

1 Characterization of sediment and granite hosted deep underground research laboratories
2 reveals diverse microbiome functions, limited temporal variation and substantial genomic
3 conservation

4

5 Yuki Amano^{1,5,*}, Rohan Sachdeva², Daniel Gittins², Karthik Anantharaman³, Shufei Lei⁴, Luis E. Valentin-
6 Alvarado², Spencer Diamond², Hikari Beppu¹, Teruki Iwatsuki⁵, Akihito Mochizuki⁵, Kazuya Miyakawa⁵,
7 Eiichi Ishii⁵, Hiroaki Murakami⁵, Alexander L. Jaffe⁶, Cindy Castelle², Adi Lavy⁴, Yohey Suzuki⁷ and Jillian
8 F. Banfield^{2,*}

8 F. Banfield^{2,*}

9

* Corresponding Authors:

¹¹ amano.yuki@jaea.go.jp, jbanfield@berkeley.edu

12

Author affiliations:

¹ Nuclear Fuel Cycle Engineering Laboratories, Japan Atomic Energy Agency, 4-33 Muramatsu Tokai,

15 Ibaraki, Japan² Innovative Genomics Institute, University of California, Berkeley, CA, 94720, USA,

¹⁶ ³Department of Bacteriology, University of Wisconsin-Madison, Madison, WI, 53706, USA, ⁴ Department

of Earth and Planetary Science, University of California Berkeley, Berkeley, CA, 94720, USA, ⁵ Horonobe

18 Underground Research Center, Japan Atomic Energy Agency, 432-2 Hokushin, Horonobe-cho, Hokkaido,

19 Japan,⁶Department of Plant and Microbial Biology, University of California, Berkeley, Berkeley, CA,

20 94720, USA, ⁷Department of Earth and Planetary Science, The University of Tokyo, 7-3-1 Hongo,

21 Bunkyo-ku, Tokyo, Japan

22

23 **Abstract**

24 Underground research laboratories (URLs) provide a window on the deep biosphere and enable
25 investigation of potential microbial impacts on nuclear waste, CO₂ and H₂ stored in the subsurface. We
26 carried out the first multi-year study of groundwater microbiomes sampled from defined intervals
27 between 140 and 400 m below the surface of the Horonobe and Mizunami URLs, Japan. We reconstructed
28 draft genomes for >90% of all organisms detected over a four year period. The Horonobe and Mizunami
29 microbiomes are dissimilar, likely because the Mizunami URL is hosted in granitic rock and the Horonobe
30 URL in sedimentary rock. Despite this, hydrogen metabolism, rubisco-based CO₂ fixation, reduction of
31 nitrogen compounds and sulfate reduction are well represented functions in microbiomes from both URLs,
32 although methane metabolism is more prevalent at the organic- and CO₂-rich Horonobe URL. High fluid
33 flow zones and proximity to subsurface tunnels select for candidate phyla radiation bacteria in the
34 Mizunami URL. We detected near-identical genotypes for approximately one third of all genetically
35 defined organisms at multiple depths within the Horonobe URL. This cannot be explained by inactivity, as
36 in situ growth was detected for some bacteria, albeit at slow rates. Given the current low hydraulic
37 conductivity and groundwater compositional heterogeneity, ongoing inter-site strain dispersal seems
38 unlikely. Alternatively, the Horonobe URL microbiome homogeneity may be explained by higher
39 groundwater mobility during the last glacial period. Genotypically-defined species closely related to those
40 detected in the URLs were identified in three other subsurface environments in the USA. Thus, dispersal
41 rates between widely separated underground sites may be fast enough relative to mutation rates to have
42 precluded substantial divergence in species composition. Species overlaps between subsurface locations
43 on different continents constrain expectations regarding the scale of global subsurface biodiversity.
44 Overall, microbiome and geochemical stability over the study period has important implications for
45 underground storage applications.

46

47 **Key words:** Underground Research Laboratory, metagenomics microbiome, stability, groundwater,
48 sedimentary rocks, granite

49

50 **Introduction**

51 The subsurface, especially hundreds of meters below ground, is one of the last biological frontiers
52 (Teske et al., 2013; Hoshino et al., 2020). Beyond basic scientific interest in organisms that live there, the
53 microbiology of these regions is an important practical consideration if they are to be used to store nuclear
54 waste and other materials (e.g., CO₂ or H₂). Microbes in proximity to underground repositories may impact
55 the stability of the stored material via container corrosion (e.g., Pedersen 1999; Rajala et al., 2015; Stroes-
56 Gascoyne et al., 2010; Stroes-Gascoyne and West, 1997), by consumption of stored resources (e.g., H₂,
57 Dopffel et al., 2023; Liu et al., 2023) or by degrading containment integrity (e.g., via increasing rock
58 porosity and/or permeability, Zeng et al., 2023).

59 Much of what we know about subsurface life (below the soil zone) comes from studies of
60 groundwater. Genome-resolved microbiome studies of 3 - 6 m depth groundwater in an aquifer adjacent
61 to the Colorado River, Rifle, Colorado brought to light dozens of new lineages of bacteria and archaea
62 (Wrighton et al., 2012; Anantharaman et al., 2016). Organisms from these groups have since been
63 described from numerous other ecosystems, including northern California groundwater aquifers (He et al.
64 2021) and groundwater delivered to the surface via eruption of a cold CO₂-driven geyser (e.g., Probst et
65 al., 2018). However, these systems provide limited insights into deep subsurface microbiology. Recovery
66 of deep underground samples for biological analysis is possible via drilling, but drilling provides samples
67 with little context, interpretation of the results may be complicated by contamination (e.g., from drilling
68 fluids), and samples represent only a single time point (Mouser et al. 2016). For this reason, the
69 construction of large, human-accessible deep subsurface research laboratories has been an important
70 development. Within these underground research laboratories microbiological samples can be collected

71 from well defined sites with minimal contamination. Important examples of such laboratories include
72 Äspö in Sweden (Hallbeck & Pedersen, 2008; Mehrshad et al., 2021), Grimsel Test Site in Switzerland
73 (Konno et al., 2013), the Mont Terri Underground Rock Laboratory in Switzerland, the Horonobe
74 Underground Research Laboratory (Horonobe URL) in Hokkaido, Japan and the Mizunami Underground
75 Research Laboratory (Mizunami URL) in Gifu, Japan. Here, we conducted the first multi-year investigation
76 of subsurface microbial diversity and metabolic capacities in microbiomes of the two Japanese URLs. This
77 work follows one prior publication that considered microbial metabolisms in some Horonobe URL
78 microbial communities (Hernsdorf et al., 2017) and another that ecologically and genomically profiled
79 anaerobic methane-oxidizing archaea in the Mizunami URL (Ino et al., 2017). Our analyses provide a
80 comprehensive genomics-based overview of microbial and metabolic diversity and spatial and temporal
81 variation. We uncover genotypic overlap within each URL and likely geologically-based differences
82 between the microbiomes of the URLs. Finally, we compare genotypes from the URLs to genotypes of
83 bacteria and archaea sampled from three underground locations in the USA. The results begin to address
84 the '(to what extent is) everything everywhere' question for microbiomes that exist way below the Earth's'
85 surface, thus providing clues to overall levels of subsurface microbial diversity.

86

87 **Results**

88

89 Our research was conducted primarily in two underground research laboratories (URLs) in Japan (**Figure**
90 **1a**). From the Mizunami URL (35°22'40.68"N, 137°14'15.63"E) we analyzed metagenomic data from 7
91 groundwater samples that were recovered from 200 to 400 m below the surface (**Table S1; Figure 1b**).
92 During URL construction, 4-7 years prior to the first sample collection in 2014, ~10 cm diameter, ~100 m
93 long boreholes were established by coring into rock from tunnel walls. At six locations in three access
94 tunnels (200 m, 300 m, 400 m below the surface), water was extracted from sampling zones at 26.9 - 96.1

95 m distances from the tunnels (**Table S2**). The seventh sample was collected from the 200 m depth site in
96 2015. Low ionic strength groundwater in these regions is organic-poor, Na^+ - Ca^{2+} - Cl^- brine (Iwatsuki et al.,
97 2005; Hayashida et al., 2016), with temperatures ranging between 15 and 23 °C. The pH ranges from 8.7
98 to 9.1 and Eh values range from -180 to -60.7 mV (**Table S3**). From the Horonobe URL (45°02'41.92"N,
99 141°51'34.20"E) we analyzed metagenomic data from 19 groundwater samples collected from five
100 locations at 4 - 57 m from the tunnels (**Table S1; S2; Figure 1c; Figure S1**). Groundwater was recovered
101 from 140 m and 250 m below the surface via tubes into defined intervals of boreholes into the surrounding
102 rock. The Horonobe groundwater is organic rich and saline, dominated by Na^+ - Cl^- - HCO_3^- (**Table S3**,
103 Amano et al., 2012; Sasamoto et al., 2015). Temperatures range from 13 - 25 °C, pH ranges from 6.4 to
104 7.2, and Eh values range from -315 to -204 mV. Unlike at Mizunami, the Horonobe groundwater is
105 saturated with both CH_4 and CO_2 (Miyakawa et al., 2017; Tamamura et al., 2018), which degas during
106 sample recovery.

107 Total DNA was extracted from the 26 samples and between 10.6 and 16.4 Gbp of 150 bp paired
108 end Illumina sequences obtained from each sample. The sequences from each sample were assembled
109 independently and genomes reconstructed. After gene prediction and functional annotation, a census of
110 organisms was performed based on phylogenetic classification of ribosomal protein S3 (rpS3) sequences.
111 The abundance of each organism was determined based on read coverage values for the rpS3-bearing
112 scaffolds. We reconstructed 1154 rpS3 sequences from the Horonobe metagenomes and 624 from
113 Mizunami metagenomes. Phylogenetic analysis revealed that the vast majority of rpS3 sequences from
114 both the Mizunami and Horonobe URLs were from Candidate Phyla, major groups that lack even a single
115 isolated representative (**Figure 2; Data S1**). Up to 21 % of Mizunami rpS3 sequences place within the
116 DPANN. No DPANN sequences were recovered from the Horonobe URL, although Altiarchaeota are
117 present (up to 90% of sequences) and they are sometimes grouped with DPANN (Adam et al., 2017). Up

118 to 47% of Mizunami sequences and up to 41% of Horonobe sequences are from bacteria of the Candidate
119 Phyla Radiation (CPR; Brown et al. 2015; **Figure 3, 4**).

120 Initially, we profiled community composition over space and time using high level (mostly phylum)
121 phylogenetic groupings (**Figure 3, Figure S2**). However, for the Horonobe URL we also evaluated class level
122 diversity and determined that the phyla with many species representatives also have species assigned to
123 more classes, predicted by a linear trend (slope of 0.39; **Figure S3a**).

124 We hypothesized that URL construction may have perturbed, and may continue to perturb, the
125 microbiomes and thus community composition may correlate with distance from the access tunnels.
126 Despite relatively consistent geological settings within each URL there are substantial differences in
127 phylum-level community composition between some sites. Within the Mizunami URL, the abundances of
128 Parcubacteria and Micrarchaeia (DPANN) vary substantially, and they were essentially absent from one
129 300 m depth site in a low-fluid flow zone. In the Horonobe microbiomes, Altarchaeales (SM1) were
130 abundant only in the two sites (140 m and 250 m) located very close to the URL tunnels. Horonobe samples
131 collected at intermediate distances from the tunnels featured abundant *Deltaproteobacteria*,
132 *Methanomicrobia*, *Firmicutes*, *Betaproteobacteria*, and *Bacteroidetes*. *Candidatus Saccharibacteria* was
133 abundantly detected only in 2013.

134 *Methanoperedens* (ANME-2d) were prevalent at the site most distant from the tunnels.
135 Parcubacteria and other CPR bacteria were abundant in samples collected from high fluid flow zones at
136 distance from the tunnels. Despite site to site differences, Horonobe samples collected over three or four
137 years from the same sites showed little variation in microbiome composition (**Figure 3b**).

138 Given the difference in surrounding rock type between the two URLs, we were interested to know
139 how similar the microbiomes of the granitic hosted Mizunami URL are to those of the sediment-hosted
140 Horonobe URL. The similar microbiome compositions within each sampling location over time made it
141 reasonable to average phylum compositions at each location to enable comparison of the URLs at a high

142 taxonomic level (**Figure 3**). For both URLs, the 15 most abundant organisms are from 15 different major
143 (mostly phylum-level) groups and the URLs share only 8 of these (Firmicutes, Alphaproteobacteria,
144 Bacteroidetes, Chloroflexi, ANME-2d (Methanoperedenceae), Betaproteobacteria and Elusimicrobia).
145 Deltaproteobacteria were far more abundant in Horonobe compared to Mizunami samples. Nanoarchaeal
146 Micrarchaeota are only prevalent in the Mizunami URL whereas Altarchaeales archaea (SM1) and
147 Saccharibacteria (TM7) are only prominent in the Horonobe URL (**Figure 2,3; Table S4**). The CPR bacterial
148 types present in Horonobe samples (Peregrinibacteria, Saccharibacteria and Dojkbacteria) are different
149 from those in the Mizunami samples (primarily diverse Parcubacteria groups) (**Figure 4; Table S4**). Given
150 these results and the relative consistency of groundwater microbiology over time, we conclude that the
151 microbiomes of the Mizunami and Horonobe URLs are distinct at high taxonomic levels. We also evaluated
152 microbial overlap between the URLs at the species level. Only 15 out of 490 species occur in both URLs (3
153 Actinobacteria, 1 Bacteroidetes, 1 Betaproteobacteria, 4 Deltaproteobacteria, 1 Firmicutes, 2
154 Elusimicrobia, 1 Ignavibacteria, 1 Spirochaete and 1 Candidatus Kuenenbacteria).

155 Dereplication of the 225 draft genomes from the Mizunami URL and 265 draft genomes from the
156 Horonobe URL yielded 489 genomes in total, implying just one genomically defined organism (Clostridia,
157 Species 4) was shared between the two URLs. The reconstructed genomes account for 93% and 90% of all
158 organisms detected at the Mizunami and Horonobe URLs, respectively, based on rpS3 analysis. Organisms
159 lacking a draft genome occur in only a small subset of samples (for Mizunami, all were detected in only
160 one sample). Thus, for both URLs, we conclude that the microbiomes of both URLs are now extensively
161 genomically sampled.

162 Analysis of the presence/absence patterns of all species across samples (**Figure 5, Table S5**)
163 showed that many organisms are present in multiple sites within each URLs. One organism, a
164 *Rhodobacterales*, occurred in 16 of the 19 Horonobe sampling sites. Hierarchical clustering of the
165 presence absence patterns for the Horonobe URL (**Figure S3b**) showed that community compositional

166 similarities are well predicted by the sampling site (i.e., samples collected at the same site in different
167 years cluster together). Clustering of Horonobe samples based on organism presence/ absence patterns
168 supports the hypothesis that microbiome distance of the access tunnels may impact the organisms
169 present.

