A metagenomic view of novel microbial and metabolic diversity found within the deep terrestrial biosphere

Running title: Metagenomic analysis of deep terrestrial fracture fluids

Authors: Lily Momper[1], Caitlin P. Casar[1], and Magdalena R. Osburn[1*]

[1]Department of Earth and Planetary Sciences, Northwestern University, Evanston IL, USA

Subject Category: Integrated genomics and post-genomics approaches in microbial ecology

Conflict of Interest Statement: the authors have no conflict of interest to report

This work was funded by the NASA Exobiology Program (NNX15AM086), David and Lucile Packard Foundation, and Canadian Institute for Advanced Research (Earth 4D).

*Correspondence: Magdalena R. Osburn

Northwestern University

Department of Earth and Planetary Sciences

2145 Sheridan Road, Technological Institute F379

Evanston IL, 60208

E-mail: maggie@northwestern.edu

ABSTRACT

The deep terrestrial subsurface is a large and diverse microbial habitat and a vast repository of biomass. However, in relation to its size and physical heterogeneity we have limited understanding of taxonomic and metabolic diversity in this realm. Here we present a detailed metagenomic analysis of samples from the Deep Mine Microbial Observatory (DeMMO) spanning depths from the surface to 1.5 km deep in the crust. From these eight geochemically and spatially distinct fluid samples we reconstructed ~600 metagenome assembled genomes (MAGs), representing 50 distinct phyla and including 18 candidate phyla. These novel clades include many members of the Patescibacteria superphylum and two new MAGs from candidate phylum OLB16, a phylum originally identified in DeMMO fluids and for which only one other MAG is currently available. We find that microbes spanning this expansive phylogenetic diversity and physical space are often capable of numerous dissimilatory energy metabolisms and are poised to take advantage of nutrients as they become available in relatively isolated fracture fluids. This metagenomic dataset is contextualized within a four-year geochemical and 16S rRNA time series, adding another invaluable piece to our knowledge of deep subsurface microbial ecology.

Keywords: Subsurface biosphere/geomicrobiology/carbon fixation/environmental metagenomics/metagenome assembled genomes

INTRODUCTION

Earth's deep biosphere is a vast repository of microbial biomass. Although the size of this biosphere has been continually refined, the most recent estimates now suggest that it is one of the largest biomes on the planet, ranging from 15-23 Pg to perhaps as much as 60 Pg of carbon

sequestered in resident microbes (Whitman *et al.*, 1998; McMahon and Parnell, 2013; Magnabosco *et al.*, 2018; Bar-On *et al.*, 2018; Flemming and Wuertz, 2019). The deep subsurface biosphere (DSB) found underlying the continents or terrestrial biosphere, has emerged as a dynamic, populated, metabolically-active environment, affecting carbon storage and global elemental cycles (Barry et al., 2019; Magnabosco et al., 2018; Flemming and Wuertz, 2019). Although some DSB environments exhibit exceptionally low diversity (Chivian et al., 2008), many of those studied harbor an abundant and diverse microbiome (e.g., Baker et al, 2016; Chivian et al, 2008; Dong et al, 2014; Lau et al, 2014; Magnabosco et al, 2015; Magnabosco et al., 2018; Nyyssönen et al, 2014; Rinke et al, 2013). Furthermore, newly identified subsurface microbial lineages are continually and commonly implicated in major geochemical cycles, highlighting the importance of uncultivated groups in understanding those cycles (Baker et al., 2016; Rasigraf et al., 2014). However, the vast, difficult to access, and heterogeneous nature of the terrestrial DSB complicate efforts to study microbially mediated processes therein. Further confounding our understanding of these processes is the growing appreciation that a high proportion of subsurface microbes have yet to be cultured, known only by culture-independent genetic evidence and colloquially referred to as 'microbial dark matter' (Castelle et al., 2015; Momper et al., 2017; Rinke et al., 2013; Wrighton et al., 2012).

Microbial observatories are portals into the deep subsurface established to monitor subsurface geochemistry and microbial ecology across the world, including in Sweden, Finland, Canada, Switzerland, and the United States (Cardace et al., 2013; Osburn et al., 2019; Pedersen, 1996; Pedersen et al., 2014; Stroes-Gascoyne et al., 2007). The advent of high-throughput shotgun DNA sequencing methods was a pivotal technological advance towards understanding

the microbial diversity of the deep terrestrial subsurface, reducing primer bias and capturing the full breadth of *in situ* diversity. Recent applications of these techniques have revealed that shallow and deep subsurface environments harbor a vast diversity of subsurface taxa particularly in uncultivated, candidate, groups and that spatial separation of taxa occurs with depth and between aquifers (Anantharaman *et al.*, 2016; Jungbluth *et al.*, 2016; Probst *et al.*, 2017; Momper et al., 2017a; Momper et al., 2017b; Probst *et al.*, 2018). Key genomic adaptations discovered within deep intraterrestrials include widespread capacity to fix carbon, particularly with the Wood Lungdal pathway (Lau et al., 2016; Magnabosco et al., 2016; Momper et al., 2017) and a dichotomy of small, ultra-streamlined genomes and larger, bulky genomes with diverse metabolic capabilities (Anantharaman et al., 2016; Jungbluth et al., 2016; Jungbluth et al., 2017; Lau *et al.*, 2016). Current comprehensive reviews of microbial life on and within Earth integrate information from these vast sequencing datasets, estimate the diversity and abundance of the DSB as a whole, and further underscore subsurface microbes' impact on global biogeochemical processes (Flemming and Wuertz, 2019; Magnabosco et al., 2018). However, to date these datasets are heavily focused on shallow subsurface realms, or individual locations within a single very deep site, often without geochemical context, leaving a gap in the range of depths and environments for which we have a clear genomic and metabolic understanding.

In this study we use whole metagenome sequencing methods to investigate microbes from the Deep Mine Microbial Observatory- DeMMO- (Osburn et al., 2019) located in the Sanford Underground Research Facility (SURF) in Lead, South Dakota, USA. We draw on a four-year geochemical and 16S rRNA tag sequencing dataset collected at DeMMO (Osburn et al., 2019, 2020) to integrate microbial, geochemical, and geological observations to expand our genomic interpretation to subsurface microbial processes. Notably, within this metagenomic

dataset, we have near complete metagenome assembled genomes (MAGs) from a rich array of previously uncultured bacteria. We analyze these MAGs to query and characterize metabolic capabilities of microbial communities at DeMMO and use phylogeny of binned MAGs to tie those metabolic capabilities to responsible taxa. We focus our analysis on MAGs from uncultivated groups, especially those for which ours are the most complete genomes to date, as this is the primary mechanism to understand the metabolic role of microbial dark matter (MDM) in the terrestrial DSB.

MATERIALS AND METHODS

Field sampling

All samples for sequencing and corresponding geochemical data were collected from DeMMO borehole fluids and controls (mine service water used for lubricant during borehole drilling and an overlying freshwater stream – Whitewood Creek) between April 16-19, 2018. Microbial cells were filtered from 10 L of fluid onto 47 mm, 0.1 μm Supor filters (Pall Corporation, Port Washington, NY, USA), frozen immediately on dry ice and stored frozen at -80^{o}C. Detailed descriptions of the host geology, borehole characteristics, the establishment of the DeMMO, and methods for geochemical analyses can be found in Osburn et al. (2014) and Osburn et al. (2019). Physical and geochemical characteristics of fluids at the time of sampling are provided in Table 1.

DNA extraction and sequencing

Whole genomic DNA was extracted using a modified phenol-chloroform method with ethanol precipitation as described in Osburn et al. (2014). Library preparation, pooling, quality

control and sequencing were performed at the Environmental Sample Preparation and Sequencing Facility at Argonne National Laboratory, Lemont, IL, USA. Sequencing was performed on an Illumina HiSeq2500 platform, resulting in ~150 bp paired end reads.

De novo assembly, read mapping and generation of metagenome assembled genomes

Trimming of paired end reads was performed using Trimmomatic 0.36 with default parameters and a minimum sequence length of 36 base pairs (Bolger et al, 2014). Reads were assembled using the MEGAHIT assembly algorithm (Li et al., 2015) with a 1,000 bp minimum contig length. Coverage depth information was then generated for scaffolds greater than 1,000 base pairs by mapping the 150 base pair paired-end reads of each sample to its respective assembly using the Burrows-Wheeler Alignment tool, BWA version 0.7.15 (Li and Durbin, 2009) with the BWA default parameters. SAMtools v0.1.17 (Li et al, 2009) was then used to convert files to binary format for downstream analysis.

Metagenome assembled genomes (MAGs) were generated using MetaBAT2, which relies on sequence composition, differential coverage and read-pair linkage (Kang et al., 2019). MAG completeness (reported as percentage of the set of single copy marker genes present) and contamination (calculated as multiple occurrences of a single copy marker gene) were calculated using the "lineage_wf" workflow in the CheckM pipeline (Parks et al, 2015). MAGs that contained >10% contaminating single copy genes were manually refined and curated using the Anvi'o program (Eren et al., 2015). MAG completeness and contamination were subsequently re-calculated using five different standard marker gene suites (Alneberg et al, 2014; Campbell et al, 2013; Creevey et al, 2011; Dupont et al, 2012; Wu & Scott, 2012). MAGs with >10% contamination after curation were removed. Here, we have included all MAGs for which

taxonomy could be assigned, even if those with low completeness (14-30%). Lowering the completeness threshold allowed greater retention of DNA generated from each sample and wider taxonomic representation. Because our analysis relies on gene presence, not absence, the inclusion of these less complete MAGs does not compromise our analyses.

Assignment of putative taxonomies and ribosomal protein tree construction

MAGs were assigned taxonomic identities first according to their placement in a phylogenomic tree using the "tree" command in CheckM (Parks et al, 2015) and refined using the Genome Taxonomy Database and the associated GTDB-Tk toolkit (Chaumeil, et al. 2019; Parks et al., 2018; Parks et al., 2020 https://github.com/Ecogenomics/GtdbTk). A ribosomal protein tree was generated using GToTree (Lee, 2019) and RAxML (Stamatakis, 2014). MAGs were queried for the 15 syntenic and phylogenetically informative ribosomal proteins identified previously (Hug et al., 2016) using HMMER v3.1b1 (Eddy, 2011). If multiple single copy target genes were identified in a single MAG, that particular gene was excluded from the alignment. Ribosomal protein sequences were aligned independently using MUSCLE v. 3.8.31 (Edgar, 2004). Automated trimming of alignments was performed with Trimal (Gutıerrez et al., 2009) and individual alignments from a given MAG were concatenated. MAGs were removed from the final alignment if <40% of the 15 queried ribosomal proteins were identified. 1,673 reference genomes were downloaded from the NCBI database and processed using the same methods to produce a concatenated ribosomal protein alignment for each genome. A maximum likelihood tree generated from this alignment was constructed using RAxML v. 8.2.12 (Stamatakis, 2014) under the LG plus gamma model of evolution (PROTGAMMALG in the RAxML model section), with 1000 bootstrap replicates.

Metabolic pathway analysis

Gene calling and metabolic pathway identification in MAGs was performed using METABOLIC (METabolic And BiogeOchemistry anaLyses In miCrobes) Version 3.0 (Zhou et al., 2019). METABOLIC uses a combination of KOFAM, Pfam, TIGRfam, MEROPs, dbCan2, and customized Hidden Markov Models (HMMs) to identify and classify coding sequences. If >75% of the genes for a given pathway were identified in a single MAG, that pathway was considered present and 'complete', regardless of that MAG's estimated genome completeness.

RESULTS

Metagenome assembled genomes and phylogenetic identification

A total of 581 MAGs were recovered from the eight assembled metagenomes. Genome statistics including MAG ID, NCBI taxon ID, completeness and contamination are listed in Supplementary Data File 1. Of all these 581, 81 were >90%, 99 were 70-89% and 154 were 50-69% complete (Fig. 1). Of the less complete genomes (< 50% complete), more than a quarter (56 of 211) belong to the Candidate Phyla Radiation (CPR). This group is known to contain streamlined genomes often lacking the full suite of 'essential' single copy marker genes, and therefore is not necessarily a reflection of poor assembly (Nelson and Stegen, 2015; Castelle and Banfield, 2018). Similarly, Whitewood Creek (WC), the stream water control, contained the greatest proportion of incomplete genomes (~75% were <50% complete) and the highest proportion of CPR MAGs at 47% (28 MAGs) (Fig. 2a and c).

