frontiers in Microbiology

ORIGINAL RESEARCH
published: 20 August 2021
doi: 10.3389/fmicb.2021.714920

Genomic Variation Influences *Methanothermococcus* Fitness in Marine Hydrothermal Systems

Michael Hoffert[1,2†], Rika E. Anderson[1†], Julie Reveillaud[3], Leslie G. Murphy[4], Ramunas Stepanauskas[5] and Julie A. Huber[6]*

[1] Biology Department, Carleton College, Northfield, MN, United States, [2] Finch Therapeutics Group, Somerville, MA, United States, [3] Maladies Infectieuses et Vecteurs: Ecologie, Génétique, Evolution et Contrôle, University of Montpellier, Institut National de la Recherche Agronomique, Centre National de la Recherche Scientifique, Institut de Recherche Pour le Développement, Montpellier, France, [4] Marine Biological Laboratory, Woods Hole, MA, United States, [5] Bigelow Laboratory for Ocean Sciences, Boothbay, ME, United States, [6] Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, MA, United States

Hydrogenotrophic methanogens are ubiquitous chemoautotrophic archaea inhabiting globally distributed deep-sea hydrothermal vent ecosystems and associated subseafloor niches within the rocky subseafloor, yet little is known about how they adapt and diversify in these habitats. To determine genomic variation and selection pressure within methanogenic populations at vents, we examined five *Methanothermococcus* single cell amplified genomes (SAGs) in conjunction with 15 metagenomes and 10 metatranscriptomes from venting fluids at two geochemically distinct hydrothermal vent fields on the Mid-Cayman Rise in the Caribbean Sea. We observed that some *Methanothermococcus* lineages and their transcripts were more abundant than others in individual vent sites, indicating differential fitness among lineages. The relative abundances of lineages represented by SAGs in each of the samples matched phylogenetic relationships based on single-copy universal genes, and genes related to nitrogen fixation and the CRISPR/Cas immune system were among those differentiating the clades. Lineages possessing these genes were less abundant than those missing that genomic region. Overall, patterns in nucleotide variation indicated that the population dynamics of *Methanothermococcus* were not governed by clonal expansions or selective sweeps, at least in the habitats and sampling times included in this study. Together, our results show that although specific lineages of *Methanothermococcus* co-exist in these habitats, some outcompete others, and possession of accessory metabolic functions does not necessarily provide a fitness advantage in these habitats in all conditions. This work highlights the power of combining single-cell, metagenomic, and metatranscriptomic datasets to determine how evolution shapes microbial abundance and diversity in hydrothermal vent ecosystems.

Keywords: hydrothermal vents, methanogen, single-cell genome, fitness, nitrogenase

OPEN ACCESS

Edited by:
Yohey Suzuki,
The University of Tokyo, Japan

Reviewed by:
Mirjam Perner,
GEOMAR Helmholtz Center for Ocean Research Kiel, Germany
Man Kit Cheung,
The Chinese University of Hong Kong, China

****Correspondence:***
Rika E. Anderson
randerson@carleton.edu

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Extreme Microbiology, a section of the journal Frontiers in Microbiology

Received: *26 May 2021*
Accepted: *31 July 2021*
Published: *20 August 2021*

Citation:
Hoffert M, Anderson RE, Reveillaud J, Murphy LG, Stepanauskas R and Huber JA (2021) Genomic Variation Influences Methanothermococcus Fitness in Marine Hydrothermal Systems. Front. Microbiol. 12:714920. doi: 10.3389/fmicb.2021.714920

INTRODUCTION

Hydrothermal vents host diverse ecosystems sustained by physical and geochemical gradients that are created by the mixing of high-temperature hydrothermal fluid with cold seawater. Several investigations have revealed the intimate relationships between geochemistry, microbial community structure, and metabolic processes in these systems (Huber et al., 2007;

Flores et al., 2012; Akerman et al., 2013; Perner et al., 2013; Reveillaud et al., 2016; Meier et al., 2017; Fortunato et al., 2018). However, little is known about how natural selection molds the genomes of metabolically and taxonomically diverse microbial species inhabiting the rocky subseafloor near hydrothermal vents, especially with respect to genomic variation and plasticity at the population level.

Microbial genomes are highly variable, facilitated by extensive gain, loss, and horizontal transfer of genes between microbes. Previous studies have shown that genome plasticity within species can allow for adaptation to diverse environments (Medini et al., 2005; Tettelin et al., 2005), including in hydrothermal vent habitats (Marteinsson et al., 1999; White et al., 2008; Brazelton et al., 2011; Meyer and Huber, 2014; Moulana et al., 2020). The accessory genome in vent microbial species has been found to include genes related to nitrogen metabolism (Meyer and Huber, 2014; Galambos et al., 2019) and nutrient uptake (Moulana et al., 2020), as well as glycosyltransferases, S-layer proteins, and other membrane-related genes (Anderson et al., 2017). Some studies suggest that the adaptation of particular microbial strains to specific environmental conditions may enable clonal expansions of these strains (Bendall et al., 2016; Shapiro, 2016). Investigation of genomic variation in vents identified patterns of genomic variation indicative of such clonal sweeps (Anderson et al., 2017).

In hydrothermal vent systems, methanogenic archaea are abundant in the subseafloor, venting fluids and the interiors of active chimneys in both basalt- and serpentinite-hosted systems (Schrenk et al., 2003; Takai et al., 2004; Brazelton et al., 2006; Ver Eecke et al., 2013). These archaea consume hydrogen and carbon dioxide, formate, or acetate to produce methane and can grow across a range of temperatures (Takai et al., 2008; Thauer et al., 2008; Stewart et al., 2019). Methanogenic archaea can vary their genomic repertoire in response to carbon sources like formate (Hansen et al., 2011), and single-species methanogenic biofilms from carbonate chimneys at vents show evidence of physiological differentiation (Brazelton et al., 2011).

At the Mid-Cayman Rise, an ultra-slow spreading ridge located in the Caribbean Sea, methanogens from the genus *Methanothermococcus* comprised the most abundant lineage of archaea observed in venting fluids (Reveillaud et al., 2016). The Mid-Cayman Rise hosts two geologically and spatially distinct vent fields: Piccard vent field, an ultra-deep (4,960 m) hydrothermal system situated in mafic rock, and Von Damm vent field, a shallower (2,350 m) ultramafic site on a nearby massif (German et al., 2010). The Von Damm hydrothermal field features fluids enriched with H_2 and CH_4 produced during serpentinization of the ultramafic bedrock in which it is situated (German et al., 2010; Hodgkinson et al., 2015; McDermott et al., 2015, 2020). In contrast, the mafic Piccard hydrothermal field is characterized by fluids with elevated concentrations of sulfide and hydrogen at lower pH and higher temperature (German et al., 2010). Previous work characterizing microbial communities in venting fluids from both vent fields at Mid-Cayman Rise demonstrated that although microbial community structure varies both within and between vent sites, the microbial communities exhibit functional redundancy across sites (Reveillaud et al., 2016). Later work revealed differences in metabolic gene expression across vent sites at the Mid-Cayman Rise, with genes related to methanogenesis more transcribed at Von Damm (Galambos et al., 2019). Patterns of fine-scale genomic variation indicate population-specific differences in selection pressure between Von Damm and Piccard (Anderson et al., 2017). More globally, the ubiquitous sulfur-oxidizing Epsilonbacteraeota genus *Sulfurovum* was shown to have an extensive pangenome, with distinct geographic patterns in gene content between *Sulfurovum* strains in the Mid-Cayman Rise and Axial Seamount in the Pacific Ocean (Moulana et al., 2020). Similar strain-level variation could occur in methanogenic archaea at the Mid-Cayman Rise. 16S rRNA gene sequencing found that different operational taxonomic units (OTUs) of *Methanothermococcus* display varying abundance across samples, indicating niche partitioning within vent sites and between vent fields (Reveillaud et al., 2016). However, niche partitioning and genomic variation have not been examined at the genome scale for thermophilic methanogens, nor is it known which genes have the strongest influence on methanogen fitness.

To gain insight into genomic variation among methanogens inhabiting hydrothermal systems, we examined five single-cell amplified genomes (SAGs) from the genus *Methanothermococcus* in conjunction with 15 metagenomes and 10 metatranscriptomes from the Von Damm and Piccard vent fields at the Mid-Cayman Rise. We examine how the differential abundance of *Methanothermococcus* lineages and their transcripts correlate with the presence and absence of particular genes. Investigation of single nucleotide variation across SAGs and across gene types provides further insights into the way differential gene content can influence fitness of this globally distributed microbial group in hydrothermal vent environments.

MATERIALS AND METHODS

Sample Collection

We collected diffuse flow hydrothermal fluid samples during cruises aboard the R/V *Atlantis* in January 2012 (FS841-FS856) and the R/V *Falkor* in June 2013 (FS866-FS881) using the ROV *Jason* and HROV *Nereus*. Further information about sample collection can be found in Reveillaud et al. (2016) and Galambos et al. (2019). Samples for single-cell genomics analysis were collected by preserving 1 mL of hydrothermal vent fluid with 100 μl glyTE solution [4 parts 11x TE to 5 parts glycerol (Stepanauskas et al., 2017)] followed by incubation for 5 min. at room temperature and storage at −80°C. For metagenomics and metatranscriptomics, we pumped ∼3–6 L of fluid through 0.22 μm Sterivex filters (Millipore) on the seafloor, flooded the filters with RNALater (Ambion) shipboard, sealed the filters with Luer Caps and stored them in sterile Falcon tubes for 18 h at 4°C, then froze them at −80°C.