170 The extensive genomic sampling of both URL microbiomes made it reasonable to use the genomes
171 to provide an overview of biogeochemical capacities in these subsurface ecosystems. Our analysis
172 targeted genes involved in CO₂ fixation, hydrogen cycling, nitrogen cycling and sulfur cycling. For example,
173 we analyzed the diversity of Rubisco large subunit sequences, focusing primarily on forms I and II that co-
174 occur with phosphoribulokinase, as these genes are indicative of the capacity for CO₂ fixation via the CBB
175 cycle (**Figure S4; Table S6**). Genes for this pathway were abundant, and detected in all samples of both
176 URLs. Hydrogenase-encoding genes were recovered from all the samples from both URLs, indicating a
177 broad distribution (**Table S7**). Most of the hydrogenases in Mizunami were assigned to groups 1a, 3b, 3c,
178 3d, 4g [NiFe]-hydrogenases and groups A3 and C3 [FeFe]-hydrogenases. Interestingly, the most abundant
179 [NiFe]-hydrogenase enzymes in Mizunami URL are from group 3b, which is predicted to oxidize NADPH
180 and evolve hydrogen, maybe reversibly (Greening et al., 2016). These complexes may have
181 sulfhydrogenase activity (elemental sulfur (S⁰) reduced to polysulfide to H₂S; Ma et al., 1993). Group 1a
182 [NiFe]-hydrogenase is predicted to oxidize H₂ and is predominantly found in anaerobic Firmicutes and
183 Deltaproteobacteria capable of H₂-dependent sulfate reduction and metal reduction. On the other hand,
184 the hydrogenases from the Horonobe URL were from all of the four major [NiFe] groups and the three
185 major [FeFe] groups. The abundant [NiFe]-hydrogenases were assigned to groups 3b, 3c, 3d, and 4f. Group
186 4 [NiFe]-hydrogenases are predicted to be membrane-bound H₂-evolving enzymes. Although the function
187 of group 4f is uncertain, it may couple oxidation of a one-carbon compound to proton reduction
188 concurrent with proton translocation. Overall, the indications are that H₂ is an important energy currency
189 in these anoxic subsurface environments.

190 Abundant genes for nitrogen cycling were identified in genomes from both URLs (**Table S8**). These
191 include genes involved in nitrogen fixation, nitrate reduction, nitrite reduction to ammonia and nitric
192 oxide reduction. In the Mizunami URL, up to 53% of genomes encode genes for nitrite reduction to
193 ammonia compared to up to 13% in the Horonobe URL. No genes for oxidation of ammonia were detected
194 in any sample from either URL, which may explain the very high concentrations of ammonia in the
195 groundwater at Horonobe (Sasamoto et al., 2018). At Mizunami, a widely distributed *Rhodocyclales* has
196 the capacity for nitrate reduction, nitrite reduction to ammonia and sulfate reduction. The
197 *Rhodobacterales* that is the most widely distributed organism within the Horonobe URL also has genes for
198 reduction of nitrate, nitrite reduction, nitric oxide and sulfur cycling. Similar capacities are predicted for
199 other widely distributed bacteria (e.g., Gammaproteobacteria).

200 Genes for sulfate reduction, sulfite reduction, sulfur oxidation, and thiosulfate oxidation, were
201 observed in genomes of organisms from both URLs, but are more prevalent in the genomes of organisms
202 from Mizunami compared to Horonobe (**Table S8**). Abundant genes for sulfur dioxygenase (sdo), which
203 could be involved in oxidation of elemental sulfur to sulfite, were detected in both URLs. Genes for sulfate
204 reduction were particularly prevalent in most *Nitrospirae* and some Deltaproteobacteria from the
205 Mizunami URL and diverse Deltaproteobacteria from the Horonobe URL. However, the two most
206 abundant *Nitrospirae* associated with the Mizunami URL apparently lack this capacity. Interestingly, genes
207 for selenate reduction were prevalent in genomes of Deltaproteobacteria and Spirochaetes from both
208 URL microbiomes. In fact, selenate reduction is the only biogeochemically relevant capacity predicted for
209 a Spirochaete that occurs in 15 of the 19 Horonobe sampling sites.

210 We asked how similar the collections of genomically defined organisms from Mizunami and
211 Horonobe were to groups of organisms from other genomically well sampled subsurface groundwater
212 systems. Our comparison set comprised organisms from the two URLs, from the cold, CO₂ driven Crystal
213 Geyser that taps into deep aquifers below the Colorado Plateau (Probst et al. 2018), the shallow

214 groundwater aquifer adjacent to the Colorado River, Rifle, Colorado (Anantharaman et al. 2016), and
215 groundwater from northern California (He et al. 2021). We chose these due to their high genomic
216 resolution and, in part, because the metagenomic datasets were processed using essentially the same
217 protocols as the current study. We did not include comparisons to soils, as there have been recent reviews
218 that show that soils worldwide are dominated by Acidobacteria, Actinobacteria and Gemmatimonadetes,
219 groups generally not prominent in the URL environments. We used each of the 1472 rpS3 sequences from
220 genomes from the five locations to approximate species and compared levels of species overlap between
221 locations. Notably ~45% of the Horonobe sequences are most similar to sequences from the sediment-
222 hosted Rifle aquifer (**Table S9**).

223 To minimize the effect of substantial differences in the number of genomes reconstructed from
224 each location on comparisons, we also evaluated species-level similarity based on average % amino acid
225 identity (aa% ID) of the best matching sequences for each rpS3 protein (**Figure 6**). Notably, the northern
226 California groundwater genome set shows statistically significant differences in similarity in all
227 comparisons (i.e., more similar to the set from Rifle than Mizunami, more similar to the set from Mizunami
228 than from Horonobe, etc.). The genomes from Horonobe are more similar to those from Rifle compared
229 to northern California groundwater ($p \leq 0.05$) and less similar to Mizunami, although the statistical
230 significance is just above our cutoff ($p > 0.05$).

231 We evaluated genome novelty at all five locations by considering only those organisms whose
232 rpS3 protein shared < 50% aa% ID with the highest scoring sequence from a comparison location. The
233 most divergent genomes sampled from the Horonobe URL are for Hadesarchaeota, Thermoplasmatales
234 or Bathyarchaeota, and from the Mizunami URL, were DPANN archaea. Notably, ~80% of most divergent
235 sequences across the five datasets were from Archaea (almost half of these are DPANN) and almost half
236 of the bacterial cases were CPR bacteria.

237 Generally lacking to date have been analyses that leverage genome collections to investigate the
238 extent of genomic overlap of highly related (i.e., rpS3 proteins with > 99% aa% ID) organisms across
239 regions distant from each other geographically (e.g., different continents) and/or physically (e.g., due to
240 depth below the surface). We sought cases where highly related organisms were shared between
241 subsurface locations in the USA and the two Japanese URLs and found 106 shared bacteria and archaea
242 from 16 different classes/phyla (Table S10). Most of the pairs were from Horonobe and Rifle or northern
243 California groundwater and Rifle. Eight genomes with identical rpS3 protein sequences (from 6 different
244 phyla/classes) are classified as representing the same species based on >95% genome average nucleotide
245 identity (**Table S10**). One Bacteroidetes species (genus *Lentimicrobium*) from Horonobe (detected from
246 167 m, 218 m and 250 m samples) shares 99 % ANI with a genome from the Rifle aquifer. The
247 *Lentimicrobium* genomes are nearly perfectly syntenous across aligned regions (**Figure S5**), consistent
248 with these organisms being very closely related. Thus, we conclude that there is some strain-level overlap
249 between geographically distant shallow and deep subsurface regions.

250 To put the results of this multi-location comparison into context we turned back to the Horonobe
251 dataset, which has sufficient sampling to provide insight into genomic similarity for organisms sampled
252 from the same site at different times and from different depths and locations. We identified 99 distinct
253 genome clusters (451 genomes), each of which is comprised of genomes that share > 99.5% ANI over >
254 95% of the genome alignment. Of the 99 Horonobe clusters, 17 consisted only of organisms from the same
255 site sampled in different years. For example, a Peregrinibacterium for which we manually curated a
256 998,424 bp complete genome (167 m sample) shares 99.997% ANI over the entire genome with a genome
257 collected a year later (and 100% ANI over 993 kbp) from the same site (this genotype was detected in all
258 four years). The remaining 83% of clusters, containing members from different depths within the URL,
259 represent 31% of all genomes reconstructed. Organisms from the shallowest depth (140 m) were least
260 commonly clustered with organisms from other depths. However, when they were clustered, it was

261 typically with organisms from the deepest site (250 m). Approximately one third of the clusters contain
262 organisms from three different depths and 12% contain organisms from the four deeper sites. Strikingly,
263 some clusters included genomes from different sites with average pairwise ANI values of >99.997 %. For
264 example, clusters of genomes for *Saccharibacterium*, *Clostridium*, *Syntrophobacterales* and a
265 *Gammaproteobacteria* from four different depths all have \geq 99.98% pairwise ANI. The complete
266 *Peregrinibacteria* genome (167 m sample) is identical to a genome comprised of three contigs from the
267 194 m depth site (98.8% genome alignment). Thus, we conclude that extremely closely related bacteria
268 occur at sites separated by large volumes of rock.

269 We considered the possibility that the very closely related species from different depths at the
270 Horonobe URL may be a distinct, possibly inactive subset of the microbial communities. However, the
271 taxonomic affiliations of the widely distributed species is well predicted by the taxonomic composition of
272 the overall Horonobe genome set. Thus, we considered the possibility that the genomic homogeneity of
273 species throughout the Horonobe URL is due to overall microbiome inactivity. We tested for in situ growth
274 by calculating an index of replication (iRep; Brown et al. 2016) for a subset of organisms present at a
275 variety of abundance levels. The *Peregrinibacteria* genome that was highly conserved in the 167 m and
276 194 m samples had an iRep value of 1.14, similar to values for other abundant microbes affiliated with the
277 Firmicutes (1.24), Methylphilales (1.12) and Rhodobacterales (1.26). Low coverage values limited the set
278 of genomes for which this calculation was possible.

279

280 **Discussion**

281

282 Underground research laboratories provide unique access to the deep subsurface biosphere. Findings
283 from such laboratories are relevant for understanding Earth's microbiomes broadly, and may inform plans
284 to use the subsurface for storage of energy resources and waste products. One question pertains to how

285 fast microbiome perturbation due to URL construction wanes as microbiomes approach a pre-
286 construction state over the following years. The Horonobe time series data indicate that the subsurface
287 microbiomes are not undergoing rapid changes that could be indicative of rebound following laboratory
288 construction in the five to nine years since the URL was constructed.

289 Any long term impact of the URLs on the subsurface should diminish with increasing distance from
290 the access shafts and tunnels. Clustering of microbiomes based on organism presence/absence patterns
291 generally is consistent with this prediction. The anomalously high abundance of SM1 archaea only close
292 to the Horonobe tunnels may be explained by degassing of CO₂ due to alteration of groundwater
293 conditions. High CO₂ conditions occur where this archaeon has been found elsewhere in high abundance
294 (e.g., the CO₂-driven Crystal Geyser, Probst et al. 2018).

295 The prevalence at Horonobe of *Methanoperedens*, an archaeon implicated in anaerobic methane
296 oxidation, only at a site that is distant from the repository is interesting, given high levels of methane
297 throughout the URL. Based on genomic information, methane oxidation is likely coupled to reduction of
298 ferric iron (Hernsdorf et al., 2017; Nishimura et al., 2023). The sample was collected at the terminus of a
299 downward inclined borehole where iron may be liberated from clay that accumulates there (Nishimura et
300 al., 2023). Supporting the possibility that clay could stimulate growth of *Methanoperedens*, only the
301 groundwater sampled from this site contained muddy suspended solids.

302 An intriguing finding regarding the Mizunami microbiomes is the high representation of CPR
303 bacteria and DPANN archaea. Consistent with many prior studies (e.g., Wrighton et al. 2012), their gene
304 inventories indicate that they are unlikely to be capable of living independently. CPR bacteria and DPANN
305 archaea may be abundant in Mizunami high groundwater flow zones because low porosity enables
306 selective mobilization of only small cells from more complex consortia attached to rock surfaces. A similar
307 conclusion was suggested previously to explain the very high representation of CPR bacteria and DPANN
308 archaea in the Crystal Geyser system (Probst et al., 2018).

309 We noted many differences in the taxa associated with the granitic vs. sediment-associated
310 microbiomes that likely relate to the rock mineralogy. Granitic minerals such as biotite (iron, manganese),
311 amphibole (iron, manganese), ilmenite (iron), chlorite (iron) (Yuguchi et al., 2010; 2018; Iwatsuki et al.,
312 1999; Ishibashi & Yuguchi., 2017), Fe-, Mn-bearing calcite (Iwatsuki et al., 2000) and pyrite (e.g., iron,
313 sulfur and selenium; Iwatsuki et al., 2002) may support microbial energy generation. Abiotic reactions
314 involving minerals can be a source of methane and H₂ (Stevens and McKinley, 1995), but carbon isotopic
315 analyses of CH₄ in the groundwater support a mixture of abiotic (Ino et al., 2017) and biogenic origin (Mills
316 et al., 2010) and microbial consortia include methanogens. In the sedimentary rocks surrounding the
317 Horonobe URL, buried organic matter, ferric iron oxides, iron-bearing clays (e.g., smectite) and pyrite may
318 provide energy resources. High levels of dissolved CO₂, H₂, NH₃ and CH₄ in the groundwater are probably
319 byproducts of microbial metabolism (Miyakawa et al., 2017). H₂, NH₃, and CH₄ may persist in the Horonobe
320 URL due to low availability of electron acceptors needed for the reactions that would consume them.

321 Given the possibility of Horonobe microbiome disruption close to the tunnels (e.g., Purkamo et al.
322 2018), communities sampled further from the tunnels may be more representative of the microbiomes
323 surrounding the URL. Our data indicate that these communities are dominated by Deltaproteobacteria,
324 especially *Syntrophobacterales*, which are predicted to be autotrophic sulfate-reducers that may also
325 produce H₂, supporting methanogenesis by abundant *Methanospirillum*. Betaproteobacteria such as
326 *Methylophilales* are also relatively abundant in these samples and are implicated in methanol oxidation,
327 the source of which may be methane oxidation linked to nitrate reduction, as performed by
328 Gammaproteobacteria such as *Methylomonas* (Kits et al., 2015). Methanol oxidation is probably coupled
329 to nitrate reduction given the lack of O₂ (analogous to the process performed by *Methylomirabilis oxyfera*;
330 Wu et al., 2015). The *Rhodocyclales* that are also abundant in these consortia are implicated in both nitrate
331 reduction and sulfur oxidation.

332 A recent study provided insights into possible relationships between the composition of
333 subsurface microbiomes and host rock type in the terrestrial subsurface (Soares et al. 2023). Despite low
334 sample sizes resulting in a lack of statistical support, distinct lithologies were suggested to host distinct
335 microbiomes, which is consistent with the taxonomic differences between the Horonobe and Mizunami
336 URLs. The authors noted the widespread occurrence of Burkholderiales, Gammaproteobacteria and
337 Clostridia, also groups represented in the URL microbiomes. However, we also note the prevalence of
338 Parcubacteria (CPR bacteria), Deltaproteobacteria, Chloroflexi, Nitrospirae, and Bacteroidetes and
339 numerous other groups, often candidate phyla (e.g., Altarchiales and *Methanoperedens*). Thus, this study
340 expands understanding of subsurface microbiome diversity, in part by providing information about
341 species diversity and spatial variability.

