ENVIRONMENTAL GENOMICS REVEALS ABUNDANT AUTOTROPHIC AND SULFATE-REDUCING NITROSPIROTA BACTERIA INHABITING DEEP SUBSEAFLOOR CRUSTAL FLUIDS FROM THE JUAN DE FUCA RIDGE FLANK

A THESIS SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAIʻI AT MĀNOA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN

OCEANOGRAPHY

DECEMBER 2021

BY

CLARISSE ELEANOR SANTOS SULLIVAN

Thesis Committee:

Michael S. Rappé, Chairperson
Craig Nelson
Grieg Steward

Keywords: Nitrospirota, Juan de Fuca Ridge, subseafloor, metagenomics

ACKNOWLEDGEMENTS

We thank the captain and crew, Andrew Fisher, Keir Becker, Geoff Wheat, and other members of the science teams on board R/V Atlantis cruise AT18-07. We also thank the pilots and crew of remote-operated vehicle Jason II. We are grateful to Sean Jungbluth, Huei-Ting Lin, Chih-Chiang Hsieh, Alberto Robador, Brian Glazer, and Jim Cowen for sample collection and processing, and Beth Orcutt and Ramunas Stepanauskas for facilitating metagenome sequencing through the Department of Energy Joint Genome Institute Community Sequencing Program (Award 987 to R. Stepanauskas). This study used samples and data provided by the Integrated Ocean Drilling Program. This research was supported by funding from the National Science Foundation (grant OCE-1851582 to MSR), C-DEBI Graduate Student Fellowship to CES, and the University of Hawai‘i Department of Oceanography and Department of Life Science to CES.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Along the eastern flank of the Juan de Fuca Ridge (JdFR), boreholes drilled through sediment and into the oceanic crust access deep subseafloor fluids that support a unique microbiome. One group of bacteria that has been previously detected in high relative abundance in gene- and genome-based surveys of JdFR crustal fluids belong to the phylum Nitrospirota. In this study, metagenomes from crustal fluids collected from boreholes along the JdFR were used to recover Nitrospirota metagenome-assembled genomes (MAGs), which were subsequently interrogated by phylogenomics, metabolic reconstruction, and for population genetic characteristics. Nine Nitrospirota MAGs ranging in size from 0.60 mega base pairs (Mbp) to 2.34 Mbp were recovered, including five high quality genomes that were at least 95% complete. Genome phylogenies based on concatenated alignments of single-copy core genes placed the JdFR Nitrospirota into three distinct lineages within the Class Thermodesulfovibrionia, and revealed that the JdFR Nitrospirota are closely related to Nitrospirota MAGs previously recovered from sulfide deposits and solid substrates incubated within a JdFR borehole. Competitive recruitment of metagenome sequence reads revealed a single lineage dominated the Nitrospirota fraction of the crustal fluid microbiome. Metabolic reconstructions indicated the shared presence of genes involved in dissimilatory sulfate reduction, carbon fixation, and gluconeogenesis. Fine-scale genetic heterogeneity was investigated within two highly similar (>99% average nucleotide identity) Nitrospirota genomes recovered from separate boreholes along the JdFR flank. Read recruitment and whole genome comparisons revealed little intrapopulation variation, but subtle genetic discontinuity in the subsurface environment in JdFR.

Referenced supplementary material belonging to Appendix A include 8 tables and can be accessed in a separate excel file. The following are the supplementary table descriptions:

Table A1 Summary of sequencing, assembly, and genome binning from 2011 JdFR crustal fluid microbiomes.

Table A2. Quality of the JdFR Nitrospirota MAGs as assessed via the CheckM v1.1.2 pipeline and their Genome Taxonomy Database classification.

Table A3. Genome information of Nitrospirota whole-genome sequences and reference taxa used in the phylogenomic analysis.

Table A4. AAI estimates between Nitrospirota genomes recovered by Jungbluth et al. 2017a and this study.

Table A5. Distribution of KEGG-annotated genes across JdFR Nitrospirota MAGs.

Table A6. Detection of miscellaneous COGS within the Nitrospirota MAGs.

Table A7. ANI and number of non-identical bases (nucleotide variation) of bacterial family JdFR-88 genomes estimated using the ANIm method in pyANI v0.2.7.

Table A8. Summary of the top 2 BlastX v.2.12.0+ results of sequences with non-identical bases between JdFRnit3A and JdFRnit7B.

1. INTRODUCTION

For nearly two decades, boreholes drilled through sediment and into basement basalt along the eastern flank of the Juan de Fuca Ridge (JdFR) in the Northeast Pacific Ocean have been providing a conduit through which to explore microbial life inhabiting Earth's deep subseafloor biome (Cowen et al. 2003, Cowen 2004). Here, a lack of fluid exchange between the aquifer that percolates through the subseafloor basalt and overlying seawater enables the accumulation of hydrothermally-heated and altered fluids that harbor chemical species capable of fueling a subsurface biosphere (Lin et al. 2012; Wheat et al. 2010; Wheat et al. 2004). Accessing this biosphere is facilitated by subseafloor observatory systems installed in boreholes that provide conduits for the collection of basement fluids and the deployment of *in situ* autonomous samplers, experiments, and data logging systems (Wheat et al. 2011; Becker and Davis 2005). Known as Circulation Obviation Retrofit Kits or CORKs, some of these observatories are situated in spatially distinct boreholes that access pristine fluids at different states of chemical alteration along the JdFR flank, and offer the opportunity to access fluids at different depth horizons within the basaltic crustal aquifer (Fisher et al. 2011). The most recent generation of CORKs installed along the JdFR flank were designed with features intended for pristine fluid sampling efforts, including investigating the microbial life inhabiting this biome (Lin et al. 2012; Cowen et al. 2012). These observatories provide opportunities to elucidate the contributions of subsurface processes to global biogeochemical cycles, and answer fundamental questions regarding the evolution and persistence of microbial populations unique to this environment.

Previous research investigating the nature of microbial life inhabiting the basement biosphere along the JdFR flank has primarily focused on attached microorganisms colonizing

solid substrates incubated in the fluid flow path of the aquifer, and microbes inhabiting the fluids themselves. For example, FLow-through Osmotic Colonization Systems (FLOCS) deployed inside CORK observatories have utilized osmotically-driven pumps to pull borehole fluids into flow-through chambers containing mineral fragments to facilitate the microbial colonization of substrates (Orcutt et al. 2010). Experiments using FLOCS installed within JdFR CORKs have enabled the characterization of resident subseafloor biofilm-forming communities (Ramirez et al. 2019; Smith et al. 2017; Smith et al. 2011; Orcutt et al. 2011). Meanwhile, customized sampling instrumentation connected to CORK fluid delivery lines has facilitated the collection of pristine borehole fluids at the seafloor, as well as the recovery of microbial and viral biomass by directly passing fluids through filters *in situ* (Nigro et al. 2017; Jungbluth et. al 2016, 2013; Lin et. al 2012).

DNA sequence-based characterizations of microorganisms inhabiting the JdFR region of the oceanic deep subsurface have revealed that it is grossly underexplored (Cowen et al. 2003, Jungbluth et al. 2013, Jungbluth et al. 2016, Jungbluth et al. 2017a). Through surveys of functional and phylogenetic gene markers and environmental genomics, observations so far have uncovered a biome dominated by uncultivated, deeply branching lineages of life that harbour ancient homologs of enzymes involved in key metabolic pathways such as methane cycling and sulfate reduction thought to be important to Earth's early microbial inhabitants, and important to microorganisms inhabiting anaerobic habitats at present (e.g. Robador et al. 2015, Anantharaman et al. 2018, Jungbluth et al. 2017a, Jungbluth et al. 2017b, Boyd et al. 2019, Carr et al. 2019, Fincker et al. 2020). For example, genomes of the novel uncultivated archaeal lineage Hydrothermoarchaeota retrieved from JdFR crustal fluids through metagenomic binning and single-cell genomics were resolved as an early-branching lineage based on their evolutionary

placement between the Euryarchaeota and DPANN, and the presence of divergent forms of sulfate and nitrate reductases for use as terminal electron acceptors for energy conservation (Carr et al. 2018; Anantharaman et. al 2018). The Hydrothermoarchaeota genomes also potentially link sulfate reduction to dissimilatory carbon monoxide oxidation to maximize energy yield, which hint to their potential role as chemolithoautotrophs in the subseafloor fluid ecosystem (Carr et al. 2018). In another study, the genome of *Ca. Polytropus marinifundus*, a novel member of the Archaeoglobi within the Euryarchaeota, was reconstructed from a JdFR crustal fluid metagenome and found to be the first Archaeoglobi to encode genes of the methyl-coenzyme M reductase (MCR) complex, a key enzyme in archaeal methanogenesis (Boyd et al. 2018). Genome predictions revealed the possible coupling of the MCR complex to alkane ß-oxidation, which is energetically favorable when linked to genes that utilize nitrate, iron or oxidized sulfur compounds as terminal electron acceptors. The genome of *Ca. Polytropus marinifundus* encodes for other metabolic pathways including glycolysis and amino acid and organic acid fermentation, underscoring the likely heterotrophic strategies (as opposed to methanogenesis) employed by this lineage to survive in this environment (Boyd et al. 2018).

Members of the bacterial phylum Nitrospirota have been repeatedly recovered in gene- and genome-based surveys of subseafloor crustal fluids recovered from the flank of the JdFR, where they have sometimes formed the most abundant bacterial group present (Jungbluth et al. 2016, 2017). The Nitrospirota contains three major lineages of cultivated microorganisms including the chemolithoautotrophic nitrite-oxidizing genus *Nitrospira*, the iron-oxidizing genus *Leptospirillum*, and the sulfate-reducing genus *Thermodesulfovibrio* (Daims, 2014). However, ribosomal RNA and environmental genomic sequence data have shown that diverse Nitrospirota lineages of poorly defined metabolism are widespread in natural systems (e.g. Lin et al. 2014;

Matsuura et al. 2016; Bhatnagar et al. 2015; Zecchin et al. 2017; Kato et al. 2018; Parks et al. 2017; Woodcroft et al. 2018; Meier et al. 2019).

In this study, a genome-centric metagenomic approach was used to comprehensively describe the potential metabolic attributes, evolutionary history, and microevolutionary characteristics of members of the bacterial phylum Nitrospirota inhabiting the basalt-hosted deep subseafloor of the JdFR flank in the Northeast Pacific Ocean. Our study reveals new insights regarding the evolution and persistence of these microorganisms in the deep subseafloor environment, as well as illuminates their role in the cycling of organic carbon and nutrients within Earth's largest subseafloor aquifer.

2. MATERIALS AND METHODS

2.1. Basement fluid sampling

In July 2011, deep subseafloor basement crustal fluids were collected from boreholes fitted with CORKs along the Juan de Fuca Ridge flank during cruise AT18-07 aboard the R/V *Atlantis*, as previously described (Jungbluth et al. 2016; Jungbluth et al. 2017a). Briefly, fluid samples were collected from Integrated Ocean Drilling Program (IODP) boreholes U1362A (47°45.6628'N, 127°45.6720'W) and U1362B (47°45.4997'N, 127°45.7312'W) (Figure B1), which are located at a water column depth of approximately 2,650 m and penetrate through approximately 240 m of sediment and into the igneous basement. The CORKs fitted to these boreholes feature polytetrafluoroethylene-lined fluid delivery lines that access fluids at different depth horizons (Fisher et al. 2011). Fluids from hole U1362B were sampled from a single isolated horizon at 30 to 117 m below the sediment-basement interface, or meters subbasement (msb), while fluids from borehole U1362A were sampled from a horizon at 193 to 292 msb.

Custom sampling equipment enabled the pumping of borehole fluids directly through 0.22 μm-pore-size Steripak-GP20 polyethersulfone filter cartridges (Millipore, Billerica, MA, USA) at the seafloor (Cowen et al. 2012; Lin et al. 2020). Approximately 124 liters and 70 liters of basement fluid were filtered from U1362A and U1362B, respectively (Jungbluth et al. 2016).

2.2. Metagenome assembly and binning

Genomic DNA was extracted and sequenced from microbial biomass collected from U1362A and U1362B crustal fluid samples as previously described (Jungbluth et al. 2016; Jungbluth et al. 2017a). Briefly, environmental DNA was extracted from the Steripak-GP20 filters using phenol-chloroform, and sequenced on an Illumina HiSeq2000 platform (2 × 150 bp reads) at the Department of Energy's Joint Genome Institute (JGI) (Jungbluth et al. 2016; Jungbluth et al. 2017a). A total of 296 and 162 million sequence reads were generated from the U1362A and U1362B libraries, respectively.

Metagenomic reads from U1362A and U1362B basement fluids were quality trimmed and separately assembled into scaffolds, which were visualized and binned using anvi'o v.5.5 (Eren et al. 2015) following a previously described metagenomic workflow (Eren et al. 2015). Illumina adapters and low-quality reads were removed using illumina-utils v1.4.6 (Eren et al., 2013), and corrected using metaSPades v3.12.0 (Nurk et al. 2017) to remove sequencing artifacts and errors. The quality-controlled sequences were assembled into scaffolds using a de Bruijn reconstruction and default parameters in metaSPades v3.12.0 (Nurk et al. 2017). A scaffold database was subsequently generated for each of the two samples from the assembled metagenomic data using '--anvi-gen-contigs-database', which stored scaffolds as 20 kb sequences or splits. Open reading frames were identified using Prodigal v2.6.3 (Hyatt et al.

2010), and kmer frequencies were quantified for each scaffold and split. The artificial splitting of scaffolds increases the resolution of information displayed during downstream interactive visualizations of genomic data (Eren et al. 2015). Hidden Markov model (HMM) profiles of single-copy bacterial (Campbell et al. 2013) and archaeal (Rinke et al. 2013) genes were identified with HMMER v3.2.1 (Eddy 2011) using the '--anvi-run-hmms' program, and stored as part of each scaffold database. The presence and absence of these genes in the scaffolds provided real-time estimates of genome completeness, redundancy and single-copy core gene (SCG) domain identities during downstream genome binning and refinement with the anvi'o platform. Metagenome sequence data was visualized with the anvi'o interactive interface by creating a sample profile using '--anvi-profile', which hierarchically clustered scaffolds according to their tetranucleotide frequency. The '--anvi-interactive' program was used to visualize and manually bin clustered sequences into metagenome-assembled genomes (MAGs) according to their GC content, mean coverage, estimated percent completion and redundancy, and assigned SCG domain identity. Genome bins without an assigned single-copy core gene domain in anvi'o likely originated from sequences of plasmids or viruses, or from extremely low completion sequences and were not included in the genome bin count (Table A1) and downstream analysis.

MAGs estimated to be >50% complete and <10% redundant based on the single-copy core gene profiles during the manual binning step were taxonomically assigned using GTDB-Tk v0.3.2 (Chaumeil et al. 2019). Genomes identified as belonging to the bacterial phylum Nitrospirota were manually curated using the '--anvi-refine' program that allows users to visualize sequences of individual bins at a higher resolution. Sequences that exhibited a divergent sequence composition, such as tetranucleotide frequency and GC content, within an existing bin were removed. The manually curated Nitrospirota genomes were then again assessed

for completion and contamination using '--anvi-summarize' and the CheckM v1.1.2 (Parks et al. 2015) lineage-specific workflow.

2.3. Phylogenomics

A phylogenomic tree was reconstructed from concatenated alignments of the JdFR Nitrospirota MAGs with publicly available Nitrospirota whole-genome sequences retrieved from the National Center for Biotechnology Information (NCBI) and European Nucleotide Archive (ENA) using a suite of programs in GToTree (Lee 2019). Nitrospirota genomes with CheckM completeness and contamination values of >50% and <10%, respectively, in the Genome Taxonomy Database (Chaumeil et al. 2019) were obtained using ncbi-genome-download v2.11 (https://github.com/kblin/ncbi-genome-download/). GToTree was run with the -T IQ-TREE and -G 0.28 parameters and utilized GToTree's bacterial SCG-set in order to generate a maximum-likelihood (ML) tree from concatenated alignments of 21single-copy bacterial marker genes using IQ-TREE v1.6.9 (Nguyen et al. 2015), which used the LG+F+R6 best fit model (Kalyaanamoorthy et al. 2017) and employed UFBoot2 (Hoang et al 2017) to calculate bootstrap values from 1000 bootstrap replicates. The final ML tree was rooted and viewed with FigTree v1.4.4 (Rambaut 2014). Percent amino acid identities (AAI) were calculated in CompareM v0.0.23 (https://github.com/dparks1134/CompareM). Average nucleotide identities were calculated using the ANIm method (Richter et al. 2009) implemented in pyANI v0.2.7 (Pritchard et al. 2015). Percent AAI between whole genome sequences in the class Thermodesulfovibrionia were formatted into a matrix data frame in R using tidyr v1.0.2 (Wickham et al. 2020). The matrix data frame was uploaded into the Morpheus online analysis software (Gould 2016) to hierarchically cluster and visualize the AAI data as a heatmap (Figure B2). The AAI percentages

clustered using the Euclidian clustering and complete linkage method in Morpheus were recursively merged according to their pair-wise distance to generate dendrograms.

2.4. Competitive recruitment of the non-redundant JdFR Nitrospirota

Metagenomic reads from U1362A and U1362B were recruited to non-redundant JdFR Nitrospirota. The survey of high sequence similarities between pairs of the JdFR Nitrospirota using the >99% ANI threshold identified two redundant genomes in the JdFR-88 family. Retaining multiple genomic representatives during competitive recruitment can underestimate relative abundance calculations due to reads being randomly distributed across redundant reference genomes (Evans & Denef 2020). Thus, JdFRnit7B was selected as the representative genome for the JdFR-88 family for this analysis. Genome sequences of the 8 JdFR Nitrospirota were merged into one FASTA file, which was used to recruit reads from 4 metagenomes using bowtie2 v2.3.5 (Langmead & Salzberg 2012). The estimated abundance of the JdFR Nitrospirota is given as a proportion of mapped reads, which is summarized in Figure B3.

2.5. Metabolic reconstruction and functional characteristics

JdFR MAGs within the phylum Nitrospirota were functionally annotated using the JGI Integrated Microbial Genomes Annotation Pipeline v5.0.3 (Hunteman et al. et al. 2015). Protein coding genes were identified using Prodigal v2.6.3 (Hyatt et al. 2010), which were assigned COG, Pfam and TIGRfam assignments using hmmsearch from the HMMERv3.1b2 package (Eddy 2011). KEGG Orthology (KO) terms were determined using last-align 983 (Kielbasa et al. 2011) and assigned using IMG-NR v20190607, a reference database comprised of isolate genomes and trusted single-cell genomes in IMG. In addition, protein coding sequences exported from the anvi'o scaffold database of the Nitrospirota genomes were assigned KO terms through

sequence similarity searches against the prokaryote KEGG database using GhostKOALA v2.2 (Kanehisa et al. 2016). The KEGG annotations from both sources were merged, formatted and exported as a csv file using R for use in KEGG-Decoder v1.1, which parsed through the KO terms to determine the completeness of canonical KEGG pathways in a given genome (Graham et al. 2018). This study used a modified KEGG-Decoder v1.1 script to also include KO terms attributed to the pentose-phosphate pathway, Wood-Ljungdahl pathway methyltransferases (*acsCDE*), glycogen metabolism (*glgBXCAP*), RNF complex, quinone-modifying oxidoreductases (*qmoABC*), fatty-acid metabolism, and sulfate, nitrate, ammonium, trace metal and macromolecule (lipooligosaccharide, lipopolysaccharide, lipoprotein) transporters to supplement the original KEGG-Decoder script.

2.6. Single nucleotide variation

Anvi'o v6.1 was used to profile and identify single-nucleotide variants (SNVs) between the closely related JdFR Nitrospirota MAGs JdFRnit3A (borehole U1362A) and JdFRNit7B (borehole U1362B) reconstructed in this study, and GCA_002376445 and GCA_002376155 previously reconstructed from metagenomes retrieved from substrates incubated inside a FLOCS apparatus within JdFR CORK U1301A (Smith et. al 2011, 2017; Parks et al. 2017). Quality filtered reads from the U1362A and U1362B metagenomes, and from the two metagenomes sequenced from microorganisms that colonized volcanic glass (SRX742692) or anorthite and bytownite (SRX742693) that was incubated in U1301A were mapped back to JdFRnit3A, JdFRNit7B, GCA_002376445, and GCA_002376155 using bowtie2 v2.3.5 (Langmead & Salzberg 2012). Resulting mapping files or BAM files were sorted and indexed using samtools v1.9 (Li et al. 2009). A sequence database was then generated for each of the four genomes using

'--anvi-gen-contigs-database' using default parameters. The mapped short read data contained in the indexed BAM files and bin sequence data stored in the sequence database were profiled in anvi'o to identify SNVs among reads (Eren et al. 2015). The SNV positions of recruited metagenomic reads were reported using default parameters in '--anvi-gen-variability-profile'.

Whole-genome alignments using the MCM algorithm of Mauve v1.1.1 (Darling et al. 2010) identified matching pairs of scaffolds from JdFR Nitrospirota MAGs JdFRnit3A and JdFRNit7B. Individual pairwise alignments of scaffold pairs were performed using MAFFT v1.4.0 (Katoh & Standley 2013) to detect nucleotide differences among JdFRnit3A and JdFRNit7B. Nucleotide differences and SNV positions detected in anvi'o were manually inspected using IGV v2.82 (Thorvaldsdóttir et al. 2012; Robinson et al. 2011). Genome regions with confirmed nucleotide differences between JdFRnit3A and JdFRnit7B were inspected in anvi'o using the anvi-interactive program. Gene sequences harboring those non-identical nucleotides were searched against the GenBank non-redundant protein database using BLASTX v 2.12.0+ (Altschul et al. 1997).

3. Results and Discussion

3.1. Recovery of JdFR Nitrospirota MAGs

Assembly of two metagenomes sequenced from crustal fluids of the eastern flank of the Juan de Fuca Ridge collected from boreholes U1362A and U1362B in 2011 (Jungbluth et al. 2016, 2017) resulted in 19844 and 26440 scaffolds of over 1500 bp in length, totaling 171 and 168 Mbp of unique sequence (Table A1). A total of 90 metagenome assembled genomes (MAGs) of >50% completion and <10% contamination were recovered from the two assemblies, including 51 Bacteria and 39 Archaea (Table A1).

Initial taxonomic identification via the Genome Taxonomy Database Tool Kit (GTDB-Tk; Chaumeil et al. 2019) revealed nine MAGs affiliated with the bacterial phylum Nitrospirota, including seven from borehole U1362A and two from U1362B (Table A1). The MAGs ranged in size from 0.734 to 2.34 Mbp, and estimated completion from and 43.9% to 100% (Table 1). Five Nitrospirota genomes with an estimated completion of >90% and size range of 1.86 to 2.34 Mbp were recovered (Table A1). Contamination estimates ranged from 0 to 1.82% while strain heterogeneity approximations were 0% for all of the Nitrospirota MAGs (Table A1), suggesting that any detected contamination was due to the presence of genomic fragments from more divergent taxa (Parks et al. 2015). Two Nitrospirota MAG genomes contained a GC content of 62.7% (Appendix A1), while the remaining seven ranged from 40.3 to 48.8% (Table A1).

When compared to Nitrospirota genomes previously recovered from the same U1362A and U1362B CORK fluid metagenomic samples (Jungbluth et al. 2017), the Nitrospirota MAGs recovered here were of similar size and GC content, but had improved genome quality. Using an automated binning and manual curation approach, Jungbluth and colleagues recovered eight Nitrospirota MAGs ranging in size from 0.74 to 2.33 Mbp, and included two genomes, JdFR-87 (GCA_002010755) and JdFR-88 (GCA_002011795), with elevated GC content (62.5 and 62.8%). The six remaining Nitrospirota MAGs recovered by Jungbluth and colleagues contained GC values ranging from 41.4 to 48.0%, which is similar to those of the remaining seven genomes assembled in the current study. An exception is JdFR-84 (GCA_002010775), which had a comparatively low GC content of 39.9%. Using CheckM to estimate completion and contamination, the MAGs assembled in the current study are of enhanced quality compared to the previous genomes of Jungbluth and colleagues (Jungbluth et al. 2017). While Jungbluth and colleagues used CheckM v1.0.5 versus v1.1.2 used in the current study, there are no substantive

changes with how genome completion, contamination and strain heterogeneity are computed between the versions (Parks et al. 2015).

Table1. Characteristics of JdFR MAGs related to the Nitrospirota recovered in this study.

Genome[1]	Completion (%)	Contamination (%)	Size (Mbp)	N50 (Kbp)	%GC	Scaffolds (#)	Quality[2]
JdFRnit1A	98.2	0.23	2.17	255	45.1	11	High
JdFRnit2A	94.6	1.82	1.89	42	48.8	74	High
JdFRnit3A	100	1.82	1.86	463	62.7	8	High
JdFRnit4A	85.9	1.09	2.18	61	40.3	49	Medium
JdFRnit5A	51.8	0.00	0.60	22	42.6	36	Medium
JdFRnit8A	59.6	0.91	1.14	22	41.5	91	Medium
JdFRnit9A	43.9	1.75	0.73	14	41.1	78	Low
JdFRnit6B	98.2	1.82	2.34	121	41.6	28	High
JdFRnit7B	100	1.82	1.86	463	62.7	7	High

[1] MAGs recovered from U1362A (A) and U1362B (B) are indicated

[2] Based on the minimum information about metagenome-assembled genome (MIMAG) standards (Bowers et al. 2017)

3.2. Phylogenomic analysis of JdFR Nitrospirota MAGs

Using a concatenated alignment of 21 single-copy core genes, a phylogenomic analysis of the JdFR Nitrospirota MAGs and reference taxa from across the phylum revealed that the nine genomes partitioned into four distinct lineages spanning three families across two orders within

the class Thermodesulfovibrionia (Figure1). Two different Nitrospirota families contained MAGs originating from each of the two boreholes. Within the family JdFR-88, MAGs JdFRnit3A and JdFRnit7B were closely related to each other, as well as to two MAGs previously recovered from solid-substrate incubations within CORK U1301A along the JdFR flank (Smith et al. 2011). Collectively, the four genomes shared >99% AAI (Figure B2). JdFRnit1A diverged recently from the genomes within family JdFR-88 (Figure1), but share an AAI of only 59.7% with JdFRnit3A and JdFRnit7B. A second lineage that included MAGs from both boreholes consisted of JdFRnit4A and JdFRnit6B in the family JdFR-85 (Figure1). JdFRnit4A and JdFRnit6B shared an AAI of 65.2%, and AAI values ranging from 58.8 – 65.5% with five MAGs recovered from a core sample of sulfide deposits from the Suiyo Sea Mount (Kato et al. 2018; Kato et al. 2015) (Table A2). The remaining four JdFR Nitrospirota MAGs were from borehole U1362A, and formed two related lineages within the order UBA6902 (Figure1). These include a lineage consisting of JdFRnit2A and JdFRnit5A (77.8% AAI), and a one consisting of JdFRnit8A and JdFRnit9A (89.5% AAI) (Table A2). These two lineages also shared a common evolutionary history with MAGs from sulfide deposits from the same study described above (Kato et al. 2018). AAI values across this lineage ranged from 65-70% (Table A2).

