Appl. Environ. Microbiol. doi:10.1128/AEM.00624-16

Comparative Single-Cell Genomics of *Chloroflexi* from the Okinawa Trough Deep Subsurface Biosphere

Heather Fullerton, Craig L. Moyer*

Department of Biology, Western Washington University, Bellingham WA 98225

*Corresponding author
E-mail address: cmoyer@hydro.biol.wwu.edu

Running Title:
Chloroflexi from a hydrothermal subsurface biosphere

Keywords:
Chloroflexi, Single-Cell Genomics, Hydrothermal Vents, Subsurface Biosphere, Reductive Dehalogenation

ABSTRACT

Chloroflexi SSU rRNA gene sequences are frequently recovered from subseafloor environments, but the metabolic potential of this phylum is poorly understood. The phylum *Chloroflexi* is represented by isolates with diverse metabolic strategies including, anoxic phototrophy, fermentation and reductive dehalogenation; therefore, function cannot be attributed to these organisms based solely on phylogeny. Single cell genomics can provide metabolic insights into uncultured organisms, like the deep-subsurface *Chloroflexi*. Nine SSU rRNA gene sequences were identified from single-cell sorts of whole-round core material collected as part of IODP Expedition 331 (Deep Hot Biosphere) from the Okinawa Trough at Iheya North hydrothermal field. Previous studies of subsurface *Chloroflexi* single amplified genomes (SAGs) suggest heterotrophic or lithotrophic metabolisms and provide no evidence for growth by reductive dehalogenation. Our nine *Chloroflexi* SAGs (of which seven are *Anaerolineales*) indicate that in addition to encoding genes for the Wood-Ljungdahl pathway, exogenous carbon sources can be actively transported into the cells. At least one subunit for the pyruvate ferredoxin oxidoreductase was found in four of the *Chloroflexi* SAGs. This protein can provide a link between the Wood-Ljungdahl pathway and other carbon anabolic pathways. Finally, one of the seven *Anaerolineales* SAGs contains a distinct reductive dehalogenase homologous (*rdhA*) gene; suggesting reductive dehalogenation is not limited to the *Dehalococcoidia* class of *Chloroflexi*.

IMPORTANCE

Through the use of single amplified genomes (SAGs) we have extended the metabolic potential of an understudied group of subsurface microbes, the *Chloroflexi*. These microbes are frequently detected in the subsurface biosphere though their metabolic capabilities have remained elusive. In contrast to previously examined *Chloroflexi* SAGs, our genomes (several are of the order *Anaerolineales*) were recovered from a hydrothermally driven system and therefore provide a unique window into the metabolic potential of this type of habitat. In addition, a reductive dehalogenase gene (*rdhA*) has been directly linked to marine subsurface *Chloroflexi* suggesting reductive dehalogenation is not limited to the class *Dehalococcoidia*. This discovery expands the nutrient cycling and metabolic potential present within the deep subsurface and provides functional gene information relating to this enigmatic group.

INTRODUCTION

The phylum *Chloroflexi*, colloquially known as the green non-sulfur bacteria, has been recognized as both ubiquitous and diverse by small subunit (SSU) rRNA from molecular studies (1). Environmental molecular microbial surveys have shown *Chloroflexi* to be abundant in marine, intertidal and freshwater surface and subsurface sediments (2); however, this phylum has only a few cultured representatives (3). The cultured representatives of this group have a wide range of metabolic activities, including fermentation, anoxygenic photosynthesis, nitrite-oxidation and reductive dehalogenation (3-7). Due to the metabolic diversity across this phylum, metabolic activity cannot be inferred based solely on phylogeny (Select cultured representatives listed in Table S1, along with their metabolic classifications).

In comparisons among *Chloroflexi* containing sites that were rich in methane hydrates and organic carbon off the Peru and Cascade Margins, *Chloroflexi* SSU rRNA gene sequences were more numerous at the organic rich sites (8). A metagenomic study of the Peru Margin showed *Chloroflexi* to be present at a broad range of depths and compose 12-16% of total genes identified (9). In addition to the Peru Margin, *Caldilineae*- and *Anaerolineae*-related sequences comprised a major portion of the subsurface bacterial community within a forearc basin off Sumatra and the Black Sea (10). Unlike their photosynthesizing or terrestrial relatives, none of these deep-sea strains are yet in culture, and so virtually nothing is known about their metabolic potential (11). *Chloroflexi* have been previously identified in a subsurface hydrothermal habitat at Iheya North, Mid-Okinawa Trough (12). It is likely that similar subsurface *Chloroflexi* have the potential for growth since *Chloroflexi* sequences were recovered from an *in situ*

colonization experiment within a borehole, indicating recent growth on previously uninhabited surfaces (13).

With the advances in single cell genomics it is no longer required to have axenic bacterial cultures to obtain an individual genome and in some cases, recovered single amplified genomes (SAGs) have been reported to be more than 90% complete (14, 15). The SAG approach can reveal the metabolic function of the abundant and frequently detected *Chloroflexi* for comparison to the sequenced photosynthetic or terrestrial isolates. Previously studied *Chloroflexi* SAGs belong to the class *Dehalococcoidia* and contain genes for autotrophic growth via the reductive acetyl-CoA pathway (16, 17). *Dehalococcoidia* SAGs from the Peru Trench, Dsc1 and DscP2, were recovered from 7.3 meters below seafloor (mbsf) at a seawater depth of 5086 m from IODP Site 1230 and a single SAG from a depth of 16.3 m in Aarhus Bay, SCGC AB-539-J10 (abbreviated to DEH-J10) was recovered at just 0.16 mbsf. Therefore, Dsc1 and DscP2 are the only deep subsurface *Chloroflexi* that have been analyzed for their metabolic potential prior to this study. Previously, *rdhA* genes have been identified from isolates within the order *Dehalococcoidales* of the phylum *Chloroflexi* (18) or from dehalogenating enrichment cultures (19). Furthermore, isolated strains of *Dehalococcoides mccartyi* depend on dehalogenation as their sole means of respiration via the RdhA proteins (18). No *rdhA* genes were identified in DscP2, Dsc1 or DEH-J10 and it was unclear if these organisms had heterotrophic or autotrophic metabolisms (16, 17). In marine subsurface sediments *rdhA* homologues have been detected by PCR and metagenomic studies from around the Pacific Ocean, as well as dehalogenation activity has been shown in sediment

78 microcosms, suggesting organohalide respiration is an important energy-yielding
79 pathway in subseafloor microbial ecosystems (20, 21).

80 Hydrothermal vents circulate reduced chemicals and seawater providing niches
81 for microbes that can support a larger biomass than the surrounding ocean floor (22)
82 making vent fields some of the most biologically active regions in the deep-sea (23).
83 These systems harbor diverse microbial communities in and around vent orifices, and
84 within surrounding subsurface environment. The Okinawa Trough contains both
85 hemipelagic and volcanic sediments, sometimes over ~1 km thick, providing
86 hydrothermal systems with abundant H_2, CO_2, CH_4, NH_4, H_2S, and CO derived from
87 sedimentary organic matter and from magmatic gases. The Iheya North hydrothermal
88 field is an area of diffuse flow surrounding a focused flow high-temperature vent system,
89 which makes it a promising environment to study functionally active and metabolically
90 diverse microbes (24, 25). The abundant supply of energy and carbon supplied through
91 hydrothermal vent circulation provide an ideal location to study elusive microbes, such as
92 the *Chloroflexi*, within the subsurface biosphere.

93 Five locations were chosen for drilling at Iheya North Hydrothermal field
94 (Figure 1) as part of the Integrated Ocean Drilling Program (IODP) expedition 331 (Deep
95 Hot Biosphere). Based on the initial heat flow results, a hydrodynamic model of fluid
96 flow throughout the Iheya North Hydrothermal field was documented, where variation in
97 chemical and physical processes including the formation of both brine and vapor-rich
98 hydrothermal fluids (i.e., phase separation and phase-partitioning) likely generated by
99 underlying high temperatures coupled with a complex hydrogeologic structure.
100 Recovered drill cores were described as interbedded hemipelagic muds with strongly

pumiceous volcanoclastic sediments; however, there was minimal hydrothermal alteration in the cooler zones, generally within the upper 30 meters below seafloor (mbsf), showing any evidence of microbial activity (24, 25). The cooler sites chosen for this study were (i) Site C0015 located ~600 meters northwest from the active vents on the crest of a hill that represented a potential diffuse (i.e., high seawater mixing) outflow migration path for hydrothermal fluids and (ii) Site C0017 located ~1550 meters to the east of the active vents representing an area of hydrothermal recharge or an inflow path for ambient seawater (24, 25).

We examined and compared nine SAGs belonging to the phylum *Chloroflexi* recovered from the Iheya North hydrothermal field in the Mid-Okinawa Trough during the IODP expedition 331 (24, 25). In contrast to previously studied *Chloroflexi* SAGs, an *rdhA* gene was identified in an *Anaerolineales* representative. This study expands the known metabolic repertoire of this elusive group of microbes and provides insights as to their metabolic potential.

MATERIAL AND METHODS

Sample Collection

Subsurface sediments were collected on IODP expedition 331 (Deep Hot Biosphere) from September 1st through October 4th, 2010 (Figure 1). On board contamination testing from Sites C0015 (126°53'E, 27°47'N, Hole B, section 1H-5, 5.6 mbsf) and C0017 (126°55'E, 27°47'N, Hole C, section 1H-7, 26.6 mbsf) found no indication of interior core contamination using fluorescent microspheres (both Holes C0015B and C0017C) and perfluorocarbon tracer (Hole C0017C only). The sample from Hole C0017C was also verified by PCR-generated phylotype comparisons based on 97%

similarity to phylotypes obtained from drilling mud at a contamination level of 1% or less (26). Subsamples were aseptically collected from the interior of whole-round cores and stored in cryovials with 27% glycerol (v/v) at -80°C.

