



Genomic Description of ‘*Candidatus* Abyssubacteria,’ a Novel Subsurface Lineage Within the Candidate Phylum Hydrogenedentes

Lily Momper^{1*}, Heidi S. Aronson² and Jan P. Amend^{2,3}

¹ Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA, United States, ² Department of Biological Sciences, University of Southern California, Los Angeles, CA, United States, ³ Department of Earth Sciences, University of Southern California, Los Angeles, CA, United States

OPEN ACCESS

Edited by:

Yiran Dong,
China University of Geosciences
Wuhan, China

Reviewed by:

Andreas Schramm,
Aarhus University, Denmark
Kim Marie Handley,
University of Auckland, New Zealand
Charles K. Lee,
University of Waikato, New Zealand

*Correspondence:

Lily Momper
momper@mit.edu

Specialty section:

This article was submitted to
Terrestrial Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 12 March 2018

Accepted: 08 August 2018

Published: 28 August 2018

Citation:

Momper L, Aronson HS and
Amend JP (2018) Genomic
Description of ‘*Candidatus*
Abyssubacteria,’ a Novel Subsurface
Lineage Within the Candidate Phylum
Hydrogenedentes.
Front. Microbiol. 9:1993.
doi: 10.3389/fmicb.2018.01993

The subsurface biosphere is a massive repository of fixed carbon, harboring approximately 90% of Earth’s microbial biomass. These microbial communities drive transformations central to Earth’s biogeochemical cycles. However, there is still much we do not understand about how complex subterranean microbial communities survive and how they interact with these cycles. Recent metagenomic investigation of deeply circulating terrestrial subsurface fluids revealed the presence of several novel lineages of bacteria. In one particular example, phylogenomic analyses do not converge on any one previously identified taxon; here we describe the first full genomic sequences of a new bacterial lineage within the candidate phylum Hydrogenedentes, ‘*Candidatus* Abyssubacteria.’ A global survey revealed that members of this proposed lineage are widely distributed in both marine and terrestrial subsurface environments, but their physiological and ecological roles have remained unexplored. Two high quality metagenome assembled genomes (SURF_5: 97%, 4%; SURF_17: 91% and 4% completeness and contamination, respectively) were reconstructed from fluids collected 1.5 kilometers below surface in the former Homestake gold mine—now the Sanford Underground Research Facility (SURF)—in Lead, South Dakota, United States. Metabolic reconstruction suggests versatile metabolic capability, including possible nitrogen reduction, sulfite oxidation, sulfate reduction and homoacetogenesis. This first glimpse into the metabolic capabilities of these cosmopolitan bacteria suggests that they are involved in key geochemical processes, including sulfur, nitrogen, and carbon cycling, and that they are adapted to survival in the dark, often anoxic, subsurface biosphere.

Keywords: subsurface biosphere, metagenomics, microbial dark matter, *Abyssubacteria*, *Hydrogenedentes*

INTRODUCTION

Although the subsurface biosphere is devoid of light and is often extremely carbon and energy-limited, it is home to the vast majority of Earth’s microbes, as much as 90% by the most recent estimates (Whitman et al., 1998; Kallmeyer et al., 2012; McMahon and Parnell, 2014; Parkes et al., 2014; Bar-On et al., 2018). Microbial metabolism is critical to the fixation, maintenance and

turnover of carbon, nitrogen, and sulfur reservoirs in the subsurface biosphere. Recent studies have shown that a high proportion of the organisms in the subsurface have yet to be cultured, known only by culture-independent genetic evidence such as metagenomic shotgun sequencing of whole genomic DNA (Wrighton et al., 2012; Castelle et al., 2015). In addition, newly identified microbial lineages are now commonly implicated in major geochemical cycles (Rasigraf et al., 2014; Baker et al., 2016). It follows that there is substantial metabolic diversity yet to be discovered within the marine and terrestrial subsurface ecosystems. To better describe and model the microbial networks that catalyze carbon, nitrogen, and sulfur cycling in the subsurface, we need to incorporate newly identified microbial lineages along with their metabolic capabilities.

Molecular environmental surveys have provided a wealth of data elucidating microbial phylogenetic diversity, especially in previously poorly sampled habitats. Sequencing of small-subunit ribosomal RNA (SSU rRNA) genes directly from the environment has vastly expanded our knowledge of the microbial tree of life (Pace, 2009; Yarza et al., 2014; Probst et al., 2018). Advances in cultivation-independent methods for examining uncultured microbes, including single-cell genomics and deep sequencing of environmental samples (metagenomics), have begun yielding complete or near-complete genomes from many novel lineages (Iverson et al., 2012; Kantor et al., 2013; Brown et al., 2015; Castelle et al., 2015; Sekiguchi et al., 2015; Anantharaman et al., 2016; Momper et al., 2017a). These candidate lineages, previously recognized only through SSU rRNA sequencing and for which we have no cultured representatives, are providing a more complete view of the tree of life and a better understanding of global microbial ecology (Rinke et al., 2013; Hug et al., 2016; Parks et al., 2017). A few of the recently discovered and thoroughly described candidate phyla include the Korarchaeota (Elkins et al., 2008), Hadesarchaea (Baker et al., 2016) Kryptonia (Eloe-Fadrosch et al., 2016), Woesarchaeota and Pacearchaeota (Castelle et al., 2015). However, in many cases, an in-depth analysis of the metabolic potential and environmental interaction of these newly discovered lineages is lacking (Parks et al., 2017; Tully et al., 2017a,b). In order to grasp the environmental relevance of lineages that are not yet cultured, we must conduct comprehensive evaluations of their metabolic potential, taking into consideration the environments and geochemical conditions in which they survive.

To that end, we present the phylogenetic and metabolic analysis of the first two genomes of a novel bacterial lineage that was recently identified in the deep terrestrial subsurface (Momper et al., 2017a). According to 16S rRNA and phylogenomic analyses, we determined that the metagenome assembled genomes (MAGs), labeled as SURF_5 and _17, constitute the first full genomes of a novel lineage. We follow proposed standards (Konstantinidis et al., 2017) to characterize this uncultivated candidate lineage and name it 'Candidatus Abyssubacteria'. *Candidatus* applies both to the organism as well as the potential new lineage. This allows us to combine environmentally derived genome sequence taxonomic classification with currently accepted nomenclature standards, as proposed by Hedlund et al. (2015). In this study, we use full

genome sequencing and metabolic reconstruction to elucidate the probable ecological importance of this newly identified candidate lineage. We show that members of this lineage are globally distributed in terrestrial and marine subsurface environments and possess putative functional adaptations that enable them to thrive in these dark, often anoxic environments. We use pairwise average nucleotide and average amino acid identities (ANI/AAI) to assess the taxonomic rank classification, and reconstructions based on concatenated ribosomal protein sequences to understand its relatedness within the broader scope of bacterial phylogeny.

