial activity.

The present study aims to describe the metabolic potential of the microbial community that develops in a drilled borehole in Opalinus Clay. This knowledge is necessary to fully assess the impact of microorganisms on the outcome of nuclear waste disposal in the geological repositories.

MATERIALS AND METHODS

The location of this study was the Underground Rock Laboratory (URL) of Mont Terri (St-Ursanne, Switzerland; Supplementary Fig. S1), located about 300 m below the ground surface. This facility allows access to the Opalinus Clay via a number of boreholes. This work focused on a single borehole (BIC-A1) that has a vertical descending orientation, is 15.6 m long and was drilled in Gallery 08 on 19 March 2012, using a rotary drilling rig and air as a fluid (Fig. 1). The borehole was maintained under anoxic

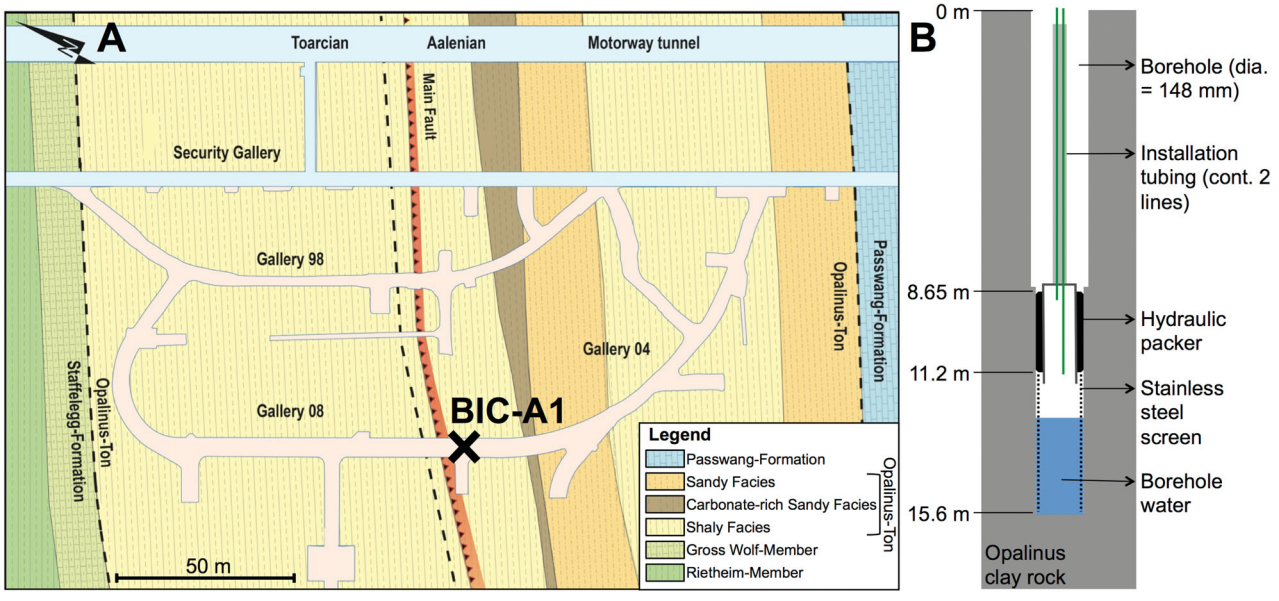


Figure 1. The BIC-A1 borehole. (**A**) Location of the BIC-A1 borehole in the URL of Mont Terri. The BIC-A1 borehole has a vertical descending orientation, indicated by the cross. (**B**) Vertical cross-section of the BIC-A1 borehole and its equipment. Modified from Swisstopo (Bossart and Thury 2008).

conditions until sampling by immediately flushing it with argon and by installing a hydraulic inflatable packer within 1 day of drilling, sealing off the rock formation from the atmosphere of the tunnel. Porewater from the formation accumulated in the borehole over time at an average rate of 44 mL/day and provided a possible environment for microbial growth. Materials in contact with the porewater were selected to minimize chemical interaction with water and included stainless steel 316L, neoprene and polyamide.

A water sample was recovered from the BIC-A1 borehole 10 months after it was drilled and sealed. Sampling was performed using a sterile glass bottle attached to an aluminum rod sterilized by ethanol and flaming. The water sample, which contained considerable suspended clay particles, was kept anoxic until filtration onsite using a sterile filtration device and a sterile 0.22 μm polycarbonate membrane. The membrane was stored in LifeGuard Soil Preservation Solution (Mo Bio, Carlsbad, CA, USA) and placed at −20°C within a few hours. As a negative control, sterile water was handled in the same way, via filtration onsite (at the Mont Terri URL) and no amplifiable DNA was extractable from that sample.

BIC-A1 filtered water was characterized for cation and anion composition using ion chromatography (DX-3000, Dionex, Sunnyvale, CA, USA). Major anions were separated with an IonPac AS11-HC column, using a KOH gradient from 0.5 to 30 mM for elution while major cations were separated with an IonPac CS16 cation-exchange column, using 40 mM of methanesulfonic acid for elution. Metals (Mn, Cr, Zn, Si, Al, Fe, Co, Cu, Sr and Ni) were measured using inductively coupled plasma–optical emission spectroscopy (ICP-OES, Shimadzu ICPE 9000), and the samples and standards were prepared in 0.1 M HNO_3 (final concentration). Fe(II) and S(−II) were measured using an assay described by Cline (1969) and Stookey (1970), respectively. The pH was measured in unfiltered water using an Orion™ ROSS Ultra™ pH electrode (Thermo Scientific).

DNA was extracted from particulate matter retained by the filtration membranes (0.2 μm mesh) using a slightly modified protocol of the DNA Spin Kit for Soil (MP Biomedicals, Santa Ana, CA, USA), including (i) a 5 min incubation of the biomass in the lysis buffer at 60°C before bead-beating, and (ii) a second extraction with new lysis buffer and another bead-beating step. Extracted DNA was further purified using the standard protocol of Genomic DNA Clean & Concentrator purification kit (Zymoresearch, Irvine, CA, USA).

The library was prepared for Illumina MiSeq sequencing using PCR primers for the hypervariable region V4 of the 16S rRNA gene, which are 515F (5′-GTG CCA GCM GCC GCG GTAA-3′) and 806R (5′-GGA CTA CHV GGG TWT CTA AT-3′), using the thermocycling conditions given by Caporaso *et al.* (2011). Sequencing was done using a 2 × 250 bp read configuration. The QC filtered reads (minimum length of 292 bp with an expected error smaller than 2) were scanned with Duk version 1.05 to remove contaminating sequences, such as Illumina adapters and human contaminants (Li, Copeland and Han 2011), trimmed to 165 bp, merged with FLASH version 1.2.6 using a maximum mismatch density of 0.3 and a minimum overlap of 20 nucleotides (Magoč and Salzberg 2011), and poor quality merged reads were discarded or trimmed (expected error of 2 bp). USEARCH version 7.0.959 (Edgar 2010) was used to define operational taxonomic units (OTUs) using a maximum dissimilatory radius of 3%. Chimeric sequences were filtered by UCHIME (Edgar *et al.* 2011). RDP Classifier version 2.5 (Wang *et al.* 2007) was used for taxonomic assignments, using Greengenes V4 database (DeSantis *et al.* 2006), using a cutoff of 0.5. Finally, QIIME version 1.7.0 (Caporaso *et al.* 2010) was run to generate the 'biom' file and the phylum-level table.

Because the first DNA extract was entirely used for 16S rRNA gene analysis, a second DNA extract was obtained from a second filtration membrane with the method described above for metagenomic sequencing. However, the purification step of the DNA extract was done by ethanol precipitation. TruSeq Nano DNA Sample Prep Kit (Illumina) was used for library preparation, using a fragment length of 400–800 bp, before Illumina HiSeq paired-end sequencing.