Single-Cell Genomics

Single cell genomics analyses were performed on one diffuse fluid sample from Ginger Castle site at the Von Damm vent field (sample FS848; see **Supplementary Table 1**). The field sample was thawed, stained with SYTO-9 nucleic acids stain

(Thermo Fisher Scientific) and pre-filtered through a 40 μm mesh-size cell strainer (Becton Dickinson); the individual cells were separated using fluorescence-activated cell sorting (FACS), lysed using a combination of freeze-thaw and alkaline treatment, their genomic DNA was amplified using multiple displacement amplification in a cleanroom, and the resulting SAGs were screened for the 16S rRNA genes as previously described (Stepanauskas et al., 2017). This work was performed at the Bigelow Laboratory Single Cell Genomics Center (SCGC).[1]

Based on 16S rRNA gene analysis, five *Methanothermococcus* SAGs were chosen for sequencing. Libraries were prepared using the Ovation Ultralow Library DR multiplex system (Nugen) and were sheared at 250 bp using a Covaris S-series sonicator. All SAGs were sequenced with NextSeq 500 (Illumina) at the W.M. Keck Facility in the Josephine Bay Paul Center at the Marine Biological Library. Each SAG was sequenced on a different plate to reduce the possibility of cross-contamination through "index-switching" or "sample bleeding" (Mitra et al., 2015; Sinha et al., 2017).

Reads were trimmed and merged using the illumina-utils software package (Eren et al., 2013). To maximize contig length while minimizing the risk of chimeras, reads were assembled using four different software packages: IDBA-UD v.1.1.2 (Peng et al., 2012), SPAdes 3.10.0 (Bankevich et al., 2012), CLC Genomics Workbench v. 7, and A5 (Coil et al., 2015). Contigs from all four assemblers were combined, compared for redundancy, and integrated into a new assembly using the Contig Integrator for Sequence Assembly of bacterial genomes (CISA) (Lin and Liao, 2013).

Pangenomic Analysis of Single Cell Genomes

A phylogenetic tree of all SAGs was constructed using RAxML v. 8.2.9 (Stamatakis, 2014) with 100 bootstraps, using the PROTGAMMAAUTO flag to allow the program to determine the best model of protein evolution. The alignment was created from universal single copy marker genes that were identified and aligned using Phylosift (Darling et al., 2014). The reference sequences were obtained from NCBI, with the following taxID numbers: *Methanococcus maripaludis* C7 (NCBI: txid426368) (designated as the outgroup), *Methanothermococcus thermolithotrophicus* DSM 2095 (NCBI: txid523845), and *Methanothermococcus okinawensis* IH1 uid51535 (NCBI: txid647113).

Open reading frames (ORFs) were identified and annotated for all SAGs using the JGI IMG/ER pipeline (Markowitz et al., 2012). Visualizations of ORFs on contigs and cross-links between ORFs were created using the Biopython collection (Cock et al., 2009) with the GenomeDiagram module. Clustering of ORFs was conducted using the Integrated Toolkit for Exploration of microbial Pangenomes (ITEP) (Benedict et al., 2014). We clustered the ORFs using an inflation value of 2 and a maxbit score of 0.3. ITEP was also used to compare ORF presence/absence across SAGs. KEGG Decoder (Graham et al., 2018) was used to determine pathway completeness for selected metabolic pathways, and the heatmap was visualized with the Python Seaborn library (Waskom et al., 2014). PAML v4.9 (Yang, 2007) was used to calculate dN/dS values for each cluster with greater than one gene assignment. We compiled the results of PAML analysis and the presence/absence table using a custom Python (v.2.7.5) script available on GitHub at https://github.com/carleton-spacehogs/SAGs.

Reads from each SAG were mapped to the contigs of each of the other SAGs using bowtie2 v. 2.2.9 (Langmead and Salzberg, 2012) using default settings. Coverage was calculated using samtools v1.3.1 (Li et al., 2009), and bedtools v2.26.0 (Quinlan, 2014). Average nucleotide identity (ANI) was calculated using the software package pyani v.0.2.7 (Pritchard et al., 2016).

Microbial Community DNA and RNA Preparation and Sequencing

To extract DNA from Sterivex filters for sequencing, we followed methods described in Reveillaud et al. (2016) and Galambos et al. (2019). We extracted total genomic DNA from half of the Sterivex filter, and we extracted RNA from the other half of the Sterivex filter using the mirVana miRNA isolation kit (Ambion), adding a bead-beating step using beads from the RNA PowerSoil kit (MoBio) as described in Fortunato and Huber (2016). RNA was treated with DNase using the Turbo-DNase kit (Ambion), and then purified and concentrated using the RNeasy MinElute kit (Qiagen). We prepared metatranscriptomic libraries using the Ovation Complete Prokaryotic RNA-Seq DR multiplex system (Nugen) following the manufacturer's instructions. We prepared the metagenomic libraries as described in Reveillaud et al. (2016). All libraries were sheared at 175 bp using a Covaris S-series sonicator, yielding paired-end reads with a 30 bp overlap. We sequenced all libraries on an Illumina HiSeq 1000 at the W.M. Keck Facility in the Josephine Bay Paul Center at the Marine Biological Laboratory.

We merged and filtered metagenomic and metatranscriptomic reads using the illumina-utils package (Eren et al., 2013) using iu-merge-pairs with the –enforce-Q30-check flag, followed by iu-filter-merged-reads with a maximum mismatch of 2 in the overlap region. This resulted in reads averaging approximately 170 bp in length. We removed rRNA *in silico* from metatranscriptomes by mapping all merged, filtered reads to the SILVA LSU and SSU Parc databases (release 111) (Pruesse et al., 2007; Quast et al., 2013) using bowtie2 v.2.2.9 with local alignment (Langmead and Salzberg, 2012) and removing all reads that mapped. We used checkm v.1.1.3 (Parks et al., 2015) to assess SAG completeness and contamination.

Metagenome and Metatranscriptome Mapping

We mapped the metagenomic and metatranscriptomic reads (**Supplementary Table 1**) of each sample to the contigs of each SAG (**Table 1**) using bowtie2 v2.2.9 (Langmead and Salzberg, 2012) with default settings. We used anvi'o v6.1 (Eren et al., 2015) to identify, quantify and characterize read coverage and single nucleotide variants (SNVs) for all of the metagenomic reads mapping to each of the SAGs. SNVs

[1]https://scgc.bigelow.org

were only counted if they had a minimum coverage of 10. Anvi'o was also used to calculate average mapping coverages of metagenomes and metatranscriptomes for each of the SAGs. A custom Python database of ORFs, contigs, SNVs, and clusters was compiled for cross-referencing annotation, variability, and homologous clustering data. From this database, SNV densities were calculated based on the length of annotated ORFs, and underlying distributions of SNV density for each sample were estimated using the Python Seaborn library (Waskom et al., 2014). Plots for SNV density vs. n2n1 (or the ratio of the second most frequent nucleotide to the consensus nucleotide), coverage, bubble plots, and average abundance were generated using lookups in this database on a sample- or genome-specific basis. SNV densities were matched to genes annotated by the JGI-IMG annotation pipeline (Markowitz et al., 2012) using in-house Python scripts, available on GitHub at https://github.com/carleton-spacehogs/SAGs. Box-and-whisker plots depicting metatranscriptome coverage of specific genes were created using the Python Seaborn library (Waskom et al., 2014).

CRISPR and Prophage Identification

CRISPR loci were identified in SAG contigs using the CRISPR Recognition Tool (CRT) (Bland et al., 2007). Prophage sequences were identified in SAG contigs using VirSorter (Roux et al., 2015) on the CyVerse platform with default settings.

Data Availability

This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank (https://www.ncbi.nlm.nih.gov) under BioProject PRJNA731123, under accessions JAHEQB000000000 (C09), JAHEQA000000000 (E23), JAHEPZ000000000 (K20), JAHEPY000000000 (M21), and JAHEPX000000000 (N22). The versions described in this paper are JAHEQB010000000 (C09), JAHEQA010000000 (E23), JAHEPZ010000000 (K20), JAHEPY000000000 (M21), and JAHEPX000000000 (N22). Reads for each SAG were deposited at ENA at https://www.ebi.ac.uk/ena/browser/view/SAMEA8800960 (C09), https://www.ebi.ac.uk/ena/browser/view/SAMEA8800961 (E23), https://www.ebi.ac.uk/ena/browser/view/SAMEA8800962 (K20), https://www.ebi.ac.uk/ena/browser/view/SAMEA8800963 (M21), and https://www.ebi.ac.uk/ena/browser/view/SAMEA8800964 (N22). Metagenomic and metatranscriptomic reads are deposited under study accession code PRJEB15541 in the EMBL-EBI European Nucleotide Archive (ENA) database (https://www.ebi.ac.uk/ena/browser/view/PRJEB15541). All code used for this project has been deposited to GitHub at https://github.com/carleton-spacehogs/SAGs.

RESULTS

Characterization of Single-Cell Amplified Genomes

Single-cell amplified genomes (SAGs) were generated from ∼47°C fluids from the Ginger Castle vent site at Von Damm in 2012 (Reveillaud et al., 2016; **Supplementary Table 1**). A summary of the sequencing and assembly of the SAGs is presented in **Table 1**. The SAGs are labeled with their full names in **Table 1**; for the remainder of the manuscript, we will refer to them simply as C09, E23, K20, M21, and N22. The N50 of the assemblies ranged from 56,319 to 8,893 bp, and assembly completeness based on the presence of single-copy universal markers ranged from 92 to 12%. A phylogenetic tree constructed from universal single-copy marker genes revealed two distinct clades, one containing C09, M21, and K20, and the other containing N22 and E23 (**Figure 1**). We mapped the reads of each SAG to the contigs of every other SAG to identify shared genomic regions, and the degree of shared genomic content generally matched the phylogenetic relationships as determined by single copy universal genes (**Supplementary Table 2**). The five SAGs showed low mapping coverage to one another in general (∼9% reads mapping on average), but we observed relatively higher average coverage in SAG-to-SAG mappings among C09, M21, and K20 (∼30% reads mapping on average). Calculation of ANI revealed that C09, M21 and K20 were most similar to each other (approximately 94% ANI on average), whereas E23 and N22 were more distantly related to all the others (approximately 84% ANI to all the other SAGs on average). This is consistent with the phylogenetic relationships illustrated by single-copy universal genes (**Figure 1**).

