

1 **Energy efficiency and biological interactions define the core microbiome of deep oligotrophic
2 groundwater**

3 Maliheh Mehrshad^{1,*}, Margarita Lopez-Fernandez^{2,*,#}, John Sundh³, Emma Bell⁴, Domenico Simone^{2,5},
4 Moritz Buck⁶, Rizlan Bernier-Latmani⁴, Stefan Bertilsson^{1,6}, and Mark Dopson²

5 ¹Department of Ecology and Genetics, Limnology and Science for Life Laboratory, Uppsala University,
6 Norbyvägen 18D, 752 36 Uppsala, Sweden

7 ²Centre for Ecology and Evolution in Microbial Model Systems (EEMiS), Linnaeus University,
8 Stuvaregatan 4, 391 82 Kalmar, Sweden

9 ³Dept of Biochemistry and Biophysics, National Bioinformatics Infrastructure Sweden, Science for Life
10 Laboratory, Stockholm University, Box 1031, SE-17121 Solna, Sweden

11 ⁴Environmental Microbiology Laboratory, Environmental Engineering Institute, School of Architecture,
12 Civil and Environmental Engineering, École Polytechnique Fédérale de Lausanne, Lausanne, 1015,
13 Switzerland

14 ⁵SLU Bioinformatics Infrastructure, Swedish University of Agricultural Sciences, Almas Allé 5, 750 07
15 Uppsala, Sweden

16 ⁶Department of Aquatic Sciences and Assessment, SLU Uppsala, Sweden

17 [#]Present address: Institute of Resource Ecology, Helmholtz-Zentrum, Dresden-Rossendorf, Bautzner
18 Landstraße 400, 01328 Dresden, Germany

19 *These authors contributed equally

20 Correspondence: Maliheh Mehrshad (maliheh.mehrshad@ebc.uu.se) and Margarita Lopez-Fernandez
21 (margarita.lopezfernandez@lnu.se).

22 **Abstract**

23 Extremely oligotrophic deep groundwaters host organisms attuned to the low-end of the bioenergetics
24 spectrum. While all domains of life along with viruses are active in this habitat, the evolutionary and
25 ecological constraints on colonization and niche shifts and their consequences for the microbiome
26 convergence are unknown. Here we provide a comparative genome-resolved analysis of the prokaryotic
27 community in disconnected fracture fluids of the Fennoscandian Shield. The data show that the
28 oligotrophic deep groundwaters flowing in similar lithologies offer fixed niches that are occupied by a
29 common deep groundwater core microbiome. Based on this high resolution “multi-omics” enabled
30 understanding of the underlying mechanisms via functional expression analysis, we conclude that deep
31 groundwater ecosystems foster highly diverse, yet cooperative microbial communities adapted to this
32 setting. The fitness of primary energy producers is increased by ecological traits such as aggregate or
33 biofilm formation. This also facilitates reciprocal promiscuous partnerships with diverse and prevalent
34 epi-bionts, alleviating the “tragedy of common goods”. Hence, instead of a lifestyle where microbes
35 predominantly invest in functions related to maintenance and survival, an episodic and cooperative
36 lifestyle ensures the subsistence of the deep groundwater microbiome. We suggest the name “halt and
37 catch fire” for this way of life.

38 **Introduction**

39 Prokaryotes are at the base of the food web in deep groundwater habitats that host all domains of life
40 as well as viruses^{1,2}. With estimated total abundances of a staggering 5×10^{27} cells^{3,4}, they are constrained
41 by factors such as bedrock lithology, available electron donors and acceptors, depth, and hydrological
42 isolation from the photosynthesis-fueled surface^{4,5}. The limited number of access points to study this
43 environment render our knowledge of deep groundwaters too patchy for robustly addressing eco-
44 evolutionary questions. Consequently, ecological strategies and factors influencing the establishment

45 and propagation of the core deep groundwater microbiome, along with its comprehensive diversity,
46 metabolic context, and adaptations remain elusive.

47 The deep disconnected biosphere is an environment of constant and frequent selection hurdles,
48 which define not only the composition of the resident community, but more importantly also its
49 strategies to cope with episodic availability of nutrients and reducing agents. In the geochemically-stable
50 and low-energy conditions of the deep biosphere, it is suggested that microbes only occasionally have
51 access to the “basal power requirement” for cell maintenance (e.g., biomass production plus synthesis
52 of biofilms and polymeric saccharides, etc.) or the costly process of duplication^{6,7}. Inspecting the
53 expression profile and metabolic context of actively transcribing microbes may reveal the dominant
54 ecological strategies in the deep groundwater and uncover the dimensions of its available niches. The
55 microbial diversity of the terrestrial subsurface has previously been probed by large-scale “omics” in
56 shallow aquifers (Rifle, USA)⁸, a CO₂ saturated geyser (Crystal Geyser, Utah, USA)⁹, and carbon-rich
57 shales (Marcellus and Utica, USA)¹⁰. However, extremely oligotrophic and disconnected deep
58 groundwater habitats lack a comprehensive comparative “multi-omics” analysis. The proterozoic
59 crystalline bedrock of the Fennoscandian Shield (~1.8 Ga. years old) hosts two deep tunnels that provide
60 access to disconnected fracture fluids (ca. 170 to 500 meters below sea level, mbsl) running through a
61 similar granite/granodiorite lithology^{4,11–14}. The two sites, located in Sweden and Finland, provide a rare
62 opportunity to place the microbiome of deep groundwater under scrutiny (**Supplementary Figure S1**
63 depicts the location of large metagenomics datasets for oligotrophic groundwaters).

64 In this study, we investigate the existence of a common core microbiome and possible community
65 convergence in the extreme and spatially heterogeneous deep groundwater biome. We leverage a large
66 “multi-omics” initiative that combines metagenomes, single cell genomes, and metatranscriptomes
67 from samples collected at two disconnected sites excavated in similar lithology. Through an extensive
68 genome-resolved view and comparative analysis of the communities, we provide strong support for the

69 existence of a common core microbiome in deep groundwaters. The metabolic context and expressed
70 functions of the microbial community were further used to elucidate the ecological and evolutionary
71 processes essential for successfully occupying and propagating in the available niches of this extreme
72 habitat.

73 **Results and discussion**

74 The Fennoscandian Shield bedrock contains an abundance of fracture zones with different groundwater
75 characteristics; including groundwater source, retention time, chemistry, and connectivity to surface-fed
76 organic compounds. The Äspö HRL and Olkiluoto drillholes were sampled over time, covering a diversity
77 of aquifers representing waters of differing ages plus planktonic versus biofilm-associated communities.

78 In order to provide a genome-resolved view of the Fennoscandian Shield bedrock prokaryotic
79 community, collected samples were used for “multi-omics” integrated analysis by combining
80 metagenomes ($n=44$), single-cell genomes ($n=564$), and metatranscriptomes ($n=9$) (**Supplementary**

81 **TableS1**). Binning of the 44 generated metagenomes (~1.3 TB sequenced data) resulted in the
82 reconstruction of 1278 metagenome-assembled genomes (MAGs; $\geq 50\%$ completeness and $\leq 5\%$
83 contamination). By augmenting this dataset with 564 sequenced single cell amplified genomes (SAGs;
84 114 with $\geq 50\%$ completeness and $\leq 5\%$ contamination), we present a comprehensive genomic database
85 for the prokaryotic diversity of these oligotrophic deep groundwaters, hereafter referred to as the
86 Fennoscandian Shield genomic database (FSGD; statistics in **Figure1A & Supplementary TableS2**).

87 Phylogenomic reconstruction using reference genomes in the Genome Taxonomy Database (GTDB-TK;
88 release 86) shows that the FSGD MAGs/SAGs span most branches on the prokaryotic tree of life
89 (**Figure2**). Harboring representatives from 53 prokaryotic phyla (152 Archaeal MAGs/SAGs in 7 phyla and
90 1240 bacterial MAGs/SAGs in 46 phyla), the FSGD highlights the remarkable diversity of these
91 oligotrophic deep groundwaters. Apart from the exceptional case of a single-species ecosystem
92 developed by *Candidatus Desulforudis audaxviator* in the fracture fluids of an African gold mine¹⁵, other

93 studies on deep groundwater as well as aquifer sediments have revealed a notable prokaryotic
94 phylogenetic diversity^{8,16}. Clustering reconstructed FSGD MAGs/SAGs into operationally defined
95 prokaryotic species ($\geq 95\%$ average nucleotide identity (ANI) and $\geq 70\%$ coverage) produced 598 genome
96 clusters. Based on the GTDB-TK affiliated taxonomy, a single FSGD cluster may represent a novel
97 phylum, whereas at the lower taxonomic levels, the FSGD harbors genome clusters representing seven
98 novel taxa at class, 58 at order, 123 at family, and 345 at the genus levels. In addition, more than 94% of
99 the reconstructed MAGs/SAGs clusters ($n=568$) represent novel species with no existing representative
100 in public databases (**Supplementary TableS2**). Mapping metagenomic reads against genome clusters
101 represented exclusively by SAGs ($n=38$, Figure 1A) revealed very low abundances for ten genome
102 clusters (16 SAGs), suggesting they might represent rare members of the microbial community in the
103 investigated deep groundwaters (**Supplementary TableS3**).

104 To explore the community composition of different groundwaters and its temporal variations,
105 presence/absence patterns were computed by competitive mapping of the metagenomics reads against
106 all reconstructed MAGs/SAGs of the FSGD and then normalized for sequencing depth in each
107 metagenome. Since metagenomes were in some cases amplified because of low DNA amounts, we only
108 discuss binary presence/absence values when referring to the community composition to avoid inherent
109 biases in abundance values calculated by counting mapped metagenomics reads. The Äspö HRL
110 metagenomics samples were collected over three years from 2013 to 2016 from five different
111 boreholes. The Olkiluoto metagenomics samples were collected between June and November 2016
112 from three different drillholes. Communities from each separate borehole cluster together and show
113 only minimal variations in prokaryotic composition over time, hinting at high stability of prokaryotic
114 community composition in the groundwater of the different aquifers. In contrast, the different
115 boreholes feature discrete community compositions (**Figures3 & 1C**). The observed compositional
116 differences are likely to be at least partially caused by the varying availability of reducing agents and

117 organic carbon in different boreholes, resulting from contrasting retention time, depth, and isolation
118 from surface inputs of organic compounds^{14,17}. In the case of Äspö HRL datasets, different treatments
119 (planktonic vs biofilm associated microbes) or size fractions (large vs small fraction) of samples
120 originating from the same boreholes also cluster separately (**Figure1C**).

