



# Single-Cell Genomics of Novel Actinobacteria With the Wood–Ljungdahl Pathway Discovered in a Serpentinizing System

Nancy Merino<sup>1,2,3\*</sup>, Mikihiko Kawai<sup>4,5</sup>, Eric S. Boyd<sup>6</sup>, Daniel R. Colman<sup>6</sup>, Shawn E. McGlynn<sup>1,7,8</sup>, Kenneth H. Nealson<sup>2</sup>, Ken Kurokawa<sup>1,9</sup> and Yuichi Hongoh<sup>1,4\*</sup>

<sup>1</sup> Earth-Life Science Institute, Tokyo Institute of Technology, Tokyo, Japan, <sup>2</sup> Department of Earth Sciences, University of Southern California, Los Angeles, CA, United States, <sup>3</sup> Biosciences and Biotechnology Division, Lawrence Livermore National Laboratory, Livermore, CA, United States, <sup>4</sup> School of Life Sciences and Technology, Tokyo Institute of Technology, Tokyo, Japan, <sup>5</sup> Graduate School of Human and Environmental Studies, Kyoto University, Kyoto, Japan, <sup>6</sup> Department of Microbiology and Immunology, Montana State University, Bozeman, MT, United States, <sup>7</sup> Biofunctional Catalyst Research Team, RIKEN Center for Sustainable Resource Science, Saitama, Japan, <sup>8</sup> Blue Marble Space Institute of Science, Seattle, WA, United States, <sup>9</sup> Department of Informatics, National Institute of Genetics, Shizuoka, Japan

## OPEN ACCESS

### Edited by:

Mirko Basen,  
University of Rostock, Germany

### Reviewed by:

Daan R. Speth,  
California Institute of Technology,  
United States  
Donald A. Bryant,  
The Pennsylvania State University,  
United States

### \*Correspondence:

Nancy Merino  
nmerino@elsi.jp  
Yuichi Hongoh  
yhongo@bio.titech.ac.jp

### Specialty section:

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

Received: 16 December 2019

Accepted: 27 April 2020

Published: 09 June 2020

### Citation:

Merino N, Kawai M, Boyd ES, Colman DR, McGlynn SE, Nealson KH, Kurokawa K and Hongoh Y (2020) Single-Cell Genomics of Novel Actinobacteria With the Wood–Ljungdahl Pathway Discovered in a Serpentinizing System. *Front. Microbiol.* 11:1031. doi: 10.3389/fmicb.2020.01031

Serpentinite-hosted systems represent modern-day analogs of early Earth environments. In these systems, water-rock interactions generate highly alkaline and reducing fluids that can contain hydrogen, methane, and low-molecular-weight hydrocarbons-potent reductants capable of fueling microbial metabolism. In this study, we investigated the microbiota of Hakuba Happo hot springs (~50°C; pH~10.5–11), located in Nagano (Japan), which are impacted by the serpentinization process. Analysis of the 16S rRNA gene amplicon sequences revealed that the bacterial community comprises *Nitrospirae* (47%), “Parcubacteria” (19%), *Deinococcus-Thermus* (16%), and *Actinobacteria* (9%), among others. Notably, only 57 amplicon sequence variants (ASV) were detected, and fifteen of these accounted for 90% of the amplicons. Among the abundant ASVs, an early-branching, uncultivated actinobacterial clade identified as RBG-16-55-12 in the SILVA database was detected. Ten single-cell genomes (average pairwise nucleotide identity: 0.98–1.00; estimated completeness: 33–93%; estimated genome size: ~2.3 Mb) that affiliated with this clade were obtained. Taxonomic classification using single copy genes indicates that the genomes belong to the actinobacterial class-level clade UBA1414 in the Genome Taxonomy Database. Based on metabolic pathway predictions, these actinobacteria are anaerobes, capable of glycolysis, dissimilatory nitrate reduction and CO<sub>2</sub> fixation via the Wood–Ljungdahl (WL) pathway. Several other genomes within UBA1414 and two related class-level clades also encode the WL pathway, which has not yet been reported for the *Actinobacteria* phylum. For the Hakuba actinobacterium, the energy metabolism related to the WL pathway is likely supported by a combination of the Rnf complex, group 3b and 3d [NiFe]-hydrogenases, [FeFe]-hydrogenases, and V-type (H<sup>+</sup>/Na<sup>+</sup> pump) ATPase. The genomes also harbor a form IV ribulose 1,5-bisphosphate carboxylase/oxygenase

(RubisCO) complex, also known as a RubisCO-like protein, and contain signatures of interactions with viruses, including clustered regularly interspaced short palindromic repeat (CRISPR) regions and several phage integrases. This is the first report and detailed genome analysis of a bacterium within the *Actinobacteria* phylum capable of utilizing the WL pathway. The Hakuba actinobacterium is a member of the clade UBA1414/RBG-16-55-12, formerly within the group “OPB41.” We propose to name this bacterium ‘*Candidatus* Hakubanella thermoalkaliphilus.’

**Keywords:** serpentinization, single-cell genomics, *Actinobacteria*, subsurface, alkaliphile, hydrogenase

## INTRODUCTION

The serpentinization reaction is fundamental to one of the leading hypotheses regarding the emergence of life on Earth, known as the submarine alkaline hydrothermal vent model (Russell et al., 2010; Branscomb and Russell, 2018). It follows that contemporary serpentinite-hosted systems might provide a window into early life. This model is based on the formation of highly reduced products (e.g., H<sub>2</sub>, CH<sub>4</sub>, and formate) from the hydration of ferromagnesian minerals in mafic and ultramafic rocks (e.g., olivine), which are subsequently mixed with solutes in comparatively more oxidized early Earth ocean waters. The resulting geochemical disequilibria could have been an energy source for the formation of early life. Importantly, this combination of alkaline pH and elevated H<sub>2</sub> concentrations of systems undergoing active serpentinization has been suggested to help overcome key biochemical bottlenecks in autotrophic metabolism, including that of acetogens and methanogens (Boyd et al., 2020), two groups of organisms commonly argued to be among the earliest evolving (Martin and Russell, 2006).

The modern-day analog of this system includes terrestrial serpentinite-hosted ecosystems, or ophiolites, created by the obduction of the oceanic lithosphere thrust onto the continental plate (Nicolas, 2012). Ophiolites are markers for the early oceanic crust, with ages ranging from 2 to 0.6 Ga (Condie, 2016). Moreover, ophiolitic terranes can be several kilometers thick (Condie, 2016), providing access to subsurface life that can persist in these reducing and alkaline (pH > 10) environments. Several studies have examined the microbial communities present in serpentinite-influenced environments, including in the Samail ophiolite (Rempfert et al., 2017; Fones et al., 2019), the Cedars (Suzuki et al., 2013), the Cabeço de Vide Aquifer (Tiago and Veríssimo, 2013), the Coast Range Ophiolite Microbial Observatory (Crespo-Medina et al., 2014; Twing et al., 2017), the Voltri Massif (Quéméneur et al., 2015; Brazelton et al., 2017), and the Zambales ophiolite (Meyer-Dombard et al., 2018). Although these can be distant locations from each other, Meyer-Dombard et al. (2018) identified a ‘principal community’ amongst serpentinizing environments, consisting of key members in the phyla *Firmicutes* (e.g., *Dethiobacter* sp.) and *Proteobacteria* (e.g., *Serpentinomonas* sp.).

The microbial communities of the Hakuba Happo hot spring (36°42′N, 137°48′E) ophiolite located along the Itoigawa–Shizuoka Tectonic Line in central Honshu, Japan have yet to be investigated. This region consists of an ultramafic rock

body that is ~580 Ma old (Sato et al., 2019) and has ongoing serpentinization activity (Suda et al., 2014, 2017). The geochemistry of the site is characteristic of a serpentinite-hosted system, with highly alkaline waters (pH > 10.6) and high concentrations of dissolved H<sub>2</sub> (201–664 μM) and CH<sub>4</sub> (124–201 μM) (Suda et al., 2014). The source of H<sub>2</sub> is likely derived from ‘low’ temperature serpentinization reactions occurring at ~50°C (Mayhew et al., 2013) while CH<sub>4</sub> could be from abiotic or biotic origins (Suda et al., 2014). Two wells (well #1 and #3) have been drilled into the Hakuba Happo ophiolite that permit acquisition of subsurface fluids for geochemical and microbiological analyses.

In the present study, we obtained single-cell genomes of an early-branching, uncultivated actinobacterial lineage from Hakuba Happo well #3 (abbreviated hereafter Happo #3), which were among the dominant taxa found in the bacterial community based on 16S rRNA gene amplicon sequences. This actinobacterial lineage was previously designated as RBG-16-55-12 in the SILVA database (Quast et al., 2012; Yilmaz et al., 2014) and approximately corresponds to the UBA1414/RBG-13-55-18/UBA9087 clade in the Genome Taxonomy Database (GTDB) (Parks et al., 2018). Herein, we predict the metabolic properties and provide the first detailed genome analysis of a bacterium in the clade UBA1414/RBG-13-55-18/UBA9087. This comes two decades after the discovery of its presence by 16S rRNA gene sequencing analysis from samples collected at Obsidian Pool in Yellowstone National Park where it acquired the name “OPB41” (Hugenholtz et al., 1998).

## MATERIALS AND METHODS

### Sample Collection and Geochemical Measurements

Samples were collected from Happo #3 (36°42′48.6″N 137°48′26.3″E) in October 2016. Detailed geochemical analysis of Happo #3 was previously described in Suda et al. (2014), including isotope compositions and ion concentrations. Happo #3 is a drilling well that extends to about 700 m depth and water is pumped to the surface for the hot spring facilities provided in Happo Town, Japan (Suda et al., 2014). In the field, water temperature (water resistant thermometer CT-430WP, CUSTOM, Japan), pH (pH meter model D-51 with electrode 9625-10D and B-712, HORIBA, Japan), oxidation-reduction potential (ORP; ORP meter model RM-30P with electrode

PST-2739C, TOA-DKK, Japan), dissolved oxygen (DO; DO meter model DO-31P with electrode OE-270AA, Japan), electrical conductivity (EC; EC meter model CM-31P with electrode CT-27112B, Japan), salinity (B-721 meter, HORIBA, Japan), calcium (B-751 meter, HORIBA, Japan), sodium (B-722 meter, HORIBA, Japan), and potassium (B-731 meter, HORIBA, Japan) ion concentrations were measured. Analysis of ions and organic acids are described in the **Supplementary Information**.

Happo #3 water was filtered using two different methods (“Total” and “Sequential”) at a flow rate of about 15 mL per min for 22 h (total water filtered ~ 19.8 L). For the “Total” method, a 0.1  $\mu\text{m}$  Omnipore membrane (25 mm diameter, Millipore, United States) was used, while the “Sequential” method used in-series filtration consisting of a 0.22  $\mu\text{m}$  Sterivex-GP (polyethersulfone, Millipore, United States), followed by a 0.1  $\mu\text{m}$  Omnipore membrane. The Omnipore membrane was housed in a PerFluoroAlkoxy filter holder (Advantec, United States). Filtered samples were aseptically placed in 100  $\mu\text{L}$  of fresh glycerol-Tris-EDTA buffer (Rinke et al., 2014) for single-cell genomics. Glycerol-Tris-EDTA consisted of 20 mL TE buffer (100 $\times$ , pH 8) and 100 mL glycerol per 180 mL, which was sterilized by passing through a 0.1  $\mu\text{m}$  filter. Samples were immediately shipped at  $-20^{\circ}\text{C}$  overnight and stored at  $-80^{\circ}\text{C}$ .

## 16S rRNA Gene Amplicon Sequencing

For 16S rRNA gene amplicon sequencing, another set of filters was collected as described above. DNA was extracted from the “Total” and “Sequential” samples, using the ZymoBIOMICS DNA/RNA Miniprep Kit (Zymo Research, United States). The V3–V4 region of the 16S rRNA genes was amplified by PCR with primers 341F (5'-CCTACGGGNGGCWGCAG) and 785R (5'-GACTACHVGGGTATCTAATCC) according to the Illumina MiSeq Protocol “16S Metagenomic Sequencing Library Preparation,” and the amplicons were used for preparation of sequencing libraries with the KOD FX Neo Kit (Toyobo Life Science, Japan). Sequencing was performed using the Illumina MiSeq platform with the V3 reagent kit (600 cycles). A total of 17,058 (“Total”) and 10,390 (“Sequential”) reads were obtained after quality filtering and trimming via DADA2 (Callahan et al., 2016). The reads were sorted to amplicon sequence variants (ASV), or unique sequences, using DADA2 and taxonomically identified (Callahan et al., 2016, 2017). Afterward, phyloseq v1.26.1 (McMurdie and Holmes, 2013) was used to prune the samples of ASVs observed in a negative control of filtered air collected during field sampling. For the remaining ASVs, a prevalence threshold of 0.1 was determined by phyloseq.

## Single-Cell Sorting, Whole Genome Amplification, and Library Preparation

A fluorescence-activated cell sorter (FACS; BD FACS Aria IIU, BD Biosciences, United States) with a 70  $\mu\text{m}$  nozzle orifice was used to sort single cells into 96-well plates. Filters stored in glycerol-Tris-EDTA stock were thawed on ice and briefly shaken to re-suspend cells from the filter, and 0.65  $\mu\text{L}$  of 1 g/L FM<sup>TM</sup> 1-43FX (Thermo Fisher Scientific, United States)

was then added to an aliquot (350  $\mu\text{L}$ ) to stain the cell membrane. The sample was incubated for at least 15 min on ice and was not pre-screened through a 70  $\mu\text{m}$  mesh-size cell strainer (BD Biosciences, United States) to prevent the loss of microbial cells since the Happo #3 water did not contain large particles or microorganisms > 70  $\mu\text{m}$ . The FACS sorting operating condition was checked by calibrating against the BD CS&T Beads (BD Biosciences, United States). A total of 5 plates were sorted for “Total” filters and 8 plates for “Sequential” filters. FACS parameters are further described in **Supplementary Information**. Targeted cells were sorted into 96-well plates with 2 wells reserved for whole genome amplification (WGA) positive controls (with added template DNA) and 8 wells were reserved for the negative control (without droplet deposition). Each plate was immediately placed at  $-80^{\circ}\text{C}$  until processed. Several single-cell lysis methods were tested and described in the **Supplementary Information**. For WGA, the Qiagen REPLI-g Single Cell Kit (Qiagen, Germany) was used with a modified protocol, as described in the **Supplementary Information**.

The WGA products were diluted (5  $\mu\text{L}$  WGA product, 95  $\mu\text{L}$  UV-sterilized H<sub>2</sub>O), mixed by pipetting 15 times, and 1  $\mu\text{L}$  was used in a qPCR reaction (SsoAdvanced<sup>TM</sup> Universal SYBR<sup>®</sup> Green Supermix, Bio-Rad Laboratories, United States) to amplify the 16S rRNA gene V6–V8 hypervariable regions with primers 926wF and 1392R (Rinke et al., 2014). The qPCR reaction contained 5  $\mu\text{L}$  SsoAdvanced<sup>TM</sup> Supermix, 0.2  $\mu\text{L}$  forward primer (10  $\mu\text{M}$  stock), 0.2  $\mu\text{L}$  reverse primer (10  $\mu\text{M}$  stock), 3.6  $\mu\text{L}$  UV-sterilized H<sub>2</sub>O, and 1  $\mu\text{L}$  of the diluted WGA product. The qPCR reaction cycle comprised 98 $^{\circ}\text{C}$  for 3 min, 35 cycles of 98 $^{\circ}\text{C}$  for 15 s and 60 $^{\circ}\text{C}$  for 1 min, a melt curve of 95 $^{\circ}\text{C}$  for 15 s, 60 $^{\circ}\text{C}$  for 1 min, with ramp of +0.3 $^{\circ}\text{C}$  to 95 $^{\circ}\text{C}$  for 15 s, followed by a 4 $^{\circ}\text{C}$  hold. Amplification of 16S rRNA genes was confirmed by gel electrophoresis, and 5  $\mu\text{L}$  of qPCR products were treated with 2  $\mu\text{L}$  ExoSAP-IT Express (ThermoFisher Scientific, United States). The cleaned qPCR products were then sent for Sanger sequencing with primer 1392R to enable cell selection for sequence library preparation. **Supplementary Table S1** describes the cells selected for sequencing, including FACS conditions, lysis and WGA reaction conditions, and single-cell genome statistics referenced against the minimum information of single amplified genome (MISAG) criteria (Bowers et al., 2017). Libraries were prepared using the TruSeq DNA PCR-Free Library Preparation Kit (Illumina, United States) and a Covaris M220 to obtain 550 bp sheared DNA.