Of the six DeMMO samples, hereafter called D1-D6, the fewest MAGs were recovered from D1 (25) and the most from D6 (100) (Fig. 1 and Fig. 2). The controls, hereafter referred to as Service Water (SW) and WC, differ from the DeMMO samples in that 1) they lack Archaea 2)

the WC fluid was enriched in the CPR (Fig. 2c) and 3) the SW contained more than twice the number of Proteobacterial MAGs of any other sample (Fig. 2d). Phylum level taxonomic assignments for MAGs can be found in Figure 2a, class break down of the Patescibacteria and Archaea in Figure 2b and 2c, and assignment to the lowest confident taxonomic level in Supplementary Data File 1. MAGs are spread across the phylogenetic tree (Fig. 3). Notably, this dataset is enriched in members of the CPR and uncultivated groups (candidate phyla and classes) compared to groups with cultivated members (Fig. 2 and 3). Currently there are 14 recognized classes within the Patescibacteria (aka CPR) (Parks et al., 2018), 12 of which are found here (Fig. 2c). Uncultured and candidate phyla are present in all sites, including controls and DeMMO boreholes, totaling more than 20 such groups (Fig. 2).

Metabolic capabilities in DeMMO fluids planktonic communities

Our objective in this analysis is to present an overview of the metabolic potential of DeMMO genomes over the represented spatial and geochemical landscape and to attribute this potential to specific taxonomic groups. To this end, we present genes for prominent energy metabolisms at all sites including WC and SW to contrast surface waters to subsurface fluids. Genes which are diagnostic for a variety of dissimilatory metabolisms were queried in MAGs and those of most interest are displayed in Fig. 4 (complete gene name, Pfam/TIGRfam are in Supplementary Data File 2). A complete list of all gene hits for all MAGs is presented in Supplementary Data File 3.

Nitrogen

196 Nitrogen is present in many oxidation states (-3 to +5) which are transformed through a
197 variety of microbially mediated redox reactions. In respiratory denitrification, nitrate (NO_3^-),
198 nitrite (NO_2^-), nitric oxide (NO) and nitrous oxide (N_2O) are sequentially reduced to dinitrogen
199 gas (N_2), with each step catalyzed by one or more metalloenzymes: *nar, nir, nor* and *nos*
200 catalyzing each step (Zumft 1997). The potential for nitrate reduction is present all samples, with
201 the intracellularly (*nar*) mediated reduction being relatively more abundant than the periplasmic
202 (*nap*) form (Fig. 4). Microbial nitrite reduction (NO_2^- to NO) is catalyzed by two related
203 respiratory periplasmic enzymes: the *nirK*-encoded copper nitrite reductase or the *nirS*-encoded
204 cytochrome cd_1 (Berks et al., 1995; Brittain et al., 1992; Green et al., 2010). These genes are
205 relatively scarce compared to those that catalyze nitrate reduction, however they were identified
206 in all samples with the exception of *nirS* in D3 MAGs (Fig. 4). The metabolic potential for the
207 final step in denitrification varies across samples as *nosD* was identified in MAGs from D3 and
208 5, but not D1, D4 of D6 and the canonical marker gene, *nosZ,* was found in D2-D6, but not D1
209 (Fig. 4). Dissimilatory nitrate reduction to ammonium (DNRA) bacteria use NrfA as their key
210 enzyme. Assisted by its redox partner NrfH, NrfA catalyzes the six-electron reduction of nitrite
211 to ammonium (Simon, 2002; Einsle, 2011; Simon and Klotz, 2012). Genes for both subunits are
212 highly abundant in all DeMMO samples (Fig. 4). The *nxrA* and *nxrB* genes are powerful
213 functional and phylogenetic markers to detect and identify uncultured nitrite oxidizing bacteria
214 (Anantharaman et al., 2016; Crane et al., 1995; Daniels et al., 1986) and these genes are most
215 prominent in D2, with lesser detection in D3, D5 and D6 (Fig. 4).

216 *Sulfur*

Sulfur is a versatile element and can be metabolically transformed between eight oxidation states (-2 to +6). Canonical genes involved in sulfite (SO_2^-), sulfate (SO_4^{3-}) and thiosulfate ($S_2O_3^{2-}$) reduction (*sat, asr,* and *phs,* respectively) were queried in all MAGs. *Sat* was highly abundant across all DeMMO samples, while *phsA* was less common, and the complete set of *asr* subunits (*A, B* and *C*) were identified exclusively in D6 (Fig. 4). Sulfur species can also be used as electron donors. Genes involved in sulfur oxidation, *sqr* and *sdo,* were identified in all DeMMO samples and *sdo* was relatively abundant (Fig. 4). SQR is known to be used for sulfide (H_2S) detoxification and therefore is not necessarily indicative of dissimilatory processes, but regardless its presence indicates the potential to transform H_2S to elemental sulfur (S^0) or sulfate (SO_4^{2-}). Microbially mediated S transformation in the environment can be quite complex as many enzymes that transform sulfur can use it as a terminal electron acceptor (TEA) or electron donor, depending on environmental conditions. Among these are dissimilatory sulfite reductase (*dsr*), adenosine-5′-phosphosulfate (*aprA*) and the Sox gene cluster (*sox*). The *dsr* genes are the most common across all sites, followed by *aprA.* Sox genes were far less common, most abundant in D1, but *soxC* was not identified at all in D3 and D6 (Fig. 4). *DsrD*, the small subunit of dissimilatory sulfite reductase that is essential for sulfate reduction, corresponds to the abundance of *dsrAB* in D1-D3 and D6, but is slightly less abundant in D4 and D5 (Fig. 4).

Hydrogen and Oxygen

Hydrogen gas is present in DeMMO fluids (Osburn et al., 2014; Osburn et al., 2019) and can be a powerful electron donor in subsurface environments (Nealson et al., 2005; Takai et al., 2004). We queried a suite of hydrogenases involved in hydrogen-transforming microbial reactions (Sondergaard et al., 2016). Among those hydrogenases, NiFe groups 1, 3ABD and 3C

were identified in all six DeMMO fluids and were the most abundant hydrogenases. FeFe group B were least abundant, with identification in only D6. FeFe group C2 and NiFe group 4H were also relatively uncommon, identified only in D3 and D1, respectively. D1 had by far the fewest number and types of hydrogenases whereas D3 and D6 had the greatest.

Aerobic bacteria and archaea use oxygen as their respiratory TEA. Oxygen-respiring genes are most abundant in the control WC and SW samples, followed closely by D1, and D4 (Figure 4). Cytochrome oxidase (Cox) catalyzes the reduction of oxygen to water and is therefore essential for aerobic metabolism (Capaldi, 1990; Chan and Li, 1990; Saraste, 1990; Babcock and Wikstrom, 1992). Subunits *coxA* and *coxB* were found in all eight samples (Figure 4). Another terminal oxidase, the bo_3 oxidase (Cyo), was found in four of the eight samples: D1, D4, WC and SW (Figure 4). In summary, oxygen respiring genes were found in all samples, but the distribution of specific genes differed between sites.

Metals and halogens

Known genes for iron, manganese and selenite reduction, as well as those involved in halogen cycling, were identified in all samples to varying degrees. Genes indicative of metal reduction, *mtrB, mtrC,* were detected in all DeMMO fluids, relatively most common in D2, but not detected in WC. Putative selenate reduction genes were found in all samples save WC and was highest in D5 and D6 (Fig. 4). The canonical gene for chlorite reduction (*cld*) was particularly abundant in the controls, whereas putative perchlorate reduction (*pcrA* and *pcrB*) was identified in samples D2-D6, but not controls. The gene implicated in reductive halogenation (*pceA*), typically associated with members of the phylum Chloroflexi (specifically

the *Dehalococcoidia*), was found in D3-D6, with highest abundances in D3 and D6 (Fig. 4) corresponding to the relative abundance of Chloroflexi.

Carbon respiration and fixation

C1 carbon metabolisms were investigated owing to their potential importance to subsurface metabolism including carbon monoxide (CO), methanol (CH_3OH), formate (CHO_2^-), and formaldehyde (CH_2O) oxidation. Genes known to be involved in CO oxidation (*coxLMS*) were abundant in all sites except WC and those used in formate oxidation (*fdhAb, fdoHG, fdwB*) were abundant across all samples (Fig. 4). Genes responsible for formaldehyde oxidation (*fae, fghA, frmA*) were variable across samples: they were not present in D3 or D6, at low abundance in D2 and D5, and most abundant in D1, D4 and controls (Fig. 4). Putative methanol oxidation (*mxfA*) was most abundant in D1, D4, SW and WC, was present but in low abundance in D2 and D6, and was not identified in D3 (Fig. 4). While an in-depth interrogation of heterotrophic pathways is beyond the scope of this analysis, genes for fermentation were queried and found in all fluids, but at noticeably lower abundance in WC, SW and D1 than D2-6.

The capability of autotrophic carbon fixation is arguably an important one in carbon limited deep subsurface environments and is widespread at DeMMO. We detected genes for five of the six empirically demonstrated carbon fixation pathways (Hugler and Sievert, 2010) across samples. A list of genes queried, and which carbon fixation mode they are associated with can be found in Supplementary Data File 4. As illustrated in Fig. 5, genes indicative of the 3-hydroxypropionate bicycle were identified in MAGs from all 8 fluids. MAGs in D6 contained a relatively high abundance of *abfD,* a key gene in the 3-Hydroxypropionate/4-Hydroxybutyrate

281 CO_2 fixation cycle, but it was less common in other DeMMO fluids, and absent from WC and
282 SW (Fig. 5). RuBisCO, which catalyzes the carboxylation and oxygenation of ribulose 1,5-
283 bisphosphate (Tabita et al., 2008) was found at quite low abundances in all fluids, as were *aclA*
284 and *aclB*, key genes in the reverse TCA cycle (Fig. 5). Importantly, genes for the Wood
285 Ljungdahl pathway (*cooS, cdhE, cdhD*) were the most abundant in DeMMO fluids when
286 compared to the other four CO_2 fixation cycles queried, but were not detected in the controls,
287 SW and WC (Fig. 5).

288 **DISCUSSION**

289 *Controls and possible contamination*

290 The control samples, WC and SW, were included in our analysis to assess their overall
291 taxonomic and metabolic similarities and differences relative to DeMMO subsurface sites. We
292 found stark differences in high-level taxonomic groups (e.g.- absence of Archaea from control
293 samples, presence of Cyanobacteria in both control samples but not DeMMO fluids- Fig. 2).
294 Differences in putative metabolisms between DeMMO and control fluids (e.g.- absence of
295 hydrogenases and metal reducing genes and relatively high abundance of oxygen respiring genes
296 in control samples, Fig. 4) were also apparent. In the service water, which is used as lubricant
297 during boring activities and was used to drill D4-D6 in 2016, we detected a high abundance of
298 Proteobacteria, compared to the DeMMO fluids. This water originates as municipal water for the
299 city of Lead, South Dakota and is stored underground in tanks until it is used for lubricant or
300 other uses in the underground mining facilities. While these are not quantitative measurements of
301 contamination or lack thereof, these differences combined with analyses from other DeMMO
302 datasets show that DeMMO fluids are distinct from laboratory controls, surface waters, and

service water (Osburn et al., 2019; Casar et al., 2020; Rowe et al., 2020). Indeed, previous 16S rRNA gene sequencing performed over a 4-year timeseries within the same DeMMO boreholes demonstrated that the service water has a significantly different microbial community compared to the DeMMO fluids, and that there is no evidence for cross contamination (Osburn et al., 2020). Because of qualitative differences observed between sites in this metagenomic dataset and those shown between controls and DeMMO fluids using 16S rRNA analyses, we are confident that the trends we describe are not driven by contamination acquired either within the mine during sampling or during laboratory processing. That said, fluid paths within the subsurface environment have the potential to exchange microbes. For instance, WC and D1 are likely hydrologically connected as are and D4 and D5 (Osburn et al., 2019) and thus a natural connection in the microbiology is to be expected. We focus our subsequent discussion specifically on the bacterial groups from the DeMMO subsurface fluids.