First, the Illumina metagenomic reads were post-processed by Casava version 1.8.2 (http://support.illumina.com/sequencing/sequencing˙software/casava.html). The assembly was performed with Ray version 2.3.1 (Boisvert *et al.* 2012) on a Cray XE6 system using 1024 cores in 4.5 h and a kmer length of 31 bp. Two other kmer lengths were tested, 41 and 51 bp, but they gave slightly worse results. Contigs were annotated using the JGI IMG pipeline (Markowitz *et al.* 2012). Reads were mapped to the contigs in order to calculate their coverage, using Bowtie 2 version 2.1.0 (Langmead and Salzberg 2012) with default parameters. Duplicate reads were removed subsequently using MarkDuplicates from Picard tools version 1.77 (http://broadinstitute.github.io/picard/). Reads were also mapped to contigs belonging to 16 other metagenomic samples, all originating from an adjacent borehole, that had been amended with hydrogen and that contained similar microbial species according to 16S rRNA gene sequencing. Binning, the clustering of contigs according to their tetranucleotide frequency and their distribution across samples (based on their coverage), was carried out with CONCOCT version 0.2 (Alneberg *et al.* 2014) run using contigs with a minimum length of 5 kb. Once the binning was completed, the completeness and contamination in each bin was evaluated using a single-copy gene analysis. This evaluation was done using the lineage mode of CheckM version 0.9.4 (Parks *et al.* 2015).

Additionally, pathways were identified manually using KO, COG, EC annotations, the KEGG database (Kanehisa *et al.* 2014), the MetaCyc database (Caspi *et al.* 2014) and textbook biochemical pathways (Kim and Gadd 2008). PSORTb version 3.0.2 was used for predicting subcellular localization of gene products (Yu *et al.* 2010). The prediction of ribosomal RNA genes in the metagenome was done by Barrnap version 0.4.2 (https://github.com/Victorian-Bioinformatics-Consortium/barrnap) and their taxonomic affiliation with the RDP Classifier version 2.7 (Wang *et al.* 2007), using a confidence threshold of 0.8. The NCBI BLASTN platform (http://blast.ncbi.nlm.nih.gov/Blast.cgi) was also used for taxonomic annotation of 16S rRNA genes. MLTreeMap version 2.061 was used for taxonomic annotation of the bins (Stark *et al.* 2010), as well as phylogenetic distribution output from the JGI IMG pipeline, which provides, for each gene, the taxonomic annotation of the organism harboring its closest homolog. Taxonomic annotation of bins deemed sufficiently pure and complete to be metagenome-assembled genomes (MAGs) was also done using average nucleotide identity (ANI) with references genomes (Goris *et al.* 2007), using the Kostas Lab platform with default parameters (window size of 1000 bp with a step of 200 bp, minimum alignment length of 700 bp and identity of 70%; http://enve-omics.ce.gatech.edu/ani/index). Proteins involved in sulfate reduction, hydrogen oxidation and the acetyl-CoA pathway of a microorganism abundant in the BIC-A1 borehole were compared with that of reference genomes, using BLASTP from BLAST version 2.2.28 (Altschul *et al.* 1990). Proteins sequence files of reference genomes were downloaded from the NCBI database (http://www.ncbi.nlm.nih.gov/genome). Synteny of these gene

loci between query and reference genomes was simply observed using each gene's number ID.

The abundance $a(B_i)$ of each bin B_i compared the abundance of the entire microbial community M as defined as follows:

$$M = \sum_{i=1}^{m} a(B_i)$$

where m is the number of bins and

$$a(B_i) = \frac{\sum_{j=1}^{n} l(C_j) \times c(C_j)}{\sum_{j=1}^{n} l(C_j)}$$

where n is the number of contigs in bin B_i, $l(C_j)$ is the length of contig C_j and $c(C_j)$ the mean coverage of C_j. That is, the abundance of a bin is expressed as its mean coverage. The relative abundance $r(B_i)$, or contribution, of each bin is defined as:

$$r(B_i) = \frac{a(B_i)}{M}$$

The following data are being or have been deposited in NCBI databases.

(i) Metagenomics raw reads: deposited at SRA NCBI; accession SRX672050.
(ii) V4 16S rRNA gene sequences:
  (a) Raw reads: deposited at SRA; accession SRX892770.
  (b) V4 16S rRNA gene sequences: deposited at NCBI; accession KP901405–KP902213.
(iii) Metagenome assembly:
  (a) Comamonadaceae bacterium BICA1-1: a whole-genome shotgun project; deposited at DDBJ/EMBL/GenBank; accession JUEA00000000. The version described in this paper is JUEA01000000.
  (b) Peptococcaceae bacterium BICA1-7: a whole genome shotgun project; deposited at DDBJ/EMBL/GenBank; accession JUEB00000000. The version described in this paper is JUEB01000000.
  (c) Pseudomonas BICA1-14: a whole genome shotgun project; deposited at DDBJ/EMBL/GenBank; accession JUEC00000000. The version described in this paper is JUEC01000000.
  (d) Desulfotomaculum sp. BICA1-6: a whole genome shotgun project; deposited at DDBJ/EMBL/GenBank; accession JUED00000000. The version described in this paper is JUED01000000.
  (e) Desulfosporosinus sp. BICA1-9: a whole genome shotgun project; deposited at DDBJ/EMBL/GenBank; accession JUEE00000000. The version described in this paper is JUEE01000000.
  (f) Peptococcaceae bacterium BICA1-8: a whole genome shotgun project; deposited at DDBJ/EMBL/GenBank; accession JUEF00000000. The version described in this paper is JUEF01000000.
  (g) All other contigs were deposited as rock porewater genome: a whole genome shotgun project; deposited at DDBJ/EMBL/GenBank; accession JUEG00000000. The version described in this paper is version JUEG01000000.

Additionally, the IMG/MER annotation can be found with genome ID 3300002735.

RESULTS

Chemical analysis of the BIC-A1 borehole porewater revealed the presence of sulfide (7.4 μM; Supplementary Table S1). Illumina sequencing of 16S rRNA gene amplicons from this water produced 127,610 high quality reads (representing 86% of the raw reads) of the V4 region of the 16S rRNA gene, grouped into 87 OTUs (Supplementary Table S2). The microbial community composition of this sample is presented in Fig. 2. Most predominant are Firmicutes, represented by *Desulfotomaculum* sp. (59% of microbial community) and *Desulfosporosinus* sp. (8%), and a *Pseudomonas* sp. (26%).

Illumina metagenomic sequencing of the BIC-A1 sample produced more than 60 million high quality reads, assembled into 2,246 contigs longer than 5000 bp (whose N50 length is 45,685 bp). These were binned into 14 genome bins using CONCOCT (Alneberg *et al.* 2014), which combines information on coverage across samples from an adjacent borehole and tetranucleotide composition (Supplementary Table S3 and Supplementary Fig. S2). Two contigs were considered to be contaminating sequences that occurred during the sequencing process, because they matched *Arabidopsis thaliana* sequences perfectly (100% similarity), and were removed (Supplementary Table S3 and Supplementary Fig. S2).

Five of the 14 bins harbored a 16S SSU rRNA gene (Supplementary Fig. S2). One of them, bin c7, contained five versions of the 16S rRNA gene. Two of them are nearly identical (97.8% similarity between BICA1a_1053282 and BICA1a_1148202), and three of them are non-identical to each other and to all other 16S rRNA genes in this bin (between 80 and 89% similarity). The inaccurate binning of contigs containing the 16S rRNA gene results from the highly conserved regions in this gene, which leads to incorrect coverage assessments. Also, since they are structural RNA genes, their nucleotide signatures are distinct from other parts of the same genome (Noble, Citek and Ogunseitan 1998). Thus, the three contigs harboring these 16S rRNA genes (originally assigned to bin c7) were reassigned to other bins (c4, c6 and c14c) as described in Supplementary Fig. S2. Other taxonomic annotations of bins, such as those provided by MLTreeMap and by the IMG pipeline, allowed the manual reassignment of these inaccurately binned 16S rRNA gene-bearing contigs to other bins, and enabled the taxonomic annotation of bins missing a 16S rRNA gene (Supplementary Tables S3 and S4 and Supplementary Fig. S2). One of these contigs, containing a 16S rRNA sequence identical to that of OTU 7 obtained from direct 16S rRNA gene sequencing analysis (Supplementary Table S2 and Fig 2), was reassigned to bin c6. The reassignment was based on similar contributions to the microbial community based on 16S rRNA gene sequencing as compared with metagenomic analysis (Table 1, Fig. 2 and Supplementary Table S2). Bin c6 could thus be annotated as a *Desulfotomaculum* sp. Bins c10 and c13, both annotated as *Streptomyces*, were merged into a newly defined bin c14c (Supplementary Fig. S2), because only a single *Streptomyces* OTU was detected by 16S rRNA gene sequencing (Supplementary Table S2), and because the two bins have complementary single-copy genes profiles (results not shown). Three bins were not considered further: bin c11, which consists of a single contig, 5733 bp in size (Supplementary Fig. S2), and bins c0 and c2, both small in size (Supplementary Fig. S2) and annotated as *Pseudomonas*. The latter were not considered because a single dominating *Pseudomonas* OTU was detected by 16S rRNA gene sequencing (Supplementary Table S2), and it corresponds to bin c5 (Fig. 2). The direct correspondence was uncovered because

