Title

Construction of Multiple Metagenome Assembled Genomes Containing Carbon Monoxide Dehydrogenases from Anaerobic Carbon Monoxide Enrichment Cultures

Shiho Nishida[a], Kimiho Omae[a, b], Masao Inoue[c, d], Yoshihiko Sako[a], Ryoma Kamikawa[a] and Takashi Yoshida[a]

[a]Laboratory of Marine Microbiology, Graduate School of Agriculture, Kyoto University, Kitashirakawa Oiwake-cho, Sakyo-ku, Kyoto 606-8502, Japan

[b]Department of Integrated Biosciences, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Chiba 277-0882, Japan

[c]R-GIRO, Ritsumeikan University, 1-1-1 Nojihigashi, Kusatsu, Shiga 525-8577, Japan

[d]College of Life Sciences, Ritsumeikan University, 1-1-1 Nojihigashi, Kusatsu, Shiga 525-8577, Japan

Correspondence

Takashi Yoshida, Laboratory of Marine Microbiology, Graduate School of Agriculture, Kyoto University, Kitashirakawa Oiwake-cho, Sakyo-ku, Kyoto 606-8502, Japan.

E-mail: yoshida.takashi.7a@kyoto-u.ac.jp

Key Words

Hydrogenogenic carbon monoxide utilizer, Carbon monoxide utilizer, Carbon monoxide dehydrogenase, Metagenome assembled genomes

Abstract

Despite its toxicity to many organisms, including most prokaryotes, carbon monoxide (CO) is utilized by some aerobic and anaerobic prokaryotes. Hydrogenogenic CO utilizers employ carbon monoxide dehydrogenase (CODH) and energy-converting hydrogenase (ECH) to oxidize CO and reduce protons to produce H_2. Those prokaryotes constitute a rare biosphere and are difficult to detect even with PCR amplification and with metagenomic analyses. In this study, anaerobic CO-enrichment cultures followed by construction of metagenome assembled genomes (MAGs) detected high-quality MAGs from potential hydrogenogenic CO utilizers. Of 32 MAGs constructed, 5 were potential CO utilizer harboring CODH genes. Of the five MAGs, two were classified into the genus *Thermolithobacter* on the basis of 16S rRNA sequence identity, related to *Carboxydocella tharmautotrophica* 41, with an average nucleotide identity (ANI) of approximately 72%. Additionally, two were related to *Geoglobus acetivorans* with ANI values ranging from 75%-77% to *G. acetivorans* SBH6, and one MAG was identified as *Desulfotomaculum kuznetsovii* with an ANI > 96% to *D. kuznetsovii* DSM 6115. The two *Thermolithobacter* MAGs identified in this study contained CODH-ECH gene clusters, and were therefore identified as potential hydrogenogenic CO utilizers. However, these MAGs harbored three CODH gene clusters that showed distinct physiological functions in addition to CODH-ECH gene clusters. In total, the five potential CO utilizer MAGs contained sixteen CODH genes. Among those CODHs, four sets did not cluster with any known CODH protein sequences (with an identity of > 90%), and the CODH database was expanded.