121 Mapping the metagenomic reads against the FSGD identified 340 MAGs/SAGs that were present in
122 groundwater samples from both sites (**Figures3 & 1D**). These prevalent MAGs/SAGs, cluster in 158
123 genome clusters taxonomically affiliated to both domain archaea (phyla Nanoarchaeota and
124 Thermoplasmata; $n=30$) and bacteria (phyla Acidobacteriota, Actinobacteriota, Bacteroidota,
125 Bipolariaulota, Caldribacteriota, Campylobacterota, CG03, Chloroflexota, Deferrisomatota,
126 Dependentiae, Desulfobacterota, Firmicutes_A, Myxococcota, Nitrospirota, Omnitrophota,
127 Patescibacteria, Proteobacteria, UBA6262, UBA9089, and Verrucomicrobiota; $n=310$). While the very
128 disconnected nature of the two sites is reflected in their discrete prokaryotic community composition
129 (**Figures3 & 1C**) the two locations harbor bedrock with similar lithologies. Consequently, they are likely
130 to provide similar niches that may result in convergent species incidence. The presence of the same
131 species in disconnected deep groundwaters, where the bedrock lithology is not the pressing divergence
132 force consolidates the existence of a deep groundwater microbiome primed to occupy the fixed niches
133 available in these habitats. The relatively high phylogenetic diversity of the prevalent species implies a
134 significant role of ecological convergence (due to e.g., availability of nutrient and reducing agents) rather
135 than evolutionary responses. However, this by no means confutes the possibility of an evolutionary
136 convergence as the community clearly undergoes adaptation over the long residence times
137 characteristic for deep groundwaters. For instance, salinity is a proxy for water retention time and it
138 ranges from 0.4 (similar to the brackish Baltic proper) to 1.8% (ca. half that of marine systems) in the
139 different groundwaters (**Supplementary TableS1**). This is reflected in a shift in the isoelectric point of
140 the predicted proteome by decreased prevalence of basic proteins as a potential adaptation strategy in

141 active groundwater microbes to their surrounding matrix over evolutionary timescales (**Figure4**). The
142 isoelectric point trend towards a reduced prevalence of basic proteins is specifically pronounced in
143 Olkiluoto drillhole OL-KR46 where salinity reaches a maximum of 1.8% (**Figure4B**). In contrast, the
144 relative frequency of calculated isoelectric points for predicted proteins in metagenomes sequenced
145 from other drillholes in Olkiluoto and Äspö HRL (ranging from 0.4 to 1.2% salinity) are similar to one
146 another, with a relatively higher frequency for basic proteins compared to OL-KR46. The OL-KR46
147 drillhole provides access to fracture fluids at ~530.6 mbsl where microbial communities have relatively
148 low species richness (15-25 genome clusters per dataset and 27 unique clusters in total), hence
149 representing a distinct community composition from other borehole communities (**Figure1B, C, & D**).
150 Inspecting the metabolic context of the reconstructed MAGs/SAGs from OL-KR46 suggests a flow of
151 carbon between sulfate-reducing bacteria as the most predominant metabolic guild in the community
152 and acetogens, methanogens, and fermenters as has already been reported¹⁸. Despite its unique
153 community composition, 85% of the genome clusters represented in the OL-KR46 drillhole are among
154 the prevalent clusters present in groundwaters collected from both sites. These prevalent genomes
155 include representatives of *Pseudodesulfovibrio aespoeensis*¹⁹, a species originally isolated from 600
156 mbsl in Äspö groundwater that is present among the FSGD MAGS from Äspö HRL borehole SA1229A and
157 Olkiluoto drillhole OL-KR13 and OL-KR46.

158 Metabolic context and ecosystem functioning of the resident prokaryotes can provide clues about
159 the features of the fixed deep groundwater niches and whether they are defined predominantly by
160 biotic interactions or abiotic forces. Prior studies prove that representatives of all domains of life are
161 actively transcribing in these deep groundwaters^{1,18,20,21}. However, a comprehensive taxonomic and
162 metabolic milieu of the transcribing constituents of the active Fennoscandian Shield community has not
163 been explored. A high-resolution and genome-resolved view of the transcription pattern by mapping
164 metatranscriptomic reads against the FSGD MAGs/SAGs was generated, where actively transcribing

165 genomes, their transcribed genes, and overall metabolic capability was catalogued. This analysis reveals
166 that the majority of FSGD MAGs/SAGs clusters are actively transcribing in the nine sequenced
167 metatranscriptomes derived from the three Äspö HRL boreholes (**Figure3**). Resources and energy costs
168 dedicated to protein synthesis (i.e., transcription and translation) appear sufficiently large for
169 prokaryotes to be recognized by natural selection and impact their fitness^{22,23}. Accordingly, evolutionary
170 adaptations such as codon usage bias and regulatory processes are in place to adjust cellular
171 investments in these processes. Unequal utilization of synonymous codons can have implications for a
172 range of cellular and interactive processes, such as mRNA degradation, translation, and protein
173 folding^{24,25}, as well as viral resistance mechanisms, and horizontal gene transfer^{26,27}. Calculating the
174 frequency of synonymous codons in the FSGD MAGs/SAGs (**Supplementary FigureS2**) and those
175 belonging to the highly expressing genome clusters (TPM>10000 arbitrary thresholds; **Supplementary**
176 **FigureS3**) reveal variable utilization of synonymous codons in different MAGs/SAGs. These variable
177 patterns are primarily related to the range of GC content (**Supplementary FigureS2 & S3**). Further
178 exploration of the variable codon utilization among highly expressing representatives of the phylum
179 Desulfobacterota by separate calculation of codon frequency for expressing CDSs (according to the
180 mapping results of metatranscriptomes) highlights cases of potential transcription efficiency controls via
181 codon usage bias in these genomes (**Figure4**). While in most cases, both expressed CDSs and the rest of
182 CDSs in the genome represent similar synonymous codon frequency and distribution, some genomes
183 display notable differences (e.g. MAGs undefined_mixed_planktonic.mb.19 and MM_PC_MetaG.mb.230
184 along with SAG 3300021839). The expressed CDSs of these reconstructed genomes encode different
185 functions related to their central role in sulfate reduction metabolism and regulatory functions
186 (**Supplementary TableS5**). Interestingly, the expressed CDSs of the SAG 3300021839 code for heritable
187 host defense functions against phages and foreign DNA (i.e., CRISPR/Cas system-associated proteins
188 Cas8a1 and Cas7). These processes are regulated to be active in case of exposure to virus. Accordingly,

189 the observed differential codon usage frequency of expressed genes versus the rest of genome could
190 hint at its potential role in regulating the efficiency of translation. The same SAG also expresses the
191 ribonuclease toxin, BrnT of type II toxin-antitoxin system. This toxin is known to respond to
192 environmental stressors and cease bacterial growth by rapid attenuation of the protein synthesis most
193 likely via its ribonuclease activity²⁸. This expression profile suggests a potential response to phage
194 infection and activation of a host defense system. We suggest that the BrnT/BrnA toxin-antitoxin system
195 could potentially be involved in attenuating the protein synthesis, consequently disrupting the lytic
196 lifecycle of phage.

197 In deep groundwater habitats, organisms respond to limited energy and nutrient availability by
198 adjusting their energy investment in different expressed traits (see above). Highly expressing cells are
199 presumed to be either equipped with efficient metabolic properties and biotic interactions that are
200 tuned to available niches (including but not limited to the dimensions of fixed niches available to the
201 common core microbiome) or alternatively represent an ephemeral bloom profiting from occasionally
202 available nutrients. To shed light on this, we explored the metabolic context and life style of 86 FSGD
203 genome clusters with high transcription levels (TPM \geq 10000, arbitrary threshold) comprising 192 MAGs
204 and 35 SAGs. The relatively large phylogenetic diversity of such highly expressing clusters (17 phyla)
205 reaffirms that a considerable fraction of the deep groundwater microbiome has competitive properties
206 with regards to both metabolism and interactions (**Figure5**).

207 The normalized count of mapped metatranscriptomic reads on genes annotated with functions
208 related to provision of “public-goods” comprise a considerable proportion of the overall transcription
209 profile (ca. 15 to 20% with one case reaching as high as 30%) in deep groundwater metatranscriptomes
210 (**Figure6 & Supplementary TableS6** for a list explored K0 identifiers). One way to alleviate the “tragedy
211 of the commons” imposed by the production and sharing of such “public-goods” is by emergence of
212 local sub-communities either through biofilm formation or aggregation. Such a lifestyle would reduce

213 the number of “cheaters” and provide an evolutionary advantage for cooperation as compared to
214 competition. However, exploitation of “public goods” seems inevitable in groundwaters considering the
215 high phylogenetic diversity and widespread presence of Patescibacteriota representatives (300 MAGs
216 and 19 SAGs forming 152 clusters) and Nanoarchaeota (100 MAGs in 56 clusters). The metabolic context
217 of these reconstructed MAGs/SAGs suggests a primarily heterotrophic and fermentative lifestyle similar
218 to prior reports on the epi-symbiotic representatives of Patescibacteriota and Nanoarchaeota²⁹, and
219 accordingly they will be highly dependent on leaky metabolites produced by adjacent cells. The high
220 RNA transcript counts for representatives of these phyla suggests that they have sufficient energy at
221 their disposal to invest in transcription (16 Patescibacteriota and 3 Nanoarchaeota clusters) (**Figure 5**).
222 Their captured transcripts were annotated as ribosomal proteins, cell division proteins (FtsZ), several
223 genes involved in central carbohydrate, lipid, nucleotide, and amino acid metabolism, as well as few
224 genes involved in cofactor and vitamin biosynthesis. To mitigate the burden of “public-goods” providers
225 and in order to stabilize their own syntrophic niche, these epi-symbiotic cells likely participate in
226 reciprocal partnerships where they supply fermentation products (e.g., lactate, acetate, and hydrogen),
227 vitamins, amino acids, and secondary metabolites to their direct or indirect partners in their immediate
228 surroundings. Representatives of these phyla are also detected in the common core microbiome across
229 both groundwater sites (25 Nanoarchaeota MAGs and 81 Patescibacteriota MAGs/SAGs; **Supplementary**
230 **Table S4**). This implies a significant role of biological interactions in the development of fixed niches in
231 the deep groundwaters. The epi-symbiotic association of Patescibacteriota and Nanoarchaeota with
232 prokaryotic hosts has already been verified for several representatives³⁰⁻³⁴. However, the level and
233 range of host/partner specificity for these associations remain understudied. The incidence of the same
234 genome clusters of Patescibacteriota and Nanoarchaeota representatives in both deep groundwaters,
235 combined with their high expression potential and inferred dependency on epi-symbiotic associations

236 for survival, underscores cooperation as a competent evolutionary strategy in oligotrophic deep
237 groundwaters.