Overall metabolic capabilities in DeMMO subsurface fluids

One key question regarding the microbial ecology of the deep terrestrial subsurface is that of geochemical cycling of labile elements (e.g.- carbon, iron, nitrogen, sulfur), and how subsurface microbes interact with those elements in their bioavailable forms (considering both assimilatory and dissimilatory processes). With increasing access to paired geochemical and biological (both DNA and RNA) information from subsurface environments around the globe, we can begin to understand the scope of subsurface microbial biogeochemical cycling. Although we cannot determine the microbial processes that are occurring with metagenomes, we can make predictions about what processes the DBS community is capable of at the DeMMO locations that vary geologically, geochemically, and spatially.

Metabolic capabilities by taxonomic group

One significant benefit to binning MAGs from metagenomic data is that metabolic pathways can be linked to discrete genome bins with accompanying taxonomy. Here we discuss a suite of 27 potential metabolic reactions (Supplementary Data File 5) and tie them to specific cultivated and uncultivated phyla (Fig. 6, Supplementary Figs. 1-15). In the following Results and Discussion we will refer to MAGs with the notation 'SAMPLE_#,' indicating the fluid sample from which they were reconstructed and the numerical rank of that MAG's relative completeness among all MAGs from that single sample.

Predicted metabolic capabilities of common cultivated phyla

Members of cultivated phyla likely play key roles in subsurface biogeochemical cycling at DeMMO. Members of the Proteobacteria have broad metabolic capabilities, mediating 25 of the 27 reactions that were queried, with the exception of methanogenesis (# 6) and sulfite reduction (# 27) (Supplementary Figs. 1-3). Nitrospira MAGs also have broad metabolic potential including for heterotrophy, nitrogen and sulfur metabolisms, but also metal reduction and hydrogen oxidation (Fig. 6, Supplementary Fig. 4). These observations are consistent with what is known from cultivated Nitrospira (Daims and Wagner, 2018).

Members of the Firmicutes have been found in global subsurface environments and often increase in abundance with depth and play key roles in S-cycling (Baker et al., 2003; Cowen et al., 2003; Chivian et al., 2008; Aüllo et al., 2013; Tiago and Veríssimo, 2013; Magnabosco et al., 2015; Jungbluth et al., 2016). Firmicutes (including the newly delineated Firmicutes groups B, D and E) are quite common in D6 (15 MAGs), and four MAGs each from Firmicutes group E and

346 B were reconstructed from D3 and D5, respectively (Figure 2). Firmicutes in D6 have a wide
347 array of potential metabolisms, including methanotrophy (#5), carbon fixation (#4), hydrogen
348 generation (#19) and selenate reduction (#16) (Supplementary Figs. 5-7). Although there are
349 only four Firmicutes group B MAGs identified in D5, they appear capable of diverse
350 metabolisms involving all elements queried (carbon, hydrogen, nitrogen, oxygen, sulfur, metals)
351 except for those dissimilatory metabolisms involving nitrogen species (Supplementary Fig. 6).

352 MAGs from the recently cultivated phylum Elusimicrobia (Geissinger et al., 2009) were
353 identified in D3-D6 fluids. The MAG in D3 (DeMMO3_73) may be capable of fermentation
354 and/or coupling carbon oxidation to selenate or arsenate reduction. Putative Elusimicrobia in D5
355 (DeMMO5_39 and _42) may be active in carbon and nitrogen cycling in those fluids, possibly
356 capable of carbon fixation, simple carbon compound oxidation with selenate reduction, and
357 possibly nitrite ammonification (#11). The single Elusimicrobia genome reconstructed from D6
358 fluids appears to be quite metabolically diverse although incomplete (50.0%/0.6%
359 completeness/contamination) containing genes for reactions involving all elements of interest,
360 save those involving hydrogen generation (#19) or oxidation (#18) (Supplementary Fig. 8).

361 *Predicted metabolic capabilities of uncultivated bacterial phyla*

362 This dataset contains a wealth of MAGs from uncultivated bacterial groups. The ability to
363 assign metabolic processes to members the candidate phyla is particularly powerful as it is one of
364 the few sources of information on these abundant and under characterized organisms. This
365 information can further assist in metagenome informed, targeted culturing of novel groups,
366 adding to its value. For example, OLB16 is a recently named bacterial phylum, for which there

367 are only two publicly available genomes (https://gtdb.ecogenomic.org/searches?s=al&q=OLB16,
368 accessed October 10, 2020). One of these genomes was recovered from D6 fluids in a 2017 study
369 (Momper et al., 2017). In this previous study we predicted that OLB16 (SURF_12) may be
370 capable of nitrate reduction, sulfite oxidation, methane oxidation, and carbon fixation via the
371 Wood Ljungdahl pathway (Momper et al., 2017). Here, we reconstructed two more OLB16
372 MAGs in D3 and D6 fluids (DeMMO3_89, DeMMO6_4). These new MAGs appear capable of
373 sulfur oxidation and/or reduction (#21-23), heterotrophy (#1), fermentation, and metal reduction
374 (#17). The more complete MAG in D6 also appears to transform hydrogen (#18, #19)
375 (Supplementary Fig. 19), indicating a wide repertoire of potential metabolisms for this newly
376 identified candidate phylum.

377 Additional newly identified phyla (Parks et al., 2017) were found in DeMMO fluids:
378 UBA 9089, UBA3054, and UBA10199. The uncultivated phylum UBA9089 was identified in
379 D1 and D3-D5 (Figure 2). Among those MAGs, sulfur and carbon cycling metabolisms are
380 ubiquitous, with MAGs in D1, D4 and D5 potentially capable of selenate reduction (#16) and
381 perhaps even capable of carbon fixation (#4) (Supplementary Fig. 10). UBA3054 (Parks et al.,
382 2017), was identified in D6 fluids (DeMMO6_5). Currently there are only three publicly
383 available genomes from this newly identified phylum, of which DeMMO6_5 is by far the most
384 complete (other available MAGs are ~54-58% compared to ~99% complete). This MAG
385 contains genes capable of heterotrophy (#1), fermentation (#3), arsenate reduction (#14),
386 elemental sulfur oxidation (#21) and hydrogen generation (#19) (Supplementary Fig. 12).
387 UBA10199 was found almost exclusively in D2 fluids (DeMMO2_63 and _70) with one MAG
388 recovered from D3. These MAGs appear capable of heterotrophy (#1-3) and elemental sulfur
389 (#21) and thiosulfate oxidation (#25) (Supplementary Fig. 12). The phylum UBP1, first reported

in 2017 (Parks et al., 2017), was found in D5-6 (DeMMO5_30, DeMMO6_31). Metabolic reconstruction indicates that members of this phylum have diverse metabolic capabilities, potentially mediating sulfur reduction and/or oxidation (reactions #21-23), nitric oxide reduction (#10), and ammonification (#11) (Supplementary Fig. 13).

A single MAG from the candidate phylum Latescibacteria, formerly known as WS3, was found in D2 fluids (DeMMO2_58). Despite its relative incompleteness (~48% complete), this MAG contains diverse metabolic potential with genes involved in 16 of the 27 reactions queried, most notably for H_2 generation and oxidation and metal reduction (reactions #17-19) (Supplementary Fig. 14). Two MAGs from the recently designated candidate phylum Margulisbacteria (formerly candidate division ZB3) were found in D3 (DeMMO3_43 and _47). These appear to be capable of elemental sulfur and carbon oxidation coupled to arsenate reduction (#14) and potentially fermentation (Supplementary Fig. 15). Establishment of fermentation and hydrogen dependent metabolisms in this candidate phylum have been well documented (Carnevali et al., 2019; Utami et al., 2019), potential sulfur oxidation and arsenate reduction have not been reported, and may point to a differing ecological niche, due to the differing metabolic role for these new MAGs.

Carbon fixation in DeMMO subsurface fluids

Although photosynthetically derived organic carbon does infiltrate into Earth's subsurface, it is often recalcitrant and at very low concentrations (Pedersen, 2000). Dissolved organic carbon (DOC) concentrations at DeMMO sites are very low (generally <1 mg/L) (Osburn et al., 2019). As this DOC is limited and potentially of low quality, autotrophic carbon

411 fixation may be a key metabolic adaptation, especially in the deeper, older DeMMO fluids. Quite
412 telling is the fact that genes involved in the Wood Ljungdahl pathway were by far most abundant
413 across all six DeMMO fluids yet not detected in the controls (Figure 5). This trend is supported
414 by the geochemical conditions in these fluids (Table 1). At the time of sampling, the oxidation
415 reduction potential in DeMMO fluids ranged from -61 to -198 mV, indicating little to no oxygen
416 present. The Wood Ljungdahl pathway requires anoxic conditions, as some of its enzymes,
417 especially the crucial acetyl-CoA synthase, are highly oxygen sensitive (Berg, 2011). In contrast,
418 the SW and WC fluids are at near oxygen saturation (~300 mV, data not shown), making it
419 highly unlikely this form of carbon fixation could be performed. The Wood Ljungdahl pathway
420 has previously been hypothesized to be a crucial source of primary production in D6 (Momper et
421 al., 2017) and other deep terrestrial subsurface environments (Magnabosco et al, 2015), a
422 conclusion supported by this study. Similarly, it follows that RuBisCo, the key enzyme which
423 catalyzes the carboxylation and oxygenation of ribulose 1,5-bisphosphate (Tabita et al., 2008)
424 was not common in the dark and reducing fluids of DeMMO (Figure 6). Other trends in carbon
425 fixation gene presence are not as straightforward. For example, the 3-hydroxypropionate bi-cycle
426 is commonly used by members of the Chloroflexi (Hügler and Sievert, 2010), yet the fluids with
427 the greatest number of Chloroflexi MAGs (D6) do not have the greatest proportion of identified
428 canonical *prpE* genes. Along the same vein, the 3-Hydroxypropionate/4-Hydroxybutyrate mode
429 of carbon fixation is typically associated with the archaea (Berg et al., 2007; Loder et al., 2016),
430 yet in this study we find the highest proportion of *abfD* genes in MAGs from D6 (Figure 6),
431 fluids from which we did not construct archaeal MAGs (Figure 2).

432 *Tying metabolisms to in situ geochemistry*

Combining geochemistry with microbiology is challenging in any environmental setting, but is especially demanding in deep, difficult to access, subsurface environments. Monitoring at the DeMMO observatory has enabled a rare opportunity to combine 16S rRNA tag sequencing, whole metagenomic DNA sequencing and assemblies, and geochemical analyses over a multi-year period. Other terrestrial subsurface studies have combined geochemical analyses with extensive metagenomic datasets to gain insight into elemental cycling, microbial community connectedness and microbial metabolic 'handoffs' (Anantharaman et al., 2016; Lau et al., 2014; Magnabosco et al., 2016). These studies found that deep (> 1 km) subsurface communities tend to rely heavily on the Wood Ljungdahl pathway for potential carbon fixation and primary production (Magnabosco et al., 2016) and that some subsurface habitats seem to preserve ancestral gene signatures in biogeochemically relevant functional genes (Lau et al., 2014).

In general terms, the distribution of metabolically functional genes identified in this MAG dataset agree well with our understanding of *in situ* geochemical conditions at DeMMO. As detailed in the results, we find broad metabolic potential for biogeochemical cycling across the DeMMO sites, but variation both between sites. For example, sulfate and nitrate are available TEAs in all fluids, and sulfur and nitrogen cycling genes are abundant across all DeMMO samples (Table 1 and Fig. 4), particularly those sites where there are abundant forms of N and S (D4, D5). Along these same lines, hydrogenases were identified in abundance in D2-6, but were far less abundant in D1, the DeMMO borehole with the lowest average dissolved hydrogen concentration, and were nearly or completely absent from MAGs binned from the oxic control samples (Table 1 and Fig. 4).