Introduction

Carbon monoxide (CO) has a high affinity for iron and copper, which are found in the active sites of many enzymes (Omaye 2002). Because of its ability to bind to metalloenzymes and inhibit the electron transport chain, CO is toxic to many organisms, including microbes (Fukuyama *et al.*, 2020; Techtmann *et al.*, 2009). Despite this, microbes that use CO for growth have been reported (King and Weber 2007; Sokolova *et al.*, 2009; Techtmann *et al.*, 2009; Oelgeschläger and Rother 2008). These microbes have the ability to oxidize CO aerobically or anaerobically using carbon monoxide dehydrogenase (CODH) and transport electrons to various terminal electron acceptors (King and Weber 2007; Oelgeschläger and Rother 2008; Svetlitchnyi *et al.*, 2001). Anaerobic CO-utilizing microorganisms use anaerobic-type CODHs to oxidize CO, and these reactions are coupled with the reduction of various electron-acceptors such as sulfate, ferric iron (Fe(III)), carbon dioxide, and protons (Oelgeschläger and Rother 2008). Anaerobic-type CODHs are categorized as either, CooS, which are predominantly found in bacteria, or Cdh, nearly all of which are found in archaea (Techtmann *et al.*, 2012). As anaerobic-type CODHs contain Ni at their active sites, they are referred to as Ni-CODHs (Dobbek *et al.*, 2001; Inoue *et al.*, 2019; Merrouch *et al.*, 2018). While Ni-CODHs are generally divided into eight clades (clades A-H) based on their phylogeny (Inoue *et al.*, 2022), the taxonomy of bacteria and archaea may not always align with the classification of their Ni-CODHs (Inoue *et al.*, 2019; Techtmann *et al.*, 2012). In addition to this, Ni-CODHs within the same clade often have different physiological roles, which are inferred from the genomic context, i.e., genes located upstream and downstream of the Ni-CODH (Inoue *et al.*, 2019; Techtmann *et al.*, 2012). Physiological roles of Ni-CODH were divided into six groups: carbon fixation (Wood-Ljungdahl pathway; WLP), energy conservation (energy converting hydrogenase; ECH), electron transport (flavin adenine dinucleotide-dependent NAD(P) oxidoreductase; FNOR), CooF, ABC transporter and Metallochaperone (Inoue *et al.*, 2022).

Especially, anaerobic hydrogenogenic CO utilizers have been focused (Fukuyama *et al.*, 2020; Omae

et al., 2019; 2021; Yoneda et al., 2013; Yoneda et al., 2015). These organisms perform CO oxidation through Ni-CODH and produce hydrogen through ECH via the water gas shift reaction (Schoelmerich and Müller 2019; Techtmann *et al.*, 2009). ECH transports protons from the cytoplasm to periplasm across the cell membrane, resulting in the conservation of energy through ATP formation via a proton gradient (Schoelmerich and Müller 2019; Svetlitchnyi *et al.*, 2001; Wu *et al.*, 2005). Genes encoding Ni-CODH, ECH, and other related genes are located adjacent to one another, collectively forming Ni-CODH-ECH gene clusters (Fox *et al.*, 1996; Techtmann *et al.*, 2012; Omae *et al.*, 2019). To date, approximately 40 hydrogenogenic CO utilizer strains have been isolated from a diverse range of environments (Imaura *et al.*, 2023; Fukuyama *et al.*, 2020). These organisms belong to the phyla Firmicutes, Proteobacteria, and Dictyoglomi in the Bacteria domain (Imaura *et al.*, 2023; Fukuyama *et al.*, 2020), or to the phyla Euryarchaeota and Crenarchaeota in the Archaea domain (Fukuyama *et al.*, 2020).

Because hydrogenogenic CO utilizers consume the CO which is toxic to other microbes, while supplying them with H_2 as an energy source, they are thought to play important environmental roles (Techtmann *et al.*, 2009). Hydrogenogenic CO utilizers are difficult to study, since, although they are widely distributed, their relative abundance in the environment is very low (Omae *et al.*, 2019), implying that they constitute a rare biosphere (Fukuyama *et al.*, 2020). This is supported by the relatively small number of genomes bearing Ni-CODH-ECH gene clusters detected even in an in-depth survey of the available metagenomic databases (Inoue *et al.*, 2022). Since the nucleotide sequences encoding Ni-CODHs within the Ni-CODH-ECH gene clusters are diverse (Inoue *et al.*, 2019; Techtmann *et al.*, 2012), it is difficult to design a universal primer set for the broad-spectrum detection of these Ni-CODH sequences in environmental DNA using PCR amplification. Although a series of primer sets for Ni-CODH genes have been designed to detect hydrogenogenic CO utilizers in hydrothermal environments and anaerobic CO enrichment cultures (Omae *et al.*, 2021), these primer sets are not able to amplify Ni-CODH genes of all phyla, and those of many non-hydrogenogenic CO utilizers that appear in anaerobic CO enrichment