238 The metabolic overview of FSGD epi-symbiotic MAGs/SAGs suggests these cells to partially depend
239 on the energetic currency provided by their partners/hosts while maintaining partnerships via reciprocal
240 metabolite exchange. Reconstructing the metabolic scheme of highly expressing genome clusters, apart
241 from small heterotrophic cells with a proposed epi-symbiotic lifestyle (**Figure5**), highlight sulfur as an
242 electron acceptor that is commonly used in the energy metabolism of the deep groundwater
243 microbiome²¹. Dissimilatory sulfur metabolism via the activity of enzymes sulfate adenylyltransferase
244 (*sat*), adenylylsulfate reductase subunits A/B (*aprAB*), and dissimilatory sulfite reductase A/B subunits
245 (*dsrAB*) is present among representative FSGD MAGs/SAGs of domain bacteria (phyla AABM5-125-24,
246 Actinobacteriota, Chloroflexota, Desulfobacterota, Firmicutes-B, Nitrospirota, Planctomycetota,
247 Proteobacteria, SAR324, UBA9089, UBP1, Verrucomicrobiota, and Zixibacteria) and domain archaea
248 (phylum JdFR-18). A total of 203 MAGs and 10 SAGs representing 95 FSGD genome clusters feature the
249 complete gene set for dissimilatory sulfate reduction or at least the *dsrA/B* genes central for the
250 pathway (absence of other genes involved in the pathway could be a consequence of incompleteness).
251 Although the listed taxa have been demonstrated to be capable of sulfite/sulfate reduction³⁵, it cannot
252 be ruled out that the *dsrA/B* genes are used in reverse for sulfur oxidation. Representatives of 26
253 clusters (ca. 30%) of the highly expressing groups represent dissimilatory sulfur metabolism (**Figure5**).
254 Their complete RNA transcript profile apart from functions related to dissimilatory sulfur metabolism
255 (*aprAB*, *dsrAB*) are related to a wide range of cellular functions. These include genetic information
256 processing, central carbohydrate turnover, lipid and protein metabolism, biofilm formation, membrane
257 transporters, and other cell maintenance functions as well as genes involved in replication and repair of
258 the genome and cell division. Sulfate-reducing bacteria of the phyla Desulfobacterota and UBA9089 also
259 contain genes encoding for the reductive acetyl-CoA pathway (Wood-Ljungdahl pathway) that is utilized

260 in reverse to compensate for the energy-consuming oxidation of acetate to H₂ and CO₂ with energy
261 derived from sulfate reduction³⁶. In this process, cells are able to use acetate as a carbon and electron
262 source (**Figure5**) with various hydrogenase types offering molecular hydrogen as an alternative electron
263 donor³⁷.

264 A notable portion of FSGD transcripts belong to motility related genes (e.g., chemotaxis and
265 flagellar assembly) and studies show that the expression of this costly trait increases in low nutrient
266 environments as an adaptation to anticipate and exploit nutrient gradients³⁸. In addition to motility
267 related genes, sulfate-reducing representatives invest in transcription of type IV secretion systems that
268 can facilitate adhesion, biofilm formation, and protein transport. In combination with their expressed
269 chemotaxis genes, we suggest that these motile cells enable cooperation in their local community by
270 attaching to surfaces or by forming aggregates.

271 Many type II toxin-antitoxin systems (TA) are expressed in the MAGs/SAGs of these oligotrophic
272 groundwaters (e.g., PemK-like, MazF-like, RelE/RelB, and BrnT/BrnA, etc.) likely to alleviate
273 environmental stressors. These molecules are prevalent in prokaryotes³⁹ where they participate in a
274 range of cellular processes such as gene regulation, growth arrest, sub-clonal persistence, and cell
275 survival. TA systems often fulfil their regulatory role by halting protein synthesis in response to
276 environmental stimuli. We propose a model in which deep groundwater microbes adjust to the very
277 nutrient-poor conditions while occasionally being flooded by short-lived pulses of nutrients by TA
278 systems triggering bacteriostasis to avoid exhausting the basal energy supply. To restart the cell
279 function in the occasion of ephemeral access to nutrients, the autoregulation of the antitoxin
280 component of the TA system is alleviated to defy the excess of toxin. Finally, we recovered error-prone
281 DNA polymerase (DnaE2) in reconstructed FSGD MAGs/SAGs (**Figure5 & Supplementary FigureS4** for
282 the phylogeny of polymerases type-C). These polymerases can be recruited to stalled replication forks
283 and are known to be involved in error-prone DNA damage tolerance⁴⁰ helping with the genome

284 replication We suggest the name “Halt and Catch Fire”¹ for this mode of subsistence for deep
285 groundwater microbiomes.

286 **Conclusions**

287 Life in deep groundwaters consisting chiefly of microbes⁴¹ represents spatial heterogeneity in response
288 to factors like bedrock lithology, depth, and available electron acceptor and donors. Spatial
289 heterogeneity together with limited access points have so far hindered our understanding of the
290 ecological and evolutionary forces governing the colonization and propagation of microbes in the deep
291 groundwater niches. By employing high-resolution exploration of the microbial community in the
292 disconnected fracture fluids running through similar lithologies at two locations, we propose the
293 existence of a common core microbiome in these deep groundwaters. The metabolic context of this
294 common core microbiome proves that both physical filters and biological interactions are involved in
295 defining the dimensions of fixed niches where sulfate-reduction and reciprocal partnerships of epi-
296 bionts are the most competent traits. We suggest an active but “Halt and Catch Fire” strategy for the
297 phylogenetically diverse microbiome of the deep groundwater in response to ephemeral nutrient
298 pulses.

299 **Methods**

300 **Sampling and multi-omics analysis.** Multiple groundwater samples were collected over several years
301 from two deep geological sites excavated in crystalline bedrock of the Fennoscandian Shield. The first is
302 the Swedish Nuclear Fuel and Waste Management Company (SKB) operated Äspö HRL located in the
303 southeast of Sweden (Lat N 57° 26' 4" Lon E 16° 39' 36"). The second site is on the island of Olkiluoto,

¹ Early operating systems were written in Assembly language that had no built-in safeguards for the hardware. Consequently, mistakes could easily cause components to overheat and be destroyed. “Halt and Catch Fire” was originally an idiom, referring to anything one had done to cause the CPU to fry itself. It got the abbreviation of HCF meaning that the system is fried and the hardware should be replaced. Later, several developers created an HCF three-letter command (or Opcode) that forced the system to halt (without damaging hardware) and it would require a restart to continue working.

304 Finland, that will also host a deep geological repository for the final disposal of spent nuclear fuel (Lat N
305 61° 14' 31", Lon E 21° 29' 23"). Water types with various ages and origins were targeted by sampling
306 fracture fluids from different depths. The Äspö HRL samples originated from boreholes SA1229A-1
307 (171.3 mbsl), KA3105A-4 (415.2 mbsl), KA2198A (294.1 mbsl), KA3385A-1 (448.4 mbsl), and KF0069A01
308 (454.8 mbsl). The Olkiluoto samples originated from drillholes OL-KR11 (366.7-383.5 mbsl), OL-KR13
309 (330.5-337.9 mbsl), and OL-KR46 (528.7-531.5 mbsl).

310 Collected samples were subjected to high-resolution "multi-omics" analysis by combining
311 metagenomics ($n=27$ from the Äspö HRL and $n=17$ from Olkiluoto), single cell genomics ($n=564$), and
312 metatranscriptomics ($n=9$) (detailed information in **Supplementary Table S1**). Single-cell amplified
313 genomes (SAGs) were captured from KA3105A-4 ($n=15$), KA3385A-1 ($n=148$), SA1229A-1 ($n=118$), OL-
314 KR11 ($n=138$), OL-KR13 ($n=117$), and OL-KR46 ($n=28$) water samples. To probe the expression pattern of
315 the resident community, metatranscriptomic datasets were generated for Äspö HRL samples^{1,20}
316 originating from boreholes KA3105A-4 ($n=2$), KA3385A-1 ($n=4$), and SA1229A-1 ($n=3$). Details of
317 sampling, filtration, DNA/RNA processing, and geochemical parameters of the water samples along with
318 statistics of the metagenomics/metatranscriptomics datasets and SAGs are available in **Supplementary**
319 **methods** and **Supplementary Table S1**.

320 **Metagenome assembly.** All datasets were separately assembled using MEGAHIT⁴² (v. 1.1.3) (--k-min
321 21 --k-max 141 --k-step 12 --min-count 2). The datasets originating from the same water type in each
322 location were also processed as co-assemblies in order to increase genome recovery rates (using the
323 same assembly parameters). A complete list of all metagenomic datasets assembled in this study ($n=44$)
324 and the co-assemblies are provided in **Supplementary Table S1**.

325 **Fennoscandian Shield Genomic Database (FSGD).** The generated "multi-omics" data were used to
326 construct a comprehensive genomic and metatranscriptomic database of the extremely oligotrophic
327 deep groundwaters. Automated binning was performed on assembled ≥ 2 kb contigs of each assembly

328 using MetaBAT2⁴³ (v. 2.12.1) with default settings. Quality and completeness of the reconstructed MAGs
329 and SAGs were estimated with CheckM⁴⁴ (v. 1.0.7). The taxonomy of MAGs/SAGs with $\geq 50\%$
330 completeness $\leq 5\%$ contamination was assigned using GTDB-tk⁴⁵ (v. 0.2.2) that identifies, aligns, and
331 concatenates marker genes in genomes. GTDB-tk then uses these concatenated alignments to place the
332 genomes (using pplacer⁴⁶) into a curated reference tree with subsequent taxonomic classification.
333 Phylogenomic trees of the archaeal and bacterial MAGs and SAGs were also created using the
334 “denovo_wf” subcommand of GTDB-tk (--outgroup_taxon p_Patescibacteria) that utilizes FastTree⁴⁷ (v.
335 2.1.10) with parameters “-wag -gamma”. Reconstructed MAGs and SAGs were de-replicated using
336 fastANI⁴⁸ (v. 1.1) at $\geq 95\%$ identity and $\geq 70\%$ coverage thresholds. A detailed description and genome
337 statistics of the Fennoscandian Shield genomic database (FSGD) is shown in the **Supplementary Table S2**
338 and **Supplementary Methods**.

339 **Functional analysis of the reconstructed genomes.** Annotation of function, computation of
340 isoelectric points and codon usage frequency⁴⁹, abundance, and expression analysis
341 (metatranscriptome) are detailed in the **Supplementary Methods**.