However, deeper examination of the MAGs in this DeMMO dataset revealed a number of instances where the most abundant functional genes in a given site were not consistent with the most abundant geochemical species at the time of DNA collection 2018, or in previous sampling campaigns over the last 4 years. For example, D4 has historically extremely low dissolved oxygen levels (~0.03 mg/L) and during sampling in 2018 the measured ORP was -200 mV, however, we identified a wide variety and relatively large abundance of oxygen respiring genes in D4 MAGs (Table 1 and Fig. 4). It is therefore not clear where these organisms are acquiring the oxygen necessary for aerobic respiration. In terms of carbon metabolisms, genes involved in C1 metabolism and fermentation have varying patterns across all DeMMO fluids, but are quite abundant in all samples (Fig. 4). But dissolved organic carbon levels are quite low in all fluids (<0.5 and <3.0e^{-6} mg/L, respectively).

In general, the average gene count for MAGs recovered from DeMMO boreholes was not small, at ~1,900-2,400 genes per MAG, with typical larger genomes containing ~4,000-6,000 genes (Supplementary Data File 6). This leads us to question why deep subsurface microbes would retain genes and entire functional pathways when the geochemical substrate for that metabolism is not present or present at such low concentrations that the reaction is not energetically favorable. We hypothesize that microbes in this deep subsurface environment gain a competitive edge by maintaining a wide variety of functional pathways. Indeed, in this study we identified diverse putative metabolisms in MAGs from little-studied groups such as the Elusimcrobia and candidate phyla such as OLB16. Our findings indicate that the capability to perform many dissimilatory energy metabolisms is the norm rather than the exception in non-CPR MAGs in this deep subsurface environment. Other groups have reported similar findings in shallow and deep subsurface environments. Large, robust genomes were found in anoxic/near-

anoxic and TEA-limited shallow aquifer environments (Anantharaman et al., 2016). When electron acceptors such as oxygen and nitrate were injected into these aquifers, there was an almost instantaneous draw down of those substrates, and they were again near detection limit within hours (Anantharaman et al., 2016). Similarly, the deep subsurface isolate *Desulforudis audaxviator* contains the genomic potential for an astonishing array of catabolic and anabolic metabolisms including sulfate reduction, carbon and nitrogen fixation, heterotrophy, and others (Chivian et al., 2008). Our findings support the growing body of evidence that in many subsurface biomes there is a dichotomy of small, ultra-streamlined genomes and larger, bulky genomes with diverse metabolic capabilities (Anantharaman et al., 2016; Jungbluth et al., 2016; Jungbluth et al., 2017; Lau *et al.*, 2016).

With time series data (Osburn et al., 2019), long term *in situ* experiments (Casar et al., 2020) and metagenomic surveys (Momper et al., 2017, this study) we are beginning to piece together how planktonic and attached microbial communities vary temporally and interact with and influence *in situ* geology and geochemistry in a deep terrestrial subsurface environment, DeMMO. In addition to long-term, incremental understanding about terrestrial subsurface processes, metagenomic studies such as these can significantly augment our understanding of novel microbial groups and their role in these environments. This understanding enables deeper questions, such as what percentage of elemental cycling can be attributed to biotic vs abiotic processes, and on what timescales these processes occur for the whole of the DSB and for environments that may be of particular interest. While no two subsurface sites are the same, there are commonalities including an enrichment of Firmicutes, uncultured candidate phyla, and abundant members of the CPR. Metabolically these groups have the potential for widespread

carbon utilization and fixation, sulfur and metal-based metabolisms, and potential roles in rapid drawdown of injected TEAs such as nitrate and oxygen.

CONCLUSIONS

A pillar of advancing environmental microbiology rests on culture independent methods to assessment of metabolic potential of microbes. Here we present nearly 600 high quality MAGs deriving from an established transect through the subsurface, from the surface to 1.5 km deep into the crust. This dataset includes a diversity of taxonomic groups including those from well-cultivated but metabolically important lineages like the Proteobacteria and Nitrospira, as well as from virtually unknown taxa known only from a small number of assembled genomes. In these genomes we find the genetic potential to mediate a range of environmentally important metabolisms including nitrogen, sulfur, and metal cycling as well as C1-metabolisms, fermentation, and carbon fixation. The spatial distribution of this metabolic potential often agrees well with available chemical species, but there are intriguing instances of disagreement which suggest that maintaining genomic plasticity is a key adaptive strategy of many intraterrestrial microbes. This work fits into a growing body of work at DeMMO which facilities integrated understanding of the deep subsurface biosphere in space, through time, and in chemical and environmental context. Further, these genomes add to and expand the growing body of subsurface genomes, informing the capabilities of novel taxa and expanding our ability to understand this biome on the global scale.

DATA DEPOSIT

519 Sequence data for metagenomic assemblies and their respective MAGs and
520 corresponding metadata can be accessed using the BioProject identifier PRJNA563685 and
521 BioSample accessions SAMN18064095, SAMN18064236, SAMN18064310, SAMN18064413,
522 SAMN18064496, SAMN18064575, SAMN18004502, and SAMN18005272 corresponding to
523 sites DeMMO1-6, Whitewood Creek, and Service Water communities, respectively. All code
524 and corresponding data used in this study are available at
525 https://github.com/CaitlinCasar/Momper2021_DeMMO_FractureFluidMetagenomes.

526 **ACKNOWLEDGEMENTS**

527 This work was supported by NASA Exobiology (NNH14ZDA001N) and grants to MRO
528 from the David and Lucille Packard Foundation and the Canadian Institute for the Advancement
529 of Research Earth 4D. We would like to recognize Michael D. Lee for helpful conversations on
530 phylogenomic analyses, phylogenetic tree building and assistance in using GToTree. We also
531 want to thank the developers of METABOLIC and Karthik Anantharaman for their help in
532 running and interpreting the data generated using that package. We want to thank especially staff
533 and personnel at SURF for access to the deep subsurface and repeated access to samples used in
534 this study and Dr. Brittany Kruger of the Desert Research Institute for coordination of field
535 sampling expeditions.

536 **Conflict of interest**

537 The authors have no conflicts of interest to report.

538 **SUPPLEMENTAL INFORMATION**

539 Supplementary information is available at ISMEJ's website

Figure 1. Statistical breakdown of MAG completeness for all 582 genomes across eight samples.

Figure 2. Taxonomic grouping of all MAGs reconstructed in this study, at the phylum and class levels. A) Phylum level taxonomic assignments for MAGs in terms of relative abundance per sample. B) Classes of Archaea identified in fluids. C) MAGs from the Candidate Phyla Radiation (CPR) identified in this study.

Figure 3. Concatenated ribosomal protein tree containing all MAGs for which at least 40% of target ribosomal proteins could be identified. Phyla with ** (Chlorobi and Ignavibacteria) have traditionally been considered separate phyla within the FCB superphylum. The GTDBTk toolkit has included them in the Bacteroidota phylum. We have kept them separate in the ribosomal protein tree for clarity and because it is not yet widely accepted that they should be classified in the same phylum.

Figure 4. Functional gene annotations indicative of potential energy-yielding metabolisms in MAGs across all eight fluid samples. Canonical genes for common electron donors and acceptors were queried in all MAGs binned from DeMMO fluids and two control samples.

Figure 5. Modes of carbon fixation in MAGs, across all eight fluid samples. Five of the six documented modes of carbon fixation are included in this figure, for which we have deterministic, essential canonical marker genes. Gene abbreviation are listed to the right of the plot, along with full pathway name. Gene abundance was normalized by number of reconstructed MAGs for each site individually.

Figure 6. Metabolic chord diagrams for 27 energy yielding metabolic reactions of interest. Diagrams are separated by sample and represent the metabolic potential for all MAGs reconstructed from each respective fluid.

References

Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, *et al.* (2014). Binning metagenomic contigs by coverage and composition. *Nature Methods* 11:1144-1146.

Anantharaman K, Brown CT, Hug LA, Sharon I, Castelle CJ, Probst AJ, Thomas BC, Singh A, Wilkins MJ, Karaoz U, Brodie EL, Williams KH, Hubbard SS, Banfield JF (2016.) Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat Commun* 7:13219. doi:10.1038/ncomms13219.

Aüllo T, Ranchou-Peyruse A, Ollivier B, Magot M. (2013). *Desulfotomaculum* spp. and related gram-positive sulfate-reducing bacteria in deep subsurface environments. *Front Microbiol* 4:362.

Baker BJ, Moser DP, MacGregor BJ, Fishbain S, Wagner M, Fry NK, *et al.* (2003). Related assemblages of sulphate□reducing bacteria associated with ultradeep gold mines of South

Africa and deep basalt aquifers of Washington State. *Environl Microbiol* 5:267-277.

Baker BJ, Lazar CS, Teske AP, Dick GJ. (2015). Genomic resolution of linkages in carbon, nitrogen, and sulfur cycling among widespread estuary sediment bacteria. *Microbiome* 3:14.

Baker BJ, Saw JH, Lind AE, Lazar CS, Hinrichs K-U, Teske AP, Ettema TJ. (2016). Genomic inference of the metabolism of cosmopolitan subsurface Archaea, Hadesarchaea. *Nature Microbiology* 1:16002.

Bar-On YM, Phillips R, Milo R. (2018). The biomass distribution on Earth. *PNAS* **115**: 6506–6511.

Barry, P.H., Nakagawa, M., Giovannelli, D. *et al.* (2019). Helium, inorganic and organic carbon isotopes of fluids and gases across the Costa Rica convergent margin. *Sci Data* **6,** 284. https://doi.org/10.1038/s41597-019-0302-4

Beal EJ, House CJ, Orphan VJ. (2009). Manganese- and iron-dependent marine methane oxidation. *Science* 325:184–187

Belay, and B. S. Rajagopal. 1986. Assimilatory reduction of sulfate and sulfite by methanogenic bacteria. *Appl. Environ. Microbiol*. 51:703-709.

Berg IA. (2011). Ecological aspects of the distribution of different autotrophic CO_2 fixation pathways. *Appl Environ Microbiol* 77:1925-1936.

Berks BC, Ferguson SJ, Moir JW, Richardson DJ (1995). Enzymes and associated electron transport systems that catalyse the respiratory reduction of nitrogen oxides and oxyanions. *Biochim Biophys Acta* 1232(3):97-173.

Bolger AM, Lohse M, Usadel B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114-2120.

Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Schulz F, Doud D, *et al.* (2017). Genome standards for single amplified genomes and genomes from metagenomes of Bacteria and Archaea. *Nature Biotechnology* in press.

Braker G, Tiedje JM (2003). Nitric oxide reductase (*norB*) genes from pure cultures and environmental samples. *Appl Environ Microb* 69(6):3476-3483. doi:10.1128/AEM.69.6.3476-3483.2003.

Brittain T, Blackmore R, Greenwood C, Thomson AJ (1992). Bacterial nitrite-reducing enzymes. *Eur J Biochem* 209(3):793-802.

Brown,CT, Hug LA, Thomas BC, Sharon I, Castelle CJ, Singh A., *et al.* (2015). Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* 523: 208–211.

Campbell JH, O'Donoghue P, Campbell AG, Schwientek P, Sczyrba A, Woyke T, *et al.* (2013). UGA is an additional glycine codon in uncultured SR1 bacteria from the human

610 microbiota. *Proc Natl Acad Sci USA* 110:5540-5545.

611 Cardace D, Hoehler T, McCollom T, Schrenk M. (2013). Establishment of the coast range
612 ophiolite microbial observatory (CROMO): drilling objectives and preliminary
613 outcomes. *Sci. Drill.* 16, 45–55. doi: 10.5194/sd-16-45-2013
614
615 Castelle CJ, Hug LA, Wrighton KC, Thomas BC, Williams KH, Wu D, *et al.* (2013).
616 Extraordinary phylogenetic diversity and metabolic versatility in aquifer sediment.
617 *Nature Communications* 4: 2120.

618 Castelle CJ, Wrighton KC, Thomas BC, Hug LA, Brown CT, Wilkins MJ. (2015). Genomic
619 expansion of domain archaea highlights roles for organisms from new phyla in anaerobic
620 carbon cycling. *Current Biology* 25:690-701.

621 Chaumeil PA, et al. 2019. GTDB-Tk: A toolkit to classify genomes with the Genome Taxonomy
622 Database. *Bioinformatics*, btz848.

