

ORIGINAL ARTICLE

Microbial metagenomes from three aquifers in the Fennoscandian shield terrestrial deep biosphere reveal metabolic partitioning among populations

Xiaofen Wu¹, Karin Holmfeldt¹, Valerie Hubalek², Daniel Lundin¹, Mats Åström³, Stefan Bertilsson² and Mark Dopson¹

¹Centre for Ecology and Evolution in Microbial Model Systems (EEMiS), Linnaeus University, Kalmar, Sweden;

²Department of Ecology and Genetics, Limnology and Science for Life Laboratory, Uppsala University,

Uppsala, Sweden and ³Department of Biology and Environmental Science, Linnaeus University, Kalmar, Sweden

Microorganisms in the terrestrial deep biosphere host up to 20% of the earth's biomass and are suggested to be sustained by the gases hydrogen and carbon dioxide. A metagenome analysis of three deep subsurface water types of contrasting age (from <20 to several thousand years) and depth (171 to 448 m) revealed phylogenetically distinct microbial community subsets that either passed or were retained by a 0.22 µm filter. Such cells of <0.22 µm would have been overlooked in previous studies relying on membrane capture. Metagenomes from the three water types were used for reconstruction of 69 distinct microbial genomes, each with >86% coverage. The populations were dominated by *Proteobacteria*, Candidate divisions, unclassified archaea and unclassified bacteria. The estimated genome sizes of the <0.22 µm populations were generally smaller than their phylogenetically closest relatives, suggesting that small dimensions along with a reduced genome size may be adaptations to oligotrophy. Shallow 'modern marine' water showed community members with a predominantly heterotrophic lifestyle. In contrast, the deeper, 'old saline' water adhered more closely to the current paradigm of a hydrogen-driven deep biosphere. The data were finally used to create a combined metabolic model of the deep terrestrial biosphere microbial community.

The ISME Journal (2016) 10, 1192–1203; doi:10.1038/ismej.2015.185; published online 20 October 2015

Introduction

The existence of a deep biosphere was postulated ~20 years ago (Gold, 1992) and it is now known that the presence of microorganisms extends for several km below the surface (Lin *et al.*, 2006). As a result of its recent discovery and the inherent difficulties in accessing samples from these systems, the deep biosphere is one of the least understood ecosystems on earth. In recent studies it has been estimated that 4.1×10^{15} g C reside in the deep marine biosphere (Kallmeyer *et al.*, 2012) whereas the deep continental subsurface is estimated to contain from 10^{16} to 10^{17} g C (McMahon and Parnell, 2014). At depths below those influenced by organic carbon from the surface, the environment is suggested to be highly oligotrophic (Hoehler and Jorgensen, 2013) and deep life has been described as proceeding 'in extreme slow motion' with generation times ranging from

hundreds to thousands of years (Jorgensen, 2011; Onstott *et al.*, 2014). Despite living under extreme energy limitation, deep subsurface microorganisms are suggested to be predominantly viable and active (Hoehler and Jorgensen, 2013). Several lines of evidence support this idea including gene expression in marine seafloor metatranscriptomes implying metabolically active, dividing cells from all three domains of life (Orsi *et al.*, 2013). Whereas information on the diversity of these deep-dwelling microorganisms is scarce, it was recently shown that microorganisms in more shallow groundwater aquifers are diverse and include representatives from many uncultivated or poorly characterized lineages (Luef *et al.*, 2015). Many of the cells in these aquifers were also characterized by small physical size and streamlined genomes (Luef *et al.*, 2015).

The SKB (Swedish Nuclear Fuel and Waste Management Co.)-run Äspö HRL (Hard Rock Laboratory) is located in Proterozoic crystalline bedrock of the Fennoscandian shield and consists of a 3.6-km long tunnel extending to 460 m below the ground. Microbiological studies of the Äspö HRL have thus far relied on a mixture of culture-dependent methods, adenosine triphosphate assays and 16S rRNA

Correspondence: M Dopson, Centre for Ecology and Evolution in Microbial Model Systems (EEMiS), Linnaeus University, 391 82 Kalmar, Sweden.

E-mail: mark.dopson@lnu.se

Received 8 May 2015; revised 1 September 2015; accepted 6 September 2015; published online 20 October 2015

gene approaches. It is suggested that the bedrock waters represent an anaerobic and oligotrophic environment colonized by nitrate-, ferric-, sulfate- and manganese-reducing microorganisms along with acetogens and methanogens (Hallbeck and Pedersen, 2012; Pedersen, 2013; Ionescu *et al.*, 2015b). A preliminary model of energy and carbon acquisition being driven by the 'geogases' hydrogen and carbon dioxide (Pedersen, 1999) is supported by the presence of the hydrogenotrophic sulfate-reducing bacterium, *Desulfovibrio aespoeensis* (Pedersen *et al.*, 2014). However, other potential energy sources that may be exploited by microorganisms in the deep biosphere should also be evaluated. These include anaerobic oxidation of inorganic sulfur compounds (Osorio *et al.*, 2013) and exudation of organic carbon by chemolithoautotrophic microorganisms that may be subsequently utilized by heterotrophs (Nancucchio and Johnson, 2010). Potential chemolithotrophic metabolisms of deep biosphere microorganisms have traditionally been evaluated based upon standard conditions rather than *in situ* conditions and reactant concentrations. Using *in situ* values suggests that energy can be harvested from the transformation of sulfur, iron, nitrogen, methane and manganese compounds, and this is in good alignment with microbial metabolic capabilities inferred from previous 16S rRNA gene inventories (Osburn *et al.*, 2014).

Metagenomics is the sequencing of environmental DNA from the entire microbial community without introducing biases from culturing or PCR amplification. This technique has been applied to the deep subsurface such as the Outokumpu drill hole where homoacetogenic, methylotrophic and methanogenic processes are present (Nyyssonen *et al.*, 2014). However, in the referenced study, the metabolic functionality was not partitioned to different taxa. A deep terrestrial biosphere metagenome has also been used to reconstruct the complete genome of *Candidatus Desulforudis audaxviator* that comprises >99.9% of the microorganisms in a fracture within a 2.8-km deep South African gold mine (Chivian, 2008). A third study identified the dominating microorganism to be *Halomonas sulfidaeris* in a 1.8-km deep subsurface Cambrian Sandstone reservoir (Dong *et al.*, 2014). The latter environments are characterized by low species diversity in contrast to more diverse communities in the Outokumpu drill hole (Nyyssonen *et al.*, 2014) and have very different geological conditions to the Fennoscandian Shield.

The present study investigates the genetic potential in microbial metagenomes from three water types at the Äspö HRL. To advance our understanding of metabolic interactions within the indigenous microbial community, the sequencing data were partitioned and assembled into amalgamated genomes from phylogenetically distinct populations. Subsequently, a metabolic reconstruction of the dominant community members and an insight into their potential dependencies and other

community-level interactions in the deep terrestrial biosphere were inferred.

Materials and methods

Site description

The Äspö HRL is situated in the south of Sweden (Lat N 57° 26' 4'' Lon E 16° 39' 36''). The bedrock is principally 1800 million years old granite and quartz monzodiorite (Hallbeck and Pedersen, 2008; Ström *et al.*, 2008) and thus is dominated by quartz and aluminosilicate minerals like feldspars and mica. The Äspö HRL tunnel contains boreholes drilled into the bedrock that interconnect with fissure systems bearing waters with different dominating characteristics. More detailed information on the chemistry, geology and hydrology of the fissures are given elsewhere (Smellie *et al.*, 1995; Laaksoharju *et al.*, 2008; Mathurin *et al.*, 2012).

Monitoring of abiotic factors

The chemical and isotopic data of the groundwaters were collected within the Groundwater Chemical Monitoring program of the Äspö HRL. The principal component analysis was generated using gdata package (Warnes *et al.*, 2014) and prcomp from stats package in R (R Development Core Team, 2011).