106 remained undetected. Further, detection of Ni-CODH genes with the primer sets does not provide any
107 additional genomic and physiological information on potential CO utilizers (pCO utilizers), of which CO
108 utilization are predicted only from the presence of Ni-CODH genes. In this study, metagenomic analysis of
109 the anaerobic CO enrichment culture was performed to overcome the above issues. The integration of the
110 anaerobic CO enrichment culture and metagenomic analyses constructed high-quality metagenome
111 assembled genomes (MAGs) from novel hydrogenogenic pCO utilizers and other pCO utilizers, regardless
112 of their low abundance in environments. The variation of Ni-CODH-ECH gene cluster in the detected
113 hydrogenogenic pCO utilizer MAGs suggest their physiological versatility of CO utilization. The findings
114 in this study that expanded the knowledge on the diversity of CO utilizers and the Ni-CODH sequence
115 database would provide a basis for cultivation and isolation of such rare, but novel CO utilizers from various
116 environmental samples in future studies.
117
118 **Materials and Methods**
119 **Anaerobic enrichment of CO utilizers and DNA extraction from cultured samples**
120 The anaerobic enrichment culture was established and DNA samples were prepared as described
121 previously (Omae *et al.*, 2021). A sediment sample was collected from a hydrothermal environment at
122 Jiunji-Onsen (JI) (60.1 ℃; pH 7.7; Oxidation-Reduction potential (ORP) 259 mV) in the Shizuoka
123 prefecture (34°38′54′′N., 138°52′00′′E), Japan in January 2015. Approximately 5 mL of sediment sample
124 and pore water from JI was dispensed into a 64 mL glass vial and duplicates were incubated under a mixture
125 of 10% v/v CO and N_2 at 65 ℃ for five days. DNA was extracted from each 0.5 g enrichment sample using
126 an Extrap Soil DNA Kit Plus V 2 (Nippon Steel and SUMIKIN Electronics, Tokyo, Japan).
127 **Metagenome sequencing**
128 DNA from two anaerobic enrichment cultures, JI_enriched_1 and JI_enriched_2, was processed for
129 shotgun metagenome sequencing using MiSeq. The DNA concentrations of the two samples were

determined using Qubit™ dsDNA HS Assay Kit (Thermo Fisher Scientific, Waltham, MA, USA). The Nextera XT DNA Library Preparation Kit (Illumina, San Diego, CA, USA) was used for library preparation according to the manufacturer's instructions. DNA libraries were quantified by a Qubit™ dsDNA HS Assay Kit (Thermo Fisher Scientific, Waltham, MA, USA) after library preparation and before sequencing. The fragment sizes of the PCR products were determined using an Agilent High Sensitivity DNA Kit (Agilent Technologies, Waldbronn, Germany) and an Agilent 2100 Bioanalyzer (Agilent Technologies), according to the manufacturer's instructions. The prepared DNA libraries were sequenced using the MiSeq Reagent Kit v3 (600 cycles, paired-end) (Illumina, San Diego, CA, USA).

Construction of metagenome assembled genomes (MAGs)

All reads were subjected to quality filtering using fastp V 0.20.1 (Chen *et al.,* 2018) and the contigs were assembled using MEGAHIT V 1.2.9 (Li *et al.*, 2016). After assembly, filtered reads were mapped to contigs using the Burrows-Wheeler Aligner (BWA) V 0.7.17 (Li 2013), and the contigs were binned using MetaBAT2 V 2.12.1 (Kang *et al.,* 2019). The quality of the genome bins (MAGs), was checked using CheckM V 1.13 (Parks *et al.*, 2015). The number of contig in each MAGs, total sequence length, and the N50 value were estimated by SeqKit V 0.13.2 (Shen *et al.*, 2016). MAGs were separated into two types: high-quality MAGs (completeness>90% and contamination<5%) and medium-quality MAGs (completeness≥50% and contamination<10%) according to criteria defined in Parks et al. 2020. MAGs that did not meet either of these criteria were excluded from the downstream analysis.