Gene and Transcript Abundance of *Methanothermococcus* Lineages Across Diffuse Sites

To determine whether the lineages represented by these *Methanothermococcus* SAGs varied in gene and transcript abundance among vent sites, we mapped reads from 15 metagenomes and 10 metatranscriptomes from different sites in the Piccard and Von Damm vent fields to each SAG (**Figure 2**). The SAGs had highest coverage in Von Damm vent sites, including Ginger Castle (FS848), Near Main Orifice (FS866), Shrimp Buttery (FS877), and Old Man Tree (FS881). With the exception of Ginger Castle (48°C), which is the site where the SAGs where sampled from, these vents were characterized by high-temperature fluids (114–131°C, **Supplementary Table 1**). We observed substantial variation in the relative abundance of the lineages represented by these SAGs between sample sites, with E23 and N22 more abundant compared to the other SAG lineages at two vent sites in particular, Old Man Tree and Near Main Orifice. In these sites, E23 had the highest coverage of all the SAGs (**Figure 2A**).

The differences in metagenomic coverage among the SAGs were reflected in differences in transcript abundance, which is an indicator of gene expression. At the whole-genome level, the *Methanothermococcus* lineages represented by these SAGs had the highest transcript abundance at Ginger Castle, the site where the SAGs were sampled from **Figure 2B**. The lineages represented by SAGs E23 and N22, which share a clade distinct from the other *Methanothermococcus* SAGs (**Figure 1**), consistently showed the highest transcript abundance, especially at Old Man Tree and Hot

TABLE 1 | Summary of sequencing and assembly statistics for five *Methanothermococcus* single cell genomes isolated from diffuse fluids at Ginger Castle in Von Damm hydrothermal field.

Name	Completeness (%)	Number of scaffolds > 1 kbp	Assembly size (bp)	GC content (%)	Number of ORFs	N50	Contamination (%)
SCGC AD-155-C09	92.25	125	1,740,890	29.10	1,474	25,888	1.90
SCGC AD-155-E23	91.67	51	1,043,924	42.97	1,221	56,319	2.66
SCGC AD-155-K20	12.85	25	188,232	30.59	205	8,893	1.87
SCGC AD-155-M21	66.32	107	932,247	30.37	937	17,756	1.46
SCGC AD-155-N22	55.12	99	856,655	35.21	866	14,681	5.33

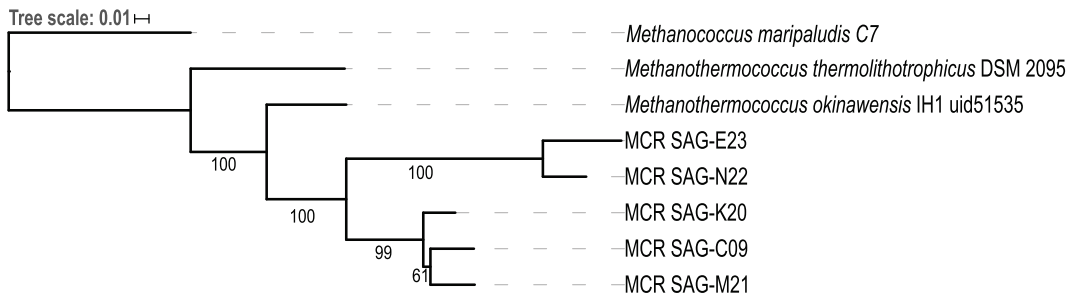

FIGURE 1 | Phylogenetic tree of five *Methanothermococcus* SAGs from the Mid-Cayman Rise (MCR). Single-copy universal marker genes were identified and aligned with PhyloSift. The tree was built with RAxML using 100 bootstraps. Bootstrap values higher than 50 are indicated on the tree. Each of the SAGs sequenced for this study was labeled as "MCR SAG" in the tree for clarity. The scale bar reflects branch lengths in terms of substitutions per site.

Cracks #2 in the Von Damm vent field. This is roughly consistent with their coverage based on metagenomic mapping (**Figure 2A**).

Analysis of Gene Presence and Absence Across *Methanothermococcus* SAGs

To determine whether variation in gene content drove the observed differences in gene and transcript abundance, we examined variable gene content across each of the SAGs. To search for shared genes across all SAGs based on sequence similarity, we clustered all ORFs based on all-v-all BLAST hit scores and compared the presence and absence of gene clusters across each SAG. It is important to note that these SAGs varied in completeness, and thus, absence of evidence for specific genes does not constitute evidence of absence. Therefore, very few genes were considered "core" (that is, found in all five SAGs). Rather than focusing on individual gene presence/absence trends, we instead searched for consistent patterns in gene presence/absence that shared similar functions across the five SAGs. We focused on groups of functional genes whose presence/absence patterns match the phylogenetic relationships shown in **Figure 1**.

Homologous clustering of the ORFs from all five SAGs produced 1,838 clusters of homologous ORFs. ORFs with identical functional annotations and identical presence/absence patterns were tabulated and summarized (**Supplementary Data File 1**). Most variable genes were annotated as hypothetical proteins. The SAGs contained several ORFs that encoded functional genes related to the metabolism of hydrogen, sulfur, and nitrogen (**Supplementary Table 3**), many of which showed consistent presence/absence patterns. With the exception of K20, which was the most incomplete, the SAGs contained genes for iron transport proteins, energy-converting hydrogenases, fumarate reductase, and methyl coenzyme M reductase. Analysis of metabolic pathway completeness revealed that all five SAGs encoded complete or nearly complete pathways for methanogenesis *via* CO_2 and acetate, but lacked the genes required for methanogenesis *via* dimethyl sulfide, methanethiol, or methylamines (**Figure 3**). We did not observe formate dehydrogenase genes in any of the *Methanothermococcus* SAGs. N22 uniquely encoded genes related to chemotaxis that were missing in the other SAGs (**Figure 3**). K20 was missing many metabolic pathways, and had largely incomplete pathways, most likely due to the incompleteness of the SAG assembly (16%).

We identified several genes whose presence/absence patterns matched the SAG phylogenetic relationships. SAGs C09 and M21 shared a large set of gene clusters, which matches their placement on the phylogenetic tree (**Figure 1**), including a cluster of CRISPR-associated proteins (**Supplementary Table 3**). The abundance of genes shared within this clade was in direct contrast to E23 and N22, a clade distinguished primarily by the absence of unique gene clusters relative to the C09/M21/K20 clade (**Supplementary Table 3** and **Supplementary Data File 1**). We identified a genomic region encoding several genes related to nitrogen metabolism, including *nif*, that was present in genomes C09/K20/M21 but not in E23 or N22, consistent with phylogenetic clustering according to single-copy universal genes (**Figure 4A** and **Supplementary Table 3**). Similarly, assessment of metabolic pathway completeness in each SAG confirmed that E23 and N22 encoded incomplete nitrogen fixation pathways, while C09/K20/M21 all encoded complete nitrogen fixation pathways (**Figure 3**). Synteny analyses revealed that the order of genes was consistent across SAGs C09/M21/K20 (**Figure 4A**). The

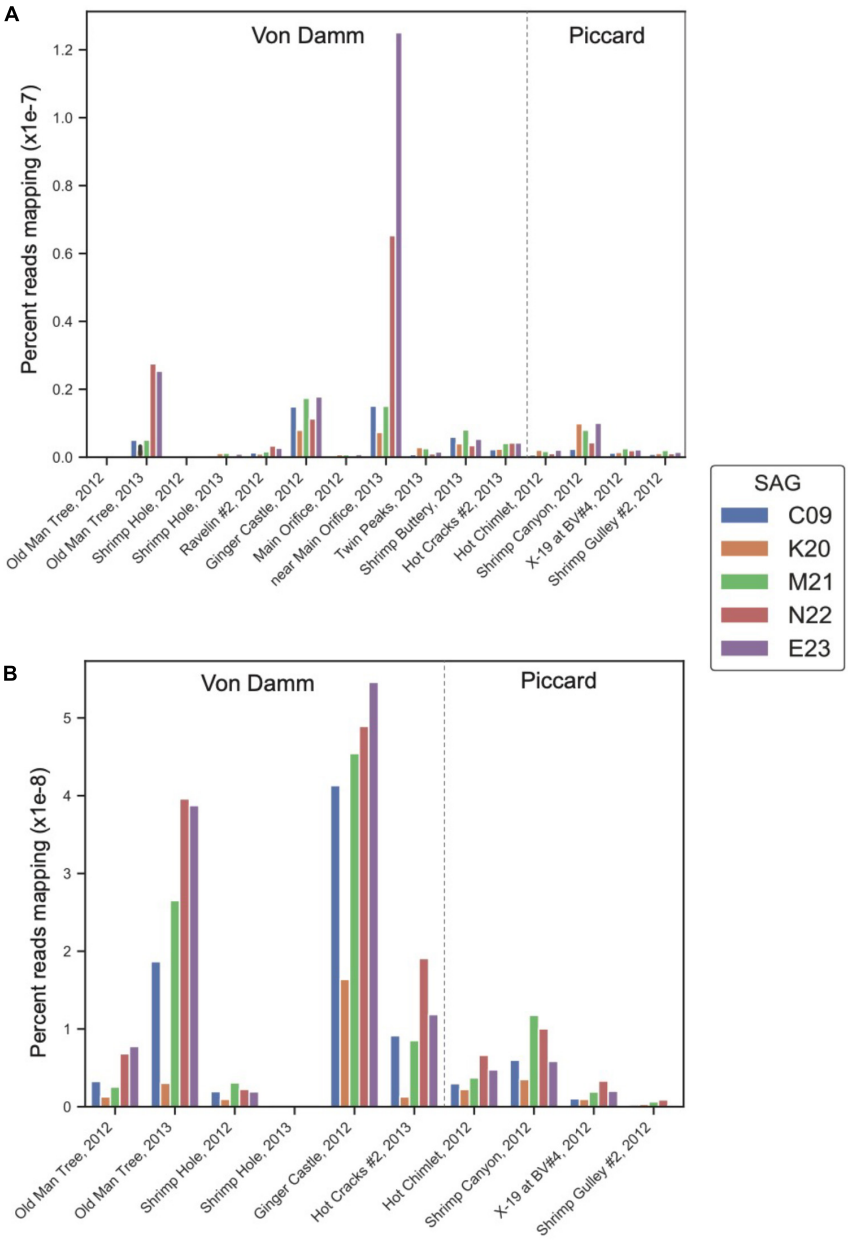

FIGURE 2 | Normalized percent mapping of **(A)** metagenomic reads and **(B)** metatranscriptomic reads for each SAG assembly. The total number of mapped reads to each SAG was divided by the number of metagenomic or metatranscriptomic reads and by total SAG length. **Supplementary Table 1** lists all samples used for mapping.

conserved genes on this island include the alpha and beta chains of a nitrogenase molybdenum-iron protein, a molybdate transport system, and the nitrogen regulatory protein PII. We identified this genomic island on three separate contigs in K20, and although BLAST comparison indicated that the island was not identical between contigs, we cannot eliminate the possibility that the placement of these sequences on three separate contigs resulted from a mis-assembly and actually represents a single region on the genome. This genomic region is surrounded by genes related to glycosyltransferases, methyltransferases, asparagine synthase and tRNA synthetase, and although similar sequences of ORFs were identified in SAG E23, the conserved set of nitrogenase genes identified in C09/M21/K20 were not found on this contig (**Figure 4A**), further suggesting that the island was missing from the E23 genome and not simply missing from the SAG sequence due to incomplete sequencing or assembly.