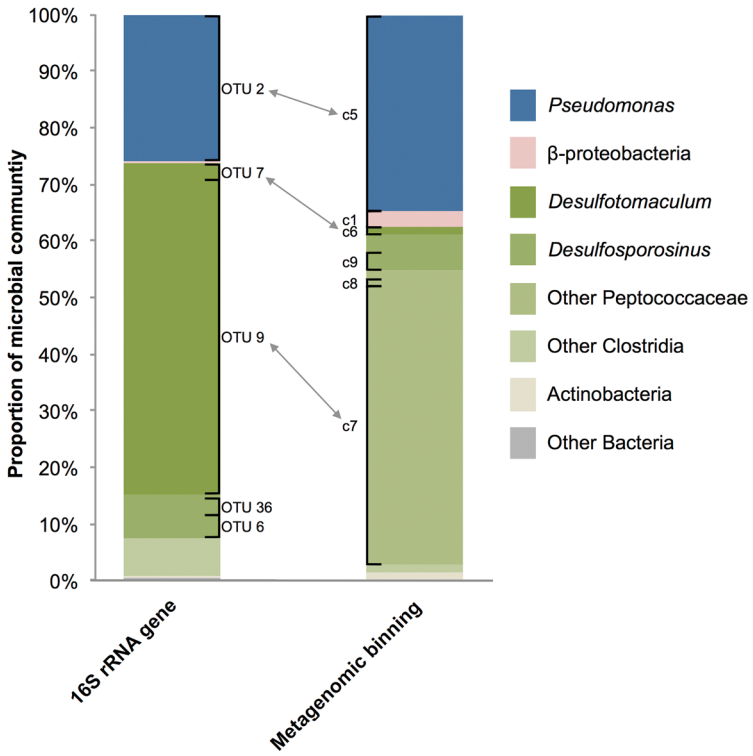


Figure 2. Microbial community composition of the BIC-A1 borehole, based on 16S rRNA gene sequencing (V4 region) and the average contig coverage of the metagenome binning. The predominant OTUs and MAGs are indicated. The arrows indicate corresponding OTUs and bins.

Table 1. Taxonomic affiliation size, number of contigs, proportion in the microbial community, average coverage, completeness and contamination (assessed by CheckM), corresponding OTU from 16S rRNA gene sequencing (if any), and the number of observations of the OTU in 22 water samples collected from seven other boreholes in the Mt-Terri URL for the selected bins.

MAG	Taxonomic affiliation	cluster length (Mbp)	Number of contigs	Proportion	Coverage	Complete-ness	Conta-mination	Corresponding OTU from 16S rRNA gene sequencing	Number of observations in other borehole water samples
c1	Comamonadaceae	2.97	128	0.03	52.52x	95.79%	12.89%		
c5	*Pseudomonas*	8.15	274	0.35	691.41x	100.00%	16.28%	OTU 2	22 / 22
c6	*Desulfotomaculum*	3.73	107	0.02	29.99x	98.09%	2.64%	OTU 7	18 / 22
c7	Peptococcaceae	4.76	36	0.50	999.06x	99.37%	2.53%	OTU 9	19 / 22
c8	Peptococcaceae	3.98	162	0.01	23.76x	93.28%	4.21%		
c9	*Desulfosporosinus*	6.99	266	0.03	60.48x	99.43%	8.48%		

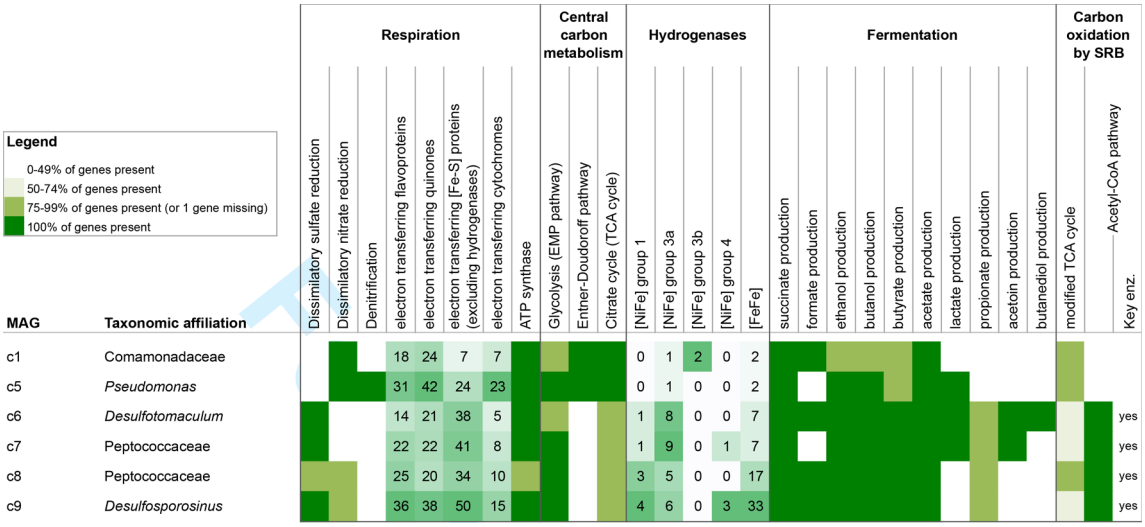
of the perfect (100%) match between 16S rRNA gene sequences obtained from direct 16S rRNA sequencing and metagenomic sequencing. We could have merged these three *Pseudomonas* bins, but this would have created an 8.15 Mbp genome, which is significantly larger than typical *Pseudomonas* genome sizes. Indeed, amongst the 1734 *Pseudomonas* genomes available on NCBI at the time of writing, only three are larger in size. A possible explanation for the existence of these two small *Pseudomonas* bins is that they constitute misassembled and chimeric contigs containing some *Pseudomonas* sequences. These bins could also represent plasmids belonging to the *Pseudomonas* organism.

Thus, after manual correction, 10 bins were identified (Supplementary Fig. S2), suggesting the dominance of 10 microorganisms in clay particles suspended in porewater. The final binning is presented in Supplementary Table S3.

Out of the 10 bins identified, six were nearly pure and complete (bins c1 and c5 to c9; Table 1 and Supplementary Table S5) based on the CheckM analysis, and thus can be considered as MAGs. They all have a completeness index greater than 90% and a contamination index lower than 20%. These bins represent four Peptococcaceae species, among them a *Desulfosporosinus* (c9) and a *Desulfotomaculum* species (c6), a *Pseudomonas* (c5) and a Comamonadaceae species (c1). The two dominant organisms of the BIC-A1 borehole based on 16S rRNA gene sequencing belong to two of these six MAGs. They are a Peptococcaceae species (bin c7 which corresponds to OTU 9) and a *Pseudomonas* species (bin c5 which corresponds to OTU 2) and represent 50 and 35% of the overall microbial community, respectively (Fig. 2 and Table 1).