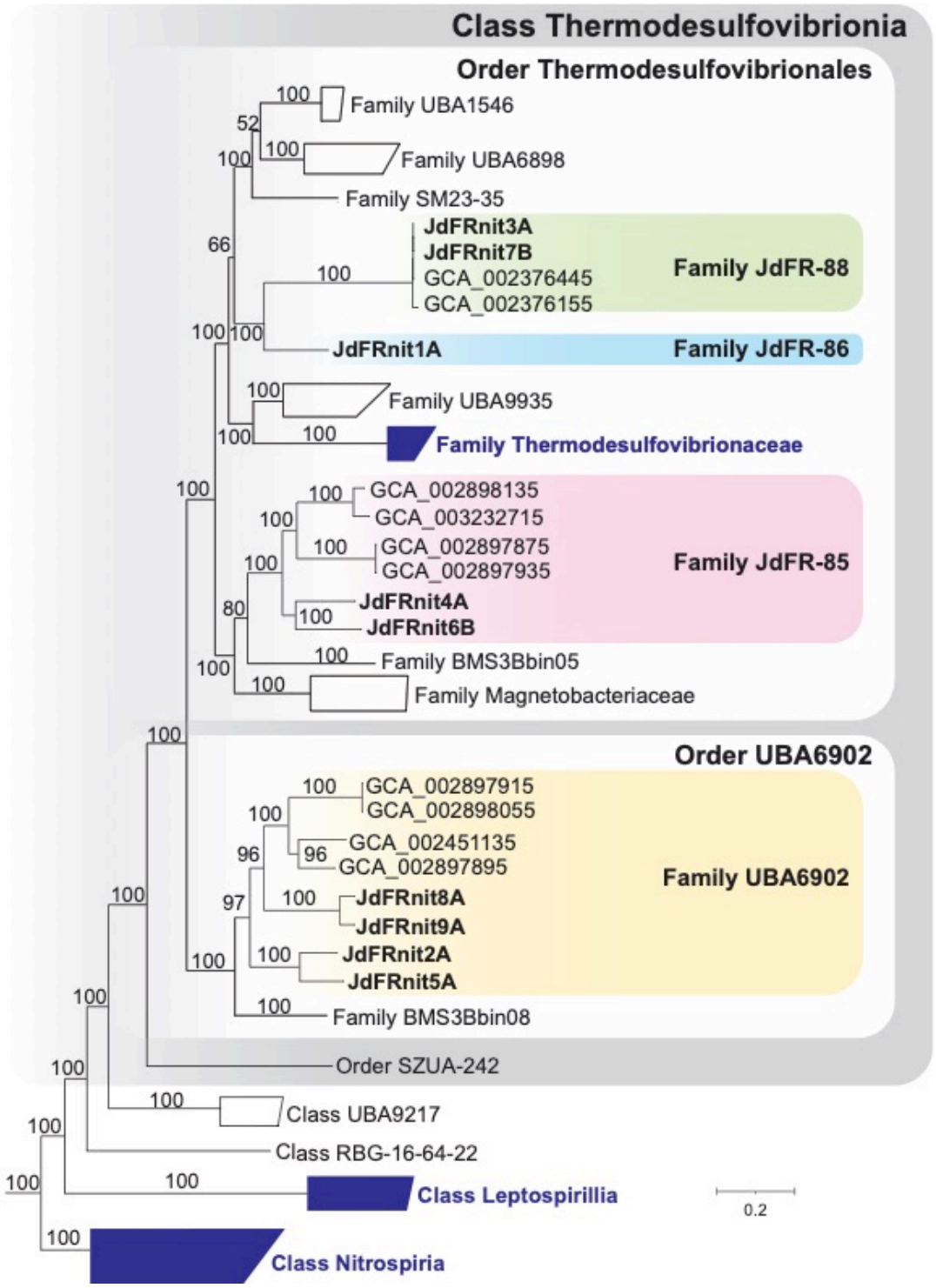

Figure 1. Phylogenomic reconstruction of the Nitrospirota phylum based on GToTree's bacterial single-copy core gene set (Lee 2019). Cultivated reference genomes are shown in dark blue. JdFR Nitrospirota MAGs are indicated in bold. The four colored boxes further group the JdFR Nitrospirota according to their family-level classification. Taxonomic classification based on the Genome Taxonomy Database (Chaumeil et al. 2019) are indicated in bold.

3.3. Competitive recruitment of the JdFR Nitrospirota in the U1362A and U1362B metagenomes

Read recruitment from U1362A and U1362B metagenomes to the non-redundant JdFR Nitrospirota MAGs indicate that the JdFR-88 lineage dominated the subseafloor environment relative to the other Nitrospirota lineages recovered in this study. High sequence similarities between pairs of the JdFR Nitrospirota determined using the >99% ANI threshold identified two redundant genomes in the JdFR-88 family. Retaining multiple genomic representatives during competitive recruitment can underestimate the relative abundance of individual genomes due to reads being randomly distributed across redundant reference genomes (Evans & Denef 2020). Thus, JdFRnit7B was selected as the representative genome for the JdFR-88 family. All JdFR Nitrospirota excluding JdFRnit3A were used as non-redundant reference genomes for the competitive recruitment of U1362A and U1362B reads. Their relative abundance in each metagenomic sample was estimated as a proportion of mapped reads (Table A8). MAG JdFRnit7B within the JdFR-88 lineage accounted for 22% of the mapped reads from U1362A and 33% from U1362B (Table A8), which is similar to a previous estimate of the abundance of Nitrospirota MAGs in the same metagenomes (Jungbluth et al. 2017). The abundances of the remaining genomes ranged from 0.3-2.1% in U1362A and 0-1.0% in U1362B, suggesting that the JdFR-88 Nitrospirota lineage likely dominates the basalt-hosted deep subsurface fluids of the JdFR compared to the other Nitrospirota lineages recovered as MAGs.

3.4. Metabolic potential of JdFR Nitrospirota

3.4.1. Carbon metabolism

Genomes within all four JdFR Nitrospirota lineages possessed a complete or nearly complete set of genes for carbon fixation via the Wood-Ljungdahl pathway (WL),

gluconeogenesis, and glycogen synthesis and degradation (Figure 2). This included the complete gene set for both the carbonyl and methyl branches of the WL pathway (Figure 2, Table A5). Additionally, investigations of Nitrospirota genomes phylogenomically characterized in this study reveal that the WL pathway is prevalent in class Thermodesulfovibrionia members (Figure 1). For example, MAGs recovered from sulfide deposits (Kato et al. 2018), which have representatives from families JdFR-85, UBA6902, BMS3Bbin05 and BMS3Bbin08 (Figure 1 and Table A3), have the genes for CO_2 fixation via the methyl and carbonyl branches of the WL pathway. Other examples of Thermodesulfovibrionia representatives that feature the WL pathway include MAGs recovered from hydrothermal sediments (Zhou et al. 2020), anoxic bioreactors (Arshad et al. 2017), and terrestrial subsurface fluids (Hernsdorf et al. 2017) (Table A3). In addition, cultivated representatives from family Thermodesulfovibrionaceae (Figure 1 and Table A3) previously recovered from activated sludge (Matsuura et al. 2016, Mukherjee et al. 2017), and freshwater hydrothermal fluids (Bhatnagar et al. 2015) (Table A3) were also shown to harbor the genes for the carbonyl branch of the WL pathway (Adam et al. 2018).

However, this ancestral pathway is found in previously isolated methanogenic archaea (Thaur et al 2008; Ladapo & Whitman 1990), acetogenic bacteria (Espositio et al. 2019; Schuchmann & Muller 2014; Drake et al. 2002), and sulfate reducing archaea (Klenk et al. 1998; Vornolt et al. 1995) and bacteria (Agostino et al 2020; Gittel et al. 2010; Schauder et al 1988). The WL pathway has also been characterized in uncultivated microbial lineages that are adapted to extreme environments. These include novel Clostridia genomes from JdFR CORK-incubated substrates (Ramirez et al. 2019; Smith et al. 2019) and from hypersaline soda lake sediments (Vavourakis et al. 2018), Actinobacteria from hotspring sediments (Jiao et al. 2021) and from serpentinite-hosted fluids (Merino et al. 2019), Hydrothermarchaeota (Carr et al. 2018;

Anantharaman et al. 2018) and Archaeoglobi (Boyd et al. 2018) from JdFR crustal fluids, Bathyarchaeota from hydrothermal vent sediments (He et al. 2016) and Lokiarchaeota from hemipelagic-glaciomarine sediments (Spang et al. 2015; Sousa et al. 2016). The pervasiveness of this pathway in anaerobic bacteria and archaea is likely due to its low ATP requirement, ability to couple with energy conservation, and diversity of usable coenzymes and electron carriers (Berg et al. 2010; Esposito et al. 2019).

All four JdFR Nitrospirota lineages possessed the potential for the conversion of pyruvate to glucose-6-phosphate (G6P) via gluconeogenesis (Figure 2). While none of the JdFR Nitrospirota possessed glucose-6-phosphatase (G6PC) needed for the conversion of G6P to glucose in the last step of gluconeogenesis (Cohen 2011), all of the lineages are likely able to convert G6P to glycogen (Figure 2). Glycogen can function as an energy reserve (Wang et al. 2020) that benefits bacteria under nutrient-poor and fluctuating conditions (Sekar et al. 2020) such as those that may be expected in the basement basalt of the JdFR flank. The potential consumption of glycogen by JdFR Nitrospirota is supported by the presence of glycogen debranching enzymes and phosphoglucomutase *pgm* (Table A5 and A6) that remake G6P. Three of four JdFR Nitrospirota lineages (JdFR-88, JdFR-86, and JdFR-85) can subsequently utilize the glycolytic Emden-Meyerhof-Parnas (EMP) pathway for G6P breakdown to pyruvate (Figure 2, Figure B4). In addition, the absence of phosphogluconolactonase (*pgl*) (Figure B4) in the JdFR-88 lineage in the glycolytic pentose phosphate (PP) pathway suggests that the most abundant lineage of Nitrospirota in the JdFR subseafloor aquifer likely breaks down G6P for energy via the EMP pathway alone.

These findings suggest that the JdFR Nitrospirota likely employ a chemolithoautotrophic strategy for growth and survival. The WL pathway converts carbon dioxide to acetyl-CoA,

which can subsequently be transformed into G6P via the gluconeogenic pathway as proposed previously (Okabe et al. 2020). When energetic sources required to drive carbon fixation are in excess, this sugar can then be stored as glycogen and consumed as needed to maintain cellular metabolism. Further research is needed to determine if the JdFR Nitrospirota are capable of simultaneously generating and consuming glycogen as in the annamox bacterium *Ca. Brocsdia sinica* (Okabe et al. 2020).

3.4.2. Sulfur and nitrogen cycling

Genes associated with both dissimilatory sulfate reduction and sulfide oxidation were evident in all four of the JdFR Nitrospirota lineages (Figure 2, Figure. 3). All the JdFR Nitrospirota lineages possess a complete pathway for dissimilatory sulfate reduction (DSR), and all lineages contain genes encoding for sulfate permeases and potential sulfate transporters (*tusA, YedE/YeeE*) (Table A5 and A6). The genomes of all four lineages also encode for the *sqr* gene, which transforms sulfide to elemental sulfur (S^0) within the periplasm of the cell (Figure 2, Figure 3). The capacity for DSR and sulfide oxidation was evident in previously recovered Nitrospirota genomes from hydrothermally-heated sediment (Zhou et al. 2020), hydrothermal sulfide deposits (Kato et al. 2018), anoxic bioreactors (Arshad et al. 2017), and terrestrial subsurface fluids and sediment (Anantharaman et al. 2016; Anatharaman et al. 2018), emphasizing the widespread role of Nitrospirota in sulfur cycling in a diversity of environments that includes the deep subseafloor basement of the JdFR. Additionally, previously recovered Hydrothermarchaeota MAGs from JdFR crustal fluids (Jungbluth et al. 2017a) featured novel and early-evolved *dsrAB* genes and CO_2-fixation genes via the WL pathway, (Anantharaman et al. 2018) that suggests their capacity to utilize sulfate as an electron carrier and to grow

autotrophically in the deep subsurface fluids of the JdFR. In another study, *dsrB* genes were characterized in mesophilic bacteria from borehole 1025C and in thermophilic archaea from CORK U1301A in the JdFR (Robador et al. 2015). Furthermore, Robador et al. (2015) demonstrated that the sulfate reduction rates of thermophilic microorganisms from U1301A were stimulated by the addition of short-chain organic acids suggesting that organotrophic sulfate reducers also play an important role in sulfur cycling in the JdFR aquifer.

Thiosulfate disproportionation via the *sox* pathway, which appears to be a shared feature among some Nitrospirota relatives (Zhou et al. 2020; Kato et al. 2018; Anantharaman et al. 2016), was found to be unique to the JdFR-85 lineage (Figure B5). We suspect that it is not likely that thiosulfate disproportionation plays a major role in sulfur cycling in the JdFR subsurface aquifer as this metabolic function is absent from the majority of Nitrospirota genomes present in this system, including the most dominant Nitrospirota lineage, JdFR-88 (Figure 2, Figure 3).

Ammonia assimilation was evident in all four Nitrospirota lineages (Figure 2). The presence of ammonia importers and the absence of ammonia oxidizing genes (*amoABC*) within the four JdFR Nitrospirota lineages indicate that ammonia is likely limited to assimilation into biomass and not utilized in energy conserving pathways (Figure 2). The capacity for the assimilatory reduction of nitrate was unique to the JdFR-88 lineage, and included a membrane-bound nitrate reductase (*nar*) and cytoplasmic nitrite reductase (*nirB*), despite the absence of nitrate and nitrite transporters (Figure B5).

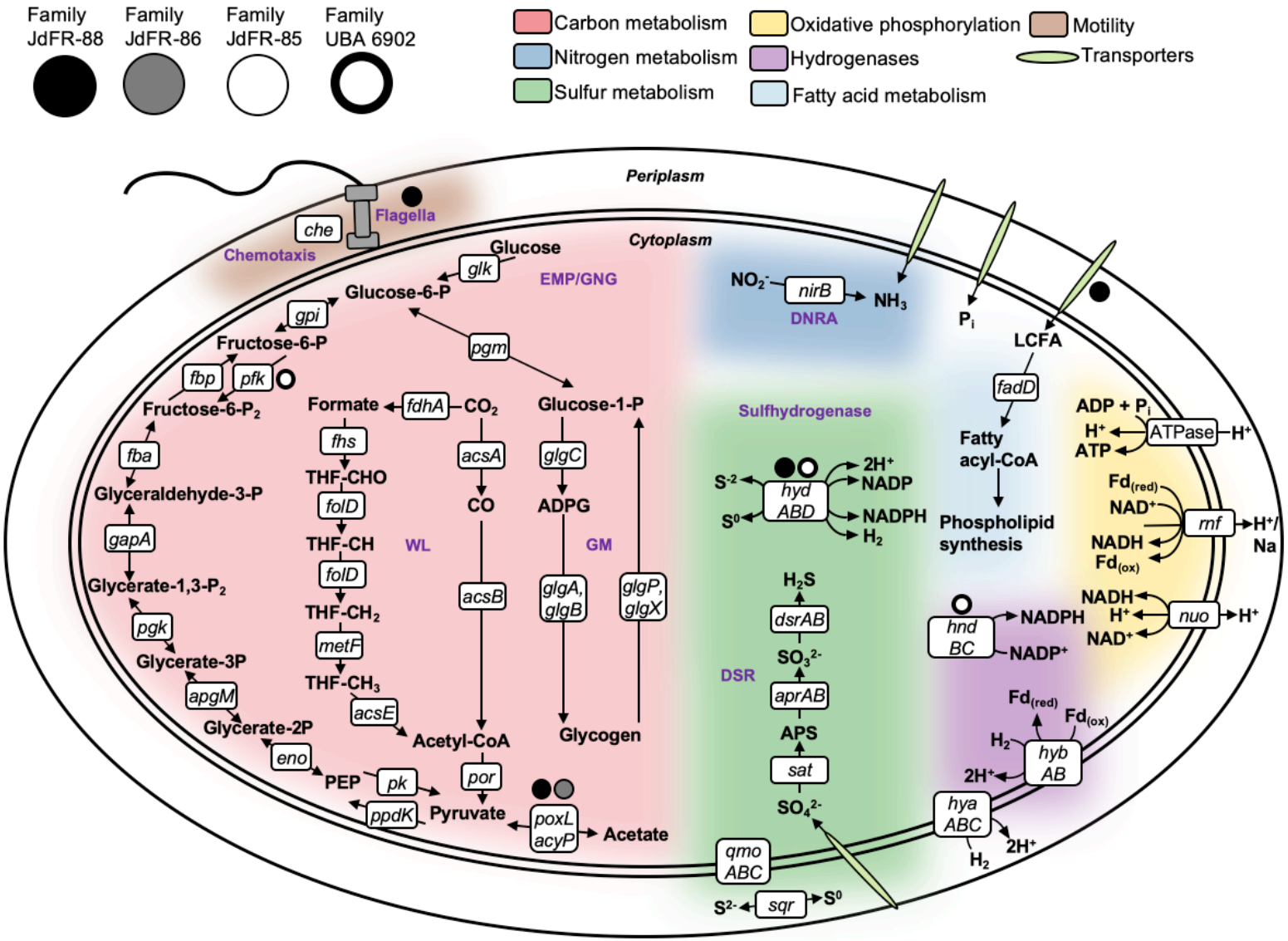

Figure 2. Metabolic reconstruction of KEGG pathways in the JdFR Nitrospirota lineages. Genes are shown as white squares. Transporters are displayed as green ovals. Transporter gene names are listed in Table A6 and A7. Circles adjacent to gene boxes indicate JdFR Nitrospirota lineages that are missing the indicated gene. EMP: Emden-Meyerhof-Parnas, GNG: Gluconeogenesis, GM: Glycogen Metabolism, WL: Wood-Ljungdahl, DNRA: Dissimilatory Nitrate Reduction to Ammonia, DSR: Dissimilatory sulfate reduction. The nickel-hydrogenase genes (*hybAB*) and the glycogen debranching gene (*glgX*) in the GM pathway were identified using COG (Table A7).

Figure 3. Summary of KEGG functions involved in carbon, nitrogen and sulfur cycling, oxidative phosphorylation, hydrogen redox reactions, vitamin biosynthesis, fatty acid metabolism, motility, and biofilm formation. Pathway completion is indicated by circle size and functional categories are indicated by color.

3.4.3. Hydrogenases and energy complexes

Nickel-Iron (NiFe) hydrogenases, energy complexes (Rnf and Nuo) and F-type ATPases were evident in all of the JdFR Nitrospirota lineages (Figure 2). The membrane-bound NiFe hydrogenase *hyb* (Figure 2, Table A6), which oxidizes hydrogen gas (H_2) within the cytoplasm (Teng et al. 2019), and *hya* (Figure 2, Table A5), which oxidizes H_2 in the periplasm were found in all of the JdFR Nitrospirota lineages from this study. While other NiFe hydrogenases were detected in the JdFR Nitrospirota genomes, none possessed the gene set required to complete the *hyd, hox*, or *hnd* hydrogenase groups (Table A5).

The Rnf complex, which was identified in all JdFR Nitrospirota lineages, energetically links cellular pools of reduced ferredoxin and NAD+ coupled with the generation of a transmembrane gradient (Figure 3), and is critical for autotrophic growth of microorganisms that exploit the WL pathway (Westphal et al 2018). Electrons transferred by reduced ferredoxin and NADH, which is generated by the Rnf complex, power the fixation of CO_2 in the carbonyl and methyl branches of the WL pathway. The presence of the membrane bound NiFe hydrogenase HybAB likely replenish the reduced ferredoxin pools (Figure2) (Teng et al. 2019) needed for the carbonyl branch of the pathway and facilitate CO_2 fixation (Westphal et al. 2018) in all 4 JdFR Nitrospirota families. All of the JdFR Nitrospirota lineages in this study harbor the Nuo complex, which couples the transfer of electrons between NADH and quinone to proton translocation (Berrisford et al. 2016), as well as F-type ATPases, which generate ATP through the translocation of H^+ protons derived from H_2 oxidation and the splitting of H^+ protons from reductants used in chemolithoautotrophy.

3.4.4. Mixed acid fermentation

The potential for acetate fermentation was detected in two of the four JdFR Nitrospirota lineages (JdFR-85 and UBA6902), while the potential for lactate fermentation was detected in only the JdFR-86 lineage. Acetate formation via genes encoding for pyruvate oxidase (*poxL*) and acylphosphatase (*acyP*) enzymes were present in the JdFR-85 and UBA6902 families (Figure2). Those genes have been previously detected in Heimdallarchaeota MAGs hypothesized to be facultative aerobes (Liu et al. 2020), which could suggest that JdFR-85 and UBA6902 members are able to tolerate microoxic conditions. While JdFR-86 was the only lineage to possess the lactate dehydrogenase (*ldhA*) gene for lactate fermentation (Figure B4, Table A5), this gene was also detected in other anaerobic microorganisms from the deep marine (Boyd et al. 2018) and terrestrial subsurface (Frank et al. 2016), and hydrothermal sediments (Zhou et al. 2020), indicating that it is a widespread fermentative strategy for energy conservation in these systems. However, further studies are needed to elucidate the JdFR Nitrospirota's ability to produce acetate and lactate.

3.4.5. Transporters

A variety of genes involved in phosphate, iron, and macromolecule transport were detected in the JdFR Nitrospirota MAGs (Table A5). All four JdFR Nitrospirota families contain the genes necessary to import inorganic phosphate via the phosphate transport system (*pst*). Furthermore, the uptake of ferrous (Fe^{2+}) iron via the *feo* transport system was only evident in the members of the Family JdFR-88. However, all 4 JdFR Nitrospirota families are likely able to import ferric (Fe^{3+}) iron due to the presence of permeases, and substrate-binding and ATP-binding proteins involved in the iron uptake complex (Table A5). Iron is an important transition

metal that is incorporated as an essential cofactor in redox chemistry, electron transfer reactions and is necessary for maintaining cellular iron homeostasis (Ferousi et al. 2017; Sestok et al. 2018). The importation of long chain fatty acids (LCFA) was evident in all families except for JdFR-88. Imported LCFAs appear likely to be repurposed by the JdFR-85, JdFR-86 and UBA6902 families for phospholipid synthesis due to the presence of key phosphate acetyltransferases (*plsCXY*) (Yao et. al 2017; Sastre et. al 2016) and the absence of LCFA β-oxidation genes (*fabABE*) (Jimenez et. al 2019).

3.4.6. Vitamin and amino acid metabolism

All four JdFR Nitrospirota families appear to fulfill riboflavin (vitamin B2) requirements by synthesizing it from ribulose-5-phosphate (Figure3). All Nitrospirota genomes contain a few genes associated with thiamine (B1) and cobalamin (B12) biosynthesis (Table A5) but are missing key genes such as RNA uracil 4-sulfurtransferase (*thiI*) and thiazole tautomerase (*tenI*), and cobalamin biosynthetic protein (*cobC*) and adenosylcobinamide-phosphate guanylyltransferase (*cobY*) that are needed to complete the respective pathways. Furthermore, transporters for vitamin B1 and B12 were not detected in all JdFR Nitrospirota families. The potential for the biosynthesis of 19 out of 20 amino acids, which were represented by the last step in the pathways in KEGG Decoder, was evident in the JdFR-88 family indicating that the JdFR Nitrospirota have the potential to biosynthesize all amino acids apart from phenylalanine (Table A5).

3.4.7. Motility and biofilm formation

The potential for motility was evident in all JdFR Nitrospirota except for JdFR-88 (Figure2, Fig. 3), while biofilm formation was evident only in the Family JdFR-85 (Fig. 3). Chemotaxis-related genes were evident in all Nitrospirota except for JdFRnit9A (Table A5). Nitrospirota from JdFR-85, JdFR-86 and UBA6902 families possessed flagellar assembly genes necessary for motility (Figure2, Figure3, Table A5). Interestingly, JdFRnit6B from JdFR-85 uniquely harbors genes for biofilm poly-β-1,6-*N*-acetyl-d-glucosamine (PGA) synthesis (*pgaC*) and export (*pgaAB*). However, due to the absence of *pgaD* required for biofilm synthesis (Itoh et al. 2008), the potential for substrate-attached growth in JdFR Nitrospirota remains unresolved.

3.5. Genome variation between JdFRnit3A and JdFRnit7B

The recovery of closely related MAGs from each of the two boreholes investigated in this study (U1362A and U1362B), as well as the availability of closely related MAGs from a third borehole nearby (U1301A), afforded us the opportunity to investigate fine scale genomic and nucleotide variation in the deep subseafloor. The four related genomes within the family JdFR-88 share a minimum average nucleotide identity (ANI) of 99.99% (Table A7).

MAUVE alignments of the four genomes revealed that JdFRnit3A and JdFRnit7B are fully syntenic. In addition, alignments using pyANI identified 64 bases that differed across the alignment of 1,859,437 nucleotides between JdFRnit3A and JdFRnit7B (Table A7). However, whole genome alignments using MAFFT estimated 63 non-identical bases instead of 64. This discrepancy is possibly due to ANIm, which is employed in pyANI, quantifying both non-identical bases and indels (Richter & Mora 2009). Nonetheless, single nucleotide variants (SNVs) identified via read recruitment of U1362A and U1362B metagenome reads revealed

additional diversity in the environment, including 380 SNVs across the JdFRnit3A genome and 517 SNVs across the JdFRnit7B genome (Table 2). Across the JdFRnit3A genome, 281 SNVs were shared between U1362A and U1362B, 30 SNVs were unique to the U1362A recruitment, and 69 SNVs were unique to U1362B (Table 2). Read recruitment to the JdFR7B genome revealed 393 shared SNVs between U1362A and U1362B, 31 SNVs unique to the U1362A genome, and 93 SNVs unique to U1362B (Table 2). Furthermore, alignments of JdFRnit3A and JdFRnit7B reveal that 63 of the non-identical nucleotides correspond to the identified SNVs, and that these positions were clustered in the genome.

Table 2. Summary of single nucleotide variants (SNVs) identified by mapping metagenomic reads from boreholes U1362A and U1362B to JdFR Nitrospirota MAGs JdFRnit3A and JdFRnit7B.

Genome	**Total SNVs**	**Shared SNVs[1]**	**Unique SNVs, U1362A**	**Unique SNVs, U1362B**
JdFRnit3A	380	281	30	69
JdFRnit7B	517	393	31	93

[1] Total number of SNVs that map to the same genomic location within mapped reads both the U1362A and U1362B metagenomes

In both JdFRnit3A and JdFRnit7B, 25 of the 63 non-identical bases were located in a region identified as an integrin protein (COG2304) based on COG annotations in anvi'o. BLASTX results of this sequence from both genomes reveal that its most similar to the integrin protein (WP_172621228.1) of *Microbulbifer sp* strain GL-2, which was isolated from marine

fish intestines (PRJDB8498) (Table A9). In eukaryotes, integrins are membrane proteins that facilitate cell-cell adhesion while in prokaryotes they likely play a role in Ca^{2+} signaling or storage (Chouhan et al. 2011). The remaining non-identical bases of JdFRnit3A and JdFRnit7B were located in regions that were not annotated into COG categories in anvi'o. BLASTX comparisons of the uncharacterized gene sequences identified them as hypothetical proteins (MBK5276770.1) from a *Desulfuromonadales bacterium* recovered from an ancient permafrost sediment core (Liang et al. 2021). The second-best BLASTX result for those sequences was the cytochrome c3 family protein (WP_199381962.1) of a *Geomesophilobacter sediminis* strain from paddy soils (Zhang et al. 2021) (Table A9). Gene sequence similarities between *G. sediminis*, and JdFRnit3A and JdFRnit7B suggest that the uncharacterized sequences could be a differentiated cytochrome c3 family protein, which facilitate ferric iron reduction in *G. sediminis* strains (Zhang et al. 2021). However, it is also possible that the JdFRnit3A and JdFRnit7B sequences are a modified c-type polyheme cytochrome, which is prevalent in the *Desulfuromonadales bacterium* and helps in mitigating osmotic stress (Liang et al. 2021).

The quantified genetic differences between JdFRnit3A and JdFRnit7B could be driven by reduced gene-flow within JdFR-88 populations due to mosaic sympatry or the patchy distribution of niches (Mallet 2008). It is plausible that the patches of nutrients and resources within the JdFR aquifer could be driving the microgeographic separation of JdFR-88 populations and subsequent gene-flow reduction. Mosaic sympatry was previously suggested to drive and maintain the ecological differentiation of sympatric and genetically similar *Vibrio* populations through the patchy distribution of resources in the open ocean (Shapiro et al. 2012; Shapiro & Polz 2015). Additionally, the genetic differences could reflect multiple Nitrospirota genotypes being represented within the JdFR-88 populations since JdFRnit3A and JdFRnit7B are metagenome-

assembled genomes and not single cells. These findings indicate clustered differences at the nucleotide level between JdFRnit3A and JdFRnit7B that could be indicative of genetic discontinuity in the subsurface fluids of the JdFR.