342 **Data availability.** Accession numbers for the metagenomes, SAGs, and metatranscriptomes are
343 provided in **Supplementary Table S1**. The reconstructed genomes of the FSGD are accessible under the
344 NCBI BioProject accession number PRJNA627556. All reconstructed genomes as well as the alignments
345 used for phylogeny reconstruction are deposited to figshare and are publicly available
346 10.6084/m9.figshare.12170313 and 10.6084/m9.figshare.12170310. all data supporting the findings of
347 this paper are available within this paper and its supplementary material. All the programs used and the
348 version and set thresholds are mentioned in the manuscript and supplementary methods.

349

350 **References**

351 1. Lopez-Fernandez, M. *et al.* Metatranscriptomes reveal that all three domains of life are active but

352 are dominated by Bacteria in the Fennoscandian crystalline granitic continental deep biosphere.

353 *MBio* **9**, 1–15 (2018).

354 2. Daly, R. *et al.* Viruses control dominant bacteria colonizing the terrestrial deep biosphere after

355 hydraulic fracturing. *Nat. Microbiol.* **4**, 352–361 (2019).

356 3. Flemming, H.-C. & Wuertz, S. Bacteria and archaea on Earth and their abundance in biofilms. *Nat.*

357 *Rev. Microbiol.* **17**, 247–260 (2019).

358 4. Mcmahon, S. & Parnell, J. Weighing the deep continental biosphere. *FEMS Microbiol. Ecol.* **87**,

359 113–120 (2014).

360 5. Hernsdorf, A. W. *et al.* Potential for microbial H₂ and metal transformations associated with

361 novel bacteria and archaea in deep terrestrial subsurface sediments. *ISME J.* **11**, 1915–1929

362 (2017).

363 6. Jørgensen, B. B., Andrén, T. & Marshall, I. P. G. Sub-seafloor biogeochemical processes and

364 microbial life in the Baltic Sea. *Environ. Microbiol.* (2020). doi:10.1111/1462-2920.14920

365 7. Hoehler, T. M. & Jørgensen, B. B. Microbial life under extreme energy limitation. *Nat. Rev.*

366 *Microbiol.* **11**, 83–94 (2013).

367 8. Castelle, C. J. *et al.* Extraordinary phylogenetic diversity and metabolic versatility in aquifer

368 sediment. *Nat. Commun.* **4**, 2120 (2013).

369 9. Probst, A. J. *et al.* Differential depth distribution of microbial function and putative symbionts

370 through sediment-hosted aquifers in the deep terrestrial subsurface. *Nat. Microbiol.* **3**, 328–336

371 (2018).

372 10. Daly, R. A. *et al.* Microbial metabolisms in a 2.5-km-deep ecosystem created by hydraulic

373 fracturing in shales. *Nat. Microbiol.* **1**, 16146 (2016).

374 11. Smellie, J. A. T., Laaksoharju, M. & Wikberg, P. Äspö, SE Sweden: a natural groundwater flow

375 model derived from hydrogeochemical observations. *J. Hydrol.* **172**, 147–169 (1995).

376 12. Laaksoharju, M., Gascoyne, M. & Gurban, I. Understanding groundwater chemistry using mixing
377 models. *Appl. Geochemistry* **23**, 1921–1940 (2008).

378 13. Mathurin, F. A., Åström, M. E., Laaksoharju, M., Kalinowski, B. E. & Tullborg, E. L. Effect of tunnel
379 excavation on source and mixing of groundwater in a coastal granitoidic fracture network.
380 *Environ. Sci. Technol.* **46**, 12779–12786 (2012).

381 14. Posiva Oy. *Olkiluoto Site Description 2011*. **31**, (2011).

382 15. Chivian, D. *et al.* Environmental Genomics Reveals a Single-Species Ecosystem Deep Within Earth.
383 *Science (80-)*. **322**, 275–278 (2008).

384 16. Momper, L., Jungbluth, S. P., Lee, M. D. & Amend, J. P. Energy and carbon metabolisms in a deep
385 terrestrial subsurface fluid microbial community. *ISME J.* **11**, 2319–2333 (2017).

386 17. Hubalek, V. *et al.* Connectivity to the surface determines diversity patterns in subsurface aquifers
387 of the Fennoscandian shield. *Isme J* **10**, 2447–2458 (2016).

388 18. Bell, E. *et al.* Biogeochemical cycling by a low-diversity microbial community in deep
389 groundwater. *Front. Microbiol.* **9**, 1–17 (2018).

390 19. Motamedi, M. & Pedersen, K. *Desulfovibrio aespoensis* sp. nov., a mesophilic sulfate-reducing
391 bacterium from deep groundwater at Aspo hard rock laboratory, Sweden. *Int. J. Syst. Bacteriol.*
392 **48**, 311–315 (1998).

393 20. Lopez-Fernandez, M., Broman, E., Simone, D., Bertilsson, S. & Dopson, M. Statistical Analysis of
394 Community RNA Transcripts between Organic Carbon and Geogas-Fed Continental Deep
395 Biosphere Groundwaters. *MBio* **10**, e01470-19 (2019).

396 21. Bell, E. *et al.* Active sulfur cycling in the terrestrial deep subsurface. *ISME J.* **14**, 1260–1272
397 (2020).

398 22. Lynch, M. & Marinov, G. K. The bioenergetic costs of a gene. *Proc. Natl. Acad. Sci. U. S. A.* **112**,
399 15690–15695 (2015).

400 23. Seward, E. A. & Kelly, S. Selection-driven cost-efficiency optimization of transcripts modulates
401 gene evolutionary rate in bacteria. *Genome Biol.* **19**, (2018).

402 24. Stoletzki, N. & Eyre-Walker, A. Synonymous codon usage in *Escherichia coli*: Selection for
403 translational accuracy. *Mol. Biol. Evol.* **24**, 374–381 (2007).

404 25. Shao, Z. Q., Zhang, Y. M., Feng, X. Y., Wang, B. & Chen, J. Q. Synonymous codon ordering: A
405 subtle but prevalent strategy of bacteria to improve translational efficiency. *PLoS One* **7**, (2012).

406 26. Bahir, I., Fromer, M., Prat, Y. & Linial, M. Viral adaptation to host: A proteome-based analysis of
407 codon usage and amino acid preferences. *Mol. Syst. Biol.* **5**, 1–14 (2009).

408 27. Tuller, T. *et al.* Association between translation efficiency and horizontal gene transfer within
409 microbial communities. *Nucleic Acids Res.* **39**, 4743–4755 (2011).

410 28. Heaton, B. E., Herrou, J., Blackwell, A. E., Wysocki, V. H. & Crosson, S. Molecular structure and
411 function of the novel BrnT/BrnA toxin-antitoxin system of *Brucella abortus*. *J. Biol. Chem.* **287**,
412 12098–12110 (2012).

413 29. Castelle, C. J. *et al.* Biosynthetic capacity, metabolic variety and unusual biology in the CPR and
414 DPANN radiations. *Nat. Rev. Microbiol.* **16**, 629–645 (2018).

415 30. Huber, H. *et al.* A new phylum of Archaea represented by a nanosized hyperthermophilic
416 symbiont. *Nature* **417**, 63–67 (2002).

417 31. Wurch, L. *et al.* Genomics-informed isolation and characterization of a symbiotic Nanoarchaeota
418 system from a terrestrial geothermal environment. *Nat. Commun.* **7**, 12115 (2016).

419 32. Golyshina, O. V. *et al.* 'ARMAN' archaea depend on association with euryarchaeal host in culture
420 and *in situ*. *Nat. Commun.* **8**, 60 (2017).

421 33. He, X. *et al.* Cultivation of a human-associated TM7 phylotype reveals a reduced genome and
422 epibiotic parasitic lifestyle. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 244–249 (2015).

423 34. Schwank, K. *et al.* An archaeal symbiont-host association from the deep terrestrial subsurface.

424 35. Anantharaman, K. *et al.* Expanded diversity of microbial groups that shape the dissimilatory
425 sulfur cycle. *ISME J.* **13**, 2135–2139 (2019).

426 36. Ragsdale, S. W. & Pierce, E. Acetogenesis and the Wood-Ljungdahl Pathway of CO₂ Fixation.
427 **1784**, 1873–1898 (2009).

428 37. Caffrey, S. M. *et al.* Function of periplasmic hydrogenases in the sulfate-reducing bacterium
429 *Desulfovibrio vulgaris hildenborough*. *J. Bacteriol.* **189**, 6159–6167 (2007).

430 38. Ni, B., Colin, R., Link, H., Endres, R. G. & Sourjik, V. Growth-rate dependent resource investment
431 in bacterial motile behavior quantitatively follows potential benefit of chemotaxis. *Proc. Natl.*
432 *Acad. Sci. U. S. A.* **117**, 595–601 (2020).

433 39. Goormaghtigh, F., Fraikin, N., Hallaert, T., Hauryliuk, V. & Garcia-pino, A. Reassessing the Role of
434 Type II Toxin-Antitoxin Systems in Formation of *Escherichia coli* Type II Persister Cells. *MBio* **9**,
435 e00640-18 (2018).

436 40. Alves, I. R. *et al.* Effect of SOS-induced levels of imuABC on spontaneous and damage-induced
437 mutagenesis in *Caulobacter crescentus*. *DNA Repair (Amst.)* **59**, 20–26 (2017).

438 41. Bar-On, Y. M., Phillips, R. & Milo, R. The biomass distribution on Earth. *Proc. Natl. Acad. Sci. U. S.*
439 *A.* **115**, 6506–6511 (2018).

440 42. Li, D., Liu, C. M., Luo, R., Sadakane, K. & Lam, T. W. MEGAHIT: An ultra-fast single-node solution
441 for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**,
442 1674–1676 (2014).

443 43. Kang, D. *et al.* MetaBAT 2: an adaptive binning algorithm for robust and efficient genome
444 reconstruction from metagenome assemblies. *PeerJ Prepr.* **7**, e27522v1 (2019).

445 44. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the
446 quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome*
447 *Res.* **10**, 137–145 (2015).

448 *Res.* **25**, 1043–55 (2015).

449 45. Parks, D. H. *et al.* A standardized bacterial taxonomy based on genome phylogeny substantially
450 revises the tree of life. *Nat. Biotechnol.* **36**, 996 (2018).

451 46. Matsen, F. A., Kodner, R. B. & Armbrust, E. V. pplacer: linear time maximum-likelihood and
452 Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics*
453 **11**, 538 (2010).

454 47. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2--approximately maximum-likelihood trees for
455 large alignments. *PLoS One* **5**, e9490 (2010).

456 48. Jain, C., Rodriguez-r, L. M. & Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes
457 reveals clear species boundaries. *Nat. Commun.* **9**, 5114 (2018).

