FEMS Microbiology Ecology, 94, 2018, fiy152

doi: 10.1093/femsec/fiy152
Advance Access Publication Date: 7 August 2018
Research Article

RESEARCH ARTICLE

A metagenomic window into the 2-km-deep terrestrial subsurface aquifer revealed multiple pathways of organic matter decomposition

Vitaly V. Kadnikov[1], Andrey V. Mardanov[1], Alexey V. Beletsky[1], David Banks[3], Nikolay V. Pimenov[4], Yulia A. Frank[2], Olga V. Karnachuk[2] and Nikolai V. Ravin[1,*]

[1]Institute of Bioengineering, Research Center of Biotechnology of the Russian Academy of Sciences, Leninsky prosp. 33-2, Moscow, 119071, Russia, [2]Laboratory of Biochemistry and Molecular Biology, Tomsk State University, Lenina prosp. 35, Tomsk, 634050, Russia, [3]School of Engineering, Systems Power & Energy, Glasgow University, Glasgow G12 8QQ, and Holymoor Consultancy Ltd., 360 Ashgate Road, Chesterfield, Derbyshire S40 4BW, UK and [4]Winogradsky Institute of Microbiology, Research Center of Biotechnology of the Russian Academy of Sciences, Leninsky prosp 33-2, Moscow, 119071, Russia

*Corresponding author: Institute of Bioengineering, Research Center of Biotechnology of the Russian Academy of Sciences, Leninsky prosp. 33-2, Moscow, 119071, Russia. Tel: +74997833264; Fax: +74991353051; E-mail: nravin@biengi.ac.ru

One sentence summary: metagenome-assembled genomes of microorganisms from the deep subsurface thermal aquifer revealed functional diversity of the community and novel uncultured bacterial lineages

Editor: Tillmann Lueders

ABSTRACT

We have sequenced metagenome of the microbial community of a deep subsurface thermal aquifer in the Tomsk Region of the Western Siberia, Russia. Our goal was the recovery of near-complete genomes of the community members to enable accurate reconstruction of metabolism and ecological roles of the microbial majority, including previously unstudied lineages. The water, obtained via a 2.6 km deep borehole 1-R, was anoxic, with a slightly alkaline pH, and a temperature around 45°C. Microbial community, as revealed by 16S rRNA gene profiling over 2 years, mostly consisted of sulfate-reducing *Firmicutes* and *Deltaproteobacteria*, and uncultured lineages of the phyla *Chlorofexi*, *Ignavibacteriae* and *Aminicenantes* (OP8). 25 composite genomes with more than 90% completeness were recovered from metagenome and used for metabolic reconstruction. Members of uncultured lineages of *Chlorofexi* and *Ignavibacteriae* are likely involved in degradation of carbohydrates by fermentation, and are also capable of aerobic and anaerobic respiration. The *Chlorofexi* bacterium has the Wood-Ljungdahl pathway of CO_2 fixation. The recently identified candidate phylum *Riflebacteria* accounted for 5%–10% of microbial community. Metabolic reconstruction of a member of *Riflebacteria* predicted that it is an anaerobe capable to grow on carbohydrates by fermentation or dissimilatory Fe(III) reduction.

Keywords: deep subsurface; metagenome; uncultured bacteria; sulfate-reducer; candidate phylum

Received: 8 May 2018; **Accepted:** 6 August 2018

INTRODUCTION

Microbial communities in deep underground ecosystems of the Earth have different composition, which depends on geochemical conditions (Gihring *et al.* 2006). These ecosystems are one of the largest habitats for microorganisms, and the total biomass in the deep subsurface constitutes a significant proportion of the entire biosphere (Whitman, Coleman and Wiebe 1998). Deep subsurface environments could remain isolated from the surface biosphere for thousands to millions of years and are not dependent on the intake of organic matter from the surface (Edwards, Becker and Colwell 2012). Typically, they are characterized by a low content of available substrates, which determines the low growth rate of microorganisms in the underground biosphere *in situ* (Jørgensen 2012; Lever *et al.* 2015). Using molecular methods for the analysis of 16S rRNA genes, various noncultured bacterial and archaeal lineages were found in such communities. In some cases they constitute the majority of communities and are specific for underground habitats (Takai *et al.* 2001; Chivian *et al.* 2008; Sahl *et al.* 2008).

Depending on available energy sources, lithoautotrophic or organotrophic microorganisms may predominate in the microbial communities of the subsurface biosphere. The main energy source for autotrophs is molecular hydrogen, which can be of abiotic origin or can be formed as a product of fermentative metabolism (Nealson, Inagaki and Takai 2005; Hallbeck and Pedersen 2008). It can be used by sulfate-reducing microorganisms or methanogenic archaea. These two groups of microorganisms are often found in the deep subsurface ecosystems (Moser *et al.* 2005). Environments associated with organic-rich sedimentary rocks such as formation waters of oil reservoirs can contain large amounts of organic substances remaining since their formation, which determines the possibility of the development of various organotrophic microorganisms. Microbial communities of the terrestrial deep subsurface ecosystems, both chemolithoautotrophic and organotrophic types, have been investigated in a number of studies (Anantharaman *et al.* 2016; Hubalek *et al.* 2016; Wu *et al.* 2016; Probst *et al.* 2017; Magnabosco *et al.* 2016; Purkamo *et al.* 2013; Miettinen *et al.* 2015).

The most accessible source of deep underground thermal waters is oil exploration boreholes, from which these waters can flow under natural pressure. The region of Western Siberia, Russia, is known as one of the world's largest oil reserves formed from marine sediments of the Mesozoic period. In addition to oil reservoirs, a large number of underground water basins, located at depths of 1–3 km, were found. Two such subsurface aquifers have been characterized previously by molecular and cultivation approaches (Frank *et al.* 2016a; Kadnikov *et al.* 2017a,b). The first one is the 3P borehole in the Parabel district. Analysis of the thermal water collected at this borehole by 16S rRNA sequencing revealed rather simple community consisting of sulfate-reducing *Firmicutes* (the genera *Desulfovirgula*, *Thermoacetogenium* and *Desulfotomaculum*) and hydrogenotrophic methanogenic archaea of the genus *Methanothermobacter* (Frank *et al.* 2016a). The members of phyla *Ignavibacteria* and *Chloroflexi* were also present but in minor amounts, and several other phylotypes appeared only accidentally in course of 5-year monitoring. The microbial community at another site, borehole 5P in the Chazhemto district, consisted of different bacterial lineages of the phyla *Firmicutes, Ignavibacteria, Chloroflexi, Bacteroidetes* and *Proteobacteria*, phylogenetically distant from cultured species, and methanogenic archaea of the genera *Methanothermobacter* and *Methanosaeta* (Kadnikov *et al.* 2017a,b). Interestingly, known lineages of sulfate reducers were not found at the 5P borehole. These studies revealed quite different microbial communities with a number of uncultured taxa, however, none of them were analyzed by metagenomics which do not only determine the composition of communities, but also the functional role of the identified microorganisms.

In this study, we sequenced a metagenome of the microbial community of the deep subsurface thermal aquifer accessed through the borehole 1-R located in the Tomsk Region of the Russian Federation. Our goal was the recovery of near-complete genomes of the community members to enable accurate reconstruction of metabolism and ecological roles of the microbial majority, including previously unstudied lineages.

MATERIALS AND METHODS

Site description

The oil-exploration borehole 1-R is located in the town Byelii Yar in Tomsk region, Russia (58.4496N, 85.0279E, 84 m asl). Location map is available from Frank *et al.* 2016b. The borehole was drilled as an oil-exploration well in 1961–1962 to 2563 m deep but no oil was found. It is located in the Western Siberian artesian basin, a synclinal structure containing Mezozoic, Tertiary and Quaternary sediments, overlaying Paleozoic basement (Banks *et al.* 2014). The borehole penetrated about 25 m of Quaternary sediments followed by approximately 150 m of Paleogene sediments and Cretaceous sediments down to 2186 m depth. Upon passing a Jurassic sedimentary sequence, at 2505 m the borehole entered Paleozoic basement, comprising sedimentary rocks up to 2534 m and then basalts (Banks *et al.* 2014). Upon completion, the borehole was tested by a bottom-up casing perforation method. In this method successive sections of casing were perforated with testing of pressure and flow. After testing the section was sealed by cement and a new upper section was tested likewise. The final tested section was from 1997 to 2005 m in the early Cretaceous sedimentary rocks, and a cement seal was placed up to 2172 m depth (Banks *et al.* 2014). Therefore at present the water flowing out of the borehole is expected to be originated from an aquifer system in this depth interval.

At present, the borehole is used by local populations as a source of thermal water through a closed wellhead set of connectors. Thus the wellhead pressure decreases and water flow increases as more users access water from the wellhead (Banks *et al.* 2014).

Sampling, field measurements, chemical analyses and DNA isolation

Samples were taken from a sampling line at the wellhead on August 23–24, 2013 and on August 04–05, 2014. Water temperature, pH and Eh were determined on site using pH-meter HI 8314 (Hanna Instruments, Germany). For chemical analysis the water samples were filtered through 0.22 μm sterile filter (Merck Millipore, Germany) using a hand-held syringe into new clean polyethylene flasks. The samples were analyzed by ICP-MS and ion chromatography at the British Geological Survey, Keyworth, UK.

For determination of dissolved hydrogen sulfide the water was fixed with 2.4% (w/v) zinc acetate solution in proportion of 1:5. H_2S was measured colorimetrically with the methylene blue method (Cline, 1969) in triplicate using a Smart Spec Plus spectrophotometer (Bio-Rad Laboratories, USA).

Gas content in the year 2013 sample was determined using a Kristall-5000.1 gas chromatograph (Russia) equipped with a catarometer detector. $\delta^{13}C$ of methane was measured on a

TRACE GC gas chromatograph (Thermo Fisher Scientific, USA) coupled to a Delta Plus isotope mass spectrometer (Thermo Fisher Scientific, USA). All isotopic analyses were carried out in triplicate.

Cells from 50 L of borehole water were collected on 0.22 μm cellulose nitrate membranes (Sartorius, Germany) using a Sartorius filtration unit. The filters were frozen in liquid nitrogen and then ground and melted with TE buffer in a water bath at 37°C. The total DNA was extracted by the CTAB/NaCl method (Wilson 1987).

16S rRNA gene sequencing and analysis

PCR amplification of the 16S ribosomal RNA gene fragments containing the V3–V6 variable regions was carried out using the universal primers PRK341F (5'-CCT ACG GGR SGC AGC AG-3') and PRK806R (5'-GGA CTA CYV GGG TAT CTA AT-3'). This primer pair targets most of Archaea and Bacteria except only a few lineages (Frank *et al.* 2016a). The PCR fragments were sequenced with a Roche Genome Sequencer (GS FLX), using the Titanium XL+ protocol according to the manufacturer's instructions. A total of 65 459 and 32 233 sequence reads were generated for the year 2013 and 2014 samples, respectively. The reads shorter than 380 nt, those with mismatches to primer sequences or containing ambiguous nucleotides were excluded from analysis using Mothur v. 1.35.1 (Schloss *et al.* 2009). Chimeric sequences were detected and removed by UCHIME algorithm implemented in Mothur (Edgar *et al.* 2011). Then all remaining singleton sequences were removed prior to Operational Taxonomic Units (OTUs) clustering. The final datasets consisted of 31 455 and 20 470 reads for the year 2013 and 2014 samples, respectively.