MATERIALS AND METHODS

Sample Collection, DNA Extraction and Sequencing

All fluid samples and corresponding geochemical data were collected from the former Homestake gold mine (now Sanford Underground Research Facility, SURF) near Lead, South Dakota, United States (44°21' N 103°45' W) in October, 2013. Geochemical measurements were made either on site, or upon return to the laboratory (Osburn et al., 2014). In this study, we examined deep fracture fluids from legacy boreholes drilled ~1.5 kilometers below surface (kmbs), accessing a deeply circulating terrestrial aquifer. A comprehensive description of the fluid sampling methods can be found in Osburn et al. (2014) and Momper et al. (2017b). In brief, the biomass in borehole fluids was collected on 47 mm, 0.2 μ m Supor filters (Pall Corporation, Port Washington, NY, United States). These filters were then stored on dry ice, transported to the University of Southern California and immediately frozen at -80°C . Whole genomic DNA was extracted using a modified phenol-chloroform extraction with ethanol precipitation as previously described in Osburn et al. (2014). DNA concentration was checked on a Qubit 2.0 fluorometer (Thermo Fisher Scientific), and purity was measured on a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific) before samples were sent for sequencing. Sequencing was performed at the University of Southern California's Genome and Cytometry Core Facility (Los Angeles, United States) on an Illumina HiSeq 2500 (San Diego, CA, United States) using a paired-end method, insert size of 500 base pairs (bp) and fragment size of 150 bp, as described in Momper et al. (2017a).

De novo Assembly and Read Mapping

Reads were quality trimmed and filtered using Trimmomatic version 0.36, with a minimum quality score of 40 and a minimum length of 36 base pairs (Bolger et al., 2014). Reads were then assembled using IDBA-UD 1.1.1 (Peng et al., 2012). Sequences from both fluid samples were co-assembled in order to implement differential coverage binning methods for genome bin analysis. Minimum contig length for the co-assembly was set at 10,000 bp. Coverage information was then attained by individually mapping the paired-end reads of each of the two samples to this co-assembly using Bowtie2 (Langmead and Salzberg, 2012). To convert alignments to the

SAM format, the BWA-SAMPE algorithm was used with default parameters. Coverage information was extracted using SAMtools 0.1.17 (Li et al., 2009).

Metagenome Assembled Genome (MAG) Reconstruction

Individual MAGs were reconstructed using sequence composition, differential coverage and read-pair linkage through the CONCOCT program (Alneberg et al., 2014). MAGs were then manually refined and curated using the interactive interface in the Anvi'o program (Eren et al., 2015). After refinement, genome bin completeness and contamination were re-calculated using five widely accepted marker gene suites compiled from Creevey et al. (2011), Dupont et al. (2012), Wu and Scott (2012), Campbell et al. (2013), and Alneberg et al. (2014). Reconstruction and identification of 16S rRNA gene sequences within each bin was completed using the CheckM pipeline (Parks et al., 2015). All requisite code for calculating completeness and contamination and identifying 16S rRNA sequences can be found at <https://github.com/ECogenomics/CheckM>. The 16S rRNA gene sequence recovered from SURF_17 was compared to existing isolates and environmental clone sequences using the SILVA Incremental Aligner (Pruesse et al., 2012). It was also compared to existing isolates and environmental clone sequences using the NCBI Basic Local Alignment Search Tool (BLAST), querying both the ref_seq and nr databases. Closest neighbors generated from these searches were used to generate a 16S phylogenetic species tree (**Supplementary Figure 1**). The 16S rRNA gene sequence could not be recovered from SURF_5 so it was not included in this portion of the analysis.

Phylogenomic Analyses and Phylogenetic Classification

A phylogenetic tree was constructed using a concatenation of 16 syntenic and co-located highly conserved ribosomal proteins according to Hug et al., 2016. A full list of all proteins used to build the phylogenetic tree can be found in **Supplementary Data File 1**. Ribosomal proteins were extracted from SURF_5 and SURF_17 genomes and ~18,000 publicly available environmental genomes using Prodigal and HMMER *hmmsearch* (Hyatt et al., 2010; Finn et al., 2011) and were aligned with reference proteins using the MUSCLE aligner (Edgar, 2004a,b; Graham and Tully, 2018). The concatenated alignment was trimmed using TrimAl (parameter -automated1) and used to build a comprehensive phylogenetic tree using FastTree with gamma and lg parameters (Capella-Gutiérrez et al., 2009; Price et al., 2009, 2010). This comprehensive tree containing ~18,000 genomes was then culled to include a distribution from all available bacterial phyla. All available genomes from the phylum 'Ca. Hydrogenedentes' were kept in the phylogenomic tree. The culled tree was then rebuilt using the RAxML maximum likelihood method with the GTR model of nucleotide substitution under the gamma- and invariable- models of rate heterogeneity (Stamatakis, 2006).

Pairwise average nucleotide identity (ANI) and average amino acid identity (AAI) were calculated for the SURF_5 and SURF_17 genomes against the three most closely related available genomes

from the 'Ca. Hydrogenedentes' phylum. ANI and AAI were analyzed using the publicly available tools provided through the ChunLab online Average Nucleotide Identity Calculator (Yoon et al., 2017), and the CompareM amino acid identity workflow (Parks, 2014), respectively. The ANI calculator estimates the average nucleotide identity using both best hits (one-way ANI) and reciprocal best hits (two-way ANI) between two genomic datasets (Goris et al., 2007). ANI values between genomes of the same species are >95%. ANI values <75% are not reliable (Rodríguez-R and Konstantinidis, 2014). In cases of low ANI values, AAI values were also calculated and reported. The percentage of conserved proteins (POCP) between the two SURF genomes was calculated via a pairwise BLAST of SURF_5 and SURF_17 protein sequences and calculations were performed as outlined by Qin et al. (2014).

RESULTS

MAG Statistics and Identification of a Novel Bacterial Candidate Phylum

Reassembly of the metagenomic data combined with differential-coverage based binning methods yielded near-complete recovery of two novel, distinct genomes: SURF_5 and SURF_17 (97 and 91% completeness, respectively, both with 4% contamination). According to recent published standards (Bowers et al., 2017), the two MAGs reported here are high quality genomes, with greater than 90% completeness and less than 5% contamination (**Table 1**). BLAST results of the 16S rRNA gene sequence from SURF_17 against the NCBI nr and ref_seq databases revealed 80–83% sequence identity to cultured isolates of the Deltaproteobacteria, Gammaproteobacteria and Firmicutes, with no consensus converging on any one of those classes or phyla. A species tree of the 16S rRNA gene sequence shows that SURF_17 putatively belongs to an uncultured group of bacteria related to the phylum Poribacteria (**Supplementary Figure 1**). Additional phylogenomic analysis using 16 concatenated ribosomal proteins showed that SURF_5 and _17 genomes are likely members of the Candidate phylum 'Hydrogenedentes' (**Figure 1**). The three most closely related genomes, according to ribosomal protein sequence similarity, are the sole members of a novel candidate phylum: *Candidatus Hydrogenedentes terephthalicus_JGI_OTU1*, *Candidatus Hydrogenedentes UBA2224* and *Candidatus Hydrogenedentes UBA6118* (**Figure 1**). This difference in phylogenetic placement between the single 16S rRNA gene and a concatenation of ribosomal proteins illustrates the limited reliability of reconstructed 16S sequences alone. Indeed, previous studies have noted the discrepancy between the percent identity of partial or reconstructed 16S rRNA sequences versus phylogeny based on full length 16S or multiple concatenated ribosomal sequences (Schloss, 2010; Parks et al., 2017). In this study, we deem the phylogeny based on multiple ribosomal proteins to be more robust, and will base our discussion of SURF_5 and _17 genomes on the phylogeny displayed in **Figure 1**.

