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 HAWAI‘I 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: Nitrospirotta, 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

<table>
<thead>
<tr>
<th>SECTION</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>ii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>v</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>vi</td>
</tr>
<tr>
<td>CHAPTER 1. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>CHAPTER 2. MATERIALS AND METHODS</td>
<td>4</td>
</tr>
<tr>
<td>2.1 Basement fluid sampling</td>
<td>4</td>
</tr>
<tr>
<td>2.2 Metagenome assembly and binning</td>
<td>5</td>
</tr>
<tr>
<td>2.3 Phylogenomics</td>
<td>7</td>
</tr>
<tr>
<td>2.4 Competitive recruitment of the non-redundant JdFR Nitrospirota</td>
<td>8</td>
</tr>
<tr>
<td>2.5 Metabolic reconstruction and functional characteristics</td>
<td>8</td>
</tr>
<tr>
<td>2.6 Single nucleotide variation</td>
<td>9</td>
</tr>
<tr>
<td>CHAPTER 3. RESULTS AND DISCUSSION</td>
<td>10</td>
</tr>
<tr>
<td>3.1 Recovery of JdFR Nitrospirota MAGs</td>
<td>10</td>
</tr>
<tr>
<td>3.2 Phylogenomic analysis of JdFR Nitrospirota MAGs</td>
<td>12</td>
</tr>
<tr>
<td>3.3 Competitive recruitment of the JdFR Nitrospirota in the U1362A and U1362B metagenomes</td>
<td>13</td>
</tr>
<tr>
<td>3.4 Metabolic potential of JdFR Nitrospirota</td>
<td>14</td>
</tr>
<tr>
<td>3.4.1 Carbon metabolism</td>
<td>14</td>
</tr>
<tr>
<td>3.4.2 Sulfur and nitrogen cycling</td>
<td>16</td>
</tr>
<tr>
<td>3.4.3 Hydrogenases and energy complexes</td>
<td>18</td>
</tr>
<tr>
<td>3.4.4 Mixed acid fermentation</td>
<td>19</td>
</tr>
<tr>
<td>3.4.5 Transporters</td>
<td>19</td>
</tr>
<tr>
<td>3.4.6 Vitamin and amino acid metabolism</td>
<td>20</td>
</tr>
<tr>
<td>3.4.7 Motility and biofilm formation</td>
<td>21</td>
</tr>
<tr>
<td>3.5 Genome variation between JdFRnit3A and JdFRnit7B</td>
<td>21</td>
</tr>
<tr>
<td>APPENDICES</td>
<td>29</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>34</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 1. Characteristics of JdFR MAGs related to the Nitrospirota recovered in this study .................................................................................................................................................12

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 ........................................................................................................................................................................26
LIST OF FIGURES

Figure 1. Phylogenomic reconstruction of the Nitrospirota phylum based on GToTree’s bacterial single-copy core gene set..........................................................................................................................14

Figure 2. Metabolic reconstruction of KEGG pathways in the JdFR Nitrospirota lineages........20

Figure 3. Summary of metabolic KEGG functions...........................................................................21
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 \textit{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 archaelal 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 archael 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;

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 Nitrospirotia MAGs with publicly available Nitrospirotia 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). Nitrospirotia 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 21 single-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 Nitrospiota 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).

Table 1. Characteristics of JdFR MAGs related to the Nitrospirota recovered in this study.