## Sequencing, Assembly, Binning, and Annotation

All single-cell amplified genome (SAG) libraries were sequenced on the Illumina MiSeq platform using 2  $\times$  300 bp paired-end sequencing (MiSeq v3 Reagent Kit). Raw reads were evaluated using FastQC v0.11.5<sup>1</sup> and trimmed and quality filtered by Trim\_galore! v0.4.1<sup>2</sup>, which uses the

<sup>1</sup><https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

<sup>2</sup>[https://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/)



cutadapt v1.9.1 program (Martin, 2011). Trim\_galore! parameters were set for paired-end files and included a stringency of 5, e 0.1 (error rate), q 20,20 (quality), with the option to retain unpaired reads. Reads were then assembled with SPAdes v3.10.1 (Bankevich et al., 2012) for single-cell samples with the “careful” option and default parameters (k-mers: 21, 33, and 55). Scaffold names were simplified for the Anvi’o v5.3 workflow (Eren et al., 2015), followed by read-mapping with Bowtie2 v2.3.2 (Langmead and Salzberg, 2012) (parameters very-sensitive-local and dovetail) with the samtools depth function to determine coverage values by searching the trimmed reads against the assembled scaffolds. The Anvi’o workflow was then used to cluster and profile the scaffolds greater than 1,000 bp and potential contaminants were removed. The ACDC program (Lux et al., 2016) was also used for contamination screening. Subsequently, gene identification was conducted using Prodigal v2.6.2 (Hyatt et al., 2010) and HMMER v3.1b2<sup>3</sup>. Functional classification was conducted using InterProScan v5.28-67.0 (with databases: TIGRFAMs, SFLD, HAMAP, ProSiteProfiles, ProSitePatterns, PANTHER, Pfam, CDD) (Jones et al., 2014) and imported into Anvi’o. Secondary metabolite biosynthetic gene clusters were identified using antiSMASH v4.1.0 with the options –clusterblast –subclusterblast –knownclusterblast –smcogs –inclusive –borderpredict –full-hmmer –asf –tta (Blin et al., 2017). MAPLE was used to obtain KEGG orthologous (KO) group assignments (Takami, 2014; Arai et al., 2018). Gas vesicle genes were annotated by using a manually curated gas vesicle hidden Markov model database, which is described in the **Supplementary Information**.

Taxonomic classification was conducted with Kaiju v1.5.0 (Menzel et al., 2016) against the NCBI non-redundant (nr) database (nr + euk database) and imported into Anvi’o. Prophage regions were detected on contigs > 2,000 bp in PHASTER (Zhou et al., 2011; Arndt et al., 2016), and clustered regularly interspaced short palindromic repeat (CRISPR) and its associated gene (Cas) regions were annotated using CRISPRCasFinder (Couvin et al., 2018). SAG sequences were manually refined through Anvi’o (anvi-interactive). CheckM v1.0.7 (Parks et al., 2015) was also used to estimate completeness, degree of contamination, and strain heterogeneity. The number of rRNA genes was determined by the Anvi’o v5.3 method (Eren et al., 2015) and Barrnap v0.6<sup>4</sup>.

## Co-assembly of SAGs

Ten SAGs (**Supplementary Table S1**) were subsequently co-assembled using SPAdes v3.10.1 with k-mers that were normalized to achieve a flat coverage distribution (target normalization depth = 100 for k-mers with at least 5 depth coverage) via BBNorm v37.95<sup>5</sup> using default parameters. A range of k-mers were tested (21, 33, 55, 77, 99, and 127) and scaffolds produced when using k-mers 21 and 33 achieved the highest N50 of 7,442 bp based on Quast v4.5 (Gurevich et al., 2013).

The generated scaffolds were subsequently placed into the Anvi’o v5.3 workflow with Bowtie2 v2.3.2 read-mapping, as described above, and after removal of contigs < 1,000 bp and potential contaminants (based on sequence composition clustering), the N50 was 8,580 bp. Functional and taxonomic classification were also conducted as described above. The number of tRNA and rRNA genes were determined using tRNAscan-SE v2.0 (Lowe and Eddy, 1997) and Anvi’o v5.3 or Barrnap v0.6, respectively. Effective DB (Eichinger et al., 2016) was used to predict the fully functional bacterial secretion systems Type III, IV, and VI.

## Phylogenetic and Comparative Genomic Analyses

The co-assembly was then placed into phylogenetic trees with reference genomes from the NCBI RefSeq and GenBank databases (O’Leary et al., 2016). The trees included *Actinobacteria* genomes from Rifle, CO (United States) (Anantharaman et al., 2016), a CO<sub>2</sub>-driven geyser (Colorado Plateau, Utah, United States) (Probst et al., 2018), the Sanford Underground Research Facility (SURF) (Momper et al., 2017), and Baltic Sea sediments (Bird et al., 2019; cleaned assemblies provided by Dr. Karen Lloyd). These genomes were the most closely related to the Hakuba SAGs, as determined by classification using the Genome Taxonomy Database Toolkit v0.2.2 (GTDB-Tk), which is a database of quality-controlled genomes that aims to standardize microbial taxonomy through genome phylogeny (Parks et al., 2018). Pyani v0.2.8<sup>6</sup> and the enveomics collection toolbox were used to calculate the pairwise average nucleotide identity (ANI) and the pairwise average amino acid identity (AAI) between the genomes, respectively (Konstantinidis and Tiedje, 2005a,b; Rodriguez-R and Konstantinidis, 2014, 2016). The occurrence of split genes was analyzed as described in **Supplementary Information**. Two phylogenetic reconstructions were conducted to evaluate the phylogenetic placement of the Hakuba *Actinobacteria* genome:

(1) A maximum likelihood (ML) tree was created using the GToTree v1.1.6 (Lee, 2019) pipeline based on 138 *Actinobacteria*-specific single copy genes<sup>7</sup>. Reference genomes from *Actinobacteria* were used, and the outgroups consisted of several genomes from each family of *Firmicutes* and *Proteobacteria* (**Supplementary Table S2**). The concatenated multiple sequence alignment of deduced amino acids was then uploaded to the CIPRES Science Gateway (Miller et al., 2010) to create a ML tree using RAXML-HPC2 on XSEDE (Stamatakis et al., 2008; Stamatakis, 2014) with options WAG PROTGAMMA model and autoMRE bootstrapping.

(2) A Bayesian phylogenetic reconstruction was conducted in Beast2 v2.5.2 (Bouckaert and Vaughan, 2019) with a subset of reference genomes used for the ML tree reconstruction (**Supplementary Table S3**). After generating a multiple sequence alignment using GToTree, a Bayesian tree was constructed using the WAG substitution model that assumed a gamma distribution with 4 categories and a relaxed clock log normal distribution with

<sup>3</sup><http://hmmer.org>

<sup>4</sup><https://github.com/tseemann/barrnap/blob/master/README.md>

<sup>5</sup><https://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-guide/bbnorm-guide/>

<sup>6</sup><https://github.com/widdowquinn/pyani>

<sup>7</sup>[https://github.com/AstroBioMike/GToTree/tree/master/hmm\\_sets](https://github.com/AstroBioMike/GToTree/tree/master/hmm_sets)

Markov chain Monte Carlo simulations (Drummond et al., 2002) set to 50,000,000 (logging every 5,000). This substitution model was selected with PartitionFinder v2.1.1 (Lanfear et al., 2016). A burn-in of 70 percent was set to combine two converging trees of Beast2, as viewed using Tracer v1.7.1 (Rambaut et al., 2018), resulting in 13,501 samples and an effective sample size of 1,278 for tree likelihood and 529 for posterior.

The phylogenetic placement of the Hakuba *Actinobacteria* co-assembled genome amongst all the genomes available in the NCBI RefSeq and Genbank database was confirmed using GTDB-Tk v0.2.2 (reference database version r86 v3). Taxonomic classification was confirmed with the classify workflow (classify\_wf), which utilizes the third-party dependencies pplacer (Matsen et al., 2010), FastANI (Jain et al., 2018), Prodigal (Hyatt et al., 2010), FastTree (Price et al., 2010), and HMMR (Eddy, 2011). The classify workflow will first identify bacterial and archaeal marker genes, followed by creating and concatenating multiple sequence alignments. After filtering the alignment to 5,000 amino acids, the workflow will then classify each genome using the GTDB-Tk reference tree and determine the relative evolutionary divergence and ANI.

Selected Protein Sequences [Carbon Monoxide Dehydrogenase / Acetyl-CoA Synthase (CODH/ACS), formylmethanofuran dehydrogenase (Fwd), V-type ATPase, adenosine-5'-phosphosulfate (APS), 3'-phosphoadenosine 5'-phosphosulfate (PAPS) reductase, nitrate reductase alpha subunit NarG, and RubisCO] were aligned using MAFFT (Katoh et al., 2002) with options `-maxiterate 1000` and default parameters. For the proteins CODH/ACS, APS, PAPS, NarG, and RubisCO, phylogenetic trees were created. Briefly, gaps were removed with trimAl v1.4.rev15 (Capella-Gutierrez et al., 2009) using option `automated1`. Manual curation was also done before creating a ML tree using either FastTree v2 (Price et al., 2010) or RAxML on the CIPRES Science Gateway (Miller et al., 2010). All phylogenetic trees were checked using Archaeopteryx (Han and Zmasek, 2009) and FigTree<sup>8</sup>.

Genes coding for [NiFe]- and [FeFe]-hydrogenases were identified by comparison against a curated in-house database (E.S. Boyd, unpublished data). Resulting catalytic subunits were checked for characteristic N- and C-terminal cysteine motifs associated with [NiFe]-hydrogenase variants and the L1, L2, and L3 motifs for [FeFe]-hydrogenase variants (Vignais and Billoud, 2007). The large catalytic subunits of the [NiFe]-hydrogenases were subjected to phylogenetic analysis, as described above, but using the IQtree ML algorithm with the LG+G amino acid substitution model and 1,000 bootstraps to evaluate node support. The phylogenetic analysis included representatives of the primary [NiFe]-hydrogenase groups (Greening et al., 2016) in addition to close representatives of the query sequences that were present in the NCBI database. Gene neighborhood analysis was conducted by surveying the co-assembly and individual SAGs for representatives associated with either Group 3 [NiFe]-hydrogenases (Peters et al., 2015) or those associated with [FeFe]-hydrogenases (Poudel et al., 2016).

<sup>8</sup><http://tree.bio.ed.ac.uk/software/figtree/>

## RESULTS AND DISCUSSION

### Bacterial Community Structure of the Hakuba Happo #3 Well

The geochemistry of Happo #3 waters from 2011 to 2016 is summarized in Table 1. The taxonomic composition of the bacterial community based on 16S rRNA gene amplicon sequencing is depicted in Supplementary Figure S1 and summarized in Supplementary Table S4. The dominant bacterial phyla were *Nitrospirae* (47%), "Parcubacteria" (19%), *Deinococcus-Thermus* (16%), and *Actinobacteria* (9%), followed by *Firmicutes* (5%), *Bacteroidetes* (2%), among others (<1%). Only 57 ASVs were detected from both the "Total" (17,058 total reads) and "Sequential" (10,390 total reads) samples, and the majority (90%) were represented by 15 ASVs. Such low bacterial diversity is consistent among serpentinite-hosted systems. For example, at The Cedars, 16 phylotypes (>99% sequence similarity cutoff) represented 84% of the 16S rRNA amplicon sequences recovered from the shallow-sourced spring

**TABLE 1** | Geochemistry of Happo #3 from 2011 to 2016.

	2011 <sup>a</sup>	2016
Temperature	48°C	47.5°C <sup>b</sup>
pH	10.7	10.95 <sup>b</sup>
ORP	n.m.	-435 mV <sup>b</sup>
EC	48.3 mS/m	47.7 mS/m <sup>b</sup>
DO	0.59 mg/L	0.10 mg/L <sup>b</sup>
Salinity	0.02%	n.d. <sup>b</sup> (detection limit 0.1%)
H <sub>2</sub>	201 μM	n.m.
CH <sub>4</sub>	124 μM	n.m.
C <sub>2</sub> H <sub>6</sub>	0.2 μM	n.m.
N <sub>2</sub>	1298 μM	n.m.
Ca <sup>2+</sup>	130 μM	220 μM
K <sup>+</sup>	70 μM	125 μM
Na <sup>+</sup>	1160 μM	1565 μM
NH <sub>3</sub>	n.m.	7 μM
Al <sub>3</sub> <sup>+</sup>	10 μM	n.m.
Li <sup>+</sup>	10 μM	n.m.
Cl <sup>-</sup>	180 μM	155 μM
SO <sub>4</sub> <sup>2-</sup>	10 μM	8 μM
F <sup>-</sup>	4 μM	n.m.
Formate	n.m.	80 μM
Acetate	n.m.	<40 μM
Not detected <sup>c</sup>		Pyruvate, lactate propionate, NO <sub>2</sub> <sup>-</sup> , NO <sub>3</sub> <sup>-</sup> , HCO <sub>3</sub> <sup>-</sup> , Mg <sup>2+</sup> , PO <sub>4</sub> <sup>2-</sup> , total Fe, nucleobases, nucleosides, and amino acids

<sup>a</sup>Published in Suda et al. (2014). <sup>b</sup>Nobu et al., unpublished data. <sup>c</sup>No peak detected; Detection limits: Pyruvate (40 μM), lactate (40 μM), propionate (40 μM), NO<sub>2</sub><sup>-</sup> (2.2 μM), NO<sub>3</sub><sup>-</sup> (1.6 μM), HCO<sub>3</sub><sup>-</sup> (0.02 μM), Mg<sup>2+</sup> (20.6 μM), PO<sub>4</sub><sup>2-</sup> (0.01 μM). CO<sub>2</sub> and total Fe not detected as described in Suda et al. (2014). Nucleobases and nucleosides include cytosine (179 nM), uracil (659 nM), adenine (740 nM), guanine (1.3 μM), C-ribonucleosides (47 nM), U-ribonucleosides (340 nM), A deoxyribonucleosides (71 nM), G deoxyribonucleosides (64 nM), C-deoxyribonucleosides (470 nM), T-deoxyribonucleosides (133 nM), A-deoxyribonucleosides (400 nM), and CH1 amino acids (50 μM). n.m. = not measured. n.d. = not detected.

and 98% of those from the deep-sourced spring (Suzuki et al., 2013). In the Cabeço de Vide Aquifer, 45 phylotypes ( $\geq 97\%$  similarity cutoff) were identified, dominated by four major taxonomic classes (Tiago and Veríssimo, 2013). Other serpentinite-hosted systems with comparatively few phylotypes include the Samail ophiolite (Rempfert et al., 2017; Fones et al., 2019), the Coast Range Ophiolite Microbial Observatory (Crespo-Medina et al., 2014; Twing et al., 2017), the Voltri Massif (Quéméneur et al., 2015; Brazelton et al., 2017), and the Zambales ophiolite (Meyer-Dombard et al., 2018). The microbial diversity and abundance of cells were previously shown to be pH-dependent within the Samail ophiolite (Rempfert et al., 2017; Fones et al., 2019). Compared to the ‘principal community’ amongst several serpentinite-hosted systems identified by Meyer-Dombard et al. (2018), Happo #3 contained few *Proteobacteria*, whereas *Nitrospirae* and “Parcubacteria” predominated.

The Happo #3 community included three ASVs affiliated with an early-branching, uncultivated *Actinobacteria* lineage that has not been previously observed in terrestrial serpentinite-hosted systems. These *Actinobacteria* ASVs shared 98% sequence identity and clustered with the clade RBG-16-55-12 in the SILVA v132 database (Quast et al., 2012; Yilmaz et al., 2014), previously classified within the clade OPB41 in the SILVA v128 database. The RBG-16-55-12 members are located in a variety of environments, including subsurface environments (Anantharaman et al., 2016), mine tailing ponds (Ramos-Padrón et al., 2011), mud volcanoes (Chang et al., 2012), hot springs (Hugenholtz et al., 1998), and deep sea sediments (Kato et al., 2009).

## General Characteristics and Taxonomic Classification of the Hakuba *Actinobacteria* SAGs

We conducted single-cell genomics of the Happo #3 samples and identified 10 SAGs belonging to the RBG-16-55-12 clade based on their 16S rRNA sequences. The general characteristics of these 10 SAGs are listed in **Supplementary Table S1**. “Low” ( $n = 6$ ), “Medium” ( $n = 3$ ), and “High” ( $n = 1$ ) quality SAGs were identified according to the MISAG standard for *Bacteria* and *Archaea* (Bowers et al., 2017). The range of completeness was between 33.1 and 92.8% with 0.7% and 6.5% contamination (median = 1.4% contamination), as estimated by the Anvi’o marker gene-based approach (Eren et al., 2015; **Supplementary Table S5**). Based on ANI (**Supplementary Table S6**) and AAI (**Supplementary Table S7**), these 10 SAGs represent the same species ( $\geq 98\%$  pairwise ANI for all 10 SAGs;  $\geq 90\%$  pairwise AAI for SAGs with  $> 50\%$  completeness) with GC content ranging from 48.5 to 49.2%. It has been suggested that species boundary is approximately 95% (ANI) and 90% (AAI) (Konstantinidis and Tiedje, 2005a; Richter and Rosselló-Móra, 2009).