623 Chen LX, Méndez-García C, Dombrowski N, Servín-Garcidueñas LE, Eloe-Fadrosh EA, Fang
624 BZ, *et al*. (2017). Metabolic versatility of small archaea Micrarchaeota and
625 Parvarchaeota. *ISME J* 12:756-775

626 Chivian D, Brodie EL, Alm EJ, Culley DE, Dehal PS, DeSantis TZ, *et al.* (2008) Environmental
627 genomics reveals a single-species ecosystem deep within Earth. *Science* 322: 275–278.

628 Cowen JP, Giovannoni SJ, Kenig F, Johnson HP, Butterfield D, Rappé MS, Hutnak M, Lam P.
629 (2003). Fluids from aging ocean crust that support microbial life. *Science* 299:120-123.

630 Crane BR, Siegel LM, Getzoff ED (1995) Sulfite reductase structure at 1.6 A: evolution and
631 catalysis for reduction of inorganic anions. *Science* 270(5233):59-67.Daniels, L., N.

632 Creevey CJ, Doerks T, Fitzpatrick DA, Raes J, Bork P. (2011). Universally distributed single-
633 copy genes indicate a constant rate of horizontal transfer. *PLoS One* 6:e22099.

634 Daims H, Wagner M. (2018). Nitrospira. *Trends in Microbiol*, 26:462-463.

635 Darling AE, Jospin G, Lowe E, Matsen IV FA, Bik HM, Eisen JA. (2014). PhyloSift:
636 phylogenetic analysis of genomes and metagenomes. *PeerJ* 9:e243.

637 Derakshani M, Lukow T, Liesack W. (2001). Novel bacterial lineages at the (sub)division level
638 as detected by signature nucleotide-targeted recovery of 16S rRNA genes from bulk soil
639 and rice Grimm F, Franz B, Dahl C. (2008). Thiosulfate and sulfur oxidation in purple
640 sulfur bacteria. In *Microbial Sulfur Metabolism* (101-116). Springer Berlin Heidelberg.

641 roots of flooded rice microcosms. *Appl Environ Microbiol* 67:623–631.

642 Dong Y, Kumar CG, Chia N, Kim PJ, Miller PA, Price ND, *et al*. (2014). *Halomonas*
643 *sulfidaeris*-dominated microbial community inhabits a 1.8 km-deep subsurface
644 Cambrian Sandstone reservoir. *Environ Microbiol* 16:1695-1708.

645 Dunham JP, Friesen ML. (2013). A cost-effective method for high-throughput construction of
646 Illumina sequencing libraries. *Cold Spring Harbor Protocols*. 9.

647 Dupont C, Rusch DB, Yooseph S, Lombardo MJ, Richter RA, Valas R, *et al*. (2012). Genomic
648 insights to SAR86, an abundant and uncultivated marine bacterial lineage. *ISME J*
649 6:1186-1199.

650 Edwards RA, Rodriguez-Brito B, Wegley L, Haynes M, Breitbart M, Peterson DM, *et al*. (2006).
651 Using pyrosequencing to shed light on deep mine microbial ecology. *BMC Genomics*,
652 7:57.

653 Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont TO. (2015).
654 Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ*, 3:e1319.

655 Farag IF, Youssef NH, Elshahed MS. (2017). Global Distribution Patterns and Pangenomic
656 Diversity of the Candidate Phylum "*Latescibacteria*" (WS3). *Applied and Environmental*
657 *Microbiology*. 10.1128/AEM.00521-17

658 Flemming, H.-C. and Wuertz, S. (2019) Bacteria and archaea on Earth and their abundance in
659 biofilms. *Nat Rev Micro* 1–14.

660 Fredrickson JK, Onstott TC. (1996). Microbes deep inside the earth. *Scientific American*,
661 275:68-73.
662 Ghosh W, Dam B. (2009). Biochemistry and molecular biology of lithotrophic sulfur oxidation
663 by taxonomically and ecologically diverse bacteria and archaea. *FEMS microbiology*
664 *reviews*, 33:999-1043.

665 Geissinger O, Herlemann DPR, Mörschel E, Maier UG, Brune A. (2009). The

666 Ultramicrobacteruim "Elusimicrobium minutum" gen. nov., sp. nov. the First Cultiaved
667 Representative of the Termite Group 1 Phylum. *Applied and Environmental*
668 *Microbiology* 75:2831-2840

669 Gihring TM, Moser DP, Lin LH, Davidson M, Onstott TC, Morgan L, *et al*. (2006). The
670 distribution of microbial taxa in the subsurface water of the Kalahari Shield, South
671 Africa. *Geomicrobiology Journal* 23:415–430.

672 Green SJ, Prakash O, Gihring TM, Akob DM, Jasrotia P, Jardine PM, Watson DB, Brown SD,
673 Palumbo AV, Kostka JE (2010) Denitrifying bacteria isolated from terrestrial subsurface
674 sediments exposed to mixed-waste contamination. *Appl Environ Microbiol* 76(10):3244-
675 54. doi:10.1128/AEM.03069-09.

676 Haveman, S.A. and Pedersen, K. (2002) Distribution of culturable microorganisms in
677 Fennoscandian Shield groundwater. *FEMS Microbiology Ecology* 39: 129–137.

678 Hu P, Tom L, Singh A, Thomas BC, Baker BJ, Piceno YM, *et al.* (2016). Genome-resolved
679 metagenomic analysis reveals roles for candidate phyla and other microbial community
680 members in biogeochemical transformations in oil reservoirs. *mBio* 7:e1669-15.

681 Hugenholtz P, Pitulle C, Hershberger KL, Pace NR. (1998). Novel division level bacterial
682 diversity in a Yellowstone hot spring. *Journal of Bacteriology* 180:366-376.

683 Hug, L.A., Baker, B.J., Anantharaman, K., Brown, C.T., Probst, A.J., Castelle, C.J., et al. (2016)
684 A new view of the tree of life. *Nature Microbiology* 1–6.

685 Hügler M, Sievert SM. (2011). Beyond the Calvin cycle: autotrophic carbon fixation in the
686 ocean. *Marine Science* 3:261-289.

687 Huntemann M, Ivanova NN, Mavromatis K, Tripp HJ, Paez-Espino D, Palaniappan K, *et al*.
688 (2015). The standard operating procedure of the DOE-JGI Microbial Genome Annotation
689 Pipeline (MGAP v. 4). *Standards in Genomic Sciences* 10:86.

690 Hyatt D, LoCascio PF, Hauser LJ, Uberbacher EC. (2012). Gene and translation initiation site
691 prediction in metagenomic sequences. *Bioinformatics* 28:2223-2230.

692 Jumas-Bilak E, Roudière, Marchandin H. (2009). Description of '*Synergistetes'* phyl. nov. and
693 emended description of the phylum *'Deferribacteres'* and of the family
694 *Syntrophomonadaceae* phylum *'Firmicutes'*. *IJSEM* 59:1028-1035

695 Jungbluth SP, Grote J, Lin HT, Cowen JP, Rappé MS. (2013). Microbial diversity within

696 basement fluids of the sediment-buried Juan de Fuca Ridge flank. *ISME J* 7:161–172.

697 Jungbluth SP, Glavina del Rio T, Tringe SG, Stepanauskas R, Rappé MS. (2016) Genomic
698 comparisons of a bacterial lineage that inhabits both marine and terrestrial deep
699 subsurface systems. *PeerJ Preprints* 4:e2592v1.

700 Kallmeyer J, Pockalny R, Adhikari RR. (2012). Global distribution of microbial abundance and
701 biomass in subseafloor sediment. *Proc Natl Acad Sci USA* 109:16213-16216.

702 Kantor RS, Wrighton KC, Handley KM, Sharon I, Hug LA, Castelle CJ, *et al.* (2013). Small
703 genomes and sparse metabolisms of sediment-associated bacteria from four candidate
704 phyla. *MBio* 4:e708-13.

705 Karnachuk OV, Lukina PA, Kadinkov VV, Sherbakova, VA, Beletsky AV, Mardanov AV, *et al.*
706 (2020). Targeted isolation based on metagenome□assembled genomes reveals a
707 phylogenetically distinct group of thermophilic spirochetes from deep biosphere. *SFAM.*
708 10.1111/1462-2920.15218

709 Kolinko S, Richter M, Glöckner FO, Brachmann A, Schüler D. (2015). Single□cell genomics of
710 uncultivated deep□branching magnetotactic bacteria reveals a conserved set of
711 magnetosome genes. *Environmental Microbiology* 18:21-37.

712 Langmead B, Salzberg SL. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*,

713 9:357-359.

714 Lau MCY, Cameron C, Magnabosco C, et al. (2014). Phylogeny and phylogeography of

functional genes shared among seven terrestrial subsurface metagenomes reveal N-cycling and microbial evolutionary relationships. *Frontiers in Microbiology*. 5:531.

Lee KCY, Herbold CW, Dunfield PF, Morgan XC, McDonald IR, Stott MB. (2013). Phylogenetic Delineation of the Novel Phylum *Armatimonadetes* (Former Candidate Division OP10) and Definition of Two Novel Candidate Divisions. *ASMJ* 79:2484-2487

Lin LH, Wang PL, Rumble D, Lippmann-Pipke J, Boice E, Pratt LM, *et al*. (2006). Long-term sustainability of a high-energy, low-diversity crustal biome. *Science* 314:479–482.

Lin X, Kennedy D, Fredrickson J, Bjornstad B, Konopka A. (2012). Vertical stratification of subsurface microbial community composition across geological formations at the Hanford Site. *Environ Microbiol* 14:414–425.

Lollar BS, Lacrampe-Couloume G, Slater GF, Ward J, Moser DP, Gihring TM, *et al*. (2006). Unravelling abiogenic and biogenic sources of methane in the Earth's deep subsurface. *Chemical Geology*, 22:328-339.

Lovley DR, and Chapelle FH. (1995). Deep subsurface microbial processes, *Rev Geophys,* 33:365-381.

Loy A, Duller S, Baranyi C, Mußmann M, Ott J, Sharon I, *et al.* (2009). Reverse dissimilatory sulfite reductase as phylogenetic marker for a subgroup of sulfur□oxidizing prokaryotes. *Environmental microbiology*, 11:289-299.

Ludwig W, Strunk O, Westram R, Richter L, Meier H, Buchner A, Brettske I. (2004). ARB: a software environment for sequence data. *Nucleic Acids Res* 32:1363-1371.

McMahon S, Parnell J. (2014). Weighing the deep continental biosphere. *FEMS Microbiol Ecol*, 87: 113-120.

Magnabosco C, Ryan K, Lau MC, Kuloyo O, Lollar BS, Kieft TL, *et al*. (2016). A metagenomic window into carbon metabolism at 3 km depth in Precambrian continental crust. *ISME J* 10:730-741.

Matsen FA, Kodner RB, Armbrust EV. (2010). pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC bioinformatics*, 11:1-16.

Markowitz VM, Ivanova NN, Szeto E, Palaniappan K, Chu K, Dalevi D, *et al.* (2008). IMG/M: a data management and analysis system for metagenomes. *Nucleic Acids Res* 36:534-538.

Markowitz VM, Chen IM, Chu K, Szeto E, Palaniappan K, Pillay M, *et al*. (2014). IMG/M 4 version of the integrated metagenome comparative analysis system. *Nucleic Acids Res* 42:568-573.

McMahon, S. and Parnell, J. (2013) Weighing the deep continental biosphere. *FEMS Microbiol.*

Ecol. 87: 113–120.

Momper LM, Reese BK, Carvalho G, Lee P, Webb EA. (2015). A novel cohabitation between two diazotrophic cyanobacteria in the oligotrophic ocean. *ISME J* 4:882-893.

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

Momper, L., Kiel Reese, B., Zinke, L., Wanger, G., Osburn, M.R., Moser, D., and Amend, J.P. (2017b) Major phylum-level differences between porefluid and host rock bacterial communities in the terrestrial deep subsurface. *Environmental Microbiology Reports* 9: 501–511.

Momper, L., Aronson, H.S., and Amend, J.P. (2018) Genomic Description of "Candidatus Abyssubacteria," a Novel Subsurface Lineage Within the Candidate Phylum Hydrogenedentes. *Frontiers in Microbiology* **9**: 1144–11.

Nancharaiah YV, Lens PNL. (2015). Ecology and Biotechnology of Selenium-Respiring Bacteria. *Microbiology and Molecular Biology Reviews.* 10.1128/MMBR.00037-14.