342 Using the findings of the current study we can predict some impacts of the granite and sediment
343 hosted microbiomes on geological disposal and storage. For example, proliferation of sulfate-reducing
344 bacteria will generate sulfide that can corrode the metal containers potentially used to store radioactive
345 waste (Pedersen, 2010; Stroes-Gascoyne, 2010; Enning and Garrels, 2014). It has been noted that
346 production of sulfide by microbial communities in the terrestrial deep subsurface may be occurring even
347 in the absence of geochemical evidence (Bell et al. 2020), highlighting the importance of microbial
348 characterisation of these systems. Methanogens could also be problematic in that they can colonize the
349 surfaces of steel containers and use the iron as the electron donor for methane production (Dinh et al.,
350 2004; Mori et al., 2010; Hirano et al., 2022). Both sulfate reduction and methane production can be
351 coupled to hydrogen metabolism, for which there are abundant genes. This raises the question of the
352 extent to which hydrogen metabolizing microbes could impact subsurface H₂ storage.

353 This study may be of interest beyond its immediate relevance to underground repositories or
354 biodiversity and metabolism in granitic and sedimentary rock environments. Our analyses of microbial
355 species overlap within subsurface locations (e.g., across depths of the Horonobe URL) and between

356 subsurface locations on different continents may constrain rates of dispersal and genome mutation and
357 calibrate intuition regarding global biodiversity (i.e., to what extent are the species or genera in every
358 spatially separated underground site different?). The answer should depend on the degree of
359 interconnection within the subsurface, the time since dispersal between distant locations, transit times
360 from the surface to the subsurface, and evolutionary rates.

361 Considering first the case of within site dispersal, our results point to substantial genotypic overlap
362 (i.e., very limited evolutionary divergence) within the Horonobe URL despite separation of the sampling
363 sites by up to hundreds of meters of solid rock. This raises the question of whether genotype overlaps
364 across sites can be explained by groundwater-mediated microbial transport. Comparison of $d^{18}\text{O}$ and
365 $d\text{D}$ between pore water and pumped groundwater indicates that there are “active” and “inactive” zones
366 of groundwater flow within the Horonobe URL (Mochizuki et al., 2022). The groundwater samples
367 collected for microbiology analyses are from “inactive” regions (**Fig. S6**). If measured hydraulic
368 conductivities (**Table S2**) are predictive of transit times, microbial dispersal on thousands of years
369 time scale might be possible. However, spatial heterogeneity in groundwater composition indicates
370 extremely low interconnectivity, probably ruling out cell movement via water flow under the present
371 hydraulic regime.

372 Alternatively, it is possible that microbial dispersal occurred under conditions that differ from
373 the present. H and O stable isotopic measurements indicate that meteoric water was intruded into
374 the Horonobe subsurface during glaciation (likely the Last Glacial Stage, between 12,000 and 70,000
375 years ago), when hydraulic gradients were higher than at present (Teramoto et al., 2010; Ishii, 2018;
376 Nakata et al., 2018; Mochizuki & Ishii, 2022). Intrusion of meteoric water probably ceased when
377 hydraulic gradients decreased due to sea level rise (Hanatani et al., 2010). If genomic overlap is due
378 to microbial dispersal via groundwater flow $>12,000$ years ago, less than one mutation was fixed on

379 average every ~400 years in the Peregrinibacteria genome. We do not attribute genome consistency to
380 inactivity because data indicate slow replication of these bacteria at the time of sampling. Some level of
381 microbial activity is unsurprising, as an investigation of the deep terrestrial biosphere in the Äspö Hard
382 Rock Laboratory in Sweden showed that subsurface populations are normally viable, with fast degradation
383 of non-viable microbes likely resulting in the absence of populations that are not adapted to grow in
384 oligotrophic, subsurface conditions (Lopez-Fernandez et al. 2018).

385 Assuming inter-site transport was possible, environmental heterogeneity could have precluded
386 colonization by transported organisms. Thus, we attribute the genomic homogeneity within the Horonobe
387 URL in part to consistent and stable geochemical and physical environments (as revealed by *in situ*
388 measurements; Miyakawa et al., 2017; Sasamoto et al., 2015; Mochizuki & Ishii, 2022; 2023). The finding
389 of biological stability is important for assessment of engineered subsurface environments, and thus long-
390 term confinement of radioactive waste. Geological, hydrological, and geochemical stability for periods of
391 at least tens of thousands of years is recommended for geological disposal (IAEA, 2003).

392 We detected closely related species overlaps between the Japanese URLs and underground
393 locations in the USA. If inter-continental dispersal via air and/or water is relatively facile, and given the
394 possibility of surface to Horonobe deep subsurface transport on the scale of tens of thousands of years,
395 and in view of likely surface to subsurface groundwater movement rates at the USA locations
396 (Supplementary Material), microbial dispersal may explain low levels of strain- to species level divergence
397 in subsurface locations on two continents. These preliminary analyses were possible because large
398 genome-resolved metagenomic datasets are available for several subsurface locations. Datasets of high
399 quality genomes are being generated for many places around the world, and will enable further analyses
400 of strain and species overlaps over geographically separated locations. Such research will advance our
401 understanding of processes that structure the microbial biosphere, and by implication, the extent of
402 microbial biodiversity in the Earth's subsurface.

403

404 **Conclusions**

405 This study was possible due to many years of planning and development to establish the two
406 Japanese URLs as subsurface microbiology research sites. This is the first study to genomically evaluate
407 the microbiomes in detail, over multiple years, and to compare the ecosystems in terms of community
408 composition and microbiome function. The URL microbiomes are dominated by little known (i.e.,
409 candidate phyla) bacteria and archaea. Hydrogen, sulfur and nitrogen metabolisms are key to ecosystem
410 survival in these dark, underground worlds. The research generated nearly comprehensive genomic
411 datasets for organisms detected in both URLs, including for sites located near and far from the access
412 tunnels, at multiple depths and over up to four years. The value of these genome collections for ecological,
413 evolutionary, biotechnological and repository engineering studies will extend far beyond the current study.
414 Our analyses provide answers to questions about microbiome stability and revealed surprising microbial
415 community compositional and genotypic overlap over sites separated by hundreds of meters of rock,
416 potentially explained by dispersal via slow groundwater flow or during a prior hydrological regime.
417 Subsurface repositories for storage of H₂ and radionuclides constructed in the future will benefit from
418 insights regarding potential microbial impacts arising from microbial metabolism and constraints on the
419 ways that the surrounding ecosystems will be impacted by repository engineering.

420

421 **Data availability**

422

423 Prior to accession via NCBI, the draft genomes can be accessed via
424 https://ggkbase.berkeley.edu/mizunami_genomes/organisms for the Mizunami URL and
425 https://ggkbase.berkeley.edu/horonobe_genomes/organisms for the Horonobe URL. Some download
426 functions may require you to sign up for a ggKbase account. Nine Horonobe datasets are publicly available

427 via BioProject ID: PRJNA321556 (e.g., V-250m-2014 = Hor_250_2014). Upon publication, the sequences
428 will be available via public data portals and the read datasets via the SRA.

429

430 **Acknowledgements**

431 We thank Chris Greening for making available up to date hydrogenase HMMs and helpful discussion, Brian
432 C. Thomas for bioinformatics support and helpful discussions, and Yusuke Watanabe, Mitsuru Kubota,
433 Toshihiro Kato for groundwater sampling at the Mizunami Underground Research Laboratory. Funding for
434 this project was provided by The Ministry of Economy, Trade and Industry of Japan, as “The project for
435 validating near-field system assessment methodology in geological disposal” and “Development of the
436 Technology for Integrating Radionuclide Migration Assessments” (2022 and 2023 FY, Grant Number:
437 JPJ007597), the Japan Atomic Energy Agency (JAEA) Fund for Exploratory Researches (Houga fund),
438 “Grant-in-Aid for Scientific Research (C) (Grant No. 19K05342) of the Japan Society for the Promotion of
439 Science, National Science Foundation grant number OCE2049478 (to KA), A University of California,
440 Berkeley Dissertation Year Fellowship (to LVA), The Berkeley Graduate Fellowship (to ALJ), Moore
441 Foundation support (to JFB), NSF “Four Networks for Geologic Hydrogen Storage” grant 2230766 (to JFB),
442 support from the Innovative Genomics Institute (to RS and JFB), the Watershed Function Scientific Focus
443 Area funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental
444 Research under Award Number DE-AC02-05CH11231 (to AL) and sabbatical support from the University
445 of California, Berkeley (to JFB).

446

447 **Author contributions**

448 This study was designed by YA and JFB, with contributions from TI and HB. YA, HB, and YS obtained the
449 metagenomic sequence datasets. RS and SL processed metagenomic data, RS, YA and JFB performed the
450 binning. Organism representation and distribution patterns were analyzed by YA, RS and JFB. YA, JFB, ALJ,

451 SD, CJC, and AL conducted the phylogenetic analyses. YA and KA took primary responsibility for metabolic
452 analyses, TI, HB, KM, AM, EI, HM analyzed geochemical and hydrologic data. Hydrogenase analyses were
453 performed by LVA, DG, YA and JFB. DG performed growth rate calculations. JFB performed genome
454 curation. JFB and YA wrote the manuscript, with input from DG. All authors reviewed and commented on
455 the manuscript prior to submission.

456

457 **Methods**

458

459 **URL description:** The Horonobe URL is located about 50 km south of Wakkanai in the northwestern
460 peninsula of Hokkaido, Japan (Figure 1). It is situated in a low-lying coastal plain where Quaternary
461 alluvium and terrace deposits overlie Tertiary and Cretaceous sediments that were deposited in the
462 Mesozoic Tempoku Basin (Waseda et al., 1996). The Tempoku Basin is an on-shore basin that is elongated
463 in the Horonobe area along a N-S axis. Seismic reflection surveys indicate that the current compressive E-
464 W neotectonic stress to the west of the Horonobe area was established at around 2-3 Ma (Ogura and
465 Kamon, 1992; Ito, 1999). Neogene strata in this area unconformably overlie Paleogene rocks. These strata
466 consist of the Miocene Onishibetsu, Masuporo and Wakkanai formations, the Miocene-Pliocene Koetoi
467 Formation and the Pliocene-Pleistocene Yuchi and Sarabetsu Formations (Iijima and Tada, 1981). All of
468 these formations were deposited in a marine environment. The Koetoi and Wakkanai Formations are the
469 main host rocks of the Horonobe URL. These formations are composed mainly of homogeneous siliceous
470 rocks. The burial and subsidence of these formations occurred throughout the Neogene and Quaternary.
471 Subsequent uplift and denudation started at about 1.3-1.0 Ma (Ishii et al., 2008). The Koetoi Formation is
472 Neogene to Quaternary diatomaceous mudstones containing opal-A, and the Wakkanai Formation is
473 Neogene siliceous mudstones containing opal-CT, with trace amount of quartz, feldspar, clay minerals,
474 pyrite, calcite and siderite (Hiraga and Ishii, 2007; Ishii et al., 2007; Tachi et al., 2011). The Horonobe URL

475 was constructed by JAEA to conduct basic geoscientific research and evaluate the feasibility and safety of
476 geological disposal in deep sedimentary environments.

477 The Mizunami URL was located in the Gifu prefecture in central Japan (Figure 1). The URL was
478 constructed by JAEA to conduct basic geoscientific research, but it was closed in 2019 as the research
479 project was completed. Around the URL site, sedimentary rock (the Mizunami Group; 20-15 Ma, consisting
480 of the Akeyo Formation) unconformably overlies Toki Granite (70 Ma). Toki granite is overlain by the
481 Tertiary sedimentary rocks at ~100 - 200 mbgl (meter below ground level) around the Mizunami URL. The
482 Toki granite has three rock facies grading from muscovite-biotite granite, hornblende-biotite granite, and
483 biotite granite. The constituent minerals are quartz, plagioclase, K-feldspar, biotite, hornblende,
484 muscovite, accessory minerals and secondary minerals such as chlorite, calcite, and pyrite (Yuguchi et al.,
485 2011, 2013). The Mizunami URL consisted of a main shaft, a ventilation shaft, sub-stages, and access
486 tunnels at 300 and 500 m below ground level.

487

488 **Sample collection:** Groundwater samples in the Mizunami URL were collected from six zones in four
489 different boreholes (07MI07, 09MI20, 09MI21, 10MI26) at three depths 200, 300, 400 mbgl during 2014
490 and 2015 (Figure 1b; Table S1). Samples in the Horonobe URL were collected from five zones in three
491 different boreholes (08E140C01, 07V140M03, 09V250M02) at five depths 140, 167, 194, 218, and 250
492 mbgl at the Horonobe URL during 2013 and 2016 (**Figure 1c**). All groundwater samples were obtained
493 using multipacker systems (Nanjo et al, 2012). To minimize the influence of drilling and the installation of
494 tools for hydrogeochemical monitoring, groundwaters were drained at least five times the section volume
495 before monitoring for geochemical and microbial studies. The groundwater chemistry has been
496 monitored since 2007, starting immediately after the drilling of these boreholes. The contamination of
497 each groundwater sample by the drilling fluids was checked by measurement of the concentration of
498 uranine and sodium naphthionate. The concentrations were below the detection limit. All but one sample

499 for microbiological analyses was collected onto 0.22 µm pore size filters (type GVWP; Merck Millipore,
500 Darmstadt, Germany) held in pressure-resistant stainless steel filter holders directly connected to the
501 tubing outlet under in-situ hydraulic pressure conditions. The other sample (Hor_218_2014_10k) was
502 collected using an ultrafiltration disc with 10,000 Da nominal molecular weight limit (type PLGC; Merck
503 Millipore) after filtering through a 0.22 µm membrane filter. The volume of groundwater samples used
504 for filtration was between 0.9 and 64 L at the Horonobe and 26 and 514 L at the Mizunami, depending on
505 the cell densities in each groundwater sample (**Table S1**).