Microbial sampling, cell counting, microscopy, DNA preparation and sequencing

Borehole waters were first flushed with three to five section volumes of water. Planktonic cells >0.22 µm and >0.1 µm were collected on membrane filters and DNA extracted using the MO BIO PowerWater DNA isolation kit as described in Supplementary Files 1 and 2. The cells passing the 0.22 µm filter (that is, the <0.22 µm fraction) were prepared by the iron chloride precipitation method followed by filtration through a 0.8 µm filter based on John *et al.* (2011). The cells were concentrated with an Amicon Ultra-15 centrifugal device and DNA extracted with the Wizard PCR Preps DNA purification System (Promega, Fitchburg, WI, USA) (Supplementary Files 1 and 2). The 16S rRNA gene tag sequencing was carried out on an Illumina MiSeq (Illumina, San Diego, CA, USA), whereas community DNA was sequenced on an Illumina MiSeq and HiSeq from the >0.22 µm and <0.22 µm borehole fractions, respectively.

Bioinformatic analysis

Metagenome analysis was carried out using an in-house pipeline incorporating CONCOCT (Alneberg *et al.*, 2014) as described in the Supplementary File 1. The near-complete genomes are given in Supplementary File 3. Metagenome data were deposited in the National Center for Biotechnology Information (NCBI) database

(BioProject IDs PRJNA279923 (>0.22 μm) and PRJNA279924 (<0.22 μm)).

Results and discussion

Geochemical characteristics of the borehole fracture waters

All three fracture groundwaters were neutral (pH 7–8), carried iron entirely as Fe^{2+} , contained dissolved sulfide (HS^-), had temporally stable chemistry and $\delta^{18}\text{O}$, and based upon the presence of Fe^{2+} and HS^- the redox potential was low and dissolved oxygen would have been absent (Table 1). However, they differed in terms of several chemical constituents. Based on their major element chemistry and $\delta^{18}\text{O}$ values, the groundwaters could be characterized with regard to origin and age (Laaksoharju *et al.*, 2008).

SA1229A had high magnesium and potassium concentrations, which are tracers of marine waters in these settings (Gimeno *et al.*, 2014), and also had similar values for the conservative variables chloride and $\delta^{18}\text{O}$ as modern Baltic Sea water (Mathurin *et al.*, 2012). This groundwater was thus composed of infiltrated brackish marine (Baltic Sea) water and was termed ‘modern marine’ (metagenome bins

defined as ‘MM’). The precise infiltration age of the groundwater was not known, but was estimated to be <20 years and is probably even more recent (Mathurin *et al.*, 2014b). The origin and age of the KA3105A:4 groundwater cannot be defined in detail. It had relatively low chloride concentrations and thus was strongly influenced by fresh waters, such as glacial meltwater and/or modern meteoric water. It may also contain appreciable amounts of Baltic Sea water or Littorina Sea water. However, it also likely contained older waters as it was sampled at 415.19 m below ground surface. It was concluded that the KA3105A:4 groundwater was composed of a mixture of several different water types and was termed an ‘undefined mixed’ (metagenome bins defined as ‘UM’). KA3385A:1 had high chloride and calcium concentrations and relatively low $\delta^{18}\text{O}$ values. These features are typical for saline groundwater with a residence time in the order of thousands of years or more (Louvart *et al.*, 1999). However, the chloride concentration of this groundwater was considerably lower than for pure old saline water at the site (Mathurin *et al.*, 2012). Therefore, this groundwater had been diluted by waters with lower salinity including glacial meltwater from the retreat of the last Pleistocene continental ice sheet and marine water intruded during the Littorina Sea stage a few

Table 1 Concentrations of abiotic factors

	SA1229A (170.84 m) ^a		KA3105A_4 (415.19 m) ^b		KA3385A_1 (448.37 m) ^c	
	Median	Min–max	Median	Min–max	Median	Min–max
pH	7.3	7.0 to 7.4	7.5	7.5 to 7.6	7.5	7.0 to 7.9
Na	1545	1500 to 1670	1120	1030 to 1260	2200	1850 to 2430
Mg	134	120 to 155	82	67 to 92	67	62 to 73
K	29	25 to 33	12	9 to 14	10	9 to 12
Ca	260	238 to 315	342	281 to 390	1900	1680 to 2270
Cl^-	3017	2850 to 3381	2341	1684 to 2400	6819	6240 to 7560
SO_4^{2-}	274	224 to 312	223	217 to 229	378	346 to 417
DOC	6.9	4.8 to 8.1	6.8	6.4 to 7.5	1.4	0.8 to 3.6
HCO_3^-	255	241 to 266	217	203 to 227	25	10 to 31
$\delta^{18}\text{O}$	–7.2	–7.1 to –8.1	–9.0	–9.3 to –8.3	–10.8	–12.5 to –10.2
Fe total	1.8	1.5 to 2.1	0.42	0.33 to 0.45	0.21	0.16 to 0.29
Fe^{2+}	1.8	1.5 to 2.1	0.42	0.29 to 0.44	0.21	0.16 to 0.28
Mn	0.72	0.62 to 0.79	0.60	0.47 to 0.62	0.46	0.39 to 0.50
HS^-	0.02	0.005 to 0.127	0.127	0.054 to 0.399	0.05	bd to 3.26
$\text{NH}_4^+_{\text{N}}$	4.92	0.98 to 6.13	0.36	0.24 to 0.37	0.06	0.02 to 0.08
$\text{PO}_4^{3-}_{\text{P}}$	0.008	0.006 to 0.022	bd	bd to bd	bd	bd to 0.001
NO_3^-	bd	bd to bd	bd	bd to bd	bd	bd to 0.001
NO_2^-	0.0007	0.0005 to 0.0008	0.0005	0.0003 to 0.0005	0.0002	bd to 0.0003
Temp (°C)	12.4 ^d	NA	14.5 ^e	14.3 to 14.6	15.1 ^e	14.6 to 15.6

Abbreviations: bd, below detection limit; NA, not available.

Data for modern marine (SA1229A) and old saline (KA3385A:1) cover 1999–2014 and undefined mixed (KA3105A:4) cover 2009–2014. Concentrations are presented as mg l^{-1} , and the isotopic ratio ($\delta^{18}\text{O}$) as $^{18}\text{O}/^{16}\text{O}$ relative to Standard Mean Ocean Water. Descriptions of the analytical techniques and precision of the chemical and isotopic variables are given elsewhere: Cl^- and $\delta^{18}\text{O}$ (Mathurin *et al.*, 2012); Na, K, Ca, Mg and NH_4^+ (Mathurin *et al.*, 2014b); Fe^{2+} , Fe total, dissolved organic carbon (DOC) and HS^- (Alakangas *et al.*, 2014); NO_3^- and NO_2^- (Rönnback and Åström 2007); and Mn, SO_4^{2-} , HCO_3^- and PO_4^{3-} (Mathurin *et al.*, 2014a). The data were retrieved on 17 October 2014 (Sicada-14-047) except temperature (Temp) that was based on values taken in 2011.

^aThree to 10 measurements.

^bThree to 6 measurements.

^cSeven to 31 measurements.

^dEstimated value taken from the nearby KA1755A borehole.

^eAverage of two data points.