Single Cell Source

Core depths were chosen from Sites C0015 and C0017 that were characterized as weakly-oxidized pumiceous gravels with no detected sulfide mineralization and less than 0.1 wt% total organic carbon, total nitrogen and total sulfur (25). Selected samples for single cell genomes were from subsurface depths of 5.6 mbsf from Hole C0015B and from 26.6 mbsf from Hole C0017C. Temperatures were estimated at ~10.5°C for C0015B and ~8.1°C for C0017C and at these depths. Details on geochemistry and lithography have been previously described (12, 24, 25).

Single-Cell Sorting, Amplification, Sequencing and Annotation

Samples from Sites C0015 and C0017 (Figure 1) were diluted with 1 ml of filter-sterilized artificial seawater (27) making a slurry, then passed through a 90 µm nylon mesh filter twice, and centrifuged at ~500 x g for 2 min to produce a particle-free cell suspension. This suspension was then processed using fluorescence-activated single-cell sorting at the Single Cell Genome Center (SCGC) at Bigelow Laboratory for Ocean Sciences. Single-cell sorting and multiple displacement amplification (MDA) has been previously described (28). The amplified SSU rRNA gene sequences (27F/907R) were classified using the RDP online classifier (28, 29). Based on their SSU rRNA gene identity, nine *Chloroflexi* SAGs (of the total 29 unique MDA reactions identified after cell sorting) were chosen for whole genome sequencing. These SAGs were sequenced, assembled and contamination checked by the SCGC, using previously well described

parameters (28, 29). Assembly was done using SPAdes v.3.0.0 (30). All contigs were compared to ensure no cross-contamination among SAGs and the NCBI nt database, which was followed by tetramer principal component analysis as previously described (31-33). These analyses revealed no contamination. The full name for each of the SAGs has been shortened, e.g., *Anaerolineales* bacterium SCGC AC-711-B22 to An-B22. Phylogeny has been shortened from *Anaerolineales* to An, *Dehalococcoidales* to De and *Thermoflexales* to Th. The assembled genomes were annotated using RAST (34). Gene annotations were compared to NCBI GenBank, via BLASTn and results can be found in Tables S2-S4.

The *Anaerolineales* SAGs were compared to the genomes of *Anaerolinea thermophila* UNI-1 (GenBank accession number NC_014960) and the single *Thermoflexales* SAG to *Thermoflexus hugenholtzii* JAD2 (NCBI BioProject PRJNA195829) as these were determined to be their closest respective relatives. Type strain *A. thermophila* UNI-1 was isolated from an anaerobic granular sludge reactor treating fried soybean-curd manufacturing waste water in Japan (35), while type strain *Thermoflexus hugenholtzii* JAD2 was isolated from the sediment of Great Boiling Spring in Nevada, USA (36). Both are considered thermophilic, Gram-negative, non-sporeforming, heterotrophic bacteria that grow in multicellular filaments (36, 37).

Phylogenetic Analysis

SSU rRNA gene sequences and phylogenetic relatives were aligned using the Silva SINA aligner (38). For the *rdhA* analysis, amino acids were aligned using ClustalW, within Geneious (39, 40). The resulting alignments were manually screened, then used to create a phylogenic consensus tree using the MrBayes within Geneious (41).

Applied and Environmental Microbiology

The HKY85 substitution model and chain length set at 1,100,000 with a subsampling frequency of 200. Priors were set with an unconstrained branch length. Average nucleotide identity (ANI) was calculated, of the SAGs and selected genomes, with the BLAST parameters as previously described (42).

Genome Completeness Estimates

Genome completeness estimates were determined with BLASTP of predicted amino acid sequences against a set of single copy core genes (43). To be considered valid, all proteins must have at least 30% identity over at least 30% of the length of the core gene (44). The core gene group is made of 66 previously established genes belonging to a non-redundant list as examined by gene ontology (GO) annotations (44, 45).

Accession Numbers

The SSU rRNA gene sequences obtained from MDA have been submitted to NCBI GenBank database (accession numbers KT119838-KT119846). All SAGs have been made public in the IMG Database (IMG Submission IDs: 69642-69649, 68650).

RESULTS AND DISCUSSION

Single Cell Sorts

Classification of SSU rRNA gene sequences by taxonomic analysis showed *Chloroflexi* to be the most abundant within these two subsurface habitats with respect to the total single cells recovered. Nine *Chloroflexi* SSU rRNA gene sequences were recovered; four from Hole C0015B and five from Hole C0017C. Of the nine *Chloroflexi* detected, seven belonged to the order *Anaerolineales*, with one each belonging to the order *Dehalococcoidales* and *Thermoflexales* (Table 1). Phylogenetic placement was not

correlated with the location from which the SAG was recovered (Figure 2). A previous study from IODP Expedition 331 examined cloned SSU rRNA gene sequences for bacterial taxonomic classification from Site C0017 and showed that *Chloroflexi* were one of the main constituents of the microbial communities from 0.7 to 74.9 mbsf crossing multiple sediment types, including oxygenated and anoxic layers as well as the gradients among these layers showing this phylum to be ubiquitous (12).

Genome Composition

Overall 6.4 Mb of *Chloroflexi* genomic data has been analyzed. The assembled sequence length was variable; from 0.15 Mb to 1.4 Mb representing 1.5% to 62.1% with an average of 20.2% completeness based on core Gene Ontology (GO) annotations (Table 1). tRNA recovery estimates reveal a slightly higher percent of recovery (4.0 - 66%, with an average of 28.3%) overall with an estimated average genome size of 3.15 Mb (Table S5). While some of these recovered SAGs were small, all the annotated genes could be conclusively linked to a phylogenetic group based on the classifications originating from the single cell sorts.

The seven *Anaerolineales* and the *Thermoflexales* SAGs had similar GC content at approximately 55%. The remaining SAG belongs to the order *Dehalococcoidales* that had a GC content of 47%. Comparison of the SSU rRNA gene sequences across the nine *Chloroflexi* showed that two SAGs, An-B22 and An-J10, had nearly identical sequences. Whole genome comparisons between An-B22 and An-J10 showed an ANI of 98% (42). Both of these SAGs were collected from Site C0015 and belong to the order *Anaerolineales*. In addition, Th-L07 had an ANI of 98% to An-B22 (Figure S1). By SSU

rRNA gene classification, strain Th-L07 is classified as a *Thermoflexales* and has an 85% identity to An-B22.

Besides An-B22 having high identity to An-J10 and Th-L07, no other comparison of SAGs yielded a greater than 91% ANI. SAG Th-L07 is only 10% complete and is roughly one-third the size of An-B22. The pairwise ANI values only take into account the genes present in the smaller of the genomes compared and though a conservative estimate, this can be misleading across variable sized genomes. It has been suggested that only genomes that are at least 20% complete will provide accurate species ANI values (46); however, ANIs have been shown to differentiate among operational taxonomic units (OTUs) when considering SAGs with a predicted completeness of <20% (47). Regardless of completeness, fuzzy boundaries have been suggested and should be considered when using ANI interpretations (48).

Carbon Metabolism

Predicted central carbon metabolic pathways are illustrated (Figure 3), as a composite of the seven *Anaerolineales* SAGs. In contrast to the photosynthetic *Chloroflexi*, such as *Chloroflexi aurantiacus*, no genes for the 3-hydroxypropionate cycle were identified (6). At least one gene for glycolysis/gluconeogenesis pathway, tricarboxylic acid (TCA) cycle, or the Wood-Ljungdahl pathway is represented in eight of the SAGs (Tables S2-S4). The single exception is De-I04, which is lacking any genes for glycolysis, the TCA cycle and the Wood-Ljungdahl pathway; however, this genome is very incomplete with only a predicted 4.5% recovery. Therefore, it is likely that while a number of genes encoding enzymes catalyzing other central metabolic reactions were recovered, genes for central carbon metabolism were not sequenced rather than were

absent from this genome. Comparisons with other strains suggest these complete pathways may be present, however they were not sequenced from our nine *Chloroflexi* SAGs.

Metabolic products of glycolysis/gluconeogenesis can either be converted into cellular components or fermented into metabolic waste products. Ethanol is one such waste product and an alcohol dehydrogenase was identified in six of the SAGs (An-E09, An-B04, An-L07, An-J10, An-B16, and An-B22). This enzyme acts on primary or secondary alcohols to produce a corresponding aldehyde and NADH. *Chloroflexi* across multiple classes contain annotated alcohol dehydrogenases, including the deep-sea strain DscP2 (16). If alcohols are present within the surrounding environment, they could passively diffuse across the membrane and could be used in central carbon metabolic processes.

In addition to using passive transport, a number of organic carbon transporters were also found. Amino acid, dipeptide and oligopeptide ABC transporters were identified in the *Thermoflexales* SAG (Th-L07) and six of the *Anaerolineales* SAGs (An-L07, An-K11, An-B04, An-J10, An-B16, and An-B22). Xylose, maltose and multiple sugar transporters were found in De-I04, An-L07, An-B16, and An-J10. Nucleoside and deoxyribose-specific transporters were identified in An-K11, An-J10, An-B16, and An-B22. Strains An-L07 and An-K11 encode a glycerol-3-phosphate ABC transporter suggesting exogenous glycerol-3-phosphate (G3P) can be utilized in lipid synthesis or glycolysis. ABC transporters are not unique to these *Chloroflexi* SAGs as *A. thermophila* UNI-1 encodes for xylose, amino acid, ribose, and carbohydrate ABC transporters. Uptake of these fixed carbon molecules could supply precursors for other anabolic

pathways or energy in catabolic pathways, depending on cellular demand. *A. thermophila* UNI-1 grows by fermentation of sugars at neutral pH (37). These *Chloroflexi* could be utilizing the dissolved amino acids, which have been identified elsewhere in an oceanic crustal aquifer (49).