Clustering and selection of representative sequences for OTU was done using RDP Tools v. 2.0.2 (Cole *et al.* 2009). For taxonomic assignment of OTU the representative sequences were searched against the NCBI NR database using BLASTN. OTU were assigned to a certain genus if the representative sequence was more than 95% identical to that of the 16S rRNA gene of a cultivated microorganism. Taxonomic classification of other OTU was performed using the RDP Classifier v. 2.11 (Cole *et al.* 2009) and searches against the SILVA database v. 132 (Quast *et al.* 2013).

Metagenome sequencing and assembly, contig binning and analysis of the composite genomes

Metagenomic DNA of the year 2014 sample was sequenced using the Illumina HiSeq2500 platform (250 nt single-end reads) according to the manufacturer's instructions (Illumina Inc., USA). Sequencing resulted in a total of 86.5 million high quality reads (a total of 18.5 Gbp) after primer and quality trimming with Cutadapt v. 1.8.3 (Martin 2011) and Sickle v. 1.33 (https://github.com/najoshi/sickle), respectively. Cutadapt was used with default settings, and Q33 score was used for Sickle. The assembly of the contigs was carried out using SPAdes Genome Assembler v. 3.7.1 (Bankevich *et al.* 2012), specifying -meta parameter indicating metagenome assembly.

Contigs longer than 2000 bp were binned into clusters representing the composite genomes of microbial community members using the program CONCOCT v. 0.4.1 (Alneberg *et al.* 2014). The completeness and contamination of the recovered genomes were estimated using CheckM v. 1.05 (Parks *et al.* 2015) with lineage-specific marker genes. Binning results were visualized using Anvi'o platform v. 4 (Eren *et al.* 2015) for manual curation of bins BY5, BY21 and BY39.

The 16S rRNA genes were found in contigs by CheckM. The search for continuations of contigs representing fragments of 16S rRNA genes using Bandage v. 0.8.0 program (Wick *et al.* 2015) allowed to link 16S rRNA genes to most of genome bins where 16S rRNA carrying contigs were missing.

For each composite genome with an estimated completeness >90%, gene search and annotation were performed using the RAST server 2.0 (Brettin *et al.* 2015), followed by manual correction by searching the National Center for Biotechnology Information (NCBI) databases.

Search for genomes related to bin BY5

In order to find genomes closely related to bin BY5, we identified homologs of 88 conserved marker genes in the NCBI NR database using BLASTP search. Genome assemblies corresponding to top 20 hits for each of the marker genes (a total of 1120) were selected. Then we performed full genome comparisons between BY5 and the selected 1120 genome assemblies using ANIcalculator v. 1.0 (Varghese *et al.* 2015). The best hits were Bacterium CG2_30_54_10 (MNYA00000000.1), Candidatus Riflebacteria bacterium HGW-Riflebacteria-1 (PGYB00000000.1), Candidatus Riflebacteria bacterium HGW-Riflebacteria-2 (PGYA00000000.1) and Candidatus Riflebacteria bacterium GWC2_50_8 (MFYY00000000.1), followed by a large gap to other organisms. In addition we used the aai.rb script from Enveomics Collection (Rodriguez-R and Konstantinidis, 2016) to calculate average amino acid identity (AAI) between BY5 and selected genomes. The same four bacteria were found to be much closer to the BY5 than all the others.

Phylogenetic analysis

The 16S rRNA sequences were aligned using MUSCLE included in MEGA 6.0 (Tamura *et al.* 2013). The maximum likelihood phylogenetic tree was computed by MEGA 6.0, using Tamura-Nei substitution model and uniform rates among sites. Bootstrap tests were performed with 100 resamplings.

We used CheckM to find single copy marker genes in the assembled genomes and to construct multiple alignment of concatenated single copy genes sequences, comprising ones from our genomes, four genomes phylogenetically related to bin BY5 and all species presented in CheckM database. We calculated distances between our genomic bins and CheckM genomes using MAFFT v. 7.221 to find and select several species close to our bins. Selected part of the CheckM multiple alignment was used for the tree construction in PhyML v. 3.2.20160531 (Guindon *et al.* 2010) using default parameters. The support values for the internal nodes were estimated by approximate Bayesian tests in PhyML.

Nucleotide sequence accession number

Pyrosequencing read data obtained for the 16S rRNA gene fragments were deposited in Sequence Read Archive (SRA) under the accession numbers SRR7102748 (year 2013 sample) and SRR7102747 (year 2014 sample). Metagenomic reads were deposited in SRA under the accession number SRR7102746.

The annotated genome sequences of Chloroflexi bacterium BY21, Ignavibacteriae bacterium BY39 and *Candidatus* Ozemobacter sibiricus BY5 have been deposited in the GenBank database under accession numbers QOQU00000000, QOQV00000000 and QOQW00000000, respectively.

Table 1. Physical and chemical characteristics of the water at borehole 1-R.

		Date and time of sampling				
		23/08/2013		24/08/2013		04/08/2014
Parameter	Unit	11:00	19:45	11:00	19:00	10:00
Wellhead pressure	kg cm^{-2}	3.3–3.5	3.1–3.2	3.7	1.7–2.0	2.5
pH		7.92	8.06	8.25	8.21	8.51
Eh	mV	-279	-341	-283	-336	nd
T	°C	40.2	42.9	41.6	44.8	42.7
t-alkalinity	meq l^{-1}	4.7	4.8	4.7	4.7	nd
Cl^-	mg l^{-1}	844	832	830	843	nd
SO_4^{2-}	mg l^{-1}	<5	<5	<5	<5	nd
NO_3^-	mg l^{-1}	<3	<3	<3	<3	nd
Br^-	mg l^{-1}	3.10	3.07	3.25	3.04	nd
F^-	mg l^{-1}	11.7	11.4	11.6	11.5	nd
Na	mg l^{-1}	661	633	639	656	nd
Mg	mg l^{-1}	0.14	0.13	0.13	0.14	nd
Ca	mg l^{-1}	9.7	9.0	8.9	9.4	nd
K	mg l^{-1}	3.02	2.94	2.93	3.14	nd
Mn	µg l^{-1}	9.4	9.0	8.6	8.3	nd
Fe (dissolved)	µg l^{-1}	119	128	105	97	nd
Sr	µg l^{-1}	1011	961	946	1003	nd
Ba	µg l^{-1}	224	229	220	227	nd
Si	µg l^{-1}	12 726	13 074	13 007	12 853	nd
S	mg l^{-1}	2	2	2	2	nd
B	µg l^{-1}	2337	2077	2284	2274	nd
As	µg l^{-1}	0.54	0.55	0.42	0.48	nd
dissolved H_2S	mg l^{-1}	nd	nd	nd	0.64	0.88

nd, not determined.

RESULTS AND DISCUSSION

Characteristics of the water chemistry

The physical and chemical characteristics of the water collected at the borehole 1-R wellhead measured in August 2013 and August 2014 are shown in Table 1. Depending on the rate of water flow, the temperature ranged from 40.2°C to 44.8°C and increased as flow increased. Considering a typical thermal gradient of 20°C–30°C per km, and taking into account that the water may cool when passing through the borehole it seems likely that the water originates from about 2 km deep horizon, as expected. The water has a slightly alkaline pH (from 7.9 to 8.5) and was highly reduced (Eh from –341 to –279 mV). Ionic content of the water was dominated by sodium as a cation, and chloride as an anion, with a moderate bicarbonate alkalinity (4.7 meq l^{-1}, Banks *et al.* 2014).

Calcium and magnesium are almost absent from the water, likely due to precipitation of calcite and dolomite in an alkaline water rich in bicarbonate (Banks *et al.* 2014). Magnesium/strontium mass ratio is very low (0.14), suggesting that strontium accumulated in water during prolonged residence time, while the magnesium was precipitated presumably by dolomitization. Another evidence of a prolonged residence time is a high concentration of fluoride that could results from dissolution of fluoride-containing minerals.

Bromide concentration is about 3 mg/L, with a Cl/Br mass ratio of 270. This value is very close to seawater (290) suggesting that the water salinity is likely to be derived from marine salts of relict ocean. However, the total mineralization of the water can be estimated by summing chloride, sodium and bicarbonate as 1.8 g L^{-1} representing only 5% of marine salinity. Therefore most of the water is derived from meteoric recharge. This proposal is consistent with the stable isotopic (^{18}O and ^{2}H) analysis indicating that the isotopic signatures of the groundwater were similar but not identical to that of meteoric water (Banks *et al.* 2014).

Rather unexpectedly, concentrations of sulfate and nitrate were very low and below detection limit (<5 mg L^{-1} and <3 mg L^{-1}). However, the groundwater contained dissolved hydrogen sulfide at concentrations of 0.64±0.35 mg L^{-1} in 2013, and 0.88±0.22 mg L^{-1} in 2014.

On exposure to atmospheric pressure, the water exsolves bubbles of gas, which were collected and analyzed for composition. Methane was the main constituent (60.2%) followed by nitrogen (32%), oxygen (6.9%), helium (0.7%) and CO (0.2%). The isotopic composition of methane ($\delta^{13}C_{av.} = -62.5‰$) indicates its mostly microbial origin with admixture of thermogenic methane. The presence of methane is typical for such highly reducing anoxic environments (for example, Frank *et al.* 2016a), but the measured content of oxygen is puzzling. It may represent atmospheric contamination during sample transport or originate from degassing of meteoric recharge upon contact with deep hot aquifer water. Rather high oxygen content in exsolved gas bubbles (up to 4.5%) has been reported for water collected at 3P borehole (Parabel district) during several years (Frank *et al.* 2016a) that supports hypothesis of its indigenous origin.

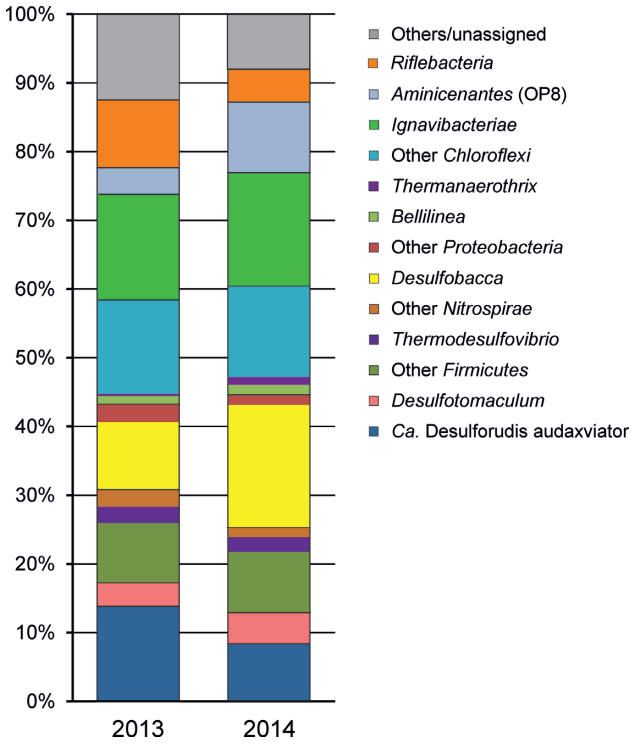

Figure 1. The relative abundance of taxonomic groups of microorganisms in the water samples collected in 2013 and 2014.