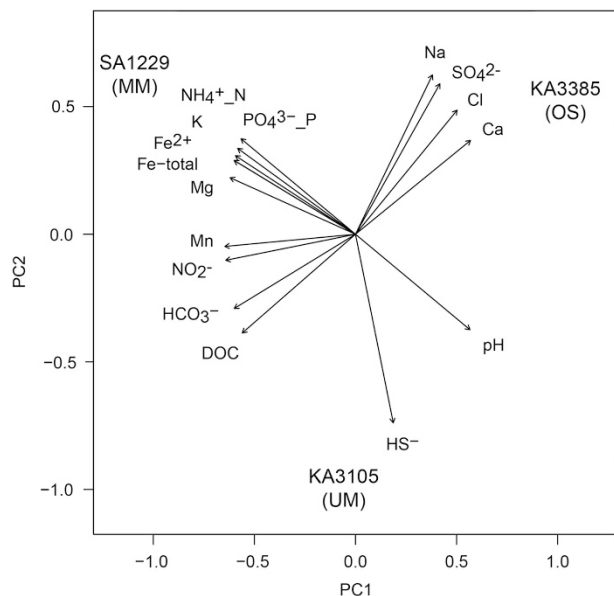


Figure 1 Principle component analysis of the chemical components from the three water types, modern marine (SA1229A), undefined mixed (KA3105A:4) and old saline (KA3385A:1).

thousands of years ago (Mathurin *et al.*, 2012). This groundwater was termed ‘old saline’ (metagenome bins defined as ‘OS’).

A principal component analysis (Figure 1) of chemical characteristics for the three water types showed that dissolved organic carbon was associated with both the modern marine and undefined mixed water types. This organic carbon is regarded to be dominated by reduced humic substances, consisting of humic and fulvic acids in variable proportions (Mathurin *et al.*, 2014a) that may be a source of electron donors for heterotrophic species. The higher amounts of dissolved organic carbon in the modern marine water was supported by cell counts showing lower cell numbers in the undefined mixed and old saline waters (127 cells per ml in the modern marine water compared with 37 and 100 cells per ml in the undefined mixed and old saline waters, respectively). These microorganisms may have been growing by nitrate reduction to nitrite and/or ammonia in both water types, as indicated by the elevated levels of nitrite in both waters and ammonium in the modern marine water. The source of oxidized nitrogen is unknown but based upon community mRNA transcript sequencing it has been hypothesized to be formed in a subsurface sediment as a by-product of anaerobic ammonium oxidation (Orsi *et al.*, 2013). In addition, previous studies at the Äspö HRL have identified nitrate/nitrite-reducing species (Nielsen *et al.*, 2006; Pedersen 2013; Ionescu *et al.*, 2015a, b). Elevated levels of hydrogen sulfide suggested that microbial sulfate reduction might be a more dominant form of respiration in the undefined mixed water compared with the modern marine water. In contrast, increased levels of ferrous iron in the modern marine water suggested anaerobic

ferric iron reduction may be a growth strategy in this water mass. However, homologs may not have been identified as the gene(s) encoding for ferric reductase has only been identified in a few species, such as from the *Geobacter* and *Shewanella* genus (Bird *et al.*, 2011), or because soluble redox mediators were used. The lack of chemical ions associated with microbial metabolism in the old saline water probably reflected a lower biomass, as supported by the lower amounts of DNA recovered from this water (Supplementary File 2).

DNA extraction and metagenome sequencing data

The volume of the different water types filtered, amount of DNA extracted and details of metagenome sequencing are provided in Supplementary File 2. Details of each approved phylogenetic bin are given in Supplementary File 4. Many of the cells in the tested Äspö HRL water types passed a 0.22 µm filter and electron microscopy confirmed their small cell size (Supplementary File 5). There was little taxonomic overlap between these small cells (<0.22 µm metagenome; metagenome bins defined as ‘S’) and those that were retained (>0.22 µm metagenome; metagenome bins defined as ‘L’) (Supplementary File 6). For duplicate metagenomes, there was general agreement between the relative percentages of the mapped reads assigned to phylogenetically closely related species (Supplementary Files 4 and 6). For instance, MMS_A4 and MMS_B4 constituted 24.1% and 22.8% of the mapped reads in their respective metagenomes.

Microbial community from metagenome binning

The deep biosphere is suggested to have cell turnover rates ranging from hundreds to thousands of years (Hoehler and Jorgensen, 2013) and as a consequence it is extremely difficult to describe processes by direct measurements. With culture-independent molecular methods, these limitations can be at least partly overcome, but only very recently have there been attempts to use such information to generate hypotheses regarding the potential metabolic capacity of individual populations within complex communities (Wrighton *et al.*, 2012). To understand how populations interact and collectively perform ecosystem processes, such information is critical. Here we partition community metagenomes into individual populations (Figure 2) to provide hypotheses on the biology and metabolic strategies (Figure 3) as well as potential interactions between populations (Figure 4) within three water types in the terrestrial deep subsurface.

The phylogeny of the dominant microbial community members from the three water types was reconstructed from nearly complete assembled population genomes (Figure 2 and Supplementary File 7). *Proteobacteria* were retrieved in all three water types, and for both size fractions but at higher

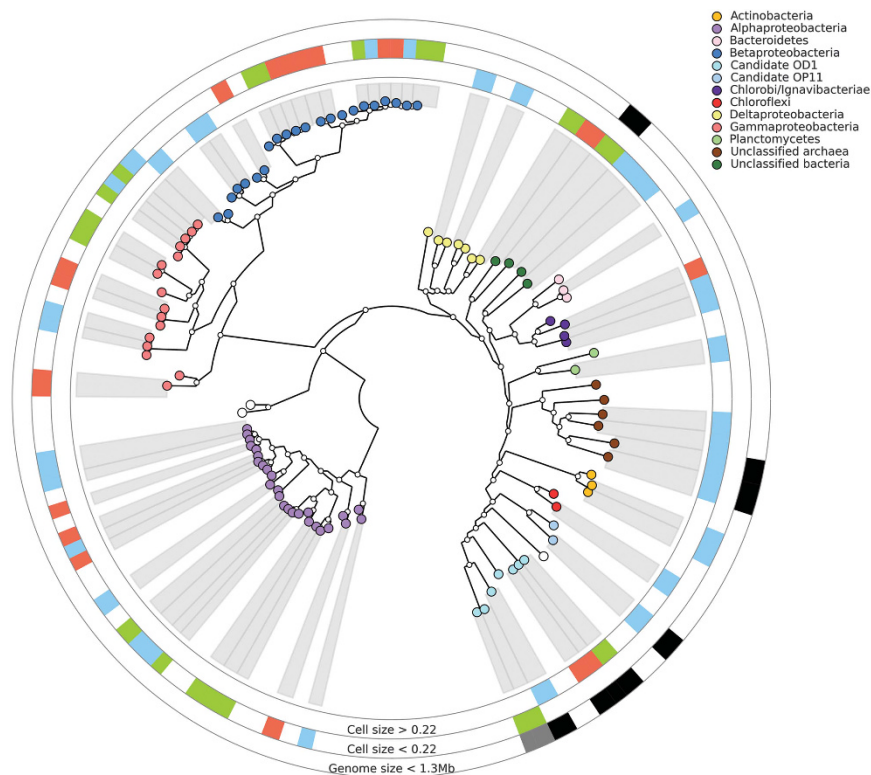


Figure 2 Whole-genome phylogenetic tree showing the CONCOCT bins from the duplicate metagenomes sequenced from the three water types. The first and second rings show $>0.22\ \mu\text{m}$ and $<0.22\ \mu\text{m}$ metagenomes, respectively, with the color coding: modern marine (orange), undefined mixed (green) and old saline (blue). The third ring shows estimated genome sizes $<1.3\ \text{Mb}$ in $>0.22\ \mu\text{m}$ (black) and $<0.22\ \mu\text{m}$ (gray) metagenomes. The clades labeled with large white circles correspond to *Sulfurovum* sp. SCGC_AAA036-F05 and *Sulfurimonas* sp. GD1 (left side of figure) and *Candidatus Saccharobacterium alaburgensis* (right side of figure).

taxonomic resolution, clearly partitioned communities were observed. At the class level, α - and γ -*Proteobacteria* were exclusively found in the $<0.22\ \mu\text{m}$ cell size, whereas δ -*Proteobacteria* were only detected in the $>0.22\ \mu\text{m}$ size fraction. This division in cell sizes between the water types at the class level was also generally supported by the 16S rRNA gene tag sequencing (data not shown). The microbial communities were also partitioned by water type, where *Actinobacteria* were only identified from the $<0.22\ \mu\text{m}$ cell size old saline water. Finally, archaea were recovered from the $>0.22\ \mu\text{m}$ size fraction old saline metagenome, but were only distantly related to any genome in the NCBI or PATRIC database. Despite the populations in this study being from pH-neutral groundwaters, the closest match was to Archaeal Richmond Mine acidophilic nanoorganisms *Candidatus Micrarchaeum acidiphilum* ARMAN-2 previously identified from an extremely acidic sulfide mineral mine (Dick *et al.*, 2009) (Supplementary File 7).