Identification of pCO utilizer MAGs

Open reading frames (ORFs) were predicted using Prodigal V 2.6.3 (Hyatt *et al.*, 2010) and protein sequences were entered into the database using blast+ V 2.10.1 (Camacho *et al.*, 2009). To identify CO utilizer genomes, the database was searched for Ni-CODH protein sequences. In this study, *Carboxydothermus hydrogenoformans* CooSII (WP_011343033.1) was used as the CooS-type CODH query sequence and *Methanosarcina barkeri* Acetyl-CoA synthaseα subunit (CdhA; WP_011305243.1) was

used as the Cdh-type CODH query sequence, as described by Inoue *et al.* (2019). Furthermore, to identify hydrogenogenic pCO utilizer MAGs that contain phylogenetically novel Ni-CODH sequences, pCO utilizer MAGs were searched for 25 Ni-CODH sequences and two parts of the Ni-CODH-ECH gene cluster sequence found by Omae *et al.* (2021). DFAST V 1.2.18 (Tanizawa *et al.*, 2018) was used to determine the G+C content and to identify coding domain sequences (CDSs), rRNAs, tRNAs, and clustered regularly interspaced short palindromic repeats (CRISPR) sequences within pCO utilizer MAGs.

Phylogenetic analysis of CO utilizer MAGs

16S rRNA gene was present in two hydrogenogenic pCO utilizer MAGs and one pCO utilizer MAG, but was not present in the remaining two pCO utilizer MAGs. Therefore, a phylogenetic analysis of MAGs based on both their 16S rRNA gene and genomes was performed. MEGA-X: Molecular Evolutionary Genetics Analysis version 10 (Tamura *et al.*, 2021) was used to construct a maximum likelihood (ML) phylogenetic tree with closely related species and other CO utilizers. When the 16S rRNA phylogeny of pCO utilizer MAGs was investigated, GTR+G+I model was adopted in all the cases.

Next, to analyze the phylogeny of MAGs, related genomes were retrieved from the National Center for Biotechnology Information (NCBI) (February 2, 2022) and Genome Taxonomy Database (GTDB) (March 5, 2022) (Parks *et al.*, 2018; 2020). Genome-based phylogenetic analysis was performed using GTDB-tk V1.7.0 (Chaumeil *et al.*, 2020; Eddy 2011; Harris *et al.*, 2020; Hyatt *et al.*, 2010; Jain *et al.*, 2018; Matsen *et al.*, 2010; Ondov *et al.*, 2016; Price *et al.*, 2010; Sukumaran and Holder 2010) and phylogenetic trees were drawn using iTOL V 6.7 (Letunic and Bork 2021). The average nucleotide identity (ANI) was calculated using ANIb in pyani V 0.2.10 (Pritchard *et al.*, 2016).

Prediction of physiological functions of Ni-CODHs encoded in MAGs

The genomic context of each CODH gene was determined to predict the function of CODHs in the MAGs. Genes within 100 bp-upstream or downstream of each CODH gene, except for k141_166150_85 and k141_166150_86 of JI_enriched_1_bin24, were manually extracted and annotated by eggNOG-mapper

V 1.0.3 (Huerta-Cepas *et al.*, 2017). When genes could not be annotated using this method, they were re-annotated with a blastp using the NCBI non-redundant protein sequence database.

Phylogenetic analysis of Ni-CODHs involved in pCO utilizer MAGs

In this study, 3,164 and 5,430 representative Ni-CODH protein sequences were retrieved from the public databases RefSeq/GenBank and MGnify respectively and used as reference sequences. These sequences were combined with the Ni-CODH sequences of the MAGs and clustered using USEARCH (Edgar 2010) with a threshold of < 90% identity. Four CODH protein sequences of pCO utilizer MAGs (J1_bin9_k141_170757_156, J1_bin24_k141_118531_26, J1_bin24_k141_190039_49, and j2_bin23_k141_18619_16) were selected as representative sequences. To construct the CODH phylogenetic tree, these four sequences and the 2,462 centroids used by Inoue et al. (2022) were merged and aligned using the E-INS-I method in MAFFT V7.471 (Katoh and Standley 2013). After alignment, the sequences were trimmed using trimAl V1.4.1 (Capella-Gutiérrez *et al.*, 2009), and a phylogenetic tree was constructed using IQ-TREE (Nguyen *et al.*, 2015).