<table>
<thead>
<tr>
<th>Genome</th>
<th>Completion (%)</th>
<th>Contamination (%)</th>
<th>Size (Mbp)</th>
<th>N50 (Kbp)</th>
<th>%GC</th>
<th>Scaffolds(#)</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>JdFRnit1A</td>
<td>98.2</td>
<td>0.23</td>
<td>2.17</td>
<td>255</td>
<td>45.1</td>
<td>11</td>
<td>High</td>
</tr>
<tr>
<td>JdFRnit2A</td>
<td>94.6</td>
<td>1.82</td>
<td>1.89</td>
<td>42</td>
<td>48.8</td>
<td>74</td>
<td>High</td>
</tr>
<tr>
<td>JdFRnit3A</td>
<td>100</td>
<td>1.82</td>
<td>1.86</td>
<td>463</td>
<td>62.7</td>
<td>8</td>
<td>High</td>
</tr>
<tr>
<td>JdFRnit4A</td>
<td>85.9</td>
<td>1.09</td>
<td>2.18</td>
<td>61</td>
<td>40.3</td>
<td>49</td>
<td>Medium</td>
</tr>
<tr>
<td>JdFRnit5A</td>
<td>51.8</td>
<td>0.00</td>
<td>0.60</td>
<td>22</td>
<td>42.6</td>
<td>36</td>
<td>Medium</td>
</tr>
<tr>
<td>JdFRnit8A</td>
<td>59.6</td>
<td>0.91</td>
<td>1.14</td>
<td>22</td>
<td>41.5</td>
<td>91</td>
<td>Medium</td>
</tr>
<tr>
<td>JdFRnit9A</td>
<td>43.9</td>
<td>1.75</td>
<td>0.73</td>
<td>14</td>
<td>41.1</td>
<td>78</td>
<td>Low</td>
</tr>
<tr>
<td>JdFRnit6B</td>
<td>98.2</td>
<td>1.82</td>
<td>2.33</td>
<td>121</td>
<td>41.6</td>
<td>28</td>
<td>High</td>
</tr>
<tr>
<td>JdFRnit7B</td>
<td>100</td>
<td>1.82</td>
<td>1.86</td>
<td>463</td>
<td>62.7</td>
<td>7</td>
<td>High</td>
</tr>
</tbody>
</table>

1 MAGs recovered from U1362A (A) and U1362B (B) are indicated
2 Based on the minimum information about metagenome-assembled genome (MIMAG)

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 (Figure 1). Two different Nitrospirotia 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 (Figure 1), 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 (Figure 1). 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 Nitrospirotia MAGs were from borehole U1362A, and formed two related lineages within the order UBA6902 (Figure 1). 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 Nitrospira MAGs are indicated in bold. The four colored boxes further group the JdFR Nitrospira MAGs 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 A5), have the genes for CO₂ 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 Nitrospirotia 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 Nitrospirotia lineages (Figure 2, Figure 3). All the JdFR Nitrospirotia 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\textsuperscript{0}) within the periplasm of the cell (Figure 2, Figure 3). The capacity for DSR and sulfide oxidation was evident in previously recovered Nitrospirotia 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 Nitrospirotia 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, \textit{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 \textit{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 (\textit{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 (\textit{nar}) and cytoplasmic nitrite reductase (\textit{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 Nitrospirotota lineages (Figure 2). The membrane-bound NiFe hydrogenase *hyb* (Figure 2, Table A6), which oxidizes hydrogen gas (H₂) within the cytoplasm (Teng et al. 2019), and *hya* (Figure 2, Table A5), which oxidizes H₂ in the periplasm were found in all of the JdFR Nitrospirotota lineages from this study. While other NiFe hydrogenases were detected in the JdFR Nitrospirotota 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 Nitrospirotota 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 CO₂ 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 (Figure 2) (Teng et al. 2019) needed for the carbonyl branch of the pathway and facilitate CO₂ fixation (Westphal et al. 2018) in all 4 JdFR Nitrospirotota families. All of the JdFR Nitrospirotota 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 H₂ 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 (Figure 2). 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 (Figure 3). 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 (Figure 2, 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 (Figure 2, Figure 3, 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.

<table>
<thead>
<tr>
<th>Genome</th>
<th>Total SNVs</th>
<th>Shared SNVs</th>
<th>Unique SNVs, U1362A</th>
<th>Unique SNVs, U1362B</th>
</tr>
</thead>
<tbody>
<tr>
<td>JdFRnit3A</td>
<td>380</td>
<td>281</td>
<td>30</td>
<td>69</td>
</tr>
<tr>
<td>JdFRnit7B</td>
<td>517</td>
<td>393</td>
<td>31</td>
<td>93</td>
</tr>
</tbody>
</table>

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 Nitrospirotta.
Figure B5. Metabolic reconstruction of sulfur and nitrogen cycling within the JdFR Nitrospirotini.
REFERENCES


https://doi.org/10.1371/journal.pone.0011147

https://doi.org/10.1023/A:1020514617738

https://doi.org/10.1371/journal.pcbi.1002195


https://doi.org/10.5281/zenodo.854445


https://doi.org/10.1101/gr.113985.110