The two 16S rRNA gene sequences of Peptococcaceae c7 are partial and don't span the V4 region covered by the 16S rRNA

Table 2. Summary of the pathway annotation of the MAGs. Numbers indicate the number of occurrences of relevant genes in the corresponding MAG. For a more complete list of annotated pathways, refer to Supplementary Tables S7–S9.



gene sequencing primers. However, this bin can be assigned to OTU 9 because both entities have a similar contribution to the microbial community (Fig. 2). The overall accuracy of the taxonomic assignment for each bin is evident from Supplementary Table S4. This is because the 16S rRNA gene (when present in the bin) best BLAST hit corresponds well with the MLTreeMap annotation and the IMG phylogenetic distribution. For example, in the case of bin c5 (*Pseudomonas*), the 16S rRNA BLAST result indicates a 99% similarity to *Pseudomonas chloridismutans*, the MLTreeMap analysis shows 100% correspondence with the Pseudomonadaceae family, and the IMG phylogenetic distribution shows 92% assignment to the *Pseudomonas* genus. In addition, the BIC-A1 microbial community profile obtained here is similar to the one obtained with the 16S rRNA gene sequencing, as shown in Fig. 2.

Because the 16S rRNA sequencing data were also obtained for 22 other water samples collected at Mont Terri URL, we compared the OTUs found in the BIC-A1 borehole to those found in other boreholes. The dominant OTUs in BIC-A1 (corresponding to bins c5, c6 and c7) are also present in the majority (>18 out of 22) of the other boreholes (Table 1). One of the boreholes, BHT-1, was drilled under sterile and anoxic conditions (Vinsot *et al.* 2014).

The similarity of the six MAGs to known genomes was calculated using their average nucleotide identity (ANI). Bins c6–c9 were compared to all known Peptococcaceae genomes, while bins c1 and c5 were contrasted with genomes harboring the highest 16S rRNA identity according to BLASTN. The results show that five of the six bins represent at minimum a new species. Four of these new MAGs, bins c1, c6, c7 and c9, may represent new genera since their highest ANI values (including at least 100 fragments) are lower than 85% as compared to existing genomes (Goris *et al.* 2007; Supplementary Table S6). Bin c5 matches both *Pseudomonas chloritidismutans* and *Pseudomonas xanthomarina*, with similarities greater than 97% (as these two *Pseudomonas* strains have an ANI > 97% by comparison with each other).

Metabolic pathway analysis of the six MAGs shows the preponderance of anaerobic metabolic processes such as dissimilatory sulfate or nitrate reduction, the Wood–Ljungdahl pathway, the Amon–Buchanan pathway and fermentation, as well as evidence for breakdown of dead biomass as evidenced by the abundance of lipases, proteases, nucleases and phosphatases in the *Pseudomonas* bin (Table 2 and Supplementary Tables S7–S9).

Because the dominant MAG c7 is phylogenetically close to the genus *Desulfotomaculum* (Supplementary Table S4) and because several species of this genus were sequenced, its genome was compared with that of *Desulfotomaculum acetoxidans*, a sulfate-reducing bacterium (SRB) capable of complete oxidation but unable to ferment (Spring *et al.* 2009), and *Desulfotomaculum reducens*, a fermenting SRB for which the mechanism of sulfate reduction was studied in some detail (Junier *et al.* 2010). Overall, the genes involved in sulfate reduction, dihydrogen oxidation and complete oxidation of acetate are remarkably conserved in terms of both identity and synteny (Table 3).

DISCUSSION

The Opalinus Clay formation is a compacted clay stone that exhibits very low hydraulic conductivity, a pore size distribution that is dominated by small pores (10–20 nm), and a porewater with a residence time of several million years (Stroes-Gascoyne *et al.* 2007). Previous attempts to characterize the microbial community have shown that the microorganisms were viable but likely metabolically inactive (dormant) and represented low biomass abundance (Stroes-Gascoyne *et al.* 2007). No evidence of microbial activity was observed when porewater was first collected from freshly drilled boreholes, and in particular no evidence for sulfide was ever found in a fresh borehole despite numerous measurements (Pearson *et al.* 2003; A. Fernandes, personal communication).

Upon drilling, porewater from the formation flows into the borehole, filling it over a period of months, and enabling direct sampling. The amount of time (10 months) elapsed between drilling and sampling allowed for accumulation of water in the borehole, but also for viable microorganisms to become active, grow and divide. The negative impact of a putative microbial contamination, which is a constant concern when working in environments where drilling is required, was shown to be limited because the dominant microorganisms

Table 3. Genome comparison of Peptococcaceae c7 with two reference genomes: *Desulfotomaculum acetoxidans* for the reductive acetyl-CoA pathway and hydrogenases involved in respiratory processes, and *Desulfotomaculum reducens* for genes involved in sulfate reduction. Grey boxes represent putative gene clusters in Peptococcaceae c7. The comparison was carried out using BLASTN. All e-values are significant ($<4 \times 10^{-20}$).