Appendix B- SUPPLEMENTARY FIGURES

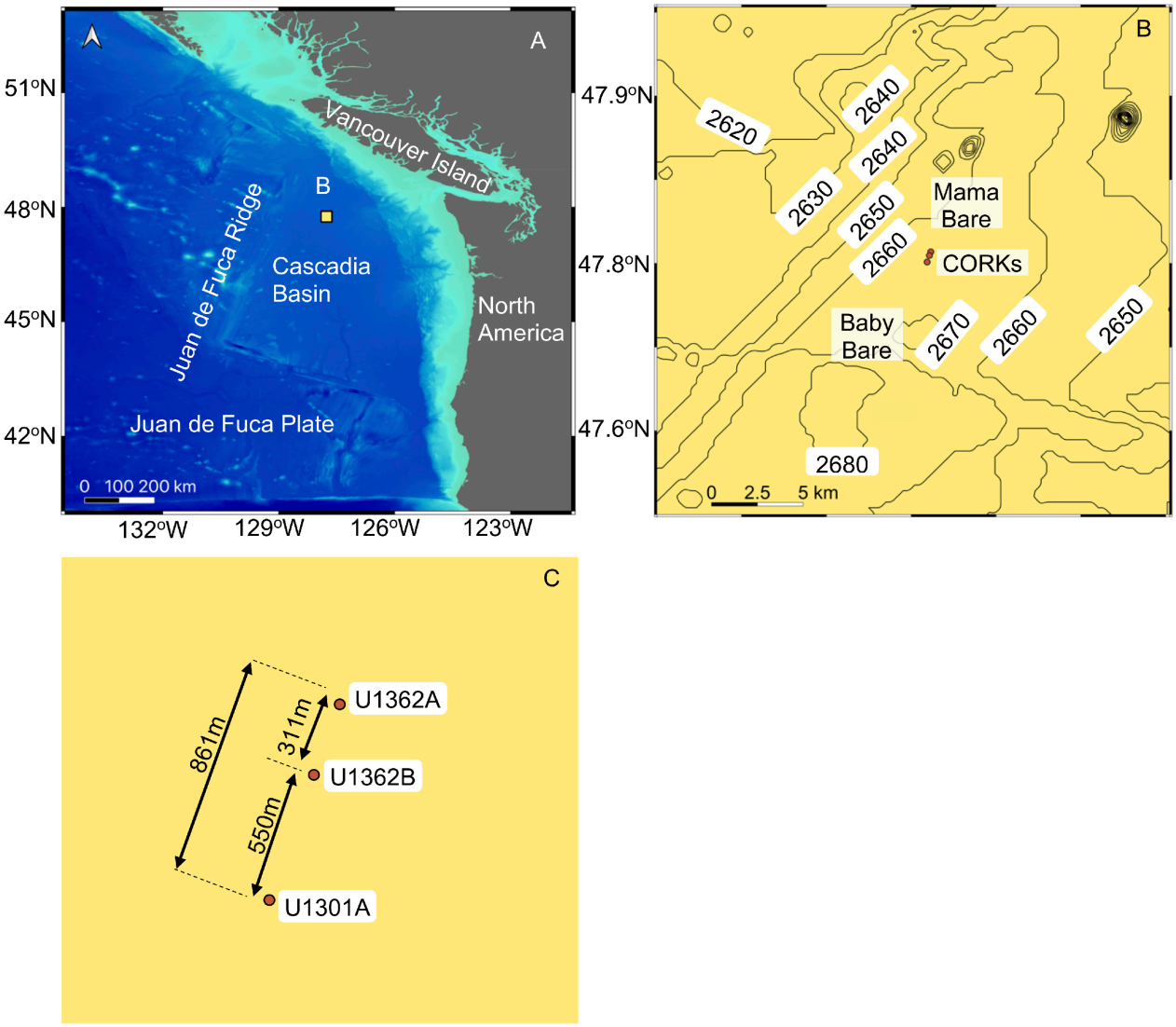

Figure B1. Location of Juan de Fuca Ridge CORKs. Maps illustrating the geographic (yellow box in Panel A) and specific locations (B and C) of IODP boreholes along the eastern flank of the Juan de Fuca Ridge.

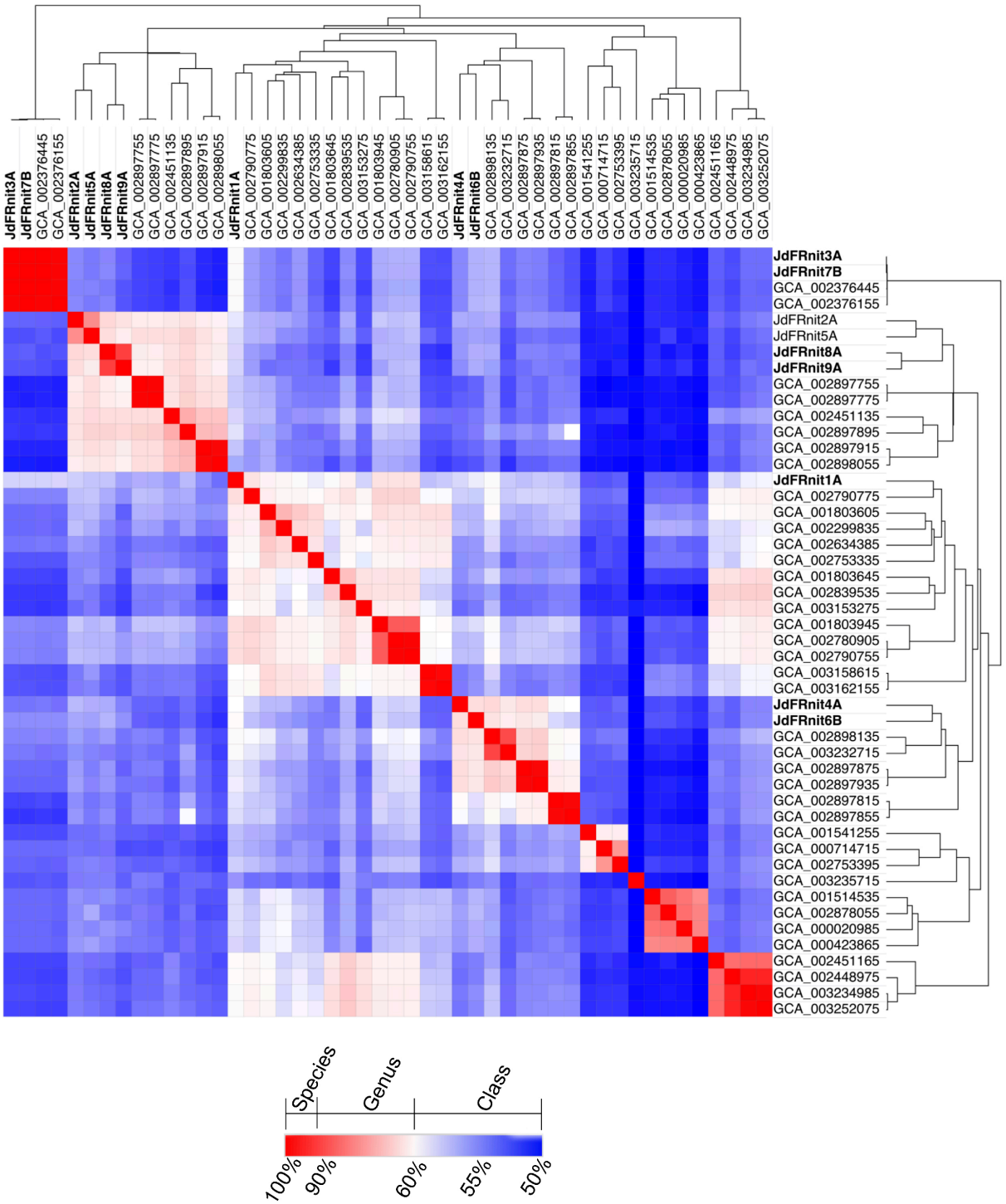

Figure B2. Heatmap representing percent amino acid identities (AAI) between whole genome sequences in the Class Thermodesulfovibrionia.

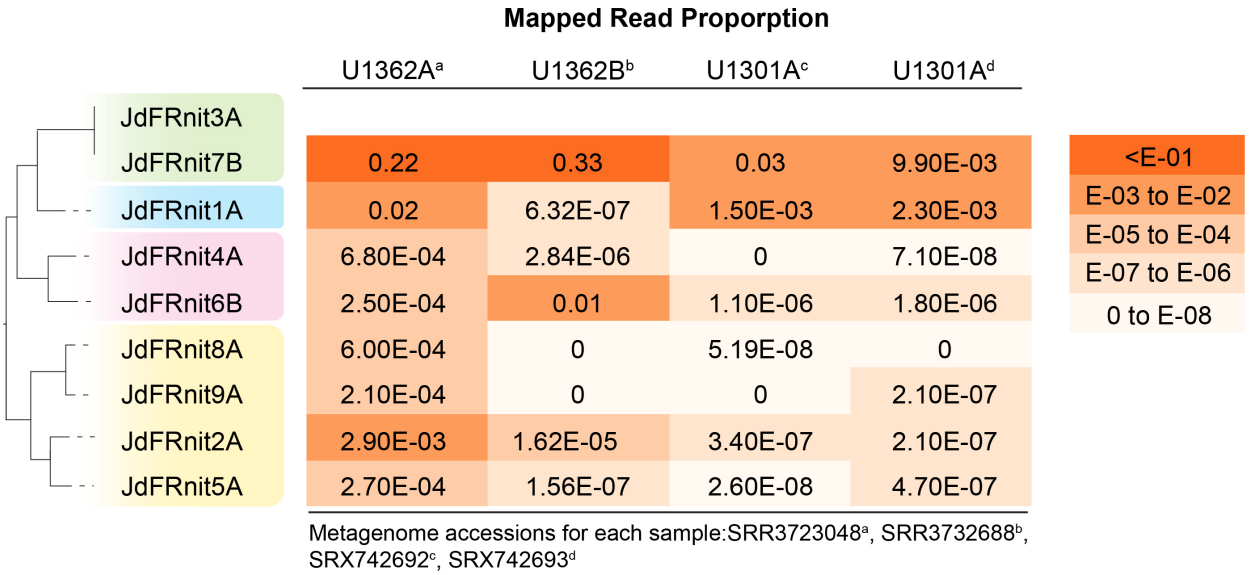

Figure B3. Heatmap representing the estimated abundance of the JdFR Nitrospirota given as a proportion of mapped reads. JdFR Nitrospirota are color coded according to their phylogenomic placement and estimated abundances are color coded

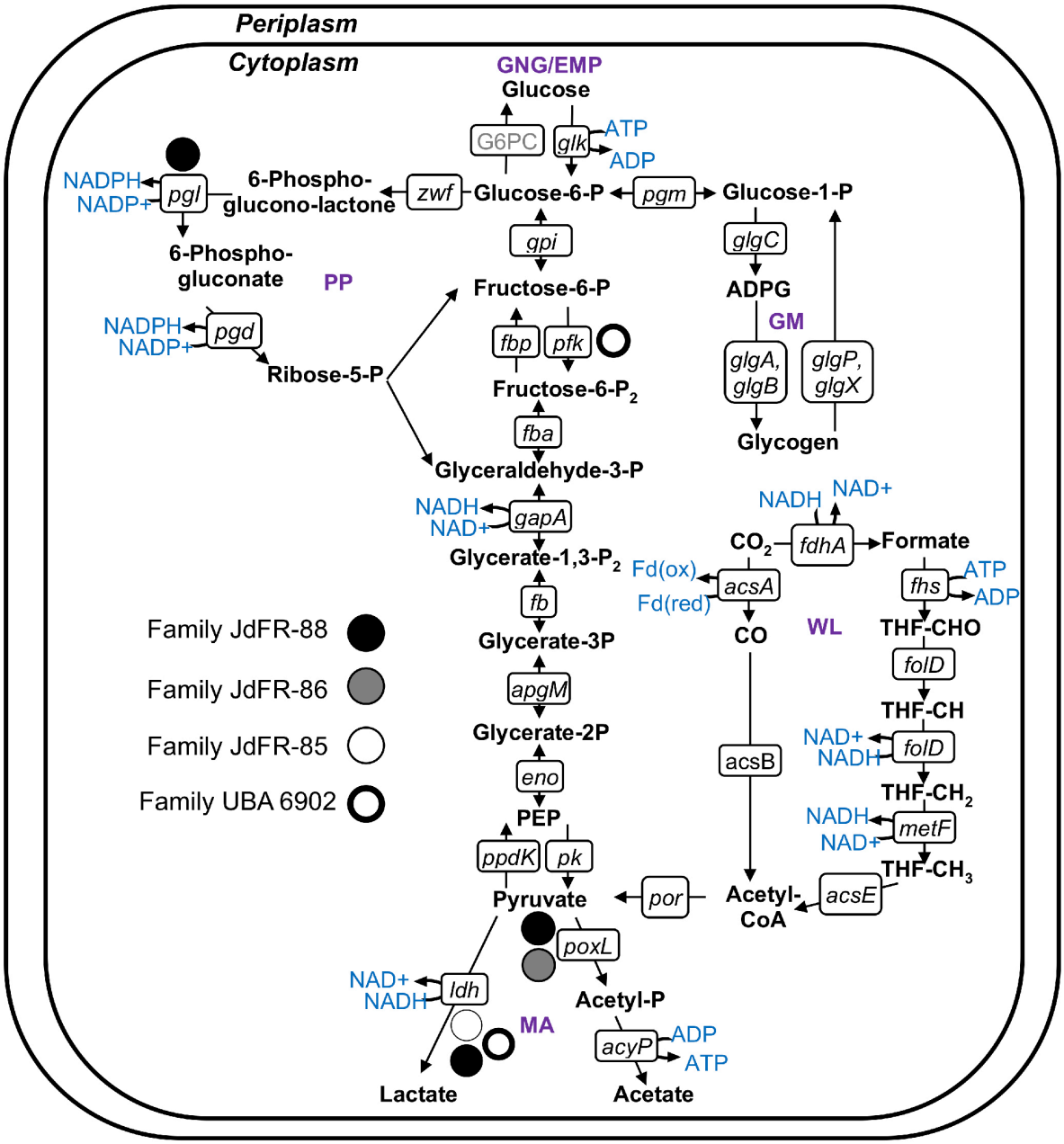

Figure B4. Reconstruction of the pathways used to metabolize carbon in JdFR Nitrospirota.

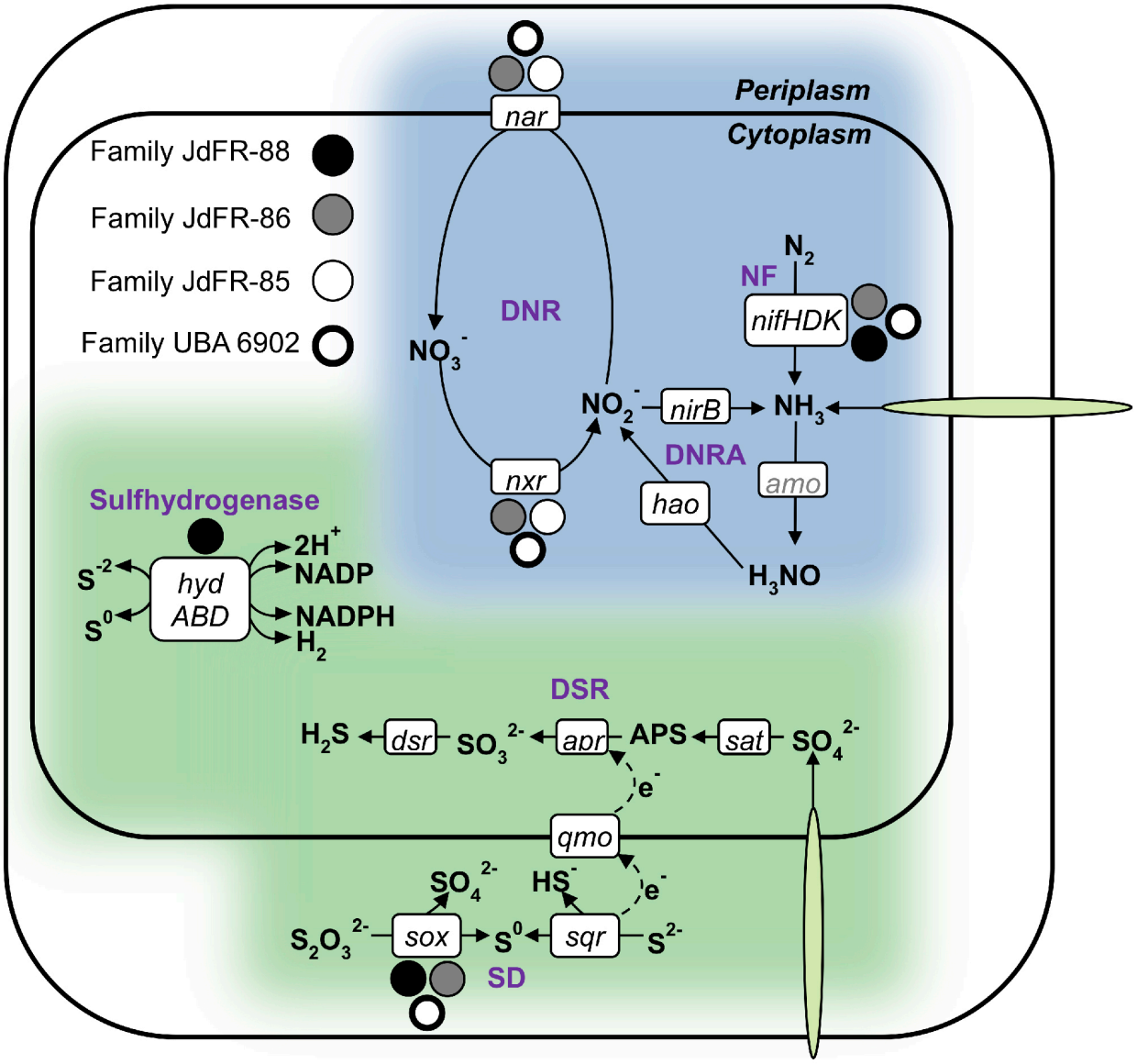

Figure B5. Metabolic reconstruction of sulfur and nitrogen cycling within the JdFR Nitrospirota.

REFERENCES

Adam, P. S., Borrel, G., & Gribaldo, S. (2018). Evolutionary history of carbon monoxide dehydrogenase/acetyl-CoA synthase, one of the oldest enzymatic complexes. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(6), E1166–E1173. https://doi.org/10.1073/pnas.1716667115

Agostino, V., Lenic, A., Bardl, B., Rizzotto, V., Phan, A. N. T., Blank, L. M., & Rosenbaum, M. A. (2020). Electrophysiology of the facultative autotrophic bacterium *Desulfosporosinus orientis*. *Frontiers in Bioengineering and Biotechnology*, *8*, 1–14. https://doi.org/10.3389/fbioe.2020.00457

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, *25*(17), 3389–3402. https://doi.org/10.2503/jjshs.58.977

Anantharaman, K., Brown, C. T., Hug, L. A., Sharon, I., Castelle, C. J., Probst, A. J., … Banfield, J. F. (2016). Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nature Communications*, *7*, 1–11. https://doi.org/10.1038/ncomms13219

Anantharaman, K., Hausmann, B., Jungbluth, S. P., Kantor, R. S., Lavy, A., Warren, L. A., … Banfield, J. F. (2018). Expanded diversity of microbial groups that shape the dissimilatory sulfur cycle. *ISME Journal*, *12*(7), 1715–1728. https://doi.org/10.1038/s41396-018-0078-0

Arshad A, Dalcin Martins P, Frank J, Jetten MSM, Op den Camp HJM, W. C. (2017). Mimicking microbial interactions under nitrate-reducing conditions in an anoxic bioreactor: enrichment of novel Nitrospirae bacteria distantly related to *Thermodesulfovibrio*. *Environ Microbiol.*, *19*(12), 1462-2920.13977. https://doi.org/10.1111/1462-2920.13977

Becker, K., & Davis, E. E. (2005). A review of CORK designs and operations during the Ocean Drilling Program. *Proceedings of the IODP, 301*, *301*. https://doi.org/10.2204/iodp.proc.301.104.2005

Berrisford, J. M., Baradaran, R., & Sazanov, L. A. (2016). Structure of bacterial respiratory complex i. *Biochimica et Biophysica Acta - Bioenergetics*, *1857*(7), 892–901. https://doi.org/10.1016/j.bbabio.2016.01.012

Bhatnagar, S., Badger, J. H., Madupu, R., Khouri, H. M., O'Connor, E. M., Robb, F. T., … Eisen, J. A. (2015). Genome sequence of the sulfate-reducing thermophilic bacterium *Thermodesulfovibrio yellowstonii* strain DSM 11347T (phylum Nitrospirae). *Genome Announcements*, *3*(1), 1994–1995. https://doi.org/10.1128/genomeA.01489-14

Berg, I. A., Kockelkorn, D., Ramos-Vera, W. H., Say, R. F., Zarzycki, J., Hügler, M., … Fuchs, G. (2010). Autotrophic carbon fixation in archaea. *Nature Reviews Microbiology*, *8*(6), 447–460. https://doi.org/10.1038/nrmicro2365

Bowers, R. M., Kyrpides, N. C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T. B. K., … Woyke, T. (2017). Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nature Biotechnology*, *35*(8), 725–731. https://doi.org/10.1038/nbt.3893

Boyd, J. A., Jungbluth, S. P., Leu, A. O., Evans, P. N., Woodcroft, B. J., Chadwick, G. L., … Tyson, G. W. (2019). Divergent methyl-coenzyme M reductase genes in a deep-subseafloor Archaeoglobi. *ISME Journal*, *13*(5), 1269–1279. https://doi.org/10.1038/s41396-018-0343-2

Campbell, J. H., O'Donoghue, P., Campbell, A. G., Schwientek, P., Sczyrba, A., Woyke, T., … Podar, M. (2013). UGA is an additional glycine codon in uncultured SR1 bacteria from the human microbiota. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(14), 5540–5545. https://doi.org/10.1073/pnas.1303090110

Carr, S. A., Jungbluth, S. P., Eloe-Fadrosh, E. A., Stepanauskas, R., Woyke, T., Rappé, M. S., & Orcutt, B. N. (2019). Carboxydotrophy potential of uncultivated Hydrothermarchaeota from the subseafloor crustal biosphere. *ISME Journal*, *13*(6), 1457–1468. https://doi.org/10.1038/s41396-019-0352-9

Cárdenas, J. P., Lazcano, M., Ossandon, F. J., Corbett, M., Holmes, D. S., & Watkin, E. (2014). Draft genome sequence of the iron-oxidizing acidophile *Leptospirillum ferriphilum* type strain DSM 14647. *Genome Announcements*, *2*(6), 2012–2013. https://doi.org/10.1128/genomeA.01153-14

Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., & Parks, D. H. (2019). GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics*, *36*(November 2019), 1925–1927. https://doi.org/10.1093/bioinformatics/btz848

Chen, M., Yan, Y., Zhang, W., Lu, W., Wang, J., Ping, S., & Lin, M. (2011). Complete genome sequence of the type strain *Pseudomonas stutzeri* CGMCC 1.1803. *Journal of Bacteriology*, *193*(21), 6095–6095. https://doi.org/10.1128/JB.06061-11

Chouhan, B., Denesyuk, A., Heino, J., Johnson, M. S., & Denessiouk, K. (2011). Conservation of the human integrin-type beta-propeller domain in bacteria. *PLoS ONE*, *6*(10). https://doi.org/10.1371/journal.pone.0025069

Cohen, G. N. (2011). Glycolysis, gluconeogenesis and glycogen synthesis. In *Microbial Biochemistry: Second Edition* (pp. 1–557). https://doi.org/10.1007/978-90-481-9437-7

Cowen, J. P., Giovannoni, S. J., Kenig, F., Johnson, H. P., Butterfield, D., Rappé, M. S., … Lam, P. (2003). Fluids from aging ocean crust that support microbial life. *Science*, *299*(5603), 120–123. https://doi.org/10.1126/science.1075653

Cowen, J. P. (2004). The microbial biosphere of sediment-buried oceanic basement. *Research in Microbiology*, *155*(7), 497–506. https://doi.org/10.1016/j.resmic.2004.03.008

Cowen, J. P., Copson, D. A., Jolly, J., Hsieh, C. C., Lin, H. T., Glazer, B. T., & Wheat, C. G. (2012). Advanced instrument system for real-time and time-series microbial geochemical sampling of the deep (basaltic) crustal biosphere. *Deep-Sea Research Part I: Oceanographic Research Papers*, *61*, 43–56. https://doi.org/10.1016/j.dsr.2011.11.004

Daims, H (2014). The Family Nitrospiraceae. In Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E., Thompson, F. (Eds.). *The Prokaryotes: Other Major Lineages of Bacteria and The Archaea* (4th ed.) (pp. 733-749). Heidelberg, Germany: Springer.

Darling, A. E., Mau, B., & Perna, N. T. (2010). progressiveMauve: Multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE*, *5*(6). https://doi.org/10.1371/journal.pone.0011147

Drake, H. L., Küsel, K., & Matthies, C. (2002). Ecological consequences of the phylogenetic and physiological diversities of acetogens. *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*, *81*, 203–213. https://doi.org/10.1023/A:1020514617738

Eddy, S. R. (2011). Accelerated profile HMM searches. *PLoS Computational Biology*, *7*(10). https://doi.org/10.1371/journal.pcbi.1002195

Eren, A. M., Vineis, J. H., Morrison, H. G., & Sogin, M. L. (2013). A filtering method to generate high quality short reads using Illumina paired-end technology. *PLoS ONE*, *8*(6), 6–11. https://doi.org/10.1371/journal.pone.0066643

Eren, A. M., Esen, O. C., Quince, C., Vineis, J. H., Morrison, H. G., Sogin, M. L., & Delmont, T. O. (2015). Anvi'o: An advanced analysis and visualization platform for 'omics data. *PeerJ*, *2015*(10), 1–29. https://doi.org/10.7717/peerj.1319

Esposito, A., Tamburini, S., Triboli, L., Ambrosino, L., Chiusano, M. L., & Jousson, O. (2019). Insights into the genome structure of four acetogenic bacteria with specific reference to the Wood-Ljungdahl pathway. *MicrobiologyOpen*, *8*(12), 1–15. https://doi.org/10.1002/mbo3.938

Evans, J. T. & Denef VJ. (2020). To dereplicate or to not dereplicate? *mSphere, 5(3),* e00971-19. https://doi.org/10.1128/mSphere.00971-19.

Ferousi, C., Lindhoud, S., Baymann, F., Kartal, B., Jetten, M. S., & Reimann, J. (2017). Iron assimilation and utilization in anaerobic ammonium oxidizing bacteria. *Current Opinion in Chemical Biology*, *37*, 129–136. https://doi.org/10.1016/j.cbpa.2017.03.009

Fincker, M., Huber, J. A., Orphan, V. J., Rappé, M. S., Teske, A., & Spormann, A. M. (2020). Metabolic strategies of marine subseafloor Chloroflexi inferred from genome

reconstructions. *Environmental Microbiology*, *22*(8), 3188–3204. https://doi.org/10.1111/1462-2920.15061

Fisher, A.T., Wheat, C.G., Becker, K., Cowen, J., Orcutt, B., Hulme, S., Inderbitzen, K., Haddad, A., Pettigrew, T.L., Davis, E.E., Jannasch, H., Grigar, K., Aduddell, R., Meldrum, R., Macdonald, R., and Edwards, K. J. (2011). Design, deployment, and status of borehole observatory systems used for single-hole and cross-hole experiments, IODP Expedition 327, eastern flank of Juan de Fuca Ridge. *Proceedings of the Integrated Ocean Drilling Program*, *327*, 1–38. doi:10.2204/iodp.proc.327.107.2011

Frank, Y. A., Kadnikov, V. V., Lukina, A. P., Banks, D., Beletsky, A. V., Mardanov, A. V., … Ravin, N. V. (2016). Characterization and genome analysis of the first facultatively alkaliphilic *Thermodesulfovibrio* isolated from the deep terrestrial subsurface. *Frontiers in Microbiology*, *7*(DEC), 1–11. https://doi.org/10.3389/fmicb.2016.02000

Fujimura, R., Sato, Y., Nishizawa, T., Oshima, K., Kim, S. W., Hattori, M., … Ohta, H. (2012). Complete genome sequence of *Leptospirillum ferrooxidans* strain C2-3, isolated from a fresh volcanic ash deposit on the Island of Miyake, Japan. *Journal of Bacteriology*, *194*(15), 4122–4123. https://doi.org/10.1128/JB.00696-12

Gittel, A., Seidel, M., Kuever, J., Galushko, A. S., Cypionka, H., & Könneke, M. (2010). *Desulfopila inferna* sp. nov., a sulfate-reducing bacterium isolated from the subsurface of a tidal sand-flat. *International Journal of Systematic and Evolutionary Microbiology*, *60*(7), 1626–1630. https://doi.org/10.1099/ijs.0.015644-0

Gonzaga, A., López-Pérez, A., Martin-Cuadrado, A. B., Ghai, R., & Rodriguez-Valera, F. (2012). Complete genome sequence of the copiotrophic marine bacterium *Alteromonas macleodii* strain ATCC 27126T. *Journal of Bacteriology*, *194*(24), 6998–6998. https://doi.org/10.1128/JB.01565-12

Gould, J. (2016). Morpheus: Versatile matrix visualization and analysis software. [WWW Document]. https://software.broadinstitute.org/morpheus. Accessed 10 Nov 21.