The metagenomes from the three water types were partitioned into draft genome bins using CONCOCT (Alneberg *et al.*, 2014) that utilizes 36 single copy housekeeping genes to estimate the genome coverage. The presence of ≥ 31 of the single copy genes (suggesting a statistical average of $\geq 86\%$ coverage of the genome) and no more than two duplicates of the

same gene were considered a near-complete genome bin. The near-complete genome bins, representing populations, were searched for genes coding for metabolic pathways involved in electron donor utilization, electron acceptors, ability to fix carbon dioxide and nitrogen and uptake of nutrients (Table 2 and Supplementary File 8). These data were then used to identify the different suggested potential growth strategies and putative population interactions within communities from the three deep biosphere water types.

Metabolism in modern marine water

The $>0.22\ \mu\text{m}$ metagenome had the highest relative percentage of reads that mapped to the bin MML_A2 that was related to *Chlorobi/Ignavibacteriae* (5.4%; all quoted values in the sections below refer to the percentage of mapped reads from the respective metagenome). This population aligned most closely to the fermentative species *Ignavibacterium album* and potentially grew via fermentation or anaerobic hydrogen and ethanol oxidation coupled to electron transport via the ferredoxin:NAD⁺ oxidoreductase Rnf complex. The recently discovered Rnf complex (Schuchmann and Muller, 2014) appears to be used in many anaerobic microorganisms to couple the oxidation of organic carbon or hydrogen to NAD⁺

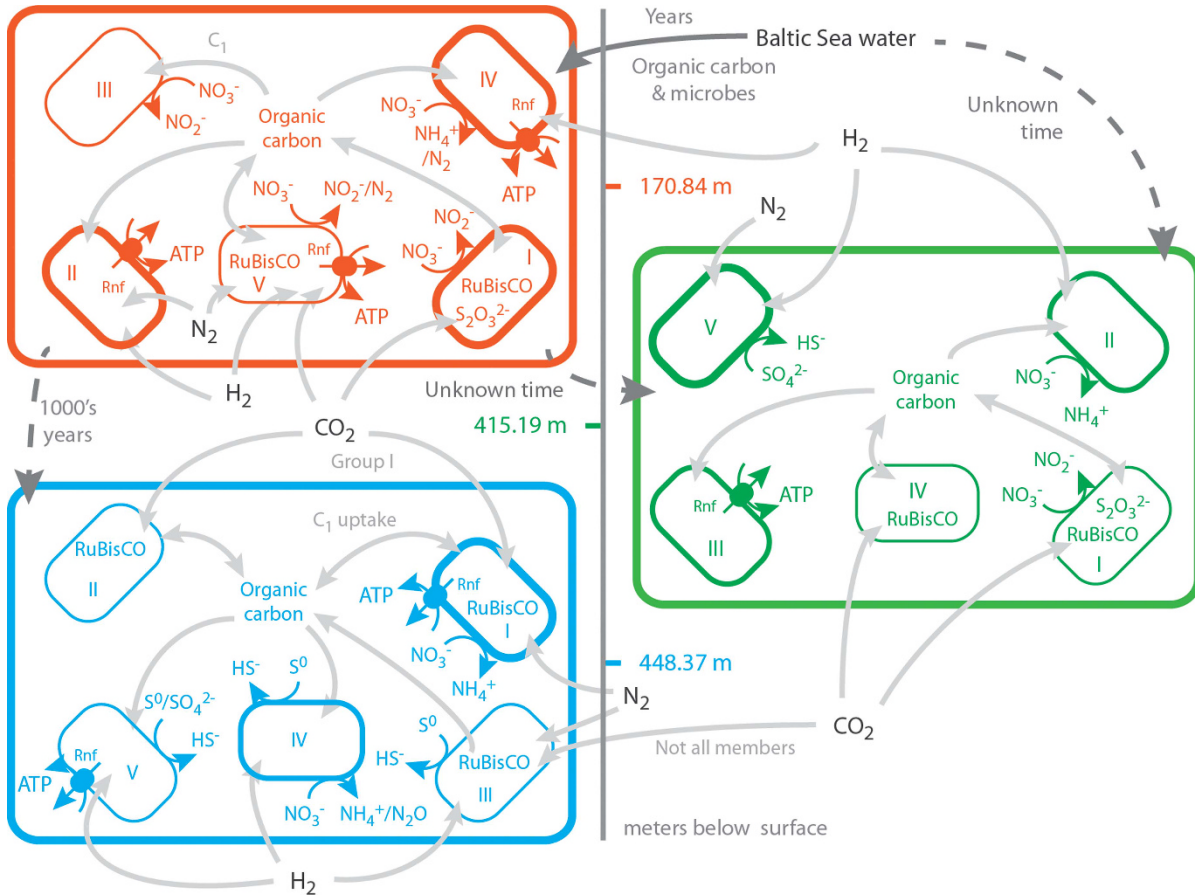


Figure 3 Model of potential metabolic pathways in the modern marine (orange), undefined mixed (green) and old saline (blue) waters. The Roman numerals refer to the groups defined in Table 2. The figure shows recharge of the modern marine water from the Baltic Sea (providing organic carbon and microorganisms) and the gases hydrogen and carbon dioxide. The cell wall thickness represents <10% (thinnest line), 10–50% (medium line) and >50% (thickest line) of the average percent mapped reads divided by the total percentage of mapped reads in all bins from the respective duplicate metagenomes (Table 2; making a total of 200%: 100% for the <0.22 μm cells plus 100% for the >0.22 μm cells).

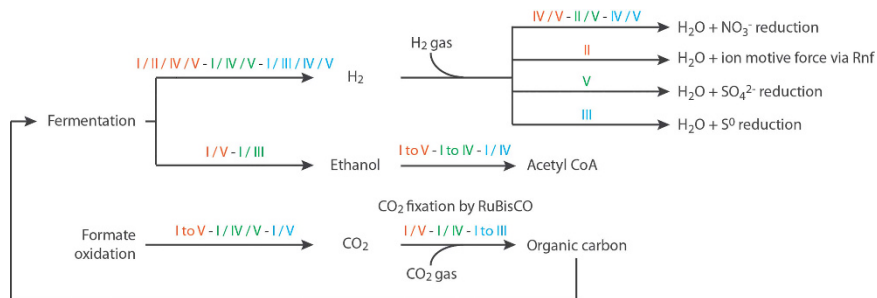


Figure 4 Model of potential metabolic interactions and dependencies among populations in the deep biosphere modern marine (orange), undefined mixed (green) and old saline (blue) waters. The Roman numerals refer to the groups defined in Table 2.