Ref. genome	Metabolic function		Ref. gene ID	*Desulfotomaculum* c7		BLAST alignment		
				Gene ID	IMG gene product annotation	Identity (%)	Ref. gene coverage	c7 gene coverage
Comparison to *Desulfotomaculum reducens* MI-1 genome	Dissimilatory sulfate reduction	NADH oxidation, H^+ translocation, quinone reduction	Dred_2036	BICA1a_1184301213	NADH:ubiquinone oxidoreductase subunit 2 (chain N)	64.36	0.997	0.997
			Dred_2037	BICA1a_1184301212	NADH:ubiquinone oxidoreductase subunit 4 (chain M)	72.48	0.989	0.985
			Dred_2038	BICA1a_1184301211	NADH:ubiquinone oxidoreductase subunit 5 (chain L)	71.1	0.996	0.997
			Dred_2039	BICA1a_1184301210	NADH:ubiquinone oxidoreductase subunit 11 or 4L	72.16	0.909	0.892
			Dred_2040	BICA1a_1184301209	NADH:ubiquinone oxidoreductase subunit 6 (chain J)	47.9	0.961	0.929
			Dred_2041	BICA1a_1184301208	NADH:ubiquinone oxidoreductase 23 kD subunit (chain I)	48.36	0.730	0.772
			Dred_2042	BICA1a_1184301207	NADH:ubiquinone oxidoreductase subunit 1 (chain H)	67.24	0.992	0.984
			Dred_2043	BICA1a_1184301206	NADH:ubiquinone oxidoreductase 49 kD subunit 7	73.35	0.990	0.995
			Dred_2044	BICA1a_1184301205	NADH:ubiquinone oxidoreductase 27 kD subunit	51.72	0.948	0.937
			Dred_2045	BICA1a_1184301204	NADH:ubiquinone oxidoreductase 20 kD subunit	76.05	0.844	0.924
			Dred_2046	BICA1a_1184301203	NADH:ubiquinone oxidoreductase subunit 3 (chain A)	64.96	0.986	0.978
		e^- transfer from quinone pool to DsrAB	Dred_3197	BICA1a_102506429	Dissimilatory sulfite reductase, gamma subunit	73.33	0.984	0.984
			Dred_3198	BICA1a_102506430	Fe-S oxidoreductase	64.75	0.968	0.979
			Dred_3199	BICA1a_102506431	Nitrate reductase gamma subunit	54.24	0.990	0.994
		Sulfite reduction to sulfide	Dred_3186	BICA1a_102506439	Dissimilatory sulfite reductase, alpha and beta subunits	58.5	0.981	0.898
			Dred_3187	BICA1a_102506440	Dissimilatory sulfite reductase, alpha and beta subunits	54.55	0.973	0.874
		Sulfate activation and reduction to sulfite	Dred_0635	BICA1a_118430176	ATP sulfurylase (sulfate adenylyltransferase)	87.86	0.991	0.986
			Dred_0636	BICA1a_118430177	Adenosine-5'-phosphosulfate reductase beta subunit	75.54	0.935	0.941
			Dred_0637	BICA1a_118430178	Succinate dehydrogenase/fumarate reductase, flavoprotein subunit	75.12	0.997	0.994
			Dred_0638	BICA1a_118430179	*No gene product annotation*	66.59	0.996	0.991
			Dred_0639	BICA1a_118430180	Methyl-viologen-reducing hydrogenase, delta subunit	66.35	0.998	0.998
		Qmo complex reduction by heterodisulfide reductase	Dred_1325	BICA1a_1172615263	*No gene product annotation*	50.3	0.869	0.855
			Dred_1326	BICA1a_1172615264	Heterodisulfide reductase, subunit B	50.18	0.987	0.980
			Dred_1327	BICA1a_1172615269	NADPH-dependent glutamate synthase beta chain	57.48	0.993	0.991
			Dred_1328	BICA1a_1172615265	Pyridine nucleotide-disulphide oxidoreductase	53.55	0.988	0.991
			Dred_1329	BICA1a_1172615275	Coenzyme F420-reducing hydrogenase, delta subunit	62.5	0.418	0.959
			Dred_1329	BICA1a_1172615268	Coenzyme F420-reducing hydrogenase, delta subunit	42.27	0.637	0.837
			Dred_1330	BICA1a_1172615273	Coenzyme F420-reducing hydrogenase, beta subunit	25.49	0.780	0.784
		Qmo complex reduction by flavoproteins	Dred_1778	BICA1a_1075448205	Electron transfer flavoprotein, beta subunit	61.09	0.993	0.994
			Dred_1779	BICA1a_1075448204	Electron transfer flavoprotein, alpha subunit	69.33	0.995	0.995
		H^+ translocation coupled to PPi hydrolysis	Dred_2985	BICA1a_118430132	Inorganic pyrophosphatase	79.61	0.995	0.972
Comparison to *Desulfotomaculum acetoxidans* DSM 771 genome	Hydrogen oxidation for respiratory processes	[NiFe]-hydrogenase	Dtox_0791	BICA1a_120211075	Formate hydrogenlyase subunit 3	68.68	0.996	0.997
			Dtox_0792	BICA1a_120211074	Formate hydrogenlyase subunit 4	74.74	0.948	0.948
			Dtox_0793	BICA1a_120211073	Hydrogenase 4 membrane component (E)	63.03	0.974	0.983
			Dtox_0794	BICA1a_120211072	Formate hydrogenlyase subunit 3	72.41	0.982	0.980
			Dtox_0795	BICA1a_120211071	Ni,Fe-hydrogenase III large subunit	64.67	0.984	0.984
			Dtox_0796	BICA1a_120211070	Ni,Fe-hydrogenase III small subunit	74.58	0.993	0.993
		[FeFe]-hydrogenase	Dtox_0168	BICA1a_1094883155	Iron only hydrogenase large subunit, C-terminal domain	75.82	0.991	0.983
			Dtox_0169	BICA1a_1094883154	Fe-S-cluster-containing hydrogenase components 1	65.77	0.986	0.990
			Dtox_0172	BICA1a_1094883150	Iron only hydrogenase large subunit, C-terminal domain	79.37	0.995	0.995
			Dtox_0173	BICA1a_1094883149	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit	80.74	0.997	0.997
			Dtox_0174	BICA1a_1094883148	Ferredoxin	53.91	0.953	0.969
			Dtox_0175	BICA1a_1094883147	Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase	53.89	0.964	0.929
			Dtox_0176	BICA1a_1094883146	NADH:ubiquinone oxidoreductase 24 kD subunit	60	0.990	0.990
			Dtox_0177	BICA1a_1094883145	DRTGG domain	64.86	0.943	0.959
			Dtox_0178	BICA1a_1094883144	Iron only hydrogenase large subunit, C-terminal domain	72.58	0.974	0.969
	Acetate oxidation to CO_2 (acetyl-CoA pathway)	methylene-H_4F + CO_2 → acetyl-CoA + H_4F	Dtox_1285	BICA1a_1172615266	5,10-methylenetetrahydrofolate reductase	59.68	0.979	0.963
			Dtox_1276	BICA1a_1172615277	Pterin binding enzyme	72.83	0.994	0.994
			Dtox_1275	BICA1a_1172615278	CO dehydrogenase/acetyl-CoA synthase delta subunit	69.62	0.995	0.995
			Dtox_1272	BICA1a_1172615280	CO dehydrogenase/acetyl-CoA synthase gamma subunit	68.31	0.996	0.996
			Dtox_1271	BICA1a_1172615281	CO dehydrogenase/acetyl-CoA synthase beta subunit	80.22	0.998	0.998
			Dtox_1270	BICA1a_1172615282	6Fe-6S prismane cluster-containing protein	61.52	0.998	0.998
		CO_2 reduction to formate	Dtox_3503	BICA1a_1089845305	Uncharacterized anaerobic dehydrogenase	54.56	0.580	0.995
			Dtox_3507	BICA1a_1089845306	Uncharacterized anaerobic dehydrogenase	68.18	0.392	0.992
		H_4F + formate → formyl-H_4F	Dtox_0231	BICA1a_109488329	Formyltetrahydrofolate synthetase	44.81	0.986	0.982
		formyl-H4F → methylene-H4F	Dtox_2589	BICA1a_1172615153	5,10-methylene-tetrahydrofolate dehydrogenase	42.76	0.980	0.987

identified in this borehole were also detected in seven other boreholes drilled at different times, with different tools and by different people (Table 1). Among these boreholes is BHT-1, which was drilled using aseptic techniques (Vinsot *et al.* 2014). Evidence of microbial growth includes the presence of soluble hydrogen sulfide (Supplementary Table S1), which, as indicated above, is ordinarily absent from pristine porewater (Pearson *et al.* 2003; A. Fernandes, personal communication). The Opalinus Clay includes organic compounds in the form of low molecular mass organic acids such as acetate (~200 μM) (Courdouan *et al.* 2007) and solid-phase organic matter in the form of humic and fulvic acids, *n*-alkanes C_{20}–C_{33}, and refractory type III kerogen (Pearson *et al.* 2003). Thus, we concluded that in the borehole itself, microbial activity was established along with the attendant growth of indigenous microorganisms. Sulfate reduction fueled by organic matter, and

potentially H_2 from radiolytic decay of water, was an obvious candidate for the active metabolism due to the detection of sulfide. Nonetheless, the goal of the detailed metagenomic study was to attempt to reconstruct a more complex microbial metabolic web.

In BIC-A1 borehole water, metagenomic sequencing and binning enabled the identification of the metabolic capabilities of six microorganisms (Table 1): one Comamonadaceae (c1), one *Pseudomonas* (c5) and four Peptococcaceae (c6–c9), together representing 93% of the microbial community. Thus, this community is of remarkably low diversity. This low diversity highlights the challenges of microbial survival in an oligotrophic environment with limited electron acceptors (sulfate and carbon dioxide are the only ones available), and in which low metabolic activity is the norm.

Based on the metagenomic sequences, we can place these organisms into two groups according to their metabolic capabilities. MAGs c1 and c5 are capable of respiration using nitrate as the final electron acceptor, reducing it to either ammonium or dinitrogen (Table 2). This respiration can be coupled to the oxidation of complex organic compounds to carbon dioxide through the Embden–Meyerhof–Parnas (EMP) or the Entner–Doudoroff (ED) pathway. However, no nitrate was detected in the BIC-A1 borehole (Supplementary Table S1), suggesting that an alternative metabolism is underway. We propose that the organisms corresponding to these two MAGs ferment rather than respire. Even though the great majority of *Pseudomonas* species are not fermenters, some are known to ferment pyruvate, producing acetate (Moore *et al*. 2006; Schreiber *et al*. 2006). For example, it was shown that *Pseudomonas aeruginosa* is able to live using this pathway after oxygen and nitrate are depleted (Eschbach *et al*. 2004).

The second group (c6, c7, c8 and c9), microorganisms all belonging to the Peptococcaceae family, are similar with respect to their metabolic abilities (Table 2). Based on their individual genome sequences, they are all able to use sulfate as a terminal electron acceptor, have an elaborate electron transfer chain and include the ATP synthase complex. This sulfate reduction capability is likely the source of the hydrogen sulfide detected in BIC-A1 borehole water (Supplementary Table S1). All sulfate-reducing bacteria identified in this sample are complete oxidizers, meaning that they include the complete set of enzymes required to oxidize acetate to carbon dioxide, including the catalytic subunit of carbon monoxide dehydrogenase. In addition to sulfate reduction, the organisms harboring these genomes appear to ferment organic carbon, through glycolysis, producing succinate, formate, ethanol, butyrate, butanol and acetate. The four sulfate-reducing MAGs all have genes coding for group 1 [NiFe]-hydrogenases, suggesting that they can use dihydrogen as a source of electrons for respiration. As they all possess the enzymes for the acetyl-CoA pathway, which is reversible, they are predicted to be autotrophic in the absence of organic carbon and in the presence of dihydrogen and carbon dioxide.