Besides using exogenous carbon sources, these SAGs encode a number of enzymes in the Wood-Ljungdahl pathway of CO_2 fixation (also known as the reductive acetyl-CoA pathway). A formyltetrahydrofolate synthase, methylenetetrahydrofolate cyclohydrolase, 5,10-methylenetetrahydrofolate dehydrogenase, 5,10-methylenetetrahydrofolate reductase, carbon monoxide dehydrogenase/acetyl-CoA synthase (CODH/ACS), carbon-monoxide dehydrogenase catalytic subunit, and formate dehydrogenase were found in all seven of the *Anaerolineales* SAGs. A recent metagenomic reconstruction of a putative aerobic *Anaerolinea* from aquifer sediments adjacent to the Colorado River lacked genes of the Wood-Ljungdahl pathway, pyruvate ferredoxin oxidoreductase and hydrogenases, which is in contrast to our seven *Anaerolineales* SAGs (50). Our *Dehalococcoidales* and *Thermoflexales* SAGs (De-I04 and Th-L07, respectively) are also lacking these genes. Though in our case, it is likely these genes are missing in the assembly rather than absent from the genome since the closest sequenced isolates to these SAGs, *Dehalogenimonas lykanthroporepellens* and *Thermoflexus hugenholtzii* JAD2, encode enzymes for the Wood-Ljungdahl pathway (51). Furthermore, an incomplete Wood-Ljungdahl pathway does not prevent these enzymes from being used in other anabolic processes (52). At least one subunit for the pyruvate ferredoxin oxidoreductase was found in four of the SAGs (An-K11, An-B04,

283 An-B16 and An-J10). This protein can provide a link between the Wood-Ljungdahl
284 pathway and other anabolic pathways (53).

285 In three of the *Anaerolineales* SAGS, An-B16, An-B22 and An-K11, the formate
286 dehydrogenases were annotated to contain selenocysteine in the active site rather than
287 cysteine. Selenocysteine-specific translation factors were identified in An-B04 and
288 An-B22, but not in An-B16. Interestingly, the encoded formate dehydrogenase of An-J10
289 is not predicted to contain a selenocysteine, although a selenocysteine-specific translation
290 factor was identified. Like An-J10, the formate dehydrogenases of *A. thermophila* UNI-1
291 does not encode a selenocysteine in its active site. Many anaerobic dehydrogenases are
292 annotated to contain selenocysteine in the order *Dehalococcoidales*; however, there was
293 no evidence of enzymes requiring selenocysteine in the *Dehalococcoidales* SAG, i.e.,
294 De-I04. *A. thermophila* UNI-1 does not have any dehydrogenases annotated to contain
295 selenocysteine, but it does encode a selenocysteine elongation factor. Selenocysteine-
296 containing formate dehydrogenases are found in many obligate anaerobes and facultative
297 aerobes, suggesting an importance under low oxygen conditions, such as the deep
298 subsurface biosphere (54). At Site C0017, the nitrate concentration at the depth of our
299 sample was less than 2 µM suggesting an anaerobic environment where oxygen was
300 presumably consumed by microbial respiration within the uppermost layer (12). Genes
301 suggesting selenocysteine usage were found at both Sites C0015 and C0017.

302 *Reductases and Hydrogenases*

303 Genes encoding subunits resembling those of the CoB-CoM heterodisulfide
304 reductase (*hdrABCD*) were identified in five of the *Anaerolineales* SAGs (An-K11,
305 An-B04, An-J10, An-B16, and An-B22) and were organized in operons. While this

306 enzyme has not been identified in *A. thermophila* UNI-1, genes for the *hdrABCD*
307 enzymes have been identified in methanogenic archaea and a number of strictly anaerobic
308 bacteria including sequenced isolates *D. lykanthroporopellens* BL-DC-9 and
309 *T. hugenholtzii* JAD2 (51, 55). Most likely due to low genome coverage, these genes
310 were not identified in the *Thermoflexales* and *Dehalococcoidales* SAGs (Th-L07 and
311 De-I04, respectively). The presence of *hrdABCD* in our SAGs supports the hypothesis
312 that these enzymes are responsible for transfer of reducing equivalents to and from
313 ferredoxins or NADH (17). Conservation across these multiple *Chloroflexi* lineages
314 suggests that these redox active enzymes play an integral role in cellular function.
315 Across all nine of our *Chloroflexi* SAGs, a number of hydrogenase and
316 hydrogenase accessory proteins were identified. The genes from the hydrogenase
317 maturation apparatus *hypABCDEF* (56) were identified in An-B16, An-L07, and De-I04.
318 These genes are required for the coordination of the NiFe cofactor and assembly of
319 mature hydrogenase. However, energy conserving hydrogenase (*ech*), cytoplasmic
320 *hymABC* hydrogenase and the uptake hydrogenase (*hup*) were not specifically identified
321 in any of these nine *Chloroflexi* SAGs.
322 *Synthesis and uptake of cofactors*
323 Hydrogenases and other redox active enzymes require specific cofactors.
324 Respiratory nitrate reductases, dimethyl sulfoxide reductases and some carbon monoxide
325 dehydrogenases are all molybdopterin oxidoreductases and therefore require the
326 molybdenum cofactor for activity. The molybdenum cofactor biosynthesis complex
327 *moaABCDE* was represented by one or more genes in six *Anaerolineales* SAGs (An-E09,
328 An-B04, An-L07, An-J10, An-B16 and An-B22). At least one gene encoding enzymes

for molybdopterin cofactor biosynthesis (*moeAB* and *mobAB*) were identified in five of the nine SAGs (De-I04, An-K11, An-B04, An-J10 and An-B16). In addition, molybdenum is required for the xanthine dehydrogenase, an enzyme in purine metabolism, which was also identified in An-B04, An-L07, An-J10 and An-B16. Molybdopterin cofactors can either coordinate the localization of molybdenum or tungsten (57). In An-B22, a tungstate ABC-transporter was identified, suggesting this organism uses tungsten rather than molybdenum. Both An-J10 and An-B22 contain a vitamin B_{12} ABC transporter.

Environmental Adaptations

In the *Dehalococcoidia* SAG DEH-J10 genes for environmental adaptions, such as osmoprotectants (trehalose and alpha-mannoslyglycerate synthases) and oxygen protection, were identified (17). A trehalose synthase gene was identified in An-B22, An-L07 and An-B16. Genes for superoxide dismutase and catalase were identified in An-B22 and only the gene for catalase was identified in Th-L07. These gene products could be used to cope with an oxygen influx introduced during ambient seawater mixing with interstitial hydrothermal fluids, as is likely in the more porous sediment layers.

Cell wall formation

Terrestrial strains of *Chloroflexi* lack genes for peptidoglycan synthesis, and *D. mccartyi* strains contain a proteinaceous surface layer, i.e., S-layer (18). The enzyme peptidoglycan glycosyltransferase is responsible for joining N-acetylglucosamine to N-acetylmuramic acid and is essential for peptidoglycan biosynthesis. This gene was identified in SAGs An-B22, An-B04, and An-K11. It is possible that the annotated peptidoglycan glycosyltransferase genes are actually responsible for glycosylation of the

S-layer proteins as suggested for DEH-J10 (17). It has also been suggested that this single cell-wall genotype and phenotype is preserved across the *Chloroflexi* phylum (58).

Reductive Dehalogenase

A reductive dehalogenase gene (*rdhA*) gene has been directly linked to a subsurface *Chloroflexi* outside of the class *Dehalococcoidia.* Unlike other SAGs within the *Chloroflexi*, An-B22 (the largest genome recovered, Table 1) has an annotated trichloroethene reductive dehalogenase gene (*tceA*); however, this annotation must be interpreted cautiously as RdhA proteins are as yet far from being well characterized. This gene is most closely related to a reductive dehalogenase of the archaeon *Ferroglobus placidus* DSM 10642 (Figure 4). Very few reductive dehalogenases have a known function and this *rdhA* is more similar to genes found among known dehalogenators of the *Clostridia, Delta-* and *Alpha-* classes of the *Proteobacteria* than the *Dehalococcoidia* (59). Like other reductive dehalogenase proteins, this *rdhA* gene (from An-B22) encodes binding motifs for a FeS cluster (Figure S2). In addition, An-B22 encodes for a vitamin B_{12} ABC transporter and the associated corrinoid is a known cofactor of reductive dehalogenases (60).

The reductive dehalogenases of *Dehalococcoides* are encoded in an operon of *rdhAB* genes. The RdhA is the functional dehalogenase and the RdhB is thought to be its associated membrane-bound anchor protein (60). Transcriptional regulators are often encoded adjacent to the *rdhAB* operon and the operon is often on genomic islands that are horizontally acquired and integrated at the single-copy tmRNA gene, *ssrA* (61). None of these conserved characteristics regarding the reductive dehalogenases of *D. mccartyi* were found in An-B22 or *F. placidus* (62). Furthermore, *F. placidus* has not been shown

375 to conserve energy via reductive dehalogenation (63). Reductive dehalogenases are
376 exported through the membrane fully folded via the TatABC export apparatus (60). A
377 *tatA* gene was identified in An-B22; however, no twin arginine signal was identified in
378 this *rdhA* or in that of *F. placidus* (62).

379 Widespread and diverse *rdhA* genes have been detected across the Pacific Ocean
380 in subsurface sediment cores and microcosms have shown dehalogenation activity (20,
381 21). Identification of an *rdhA* in An-B22 provides phylogenetic linkage to an organism
382 within the phylum *Chloroflexi,* but outside the order *Dehalococcoidales*. Whether or not
383 this Rdh is involved with organohalide respiration or more simply dehalogenation of
384 organic compounds enhancing heterotrophic growth or perhaps a mixotrophic
385 combination depending upon ephemeral conditions (e.g., oxygen levels), these results,
386 along with previous research (20, 21), supports the hypothesis that reductive
387 dehalogenation is an important biogeochemical processes within the oceanic deep
388 subsurface biosphere.