(via ferredoxin) for adenosine triphosphate production at redox potentials below -320 mV. A second phylogenetic population was represented by MML_B2 (most related to *Candidatus Endomicrobium* sp.) that also encoded genes assigned to anaerobic hydrogen oxidation and the electron transport Rnf complex. Both of these populations have a reversible type III anaerobic, NADP-dependent hydrogenase that we suggest functions for hydrogen oxidation as the populations also

contain the Rnf complex. An example of potential metabolic interactions (Figure 4) is hydrogen produced as a fermentation product that can subsequently be utilized by a hydrogen oxidizing population, thus creating a cryptic hydrogen pathway, whereby any produced hydrogen would likely be immediately consumed. MML_A1 and MML_B1 (2.2% and 1.3%, respectively) from the replicate metagenomes both grouped with Candidate division OD1. Based on the number of identified single copy

Table 2 Key metabolic characteristics of the dominant populations from the three water types

Group	Cell size ^a	Bins	Metabolic pathways	Respiration/electron transport	Nutrient fixation	% Of total ^b
<i>Modern marine</i>						
I	<	MMS_A2/MMS_B2	Fermentation and formate oxidation	Nitrate reduction	CO ₂	32.5/35.0
	<	MMS_A5/MMS_B5				2.0/3.2
II	>	MML_A2/MML_B2	Fermentation and H ₂ oxidation	(Rnf) ^c	(N ₂)	70.9/27.6
	<	MMS_A6/MMS_A8				1.7/1.3
III	<	MMS_A3/MMS_B3	Fermentation; (ethanol), formate and methane oxidation	Nitrate reduction	None ^d	4.3/3.5
IV	<	MMS_A4/MMS_B4	Fermentation; ethanol, formate and H ₂ oxidation	Nitrate reduction, denitrification and Rnf	None	53.0/55.1
V	<	MMS_A7	Fermentation; ethanol, formate and H ₂ oxidation	Nitrate reduction and Rnf	CO ₂ and N ₂	2.7
<i>Undefined mixed</i>						
I	<	UMS_A2/UMS_B2	(Fermentation); (ethanol) and (formate) oxidation	Nitrate reduction	CO ₂	6.6/4.6
II	<	UMS_A1/UMS_B1	Fermentation; ethanol, formate and H ₂ oxidation	Nitrate reduction	None	36.7/28.0
III	<	UMS_A3/UMS_B3	Fermentation and ethanol oxidation	Rnf complex	None	28.9/20.6
IV	<	UMS_B9	Fermentation; ethanol, formate and H ₂ oxidation	None	CO ₂	4.9
V	>	UML_B1	Fermentation; formate and H ₂ oxidation	Sulfate reduction and nitrate reduction	N ₂	89.8
<i>Old saline</i>						
I	>	OSL_A1/OSL_B1	Fermentation; ethanol, formate and methane oxidation	Nitrate reduction, denitrification and Rnf	CO ₂ and N ₂	20.4/89.8
	>	OSL_B3				4.2
II	>	OSL_A4/OSL_A12	Fermentation	None	CO ₂	3.2/1.1
III	>	OSL_A5/OSL_A9	Fermentation and H ₂ oxidation	Sulfur reduction	(CO ₂)	3.1/4.2
IV	>	OSL_A2/OSL_B2	Fermentation; (ethanol oxidation), (formate oxidation) and H ₂ oxidation	Nitrate reduction; (sulfur reduction); denitrification to N ₂ O	None	26.0/4.6
V	<	OSS_A4/OSS_B4	Fermentation; formate oxidation and H ₂ oxidation	Sulfate reduction; (nitrate reduction); & Rnf	None	19.6/20.6
	>	OSL_A7/OSL_A10				7.8/10.4

Detailed information is provided in Supplementary File 8.

^aCell size > 0.22 µm or < 0.22 µm.

^bMapped percentage of reads for each bin divided by the total percentage of mapped reads in all bins from the respective duplicate metagenomes.

^cProcesses in brackets denote that not all populations in the group possess the genes.

^dNone, no gene homologs for complete pathways detected.

genes, these genome bins were estimated to have >86% genome coverage (Supplementary File 4). Despite the high estimated genome coverage, only genes coding for fermentation of simple organic carbon compounds could be identified (for example, the cells appeared to lack tricarboxylic acid cycle or electron transport components). Neither carbon dioxide nor nitrogen fixation could be identified from the >0.22 μm metagenome bins.

The <0.22 μm metagenome for the same water type generated 13 bins (totaling 41.5% and 41.3% in the replicate metagenomes) representing populations that were predicted to ferment pyruvate to ethanol, propionate, lactate or hydrogen. Further possible electron donors included ethanol that could be degraded to acetyl-CoA (via pathways I to IV) that was present in the majority of the reconstructed genomes. Methane is also a potential electron donor via anaerobic oxidation coupled to nitrate reduction (Haroon *et al.*, 2013; Orcutt *et al.*, 2013) and the phylogenetically similar MMS_A3 and MMS_B3 (both related to *Mesorhizobium alhagi*) that contains genes assigned as the respiratory nitrate reductase *napAB* (2.0% and 1.4%, respectively) and conversion of methane to methanol. However, the encoded proteins may also mediate reversed methanogenesis. The final step in conversion of one carbon (C_1) compounds to carbon dioxide is catalyzed by formate dehydrogenase and homologs were identified in 13 populations (45.5% and 41.0%), suggesting these species were able to grow utilizing C_1 compounds. A potential respiratory electron acceptor was nitrate that was converted to nitrite (seven bins totaling 18.9% and 17.2%), ammonia (four bins totaling 25.2% and 23.8%) or nitrogen gas (24.1% and 22.8%) as the end products. This was in agreement with the chemical signature of the marine water mass, where ammonia and nitrite concentrations were elevated (Figure 1). Furthermore, five bins (16.9% and 15.8%) were suggested to fix carbon dioxide via the Calvin–Benson–Bassham (CBB) cycle and two bins (24.1% and 24.1%) to fix nitrogen. No evidence was identified in any of the water types for other carbon dioxide fixation pathways than the energetically expensive CBB cycle, suggesting alternative autotrophic processes were not dominant in the communities. Surprisingly, this included the Wood–Ljungdahl pathway that operates in conjunction with the Rnf complex in acetogenic microorganisms for autotrophic growth in thermodynamically limited environments (Schuchmann and Muller, 2014). It was also surprising that genes supporting the presence of the reductive tricarboxylic acid cycle and reductive acetyl CoA pathway were not identified as these methods of carbon dioxide fixation are more prevalent in anaerobic microorganisms (Hugler *et al.*, 2003). The Rnf complex has additional roles such as during anaerobic nitrogen fixation that may be occurring in some populations, such as MMS_A7, or to create an ion motive force for transport across the membrane (Schuchmann and Muller, 2014).

Finally, the presence of gene homologs in several populations to oxidize thiosulfate and reduce nitrate along with CBB genes for carbon dioxide fixation (for example, MMS_B2) suggested the potential to grow chemoautotrophically.

The current paradigm for life in the deep terrestrial biosphere points toward an autotrophic system fueled mainly by hydrogen of geological origin (Hallbeck and Pedersen, 2008). Our data partially support this model but also point to an important role of heterotrophic growth via organic carbon potentially recharged from the Baltic Sea over years to decades. This results in many abundant community members that seem to rely exclusively on fermentation.

Metabolism in undefined mixed water

Three near-complete genome bins were reconstructed from the undefined mixed water >0.22 μm metagenome. Two of these bins were phylogenetically most related to Candidate division OD1 (UML_A1) and unclassified bacteria (UML_B2). However, these bins only represented a minor portion of the community as they contributed 1.4% and 0.3% of the mapped reads, respectively. These populations also had a reduced genome size and we suggest they have a similar, simple fermentative growth strategy as described for MML_A1 and MML_B1 from the modern marine water. The third bin, UML_B1 has the suggested metabolic strategy of fermenting pyruvate to propionate, anaerobically oxidizing hydrogen, and oxidizing formate to carbon dioxide. Bin UML_B1 was also suggested to utilize nitrate reduction to nitrite and sulfate to sulfide during anaerobic respiration and fix nitrogen for cellular growth. The ability to reduce nitrate was supported by an elevated concentration of nitrite in this water type, analogous to the modern marine water. Hydrogen sulfide was strongly elevated in the undefined mixed water, despite that only a single phylogenetic bin was identified containing genes coding for the dissimilatory sulfite reductase. This may be explained by either a disproportionately high sulfate reduction activity of this population compared with sulfate reducers in the other waters (that novel, unidentified sulfate-reducing mechanisms were present) or that genomes from less abundant sulfate-reducing populations could not be assembled.