To construct a separate phylogenetic tree of each clade in which Ni-CODHs of pCO utilizer MAGs were distributed, Ni-CODH protein sequences of clades B, D, and F were retrieved from 2,462 representative sequences, and alignments and phylogenetic tree construction were performed as described above.

Data availability

The raw reads for paired-end sequencing of the JI_enriched_1 and JI_enriched_2 samples were deposited in DRA under BioProject PRJDB15732. Accession numbers of JI_enriched_1 and JI_enriched_2 were DRR461309 and DRR461310, respectively. The genome sequence of 5 MAGs, JI_enriched_1_bin2, JI_enriched_1_bin9, JI_enriched_1_bin24, JI_enriched_2_bin9 and JIenriched_2_bin23, were deposited in figshare under DOI: 10.6084/m9.figshare.22547407.

Results and Discussion

Genomic properties of pCO utilizer MAGs

Shotgun metagenome sequencing produced 27,532,830 raw paired-end reads for JI_enriched_1, and 26,259,718 for JI_enriched_2. After filtering for quality, 15,534,972, and 16,503,906 paired-end reads remained, respectively (Table 1). Using assembly and binning, nine high-quality and six medium-quality MAGs were generated from JI_enriched_1 (Table 1), while eight high-quality and nine medium-quality MAGs were obtained from JI_enriched_2 (Table 1). From the 32 MAGs, 5 MAGs (JI_enriched_1_bin2, JI_enriched_1_bin9, JI_enriched_1_bin24, JI_enriched_2_bin9, JI_enriched_2_bin23) contained Ni-CODH gene sequences and were identified as pCO utilizer MAGs (Table 2). Completeness and contamination percentages for all pCO utilizer MAGs provided in Table 2 demonstrated that they were of high- or medium-quality according to Parks *et al.* (2020).

The number of Ni-CODH genes in each pCO utilizer MAG can be found in Table 2. Two of the pCO utilizer MAGs, JI_enriched_1_bin9, and JI_enriched_2_bin9, contained one CooS-type Ni-CODH and one Cdh-type CODH gene (Table 2). Because of the presence of Cdh-type Ni-CODH genes, JI_enriched_1_bin9 and JI_enriched_2_bin9 could be presumed to contain archaeal genomes.

According to the previous amplicon analyses for 16S rRNA genes and Ni-CODH-ECH gene clusters as well as metagenomic database survey (Omae *et al.*, 2019; Fukuyama *et al.*, 2020; Omae *et al.*, 2021; Inoue *et al.*, 2022), hydrogenogenic CO utilizers constitute a rare biosphere. Thus, the high-quality or medium-quality MAGs derived from pCO utilizers (Table 2) might imply usefulness of the anaerobic CO enrichment followed by the metagenomic analyses for unveiling the diversity of CO utilizers and their CO metabolisms from diverse environments where CO utilizers less abundantly distribute.

Identification of hydrogenogenic pCO utilizer MAGs

Hydrogenogenic CO utilizers usually have a Ni-CODH-ECH gene cluster and are thought to conserve

energy using CO oxidation and H_2 production (Fox *et al.*, 1996; Inoue *et al.*, 2019; Omae *et al.*, 2019; Techtmann *et al.*, 2012). Among the five pCO utilizers, JI_enriched_1_bin24 and JI_enriched_2_bin23 contained Ni-CODH-ECH gene sequences. These two sequences showed 100% identity with the Ni-CODH sequences identified by Omae *et al.* (2021). The nucleotide sequence identities between the partial Ni-CODH-ECH gene sequences in the previous study (Omae *et al.*, 2021) and the same regions of JI_enriched_1_bin24 and JI_enriched_2_bin23 were 99.97% and 100%, respectively. Therefore, JI_enriched_1_bin24 and JI_enriched_2_bin23 were identified as hydrogenogenic pCO utilizers. Additionally, Table 2 shows that almost complete draft genomes of the hydrogenogenic pCO utilizers were acquired in this study. JI_enriched_1_bin24 and JI_enriched_2_bin23 were found to contain five and four CooS-type CODH genes, respectively (Table 2). These multiple Ni-CODH genes suggested that the two MAGs were highly likely derived from prokaryotes that are physiologically versatile in CO utilization (Fukuyama *et al.*, 2020).