389 **CONCLUSION**

390 This study used single cell genomics to examine an active subsurface
391 hydrothermal system and to expand the known metabolic function of uncultured
392 organisms within the phylum *Chloroflexi*. Photosynthetic *Chloroflexi* (e.g., *Chloroflexus*
393 *aurantiacus*) can fix carbon through the 3-hydroxypropionate pathway (6), whereas other
394 *Chloroflexi* isolates (e.g., *D. lykanthroporepellens* and *A. thermophila*) use exogenous
395 sources of carbon (35, 51). Although the two sites from the Iheya North hydrothermal
396 field demonstrated hydrodynamic differences, this was not reflected in the SAGs we
397 analyzed. From this study it remains unclear if these subsurface strains are strictly

heterotrophic or autotrophic. Hydrothermal vents can circulate dissolved organic carbon in addition to inorganic chemicals into the subsurface biosphere (49). Due to the number of ABC transporters identified across all nine SAGs, it is likely these organisms are living heterotrophically. This hypothesis is further supported by *A. thermophila*, which also grows heterotrophically by fermentation and contains a number of shared transporters and carbon utilization genes.

These SAGs yield compelling insights to central carbon metabolism and the potential for reductive dehalogenation in an enigmatic group of subsurface microbes. In contrast to previous *Chloroflexi* SAGs, these genomes were recovered from hydrothermally driven habitats and therefore provide an extended window into the metabolic potential of the deep subsurface biosphere. Even though the SAGs examined here range in genome completeness, collectively they have enabled insights into the metabolic potential of subsurface *Chloroflexi* providing useful functional gene information with special emphasis regarding the *Anaerolineales*.

FUNDING INFORMATION

This work has been funded in part by Western Washington University's Office of Research and Sponsored Programs, by the National Science Foundation OCE 1260710 (to CLM), and the Center for Dark Energy Biosphere Investigations. This is C-DEBI Contribution 322.

ACKNOWLEDGEMENTS

The authors wish to acknowledge and thank Yuki Morono from the Kochi Institute for Core Sample Research (Japan Agency for Marine-Earth Science and Technology, Japan) as well as Joyce McBeth and Erin Field from Bigelow Laboratory for Ocean Sciences for their help and guidance with sample preparation for cell sorting. We also thank Benjamin Tully for his technical expertise and advice in determining genome completeness estimates. We are also grateful to both David Emerson (Bigelow Laboratory) and Stephen H. Zinder (Cornell University) for their insightful comments.

REFERENCES

1. **Rappé MS, Giovannoni SJ.** 2003. The uncultured microbial majority. Annu Rev Microbiol **57:**369-394.
2. **Wang Y, Sheng HF, He Y, Wu JY, Jiang YX, Tam NFY, Zhou HW.** 2012. Comparison of the levels of bacterial diversity in freshwater, intertidal wetland, and marine sediments by using millions of illumina tags. Appl Environ Microbiol **78:**8264-8271.
3. **Yamada T, Sekiguchi Y.** 2009. Cultivation of uncultured *Chloroflexi* subphyla: Significance and ecophysiology of formerly uncultured *Chloroflexi* 'subphylum I' with natural and biotechnological relevance. Microbes Environ **24:**205-216.
4. **Taş N, van Eekert MHA, de Vos WM, Smidt H.** 2009. The little bacteria that can - diversity, genomics and ecophysiology of '*Dehalococcoides*' spp. in contamnated environments. Microb Biotechnol **3:**389-402.
5. **Krzmarzick MJ, Crary BB, Harding JJ, Oyerinde OO, Leri AC, Myneni SCB, Novak PJ.** 2012. Natural niche for organohalide-respiring *Chloroflexi*. Appl Environ Microbiol **78:**393-401.
6. **Bryant DA, Liu Z, Li T, Zhao F, Costas AMG, Klatt CG, Ward DM, Frigaard N-U, Overmann J.** 2012. Comparative and functional genomics of anoxygenic green bacteria from the taxa *Chlorobi*, *Chloroflexi*, and *Acidobacteria*, p 47-102. *In* Burnap RL, Vermaas, W (eds), Functional genomics and evolution of photosynthetic systems. Advances in photosynthesis and respiration, vol. 33. Springer, Netherlands.
7. **Sorokin DY, Lücker S, Vejmelkova D, Kostrikina NA, Kleerebezem R, Rijpstra WIC, Damsté JSS, Le Paslier D, Muyzer G, Wagner M, van Loosdrecht MCM, Daims H.** 2012. Nitrification expanded: discovery, physiology and genomics of a nitrite-oxidizing bacterium from the phylum *Chloroflexi*. ISME J. **6:**2245-2256.
8. **Inagaki F, Nunoura T, Nakagawa S, Teske A, Lever M, Lauer A, Suzuki M, Takai K, Delwiche M, Colwell FS, Nealson KH, Horikoshi K, D'Hondt S, Jørgensen BB.** 2006. Biogeographical distribution and diversity of microbes in methane hydrate-bearing deep marine sediments on the Pacific Ocean Margin. Proc Natl Acad Sci USA **103:**2815-2820.
9. **Biddle JF, Fitz-Gibbon S, Schuster SC, Brenchley JE, House CH.** 2008. Metagenomic signatures of the Peru Margin subseafloor biosphere show a genetically distinct environment. Proc Natl Acad Sci USA **105:**10583-10588.
10. **Blazejak A, Schippers A.** 2010. High abundance of JS-1- and *Chloroflexi*-related *Bacteria* in deeply buried marine sediments revealed by quantitative, real-time PCR. FEMS Microbiol Ecol **72**:198-207.
11. **Biddle JF, Sylvan JB, Brazelton WJ, Tully BJ, Edwards KJ, Moyer CL, Heidelberg JF, Nelson WC.** 2011. Prospects for the study of evolution in the deep biosphere. Front Microbiol **2:**285.
12. **Yanagawa K, Breuker A, Schippers A, Nishizawa M, Ijiri A, Hirai M, Takaki Y, Sunamura M, Urabe T, Nunoura T, Takai K.** 2014. Microbial community stratification controlled by the subseafloor fluid flow and geothermal gradient at the Iheya North hydrothermal field in the Mid-Okinawa Trough

(Integrated Ocean Drilling Program Expedition 331). Appl Environ Microbiol **80:**6126-6135.

13. **Orcutt BN, Bach W, Becker K, Fisher AT, Hentscher M, Toner BM, Wheat CG, Edwards KJ.** 2010. Colonization of subsurface microbial observatories deployed in young ocean crust. ISME J **5:**692-703.
14. **Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J-F, Darling A, Malfatti S, Swan BK, Gies EA, Dodsworth JA, Hedlund BP, Tsiamis G, Sievert SM, Liu W-T, Eisen JA, Hallam SJ, Kyrpides NC, Stepanauskas R, Rubin EM, Hugenholtz P, Woyke T.** 2013. Insights into the phylogeny and coding potential of microbial dark matter. Nature **499:**431-437.
15. **Stepanauskas R.** 2012. Single cell genomics: an individual look at microbes. Curr Opin Microbiol **15:**613-620.
16. **Kaster A-K, Mayer-Blackwell K, Pasarelli B, Spormann AM.** 2014. Single cell genomic study of *Dehalococcoidetes* species from deep-sea sediments of the Peruvian Margin. ISME J **8:**1831-1842.
17. **Wasmund K, Schreiber L, Lloyd KG, Petersen DG, Schramm A, Stepanauskas R, Jørgensen BB, Adrian L.** 2014. Genome sequencing of a single cell of the widely distributed marine subsurface *Dehalococcoidia*, phylum *Chloroflexi*. ISME J **8:**383-397.
18. **Löffler FE, Yan J, Ritalahti KM, Adrian L, Edwards EA, Konstantinidis KT, Müller JA, Fullerton H, Zinder SH, Spormann AM.** 2013. *Dehalococcoides mccarty*i gen. nov., sp. nov., obligately organohalide-respiring anaerobic bacteria relevant to halogen cycling and bioremediation, belong to a novel bacterial class, *Dehalococcoidia* classis nov., order *Dehalococcoidales* ord. nov. and family *Dehalococcoidaceae* fam. nov., within the phylum *Chloroflexi*. Int J Syst Evol Microbiol **63:**625-635.
19. **Hug LA, Edwards EA.** 2013. Diversity of reductive dehalogenase genes from environmental samples and enrichment cultures identified with degenerate primer PCR screens. Front Microbiol **4:**341.
20. **Futagami T, Morono Y, Terada T, Kaksonen AH, Inagaki F.** 2009. Dehalogenation activities and distribution of reductive dehalogenase homologous genes in marine subsurface sediments. Appl Environ Microbiol **75:**6905-6909.
21. **Kawai M, Futagami T, Toyoda A, Takaki Y, Nishi S, Hori S, Arai W, Tsubouchi T, Morono Y, Uchiyama I, Ito T, Fujiyama A, Inagaki F, Takami H.** 2014. High frequency of phylogenetically diverse reductive dehalogenase-homologous genes in deep subseafloor sedimentary metagenomes. Front Microbiol **5:**80.
22. **Holden JF, Breier JA, Rogers KL, Schulte MD, Toner BM.** 2012. Biogeochemical processes at hydrothermal vents: Microbes and minerals, bioenergetics, and carbon fluxes. Oceanography **25:**196-208.
23. **Nakagawa S, Takai K.** 2008. Deep-sea vent chemoautotrophs: diversity, biochemistry and ecological significance. FEMS Microbiol Ecol **65:**1-14.
24. **Takai K, Mottl MJ, Nielsen SHH, Expedition 331 Scientists.** 2012. IODP Expedition 331: Strong and expansive subseafloor hydrothermal activities in the Okinawa Trough. Scientific Drilling **13:**19-27.