Graham, E. D., Heidelberg, J. F., & Tully, B. J. (2018). Potential for primary productivity in a globally-distributed bacterial phototroph. *ISME Journal*, *12*(7), 1861–1866. https://doi.org/10.1038/s41396-018-0091-3

He, Y., Li, M., Perumal, V., Feng, X., Fang, J., Xie, J., … Wang, F. (2016). Genomic and enzymatic evidence for acetogenesis among multiple lineages of the archaeal phylum Bathyarchaeota widespread in marine sediments. *Nature Microbiology*, *1*(6), 1–9. https://doi.org/10.1038/nmicrobiol.2016.35

Hernsdorf, A. W., Amano, Y., Miyakawa, K., Ise, K., Suzuki, Y., Anantharaman, K., … Banfield, J. F. (2017). Potential for microbial H 2 and metal transformations associated with novel bacteria and archaea in deep terrestrial subsurface sediments. *ISME Journal*, *11*(8), 1915–1929. https://doi.org/10.1038/ismej.2017.39

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018).

UFBoot2: Improving the Ultrafast Bootstrap Approximation. Molecular biology and evolution. *Molecular Biology and Evolution*, *35*(2), 518–522. https://doi.org/10.5281/zenodo.854445

Huntemann, M., Ivanova, N. N., Mavromatis, K., James Tripp, H., Paez-Espino, D., Palaniappan, K., … Kyrpides, N. C. (2015). The standard operating procedure of the DOE-JGI Microbial Genome Annotation Pipeline (MGAP v.4). *Standards in Genomic Sciences*, *10*(1), 1–6. https://doi.org/10.1186/s40793-015-0077-y

Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, *11*(1), 119. https://doi.org/10.1186/1471-2105-11-119

Itoh, Y., Rice, J. D., Goller, C., Pannuri, A., Taylor, J., Meisner, J., … Romeo, T. (2008). Roles of pga ABCD genes in synthesis, modification, and export of the *Escherichia coli* biofilm adhesin Poly-β-1,6-N-Acetyl-d-Glucosamine. *Journal of Bacteriology*, *190*(10), 3670 – 3680. https://doi.org/10.1128/JB.01920-07

Jiao, J. Y., Fu, L., Hua, Z. S., Liu, L., Salam, N., Liu, P. F., … Li, W. J. (2021). Insight into the function and evolution of the Wood-Ljungdahl pathway in *Actinobacteria*. *ISME Journal*, *15*(10), 3005–3018. https://doi.org/10.1038/s41396-021-00935-9

Jimenez-diaz, L., Caballero, A., & Segura, A. (2019). Pathways for the Degradation of Fatty Acids in Bacteria. In Rojo F. (Ed.), *Aerobic Utilization of Hydrocarbons, Oils, and Lipids. Handbook of Hydrocarbon and Lipid Microbiology.* (pp. 291–313). Springer. https://doi.org/10.1007/978-3-319-50418-6

Jungbluth, S. P., Grote, J., Lin, H. T., Cowen, J. P., & Rappé, M. S. (2013). Microbial diversity within basement fluids of the sediment-buried Juan de Fuca Ridge flank. *ISME Journal*, *7*(1), 161–172. https://doi.org/10.1038/ismej.2012.73

Jungbluth, S. P., Bowers, R. M., Lin, H. T., Cowen, J. P., & Rappé, M. S. (2016). Novel microbial assemblages inhabiting crustal fluids within mid-ocean ridge flank subsurface basalt. *ISME Journal*, *10*(8), 2033–2047. https://doi.org/10.1038/ismej.2015.248

Jungbluth, S. P., Amend, J. P., & Rappé, M. S. (2017a). Metagenome sequencing and 98 microbial genomes from Juan de Fuca Ridge flank subsurface fluids. *Scientific Data*, *4*, 170080. https://doi.org/10.1038/sdata.2017.80

Jungbluth, S. P., del Rio, T. G., Tringe, S. G., Stepanauskas, R., & Rappé, M. S. (2017b). Genomic comparisons of a bacterial lineage that inhabits both marine and terrestrial deep subsurface systems. *PeerJ*, *2017*(4), 1–22. https://doi.org/10.7717/peerj.3134

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A., & Jermiin, L. S.

(2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587–589. https://doi.org/10.1038/nmeth.4285

Kanehisa, M., Sato, Y., & Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG Tools for functional characterization of genome and metagenome sequences. *Journal of Molecular Biology*, *428*(4), 726–731. https://doi.org/10.1016/j.jmb.2015.11.006

Kato, S., Ikehata, K., Shibuya, T., Urabe, T., Ohkuma, M., & Yamagishi, A. (2015). Potential for biogeochemical cycling of sulfur, iron and carbon within massive sulfide deposits below the seafloor. *Environmental Microbiology*, *17*(5), 1817–1835. https://doi.org/https://doi.org/10.1111/1462-2920.12648

Kato, S., Shibuya, T., Takaki, Y., Hirai, M., Nunoura, T., & Suzuki, K. (2018). Genome-enabled metabolic reconstruction of dominant chemosynthetic colonizers in deep-sea massive sulfide deposits. *Environmental Microbiology*, *20*(2), 862–877. https://doi.org/10.1111/1462-2920.14032

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Kiełbasa, S. M., Wan, R., Sato, K., Horton, P., & Frith, M. C. (2011). Adaptive seeds tame

genomic sequence comparison. *Genome Research*, *21*(3), 487–493. https://doi.org/10.1101/gr.113985.110

Klenk, H., Clayton, R. A., Tomb, J., Dodson, R. J., Gwinn, M., Hickey, E. K., … Adams, M. D. (1997). The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon. *Nature*, *390*, 364–370. https://doi.org/10.1038/37052

Koch, H., Lücker, S., Albertsen, M., Kitzinger, K., Herbold, C., Spieck, E., … Daims, H. (2015). Expanded metabolic versatility of ubiquitous nitrite-oxidizing bacteria from the genus *Nitrospira*. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(36), 11371–11376. https://doi.org/10.1073/pnas.1506533112

Ladapo, J., & Whitman, W. B. (1990). Method for isolation of auxotrophs in the methanogenic archaebacteria: Role of the acetyl-CoA pathway of autotrophic CO2 fixation in *Methanococcus maripaludis*. *Proceedings of the National Academy of Sciences of the United States of America*, *87*(15), 5598–5602. https://doi.org/10.1073/pnas.87.15.5598

Lang, S. Q., Früh-Green, G. L., Bernasconi, S. M., Brazelton, W. J., Schrenk, M. O., & McGonigle, J. M. (2018). Deeply-sourced formate fuels sulfate reducers but not methanogens at Lost City hydrothermal field. *Scientific Reports*, *8*(1), 1–10. https://doi.org/10.1038/s41598-017-19002-5

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Lee, M. D. (2019). GToTree: A user-friendly workflow for phylogenomics. *Bioinformatics*, *35*(20), 4162–4164. https://doi.org/10.1093/bioinformatics/btz188

Liang, R., Li, Z., Lau Vetter, M. C. Y., Vishnivetskaya, T. A., Zanina, O. G., Lloyd, K. G., … Onstott, T. C. (2021). Genomic reconstruction of fossil and living microorganisms in ancient Siberian permafrost. *Microbiome*, *9*(1), 1–20. https://doi.org/10.1186/s40168-021-01057-2

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lin, H. T., Cowen, J. P., Olson, E. J., Amend, J. P., & Lilley, M. D. (2012). Inorganic chemistry, gas compositions and dissolved organic carbon in fluids from sedimented young basaltic crust on the Juan de Fuca Ridge flanks. *Geochimica et Cosmochimica Acta*, *85*, 213–227. https://doi.org/10.1016/j.gca.2012.02.017

Lin, W., Deng, A., Wang, Z., Li, Y., Wen, T., Wu, L. F., … Pan, Y. (2014). Genomic insights into the uncultured genus "Candidatus Magnetobacterium" in the phylum Nitrospirae. *ISME Journal*, *8*(12), 2463–2477. https://doi.org/10.1038/ismej.2014.94

Lin, W., Paterson, G. A., Zhu, Q., Wang, Y., Kopylova, E., Li, Y., … Pan, Y. (2017). Origin of microbial biomineralization and magnetotaxis during the Archean. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(9), 2171–2176. https://doi.org/10.1073/pnas.1614654114

Lin, W., Zhang, W., Zhao, X., Roberts, A. P., Paterson, G. A., Bazylinski, D. A., & Pan, Y. (2018). Genomic expansion of magnetotactic bacteria reveals an early common origin of magnetotaxis with lineage-specific evolution. *ISME Journal*, *12*(6), 1508–1519. https://doi.org/10.1038/s41396-018-0098-9

Lin, H. T., Hsieh, C. C., Repeta, D. J., & Rappé, M. S. (2020). Sampling of basement fluids via Circulation Obviation Retrofit Kits (CORKs) for dissolved gases, fluid fixation at the seafloor, and the characterization of organic carbon. *MethodsX*, *7*(1), 101033. https://doi.org/10.1016/j.mex.2020.101033

Liu, C., Liu, Y. M., Sun, Q. L., Jiang, C. Y., & Liu, S. J. (2015). Unraveling the kinetic diversity of microbial 3-dehydroquinate dehydratases of shikimate pathway. *AMB Express*, *5*(1), 1–9. https://doi.org/10.1186/s13568-014-0087-y

Lücker, S., Wagner, M., Maixner, F., Pelletier, E., Koch, H., Vacherie, B., … Daims, H. (2010). A *Nitrospira* metagenome illuminates the physiology and evolution of globally important nitrite-oxidizing bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(30), 13479–13484. https://doi.org/10.1073/pnas.1003860107

Mallet, J. (2008). Hybridization, ecological races and the nature of species: Empirical evidence for the ease of speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1506), 2971–2986. https://doi.org/10.1098/rstb.2008.0081

Matsuura, N., Ohashi, A., Tourlousse, D. M., & Sekiguchi, Y. (2016). Draft genome sequence of *Thermodesulfovibrio aggregans* TGE-P1T, an obligately anaerobic, thermophilic, sulfate-reducing bacterium in the phylum Nitrospirae. *Genome Announcements*, *4*(2), 4–5. https://doi.org/10.1128/genomeA.00089-16

Meier, D. V., Pjevac, P., Bach, W., Markert, S., Schweder, T., Jamieson, J., … Meyerdierks, A. (2019). Microbial metal-sulfide oxidation in inactive hydrothermal vent chimneys suggested by metagenomic and metaproteomic analyses. *Environmental Microbiology*, *21*(2), 682–701. https://doi.org/10.1111/1462-2920.14514

Meier-Kolthoff, J. P., Hahnke, R. L., Petersen, J., Scheuner, C., Michael, V., Fiebig, A., … Klenk, H.-P. (2014). Complete genome sequence of DSM 30083T, the proposal for delineating subspecies in microbial taxonomy. *Standards in Genomic Sciences*, *9*(2). https://doi.org/10.1186/1944-3277-9-2

Merino, N., Kawai, M., Boyd, E. S., Colman, D. R., McGlynn, S. E., Nealson, K. H., Kurokawa, K., & Hongoh, Y. (2020). Single-cell genomics of novel *Actinobacteria* with the Wood-Ljungdahl pathway discovered in a serpentinizing system. *Frontiers in Microbiology*, *11*. Retrieved from https://doi.org/10.3389/fmicb.2020.01031

Mukherjee, S., Seshadri, R., Varghese, N. J., Eloe-Fadrosh, E. A., Meier-Kolthoff, J. P., Göker, M., … Kyrpides, N. C. (2017). 1,003 Reference genomes of bacterial and archaeal isolates expand coverage of the tree of life. *Nature Biotechnology*, *35*(7), 676–683. https://doi.org/10.1038/nbt.3886

Nierman, W. C., DeShazer, D., Kim, H. S., Tettelin, H., Nelson, K. E., Feldblyum, T., … Fraser, C. M. (2004). Structural flexibility in the *Burkholderia mallei* genome. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(39), 14246–14251. https://doi.org/10.1073/pnas.0403306101

Nigro, O. D., Jungbluth, S. P., Lin, H. T., Hsieh, C. C., Miranda, J. A., Schvarcz, C. R., … Steward, G. F. (2017). Viruses in the oceanic basement. *mBio*, *8*(2), 1–15. https://doi.org/10.1128/mBio.02129-16

Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274. https://doi.org/10.1093/molbev/msu300

Nurk, S., Meleshko, D., Korobeynikov, A., & Pevzner, P. A. (2017). MetaSPAdes: A new versatile metagenomic assembler. *Genome Research*, *27*(5), 824–834. https://doi.org/10.1101/gr.213959.116

Okabe, S., Shafdar, A. A., Kobayashi, K., Zhang, L., & Oshiki, M. (2020). Glycogen metabolism of the anammox bacterium "*Candidatus* Brocadia sinica." *ISME Journal*, *15*, 1287–1301. https://doi.org/10.1038/s41396-020-00850-5

Orcutt, B., Wheat, C. G., & Edwards, K. J. (2010). Subseafloor ocean crust microbial observatories: Development of FLOCS (FLow-through Osmo Colonization System) and evaluation of borehole construction materials. *Geomicrobiology Journal*, *27*(2), 143–157. https://doi.org/10.1080/01490450903456772

Orcutt, B. N., Bach, W., Becker, K., Fisher, A. T., Hentscher, M., Toner, B. M., … Edwards, K. J. (2011). Colonization of subsurface microbial observatories deployed in young ocean crust. *ISME Journal*, *5*(4), 692–703. https://doi.org/10.1038/ismej.2010.157

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015). CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, *25*(7), 1043–1055. https://doi.org/10.1101/gr.186072.114

Parks, D. H., Rinke, C., Chuvochina, M., Chaumeil, P. A., Woodcroft, B. J., Evans, P. N., … Tyson, G. W. (2017). Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nature Microbiology*, *2*(11), 1533–1542. https://doi.org/10.1038/s41564-017-0012-7

Pritchard, L., Glover, R. H, Humphris, S., Elphinstone, J. G, & Toth, I. K. (2015). Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Analytical Methods, 8*(1), 12-24. https://doi.org/10.1039/C5AY02550H

Probst, A. J., Ladd, B., Jarett, J. K., Geller-Mcgrath, D. E., Sieber, C. M. K., Emerson, J. B., … Banfield, J. F. (2018). Differential depth distribution of microbial function and putative symbionts through sediment-hosted aquifers in the deep terrestrial subsurface. *Nature Microbiology*, *3*(3), 328–336. https://doi.org/10.1038/s41564-017-0098-y

Rambaut, A. (2018) FigTree v1.4.4. Institute of Evolutionary Biology, University of Edinburgh, Edinburgh. http://tree.bio.ed.ac.uk/software/figtree/

Ramírez, G. A., Garber, A. I., Lecoeuvre, A., D'angelo, T., Wheat, C. G., & Orcutt, B. N. (2019). Ecology of subseafloor crustal biofilms. *Frontiers in Microbiology*, *10*(AUG), 1–17. https://doi.org/10.3389/fmicb.2019.01983

Richter, M., & Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences, 106*(45), 19126 LP – 19131. https://doi.org/10.1073/pnas.0906412106

Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N. N., Anderson, I. J., Cheng, J. F., … Woyke, T. (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature*, *499*(7459), 431–437. https://doi.org/10.1038/nature12352

Robador, A., Jungbluth, S. P., LaRowe, D. E., Bowers, R. M., Rappé, M. S., Amend, J. P., & Cowen, J. P. (2015). Activity and phylogenetic diversity of sulfate-reducing microorganisms in low-temperature subsurface fluids within the upper oceanic crust. *Frontiers in Microbiology*, *6*(JAN), 1–13. https://doi.org/10.3389/fmicb.2014.00748

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative Genome Viewer. *Nature Biotechnology*, *29*(1), 24–26. https://doi.org/10.1038/nbt.1754.Integrative

Sastre, D. E., Bisson-Filho, A., de Mendoza, D., & Gueiros-Filho, F. J. (2016). Revisiting the cell biology of the acyl-ACP:phosphate transacylase PlsX suggests that the phospholipid synthesis and cell division machineries are not coupled in *Bacillus subtilis*. *Molecular Microbiology*, *100*(4), 621–634. https://doi.org/https://doi.org/10.1111/mmi.13337

Sekar, K., Linker, S. M., Nguyen, J., Stocker, R., & Sauer, U. (2020). Bacterial glycogen

provides short-term benefits in changing environments. *Applied and Environmental Microbiology*, *86*(9). Retrieved from https://doi.org/10.1128/AEM.00049-20

Schauder, R., Preuß, A., Jetten, M., & Fuchs, G. (1988). Oxidative and reductive acetyl CoA/carbon monoxide dehydrogenase pathway in *Desulfobacterium autotrophicum* - 2. Demonstration of the enzymes of the pathway and comparison of CO dehydrogenase. *Archives of Microbiology*, *151*(1), 84–89. https://doi.org/10.1007/BF00444674

Schuchmann, K., & Müller, V. (2014). Autotrophy at the thermodynamic limit of life: A model for energy conservation in acetogenic bacteria. *Nature Reviews Microbiology*, *12*(12), 809–821. https://doi.org/10.1038/nrmicro3365

Sestok, A. E., Linkous, R. O., & Smith, A. T. (2018). Toward a mechanistic understanding of Feo-mediated ferrous iron uptake. *Metallomics*, *10*(7), 887–898. https://doi.org/10.1039/C8MT00097B

Shapiro, B. J., Friedman, J., Cordero, O. X., Preheim, S. P., Timberlake, S. C., Szabó, G., … Alm, E. J. (2012). Population genomics of early events in the ecological differentiation of bacteria. *Science*, *335*(6077), 48–51. https://doi.org/10.1126/science.1218198

Shapiro, B. J., & Polz, M. F. (2015). Microbial speciation. *Cold Spring Harbor Perspectives in Biology*, *7*(10), 1–22. https://doi.org/10.1101/cshperspect.a018143

Smith, A., Popa, R., Fisk, M., Nielsen, M., Wheat, C. G., Jannasch, H. W., … Flores, G. (2011). In situ enrichment of Ocean crust microbes on igneous minerals and glasses using an osmotic flow-through device. *Geochemistry, Geophysics, Geosystems*, *12*(6), 1–19. https://doi.org/10.1029/2010GC003424

Smith, A. R., Fisk, M. R., Thurber, A. R., Flores, G. E., Mason, O. U., Popa, R., & Colwell, F. S. (2017). Deep crustal communities of the Juan de Fuca Ridge are governed by mineralogy. *Geomicrobiology Journal*, *34*(2), 147–156. https://doi.org/10.1080/01490451.2016.1155001

Smith, A. R., Kieft, B., Mueller, R., Fisk, M. R., Mason, O. U., Popa, R., & Colwell, F. S. (2019). Carbon fixation and energy metabolisms of a subseafloor olivine biofilm. *ISME Journal*, *13*(7), 1737–1749. https://doi.org/10.1038/s41396-019-0385-0

Sousa, F. L., Neukirchen, S., Allen, J. F., Lane, N., & Martin, W. F. (2016). Lokiarchaeon is hydrogen dependent. *Nature Microbiology*, *1*(5), 1–3. https://doi.org/10.1038/nmicrobiol.2016.34

Spang, A., Saw, J. H., Jørgensen, S. L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A. E., … Ettema, T. J. G. (2015). Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature*, *521*(7551), 173–179. https://doi.org/10.1038/nature14447

Teng, Y., Xu, Y., Wang, X., & Christie, P. (2019). Function of biohydrogen metabolism and

related microbial communities in environmental bioremediation. *Frontiers in Microbiology*, *10*(106), 1–14. https://doi.org/10.3389/fmicb.2019.00106

Thauer, R. K., Kaster, A. K., Seedorf, H., Buckel, W., & Hedderich, R. (2008). Methanogenic archaea: Ecologically relevant differences in energy conservation. *Nature Reviews Microbiology*, *6*(8), 579–591. https://doi.org/10.1038/nrmicro1931

Thorvaldsdóttir, H., Robinson, J. T., & Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, *14*(2), 178–192. https://doi.org/10.1093/bib/bbs017

Umezawa, K., Kojima, H., Kato, Y., & Fukui, M. (2020). Disproportionation of inorganic sulfur compounds by a novel autotrophic bacterium belonging to *Nitrospirota*. *Systematic and Applied Microbiology*, *43*(5), 126110. https://doi.org/10.1016/j.syapm.2020.126110

Ushiki, N., Fujitani, H., Shimada, Y., Morohoshi, T., Sekiguchi, Y., & Tsuneda, S. (2018). Genomic analysis of two phylogenetically distinct *Nitrospira* species reveals their genomic plasticity and functional diversity. *Frontiers in Microbiology*, *8*(JAN), 1–12. https://doi.org/10.3389/fmicb.2017.02637

Van Kessel, M. A. H. J., Speth, D. R., Albertsen, M., Nielsen, P. H., Op Den Camp, H. J. M., Kartal, B., … Lücker, S. (2015). Complete nitrification by a single microorganism. *Nature*, *528*(7583), 555–559. https://doi.org/10.1038/nature16459

Vavourakis, C. D., Andrei, A. S., Mehrshad, M., Ghai, R., Sorokin, D. Y., & Muyzer, G. (2018). A metagenomics roadmap to the uncultured genome diversity in hypersaline soda lake sediments. *Microbiome*, *6*(1), 1–18. https://doi.org/10.1186/s40168-018-0548-7

Vornolt, J., Kunow, J., Stetter, K. O., & Thauer, R. K. (1995). Enzymes and coenzymes of the carbon monoxide dehydrogenase pathway for autotrophic CO2 fixation in *Archaeoglobus lithotrophicus* and the lack of carbon monoxide dehydrogenase in the heterotrophic *A. profundus*. *Archives of Microbiology*, *163*(2), 112–118. https://doi.org/10.1007/BF00381784

Wang, L., Wang, M., Wise, M. J., Liu, Q., Yang, T., Zhu, Z., … Wang, W. (2020). Recent progress in the structure of glycogen serving as a durable energy reserve in bacteria. *World Journal of Microbiology and Biotechnology*, *36*(1), 1–12. https://doi.org/10.1007/s11274-019-2795-6

Westphal, L., Wiechmann, A., Baker, J., Minton, NP., & Müller, V. (2018). The Rnf Complex is an Energy-coupled transhydrogenase essential to reversibly link cellular NADH and ferredoxin pools in the acetogen *Acetobacterium woodii*. *Journal of Bacteriology*, *200*(21), 1–13. Retrieved from https://doi.org/10.1128/JB.00357-18.

Wickham, H & Henry, L. (2020). tidyr: Tidy Messy Data. R package version 1.0.2. https://CRAN.R-project.org/package=tidyr

Wilkins, L. G. E., Ettinger, C. L., Jospin, G., & Eisen, J. A. (2019). Metagenome-assembled genomes provide new insight into the microbial diversity of two thermal pools in Kamchatka, Russia. *Scientific Reports*, *9*(1), 1–15. https://doi.org/10.1038/s41598-019-39576-6

Wheat, C. G., Mottl, M. J., Fisher, A. T., Kadko, D., Davis, E. E., & Baker, E. (2004). Heat flow through a basaltic outcrop on a sedimented young ridge flank. *Geochemistry, Geophysics, Geosystems*, *5*(12). https://doi.org/10.1029/2004GC000700

Wheat, C. G., Jannasch, H. W., Fisher, A. T., Becker, K., Sharkey, J., & Hulme, S. (2010). Subseafloor seawater-basalt-microbe reactions: Continuous sampling of borehole fluids in a ridge flank environment. *Geochemistry, Geophysics, Geosystems*, *11*(7), 1–18. https://doi.org/10.1029/2010GC003057

Wheat, C. G., Jannasch, H. W., Kastner, M., Hulme, S., Cowen, J., Edwards, K. J., … Glazer, B. (2011). Fluid sampling from oceanic borehole observatories: design and methods for CORK activities (1990–2010), *327*. https://doi.org/10.2204/iodp.proc.327.109.2011

Woodcroft, B. J., Singleton, C. M., Boyd, J. A., Evans, P. N., Emerson, J. B., Zayed, A. A. F., … Tyson, G. W. (2018). Genome-centric view of carbon processing in thawing permafrost. *Nature*, *560*(7716), 49–54. https://doi.org/10.1038/s41586-018-0338-1

Yao, J., & Rock, C. O. (2017). Exogenous fatty acid metabolism in bacteria. *Biochimie*, *141*, 30–39. https://doi.org/10.1016/j.biochi.2017.06.015

Zecchin, S., Mueller, R. C., Seifert, J., Stingl, U., Anantharaman, K., von Bergen, M., … Pester, M. (2018). Rice paddy *Nitrospirae* carry and express genes related to sulfate respiration: Proposal of the new genus "*Candidatus* Sulfobium." *Applied and Environmental Microbiology*, *84*(5), 1–15. https://doi.org/10.1128/AEM.02224-17

Zhang, Z., Xu, Z., Masuda, Y., Wang, X., Ushijima, N., Shiratori, Y., … Itoh, H. (2021). *Geomesophilobacter sediminis* gen. nov., sp. nov., *Geomonas propionica* sp. nov. and *Geomonas anaerohicana* sp. nov., three novel members in the family Geobacterecace isolated from river sediment and paddy soil. *Systematic and Applied Microbiology*, *44*(5), 126233. https://doi.org/10.1016/j.syapm.2021.126233

Zhou, Z., Liu, Y., Xu, W., Pan, J., Luo, Z.-H., & Li, M. (2020). Genome- and community-level interaction insights into carbon utilization and element cycling functions of *Hydrothermarchaeota* in hydrothermal sediment. *mSystems*, *5*(1), e00795-19. https://doi.org/10.1128/mSystems.00795-19

Supplementary Materials

Table of Contents

Table A1. Summary of sequencing, assembly, and genome binning from 2011 JdFR crustal fluid microbiomes.

	U1362A	U1362B
Total sequence reads (raw)	296,383,518	162,069,852
Total sequence reads (post quality filtering and error correction)	269,316,754	140,739,522
Total bases (raw)	45,424,726,500	24,851,414,400
Total bases (post quality filtering and error correction)	40,158,401,330	20,836,645,488
Total assembly length[a]	171,066,974	168,068,444
No. scaffolds[b]	19,844	26,440
Longest scaffold (Mbp)	1.92	1.81
L50	973	1325
L75	4,269	7,612
L90	10,875	16,857
N50	29,031	16,658
N75	6,550	3,757
N90	2,518	2,082
No. genes	178,638	183,069
Read Recruitment (%)[c]	96.2	88.5
Genome bins (with assigned anvi'o SCG domains)	63	50
Genome bins, >50% completion & <10% redundancy	54	36
Archaeal genome bins	23	21
Archaeal genome bins, >50% completion & <10% redundancy	21	18
Bacterial genome bins	41	29
Bacterial genome bins, >50% completion & <10% redundancy	33	18
Nitrospirota genome bins	7	2
Nitrospirota genome bins, >50% completion & <10% redundancy	6	2

[a]1500 bp minimum contig

[b]1500 bp minimum

[c]Source metagenome sequence reads recruited to their respective assemblies

Table A2. Quality of the JdFR Nitrospirota MAGs as assessed via the CheckM v1.1.2 pipeline and their Genome Taxonomy Database classification.