Here we will further describe the metabolic potential of the two dominant MAGs (c5 and c7, which, when combined, correspond to 85% of the microbial community), as they are representative of the two groups of organisms identified.

The *Pseudomonas* MAG (c5), which represent 35% of the microbial community of sample BIC-A1 (Table 1), includes genes encoding proteins involved in extracellular phosphate hydrolysis (Supplementary Table S9): phytases, specific and non-specific organic phosphatases (Rodríguez and Fraga 1999), polyphosphatase, inorganic pyrophosphatases (Gobat, Aragno and Matthey 2004) and the machinery to produce gluconic acid, which solubilizes inorganic phosphate from phosphate-bearing minerals in Opalinus Clay (apatite or monazite) (Pearson *et al*. 2003). Extracellular degradation of complex organic matter is also part of the metabolic capabilities in this MAG (Supplementary Table S9): genes encoding proteins for lipid degradation (ABC transport, chemotaxis and esterase/lipase), protein degradation (membrane-bound, secreted, ABC transport), nucleic acid degradation (other than restriction nuclease, ribonucleases, polymerases and repair nucleases), amino acid mineralization (asparaginases and glutaminases), urea mineralization (Gobat, Aragno and Matthey 2004), and starch and amorphous cellulose degradation (Kim and Gadd 2008) are all identifiable in the MAG. However, this organism does not appear to feed on alkanes found in Opalinus Clay, because it lacks genes coding for both benzyl- and alkylsuccinate synthase, typically involved in anaerobic alkane degradation (Callaghan *et al*. 2010). Figure 3A shows the putative metabolic pathways of this MAG (c5) in the BIC-A1 porewater. From these observations, it appears that this organism can use complex organic carbon compounds that are broken down extracellularly prior to transport inside the cell and catabolized through diverse fermentation pathways. *Pseudomonas* c5 can generate dihydrogen gas using [FeFe]-hydrogenases or by group 3a [NiFe]-hydrogenases, or it can produce lactate, ethanol, butanol or succinate through fermentation.

The dominant MAG in the Peptococcaceae family is bin c7, which represents 50% of the microbial community (Table 1). As with the other MAGs belonging to this family, it represents sulfate-reducing bacteria, capable of oxidizing acetate to CO_2 and of autotrophic growth. According to the metagenomic annotation, Peptococcaceae c7 can generate reducing equivalents from the oxidation of acetate, formate, lactate, ethanol, butanol, butyrate and propionate, using reverse fermentation pathways (Fig. 3B). Interestingly, the propionate oxidation pathway is only found in sulfate-reducing bacteria in this study, but all of them are missing methylmalonyl-CoA carboxytransferase (EC 2.1.3.1), which can generate pyruvate and methylmalonyl-CoA from oxaloacetate and propionyl-CoA (Stams *et al*. 1984; Supplementary Fig. S3). This metabolism appears to be catalysed instead by propionyl-CoA carboxylase (EC 6.4.1.3), for which both MAGs c6 and c7 have a gene. This alternative reaction can transform propionyl-CoA to methylmalonyl-CoA, and pyruvate can be produced from oxaloacetate by an oxaloacetate decarboxylase (EC 4.1.1.3; Supplementary Fig. S3). Additionally, *Desulfotomaculum* c7 is able to catabolize glucose via the EMP pathway followed by the acetyl-CoA pathway for complete oxidation of the acetyl-CoA produced (Fig. 3B). Finally, the MAG also includes two versions of a benzylsuccinate synthase, suggesting the ability to degrade alkanes.

As indicated above, organic carbon can be oxidized to carbon dioxide through the acetyl-CoA pathway. A BLASTN comparative analysis to the genome of *Desulfotomaculum acetoxidans* (Spring *et al*. 2009) showed that genes coding for enzymes of this pathway are strongly conserved at the gene cluster level (Table 3). This genome comparison also revealed that genes coding for two membrane-bound hydrogenase complexes (a [NiFe]- and an [FeFe]-hydrogenase) were similar between these organisms (Table 3). The [NiFe]-hydrogenase belongs to group 3 hydrogenases, and is part of a membrane-bound complex that also contains formate hydrogenlyase (Bagramyan and Trchounian 2003). Therefore, it is likely a hydrogenase that transfers reducing equivalents from formate to dihydrogen, which subsequently is oxidized by a membrane-bound [NiFe]-hydrogenase type 1 (Fig. 3B).

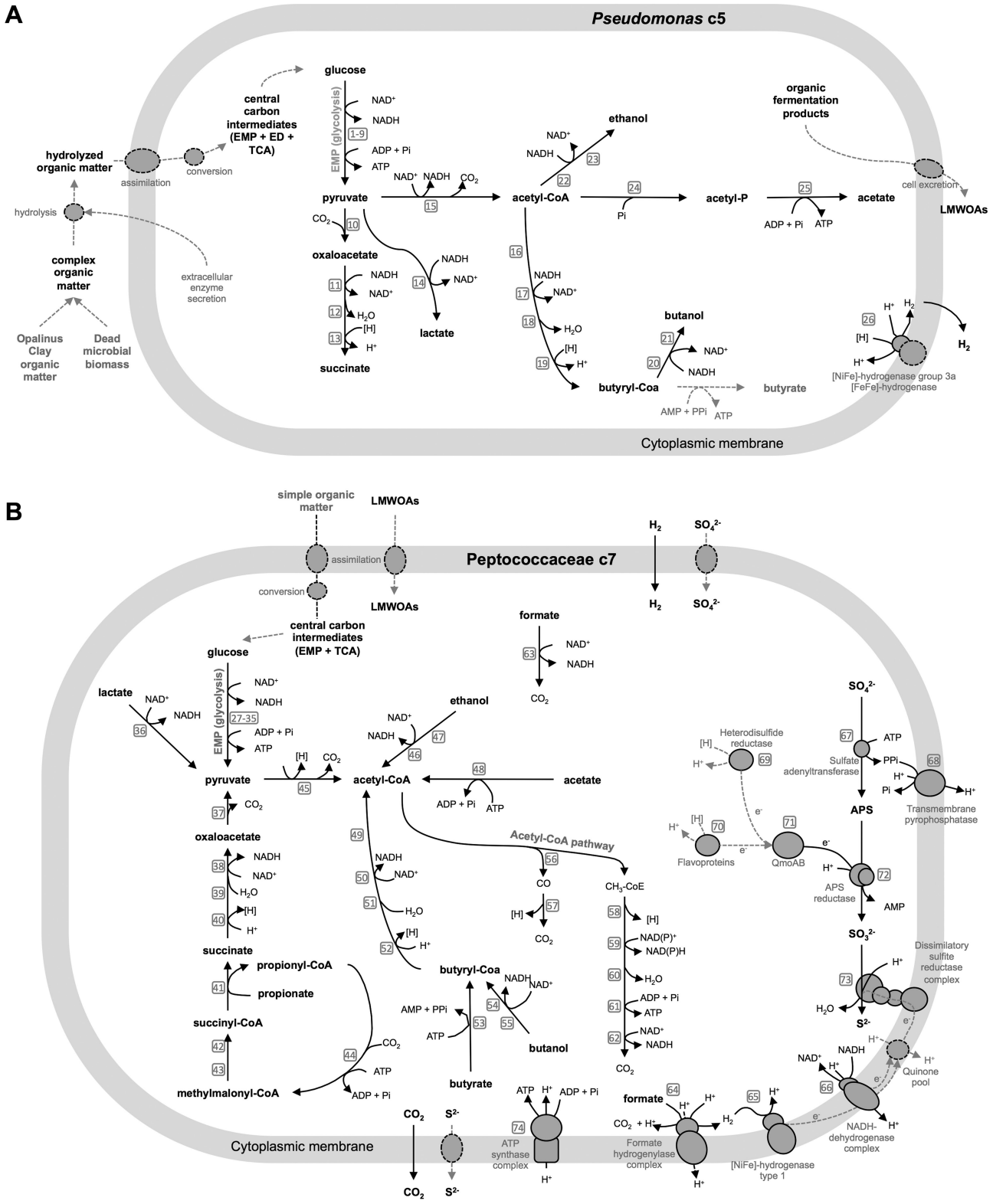


Figure 3. Putative energetic metabolism of (**A**) *Pseudomonas* MAG c5 and (**B**) Peptococcaceae MAG c7. All dashed elements represent biological processes that have not been formally identified in these two MAGs. For more information concerning genes coding for proteins involved in each of these reactions, refer to Supplementary Table S10 (for *Pseudomonas* c5) and Supplementary Table S11 (for Peptococcaceae c7). *Pseudomonas* c5 is thought to hydrolyse complex organic matter extracellularly, and to metabolize the products through fermentation, producing acetate, lactate, ethanol, succinate and dihydrogen gas. Peptococcaceae MAG c7 is thought to oxidize low-molecular mass organic acids (LMWOAs) such as formate, acetate, lactate, propionate and butyrate, and alcohols such as ethanol and butanol all the way to CO_2. Also, dihydrogen oxidation is coupled to sulfate reduction.