While all of the SAGs encoded nifH-like genes, most of these nitrogenases fell into a clade containing cluster IV nitrogenases on a phylogenetic tree, which include divergent nitrogenases from archaea that cannot fix N_2 (Chien and Zinder, 1994;

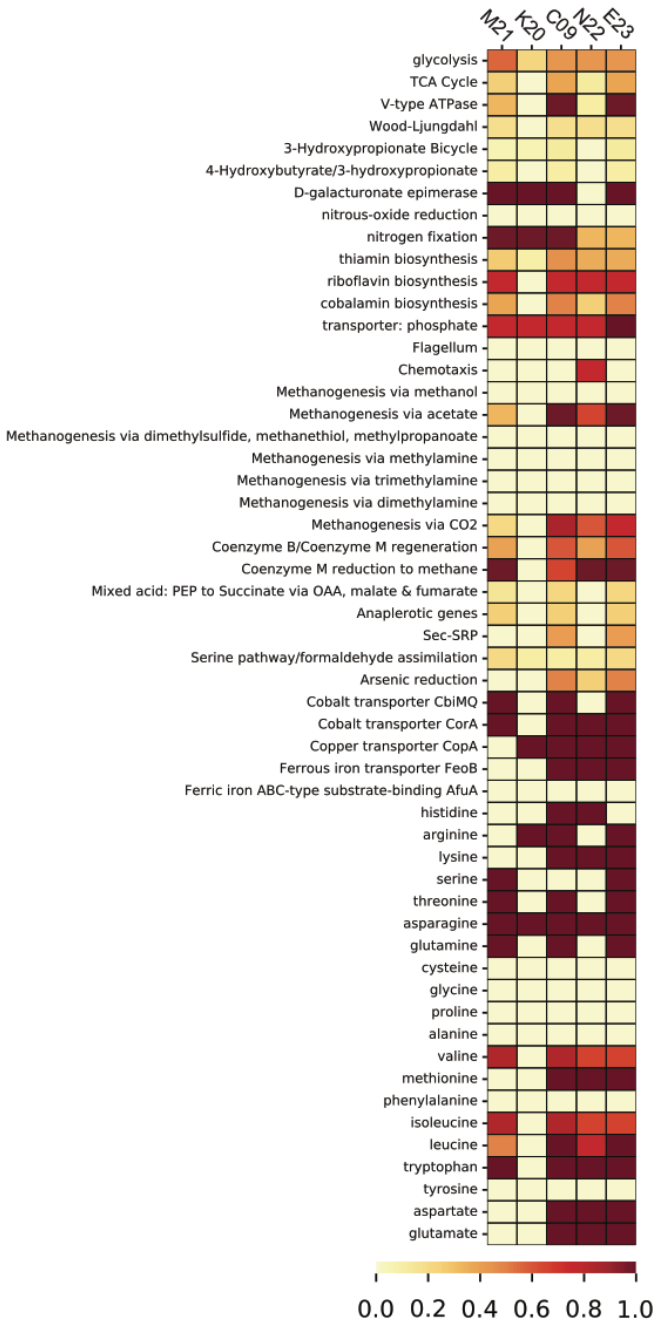

FIGURE 3 | Heatmap depicting pathway completeness of selected metabolic pathways in each of the five *Methanothermococcus* SAGs. Darker values indicate more complete pathways.

Figure 4B). Only the *Methanothermococcus* SAGs K20 and M21 encoded nifH genes from cluster II (**Figure 4B**), a clade which is known to include functional nitrogenases. These cluster II nifH genes were part of the genomic region that was conserved among C09, K20, and M21 (**Figure 4A**).

There is evidence that nitrogenase genes were expressed in the *Methanothermococcus* lineages encoding them. We observed high transcript abundance for the nitrogenase genes in SAGs C09, M21, and K20 at Ginger Castle in Von Damm (where these SAGs were isolated), though the transcript abundance for the nitrogenases was not as high as for genes related to methanogenesis (**Supplementary Figure 1**). Finally, although N22 was distinguished from other SAGs by the presence of chemotaxis-related genes on its genome, we observed very few

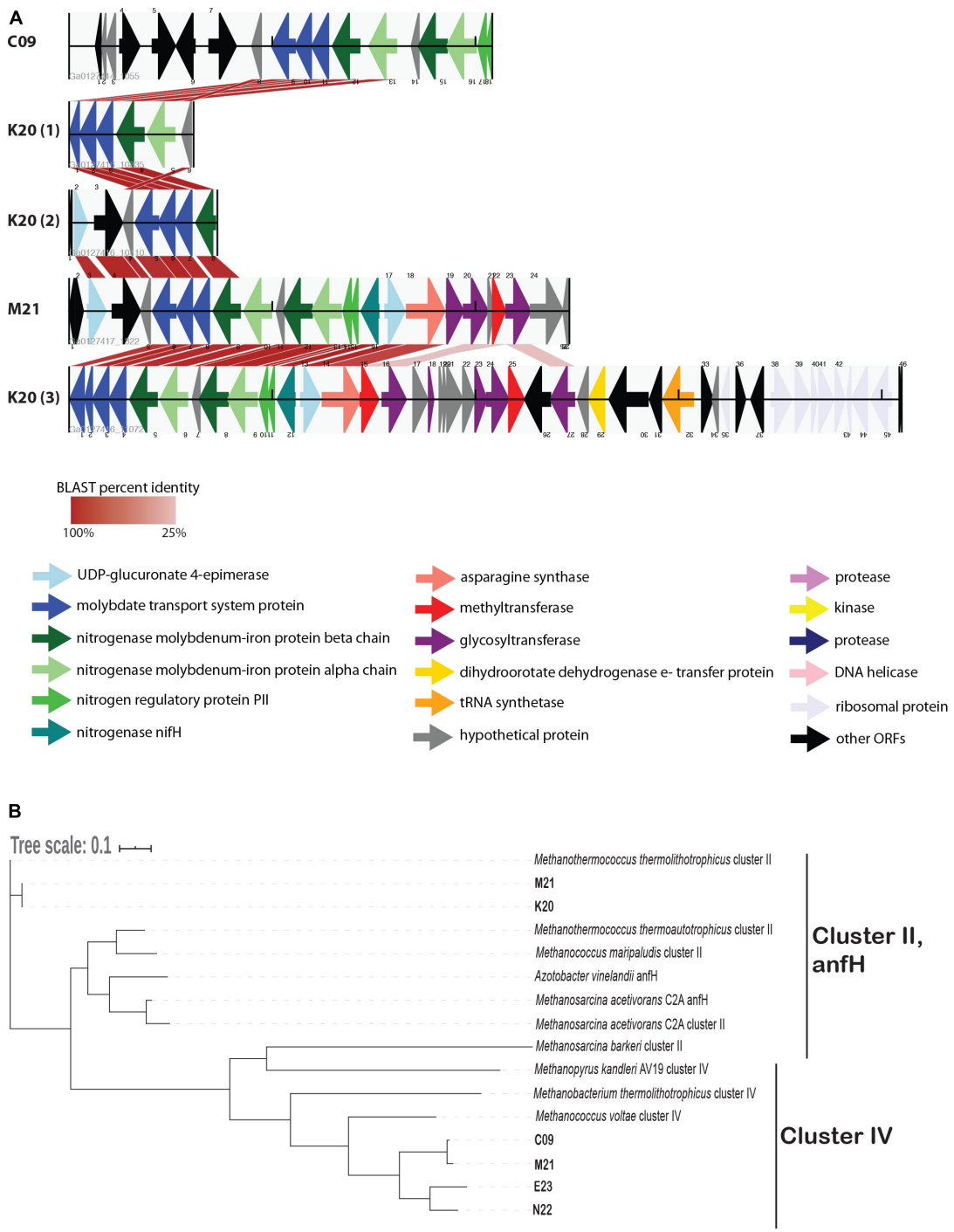

FIGURE 4 | (A) Alignment of open reading frames (ORFs) associated with nitrogenase genes encoded in *Methanothermococcus* SAGs. The ORFs are represented by arrows indicating the direction of transcription. ORFs are color-coded according to their predicted function as determined by the JGI IMG annotation pipeline. Cross-links indicate the best match to the ORFs on the corresponding contig as determined by BLAST analysis; shading of the cross-link indicates the BLAST percent identity, with darker red colors corresponding to higher percent identity. **(B)** Maximum likelihood tree depicting phylogenetic relationships among nifH genes identified in *Methanothermococcus* SAGs. SAG gene numbers are included in the SAG leaf name, and GenBank taxID for reference sequences are included in the Methods. The tree depicts sequences from cluster II and cluster IV nitrogenases. Note that the cluster II nitrogenases included in the tree appear as teal arrows in the ORF diagram. The tree was constructed with RAxML with 100 bootstraps. The scale bar reflects branch lengths in terms of substitutions per site.

transcripts mapping to these chemotaxis genes at Ginger Castle in Von Damm.