TABLE 1 | General characteristics of metagenome assembled genomes analyzed in this study.

| | SURF_5 | SURF_17 |
|-----------------------------|---------|---------|
| Completeness (%) | 96 | 91 |
| Contamination (%) | 4 | 4 |
| GC (%) | 54.7 | 54.8 |
| Genome size (Mbp) | 5.1 | 4.6 |
| Number scaffolds | 145 | 144 |
| Longest scaffold (bp) | 208,377 | 111,957 |
| Number of genes | 4482 | 4105 |
| Fraction unclassified genes | 0.60 | 0.60 |
| 16S rRNA gene? (y/n) | n | y |
| Collection temperature C | 22 | 22 |
| Collection depth (mbs) | 1,500 | 1,500 |

Mbp, million base pairs

mbs, meters below surface

Further whole genome comparison at the nucleotide and amino acid sequence levels consistently indicated that SURF_5 and SURF_17 were too divergent from the three previously sequenced 'Ca. Hydrogenedentes' genomes to have meaningful ANI (**Supplementary Data File 2**). The ANI values between 'Ca. Abyssubacteria' and other similar genomes were all below 70% and hence were not reliable for direct comparison (Rodríguez-R and Konstantinidis, 2014). The AAI values between 'Ca. Abyssubacteria' and other genomes within the 'Ca. Hydrogenedentes' phylum were all in the 45–55% range (**Supplementary Data File 2**), a relatively low value that could indicate the SURF genomes represent a novel phylum or order (Luo et al., 2014), although the undersampling and lack of available genomes from the candidate phylum Hydrogenedentes makes it difficult to determine the taxonomic level represented by the SURF genomes with confidence. Genomes SURF_5 and SURF_17 were also compared to each other. The AAI value between these two genomes was ~65%, and the POCP value was 67%, indicating that these two genomes belong to the same family or genus-level taxonomic classification (Konstantinidis and Tiedje, 2005; Qin et al., 2014).

Metabolic Reconstruction

The availability of two nearly complete 'Ca. Abyssubacteria' MAGs enabled metabolic and putative functional predictions for this novel candidate lineage (**Figures 2, 3**). 'Ca. Abyssubacteria' contains genes that encode proteins for a complete tricarboxylic acid (TCA) cycle and all key metabolic enzymes for Embden–Meyerhof glycolysis and the pentose phosphate pathways. In both 'Ca. Abyssubacteria' genomes we found genes for the complete reductive acetyl-CoA carbon fixation (Wood-Ljungdahl) pathway (**Figures 2, 3** and **Supplementary Data File 3**) and multiple copies of the carbon dioxide transporter, carbonic anhydrase, were found in both genomes, possibly facilitating import of carbon dioxide gas into the cell for subsequent fixation. Genes for nitrate transport (*ntrABCD*) into the cell and subsequent chemotrophic nitrate reduction (*narIHGK*) were also identified. The SURF_5 genome contains the canonical gene for

dissimilatory nitrate reduction to ammonium (DNRA, *nrfA*) and both genomes are putatively capable of nitric oxide reduction to nitrous oxide (qNOR). Pathways for both assimilatory and dissimilatory sulfur metabolisms were identified. We also found abundant metal and polysaccharide transporters, including those for molybdate (*modABCM*), tungstate (*tupABC*), lipoproteins (*lolCDE*), and lipopolysaccharides (*rfbAB*). The genomes from SURF do not contain complete gene sets for any of the six well characterized secretion systems (**Figure 2**), but contain a mostly complete Sec-SRP system consisting of *secABDEFGY* (**Figures 2, 3**). 'Ca. Abyssubacteria' is a motile bacterium with putative chemotaxis proteins. An enlarged version of the flagellar assembly, chemotaxis cassette and complete gene annotations can be found in **Supplementary Figure 2** and **Supplementary Data File 3**.

DISCUSSION

Distribution, Habitat and Phylogeny of 'Candidatus Abyssubacteria'

A BLAST search of the NCBI nr database using the 16S rRNA gene sequence from SURF_17 revealed that members of the proposed candidate lineage, 'Ca. Abyssubacteria', are globally distributed (**Figure 4**) in marine and terrestrial, shallow and deep, subsurface environments (**Supplementary Data File 4**). Interestingly, BLAST results >90% identical to SURF_17 (the generally accepted cutoff for family level lineage) are all from deep subsurface environments, including freshwater aquifers (Flynn et al., 2013), gas hydrates, deep-sea hydrothermal sediments and deep-sea sediments from the Mariana Trough (Kato et al., 2015). Indeed, the only 16S rRNA gene sequence >98% identical to that for SURF_17, the accepted cutoff for same species lineage (Yarza et al., 2014), was collected from the world's deepest sinkhole in Zacatón, Mexico (Sahl et al., 2010). Considering the global distribution and apparent habitat restriction to subsurface environments, we propose that 'Ca. Abyssubacteria' is a lineage metabolically suited to dark and often anoxic subsurface environments.

Phylogenomic and amino acid identity analyses indicate that the two SURF genomes analyzed here are a relatively divergent lineage, likely constituting a novel class or order within the proposed candidate phylum, 'Ca. Hydrogenedentes' (**Figure 1**, **Supplementary Figure 1**, and **Supplementary Data File 4**). To date, only a handful of 'Ca. Hydrogenedentes' genomes have been sequenced and made publicly available. The first genome that corresponded to the unclassified group 'NKB19' according to 16S sequence identity was a compilation of four single cell amplified genomes (SAGs), combined into one genome, identified as *Candidatus_Hydrogenedentes_terephthalicus_JGI_OTU1* (Rinke et al., 2013) (**Figure 1**, **Supplementary Figure 1**, and **Supplementary Data File 2**). These cells were collected from an anaerobic terephthalate-degrading sludge bioreactor (Rinke et al., 2013). Very recently, two more genomes belonging to this candidate phylum were sequenced, UBA2224 and UBA6118 (Parks et al., 2017). These genomes improved upon the completeness and contamination of the combined