The genomes were subsequently co-assembled into one composite genome assembly (“Hakuba co-assembly”) of all 10 SAGs combined, resulting in 93.5% completeness and 6.5% contamination (**Table 2**). The co-assembly was generated to guide genome analysis of the SAGs to

supplement the inherent biases of single-cell genomics caused during WGA (e.g., chimeric DNA, uneven genome coverage, low completeness) (Xu and Zhao, 2018). Sequence similarity analysis of the 588 co-assembled contigs using the Kaiju taxonomic classifier (Menzel et al., 2016) with the NCBI nr database did not provide confident placement of the taxonomic position of this genome. The taxonomic affiliation of the contigs was not consistent (**Supplementary Table S8**): the contigs were affiliated with “unclassified” (43%), *Firmicutes* (11%), *Proteobacteria* (9%), *Actinobacteria* (5%), *Chloroflexi* (4%), *Nitrospirae* (4%), *Euryarchaeota* (2%), “Omnitrophica” (2%), and others ( $< 1\%$ ). The contigs not taxonomically identified as *Actinobacteria* were not removed for two reasons: (1) the SAG and co-assembled genome redundancy (Anvi’o) and contamination (CheckM) were  $< 6.5\%$  (median = 1.4% contamination) (**Supplementary Table S1**) and (2) contig clustering by sequence composition on Anvi’o and ACDC did not reveal that these taxa contributed to contamination (**Supplementary Figure S2**). Similar inconsistent results of taxonomic affiliation were reported for genomes of *Bacteria* belonging to deeply branching lineages with limited reference sequences, such as members within the candidate bacterial phylum OP9 (Dodsworth et al., 2013).

Based on taxonomic analysis using GTDB-Tk (**Supplementary Table S9**), the Hakuba co-assembly and the 10 SAGs were classified to the uncultured, class-level clade “UBA1414” in the *Actinobacteria* phylum. This clade includes four Baltic Sea SAGs and one metagenome-assembled genome (MAG) from the Rifle aquifer (GCA\_001767735). GTDB-Tk was also used to estimate the novelty of the Hakuba *Actinobacteria* genomes by calculating a relative evolutionary divergence metric and comparing against the GTDB rank normalized taxonomy. This metric is more robust than pairwise AAI to assign taxonomic rank as it considers the variation in the evolutionary tempos amongst different lineages (Hugenholtz et al., 2016; Parks et al., 2018). Based on this metric, the Hakuba SAGs and co-assembly could represent a new order within the UBA1414 class while the Baltic Sea SAGs and Rifle MAG represent new species within the genus currently named “20-14-0-20-35-9” in GTDB. According to the classification of GTDB, two closely related class-level clades to UBA1414 are “UBA9087” and “RBG-13-55-18,” which consist of several MAGs from the Rifle aquifer, one MAG from Crystal Geyser, and one MAG from SURF (**Table 3**). These three clades (UBA1414, UBA9087, and RBG-13-55-18) correspond to two 16S rRNA-based clades in the SILVA v132 database: WCHB1-81 and the above-mentioned clade, RBG-16-55-12.

The relationship of the clades UBA1414, UBA9087, and RBG-13-55-18 was further examined by ML and Bayesian phylogenetic analyses of a concatenated protein sequence (**Figure 1** and **Supplementary Figure S3**). The monophyly of the three clades was confirmed by both methods, which generated identical topologies with a high confidence level for the clades. The Hakuba co-assembly formed a clade with the Baltic Sea SAGs and one Rifle MAG (GCA\_001767735), similar to the analysis using GTDB-Tk (**Supplementary Table S9**).

**TABLE 2** | Basic information for the co-assembled Hakuba genome and SAG S34.

Analysis project type	Co-assembled genome	SAG ID: S34
DDBJ BioProject	PRJDB8357	PRJDB8357
DDBJ Accession Number Co-assembly	BLSE01000001–BLSE01000587	
DDBJ Accession Number SAGs	BLRU01000000–BLRZ01000000, BLSA01000000–BLSO01000000	BLRZ01000001–BLRZ01000510
<b>Co-assembly/SAG information</b>		
Cell isolation approach	FACS	
Single cell lysis approach	Chemical and enzymatic	
Single cell kit	Qiagen REPLI-g Single Cell Kit (multiple displacement amplification)	
Assembly software	SPAdes v3.10.1	
Estimation of completeness	Anvi'o v5.3 (marker gene-based approach)	
Assembly quality <sup>a</sup>	Medium quality	High quality
Estimated completeness	93.5%	92.8%
Contamination	6.5%	1.4%
<b>Genome information</b>		
Genome size	2,947,136 bp	2,120,563 bp
Number of contigs	588	316
N50	8,580	13,695
Max contig length	37,255 bp	55,032 bp
rRNAs	2 (16S), 2 (23S), 1 (5S)	1 (16S), 1 (23S)
tRNAs	45	47
GC Content	48.5%	48.6%

The genome was co-assembled from 10 single-cell amplified genomes (SAGs) collected from Happo #3 (see **Supplementary Table S1** for individual SAGs). The SAG S34 represents the highest quality SAG obtained. <sup>a</sup>Based on the minimum information of single amplified genome (MISAG) criteria (Bowers et al., 2017).

**TABLE 3** | Information for the metagenome-assembled genomes (MAGs) and single-cell amplified genomes (SAGs) analyzed in this paper.

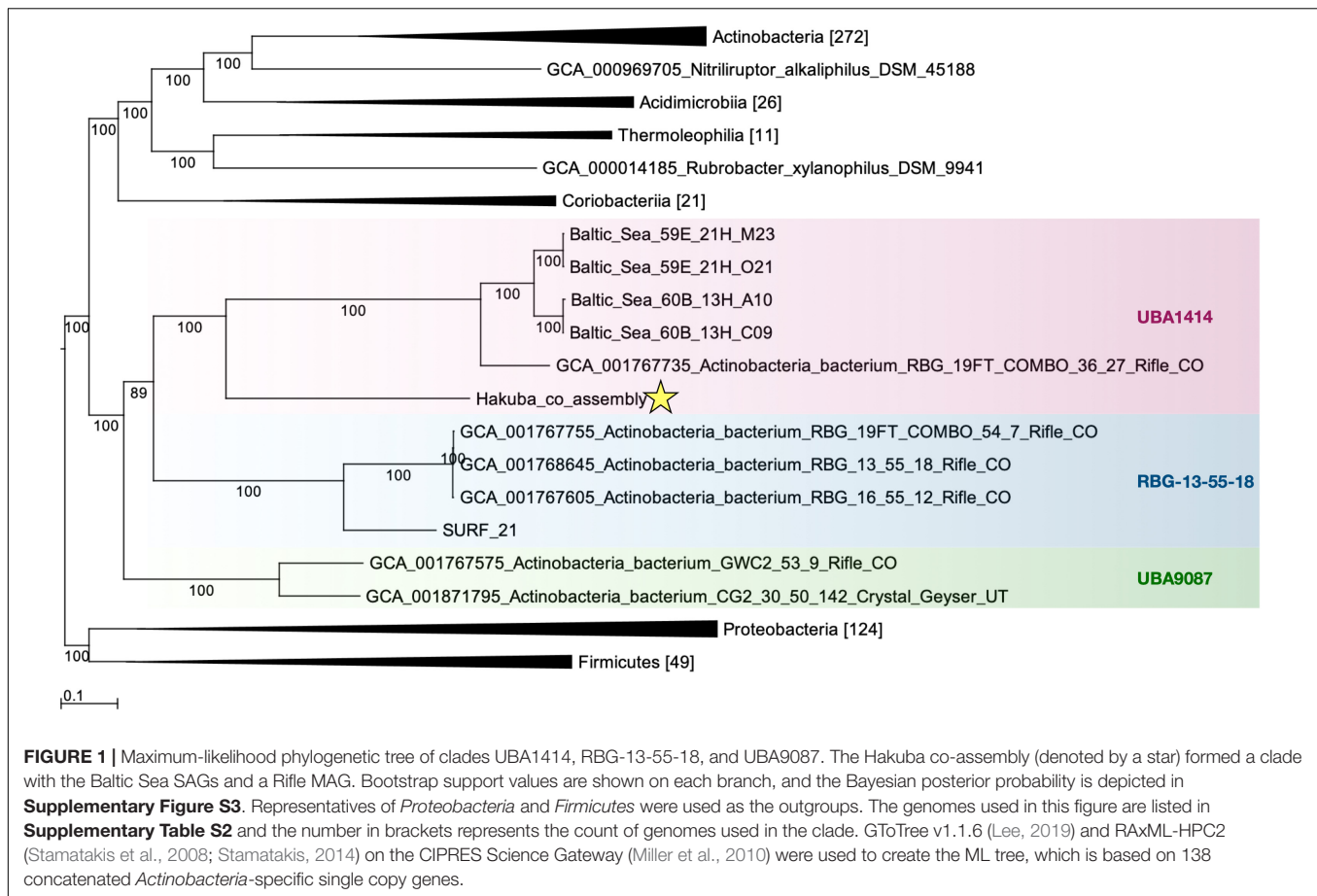
Clade	ID	NCBI BioProject (BioSample)	Completeness (%)	Contamination (%)	References
UBA1414	Baltic Sea_59E_21H_M23	PRJNA417388 <sup>a</sup>	54.7	1.4	Bird et al., 2019
	Baltic Sea_59E_21H_O21	PRJNA417388 <sup>a</sup>	52.5	1.4	Bird et al., 2019
	Baltic Sea_60B_13H_A10	PRJNA417388 <sup>a</sup>	69.8	2.9	Bird et al., 2019
	Baltic Sea_60B_13H_C09	PRJNA417388 <sup>a</sup>	71.9	1.4	Bird et al., 2019
	GCA_001767735_Rifle_CO	PRJNA288027 (SAMN04314056)	63.3	2.9	Anantharaman et al., 2016
RBG-13-55-18	Hakuba_co-assembly <sup>b</sup>	PRJDB8357	93.5	6.5	This study
	GCA_001767755_Rifle_CO	PRJNA288027 (SAMN04313673)	60.4	0.0	Anantharaman et al., 2016
	GCA_001768645_Rifle_CO	PRJNA288027 (SAMN04313996)	97.1	0.7	Anantharaman et al., 2016
	GCA_001767605_Rifle_CO	PRJNA288027 (SAMN04313722)	97.8	0.0	Anantharaman et al., 2016
UBA9087	SURF_21	PRJNA355136 <sup>c</sup>	89.2	0.7	Momper et al., 2017
	GCA_001767575_Rifle_CO	PRJNA288027 (SAMN04314195)	78.4	1.4	Anantharaman et al., 2016
	GCA_001871795_Crystal_Geyser	PRJNA297582 (SAMN04328288)	95.7	0.7	Burstein et al., 2016

Completeness and contamination were determined by Anvi'o marker gene-based approach (Eren et al., 2015). <sup>a</sup>The cleaned genome assemblies were provided by Dr. Karen Lloyd. <sup>b</sup>The highest quality Hakuba SAG S34 has 92.8% completeness and 1.4% contamination. <sup>c</sup>The cleaned genome assemblies were obtained from the data depository referenced in Momper et al. (2017).

However, no genomes showed > 45% pairwise AAI and > 76% pairwise ANI to the Hakuba co-assembly and 10 SAGs (**Supplementary Tables S6, S7**). It has been suggested that the

genus-level boundary is  $\geq 60\%$  for pairwise AAI (Rodríguez-R and Konstantinidis, 2014). This further demonstrates the novelty of the Hakuba SAGs, possibly as a new order





within the UBA1414 class, as suggested by the GTDB-Tk analysis.

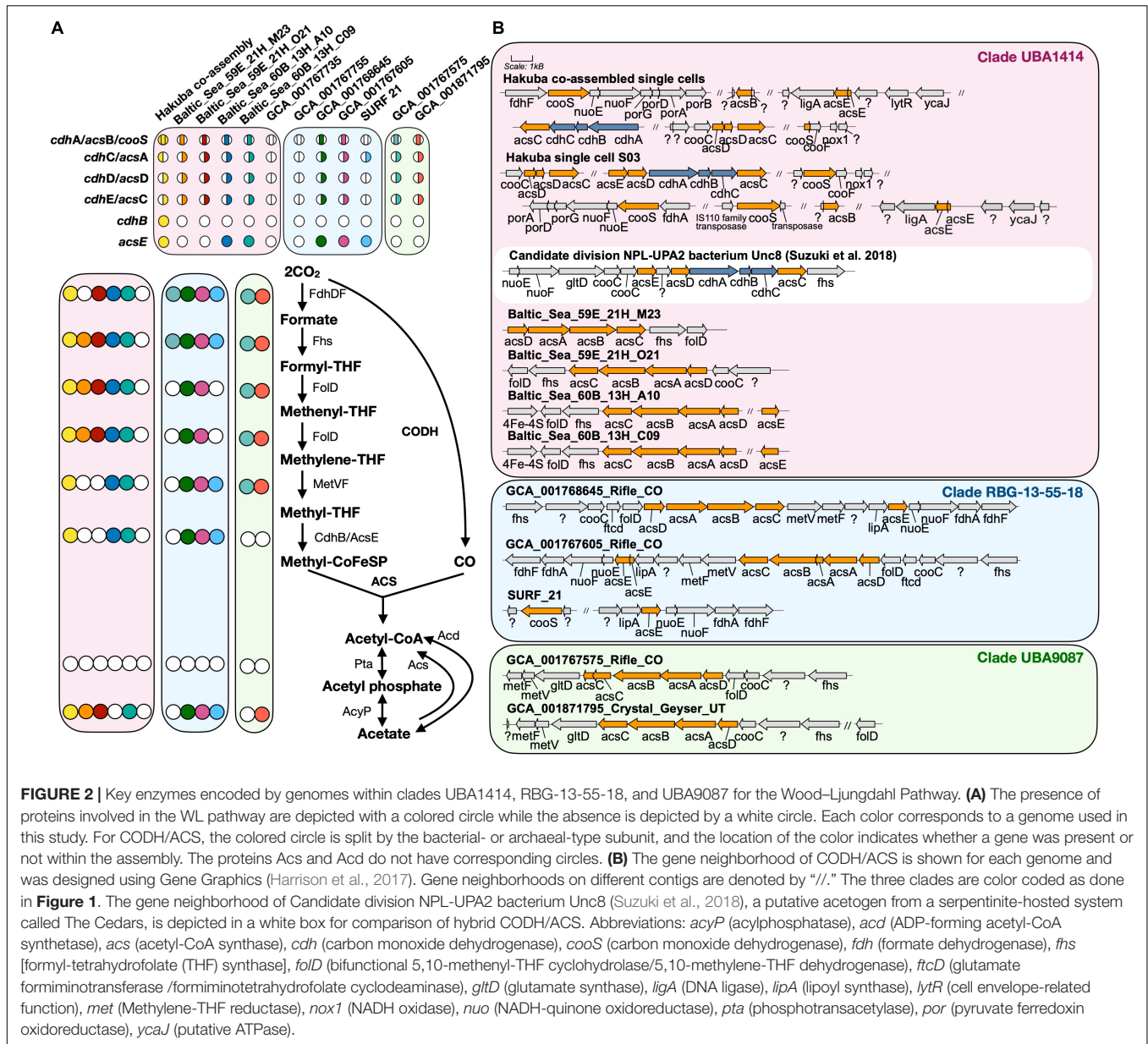
## The Hakuba Actinobacteria Genomes Encode the Wood–Ljungdahl Pathway

The Hakuba co-assembly contains all key genes for the Wood–Ljungdahl (WL) pathway for CO<sub>2</sub> fixation (Ragsdale, 2008; Figure 2). The presence of this pathway has yet to be reported in *Actinobacteria* (Adam et al., 2018); only genes homologous to CODH (*cooS/cdhA/acsB*) in five actinobacterial genomes have been reported (Inoue et al., 2019). The genes for the WL pathway found in the Hakuba co-assembly are: CODH/acetyl-CoA synthase complex (*acs*), formate dehydrogenase (*fdhDF*), formyl-tetrahydrofolate (THF) synthase (*fts*), bifunctional 5,10-methenyl-THF cyclohydrolase / 5,10-methylene-THF dehydrogenase (*folD*), and methylene-THF reductase (*metF*). These genes, including those for the CODH/ACS complex, are also found in four other genomes within the clades UBA1414 and RBG-13-55-18, including two genomes from the Baltic Sea (60B\_13H\_A10 and 60B\_13H\_C09) and two genomes from Rifle (GCA\_001768645 and GCA\_001767605).

The Hakuba co-assembly harbors the genes *fdhD* and *fdhF* that encode proteins involved in the reduction of CO<sub>2</sub> to formate. The source of inorganic carbon could be from the environment or from pyruvate oxidation by the action of

pyruvate ferredoxin oxidoreductase (PorABDG) (Ragsdale, 2003) or pyruvate dehydrogenase (PdhABCD) (de Kok et al., 1998). In serpentinite-hosted environments, there is limited dissolved inorganic carbon, and due to the alkaline pH and presence of high concentrations of divalent cations (e.g., Ca<sup>2+</sup>), inorganic carbon, such as CO<sub>2</sub>, is rapidly sequestered into mineral carbonates (Matter and Kelemen, 2009). Indeed, the total inorganic carbon in Haplo #3 was undetectable (Suda et al., 2014). One candidate source of inorganic carbon is carbon monoxide (CO). CO can be synthesized in these types of environments (Seewald et al., 2006; McCollom and Seewald, 2007) and has been detected in other serpentinite-hosted systems, such as the Coast Range Ophiolite Microbial Observatory (Twing et al., 2017). Furthermore, CO can be utilized by the microbial community in these ecosystems (Morrill et al., 2014; Fones et al., 2019). The Hakuba co-assembly contains genes for anaerobic-type CO dehydrogenase (*cooS* and *cooF*), and the CO dehydrogenase maturation protein (*cooC*). Bicarbonate, if present, could also be another source of inorganic carbon as the Hakuba co-assembly encodes two Na<sup>+</sup>-dependent bicarbonate transporters, indicating the potential to uptake HCO<sub>3</sub><sup>-</sup>, similar to acetogens (Braus-Stromeyer et al., 1997; Smith and Ferry, 2000; Pander et al., 2019). However, homologs of genes coding for carbonic anhydrase that converts HCO<sub>3</sub><sup>-</sup> to CO<sub>2</sub> were not detected in the genome assembly.