25. **Takai K, Mottl MJ, Nielsen SHH, Expedition 331 Scientists.** 2011. Proceeding of the IODP, 331: Washington DC. Integrated Ocean Drilling Program Management International, Inc, La Jolla, CA.
26. **Yanagawa K, Nunoura T, McAllister SM, Hirai M, Breuker A, Brandt L, House CH, Moyer CL, Birrien J-L, Aoike K, Sunamura M, Urabe T, Mottl MJ, and Takai K.** 2013. The first microbiological contamination assessment by deep-sea drilling and coring by the D/V *Chikyu* at the Iheya North hydrothermal field in the Mid-Okinawa Trough (IODP Expedition 331). Front Microbiol **4:**327.
27. **Kester DR, Duedall IW, Connors DN, Pytkowicz RM.** 1967. Preparation of artificial seawater. Limnol Oceanogr **12**:176-179.
28. **Stepanauskas R, Sieracki ME.** 2007. Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. Proc Natl Acad Sci USA **104:**9052-9057.
29. **Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, Kulam-Syed-Mohideen AS, McGarrell DM, Marsh T, Garrity GM, Tiedje JM.** 2009. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res **37:**D141-D145.
30. **Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA.** 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol **19:**455-477.
31. **Wilkins MJ, Kennedy DW, Castelle CJ, Field EK, Stepanauskas R, Fredrickson JK, Konopka AE.** 2014. Single-cell genomics reveals metabolic strategies for microbial growth and survival in an oligotrophic aquifer. Microbiol **160:**362-372.
32. **Swan BK, Tupper B, Sczyrba A, Lauro FM, Martinez-Garcia M, González JM, Luo H, Wright JJ, Landry ZC, Hanson NW, Thompson BP, Poulton NJ, Schwientek P, Acinas SG, Giovannoni SJ, Moran MA, Hallam SJ, Cavicchioli R, Woyke T, Stepanauskas R.** 2013. Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. Proc Natl Acad Sci USA **110:**11463-11468.
33. **Woyke T, Xie G, Copeland A, Gonzalez JM, Han C, Kiss H, Saw JH, Senin P, Yang C, Chatterji S, Cheng J-F, Eisen JA, Sieracki ME, Stepanauskas, R.** 2009. Assembling the marine metagenome, one cell at a time. PLoS One **4:**e5299.
34. **Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O.** 2008. The RAST Server: Rapid Annotations using Subsystems Technology. BMC Genomics **9:**75.
35. **Sekiguchi Y, Yamada T, Hanada S, Ohashi A, Harada H, Kamagata Y.** 2003. *Anaerolinea thermophila* gen. nov., sp. nov. and *Caldilinea aerophila* gen. nov., sp. nov., novel filamentous thermophiles that represent a previously uncultured lineage of the domain *Bacteria* at the subphylum level. Int J Syst Evol Microbiol **53:**1843-1851.

36. **Dodsworth JA, Gevorkian J, Despujos F, Cole JK, Murugapiran SK, Ming H, Li W-J, Zhang G, Dohnalkova A, Hedlund BP.** 2014. *Thermoflexus hugenholtzii* gen. nov., sp. nov., a thermophilic, microaerophilic, filamentous bacterium representing a novel class in the *Chloroflexi*, *Thermoflexia* classis nov., and description of *Thermoflexaceae* fam. nov. and *Thermoflexales* ord. nov. Int J Syst Evol Microbiol **64:**2119-2127.
37. **Yamada T, Sekiguchi Y, Hanada S, Imachi H, Ohashi A, Harada H, Kamagata Y.** 2006. *Anaerolinea thermolimosa* sp. nov., *Levilinea saccharolytica* gen. nov., sp. nov. and *Leptolinea tardivitalis* gen. nov., sp. nov., novel filamentous anaerobes, and description of the new classes *Anaerolineae* classis nov. and *Caldilineae* classis nov. in the bacterial phylum *Chloroflexi*. Int J Syst Evol Microbiol **56:**1331-1340.
38. **Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glockner 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.
39. **Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG.** 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol **7:**539.
40. **Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A.** 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics **28:**1647-1649.
41. **Huelsenbeck JP, Ronquist F.** 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics **17:**754-755.
42. **Konstantinidis KT, Tiedje JM.** 2005. Genomic insights that advance the species definition for prokaryotes. Proc Natl Acad Sci USA **102:**2567-2572.
43. **Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res **25:**3389-3402.
44. **Mitreva M.** 2008. Bacterial core gene set. NIH Human Microbiome Project. [http://www.hmpdacc.org/doc/sops/reference_genomes/metrics/Bacterial_CoreGenes_SOP.pdf]
45. **Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G.** 2000. Gene Ontology: tool for the unification of biology. Nat Genet **25:**25-29.
46. **Richter M, Rosselló-Móra R.** 2009. Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci USA **106:**19126-19131.
47. **Field EK, Sczyrba A, Lyman AE, Harris CC, Woyke T, Stepanauskas R, Emerson D.** 2015. Genomic insights into the uncultivated marine *Zetaproteobacteria* at Loihi Seamount. ISME J **9:**857-870.

48. **Rosselló-Móra R, Amann R.** 2015. Past and future species definitions for *Bacteria* and *Archaea*. Syst Appl Microbiol **38:**209-216.
49. **Lin H-T, Amend JP, LaRowe DE, Bingham J-P, Cowen JP.** 2015. Dissolved amino acids in oceanic basaltic basement fluids. Geochim Cosmochim Acta **164:**175-190.
50. **Hug LA, Castelle CJ, Wrighton KC, Thomas BC, Sharon I, Frischkorn KR, Williams KH, Tringe SG, Banfield JF.** 2013. Community genomic analyses constrain the distribution of metabolic traits across the Chloroflexi phylum and indicate roles in sediment carbon cycling. Microbiome **1:**22.
51. **Siddaramappa S, Challacombe JF, Delano SF, Green LD, Daligault H, Bruce D, Detter C, Tapia R, Han S, Goodwin L, Han J, Woyke T, Pitluck S, Pennacchio L, Nolan M, Land M, Chang Y-J, Kyrpides NC, Ovchinnikova G, Hauser L, Lapidus A, Yan J, Bowman KS, da Costa MS, Rainey FA, Moe WM.** 2012. Complete genome sequence of *Dehalogenimonas lykanthroporepellens* type strain (BL-DC-9^T) and comparison to "*Dehalococcoides*" strains. Stand Genomic Sci **6:**251-264.
52. **Zhuang W-Q, Yi S, Bill M, Brisson VL, Feng X, Men Y, Conrad ME, Tang YJ, Alvarez-Cohen L.** 2014. Incomplete Wood-Ljungdahl pathway facilitates one-carbon metabolism in organohalide-respiring *Dehalococcoides mccartyi*. Proc Natl Acad Sci USA **111:**6419-6424.
53. **Furdui C, Ragsdale SW.** 2000. The role of pyruvate ferredoxin oxidoreductase in pyruvate synthesis during autotrophic growth by the Wood-Ljungdahl pathway. J Biol Chem **275:**28494-28499.
54. **Romero H, Zhang Y, Gladyshev VN, Salinas G.** 2005. Evolution of selenium utilization traits. Genome Biol **6:**R66.
55. **Thauer RK, Kaster A-K, Seedorf H, Buckel W, Hedderich R.** 2008. Methanogenic archaea: ecologically relevant differences in energy conservation. Nat Rev Microbiol **6:**579-591.
56. **Constant P, Chowdhury SP, Hesse L, Pratscher J, Conrad R.** 2011. Genome data mining and soil survey for the novel group 5 [NiFe]-hydrogenase to explore the diversity and ecological importance of presumptive high-affinity H_2-oxidizing Bacteria. Appl Environ Microbiol **77:**6027-6035.
57. **Kisker C, Schindelin H, Baas D, Rétey J, Meckenstock RU, Kroneck PMH.** 1999. A structural comparison of molybdenum cofactor-containing enzymes. FEMS Microbiol Rev **22:**503-521.
58. **Sutcliffe IC.** 2011. Cell envelope architecture in the *Chloroflexi*: a shifting frontline in a phylogenetic turf war. Environ Microbiol **13:**279-282.
59. **Hug LA, Maphosa F, Leys D, Löffler FE, Smidt H, Edwards EA, Adrian L.** 2013. Overview of organohalide-respiring bacteria and a proposal for a classification system for reductive dehalogenases. Philos Trans R Soc Lond B Biol Sci **368:**20120322.
60. **Smidt H, de Vos WM.** 2004. Anaerobic microbial dehalogenation. Annu Rev Microbiol **58:**43-73.
61. **McMurdie PJ, Hug LA, Edwards EA, Holmes S, Spormann AM.** 2011. Site-specific mobilization of vinyl chloride respiration islands by a mechanism common in *Dehalococcoides*. BMC Genomics **12:**287.

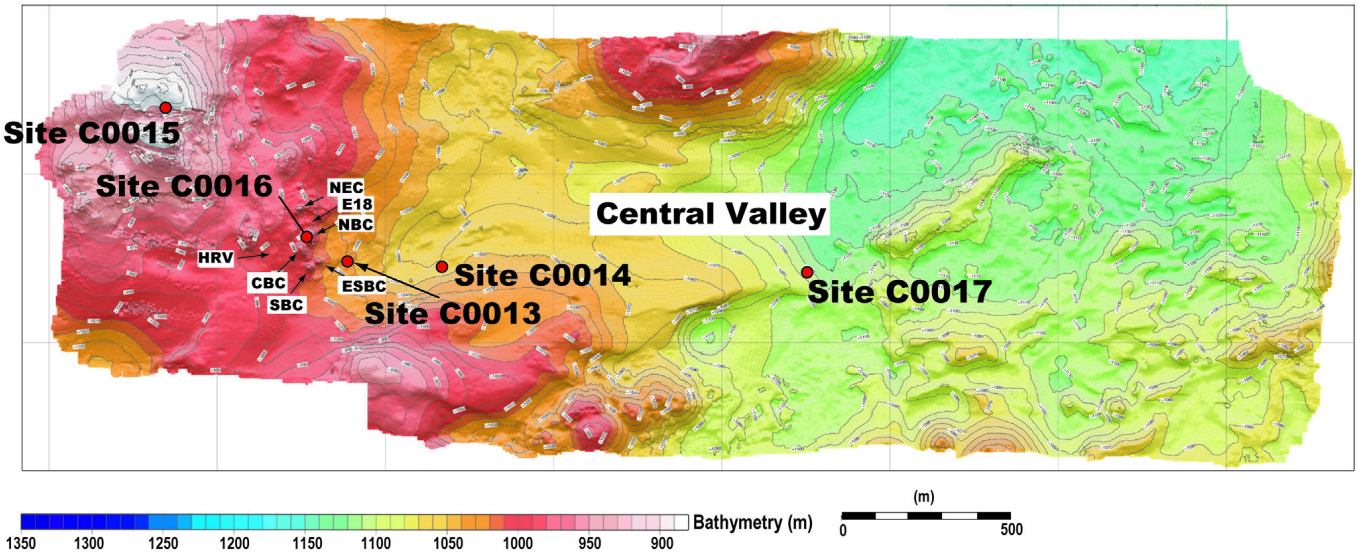
62. **Anderson I, Risso C, Holmes D, Lucas S, Copeland A, Lapidus A, Cheng J-F, Bruce D, Goodwin L, Pitluck S, Saunders E, Brettin T, Detter JC, Han C, Tapia R, Larimer F, Land M, Hauser L, Woyke T, Lovley D, Kyrpides N, Ivanova N.** 2011. Complete genome sequence of *Ferroglobus placidus* AEDII12DO. Stand Genomic Sci **5:**50-60.