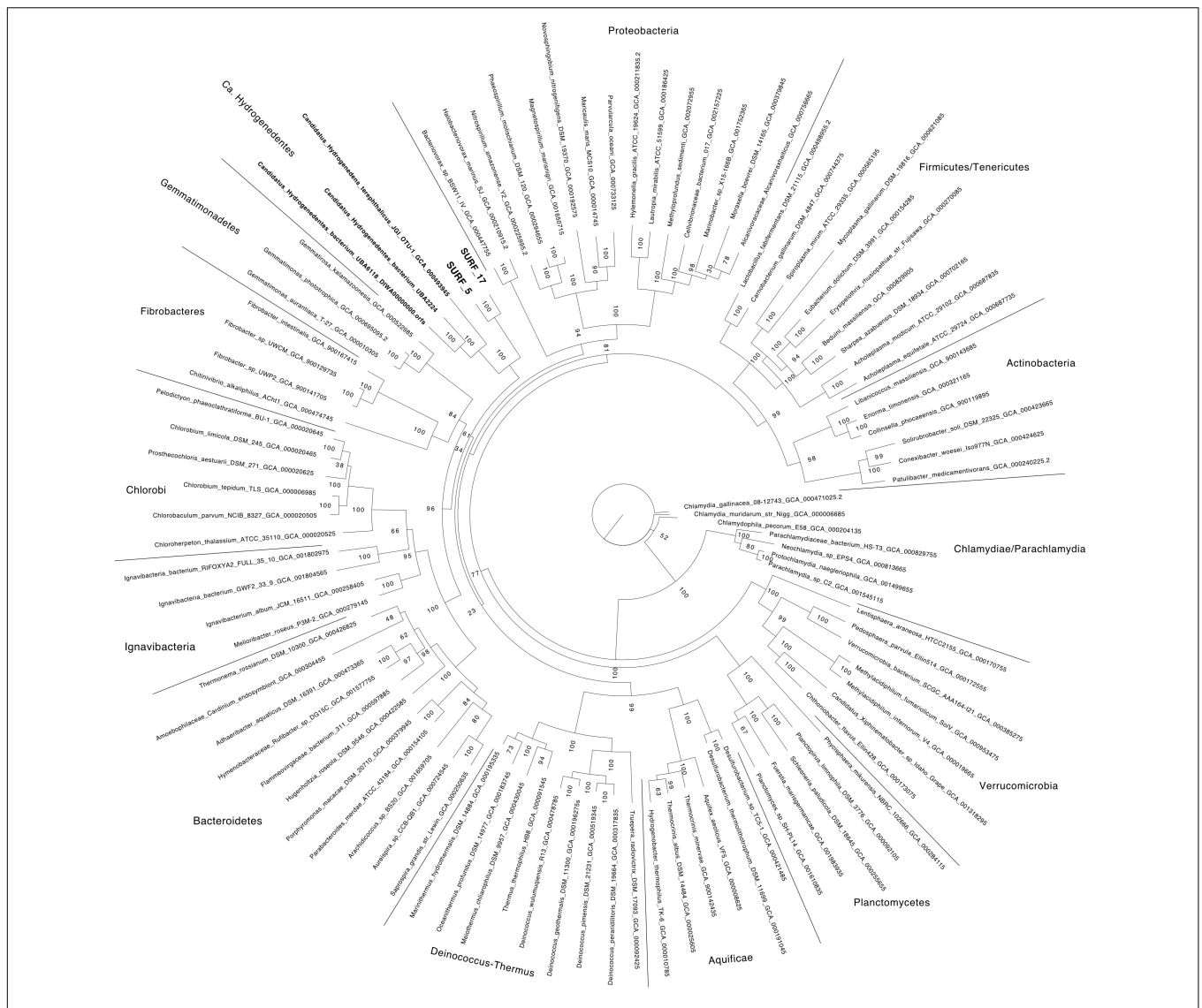


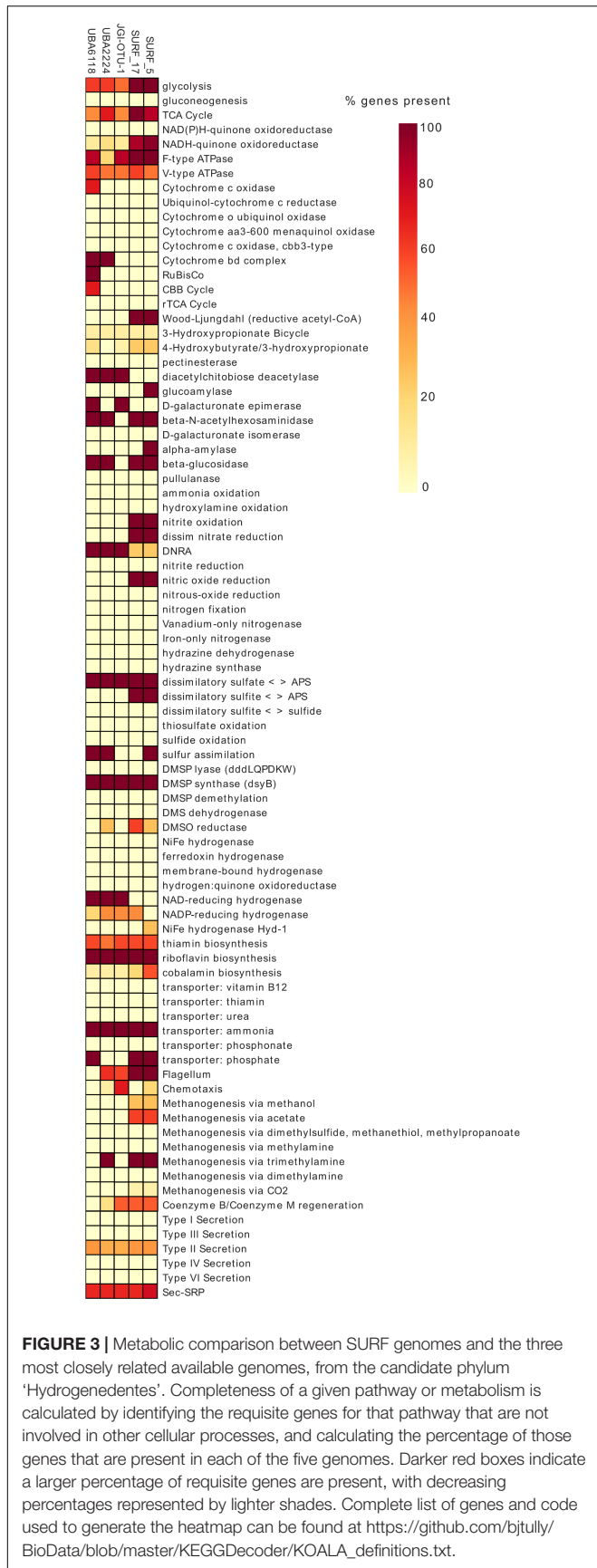
FIGURE 1 | Bacterial phylogeny estimated from the maximum likelihood phylogeny of 16 concatenated ribosomal proteins is shown. SURF_5 and SURF_17 collected from SURF fluids are indicated with bold-face font. Scale bar indicates 0.2 amino acid substitutions. Support values are reported from 1000 replicates, with a value of '100' shortened to '1' for brevity. The list of proteins used to build the tree can be found in **Supplementary Data File 1**.

SAGs (~85% complete to ~98% complete and ~4% versus ~1% contaminated, respectively), but no functional analysis was provided for these new MAGs. Here, we perform an in depth functional analysis of the SURF_5 and _17 MAGs and compare and contrast them to the previously sequenced 'Ca. Hydrogenedentes' genomes: *Candidatus Hydrogenedens terephthalicus_JGI_OTU1*, UBA2224 and UBA6118.

Putative Carbon Metabolism

Both SURF_5 and SURF_17 have all the genes necessary for the reductive acetyl-CoA pathway, the only carbon fixation pathway known to be used by both Archaea and Bacteria (Hügler and Sievert, 2011). It also requires the lowest energy input of all the six known carbon fixation pathways (Berg et al., 2010; Hügler and Sievert, 2011), making it ideal for organisms operating in

energy-deplete subsurface environments where nutrients have been highly recycled. Furthermore, the reductive acetyl-CoA pathway requires anoxic conditions, as some of its enzymes, especially the crucial acetyl-CoA synthase, are highly oxygen sensitive (Berg et al., 2010). Because of energetic efficiency and the necessity for anoxia, the reductive acetyl-CoA pathway is the ideal mode of inorganic carbon fixation in highly reducing, aphotic and energy-deplete deep subsurface fluids, including those encountered at SURF, where the oxidation-reduction potential was assessed at -235 to -276 mV (Osburn et al., 2014). Interestingly, the SURF genomes are the first members of the 'Ca. Hydrogenedentes' reported to have carbon fixing capability via the reductive acetyl-CoA pathway (Figure 3). The only other 'Ca. Hydrogenedentes' genome having any possibility of carbon fixation is UBA6118, containing the canonical gene for RuBisCo



energy metabolisms, including homoacetogenesis, as discussed previously.