**FIGURE 2 |** Key enzymes encoded by genomes within clades UBA1414, RBG-13-55-18, and UBA9087 for the Wood-Ljungdahl Pathway. **(A)** The presence of proteins involved in the WL pathway are depicted with a colored circle while the absence is depicted by a white circle. Each color corresponds to a genome used in this study. For CODH/ACS, the colored circle is split by the bacterial- or archaeal-type subunit, and the location of the color indicates whether a gene was present or not within the assembly. The proteins Acs and Acd do not have corresponding circles. **(B)** The gene neighborhood of CODH/ACS is shown for each genome and was designed using Gene Graphics (Harrison et al., 2017). Gene neighborhoods on different contigs are denoted by “//.” The three clades are color coded as done in Figure 1. The gene neighborhood of Candidate division NPL-UPA2 bacterium Unc8 (Suzuki et al., 2018), a putative acetogen from a serpentinite-hosted system called The Cedars, is depicted in a white box for comparison of hybrid CODH/ACS. Abbreviations: *acyP* (acylphosphatase), *acd* (ADP-forming acetyl-CoA synthetase), *acs* (acetyl-CoA synthase), *cdh* (carbon monoxide dehydrogenase),

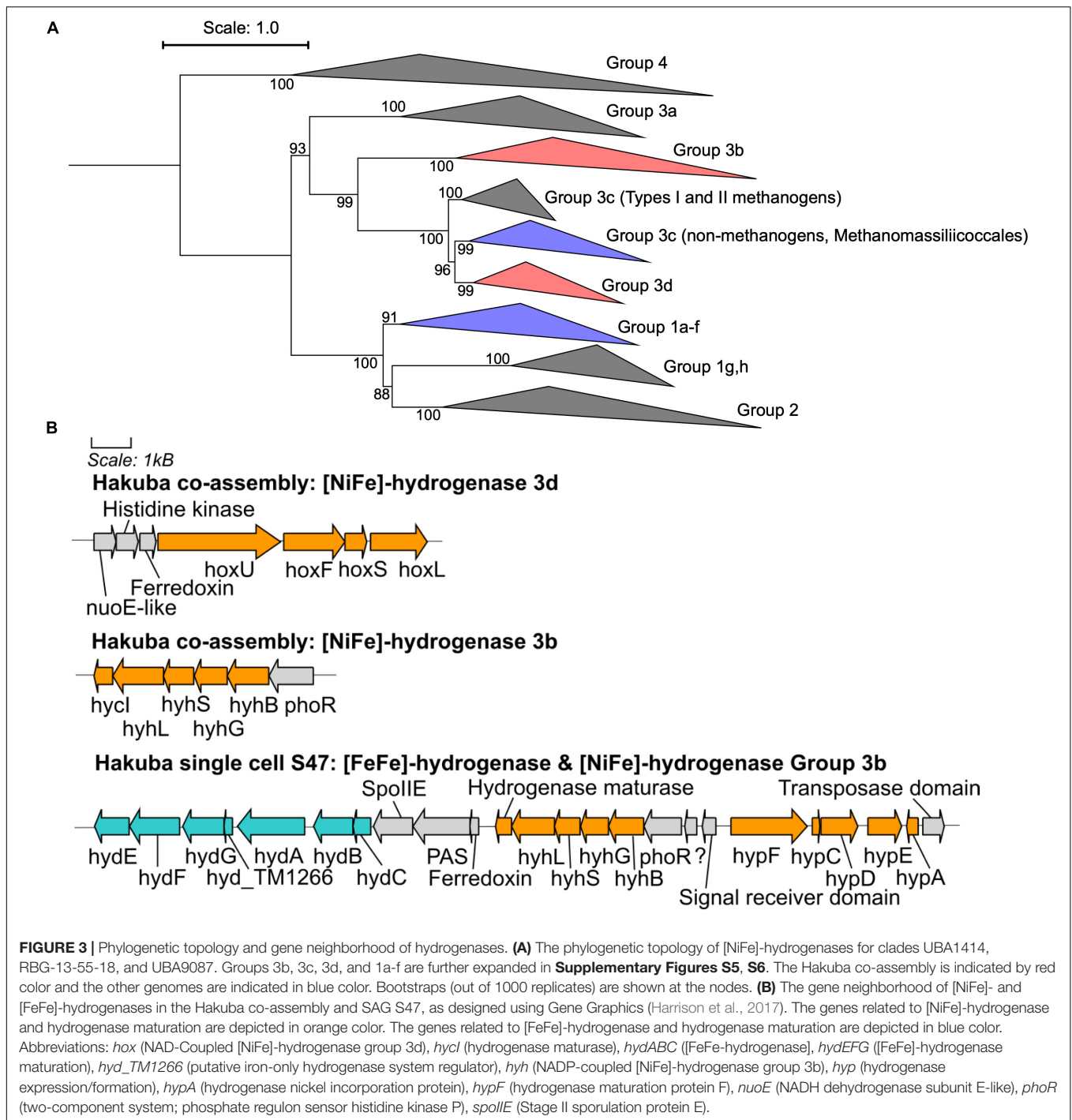
(Pomper et al., 2002; Adam et al., 2019; Hemmann et al., 2019). However, the genes to synthesize methanopterin derivatives, such as dihydromethanopterin reductase and tetrahydromethanopterin:alpha-L-glutamate ligase (Xu et al., 1999; Maden, 2000; Scott and Rasche, 2002), were not observed in the Hakuba actinobacterium genome. In addition, compared to *fhcB* from the methylophilic *Methylorubrum extorquens* (Hemmann et al., 2019), the amino acid sequence of the catalytic subunit *fwdB* from the Hakuba SAGs lacks the sequence motifs for a tungstopterin cofactor and contains two necessary components for a functioning *fwdB*: a N-terminal domain with [4Fe-4S] cluster and a catalytic Cys118 (**Supplementary Figure S4**).

In methanogens, FwdABD comprise the catalytic subcomplex while the function of FwdC remains unknown (Wagner et al., 2016). FwdE is an iron-sulfur protein (Hochheimer et al., 1995) and is hypothesized to function as a DNA-binding protein in acetogens (Shin et al., 2016). The gene cluster *fwdABCD* without *fwdE* was also identified in the methanogen OP bin 54 (*Methanomethyliales*) (Berghuis et al., 2019). In *Methanosarcina acetivorans*, the synthesis of FwdDBAC is likely important for carboxydrotrophic growth and is potentially involved in the production of formate (Matschiavelli and Rother, 2015). In comparison, in the genomes of 14 cultivated acetogens, only the subunit *fwdE* is observed (Shin et al., 2016). Similarly, the Rifle MAGs within sub-clades RBG-13-55-18 and UBA9087 harbor *fwdE* but not *fwdABCD*. The presence of the catalytic subunits *fwdABD* in the Hakuba co-assembly or SAGs suggests that the Fwd-like complex could be active, with potential function during growth on CO, as demonstrated for *Methanosarcina acetivorans* (Matschiavelli and Rother, 2015).

CO<sub>2</sub> reduction could be driven by reducing equivalents generated by hydrogen (H<sub>2</sub>) oxidation via [NiFe]- or [FeFe]-hydrogenases (**Figure 3** and **Supplementary Figures S5, S6**). Although known acetogens encode hydrogenase modules in a gene cluster containing a *fdh* gene (Shin et al., 2016), the hydrogenase genes in genomes of clades UBA1414, UBA9087, and RBG-13-55-18 are not clustered with *fdh*, with the exception of the Rifle MAG GCA\_001767575. Considering that these are incomplete genomes derived from metagenomic or single-cell genomic assemblies (**Table 3**), it is possible that a gene cluster containing *fdh* and hydrogenases was fragmented. Two [NiFe]-hydrogenase genes were identified in the Hakuba co-assembly and were phylogenetically affiliated with Group 3b and Group 3d [NiFe]-hydrogenases (**Figure 3A**) that are coupled to the bidirectional reduction of NADP<sup>+</sup> and NAD<sup>+</sup>, respectively, in other organisms (Vignais et al., 2001; Peters et al., 2015). Accordingly, the diaphorase (*hyhG* or *hoxF*) and Fe-S (*hyhB* or *hoxU*) cluster modules associated with these two enzymes were also co-localized with the large (*hyhL* or *hoxL*) and small (*hyhS* or *hoxS*) subunits of the respective [NiFe]-hydrogenase groups (**Figure 3B**; Peters et al., 2015). The other genome assemblies in the three clades also encoded [NiFe]-hydrogenases that were variably affiliated with either group 1, group 3b, group 3c, or group 3d (**Supplementary Figures S5, S6**). Amongst all the genome assemblies in the three clades, the Hakuba co-assembly is the only genome to encode a [FeFe]-hydrogenase.

The Hakuba co-assembly encoded the catalytic subunit HydA, and the Hakuba SAG S47 also encoded the cluster HydABC along with [FeFe]-hydrogenase accessory proteins, including HydEFG (Posewitz et al., 2004). The HydABC cluster was present within the gene neighborhood of the Group 3b-like [NiFe]-hydrogenase (**Figure 3**). Homologs of HydABC have been suggested to be involved in electron bifurcation (Schut and Adams, 2009; Poudel et al., 2016), a process where reversible H<sub>2</sub> oxidation is coupled to simultaneous reduction of NAD<sup>+</sup> and ferredoxin (Fd) (Buckel and Thauer, 2013). Electron bifurcating hydrogenases are implicated in the energy conservation of model acetogens like *Acetobacterium woodii* (Wiechmann et al., 2020) via coupling of Fd (oxidation/reduction) and NAD<sup>+</sup>/NADH reduction and oxidation with H<sub>2</sub> (oxidation/reduction). It is possible that such activities are also catalyzed by the Hakuba actinobacterial cells in conjunction with the reduction/oxidation of NAD(P)<sup>+</sup>/NAD(P)H via the [NiFe]-hydrogenases.

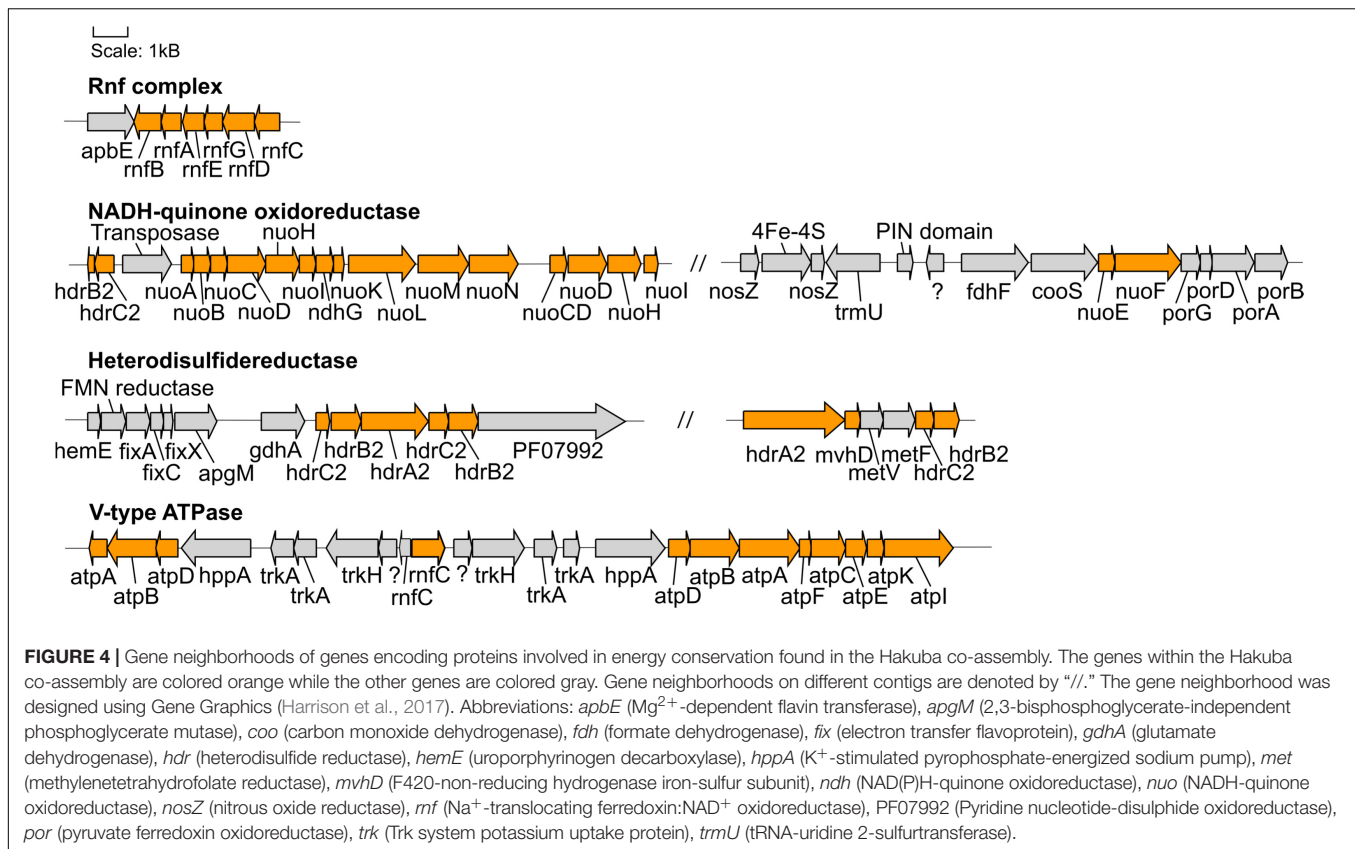
Over the next few steps of the WL pathway, formate is likely converted to methyl-THF. From formyl-THF to methenyl-THF, the bifunctional 5,10-methenyl-THF cyclohydrolase / 5,10-methylene-THF dehydrogenase (FolD) is likely the enzyme used by the Hakuba bacterium and those within clades UBA1414, UBA9087, and RBG-13-55-18. In comparison, pan-genome analysis of 14 cultivated acetogens demonstrated that this conversion mainly occurs with the enzyme formyl-THF cyclohydrolase (Fch) (Shin et al., 2016). However, in the acetogen *Moorella thermoacetica*, FolD is used as a bifunctional protein with cyclohydrolase and dehydrogenase activity for the two-step conversion of formyl-THF to methylene-THF (O'Brien et al., 1973). Given the absence of *fch* and presence of *folD* in the clades UBA1414, UBA9087, and RBG-13-55-18, these genomes likely contain a bifunctional *folD* similar to *Moorella thermoacetica*. Methylene-THF reductase subunits V and F (MetVF) then catalyze the reduction from methylene-THF to methyl-THF (**Figure 4**). Interestingly, *metVF* in the Hakuba co-assembly is located with the F<sub>420</sub>-non-reducing hydrogenase iron-sulfur subunit D (*mvhD*) and heterodisulfide reductase subunits *hdrA*, *hdrB*, and *hdrC* (*hdrA-mvhD-metVF-hdrCB*). In comparison, genome assemblies within clade RBG-13-55-18 encode a gene cluster with only *mvhD/hdrABC* while the other MAGs/SAGs of clades UBA1414, UBA9087, and RBG-13-55-18 do not harbor this gene cluster. The genes *hdrABC* encode a key enzyme in methanogens that is usually complexed with a [NiFe]-hydrogenase (HdrABC-MvhAGD) and functions in electron bifurcation to oxidize H<sub>2</sub> coupled with the reduction of Fd and CoM-S-S-CoB (a final product of methanogenesis; heterodisulfide coenzyme M and coenzyme B) (Kaster et al., 2011; Wagner et al., 2017). The *metVF/mvhD/hdrABC* gene cluster has been observed in acetogens, such as *Moorella thermoacetica* (Mock et al., 2014) and '*Candidatus* Adiatrix intracellularis' (Ikeda-Ohtsubo et al., 2016). In *Moorella thermoacetica*, this complex can reduce methylene-THF with benzyl viologen, and benzyl viologen can be reduced by NADH (Mock et al., 2014). Although the second electron acceptor remains unknown, it is likely that MetVF/MvhD/HdrABC from *Moorella thermoacetica* is capable of electron bifurcation via an electron-bifurcating flavin in the subunit HdrA (Mock et al., 2014). The MvhD



subunit of MvhAGD contains a [2Fe-2S] cluster (Wagner et al., 2017) and could potentially function to donate electrons to HdrABC. It is possible that this *metVF/mvhD/hdrABC* cluster is regulated by the same mechanism for the catalysis of methylene-THF to methyl-THF.