CRISPR Loci and Prophage Sequences

To examine viral-driven genomic differentiation among the SAGs, we searched for CRISPR loci and integrated prophage. Several groups of CRISPR-related proteins as well as CRISPR loci were identified in C09, M21, and K20. C09 and M21 each contained six CRISPR loci, with a total of 237 and 109 spacers, respectively, and K20 contained one CRISPR locus with just two spacers (**Table 2** and **Supplementary Figure 2**). No CRISPR genes or CRISPR loci were observed in E23 or N22. We observed substantial variation in spacer content among the CRISPR loci. The vast majority of CRISPR loci in C09, M21 and K20 were unique, with only 18 spacers matching between C09 and M21 (using an e-value cutoff of 0.001 to constitute a "match," **Supplementary Figure 2**). These spacer matches did not occur in the same order, and occasionally matched other spacers within the same SAG. Moreover, many of these spacers matched CRISPR spacers from the cultured isolate *Methanothermococcus okinawensis* (**Supplementary Figure 2**). CRISPR spacers in SAGs C09 and M21 had matches to contigs in several of the metagenomes, with the most matches to sites in Von Damm, including Ginger Castle (FS848), near Main Orifice (FS866), Shrimp Buttery (FS877), Hot Cracks #2 (FS879), and Old Man Tree (FS881) (**Table 2**). Transcripts related to CRISPRs were detected at relatively low levels in SAGs C09 and M21 at Ginger Castle vent, where they were originally isolated (**Supplementary Figure 1**).

In addition to the CRISPR sequences, several putative prophage sequences were identified in *Methanothermococcus* SAG C09. These genes contained phage-like proteins related to phage tail and phage head proteins, many of which match several prophage genes on the previously sequenced *Methanothermococcus okinawensis* (e-value cutoff 1E-05) (**Table 2**). Genes in the integrated prophage region did not appear to be expressed at the time of sampling, as metatranscriptomic mapping to C09 indicated that the transcripts for prophage genes identified in C09 were present at low levels (with only ∼36 reads mapping to the prophage region 1 and ∼19 mapping to prophage region 2) at Ginger Castle in Von Damm (the site from which the SAGs were originally sampled).

Selective Signatures Between SAGs

The dN/dS ratio, or ω, is generally used as an indicator to quantify the strength of selection by comparing the rate of substitution of synonymous sites (dS), which are ostensibly neutral, to the rate of non-synonymous substitutions (dN), which are subject to selection. The ratio depends on variables such as effective population size and the divergence of the microbial populations compared (Kryazhimskiy and Plotkin, 2008). To determine the strength of selection on conserved and variable genes among the *Methanothermococcus* SAGs, we calculated ω for homologous ORF clusters containing sequences from two or more SAGs. ω values varied widely among the ORFs (**Figure 5**), ranging from 0.03 to 3.4, with an average of 0.16. In general, although ω values were somewhat higher than observed in some other microbial populations (e.g., Hu and Blanchard, 2008; Luo et al., 2014; Martinez-Gutierrez and Aylward, 2019), The vast majority of ORFs had ω values less than 1, indicating purifying selection, which is consistent with previous work on signatures of selection in microbial genomes (Kryazhimskiy and Plotkin, 2008). However, several sets of gene clusters containing genes from SAGs in the C09/M21/K20 clade had elevated ω (closer to 1) (**Figure 5**). These gene clusters formed a group distinct from those that were shared between the C09/M21/K20 and E23/N22 clades (**Figure 5**). Very few SNPs were identified in this group of ORFs (highlighted in **Figure 5**), but the average ω for ORFs in this genomic region was close to 1 (**Figure 5**), indicating relaxed purifying selection or mild positive selection. Included in this group of ORFs were nitrogenases encoded on the genomic region that was unique to C09, M21, and K20.

Nucleotide Variation Within Populations

Genomic variation within natural microbial populations can reveal whether specific populations have recently been subjected to selective sweeps or to clonal expansions (Bendall et al., 2016; Shapiro, 2016; Anderson et al., 2017). To characterize genomic variation within natural populations of *Methanothermococcus*, we measured SNV density and allele frequency (measured as n2n1) for each metagenome mapping to each SAG and compared this to coverage, which is a measure of relative abundance (**Supplementary Figure 3**). SNV density varied substantially within sample sites. Generally, SAGs with low average coverage had low SNV density (<10) and low n2n1 (<0.1) at those sample sites. In general, we consistently observed high SNV density associated with high coverage mappings (**Supplementary Figure 3**). The relationship between mapping coverage and SNVs/kbp was positive, with R^2 values above 0.80, indicating that increased mapping coverage also increased the number of SNVs observed (**Supplementary Figure 4**). However, the slope of the relationship varied between SAGs (**Supplementary Figure 4**), suggesting differential rates of SNV accumulation among the *Methanothermococcus* lineages. SAGs with high phylogenetic relatedness showed similar trends of SNV density (**Supplementary Figure 4**). We did not observe cases in which SAGs had both high abundance and low SNV density, which would be interpreted as evidence of a clonal expansion or selective sweep.

DISCUSSION

Methanothermococcus Lineages Exhibited Varying Gene and Transcript Abundance Across Vent Sites

Although methanogens are among the most common chemoautotrophic archaea observed in hydrothermal systems and constitute an important part of the microbial community, little is known about how they adapt and evolve in these globally distributed deep-sea ecosystems. Our results suggest that specific lineages of *Methanothermococcus* have consistently higher

TABLE 2 | Prophage and CRISPR loci in *Methanothermococcus* SAGs from the Mid-Cayman Rise (MCR).

	CRISPR loci	
	Number of CRISPR loci	**Number of matches to MCR metagenome contigs**
C09	6	FS848 (10); FS854 (1); FS866 (26); FS877 (1); FS879 (4); FS881 (8)
E23	0	–
K20	1	0
M21	6	FS848 (1); FS866 (4); FS879 (1)
N22	0	–
	Prophage	
	Prophage genes identified	**Number of matches to *Methanothermococcus okinawensis* phage**
C09	• Category 2 putative phage on 2 contigs. 27 genes total. • 20 hypothetical proteins • Terminase-like family protein • Phage putative head morphogenesis protein, SPP1 gp7 family • Phage protein, HK97 gp10 family • Phage tail tape measure protein • Phage major capsid protein E • Bacteriophage lambda head decoration protein D	22
E23	None	–
K20	None	–
M21	None	–
N22	None	–

Phage clusters listed are as defined by the VirSorter database (Roux et al., 2015). BLAST matches correspond to a maximum e-value cutoff of 1e-05. M. okinawensis phage were identified using VirSorter.

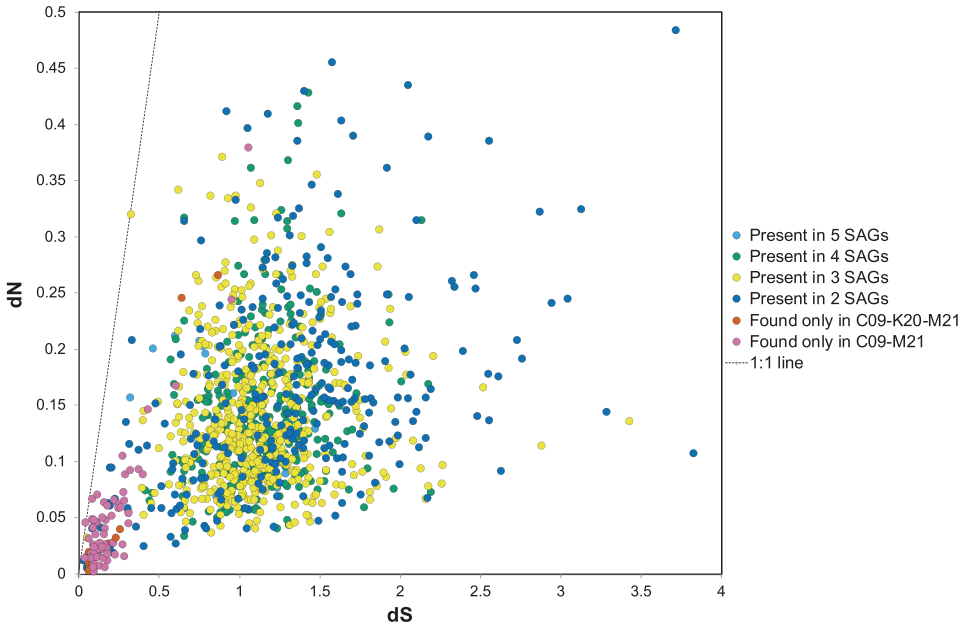

FIGURE 5 | dN relative to dS for open reading frame (ORF) clusters. The dots in orange or pink represent ORFs shared between C09/M21 or C09/M21/K20, which includes the nitrogenase island of genes. Clusters of genes found in single genomes are excluded from this plot. The dotted line represents dN/dS = 1.

genomic and transcript abundance in ultramafic hydrothermal vent settings (at Von Damm vent field) compared to mafic (at Piccard). Previous work has shown that methanogens in general, and *Methanothermococcus* in particular, were abundant in the Von Damm vent field (Reveillaud et al., 2016), and methanogens have been isolated from Von Damm (Sakai et al., 2019). The geochemistry of venting fluids at Von Damm is influenced by serpentinization, and thermodynamics calculations have suggested slightly higher catabolic energy availability for methanogenesis at Von Damm compared to

Piccard (Reveillaud et al., 2016). Therefore, conditions at Von Damm are more favorable for elevated methanogen abundance and activity. In the present study, some *Methanothermococcus* lineages consistently displayed higher abundance relative to all others, indicating differences in fitness among the lineages represented by the SAGs, raising the question of what genomic features enabled some SAGs to have higher transcript abundance than others in these vent habitats. Although it is difficult to identify the specific genomic features that provide a selective advantage, notable differences among the lineages studied here included genes related to nitrogen fixation, chemotaxis, CRISPR proteins, and viruses. The genes and genomes also show distinct patterns of nucleotide variation across environments, suggesting that selection operates differently on these lineages depending on the environmental context.