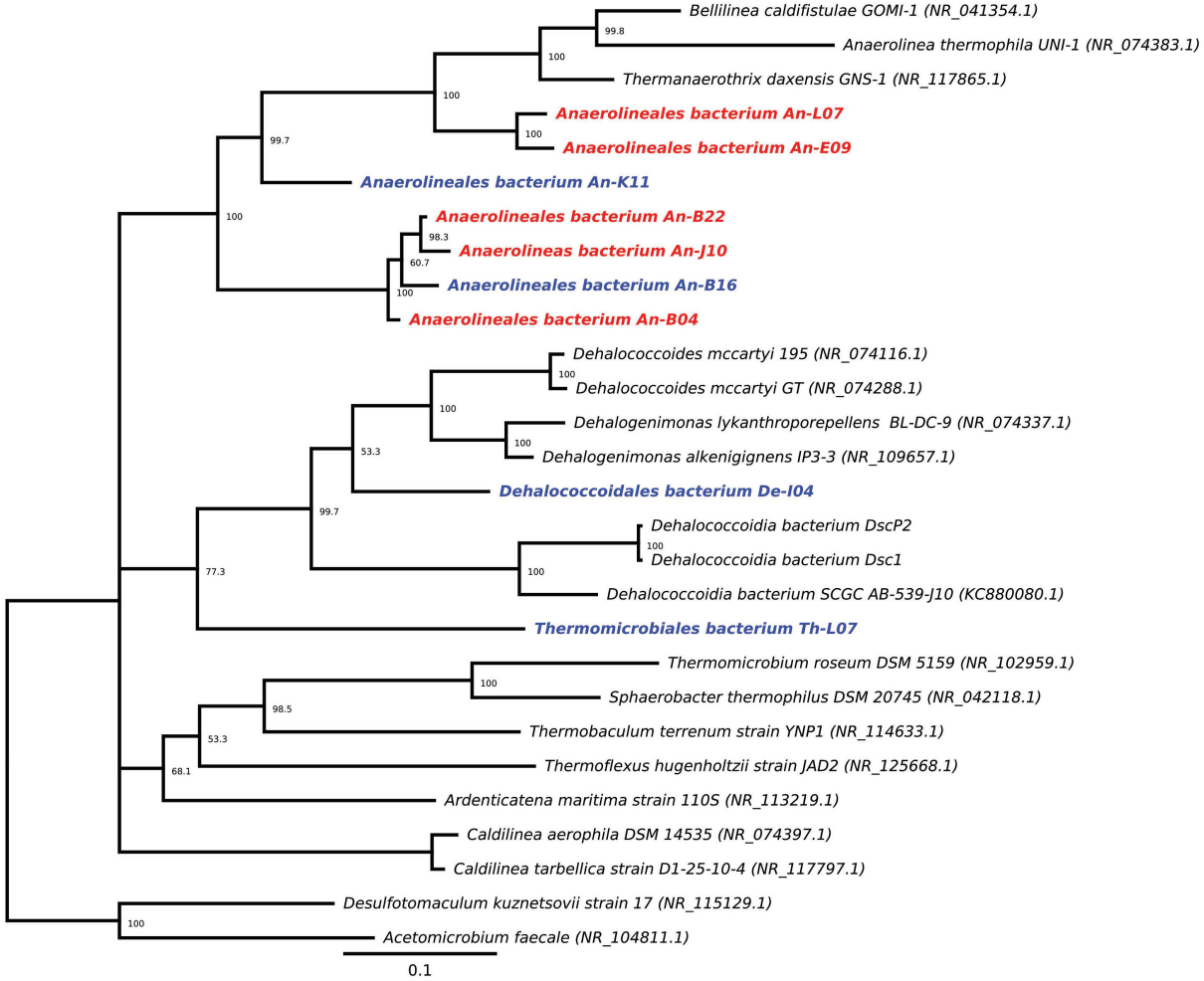
63. **Hafenbradl D, Keller M, Dirmeier R, Rachel R, Roßnagel P, Burggraf S, Huber H, Stetter KO.** 1996. *Ferroglobus placidus* gen. nov., sp. nov., a novel hyperthermophilic archaeum that oxidizes Fe^{2+} at neutral pH under anoxic conditions. Arch Microbiol **166:**308-314.

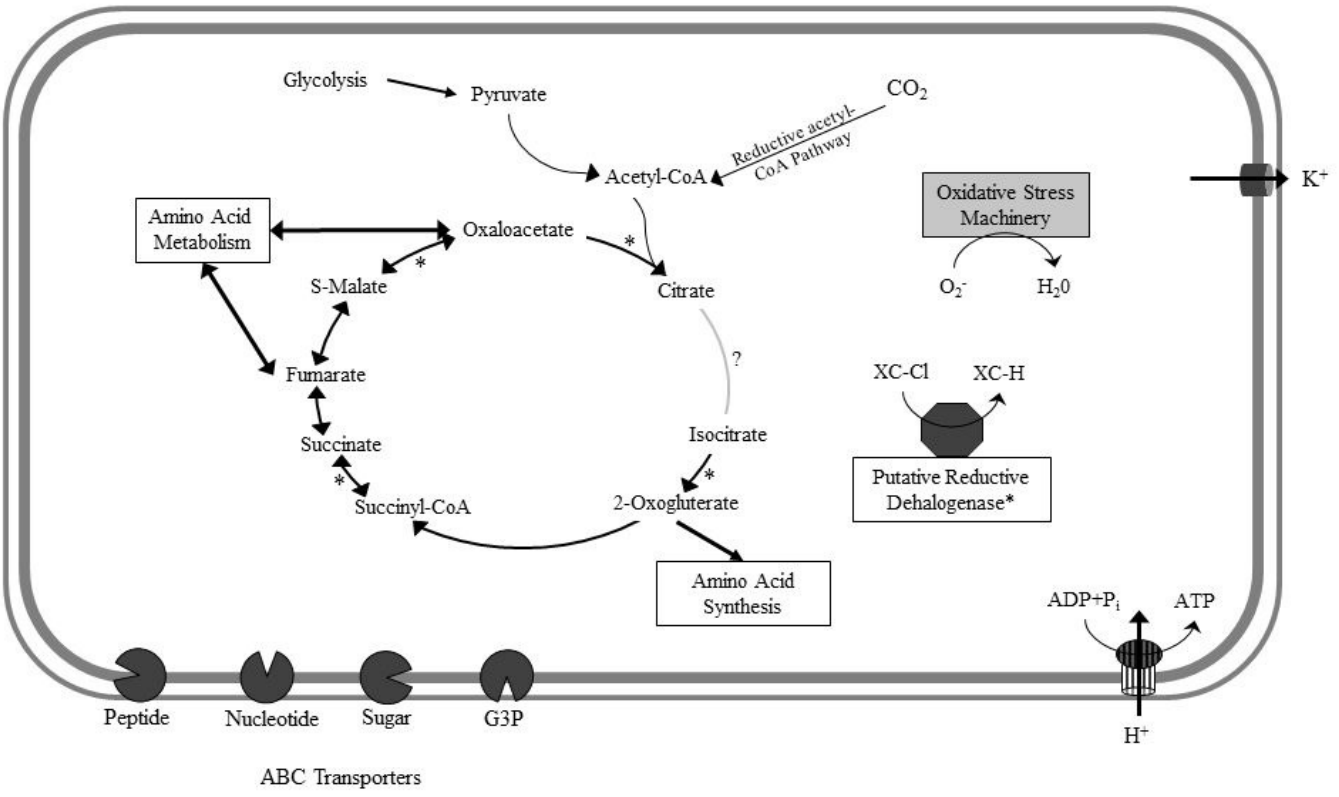
Table 1. Assembly Statistics and phylogenetic classification as determined by SSU rRNA gene sequences from *Chloroflexi* SAGs.