MAG	Marker lineage	# genomes[a]	#markers[b]	# marker sets[c]	# marker gene occurrence[d] 0	1	2	3+	Completeness[e]	Contamination[f]	Strain heterogeneity[g]
JdFRnit1A	k__Bacteria (UID3187)	2258	181	110	2	178	1	0	98.18	0.23	0.00
JdFRnit2A	k__Bacteria (UID3187)	2258	181	110	7	172	2	0	94.55	1.82	0.00
JdFRnit3A	k__Bacteria (UID3187)	2258	181	110	0	179	2	0	100.00	1.82	0.00
JdFRnit4A	k__Bacteria (UID3187)	2258	181	110	36	143	2	0	85.91	1.09	0.00
JdFRnit5A	k__Bacteria (UID3187)	2258	181	110	63	118	0	0	51.82	0.00	0.00
JdFRnit8A	k__Bacteria (UID3187)	2258	181	110	67	113	1	0	59.64	0.00	0.00
JdFRnit9A	k__Bacteria (UID203)	5499	103	57	74	28	1	0	43.86	1.75	0.00
JdFRnit6B	k__Bacteria (UID3187)	2258	181	110	2	177	2	0	98.18	1.82	0.00
JdFRnit7B	k__Bacteria (UID3187)	2258	181	110	0	179	2	0	100.00	1.82	0.00

[a] number of reference genomes used to infer lineage-specific marker set

[b] number of marker genes included in the lineage-specific marker set`

[c] number of inferred co-located marker sets

[d] number of times each marker gene is identified

[e] determined from the presence/absence of marker genes and the expected collocalization of these genes

[f] determined by the presence of multi-copy marker genes and the expected collocalization of these genes

[g] determined from the number of multi-copy marker pairs which exceed a specified amino acid identity threshold (default = 90%)

Table A2. continued.

MAG	GTDB-Tk Classification
JdFRnit1A	d__Bacteria;p__Nitrospirota;c__Thermodesulfovibrionia;o__Thermodesulfovibrionales;f__JdFR-86;g__JdFR-86;s__JdFR-86 sp002011815
JdFRnit2A	d__Bacteria;p__Nitrospirota;c__Thermodesulfovibrionia;o__UBA6902;f__UBA6902;g__JdFR-81;s__JdFR-81 sp002011735
JdFRnit3A	d__Bacteria;p__Nitrospirota;c__Thermodesulfovibrionia;o__Thermodesulfovibrionales;f__JdFR-88;g__JdFR-88;s__JdFR-88 sp002011795
JdFRnit4A	d__Bacteria;p__Nitrospirota;c__Thermodesulfovibrionia;o__Thermodesulfovibrionales;f__JdFR-85;g__;s__
JdFRnit5A	d__Bacteria;p__Nitrospirota;c__Thermodesulfovibrionia;o__UBA6902;f__UBA6902;g__JdFR-81;s__
JdFRnit8A	d__Bacteria;p__Nitrospirota;c__Thermodesulfovibrionia;o__UBA6902;f__UBA6902;g__;s__
JdFRnit9A	d__Bacteria;p__Nitrospirota;c__Thermodesulfovibrionia;o__UBA6902;f__UBA6902;g__;s__
JdFRnit6B	d__Bacteria;p__Nitrospirota;c__Thermodesulfovibrionia;o__Thermodesulfovibrionales;f__JdFR-85;g__JdFR-85;s__JdFR-85 sp002011745
JdFRnit7B	d__Bacteria;p__Nitrospirota;c__Thermodesulfovibrionia;o__Thermodesulfovibrionales;f__JdFR-88;g__JdFR-88;s__JdFR-88 sp002011795

Table A3. Genome information of Nitrospirota whole-genome sequences and reference taxa used in the phylogenomic analysis.

Name	NCBI BioProject ID	GenBank assembly accession	CheckM completeness	CheckM contamination	Sequence length (Mbp)	GC Content	Type	Class	Family	Source	Reference
Alteromonas macleodii ATCC 27126	PRJNA29793	GCA_000172635	99.66	0.51	4.65	44.71	Strain	Gammaproteobacteria	Alteromonadaceae	Surface ocean seawater	Gonzaga et al. 2012
Burkholderia mallei ATCC 23344	PRJNA171	GCA_000011705	99.95	0.00	5.84	68.49	Strain	Gammaproteobacteria	Burkholderiaceae	Human	Nierman et al. 2004
Escherichia coli ATCC 11775	PRJNA50621	GCA_000690815	99.97	0.39	5.04	50.64	Strain	Gammaproteobacteria	Enterobacteriaceae	Human	Meier-Kolthoff et al. 2014
Pseudomonas stutzeri ATCC 17588	PRJNA68131	GCA_000219605	99.86	0.55	4.55	63.93	Strain	Gammaproteobacteria	Pseudomonadaceae	Human	Chen et al. 2011
Methylophilus methylotrophus ATCC 53528	PRJNA181379	GCA_000378225	100.00	0.00	2.86	49.61	Strain	Gammaproteobacteria	Methylophilaceae	Activated sludge	Mukherjee et al. 2017
Leptospirillum ferriphilum	PRJNA248540	GCA_000755505	94.17	2.73	2.41	54.05	MAG	Leptospirillia	Leptospirillaceae	Acid mine	Cardenas et al. 2014
Leptospirillum ferrooxidans	PRJDB73	GCA_000284315	94.49	1.82	2.56	50.05	Strain	Leptospirillia	Leptospirillaceae	Volcanic ash deposit	Fujimura et al. 2014
Nitrospira defluvii	PRJEA46433	GCA_000196815	97.67	2.27	4.32	59.03	Strain	Nitrospiria	Nitrospiraceae	Activated sludge	Lücker et al. 2010
Nitrospira japonica	PRJEB18128	GCA_900169565	96.82	3.92	4.08	58.96	Strain	Nitrospiria	Nitrospiraceae	Activated sludge	Ushiki et al. 2017
Nitrospira moscoviensis	PRJNA283178	GCA_001273775	95.91	6.55	4.59	61.99	Strain	Nitrospiria	Nitrospiraceae	Corroded iron pipe	Koch et al. 2015
Candidatus Nitrospira nitrificans	PRJEB11446	GCA_001458775	96.76	2.73	4.12	56.59	MAG	Nitrospiria	Nitrospiraceae	Anoxic bioreactor	van Kessel et al. 2015

Table A3 continued.

Name	NCBI BioProject ID	GenBank assembly accession	CheckM completeness	CheckM contamination	Sequence length (Mbp)	GC Content	Type	Class	Family	Source	Reference
Nitrospirae bacterium RBG	PRJNA288027	GCA_001803795	89.55	0.91	2.38	64.57	MAG	RBG-16-64-22	RBG-16-64-22	Terrestrial subsurface	Ananthaman et al. 2016 & 2018
Candidatus Magnetobacterium casensis	PRJNA236065	GCA_000714715	92.42	0.91	3.42	48.87	MAG	Thermodesulfovibrionia	Magnetobacteriaceae	Freshwater lake	Lin et al. 2014
Nitrospira bacterium HGW	PRJNA321556	GCA_002839535	96.70	0.00	1.66	46.15	MAG	Thermodesulfovibrionia	UBA6898	Terrestrial subsurface fluids	Hernsdorf et al. 2017
Nitrospira sp.	PRJNA397647	GCA_002634385	100.00	4.98	4.31	55.93	MAG	Thermodesulfovibrionia	UBA9935	Anoxic bioreactor	Arshad et al. 2017
Nitrospiraceae bacterium	PRJNA386568	GCA_003162155	95.00	0.91	2.95	49.26	MAG	Thermodesulfovibrionia	UBA9935	Permafrost soil	Woodcroft et al. 2018
Nitrospiraceae bacterium	PRJNA386568	GCA_003158615	94.00	1.82	3.56	48.85	MAG	Thermodesulfovibrionia	UBA9935	Permafrost soil	Woodcroft et al. 2018
Nitrospiraceae bacterium	PRJNA386568	GCA_003153275	79.34	1.82	2.52	49.59	MAG	Thermodesulfovibrionia	UBA6898	Permafrost soil	Woodcroft et al. 2018
Nitrospiraceae bacterium UBA3562	PRJNA348753	GCA_002376155	93.64	3.36	1.77	62.98	MAG	Thermodesulfovibrionia	JdFR-88	CORK-incubated minerals	Smith et al. 2011; Parks et al. 2017
Nitrospiraceae bacterium UBA3568	PRJNA348753	GCA_002376445	86.82	2.27	1.66	63.04	MAG	Thermodesulfovibrionia	JdFR-88	CORK-incubated minerals	Smith et al. 2011; Parks et al. 2017
Nitrospiraceae bacterium UBA665	PRJNA348753	GCA_002299835	94.41	0.71	2.02	40.35	MAG	Thermodesulfovibrionia	UBA9935	Hydrothermal vent fluids	Lang et al. 2016; Parks et al. 2017
Nitrospiraceae bacterium UBA6898	PRJNA348753	GCA_002448975	82.90	3.64	3.10	52.37	MAG	Thermodesulfovibrionia	UBA6898	Freshwater river water	Parks et al. 2017

Table A3 continued.

Name	NCBI BioProject ID	GenBank assembly accession	CheckM completeness	CheckM contamination	Sequence length (Mbp)	GC Content	Type	Class	Family	Source	Reference
Nitrospiraceae bacterium UBA6902	PRJNA348753	GCA_002451135	94.49	1.19	3.05	46.5	MAG	Thermodesulfovibrionia	UBA6902	Freshwater river sediment	Parks et al. 2017
Nitrospiraceae bacterium UBA6905	PRJNA348753	GCA_002451165	86.26	5.45	2.98	51.5	MAG	Thermodesulfovibrionia	UBA6898	Freswater river sediment	Parks et al. 2017
Nitrospirae bacterium	PRJNA385762	GCA_003252075	89.09	2.73	2.86	52.15	MAG	Thermodesulfovibrionia	UBA6898	Hydrothermal sulfide sediment	Zhou et al. 2020
Nitrospirae bacterium	PRJNA385762	GCA_003235715	94.55	4.55	3.24	57.86	MAG	Thermodesulfovibrionia	SZUA-242	Hydrothermal sulfide sediment	Zhou et al. 2020
Nitrospirae bacterium	PRJNA385762	GCA_003234985	99.09	7.73	3.38	52.33	MAG	Thermodesulfovibrionia	UBA6898	Hydrothermal sulfide sediment	Zhou et al. 2020
Nitrospirae bacterium	PRJNA385762	GCA_003232715	84.26	0.00	2.06	46.04	MAG	Thermodesulfovibrionia	JdFR-85	Hydrothermal chimney wall	Zhou et al. 2020
Nitrospirae bacterium	PRJNA300727	GCA_001541255	98.18	0.91	3.59	45.37	MAG	Thermodesulfovibrionia	Magnetobacteriaceae	Freshwater sediment	Lin et al. 2017
Nitrospirae bacterium BMS3Abin06	PRJDB5792	GCA_002897895	87.53	4.55	2.98	43.06	MAG	Thermodesulfovibrionia	UBA6902	Sulfide deposits	Kato et al. 2018
Nitrospirae bacterium BMS3Abin07	PRJDB5792	GCA_002897815	85.45	0.00	2.69	44.44	MAG	Thermodesulfovibrionia	BMS3Bbin05	Sulfide deposits	Kato et al. 2018
Nitrospirae bacterium BMS3Abin08	PRJDB5792	GCA_002897935	91.82	0.91	2.61	47.88	MAG	Thermodesulfovibrionia	JdFR-85	Sulfide deposits	Kato et al. 2018
Nitrospirae bacterium BMS3Abin09	PRJDB5792	GCA_002898055	54.55	0.91	1.33	46.22	MAG	Thermodesulfovibrionia	UBA6902	Sulfide deposits	Kato et al. 2018

Table A3 continued.

Name	NCBI BioProject ID	GenBank assembly accession	CheckM completeness	CheckM contamination	Sequence length (Mbp)	GC Content	Type	Class	Family	Source	Reference
Nitrospirae bacterium BMS3Abin10	PRJDB5792	GCA_002897755	92.73	2.73	2.49	45.40	MAG	Thermodesulfovibrionia	BMS3Bbin08	Sulfide deposits	Kato et al. 2018
Nitrospirae bacterium BMS3Bbin05	PRJDB5792	GCA_002897855	91.82	0.00	2.46	44.67	MAG	Thermodesulfovibrionia	BMS3Bbin05	Sulfide deposits	Kato et al. 2018
Nitrospirae bacterium BMS3Bbin06	PRJDB5792	GCA_002897875	91.82	0.91	2.50	48.04	MAG	Thermodesulfovibrionia	JdFR-85	Sulfide deposits	Kato et al. 2018
Nitrospirae bacterium BMS3Bbin07	PRJDB5792	GCA_002898135	76.13	0.00	1.61	46.98	MAG	Thermodesulfovibrionia	JdFR-85	Sulfide deposits	Kato et al. 2018
Nitrospirae bacterium BMS3Bbin08	PRJDB5792	GCA_002897775	96.36	3.64	2.71	45.24	MAG	Thermodesulfovibrionia	BMS3Bbin08	Sulfide deposits	Kato et al. 2018
Nitrospirae bacterium BMS3Bbin09	PRJDB5792	GCA_002897915	86.31	0.00	1.85	45.96	MAG	Thermodesulfovibrionia	UBA6902	Sulfide deposits	Kato et al. 2018
Nitrospirae bacterium CG	PRJNA362739	GCA_002790775	96.59	0.91	1.97	40.63	MAG	Thermodesulfovibrionia	SM23-35	Terrestrial subsurface fluids	Probst et al. 2018
Nitrospirae bacterium CG	PRJNA362739	GCA_002790755	95.23	0.00	1.94	44.09	MAG	Thermodesulfovibrionia	UBA1546	Terrestrial subsurface fluids	Probst et al. 2018
Nitrospirae bacterium CG02	PRJNA362739	GCA_002780905	97.27	0.91	2.05	43.94	MAG	Thermodesulfovibrionia	UBA1546	Terrestrial subsurface fluids	Probst et al. 2018
Nitrospirae bacterium GWA2	PRJNA288027	GCA_001803605	92.27	0.91	2.29	46.70	MAG	Thermodesulfovibrionia	UBA9935	Terrestrial subsurface fluids	Ananthaman et al. 2018
Nitrospirae bacterium GWC2	PRJNA288027	GCA_001803645	69.48	0.97	2.58	42.06	MAG	Thermodesulfovibrionia	UBA6898	Terrestrial subsurface fluids	Ananthaman et al. 2016

Table A3 continued.

Name	NCBI BioProject ID	GenBank assembly accession	CheckM completeness	CheckM contamination	Sequence length (Mbp)	GC Content	Type	Class	Family	Source	Reference
Nitrospirae bacterium RIFOXYA2	PRJNA288027	GCA_001803945	94.55	0.91	1.93	43.67	MAG	Thermodesulfovibrionia	UBA1546	Terrestrial subsurface	Ananthaman et al. 2018
Nitrospirae bacterium MYbin3	PRJNA400260	GCA_002753335	96.49	0.91	2.93	44.36	MAG	Thermodesulfovibrionia	UBA9935	Freshwater sediment	Lin et al. 2018
Nitrospirae bacterium MYbinv3	PRJNA400260	GCA_002753395	94.24	3.18	3.71	44.44	MAG	Thermodesulfovibrionia	Magnetobacteriaceae	Freshwater sediment	Lin et al. 2018
Thermodesulfovibrio aggregans	PRJNA419931	GCA_002878055	76.36	0.00	1.29	34.90	MAG	Thermodesulfovibrionia	Thermodesulfovibrionaceae	Freshwater hot spring	Wilkins et al. 2019
Thermodesulfovibrio aggregans DSM 17283	PRJDB4394	GCA_001514535	100.00	0.91	2.00	34.89	Strain	Thermodesulfovibrionia	Thermodesulfovibrionaceae	Activated sludge	Matsuura et al. 2016
Thermodesulfovibrio thiophilus DSM 17215	PRJNA185673	GCA_000423865	98.11	0.00	1.87	34.43	Strain	Thermodesulfovibrionia	Thermodesulfovibrionaceae	Activated sludge	Mukherjee et al. 2017
Thermodesulfovibrio yellowstonii DSM 11347	PRJNA30733	GCA_000020985	99.08	3.67	2.00	34.13	Strain	Thermodesulfovibrionia	Thermodesulfovibrionaceae	Freshwater hydrothermal fluid	Bhatnagar et al. 2015
Nitrospirae bacterium GWC2	PRJNA288027	GCA_001803705	67.05	0.91	3.16	56.44	MAG	UBA9217	UBA9217	Terrestrial subsurface fluids	Ananthaman et al. 2016
Nitrospirae bacterium GWC2	PRJNA288027	GCA_001805055	93.18	1.42	3.32	57.15	MAG	UBA9217	UBA9217	Terrestrial subsurface fluids	Ananthaman et al. 2016 & 2018

Table A4. AAI between Nitrospirota genomes recovered by Jungbluth et al. 2017a and this study.

Nitrospirota MAGS, Jungbluth et al. (2017)	Source fluids	JdFRnit1A	JdFRnit2A	JdFRnit3A	JdFRnit4A	JdFRnit5A	JdFRnit8A	JdFRnit9A	JdFRnit6B	JdFRnit7B
GCA_002011735	U1362A	59.58	**99.99**	56.17	57.99	77.90	66.05	64.79	57.91	56.15
GCA_002011775	U1362A	62.27	78.88	58.15	60.10	**94.88**	71.04	66.19	60.30	58.15
GCA_002011715	U1362A	59.25	69.40	55.79	57.50	76.38	**96.59**	**86.17**	58.12	55.84
GCA_002010775	U1362A	57.73	69.46	53.83	**68.56**	80.83	80.32	72.33	59.07	53.87
GCA_002011745	U1362B	60.28	57.87	57.15	65.19	57.82	58.29	57.43	**99.99**	57.16
GCA_002011815	U1362A	**99.98**	59.58	59.71	59.80	59.18	59.24	58.44	60.40	59.70
GCA_002010755	U1362B	59.57	56.01	99.69	55.98	55.78	55.78	54.70	56.99	**99.69**
GCA_002011795	U1362A	59.50	55.82	**99.73**	55.98	55.48	55.77	54.79	57.00	99.73

Table A5. Distribution of KEGG-annotated genes across JdFR Nitrospirota MAGs. Pathway abbreviation: EMP: Emden-Meyerhof-Parnas.

Category	Pathway	KEGG no.	Gene	JdFR nit3A	JdFR nit7B	JdFR nit1A	JdFR nit4A	JdFR nit6B	JdFR nit8A	JdFR nit9A	JdFR nit2A	JdFR nit5A
Carbon Metabolism-Glycolysis	EMP	K00845	glk; glucokinase [EC:2.7.1.2]	1	1	1	1	1	1	1	1	1
	EMP	K01810	GPI, pgi; glucose-6-phosphate isomerase [EC:5.3.1.9]	1	1	1	1	1	1	0	1	0
	EMP	K00850	6-phosphofructokinase 1 [EC:2.7.1.11] (pfkA, PFK)	1	1	1	1	1	0	0	0	0
	EMP	K01622	fructose 1,6-bisphosphate aldolase/phosphatase [EC:4.1.2.13 3.1.3.11]	1	1	1	1	1	0	0	1	0
	EMP	K01803	TPI, tpiA; triosephosphate isomerase (TIM) [EC:5.3.1.1]	1	1	1	1	1	1	0	1	0
	EMP	K00134	GAPDH, gapA; glyceraldehyde 3-phosphate dehydrogenase [EC:1.2.1.12]	1	1	1	1	1	1	1	1	0
	EMP	K00927	PGK, pgk; phosphoglycerate kinase [EC:2.7.2.3]	1	1	1	1	1	1	1	1	0
	EMP	K15635	apgM; 2,3-bisphosphoglycerate-independent phosphoglycerate mutase [EC:5.4.2.12]	1	1	1	1	1	0	0	1	1
	EMP	K01689	ENO, eno; enolase [EC:4.2.1.11]	1	1	1	1	1	0	0	1	0
	EMP	K00873	PK, pyk; pyruvate kinase [EC:2.7.1.40]	1	1	1	1	1	1	1	1	1

Table A5 continued. Pathway abbreviation: PP: Pentose Phosphate; ED 1: Entner-Doudoroff 1.

Category	Pathway	KEGG no.	Gene	JdFR nit3A	JdFR nit7B	JdFR nit1A	JdFR nit4A	JdFR nit6B	JdFR nit8A	JdFR nit9A	JdFR nit2A	JdFR nit5A
Carbon Metabolism-Glycolysis	PP	K00036	G6PD, zwf; glucose-6-phosphate 1-dehydrogenase [EC:1.1.1.49 1.1.1.363]	1	1	1	1	1	1	1	1	0
	PP	K01057	PGLS, pgl, devB; 6-phosphogluconolactonase [EC:3.1.1.31]	0	0	1	1	1	1	1	1	1
	PP	K00033	PGD, gnd, gntZ; 6-phosphogluconate dehydrogenase [EC:1.1.1.44 1.1.1.343]	1	1	1	1	1	1	1	1	0
	PP	K01808	rpiB; ribose 5-phosphate isomerase B [EC:5.3.1.6]	1	1	1	1	1	0	0	1	0
	PP	K01783	rpe, RPE; ribulose-phosphate 3-epimerase [EC:5.1.3.1]	1	1	1	1	1	1	1	1	0
	PP	K00615	tktA, tktB; transketolase [EC:2.2.1.1]	1	1	1	1	1	1	0	1	0
	PP	K00616	talA, talB; transaldolase [EC:2.2.1.2]	1	1	1	0	1	1	0	1	0
	ED 1	K00036	G6PD, zwf; glucose-6-phosphate 1-dehydrogenase [EC:1.1.1.49 1.1.1.363]	1	1	1	1	1	1	1	1	0
	ED 1	K01057	PGLS, pgl, devB; 6-phosphogluconolactonase [EC:3.1.1.31]	0	0	1	1	1	1	1	1	1
	ED 1	K01690	edd, phosphogluconate dehydratase	0	0	0	0	0	0	0	0	0
	ED 1	K01625	eda; 2-dehydro-3-deoxyphosphogluconate aldolase / (4S)-4-hydroxy-2-oxoglutarate aldolase [EC:4.1.2.14 4.1.3.42]	0	0	1	0	0	0	0	0	0

Table A5 continued. Pathway abbreviation: ED 2: Entner-Doudoroff 2; GNG: Gluconeogenesis.

Category	Pathway	KEGG no.	Gene	JdFR nit3A	JdFR nit7B	JdFR nit1A	JdFR nit4A	JdFR nit6B	JdFR nit8A	JdFR nit9A	JdFR nit2A	JdFR nit5A
Carbon Metabolism-Glycolysis	ED 2	K13937	H6PD,Hexose-6-phosphate dehydrogenase	0	0	0	0	0	0	0	0	0
	ED 2	K01690	edd, phosphogluconate dehydratase	0	0	0	0	0	0	0	0	0
	ED 2	K01625	eda; 2-dehydro-3-deoxyphosphogluconate aldolase / (4S)-4-hydroxy-2-oxoglutarate aldolase [EC:4.1.2.14 4.1.3.42]	0	0	1	0	0	0	0	0	0
Carbon Metabolism-GNG	GNG	K01084	G6PC; glucose-6-phosphatase [EC:3.1.3.9]	0	0	0	0	0	0	0	0	0
	GNG	K01006	pdK; pyruvate, orthophosphate dikinase [EC:2.7.9.1]	1	1	1	1	1	1	0	1	0
	GNG	K01689	ENO, eno; enolase [EC:4.2.1.11]	1	1	1	1	1	0	0	1	0
	GNG	K15635	apgM; 2,3-bisphosphoglycerate-independent phosphoglycerate mutase [EC:5.4.2.12]	1	1	1	1	1	0	0	1	1
	GNG	K00927	PGK, pgk; phosphoglycerate kinase [EC:2.7.2.3]	1	1	1	1	1	1	1	1	0
	GNG	K00134	GAPDH, gapA; glyceraldehyde 3-phosphate dehydrogenase [EC:1.2.1.12]	1	1	1	1	1	1	1	1	0

Table A5 continued. Pathway abbreviation: GNG: Gluconeogenesis.

Category	Pathway	KEGG no.	Gene	JdFR nit3A	JdFR nit7B	JdFR nit1A	JdFR nit4A	JdFR nit6B	JdFR nit8A	JdFR nit9A	JdFR nit2A	JdFR nit5A
Carbon Metabolism-GNG	GNG	K01622	fructose 1,6-bisphosphate aldolase/phosphatase [EC:4.1.2.13 3.1.3.11]	1	1	1	1	1	0	0	1	0
	GNG	K01803	TPI, tpiA; triosephosphate isomerase (TIM) [EC:5.3.1.1]	1	1	1	1	1	1	0	1	0
	GNG	K01810	GPI, pgi; glucose-6-phosphate isomerase [EC:5.3.1.9]	1	1	1	1	1	1	0	1	0
Carbon Metabolism-Glycogen Metabolism	Glycogen Synthesis/ Breakdown	K15778	pmm-pgm; phosphomannomutase / phosphoglucomutase [EC:5.4.2.8 5.4.2.2]	1	1	1	1	1	1	0	0	0
	Glycogen Synthesis	K00975	glgC; glucose-1-phosphate adenylyltransferase [EC:2.7.7.27]	1	1	1	1	1	1	1	0	0
	Glycogen Synthesis	K00703	glgA; glucose synthase [EC:2.4.1.21]	1	1	1	1	1	1	1	1	1
	Glycogen Synthesis	K00700	GBE1, glgB; 1,4-alpha-glucan branching enzyme [EC:2.4.1.18]	1	1	1	1	1	1	1	0	1
	Glycogen Breakdown	K00688	PYG, glgP; glycogen phosphorylase [EC:2.4.1.1]	1	1	1	0	1	1	0	1	0
	Glycogen Breakdown	K02438	glgX; glycogen debranching enzyme [EC:3.2.1.196]	0	0	0	0	0	0	0	0	0

Table A5 continued. TCA: Tricarboxylic Acid.

Category	Pathway	KEGG no.	Gene	JdFR nit3A	JdFR nit7B	JdFR nit1A	JdFR nit4A	JdFR nit6B	JdFR nit8A	JdFR nit9A	JdFR nit2A	JdFR nit5A
Carbon Metabolism-TCA Cycle	TCA Cycle	K01681	ACO, acnA; aconitate hydratase [EC:4.2.1.3]	1	1	1	1	1	1	1	1	0
	TCA Cycle	K00030	IDH3; isocitrate dehydrogenase (NAD+) [EC:1.1.1.41]	1	1	1	1	1	1	0	1	0
	TCA Cycle	K00174	korA, oorA, oforA; 2-oxoglutarate/2-oxoacid ferredoxin oxidoreductase subunit alpha [EC:1.2.7.3 1.2.7.11]	1	1	1	1	1	0	0	1	0
	TCA Cycle	K00175	korB, oorB, oforB; 2-oxoglutarate/2-oxoacid ferredoxin oxidoreductase subunit beta [EC:1.2.7.3 1.2.7.11]	1	1	1	1	1	0	0	1	0
	TCA Cycle	K01899, K01902	succinyl-CoA synthetase alpha subunit [EC 6.2.1.5]	0	0	0	0	0	0	0	0	0
	TCA Cycle	K01900, K01903	succinyl-CoA synthetase beta subunit [EC 6.2.1.5]	0	0	0	0	0	0	0	0	0
	TCA Cycle	K00239	sdhA, frdA; succinate dehydrogenase / fumarate reductase, flavoprotein subunit [EC:1.3.5.1 1.3.5.4]	0	0	0	0	1	0	0	1	0
	TCA Cycle	K00240	sdhB, frdB; succinate dehydrogenase / fumarate reductase, iron-sulfur subunit [EC:1.3.5.1 1.3.5.4]	0	0	0	0	1	0	0	1	0
	TCA Cycle	K00241	sdhC, frdC; succinate dehydrogenase / fumarate reductase, cytochrome b subunit	0	0	0	0	1	0	0	1	0

Table A5 continued. TCA: Tricarboxylic Acid; CBB: Calvin-Benson-Bassham.