Presence of Accessory Functions in the Genome Does Not Confer Selective Advantages

We identified several genomic features that matched the phylogenetic placement of these SAGs according to single-copy universal genes, and which may have contributed to the differences in fitness between lineages. The lineages represented by the E23 and N22 SAGs, which were more abundant than the other lineages, were distinguished by the absence, rather the presence, of unique genes. In contrast, SAGs in the C09/K20/M21 clade shared a set of nitrogen-fixing genes in the *nif* family that are known to be part of the functional clade of cluster II nitrogenases. Diazotrophic growth is widespread among methanogens, including *Methanothermococcus* (Leigh, 2000; Nishizawa et al., 2014). Genomic islands encoding nitrogenase have previously been observed in the *Lebetimonas* (order *Nautiliales*) strains isolated from the subseafloor of hydrothermal vents (Meyer and Huber, 2014). In addition, nitrogen-fixing genes as well as demonstrated nitrogen fixing ability have been observed in methanogens from hydrothermal vents previously (Mehta et al., 2003; Mehta and Baross, 2006; Ver Eecke et al., 2013), including from a *Methanofervidicoccus* (order *Methanococcales*) strain isolated from the Mid-Cayman Rise (Wang et al., 2020). Although there is some ammonium enrichment relative to deep background seawater in Mid-Cayman Rise vent fluids, the maximum concentration is about 15–20 μmol, and nitrate is undetectable (Reeves et al., 2014; Charoenpong et al., 2015). Thus, fixed nitrogen sources are somewhat limited. Moreover, nitrogen-fixing genes are known to be part of the accessory genome in other microbial lineages in the vent environment (Meyer and Huber, 2014). The *nif* genes in these *Methanothermococcus* SAGs were transcribed and highly conserved. However, these genes were under relaxed purifying selection or mild positive selection, suggesting that natural selection allowed for some nucleotide diversification in these genes. Moreover, the *nif* genes do not appear to have provided a fitness advantage to these lineages at the time of sampling, as SAGs encoding this genomic region had distinctly lower transcript abundance and lower overall abundance compared to the SAGs missing this genomic region. Thus, our results suggest that the possession of an accessory metabolic function (in this case, nitrogen fixation) does not necessarily confer an adaptive advantage to the lineages possessing it with respect to population size, at least at the time of sampling. However, as stated previously, given the incompleteness of the SAGs, absence of evidence for certain genes does not constitute evidence of absence, and thus these results must be considered with this caveat. We focused on genes that matched the phylogenetic placement of the SAGs because these genes are slightly more likely to have been truly absent or present on the genomes of their respective lineages.

Our results also indicate that *Methanothermococcus* lineages in hydrothermal systems are actively infected by viruses. Previous research has shown that the genomes of thermophilic methanogens are particularly enriched with CRISPR loci (Anderson et al., 2011), and we identified several CRISPR loci in these SAGs. C09 and M21 encoded several distinct CRISPR-related proteins and lysogenic components of viruses, suggesting that these lineages may be actively infected by viruses in the environment. CRISPR matches between the SAGs and the metagenomic contigs, transcripts of CRISPR and prophage-related genes, and the diversity and lack of conservation between spacers in the CRISPR loci all suggest that *Methanothermococcus* are susceptible to frequent viral infection in these vent systems. Previous work on the isolation of methanogen-targeting archaeal viruses from hydrothermal systems (Thiroux et al., 2021) and the high abundance of CRISPR loci in vent methanogens (Anderson et al., 2011) is consistent with this observation. However, we did not identify any CRISPR loci in the *Methanothermococcus* lineages with the highest abundance and gene transcription across vent sites (E23 and N22). It is possible that the CRISPR loci were not present due to incomplete sequencing and assembly, but the absence of any detectible CRISPR loci or CRISPR-related proteins on these two closely related lineages suggests that, as with the nitrogenase cassette, possession of accessory functions such as CRISPR antiviral immune systems does not necessarily provide a selective advantage. Instead, both E23 and N22 encoded some restriction modification genes, membrane proteins, and glycosyltransferases that were absent in other lineages. Both restriction enzymes and modifications to the cell wall and outer membrane pose alternative strategies by which these organisms may evade viral predation. Thus, the presence of antiviral mechanisms and variable CRISPR loci indicates that some of these lineages may be actively infected by viruses in the environment, and that differential susceptibility to viral infection could contribute to differential fitness of the *Methanothermococcus* lineages.

Nucleotide Variation in Naturally Occurring *Methanothermococcus* Populations Suggests a Lack of Clonal Expansions

Finally, mapping of metagenomic reads to the SAGs allowed us to examine fine-scale variation at the nucleotide and amino acid levels, which revealed insights into selection pressure in

the natural environment. We observed that closely related *Methanothermococcus* lineages accumulated variation at similar rates across different vent sites, indicating that there are consistent patterns of mutation accumulation that match *Methanothermococcus* phylogeny regardless of sample site. Thus, patterns of mutation accumulation or loss due to selection or drift appear to correlate closely to phylogeny. The density of SNVs was relatively high across all sample sites, suggesting that mutation rates were high or that clonal sweeps were not common in these methanogenic populations, in contrast to observations of microbial populations in previous studies (Bendall et al., 2016; Anderson et al., 2017). It appears that *Methanothermococcus* lineages at these two hydrothermal vent sites are not susceptible to blooms or extinction events and are instead maintained in high abundances consistently over time, with specific lineages maintaining higher abundance than others.

CONCLUSION

Together, our analysis of 5 *Methanothermococcus* SAGs alongside 15 metagenomes and 10 metatranscriptomes from diffuse flow fluids at the Mid-Cayman Rise reveals that the genomes of methanogens in vent environments exhibit a plasticity that can lead to differences in growth potential and gene expression across vent habitats. Some *Methanothermococcus* lineages exhibited higher gene and transcript abundance in some vent sites than in others. However, in contrast to observations of other populations in these vent fields (Anderson et al., 2017), we did not observe evidence for frequent clonal sweeps. Instead, the *Methanothermococcus* lineages with the highest fitness appeared to have high rates of natural variation, indicating that they had been maintained in high abundance for long enough to accumulate variation. The most successful *Methanothermococcus* lineages were notable for the absence of genes related to nitrogen fixation, prophage sequences and the CRISPR/Cas immune response, suggesting that the possession of accessory functions related to nitrogen fixation and viral immunity may not necessarily provide a selective advantage in these conditions. Continued work combining genomic and transcriptomic approaches will be key to elucidating the mechanisms by which methanogens co-evolve with their environment in hydrothermal vent systems.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/, JAHEQB000000000; https://www.ncbi.nlm.nih.gov/, JAHEQA000000000; https://www.ncbi.nlm.nih.gov/, JAHEPZ000000000; https://www.ncbi.nlm.nih.gov/, JAHEPY000000000; https://www.ncbi.nlm.nih.gov/, JAHEPX000000000; https://www.ebi.ac.uk/ena, ERS6485300; https://www.ebi.ac.uk/ena, ERS6485304; https://www.ebi.ac.uk/ena, ERS6485303; https://www.ebi.ac.uk/ena, ERS6485302; and https://www.ebi.ac.uk/ena, ERS648530.

AUTHOR CONTRIBUTIONS

MH, RA, and JH designed the study and wrote the manuscript. JR and LM collected and prepared samples for sequencing. RS sorted and amplified the cells. All authors contributed to the creation of the manuscript and study design.

FUNDING

Support for MH was provided by Carleton College. RA was supported by a NASA Postdoctoral Fellowship with the NASA Astrobiology Institute. This work was supported by grants from the Deep Carbon Observatory's Deep Life Initiative to RA and JH, NASA Exobiology Grant Number 80NSSC18K1076 to RA and JH, NASA Astrobiology Science and Technology for Exploring Planets (ASTEP) Grant NNX-327 09AB75G to JH, the NSF Science and Technology Center for Dark Energy Biosphere Investigations (C-DEBI) to JH, and NSF Grants OCE-1136488 and OIA-1826734 to RS. Data collected in this study in 2013 are based upon work supported by the Schmidt Ocean Institute during cruise FX008-2013 aboard R/V Falkor. Ship and vehicle time in 2012 were supported by the NSF-OCE Grant OCE-1061863. Funds for open access publication were provided by the Schmidt Ocean Institute. This is C-DEBI contribution number 574.

ACKNOWLEDGMENTS

We would like to thank Victor Huerta and Jackson Raynor for assistance with bioinformatics analysis, and Chip Brier, Chris German, Jeff Seewald, Cindy Lee Van Dover, Jill McDermott, and Emily Reddington for assistance with sample collection, processing, and analysis. We would also like to thank the captains and crew of R/V *Atlantis* and R/V *Falkor* as well as ROVs *Jason* and *Nereus* for assistance with sample collection and preparation. We also thank the SCGC staff for single cell genomics analyses.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.714920/full#supplementary-material

Supplementary Figure 1 | Box-and-whisker and swarm plots indicating the metatranscriptomic coverage of specific genes in *Methanothermococcus* single cell-amplified genomes in sample FS848, Ginger Castle, at Von Damm hydrothermal field. Coverage values are not normalized between SAGs, and thus relative transcript abundance of specific gene types should only be compared within individual SAGs. The boxes show the quartiles of the dataset, and the whiskers extend to show the rest of the distribution. Outliers are indicated by points outside of the whiskers. Individual dots indicate individual gene coverage values for genes within that category. Y axes are split to depict the full range of coverage values. A list of the KEGG numbers that were included within each category is listed in **Supplementary Data File 1**.

Supplementary Figure 2 | Summary of CRISPR loci in the 5 *Methanothermococcus* SAGs from Ginger Castle vent from Von Damm vent field

on the Mid-Cayman Rise, sampled in 2012, plus *Methanothermococcus okinawensis* IH1 for reference. Each of the CRISPR loci is represented by a line with a series of boxes, in which each box represents a spacer sequence. Matching spacers are shown in the same color; unique spacers are shown in black.