458 49. Rice, P., Longden, I. & Bleasby, A. EMBOSS: The European Molecular Biology Open Software
459 Suite. *Trends Genet.* **16**, 276–277 (2000).

460

461 **Acknowledgements**

462 The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science
463 User Facility, is supported under Contract No. DE-AC02-05CH11231. The Swedish Research Council
464 contracts 2018-04311, 2017-04422, and 2014-4398) and The Swedish Nuclear Fuel and Waste
465 Management Company (SKB) supported the study. M.D. thanks the Crafoord Foundation (contracts
466 20180599 and 20130557), the Nova Center for University Studies, Research and Development, and
467 Familjen Hellmans Stiftelse for financial support. M.D. and D.S. thank the Carl Tryggers Foundation
468 (grant KF16: 18) for financial support. S.B. and M.M. acknowledge financial support from the Swedish
469 Research Council and Science for Life Laboratory. High-throughput sequencing was also carried out at
470 the National Genomics Infrastructure hosted by the Science for Life Laboratory. Bioinformatics analyses
471 were carried out utilizing the Uppsala Multidisciplinary Center for Advanced Computational Science

472 (UPPMAX) at Uppsala University (projects b2013127 and SNIC 2019/3-22) with support from a
473 SciLifeLab-WABI bioinformatics grant.

474

475 **Author contributions**

476 M.D., S.B., and R.B.-L. devised the study; M.L.-F. and E.B. collected and processed the samples; M.M.,
477 J.S., D.S. and M. B. analyzed the data; M.M. and M.D. interpreted the data and drafted the manuscript;
478 and all authors read and approved the final manuscript.

479

480 **Competing interests**

481 The authors declare no competing financial interests.

482

483 **Figure legends**

484 **Figure 1- Overview of the FSDB MAGs and SAGs.** Statistics of the metagenome assembled genomes
485 (MAGs) and single cell amplified genomes (SAGs) of the Fennoscandian Shield Genomic Database [A].
486 The species richness as the number of genome clusters present in each borehole/borehole treatment.
487 Numbers on top of each box plot represent the number of metagenomes for each borehole/borehole
488 treatment [B]. NMDS plot of unweighted binary Jaccard beta-diversities of presence/absence of all FSGD
489 reconstructed MAGs/SAGs [C]. MAGs and SAGs belonging to the prevalent clusters present in Äspö HRL
490 and Olkiluoto[D].

491 **Figure 2- Phylogenetic diversity of reconstructed MAGs and SAGs of the Fennoscandian Shield**
492 **genomic database (FSGD).** Genomes present in genome taxonomy database (GTDB) release 86 were
493 used as reference. Archaea and bacteria phylogeny is represented separately in the top and bottom
494 panels respectively. MAGs and SAGs of the FSGD are highlighted in red.**Figure 3- Distribution pattern**
495 **and transcription status of the FSGD genome clusters along different metagenomes and**

496 **metatranscriptomes.** Each column represents a genome cluster of reconstructed MAGs/SAGs of the
497 deep groundwater datasets of this study. The top heat map depicts the distribution of the
498 representatives of each genome cluster along metagenomics datasets originating from different
499 boreholes of the two Fennoscandian shield deep repositories. The bottom heat map represents the
500 transcription status of the genome cluster representatives in the metatranscriptomes originating from
501 different boreholes of the Äspö HRL. Clusters with all zero values have been removed from the plot (in
502 total 10 clusters that are solely represented by SAGs).

503 **Figure 4- Adaptations of the coding sequences of the deep groundwater microbiota.** Relative
504 frequency of isoelectric points in the predicted proteins of assembled metagenomes from Äspö HRL
505 boreholes [A] and Olkiluoto [B]. Salinity of the water flowing in each borehole is shown on the top left
506 legend. Representation of frequency (the expected number of codons, given the input sequences, per
507 1000 bases) of utilization of synonymous codons across MAGs and SAGs of different genomes clusters of
508 highly expressing Desulfobacterota. Codon usage frequencies are calculated separately for the CDSs for
509 which RNA transcripts are detected (green) and the rest of CDSs not actively being transcribed in the
510 sequenced metatranscriptomes (orange) [C]. Stars are showing cases of potential transcription
511 efficiency control via codon usage bias.

512 **Figure 5- Expression profile and metabolic context of the highly expressing genomic clusters.** The
513 expression profile of genome clusters with commutative expression ≥ 10000 TPM and their metabolic
514 potential for nitrogen and sulfur energy metabolism as well as carbon fixation is represented as KEGG
515 modules presence only if all genes of the module or the key genes of the process are present. Nitrogen
516 fixation (M00175), dissimilatory nitrate reduction (M00530), nitrification (M00528), and denitrification
517 (M00529). Dissimilatory sulfate reduction (M00596) and thiosulfate oxidation by SOX complex
518 (M00595). Reductive pentose phosphate cycle, ribulose-5P \Rightarrow glyceraldehyde-3P (M00166), reductive
519 pentose phosphate cycle, glyceraldehyde-3P \Rightarrow ribulose-5P (M00167), Crassulacean acid metabolism,

520 dark (M00168) and light (M00169), reductive acetyl-CoA pathway (Wood-Ljungdahl pathway) (M00377),
521 and phosphate acetyltransferase-acetate kinase pathway, acetyl-CoA => acetate (M00579). Nickel-
522 dependent hydrogenase (PF00374), iron hydrogenase (PF02256 and PF02906), coenzyme F420
523 hydrogenase/dehydrogenase (PF04422 and PF04432), and NiFe/NiFeSe hydrogenase (PF14720).

524 **Figure 6- Expression level of functional classes involved in public good provision in the sequenced
525 metatranscriptomes.** The list of screened K0 identifiers are shown in **Supplementary Table S6.**

526

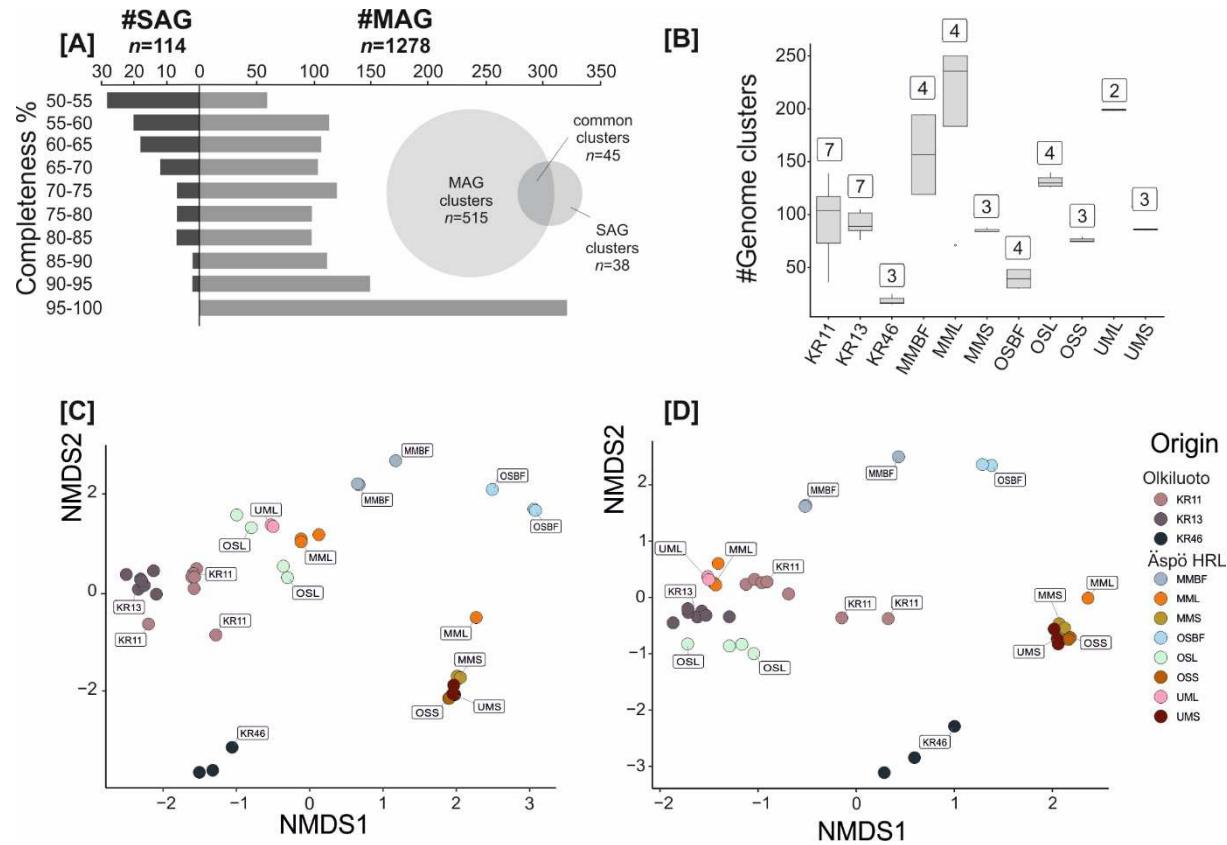


Figure 1- Overview of the FSDB MAGs and SAGs. Statistics of the metagenome assembled genomes (MAGs) and single cell amplified genomes (SAGs) of the Fennoscandian Shield Genomic Database [A]. The species richness as the number of genome clusters present in each borehole/borehole treatment. Numbers on top of each box plot represent the number of metagenomes for each borehole/borehole treatment [B]. NMDS plot of unweighted binary Jaccard beta-diversities of presence/absence of all FSGD reconstructed MAGs/SAGs [C]. MAGs and SAGs belonging to the prevalent clusters present in Äspö HRL and Olkiluoto [D].