Nelson, WC, Stegen, JC. (2015). The reduced genomes of Parcubacteria (OD1) contain signatures of a symbiotic lifestyle. *FMICB* 6:713

Noha YH, Ibrahim FF, Rinke C, Hallam SJ, Woyke T, Elshahed MS. (2015). In *Silico* Analysis of the Metabolic Potential and Niche Specialization of Candidate Phylum *"Latescibacteria"* (WS3). *PLoS ONE* 10(6): e0127499.doi:10.1371/journalpone.0127499

Nyyssönen M, Hultman J, Ahonen L, Kukkonen I, Paulin L, Laine P, Itävaara M, Auvinen P. (2014). Taxonomically and functionally diverse microbial communities in deep crystalline rocks of the Fennoscandian shield. *ISME J*, 8:126-138.

Onstott TC, Phelps TJ, Colwell FS, Ringelberg D, White DC, Boone DR, *et al*. (1998). Observations pertaining to the origin and ecology of microorganisms recovered from the deep subsurface of Taylorsville Basin, Virginia. *Geomicrobiology Journal* 15:353-585.

Orcutt, BN, Sylvan, JB, Knab, NJ, Edwards, KJ. (2011). Microbial ecology of the dark ocean above, at, and below the seafloor. *Microbiology and Molecular Biology Reviews*, 75:361-422.

Osburn MR, LaRowe DE, Momper L, Amend JP. (2014). Chemolithotrophy in the continental deep subsurface: Sanford Underground Research Facility (SURF), USA. *Front Extr Microbiol* 5:610.

Osburn, M.R., Kruger, B., Masterson, A., Casar, C., Amend, J.P., and Amend, J. (2019) Establishment of the Deep Mine Microbial Observatory (DeMMO), South Dakota, USA, a Geochemically Stable Portal Into the Deep Subsurface. *Front. Earth Sci.* 7: 1–17.

Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes.

Gen Res 25:1043-1055.

Parks D, Rinke C, Chuvochina M, Chaumiel PA, Woodcroft B, Evans PN, *et al.* (2017). Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nature Microbiology.* 2:1533–1542.

Parks DH, et al. 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nature Biotechnology*, http://dx.doi.org/10.1038/nbt.4229.

Parks, D.H., et al. 2020. A complete domain-to-species taxonomy for Bacteria and Archaea. *Nature Biotechnology*, https://doi.org/10.1038/s41587-020-0501-8.

Pedersen, K. (1996). 16S rRNA gene diversity of attached and unattached bacteria in boreholes along the access tunnel to the Äspö hard rock laboratory, Sweden. *FEMS Microbiol. Ecol.* 19, 249–262. doi: 10.1016/0168-6496(96)00017-7

Pedersen K. (2000). Exploration of deep intraterrestrial microbial life: current perspectives *FEMS Microbiology Letters,* 185:9-16

Pedersen, K. (2001) Diversity and activity of microorganisms in deep igneous rock aquifers of the Fennoscandian Shield. In, Fredrickson,J.K. and Fletcher,M. (eds), *Subsurface Microbiology and Biogeochemistry*. Wiley-Liss, pp. 97–139.

Peng Y, Leung HC, Yiu SM, Chin FY. (2012). IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*, 28:1420-1428.

Pfiffner SM, Cantu JM, Smithgall A, Peacock AD, White DC, Moser, DP, *et al*. (2006). Deep subsurface microbial biomass and community structure in Witwatersrand Basin mines. *Geomicrobiology Journal*, 23:431-442.

Pilhofer M, Rappl K, Eckl C, Bauer AP, Ludwig W, Schleifer KH, *et al*. (2008). Characterization and evolution of cell division and cell wall synthesis genes in the bacterial phyla *Verrucomicrobia, Lentisphaerae, Chlamydiae*, and *Planctomycetes* and phylogenetic comparison with rRNA genes. *J Bacteriol* 190:3192–3202.

Price MN, Dehal PS, Arkin AP. (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. *PloS one* 5:e9490.

Pruesse E, Queast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glöckner FO. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* 35:7188-7196.

Pruesse E, Peplies J, Glöckner FO. (2012). SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics*, 28:1823-1829.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, *et al*. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools.

Nucleic Acids Res 41:590-596.

Reed AJ, Lutz RA, Vetriani C. (2006). Vertical distribution and diversity of bacteria and archaea in sulfide and methane-rich cold seep sediments located at the base of the Florida Escarpment. *Extremophiles* 10:199–211.

Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng JF, *et al.* (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature*, 499:431-437.

Sander J, Engels-Schwarzlose S, Dahl C. (2006). Importance of the DsrMKJOP complex for sulfur oxidation in *Allochromatium vinosum* and phylogenetic analysis of related complexes in other prokaryotes. *Archives of microbiology*, 186:357-366.

Schauder, R, Preuß, A, Jetten, M, Fuchs, G. (1988). Oxidative and reductive acetyl CoA/carbon monoxide dehydrogenase pathway in *Desulfobacterium autotrophicum*. *Archives of Microbiology* 151:84-89.

Seitz KW, Lazar CS, Hinrichs K-U, Teske AP, Baker BJ. (2016). Genomic reconstruction of a novel, deeply branched sediment archaeal phylum with pathways for acetogenesis and sulfur reduction. *ISME J* 10:1696–1705.

Simkus DN, Slater GF, Lollar BS, Wilkie K, Kieft TL, Magnabosco C, *et al*. (2016). Variations in microbial carbon sources and cycling in the deep continental subsurface. *Geochimica et Cosmochimica Acta* 173:264-283.

Speth DR, Guerrero-Cruz S, Dutilh BE, Jetten MS. (2016). Genome-based microbial ecology of anammox granules in a full-scale wastewater treatment system. *Nature Communications* 7:1-10.

Stamatakis A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688-2690.

Stevens T. (1997) Lithoautotrophy in the subsurface. *FEMS Microbiology Reviews*, 20:327-337.

Stevens TO, McKinley JP. (1995). Lithoautotrophic microbial ecosystems in deep basalt aquifers. *Science* 270:450-454.

Stroes-Gascoyne, S., Schippers, A., Schwyn, B., Poulain, S., Sergeant, C., and Simonoff, M., et al. (2007). Microbial community analysis of opalinus clay drill core samples from the mont terri underground research laboratory, Switzerland. *Geomicrobiol. J.* 24, 1–17. doi: 10.1080/01490450601134275

Tamaki H, Tanaka Y, Matsuzawa H, Muramatsu M, Meng XY, Hanada S, *et al*. (2011). *Armatimonas rosea* gen. nov., sp. nov., of a novel bacterial phylum, *Armatimonadetes* phyl. nov., formally called the candidate phylum OP10. *IJSEM* 61:1442-1447

Tabita FR, Satagopan S, Hanson TE, Kreel NE, Scott SS. (2008). Distinct form I, II, III, and IV RuBisCo proteins from the three kingdoms of life provide clues about RuBisCo evolution and structure/function relationships. *Journal of Experimental Botany* 59:1515-1524.

859 Tiago I, Veríssimo A. (2013). Microbial and functional diversity of a subterrestrial high pH
860 groundwater associated to serpentinization. *Environ Microbiol* 15:1687-1706.

861 Wagner M, Horn M. (2006). The *Planctomycetes, Verrucomicrobia, Chlamydiae* and sister phyla
862 comprise a superphylum with biotechnological and medical relevance. *Curr Opin*
863 *Biotechnol* 17: 241–249.

864 Whitman WB, Coleman DC, Wiebe WJ. (1998). Prokaryotes: the unseen majority. *Proc Natl*
865 *Acad Sci USA* 95:6578–6583.

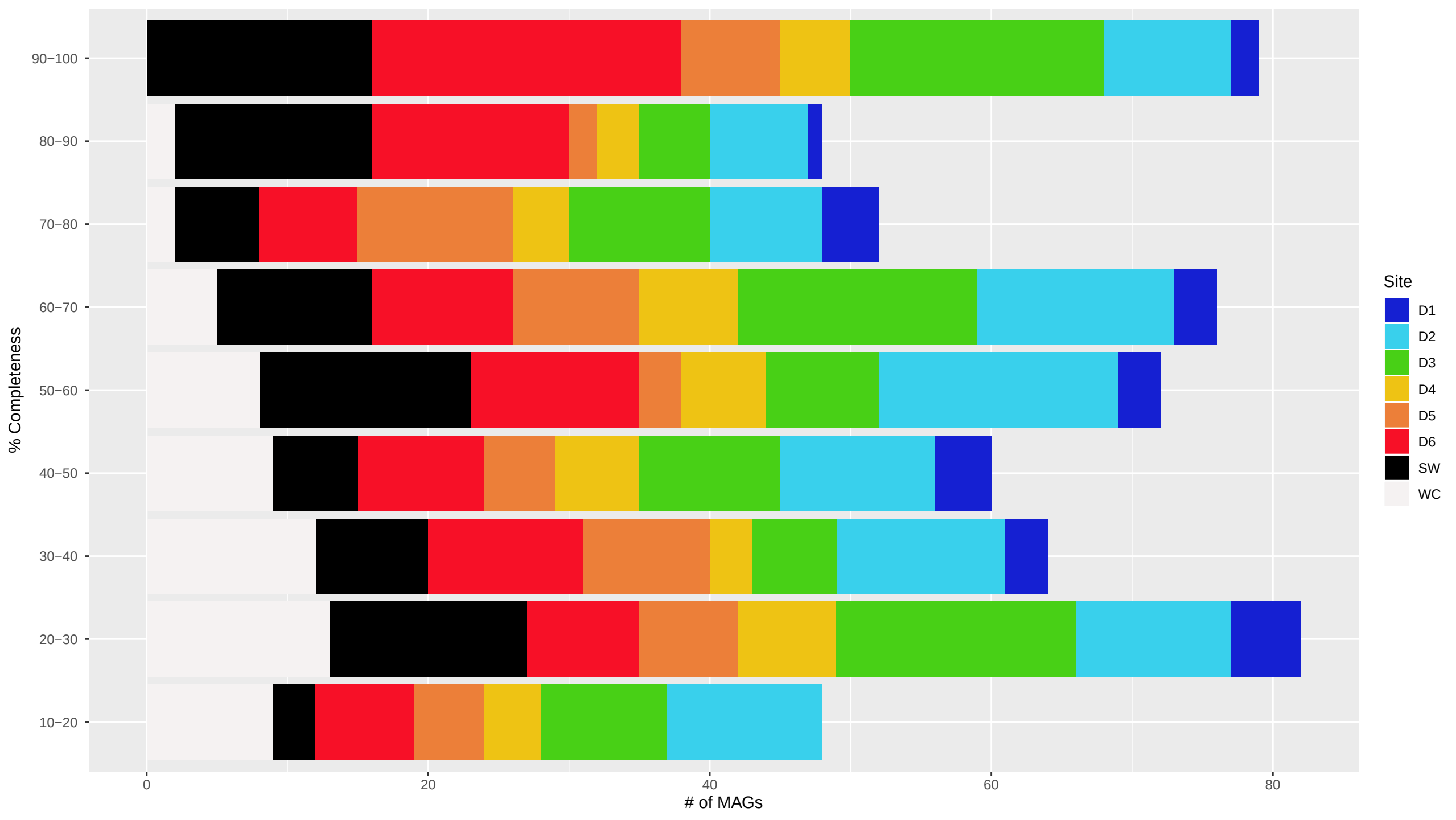
866 Wrighton KC, Thomas BC, Sharon I, Miller CS, Castelle CJ, VerBerkmoes NC, *et al*. (2012).
867 Fermentation, Hydrogen, and Sulfur Metabolism in Multiple Uncultivated Baterial Phyla.
868 *Science Magazine* 337:1661-1665