Microbial community structures revealed by pyrosequencing of 16S rRNA gene fragments

A total of 31 455 and 20 470 16S rRNA gene sequences were obtained after applying quality filters for the year 2013 and 2014 samples, respectively. The results of the taxonomic classification of the obtained OTUs are shown in Fig. 1. In the year 2014 the microbial community was dominated by five major bacterial lineages each comprising more than 10% of the 16S rRNA gene sequences. Members of archaea accounted for less than 0.1% of microorganisms and belonged to the genus *Methanothermobacter*.

The most abundant phylum, *Firmicutes*, accounted for 21.8% of the community and was represented by *Candidatus* Desulforudis audaxviator (8.4% of reads), the genera *Desulfotomaculum* (4.5% of 16S rRNA reads) and *Pelotomaculum* (0.5%) and several uncultured lineages phylogenetically distant from known species. Both *Ca*. Desulforudis audaxviator and *Desulfotomaculum* spp. are sulfate-reducing bacteria able to use molecular hydrogen as an energy source (Chivian *et al.* 2008; Aüllo *et al.* 2013). *Desulfotomaculum* sp. is typical for deep subsurface environments worldwide while distribution of *Ca*. Desulforudis audaxviator is more limited (Chivian *et al.* 2008; Jungbluth *et al.* 2017). Members of *Pelotomaculum* are usually known to be fermenters, undergoing syntrophy with anaerobic respiring microorganisms such as sulfate reducers, but sulfate reducing organisms were also found in this genus (Dong *et al.* 2017).

The second most frequent lineage was *Deltaproteobacteria* (19.3%), represented by the genus *Desulfobacca* (17.8%) and an uncultured lineage of the family *Desulfobacteraceae*. The cultured member of this genus, *D. acetoxidans*, is an acetate-degrading sulfate reducer, able to grow with acetate as sole carbon and energy source (Oude Elferink *et al.* 1999).

Three other most abundant bacterial divisions, *Chloroflexi*, *Ignavibacteriae* and candidate phylum *Aminicenantes* (OP8), accounted for 15.8%, 16.4% and 10.3% of 16S rRNA reads, respectively. All of them were represented by lineages only distantly related to cultured species, with the only exception being *Thermanaerothrix daxensis* (1.1%)—thermophilic chemoorganotrophic bacterium of the phylum *Chloroflexi* isolated from a deep hot aquifer in the Aquitaine Basin (Grégoire *et al.* 2011). The phylum *Chloroflexi* contains metabolically diverse microorganisms, including aerobic and anaerobic organotrophs, phototrophs, nitrate and organohalide respirers (Bryant *et al.* 2012; Krzmarzick *et al.* 2012).

The phylum *Ignavibacteriae* is currently represented by only two cultured species—*Ignavibacterium album* (Iino *et al.* 2010) and *Melioribacter roseus* (Podosokorskaya *et al.* 2013). The later bacterium was isolated from microbial mat developing on the wooden surface of a chute under the flow of hot water coming out of an oil exploration well 3P (Parabel district). Both species are moderate thermophiles and facultative anaerobes growing on mono-, di- or polysaccharides by aerobic respiration, fermentation or by reducing diverse electron acceptors (Liu *et al.* 2012; Kadnikov *et al.* 2013).

The candidate phylum *Aminicenantes*, still lacking cultivated species, exhibited the highest relative abundance in hydrocarbon-impacted environments, followed by marine hydrothermal habitats, and aquatic, non-marine habitats such as terrestrial springs and groundwater (Farag *et al.* 2014). Metabolic reconstruction of near-complete genomes obtained from metagenomes indicated a capability of utilizing a variety of complex sugar polymers in fermentative metabolism, while known pathways of aerobic or anaerobic respiration were missing (Robbins *et al.* 2016).

About 3.6% of 16S rRNA reads were assigned to the phylum *Nitrospirae*, of them 2.2% represented the genus *Thermodesulfovibrio*. A member of this genus, *Thermodesulfovibrio* sp. N1, was previously isolated from water collected at this borehole (Frank *et al.* 2016b). This strain is able to reduce sulfate, thiosulfate or sulfite with a limited range of electron donors, such as formate, pyruvate and lactate. Other members of *Nitrospirae* found in the microbial community are phylogenetically distant from cultured species. Among other recognized phyla, members of *Actinobacteria* and *Elusimicrobia* were found in minor amounts (<1% of reads), all of them phylogenetically distant from known species. About 4.9% of microorganisms detected by the 16S rRNA genes sequences belonged to a single non-cultivated lineage related to the candidate phylum *Riflebacteria*, recently identified in a deep subsurface environments (Anantharaman *et al.*, 2016).

Analysis of the composition of microbial community of the water sample, collected from 1-R borehole 1 year earlier (2013) revealed the same groups of bacteria and likewise absence of archaea (Fig. 1). The same major groups were found but some variations in their relative abundances were observed. Among sulfate reducers, the share of *Ca*. Desulforudis audaxviator in 2013 was higher than in 2014 (13.8% vs 8.4%), while the opposite was observed for *Desulfobacca* (9.8% vs 17.8%). Considerable variations were also observed for *Aminicenantes* (3.8% vs 10.3%) and *Riflebacteria* (9.9% vs 4.9%). For some groups of bacteria, namely *Chloroflexi*, *Ignavibacteriae and Thermodesulfovibrio*, only limited variations were observed and their shares in the community remained almost unchanged.

Metagenome assembly and binning results

In order to assemble the composite genomes of the most numerous members of the microbial community, we sequenced the metagenome of a microbial community from a water sample collected in 2014. A total of 18.5 Gbp metagenomic sequences were assembled into contigs, which were distributed among 40 major genome bins. Twenty five genomic bins represented composite genomes, having a completeness of at least 90% (Table 2). Altogether these bins accumulated more than 91% of metagenome sequences.

Analysis of the presence of a set of single-copy marker genes showed that 16 of 25 bins meet the recently proposed criteria (Bowers *et al.* 2017) for the high quality metagenome-assembled genomes (>90% completeness with <5% contamination) and six bins were slightly below this cutoff with an estimated contamination of 5%–6%. Two genomes (BY45 and BY32) apparently represented mixtures of two bacteria each since near 100% contamination was predicted by CheckM. 23 of 25 genomes contained or were linked to near-complete (>1400 bp) sequences of the 16S ribosomal RNA gene.

Phylogenetic identification of the genomic bins based on full-size 16S rRNA genes revealed the same lineages detected by 16S rRNA profiling. The relative abundances of phylotypes, defined as a fraction of a given bin in the whole metagenome, was in good agreement with 16S rRNA based data, taking into account different numbers of 16S rRNA gene copies in the genomes (Table 2). To investigate the phylogenetic position of genomic bins in more detail, we constructed phylogenetic trees based on concatenated conserved marker genes (Fig. 2).

All groups of sulfate-reducing bacteria identified by 16S rRNA profiling were found among 25 major genomic bins, namely *Ca.* Desulforudis audaxviator (BY57, 100% 16S rRNA identity to *Ca.* Desulforudis audaxviator MP104C), *Desulfotomaculum* (BY29, 96% 16S rRNA identity to *D. kuznetsovii*), *Desulfobacca* (BY47, 96% 16S rRNA identity to *D. acetoxidans*), and *Thermodesulfovibrio* (BY55, 97% 16S rRNA identity to *Thermodesulfovibrio* sp. N1). Another lineage of the family *Desulfobacteraceae* distantly related to cultured species (BY8, 92% 16S rRNA identity to *Desulfococcus biacutus*) is likely a sulfate reducing bacterium as well since this genome contained a full set of genes encoding dissimilatory sulfate reduction pathway.

Besides bins BY57 and BY29, five other genomic bins were assigned to the phylum *Firmicutes*—BY15 representing *Pelotomaculum* (95% 16S rRNA identity to *P. thermopropionicum*) and four bins distantly related to cultured species. Six genomic bins represented the phylum *Chloroflexi*. Only BY32 is closely related to cultured species, with 99% 16S rRNA sequence identity to *T. daxensis*. Other genomic bins were phylogenetically distant from known isolates but all fell into the order *Anaerolineales* (Figs 2 and 3). On the contrary, only a single genomic bin (BY39) was assigned to the phylum *Ignavibacteriae*, one of the most abundant groups in the microbial community. Single genomic bins were also identified for *Actinobacteria* (BY27), *Spirochaetes* (BY33), *Elusimicrobia* (BY11) and the candidate phyla *Aminicenantes* (BY38) and BRC1 (BY40). BY19 on the phylogenetic tree formed a deep lineage adjacent to *Firmicutes* (Fig. 2) and could belong to this phylum. BY5 is phylogenetically distant form any cultured bacterial lineages and belongs to the recently proposed candidate phylum *Riflebacteria* (Anantharaman *et al.* 2016).

To characterize the metabolic capabilities of microorganisms and their possible ecological roles in the community, we analyzed the genomes of dominant microbial lineages phylogenetically remote from known species—BY21 (*Chloroflexi*), BY39 (*Ignavibacteriae*) and BY5 (*Riflebacteria*).

Chloroflexi bacterium BY21

Genome of BY21 bacterium was sequenced to 484x average coverage and assembled into 20 contigs with a total length of 3896 107 bp. CheckM estimated completeness of this genome as 92% with 6% possible contamination. Phylogenetically BY21 was placed within the *Chloroflexi* class *Anaerolinea* (Fig. 3), members of which have been identified from diverse environments, including permafrost, marine and freshwater sediments, anaerobic sludge bioreactors and deep subsurface aquifers (Yamada and Sekiguchi, 2009). The BY21 16S rRNA gene shares 89% sequence identity with the type organism for this class, *A. thermophila* UNI-1, and 99% sequence identity with the nearest 16S rRNA gene, from an uncultured bacterium msunder68 identified in hot spring microbial mat (Thiel *et al.* 2016).

Genome analysis revealed a set of genes encoding a bacterial flagellar apparatus and chemotaxis functions suggesting that BY21 bacterium, unlike most described *Chloroflexi*, could be motile. Analysis of metabolic pathways revealed that this bacterium is metabolically versatile. The presence of complete Embden-Meyerhof pathway of glycolysis and downstream enzymes, pyruvate-flavodoxin oxidoreductase and acetyl-CoA synthetase (ADP-forming), enables fermentation of carbohydrates with the concomitant formation of ATP. Pathways for the utilization of other sugars are linked to the Embden-Meyerhof pathway. For example, xylose could be metabolized by the action of a putative xylose isomerase and a xylulose kinase. Enzymes of the nonoxidative branch of the pentose phosphate pathway, ribulose phosphate 3-epimerase, ribose-5-phosphate isomerase, transketolase and transaldolase are also encoded.

The search for potential hydrolytic enzymes revealed a very limited range of glycoside hydrolases. Extracellular hydrolysis of beta-linked polysaccharides could be enabled by an endoglucanase of the GH44 family and GH26 family beta-mannanase. Both enzymes were predicted to carry N-terminal signal peptide enabling their export from the cell. On the contrary, two encoded alpha amylases lack signal peptides and likely operates intracellularly. A high number of sugar ABC-type transporters encoded by the genome suggest that BY21 bacterium most likely relies on utilization of simple sugars and short oligosaccharides rather than hydrolysis of complex polymers.