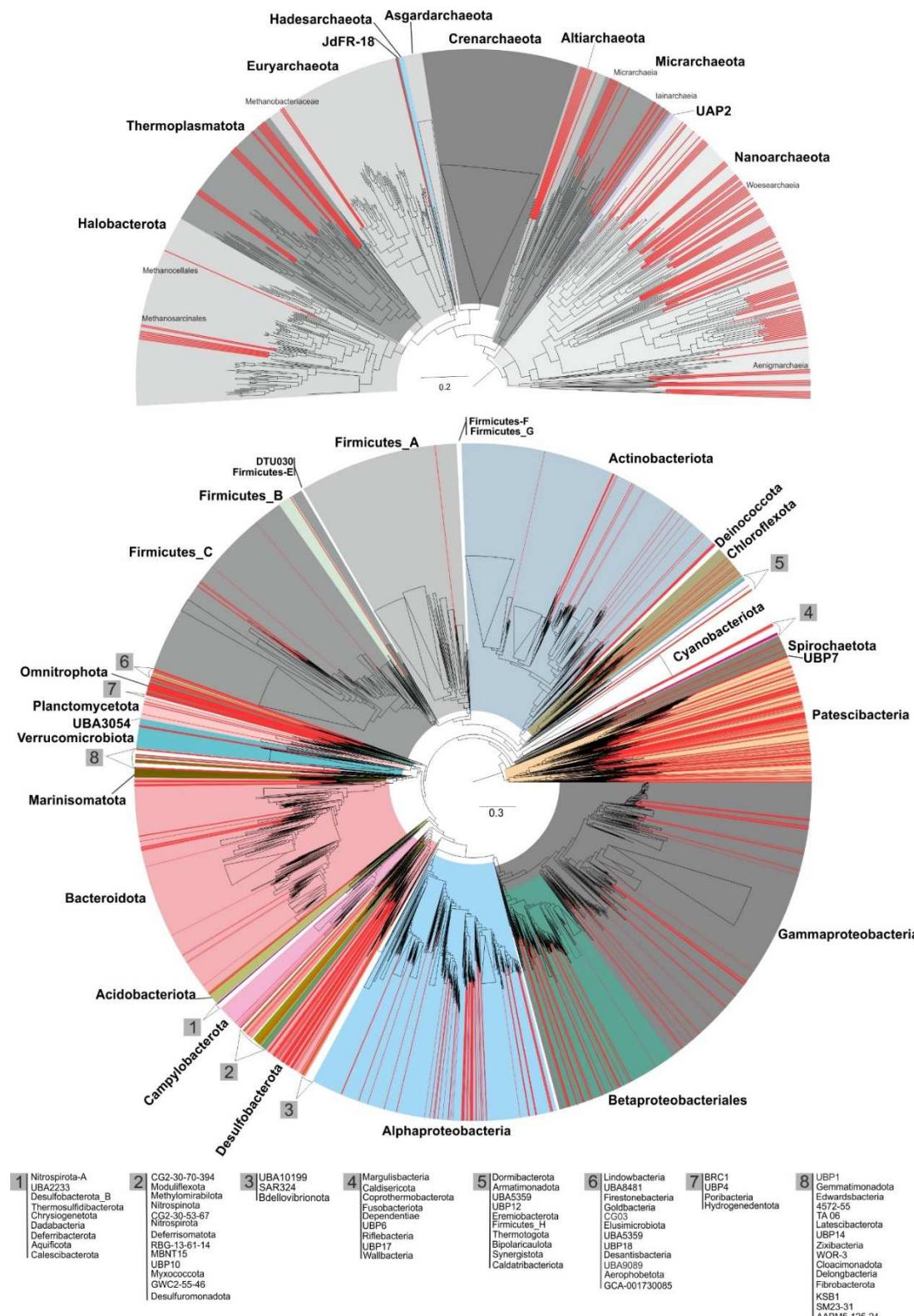


Figure 2- Phylogenetic diversity of reconstructed MAGs and SAGs of the Fennoscandian Shield genomic database (FSGD). Genomes present in genome taxonomy database (GTDB) release 86 were used as reference. Archaea and bacteria phylogeny is represented separately in the top and bottom panels respectively. MAGs and SAGs of the FSGD are highlighted in red.

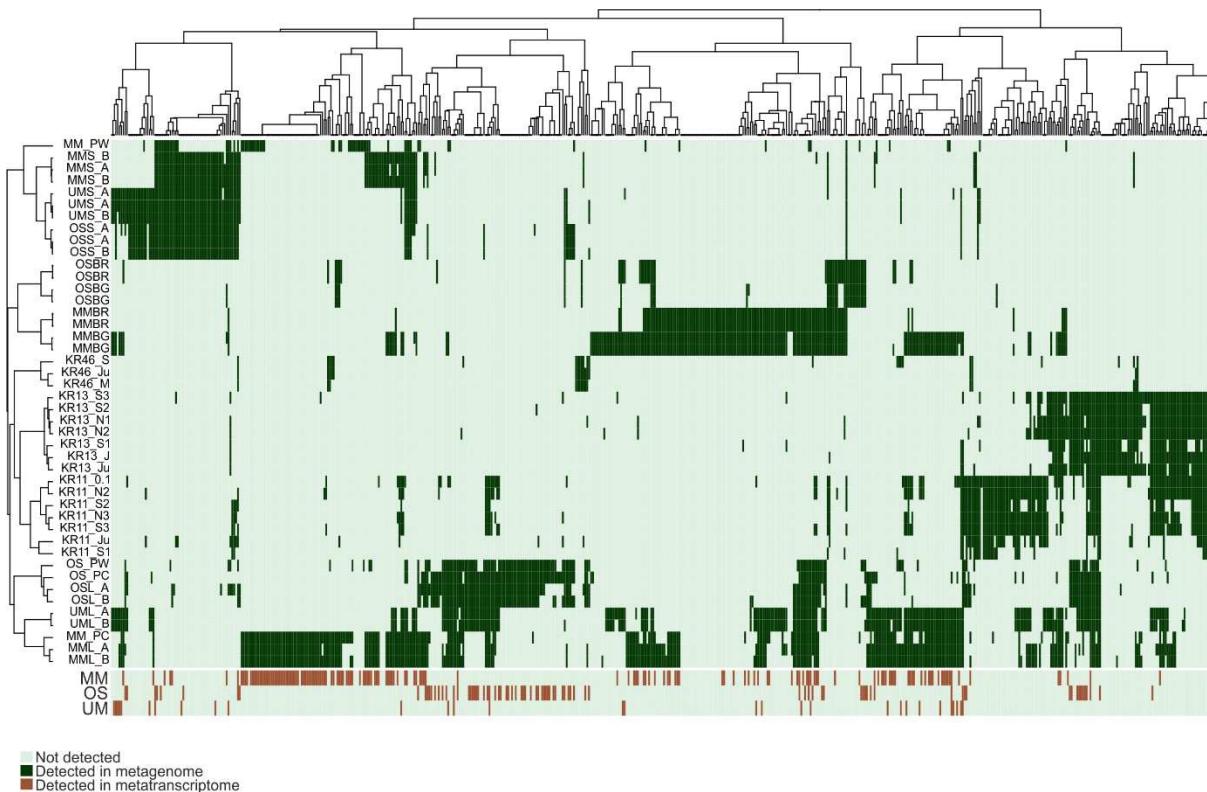


Figure 3- Distribution pattern and transcription status of the FSGD genome clusters along different metagenomes and metatranscriptomes. Each column represents a genome cluster of reconstructed MAGs/SAGs of the deep groundwater datasets of this study. The top heat map depicts the distribution of the representatives of each genome cluster along metagenomics datasets originating from different boreholes of the two Fennoscandian shield deep repositories. The bottom heat map represents the transcription status of the genome cluster representatives in the metatranscriptomes originating from different boreholes of the Äspö HRL. Clusters with all zero values have been removed from the plot (in total 10 clusters that are solely represented by SAGs).

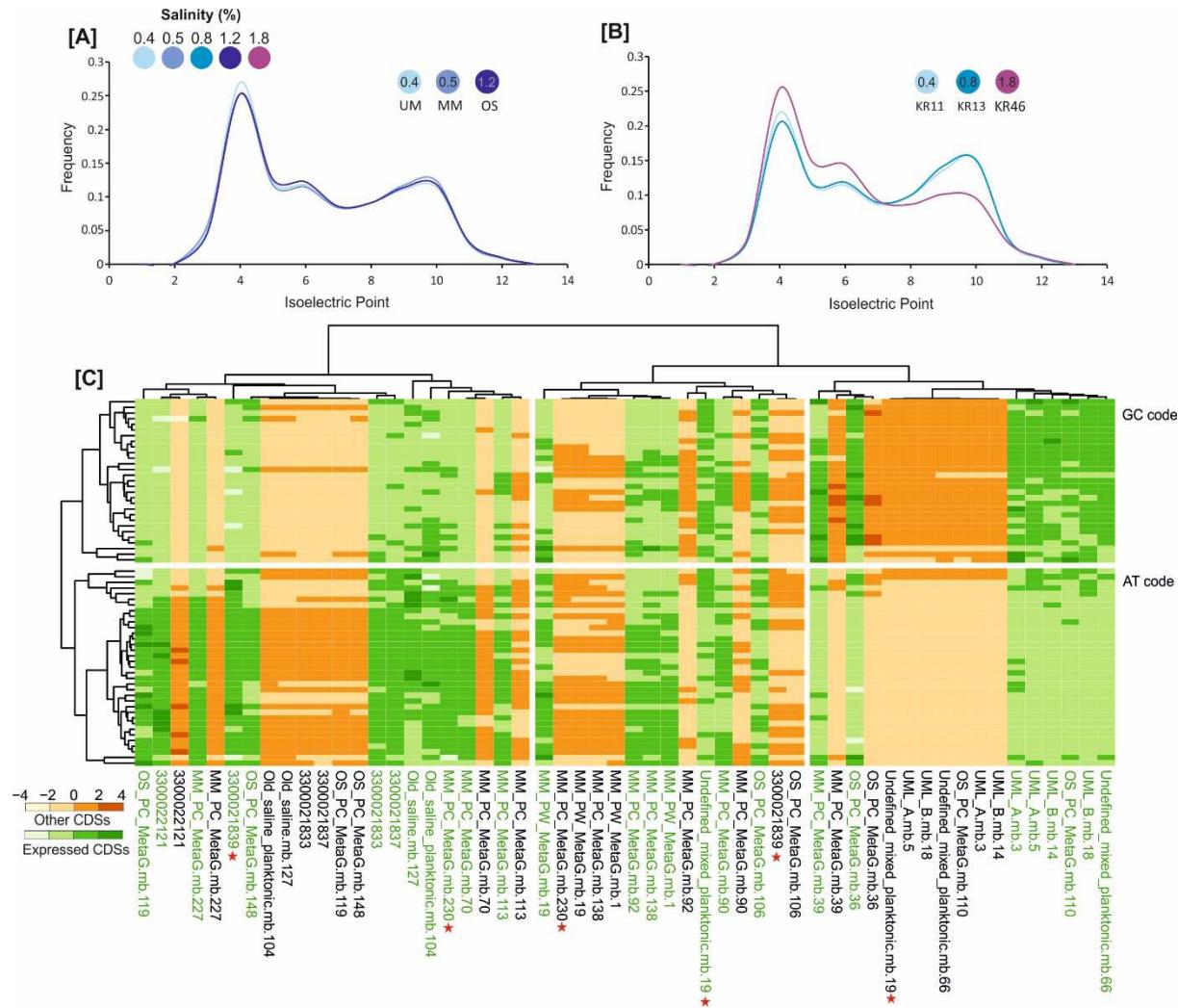


Figure 4- Adaptations of the coding sequences of the deep groundwater microbiota. Relative frequency of isoelectric points in the predicted proteins of assembled metagenomes from Äspö HRL boreholes [A] and Olkiluoto [B]. Salinity of the water flowing in each borehole is shown on the top left legend. Representation of frequency (the expected number of codons, given the input sequences, per 1000 bases) of utilization of synonymous codons across MAGs and SAGs of different genomes clusters of highly expressing Desulfobacterota. Codon usage frequencies are calculated separately for the CDSs for which RNA transcripts are detected (green) and the rest of CDSs not actively being transcribed in the sequenced metatranscriptomes (orange) [C]. Stars are showing cases of potential transcription efficiency control via codon usage bias.