506

507 **DNA extraction and sequencing:** DNA was extracted from the biomass collected on filters using the Extrapol
508 Soil DNA Kit Plus ver. 2 (Nippon Steel and Sumikin EcoTech Corporation, Tsukuba Japan). Genomic DNA
509 libraries were prepared using TruSeq Nano DNA Sample Prep Kit (Illumina, San Diego, CA, USA) according
510 to the manufacturer's instructions. Quality of the library was examined using an Agilent 2100 bioanalyzer
511 (Agilent Technologies) and paired-end 150-bp reads with a 550 bp insert size were sequenced by Hokkaido
512 System Science Co., Ltd (Hokkaido, Japan), using an Illumina HiSeq 2500 (San Diego, CA, USA). DNA
513 concentrations and sequencing information are presented in Table S1. Read datasets were assembled
514 using IDBA_UD with the following parameters: -mink 40, -maxk 100, -step 20, and -pre_correction (Peng
515 et al., 2012). Trimmed shotgun sequencing reads from each sample were mapped to all scaffolds >1000
516 bp, using Bowtie2 with default parameters (Langmead et al., 2009). For all scaffolds over 1000 bp, open
517 reading frames were predicted with Prodigal using the meta setting (Hyatt et al., 2010). Functional
518 annotations for all open reading frames were predicted using USEARCH (Edgar, 2010) searches against
519 the Uniref100 (Suzek et al., 2007), Uniprot (Magrane and UniProt Consortium, 2011) and KEGG (Graham
520 et al. 2018) to parse genes annotated with KEGG Orthology using Kofamscan (Aramaki et al. 2020), tRNA
521 sequences were predicted using tRNAscan-SE (Schattner et al., 2005).

522

523 **Genome reconstruction:** After assembly, scaffolds >1000 bp were binned by combination of phylogenetic
524 profiles, read coverage and nucleotide content (GC proportion and tetranucleotide signatures) using
525 ggKbase binning tools (<https://ggkbase.berkeley.edu>). Genomes that shared $\geq 95\%$ average nucleotide
526 identity were clustered to approximate species and the highest quality genome selected from each cluster.
527 This analysis was performed using dRep (Olm et al. 2017).

528

529 **Organism distribution patterns:** We conducted a census of organism types and abundances by identifying
530 a non-redundant set of scaffolds encoding the ribosomal protein S3 (rpS3). This phylogenetically
531 informative sequence, unlike 16S rRNA genes, tends to be well reconstructed from short read datasets
532 (Olm et al. 2020). Reads were mapped to each scaffold encoding an rpS3 gene so that organism relative
533 abundances in each sample could be determined. Species level groupings used rpS3 sequence
534 comparisons with a $\geq 98\%$ amino acid identity for classification of two organisms as approximately the
535 same species.

536

537 **Phylogenetic analysis:** The phylogeny of ribosomal protein S3 was calculated by aligning the sequences
538 with select reference sequences using MUSCLE (Edgar, 2004). The alignments were manually trimmed in
539 Geneious v.8 (BioMatters Ltd., San Francisco, CA, USA) to remove poorly aligned positions and columns
540 composed of over 90% gaps before concatenation of protein sequences. Trees were built using RAxML v.
541 8.2.10 (Stamatakis, 2014) (as implemented on the CIPRES web server) (Miller et al., 2010), under the LG
542 plus gamma model of evolution (PROTGAMMALG in the RAxML model section), and with the number of
543 bootstraps automatically determined. RpS3 sequences were taxonomically classified using phylogenetic
544 trees and NCBI protein BLAST searches. The in situ replication rates of microbial genomes were calculated
545 using iRep v.1.1.14 (Brown et al. 2016) with default parameters.

546

547 **Functional gene analysis:** Metabolic coding potential of all non-redundant sequences was explored by
548 HMM searches against protein families downloaded from FunGene (Fish et al., 2013), TIGRFAM (Haft et
549 al., 2003) and Pfam (Finn et al., 2014) as well as custom-built profile HMMs for several target genes (Eddy,
550 2011). Hydrogenases were annotated using profile hidden Markov model (HMM) searches (Eddy, 2011)
551 with a custom set of [NiFe]-, [FeFe]- and [Fe]-hydrogenase HMMs (Søndergaard et al. 2016). Rubisco
552 protein sequences were classified via phylogenetic analyses that used a custom dataset of reference
553 sequences, including a subset from biochemically characterized types (**Data S2**).

554

555 **Similarity of organisms in each site:** To evaluate the extent to which organisms present at each URL are
556 now represented by genomes we identified all rpS3 genes from the combined dataset, regardless of
557 whether or not they were in a draft genome bin and determined which samples contained sequences
558 belonging to the dereplicated sequence clusters. Both dereplication and comparison of the dereplicated
559 set to rpS3 sequences in each sample used an 99% nucleotide ANI threshold. This approach generated a
560 presence/absence matrix that was sorted by the number of times each genome or unbinned sequence
561 was detected across the sample series.

562 To evaluate how similar the collections of organisms represented by genomes from Mizunami and
563 Horonobe were to each other and to groups of genetically defined organisms from other well sampled
564 subsurface systems we first compared organisms from each site based on their rpS3 sequences encoded
565 in their genomes, then based on average nucleotide identity (ANI) across alignable portions of their
566 genomes. Different ANI cutoffs were used for different parts of the analysis. For the most closely related
567 organisms, genome sequences and synteny were evaluated using Mauve (Darling et al., 2004). For one
568 especially well assembled bacterial genome, manual curation was performed. This involved verification of
569 overall assembly accuracy using mapped paired reads, removal of local assembly errors by excision of the
570 portion not supported by reads and gap filling using reads and unplaced paired reads. Scaffold ends were

571 extended using unplaced read pairs until circularization of the fully verified sequence was accomplished.

572 The complete genome structure was evaluated using GC skew and cumulative GC skew using using

573 `gc_skew.py` (<https://github.com/christophertbrown/iRep>)

574 (Brown et al. 2016).

575 **References**

- 576 Teske A, Biddle JF, Edgcomb VP, Schippers A. Deep subsurface microbiology: a guide to the research topic
577 papers. *Front Microbiol.* 2013;4:1–3.
- 578 Hoshino T, Doi H, Uramoto GI, Wörmer L, Adhikari RR, Xiao N, Morono Y, D'Hondt S, Hinrichs KU, Inagaki
579 F. Global diversity of microbial communities in marine sediment. *Proc Natl Acad Sci USA.* 2020;44:27587–
580 97.
- 581 Pedersen K. Subterranean microorganisms and radioactive waste disposal in Sweden. *Eng Geol.*
582 1999;52:163–76.
- 583 Rajala P, Carpén L, Vepsäläinen M, Raulio M, Sohlberg E, Bomberg M. Microbially induced corrosion of
584 carbon steel in deep groundwater environment. *Front Microbiol.* 2015;6:647.
- 585 Stroes-Gascoyne, S. & Hamon, C. & Maak, P. & Russell, S. The effects of the physical properties of highly
586 compacted smectitic clay (bentonite) on the culturability of indigenous microorganisms. *Applied Clay
587 Science.* 2010;47:155-162.
- 588 Stroes-Gascoyne S, West JM. Microbial studies in the Canadian nuclear fuel waste management program.
589 *FEMS Microbiol Rev.* 1997;20:573–90.
- 590 Dopffel N, Mayers K, Kedir A, et al. Microbial hydrogen consumption leads to a significant pH increase
591 under high-saline-conditions: implications for hydrogen storage in salt caverns. *Sci Rep.* 2023;13:10564.
- 592 Liu N, Kovscek AR, Fernø MA, Dopffel N. Pore-scale study of microbial hydrogen consumption and
593 wettability alteration during underground hydrogen storage. *Front Energy Res.* 2023;11.
- 594 Zeng L, Sarmadivaleh M, Saeedi A, Chen Y, Zhong Z, Xie Q. Storage integrity during underground hydrogen
595 storage in depleted gas reservoirs. *Earth-Sci Rev.* 2023;247:104625.
- 596 Wrighton KC, Thomas BC, Sharon I, Miller CS, Castelle CJ, VerBerkmoes NC, Wilkins MJ, Hettich RL, Lipton
597 MS, Williams KH, Long PE, Banfield JF. Fermentation, hydrogen, and sulfur metabolism in multiple
598 uncultivated bacterial phyla. *Science.* 2012;337:1661–65.
- 599 Anantharaman K, Brown CT, Hug LA, Sharon I, Castelle CJ, Probst AJ, et al. Metabolic handoffs shape
600 biogeochemical cycles mediated by complex microbial communities. *Nat Commun.* 2016;7:13219.
- 601 He C, Keren R, Whittaker ML, et al. Genome-resolved metagenomics reveals site-specific diversity of
602 episymbiotic CPR bacteria and DPANN archaea in groundwater ecosystems. *Nat Microbiol.* 2021;6:354–
603 65.
- 604 Probst AJ, Ladd B, Jarett JK, Geller-McGrath DE, Sieber CMK, Emerson JB, Anantharaman K, Thomas BC,
605 Malmstrom RR, Stiegmeier M, Klingl A, Woyke T, Ryan MC, Banfield JF. Differential depth distribution of

- 606 microbial function and putative symbionts through sediment-hosted aquifers in the deep terrestrial
607 subsurface. *Nat Microbiol.* 2018;3:328–36.
- 608 Hallbeck L, Pedersen K. Characterization of microbial processes in deep aquifers of the Fennoscandian
609 Shield. *Appl Geochem.* 2008;23:1796–819.
- 610 Mehrshad M, Lopez-Fernandez M, Sundh J, Bell E, Simone D, Buck M, et al. Energy efficiency and biological
611 interactions define the core microbiome of deep oligotrophic groundwater. *Nat Commun.* 2021;12:4253.
612
- 613 Konno U, Kouduka M, Komatsu DD, Ishii K, Fukuda A, Tsunogai U, et al. Novel microbial populations in
614 deep granitic groundwater from Grimsel Test Site, Switzerland. *Microbial Ecol.* 2013;65:1–12.
615
- 616 Hernsdorf AW., Amano Y, Miyakawa K, Ise K, Suzuki Y, Anantharaman, K, Probst A., Burstein D., Thomas
617 B.C., Banfield J.F. Potential for microbial H₂ and metal transformations associated with novel bacteria and
618 archaea in deep terrestrial subsurface sediments. *The ISME Journal.* 2017;11:1915–1929.
619
- 620 Ino K, Hernsdorf A, Konno U, Kouduka M, Yanagawa K, Kato S, et al. Ecological and genomic profiling of
621 anaerobic methane-oxidizing archaea in a deep granitic environment. *ISME J.* 2017;12:31–47.
- 622 Iwatsuki T, Furue R, Mie H, Ioka S, Mizuno T. Hydrochemical baseline condition of groundwater at the
623 Mizunami underground research laboratory (MIU). *Applied Geochemistry.* 2005;20:2283–2302.
- 624 Hayashida K, Munemoto T, Aosai D, Inui M, Iwatsuki T. Hydrochemical Investigation at the Mizunami
625 Underground Research Laboratory -Compilation of Groundwater Chemistry Data in the Mizunami Group
626 and the Toki Granite- (Fiscal Year 2014). JAEA-Data/Code 2016-001 (Technical Report by Japan Atomic
627 Energy Agency). 2016. <https://doi.org/10.11484/jaea-data-code-2016-001> (in Japanese with English
628 abstract).
629
- 630 Amano Y, Yamamoto Y, Nanjo I, Murakami H, Yokota H, Yamazaki M, et. Data of Groundwater from
631 Boreholes, River Water and Precipitation for the Horonobe Underground Research Laboratory Project
632 (2001–2010). JAEA-Data/Code 2011-023 (Technical Report by Japan Atomic Energy Agency). 2012.
633 <https://doi.org/10.11484/jaea-data-code-2011-023> (in Japanese with English abstract).
634
- 635 Sasamoto H, Yamamoto N, Miyakawa K, Mizuno. Data of Groundwater Chemistry Obtained in the
636 Horonobe Underground Research Laboratory Project (2011–2013). JAEA-Data/Code 2014-033 (Technical
637 Report by Japan Atomic Energy Agency). 2015. <https://doi.org/10.11484/jaea-data-code-2014-033> (in
638 Japanese with English abstract).
- 639 Miyakawa K, Ishii E, Hirota A, Komatsu D.D, Ikeya K, Tsunogai U. The role of low-temperature organic
640 matter diagenesis in carbonate precipitation within a marine deposit. *Appl. Geochem.*, 2017;76: 218–31.
- 641 Tamamura S, Miyakawa K, Aramaki N, Igarashi T, Kaneko K. A proposed method to estimate in situ
642 dissolved gas concentrations in gas-saturated groundwater. *Groundwater.* 2018;56(1):118–30.

- 643 Adam PS, Panagiotis S, Borrel G, Brochier-Armanet C, Gribaldo S. The growing tree of Archaea: New
644 perspectives on their diversity, evolution, and ecology. ISME J. 2017;11.
- 645 Brown CT, Hug LA, Thomas BC, Sharon I, Castelle CJ, Singh A, Wilkins MJ, Wrighton KC, Williams KH,
646 Banfield JF. Unusual biology across a group comprising more than 15% of domain Bacteria. Nature.
647 2015;523:208–11.
- 648 Greening C, Biswas A, Carere CR, et al. Genomic and metagenomic surveys of hydrogenase distribution
649 indicate H₂ is a widely utilised energy source for microbial growth and survival. ISME J. 2016;10:761–77.
- 650 Ma K, Schicho R, Kelly R, Adams MW. Hydrogenase of the hyperthermophile *Pyrococcus furiosus* is an
651 elemental sulfur reductase or sulfhydrogenase: evidence for a sulfur-reducing hydrogenase ancestor. Proc
652 Natl Acad Sci USA. 1993;90:5341–5344.
- 653 Sasamoto H, Sato H, Arthur RC. Characterization of mineralogical controls on ammonium concentrations
654 in deep groundwaters of the Horonobe area, Hokkaido. J Geochemical Exploration. 2018;188:318–325.
- 655 Brown CT, Olm MR, Thomas BC, Banfield JF. Measurement of bacterial replication rates in microbial
656 communities. Nat Biotechnol. 2016;34(12):1256–63.
- 657
- 658 Nishimura H, Kouduka M, Fukuda A, Ishimura T, Amano Y, Beppu H, et al. Anaerobic methane-oxidizing
659 activity in a deep underground borehole dominantly colonized by Ca. Methanoperedenaceae. Environ
660 Microbiol Rep. 2023;15(3):197–205.
- 661
- 662 Yuguchi T, Tsuruta T, Nishiyama T. Zoning of rock facies and chemical composition in the Toki granitic body,
663 Central Japan. GKK. 2010;39:50-70 (in Japanese with English abstract).
- 664
- 665 Yuguchi T, Sueoka S, Iwano H, Izumino Y, Ishibashi M, Danhara T, et al. Position-by-position cooling paths
666 within the Toki granite, central Japan: Constraints and the relation with fracture population in a pluton.
667 Journal of Asian Earth Sciences. 2018;169:47-66.
- 668
- 669 Iwatsuki T, Yoshida H. Groundwater chemistry and fracture mineralogy in the basement granitic rock in
670 the Tono area, Gifu Prefecture, Japan -Groundwater composition, Eh evolution analysis by fracture filling
671 minerals-. Geochem Jour. 1999;33:19-32.
- 672 Ishibashi M, Yuguchi T. Development of New Method for Evaluating the Mineral Distribution and Mode:
673 Quantitative Image Analysis Using the Elemental Maps Obtained by the Scanning X-ray Analytical
674 Microscope. Journal of the Japan Society of Engineering Geology. 2017;58: 80-93.
- 675 Iwatsuki T, Xu S, Itoh S, Abe M, Watanabe M. Estimation of relative groundwater age in the granite at the
676 Tono research site, central Japan. Nuclear Instruments and Methods in Physics Research Section B: Beam
677 Interactions with Materials and Atoms. 2000;172:524-529.
- 678