We searched for energy-yielding pathways and found evidence for a complex oxidative phosphorylation pathway in both MAGs, which suggests aerobic respiration and the transfer of electrons to molecular oxygen (Figures 3, 4 and Supplementary Data File 3). However, oxygen levels were below detection in subsurface fluids at SURF (Osburn et al., 2014). We searched for further evidence of oxygen respirations, such as cytochrome c oxidases, but could not identify any in either MAG. We think it is more likely that 'Ca. Abyssubacteria' at SURF are using anaerobic metabolisms such as nitrogen reduction, sulfate reduction or sulfite oxidation as their predominant energy metabolisms *in situ*, discussed in more detail below.

Putative Energy Metabolisms

In denitrification, energy is conserved as nitrate (NO₃⁻), nitrite (NO₂⁻), nitric oxide (NO) and nitrous oxide (N₂O) are sequentially reduced to dinitrogen gas (N₂), each step catalyzed by one or more metalloenzymes (Mahne and Tiedje, 1995; Zumft, 1997). 'Ca. Abyssubacteria' genomes contained genes (*nar*) encoding enzymes that catalyze the first step of this pathway (Figure 2). The first step, the reduction of nitrate to nitrite, is enabled by the transport of NO₃⁻ into the cell by an ATP-binding cassette (ABC)-type NRT in 'Ca. Abyssubacteria', which putatively provides NO₃⁻ for reduction by *Nar* enzymes (Figure 2). The NO₂⁻ produced by this reaction is also a highly reactive possible electron acceptor. In the three publicly available 'Ca. Hydrogenedentes' genomes (Candidatus_Hydrogenedentes_tereophthalicus_JGI_OTU1, UBA2224 and UBA6118) it appears that nitrite could be reduced to ammonium via DNRA. The requisite genes for this metabolism are *nrfABEFG*, which are present in Hydrogenedentes_tereophthalicus_JGI_OTU1, UBA2224 and UBA6118 (Figure 3). However, we were only able to identify the gene for the canonical cytochrome c, *nrfA*, in SURF_5 but not in SURF_17 (Figures 2, 3). We hypothesize that the other *nrf* genes have been lost in SURF genomes, and instead of DNRA they are likely performing other nitrogen, sulfur or carbon based metabolisms that are more exergonic in the subsurface fluids from which they were collected. Dissimilatory NO reduction to N₂O can be catalyzed by two classes of the *NorB* enzyme: a quinol-oxidizing single subunit (*qNorB*) or a cytochrome *bc*-type multiplex (*cNorB*) (Zumft, 1997). SURF_5 and SURF_17 possess the quinol-oxidizing nitric oxide reductase, bound to the inner membrane of the cell (Figure 2). However, because NO is a highly toxic molecule, interfering with cellular processes, we cannot rule out the possibility that the function of the nitric oxide reductase, *qNor*, is for detoxification rather than an energy yielding process. Considering the measured nitrate levels in SURF fluids (10–25 μM, Osburn et al., 2014), it appears likely that 'Ca. Abyssubacteria' could be performing the first step of denitrification in the subsurface at SURF, reduction of nitrate to nitrite. Given the presence of methyltransferase genes in SURF_5 and SURF_17 genomes, it is possible that these bacteria are coupling the anaerobic oxidation of methylated

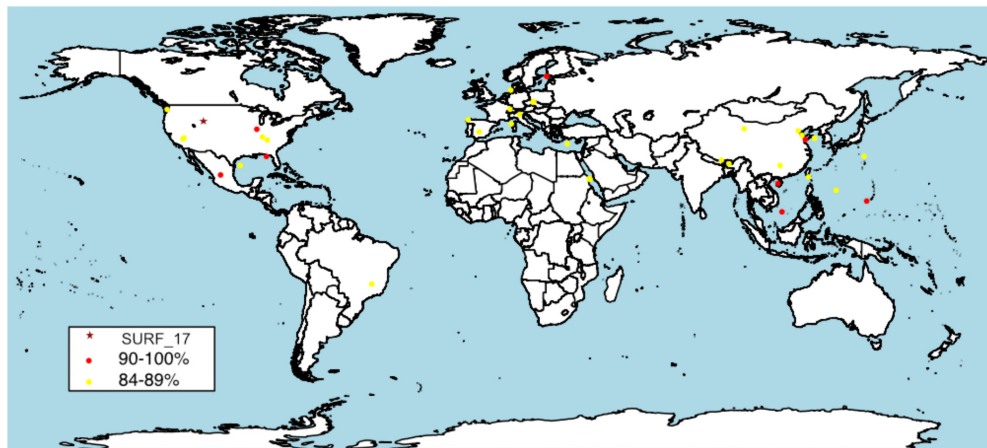


FIGURE 4 | Global distribution of SURF_5 and SURF_17 relatives as measured by 16S rRNA clone nucleotide identity. The reconstructed SURF_17 16S gene was used as query sequence in the NCBI nr database. Top 50 results were included, results range from 98 to 84% identity. A complete list of 16S gene results, including accession numbers, query coverage, identity and collection source can be found in **Supplementary Data File 4**.

compounds, such as methanol or methylamine, using nitrate as an electron acceptor. Although this metabolic strategy is not widespread, anaerobic methylotrophy coupled to denitrification is exhibited in several marine and wastewater-associated species, including *Methylophaga nitratreducentescens*, *Hyphomicrobium denitrificans*, and *Methyloversatilis* spp. (Urakami et al., 1995; Baytshok et al., 2008; Mauffrey et al., 2017).

Another possibility for energy yielding chemotrophic metabolism in 'Ca. Abyssubacteria' is the dissimilatory oxidation or reduction of sulfur compounds. 'Ca. Abyssubacteria' genomes contain putative dissimilatory capabilities, with genes for *sat* and *aprAB* possibly oxidizing sulfite to APS and then to sulfate, respectively (Figure 2). This pathway can also be run in reverse, catalyzing the stepwise reduction of sulfate to sulfite and then sulfide. However, a genome-wide search revealed only the presence of assimilatory sulfite reductases. Neither the dissimilatory sulfite reductase genes (*dsrAB*), nor the requisite companion gene, *dsrD*, was identified. Therefore, the SURF genomes appear capable of only the first step in sulfate reduction, that of SO_4^{2-} reduction to SO_3^- via *sat* and *aprAB*, and not further reduction to sulfide via *dsrABD*.

Although both genomes contained the *SoxD* gene, 'Ca. Abyssubacteria' likely cannot oxidize sulfide, as other *Sox* genes necessary to complete the pathway were not present. The presence of heterodisulfide reductases and other reductive enzymes suggests that 'Ca. Abyssubacteria' could be using H_2 as an electron donor when available *in situ*. However, given that sulfate reduction with H_2 is only moderately exergonic in SURF fluids, and that sulfate reduction fell near anaerobic heterotrophy in a principal component analysis of thermodynamic favorability using *in situ* fluid geochemistry (Osburn et al., 2014), it is also possible that 'Ca. Abyssubacteria' utilizes organic matter as an electron donor. Given the presence of genes that encode enzymes for hydrogenases

and methylotrophy, electron donor utilization appears to be versatile in SURF genomes, is most likely varied depending upon the electron acceptor concentration, and on subsurface geochemical conditions *in situ*. The retention of genes encoding multiple putative methods of carbon, nitrogen and sulfur metabolisms may be a way for 'Ca. Abyssubacteria' to cope with variable geochemistry and intermittent energetically unfavorable conditions.