The next step in the WL pathway utilizes CODH/ACS to produce acetyl-CoA. A complete or partial gene cluster for the CODH/ACS complex was identified in all genomes

in the three clades except for GCA\_001767735 (Rifle; clade UBA1414) and GCA\_001767755 (Rifle; clade RBG-13-55-18) (**Figure 2A**). The CODH/ACS enzyme complex consists of five subunits, in which four share homology between *Bacteria* and *Archaea* (Adam et al., 2018; Inoue et al., 2019): *cdhA* (*acsB* in *Bacteria*;  $\alpha$ -subunit), *cdhC* (*acsA*;  $\beta$ -subunit), *cdhD* (*acsD*;  $\delta$ -subunit), *cdhE* (*acsC*;  $\gamma$ -subunit). The gene only found in *Archaea* is *cdhB* ( $\epsilon$ -subunit) while *acsE* is unique to *Bacteria*.



Notably, the Hakuba co-assembly is the only genome amongst the three clades to harbor a hybrid CODH/ACS consisting of both archaeal- (*cdhABC*) and bacterial-type (*acsCDE*) subunits (Figure 2 and Supplementary Figure S7). Although the Hakuba co-assembly did not contain the entire gene cluster for the CODH/ACS complex on one contig, the Hakuba SAG S03 had the full operon of *acsED-cdhABC-acsC*, in addition to separate gene clusters containing another *acsC*, a split *acsD*, and a split *acsE* (Figure 2B). The hybrid CODH/ACS has been observed in some putative acetogens, and several subunits have sequence similarity to subunits identified in a MAG of the candidate phylum NPL-UPA2 from the serpentinizing environment of The Cedars (Suzuki et al., 2018; Supplementary Figure S7). The hybrid CODH/ACS could be a result of horizontal transfer from *Archaea* to *Bacteria* (Adam et al., 2018; Suzuki et al., 2018). The biochemical properties of a hybrid CODH/ACS remain unknown and could provide insight into metabolisms present in serpentinizing systems and subsurface environments.

In the final steps of the WL pathway, acetyl-CoA is converted to acetate. Although none of the genomes within clades UBA1414, UBA9087, and RBG-13-55-18 harbor phosphotransacetylase (*pta*; acetyl-CoA to acetyl phosphate) or acetate kinase (*ack*; acetyl phosphate to acetate), several genome assemblies within all three clades contain homologs of acylphosphatase (*acyP*), suggesting the likely conversion of acetyl phosphate to acetate. It remains unclear whether

the bacteria in all three clades can convert acetyl-CoA to acetyl phosphate, and potential genes which could replace *pta*, such as phosphotransbutyrylase (*ptb*), butyrate kinase (*buk*), or propanediol utilization protein (*pduL*) (Köpke et al., 2010; Poehlein et al., 2015), were not detected amongst any of the genomes. All genomes, except for three (Hakuba co-assembly/SAGs, BS\_59E\_21H\_M23, and GCA\_001767575), have the ADP-forming acetyl-CoA synthetase (*acd*) (Supplementary Table S10). The presence of the WL pathway and the absence of *ack* and *acd* in the Hakuba co-assembly/SAGs suggests that this bacterium likely cannot autotrophically fix CO<sub>2</sub>, although the genome is incomplete. It is known that autotrophic growth by acetogenesis requires the ATP generated by the action of *ack* or *acd* (Musfeldt et al., 1999; Schuchmann and Müller, 2014).

Several genomes within clades UBA1414, UBA9087, and RBG-13-55-18 harbor genes that can convert acetate to other products, such as acetyl-CoA (acetyl-CoA synthetase, *acs*), acetaldehyde (aldehyde ferredoxin oxidoreductase, *aor*), and ethanol (aldehyde-alcohol dehydrogenase, *adhE*; alcohol dehydrogenase, *adh*) (Supplementary Table S10). Acetate may be derived from several sources, including the WL pathway, the L-cysteine synthesis pathway, and the environment via a putative acetate transporter. However, it remains unclear whether the Hakuba actinobacterium is a bonafide acetogen as it likely cannot convert acetyl-CoA to the major end products (acetate, acetone, ethanol, and butyrate) of the WL pathway, and several genes are missing for these pathways, as mentioned above, including genes



involved in acetone or butyrate synthesis (e.g., 3-hydroxybutyryl-CoA dehydrogenase and enoyl-CoA hydratase).

## Energy Conservation Mechanisms in the Hakuba Co-assembly

The Hakuba co-assembly encodes enzymes involved in energy conservation used by homoacetogens that exploit the WL pathway (Figures 3, 4 and Supplementary Table S11). In addition to hydrogenases and heterodisulfide reductase, genes coding for several subunits of the Na<sup>+</sup>-dependent V-type ATPase (*atpABCDEFIK*) were identified with key amino acid residues within subunit AtpK implicated in Na<sup>+</sup> translocation (Mulikidjanian et al., 2008; Supplementary Figure S8). The Na<sup>+</sup>-dependent V-type ATPase can also translocate H<sup>+</sup> (Dimroth, 1997; von Ballmoos and Dimroth, 2007). There were no genes encoding the H<sup>+</sup>-dependent F-type ATPase and only one subunit for the Ca<sup>2+</sup>/Mg<sup>2+</sup>-dependent P-type ATPase was identified. For the Hakuba bacterium, the translocation of Na<sup>+</sup> compared to other ions (H<sup>+</sup>, Ca<sup>2+</sup>, or Mg<sup>2+</sup>) is likely to occur since there is about 1–1.6 mM Na<sup>+</sup> in Haplo #3 (Table 1), which is ten times higher than the concentration of Ca<sup>2+</sup>. Magnesium ions were not detected, and protons in serpentinite-hosted ecosystems are expected to be extremely low in concentration and are consumed during serpentinitization, leading to alkaline pH (Okland et al., 2012).

The Hakuba co-assembly also encodes all subunits of the Rnf complex (RnfABCDEG). In model acetogens, such as *A. woodii*, the Rnf complex couples Fd oxidation to NAD<sup>+</sup> reduction and Na<sup>+</sup> translocation across the membrane, creating a sodium ion gradient that is subsequently utilized by the V-type ATPase for ATP synthesis (Biegel and Müller, 2010; Biegel et al., 2011; Schuchmann and Müller, 2014, 2016). The Hakuba bacterium may generate ATP via the combination of the Rnf complex and the V-type ATPase, as described above (Schuchmann and Müller, 2014, 2016). The lack of *ack* or *acd* suggests that the bacterium likely cannot conduct net ATP production via only the WL pathway. Thus, the bacterium would require ATP production via glycolysis coupled with the H<sup>+</sup>-translocating NADH-quinone oxidoreductase (NuoABCDEFGHIKLMN, NuoE, NuoF and NuoG) and dissimilatory nitrate reduction with nitrate reductase, NarGH. Reduction of nitrate via nitrate reductase may be coupled with the oxidation of hydrogen via hydrogenases, although *narG* is pseudogenized in some SAGs, as discussed below. The genome assemblies of potential acetogens with the complete WL pathway and *acd* (GCA\_001768645\_Rifle\_CO, and GCA\_001767605\_Rifle\_CO) also encode the V-type ATPase and Rnf complex, but not NarGH and the NADH-quinone oxidoreductase (Supplementary Table S10).

## The Hakuba Actinobacterium Is a Possible Heterotroph

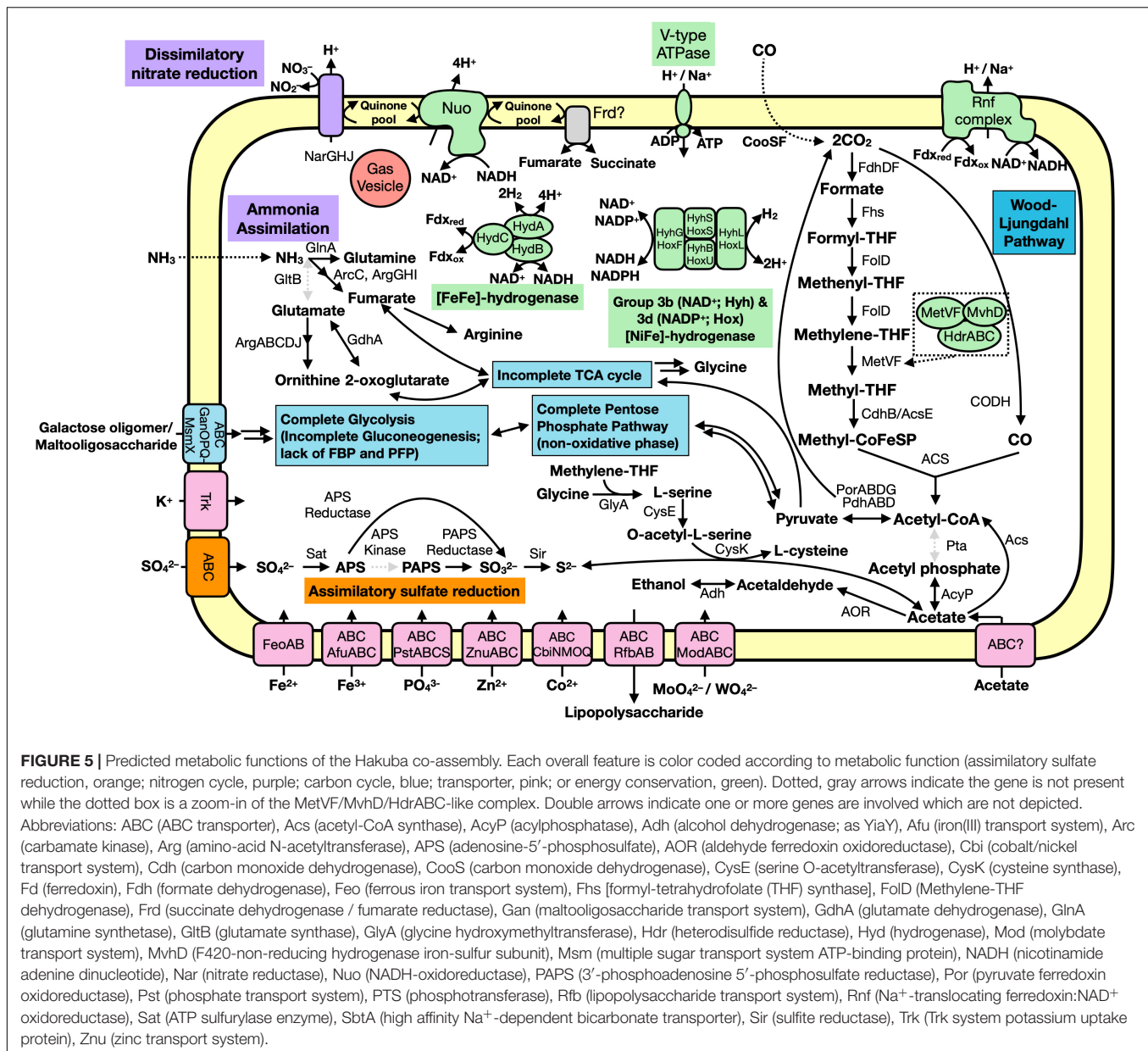
We further characterized the Hakuba co-assembly to ascertain the metabolic capabilities other than utilizing the WL pathway (Figure 5). The Hakuba co-assembly is capable of assimilatory sulfate reduction (see Supplementary Figure S9 and Supplementary Information) and has the complete set

of genes for glycolysis (Embden-Meyerhof-Parnas pathway), converting glucose to acetyl-CoA (Supplementary Table S10). Sugars can be imported via the ABC transporter GanOPQ-MsmX and subsequently converted to glucose with genes, such as galactokinase (*galK*) and galactose-1-phosphate uridylyltransferase (*galT*). However, the Hakuba co-assembly and all SAGs lack fructose-1,6-bisphosphatase or diphosphate-dependent phosphofructokinase; therefore, it may not complete gluconeogenesis. This suggests that this bacterium cannot grow solely by carbon fixation via the WL pathway and is likely a heterotroph. In addition, the bacterium is likely not capable of producing major WL pathway end products from acetyl-CoA, as mentioned above, further supporting its dependence on glycolysis for energy generation. As Haplo #3 is nutrient limited and organic carbon may not always be present, the Hakuba bacterium may supplement anabolic processes via the WL pathway.

The Hakuba co-assembly harbors all the genes for the non-oxidative pentose-phosphate pathway. The tricarboxylic acid cycle is incomplete, as seen in other known acetogens (Shin et al., 2016). Pyruvate can be reduced to malate using malate dehydrogenase, which is further converted to fumarate using fumarate hydratase. The Hakuba co-assembly contains genes for fumarate reductase catalytic, cytosolic subunits (*frdAB*), but the membrane-bound subunits are missing. The Hakuba co-assembly also contains glutamine synthetase indicating that it can potentially assimilate ammonia as a nitrogen source.

## Characterization of Nitrate Reductase and Pseudogenization of *narG* With Intraspecies-Variations

Nitrate reductase may confer the ability to respire nitrate for the Hakuba bacterium. Its genes, *narGHJ*, were not found in the same operon within the Hakuba co-assembly, but the intact full-length operon was identified in the Hakuba SAGs S03, S09, S34, and S42. Although the co-assembly and SAGs are missing the integral membrane subunit *narI*, which is often observed with *narGHJ* (Philippot, 2002; Cabello et al., 2004), there is at least one known denitrifying microorganism that encodes only the subunits NarGH and NarJ, *Haloarcula marismortui* (Yoshimatsu et al., 2000, 2002). It is also possible that *narI* was not sequenced, or there is a yet to be identified putative *narI*, such as observed in the secretome of *Aeropyrum pernix* K1 (Palmieri et al., 2009) but not in the genome (Kawarabayasi et al., 1999). The subunit NarH is responsible for electron transfer (Blasco et al., 1989) while NarJ is a chaperone protein necessary for assembling an active NarGH complex (Dubourdieu and DeMoss, 1992; Liu and DeMoss, 1997; Blasco et al., 1998). The subunit NarG is the catalytic subunit and can be located in either the cytoplasm or the periplasm (Richardson et al., 2001). The Hakuba co-assembly NarG has a canonical twin-arginine motif, [S/T]RR, at the N-terminal region, which is responsible for protein export to the periplasm (Martinez-Espinosa et al., 2007; Kameya et al., 2017; Supplementary Figure S10), and it phylogenetically clustered with NarG from *Hydrogenobacter thermophilus* (Supplementary Figure S11). The NarG from *H. thermophilus* was the first known



bacterial NarG to be localized on the periplasmic side of the cell membrane (Kameya et al., 2017). Similar to *H. thermophilus*, the Hakuba co-assembly genome also does not contain genes for any nitrate/nitrite transporters, further supporting the periplasmic localization of NarG. After nitrate reduction, nitrite does not appear to be utilized by the Hakuba organisms. Although the Hakuba co-assembly does not encode nitrite reductase (Nir) and nitric oxide reductase (Nor), the gene neighborhood of *narGHJ* contains several Fe-S cluster-containing proteins, putative ubiquinol oxidases, and two putative *nosZ* genes. These proteins may be involved in preventing toxicity from nitrite and potentially, nitric oxide.

The gene *narG* appears to be pseudogenized by a nonsense mutation (“TGG” to “TAG”) in the Hakuba co-assembly and

several SAGs (S03, S09, and S34), whereas SAGs S06, S33, and S42 have a complete *narG* sequence (Supplementary Figure S12). Examination of raw reads mapped to the region suggest that the pseudogenized *narG* was not derived from sequencing errors because the same mutation occurred at the same position on orthologous contigs among three independent samples. Moreover, this mutation was not within a homopolymeric region, which could have easily resulted in indel errors during sequencing. The potential loss of function for dissimilatory nitrate reduction in SAGs S03, S09, and S34 may arise from either functional heterogeneity amongst the individual strains of this species or an adaptation to the serpentinite-hosted environment, which has no detectable amount of nitrate or nitrite (Table 1) and may lead to the

loss of NarGHJ amongst the whole population of this species. Without the utilization of nitrate, the Hakuba actinobacterium may use fumarate as an electron acceptor, although it is unclear whether the putative fumarate reductase is membrane-bound or not.

The split *narG* amongst the 10 SAGs coincide with two intraspecies-level phylotypes observed by ANI (**Supplementary Table S6**) and AAI (**Supplementary Table S7**). Based on the criteria of ANI  $\geq 99\%$  and AAI  $\geq 94\%$ , one phylotype consists of SAGs S03, S09, S34, S44, and S47 (“first phylotype”) and another phylotype consists of S25, S33, and S43 (“second phylotype”). The SAGs S06 and S42 are closer to the second phylotype, but the similarity is lower (AAI  $< 94\%$ ). Further examination of the genome assemblies identified 12 split genes (**Supplementary Figure S13** and **Supplementary Table S12**), and the two phylotypes coincided with 9 out of the 12 genes. Notably, the bacterial subunit *acsE* of CODH/ACS (**Figure 2B**) was split only in the first phylotype (**Supplementary Figure S13C**). The discovery of split genes coinciding with two phylotypes clarified the strength of our approach to analyze multiple SAGs of the same species, while the co-assembly provided higher completeness and facilitated analysis of potential metabolic traits of the Hakuba actinobacterium as a species.