In addition to fermentation, BY21 bacterium was predicted to be capable of complete oxidation of organic substrates in course of aerobic and anaerobic respiration. In the presence of external electron acceptors, acetyl-CoA formed in the Embden-Meyerhof pathway, may be completely oxidized via the tricarboxylic acid cycle, which is encoded by the genome. The genome encodes all of the major components of the electron transfer chain necessary for energy generation via oxidative phosphorylation, namely, the proton-translocating NADH-dehydrogenase complex, membrane-bound succinate dehydrogenase (complex II), cytochrome bc_1 complex III and membrane-linked terminal oxidoreductases. The resulting transmembrane proton gradient may be used for ATP generation by the encoded membrane F_0F_1-type ATP synthase. Terminal oxygen reductases in BY21 bacterium are represented by a cytochrome *bd* ubiquinol oxidase complex and several proton-translocating cytochrome c oxidases of different types. Consistently with the predicted ability of BY21 bacterium to grow under aerobic conditions, its genome encodes a catalase/peroxidase, nickel and iron/manganese superoxide dismutases.

Genome analysis predicted the ability of the bacterium to grow by anaerobic respiration. Dissimilatory reduction of nitrate

Table 2. General characteristics of genomes obtained in this study.

Bin ID	Phylogenetic assignment	MAG quality*	16S rRNA gene**	Completeness (%)	Contamination (%)	Contigs	Genome size (Mbp)***	Sequence coverage	Share in the whole metagenome
57	*Firmicutes*, *Ca* Desulforudis audaxviator	HQ	F	98,1	0,3	48	2.18	413	6,04%
29	*Firmicutes*, *Desulfotomaculum*	HQ	P	100,0	4,6	83	3.26	102	2,24%
45	*Firmicutes*, unclassified		F (x2)	100,0	103,1	751	6.76	14	0,62%
12	*Firmicutes*, unclassified	HQ	F	97,4	1,4	190	3.85	17	0,45%
9	*Firmicutes*, unclassified	HQ	F	96,6	0,6	39	2.35	33	0,52%
15	*Firmicutes*, *Pelotomaculum*	HQ	F	93,4	0,6	330	2.16	9	0,13%
50	*Firmicutes*, unclassified	HQ	F	90,9	2,3	119	3.23	18	0,39%
55	*Nitrospirae*, *Thermodesulfovibrio*	HQ	F	99,0	0,9	17	1.90	140	1,79%
2	*Nitrospirae*, unclassified	HQ	F	97,3	3,6	74	2.89	36	0,71%
47	*Deltaproteobacteria*, *Desulfobacca*	HQ	F	98,1	1,3	32	3.09	1383	28,67%
8	*Deltaproteobacteria*, unclassified *Desulfobacteraceae*	HQ	F	96,8	0,6	26	3.69	37	0,91%
32	*Chloroflexi*, *Thermanaerothrix*		F	100,0	95,8	71	6.29	25	1,06%
23	*Chloroflexi*, unclassified *Anaerolineales*	MQ	F	92,7	5,6	32	3.67	36	0,90%
34	*Chloroflexi*, unclassified *Anaerolineales*		F	96,6	12,3	513	4.08	25	0,69%
1	*Chloroflexi*, unclassified	MQ	F	96,3	5,5	185	4.79	174	5,60%
13	*Chloroflexi*, unclassified *Anaerolineales*	MQ	N	92,7	5,6	32	3.97	51	1,36%
21	*Chloroflexi*, unclassified *Anaerolineales*	MQ	F	91,8	6,1	20	3.90	484	12,68%
39	*Ignavibacteriae*, unclassified	HQ	F	95,6	1,1	23	2.63	826	14,58%
27	*Actinobacteria*, unslassified	HQ	F	95,4	0,8	20	1.84	18	0,22%
33	*Spirochaetes*, unclassified	HQ	F	91,0	0	83	2.32	30	0,46%
11	*Elusimicrobia*, unclassified	HQ	F	96,6	0	88	2.11	32	0,45%
38	Candidate phylum Aminicenantes (OP8)	MQ	F	94,7	5,1	25	2.90	341	6,64%
40	Candidate phylum BRC1	MQ	F	99,4	6,2	4	3.29	37	0,81%
5	Candidate phylum Riflebacteria	HQ	F	98,3	2,5	60	5.70	85	3,24%
19	Unclassified bacterium	HQ	F	96,1	0,5	159	2.56	13	0,23%

*HQ, high-quality draft; MQ, medium-quality draft, according to standards described in Bowers *et al.*, 2017. MAGs with >10% contamination were not classified.
**F, full length (>1400 bp); P, partial; N, missing.
***total length of all contigs.

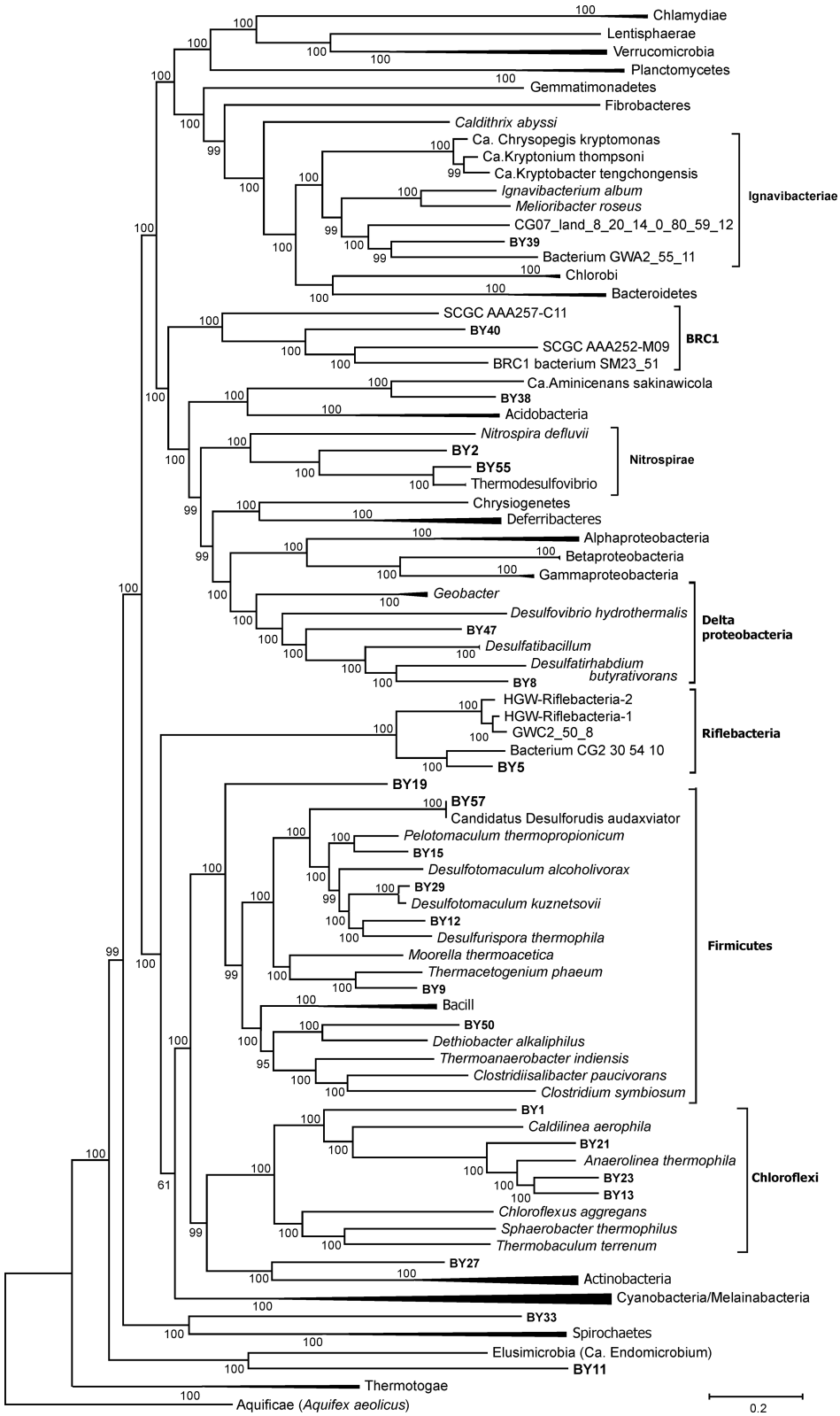

Figure 2. Position of the new genomes in the maximum likelihood concatenated protein phylogeny. Selected part of the CheckM multiple alignment was used for the tree construction in PhyML using default parameters. The tree was inferred from the concatenation of 43 conserved marker genes and incorporates 75 genomes form the CheckM database, 13 other genomes (*Candidatus* Aminicenans, 3 BRC1 bacteria, 5 Ignavibacteriae and 4 Riflebacteria), and 22 new genome bins. The support values for the internal nodes were estimated by approximate Bayes tests in PhyML. Note that genomes with >10% contamination (BY32, 34 and 45) were not included.

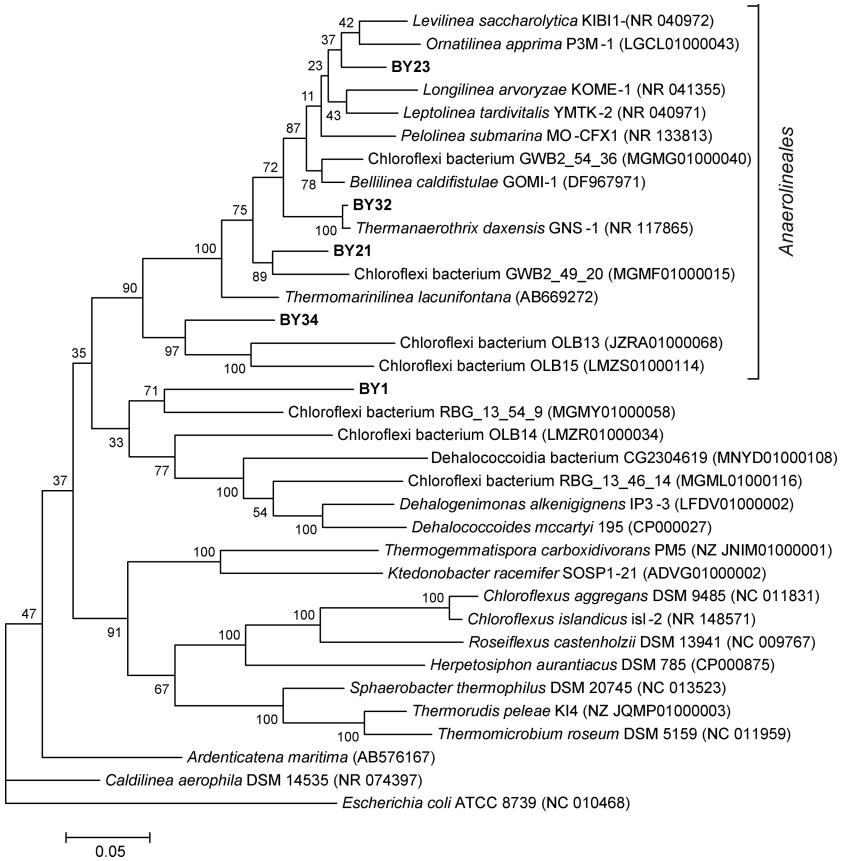

Figure 3. Maximum likelihood 16S rRNA gene phylogenetic tree of the phylum *Chloroflexi*. The tree was computed by MEGA 6.0, using Tamura-Nei substitution model and uniform rates among sites. Bootstrap tests were performed with 100 resamplings. The scale bar represents substitutions per nucleotide base. Note that BY13, assigned to *Chloroflexi* on the basis of concatenated protein phylogeny lacked the 16S rRNA gene and not shown here.

and nitrite could be enabled by the functions of a heterotrimeric nitrate reductase and an ammonia-forming cytochrome c nitrite reductase. The genome analysis revealed gene cluster encoding putative membrane-bound molybdopterin oxidoreductase of the Psr/Phs family. Although phylogenetic analysis of the catalytic subunit of this complex indicated its specificity as a tetrathionate reductase (data not shown), such complexes could also act as thiosulfate-, polysulfide-, nitrate- and arsenate-reductases. Notable in the BY21 genome is the diversity of cytochromes: beyond the cytochrome oxidases, the genome encodes several predicted multiheme c-type cytochromes, including octaheme tetrathionate reductase.