The candidate phylum 'Hydrogenedentes' was named for the abundance of hydrogenases and putative H_2 -utilizing pathways in the four SAGs that were the first partial genomes representing the phylum (Rinke et al., 2013). Later, metagenomic and metatranscriptomic analyses of the microbial community in a methanogenic bioreactor identified abundant read mapping to putative 'Ca. Hydrogenedentes' and identified them as lipolytic glycerol degraders (Nobu et al., 2015). Indeed, to date, all of the available genomes belonging to the 'Ca. Hydrogenedentes' have been collected from engineered environments such as anaerobic sludge bioreactors (Rinke et al., 2013; Nobu et al., 2015). In terms of 16S rRNA sequences, after the collection of the original 'NKB19' from the depths of the Mariana trench, subsequent published 16S sequences identifying with this group have been collected from phthalate-degrading bioreactors (Rinke et al., 2013) and methanogenic sludge bioreactors (Rivière et al., 2009; Narihiro et al., 2015; Nobu et al., 2015). In contrast, the two genomes analyzed here were collected from deep continental subsurface fluids, in a relatively natural and undisturbed environment, compared to anaerobic sludge bioreactors. Their closest relatives, according to the vast repository of 16S rRNA clone sequences, reside in similar continental and marine subsurface environments (Figure 4 and Supplementary Data File 4). This discrepancy in habitat is reflected in the putative metabolisms encoded by available genomes (Figure 3). The three genomes that comprise the 'Ca. Hydrogenedentes' phylum are putative organic carbon degraders, potentially hydrolyzing carbon compounds such as phthalates, lipids and glycerols.

In contrast, the genomes from SURF contain all genes necessary for autotrophic carbon fixation via the reductive acetyl-CoA pathway (Figures 2, 3). Although genome reconstruction was not possible, Nobu et al. were able to construct a pangenome and metatranscriptome putatively belonging to *Ca. Hydrogenedentes* and found that they were likely metabolizing lipids, hydrolyzing triacylglycerols to glycerol and long chain fatty acids, and/or syntrophically oxidizing glycerol to carbon dioxide and acetate (2015). None of these appear to be the most likely metabolisms for the genomes collected from SURF. In our metabolic analysis of all five currently available genomes in the 'Ca. Hydrogenedentes' phylum, the only energy metabolisms they may have in common are homoacetogenesis and sulfate reduction to sulfite (Figure 3).

Concluding Remarks and Description of 'Ca. Abyssubacteria'

Based on our phylogenomic analysis using a concatenated alignment of single-copy marker genes (Figure 1), phylogenetic analysis of 16S rRNA gene sequences (Supplementary Figure 1 and Supplementary Data File 4), and AAI and POCP analyses (Supplementary Data File 2), the two MAGs identified in this study are proposed as a novel class or order-level lineage that most likely falls within the candidate phylum "Hydrogenedentes." We designate this new candidate lineage 'Candidatus Abyssubacteria', from the Latin prefix meaning deep, owing to their collection 1.5 km below surface. Cells are motile, and have versatile carbon and energy-yielding chemotrophic metabolic potential, enabling them to survive in dark, energy deplete subsurface environments under varying levels of oxygen, nitrogen, carbon, and sulfur.

DATA DEPOSITION

Sequence data for metagenomic reads from SURF, South Dakota, contigs and genes were submitted to the JGI-IMG under accession number IMG 3300007351. Sample metadata for those sequences can be accessed using the BioProject identifier PRJNA355136. Genomes comprising the proposed candidate

phylum Abyssubacteria, SURF_5 and SURF_17, can be found on the publicly accessible database, NCBI, under accession numbers SAMN08498999 and SAMN08499011, respectively.

AUTHOR CONTRIBUTIONS

LM analyzed the metagenomic data, performed the phylogenetic analyses, and reconstructed the metabolism of genomes analyzed in the study and wrote the manuscript. HA analyzed the global distribution, performed the phylogenomic analyses, POCP calculations and contributed to results and discussion. JA funded sample retrieval and provided guidance during analysis and writing of the manuscript.

FUNDING

This work, including sample collection from SURF, was funded by the NASA Astrobiology Institute under cooperative agreement NNA13AA92A and by a W. O. Crosby Postdoctoral Fellowship at the Massachusetts Institute of Technology to LM. This is an NAI *Life Underground* publication, contribution number 132.

ACKNOWLEDGMENTS

We thank the staff at the Sanford Underground Research Facility (SURF) for their support, assistance, and coordination during sample collection. Many thanks to Dr. Laura Zinke for insightful discussion of microbial metabolisms. We also thank Elaina Graham and Dr. Benjamin Tully at the University of Southern California for their invaluable assistance with bioinformatic analyses.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2018.01993/full#supplementary-material>

REFERENCES

- Alneberg, J., Bjarnason, B. S., de Bruijn I, Schirmer, M., Quick, J., Ijaz, U. Z., et al. (2014). Binning metagenomic contigs by coverage and composition. *Nat. Methods* 11, 1144–1146. doi: 10.1038/nmeth.3103
- Anantharaman, K., Brown, C. T., Hug, L. A., Sharon, I., Castelle, C. J., Probst, A. J., et al. (2016). Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat. Commun.* 7:13219. doi: 10.1038/ncomms13219
- Baker, B. J., Saw, J. H., Lind, A. E., Lazar, C. S., Hinrichs, K. U., Teske, A. P., et al. (2016). Genomic inference of the metabolism of cosmopolitan subsurface Archaea, Hadesarchaea. *Nat. Microbiol.* 1:16032. doi: 10.1038/nmicrobiol.2016.2
- Bar-On, Y. M., Phillips, R., Milo, R., and Falkowski, P. G. (2018). "The biomass distribution on earth," in *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 115, (New Jersey, NJ), 6506–6511. doi: 10.1073/pnas.1711842115
- Baytshtok, V., Kim, S., Yu, R., Park, H., and Chandran, K. (2008). Molecular and biokinetic characterization of methylotrophic denitrification using nitrate and nitrite as terminal electron acceptors. *Water Sci. Technol.* 58, 359–365. doi: 10.2166/wst.2008.391
- Berg, I. A., Kockelkorn, D., Ramos-Vera, W. H., Say, R. F., Zarzycki, J., Hügler, M., et al. (2010). Autotrophic carbon fixation in archaea. *Nat. Rev. Microbiol.* 8, 447–460. doi: 10.1038/nrmicro2365
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Bowers, R. M., Kyrpides, N. C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T. B. K., et al. (2017). Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat. Biotechnol.* 35, 725–731. doi: 10.1038/nbt.3893
- Brown, C., Hug, L., Thomas, B., Sharon, I., and Nature, C. C. - (2015). *Undefined Unusual Biology Across a Group Comprising More than 15% of*