A total of 15 near-complete bins were constructed from the undefined mixed <0.22 μm metagenome with a relatively uniform abundance distribution and altogether contributing 9.1% and 12.8% of the mapped reads. Of the 15 reconstructed bins, 9 were suggested to be able to ferment pyruvate to ethanol (3 bins; 3.2% and 2.6%), propionate (5 bins; 2.0% and 5.5%), lactate (6 bins; 7.4% and 4.8%) or hydrogen (2 bins; 0% and 4.2%). Eight bins had genes assigned to mediate ethanol degradation to acetyl-CoA (6.6% and 9.8%) and a further three bins

were suggested to be able to further ferment the acetyl-CoA to butyrate (2.6% and 3.3%). Four of the undefined mixed populations were suggested to utilize formate. None of the bins were suggested to reduce sulfate or sulfur, but five bins were suggested to reduce nitrate to either nitrite (0.6% and 1.9%) or ammonia (3.4% and 3.6%). The phylogenetically similar UMS_A3 and UMS_B3 (2.6% and 2.6%) were most similar to *Marinomonas* sp. and were suggested to generate an ion motive force via the Rnf complex during heterotrophic growth. Only two populations contained genes assigned to the CBB cycle for carbon dioxide fixation (UMS_A2 and UMS_B2 related to *Limnobacter* sp.; both 0.6%), suggesting these species may have been sufficiently active to support the community or that novel carbon dioxide pathways were not identified. The ability to oxidize the inorganic sulfur compound thiosulfate might also allow the nitrate-reducing populations to grow chemoautotrophically.

Metabolism in old saline water

A total of 17 bins were identified from the old saline water >0.22 μm metagenome (totaling 23.8% and 64.3%). One of the >0.22 μm replicate metagenomes was dominated by population OSL_B1 that was most similar to *Dechloromonas aromatica* and contributed 57.7% of the mapped reads. For the replicate metagenome, the phylogenetically related population (OSL_A1) merely contributed 4.9% of the mapped reads. Gene homologs in these bins implied a capacity to utilize several organic carbon sources including pyruvate, acetyl-CoA, ethanol, formate and methane. A potential anaerobic terminal electron sink was nitrate reduction to ammonia or nitrogen gas coupled to generation of an ion motive force via the Rnf complex. These two populations were also suggested to fix nitrogen and carbon dioxide. The other bins in the >0.22 μm metagenome also contained populations suggested to ferment pyruvate and acetyl-CoA, degrade ethanol, oxidize formate and convert methane into methanol. In addition to the two bins most similar to *D. aromatica*, a further eight populations were suggested to utilize nitrate and/or sulfate as terminal electron acceptor. Despite the presence of potential sulfate-reducing bacteria, the principal component analysis suggested the old saline water was enriched with sulfate. This may be because of insufficient energy supply (electron donor inputs) to effectively drive sulfate reduction. Alternatively, the bacterial (OSL_A10 and OSL_A11) and archaeal (OSL_A2, OSL_A5, and OSL_A9) populations reduced the sulfur (but not sulfate) to sulfide. Three populations (4.9% and 60.5%) were suggested to be able to denitrify to the level of nitrous oxide or nitrogen gas that could also fix nitrogen. Eight of the old saline >0.22 μm metagenome bins contained homologs for hydrogen oxidation (13.5% and 3.0%), whereas 6 populations contained homologs for carbon dioxide

fixation via the CBB cycle (6.9% and 60.5%), potentially suggesting a greater dependence on the gases hydrogen and carbon dioxide from geological origin to support growth in the deeper and older fracture waters. Finally, a population most similar to Candidate division OP11 was identified that, similar to the other detected candidate divisions, had a fermentative metabolism while lacking the tricarboxylic acid cycle and ability to respire.

The <0.22 μm old saline metagenome contained 16 populations totaling 11.9% and 16.3% of the mapped reads. Of these 16 populations, 13 were suggested to ferment pyruvate to various final electron acceptors (9.5% and 14.1%); 11 had the potential to degrade ethanol (10.8% and 13.7%); 7 could possibly perform formate oxidation (3.8% and 7.5%); and 1 population was suggested to be able to oxidize methane to methanol. The potential final electron acceptor was nitrate that was reduced to nitrite, ammoni, or nitrous oxide by, for example, OSS_A4 and OSS_B4 or sulfur reduction to sulfate by OSS_A6 and OSS_B6. In addition, three bins (all γ -*Proteobacteria*; 0.7% to 1.9%) were suggested to be able to generate an ion motive force via the Rnf complex.

In contrast to the modern marine water, the deep old saline water more closely adheres to the present metabolic model with populations growing via chemolithotrophic processes. These putative autotrophs also include unclassified and poorly understood archaeal groups, in agreement with the tendency for archaea to become more prevalent as the depth below the surface increases (Hoehler and Jorgensen, 2013). Fermentative populations from candidate divisions constituted 3.5% of the mapped reads in the modern marine water compared with 0.8% in the old saline water, further pointing to the importance of heterotrophy in the younger water versus autotrophy in the old saline water. The old saline water populations were also suggested to have a greater tendency to sulfate/sulfur reduction than the shallower water types.

Potential adaptations to extreme oligotrophy

One feature of the communities was the apparent existence of extremely small microbes that passed the 0.22 μm membrane filter, and electron microscope observations indeed supported the presence of very small cells (Supplementary File 5). These small cells included genome bins assigned to the α -, β - and γ -*Proteobacteria* (Figure 2). The cell sizes of all the phylogenetically closest published type species related to the populations in the <0.22 μm metagenomic bins were >0.3 \times 1.0 μm . Because of the small cell size, low cell density and consequent challenges in microscopic analyses, the portion of small cells (<0.22 μm) were estimated with a PCR and sequencing approach. The abundance of operational taxonomic units that were retained by a 0.22 μm filter (designated as large cells) and those that were

retained by a 0.1 μm filter but not present in the large cells (designated as small cells) was quantitatively compared (Supplementary Files 1 and 9). These data suggested a median percentage of 47%, 54% and 50% of the modern marine, undefined mixed and old saline communities were made up of cells that were unique to the fraction that passed the 0.22 μm filter. Because of the 0.1 μm filter size being used to select for the small cells (as compared with iron chloride precipitation), the estimated percentage of small cells may be conservative. Although these PCR-based estimates are associated with some uncertainty, the values of $\sim 50\%$ small cells in the three water types warrants further investigation of their role in the deep terrestrial biosphere. An abundance of microorganisms with small cell sizes have recently been described for microbes in shallow aquifers (Luef *et al.*, 2015), but these are the first data pointing to their existence in the terrestrial deep biosphere.

Another potential adaptation to oligotrophy is reduced genome size (Swan *et al.*, 2013; Giovannoni *et al.*, 2014). Although most of the reconstructed genomes were >1.3 Mb, a few streamlined genomes were identified. These included bins assigned to Candidate division OD1 found in the modern marine and undefined mixed waters along with Candidate division OP3 found in the undefined mixed water (Figure 2). Comparing the estimated size of the reconstructed genomes to the closest matching reference genomes in the databases revealed that overall 63% of the genomes in the <0.22 μm fraction were reduced in comparison with the reference genomes in the NCBI database. One example was for the replicate populations OSS_A5 and OSS_B5 that both had genomes 37% smaller than the *Microbacterium barkeri* reference genome. However, there did not appear to be a trend of smaller genome sizes with depth. As a reference, the sizes of reconstructed genomes in the >0.22 μm metagenomes were on average larger than the closest matching reference genomes in the NCBI database.