Category	Pathway	KEGG no.	Gene	JdFR nit3A	JdFR nit7B	JdFR nit1A	JdFR nit4A	JdFR nit6B	JdFR nit8A	JdFR nit9A	JdFR nit2A	JdFR nit5A
Carbon Metabolism-TCA Cycle	TCA Cycle	K00242	sdhD, frdD; succinate dehydrogenase / fumarate reductase, membrane anchor subunit	0	0	0	0	1	0	0	0	0
	TCA Cycle	K01676	E4.2.1.2A, fumA, fumB; fumarate hydratase, class I [EC:4.2.1.2]	1	1	1	1	1	1	0	0	0
	TCA Cycle	K00024	mdh; malate dehydrogenase [EC:1.1.1.37]	1	1	1	1	1	0	1	1	0
	TCA Cycle	K01647	CS, gltA; citrate synthase [EC:2.3.3.1]	0	0	0	1	0	0	0	0	0
Carbon Metabolism-CBB Cycle	CBB Cycle	K01601	ribulose-bisphosphate carboxylase large chain [EC:4.1.1.39] (rbcL)	1	1	1	1	1	0	0	1	0
	CBB Cycle	K00927	PGK, pgk; phosphoglycerate kinase [EC:2.7.2.3]	1	1	1	1	1	1	1	1	0
	CBB Cycle	K00134	GAPDH, gapA; glyceraldehyde 3-phosphate dehydrogenase [EC:1.2.1.12]	1	1	1	1	1	1	1	1	0
	CBB Cycle	K00855	prkB, phosphoribulokinase [EC:2.7.1.19]	0	0	0	0	0	0	0	0	0
	CBB Cycle	K01783	rpe, RPE; ribulose-phosphate 3-epimerase [EC:5.1.3.1]	1	1	1	1	1	1	1	1	0
	CBB Cycle	K01621	xfp, xpk; xylulose-5-phosphate/fructose-6-phosphate phosphoketolase [EC:4.1.2.9 4.1.2.22]	0	0	0	0	0	1	1	0	0

Table A5 continued. TCA: Tricarboxylic Acid; CBB: Calvin-Benson-Bassham; WLMB: Wood-Ljungdahl Methyl Branch.

Category	Pathway	KEGG no.	Gene	JdFR nit3A	JdFR nit7B	JdFR nit1A	JdFR nit4A	JdFR nit6B	JdFR nit8A	JdFR nit9A	JdFR nit2A	JdFR nit5A
Carbon Metabolism-CBB Cycle	CBB Cycle	K00615	tktA, tktB; transketolase [EC:2.2.1.1]	1	1	1	1	1	1	0	1	0
	CBB Cycle	K01807	rpiA; ribose 5-phosphate isomerase A [EC:5.3.1.6]	0	0	0	0	0	0	0	0	0
	CBB Cycle	K01622	fructose 1,6-bisphosphate aldolase/phosphatase [EC:4.1.2.13 3.1.3.11]	1	1	1	1	1	0	0	1	0
	CBB Cycle	K03841	FBP, fbp; fructose-1,6-bisphosphatase I [EC:3.1.3.11]	0	0	1	1	0	0	0	0	0
Carbon Metabolism Reverse-TCA Cycle	Reverse-TCA Cycle	K15230	aclA, ATP-citrate lyase alpha-subunit [EC:2.3.3.8]	0	0	0	0	0	0	0	0	0
	Reverse-TCA Cycle	K15231	aclB, ATP-citrate lyase beta-subunit [EC:2.3.3.8]	0	0	0	0	0	0	0	0	0
	Reverse-TCA Cycle	K15232	ccsA, citryl-CoA synthetase large subunit [EC:6.2.1.18]	0	0	0	0	0	0	0	0	0
	Reverse-TCA Cycle	K15233	ccsB, citryl-CoA synthetase small subunit [EC:6.2.1.18]	0	0	0	0	0	0	0	0	0
	Reverse-TCA Cycle	K15234	citryl-CoA lyase [EC:4.1.3.34]	0	0	0	0	0	0	0	0	0
Carbon Metabolism-Carbon Fixation	WLMB	K05299	fdhA; formate dehydrogenase alpha subunit [EC:1.2.1.43]	1	1	1	1	1	1	0	1	0
	WLMB	K15022	fdhB; formate dehydrogenase beta subunit [EC:1.2.1.43]	0	0	0	0	0	0	0	0	0
	WLMB	K01938	formate--tetrahydrofolate ligase [EC:6.3.4.3] (fhs)	1	1	1	1	1	0	0	1	0

Table A5 continued. WLMB: Wood-Ljungdahl Methyl Branch; WLCB: Wood-Ljungdahl Carbonyl Branch.

Category	Pathway	KEGG no.	Gene	JdFR nit3A	JdFR nit7B	JdFR nit1A	JdFR nit4A	JdFR nit6B	JdFR nit8A	JdFR nit9A	JdFR nit2A	JdFR nit5A
Carbon Metabolism-Carbon Fixation	WLMB	K01491	folD; methylenetetrahydrofolate dehydrogenase (NADP+) / methenyltetrahydrofolate cyclohydrolase [EC:1.5.1.5 3.5.4.9]	1	1	1	1	1	0	0	1	0
	WLMB	K00297	metF, MTHFR; methylenetetrahydrofolate reductase (NADPH) [EC:1.5.1.20]	1	1	1	1	1	1	1	1	0
	WLMB	K00197	cdhE, acsC; acetyl-CoA decarbonylase/synthase complex subunit gamma [EC:2.1.1.245]	1	1	1	0	1	0	1	1	0
	WLMB	K00194	cdhD, acsD; acetyl-CoA decarbonylase/synthase complex subunit delta [EC:2.1.1.245]	1	1	1	0	1	0	1	1	0
	WLMB	K15023	acsE; 5-methyltetrahydrofolate corrinoid/iron sulfur protein methyltransferase [EC:2.1.1.258]	1	1	1	0	1	0	1	1	0
	WLCB	K14138	acsB; acetyl-CoA synthase [EC:2.3.1.169]	1	1	1	0	1	0	1	1	0
	WLCB	K00198	cooS, acsA; anaerobic carbon-monoxide dehydrogenase catalytic subunit [EC:1.2.7.4]	1	1	1	1	1	0	1	1	0

Table A5 continued. 3HPB: 3-Hydroxypropionate Bicycle.

Category	Pathway	KEGG no.	Gene	JdFR nit3A	JdFR nit7B	JdFR nit1A	JdFR nit4A	JdFR nit6B	JdFR nit8A	JdFR nit9A	JdFR nit2A	JdFR nit5A
Carbon Metabolism-Carbon Fixation	3HPB	K00169	porA; pyruvate ferredoxin oxidoreductase alpha subunit [EC:1.2.7.1]	1	1	1	1	1	1	0	1	0
	3HPB	K00170	porB; pyruvate ferredoxin oxidoreductase beta subunit [EC:1.2.7.1]	1	1	1	1	1	1	0	1	0
	3HPB	K00171	porD; pyruvate ferredoxin oxidoreductase delta subunit [EC:1.2.7.1]	1	1	1	1	1	1	0	1	0
	3HPB	K00172	porG; pyruvate ferredoxin oxidoreductase gamma subunit [EC:1.2.7.1]	1	1	1	1	1	1	0	1	0
	3HPB	K01006	ppdK; pyruvate, orthophosphate dikinase [EC:2.7.9.1]	1	1	1	1	1	1	0	1	0
	3HPB	K01595	ppc; phosphoenolpyruvate carboxylase [EC:4.1.1.31]	1	1	0	0	0	0	0	0	0
	3HPB	K00024	mdh; malate dehydrogenase [EC:1.1.1.37]	1	1	1	1	1	0	1	1	0
	3HPB	K14471	smtA1, succinyl-CoA:(S)-malate CoA-transferase subunit A [EC:2.8.3.22]	0	0	0	0	0	0	0	0	0
	3HPB	K14472	smtB, succinyl-CoA:(S)-malate CoA-transferase subunit B [EC:2.8.3.22]	0	0	0	0	0	0	0	0	0
	3HPB	K08691	mcl, malyl-CoA/(S)-citramalyl-CoA lyase [EC:4.1.3.24 4.1.3.25]	0	0	0	0	0	0	0	0	0
	3HPB	K02160	accB, bccP; acetyl-CoA carboxylase biotin carboxyl carrier protein	1	1	0	1	1	0	0	1	1

Table A5 continued. 3HPB: 3-Hydroxypropionate Bicycle.

Category	Pathway	KEGG no.	Gene	JdFR nit3A	JdFR nit7B	JdFR nit1A	JdFR nit4A	JdFR nit6B	JdFR nit8A	JdFR nit9A	JdFR nit2A	JdFR nit5A
Carbon Metabolism-Carbon Fixation	3HPB	K01961	accC; acetyl-CoA carboxylase, biotin carboxylase subunit [EC:6.4.1.2 6.3.4.14]	1	1	1	1	1	1	0	1	1
	3HPB	K01962	accA; acetyl-CoA carboxylase carboxyl transferase subunit alpha [EC:6.4.1.2]	1	1	1	1	1	1	0	1	1
	3HPB	K01963	accD; acetyl-CoA carboxylase carboxyl transferase subunit beta [EC:6.4.1.2]	1	1	1	1	1	0	1	0	0
	3HPB	K14468	mcr, malonyl-CoA reductase / 3-hydroxypropionate dehydrogenase (NADP+) [EC:1.2.1.75 1.1.1.298]	0	0	0	0	0	0	0	0	0
	3HPB	K15017	malonyl-CoA/succinyl-CoA reductase (NADPH) [EC:1.2.1.751.2.1.76]	0	0	0	0	0	0	0	0	0
	3HPB	K15039	3-hydroxypropionate dehydrogenase (NADP+) [EC:1.1.1.298]	0	0	0	0	0	0	0	0	0
	3HPB	K14469	acrylyl-CoA reductase (NADPH) / 3-hydroxypropionyl-CoA dehydratase / 3-hydroxypropionyl-CoA synthetase [EC:1.3.1.844.2.1.116 6.2.1.36]	0	0	0	0	0	0	0	0	0

Table A5 continued. 3HPB: 3-Hydroxypropionate Bicycle; 4HB/3HP: 4-Hydroxybutyrate/3-hydroxypropionate.

Category	Pathway	KEGG no.	Gene	JdFR nit3A	JdFR nit7B	JdFR nit1A	JdFR nit4A	JdFR nit6B	JdFR nit8A	JdFR nit9A	JdFR nit2A	JdFR nit5A
Carbon Metabolism-Carbon Fixation	3HPB	K15018	3-hydroxypropionyl-coenzyme A synthetase [EC:6.2.1.36]	0	0	0	0	0	0	0	0	0
	3HPB	K15019	3-hydroxypropionyl-coenzyme A dehydratase [EC:4.2.1.116]	0	0	0	0	0	0	0	0	0
	3HPB	K15020	acryloyl-coenzyme A reductase [EC:1.3.1.84]	0	0	0	0	0	0	0	0	0
	3HPB	K14449	mch, 2-methylfumaryl-CoA hydratase [EC:4.2.1.148]	0	0	0	0	0	0	0	0	0
	3HPB	K14470	mct 2-methylfumaryl-CoA isomerase [EC:5.4.1.3]	0	0	0	0	0	0	0	0	0
	3HPB	K09709	meh, 3-methylfumaryl-CoA hydratase [EC:4.2.1.153]	0	0	0	0	0	0	0	0	0
	4HB/3HP	K02160	accB, bccP; acetyl-CoA carboxylase biotin carboxyl carrier protein	1	1	0	1	1	0	0	1	1
	4HB/3HP	K01961	accC; acetyl-CoA carboxylase, biotin carboxylase subunit [EC:6.4.1.2 6.3.4.14]	1	1	1	1	1	1	0	1	1
	4HB/3HP	K01962	accA; acetyl-CoA carboxylase carboxyl transferase subunit alpha [EC:6.4.1.2]	1	1	1	1	1	1	0	1	1
	4HB/3HP	K01963	accD; acetyl-CoA carboxylase carboxyl transferase subunit beta [EC:6.4.1.2]	1	1	1	1	1	0	1	0	0

Table A5 continued. 4HB/3HP: 4-Hydroxybutyrate/3-Hydroxypropionate.

Category	Pathway	KEGG no.	Gene	JdFR nit3A	JdFR nit7B	JdFR nit1A	JdFR nit4A	JdFR nit6B	JdFR nit8A	JdFR nit9A	JdFR nit2A	JdFR nit5A
Carbon Metabolism-Carbon Fixation	4HB/3HP	K18594	3-hydroxypropionyl-CoA synthetase (ADP-forming) [EC:6.2.1.-]	0	0	0	0	0	0	0	0	0
	4HB/3HP	K14469	acrylyl-CoA reductase (NADPH) / 3-hydroxypropionyl-CoA dehydratase / 3-hydroxypropionyl-CoA synthetase [EC:1.3.1.844.2.1.116 6.2.1.36]	0	0	0	0	0	0	0	0	0
	4HB/3HP	K15019	3-hydroxypropionyl-coenzyme A dehydratase [EC:4.2.1.116]	0	0	0	0	0	0	0	0	0
	4HB/3HP	K05606	methylmalonyl-CoA/ethylmalonyl-CoA epimerase [EC:5.1.99.1]	0	0	0	0	0	0	0	0	0
	4HB/3HP	K01847	methylmalonyl-CoA mutase [EC:5.4.99.2]	0	0	0	0	0	0	0	0	0
	4HB/3HP	K01848	methylmalonyl-CoA mutase, N-terminal domain [EC:5.4.99.2]	0	0	0	0	0	0	0	0	0
	4HB/3HP	K01849	methylmalonyl-CoA mutase, C-terminal domain [EC:5.4.99.2]	0	0	0	0	0	0	0	0	0
	4HB/3HP	K18593	4-hydroxybutyrate---CoA ligase (ADP-forming) [EC:6.2.1.56]	0	0	0	0	0	0	0	0	0
	4HB/3HP	K14534	4-hydroxybutyryl-CoA dehydratase / vinylacetyl-CoA-Delta-isomerase [EC:4.2.1.120 5.3.3.3]	0	0	0	0	0	0	0	0	0

Table A5 continued. 4HB/3HP: 4-Hydroxybutyrate/3-Hydroxypropionate.

Category	Pathway	KEGG no.	Gene	JdFR nit3A	JdFR nit7B	JdFR nit1A	JdFR nit4A	JdFR nit6B	JdFR nit8A	JdFR nit9A	JdFR nit2A	JdFR nit5A
Carbon Metabolism-Carbon Fixation	4HB/3HP	K15016	enoyl-CoA hydratase / 3-hydroxyacyl-CoA dehydrogenase [EC:4.2.1.17 1.1.1.35]	0	0	0	0	0	0	0	0	0
	4HB/3HP	K00626	acetyl-CoA C-acetyltransferase [EC:2.3.1.9]	0	0	0	0	0	0	0	0	0
Carbon Metabolism-Mixed Acid Fermentation	Pyruvate to lactate	K03778	ldhA; D-lactate dehydrogenase [EC:1.1.1.28]	0	0	1	0	0	0	0	0	0
	Formate dehydrogenase	K00123	fdoG, fdhF; formate dehydrogenase major subunit [EC:1.2.1.2]	0	0	1	1	1	0	0	0	0
	Formate dehydrogenase	K00124	fdoH; formate dehydrogenase iron-sulfur subunit	1	1	1	1	1	0	0	0	0
	Formate dehydrogenase	K00127	fdoI; formate dehydrogenase subunit gamma	0	0	0	1	1	0	0	0	0
	Pyruvate to Acetyl phosphate to acetate	K00158	pyruvate oxidase [EC:1.2.3.3] (E1.2.3.3, poxL)	0	0	0	1	1	0	0	0	1
	Pyruvate to Acetyl phosphate to acetate	K01512	acyP; acylphosphatase [EC:3.6.1.7]	1	1	1	1	1	0	0	0	1
	PEP to Succinate via OAA, malate & fumarate	K00024	mdh; malate dehydrogenase [EC:1.1.1.37]	1	1	1	1	1	0	1	1	0

Table A5 continued.

Category	Pathway	KEGG no.	Gene	JdFR nit3A	JdFR nit7B	JdFR nit1A	JdFR nit4A	JdFR nit6B	JdFR nit8A	JdFR nit9A	JdFR nit2A	JdFR nit5A
Carbon-Metabolism-Carbon Degradation	Carbon Degradation	K05349	bglX; beta-glucosidase [EC:3.2.1.21]	0	0	1	0	0	0	0	1	0
	Carbon Degradation	K08679	E5.1.3.6; UDP-glucuronate 4-epimerase [EC:5.1.3.6]	1	1	0	0	1	0	0	1	0
	Carbon Degradation	K01207	nagZ; beta-N-acetylhexosaminidase [EC:3.2.1.52]	0	0	0	0	0	1	0	0	0
Carbon Metabolism-Methanogenesis	Methanol to methane	K14080	mtaA; [methyl-Co(III) methanol-specific corrinoid protein]:coenzyme M methyltransferase [EC:2.1.1.246]	0	0	0	1	1	0	0	0	0
	Methanol to methane	K04480	mtaB; methanol---5-hydroxybenzimidazolylcobamide Co-methyltransferase [EC:2.1.1.90]	0	0	0	0	0	0	0	0	0
	Methanol to methane	K14081	mtaC ;methanol corrinoid protein	0	0	0	0	0	0	0	0	0
	Acetate to methane	K00194	cdhD, acsD; acetyl-CoA decarbonylase/synthase complex subunit delta [EC:2.1.1.245]	1	1	1	0	1	0	1	1	0
	Acetate to methane	K00197	cdhE, acsC; acetyl-CoA decarbonylase/synthase complex subunit gamma [EC:2.1.1.245]	1	1	1	0	1	0	1	1	0

Table A5 continued. Pathway abbreviation: DNR: Dissimilatory Nitrate Reduction; ANR: Assimilatory Nitrate Reduction.

Category	Pathway	KEGG no.	Gene	JdFR nit3A	JdFR nit7B	JdFR nit1A	JdFR nit4A	JdFR nit6B	JdFR nit8A	JdFR nit9A	JdFR nit2A	JdFR nit5A
Carbon Metabolism-Methanogenesis	CO2 to methane	K11261	fwdE, fmdE; formylmethanofuran dehydrogenase subunit E [EC:1.2.7.12]	1	1	1	1	1	1	1	1	0
	CO2 to methane	K00577	mtrA; tetrahydromethanopterin S-methyltransferase subunit A [EC:2.1.1.86]	0	0	0	0	0	0	0	1	1
	CO2 to methane	K00584	mtrH; tetrahydromethanopterin S-methyltransferase subunit H [EC:2.1.1.86]	0	0	0	1	1	0	0	0	0
Carbon Metabolism-Coenzyme B/Coenzyme M regeneration	Coenzyme B/Coenzyme M regeneration	K08264	hdrD; heterodisulfide reductase subunit D [EC:1.8.98.1]	0	0	0	1	1	0	0	0	0
Nitrogen Metabolism-DNR	DNR	K00370	narG, narZ, nxrA; nitrate reductase / nitrite oxidoreductase, alpha subunit [EC:1.7.5.1 1.7.99.-]	1	1	0	0	0	0	0	0	0
	DNR	K00371	narH, narY, nxrB; nitrate reductase / nitrite oxidoreductase, beta subunit [EC:1.7.5.1 1.7.99.-]	1	1	0	0	0	0	0	0	0
	DNR	K00374	narI, narV; nitrate reductase gamma subunit [EC:1.7.5.1 1.7.99.-]	1	1	1	1	1	1	1	1	0

Table A5 continued. Pathway abbreviation: DNR: Dissimilatory Nitrate Reduction; ANR: Assimilatory Nitrate Reduction.

Category	Pathway	KEGG no.	Gene	JdFR nit3A	JdFR nit7B	JdFR nit1A	JdFR nit4A	JdFR nit6B	JdFR nit8A	JdFR nit9A	JdFR nit2A	JdFR nit5A
Nitrogen Metabolism-DNR	DNR	K02567	napA; periplasmic nitrate reductase NapA [EC:1.7.99.-]	1	1	0	1	0	1	1	0	0
	DNR	K02568	napB; cytochrome c-type protein NapB	0	0	0	0	0	1	0	0	0
Nitrogen Metabolism-ANR	ANR	K00362	nirB; nitrite reductase (NADH) large subunit [EC:1.7.1.15]	1	1	1	1	0	1	0	0	0
	ANR	K00363	nirD; nitrite reductase (NADH) small subunit [EC:1.7.1.15]	0	0	0	0	0	0	0	0	0
Nitrogen Metabolism-Nitric Oxide Reduction	Nitric oxide reduction	K02305	norC; nitric oxide reductase subunit C	1	1	0	1	0	0	0	1	0
	Nitric oxide reduction	K04561	norB; nitric oxide reductase subunit B [EC:1.7.2.5]	1	1	0	1	0	0	0	1	0
Nitrogen Metabolism-Nitrogen fixation	Nitrogen fixation	K02586	nifD; nitrogenase molybdenum-iron protein alpha chain [EC:1.18.6.1]	0	0	0	0	1	0	0	0	0
	Nitrogen fixation	K02588	nifH; nitrogenase iron protein NifH [EC:1.18.6.1]	0	0	0	0	1	0	0	0	0
	Nitrogen fixation	K02591	nifK; nitrogenase molybdenum-iron protein beta chain [EC:1.18.6.1]	0	0	0	0	1	0	0	0	0
Nitrogen Metabolism-Hydroxylamine Oxidation	Hydroxylamine oxidation	K10535	hao; hydroxylamine dehydrogenase [EC:1.7.2.6]	1	1	1	1	1	1	0	1	0

Table A5 continued. Pathway abbreviation: DSR: Dissimilatory Sulfate Reduction.

Category	Pathway	KEGG no.	Gene	JdFR nit3A	JdFR nit7B	JdFR nit1A	JdFR nit4A	JdFR nit6B	JdFR nit8A	JdFR nit9A	JdFR nit2A	JdFR nit5A
Sulfur Metabolism-DSR	DSR	K00958	sat, met3; sulfate adenylyltransferase [EC:2.7.7.4]	1	1	1	1	1	0	0	1	0
	DSR	K00394	aprA; adenylylsulfate reductase, subunit A [EC:1.8.99.2]	1	1	1	1	1	0	0	1	0
	DSR	K00395	aprB; adenylylsulfate reductase, subunit B [EC:1.8.99.2]	1	1	1	1	1	0	0	1	0
	DSR	K11180	dsrA; dissimilatory sulfite reductase alpha subunit [EC:1.8.99.5]	1	1	1	1	1	1	1	1	0
	DSR	K11181	dsrB; dissimilatory sulfite reductase beta subunit [EC:1.8.99.5]	1	1	1	1	1	1	1	1	0
	DSR	K16885	quinone-modifying oxidoreductase, subunit QmoA (qmoA)	1	1	1	1	1	0	0	1	0
	DSR	K16886	quinone-modifying oxidoreductase, subunit QmoB (qmoB)	1	1	1	1	1	0	0	1	0
	DSR	K16887	quinone-modifying oxidoreductase, subunit QmoC (qmoC)	1	1	1	1	1	0	0	1	0
Sulfur Metabolism-Sulfur oxidation	Sulfur oxidation	K17218	sqr; sulfide:quinone oxidoreductase [EC:1.8.5.4]	1	1	1	1	1	0	0	1	0
Sulfur Metabolism-Thiosulfate oxidation	Thiosulfate oxidation	K17222	soxA; sulfur-oxidizing protein SoxA	0	0	0	1	0	0	0	0	0

Table A5 continued. Pathway abbreviation: DSR: Dissimilatory Sulfate Reduction.

Category	Pathway	KEGG no.	Gene	JdFR nit3A	JdFR nit7B	JdFR nit1A	JdFR nit4A	JdFR nit6B	JdFR nit8A	JdFR nit9A	JdFR nit2A	JdFR nit5A
Sulfur Metabolism-Thiosulfate oxidation	Thiosulfate oxidation	K17223	soxX; sulfur-oxidizing protein SoxX	0	0	0	1	0	0	0	0	0
	Thiosulfate oxidation	K17224	soxB; sulfur-oxidizing protein SoxB	0	0	0	1	0	0	0	0	0
	Thiosulfate oxidation	K17226	soxY; sulfur-oxidizing protein SoxY	0	0	0	1	0	0	0	0	0
	Thiosulfate oxidation	K17227	soxZ; sulfur-oxidizing protein SoxZ	0	0	0	1	0	0	0	0	0
	Thiosulfate oxidation	K19713	thiosulfate dehydrogenase [EC:1.8.2.2] (tsdA)	0	0	0	0	0	0	0	1	0
	Sulfur reduction to H_2S	K17993	hydA; sulfhydrogenase subunit alpha [EC:1.12.1.3 1.12.1.5]	1	1	1	1	1	0	0	0	0
Sulfur Metabolism-Sulfur reduction to H_2S	Sulfur reduction to H_2S	K17994	hydD; sulfhydrogenase subunit delta [EC:1.12.1.3 1.12.1.5]	1	1	1	1	1	0	0	0	0
	Sulfur reduction to H_2S	K17996	hydB; sulfhydrogenase subunit beta (sulfur reductase) [EC:1.12.98.4]	0	0	1	1	1	0	0	0	0
Hydrogenase	NiFe Hydrogenase	K00437	[NiFe] hydrogenase large subunit [EC:1.12.2.1] (hydB2)	0	1	0	1	1	0	0	0	0
	NiFe Hydrogenase	K03620	hyaC; Ni/Fe-hydrogenase 1 B-type cytochrome subunit	1	1	1	1	1	0	0	1	0
	NiFe Hydrogenase	K06281	hyaB, hybC; hydrogenase large subunit [EC:1.12.99.6]	1	1	1	1	1	0	0	1	0
	NiFe Hydrogenase	K06282	hyaA, hybO; hydrogenase small subunit [EC:1.12.99.6]	1	1	1	1	1	0	0	1	0

Table A5 continued. Pathway abbreviation: RNF: *Rhodobacter* Nitrogen Fixation.

Category	Pathway	KEGG no.	Gene	JdFR nit3A	JdFR nit7B	JdFR nit1A	JdFR nit4A	JdFR nit6B	JdFR nit8A	JdFR nit9A	JdFR nit2A	JdFR nit5A
Hydrogenase	NiFe Hydrogenase	K00436	NAD-reducing hydrogenase large subunit [EC:1.12.1.2] (hoxH)	0	0	0	0	1	0	0	0	0
	FeFe Hydrogenase	K17992	NADP-reducing hydrogenase subunit HndB [EC:1.12.1.3] (hndB)	1	1	1	1	1	0	0	0	0
	FeFe Hydrogenase	K18331	NADP-reducing hydrogenase subunit HndC [EC:1.12.1.3] (hndC)	0	1	1	0	1	0	0	1	0
Oxidative Phosphorylation	RNF Complex	K03612	Na+-translocating ferredoxin:NAD+ oxidoreductase (rnfG)	1	1	1	0	0	0	0	1	0
	RNF Complex	K03613	Na+-translocating ferredoxin:NAD+ oxidoreductase (rnfE)	1	1	1	0	1	0	0	1	0
	RNF Complex	K03614	Na+-translocating ferredoxin:NAD+ oxidoreductase (rnfD)	1	1	1	0	1	0	0	1	0
	RNF Complex	K03615	Na+-translocating ferredoxin:NAD+ oxidoreductase (rnfC)	1	1	1	0	1	0	0	1	0
	RNF Complex	K03616	Na+-translocating ferredoxin:NAD+ oxidoreductase (rnfB)	1	1	1	0	1	0	0	1	0
	RNF Complex	K03617	Na+-translocating ferredoxin:NAD+ oxidoreductase (rnfA)	1	1	1	0	1	0	0	1	0

Table A5 continued. Pathway abbreviation: NUO: NADH-Quinone Oxidoreductase.

Category	Pathway	KEGG no.	Gene	JdFR nit3A	JdFR nit7B	JdFR nit1A	JdFR nit4A	JdFR nit6B	JdFR nit8A	JdFR nit9A	JdFR nit2A	JdFR nit5A
Oxidative Phosphorylation	NUO Complex	K00330	nuoA; NADH-quinone oxidoreductase subunit A [EC:1.6.5.3]	1	1	1	1	1	1	0	1	0
	NUO Complex	K00331	nuoB; NADH-quinone oxidoreductase subunit B [EC:1.6.5.3]	1	1	1	1	1	1	0	1	0
	NUO Complex	K00332	nuoC; NADH-quinone oxidoreductase subunit C [EC:1.6.5.3]	1	1	1	1	1	1	0	1	0
	NUO Complex	K00333	nuoD; NADH-quinone oxidoreductase subunit D [EC:1.6.5.3]	1	1	1	1	1	1	0	1	0
	NUO Complex	K00334	nuoE; NADH-quinone oxidoreductase subunit E [EC:1.6.5.3]	1	1	1	1	1	1	1	1	0
	NUO Complex	K00335	nuoF; NADH-quinone oxidoreductase subunit F [EC:1.6.5.3]	1	1	1	1	1	1	0	1	0
	NUO Complex	K00337	nuoH; NADH-quinone oxidoreductase subunit H [EC:1.6.5.3]	1	1	1	1	1	1	0	1	0
	NUO Complex	K00338	nuoI; NADH-quinone oxidoreductase subunit I [EC:1.6.5.3]	1	1	1	1	1	1	0	1	0
	NUO Complex	K00339	nuoJ; NADH-quinone oxidoreductase subunit J [EC:1.6.5.3]	1	1	1	1	1	1	0	1	0
	NUO Complex	K00340	nuoK; NADH-quinone oxidoreductase subunit K [EC:1.6.5.3]	1	1	1	1	1	1	0	1	0

Table A5 continued. Pathway abbreviation: NUO: NADH-Quinone Oxidoreductase.