The peculiar property of the BY21 genome is the presence of the complete Wood-Ljungdahl pathway of carbon fixation (Ragsdale and Pierce 2008), including formate dehydrogenase, formate-tetrahydrofolate ligase, the bifunctional enzyme methenyl-THF cyclohydrolase/methylene-THF dehydrogenase, 5,10-methylenetetrahydrofolate reductase, 5-methyltetrahydrofolate:corrinoid iron-sulfur protein methyltransferase and a CO dehydrogenase/acetyl-CoA synthase complex (CODH/ACS). Altogether, two CO_2 molecules are reduced to form acetyl-CoA. The Wood-Ljungdahl pathway is utilized by a wide variety of anaerobic microorganisms for both C1 metabolism and energy conservation by coupling C1 metabolism to either CO_2 reduction (forward) or acetate oxidation (reverse). In the forward direction, homoacetogenic bacteria such as *Moorella thermoacetica* (Pierce *et al.* 2008) and *Acetobacterium woodii* (Schuchmann and Müller 2014) utilize this pathway for acetate production from two CO_2 with H_2 as electron donor. In some sulfate-reducing bacteria, the Wood-Ljungdahl pathway operates in reverse by coupling the oxidation of acetate to sulfate reduction. Previously the Wood-Ljungdahl pathway was found in the genomes of some uncultured *Chloroflexi*, including the members of *Anaerolinea* (Hug *et al.* 2013; Sewell, Kaster and Spormann 2017).

Ignavibacteriae bacterium BY39

Genome of BY39 bacterium was sequenced to 826x average coverage and assembled into 23 contigs with a total length of 2626 338 bp. CheckM estimated completeness of this genome as 96% with a minimal possible contamination. The BY39 16S rRNA gene shares only 85% sequence identity with *I. album*, and 87% sequence identity with the nearest database 16S rRNA gene, from an uncultured bacterium ART_eB17 identified in the deep subsurface aquifer in Japan (AB924432). Several 16S rRNA clones with about 94% identity (KX163592–KX163599) were detected in basaltic subsurface ecosystems.

Analysis of the BY39 genome revealed that this bacterium is not only phylogenetically distant from *I. album* and *M. roseus* but also has important difference of lifestyle and metabolic properties. At first, unlike *I. album* and *M. roseus*, BY39 bacterium is likely nonmotile because it lacked flagellar machinery and most of chemotaxis genes. It has limited enzymatic capacity to deconstruct complex polysaccharides compared with *I. album* and especially *M. roseus*. While the later bacterium has more than 100 genes that were classified into 31 different families of

glycoside hydrolases (GHs), carbohydrate esterases and polysaccharide lyases (Kadnikov *et al.* 2013), only 7 glycoside hydrolases were found in the genome of BY39 bacterium. The utilization of starch may be enabled by three alpha-amylases of GH57 family. Two of them were predicted to be secreted out of the cell and one of also carries a CMB2 starch binding domain. Products of starch hydrolysis could be imported into the cell using maltose/maltodextrin ABC transporters. Other carbohydrate-active enzymes, predicted to contain N-terminal secretion signals, are GH3 family beta-hexosaminidase, GH36 family hydrolase, pectate lyase and an unsaturated glucuronyl hydrolase of GH88 family. Contrary to *M. roseus*, BY39 bacterium lacks enzymes that could enable hydrolysis of cellulose and other beta-linked polysaccharides. Genome analysis also revealed six signal-peptide containing proteases, including three subtilisin-family enzymes, which could enable extracellular hydrolysis of proteinaceous substrates.

The BY39 genome encodes a complete set of genes for glycolysis, gluconeogenesis, the non-oxidative branch of the pentose phosphate pathway and the tricarboxylic acids cycle. The pyruvate produced in the glycolysis could be oxidized by pyruvate-flavodoxin oxidoreductase, followed by conversion of acetyl-CoA to acetate with the concomitant production of ATP via a two-step reaction by phosphate acetyltransferase and acetate kinase. Genes for the fermentative production of ethanol are also present. Several cytoplasmic hydrogenases of the NiFe—family could couple the oxidation of reduced ferredoxin and NADH to the reduction of protons to H_2 in course of fermentative growth.

The major electron transport chain components were identified in the BY39 genome, namely NADH ubiquinone oxidoreductase, membrane-linked succinate dehydrogenase, cytochrome bc_1 complex, cytochrome *c* oxidase and F_0F_1-type H^+-transporting ATPase. Interestingly, in spite of the presence of standard Complex III, the genome of BY39 bacterium also contains an operon encoding quinol-oxidizing alternative Complex III (Yanyushin *et al.* 2005). Unlike *I. album* and *M. roseus*, BY39 bacterium lacks membrane-bound electron transport complex Rnf. The terminal reductases in BY39 bacterium are represented by a single cbb_3-type cytochrome c oxidase, while a quinol oxidase *bd* complex and $cc(o/b)o_3$- type cytochrome c oxidase, found in *M. roseus* (Kadnikov *et al.* 2013), are missing. The cbb_3-type oxidases usually have a very high affinity for oxygen, enabling respiration at microaerobic conditions (Pitcher, Brittain and Watmugh 2002).

Analysis of the BY39 genome revealed limited capacities for anaerobic respiration. Genome analysis revealed no pathways for dissimilatory reduction of sulfate, Fe(III), nitrate and nitrite. However, the genome encodes a putative membrane bound molybdopterin oxidoreductases of the Psr/Phs family, consisting three subunits: a catalytic subunit with molybdopterin, an electron transfer subunit with the [Fe–S] cluster and a membrane anchor NrfD-like subunit, participating in the transfer of electrons from the quinone pool. Phylogenetic analysis of the catalytic subunit predicted its function as a polysulfide reductase. A similar enzyme is encoded in the genome of *M. roseus* (Mros_1774). Since *M. roseus* can use arsenate, but not elemental sulfur and thiosulfate in anaerobic respiration, it was proposed that this oxidoreductase can reduce arsenate and/or another yet unknown electron acceptor (Kadnikov *et al.* 2013).

Candidate phylum *Riflebacteria* bacterium BY5

Genome of BY5 bacterium was sequenced to 85x average coverage and assembled into 60 contigs with a total length of 5695 728 bp. Completeness of this genome was estimated by CheckM as 98% with 2.5% possible contamination. The 16S rRNA of BY5 bacterium have only 82% sequence identity with the nearest cultured bacteria (*Thermaerobacter subterraneus*, phylum *Firmicutes*), indicating a phylum-level position of this bacterial group (Rosselló-Móra and Amann 2015; Konstantinidis, Rosselló-Móra and Amann 2017). Even considering uncultured bacteria, GenBank searches (Jan 27, 2018) revealed no sequences with more than 92% identity to 16S rRNA of BY5 bacterium further supporting the novelty of this lineage. Search for relatives of BY5 among uncultured microorganisms on the basis of genome-to-genome distance evaluation revealed that BY5 is most related to uncultured bacterium CG2_30_54_10 (Probst *et al.* 2017) and three members of the candidate phylum *Riflebacteria*,—GWC2_50_8 (Anantharaman *et al.* 2016), HGW-Riflebacteria-1 and HGW-Riflebacteria-2 (Hernsdorf *et al.* 2017). All these organisms were found in terrestrial subsurface environments—Crystal Geyser (Utah, USA), an aquifer adjacent to the Colorado River near Rifle (Colorado, USA) and a groundwater at the Horonobe Underground Research Laboratory (Hokkaido, Japan). Pairwise comparisons of BY5 genome with CG2_30_54_10, GWC2_50_8, HGW-Riflebacteria-1 and HGW-Riflebacteria-2 genomes showed the AAI values of 64.9%, 50.9%, 51.5% and 51.4%, respectively, while AAI with other genomes was below 42%.

The genomic information from BY5 was used alongside those from other bacterial genomes to perform phylogenetic analysis based on concatenated conservative marker genes. The results show that BY5 together with CG2_30_54_10 and three members of *Riflebacteria*, formed a monophyletic distinct phylum-level lineage (Fig. 2). Consistently with AAI data, the closest relative of BY5 appeared to be bacterium CG2_30_54_10, although the lack of 16S rRNA sequence in the CG2_30_54_10 assembly does not allow determining their taxonomical relationship. Among three other *Riflebacteria*, the 16S rRNA sequence is available only for HGW-Riflebacteria-1; it is 89.4% identical to the 16S rRNA sequence of BY5 bacterium indicating that these two organisms could represents different classes in the candidate phylum *Riflebacteria*.

Analysis of the BY5 genome revealed a set of genes encoding bacterial flagellar machinery and chemotaxis proteins, as well as a set of type IV pilin proteins enabling twitching motility of this bacterium. The search for hydrolytic enzymes predicted that BY5 bacterium is able to degrade some oligo- and polysaccharides. Utilization of starch and similar polymers could be enabled by nine alpha-amylases and a pullulanase. Three alpha-amylases and a pullulanase were predicted to contain N-terminal secretion signal peptide, indicating their extracellular operation. Consistently, the genome encodes maltose/maltodextrin ABC transport system, and enzymes for downstream maltose metabolism. Enzymes responsible for hydrolysis of beta-linked polysaccharides, like cellulose and xylan, were not found, as well as carbohydrate esterases, and polysaccharide lyases. However, two extracellular and one intracellular GH1 family glycoside hydrolases are encoded. These enzymes usually have activities of β-glucosidases and β-galactosidases, β-mannosidases, β-D-fucosidases and β-glucuronidases were also found in this family. The genome also encodes two β-N-acetylglucosaminidases of the GH3 family, one of which was predicted to act outside the cell. This enzyme could hydrolyze chitooligosaccharides and produce N-acetyl-β-glucosamine (Hutcheson, Zhang and Suvorov. 2011). Upon import into the cytoplasm

via specific ABC transport system N-Acetyl-D-glucosamine may be phosphorylated by N-acetylglucosamine kinase NagC. At the final step, N-acetylglucosamine-6-phosphate deacetylase and glucosamine-6-phosphate deaminase transforms the N-Acetyl-D-glucosamine-6-P into fructose-6-P that enters central carbohydrate metabolism.

The BY5 genome contains a complete set of genes encoding enzymes of the Embden-Meyerhof glycolytic pathway and both oxidative and nonoxidative stages of pentose phosphate pathway enabling complete oxidation of sugars (Fig. 4). Glycogen may be used as storage polymer as suggested by the presence of genes for the key enzymes involved in its synthesis and degradation, glucose-1-phosphate adenylyltransferase, glycogen synthase, glycogen branching enzyme, glycogen phosphorylase and glycogen debranching enzyme Production of another storage sugar and osmolyte, trehalose, could be enabled by trehalose synthetase.