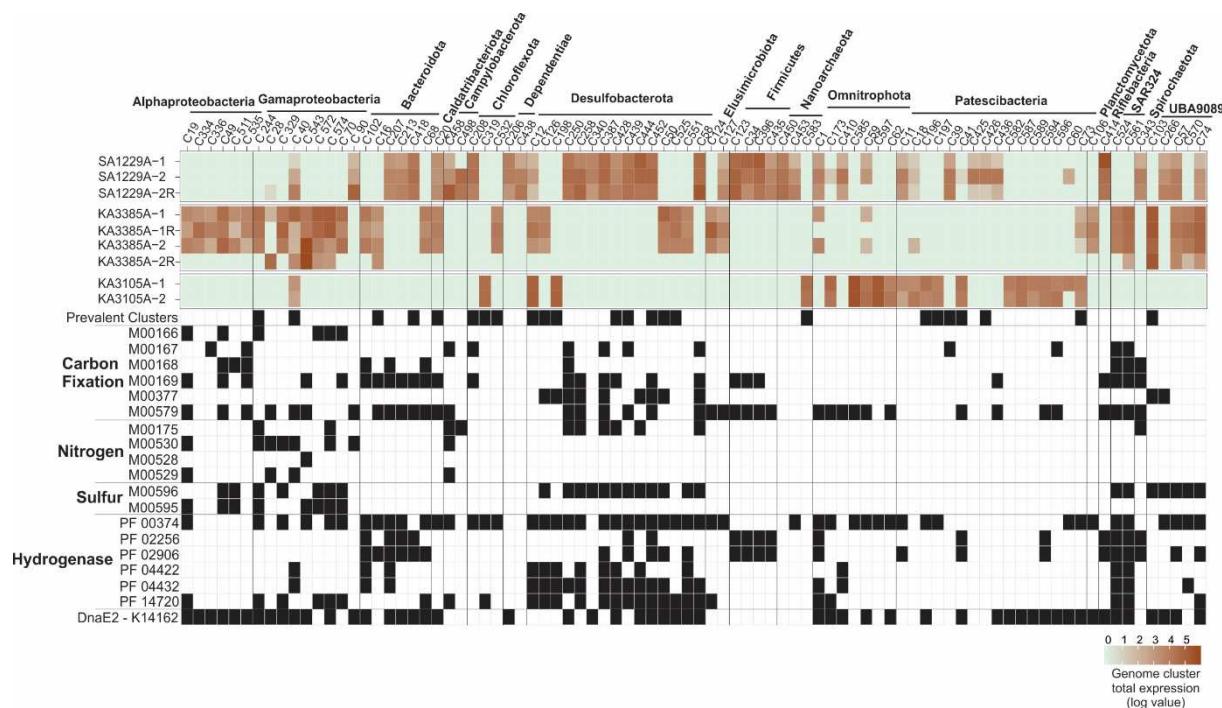


Figure 5- Expression profile and metabolic context of the highly expressing genomic clusters. The expression profile of genome clusters with commutative expression ≥ 10000 TPM and their metabolic potential for nitrogen and sulfur energy metabolism as well as carbon fixation is represented as KEGG modules presence only if all genes of the module or the key genes of the process are present. Nitrogen fixation (M00175), dissimilatory nitrate reduction (M00530), nitrification (M00528), and denitrification (M00529). Dissimilatory sulfate reduction (M00596) and thiosulfate oxidation by SOX complex (M00595). Reductive pentose phosphate cycle, ribulose-5P \Rightarrow glyceraldehyde-3P (M00166), reductive pentose phosphate cycle, glyceraldehyde-3P \Rightarrow ribulose-5P (M00167), Crassulacean acid metabolism, dark (M00168) and light (M00169), reductive acetyl-CoA pathway (Wood-Ljungdahl pathway) (M00377), and phosphate acetyltransferase-acetate kinase pathway, acetyl-CoA \Rightarrow acetate (M00579). Nickel-dependent hydrogenase (PF00374), iron hydrogenase (PF02256 and PF02906), coenzyme F420 hydrogenase/dehydrogenase (PF04422 and PF04432), and NiFe/NiFeSe hydrogenase (PF14720).

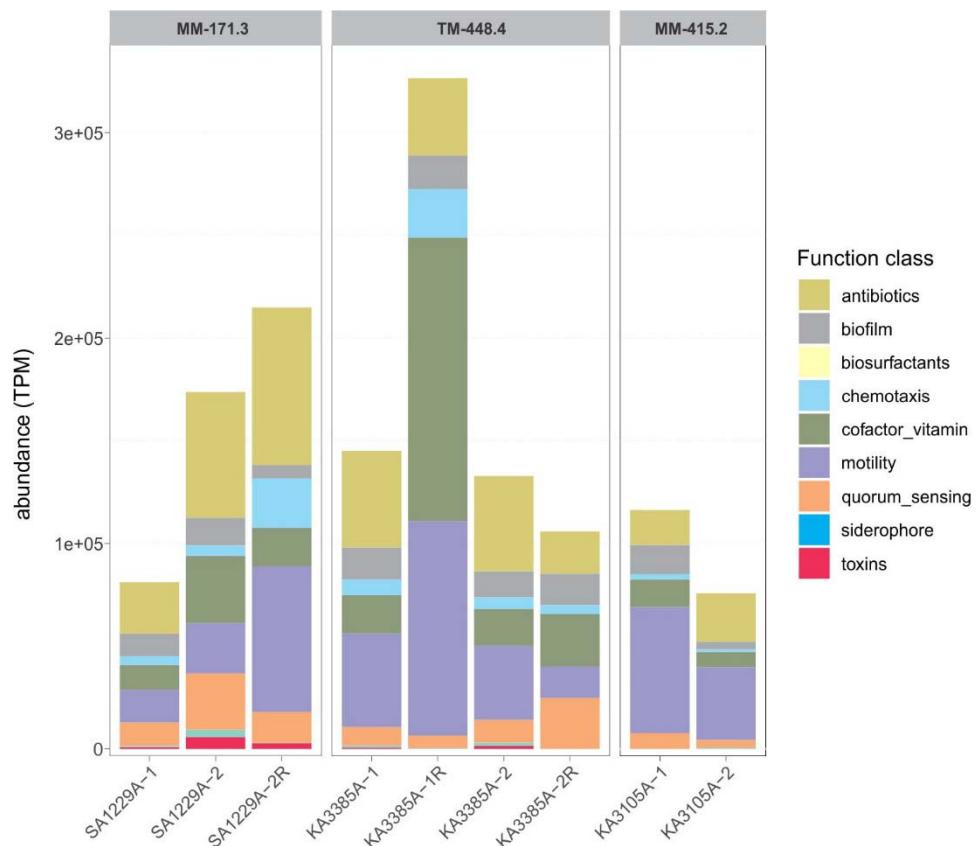
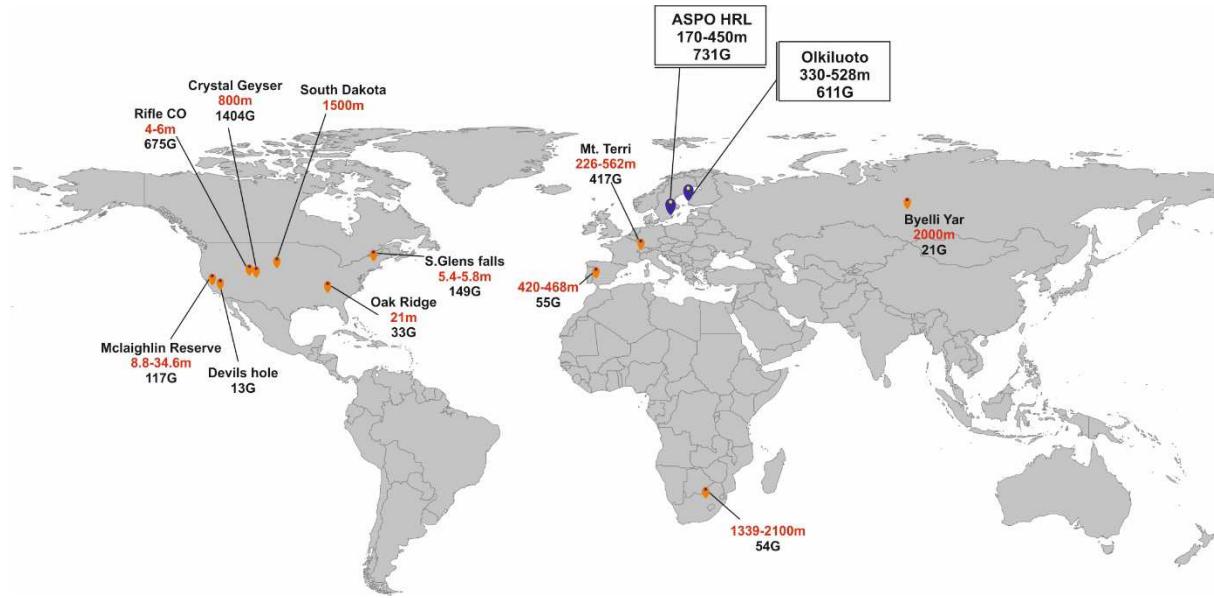
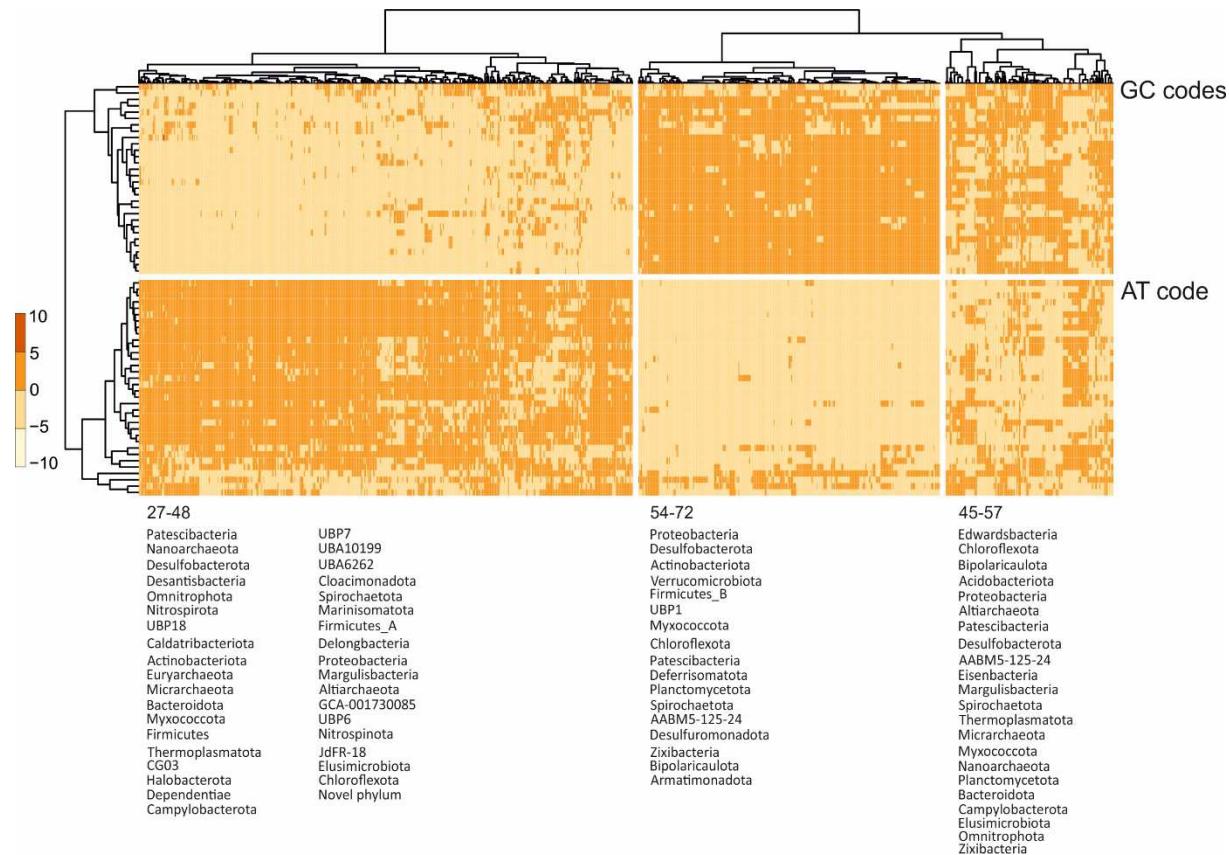