- Domain Bacteria. *Nature.Com*. Available at: <https://www.nature.com/articles/nature14486> [Accessed February 11, 2018].
- Campbell, J. H., O'Donoghue, P., Campbell, A. G., Schwientek, P., Sczyrba, A., Woyke, T., et al. (2013). UGA is an additional glycine codon in uncultured SR1 bacteria from the human microbiota. *Proc. Natl. Acad. Sci. U.S.A.* 110, 5540–5545. doi: 10.1073/pnas.1303090110
- Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348
- Castelle, C. J., Wrighton, K. C., Williams, K. H., Banfield, J. F., Thomas, B. C., Hug, L. A., et al. (2015). Article genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. *Curr. Biol.* 25, 690–701. doi: 10.1016/j.cub.2015.01.014
- Creevey, C. J., Doerks, T., Fitzpatrick, D. A., Raes, J., and Bork, P. (2011). Universally distributed single-copy genes indicate a constant rate of horizontal transfer. *PLoS One* 6:e22099. doi: 10.1371/journal.pone.0022099
- Dupont, C. L., Rusch, D. B., Yooshep, S., Lombardo, M. J., Alexander Richter, R., Valas, R., et al. (2012). Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. *ISME J.* 6, 1186–1199. doi: 10.1038/ismej.2011.189
- Edgar, R. C. (2004a). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113. doi: 10.1186/1471-2105-5-113
- Edgar, R. C. (2004b). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- Elkins, J. G., Podar, M., Graham, D. E., Makarova, K. S., Wolf, Y., Randau, L., et al. (2008). A korarchaeal genome reveals insights into the evolution of the Archaea. *Proc. Natl. Acad. Sci. U.S.A.* 105, 8102–8107. doi: 10.1073/pnas.0801980105
- Eloe-Fadros, E. A., Paez-Espino, D., Jarett, J., Dunfield, P. F., Hedlund, B. P., Dekas, A. E., et al. (2016). Global metagenomic survey reveals a new bacterial candidate phylum in geothermal springs. *Nat. Commun.* 7, 1–10. doi: 10.1038/ncomms10476
- Eren, A. M., Esen, Ö.C., Quince, C., Vineis, J. H., Morrison, H. G., Sogin, M. L., et al. (2015). Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* 3:e1319. doi: 10.7717/peerj.1319
- Finn, R. D., Clements, J., and Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 39, W29–W37. doi: 10.1093/nar/gkr367
- Flynn, T. M., Sanford, R. A., Ryu, H., Bethke, C. M., Levine, A. D., Ashbolt, N. J., et al. (2013). Functional microbial diversity explains groundwater chemistry in a pristine aquifer. *BMC Microbiol.* 13:146. doi: 10.1186/1471-2180-13-146
- Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., and Tiedje, J. M. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* 57, 81–91. doi: 10.1099/ijs.0.64483-0
- Graham, E. D., and Tully, B. J. (2018). *Building a Phylogenetic Tree*. Los Angeles, CA: University of Southern California, doi: 10.17504/protocols.io.mp5c5q6
- Hedlund, B. P., Murugapiran, S. K., Alba, T. W., Levy, A., Dodsworth, J. A., Goertz, G. B., et al. (2015). Uncultivated thermophiles: current status and spotlight on "Aigarchaeota." *Curr. Opin. Microbiol.* 25, 136–145. doi: 10.1016/j.mib.2015.06.008
- Hug, L. A., Baker, B. J., Anantharaman, K., Brown, C. T., Probst, A. J., Castelle, C. J., et al. (2016). A new view of the tree of life. *Nat. Microbiol.* 1:16048. doi: 10.1038/nmicrobiol.2016.48
- Hügler, M., and Sievert, S. M. (2011). Beyond the calvin cycle: autotrophic carbon fixation in the ocean. *Ann. Rev. Mar. Sci.* 3, 261–289. doi: 10.1146/annurev-marine-120709-142712
- Hyatt, D., Chen, G. L., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119
- Iverson, V., Morris, R. M., Frazar, C. D., Berthiaume, C. T., Morales, R. L., and Armbrust, E. V. (2012). Untangling genomes from metagenomes: revealing an uncultured class of marine euryarchaeota. *Science* 335, 587–590. doi: 10.1126/science.1212665
- Kallmeyer, J., Pockalny, R., Adhikari, R. R., Smith, D. C., and D'Hondt, S. (2012). Global distribution of microbial abundance and biomass in seafloor sediment. *Proc. Natl. Acad. Sci. U.S.A.* 109, 16213–16216. doi: 10.1073/pnas.1203849109
- Kantor, R. S., Wrighton, K. C., Handley, K. M., Sharon, I., Hug, L. A., Castelle, C. J., et al. (2013). Small genomes and sparse metabolisms of sediment-associated bacteria from four candidate phyla. *mBio* 4:e708–e713. doi: 10.1128/mBio.00708-13
- Kato, S., Ikehata, K., Shibuya, T., Urabe, T., Ohkuma, M., and Yamagishi, A. (2015). Potential for biogeochemical cycling of sulfur, iron and carbon within massive sulfide deposits below the seafloor. *Environ. Microbiol.* 17, 1817–1835. doi: 10.1111/1462-2920.12648
- Konstantinidis, K. T., Rosselló-Móra, R., and Amann, R. (2017). Uncultivated microbes in need of their own taxonomy. *ISME J.* 11, 2399–2406. doi: 10.1038/ismej.2017.113
- Konstantinidis, K. T., and Tiedje, J. M. (2005). Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci. U.S.A.* 102, 2567–2572. doi: 10.1073/pnas.0409727102
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Li, H., Handsaker, B., and Wysoker, A. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Luo, C., Rodriguez-R, L. M., and Konstantinidis, K. T. (2014). MyTaxa: an advanced taxonomic classifier for genomic and metagenomic sequences. *Nucleic Acids Res.* 42:e73. doi: 10.1093/nar/gku169
- Mahne, I., and Tiedje, J. M. (1995). Criteria and Methodology for Identifying Respiratory Denitrifiers. *Appl. Environ. Microbiol.* 61, 1110–1115.
- Mauffrey, F., Cucaita, A., Constant, P., and Villemur, R. (2017). Denitrifying metabolism of the methylotrophic marine bacterium *Methylophaga nitratireducentis* strain JAM1. *PeerJ* 5:e4098. doi: 10.7717/peerj.4098
- McMahon, S., and Parnell, J. (2014). Weighing the deep continental biosphere. *FEMS Microbiol. Ecol.* 87, 113–120. doi: 10.1111/1574-6941.12196
- 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. doi: 10.1038/ismej.2017.94
- Momper, L., Kiel Reese, B., Zinke, L., Wanger, G., Osburn, M. R., Moser, D., et al. (2017b). Major phylum-level differences between porefluid and host rock bacterial communities in the terrestrial deep subsurface. *Environ. Microbiol. Rep.* 9, 501–511. doi: 10.1111/1758-2229.12563
- Müller, V. (2003). Energy Conservation in Acetogenic. *Appl. Environ. Microbiol.* 69, 6345–6353. doi: 10.1128/AEM.69.11.6345
- Narihiro, T., Nobu, M. K., Kim, N. K., Kamagata, Y., and Liu, W. T. (2015). The nexus of syntrophy-associated microbiota in anaerobic digestion revealed by long-term enrichment and community survey. *Environ. Microbiol.* 17, 1707–1720. doi: 10.1111/1462-2920.12616
- Nobu, M. K., Narihiro, T., Rinke, C., Kamagata, Y., Tringe, S. G., Woyke, T., et al. (2015). Microbial dark matter ecogenomics reveals complex synergistic networks in a methanogenic bioreactor. *ISME J.* 9, 1710–1722. doi: 10.1038/ismej.2014.256
- Osburn, M. R., LaRowe, D. E., Momper, L. M., and Amend, J. P. (2014). Chemolithotrophy in the continental deep subsurface: sanford underground research facility (SURF). USA. *Front. Microbiol.* 5:610. doi: 10.3389/fmicb.2014.00610
- Pace, N. R. (2009). Mapping the tree of life: progress and prospects. *Microbiol. Mol. Biol. Rev.* 73, 565–576. doi: 10.1128/MMBR.00033-09
- Parkes, R. J., Cragg, B., Roussel, E., Webster, G., Weightman, A., and Sass, H. (2014). A review of prokaryotic populations and processes in sub-seafloor sediments, including biosphere: geosphere interactions. *Mar. Geol.* 352, 409–425. doi: 10.1016/j.margeo.2014.02.009
- Parks, D. H. (2014). *CompareM*. Available at: <https://github.com/dparks1134/CompareM>
- Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055. doi: 10.1101/gr.186072.114