Pyruvate could be decarboxylated to acetyl-CoA by a four-subunit pyruvate:ferredoxin oxidoreductase and another, single subunit pyruvate-flavodoxin oxidoreductase. Further conversion of acetyl-CoA to acetate with the production of ATP is performed by two-step reaction by phosphate acetyltransferase and acetate kinase, while acetyl-CoA synthetase was not found. The presence of aldehyde and alcohol dehydrogenases suggested that besides acetate alcohols could be among the fermentation products. Acetyl-CoA generated from pyruvate may enter the tricarboxylic acid cycle. However, this cycle is incomplete and is likely used only for biosynthetic purposes. Reduced products generated in fermentation pathways, NADH and reduced ferredoxin, could be re-oxidized by three cytoplasmic [Fe-Fe] group A3 hydrogenases (Greening *et al.* 2016).

Analysis of the BY5 genome did not reveal the presence of the major components of aerobic electron transfer chains: proton-translocating NADH-dehydrogenase complex, a membrane-bound succinate dehydrogenase, cytochrome bc_1 complex or quinol-oxidizing alternative Complex III, and terminal cytochrome oxidases. However, BY5 bacterium still has several mechanisms to generate a transmembrane ion gradient that could be used by membrane F_0F_1-type ATP synthase for ATP production. The genome of BY5 bacterium contains genes encoding membrane-bound ion-transporting complex Rnf. This complex could action either anabolically to reduce ferredoxin and oxidize NADH for biosynthetic purposes, or catabolically, coupling oxidation of ferredoxin to NAD^+ reduction and pumping of protons or sodium ions across the cytiplasmic membrane (Biegel *et al.* 2011). Based on gene order, *rnfCDGEAB*, the Rnf operon of BY5 bacterium belongs to type 3, typically found in *Firmicutes* where it acts catabolically, as an energy-conserving ferredoxin:NAD^+ oxidoreductase. Two other enzymes could contribute to generation of transmembrane ion gradient—V-type ATPase that use the hydrolysis of ATP produced in fermentative pathways for the transportatiopn of protons, and membrane-bound pyrophosphatase coupling the hydrolysis of pyrophosphate to the translocation of H^+ or Na^+ ions. The absence of lysine in the GNXX(K/A) signature sequence (Malinen *et al.* 2007) suggests that this pyrophosphatase transports sodium ions rather than protons. The balance between H^+ and Na^+ concentrations could be maintained by a multisubunit H^+/Na^+ antiporter of the Mnh familty (Ito *et al.* 1999).

Another energy-conserving complex, identified in BY5 genome, is a Na^+ -translocating NADH-quinone oxidoreductase, Na-NQR. This enzymatic complex is the main ion pump and the primary site where electrons enter the respiratory chain in different pathogenic and free-living bacteria (Reyes-Prieto, Barquera and Juarez 2014). Na-NQR is a sodium-specific ion pump evolutionary related to Rnf complex and composed of six subunits (NqrA-F). Two *nqr* operons were predicted in the BY5 genome, both encoding NqrB, C, D, E and F but not the A subunit. In contrast with other subunits, NqrA is cytosolic, does not contain cofactor binding sites and its function is unclear. It is possible that NqrA was not found or it is dispensable in BY5 bacterium. One of *nqr* operons is clustered with genes that could encode a reductase consuming electrons provided from Na-NQR. Besides *nrqBCDEF*, this cluster comprises five c-type cytochromes with 3 to 12 heme c-binding motifs and a ferritin-like protein. All cytochromes carry N-terminal signal peptides and membrane anchor regions, indicating their export from the cell and membrane localization that are necessary to contact the insoluble electron acceptor. Such c-type cytochromes play a key role in electron transfer out of the cell to the extracellular electron acceptor in well studied dissimilatory Fe(III) reducing Gram-negative bacteria *Shewanella* and *Geobacter* (Shi *et al.* 2007; Richter, Schicklberger and Gescher 2012). Particularly, one of these cytochromes contains eight heme-binding sites and a Ser/Thr-Pro-Ser/Thr motif enabling binding to hematite, the mineral form of iron(III) oxide (Fe_2O_3), as described previously for a putative terminal Fe(III)-oxide reductase of *S. oneidensis* (Lower *et al.* 2008). Adhesion to Fe(III) minerals could be also facilitated by the type IV pili, encoded by the BY5 genome. Therefore, this oxidoreductase could accept electrons from the quinone pool and perform extracellular reduction of Fe(III). Since the known pathways for dissimilatory reduction of sulfate, sulfite, nitrate, nitrite, arsenate and sulfur compounds were not found in the genome, iron reduction seems to be the only respiratory pathway of BY5 bacterium. The predicted metabolic pathways are summarized in Fig. 4.

Very limited information is available about the metabolic capabilities of other *Riflebacteria*. Only HGW-Riflebacteria-1 genome is near complete, while three other genomes were obtained as medium-quality drafts, which complicate metabolic reconstructions. Genomes of GWC2_50_8, HGW-Riflebacteria-1 and HGW-Riflebacteria-2, contain anaerobic sulfite reductase (asr) genes and it was suggested that they could perform dissimilatory sulfite reduction using the Asr complex as observed in *Clostridium difficile* (Anantharaman *et al.* 2018). However, the *asr* genes are absent in the BY5 genome, as well as in the draft (63% complete, as estimated by CheckM) genome of its closest relative, bacterium CG2_30_54_10. Like BY5, the later bacterium lacked the aerobic respiratory pathways and could be capable of dissimilatory iron reduction, since its genome encodes multihaem c-type cytochromes, homologues to putative iron reductase of BY5.

Phylogenetic and genomic analysis of the new BY5 genome classified it within a candidate phylum *Riflebacteria*. The genome meets the criteria, recently suggested for description of new taxa of uncultivated Bacteria and Archaea (Konstantinidis, Rosselló-Móra and Amann 2017), and we propose the following taxonomic names for the novel genus and species of BY5:

- Description of *Candidatus* Ozemobacter gen. nov. (O.ze.mo.bac'ter. N.L. masc. n. *Ozem*, a name for the underground deity in Slavic mythology; L. masc. n. *bacter*, a rod).
- Description of *Candidatus* Ozemobacter sibiricus sp. nov.

Ozemobacter sibiricus (si.bi'ri.cus N.L. masc. adj. *sibiricus*, originating from Siberia). Not cultivated. Inferred to be anaerobic, motile, obligate organotroph, obtains energy by fermentation or

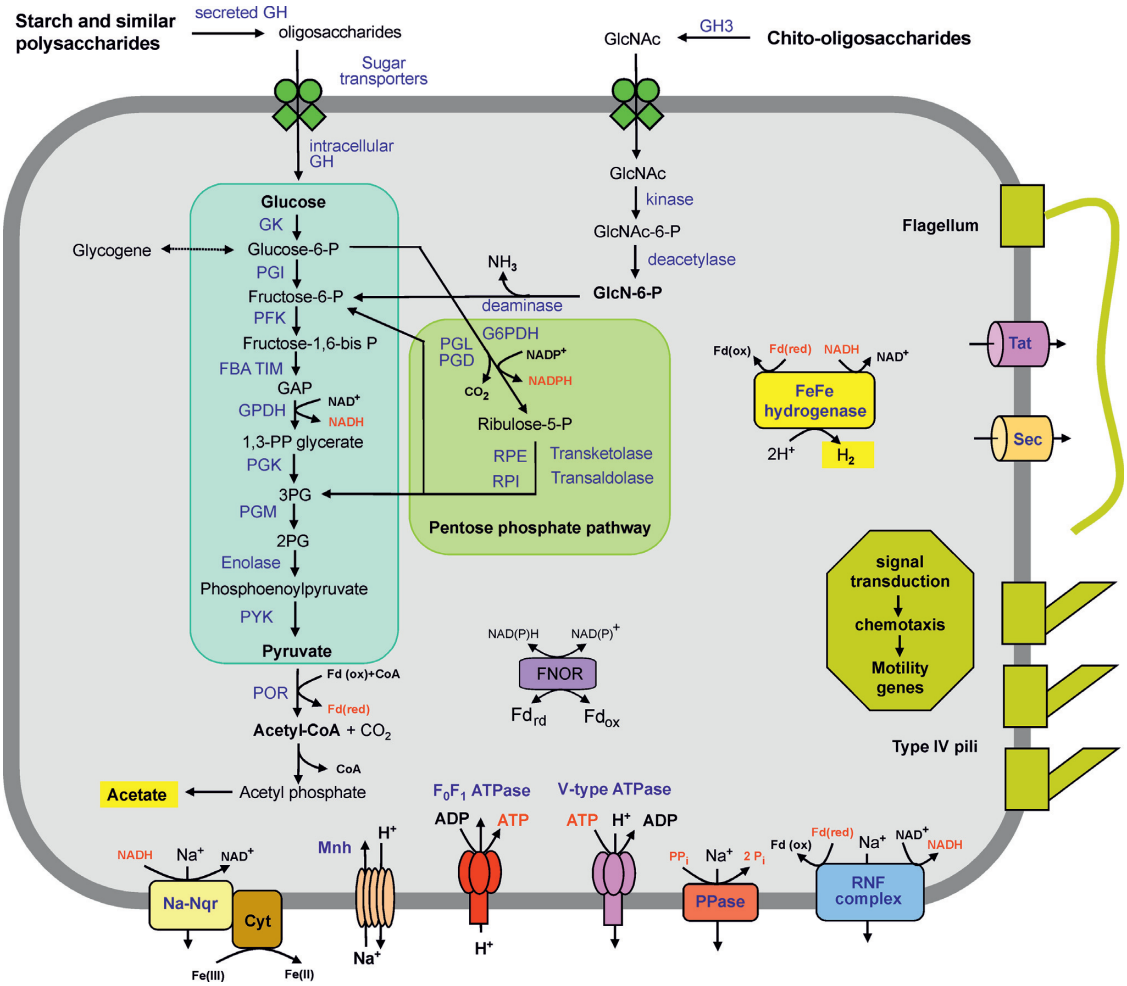

Figure 4. An overview of the metabolism of Bin BY5 bacterium (*Candidatus* Ozemobacter sibiricus) reconstructed from its genome. Enzyme abbreviations: GH, glycoside hydrolase; GK, glukokinase; PGI, glucose-6-phosphate isomerase; PFK, 6-phosphofructokinase; FBA, fructose-bisphosphate aldolase; TIM, triosephosphate isomerase; GPDH, glyceraldehyde 3-phosphate dehydrogenase; PGK, phosphoglycerate kinase; PGM, phosphoglycerate mutase; PYK, pyruvate kinase; G6PDH, glucose-6-phosphate 1-dehydrogenase; PGL, 6-phosphogluconolactonase; PGD, 6-phosphogluconate dehydrogenase; RPE, Ribulose-phosphate 3-epimerase; RPI, Ribose 5-phosphate isomerase; POR, pyruvate ferredoxin oxidoreductase; FNOR, ferredoxin-NAD(P)$^+$ reductase; PPase, pyrophosphatase; Mnh, a multisubunit Na^+/H^+ antiporter of the Mnh family; Na-Nqr, Na^+ -translocating NADH-quinone oxidoreductase. Other abbreviations: ox/red, oxidized and reduced forms; cyt, cytochrome c; Pi, phosphate; PPi, pyrophosphate; CoA, coenzyme A.

respiration with Fe(III) and able to use starch as a growth substrate. Represented by near-complete genome (GenBank acc. no. QOQW00000000) obtained from metagenomes of a deep subsurface thermal aquifer in Western Siberia, Russia.