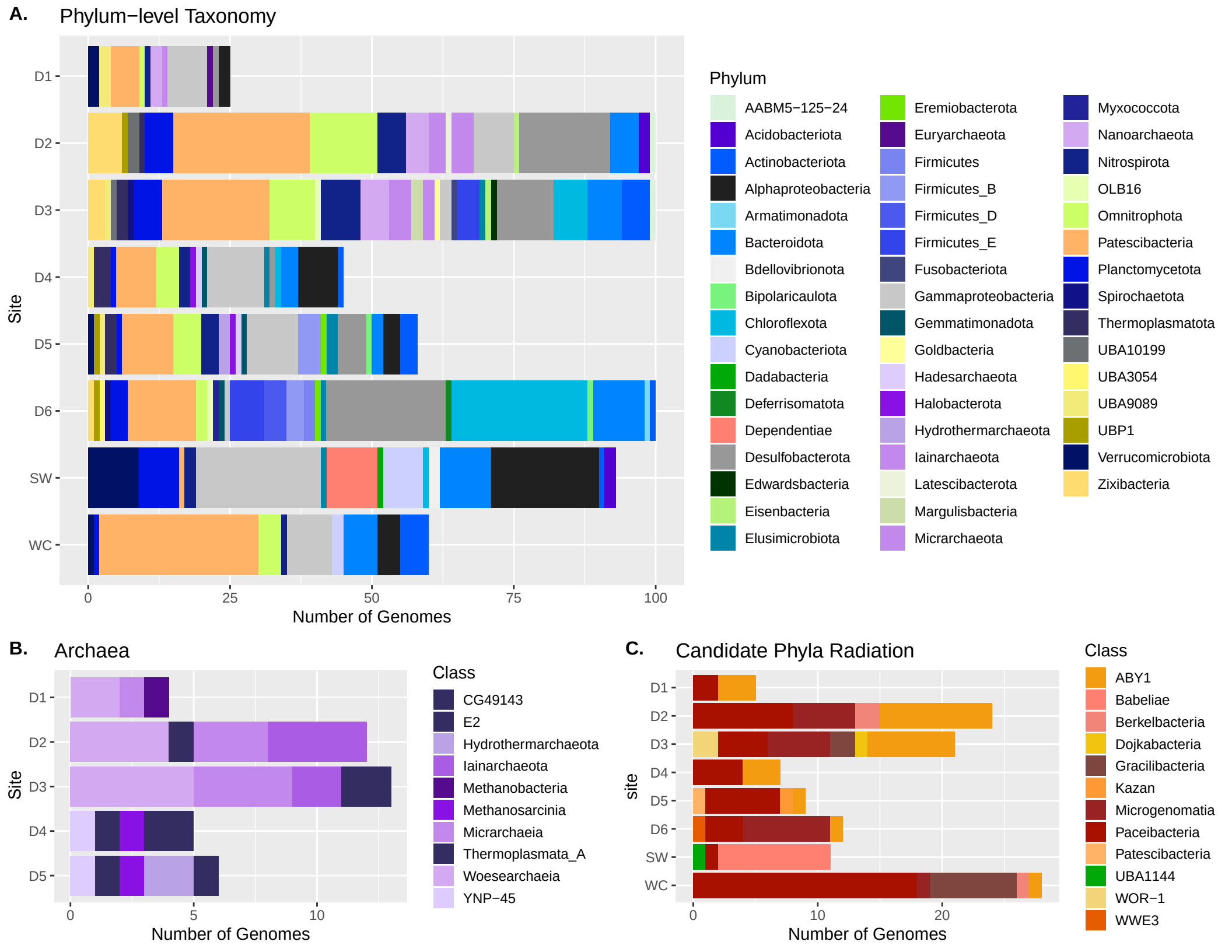
869 Wu M, Scott AJ. (2012). Phylogenomic analysis of bacterial and archaeal sequences with
870 AMPHORA2. *Bioinformatics* 28:1033-1034.

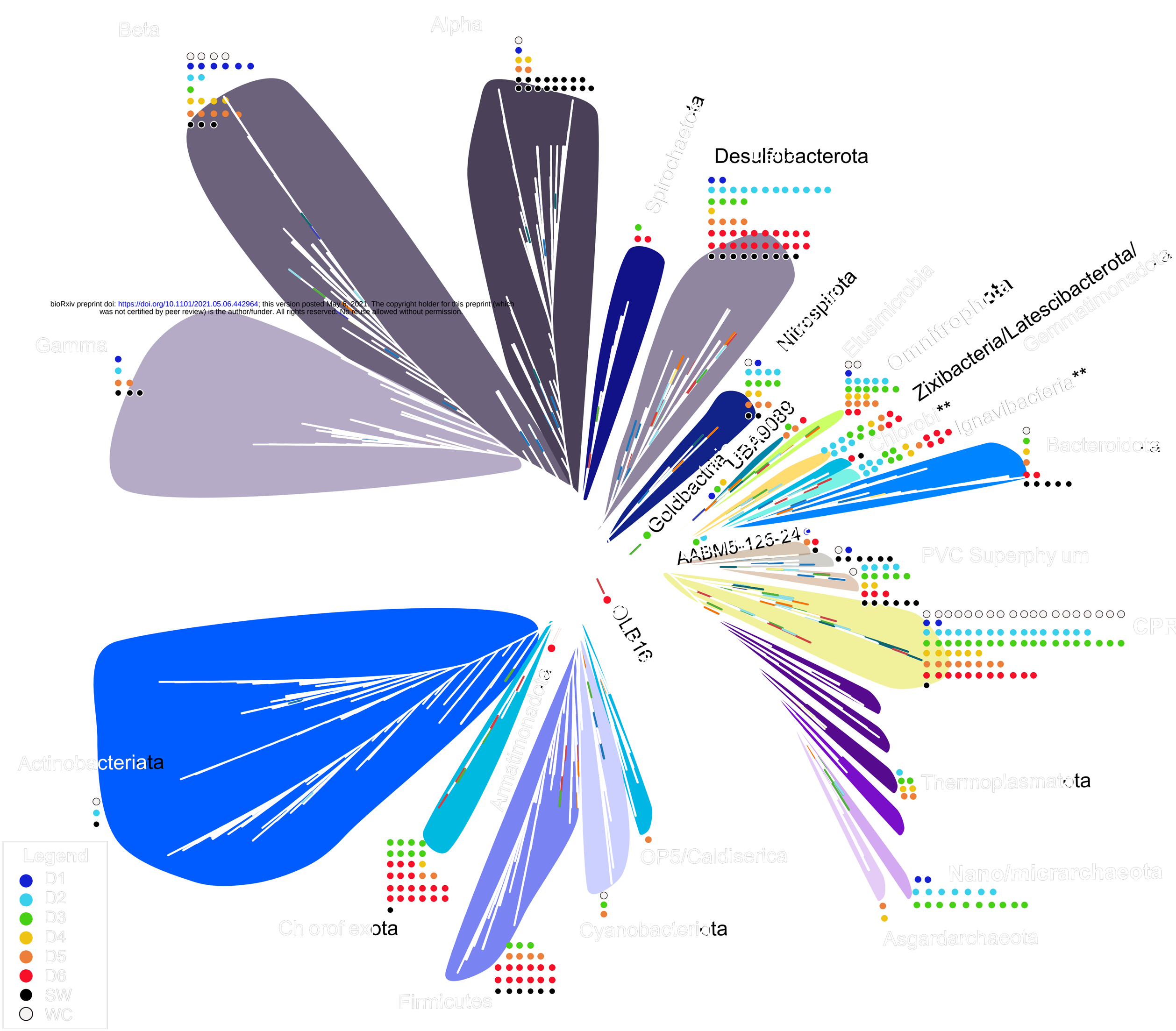
871 Yarza P, Richter M, Peplies J, Euzeby J, Amann R, Schleifer KH, *et al*. (2008). The All-Species
872 Living Tree project: A 16S rRNA-based phylogenetic tree of all sequenced type strains.
873 *Systematic and Applied Microbiology* 31:241-250.

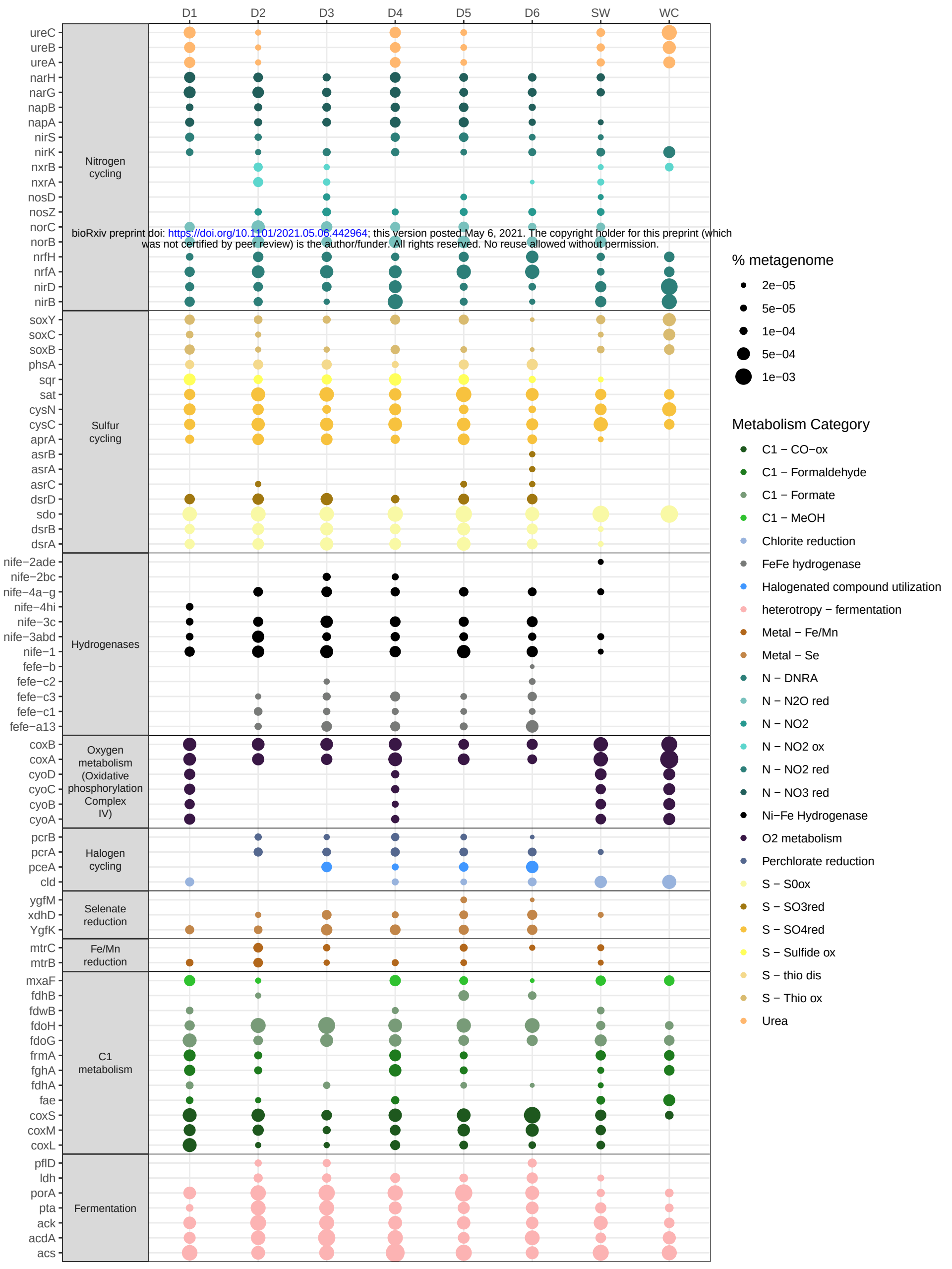
874 Youssef NH, Rinke C, Stepanauskas R, Farag I, Woyke T, Elshahed MS (2015a). Insights into
875 the metabolism, lifestyle and putative evolutionary history of the novel archaeal phylum
876 *'Diapherotrites'*. *ISME J* 9:447–460.