- Parks, D. H., Rinke, C., Chuvochina, M., Chaumeil, P., Woodcroft, B. J., Evans, P. N., et al. (2017). Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat. Microbiol.* 2, 1533–1542. doi: 10.1038/s41564-017-0012-7
- Peng, Y., Leung, H. C., Yiu, S. M., and Chin, F. Y. (2012). IDBA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28, 1420–1428. doi: 10.1093/bioinformatics/bts174
- Price, M., Dehal, P., and Arkin, A. (2010). FastTree 2-Approximately maximum likelihood trees for large alignments. *PLoS One* 5:28. doi: 10.1371/journal.pone.0009490
- Price, M. N., Dehal, P. S., and Arkin, A. P. (2009). Fasttree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* 26, 1641–1650. doi: 10.1093/molbev/msp077
- Probst, A. J., Ladd, B., Jarett, J. K., Geller-Mcgrath, D. E., Sieber, C. M. K., Emerson, J. B., et al. (2018). Differential depth distribution of microbial function and putative symbionts through sediment-hosted aquifers in the deep terrestrial subsurface. *Nat. Microbiol.* 3, 328–336. doi: 10.1038/s41564-017-0098-y
- Pruesse, E., Peplies, J., and Glöckner, F. O. (2012). SINA: accurate high throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* 28, 1823–1829. doi: 10.1093/bioinformatics/bts252
- Qin, Q. L., Xie, B. B., Zhang, X. Y., Chen, X. L., Zhou, B. C., Zhou, J., et al. (2014). A proposed genus boundary for the prokaryotes based on genomic insights. *J. Bacteriol.* 196, 2210–2215. doi: 10.1128/JB.01688-14
- Rabus, R., Hansen, T. A., and Widdel, F. (2006). "Dissimilatory sulfate- and sulfur-reducing prokaryotes," in *The Prokaryotes: Prokaryotic Physiology and Biochemistry*, eds E. F. DeLong, E. Stackebrandt, S. Lory, and F. Thompson (New York, NY: Springer-Verlag Berlin Heidelberg), 309–404.
- Ragsdale, S. W., and Pierce, E. (2008). Acetogenesis and the Wood-Ljungdahl pathway of CO₂ fixation. *Biochim. Biophys. Acta Proteins Proteomics* 1784, 1873–1898. doi: 10.1016/j.bbapap.2008.08.012
- Rasigraf, O., Kool, D. M., Jetten, M. S. M., Sinninghe Damsté, J. S., and Ettwig, K. F. (2014). Autotrophic carbon dioxide fixation via the Calvin-Benson-Bassham cycle by the denitrifying methanotroph "*Candidatus* Methyloirabilis oxyfera". *Appl. Environ. Microbiol.* 80, 2451–2460. doi: 10.1128/AEM.04199-13
- Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N. N., Anderson, I. J., Cheng, J. F., et al. (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 499, 431–437. doi: 10.1038/nature12352
- Rivière, D., Desvignes, V., Pelletier, E., Chaussonnerie, S., Guermazi, S., Weissenbach, J., et al. (2009). Towards the definition of a core of microorganisms involved in anaerobic digestion of sludge. *ISME J.* 3, 700–714. doi: 10.1038/ismej.2009.2
- Rodriguez-R, L. M., and Konstantinidis, K. T. (2014). Estimating coverage in metagenomic data sets and why it matters. *ISME J.* 8, 2349–2351. doi: 10.1038/ismej.2014.76
- Sahl, J. W., Fairfield, N., Harris, J. K., Wettergreen, D., Stone, W. C., and Spear, J. R. (2010). Novel Microbial Diversity Retrieved by Autonomous Robotic Exploration of the World's Deepest Vertical Phreatic Sinkhole. *Astrobiology* 10, 201–213. doi: 10.1089/ast.2009.0378
- Schloss, P. D. (2010). The effects of alignment quality, distance calculation method, sequence filtering, and region on the analysis of 16S rRNA gene-based studies. *PLoS Comput. Biol.* 6:e1000844. doi: 10.1371/journal.pcbi.1000844
- Sekiguchi, Y., Ohashi, A., Parks, D. H., Yamauchi, T., Tyson, G. W., and Hugenholtz, P. (2015). First genomic insights into members of a candidate bacterial phylum responsible for wastewater bulking. *PeerJ* 3:e740. doi: 10.7717/peerj.740
- Stamatakis, A. (2006). RAxML-VI-HPC: maximum-likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22, 2688–2690. doi: 10.1093/bioinformatics/btl446
- Tully, B. J., Graham, E. D., and Heidelberg, J. F. (2017a). Data descriptor: the reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. *Sci. Data* 5:170203. doi: 10.1038/sdata.2017.203
- Tully, B. J., Sachdeva, R., Graham, E. D., and Heidelberg, J. F. (2017b). 290 metagenome-assembled genomes from the Mediterranean Sea: a resource for marine microbiology. *PeerJ* 5:e3558. doi: 10.7717/peerj.3558
- Urakami, T., Sasaki, J., Suzuki, K. I., and Komagata, K. (1995). Characterization and Description of *Hyphomicrobium denitrificans* sp. nov. *Int. J. Syst. Evol. Microbiol.* 45, 528–532. doi: 10.1099/00207713-45-3-528
- Whitman, W. B., Coleman, D. C., and Wiebe, W. J. (1998). Prokaryotes: the unseen majority. *Proc. Natl. Acad. Sci. U.S.A.* 95, 6578–6583. doi: 10.1073/pnas.95.12.6578
- Wrighton, K. C., Thomas, B. C., Sharon, I., Miller, C. S., Castelle, C. J., VerBerkmoes, N. C., et al. (2012). Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. *Science* 337, 1661–1665. doi: 10.1126/science.1224041
- Wu, M., and Scott, A. J. (2012). Phylogenomic Analysis of Bacterial and Archaeal Sequences with AMPHORA2. *Bioinformatics* 28, 1033–1034. doi: 10.1093/bioinformatics/bts079
- Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F. O., Ludwig, W., Schleifer, K. H., et al. (2014). Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat. Rev. Microbiol.* 12, 635–645. doi: 10.1038/nrmicro3330
- Yoon, S.-H., Ha, S., Lim, J., Kwon, S., and Chun, J. (2017). A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie Van Leeuwenhoek* 110, 1281–1286. doi: 10.1007/s10482-017-0844-4
- Zumft, W. G. (1997). Cell biology and molecular basis of denitrification. *Microbiol. Mol. Biol. Rev.* 61, 533–616.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Momper, Aronson and Amend. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.