Supplementary Figure 3 | Plot of SNVs per kilobase pair vs. n2n1 (majority allele frequency) for metagenome mappings to single cell amplified genomes from Piccard (yellow) and Von Damm (blue) vent fields. All y-axes reflect n2n1, or the ratio of the second most common allele to the most common allele. Bubble diameter corresponds to the average coverage of each SAG-metagenome mapping normalized by number of metagenomic reads. Bubbles are colored according to which vent field they came from; sample numbers are indicated.

Supplementary Figure 4 | Scatterplots of average mapping coverage vs. average number of SNVs per kilobase pair of DNA for metagenomic mappings, colored by SAG. Each point represents data from a different metagenome mapped to each SAG.

Supplementary Table 1 | Data regarding quality filtering and assembly of metagenomic and metatranscriptomic reads. No MT, no metatranscriptome. NA, not available due to loss of USBL navigation by the ROV.

Supplementary Table 2 | Percent reads mapping for all SAG-to-SAG mappings. SAG contigs used as reference are listed on the left, SAG reads mapped to the reference are listed along the top of the table. Numbers are listed as percent of merged reads mapped using bowtie2.

Supplementary Table 3 | Selected gene counts for genes present and/or absent in each SAG from the Mid-Cayman Rise.

Supplementary Data File 1 | An Excel spreadsheet that lists all ORF clusters as well as a list of the KO categories used to make the box-and-whisker plot.

REFERENCES

Akerman, N. H., Butterfield, D. A., and Huber, J. A. (2013). Phylogenetic diversity and functional gene patterns of sulfur-oxidizing subseafloor *Epsilonproteobacteria* in diffuse hydrothermal vent fluids. *Front. Microbiol.* 4:185. doi: 10.3389/fmicb.2013.00185

Anderson, R. E., Brazelton, W. J., and Baross, J. A. (2011). Using CRISPRs as a metagenomic tool to identify microbial hosts of a diffuse flow hydrothermal vent viral assemblage. *FEMS Microbiol. Ecol.* 77, 120–133.

Anderson, R. E., Reveillaud, J., Reddington, E., Delmont, T. O., Eren, A. M., McDermott, J. M., et al. (2017). Genomic variation in microbial populations inhabiting the marine subseafloor at deep-sea hydrothermal vents. *Nat. Commun.* 8:1114. doi: 10.1038/s41467-017-01228-6

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021

Bendall, M. L., Stevens, S. L., Chan, L.-K., Malfatti, S., Schwientek, P., Tremblay, J., et al. (2016). Genome-wide selective sweeps and gene-specific sweeps in natural bacterial populations. *ISME J.* 10, 1589–1601. doi: 10.1038/ismej.2015.241

Benedict, M. N., Henriksen, J. R., Metcalf, W. W., Whitaker, R. J., and Price, N. D. (2014). ITEP: an integrated toolkit for exploration of microbial pan-genomes. *BMC Genomics* 15:8. doi: 10.1186/1471-2164-15-8

Bland, C., Ramsey, T. L., Sabree, F., Lowe, M., Brown, K., Kyrpides, N. C., et al. (2007). CRISPR Recognition Tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* 8:209. doi: 10.1186/1471-2105-8-209

Brazelton, W. J., Mehta, M. P., Kelley, D. S., and Baross, J. A. (2011). Physiological differentiation within a single-species biofilm fueled by serpentinization. *mBio* 2:e00127-11. doi: 10.1128/mBio.00127-11

Brazelton, W. J., Schrenk, M. O., Kelley, D. S., and Baross, J. A. (2006). Methane- and sulfur-metabolizing microbial communities dominate the Lost City hydrothermal field ecosystem. *Appl. Environ. Microbiol.* 72, 6257–6270. doi: 10.1128/AEM.00574-06

Charoenpong, C., Wankel, S. D., and Seewald, J. (2015). *Stable Isotope Evidence for Abiotic Ammonium Production in the Hydrothermal Vent Fluids from the Mid-Cayman Rise. AGUFM* 2015, OS43A-2015. Available online at: https://ui.adsabs.harvard.edu/abs/2015AGUFMOS43A2014C/abstract (accessed May 23, 2021).

Chien, Y. T., and Zinder, S. H. (1994). Cloning, DNA sequencing, and characterization of a nifD-homologous gene from the archaeon *Methanosarcina barkeri* 227 which resembles nifD1 from the eubacterium *Clostridium pasteurianum*. *J. Bacteriol.* 176, 6590–6598. doi: 10.1128/jb.176.21.6590-6598.1994

Cock, P. J. A., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., et al. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25, 1422–1423. doi: 10.1093/bioinformatics/btp163

Coil, D., Jospin, G., and Darling, A. E. (2015). A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics* 31, 587–589. doi: 10.1093/bioinformatics/btu661

Darling, A. E., Jospin, G., Lowe, E., Matsen, F. A., Bik, H. M., and Eisen, J. A. (2014). PhyloSift: phylogenetic analysis of genomes and metagenomes. *PeerJ* 2:e243. doi: 10.7717/peerj.243

Eren, A. M., Esen, Ö. C., Quince, C., Vineis, J. H., Morrison, H. G., Sogin, M. L., et al. (2015). Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* 3:e1319. doi: 10.7717/peerj.1319

Eren, A. M., Vineis, J. H., Morrison, H. G., and Sogin, M. L. (2013). A filtering method to generate high quality short reads using illumina paired-end technology. *PLoS One* 8:e66643. doi: 10.1371/journal.pone.0066643

Flores, G. E., Shakya, M., Meneghin, J., Yang, Z. K., Seewald, J. S., Wheat, G. C., et al. (2012). Inter-field variability in the microbial communities of hydrothermal vent deposits from a back-arc basin. *Geobiology* 10, 333–346. doi: 10.1111/j.1472-4669.2012.00325.x

Fortunato, C. S. and Huber, J. A. (2016). Coupled RNA-SIP and metatranscriptomics of active chemolithoautotrophic communities at a deep-sea hydrothermal vent. *ISMEJ* 10, 1925–1938. doi: 10.1038/ismej.2015.258

Fortunato, C. S., Larson, B., Butterfield, D. A., and Huber, J. A. (2018). Spatially distinct, temporally stable microbial populations mediate biogeochemical cycling at and below the seafloor in hydrothermal vent fluids. *Environ. Microbiol.* 20, 769–784. doi: 10.1111/1462-2920.14011

Galambos, D., Anderson, R. E., Reveillaud, J., and Huber, J. A. (2019). Genome-resolved metagenomics and metatranscriptomics reveal niche differentiation in functionally redundant microbial communities at deep-sea hydrothermal vents. *Environ. Microbiol.* 21, 4395–4410. doi: 10.1111/1462-2920.14806

German, C. R., Bowen, A., Coleman, M. L., Honig, D. L., Huber, J. A., Jakuba, M. V., et al. (2010). Diverse styles of submarine venting on the ultraslow spreading Mid-Cayman Rise. *Proc. Natl. Acad. Sci. U.S.A.* 107, 14020–14025. doi: 10.1073/pnas.1009205107

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

Hansen, E. E., Lozupone, C. A., Rey, F. E., Wu, M., Guruge, J. L., Narra, A., et al. (2011). Pan-genome of the dominant human gut-associated archaeon, *Methanobrevibacter smithii*, studied in twins. *Proc. Natl. Acad. Sci. U.S.A.* 108(Suppl.), 4599–4606. doi: 10.1073/pnas.1000071108

Hodgkinson, M. R. S., Webber, A. P., Roberts, S., Mills, R. A., Connelly, D. P., and Murton, B. J. (2015). Talc-dominated seafloor deposits reveal a new class of hydrothermal system. *Nat. Commun.* 6:10150. doi: 10.1038/ncomms10150

Hu, J., and Blanchard, J. L. (2008). Environmental sequence data from the Sargasso sea reveal that the characteristics of genome reduction in prochlorococcus are not a harbinger for an escalation in genetic drift. *Mol. Biol. Evol.* 26, 5–13. doi: 10.1093/molbev/msn217

Huber, J. A., Mark Welch, D. B., Morrison, H. G., Huse, S. M., Neal, P. R., Butterfield, D. A., et al. (2007). Microbial population structures in the deep marine biosphere. *Science* 318, 97–100. doi: 10.1126/science.1146689

Kryazhimskiy, S., and Plotkin, J. B. (2008). The population genetics of dN/dS. *PLoS Genet.* 4:e1000304. doi: 10.1371/journal.pgen.1000304

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

Leigh, J. A. (2000). Nitrogen fixation in methanogens: the archaeal perspective. *Curr. Issues Mol. Biol.* 2, 125–131.

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

Lin, S.-H., and Liao, Y.-C. (2013). CISA: contig integrator for sequence assembly of bacterial genomes. *PLoS One* 8:e60843. doi: 10.1371/journal.pone.0060843

Luo, H., Swan, B. K., Stepanauskas, R., Hughes, A. L., and Moran, M. A. (2014). Comparing effective population sizes of dominant marine alphaproteobacteria lineages. *Environ. Microbiol. Rep.* 6, 167–172. doi: 10.1111/1758-2229.12129

Markowitz, V. M., Chen, I. M. A., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., et al. (2012). IMG: the integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res.* 40, D115–D122. doi: 10.1093/nar/gkr1044

Marteinsson, V. T., Birrien, J. L., Reysenbach, A. L., Vernet, M., Marie, D., Gambacorta, A., et al. (1999). *Thermococcus barophilus* sp. nov., a new barophilic and hyperthermophilic archaeon isolated under high hydrostatic pressure from a deep-sea hydrothermal vent. *Int. J. Syst. Bacteriol.* 49, 351–359.