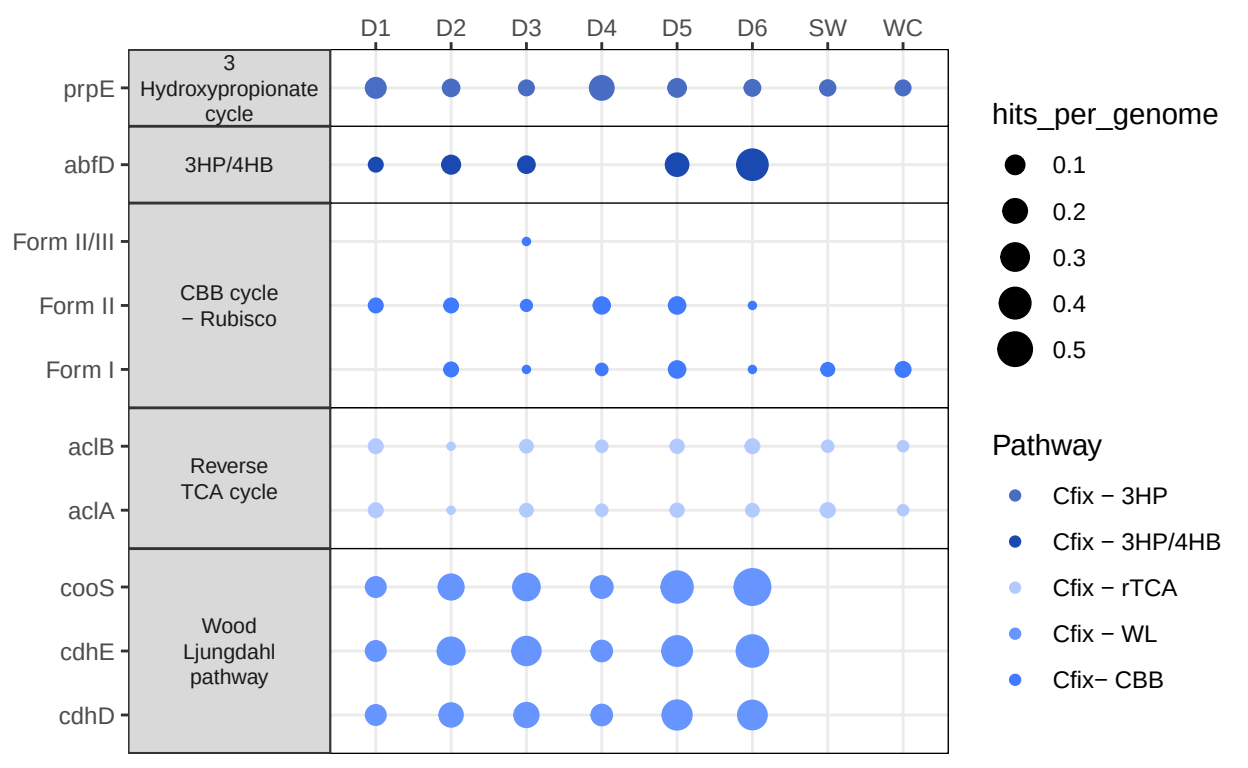
877 Zumft WG. 1997. Cell biology and molecular basis of denitrification. *Microbiol Mol Biol Rev*
878 61(4):533–616.


90−100
80−90
70−80
60−70
50−60
40−50
30−40
20−30
10−20
% Completeness
0
20
40
60
80
# of MAGs
Site
D1
D2
D3
D4
D5
D6
SW
WC


A.
Phylum−level Taxonomy
Site
D1
D2
D3
D4
D5
D6
SW
WC
0
25
50
75
100
Number of Genomes
Phylum
AABM5−125−24
Acidobacteriota
Actinobacteriota
Alphaproteobacteria
Armatimonadota
Bacteroidota
Bdellovibrionota
Bipolaricaulota
Chloroflexota
Cyanobacteriota
Dadabacteria
Deferrisomatota
Dependentiae
Desulfobacterota
Edwardsbacteria
Eisenbacteria
Elusimicrobiota
Eremiobacterota
Euryarchaeota
Firmicutes
Firmicutes_B
Firmicutes_D
Firmicutes_E
Fusobacteriota
Gammaproteobacteria
Gemmatimonadota
Goldbacteria
Hadesarchaeota
Halobacterota
Hydrothermarchaeota
Iainarchaeota
Latescibacterota
Margulisbacteria
Micrarchaeota
Myxococcota
Nanoarchaeota
Nitrospirota
OLB16
Omnitrophota
Patescibacteria
Planctomycetota
Spirochaetota
Thermoplasmatota
UBA10199
UBA3054
UBA9089
UBP1
Verrucomicrobiota
Zixibacteria
B.
Archaea
Site
D1
D2
D3
D4
D5
0
5
10
Number of Genomes
Class
CG49143
E2
Hydrothermarchaeota
Iainarchaeota
Methanobacteria
Methanosarcinia
Micrarchaeia
Thermoplasmata_A
Woesearchaeia
YNP−45
C.
Candidate Phyla Radiation
site
D1
D2
D3
D4
D5
D6
SW
WC
0
10
20
Number of Genomes
Class
ABY1
Babeliae
Berkelbacteria
Dojkabacteria
Gracilibacteria
Kazan
Microgenomatia
Paceibacteria
Patescibacteria
UBA1144
WOR−1
WWE3






D1
D2
D3
D4
D5
D6
SW
WC
prpE
abfD
Form II/III
Form II
Form I
aclB
aclA
cooS
cdhE
cdhD
3 Hydroxypropionate cycle
3HP/4HB
CBB cycle − Rubisco
Reverse TCA cycle
Wood Ljungdahl pathway
hits_per_genome
0.1
0.2
0.3
0.4
0.5
Pathway
Cfix − 3HP
Cfix − 3HP/4HB
Cfix − rTCA
Cfix − WL
Cfix− CBB

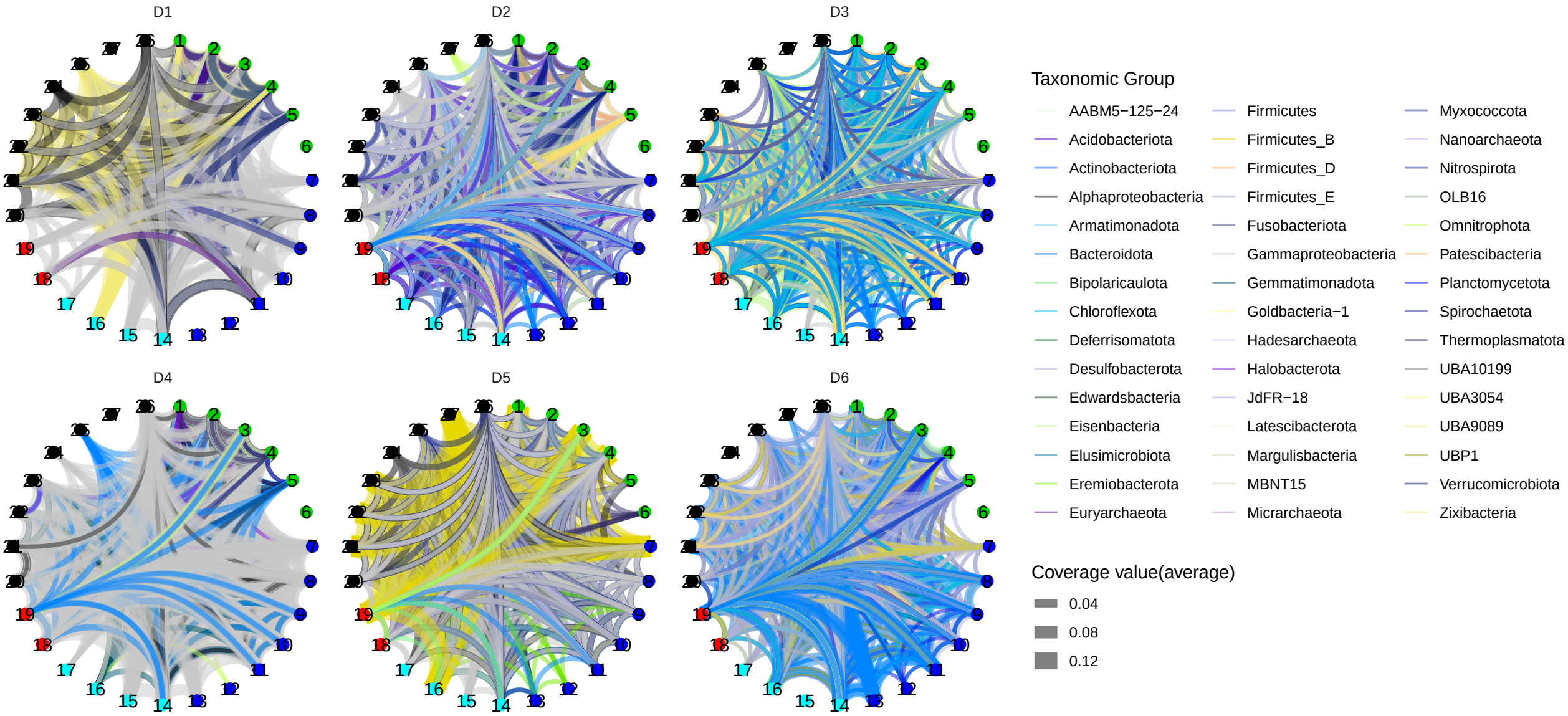

D1
D2
D3
D4
D5
D6
Taxonomic Group
AABM5−125−24
Acidobacteriota
Actinobacteriota
Alphaproteobacteria
Armatimonadota
Bacteroidota
Bipolaricaulota
Chloroflexota
Deferrisomatota
Desulfobacterota
Edwardsbacteria
Eisenbacteria
Elusimicrobiota
Eremiobacterota
Euryarchaeota
Firmicutes
Firmicutes_B
Firmicutes_D
Firmicutes_E
Fusobacteriota
Gammaproteobacteria
Gemmatimonadota
Goldbacteria−1
Hadesarchaeota
Halobacterota
JdFR−18
Latescibacterota
Margulisbacteria
MBNT15
Micrarchaeota
Myxococcota
Nanoarchaeota
Nitrospirota
OLB16
Omnitrophota
Patescibacteria
Planctomycetota
Spirochaetota
Thermoplasmatota
UBA10199
UBA3054
UBA9089
UBP1
Verrucomicrobiota
Zixibacteria
Coverage value(average)
0.04
0.08
0.12

Table 1. Measurements of geochemical constituents collected concomitantly with biologcial samples for DNA e

Site	Date	Flow rate	Temp	Conductivity	TDS	pH	ORP	NO_3^-	NH_4^+
		mL/min	(°C)	(µS)	ppm		(mV)	mg/L	mg/L
DeMMO 1	4/18/2018	3000	10.5	1045	756	5.83	-68	0.3	0.06
DeMMO 2	4/18/2018	300	12	609.4	430.4	7.44	-100	0.3	0.03
DeMMO 3	4/17/2018	4600	16.1	3082	2349	7.13	-61	0.3	0.21
DeMMO 4	4/19/2018	300	22.2	1781	1302	7.46	-200	1.1	1.36
DeMMO 5	4/16/2018	15600	31.6	1534	1103	8.45	-149	0.7	0.48
DeMMO 6	4/16/2018	0	20.1	7933	6602	6.61	-198	0.3	0

*Dissolved gas concentrations were not measured on the day of collection in April of 2018. We have used an av

xtraction and sequencing

Fe^{2+}	Total sulfide	DO	SO_4^{2-}	Total Fe	Total Mn	DOC	DIC	H_2 *	CH_4 *
mg/L	µg/L	mg/L	mg/L	mg/L	mg/L	mg/L	mM	mg/L	mg/L
2.32	6	0.025	393	6.08	0.288	0.468	3.81	1.19E-07	7.75E-06
0.31	0	0.054	84.8	0.35	0.052	0.393	4.02	1.42E-07	5.85E-06
2.86	5	0.029	1800	3.3	0.297	0.25	9.46	2.68E-07	6.88E-05
0	582	0.031	315	0.03	0.005	0.244	11.31	2.16E-07	0.000719
0	254	0.015	186	0	0	0.316	10.96	1.98E-07	0.000346
1.47	13	0	4110	2.44	0.375	0.286	1.86	3.05E-07	0.005041

eraged value from measurements taken December 2015-December 2019 in corresponding boreholes.

CO_2 *	CO *
mg/L	mg/L
0.022831	5.56E-06
0.017753	4.94E-06
0.091982	8.19E-06
0.013094	3.54E-06
0.00477	4.14E-06
0.002487	2.52E-06