Figure 6- Expression level of functional classes involved in public good provision in the sequenced metatranscriptomes. The list of screened K0 identifiers are shown in **Supplementary Table S6**.

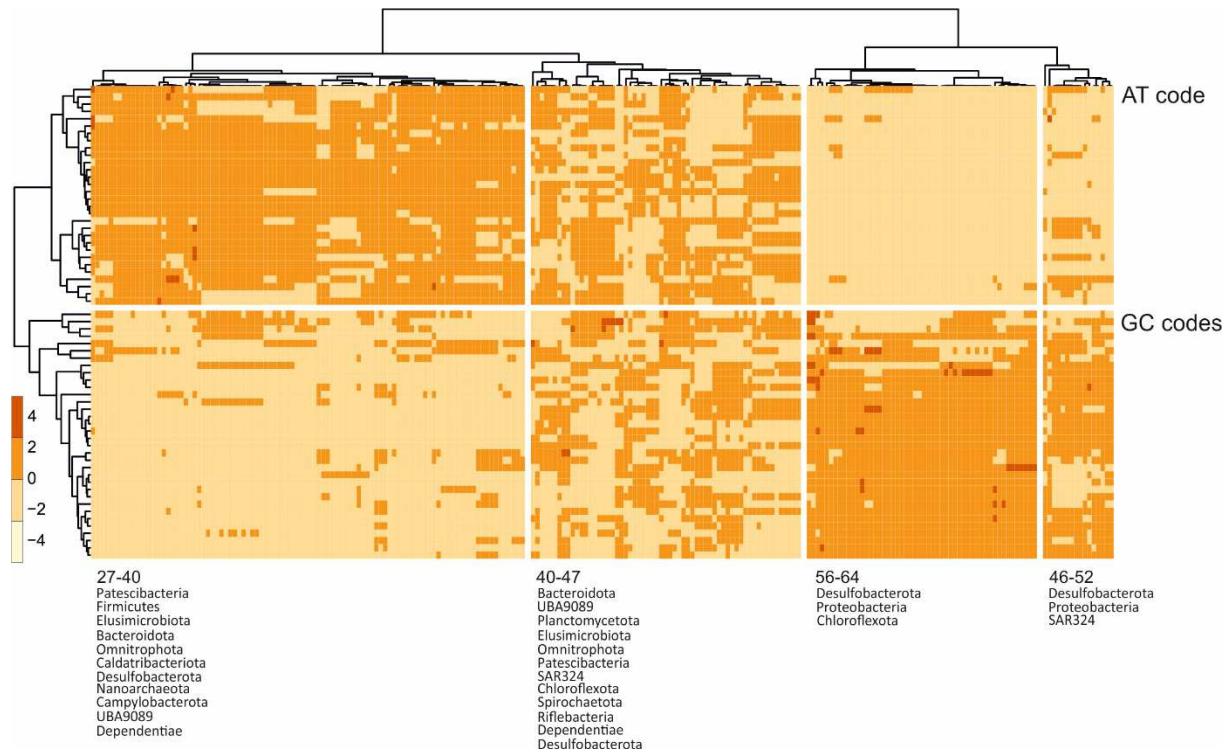
Supplementary Figures



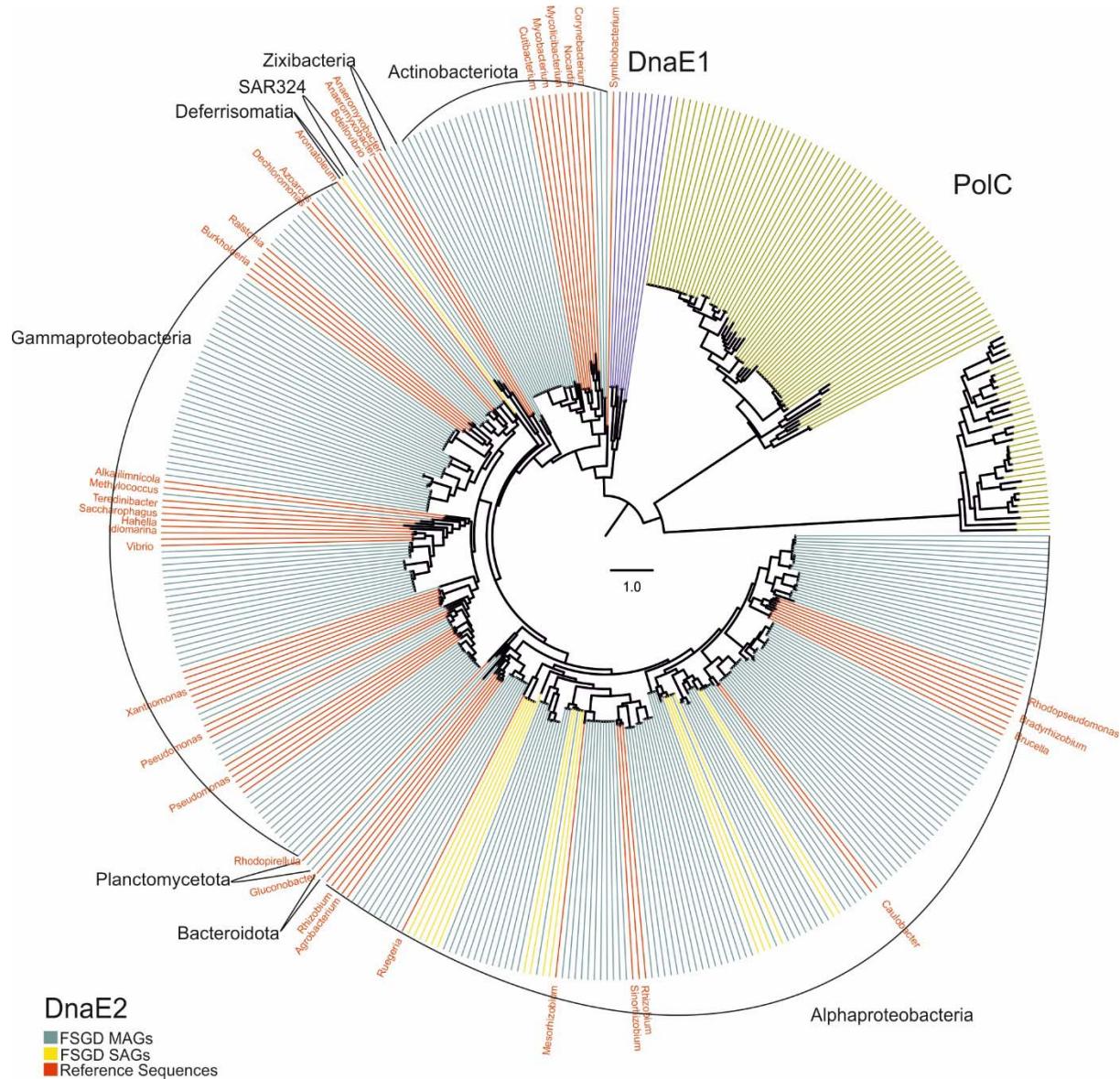
Supplementary Figure S1- Geographical distribution of publicly available metagenomic datasets sequenced from oligotrophic groundwater samples (landfill groundwater, oil-influenced, and shale samples are not included in this representation). The depth ranges (as mentioned in the corresponding publication) of the samples (in red) and the amount of publicly available sequenced data (presented as Gb) are shown for each location.



Supplementary Figure S2- Representation of frequency (the expected number of codons, given the input sequences, per 1000 bases) of utilization of synonymous codons across MAGs and SAGs of the FSGD.



Supplementary Figure S3- Representation of frequency (the expected number of codons, given the input sequences, per 1000 bases) of utilization of synonymous codons across highly expressed (TPM >10000 arbitrary threshold) MAGs and SAGs of the FSGD.



Supplementary Figure S4- Reconstructed phylogeny of the C-family polymerases. The tree is arbitrarily rooted the PolC branch. Taxonomy of the leafs are written for each group. The reference genomes are the reviewed sequences retrieved from uniprot for each polymerase type (red).