Based on this, we propose the following names for the class, order and family:

- *Candidatus* Ozemobacteria classis nov.
- *Candidatus* Ozemobacterales ord. nov.
- *Candidatus* Ozemobacteraceae fam. nov.

An ecological model of the deep subsurface aquifer

Taken together, these genome-based metabolic predictions provide insight into the links between the resources of the deep subsurface aquifer and microbial ecology (Fig. 5). The 16S rRNA molecular profiling and metagenomic analysis revealed functionally diverse community with different types of metabolism. The abundant organics buried in Mesozoic sediments of marine origin provides substrates for respiratory organisms and fermenters. The later produces CO_2, hydrogen, acetate and other simple organic compounds that can be used as electron donors by dissimilatory sulfate and iron reducers. The latter group was represented by BY5 lineage.

Sulfate-reducing bacteria accounted for about one third of microorganisms and included species able to utilize all major fermentation products—hydrogenotrophic chemolithoautotrophes (*Ca.* Desulforudis audaxviator, *Desulfotomaculum*) and heterotrophes, growing on acetate (*Desulfobacca*) and other low molecular weight organic compounds (*Thermodesulfovibrio*). The abundance of sulfate reducers was not anticipated, given that sulfate concentration in the groundwater is below 5 mg L^{-1}, but several explanations could be proposed. First, local concentration of sulfate could be much higher in porewater of the rock matrix due to oxidation of sulfur-containing minerals and/or dissolution of sulfate-containing minerals such as gypsum. Second, sulfate could remain in the water of relict ocean origin. Taking into account mineralization of groundwater, relict marine water could account for about 5% of the whole, but it also could be localized in spatially isolated sulfate-rich parts of the aquifer system.

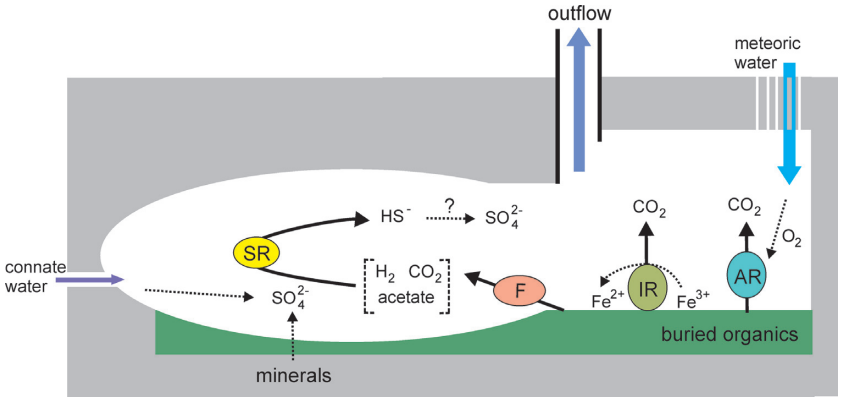

Figure 5. An ecological model of the deep subsurface aquifer. AR, aerobic respiration; F, fermentation; SR, sulfate reduction; IR, anaerobic respiration by the example of iron reduction.

Analysis of assembled genomes representing other abundant groups—*Chloroflexi*, *Ignavibacteriae* and *Riflebacteria* revealed their different metabolic specialization. All of them have fermentative pathways enabling production of acetate and hydrogen, but rather limited capacities for the hydrolysis of complex polymers, restricted to some beta-linked polysaccharides and proteinaceous substrates (*Chloroflexi* BY21) or starch (*Ignavibacteriae* BY39 and *Riflebacteria* BY5). Utilization of complex polysaccharides in fermentative pathways could be also performed by members of another abundant group, the candidate phylum *Aminicenantes* (Rinke *et al.* 2013; Robbins *et al.* 2016). The presence of bacteria capable to degrade complex polysaccharides corroborates with the previous reports on isolation of such organisms from Western Siberian deep subsurface aquifers (e.g. Podosokorskaya *et al.* 2013).

Both *Chloroflexi* BY21 and *Ignavibacteriae* BY39 in addition to fermentation has the capacities for aerobic respiration, and while the first bacterium has different cytochrome oxidases and seems to be adapted to various concentrations of oxygen, BY39 is likely restricted to the microaerobic conditions. The capacity for aerobic respiration was rather surprizing for microorganisms from a deep subsurface, but corresponded to the finding of oxygen in the gas samples. Likewise, oxygen has been found in gas samples collected at 3P borehole during several years (Frank *et al.* 2016a). We could speculate that oxygen was delivered to the subsurface aquifer with meteoric recharge waters. The presence of aerobic microorganisms in subsurface environments was also reported in other studies, for example, in Alberta oil sands where the presence of oxygen was suggested to result from the influx of oxygenated meteoric waters (An *et al.* 2013).

Overall, the findings presented in this study provide insight to the diversity and metabolism of microorganisms of a deep subsurface sedimentary environment. The availability of buried organics and spatial heterogeneity of the subsurface aquifer system provide conditions for the development of microbial communities that are taxonomically and functionally more diverse as compared to those found in oligotrophic hydrogen-fueled subsurface environments. The metabolic reconstruction results will motivate targeted cultivation strategies for uncultured bacterial groups such as *Riflebacteria*, a lineage indigenous for the deep subsurface.

ACKNOWLEDGMENTS

This work was performed using the scientific equipment of the Core Research Facility 'Bioengineering' (Research Center of Biotechnology RAS).

FUNDING

The work was supported by the Russian Science Foundation (grant 14–14-01016).

Conflicts 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.[CrossRef] PMID: 25218180

Anantharaman K, Brown CT, Hug LA *et al.* Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat Commun* 2016;**7**:13219.[CrossRef] PMID: 27774985

Anantharaman K, Hausmann B, Jungbluth SP *et al.* Expanded diversity of microbial groups that shape the dissimilatory sulfur cycle. *ISME J.* 2018;**12**:1715–28.[CrossRef] PMID: 29467397

An D, Caffrey SM, Soh J *et al.* Metagenomics of hydrocarbon resource environments indicates aerobic taxa and genes to be unexpectedly common. *Environ Sci Technol* 2013;**47**:10708–17.[CrossRef] PMID: 23889694

Aüllo T, Ranchou-Peyruse A, Ollivier B *et al.* Desulfotomaculum spp. and related gram-positive sulfate-reducing bacteria in deep subsurface environments. *Front Microbiol* 2013;**4**:362.[CrossRef] PMID: 24348471

Bankevich A, Nurk S, Antipov D *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;**19**:455–77.[CrossRef] PMID: 22506599

Banks D, Frank YA, Kadnikov VV *et al. Hydrochemical Data Report from Sampling of Two Deep Abandoned Hydrocarbon Exploration Wells: Byelii Yar and Parabel', Tomsk oblast', Western Siberia, Russian Federation*, Trondheim: NGU Report, 2014.034, Geological Survey of Norway, 2014.

Biegel E, Schmidt S, González JM *et al.* Biochemistry, evolution and physiological function of the Rnf complex, a novel ion-motive electron transport complex in prokaryotes. *Cell Mol Life Sci* 2011;**68**:613–34.[CrossRef] PMID: 21072677

Bowers RM, Kyrpides NC, Stepanauskas R *et al.* Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol* 2017;**35**:725–31.[CrossRef] PMID: 28787424

Brettin T, Davis JJ, Disz T *et al.* RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci Rep* 2015;**5**:8365.[CrossRef] PMID: 25666585

Bryant DA, Liu Z, Li T *et al.* Comparative and functional genomics of anoxygenic green bacteria from the taxa Chlorobi, Chloroflexi, and Acidobacteria. In: Burnap R Vermaas W (eds.). *Functional Genomics and Evolution of Photosynthetic Systems*, Netherlands: Springer, 2012, 47–102.

Chivian D, Brodie EL, Alm EJ *et al.* Environmental genomics reveals a single-species ecosystem deep within Earth. *Science* 2008;**322**:275–8.[CrossRef] PMID: 18845759

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

Cole JR, Wang Q, Cardenas E *et al.* The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res* 2009;**37**(Database issue): D141–5.[CrossRef] PMID: 19004872

Dong X, Dröge J, von Toerne C *et al.* Reconstructing metabolic pathways of a member of the genus Pelotomaculum suggesting its potential to oxidize benzene to carbon dioxide with direct reduction of sulfate. *FEMS Microbiol Ecol* 2017;**93**:fiw254.

Edgar RC, Haas BJ, Clemente JC *et al.* UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 2011;**27**:2194–200.[CrossRef] PMID: 21700674

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

Eren AM, Esen ÖC, Quince C *et al.* Anvi'o: an advanced analysis and visualization platform for 'omics data. *Peer J* 2015;**3**:e1319.[CrossRef] PMID: 26500826

Farag IF, Davis JP, Youssef NH *et al.* Global patterns of abundance, diversity and community structure of the Aminicenantes (candidate phylum OP8). *PloS One* 2014;**9**:e92139.[CrossRef] PMID: 24637619

Frank YA, Kadnikov VV, Gavrilov SN *et al.* Stable and variable parts of microbial community in siberian deep subsurface thermal aquifer system revealed in a long-term monitoring study. *Front Microbiol* 2016a;**7**:2101.[CrossRef]

Frank YA, Kadnikov VV, Lukina AP *et al.* Characterization and genome analysis of the first facultatively alkaliphilic Thermodesulfovibrio isolated from the deep terrestrial subsurface. *Front Microbiol* 2016b;**7**:2000.[CrossRef]

Gihring TM, Moser DP, Lin LH *et al.* The distribution of microbial taxa in the subsurface water of the Kalahari shield, South Africa. *Geomicrobiol J* 2006;**23**:415–30.[CrossRef]

Greening C, Biswas A, Carere CR *et al.* Genomic and metagenomic surveys of hydrogenase distribution indicate H2 is a widely utilised energy source for microbial growth and survival. *ISME J* 2016;**10**:761–77.[CrossRef] PMID: 26405831

Grégoire P, Fardeau ML, Joseph M *et al.* Isolation and characterization of Thermanaerothrix daxensis gen. nov., sp. nov., a thermophilic anaerobic bacterium pertaining to the phylum "Chloroflexi", isolated from a deep hot aquifer in the Aquitaine Basin. *Syst Appl Microbiol* 2011;**34**:494–7.[CrossRef] PMID: 21621938

Guindon S, Dufayard JF, Lefort V *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 2010;**59**:307–21.[CrossRef] PMID: 20525638

Hallbeck L, Pedersen K. Characterization of microbial processes in deep aquifers of the Fennoscandian Shield. *Appl Geochem* 2008;**23**:1796–819.[CrossRef]

Hernsdorf AW, Amano Y, Miyakawa K *et al.* Potential for microbial H2 and metal transformations associated with novel bacteria and archaea in deep terrestrial subsurface sediments. *ISME J* 2017;**11**:1915–29.[CrossRef] PMID: 28350393

Hubalek V, Wu X, Eiler A *et al.* Connectivity to the surface determines diversity patterns in subsurface aquifers of the Fennoscandian shield. *ISME J* 2016;**10**:2447–58.[CrossRef] PMID: 27022994