- 679 Iwatsuki T, Satake H, Metcalfe R, Yoshida H, Hama K. Isotopic and morphological features of fracture
680 calcite from granitic rocks of the Tono area, Japan: a promising palaeohydrogeological tool. *Appl Geochem.*
681 2002;17(9):1241-57.
- 682 Stevens TO, McKinley JP. Lithoautotrophic microbial ecosystems in deep basalt aquifers. *Science.*
683 1995;270(5235):450-5.
- 684 Ino K, Hernsdorf A, Konno U, Kouduka M, Yanagawa K, Kato S, et al. Ecological and genomic profiling of
685 anaerobic methane-oxidizing archaea in a deep granitic environment. *ISME J.* 2017;12:31–47.
686
- 687 Mills CT, Amano Y, Slater GF, Dias RF, Iwatsuki T, Mandernack KW. Microbial carbon cycling in oligotrophic
688 regional aquifers near the Tono Uranium Mine, Japan as inferred from $\delta^{13}\text{C}$ and $\Delta^{14}\text{C}$ values of in situ
689 phospholipid fatty acids and carbon sources. *Geochimica et Cosmochimica Acta.* 2010;74(13):3785-805.
- 690 Miyakawa K, Ishii E, Hirota A, Komatsu DD, Ikeya K, Tsunogai U. The role of low-temperature organic
691 matter diagenesis in carbonate precipitation within a marine deposit. *Appl. Geochem.* 2017;76:218-231.
- 692 Purkamo L, Kietäväinen R, Miettinen H, Sohlberg E, Kukkonen I, Itävaara M, Bomberg M. Diversity and
693 functionality of archaeal, bacterial and fungal communities in deep Archaean bedrock groundwater. *FEMS*
694 *microbiology ecology,* 2018;94(8):fiy116.
695
- 696 Kits K, Klotz M, Stein L. Methane oxidation coupled to nitrate reduction under hypoxia by the
697 Gammaproteobacterium *Methylomonas denitrificans*, sp. nov. Type Strain FJG1. *Environmental*
698 *Microbiology.* 2015;17(9):3219-32.
- 699 Wu ML, Wessels JC, Pol A, Op den Camp HJ, Jetten MS, van Niftrik L. XoxF-type methanol dehydrogenase
700 from the anaerobic methanotroph “*Candidatus Methylomirabilis oxyfera*”. *Appl Environ Microbiol.*
701 2015;81(4):1442-51.
- 702 Soares A, Edwards A, An D, Bagnoud A, Bradley J, Barnhart E, et al. A global perspective on bacterial
703 diversity in the terrestrial deep subsurface. *Microbiology.* 2023;169.
704
- 705 Pedersen K. Analysis of copper corrosion in compacted bentonite clay as a function of clay density and
706 growth conditions for sulfate-reducing bacteria. *J Appl Microbiol.* 2010;108:1094-1104.
707
- 708 Enning D, Garrels J. Corrosion of iron by sulfate-reducing bacteria: new views of an old problem. *Appl*
709 *Environ Microbiol.* 2014;80(4):1226–36.
- 710 Bell E, Lamminmäki T, Alneberg J, Andersson AF, Qian C, Xiong W, Hettich RL, Frutschi M, Bernier-Latmani
711 R. Active sulfur cycling in the terrestrial deep subsurface. *ISME J.* 2020 May;14(5):1260–72.
- 712 Dinh HT, Kuever J, Mußmann M, Hassel AW, Stratmann M, Widdel F. Iron corrosion by novel anaerobic
713 microorganisms. *Nature.* 2004;427:829–32.

- 714 Mori K, Tsurumaru H, Harayama S. Iron corrosion activity of anaerobic hydrogen-consuming
715 microorganisms isolated from oil facilities. *J Biosci Bioeng.* 2010;110:426–30.
- 716 Hirano, S.-i.; Ihara, S.; Wakai, S.; Dotsuta, Y.; Otani, K.; Kitagaki, T.; Ueno, F.; Okamoto, A. Novel
717 *Methanobacterium* Strain Induces Severe Corrosion by Retrieving Electrons from Fe⁰ under a Freshwater
718 Environment. *Microorganisms* 2022;10:270.
- 719 Mochizuki, Akihito & Ishii, Eiichi. Assessment of the level of activity of advective transport through
720 fractures and faults in marine deposits by comparison between stable isotope compositions of fracture
721 and pore waters. *Hydrogeology Journal.* 2022;30:813-827.
- 722 Teramoto M, Shimada J, Kunimaru T. Understanding groundwater flow regimes in low permeability rocks
723 using stable isotope paleo records in porewaters. In: Taniguchi M, Holman IP, editors. *Groundwater*
724 *Response to Changing Climate.* Florida: CRC Press; 2010. p. 79-86.
- 725
- 726 Ishii E. Assessment of Hydraulic Connectivity of Fractures in Mudstones by Single-Borehole Investigations.
727 *Water Resources Research.* 2018;54.
- 728
- 729 Nakata K, Hasegawa T, Oyama T, Ishii E, Miyakawa K, Sasamoto H. An Evaluation of the Long-Term
730 Stagnancy of Porewater in the Neogene Sedimentary Rocks in Northern Japan. *Geofluids.* 2018;1-21.
- 731 Mochizuki A, Ishii E. Assessment of the level of activity of advective transport through fractures and faults
732 in marine deposits by comparison between stable isotope compositions of fracture and pore waters.
733 *Hydrogeology Journal.* 2022;30:813-827.
- 734 Hanatani I, Munakata M, Kimura H, Sanga T. An analytic investigation of periglacial topography in
735 Horonobe region Hokkaido. *J. Nucl. Fuel Cycle Environ.* 2010;17:55-70.
736 <https://doi.org/10.3327/jnuce.17.55> (in Japanese with English abstract).
- 737 Lopez-Fernandez M, Broman E, Turner S, Wu X, Bertilsson S, Dopson M. Investigation of viable taxa in the
738 deep terrestrial biosphere suggests high rates of nutrient recycling. *FEMS Microbiol Ecol.*
739 2018;94(8):fiy121.
- 740 Miyakawa K, Ishii E, Hirota A, Komatsu DD, Ikeya K, Tsunogai U. The role of low-temperature organic
741 matter diagenesis in carbonate precipitation within a marine deposit. *Appl Geochem.* 2017;76:218–31.
- 742 Sasamoto H, Yamamoto N, Miyakawa K, Mizuno. Data of Groundwater Chemistry Obtained in the
743 Horonobe Underground Research Laboratory Project (2011–2013). JAEA-Data/Code 2014-033 (Technical
744 Report by Japan Atomic Energy Agency). 2015. <https://doi.org/10.11484/jaea-data-code-2014-033> (in
745 Japanese with English abstract).
- 746 Mochizuki A, Ishii E. Paleohydrogeology of the Horonobe area, Northern Hokkaido, Japan: Groundwater
747 flow conditions during glacial and postglacial periods estimated from chemical and isotopic data for
748 fracture and pore water. *Appl. Geochem.* 2023;155:105737.

- 749 International Atomic Energy Agency (IAEA) Scientific and technical basis for geological disposal of
750 radioactive wastes. Technical reports series, 2003;TRS no. 413.
- 751 Waseda A, Kajiwara Y, Nishita H, Iwano H. Oil-source rock correlation in the Tempoku basin of northern
752 Hokkaido, Japan. *Organic Geochemistry*. 1996;24:351-362.
- 753 Ogura N, Kamon M. The subsurface structures and hydrocarbon potentials in the Tenpoku and haboro
754 area, the northern Hokkaido, Japan. *J. Jpn. Assoc. Petroleum Technol.* 1992;57:32-44 (in Japanese with
755 English abstract).
- 756 Iijima A, Tada R. Silica diagenesis of neogene diatomaceous and volcaniclastic sediments in northern Japan.
757 *Sedimentology*. 1981;28:185e200.
- 758 Ishii E, Yasue K, Ohira H, Furusawa A, Hasegawa T, Nakagawa M. Inception of anticline growth near the
759 Omagari Fault, northern Hokkaido, Japan. *J Geol Soc Japan*. 2008;114(6):286–99.
- 760 Hiraga N, Ishii E. Mineral and chemical composition of rock core and surface gas composition in Horonobe
761 Underground Research Laboratory Project (Phase 1). JAEA-Data/Code 2007-022 (Technical Report by
762 Japan Atomic Energy Agency). 2007.
- 763 Ishii E, Hama K, Kunimaru T, Sato H. Change in groundwater pH by infiltration of meteoric water into
764 shallow part of marine deposits. *J Geol Soc Japan*. 2007;113:41–50.
- 765 Tachi Y, Yotsuji K, Seida Y, Yui M. Diffusion and sorption of Cs+, I- and HTO in samples of the argillaceous
766 Wakkanai Formation from the Horonobe URL, Japan: A clay-based modeling approach. *Geochim
767 Cosmochim Acta*. 2011;75:6742–59.
- 768 Yuguchi T, Amano K, Tsuruta T, Danhara T, Nishiyama T. Thermochronology and the three-dimensional
769 cooling pattern of a granitic pluton: An example of the Toki granite, Central Japan. *Contrib Miner Petrol*.
770 2011;162:1063–77.
- 771 Yuguchi T, Tsuruta T, Hama K, Nishiyama T. The spatial variation of initial $^{87}\text{Sr}/^{86}\text{Sr}$ ratios in the Toki granite,
772 Central Japan: Implications for the intrusion and cooling processes of a granitic pluton. *J Miner Petrol Sci*.
773 2013;108:1–12.
- 774 Nanjo I, Amano Y, Iwatsuki T, Kunimaru T, Murakami H, Hosoya S, Morikawa K. Development of a
775 groundwater monitoring system at Horonobe Underground Research Center. JAEA Research 2011-048
776 (Technical Report by Japan Atomic Energy Agency). 2012. <https://doi.org/10.11484/jaea-research-2011-048> (in Japanese with English abstract).
- 778 Peng Y, Leung HCM, Yiu SM, Chin FYL. IDBA-UD: a de novo assembler for single-cell and metagenomic
779 sequencing data with highly uneven depth. *Bioinformatics*. 2012;28:1420–28.

- 780 Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA
781 sequences to the human genome. *Genome Biol* 2009;10:R25.
- 782 Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition
783 and translation initiation site identification. *BMC Bioinformatics*. 2010;11:1–11.
- 784 Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. 2010;26(19):2460–
785 61.
- 786 Suzek BE, Huang H, McGarvey P, Mazumder R, Wu CH. UniRef: comprehensive and non-redundant Uni-
787 Prot reference clusters. *Bioinformatics*. 2007;23:1282–1288.
- 788 Magrane M, UniProt Consortium. UniProt Knowl- edgebase: a hub of integrated protein data. *Database*.
789 2011:bar009.
- 790 Graham ED, Heidelberg JF, Tully BJ. Potential for primary productivity in a globally-distributed bacterial
791 phototroph. *ISME J*. 2018;12(7):1861–66.
- 792 Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, Ogata H. KofamKOALA: KEGG
793 Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics*.
794 2020;36(7):2251–52.
- 795 Schattner P, Brooks AN, Lowe TM. The tRNAscan- SE, snoScan and snoGPS web servers for the detection
796 of tRNAs and snoRNAs. *Nucleic Acids Res*. 2005;33:W686–W689.
- 797 Olm, M., Brown, C., Brooks, B. et al. dRep: a tool for fast and accurate genomic comparisons that enables
798 improved genome recovery from metagenomes through de-replication. *ISME J* 11, 2864–2868 (2017).
799 <https://doi.org/10.1038/ismej.2017.126>
- 800 Olm MR, Crits-Christoph A, Diamond S, Lavy A, Carnevali PBM, Banfield JF. Consistent metagenome-
801 derived metrics verify and delineate bacterial species boundaries. *mSystems*. 2020;14(5):e00731-19. DOI:
802 10.1128/mSystems.00731-19.
- 803 Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids
804 Res*. 2004;32:1792–97.
- 805 Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies.
806 *Bioinformatics*. 2014;30:1312–13.
- 807 Miller MA, Pfeiffer W, Schwartz T. Creating the CIPRES Science Gateway for inference of large phylogenetic
808 trees. In 2010 Gateway Computing Environments Workshop (GCE). 2010;1–8.
- 809 Brown CT, Olm MR, Thomas BC, Banfield JF. Measurement of bacterial replication rates in microbial
810 communities. *Nat Biotechnol*. 2016;34(12):1256–63.

- 811 Fish JA, Chai B, Wang Q, Sun Y, Brown CT, Tiedje JM et al. FunGene: the functional gene pipeline and
812 repository. *Front Microbiol.* 2013;4:00291.
- 813 Haft DH, Selengut JD, White O. The TIGRFAMs data- base of protein families. *Nucleic Acids Res.*
814 2003;31:371–373.
- 815 Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR et al. Pfam: the protein families database.
816 *Nucleic Acids Res.* 2014;42:D222–D230.
- 817 Eddy SR. Accelerated profile HMM searches. *PLoS Comput Biol.* 2011;7(10):e1002195.
- 818 Søndergaard D, Pedersen CN, Greening C. HydDB: A web tool for hydrogenase classification and analysis.
819 *Sci Rep.* 2016;6:34212.
- 820 Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence
821 with rearrangements. *Genome research.* 2004 Jul 1;14(7):1394-403.
- 822 Olm, M.R., Brown, C.T., Brooks, B., and Banfield, J.F. dRep: A tool for fast and accurate genome de-
823 replication that enables tracking of microbial genotypes and improved genome recovery from
824 metagenomes. *The ISME Journal.* 2017;11:2864-68. doi: 10.1038/ismej.2017.126.
- 825 Wilkinson M, Gilfillan MV, Haszeldine RS, Ballentine CJ. Plumbing the depths: Testing natural tracers of
826 subsurface CO₂ origin and migration, Utah, USA. In: Grobe M, Pashin JC, Dodge RL. (Eds.), Carbon dioxide
827 sequestration in geologic media -State of the science. American Association of Petroleum Geologists,
828 AAPG Studies in Geology. 2009;59:619-634.
- 829 Hagiwara H, Iwatsuki T, Hasegawa T, Nakaga K, Tomioka Y. Shallow groundwater intrusion to deeper
830 depths caused by construction and drainage of a large underground facility. *Journal of Japanese
831 Association of Hydrological Sciences.* 2015;45:21-38
- 832 Matsuoka T, Hama K. Mizunami Underground Research Laboratory Project -Synthesis report on the R & D
833 concerning important issues-. JAEA-Research 2019-2019-012. (Technical Report by Japan Atomic Energy
834 Agency). 2019. <https://doi.org/10.11484/jaea-research-2019-012> (in Japanese with English abstract).
- 835