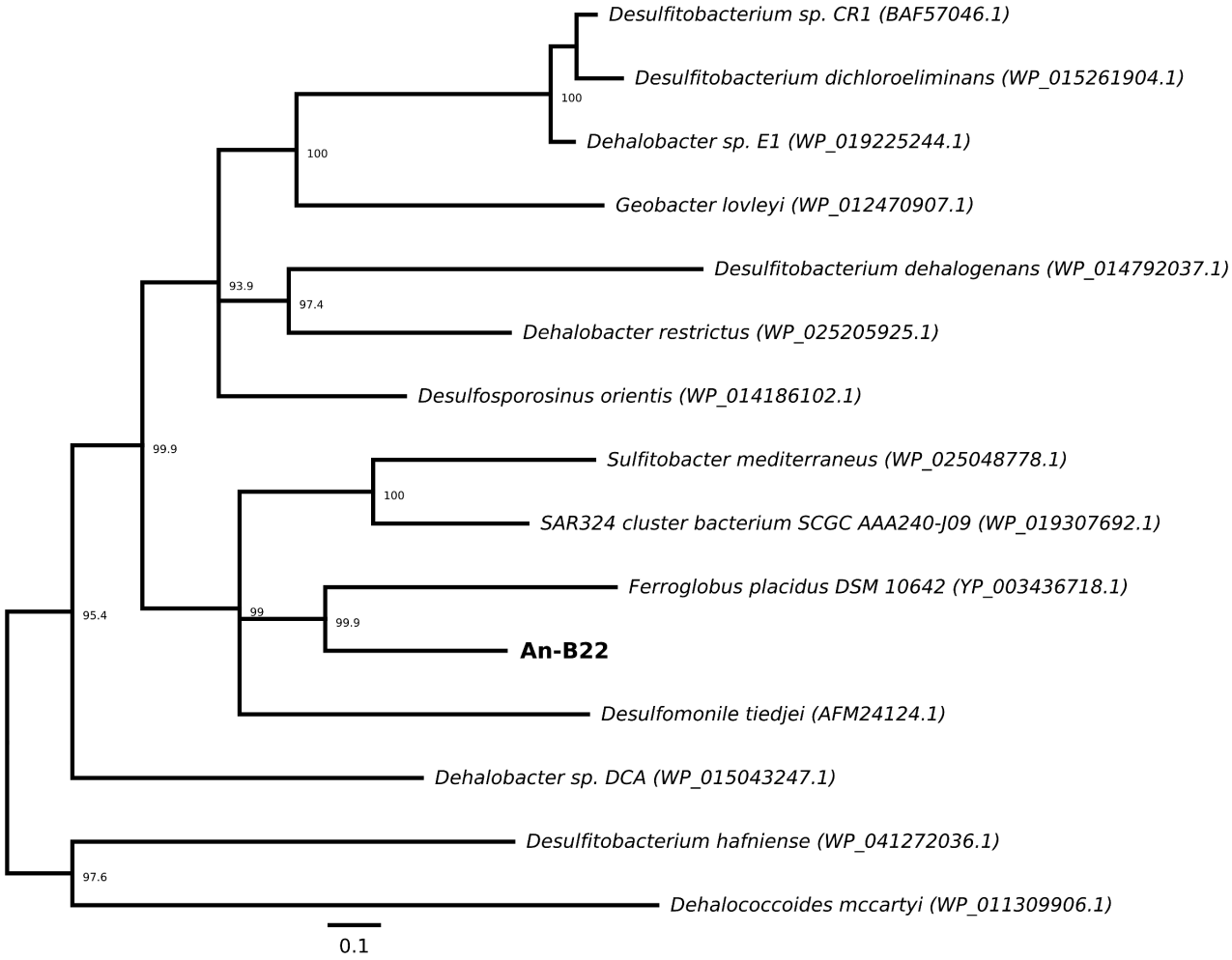
SAG Name	Order	Collection Site	Assembled Length	Estimated genome recovery (%)	Max Contig Length	Contig Count	GC content (%)	No. predicted genes	No. of RNAs	No. of tRNAs
An-E09	*Anaerolineales*	C0015	153,247	1.5%	50,084	11	55.6	175	2	2
De-I04	*Dehalococcoidales*	C0017	187,486	4.5%	30,599	21	47.8	211	5	5
Th-L07	*Thermoflexales*	C0017	462,464	10.6%	46,198	46	57.63	447	10	7
An-K11	*Anaerolineales*	C0017	493,899	18.2%	46,925	50	56.5	455	18	18
An-B04	*Anaerolineales*	C0017	615,816	16.7%	66,121	63	56.6	588	5	4
An-L07	*Anaerolineales*	C0015	800,722	16.7%	125,785	72	58.4	818	11	11
An-J10	*Anaerolineales*	C0015	1,010,824	19.7%	84,943	62	56.9	967	21	20
An-B16	*Anaerolineales*	C0017	1,258,475	31.8%	101,499	82	56.9	1214	35	33
An-B22	*Anaerolineales*	C0015	1,423,460	62.1%	67,055	134	56.8	1411	33	26

663 **FIGURE LEGENDS**
664
665 **Figure 1.** Detailed bathymetry of the Iheya North hydrothermal field and the central
666 valley at Iheya North Knoll, with the location of Sites C0013 through C0017 drilled
667 during IODP Expedition 331. HRV = High Radioactive Vent mound, CBC = Central Big
668 Chimney, SBC = South Big Chimney mound, NEC = North Edge Chimney mound, E18
669 = Event Marker 18 mound, NBC = North Big Chimney mound, ESBC = "Ese" South Big
670 Chimney mound (24, 25).
671
672 **Figure 2.** An unrooted Bayesian tree of *Chloroflexi* SSU rRNA gene sequences from the
673 Okinawa Trough. Samples are color coded for collection Site, from C0015 and C0017 in
674 red and blue, respectively. Numbers at nodes represent % consensus support. Scale bar
675 represents one nucleotide substitution per ten positions.
676
677 **Figure 3.** Schematic representing the composite metabolic and transport proteins
678 hypothesized from genome analysis of seven *Anaerolineales* SAGs. The genes marked
679 with an (*) were identified in only one of the *Anaerolineales* SAGs.
680
681 **Figure 4.** An unrooted Bayesian tree of reductive dehalogenase homologous (RdhA)
682 sequences, including the one identified in *Anaerolineales* SAG An-B22. Numbers at
683 nodes represent % consensus support. Scale bar represents one amino acid substitution
684 per ten positions.









Supplementary Information for

Comparative single-cell genomics of *Chloroflexi* from the Okinawa Trough deep subsurface biosphere

H. Fullerton and C. L. Moyer*

*To whom correspondence should be addressed: cmoyer@hydro.biol.wwu.edu

This file contains:

Supplementary Tables S1 - S5

Supplementary Figures S1 - S2

Table S1. Phylogeny and metabolism of select organisms in the phylum *Chloroflexi* .

Organism	Order	Metabolic Classification	Accession Number
Anaerolinea thermophila	*Anaerolineales*	Anaerobic Heterotroph	NC_014960
Caldilinea aerophila	*Caldilineales*	Facultative Heterotroph	NC_017079
Chloroflexus aurantiacus	*Chloroflexales*	Anoxygenic Photoautotroph	NC_010175
Dehalogenimonas lykanthroporepellens	*Dehalococcoidales*	Anaerobic Dehalorespirer	GCA_000143165
Nitrolancetus hollandicus	*Sphaerobacterales*	Chemoautotrophic Nitrite Oxidizer	CAGS01000001
Thermomicrobium roseum	*Thermomicrobiales*	Aerobic Chemoheterotroph	CP001275 (chromosome) CP001276 (Megaplasmid)

Table S2. One Carbon Pool Genes

Carbon Utilization Pathway	Enzyme	EC number	Dehalococcoidales	Thermoflexales	Anaerolineales						
			AB-790-I04	AB-790-L07	AB-790-K11	AC-715-B04	AB-790-B16	AC-711-E09	AC-711-L07	AC-711-J10	AC-711-B22
One Carbon pool by Folate	Dihydrofolate reductase	1.5.1.3						0115			
One Carbon pool by Folate	Formyltetrahydrofolate dehydrogenase	1.5.1.6									
One Carbon pool by Folate	Thymidylate synthase	2.1.1.45						0116			
One Carbon pool by Folate	Phosphoribosylglycinamide formyltransferase	2.1.2.3/2.1.2.2	0159c						0803		
One Carbon pool by Folate	5-methyltetrahydrofolate--homocysteine methyltransferase	2.1.1.13					0599, 0651c, 0657c, 0998c		0365	0400	0307
One Carbon pool by Folate	Formimidoyltetrahydrofolate cyclodeaminase	4.3.1.4			0064c					0682	
One Carbon pool by Folate	methionyl-tRNA formyltransferase	2.1.2.9	0081c			0409c				0261	
Wood-Ljungdahl	carbon-monoxide dehydrogenase catalytic subunit	1.2.99.2/1.2.7.4			CooS=0059c				Large Chain=0723		
Wood-Ljungdahl	5,10-Methylenetetrahydrofolate reductase (NAD(P))	1.5.1.20				0343, 0347	0309				0375c
Wood-Ljungdahl	CO dehydrogenase/acetyl-CoA synthase, acetyl-CoA synthase subunit	2.3.1.169			0055c						
Wood-Ljungdahl	Formate dehydrogenase, alpha and delta subunit	1.2.1.2			0279		0768c			0063	0854
Wood-Ljungdahl	Formyltetrahydrofolate synthetase	6.3.4.3			0062c						
Wood-Ljungdahl	5,10-methylenetetrahydrofolate dehydrogenase	1.5.1.5			0063c	0351				0311	0370c
Wood-Ljungdahl	cyclohydrolase	3.5.4.9			0063c	0351				0311	0370c