Martinez-Gutierrez, C. A., and Aylward, F. O. (2019). Strong purifying selection is associated with genome streamlining in Epipelagic *Marinimicrobia*. *Genome Biol. Evol.* 11, 2887–2894. doi: 10.1093/gbe/evz201

McDermott, J. M., Seewald, J. S., German, C. R., and Sylva, S. P. (2015). Pathways for abiotic organic synthesis at submarine hydrothermal fields. *Proc. Natl. Acad. Sci. U.S.A.* 112, 7668–7672. doi: 10.1073/pnas.1506295112

McDermott, J. M., Sylva, S. P., Ono, S., German, C. R., and Seewald, J. S. (2020). Abiotic redox reactions in hydrothermal mixing zones: decreased energy availability for the subsurface biosphere. *Proc. Natl. Acad. Sci. U.S.A.* 117, 20453–20461. doi: 10.1073/pnas.2003108117

Medini, D., Donati, C., Tettelin, H., Masignani, V., and Rappuoli, R. (2005). The microbial pan-genome. *Curr. Opin. Genet. Dev.* 15, 589–594. doi: 10.1016/j.gde.2005.09.006

Mehta, M. P., and Baross, J. A. (2006). Nitrogen fixation at 92°C by a hydrothermal vent archaeon. *Science* 314, 1783–1786. doi: 10.1126/science.1134772

Mehta, M. P., Butterfield, D. A., and Baross, J. A. (2003). Phylogenetic diversity of nitrogenase (nifH) genes in deep-sea and hydrothermal vent environments of the Juan de Fuca Ridge. *Appl. Environ. Microbiol.* 69, 960–970. doi: 10.1128/AEM.69.2.960-970.2003

Meier, D. V., Pjevac, P., Bach, W., Hourdez, S., Girguis, P. R., Vidoudez, C., et al. (2017). Niche partitioning of diverse sulfur-oxidizing bacteria at hydrothermal vents. *ISME J.* 11, 1545–1558. doi: 10.1038/ismej.2017.37

Meyer, J. L., and Huber, J. A. (2014). Strain-level genomic variation in natural populations of Lebetimonas from an erupting deep-sea volcano. *ISME J.* 8, 867–880. doi: 10.1038/ismej.2013.206

Mitra, A., Skrzypczak, M., Ginalski, K., and Rowicka, M. (2015). Strategies for achieving high sequencing accuracy for low diversity samples and avoiding sample bleeding using Illumina Platform. *PLoS One* 10:e0120520. doi: 10.1371/journal.pone.0120520

Moulana, A., Anderson, R. E., Fortunato, C. S., and Huber, J. A. (2020). Selection is a significant driver of gene gain and loss in the pangenome of the bacterial genus Sulfurovum in geographically distinct deep-sea hydrothermal vents. *mSystems* 5:e00673-19. doi: 10.1128/mSystems.00673-19

Nishizawa, M., Miyazaki, J., Makabe, A., Koba, K., and Takai, K. (2014). Physiological and isotopic characteristics of nitrogen fixation by hyperthermophilic methanogens: key insights into nitrogen anabolism of the microbial communities in Archean hydrothermal systems. *Geochim. Cosmochim. Acta* 138, 117–135. doi: 10.1016/J.GCA.2014.04.021

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

Peng, Y., Leung, H. C. M., Yiu, S. M., and Chin, F. Y. L. (2012). IDBA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28, 1420–1428. doi: 10.1093/bioinformatics/bts174

Perner, M., Gonnella, G., Hourdez, S., Böhnke, S., Kurtz, S., and Girguis, P. (2013). In situ chemistry and microbial community compositions in five deep-sea hydrothermal fluid samples from Irina II in the Logatchev field. *Environ. Microbiol.* 15, 1551–1560. doi: 10.1111/1462-2920.12038

Pritchard, L., Glover, R. H., Humphris, S., Elphinstone, J. G., and Toth, I. K. (2016). Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Anal. Methods* 8, 12–24. doi: 10.1039/c5ay02550h

Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., et al. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* 35, 7188–7196. doi: 10.1093/nar/gkm864

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590–D596. doi: 10.1093/nar/gks1219

Quinlan, A. R. (2014). BEDTools: the Swiss-army tool for genome feature analysis. *Curr. Protoc. Bioinformatics* 47, 11.12.1–11.12.34. doi: 10.1002/0471250953.bi1112s47

Reeves, E. P., McDermott, J. M., and Seewald, J. S. (2014). The origin of methanethiol in midocean ridge hydrothermal fluids. *Proc. Natl. Acad. Sci. U.S.A.* 111, 5474–5479. doi: 10.1073/pnas.1400643111

Reveillaud, J., Reddington, E., McDermott, J., Algar, C., Meyer, J. L., Sylva, S., et al. (2016). Subseafloor microbial communities in hydrogen-rich vent fluids from hydrothermal systems along the Mid-Cayman Rise. *Environ. Microbiol.* 18, 1970–1987. doi: 10.1111/1462-2920.13173

Roux, S., Enault, F., Hurwitz, B. L., and Sullivan, M. B. (2015). VirSorter: mining viral signal from microbial genomic data. *PeerJ* 3:e985. doi: 10.7717/peerj.985

Sakai, S., Takaki, Y., Miyazaki, M., Ogawara, M., Yanagawa, K., Miyazaki, J., et al. (2019). *Methanofervidicoccus abyssi* gen. Nov., sp. nov., a hydrogenotrophic methanogen, isolated from a hydrothermal vent chimney in the mid-cayman spreading center, the caribbean sea. *Int. J. Syst. Evol. Microbiol.* 69, 1225–1230. doi: 10.1099/ijsem.0.003297

Schrenk, M. O., Kelley, D. S., Delaney, J. R., and Baross, J. A. (2003). Incidence and diversity of microorganisms within the walls of an active deep-sea sulfide chimney. *Appl. Environ. Microbiol.* 69, 3580–3592. doi: 10.1128/AEM.69.6.3580-3592.2003

Shapiro, B. J. (2016). How clonal are bacteria over time? *Curr. Opin. Microbiol.* 31, 116–123. doi: 10.1016/j.mib.2016.03.013

Sinha, R., Stanley, G., Gulati, G. S., Ezran, C., Travaglini, K. J., Wei, E., et al. (2017). Index Switching Causes "Spreading-Of-Signal" Among Multiplexed Samples In Illumina HiSeq 4000 DNA Sequencing. *bioRxiv* [Preprint]. doi: 10.1101/125724

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Stepanauskas, R., Fergusson, E. A., Brown, J., Poulton, N. J., Tupper, B., Labonté, J. M., et al. (2017). Improved genome recovery and integrated cell-size analyses of individual uncultured microbial cells and viral particles. *Nat. Commun.* 8:840. doi: 10.1038/s41467-017-00128-z

Stewart, L. C., Algar, C. K., Fortunato, C. S., Larson, B. I., Vallino, J. J., Huber, J. A., et al. (2019). Fluid geochemistry, local hydrology, and metabolic activity define methanogen community size and composition in deep-sea hydrothermal vents. *ISME J.* 13, 1711–1721. doi: 10.1038/s41396-019-0382-3

Takai, K., Gamo, T., Tsunogai, U., Nakayama, N., Hirayama, H., Nealson, K. H., et al. (2004). Geochemical and microbiological evidence for a hydrogen-based, hyperthermophilic subsurface lithoautotrophic microbial ecosystem (HyperSLiME) beneath an active deep-sea hydrothermal field. *Extremophiles* 8, 269–282. doi: 10.1007/s00792-004-0386-3

Takai, K., Nakamura, K., Toki, T., Tsunogai, U., Miyazaki, M., Miyazaki, J., et al. (2008). Cell proliferation at 122°C and isotopically heavy CH4 production by a hyperthermophilic methanogen under high-pressure cultivation. *Proc. Natl. Acad. Sci. U.S.A.* 105, 10949–10954. doi: 10.1073/pnas.0712334105

Tettelin, H., Masignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., et al. (2005). Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome". *Proc. Natl. Acad. Sci. U.S.A.* 102, 13950–13955. doi: 10.1073/pnas.0506758102

Thauer, R. K., Kaster, A. K., Seedorf, H., Buckel, W., and Hedderich, R. (2008). Methanogenic archaea: ecologically relevant differences in energy conservation. *Nat. Rev. Microbiol.* 6, 579–591. doi: 10.1038/nrmicro1931

Thiroux, S., Dupont, S., Nesbø, C. L., Bienvenu, N., Krupovic, M., L'Haridon, S., et al. (2021). The first head-tailed virus, *MFTV1*, infecting hyperthermophilic

methanogenic deep-sea archaea. *Environ. Microbiol.* 23, 3614–3626. doi: 10.1111/1462-2920.15271

Ver Eecke, H. C., Akerman, N. H., Huber, J. A., Butterfield, D. A., and Holden, J. F. (2013). Growth kinetics and energetics of a deep-sea hyperthermophilic methanogen under varying environmental conditions. *Environ. Microbiol. Rep.* 5, 665–671. doi: 10.1111/1758-2229.12065

Wang, L., Lai, Q., Jebbar, M., Shao, Z., and Alain, K. (2020). Complete genome sequence of *Methanofervidicoccus* sp. A16, a thermophilic methanogen isolated from Mid Cayman Rise hydrothermal vent. *Mar. Genomics* 53:100768. doi: 10.1016/j.margen.2020.100768

Waskom, M., Botvinnik, O., Hobson, P., Cole, J. B., Halchenko, Y., Hoyer, S., et al. (2014). *seaborn: v0.5.0 (November 2014)*. Available online at: https://zenodo.org/record/12710#.YQmI5L0zbIU (accessed April, 2020).

White, J. R., Escobar-Paramo, P., Mongodin, E. F., Nelson, K. E., and DiRuggiero, J. (2008). Extensive genome rearrangements and multiple horizontal gene transfers in a population of *Pyrococcus* isolates from Vulcano Island, Italy. *Appl. Environ. Microbiol.* 74, 6447–6451. doi: 10.1128/AEM.01024-08

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. doi: 10.1093/molbev/msm088

Conflict of Interest: MH was employed by company Finch Therapeutics Group.

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

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Hoffert, Anderson, Reveillaud, Murphy, Stepanauskas and Huber. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.