836

Supplementary Information

837

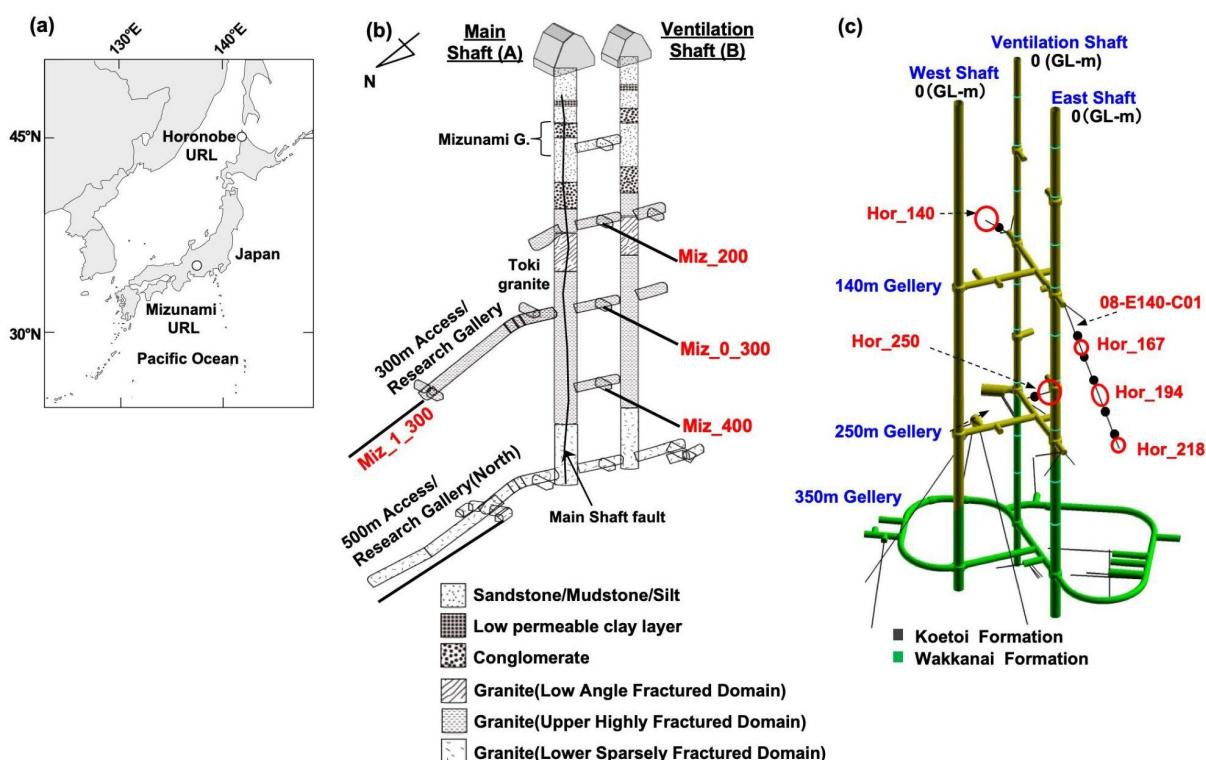
838 **Mizunami and USA site hydrology constraints on cross-continent comparisons**

839

840 The Rifle aquifer is shallow and impacted by inflow from the surface over weeks to months. Northern
841 California groundwater wells respond to rainfall inputs over months. Although 10-20% of the Crystal
842 Geyser water is derived from a Paleozoic aquifer, most is supplied by meteoric recharge (from a region
843 40-50 km away; Wilkinson et al., 2009). Residence time of approximately 9300 years for the groundwater
844 was determined at the deeper part of sedimentary rocks overlying the Toki granite at the Mizunami URL
845 (Iwatsuki et al., 2005), and the recent decrease in the water pressure following construction of the URL
846 likely moved this water through the granitic URL (Hagiwara et al., 2015; Matsuoka & Hama, 2019). This
847 groundwater intrusion provides a potential mechanism for microbial dispersal throughout the Mizunami
848 URL. The case of Horonobe, results suggest that movement of surface water into the subsurface was
849 possible ~12,000 years ago. Thus, organisms could potentially have been distributed across all five
850 locations ~12,000 - 20,000 years ago.

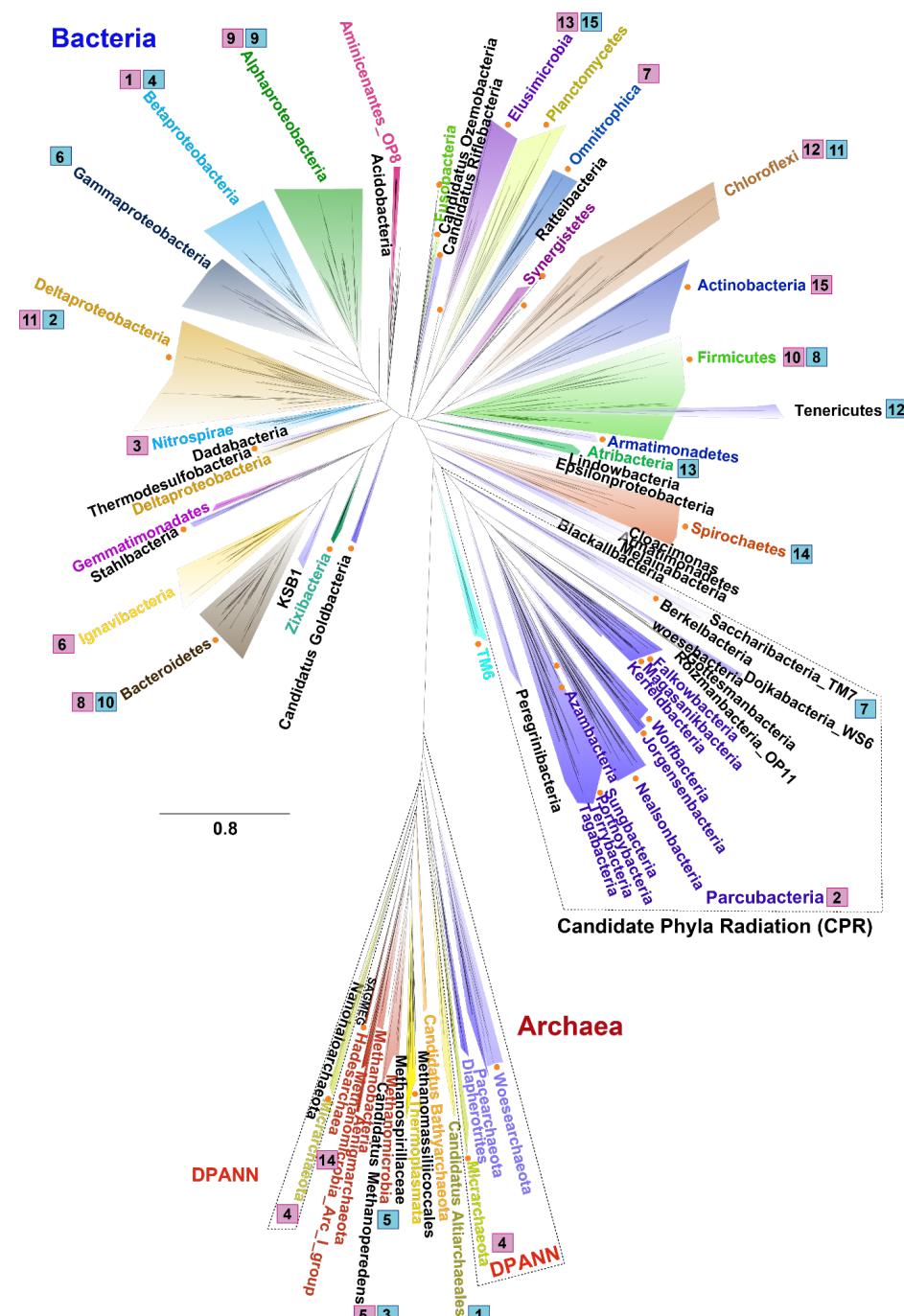
851

852



853

854 **Figure 1.** (a) Map of the Mizunami and Horonobe URL locations in Japan. (b) Layout of boreholes in
855 shafts and galleries in the Mizunami URL and (c) the Horonobe URL.

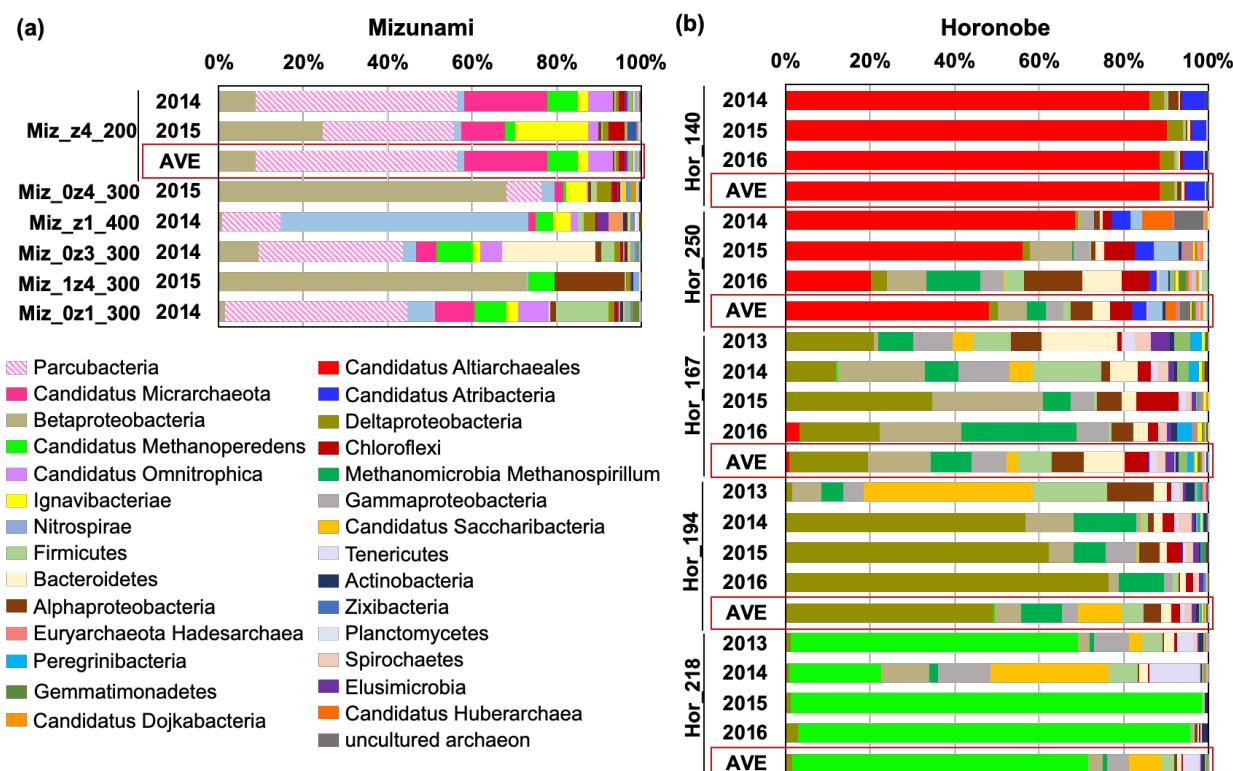


856

857

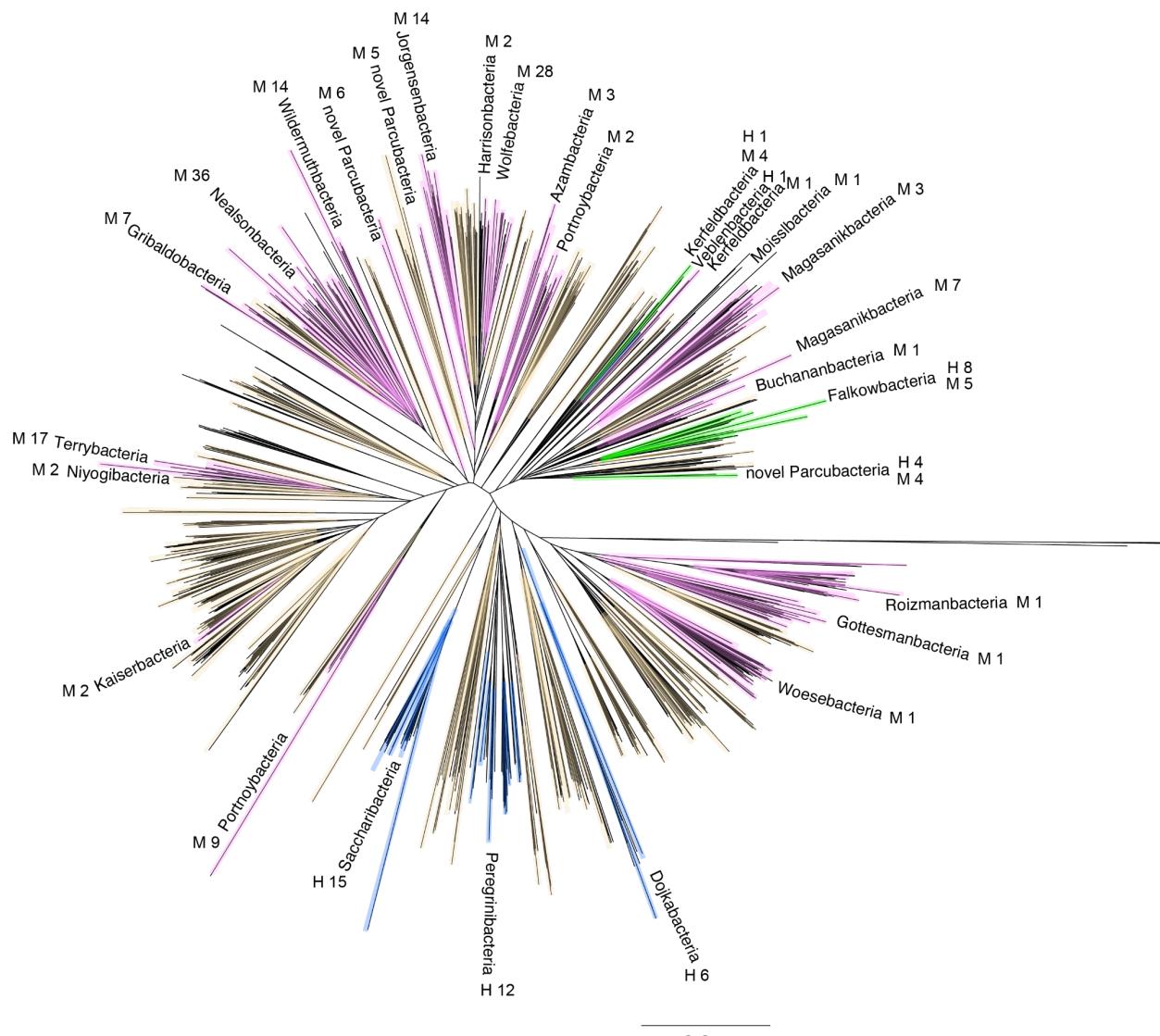
858 **Figure 2.** Phylogenetic tree of 956 representative sequences of ribosomal protein S3 from the Horonobe
859 and Mizunami samples, along with reference sequences. Orange dots indicate individual sequences with
860 <70% rpS3 amino acid identity to sequences in NCBI. Numbers in squares show the average rank

861 abundance of each phylum (or class in the case of Proteobacteria). Blue squares indicate organisms from
862 the Horonobe URL and pink squares indicate samples from the Mizunami URL.