## Transporters, Stress Response, Motility, and RubisCO-Like Protein of the Hakuba Co-assembly

The Hakuba actinobacterium genome encodes several mechanisms that may allow it to survive in the nutrient-limited, serpentinite-hosted ecosystem of Happo #3. Several transporters are encoded by the Hakuba co-assembly, including organic carbon and inorganic ion transporters ( $K^+$ , Fe(II/III),  $SO_4^{2-}$ ,  $PO_4^{3-}$ ,  $Zn^{2+}$ ,  $Co^{2+}$ , and  $MoO_4^{2-}/WO_4^{2-}$ ). In addition to the  $H^+/Na^+$ -dependent V-type ATPase and  $Na^+/H^+$ -translocating Rnf complex, the  $K^+/H^+$ -symporter (Trk) is needed to maintain homeostasis in an alkaline environment and to create a  $Na^+/K^+$  gradient, which likely has greater electrochemical storage capacity than a proton gradient (Dibrova et al., 2015). Moreover, the Happo #3 alkaline environment contains high concentrations of  $K^+$  and  $Na^+$  compared to protons (**Table 1**). The genome also encodes several secondary metabolite biosynthetic gene clusters related to putative saccharide and fatty acid biosynthesis with unknown products (**Supplementary Table S13**). Although the function remains unknown for these pathways, secondary metabolites are known to exhibit diverse biological activities and play an important role in community interactions (Cimermancic et al., 2014).

The genome encodes a range of anti-stress and defense mechanisms, including cold and heat shock proteins (YfiA, CspA, HspR) and defense against phage infection (AbiEii toxin-antitoxin Type IV system and CRISPR/Cas system). A Cas Type IIIB operon is located near a short (99 bp) putative CRISPR sequence, which was identified at the end of the contig and could be truncated (**Supplementary Figure S14**). The Type III CRISPR/Cas defense mechanism produces a

complex for targeted search and elimination (Wright et al., 2016). However, the corresponding virus remains unknown as the spacer within the CRISPR sequence did not match any known sequences in the CRISPRCasFinder database (Couvin et al., 2018) or the NCBI nr database. There are also three putative tyrosine-type phage integrases in the Hakuba co-assembly. One integrase is located next to a tRNA gene. In general, tRNA genes are the preferred integration site of prophages (Williams, 2002), and accordingly, this integrase region located next to a tRNA gene could be a part of a prophage. Within the gene neighborhood, there was also a Type IIG restriction-modification gene, which is one defense mechanism against ‘non-self’ DNA (Naito et al., 1995; Kobayashi, 2001).

Flagellar motility for the cells represented by the Hakuba co-assembly is unclear as there are only a few genes for biosynthesis of flagella (*fliAD*) and Type IV pili. On the other hand, the Hakuba actinobacterium may be capable of flotation using gas vesicles. The Hakuba co-assembly contains genes for several gas vesicle proteins (*gvpGK[L/F]MNOVY*) spread across three contigs, and the SAG S33 has the *gvp* genes (*gvpAGHJK[L/F]MNOV*) located on one contig (**Supplementary Table S14**). Gas vesicles are proteinaceous organelles, generally in the shape of a spindle or cylinder, found in both *Bacteria* and *Archaea*, and impart buoyancy to cells by allowing passive gas diffusion (e.g.,  $O_2$ ,  $N_2$ ,  $H_2$ ,  $CO_2$ ,  $CO$ , and  $CH_4$ ) (Walsby, 1994; Oesterhelt, 1998; Oren et al., 2006; Coker and DasSarma, 2007; Hechler and Pfeifer, 2009; Pfeifer, 2012; Tashiro et al., 2016). The major gas vesicle protein GvpA is known to have an important role in assembling a gas vesicle while the other proteins play minor (e.g., GvpCG), regulatory (e.g., GvpDE), or unknown (e.g., GvpHI) roles (DasSarma and DasSarma, 2015). A single cell can contain several gas vesicles and likely produces gas vesicles in response to stress or environmental stimuli, such as light, oxygen concentrations, and available nutrients (Pfeifer, 2012). The Hakuba actinobacterium may synthesize gas vesicles in response to heat shock since *hsp20* (heat shock protein) is located in the same gene neighborhood as the *gvp* genes (**Supplementary Table S14**).

The Hakuba co-assembly encodes RubisCO (ribulose 1,5-bisphosphate carboxylase/oxygenase) Form IV protein, also known as a RubisCO-like protein (**Supplementary Figure S15**), which was identified in six SAGs S03, S25, S34, S42, S43, and S44. RubisCO is one of the enzymes involved in carbon fixation and is categorized into four forms. However, Form IV is known as a RubisCO-like protein because it lacks the catalytic site residue involved in the carboxylation reaction and is thought to be the ancestral form of RubisCO, arising before the great oxygenation event (Kacar et al., 2017; Erb and Zarzycki, 2018). Some RubisCO-like proteins are known to be involved in the methionine salvage pathway (Ashida, 2003; Erb et al., 2012) and the degradation of four carbon sugar acids (Zhang et al., 2016). The full metabolic range of RubisCO-like proteins remains unknown, although most are likely isomerases and/or epimerases (Erb and Zarzycki, 2018). In the Hakuba actinobacterium, the RubisCO-like protein may function as an epimerase acting on



sugars, as the gene neighborhood of the RubisCO-like protein harbors genes with an epimerase conserved domain (cd09023) and a sugar substrate binding site (DeoR C-terminal sensor domain, PF00455). Closely related RubisCO-like proteins to the Hakuba co-assembly, as determined by BLASTp searches, include those from the Candidate Phyla Radiation, *Spirochaetes*, *Planctomycetes*, and the model acetogen *Moorella thermoacetica*. The functions of RubisCO-like proteins in these genomes are also unknown.

## CONCLUSION

Terrestrial serpentinite-hosted ecosystems are important modern-day analogs of early Earth and can also provide insights into processes that may have supported life at that time. Here, we present a genomic characterization of a dominant member in the Hakuba Happo hot spring ecosystem that belongs to the early-branching actinobacterial clade UBA1414. Single-cell genomics revealed that the bacterium utilizes the WL pathway for converting CO<sub>2</sub> to acetyl-CoA and could be represented by two phylotypes within a single species. We also identified related genome assemblies that encode the WL pathway; these bacteria are the first known to encode the WL pathway within the *Actinobacteria* phylum. Within Happo #3, examination of other dominant members, such as “Parcubacteria” and *Nitrospirae*, will further reveal the characteristics of this ecosystem. On the basis of the single-cell genome sequences, we propose a novel order ‘*Candidatus* Hakubanellales’ and novel family ‘*Candidatus* Hakubanellaceae.’ We propose to name this bacterium ‘*Candidatus* Hakubanella thermoalkaliphilus’ as described below.

### Description of ‘*Candidatus* Hakubanella’ gen. nov

*Hakubanella* [ha.ku.ba.nel’la, N.L. fem. dim. n. *Hakubanella* of Hakuba Happo, a serpentinite-hosted environment located in Nagano (Japan) from where the single-cell genome assemblies were obtained. The type species is ‘*Candidatus* Hakubanella thermoalkaliphilus’ with the single-cell genome assemblies as the type material.

### Description of ‘*Candidatus* Hakubanella thermoalkaliphilus’ sp. nov

*Hakubanella thermoalkaliphilus* (ther.mo.al.ka.liphil.us, Gr. adj. *thermos* hot; N.L. n. *alkali* from Arabic *al-qaliy* the ashes of saltwort; Gr. adj. *philos* friend, loving; N.L. adj. *alkaliphilus* liking alkaline environments). The genome of the bacterium was discovered in Hakuba Happo hot springs, where temperatures reach about 50°C and pH~11. Based on genome analysis, the bacterium is anaerobic and possesses glycolysis for energy harvesting and the WL pathway for anabolic processes. The bacteria can potentially utilize sugars and accommodate two phylotypes: one is capable of dissimilatory nitrate reduction and another has lost the activity. The bacteria probably cannot grow autotrophically. The assignment is based on single-copy taxonomic marker genes.

## DATA AVAILABILITY STATEMENT

The 16S rRNA amplicon sequences (DRR198702–DRR198704 under DRA009263), the raw fastq files (DRR198705–DRR198714 under DRA009264), and cleaned assemblies for SAGs (BLRU01000000–BLRZ01000000, BLSA01000000–BLSD01000000) and the co-assembly (BLSE01000000) were deposited into DDBJ under BioProject accession no. PRJDB8357. The gas vesicle hidden Markov model database is available on GitHub at <https://github.com/Arkadiy-Garber/MagicCave/tree/master/hmms/gas>.

## AUTHOR CONTRIBUTIONS

NM, KK, and YH designed the study. NM contributed to the field sampling and the collection, sequencing, and bioinformatics analyses of the single-cell genomes. MK and YH contributed to the sequencing and bioinformatics analyses. EB and DC contributed to analysis of hydrogenases. NM, MK, EB, DC, SM, KN, KK, and YH wrote the manuscript.

## FUNDING

NM was supported by the ELSI Origins Network (EON) research fellowship, which is supported by a grant from the John Templeton Foundation. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation. NM was also partially supported by the NASA Grant NNA13AA92A and Air Force Office of Scientific Research Grant FA9550-14-1-0114. EB acknowledges support from the NASA Astrobiology Institute (NNA15BB02A). KN was supported in part by the NSF Grant #1638216. Lawrence Livermore National Laboratory is operated by Lawrence Livermore National Security, LLS, for the U.S. Department of Energy, National Nuclear Security Administration under Contract DE-AC52-07NA27344 (LLNL-JRNL-800258).

## ACKNOWLEDGMENTS

We thank the members of Yoichi Kamagata’s lab (Hideyuki Tamaki, Masaru K. Nobu, Ryosuke Nakai, Satoshi Tamazawa, and Yoichi Kamagata) for field sampling and geochemical measurements, which was supported by the JSPS KAKENHI grant JP26106004, and advice on co-assembly of the single cells. We thank members of Yuichi Hongoh’s lab (Hirokazu Kuwahara, Yuniar Devi Utami, Takumi Murakami, and Katsura Igai) for advice and help with collection and sequencing of the single-cells. Sanger sequencing was done at the Biomaterial Analysis Center of the Tokyo Institute of Technology. We thank Yayoi Hongo and Yoko Shibamoto for analyzing the nucleobases, nucleotides, and amino acids. We thank Shino Suzuki and Arkadiy Garber for advice on bioinformatics analyses and John Yu for setup of bioinformatics tools on the University of Southern California server. We thank Karen Lloyd for providing cleaned assemblies



collected from the Baltic Sea. We also thank Tanja Woyke for advice on single cell collection and Philip Hugenholtz for advice on GTDB-Tk. Computational resources were partly provided by the Data Integration and Analysis Facility, National Institute for Basic Biology, Japan.

## REFERENCES

- Adam, P. S., Borrel, G., and Gribaldo, S. (2018). Evolutionary history of carbon monoxide dehydrogenase/acetyl-CoA synthase, one of the oldest enzymatic complexes. *Proc. Natl. Acad. Sci. U.S.A.* 115, E1166–E1173. doi: 10.1073/pnas.1807540115
- Adam, P. S., Borrel, G., and Gribaldo, S. (2019). An archaeal origin of the Wood–Ljungdahl H4MPT branch and the emergence of bacterial methylophily. *Nat. Microbiol.* 4, 2155–2163. doi: 10.1038/s41564-019-0534-2
- Anantharaman, K., Brown, C. T., Hug, L. A., Sharon, I., Castelle, C. J., Probst, A. J., et al. (2016). Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat. Commun.* 7:13219. doi: 10.1038/ncomms13219
- Arai, W., Taniguchi, T., Goto, S., Moriya, Y., Uehara, H., Takemoto, K., et al. (2018). MAPLE 2.3.0: an improved system for evaluating the functionomes of genomes and metagenomes. *Biosci. Biotechnol. Biochem.* 82, 1515–1517. doi: 10.1080/09168451.2018.1476122
- Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., et al. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* 44, W16–W21. doi: 10.1093/nar/gkw387
- Ashida, H. (2003). A functional link between RuBisCO-like protein of *Bacillus* and photosynthetic RuBisCO. *Science* 302, 286–290. doi: 10.1126/science.1086997
- 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
- Barker, H. A., and Kamen, M. D. (1945). Carbon dioxide utilization in the synthesis of acetic acid by *Clostridium thermoaceticum*. *Proc. Natl. Acad. Sci. U.S.A.* 31, 219–225. doi: 10.1073/pnas.31.8.219
- Berghuis, B. A., Yu, F. B., Schulz, F., Blainey, P. C., Woyke, T., and Quake, S. R. (2019). Hydrogenotrophic methanogenesis in archaeal phylum *Verstraetearchaeota* reveals the shared ancestry of all methanogens. *Proc. Natl. Acad. Sci. U.S.A.* 116, 5037–5044. doi: 10.1073/pnas.1815631116
- Biegel, E., and Müller, V. (2010). Bacterial Na<sup>+</sup>-translocating ferredoxin:NAD<sup>+</sup> oxidoreductase. *Proc. Natl. Acad. Sci. U.S.A.* 107, 18138–18142. doi: 10.1073/pnas.1010318107
- Biegel, E., Schmidt, S., González, J. M., and Müller, V. (2011). Biochemistry, evolution and physiological function of the Rnf complex, a novel ion-motive electron transport complex in prokaryotes. *Cell. Mol. Life Sci.* 68, 613–634. doi: 10.1007/s00018-010-0555-8
- Bird, J. T., Tague, E. D., Zinke, L., Schmidt, J. M., Steen, A. D., Reese, B., et al. (2019). Uncultured microbial phyla suggest mechanisms for multi-thousand-year subsistence in Baltic Sea Sediments. *mBio* 10:e02376-18. doi: 10.1128/mBio.02376-18
- Blasco, F., Dos Santos, J., Magalon, A., Frixion, C., Guigliarelli, B., Santini, C., et al. (1998). NarJ is a specific chaperone required for molybdenum cofactor assembly in nitrate reductase A of *Escherichia coli*. *Mol. Microbiol.* 28, 435–447. doi: 10.1046/j.1365-2958.1998.00795.x
- Blasco, F., Lobb, C., Giordano, G., Chippaux, M., and Bonnefoy, V. (1989). Nitrate reductase of *Escherichia coli*: completion of the nucleotide sequence of the nar operon and reassessment of the role of the  $\alpha$  and  $\beta$  subunits in iron binding and electron transfer. *Mol. Gen. Genet.* 218, 249–256. doi: 10.1007/BF00331275
- Blin, K., Wolf, T., Chevrette, M. G., Lu, X., Schwalen, C. J., Kautsar, S. A., et al. (2017). antiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* 45, W36–W41. doi: 10.1093/nar/gkx319
- Bouckaert, R., and Vaughan, T. G. (2019). BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* 15:e1006650. doi: 10.1371/journal.pcbi.1006650
- Bowers, R. M., Kyrpides, N. C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T. B. K., et al. (2017). Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat. Biotechnol.* 35, 725–731. doi: 10.1038/nbt.3893
- Boyd, E. S., Amenabar, M. J., Poudel, S., and Templeton, A. S. (2020). Bioenergetic constraints on the origin of autotrophic metabolism. *Philos. Trans. R. Soc. A* 378:2019015. doi: 10.1098/rsta.2019.0151
- Branscomb, E., and Russell, M. J. (2018). Frankenstein or a submarine alkaline vent: who is responsible for abiogenesis: Part 2: as life is now, so it must have been in the beginning. *Bioessays* 40:1700182. doi: 10.1002/bies.201700182
- Braus-Stromeier, S. A., Schnappauf, G., Braus, G. H., Gössner, A. S., and Drake, H. L. (1997). Carbonic anhydrase in *Acetobacterium woodii* and other acetogenic bacteria. *J. Bacteriol.* 179, 7197–7200. doi: 10.1128/jb.179.22.7197-7200.1997
- Brazelton, W. J., Thornton, C. N., Hyer, A., Twing, K. I., Longino, A. A., Lang, S. Q., et al. (2017). Metagenomic identification of active methanogens and methanotrophs in serpentine springs of the Voltri Massif, Italy. *PeerJ* 5:e2945. doi: 10.7717/peerj.2945
- Buckel, W., and Thauer, R. K. (2013). Energy conservation via electron bifurcating ferredoxin reduction and proton/Na<sup>+</sup> translocating ferredoxin oxidation. *Biochim. Biophys. Acta.* 1827, 94–113. doi: 10.1016/j.bbabi.2012.07.002
- Burstein, D., Sun, C. L., Brown, C. T., Sharon, I., Anantharaman, K., Probst, A. J., et al. (2016). Major bacterial lineages are essentially devoid of CRISPR-Cas viral defence systems. *Nat. Commun.* 7:10613. doi: 10.1038/ncomms10613
- Cabello, P., Roldán, M. D., and Moreno-Vivián, C. (2004). Nitrate reduction and the nitrogen cycle in *Archaea*. *Microbiology* 150, 3527–3546. doi: 10.1099/mic.0.27303-0
- Callahan, B. J., McMurdie, P. J., and Holmes, S. P. (2017). Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J.* 11, 2639–2643. doi: 10.1038/ismej.2017.119
- Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., and Holmes, S. P. (2016). DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13, 581–583. doi: 10.1038/nmeth.3869
- Capella-Gutierrez, S., Silla-Martinez, J. M., and Gabaldon, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348
- Chang, Y.-H., Cheng, T.-W., Lai, W.-J., Tsai, W.-Y., Sun, C.-H., Lin, L.-H., et al. (2012). Microbial methane cycling in a terrestrial mud volcano in eastern Taiwan: Microbial methane cycling in mud volcano. *Environ. Microbiol.* 14, 895–908. doi: 10.1111/j.1462-2920.2011.02658.x
- Cimermancic, P., Medema, M. H., Claesen, J., Kurita, K., Wieland Brown, L. C., Mavrommatis, K., et al. (2014). Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* 158, 412–421. doi: 10.1016/j.cell.2014.06.034
- Coker, J. A., and DasSarma, S. (2007). Genetic and transcriptomic analysis of transcription factor genes in the model halophilic Archaeon: coordinate action of TbpD and TfbA. *BMC Genet.* 8:61. doi: 10.1186/1471-2156-8-61
- Condie, K. C. (ed.). (2016). “Chapter 3 - Tectonic Settings,” in *Earth as an Evolving Planetary System* (Amsterdam: Elsevier), 43–88.
- Couvin, D., Bernheim, A., Toffano-Nioche, C., Touchon, M., Michalik, J., Néron, B., et al. (2018). CRISPRCasFinder, an update of CRISPRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res.* 46, W246–W251. doi: 10.1093/nar/gky425
- Crespo-Medina, M., Twing, K. I., Kubo, M. D. Y., Hoehler, T. M., Cardace, D., McCollom, T., et al. (2014). Insights into environmental controls on microbial communities in a continental serpentine aquifer using a microcosm-based approach. *Front. Microbiol.* 5:604. doi: 10.3389/fmicb.2014.00604
- DasSarma, S., and DasSarma, P. (2015). Gas vesicle nanoparticles for antigen display. *Vaccines* 3, 686–702. doi: 10.3390/vaccines3030686