Assessment of potential sample contamination

Many potential sources of contamination in groundwater communities are associated with drilling and deep biosphere sampling. The Äspö HRL circumvents the majority of these issues because: (1) the Äspö HRL tunnel was constructed decades ago and the studied fissures are reached by boreholes from the tunnel wall and are thus far away from an oxidizing environment and (2) the water from the boreholes flow into the tunnel, where the sampling is carried out, by gravity and therefore water-pumping activities that can induce a variety of contamination and sampling errors are avoided. However, the addition of materials to boreholes in order to enclose the targeted fissure (called 'sections') can be a source of electron donors for the microbes (Drake *et al.*, 2015). For this reason, three to five section water

volumes were drained from the borehole to allow pristine groundwater to flow. Recently, Salter *et al.* (2014) presented an analysis of the impact of contamination of DNA derived from laboratory chemicals in sequence-based community analyses. This should not be a major issue in the present study as the recovered amount of DNA was high (0.17 to 2.10 μg). Nevertheless, an evaluation of potential contamination was performed, showing no evidence of reagent contamination (Supplementary Files 10 and 11). In addition, in a different project with samples prepared with the same protocol, laboratory chemicals and in the same laboratory as the <0.22 μm fraction for the data reported here, none of the taxa assigned to the bins in this study were detected among contigs longer than 100 kb (data not shown). This was consistent with sample preparation or reagents used not significantly contributing to the recovered DNA from the boreholes.

Conclusions

One striking feature for all three water types was the potential ability of the major populations to use a plethora of metabolic processes to sustain their energy and nutrient requirements. For example, several populations were suggested to be able to grow mixotrophically including the modern marine group I (Figure 3) that potentially ferments organic carbon, couples inorganic sulfur compound oxidation to nitrate reduction and fixes carbon dioxide via the CBB cycle. An exception to this was several poorly understood microbial taxa associated with candidate division populations that were suggested to grow via fermentation of simple organic carbon compounds or amino acids. These candidate division populations show a greater dependency on heterotrophy in these waters than previously reported.

The metabolic model of the dominant populations in the microbiome at 171 m below the surface of the earth at the Äspö HRL has a greater dependency on organic carbon than the paradigm of an autotrophic, hydrogen-driven environment. This organic carbon could potentially be provided from the surface waters of the Baltic Sea. In contrast, the dominant microbial community in the deeper and more anciently formed saline waters appears to be more extensively maintained by chemolithoautotrophic processes. The ability to reduce sulfate (or sulfur) by known pathways was primarily found in the deepest, old saline waters but was less widespread in the modern marine water compared with previous models of the Fennoscandian deep biosphere based upon growth experiments and 16S rRNA gene-based methods (Itavaara *et al.*, 2011; Pedersen 2012, 2013). Furthermore, the dominant populations have no genomic signal for ferric or manganese reduction or methanogenesis, despite that these microorganisms have been cultured from the Äspö HRL (Hallbeck

and Pedersen, 2008) and that key genes in methanogenesis have been amplified from the Outokumpu deep borehole waters (Purkamo *et al.*, 2015).

Finally, the Äspö HRL microbial communities included cells with a small cell size that also had a tendency to have smaller genomes than their closest sequenced relatives. This may be a physiological adaptation to life in highly oligotrophic deep biosphere groundwaters. A consequence of their small size makes these cells likely to have been overlooked in earlier studies relying on membrane capture.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

We acknowledge the Swedish Research Council (contracts 621-2014-4398 and 2012-3892) and SKB; MD and KH thank The Crafoord Foundation (contract 20130557); and MD thanks the Nova Center for University Studies, Research and Development and Familjen Hellmans stiftelse for financial support. Sequencing was carried out at the National Genomics Infrastructure hosted by Science for Life Laboratory. Bioinformatics was carried out utilizing the Uppsala Multidisciplinary Center for Advanced Computational Science resource at Uppsala University (project b2013127). Support by Ino de Bruijn and Anders Andersson is gratefully acknowledged under the auspices of BILS (Bioinformatics Infrastructure for Life Sciences) for bioinformatic identification and separation of the bacterial fraction of the <0.22 micrometer metagenomes.

References

- Alakangas LJ, Mathurin FA, Faarinen M, Wallin B, Åström ME. (2014). Sampling and characterizing rare earth elements in groundwater in deep-lying fractures in granitoids under in situ high-pressure and low-redox conditions. *Aquat Geochem* **20**: 405–418.
- Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ *et al.* (2014). Binning metagenomic contigs by coverage and composition. *Nat Methods* **11**: 1144–1146.
- Bird LJ, Bonnefoy V, Newman DK. (2011). Bioenergetic challenges of microbial iron metabolisms. *Trends Microbiol* **19**: 330–340.
- Chivian D. (2008). Environmental genomics reveals a single-species ecosystem deep within earth. *Science* **322**: 275–278.
- Dick GJ, Andersson AF, Baker BJ, Simmons SL, Thomas BC, Yelton AP *et al.* (2009). Community-wide analysis of microbial genome sequence signatures. *Genome Biol* **10**: R85.
- Dong Y, Kumar CG, Chia N, Kim PJ, Miller PA, Price ND *et al.* (2014). Halomonas sulfidaeris-dominated microbial community inhabits a 1.8 km-deep subsurface Cambrian Sandstone reservoir. *Environ Microbiol* **16**: 1695–1708.
- Drake H, Tullborg E-L, Whitehouse M, Sandberg B, Blomfeldt T, Åström ME. (2015). Extreme fractionation and micro-scale variation of sulphur isotopes during bacterial sulphate reduction in deep groundwater systems. *Geochim Cosmochim Acta* **161**: 1–18.
- Gimeno MJ, Auqué LF, Acero P, Gómez JB. (2014). Hydrogeochemical characterisation and modelling of groundwaters in a potential geological repository for spent nuclear fuel in crystalline rocks (Laxemar, Sweden). *Appl Geochem* **45**: 50–71.
- Giovannoni SJ, Cameron Thrash J, Temperton B. (2014). Implications of streamlining theory for microbial ecology. *ISME J* **8**: 1553–1565.
- Gold T. (1992). The deep, hot biosphere. *Proc Natl Acad Sci USA* **89**: 6045–6049.
- Hallbeck L, Pedersen K. (2008). Characterization of microbial processes in deep aquifers of the Fennoscandian Shield. *Appl Geochem* **23**: 1796–1819.
- Hallbeck L, Pedersen K. (2012). Culture-dependent comparison of microbial diversity in deep granitic groundwater from two sites considered for a Swedish final repository of spent nuclear fuel. *FEMS Microbiol Ecol* **81**: 66–77.
- Haroon MF, Hu S, Shi Y, Imelfort M, Keller J, Hugenholtz P *et al.* (2013). Anaerobic oxidation of methane coupled to nitrate reduction in a novel archaeal lineage. *Nature* **500**: 567–570.
- Hoehler TM, Jorgensen BB. (2013). Microbial life under extreme energy limitation. *Nat Rev Micro* **11**: 83–94.
- Hugler M, Huber H, Stetter KO, Fuchs G. (2003). Autotrophic CO₂ fixation pathways in archaea (Crenarchaeota). *Arch Microbiol* **179**: 160–173.
- Ionescu D, Heim C, Polerecky L, Thiel V, de Beer D. (2015a). Biotic and abiotic oxidation and reduction of iron at circumneutral pH are inseparable processes under natural conditions. *Geomicrobiol J* **32**: 221–230.
- Ionescu D, Heim C, Thiel V, Ramette A, Reitner J, de Beer D. (2015b). Diversity of iron oxidizing and reducing bacteria in bioreactors set in the Äspö Hard Rock Laboratory. *Geomicrobiology* **32**: 207–220.
- Itavaara M, Nyssonson M, Kapanen A, Nousiainen A, Ahonen L, Kukkonen I. (2011). Characterization of bacterial diversity to a depth of 1500 m in the Outokumpu deep borehole, Fennoscandian Shield. *FEMS Microbiol Ecol* **77**: 295–309.
- John SG, Mendez CB, Deng L, Poulos B, Kauffman AK, Kern S *et al.* (2011). A simple and efficient method for concentration of ocean viruses by chemical flocculation. *Environ Microbiol Rep* **3**: 195–202.
- Jorgensen BB. (2011). Deep seafloor microbial cells on physiological standby. *Proc Natl Acad Sci USA* **108**: 18193–18194.
- Kallmeyer J, Pockalny R, Adhikari RR, Smith DC, D'Hondt S. (2012). Global distribution of microbial abundance and biomass in seafloor sediment. *Proc Natl Acad Sci USA* **109**: 16213–16216.
- Laaksoharju M, Gascoyne M, Gurban I. (2008). Understanding groundwater chemistry using mixing models. *Appl Geochem* **23**: 1921–1940.
- Lin LH, Wang PL, Rumble D, Lippmann-Pipke J, Boice E, Pratt LM *et al.* (2006). Long-term sustainability of a high-energy, low-diversity crustal biome. *Science* **314**: 479–482.
- Louvat D, Michelot JL, Aranyossy JF. (1999). Origin and residence time of salinity in the Äspö groundwater system. *Appl Geochem* **14**: 917–925.