863

864 **Figure 3.** Overview of URL microbial diversity based on the 15 most abundant organisms in each sample,
 865 classified mostly at the phylum/class level. Samples are listed in order of increasing the distance between
 866 the sampling site and the closest access tunnel to seek evidence of perturbation due to the presence of
 867 the tunnel. **(a)** 7 Mizunami 0.2 µm-filter URL samples collected between 2014 and 2015. **(b)** 18 Horonobe
 868 URL 0.2 µm-filter samples collected between 2013 and 2016. Given the general consistency in community
 869 composition over time, we also averaged the compositional data from the same site for subsequent
 870 analyses (red boxes).

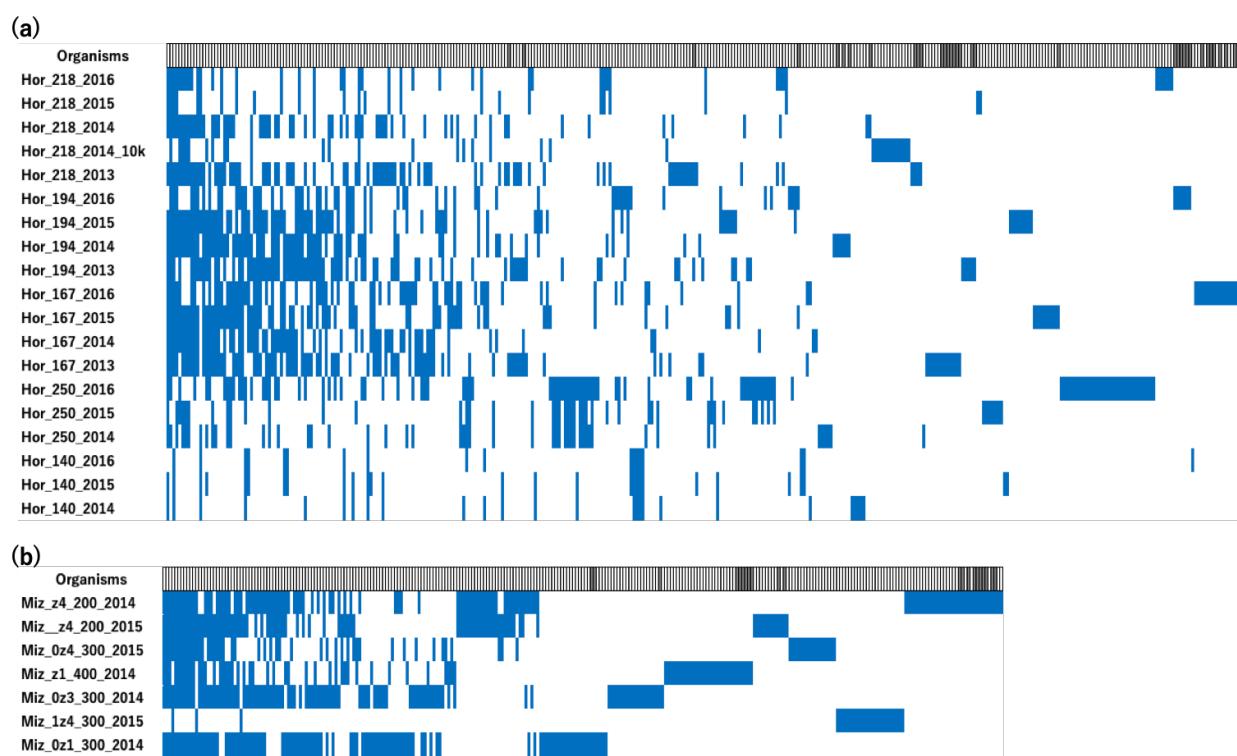


871

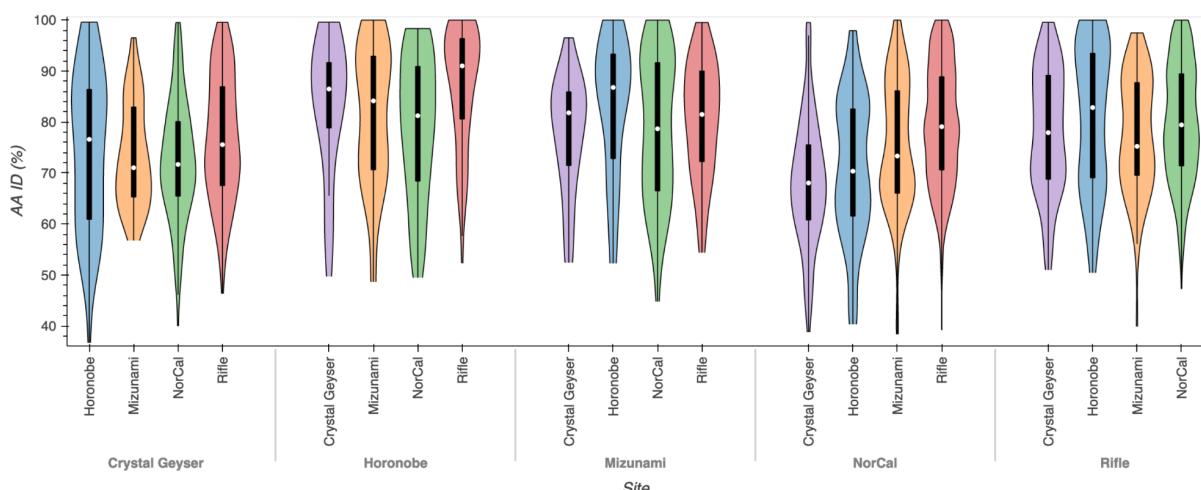
0.2

872

873 **Figure 4.** Phylogenetic tree for CPR bacteria constructed using ribosomal protein S3 sequences. Pink and
874 blue shadings indicate sequences from the Mizunami and Horonobe URLs, respectively. Green branches
875 indicate lineages with sequences from both the Mizunami and Horonobe URLs. The numbers after M
876 (Mizunami) and H (Horonobe) indicate the number of sequences in each named lineage. Brown
877 highlighted sequences are reference sequences. The long branches without color indicate Archaea,
878 which were used as the outgroup.

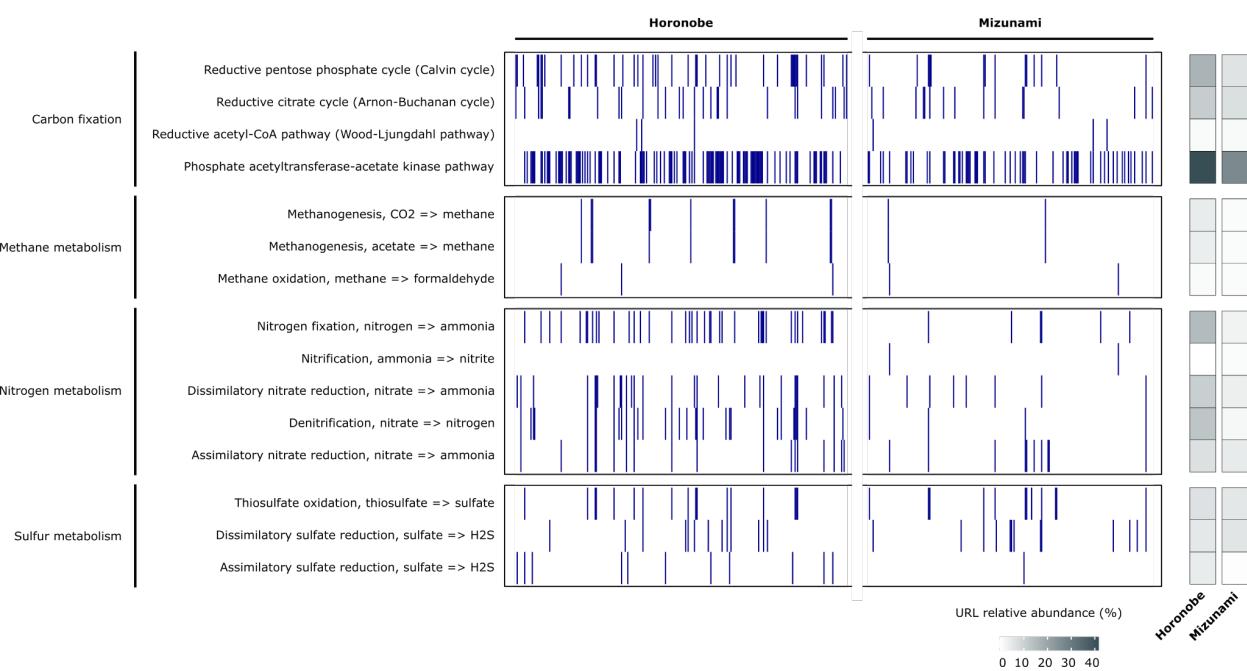


879 **Figure 5.** Detection (blue bars) of organisms (columns) in samples (rows) listed in approximate order of
880 decreasing distance from the access tunnels. Organisms lacking genomes are indicated by a dark gray box
881 in the Organisms bar. (a) Almost one quarter of all organisms detected within the Horonobe URL were
882 present in >25% of the samples. 90.389.2% of all organisms are represented by draft genomes. Organisms
883 lacking genomes were all detected in ≤ 3 samples. (b) Within the Mizunami URL, 45% of organisms were
884 detected in at least 25% of the samples and 10% of all organisms detected were present in >70% of the
885 samples. 92.0% of organisms are represented by draft genomes. All organisms lacking genomes were
886 detected in just one sample. For details see **Table S3**.



888

889 **Figure 6:** Comparison using the dereplicated genome sets from five genetically well sampled terrestrial
890 subsurface ecosystems to seek instances where two ecosystems are similar or significantly different from
891 each other. The highest scoring pairwise hit for each of 1472 sequences from genomes was assigned to
892 an ecosystem comparison category and the aa ID (%) inventoried. For example, when all 503 sequences
893 from Crystal Geyser were compared to all sequences from the four comparison datasets, there were 59
894 instances where the closest sequence was found in the Horonobe dataset. The 59 aa ID % values are
895 represented by column 1. Of the comparisons, those that were significantly different are Horonobe
896 compared to northern California (NorCal) vs. Rifle (p value of 1×10^{-3}), Crystal Geyser compared to Rifle
897 vs. NorCal (p value of 5×10^{-3}), and NorCal compared to all other ecosystems ($p < 1 \times 10^{-5}$. For details see
898 **Table S5.**



899

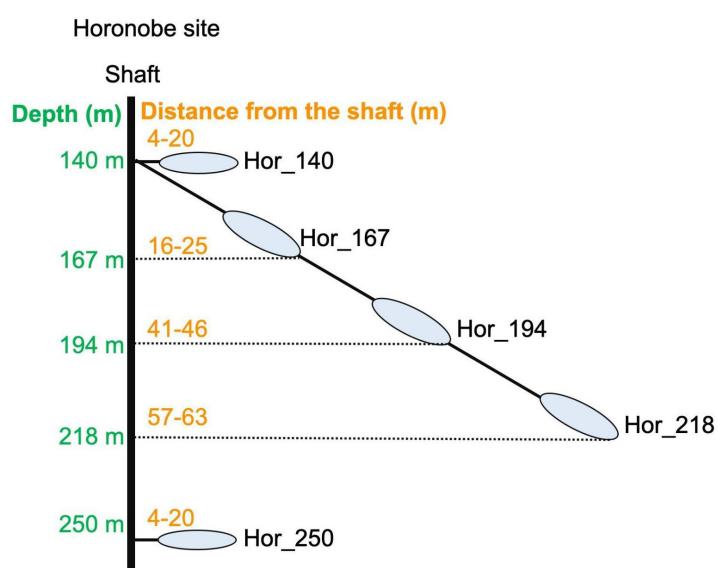
900

901 **Figure 7.** Key metabolisms across the Mizunami and Horonobe URLs. Presence/absence of each metabolic
902 pathway based on the occurrence of indicative marker genes annotated with KEGG Orthology using
903 Kofamscan (y-axis) in each of the recovered genomes (x-axis). The URL relative abundance (%) shows the
904 proportion of the 265 Horonobe and 225 Mizunami genomes with that metabolism.