Category	Pathway	KEGG no.	Gene	JdFR nit3A	JdFR nit7B	JdFR nit1A	JdFR nit4A	JdFR nit6B	JdFR nit8A	JdFR nit9A	JdFR nit2A	JdFR nit5A
Oxidative Phosphorylation	NUO Complex	K00341	nuoL; NADH-quinone oxidoreductase subunit L [EC:1.6.5.3]	1	1	1	1	1	1	0	1	0
	NUO Complex	K00342	nuoM; NADH-quinone oxidoreductase subunit M [EC:1.6.5.3]	1	1	1	1	1	1	0	1	0
	NUO Complex	K00343	nuoN; NADH-quinone oxidoreductase subunit N [EC:1.6.5.3]	1	1	1	1	1	1	1	1	0
	F-Type ATPase	K02108	ATPF0A, atpB; F-type H+-Transporting ATPase subunit a	1	1	1	1	1	0	0	0	0
	F-Type ATPase	K02109	ATPF0B, atpF; F-type H+-Transporting ATPase subunit b	1	1	1	1	1	1	1	1	1
	F-Type ATPase	K02110	ATPF0C, atpE; F-type H+-Transporting ATPase subunit c	1	1	1	1	1	0	0	0	0
	F-Type ATPase	K02111	ATPF1A, atpA; F-type H+-Transporting ATPase subunit alpha [EC:3.6.3.14]	1	1	1	1	1	1	1	1	1
	F-Type ATPase	K02112	ATPF1B, atpD; F-type H+-Transporting ATPase subunit beta [EC:3.6.3.14]	1	1	1	1	1	1	1	1	1
	F-Type ATPase	K02113	ATPF1D, atpH; F-type H+-Transporting ATPase subunit delta	1	1	1	1	1	1	1	1	1
	F-Type ATPase	K02114	ATPF1E, atpC; F-type H+-Transporting ATPase subunit epsilon	1	1	1	1	1	1	1	1	1

Table A5 continued.

Category	Pathway	KEGG no.	Gene	JdFR nit3A	JdFR nit7B	JdFR nit1A	JdFR nit4A	JdFR nit6B	JdFR nit8A	JdFR nit9A	JdFR nit2A	JdFR nit5A
Oxidative Phosphorylation	F-Type ATPase	K02115	ATPF1G, atpG; F-type H+-Transporting ATPase subunit gamma	1	1	1	1	1	1	1	1	1
	Ubiquinol-cytochrome c reductase	K00412	CYTB, petB; ubiquinol-cytochrome c reductase cytochrome b subunit	0	0	0	1	0	0	0	1	0
	Cytochrome c oxidase, cbb3-type	K00404	ccoN; cytochrome c oxidase cbb3-type subunit I [EC:1.9.3.1]	0	0	1	1	0	1	0	1	0
	Cytochrome c oxidase, cbb3-type	K00405	ccoO; cytochrome c oxidase cbb3-type subunit II	0	0	0	0	0	1	0	1	0
	Cytochrome c oxidase, cbb3-type	K00406	ccoP; cytochrome c oxidase cbb3-type subunit III	0	0	0	0	0	0	0	1	0
	Cytochrome bd complex	K00425	cydA; cytochrome d ubiquinol oxidase subunit I [EC:1.10.3.14]	1	1	0	0	0	0	0	0	0
	Cytochrome c oxidase	K02274	coxA; cytochrome c oxidase subunit I [EC:1.9.3.1]	0	0	0	1	0	0	0	0	0
Transporters	Phosphonate Transport	K02044	phnD; phosphonate Transport system substrate-binding protein	1	1	1	1	1	1	1	1	0
	Phosphate Transport	K02036	pstB; phosphate Transport system ATP-binding protein [EC:3.6.3.27]	1	1	1	1	1	0	1	1	1
	Phosphate Transport	K02037	pstC; phosphate Transport system permease protein	1	1	1	1	1	0	1	1	1

Table A5 continued. Pathway abbreviation: PMST: Putative Multi-Sugar Transporter.

Category	Pathway	KEGG no.	Gene	JdFR nit3A	JdFR nit7B	JdFR nit1A	JdFR nit4A	JdFR nit6B	JdFR nit8A	JdFR nit9A	JdFR nit2A	JdFR nit5A
Transporters	Phosphate Transport	K02038	pstA; phosphate Transport system permease protein	1	1	1	1	1	0	1	1	1
	Phosphate Transport	K02040	pstS; phosphate Transport system substrate-binding protein	1	1	1	1	1	0	1	1	1
	Sulfate /Phosphate Transport	K03306	inorganic phosphate Transporter, PiT family (TC.PIT)	0	0	1	1	1	1	0	0	0
	Ammonium Transporter	K03320	ammonium Transporter, Amt family (amt, AMT, MEP)	1	1	1	1	1	0	0	1	0
	Nitrate Transport	K02051	NitT/TauT family Transport system substrate-binding protein (ABC.SN.S)	0	0	0	1	1	0	0	0	0
	Magnesium Transport	K06213	mgtE; magnesium Transporter	1	1	1	1	1	1	0	1	0
	PMST	K02025	multiple sugar Transport system permease protein (ABC.MS.P)	0	0	1	1	1	0	0	1	0
	PMST	K02026	multiple sugar Transport system permease protein (ABC.MS.P1)	0	0	1	1	1	0	0	1	0
	PMST	K02027	multiple sugar Transport system substrate-binding protein (ABC.MS.S)	0	0	1	1	1	0	0	1	0
	PMST	K10112	msmX, msmK, malK, sugC, ggtA, msiK; multiple sugar Transport system ATP-binding protein	1	1	1	0	1	1	0	1	0

Table A5 continued.

Category	Pathway	KEGG no.	Gene	JdFR nit3A	JdFR nit7B	JdFR nit1A	JdFR nit4A	JdFR nit6B	JdFR nit8A	JdFR nit9A	JdFR nit2A	JdFR nit5A
Transporters	Lipooligosaccharide Export	K09694	nodJ; lipooligosaccharide Transport system permease protein	0	0	0	0	0	1	0	1	0
	Lipooligosaccharide Export	K09695	nodI; lipooligosaccharide Transport system ATP-binding protein	0	0	0	0	0	1	0	1	0
	Lipopolysaccharide Export	K11085	msbA; ATP-binding cassette, subfamily B, bacterial MsbA [EC:3.6.3.-]	1	1	1	1	1	0	0	1	1
	Lipopolysaccharide Export	K09774	lipopolysaccharide export system protein LptA (lptA)	0	0	1	0	1	0	0	1	1
	Lipopolysaccharide export	K06861	lptB; lipopolysaccharide export system ATP-binding protein [EC:3.6.3.-]	1	1	1	1	1	1	1	1	1
	Lipopolysaccharide export	K04744	LPS-assembly protein (lptD, imp, ostA)	0	0	1	1	1	0	0	0	0
	Lipopolysaccharide export	K07091	lptF; lipopolysaccharide export system permease protein	1	1	1	1	1	0	1	0	0
	Lipopolysaccharide export	K11720	lptG; lipopolysaccharide export system permease protein	1	1	1	1	1	1	1	0	0
	Lipoprotein export	K13292	lgt; phosphatidylglycerol:prolipoprotein diacylglycerol transferase [EC:2.-.-.-]	1	1	1	1	1	0	0	1	0
	Lipoprotein export	K03101	lspA; signal peptidase II [EC:3.4.23.36]	1	1	1	1	1	1	0	1	1
	Lipoprotein export	K03820	lnt; apolipoprotein N-acyltransferase [EC:2.3.1.-]	1	1	1	0	1	0	0	1	1
	Lipoprotein export	K03634	lolA; outer membrane lipoprotein carrier protein	0	0	1	0	1	0	0	1	0

Table A5 continued.

Category	Pathway	KEGG no.	Gene	JdFR nit3A	JdFR nit7B	JdFR nit1A	JdFR nit4A	JdFR nit6B	JdFR nit8A	JdFR nit9A	JdFR nit2A	JdFR nit5A
Transporter	Lipoprotein export	K09808	lolC_E; lipoprotein-releasing system permease protein	1	1	1	1	1	1	1	1	1
	Lipoprotein export	K09810	lolD; lipoprotein-releasing system ATP-binding protein [EC:3.6.3.-]	1	1	1	1	1	1	1	1	1
Metal Transporter	Cobalt Transporter	K02007	cbiM; cobalt/nickel Transport system permease protein	1	1	1	0	1	1	1	1	0
	Copper Transporter	K17686	copA, ATP7; Cu+-exporting ATPase [EC:3.6.3.54]	1	1	1	1	1	0	0	1	1
	Copper Transporter	K01533	Cu2+-exporting ATPase [EC:3.6.3.4] (copB)	0	0	0	0	1	0	0	1	0
	Ferrous iron Transport	K04758	ferrous iron Transport protein A (feoA)	1	1	0	0	0	0	0	0	0
	Ferrous iron Transport	K04759	ferrous iron Transport protein B (feoB)	1	1	1	0	1	0	0	1	0
	Ferric iron Transport	K02013	ABC.FEV.A; iron complex Transport system ATP-binding protein [EC:3.6.3.34]	1	1	1	1	1	1	1	1	0
	Ferric iron Transport	K02014	iron complex outermembrane recepter protein (TC.FEV.OM)	0	0	0	1	0	0	0	0	0
	Ferric iron Transport	K02015	ABC.FEV.P; iron complex Transport system permease protein	1	1	1	1	1	1	1	1	0
	Ferric iron Transport	K02016	ABC.FEV.S; iron complex Transport system substrate-binding protein	1	1	1	1	1	1	1	1	0

Table A5 continued. Pathway abbreviation: TB: Thiamine biosynthesis.

Category	Pathway	KEGG no.	Gene	JdFR nit3A	JdFR nit7B	JdFR nit1A	JdFR nit4A	JdFR nit6B	JdFR nit8A	JdFR nit9A	JdFR nit2A	JdFR nit5A
Metal Transporter	Tungstate Transport	K05772	tupA, vupA; tungstate Transport system substrate-binding protein	1	1	1	1	1	0	0	0	0
	Tungstate Transport	K05773	tupB, vupB; tungstate Transport system permease protein	1	1	1	1	1	0	0	0	0
	Tungstate Transport	K06857	tupC, vupC; tungstate Transport system ATP-binding protein [EC:3.6.3.55]	1	1	1	1	1	0	0	0	0
	Zinc Transport	K09815	znuA; zinc Transport system substrate-binding protein	1	1	1	0	0	1	0	0	0
	Zinc Transport	K09816	znuB; zinc Transport system permease protein	1	1	1	0	0	1	0	0	0
	Zinc Transport	K09817	znuC; zinc Transport system ATP-binding protein [EC:3.6.3.-]	1	1	1	0	0	1	0	0	0
	Zinc and cadmium Transport	K16267	zipB; zinc and cadmium Transporter	0	0	0	0	0	0	0	1	1
	Zinc and cadmium Transport	K01534	Cd2+/Zn2+-exporting ATPase [EC:3.6.3.3 3.6.3.5] (zntA)	0	0	0	1	1	0	0	1	0
Vitamin Biosynthesis	TB	K03154	thiS; sulfur carrier protein	1	1	1	1	1	0	0	1	0
	TB	K04487	iscS, NFS1; cysteine desulfurase [EC:2.8.1.7]	1	1	1	1	1	1	0	1	0
	TB	K03150	thiH; 2-iminoacetate synthase [EC:4.1.99.19]	0	0	0	0	0	0	0	1	0

Table A5 continued. Pathway abbreviation: TB: Thiamine Biosynthesis; RB: Riboflavin Biosynthesis.

Category	Pathway	KEGG no.	Gene	JdFR nit3A	JdFR nit7B	JdFR nit1A	JdFR nit4A	JdFR nit6B	JdFR nit8A	JdFR nit9A	JdFR nit2A	JdFR nit5A
Vitamin Biosynthesis	TB	K01662	dxs; 1-deoxy-D-xylulose-5-phosphate synthase [EC:2.2.1.7]	1	1	1	0	0	1	1	1	1
	TB	K03149	thiG; thiazole synthase [EC:2.8.1.10]	1	1	1	0	1	0	0	1	0
	TB	K03147	thiC; phosphomethylpyrimidine synthase [EC:4.1.99.17]	1	1	1	1	1	0	0	1	0
	TB	K00941	thiD; hydroxymethylpyrimidine/ phosphomethylpyrimidine kinase [EC:2.7.1.49 2.7.4.7]	1	1	1	1	1	1	0	1	0
	TB	K00788	thiE; thiamine-phosphate pyrophosphorylase [EC:2.5.1.3]	1	1	1	1	1	1	0	1	1
	TB	K00946	thiL; thiamine-monophosphate kinase [EC:2.7.4.16]	1	1	1	1	1	1	1	1	0
	RB	K14652	ribBA; 3,4-dihydroxy 2-butanone 4-phosphate synthase / GTP cyclohydrolase II [EC:4.1.99.12 3.5.4.25]	1	1	1	1	1	1	1	1	1
	RB	K11752	ribD; diaminohydroxyphosphoribosylaminopyrimidine deaminase / 5-amino-6-(5-phosphoribosylamino)uracil reductase [EC:3.5.4.26 1.1.1.193]	1	1	1	1	1	0	0	1	0

Table A5 continued. Pathway abbreviation: RB: Riboflavin Biosynthesis; CB: Cobalamin Biosynthesis.

Category	Pathway	KEGG no.	Gene	JdFR nit3A	JdFR nit7B	JdFR nit1A	JdFR nit4A	JdFR nit6B	JdFR nit8A	JdFR nit9A	JdFR nit2A	JdFR nit5A
Vitamin Biosynthesis	RB	K00794	ribH, RIB4; 6,7-dimethyl-8-ribityllumazine synthase [EC:2.5.1.78]	1	1	1	1	1	1	1	1	0
	RB	K00793	ribE, RIB5; riboflavin synthase [EC:2.5.1.9]	1	1	1	1	1	1	1	1	1
	RB	K11753	ribF; riboflavin kinase / FMN adenylyltransferase [EC:2.7.1.26 2.7.7.2]	1	1	1	1	1	1	0	0	1
	CB	K00798	MMAB, pduO; cob(I)alamin adenosyltransferase [EC:2.5.1.17]	0	0	0	1	0	0	0	0	0
	CB	K02232	cobQ, cbiP; adenosylcobyric acid synthase [EC:6.3.5.10]	1	1	1	1	1	0	0	1	0
	CB	K02227	cbiB, cobD; adenosylcobinamide-phosphate synthase [EC:6.3.1.10]	1	1	1	1	1	0	0	0	0
	CB	K02231	cobP, cobU; adenosylcobinamide kinase / adenosylcobinamide-phosphate guanylyltransferase [EC:2.7.1.156 2.7.7.62]	1	1	1	1	1	1	0	1	0
	CB	K02233	E2.7.8.26, cobS, cobV; adenosylcobinamide-GDP ribazoletransferase [EC:2.7.8.26]	1	1	1	1	1	1	0	1	0
	CB	K02226	cobC, phpB; alpha-ribazole phosphatase [EC:3.1.3.73]	1	1	1	0	1	0	0	1	0

Table A5 continued. Pathway abbreviation: CB: Cobalamin Biosynthesis.

Category	Pathway	KEGG no.	Gene	JdFR nit3A	JdFR nit7B	JdFR nit1A	JdFR nit4A	JdFR nit6B	JdFR nit8A	JdFR nit9A	JdFR nit2A	JdFR nit5A
Vitamin Biosynthesis	CB	K00768	E2.4.2.21, cobU, cobT; nicotinate-nucleotide--dimethylbenzimidazole phosphoribosyltransferase [EC:2.4.2.21]	1	1	1	1	1	1	0	1	0
Amino Acid Biosynthesis	Histidine biosynthesis	K00013	hisD; histidinol dehydrogenase [EC:1.1.1.23]	1	1	1	1	1	1	0	1	1
	Arginine biosynthesis	K01755	argH, ASL; argininosuccinate lyase [EC:4.3.2.1]	1	1	1	1	1	1	0	1	1
	Aspargine biosynthesis	K01953	asnB, ASNS; asparagine synthase (glutamine-hydrolysing) [EC:6.3.5.4]	1	1	1	1	1	0	0	0	0
	Lysine biosynthesis	K01586	lysA; diaminopimelate decarboxylase [EC:4.1.1.20]	1	1	1	1	1	0	0	1	0
	Serine biosynthesis	K00600	glyA, SHMT; glycine hydroxymethyltransferase [EC:2.1.2.1]	1	1	1	1	1	0	0	1	0
	Threonine biosynthesis	K01733	thrC; threonine synthase [EC:4.2.3.1]	1	1	1	1	1	1	1	1	1
	Glycine biosynthesis	K00600	glyA, SHMT; glycine hydroxymethyltransferase [EC:2.1.2.1]	1	1	1	1	1	0	0	1	0
	Glutamine biosynthesis	K01915	glnA, GLUL; glutamine synthetase [EC:6.3.1.2]	1	1	1	0	1	1	1	1	0
	Cysteine biosynthesis	K01738	cysK; cysteine synthase A [EC:2.5.1.47]	1	1	1	1	1	1	1	1	0
	Proline biosynthesis	K00286	proC; pyrroline-5-carboxylate reductase [EC:1.5.1.2]	1	1	1	1	1	1	1	1	1

Table A5 continued.

Category	Pathway	KEGG no.	Gene	JdFR nit3A	JdFR nit7B	JdFR nit1A	JdFR nit4A	JdFR nit6B	JdFR nit8A	JdFR nit9A	JdFR nit2A	JdFR nit5A
Amino Acid Biosynthesis	Alanine biosynthesis	K00259	ald; alanine dehydrogenase [EC:1.4.1.1]	1	1	0	0	0	0	0	0	0
	Valine and Isoleucine biosynthesis	K00826	E2.6.1.42, ilvE; branched-chain amino acid aminotransferase [EC:2.6.1.42]	1	1	1	1	1	1	0	0	0
	Valine and Isoleucine biosynthesis	K01687	ilvD; dihydroxy-acid dehydratase [EC:4.2.1.9]	1	1	1	1	1	1	0	1	0
	Valine and Isoleucine biosynthesis	K00053	ilvC; ketol-acid reductoisomerase [EC:1.1.1.86]	1	1	1	0	1	1	1	1	0
	Valine and Isoleucine biosynthesis	K01652	E2.2.1.6L, ilvB, ilvG, ilvI; acetolactate synthase I/II/III large subunit [EC:2.2.1.6]	1	1	1	0	1	1	1	1	1
	Valine and Isoleucine biosynthesis	K01653	E2.2.1.6S, ilvH, ilvN; acetolactate synthase I/III small subunit [EC:2.2.1.6]	1	1	1	0	1	1	1	1	0
	Leucine biosynthesis	K00826	E2.6.1.42, ilvE; branched-chain amino acid aminotransferase [EC:2.6.1.42]	1	1	1	1	1	1	1	1	0
	Leucine biosynthesis	K00052	leuB, IMDH; 3-isopropylmalate dehydrogenase [EC:1.1.1.85]	1	1	1	1	1	1	1	1	0
	Leucine biosynthesis	K01703	leuC, IPMI-L; 3-isopropylmalate/(R)-2-methylmalate dehydratase large subunit [EC:4.2.1.33 4.2.1.35]	1	1	1	1	1	1	1	1	1

Table A5 continued. Pathway abbreviation: LCFA: Long-Chain Fatty Acid.

Category	Pathway	KEGG no.	Gene	JdFR nit3A	JdFR nit7B	JdFR nit1A	JdFR nit4A	JdFR nit6B	JdFR nit8A	JdFR nit9A	JdFR nit2A	JdFR nit5A
Amino Acid Biosynthesis	Leucine biosynthesis	K01649	leuA, IMS; 2-isopropylmalate synthase [EC:2.3.3.13]	1	1	1	1	1	1	1	1	1
	Methionine biosynthesis	K00548	5-methyltetrahydrofolate--homocysteine methyltransferase [EC:2.1.1.13] (metH, MTR)	1	1	1	0	0	0	0	0	0
	Phenylalanine biosynthesis	K00832 K00838 K04518 K05359 K01713	Aromatic amino acid aminotransferase	0	0	0	0	0	0	0	0	0
	Tyrosine biosynthesis	K00220	tyrC; cyclohexadieny/prephenate dehydrogenase [EC:1.3.1.43 1.3.1.12]	1	1	1	0	0	0	0	0	0
	Tryptophan biosynthesis	K01695	trpA; tryptophan synthase alpha chain [EC:4.2.1.20]	1	1	1	1	1	0	1	0	0
	Tryptophan biosynthesis	K01696	trpB; tryptophan synthase beta chain [EC:4.2.1.20]	1	1	1	1	0	0	1	0	0
	Aspartate and Glutamate	K00812	aspB; aspartate aminotransferase [EC:2.6.1.1]	1	1	1	1	1	1	1	1	0
Fatty Acid Metabolism	LCFA import	K01897	ACSL, fadD; long-chain acyl-CoA synthetase [EC:6.2.1.3]	1	1	1	1	0	0	0	1	1
	LCFA import	K06076	long-chain fatty acid Transport protein (fadL)	0	0	1	1	0	0	0	1	0
	Phospholipid synthesis via LCFA	K00655	plsC; 1-acyl-sn-glycerol-3-phosphate acyltransferase [EC:2.3.1.51]	1	1	1	1	1	1	0	0	0

Table A5 continued. Pathway abbreviation: LCFA: Long-Chain Fatty Acid.

Category	Pathway	KEGG no.	Gene	JdFR nit3A	JdFR nit7B	JdFR nit1A	JdFR nit4A	JdFR nit6B	JdFR nit8A	JdFR nit9A	JdFR nit2A	JdFR nit5A
Fatty Acid Metabolism	Phospholipid synthesis via LCFA	K03621	plsX; glycerol-3-phosphate acyltransferase PlsX [EC:2.3.1.15]	1	1	1	0	1	0	0	1	0
	Phospholipid synthesis via LCFA	K08591	plsY; glycerol-3-phosphate acyltransferase PlsY [EC:2.3.1.15]	1	1	1	1	1	0	0	1	0
Biofilm Synthesis	Biofilm PGA Synthesis protein	K11931	pgaB; poly-beta-1,6-N-acetyl-D-glucosamine N-deacetylase [EC:3.5.1.-]	0	0	0	0	1	0	0	0	0
	Biofilm PGA Synthesis protein	K11935	pgaA; biofilm PGA synthesis protein PgaA	0	0	0	0	1	0	0	0	0
	Biofilm PGA Synthesis protein	K11936	pgaC, icaA; poly-beta-1,6-N-acetyl-D-glucosamine synthase [EC:2.4.1.-]	0	0	0	0	1	0	0	0	0
Motility	Flagella	K02387	flgB; flagellar basal-body rod protein FlgB	0	0	1	1	1	1	0	0	0
	Flagella	K02389	flgD; flagellar basal-body rod modification protein FlgD	0	0	1	1	1	1	0	0	0
	Flagella	K02390	flgE; flagellar hook protein FlgE	0	0	1	1	1	1	0	0	0
	Flagella	K02392	flgG; flagellar basal-body rod protein FlgG	0	0	1	1	1	0	0	0	0
	Flagella	K02393	flgH; flagellar L-ring protein precursor FlgH	0	0	1	1	1	0	0	0	0
	Flagella	K02394	flgI; flagellar P-ring protein precursor FlgI	0	0	1	1	1	0	0	0	0
	Flagella	K02396	flgK; flagellar hook-associated protein 1 FlgK	0	0	1	1	1	0	0	0	0

Table A5 continued.

Category	Pathway	KEGG no.	Gene	JdFR nit3A	JdFR nit7B	JdFR nit1A	JdFR nit4A	JdFR nit6B	JdFR nit8A	JdFR nit9A	JdFR nit2A	JdFR nit5A
Motility	Flagella	K02397	flgL; flagellar hook-associated protein 3 FlgL	0	0	1	1	1	0	0	0	0
	Flagella	K02406	fliC; flagellin	0	0	1	1	1	0	0	0	0
	Flagella	K02407	fliD; flagellar hook-associated protein 2	0	0	1	1	1	0	0	0	0
	Flagella	K02409	fliF; flagellar M-ring protein FliF	0	0	1	1	1	1	0	0	0
	Flagella	K02410	fliG; flagellar motor switch protein FliG	0	0	1	1	1	1	0	0	0
	Flagella	K02411	fliH; flagellar assembly protein FliH	0	0	0	0	1	1	0	0	0
	Flagella	K02412	fliI; flagellum-specific ATP synthase [EC:3.6.3.14]	0	0	1	1	1	1	0	0	0
	Flagella	K02557	motB; chemotaxis protein MotB	0	0	1	1	1	1	1	0	0
	Flagella	K02416	fliM; flagellar motor switch protein FliM	0	0	1	1	1	0	0	0	0
	Flagella	K02417	fliNY, fliN; flagellar motor switch protein FliN/FliY	0	0	1	1	1	0	0	0	0
	Flagella	K02556	motA; chemotaxis protein MotA	0	0	1	1	1	1	1	0	0
	Chemotaxis	K00575	cheR; chemotaxis protein methyltransferase CheR [EC:2.1.1.80]	1	1	1	1	1	0	0	1	1
	Chemotaxis	K03406	mcp; methyl-accepting chemotaxis protein	1	1	1	1	1	1	0	1	1
	Chemotaxis	K03407	cheA; two-component system, chemotaxis family, sensor kinase CheA [EC:2.7.13.3]	1	1	1	1	1	0	0	1	1

Table A5 continued.

Category	Pathway	KEGG no.	Gene	JdFR nit3A	JdFR nit7B	JdFR nit1A	JdFR nit4A	JdFR nit6B	JdFR nit8A	JdFR nit9A	JdFR nit2A	JdFR nit5A
Motility	Chemotaxis	K03408	cheW; purine-binding chemotaxis protein CheW	1	1	1	1	1	0	0	1	1
	Chemotaxis	K03412	cheB; two-component system, chemotaxis family, response regulator CheB [EC:3.1.1.61]	1	1	1	1	1	0	0	1	1
	Chemotaxis	K03413	cheY; two-component system, chemotaxis family, response regulator CheY	1	1	1	1	1	1	0	1	1

Table A6. Detection of miscellaneous COGS within the Nitrospirota MAGs. Pathway abbreviation: DNR: Dissimilatory Nitrate Reduction.

Category	Pathway	COG no.	Gene	JdFR nit3A	JdFR nit7B	JdFR nit1A	JdFR nit4A	JdFR nit6B	JdFR nit8A	JdFR nit9A	JdFR nit2A	JdFR nit5A
Nitrogen Metabolism-DNR	DNR	COG3005	Tetraheme cytochrome c subunit of nitrate reductase (nap C)	1	1	1	1	1	1	1	1	0
	DNR	COG3043	Nitrate reductase cytochrome c-type subunit (napB)	1	1	1	0	0	1	0	0	0
	DNR	COG3062	Cytoplasmic chaperone NapD for the signal peptide of periplasmic nitrate reductase NapAB	1	1	0	1	0	0	1	1	0
Carbon Metabolism-Glycogen Metabolism	Glycogen Degradation	COG3408	Glycogen debranching enzyme (alpha-1,6-glucosidase)	1	1	1	0	1	0	0	1	0
Hydrogenase	NiFe Hydrogenase	COG0437	Fe-S-cluster containing dehydrogenase component HybA	1	1	1	1	1	1	1	1	0
	NiFe Hydrogenase	COG5557	Ni/Fe-hydrogenase 2 integral membrane subunit HybB	1	1	1	1	1	1	1	1	0
Transporters	Nitrate transporter	COG0715	ABC-type nitrate/sulfonate/bicarbonate transport system, periplasmic component	0	0	0	1	1	0	0	0	0
	Sulfate transporter	COG0306	Phosphate/sulfate permease	0	0	1	1	1	1	1	0	0
	Sulfur transporter[1]	COG0425	TusA-related sulfurtransferase	1	1	1	1	1	1	1	1	1
	Sulfur transporter[1]	COG2391	Uncharacterized membrane protein YedE/YeeE	1	1	1	1	1	0	1	1	0

Table A6. continued.