Table S3. Glycolysis Genes

			Dehalococcoidales	Thermoflexales	Anaerolineales						
Carbon Utilization Pathway	Enzyme	EC number	AB-790-I04	AB-790-L07	AB-790-K11	AC-715-B04	AB-790-B16	AC-711-E09	AC-711-L07	AC-711-J10	AC-711-B22
Glycolysis	Phosphoglucomutase	5.4.2.2				0194	0718c				0165c
Glycolysis	glucokinase	2.7.1.2			0102				0287c		
Glycolysis	isomerase/mannose-6-phosphate isomerase, archaeal	5.3.1.9/5.3.1.8		0333						0652c	1150c
Glycolysis	6-phosphofructokinase	2.7.1.11			0068c		0943c				
Glycolysis	fructose-bisphosphatase	3.1.3.11					0505				
Glycolysis	fructose-bisphosphate aldolase	4.1.2.13					0817c				
Glycolysis	Triosephosphate isomerase	5.3.1.1					0095c				
Glycolysis	glyceraldehyde-3-phosphate dehydrogenase	1.2.1.12					0944c				
Glycolysis	glyceraldehyde-3-phosphate dehydrogenase	1.2.1.59									1092
Glycolysis	Phosphoglycerate kinase	2.7.2.3							0775c		
Glycolysis	2,3-bisphosphoglycerate-independent phosphoglycerate mutase, archaeal type	5.4.2.1			0018c		0606c		0541	0831c	1279
Glycolysis	Enolase	4.2.1.11				0143c	0294				
Glycolysis	Pyruvate Kinase	2.7.1.40					0945c				
Glycolysis	Pyruvate dehydrogenase E1 component	1.2.4.1						α=0146c, β=0147c	α=0012c, β=0013c		
Glycolysis	Pyruvate:ferredoxin oxidoreductase	1.2.7.1			α=0317, β=0318	α=0532c, β=0531c, γ=0534c, δ=0533c	β=1197c			α=0200c, β=0199c, γ=0202c, δ=0201c	
Glycolysis	Dihydrolipoamide dehydrogenase	1.8.1.4			0338		0174c		0297c		
Glycolysis	Acetate-CoA ligase	6.2.1.1				0158	0222c			0250c	

Table S4. TCA Genes

			Dehalococcoidales	Thermoflexales	Anaerolineales						
Carbon Utilization Pathway	Enzyme	EC number	AB-790-I04	AB-790-L07	AB-790-K11	AC-715-B04	AB-790-B16	AC-711-E09	AC-711-L07	AC-711-J10	AC-711-B22
TCA	Citrate Synthase	2.3.3.1						0089			
TCA	Aconitase	4.2.1.3									
TCA	Isocitrate dehydrogenase	1.1.1.42/1.1.1.41		0084							0381c
TCA	2-Oxoacid:acceptor oxidoreductase	1.2.7.3				α=0512c, β=0488c/0513, γ=514c, δ=0511	α=0609/0610, β=0134c/0611, γ=0612, δ=0608			α=0075/0544c, β=0076, γ=0077, δ=0074	α=1089, β=1090, γ=0588, δ=0587
TCA	Succinyl-CoA synthetase (ADP-forming)	6.2.1.5									α=1198, β=1196
TCA	Fumerate Reductase	1.3.5.4							C=0522c, D=0521c		
TCA	Succinate Dehydrogenase	1.3.5.1							0523c, 0524c,	0756c	1176c
TCA	Fumarase, class I, aerobic	4.2.1.2				0072, 0073	0778, 0780				0317, 0319
TCA	Malate Dehydrogenase	1.1.1.37		0085							0249c
TCA	Pyruvate/oxaloacetate carboxyltransferase	6.4.1.1					0801c				

Table S5. Estimated genome size based percent genome recovery as determined by tRNAs of nearest neighbor.

SAG Name	Nearest Sequenced Neighbor	SSU % ID to Nearest Neighbor	Estimated Genome Recovery (Based on Nearest Neighbor tRNA Count)	Estimated Size of Genome (Based on tRNAs)
AC-711-E09	*Anaerolinea thermophila* UNI-1	82.54%	4%	3,831,175
AB-790-I04	*Dehalogenimonas lykanthroporepellens* BL-DC-9	87.41%	11%	1,768,736
AB-790-L07	*Thermoflexus hugenholtzii JAD2*	85.62%	16%	2,908,579
AB-790-K11	*Anaerolinea thermophila* UNI-1	80.72%	36%	1,371,942
AC-715-B04	*Anaerolinea thermophila* UNI-1	78.50%	8%	7,697,700
AC-711-L07	*Anaerolinea thermophila* UNI-1	84.27%	22%	3,639,645
AC-711-J10	*Anaerolinea thermophila* UNI-1	78.95%	40%	2,527,060
AB-790-B16	*Anaerolinea thermophila* UNI-1	78.56%	66%	1,906,780
AC-711-B22	*Anaerolinea thermophila* UNI-1	78.47%	52%	2,737,423

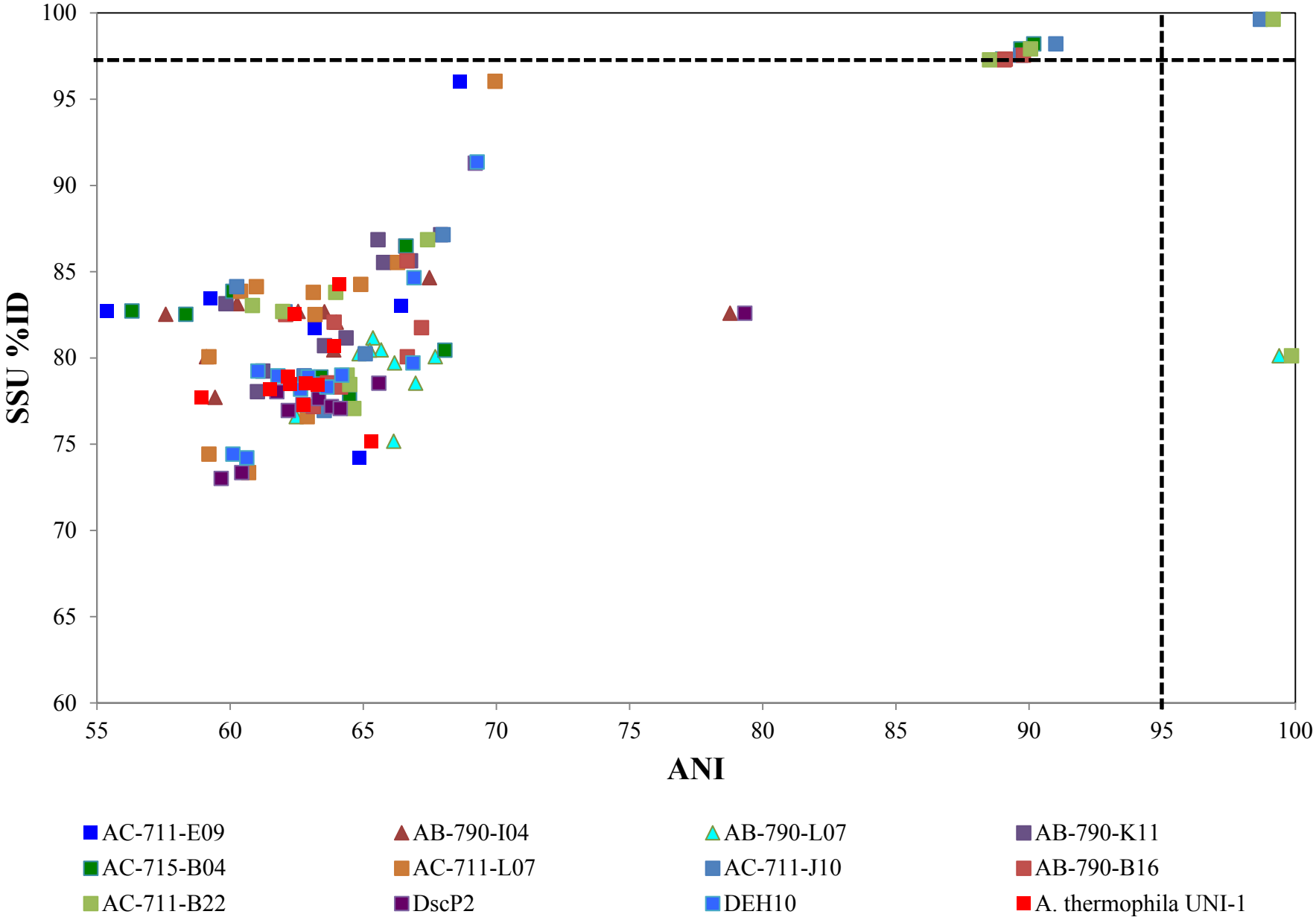


Figure S1. Pairwise comparisons of *Chloroflexi* genomes ANI vs. SSU percent identity. Dashed lines represent species level cutoff values.

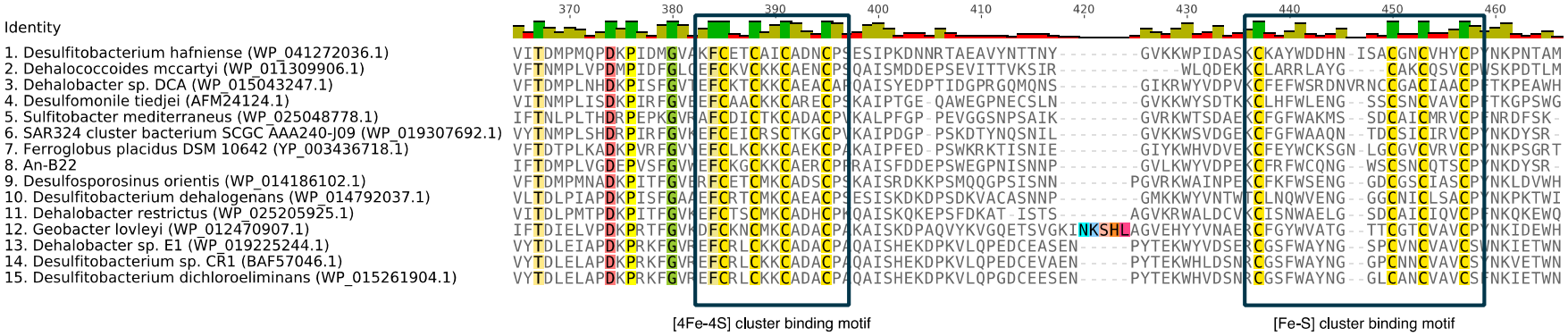


Figure S2. N-terminal amino acid alignment of RdhA sequences, including the one found in SAG An-B22 with boxes highlighting the Fe-S cluster binding motif. This is a subset of the amino acid alignment that was used to create Figure 4.