905

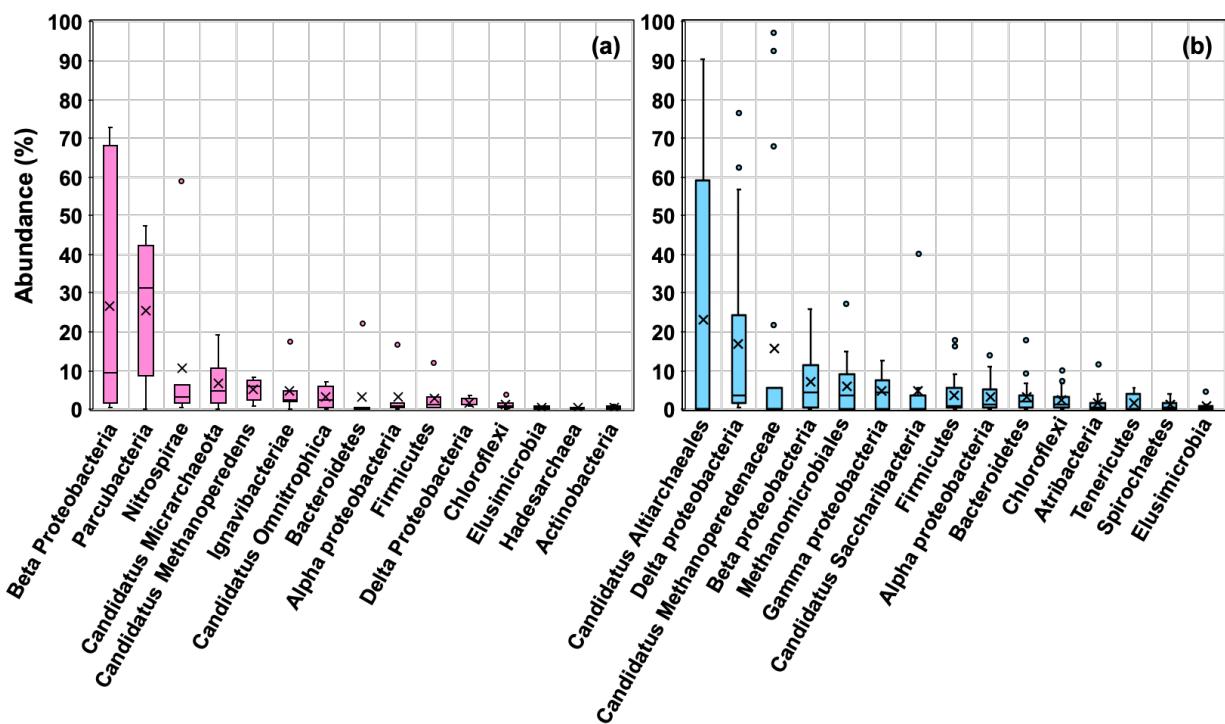
Supplementary Information

906

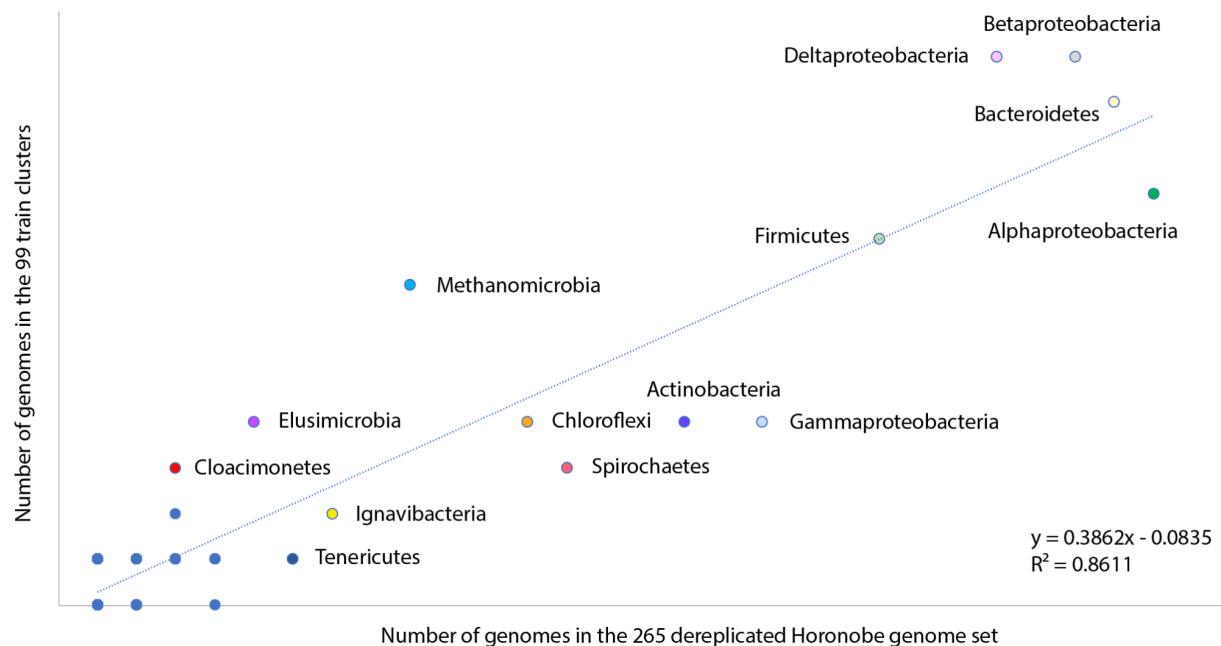


907

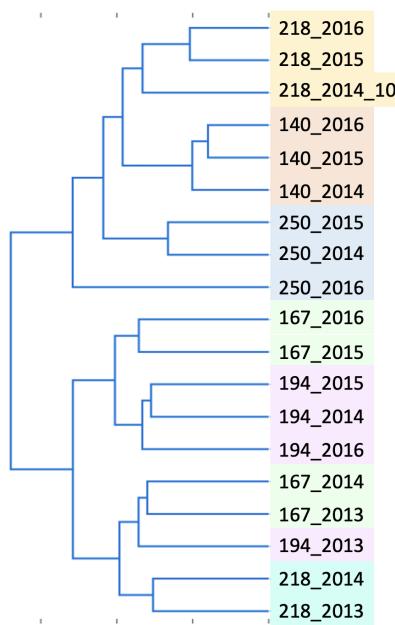
908 **Figure S1.** Diagram showing the layout of the Horonobe sampling sites. The y-axis indicates depth below
909 the surface and the ovals indicate the distances that the groundwater sampling volumes are from the
910 access tunnels.



911
912 **Figure S2.** Relative abundances of the 15 most abundant organisms at the assigned rank based on
913 normalized coverage values averaged over all samples. **(a)** Mizunami URL and **(b)** Horonobe URL. The
914 edges of the box are the first and third quartile, crosses in the boxes are average proportions, circles are
915 outliers.



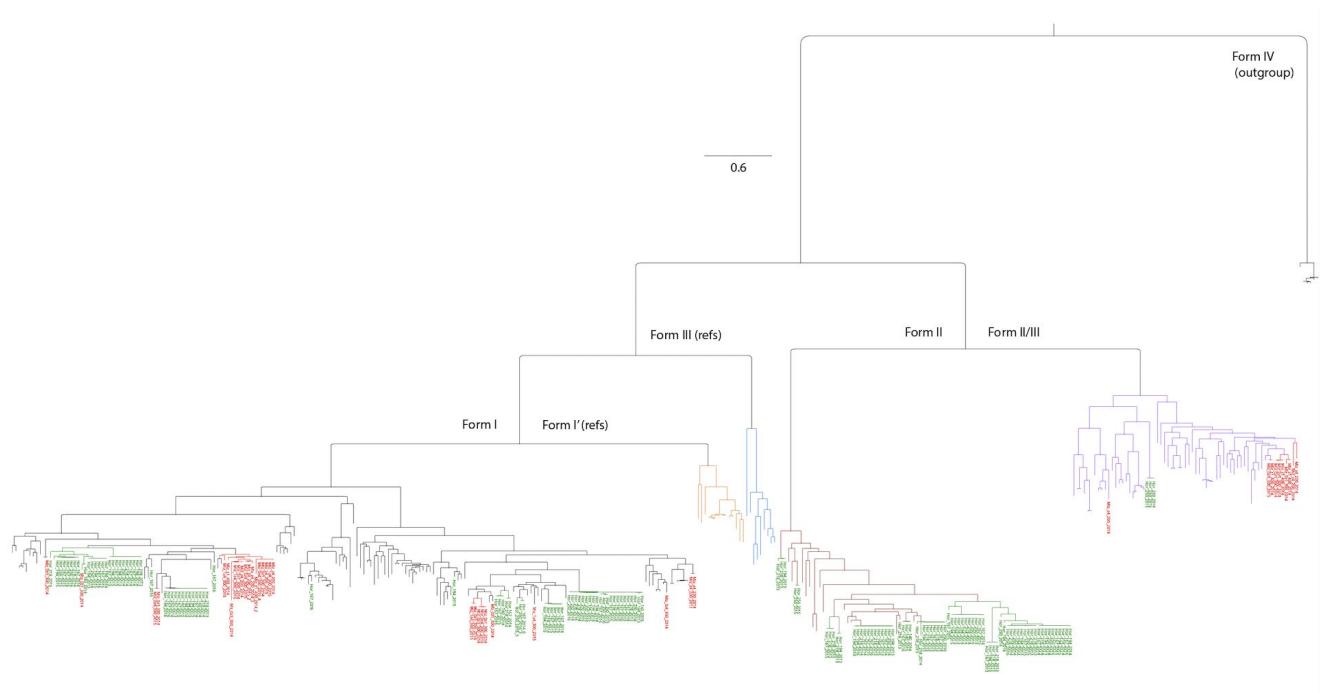
916



917

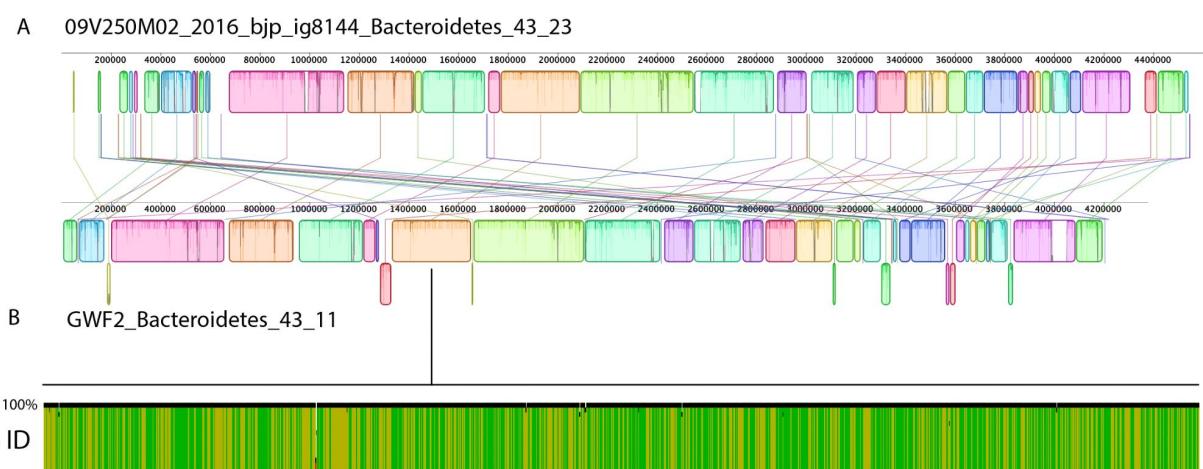
918 **Figure S3.** Analysis of the diversity structure and compositional similarity across sites and years for the
919 Horonobe URL (a) The extent of within class-level diversity for each major taxonomic (approximately,
920 phylum-level) group (dots) for 99 different species (40 groups represented by 265 genomes). Many points
921 with low numbers of genomes assigned to a few classes are superimposed. Taxonomic affiliations are
922 indicated for groups found at more than one site or depth (83% occur at > 1 depth). Generally, groups

923 with many genomes (e.g., Alphaproteobacteria) have the genomes assigned to many classes (overall linear
924 trend). **(b)** Hierarchical clustering based species presence/absence patterns for the Horonobe URL.
925 Clustering mostly groups samples from the same site in different years. Samples from intermediate
926 distances from the tunnels cluster with each other and with the earliest samples collected from the most
927 distant site. Samples from 140 m and 250 m depths may be similar due to site locations very close to the
928 access tunnels. The major subdivision in the sites is predicted by the presence of *Candidatus* groups
929 *Altarchaeales*, *Methanoperedens*, *Saccharibacteria*, *Deltaproteobacteria*, and *Betaproteobacteria*.
930 *Saccharibacteria* was detected in 2013 and 2014 at 167, 194 and 218 m depth sites, but not detected in
931 2015, contributing to the split in clustering for the 2015 samples.



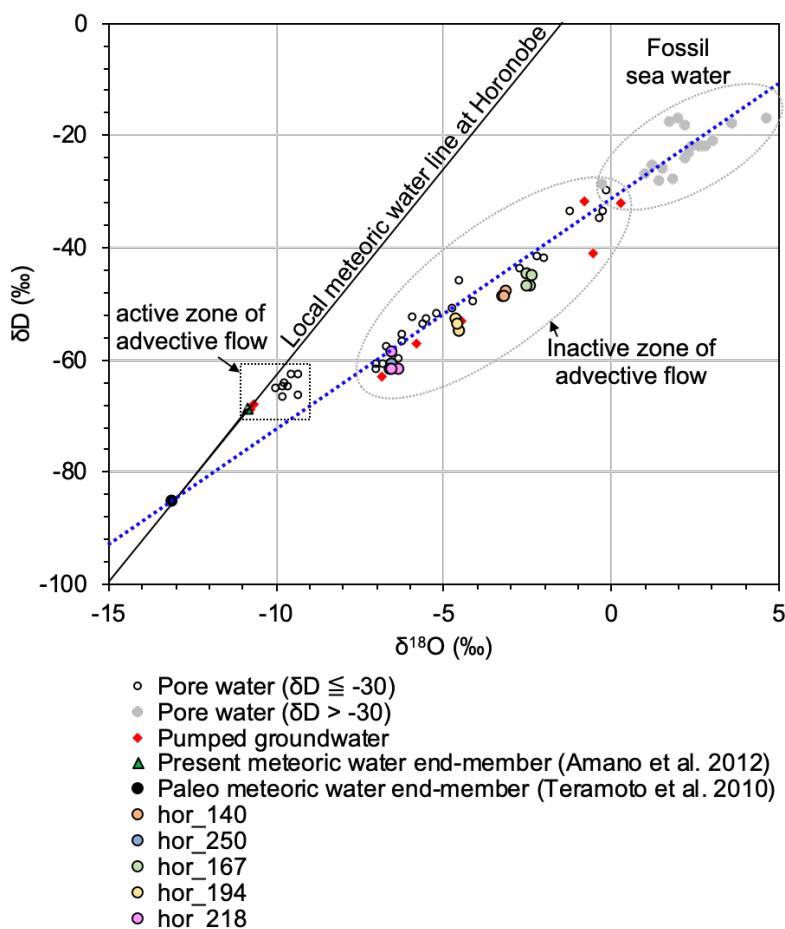
932

933 **Figure S4.** Phylogenetic tree showing the diversity of Rubisco forms I, II, and II/III. Forms I and II are likely
934 involved in CO₂ fixation via the CBB pathway. In most cases, phosphoribulokinase, another marker gene
935 for this pathway, was identified (**Table S7**). Orange and blue branches are form 1' and some form III
936 sequences (included for context).



937

938 **Figure S5.** Comparison of the genome of a *Lentimicrobium* (Bacteroidetes) from the Horonobe URL
939 (Hor_250_2016) to a genome from Rifle, CO groundwater (GWF2). Despite geographic and depth
940 separation of these subsurface sites, the genome-wide average nucleotide identity determined based on
941 alignable segments (which comprised > 85% of each genome bin) is 99.1%. **(a)** Alignment of the
942 concatenated GWF2 *Lentimicrobium* genome bin scaffolds to the concatenated *Lentimicrobium* genome
943 bin from Hor_250_2016. The colored blocks linked by lines indicate alignable sequence and the vertical
944 axis of each box is % identity. **(b)** Segment of the aligned genomes showing nearly 100% sequence identity
945 throughout.



946

947 **Figure S6.** The relationship between $d^{18}\text{O}$ and $d\text{D}$ in pore water and groundwater collected from the
948 Horonobe URL (modified from Mochizuki & Ishii, 2022). Based on the comparison of $d^{18}\text{O}$ and $d\text{D}$ between
949 pore water and pumped groundwater, “active” and “inactive” zones of advective flow of meteoric water
950 through fractures at present are suggested for the Horonobe samples (colored circles). The groundwater
951 samples collected for microbiology analyses are classified as from currently inactive advective flow zones.

952 https://docs.google.com/spreadsheets/d/16RG_b07TiZx0E-

953 [U7qR5j7lZOsc08WVQS/edit#gid=1381792296](https://docs.google.com/spreadsheets/d/1U7qR5j7lZOsc08WVQS/edit#gid=1381792296)

954

955 **Supplementary Tables:**

956

957 https://docs.google.com/spreadsheets/d/16RG_b07TiZx0E-

958 [U7qR5j7lZOsc08WVQS/edit#gid=1381792296](https://docs.google.com/spreadsheets/d/1U7qR5j7lZOsc08WVQS/edit#gid=1381792296)

959

960

961 **Data S1. Non-redundant Newick format RPS3 protein phylogenetic tree**

962 (Data_S1_Newick format rpS3 protein phylogenetic tree.newick)