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2020.01031/full#supplementary-material>

- de Kok, A., Hengeveld, A. F., Martin, A., and Westphal, A. H. (1998). The pyruvate dehydrogenase multi-enzyme complex from Gram-negative bacteria. *Biochim. Biophys. Acta* 1385, 353–366. doi: 10.1016/S0167-4838(98)00079-X
- Dibrova, D. V., Galperin, M. Y., Koonin, E. V., and Mulikidjanian, A. Y. (2015). Ancient systems of sodium/potassium homeostasis as predecessors of membrane bioenergetics. *Biochemistry* 80, 495–516. doi: 10.1134/S0006297915050016
- Dimroth, P. (1997). Primary sodium ion translocating enzymes. *Biochim. Biophys. Acta* 1318, 11–51. doi: 10.1016/S0005-2728(96)00127-2
- Dodsworth, J. A., Blainey, P. C., Murugapiran, S. K., Swingley, W. D., Ross, C. A., Tringe, S. G., et al. (2013). Single-cell and metagenomic analyses indicate a fermentative and saccharolytic lifestyle for members of the OP9 lineage. *Nat. Commun.* 4:1854. doi: 10.1038/ncomms2884
- Drake, H. L., Hu, S., and Wood, H. G. (1981). Purification of Five Components from *Clostridium thermoaceticum* Which Catalyze Synthesis of Acetate from Pyruvate and Methyltetrahydrofolate. *J. Biol. Chem.* 256, 11137–11144.
- Drummond, A. J., Nicholls, G. K., Rodrigo, A. G., and Solomon, W. (2002). Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics* 161, 1307–1320.
- Dubourdieu, M., and DeMoss, J. A. (1992). The *narJ* gene product is required for biogenesis of respiratory nitrate reductase in *Escherichia coli*. *J. Bacteriol.* 174, 867–872. doi: 10.1128/JB.174.3.867-872.1992
- Eddy, S. R. (2011). Accelerated profile HMM searches. *PLoS Comput. Biol.* 7:e1002195. doi: 10.1371/journal.pcbi.1002195
- Eichinger, V., Nussbaumer, T., Platzer, A., Jehl, M.-A., Arnold, R., and Rattei, T. (2016). EffectiveDB—updates and novel features for a better annotation of bacterial secreted proteins and Type III, IV, VI secretion systems. *Nucleic Acids Res.* 44, D669–D674. doi: 10.1093/nar/gkv1269
- Erb, T. J., Evans, B. S., Cho, K., Warlick, B. P., Sriram, J., Wood, B. M., et al. (2012). A RubisCO-like protein links SAM metabolism with isoprenoid biosynthesis. *Nat. Chem. Biol.* 8, 926–932. doi: 10.1038/nchembio.1087
- Erb, T. J., and Zarzycki, J. (2018). A short history of RubisCO: the rise and fall (?) of Nature's predominant CO<sub>2</sub> fixing enzyme. *Curr. Opin. Biotechnol.* 49, 100–107. doi: 10.1016/j.copbio.2017.07.017
- 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
- Fones, E. M., Colman, D. R., Kraus, E. A., Nothaft, D. B., Poudel, S., Rempfert, K. R., et al. (2019). Physiological adaptations to serpentinization in the Samail Ophiolite, Oman. *ISME J* 13, 1750–1762. doi: 10.1038/s41396-019-0391-2
- Fontaine, F. E., Peterson, W. H., McCoy, E., Johnson, M. J., and Ritter, G. J. (1942). A new type of glucose fermentation by *Clostridium thermoaceticum* n.sp. *J. Bacteriol.* 43, 701–715.
- Greening, C., Biswas, A., Carere, C. R., Jackson, C. J., Taylor, M. C., Stott, M. B., et al. (2016). Genomic and metagenomic surveys of hydrogenase distribution indicate H<sub>2</sub> is a widely utilised energy source for microbial growth and survival. *ISME J.* 10, 761–777. doi: 10.1038/ismej.2015.153
- Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi: 10.1093/bioinformatics/btt086
- Han, M. V., and Zmasek, C. M. (2009). phyloXML: XML for evolutionary biology and comparative genomics. *BMC Bioinformatics* 10:356. doi: 10.1186/1471-2105-10-356
- Harrison, K. J., de Crecy-Lagard, V., and Zallot, R. (2017). Gene Graphics: a genomic neighborhood data visualization web application. *Bioinformatics* 34, 1406–1408. doi: 10.1093/bioinformatics/btx793
- Hechler, T., and Pfeifer, F. (2009). Anaerobiosis inhibits gas vesicle formation in halophilic *Archaea*. *Mol. Microbiol.* 71, 132–145. doi: 10.1111/j.1365-2958.2008.06517.x
- Hemmann, J. L., Wagner, T., Shima, S., and Vorholt, J. A. (2019). Methylolofuran is a prosthetic group of the formyltransferase/hydrolase complex and shuttles one-carbon units between two active sites. *Proc. Natl. Acad. Sci. U.S.A.* 116, 25583–25590. doi: 10.1073/pnas.1911595116
- Hochheimer, A., Linder, D., Thauer, R. K., and Hedderich, R. (1996). The molybdenum formylmethanofuran dehydrogenase operon and the tungsten formylmethanofuran dehydrogenase operon from *Methanobacterium Thermoautotrophicum*. *Eur. J. Biochem.* 242, 156–162. doi: 10.1111/j.1432-1033.1996.0156r.x
- Hochheimer, A., Schmitz, R. A., Thauer, R. K., and Hedderich, R. (1995). The tungsten formylmethanofuran dehydrogenase from *Methanobacterium Thermoautotrophicum* contains sequence motifs characteristic for molybdopterins containing molybdopterin dinucleotide. *Eur. J. Biochem.* 234, 910–920. doi: 10.1111/j.1432-1033.1995.910\_a.x
- Hughenoltz, P., Pitulle, C., Hershberger, K. L., and Pace, N. R. (1998). Novel division level bacterial diversity in a yellowstone hot spring. *J. Bacteriol.* 180, 366–376.
- Hughenoltz, P., Skarszewski, A., and Parks, D. H. (2016). Genome-based microbial taxonomy coming of age. *Cold Spring Harb. Perspect. Biol.* 8:a018085. doi: 10.1101/cshperspect.a018085
- Hyatt, D., Chen, G.-L., LoCasio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119
- Ikeda-Ohtsubo, W., Strasser, J. F. H., Köhler, T., Mikaelyan, A., Gregor, I., McHardy, A. C., et al. (2016). 'Candidatus Adiatrix intracellularis', an endosymbiont of termite gut flagellates, is the first representative of a deep-branching clade of *Deltaproteobacteria* and a putative homoacetogen: a homoacetogenic *deltaproteobacterium* from termite guts. *Environ. Microbiol.* 18, 2548–2564. doi: 10.1111/1462-2920.13234
- Inoue, M., Nakamoto, I., Omae, K., Oguro, T., Ogata, H., Yoshida, T., et al. (2019). Structural and phylogenetic diversity of anaerobic carbon-monoxide dehydrogenases. *Front. Microbiol.* 9:3353. doi: 10.3389/fmicb.2018.03353
- Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* 9:5114. doi: 10.1038/s41467-018-07641-9
- Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031
- Kacar, B., Hanson-Smith, V., Adam, Z. R., and Boekelheide, N. (2017). Constraining the timing of the Great Oxidation Event within the RubisCO phylogenetic tree. *Geobiology* 15, 628–640. doi: 10.1111/gbi.12243
- Kameya, M., Kanbe, H., Igarashi, Y., Arai, H., and Ishii, M. (2017). Nitrate reductases in *Hydrogenobacter thermophilus* with evolutionarily ancient features: distinctive localization and electron transfer: nitrate reductases of *Hydrogenobacter thermophilus*. *Mol. Microbiol.* 106, 129–141. doi: 10.1111/mmi.13756
- Kaster, A.-K., Moll, J., Parey, K., and Thauer, R. K. (2011). Coupling of ferredoxin and heterodisulfide reduction via electron bifurcation in hydrogenotrophic methanogenic archaea. *Proc. Natl. Acad. Sci. U.S.A.* 108, 2981–2986. doi: 10.1073/pnas.1016761108
- Kato, S., Yanagawa, K., Sunamura, M., Takano, Y., Ishibashi, J., Kakegawa, T., et al. (2009). Abundance of *Zetaproteobacteria* within crustal fluids in back-arc hydrothermal fields of the Southern Mariana Trough. *Environ. Microbiol.* 11, 3210–3222. doi: 10.1111/j.1462-2920.2009.02031.x
- Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059–3066. doi: 10.1093/nar/gkf436
- Kawarabayasi, Y., HINO, Y., Horikawa, H., Yamazaki, S., Haikawa, Y., Jin-No, K., et al. (1999). Complete Genome Sequence of an Aerobic Hyper-thermophilic Crenarchaeon, *Aeropyrum pernix* K1. *DNA Res.* 6, 83–101. doi: 10.1093/dnares/6.2.83
- Kobayashi, I. (2001). Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. *Nucleic Acids Res.* 29, 3742–3756. doi: 10.1093/nar/29.18.3742
- Konstantinidis, K. T., and Tiedje, J. M. (2005a). Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci. U.S.A.* 102, 2567–2572. doi: 10.1073/pnas.0409727102
- Konstantinidis, K. T., and Tiedje, J. M. (2005b). Towards a genome-based taxonomy for prokaryotes. *J. Bacteriol.* 187, 6258–6264. doi: 10.1128/JB.187.18.6258-6264.2005
- Köpke, M., Held, C., Hujer, S., Liesegang, H., Wiezer, A., Wollherr, A., et al. (2010). *Clostridium ljungdahlii* represents a microbial production platform based on syngas. *Proc. Natl. Acad. Sci. U.S.A.* 107, 13087–13092. doi: 10.1073/pnas.1004716107

- Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., and Calcott, B. (2016). PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol.* 34, 772–773. doi: 10.1093/molbev/msw260
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Lee, M. D. (2019). GToTree: a user-friendly workflow for phylogenomics. *Bioinformatics* 35, 4162–4164. doi: 10.1093/bioinformatics/btz188
- Liu, X., and DeMoss, J. A. (1997). Characterization of NarJ, a system-specific chaperone required for nitrate reductase biogenesis in *Escherichia coli*. *J. Biol. Chem.* 272, 24266–24271. doi: 10.1074/jbc.272.39.24266
- Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964. doi: 10.1093/nar/25.5.955
- Lux, M., Krüger, J., Rinke, C., Maus, I., Schlüter, A., Woyke, T., et al. (2016). acdc – automated contamination detection and confidence estimation for single-cell genome data. *BMC Bioinformatics* 17:543. doi: 10.1186/s12859-016-1397-7
- Maden, B. E. H. (2000). Tetrahydrofolate and tetrahydromethanopterin compared: functionally distinct carriers in C1 metabolism. *Biochem. J.* 350, 609–629.
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 17, 10–12. doi: 10.14806/ej.17.1.200
- Martin, W., and Russell, M. J. (2006). On the origin of biochemistry at an alkaline hydrothermal vent. *Philos. Trans. R. Soc. B Biol. Sci.* 362, 1887–1926. doi: 10.1098/rstb.2006.1881
- Martinez-Espinosa, R. M., Dridge, E. J., Bonete, M. J., Butt, J. N., Butler, C. S., Sargent, F., et al. (2007). Look on the positive side! The orientation, identification and bioenergetics of “Archaeal” membrane-bound nitrate reductases. *FEMS Microbiol. Lett.* 276, 129–139. doi: 10.1111/j.1574-6968.2007.00887.x
- Matschiavelli, N., and Rother, M. (2015). Role of a putative tungsten-dependent formylmethanofuran dehydrogenase in *Methanosarcina acetivorans*. *Arch. Microbiol.* 197, 379–388. doi: 10.1007/s00203-014-1070-3
- Matsen, F. A., Kodner, R. B., and Armbrust, E. V. (2010). pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics* 11:538. doi: 10.1186/1471-2105-11-538
- Matter, J. M., and Kelemen, P. B. (2009). Permanent storage of carbon dioxide in geological reservoirs by mineral carbonation. *Nat. Geosci.* 2, 837–841. doi: 10.1038/ngeo683
- Mayhew, L. E., Ellison, E. T., McCollom, T. M., Trainor, T. P., and Templeton, A. S. (2013). Hydrogen generation from low-temperature water–rock reactions. *Nat. Geosci.* 6, 478–484. doi: 10.1038/ngeo1825
- McCollom, T. M., and Seewald, J. S. (2007). Abiotic synthesis of organic compounds in deep-sea hydrothermal environments. *Chem. Rev.* 107, 382–401. doi: 10.1021/cr0503660
- McMurdie, P. J., and Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 8:e61217. doi: 10.1371/journal.pone.0061217
- Menon, S., and Ragsdale, S. W. (1996). Evidence that carbon monoxide is an obligatory intermediate in anaerobic Acetyl-CoA synthesis. *Biochemistry* 35, 12119–12125. doi: 10.1021/bi961014d
- Menzel, P., Ng, K. L., and Krogh, A. (2016). Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat. Commun.* 7:11257. doi: 10.1038/ncomms11257
- Meyer-Dombard, D. R., Casar, C. P., Simon, A. G., Cardace, D., Schrenk, M. O., and Arcilla, C. A. (2018). Biofilm formation and potential for iron cycling in serpentinization-influenced groundwater of the Zambales and Coast Range ophiolites. *Extremophiles* 122, 407–431. doi: 10.1007/s00792-018-1005-z
- Miller, M. A., Pfeiffer, W., and Schwartz, T. (2010). “Creating the CIPRES Science Gateway for inference of large phylogenetic trees,” in *Proceedings of the 2010 Gateway Computing Environments Workshop (GCE)*, (New Orleans, LA: IEEE), 1–8. doi: 10.1109/GCE.2010.5676129
- Mock, J., Wang, S., Huang, H., Kahnt, J., and Thauer, R. K. (2014). Evidence for a Hexaheteromeric Methylene-tetrahydrofolate Reductase in *Moorella thermoacetica*. *J. Bacteriol.* 196, 3303–3314. doi: 10.1128/JB.01839-14
- Momper, L., Jungbluth, S. P., Lee, M. D., and Amend, J. P. (2017). Energy and carbon metabolisms in a deep terrestrial subsurface fluid microbial community. *ISME J.* 11, 2319–2333. doi: 10.1038/ismej.2017.94
- Morrill, P. L., Brazelton, W. J., Kohl, L., Rietze, A., Miles, S. M., Kavanagh, H., et al. (2014). Investigations of potential microbial methanogenic and carbon monoxide utilization pathways in ultra-basic reducing springs associated with present-day continental serpentinization: the Tablelands, NL, CAN. *Front. Microbiol.* 5:613. doi: 10.3389/fmicb.2014.00613
- Mulkidjanian, A. Y., Galperin, M. Y., Makarova, K. S., Wolf, Y. I., and Koonin, E. V. (2008). Evolutionary primacy of sodium bioenergetics. *Biol. Direct* 3:13. doi: 10.1186/1745-6150-3-13
- Musfeldt, M., Selig, M., and Schönheit, P. (1999). Acetyl Coenzyme A Synthetase (ADP Forming) from the Hyperthermophilic Archaeon *Pyrococcus furiosus*: identification, cloning, separate expression of the encoding genes, *acdAI* and *acdBI*, in *Escherichia coli*, and *in vitro* reconstitution of the active heterotetrameric enzyme from its recombinant subunits. *J. Bacteriol.* 181, 5885–5888. doi: 10.1128/JB.181.18.5885-5888.1999
- Naito, T., Kusano, K., and Kobayashi, I. (1995). Selfish behavior of restriction-modification systems. *Science* 267, 897–899. doi: 10.1126/science.7846533
- Nicolas, A. (2012). “27 - Ophiolites and oceanic lithosphere,” in *Regional Geology and Tectonics: Principles of Geologic Analysis*, eds D. G. Roberts, and A.W. Bally (Amsterdam: Elsevier), 820–835.
- O’Brien, W. E., Brewer, J. M., and Ljungdahl, L. G. (1973). Purification and characterization of thermostable 5,10-methylenetetrahydrofolate dehydrogenase from *Clostridium thermoaceticum*. *J. Biol. Chem.* 248, 403–408.
- Oesterheld, D. (1998). The structure and mechanism of the family of retinal proteins from halophilic archaea. *Curr. Opin. Struct. Biol.* 8, 489–500. doi: 10.1016/s0959-440x(98)80128-0
- Okland, I., Huang, S., Dahle, H., Thorseth, I. H., and Pedersen, R. B. (2012). Low temperature alteration of serpentinized ultramafic rock and implications for microbial life. *Chem. Geol.* 318–319, 75–87. doi: 10.1016/j.chemgeo.2012.05.015
- O’Leary, N. A., Wright, M. W., Brister, J. R., Ciuffo, S., Haddad, D., McVeigh, R., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 44, D733–D745. doi: 10.1093/nar/gkv1189
- Oren, A., Pri-El, N., Shapiro, O., and Siboni, N. (2006). Buoyancy studies in natural communities of square gas-vacuolate archaea in saltern crystallizer ponds. *Saline Syst.* 2:4. doi: 10.1186/1746-1448-2-4
- Palmieri, G., Cannio, R., Fiume, I., Rossi, M., and Pocsfalvi, G. (2009). Outside the unusual cell wall of the hyperthermophilic archaeon *Aeropyrum pernix* K1. *Mol. Cell. Proteomics* 8, 2570–2581. doi: 10.1074/mcp.M900012-MCP200
- Pander, B., Harris, G., Scott, D. J., Winzer, K., Köpke, M., Simpson, S. D., et al. (2019). The carbonic anhydrase of *Clostridium autoethanogenum* represents a new subclass of  $\beta$ -carbonic anhydrases. *Appl. Microbiol. Biotechnol.* 103, 7275–7286. doi: 10.1007/s00253-019-10015-w
- Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarshewski, A., Chaumeil, P.-A., et al. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* 36, 996–1004. doi: 10.1038/nbt.4229
- 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
- Peters, J. W., Schut, G. J., Boyd, E. S., Mulder, D. W., Shepard, E. M., Broderick, J. B., et al. (2015). [FeFe]- and [NiFe]-hydrogenase diversity, mechanism, and maturation. *Biochim. Biophys. Acta* 1853, 1350–1369. doi: 10.1016/j.bbamer.2014.11.021
- Pfeifer, F. (2012). Distribution, formation and regulation of gas vesicles. *Nat. Rev. Microbiol.* 10, 705–715. doi: 10.1038/nrmicro2834
- Philippot, L. (2002). Denitrifying genes in bacterial and Archaeal genomes. *Biochim. Biophys. Acta* 1577, 355–376. doi: 10.1016/S0167-4781(02)00420-7
- Poehlein, A., Cebulla, M., Ilg, M. M., Bengelsdorf, F. R., Schiel-Bengelsdorf, B., Whited, G., et al. (2015). The Complete Genome Sequence of *Clostridium acetium*: a Missing Link between Rnf- and Cytochrome-Containing Autotrophic Acetogens. *mBio* 6:e01168-15. doi: 10.1128/mBio.01168-15
- Pomper, B. K., Saurel, O., Milon, A., and Vorholt, J. A. (2002). Generation of formate by the formyltransferase/hydrolase complex (Fhc) from *Methylobacterium extorquens* AM1. *FEBS Lett.* 523, 133–137. doi: 10.1016/S0014-5793(02)02962-9