The machinery for reducing sulfate was compared to *Desulfotomaculum reducens* (Junier *et al.* 2010) and all its constituents were strongly conserved at the gene cluster level (Table 3). Adenosine triphosphate (ATP) consumed by sulfate adenylyltransferase (EC 2.7.7.4) for sulfate activation can be partially recovered through a transmembrane pyrophosphatase that translocates a proton while hydrolyzing PP_i (Fig. 3B). Unlike *Desulfotomaculum reducens*, Peptococcaceae c7 does not possess a membrane-bound adenylyl-sulfate reductase (EC 1.8.99.2), but a cytoplasmic one. Before reducing adenylyl-sulfate, electrons coming from acetate oxidation are transferred to APS reductase, through flavoproteins and/or heterodisulfide reductase, and then through the QmoAB complex, without transiting into the cytoplasmic membrane (Fig. 3). Similarly to *Desulfotomaculum reducens* and in contrast to Gram-negative SRB, no membrane-bound Qmo subunit could be identified. Electrons coming from dihydrogen, formate and NADH oxidation are transferred in the cytoplasmic membrane, via the quinone pool, generating a proton motive force, and are used by dissimilatory sulfite reductase (EC 1.8.99.3).

Thus, Peptococcaceae c7 is able to use sulfate as a terminal electron acceptor, producing hydrogen sulfide. It can oxidize fermentation products as well as monomers of others organic molecules to carbon dioxide, and is capable of autotrophic growth when carbon dioxide and dihydrogen are available. The genome annotation of Peptococcaceae c7 clearly shows that key processes for energy conservation are highly conserved among Gram-positive SRB, since they were consistent with the ones described for *Desulfotomaculum reducens* (Junier *et al.* 2010) and *Desulfotomaculum acetoxidans* (Spring *et al.* 2009), as shown in Table 3. It is remarkable that although Peptococcaceae c7 is not similar to these two others Peptococcaceae members according to ANI (Supplementary Table S6), genes involved in sulfate reduction, hydrogen oxidation and the acetyl-CoA pathway are almost perfectly conserved at the gene cluster level among these three organisms. This suggests that the genes coding for these metabolic capabilities are orthologous and were present in the common ancestor of these organisms.

The two dominating microorganisms, Peptococcaceae c7 and *Pseudomonas* c5, seem to be indigenous to Opalinus Clay, as indicated by their presence in almost all other samples investigated in Mont Terri URL (Table 1). In addition, they were both detected in the BHT-1 borehole, which was drilled after sterilizing the drilling equipment (Vinsot *et al.* 2014), which indicates that their origin is not a contamination that occurred during the drilling process.

From the annotations of the *Pseudomonas* c5 and Peptococcaceae c7 MAGs, we reconstructed potential interactions between these two species in the BIC-A1 borehole. As shown previously (Courdouan *et al.* 2007), acetate is an organic acid commonly found in Opalinus Clay porewater, at concentrations that can reach 200 μM. Acetate can serve as a carbon and energy source for the Peptococcaceae c7 for sulfate reduction to hydrogen sulfide. After the death of Peptococcaceae c7 cells, complex organic matter from the dead biomass is available to the *Pseudomonas* MAG for energy conservation through fermentative pathways, likely producing acetate and dihydrogen, as these products generate the highest yield of ATP. In turn, these compounds can serve as electron donors and energy sources by the Peptococcaceae MAG c7. However, this putative carbon loop is not sustainable if the only energy source is acetate found in the Opalinus Clay porewater. Over time, after each cycle, because energy conversion efficiency in heterotrophic bacteria is typically 60% (Calow 1977), acetate concentration would continuously decrease.

Another potential energy source that can feed this carbon loop and render it sustainable would be the buried organic matter contained in Opalinus Clay. The question is whether *Pseudomonas* c5 is able to utilize this organic carbon. As mentioned above, organic matter in Opalinus Clay is thermally immature and largely composed of humic and fulvic acids, while other organic compounds such as kerogen type III (humic type) and *n*-alkanes C_{20}–C_{33} were also identified (Jones and Tipping 1998). *Pseudomonas* species are known to utilize this type of organic carbon for respiration (De Haan 1974; Tikhonov *et al.* 2010), but it remains unclear whether these carbon sources can support fermentation. According to the metagenomic analysis, it seems that hydrocarbon cannot be degraded by *Pseudomonas* c5, as benzyl- and alkylsuccinate synthase genes are both missing, but it could be degraded by the Peptococcaceae.

According to reconstructed genomes, pathway annotation and observed environmental parameters, Gram-positive, complete-oxidizer sulfate-reducing bacteria are able to grow using organic compounds present in Opalinus Clay porewater (e.g. acetate) to respire sulfate. Gram-negative bacteria breakdown and ferment dead biomass from the SRB and potentially the Opalinus Clay organic matter to produce low molecular mass organic acids that further sustain the sulfate-reducing bacteria. These microbial metabolisms will likely occur in Opalinus Clay nuclear waste repositories and they should be considered when assessing the long-term safety of such an undertaking. For instance, SRB can produce CO_2 from organic carbon oxidation that is, depending where it is produced in a geological repository, likely to dissolve in the porewater, precipitate while reacting with the cement or less likely, diffuse as a gas phase. Sulfide production by SRB could also increase the rate of anoxic steel corrosion of the drums containing waste. Also, microbial activity could lead to local rock weathering, because active organisms may gather nutrients from the rock itself by secreting protons, organic acids and complexing agents that will complex metal ions (Uroz *et al.* 2009). On the other hand, beneficial impacts of microbial activity include the consumption of H_2 gas that will be produced in the repository through anoxic steel corrosion.

SUPPLEMENTARY DATA

Supplementary data are available at FEMSEC online.

ACKNOWLEDGEMENTS

The work conducted by the U.S. Department of Energy Joint Genome Institute (JGI) is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. JGI is acknowledged for providing resources for the sequencing analysis through the grant CSP 1505. A.F.A. is supported by a grant from the Swedish Research Council VR (grant 2011–5689). Computations were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC). The Swisstopo team of the Mont Terri Rock Laboratory provided the best possible conditions for scientific research in the URL and we are deeply indebted to them for their help. We thank Manon Frutschi for her invaluable help during sampling. Elena Rossel, Karin Vernez Thomas, Sylvain Coudret and Pierre Rossi are acknowledged for their help with laboratory analyses.

FUNDING

This work was supported by the National Cooperative for the Disposal of Radioactive Waste [order no 8620 and order no. 11,761].

Conflict of interest. None declared.

REFERENCES

Alneberg J, Bjarnason BS, de Bruijn I, *et al.* Binning metagenomic contigs by coverage and composition. *Nat Methods* 2014;**11**:1144–6.

Altschul SF, Gish W, Miller W, *et al.* Basic local alignment search tool. *J Mol Biol* 1990;**215**:403–10.

Bagramyan K, Trchounian A. Structural and functional features of formate hydrogen lyase, an enzyme of mixed-acid fermentation from *Escherichia coli*. *Biochem Mosc* 2003;**68**:1159–70.