- Luef B, Frischkorn KR, Wrighton KC, Holman H-YN, Birarda G, Thomas BC *et al.* (2015). Diverse uncultivated ultra-small bacterial cells in groundwater. *Nat Commun* **6**: 6372.
- Mathurin FA, Astrom ME, Laaksoharju M, Kalinowski BE, Tullborg EL. (2012). Effect of tunnel excavation on source and mixing of groundwater in a coastal granitoidic fracture network. *Environ Sci Technol* **46**: 12779–12786.
- Mathurin FA, Åström ME, Drake H, Maskenskaya OM, Kalinowski BE. (2014a). REE and Y in groundwater in the upper 1.2 km of Proterozoic granitoids (Eastern Sweden) – assessing the role of composition and origin of groundwaters, geochemistry of fractures, and organic/inorganic aqueous complexation. *Geochim Cosmochim Acta* **144**: 342–378.
- Mathurin FA, Drake H, Tullborg E-L, Berger T, Peltola P, Kalinowski BE *et al.* (2014b). High cesium concentrations in groundwater in the upper 1.2 km of fractured crystalline rock – influence of groundwater origin and secondary minerals. *Geochim Cosmochim Acta* **132**: 187–213.
- McMahon S, Parnell J. (2014). Weighing the deep continental biosphere. *FEMS Microbiol Ecol* **87**: 113–120.
- Nancucheo I, Johnson DB. (2010). Production of glycolic acid by chemolithotrophic iron- and sulfur-oxidizing bacteria and its role in delineating and sustaining acidophilic sulfide mineral-oxidizing consortia. *Appl Environ Microbiol* **76**: 461–467.
- Nielsen ME, Fisk MR, Istok JD, Pedersen K. (2006). Microbial nitrate respiration of lactate at in situ conditions in ground water from a granitic aquifer situated 450 m underground. *Geobiology* **4**: 43–52.
- Nyyssonen M, Hultman J, Ahonen L, Kukkonen I, Paulin L, Laine P *et al.* (2014). Taxonomically and functionally diverse microbial communities in deep crystalline rocks of the Fennoscandian shield. *ISME J* **8**: 126–138.
- Onstott TC, Magnabosco C, Aubrey AD, Burton AS, Dworkin JP, Elsila JE *et al.* (2014). Does aspartic acid racemization constrain the depth limit of the subsurface biosphere? *Geobiology* **12**: 1–19.
- Orcutt BN, LaRowe DE, Biddle JF, Colwell FS, Glazer BT, Reese BK *et al.* (2013). Microbial activity in the marine deep biosphere: progress and prospects. *Front Microbiol* **4**: 189.
- Orsi WD, Edgcomb VP, Christman GD, Biddle JF. (2013). Gene expression in the deep biosphere. *Nature* **499**: 205–208.
- Osburn MR, LaRowe DE, Momper LM, Amend JP. (2014). Chemolithotrophy in the continental deep subsurface: Sanford Underground Research Facility (SURF), USA. *Front Microbiol* **5**: 610.
- Osorio H, Mangold S, Denis Y, Nancucheo I, Johnson DB, Bonnefoy V *et al.* (2013). Anaerobic sulfur metabolism coupled to dissimilatory iron reduction in the extremophile *Acidithiobacillus ferrooxidans*. *Appl Environ Microbiol* **79**: 2172–2181.
- Pedersen K. (1999). Subterranean microorganisms and radioactive waste disposal in Sweden. *Engin Geol* **52**: 163–176.
- Pedersen K. (2012). Influence of H₂ and O₂ on sulphate-reducing activity of a subterranean community and the coupled response in redox potential. *FEMS Microbiol Ecol* **82**: 653–665.
- Pedersen K. (2013). Metabolic activity of subterranean microbial communities in deep granitic groundwater supplemented with methane and H₂. *ISME J* **7**: 839–849.
- Pedersen K, Bengtsson A, Edlund J, Rabe L, Hazen T, Chakraborty R *et al.* (2014). Complete genome sequence of the subsurface, mesophilic sulfate-reducing bacterium *Desulfovibrio aespoensis* Aspo-2. *Genome Announc* **2**: e00509–14.
- Purkamo L, Bomberg M, Nyyssonen M, Kukkonen I, Ahonen L, Itavaara M. (2015). Heterotrophic communities supplied by ancient organic carbon predominate in deep Fennoscandian bedrock fluids. *Microb Ecol* **69**: 319–332.
- R Development Core Team. (2011). *R: A Language and Environment for Statistical Computing*. The R Foundation for Statistical Computing: Vienna, Austria, <http://www.R-project.org/>.
- Rönneback P, Åström M. (2007). Hydrochemical patterns of a small lake and a stream in an uplifting area proposed as a repository site for spent nuclear fuel, Forsmark, Sweden. *J Hydrol* **344**: 223–235.
- Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF *et al.* (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol* **12**: 87.
- Schuchmann K, Muller V. (2014). Autotrophy at the thermodynamic limit of life: a model for energy conservation in acetogenic bacteria. *Nat Rev Microbiol* **12**: 809–821.
- Smellie JAT, Laaksoharju M, Wikberg P. (1995). Äspö, SE Sweden - a natural groundwater-flow model derived from hydrogeological observations. *J Hydrol* **172**: 147–169.
- Ström A, Andersson J, Skagius K, Winberg A. (2008). Site descriptive modelling during characterization for a geological repository for nuclear waste in Sweden. *Appl Geochem* **23**: 1747–1760.
- Swan BK, Tupper B, Sczyrba A, Lauro FM, Martinez-Garcia M, González JM *et al.* (2013). Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. *Proc Natl Acad Sci USA* **110**: 11463–11468.
- Warnes GR, Bolker B, Goryanc G, Grothendieck G, Korosec A, Lumley T *et al.* (2014), gdata: Various R programming tools for data manipulation. R package version 2.13.3 <http://CRAN.R-project.org/package=gdata>.
- Wrighton KC, Thomas BC, Sharon I, Miller CS, Castelle CJ, VerBerkmoes NC *et al.* (2012). Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. *Science* **337**: 1661–1665.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Supplementary Information accompanies this paper on The ISME Journal website (<http://www.nature.com/ismej>)