- Posewitz, M. C., King, P. W., Smolinski, S. L., Zhang, L., Seibert, M., and Ghirardi, M. L. (2004). Discovery of two novel radical S-Adenosylmethionine proteins required for the assembly of an active [Fe] Hydrogenase. *J. Biol. Chem.* 279, 25711–25720. doi: 10.1074/jbc.M403206200
- Poudel, S., Tokmina-Lukaszewska, M., Colman, D. R., Refai, M., Schut, G. J., King, P. W., et al. (2016). Unification of [FeFe]-hydrogenases into three structural and functional groups. *Biochim. Biophys. Acta* 1860, 1910–1921. doi: 10.1016/j.bbagen.2016.05.034
- Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490. doi: 10.1371/journal.pone.0009490
- Probst, A. J., Ladd, B., Jarett, J. K., Geller-McGrath, D. E., Sieber, C. M. K., Emerson, J. B., et al. (2018). Differential depth distribution of microbial function and putative symbionts through sediment-hosted aquifers in the deep terrestrial subsurface. *Nat. Microbiol.* 3, 328–336. doi: 10.1038/s41564-017-0098-y
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yara, P., et al. (2012). 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
- Quéméneur, M., Palvadeau, A., Postec, A., Monnin, C., Chavagnac, V., Ollivier, B., et al. (2015). Endolithic microbial communities in carbonate precipitates from serpentinite-hosted hyperalkaline springs of the Voltri Massif (Ligurian Alps, Northern Italy). *Environ. Sci. Pollut. Res.* 22, 13613–13624. doi: 10.1007/s11356-015-4113-7
- Ragsdale, S. W. (2003). Pyruvate ferredoxin oxidoreductase and its radical intermediate. *Chem. Rev.* 103, 2333–2346. doi: 10.1021/cr020423e
- Ragsdale, S. W. (2008). Enzymology of the wood–ljungdahl pathway of acetogenesis. *Ann. N. Y. Acad. Sci.* 1125, 129–136. doi: 10.1196/annals.1419.015
- Rambaut, A., Drummond, A. J., Xie, D., Baele, G., and Suchard, M. A. (2018). Posterior summarization in bayesian phylogenetics using tracer 1.7. *Syst. Biol.* 67, 901–904. doi: 10.1093/sysbio/syy032
- Ramos-Padrón, E., Bordenave, S., Lin, S., Bhaskar, I. M., Dong, X., Sensen, C. W., et al. (2011). Carbon and sulfur cycling by microbial communities in a gypsum-treated oil sands tailings pond. *Environ. Sci. Technol.* 45, 439–446. doi: 10.1021/es1028487
- Rempfert, K. R., Miller, H. M., Bompard, N., Nothhaft, D., Matter, J. M., Kelemen, P., et al. (2017). Geological and Geochemical Controls on Subsurface Microbial Life in the Samail Ophiolite, Oman. *Front. Microbiol.* 8:56. doi: 10.3389/fmicb.2017.00056
- Richardson, D. J., Berks, B. C., Russell, D. A., Spiro, S., and Taylor, C. J. (2001). Functional, biochemical and genetic diversity of prokaryotic nitrate reductases. *Cell. Mol. Life Sci.* 58, 165–178. doi: 10.1007/PL00000845
- Richter, M., and Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. U.S.A.* 106, 19126–19131. doi: 10.1073/pnas.0906412106
- Rinke, C., Lee, J., Nath, N., Goudeau, D., Thompson, B., Poulton, N., et al. (2014). Obtaining genomes from uncultivated environmental microorganisms using FACS-based single-cell genomics. *Nat. Protoc.* 9, 1038–1048. doi: 10.1038/nprot.2014.067
- Rodriguez-R, L. M., and Konstantinidis, K. T. (2014). Bypassing Cultivation To Identify Bacterial Species: culture-independent genomic approaches identify credibly distinct clusters, avoid cultivation bias, and provide true insights into microbial species. *Microbe Mag.* 9, 111–118. doi: 10.1128/microbe.9.111.1
- Rodriguez-R, L. M., and Konstantinidis, K. T. (2016). The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. *PeerJ* 4:e1900v1. doi: 10.7287/peerj.preprints.1900v1
- Russell, M. J., Hall, A. J., and Martin, W. (2010). Serpentinization as a source of energy at the origin of life. *Geobiology* 8, 355–371. doi: 10.1111/j.1472-4669.2010.00249.x
- Sato, T., Yoshiya, K., and Maruyama, S. (2019). History of the Hadean “Living Microfossil” OD1 and Ultra-reducing Environments. *J. Geogr.* 128, 571–596. doi: 10.5026/jgeography.128.571
- Schuchmann, K., and Müller, V. (2014). Autotrophy at the thermodynamic limit of life: a model for energy conservation in acetogenic bacteria. *Nat. Rev. Microbiol.* 12, 809–821. doi: 10.1038/nrmicro3365
- Schuchmann, K., and Müller, V. (2016). Energetics and application of heterotrophy in Acetogenic Bacteria. *Appl. Environ. Microbiol.* 82, 4056–4069. doi: 10.1128/AEM.00882-16
- Schut, G. J., and Adams, M. W. W. (2009). The Iron-Hydrogenase of *Thermotoga maritima* Utilizes Ferredoxin and NADH Synergistically: a New Perspective on Anaerobic Hydrogen Production. *J. Bacteriol.* 191, 4451–4457. doi: 10.1128/JB.01582-08
- Scott, J. W., and Rasche, M. E. (2002). Purification, Overproduction, and Partial Characterization of  $\beta$ -RFAP Synthase, a Key Enzyme in the Methanopterin Biosynthesis Pathway. *J. Bacteriol.* 184, 4442–4448. doi: 10.1128/JB.184.16.4442-4448.2002
- Seewald, J. S., Zolotov, M. Y., and McCollom, T. (2006). Experimental investigation of single carbon compounds under hydrothermal conditions. *Geochim. Cosmochim. Acta* 70, 446–460. doi: 10.1016/j.gca.2005.09.002
- Shin, J., Song, Y., Jeong, Y., and Cho, B.-K. (2016). Analysis of the core genome and pan-genome of autotrophic Acetogenic Bacteria. *Front. Microbiol.* 7:1531. doi: 10.3389/fmicb.2016.01531
- Smith, K. S., and Ferry, J. G. (2000). Prokaryotic carbonic anhydrases. *FEMS Microbiol. Rev.* 24, 335–366. doi: 10.1111/j.1574-6976.2000.tb00546.x
- 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
- Stamatakis, A., Hoover, P., and Rougemont, J. (2008). A rapid bootstrap algorithm for the RAXML web servers. *Syst. Biol.* 57, 758–771. doi: 10.1080/10635150802429642
- Suda, K., Gilbert, A., Yamada, K., Yoshida, N., and Ueno, Y. (2017). Compound- and position-specific carbon isotopic signatures of abiogenic hydrocarbons from on-land serpentinite-hosted Hakuba Happo hot spring in Japan. *Geochim Cosmochim. Acta* 206, 201–215. doi: 10.1016/j.gca.2017.03.008
- Suda, K., Ueno, Y., Yoshizaki, M., Nakamura, H., Kurokawa, K., Nishiyama, E., et al. (2014). Origin of methane in serpentinite-hosted hydrothermal systems: The CH<sub>4</sub>-H<sub>2</sub>-H<sub>2</sub>O hydrogen isotope systematics of the Hakuba Happo hot spring. *Earth Planet. Sci. Lett.* 386, 112–125. doi: 10.1016/j.epsl.2013.11.001
- Suzuki, S., Ishii, S., Wu, A., Cheung, A., Tenney, A., Wanger, G., et al. (2013). Microbial diversity in The Cedars, an ultrabasic, ultrareducing, and low salinity serpentinitizing ecosystem. *Proc. Natl. Acad. Sci. U.S.A.* 110, 15336–15341. doi: 10.1073/pnas.1302426110
- Suzuki, S., Nealson, K. H., and Ishii, S. (2018). Genomic and in-situ transcriptomic characterization of the candidate phylum NPL-UPL2 from highly alkaline highly reducing serpentinitized groundwater. *Front. Microbiol.* 9:3141. doi: 10.3389/fmicb.2018.03141
- Takami, H. (2014). “New method for comparative functional genomics and metagenomics using KEGG MODULE,” in *Encyclopedia of Metagenomics*, ed. K. E. Nelson (New York, NY: Springer), 1–15. doi: 10.1007/978-1-4614-6418-1\_183-3
- Tashiro, Y., Monson, R. E., Ramsay, J. P., and Salmond, G. P. C. (2016). Molecular genetic and physical analysis of gas vesicles in buoyant enterobacteria. *Environ. Microbiol.* 18, 1264–1276. doi: 10.1111/1462-2920.13203
- Tiago, I., and Veríssimo, A. (2013). Microbial and functional diversity of a subterrestrial high pH groundwater associated to serpentinization: microbial and functional diversity of subsurface aquifer. *Environ. Microbiol.* 15, 1687–1706. doi: 10.1111/1462-2920.12034
- Twing, K. I., Brazelton, W. J., Kubo, M. D. Y., Hyer, A. J., Cardace, D., Hoehler, T. M., et al. (2017). Serpentinization-influenced groundwater harbors extremely low diversity microbial communities adapted to high pH. *Front. Microbiol.* 8:308. doi: 10.3389/fmicb.2017.00308
- Vignais, P. M., and Billoud, B. (2007). Occurrence, classification, and biological function of hydrogenases: an overview. *Chem. Rev.* 107, 4206–4272. doi: 10.1021/cr050196r
- Vignais, P. M., Billoud, B., and Meyer, J. (2001). Classification and phylogeny of hydrogenases. *FEMS Microbiol. Rev.* 25, 455–501. doi: 10.1111/j.1574-6976.2001.tb00587.x
- von Ballmoos, C., and Dimroth, P. (2007). Two distinct proton binding sites in the ATP synthase family. *Biochemistry* 46, 11800–11809. doi: 10.1021/bi701083v
- Wagner, T., Ermler, U., and Shima, S. (2016). The methanogenic CO<sub>2</sub> reducing- and fixing enzyme is bifunctional and contains 46 [4Fe-4S] clusters. *Science* 354, 114–117. doi: 10.1126/science.aaf9284
- Wagner, T., Koch, J., Ermler, U., and Shima, S. (2017). Methanogenic heterodisulfide reductase (HdrABC-MvhAGD) uses two noncubane [4Fe-4S] clusters for reduction. *Science* 357, 699–703. doi: 10.1126/science.aan0425
- Walsby, A. E. (1994). Gas vesicles. *Microbiol. Rev.* 58, 94–144.



- Wiechmann, A., Ciurus, S., Oswald, F., Seiler, V. N., and Müller, V. (2020). It does not always take two to tango: "Syntrophy" via hydrogen cycling in one bacterial cell. *ISME J.* doi: 10.1038/s41396-020-0627-1 [Epub ahead of print].
- Williams, K. P. (2002). Integration sites for genetic elements in prokaryotic tRNA and tmRNA genes: sublocation preference of integrase subfamilies. *Nucleic Acids Res.* 30, 866–875. doi: 10.1093/nar/30.4.866
- Wright, A. V., Nuñez, J. K., and Doudna, J. A. (2016). Biology and applications of CRISPR systems: harnessing nature's toolbox for genome engineering. *Cell* 164, 29–44. doi: 10.1016/j.cell.2015.12.035
- Xu, H., Aurora, R., Rose, G. D., and White, R. H. (1999). Identifying two ancient enzymes in Archaea using predicted secondary structure alignment. *Nat. Struct. Biol.* 6, 750–754. doi: 10.1038/11525
- Xu, Y., and Zhao, F. (2018). Single-cell metagenomics: challenges and applications. *Protein Cell* 9, 501–510. doi: 10.1007/s13238-018-0544-5
- Yilmaz, P., Parfrey, L. W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., et al. (2014). The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Res.* 42, D643–D648. doi: 10.1093/nar/gkt1209
- Yoshimatsu, K., Iwasaki, T., and Fujiwara, T. (2002). Sequence and electron paramagnetic resonance analyses of nitrate reductase NarGH from a denitrifying halophilic euryarchaeote *Haloarcula marismortui*. *FEBS Lett.* 516, 145–150. doi: 10.1016/S0014-5793(02)02524-3
- Yoshimatsu, K., Sakurai, T., and Fujiwara, T. (2000). Purification and characterization of dissimilatory nitrate reductase from a denitrifying halophilic archaeon, *Haloarcula marismortui*. *FEBS Lett.* 470, 216–220. doi: 10.1016/s0014-5793(00)01321-1
- Zhang, X., Carter, M. S., Vetting, M. W., San Francisco, B., Zhao, S., Al-Obaidi, N. F., et al. (2016). Assignment of function to a domain of unknown function: DUF1537 is a new kinase family in catabolic pathways for acid sugars. *PNAS* 113, E4161–E4169. doi: 10.1073/pnas.1605546113
- Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J., and Wishart, D. S. (2011). PHAST: a fast phage search tool. *Nucleic Acids Res.* 39, W347–W352. doi: 10.1093/nar/gkr485

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

Copyright © 2020 Merino, Kawai, Boyd, Colman, McGlynn, Neelson, Kurokawa and Hongoh. 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.