Boisvert S, Raymond F, Godzaridis É, *et al.* Ray Meta: scalable de novo metagenome assembly and profiling. *Genome Biol* 2012;**13**:R122.

Bossart P, Meier PM, Moeri A, *et al.* Geological and hydraulic characterisation of the excavation disturbed zone in the Opalinus Clay of the Mont Terri rock laboratory. *Eng Geol* 2002;**66**:19–38.

Bossart P, Thury M. *Mont Terri Rock Laboratory. Project, Programme 1996 to 2007 and Results*. Bern: Federal Office of Topography – Swisstopo, 2008.

Callaghan AV, Davidova IA, Savage-Ashlock K, *et al.* Diversity of benzyl- and alkylsuccinate synthase genes in hydrocarbon-impacted environments and enrichment cultures. *Environ Sci Technol* 2010;**44**:7287–94.

Calow P. Conversion efficiencies in heterotrophic organisms. *Biol Rev* 1977;**52**:385–409.

Caporaso JG, Kuczynski J, Stombaugh J, *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 2010;**7**:335–6.

Caporaso JG, Lauber CL, Walters WA, *et al.* Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci USA* 2011;**108**:4516–22.

Caspi R, Altman T, Billington R, *et al.* The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res* 2014;**42**:D459–71.

Cline JD. Spectrophotometric determination of hydrogen sulfide in natural waters. *Limnol Oceanogr* 1969;**14**:454–8.

Colwell FS, Smith RP. Unifying principles of the deep terrestrial and deep marine biospheres. In: Wilcock WSD, Delong EF, Kelley DS, *et al.* (eds). *The Subseafloor Biosphere at Mid-Ocean Ridges*. Washington, DC: American Geophysical Union, 2004, 355–67.

Courdouan A, Christl I, Meylan S, *et al.* Characterization of dissolved organic matter in anoxic rock extracts and *in situ* pore water of the Opalinus Clay. *Appl Geochem* 2007;**22**:2926–39.

De Haan H. Effect of a fulvic acid fraction on the growth of a *Pseudomonas* from Tjeukemeer (The Netherlands). *Freshw Biol* 1974;**4**:301–10.

DeSantis TZ, Hugenholtz P, Larsen N, *et al.* Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 2006;**72**:5069–72.

Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 2010;**26**:2460–1.

Edgar RC, Haas BJ, Clemente JC, *et al.* UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 2011;**27**:2194–200.

Edwards KJ, Becker K, Colwell F. The deep, dark energy biosphere: intraterrestrial life on earth. *Annu Rev Earth Planet Sci* 2012;**40**:551–68.

Eschbach M, Schreiber K, Trunk K, *et al.* Long-term anaerobic survival of the opportunistic pathogen *Pseudomonas aeruginosa* via pyruvate fermentation. *J Bacteriol* 2004;**186**:4596–604.

Gobat J-M, Aragno M, Matthey W. *The Living Soil: Fundamentals of Soil Science and Soil Biology*. Enfield, NH: Science Publishers, 2004.

Goris J, Konstantinidis KT, Klappenbach JA, *et al.* DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* 2007;**57**:81–91.

IAEA. *Scientific and Technical Basis for the Geological Disposal of Radioactive Wastes*. New York: International Atomic Energy Agency, 2003.

Jones DM, Tipping E. *Organic Geochemical Analysis of Opalinus Clay Samples from Mont Terri*. St-Ursanne, Switzerand: Mont Terri Project, 1998.

Junier P, Junier T, Podell S, *et al.* The genome of the Gram-positive metal- and sulfate-reducing bacterium *Desulfotomaculum reducens* strain MI-1. *Environ Microbiol* 2010;**12**:2738–54.

Kanehisa M, Goto S, Sato Y, *et al.* Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res* 2014;**42**:D199–205.

Kim BH, Gadd GM. *Bacterial Physiology and Metabolism*, 1st edn. New York: Cambridge University Press, 2008.

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;**9**:357–9.

Li M, Copeland A, Han J. *DUK – A Fast and Efficient Kmer Based Sequence Matching Tool*. Lawrence Berkeley National Laboratory 2011.

Magoč T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 2011;**27**:2957–63.

Markowitz VM, Chen I-MA, Palaniappan K, *et al.* IMG: the integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res* 2012;**40**:D115–22.

Mazurek M. Mineralogy of the Opalinus clay. In: Thury M, Bossart P (eds). *Mont Terri Rock Laboratory: Results of the Hydrogeological, Geochemical and Geotechnical Experiments Performed in 1996 and 1997*, vol 23. Bern: Federal Office of Topography Swisstopo, 1999, 15–8.

Moore ERB, Tindall, Dos Santos VAPM BJ, *et al.* Nonmedical: Pseudomonas. In: Dworkin M, Falkow S, Rosenberg E, *et al.* (eds). *The Prokaryotes*, vol 6, 3rd ed. New York: Springer, 2006, 646–703.

Noble PA, Citek RW, Ogunseitan OA. Tetranucleotide frequencies in microbial genomes. *Electrophoresis* 1998;**19**:528–35.

Parks DH, Imelfort M, Skennerton CT, *et al.* CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 2015:**25**:1043–55.

Pearson FJ, Acros D, Boisson J-Y, *et al. Mont Terri Project: Geochemistry of Water in the Opalinus Clay Formation at the Mont Terri Rock Laboratory*. Bern: Federal Office for Water and Geology, 2003.

Pedersen K. Exploration of deep intraterrestrial microbial life: current perspectives. *FEMS Microbiol Lett* 2000;**185**:9–16.

Poulain S, Sergeant C, Simonoff M, *et al.* Microbial investigations in Opalinus clay, an argillaceous formation under evaluation as a potential host rock for a radioactive waste repository. *Geomicrobiol J* 2008;**25**:240–9.

Rodríguez H, Fraga R. Phosphate solubilizing bacteria and their role in plant growth promotion. *Biotechnol Adv* 1999;**17**:319–39.

Schreiber K, Boes N, Eschbach M, *et al*. Anaerobic survival of *Pseudomonas aeruginosa* by pyruvate fermentation requires an usp-type stress protein. *J Bacteriol* 2006;**188**:659–68.

Spring S, Lapidus A, Schroder M, *et al*. Complete genome sequence of *Desulfotomaculum acetoxidans* type strain (5575T). *Stand Genomic Sci* 2009;**1**:242–53.

Stams AJM, Kremer DR, Nicolay K, *et al*. Pathway of propionate formation in *Desulfobulbus propionicus*. *Arch Microbiol* 1984;**139**:167–73.

Stark M, Berger SA, Stamatakis A, *et al*. MLTreeMap - accurate maximum likelihood placement of environmental DNA sequences into taxonomic and functional reference phylogenies. *BMC Genomics* 2010;**11**:461.

Stookey LL. Ferrozine – a new spectrophotometric reagent for iron. *Anal Chem* 1970;**42**:779–81.

Stroes-Gascoyne S, Schippers A, Schwyn B, *et al*. Microbial community analysis of Opalinus Clay drill core samples from the Mont Terri underground research laboratory. *Geomicrobiol J* 2007;**24**:1–17.

Thury M, Bossart P. The Mont Terri rock laboratory, a new international research project in a Mesozoic shale formation, in Switzerland. *Eng Geol* 1999;**52**:347–59.

Tikhonov VV, Yakushev AV, Zavgorodnyaya YA, *et al*. Effects of humic acids on the growth of bacteria. *Eurasian Soil Sci* 2010;**43**:305–13.

Uroz S, Calvaruso C, Turpault M-P, *et al*. Mineral weathering by bacteria: ecology, actors and mechanisms. *Trends Microbiol* 2009;**17**:378–87.

Vinsot A, Appelo CAJ, Lundy M, *et al*. *In situ* diffusion test of hydrogen gas in the Opalinus Clay. *Geol Soc Lond Spec Publ* 2014;**400**:563–78.

Wang Q, Garrity GM, Tiedje JM, *et al*. Naïve bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* 2007;**73**:5261–7.

Yu NY, Wagner JR, Laird MR, *et al*. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* 2010;**26**:1608–15.