Hug LA, Castelle CJ, Wrighton KC *et al.* Community genomic analyses constrain the distribution of metabolic traits across the Chloroflexi phylum and indicate roles in sediment carbon cycling. *Microbiome* 2013;**1**:22.[CrossRef] PMID: 24450983

Hutcheson SW, Zhang H, Suvorov M. Carbohydrase systems of Saccharophagus degradans degrading marine complex polysaccharides. *Mar Drugs* 2011;**9**:645–65.[CrossRef] PMID: 21731555

Iino T, Mori K, Uchino Y *et al.* Ignavibacterium album gen. nov., sp. nov., a moderately thermophilic anaerobic bacterium isolated from microbial mats at a terrestrial hot spring and proposal of Ignavibacteria classis nov., for a novel lineage at the periphery of green sulfur bacteria. *Int J Syst Evol Microbiol* 2010;**60**:1376–82.[CrossRef] PMID: 19671715

Ito M, Guffanti AA, Oudega B *et al.* mrp, a multigene, multifunctional locus in Bacillus subtilis with roles in resistance to cholate and to Na+ and in pH homeostasis. *J Bacteriol* 1999;**181**:2394–402. PMID: 10198001

Jungbluth SP, del Rio TG, Tringe SG *et al.* Genomic comparisons of a bacterial lineage that inhabits both marine and terrestrial deep subsurface systems. *PeerJ* 2017;**5**:e3134.[CrossRef] PMID: 28396823

Jørgensen BB. Shrinking majority of the deep biosphere. *Proc Natl Acad Sci USA* 2012;**109**:15976–7.[CrossRef] PMID: 23012471

Kadnikov VV, Frank YA, Mardanov AV *et al.* Uncultured bacteria and methanogenic archaea predominate in the microbial community of Western Siberian deep subsurface aquifer. *Microbiology* 2017a;**86**:412–5.[CrossRef]

Kadnikov VV, Frank YA, Mardanov AV *et al.* Variability of the composition of the microbial community of the deep subsurface thermal aquifer in Western Siberia. *Microbiology* 2017b;**86**:765–72.[CrossRef]

Kadnikov VV, Mardanov AV, Podosokorskaya OA *et al.* Genomic analysis of Melioribacter roseus, facultatively anaerobic organotrophic bacterium representing a novel deep lineage within Bacteriodetes/Chlorobi group. *PLoS One* 2013;**8**:e53047.[CrossRef] PMID: 23301019

Konstantinidis KT, Rosselló-Móra R, Amann R. Uncultivated microbes in need of their own taxonomy. *ISME J* 2017;**11**:2399–406.[CrossRef] PMID: 28731467

Krzmarzick MJ, Crary BB, Harding JJ *et al.* Natural niche for organohalide-respiring Chloroflexi. *Appl Environ Microbiol* 2012;**78**:393–401.[CrossRef] PMID: 22101035

Lever MA, Rogers KL, Lloyd KG *et al.* Life under extreme energy limitation: a synthesis of laboratory- and field-based investigations. *FEMS Microbiol Rev* 2015;**39**:688–728.[CrossRef] PMID: 25994609

Liu Z, Frigaard NU, Vogl K *et al.* Complete genome of Ignavibacterium album, a metabolically versatile, flagellated, facultative anaerobe from the phylum Chlorobi. *Front Microbiol* 2012;**3**:185.[CrossRef] PMID: 22661972

Lower BH, Lins RD, Oestreicher Z *et al.* In vitro evolution of a peptide with a hematite binding motif that may constitute

a natural metal-oxide binding archetype. *Environ Sci Technol* 2008;**42**:3821–7.[CrossRef] PMID: 18546729

Magnabosco C, Ryan K, Lau MC *et al.* A metagenomic window into carbon metabolism at 3 km depth in Precambrian continental crust. *ISME J* 2016;**10**:730–41.[CrossRef] PMID: 26325359

Malinen AM, Belogurov GA, Baykov AA *et al.* Na+-pyrophosphatase: a novel primary sodium pump. *Biochemistry* 2007;**46**:8872–8.[CrossRef] PMID: 17605473

Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 2011;**17**:10–12.[CrosRef]

Miettinen H, Kietäväinen R, Sohlberg E *et al.* Microbiome composition and geochemical characteristics of deep subsurface high-pressure environment, Pyhäsalmi mine Finland. *Front Microbiol* 2015;**6**:1203.[CrossRef] PMID: 26579109

Moser DP, Gihring TM, Brockman FJ *et al.* Desulfotomaculum and Methanobacterium spp. dominate a 4- to 5-kilometer-deep fault. *Appl Environ Microbiol* 2005;**71**:8773–83.[CrossRef] PMID: 16332873

Nealson KH, Inagaki F, Takai K. Hydrogen-driven subsurface bial community rust lithoautotrophic microbial ecosystems (SLiMEs): do they exist and why should we care ? *Trends Microbiol* 2005;**13**:405–10.[CrossRef] PMID: 16054814

Oude Elferink SJ, Akkermans-van Vliet WM, Bogte JJ *et al.* Desulfobacca acetoxidans gen. nov., sp. nov., a novel acetate-degrading sulfate reducer isolated from sulfidogenic granular sludge. *Int J Syst Bacteriol* 1999;**49**:345–50.[CrossRef] PMID: 10319454

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.[CrosRef] PMID: 25977477

Pierce E, Xie G, Barabote RD *et al.* The complete genome sequence of Moorella thermoacetica (f. Clostridium thermoaceticum). *Environ Microbiol* 2008;**10**: 2550–73.[CrossRef] PMID: 18631365

Pitcher RS, Brittain T, Watmugh NJ. Cytochrome cbb3 oxidase and bacterial microaerobic metabolism. *Biochem Soc Trans* 2002;**30**:653–8.[CrossRef] PMID: 12196157

Podosokorskaya OA, Kadnikov VV, Gavrilov SN *et al.* Characterization of Melioribacter roseus gen. nov., sp. nov., a novel facultatively anaerobic thermophilic cellulolytic bacterium from the class Ignavibacteria, and a proposal of a novel bacterial phylum Ignavibacteriae. *Environ Microbiol* 2013;**15**:1759–71.[CrossRef] PMID: 23297868

Probst AJ, Castelle CJ, Singh A *et al.* Genomic resolution of a cold subsurface aquifer community provides metabolic insights for novel microbes adapted to high CO2 concentrations. *Environ Microbiol* 2017;**19**:459–74.[CrossRef] PMID: 27112493

Purkamo L, Bomberg M, Nyyssönen M *et al.* Dissecting the deep biosphere: retrieving authentic microbial communities from packer-isolated deep crystalline bedrock fracture zones. *FEMS Microbiol Ecol* 2013;**85**:324–37.[CrossRef] PMID: 23560597

Quast C, Pruesse E, Yilmaz P *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 2013;**41**(Database issue):D590–6. PMID: 23193283

Ragsdale SW, Pierce E. Acetogenesis and the Wood–Ljungdahl pathway of CO(2) fixation. *Biochim Biophys Acta* 2008;**1784**:1873–98.[CrossRef] PMID: 18801467

Reyes-Prieto A, Barquera B, Juarez O. Origin and evolution of the sodium-pumping NADH: ubiquinone oxidoreductase. *PLoS One* 2014;**9**:e96696.[CrossRef] PMID: 24809444

Richter K, Schicklberger M, Gescher J. Dissimilatory reduction of extracellular electron acceptors in anaerobic respiration. *Appl Environ Microbiol* 2012;**78**:913–21.[CrossRef] PMID: 22179232

Rinke C, Schwientek P, Sczyrba A *et al.* Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 2013,**499**:431–7.[CrossRef] PMID: 23851394

Robbins SJ, Evans PN, Parks DH *et al.* Genome-centric analysis of microbial populations enriched by hydraulic fracture fluid additives in a coal bed methane production well. *Front Microbiol* 2016;**7**:731.[CrossRef] PMID: 27375557

Rodriguez-R LM, Konstantinidis KT. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. *PeerJ Preprints* 2016;**4**:e1900v1.

Rosselló-Móra R, Amann R. Past and future species definitions for Bacteria and Archaea. *Syst Appl Microbiol* 2015;**38**:209–16.[CrossRef] PMID: 25747618

Sahl JW, Schmidt R, Swanner ED *et al.* Subsurface microbial diversity in deep-granitic-fracture water in Colorado. *Appl Environ Microbiol* 2008;**74**:143–52.[CrossRef] PMID: 17981950

Schloss PD, Westcott SL, Ryabin T *et al.* Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 2009;**75**:7537–41.[CrossRef] PMID: 19801464

Schuchmann K, Müller V. Autotrophy at the thermodynamic limit of life: a model for energy conservation in acetogenic bacteria. *Nat Rev Microbiol* 2014;**12**:809–21.[CrossRef] PMID: 25383604

Sewell HL, Kaster AK, Spormann AM. Homoacetogenesis in deep-sea chloroflexi, as inferred by single-cell genomics, provides a link to reductive dehalogenation in terrestrial dehalococcoidetes. *mBio* 2017;**8**:e02022–17.[CrossRef] PMID: 29259088

Shi L, Squier TC, Zachara JM *et al.* Respiration of metal (hydr) oxides by Shewanella and Geobacter: a key role for multihaem c-type cytochromes. *Mol Microbio* l 2007;**65**:12–20.[CrosRef]

Takai K, Moser DP, DeFlaun M *et al.* Archaeal diversity in waters from deep south african gold mines. *Appl Environ Microbiol* 2001;**67**:750–60.[CrossRef] PMID: 11157240

Tamura K, Stecher G, Peterson D *et al.* MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 2013;**30**:2725–9.[CrossRef] PMID: 24132122

Thiel V, Wood JM, Olsen MT *et al.* The dark side of the mushroom spring microbial mat: life in the shadow of chlorophototrophs. I. Microbial diversity based on 16S rRNA gene amplicons and metagenomic sequencing. *Front Microbiol* 2016;**7**:919.[CrossRef] PMID: 27379049

Varghese NJ, Mukherjee S, Ivanova N *et al.* Microbial species delineation using whole genome sequences. *Nucleic Acids Res* 2015;**43**:6761–71.[CrossRef] PMID: 26150420

Whitman WB, Coleman DC, Wiebe WJ. Prokaryotes: the unseen majority. *Proc Natl Acad Sci USA* 1998;**95**:6578–83.[CrossRef]

Wick RR, Schultz MB, Zobel J *et al.* Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* 2015;**31**:3350–2.[CrossRef] PMID: 26099265

Wilson K. Preparation of genomic DNA from bacteria. *Curr Protoc Mol Biol* 1987;**2**:2–4.

Wu X, Holmfeldt K, Hubalek V *et al.* Microbial metagenomes from three aquifers in the Fennoscandian shield terrestrial deep biosphere reveal metabolic partitioning among populations. *ISME J* 2016;**10**:1192–203.[CrossRef] PMID: 26484735

Yamada T, Sekiguchi Y. Cultivation of uncultured chloroflexi subphyla: significance and ecophysiology of formerly uncultured chloroflexi 'subphylum i' with natural and biotechnological relevance. *Microbes Environ* 2009;**24**:205–16.[CrossRef] PMID: 21566375

Yanyushin MF, del Rosario MC, Brune DC *et al*. New class of bacterial membrane oxidoreductases. *Biochemistry* 2005;**44**:10037–45.[CrossRef] PMID: 16042380