Category	Pathway	COG no.	Gene	JdFR nit3A	JdFR nit7B	JdFR nit1A	JdFR nit4A	JdFR nit6B	JdFR nit8A	JdFR nit9A	JdFR nit2A	JdFR nit5A
Transporters	Ferric iron transporter	COG1629	Outer membrane receptor proteins, mostly Fe transport	0	0	0	1	1	0	0	0	0
Sulfur Metabolism	Sulfur oxidation	COG1416	Intracellular sulfur oxidation protein, DsrE/DsrF family	0	0	0	1	1	0	0	0	0

[1] Based on Umezawa et al. 2020

Table A7. ANI and number of non-identical bases (nucleotide variation) of bacterial family JdFR-88 genomes estimated using the ANIm method in pyANI v0.2.7.

	ANI				Nucleotide Variation			
Genomes	JdFRnit3A	JdFRnit7B	GCA_002376155	GCA_002376445	JdFRnit3A	JdFRnit7B	GCA_002376155	GCA_002376445
JdFRnit3A	1	0.99997	0.99995	0.99989	0	64	82	178
JdFRnit7B	0.99997	1	0.99996	0.99987	64	0	64	212
GCA_002376155	0.99995	0.99996	1	0.99992	82	64	0	128
GCA_002376445	0.99989	0.99987	0.99992	1	178	212	128	0

Table A8. Summary of the top 2 BlastX v.2.12.0+ results of sequences with non-identical bases between JdFRnit3A and JdFRnit7B.

Sequence 1a

>JdFRnit3A_Scaffold1020|start:2|stop:1379|direction:f|rev_compd:False|length:1377

AGCAGCACCGACAGCACGGGCACCACCTATACCCTGGCCATACCGCTTCCTGCTGCGCCTGCGGCAGGAGATCACTACGAGGTGCTCAGGCATGA
CGGGGCCACCGGCTCGGGCGGAAAGGGTGTCTGGGGCGTGGAGATAACCGGCTACAGCTCTGCCGGGCAGGACCCCAGCACCCTTACCTACACCA
AGAAGTACGATGCCGGCGGCGATGCCGGAGGCAAGCAGTACGAGCAGTGCCTGAGGTGCCACAGCTACTACAGCTACCTCAACGCCATACCCAG
CACGCCCAGCGGTGGGCCCGACGGCTCGGCGGCGGCGGAGACCGACATAGCCAAGGAGATAAACCCCCAGAATTACGCCCACCACGCCATCTAT
GCGGTGGGCAGGAACCAGCCCATCGTGCCTGACGGCTCCACCGCGCCGGTGGTGAGCCAGTACTGGCCTGTGCAGTCCAACACCATCTCCAGCTA
CGACTCCGGCACGGGTGTGGCCACACTGGGGAGCAGCCTGCCGGATACGGCCCTTCCTGGCTGGTATGTGATAGACACGGGGGCCAATGTGGCCT
ACCAGATAGTGGAGATACTGAGCACAACCCAGATAAGGATAGCTGCCACCAACGGCAACCCCTGGGATGCCAATGCCACCTACCCGGCCAGCAT
AGGCAGCAATGTGGACATCACCGCCGGGCTGGGAAACACCTTCGTGCCGCCCTACGGGCCCTGGGCCGTAATAAGGTGCTCCGACTGCCACCGCT
CAGCGGAGACGGACCCCCTGGGGCCCCATGCCTCGGTAAACAGGTGGCTTCTCAAGAGCCTGGACACGGCCCTGAAGTTTGAATTCTGGGACGGC
TCCTCCGTAGTGGAGCCTGCGCCCAACTCTGGGGCGGACACGGCCATCTGGTGCTTCAACTGCCACCGAAGGGATGTCTACGGCGATGGCAACAA
CTACGATGTGGACTCGGAGGTGCCTTACAGGATTTACTCCAGGGTGCCGCATTCAGGCACCACCAAGCACCCCGACGGCACGGGCATGGAGGGAA
TGTTCGTGGATGATCCCAACATAGTAGACCCTAAGTATGCCACCGTGTGGCCTGAGATATGCCGGCACTGCCACCTGGGCTATAGATTAGGAGGCT
CTCACGGCACCTGGCTGGTAGACCCTGACCCGGGCGACGGCTTCAGCAGCGAGCTGGGGGAGCAGGGCAGGAGGTTCCTCAACGGCGCCACCTG
GGATGCCTACGATGCCCCCACGACGGGCAGCACCATTACCTGCTACACCATTGGCTCTGCCACCAGCGTTTCTTCCTGTACCAAACACGACGGAGG
AGAAGGCGGCAAAACCAAAGCCACCTACGACTACACGGGGATATACTGA

Sequence 1b

>JdFRnit7B_Scaffold0001|start:26599|stop:27976|direction:r|rev_compd:True|length:1377AGCAGCACCGACAGCACGGGCACCACCTATACCCTGGCCATA
CCGCTTCCTGCTGCGCCTGCGGCAGGAGATCACTACGAGGTGCTCAGGCATGACGGGGCCACCGGCTCGGGCGGAAAGGGTGTCTGGGGCGTGGA
GATAACCGGCTACAGCTCTGCCGGGCAGGACCCCAGCACCCTTACCTACACCAAGAAGTACGATGCCGGCGGCGATGCCGGAGGCAAGCAGTAC
GAGCAGTGCCTGAGGTGCCACAGCTACTACAGCTACCTCAACGCCATACCCAGCACGCCCAGCGGTGGGCCCGACGGCTCGGCGGCGGCGGAGA
CCGACATAGCCAAGGAGATAAACCCCCAGAATTACGCCCACCACGCCATCTATGCGGTGGGCAGGAACCAGCCCATCGTGCCTGACGGCTCCACC
GCGCCGGTGGTGAGCCAGTACTGGCCTGTGCAGTCCAACACCATCTCCAGCTACGACTCCGGCACGGGTGTGGCCACACTGGGGAGCAGCCTGCC
GGATACGGCCCTTCCTGGCTGGTATGTGATAGACACGGGGGCCAATGTGGCCTACCAGATAGTGGAGATACTGAGCACAACCCAGATAAGGATAG
CTGCCACCAACGGCAACCCCTGGGATGCCAATGCCACCTACCCGGCCAGCATAGGCAGCAATGTGGACATCACCGCCGGGCTGGGCAACACCTTC
GTGCCTCCTTACGGGCCCTGGAGCATCATAAGGTGCACCGACTGCCACGGCTCCACCTTGCAGGACCCCCTGGGGCCCCATGCCTCGGTAAACAA
GTGGCTTCTCAAGAGCCTGGACACGGCCCTGAAGTTTGAATTCTGGGACGGCTCCTCCGTAGTGGAGCCTGTGCCCAACTCTGGGGCGGACACGG
CCATCTGGTGCTTCAACTGCCACCGAAGGGATGTCTACGGCGATGGCAACAACTACGATGTGGACTCGGAGGTGCCTTACAGGATTTACTCCAGG
GTGCCGCATTCAGGCACCACCAAGCACCCCGACGGCACGGGCATGAAGGGAATGTTCGTGGATGATCCCAACATAGTAGACCCTAAGTATGCCAC
CGTGTGGCCTGAGATATGCCGGCACTGCCACCTGGGCTATAGATTAGGAGGCTCTCACGGCACCTGGCTGGTAGACCCTGACCCGGGCGACGGCT
TCAGCAGCGAGCTGGGGGAGCAGGGCAGGAGGTTCCTCAACGGCGCCACCTGGGATGCCTACGATGCCCCCACGACGGGCAGCACCATTACCTGC
TACACCATTGGCTCTGCCACCAGCGTTTCTTCCTGTACCAAACACGACGGAGGAGAAGGCGGCAAAACCAAAGCCACCTACGACTACACGGGGAT
ATACTGA

Table A8. continued.

Sequence 2a

>JdFRnit3A_Scaffold0016|start:420767|stop:422621|direction:f|rev_compd:False|length:1854

ATGAAGATGAAGGAAGAGGTGTTTCATAGAGCGGGGAGTCTGGGAAGGGTAGTTTTTCTACCTCTCCTGCTCCTGGCGGTGCTTCTGCTTGCCGC
ATGCGGCGATGGCGACGGAGTTGCCGTTCAGGAGGAGGCTCCAGAGGCCGCACCCACCCTCCAGTGGGAGCCCCAGGCCGTCAAGACCTTCCAC
TTCAGCTGGACAGATGTAGATGACGCCACCTACTACAACCTCTTGGAGGACCCTGACGGGGCGTCAGGCTTTACTGCGGTGATTACCAATATCCC
CCAGGGCACTCAGCAGGCAGACCACCAGGTGCCCCTCTACTTTCGCCTCAACGCCCTCTACATCCTGCAGGCCTGCAATGACGCCGGCTGCGTGA
ACTCCCCTGAGCTTGCCGTAAGCGGCAACCTGGCCAATGCCGTGGGCTACTTCAAGGCAAGCAACACCGATGGCGATGATTTCGGCTATGCCCTT
TCCATCTCTGGTGACGGAAGCACCCTGGCCGTGGGGGCATACGGGGAGGACAGCAGTGCCACCGGCATAAACGGCGACCAGAGCAATGACGAT
GCTTCAAACGCCGGTGCGGTCTATGTGTTCATCCGTGATGAGGACGGTGCCTGGGCGCAGCAGGCATACATCAAGGCAAGCAATACCGATGCTT
ACGACTACTTCGGCTATGCCCTTTCCATCTCTGGTGACGGAAGCACCCTGGCCGTGGGGGCATACGGGGAGGACAGCAATGCCACCGGCGTGGA
TGGCGACCAGGGCAATAACGATACACCGTATTCCGGTGCAGTCTATGTGTTCGTCCGTGATGAGGACGGTGCCTGGGTGCAGGAGGCATACATC
AAGGCAAGCAACACCGGTGGAGGCGACATGTTCGGCTATGCCCTTTCCATCTCTGGTGACGGAAGCACCCTGGCCGTGGGGGCATATCTGGAGG
ACAGCAGTGCCACCGGCATAGACGGCGACCAGAGCAATGACGATGCCACGGATTCCGGTGCAGTCTATGTGTTCGTCCGTGATGAGGACGGTGC
CTGGGCGCAGGAGGCATACATCAAGGCAAGCAACACCGGTGGAGAAGACTTATTTGGCTGGGCCCTTTCCGTCTCTATGGATGGCAGCATCATA
GCCGTGGGGGCATTGTGGGAGGACAGCAGTGCCACCGGCATAGACGGCGACCAGAGCAATGACGATGCTGAAAACTCCGGTGCAGTCTATGTGT
TCGTCCGTGATGGCGGTGCCTGGGCGCAGCAGGCATACGTCAAGGCAAGCAACACCGATGGAGGCGATGAGTTCGGCTATGCCCTTTCCATCTCT
GGTGACGGAAGCACCCTGGCCGTGGGGGCATATCTGGAGGACAGCAGTGCCACCGGCATAGACGGCGACCAGAGCAATGACGATGCTGAAAAC
TCCGGTGCAGTCTATGTGTTCGTCCGTGATGGCGGTGCCTGGGCGCAGCAGGCATACGTCAAGGCAAGCAACACCGATGGAGGCGATGAGTTCG
GCTATGCCCTTTCCATCTCTGGTGACGGAAGCACCCTGGCCGTGGGGGCATACGGGGAGAACAGCAATGCCACCGGCGTGGATGGCGACCAGGG
CAATAACGATACACCGGATTCCGGTGCAGTCTATGTGTTCGTCCGTGATGAGGGCGGTGCCTGGGCACAGCAGGCATACGTCAAGGCAAGCAAT
ACCGAGACTGACGACTACTTCGGCTATGCCCTTTCCATCTCTGGTGACGGAAGCACCCTGGCCGTGGGGGCATACGGGGAGGACAGCAGCGCCA
CCGGCATAGACGGCGACCAGAGCAATAACGATGCTTCGGGCTCAGGGGCAGCCTATCTTTACTGA

Sequence 2b

>JdFRnit7B_Scaffold0004|start:58337|stop:59267|direction:r|rev_compd:True|length:930

GTGGGGGCATATCTGGAGGACAGCAGTGCCACCGGCATAGACGGCGACCAGAGCAATGACGATGCCACGGATTCCGGTGCAGTCTATGTGTTCG
TCCGTGATGAGGACGGTGCCTGGGCGCAGGAGGCATACATCAAGGCAAGCAACACCGGTGGAGAAGACTTATTTGGCTGGGCCCTTTCCGTCTC
TATGGATGGCAGCATCATAGCCGTGGGGGCATTGTGGGAGGACAGCAGTGCCACCGGCATAGACGGCGACCAGAGCAATGACGATGCTGAAAA
CTCCGGTGCAGTCTATGTGTTCGTCCGTGATGGCGGCGCCTGGGCACAGGAGGCGTACATCAAGGCAAGCAACACCGATGCTCACGACTACTTCG
GCCATGCCCTTTCCATCTCTGGTGACGGAAGCACCCTGGCCGTGGGGGCACATAAGGAGGACAGCAATGCCACCGGCATAGACGGCGACCAGAA
CAATGACGATGCTTCACACGCCGGTGCGGTCTATGTGTTTGCCCGTGATGGCGGTGCCTGGGCGCAGCAGGCATACGTCAAGGCAAGCAACACC
GATGGAGGCGATGAGTTCGGCTATGCCCTTTCCATCTCTGGTGACGGAAGCACCCTGGCCGTGGGGGCATACGGGGAGAACAGCAATGCCACCG
GCGTGGATGGCGACCAGGGCAATAACGATACACCGGATTCCGGTGCAGTCTATGTGTTCGTCCGTGATGAGGGCGGTGCCTGGGCACAGCAGGC
ATACGTCAAGGCAAGCAATACCGAGACTGACGACTACTTCGGCTATGCCCTTTCCATCTCTGGTGACGGAAGCACCCTGGCCGTGGGGGCATACG
GGGAGGACAGCAGCGCCACCGGCATAGACGGCGACCAGAGCAATAACGATGCTTCGGGCTCAGGGGCAGCCTATCTTTACTGA

Table A8. continued.

Sequence 3a

>JdFRnit3A_Scaffold0107|start:152540|stop:153899|direction:r|rev_compd:True|length:1359

AGCAGCACCGACAGCACGGGCACCACCTATACCCTGGCCATACCGCTTCCTGCAGCCCCCTCAGGGGGAGAGTCCTACCGGGTGCTCAAGCTG
GATGGCTCGGGCGCCGAAAAGGGGGTATGGGGGGTTCAGCTCAGCGGCTACAGCTCGGCAGGGCAGGAGCCCAGCACCCTTACTTACGCGAAG
AAGTACAGCCGCACAAAACAGTACGAGCAGTGCCTCCGCTGCCACAGCTACTATGGATACTACACTGCCCCGCCCAGCACGCCCAGCGGCGGG
CCCGACGGCTCGGCGGCGGCCCAGACCGACCTGGCCAAGGAGATAAACCCCCAGAATTACGCCCACCACGCCATCTATGCGGTGGGCAGGAAC
CAGCCCATCGTGCCTGACGGCTCCACCGCGCCGGTGGTGAGCCAGTACTGGCCTGTGCAGTCCAACACCATCTCCAGCTACGACTCCGGCACGG
GTGTGGCCACACTGGGGAGCAGCCTGCCGGATACGGCCCTTCCTGGCTGGTATGTGATAGACACGGGGGCCAATGTGGCCTACCAGATAGTGG
AGATACTGAGCACAACCCAGATAAGGATAGCTGCCACCAACGGCAACCCCTGGGATGCCAATGCCACCTACCCGGCCAGCATAGGCAGCAATG
TGGACATCACCGCCGGGCTGGGAAACACCTTCGTGCCGCCCTACGGGCCCTGGGCCGTAATAAGGTGCTCCGACTGCCACCGCTCAGCGGAGA
CGGACCCCCTGGGGCCCCATGCCTCGGTAAACAGGTGGCTTCTCAAGAGCCTGGACACGGCCCTGAAGTTTGAATTCTGGGACGGCTCCTCCGT
AGTGGAGCCTGCGCCCAACTCTGGGGCGGACACGGCCATCTGGTGCTTCAACTGCCACCGAAGGGATGTCTACGGCGATGGCAACAACTACGA
TGTGGACTCGGAGGTGCCTTACAGGATTTACTCCAGGGTGCCGCATTCAGGCACCACCAAGCACCCCGACGGCACGGGCATGGAGGGAATGTT
CGTGGATGATCCCAACATAGTAGACCCTAAGTATGCCACCGTGTGGCCTGAGATATGCCGGCACTGCCACCTGGGCTATAGATTAGGAGGCTCT
CACGGCACCTGGCTGGTAGACCCTGACCCGGGCGACGGCTTCAGCAGCGAGCTGGGGGAGCAGGGCAGGAGGTTCCTCAACGGCGCCACCTGG
GATGCCTACGATGCCCCCACGACGGGCAGCACCATTACCTGCTACACCATTGGCTCTGCCACCAGCGTTTCTTCCTGTACCCACCACGACGACC
GAGGAGCAAGCGGCGACACCAAAGCCACCTACGACTACACCGGGATATACTAA

Sequence3b

>JdFRnit7B_Scaffold0006|start:152540|stop:156392|direction:r|rev_compd:True|length:3852

ATGAAGAGGCCCTTCCTCATAGGCCCGCTGGCAGTGCTCTTGTCTCTCCTTGCGGCCCGTGCCGCCTCGCCGCCTCCCATGGTGAGCATAATCAA
CCCGGCCCCAGGCACCCAGGCGGCCCCCTCGGTGGTCTCCAACAGCCCCACCGCGGGGGGATTCAAGGTGCAGGTGCAGGTGCACAACGATGT
CCAGGACATTACCGAGGTCTCCATAGGGGTGTGCCCTGGAGCCGACGACCCCTCCACCTGCTCCTGGAGCTGGAGTCCCCTGACGCTCAATACC
AATTACGACTGCGGAGAGGGCTGTGGCATCTACGAGGCACTGCTGGGCCTCCAGGAAGGGGACTGGTGGCTTCAGGCCAGGGCGGTCTCCAGC
GCCGATGGCACCGGCTACAGCTCCGACCAGCGCACAGGAAACGATGCTAGGTACATCTATCTGAGGGTAAGGGCCGCCAAGCGGGGCACGGG
AAGCCTTCTGGTCAGGGACCGCTCGAGCCGCATGTGCATGGACTGCCACCAGGGCGAGATTTTAACCCATTCAAGCCAGAGCACCGACACCTCC
TACGGCAACTGGCAGACCGTGTGCCTGGAATGCCACACCCCCCACAACACCTTCAACATCTACCTCATAAGGAACGCCATCCAGACCCCTAACA
GCGGCCTCAAGGAGGTAGTGTTCTATAACACCACGGGCGATGCCCCCCAGAGCTATGTAGACTCCTCCGCAGGCTCGGCCACCAAGGGGGTGT
GCCAGGTGTGCCACACCCAGACGGGAGGGGCCGAGGCCAGATGGCGAAACACCGGGGACGAGAGCAACCATTACGGGGGCAGCAGCCCTACT
CAGCGCTGCACCAACTGCCATACCCACGAGCTGGGCTTCGCAGGGGGAGAGTCCACGGGAGGGCAGAGCTGCTCCAGCTGCCACGCCGACCTC
TGGGGCATGATGCACTCCGGCATCGCCAACATGGTATACAAGCATTACCTGGCAAGCGACGCTGCCACCTACCTGGTGAACTCCACCCTGCCCC
ATGTCCTGGGCAGCCAGGACGATACAGACCGCAACTGCCTCATGTGCCACGCCGACCACGACATCTTCAGGCCCGACATAAACCCCGAGGGCC
AGAGGGCGGCCAACCTCCGCAGCACCCTCTCGGAGGCCCCTGACGCCGGGGCGGACACCGGCTTTGCAAACACCGACTTTCTGGCCAGCACAG
GGGGCATATGCCTGAGCTGCCACAGGGCCAGGCAACAGAAGGCCCTCTCCAGTCCCAACGGACTGGGCCAGGTGCAGCCCATACCCTATTCCG
ACGAGGCCAACCCCTCTGAGTCCCCCCCTGATGCAGCCCTGGTGGTCAATGCCTCGGCTCACGGCTACGAGGTGGCCTCCGGCGGCTTCGGCGA
CGGCTCCACCTTCCGGGCCGTCTGCACCAAGTGCCACAACGACACCGTGGGGGGCGAGGCAGGGGCCAAAAGCTCTGTGGCAGCCCAGGCTGC
CGAGCCCAGGTTCGGGCTTCACCAGAGCTCCAGGGCAGGCCTCCTGGCCCCCGATGGAGAGGCCAATGCAGGGGAGTACCTCCAGGGCAGCAT
CACCGCCGTGGCAGGCACCACCGTGACGGTGGACGAGGACCTCACAAAGGACTACACAGGCTACGCCCTGGTAGTGATAAGCGCTGCTACCTA
CGAGAGCCAGCGGGCCCTCATTACCGCCACGGACACCGCCAACGAGACCT

Table A8. continued.

Sequence 3b continued

TCACGGTGGCCTTCTGGCCCCTTTATGAACCCAATGTGGGCGACACCTTTGAGACAACCCCTGAGCATACCGCCTCGCAGGGCCTCTGCTACAGCTG
CCACAGCCAGGGCTCCAACTCAGGTGGCACAGGCCTGGATGCCCCCAACAGCTACAAGCCCTATGACGGCGTAGACATCTACAACACCGTGGCCAT
GAGAAATCCCATGGAGCAGATTCTAAACCTCTTTGGCTACTCCTACAGCCATCCCCTGGATGAGGAAGGTCGGCACCGGCCCGATGAGCCGGCCAC
GGCGCCTTCCTCGGCCCCGGCAGGCCTCTCGGAGGCCTGGAATGTGGGCGACATGGGCACCTGCGATGCCGCTGCCGACCAGACCACCTGCCAGGA
TTCCACCAAGGCCTGGCCCAACACCGACCAGTTCGTGGGCCTGAGCATCACCTTCTACACGGGCGACTGCTCGGGGCAGAGCTTCAGCATCACTGCC
AGCAGTACCGACACGGTCAGCTTTGACACCGGCGGGGCCTGCACCCCTGCGGTGGGCGATAGCTACTACATAGGCACCAGGCATGTGAGCTGCTCC
GACTGCCACAACACCCATGCGGCGCAGGTCAACCCCATGGGCACGGTAAGCGGCGGCACCACTACCACCGTACAGGACGGCACCAAGACCTTTGA
GAAGACCGGCTGGGCCAACGACAGGTGGAAGGGCTACCTCATGGTGTTTACGGACTCAAACGGCGTAAGGAGGGTGCGTCCCATAGCGGGGAGCA
GCACCGACAGCACGGGCACCACCTATACCCTGGCCATACCGCTTCCTGCAGCCCCCTCAGGGGGAGAGTCCTACCGGGTGCTCAAGCTGGATGGCT
CGGGCGCCGAAAAGGGGGTATGGGGGGTTCAGCTCAGCGGCTACAGCTCGGCAGGGCAGGAGCCCAGCACCCTTACTTACGCGAAGAAGTACAGC
CGCACAAAACAGTACGAGCAGTGCCTCCGCTGCCACAGCTACTATGGATACTACACTGCCCCGCCCAGCACGCCCAGCGGCGGGCCCGACGGCTCG
GCGGCGGCCCAGACCGACCTGGCCAAGGAGATAAACCCCCAGAATTACGCCCACCACGCCATCTATGCGGTGGGCAGGAACCAGCCCATCGTGCCT
GACGGCTCCACCGCGCCGGTGGTGAGCCAGTACTGGCCTGTGCAGTCCAACACCATCTCCAGCTACGACTCCGGCACGGGTGTGGCCACACTGGGG
AGCAGCCTGCCGGATACGGCCCTTCCTGGCTGGTATGTGATAGACACGGGGGCCAATGTGGCCTACCAGATAGTGGAGATACTGAGCACAACCCAG
ATAAGGATAGCTGCCACCAACGGCAACCCCTGGGATGCCAATGCCACCTACCCGGCCAGCATAGGCAGCAATGTGGACATCACCGCCGGGCTGGGC
AACACCTTCGTGCCTCCTTACGGGCCCTGGAGCATCATAAGGTGCACCGACTGCCACGGCTCCACCTTGCAGGACCCCCTGGGGCCCCATGCCTCGG
TAAACAAGTGGCTTCTCAAGAGCCTGGACACGGCCCTGAAGTTTGAATTCTGGGACGGCTCCTCCGTAGTGGAGCCTGTGCCCAACTCTGGGGCGGA
CACGGCCATCTGGTGCTTCAACTGCCACCGAAGGGATGTCTACGGCGATGGCAACAACTACGATGTGGACTCGGAGGTGCCTTACAGGATTTACTCC
AGGGTGCCGCATTCAGGCACCACCAAGCACCCCGACGGCACGGGCATGAAGGGAATGTTCGTGGATGATCCCAACATAGTAGACCCTAAGTATGCC
ACCGTGTGGCCTGAGATATGCCGGCACTGCCACCTGGGCTATAGATTAGGAGGCTCTCACGGCACCTGGCTGGTAGACCCTGACCCGGGCGACGGC
TTCAGCAGCGAGCTGGGGGAGCAGGGCAGGAGGTTCCTCAACGGCGCCACCTGGGATGCCTACGATGCCCCCACGACGGGCAGCACCATTACCTGC
TACACCATTGGCTCTGCCACCAGCGTTTCTTCCTGTACCCACCACGACGACCGAGGAGCAAGCGGCGACACCAAAGCCACCTACGACTACACCGGG
ATATACTAA

Table A8. continued.

Scaffold Sequence	Description	Taxonomy	Max Score	Total Score	Query Cover (%)	E value	Percent Identity	Accession Length	Protein Accession	Project Accession	Non-identical bases between corresponding scaffolds
1a	Hypothetical protein	*Desulfuromonadales*	322	322	79	7.00E-95	45.65	1263	MBK5276770.1	PRJNA680161	19
1a	Cytochrome c3 family protein	*G. sediminis*	320	320	79	3.00E-94	43.81	1262	WP_199381962.1	MW386414	
1b	Hypothetical protein	*Desulfuromonadales*	327	327	79	8.00E-97	46.86	1263	MBK5276770.1	PRJNA680161	
1b	Cytochrome c3 family protein	*G. sediminis*	323	323	79	3.00E-95	44.31	1262	WP_199381962.1	MW386414	
2a	Integrin	*Microbulbifer sp.* GL-2	644	644	90	0	62.41	607	WP_172621228.1	PRJDB8498	25
2a	Hypothetical protein	*T.denitrificans*	620	620	90	0	61.99	618	WP_092999382.1	PRJNA263057	
2b	Integrin	*Microbulbifer sp.* GL-2	376	1435	99	2.00E-123	65.70	607	WP_172621228.1	PRJDB8498	
2b	FG-GAP repeat protein	*Pseudenhygromyxa sp.* WMMC2535	366	1044	99	1.00E-119	63.75	579	WP_172308291.1	PRJNA608260	
3a	Hypothetical protein	*Desulfuromonadales*	301	301	84	2.00E-87	44.80	1263	MBK5276770.1	PRJNA680161	19
3a	Cytochrome c3 family protein	*G. sediminis*	299	299	79	1.00E-86	43.78	1262	WP_199381962.1	MW386414	
3b	Hypothetical protein	*Desulfuromonadales*	372	550	86	7.00E-105	39.16	1263	MBK5276770.1	PRJNA680161	
3b	Cytochrome c3 family protein	*G. sediminis*	365	558	84	4.00E-102	37.23	1262	WP_199381962.1	MW386414	