Phylogenetic analysis of pCO utilizer MAGs

The phylogenetic analyses of pCO utilizers based on 16S rRNA sequences was performed. JI_enriched_1_bin24, JI_enriched_2_bin23, and JI_enriched_1_bin9 all contained one 16S rRNA gene, whereas JI_enriched_1_bin2 and JI_enriched_2_bin9 did not contain any. Phylogenetic analysis revealed that JI_enriched_1_bin24 and JI_enriched_2_bin23 shared sequence homology with *Thermolithobacter ferrireducens* strain JW/KA-2 and *T. carboxydivorans* strain R1 belonging to the phylum Firmicutes (Fig. 1A). In contrast to strains of *T. ferrireducens*, *T. carboxydivorans* R1 is known to exhibit its CO utilization (Sokolova *et al.*, 2007). JI_enriched_1_bin24 formed a monophyletic group with *T. ferrireducens* JW/KA-2 (a), *T. ferrireducens* JW/KA-2 (b) and *T. carboxydivorans* R1. JI_enriched_1_bin24 shared a 16S rRNA gene sequence identity of over 99% with *T. ferrireducens* strain JW/KA-2 (a and b) and 98.96% with *T. carboxydivorans* strain R1. JI_enriched_2_bin23 formed a monophyletic group with *T. ferrireducens* strain JW/JH-Fiji-2. The sequence identity between JI_enriched_2_bin23 and that of *T. ferrireducens* strain

JW/JH-Fiji-2 was over 99%, while it shared approximately 98.61% identity with *T. ferrireducens* strain JW/KA-2 (a), 99.00% identity with *T. ferrireducens* strain JW/KA-2 (b), and 98.90% with *T. carboxydivorans* strain R1. *T. ferrireducens* strains JW/KA-2 and JW/JH-Fiji-2 and *T. carboxydivorans* strain R1 have been isolated previously (Sokolova *et al.*, 2007), however, the complete genomes of these organisms have not been sequenced. The 16S rRNA gene sequence of JI_enriched_1_bin9 showed 98.87 % identity with *Geoglobus acetivorans* strain SBH6. Therefore, it was concluded that JI_enriched_1_bin9 was related to *G. acetivorans* strain SBH6, which belongs to the phylum Halobacteriota (Fig. 1B).

The phylogeny of the pCO utilizer MAGs based on their genome sequences was also analyzed in this study. These analyses revealed that the hydrogenogenic pCO utilizers were closest to *C. thermautotrophica* and *C. hydrogenoformans* in genome-based phylogeny (Fig. 2A). The ANI between JI_enriched_1_bin24 and *C. thermautotrophica* 41 was approximately 71.7%, whereas ANI between JI_enriched_2_bin23 and *C. hydrogenoformans* Z-2901 was approximately 71.6%. Furthermore, the ANI between JI_enriched_1_bin24 and JI_enriched_2_bin23 was > 96%. The genomic properties of these pCO utilizer MAGs were compared to those of the phylogenetically related microbes (Table 3). Differences in the total length of genome, G+C content, the number of coding sequences (CDSs) and rRNA and tRNA were found (Table 3). The number of CDSs in JI_enriched_1_bin24 and JI_enriched_2_bin23 was lower than *C. thermautotrophica* 41 and *C. hydrogenoformans* Z-2901, respectively. The total genome lengths of both hydrogenogenic pCO utilizer genomes (ca. 2.0 Mbp) were shorter than their respective related species (Table 3). Furthermore, the number of rRNA and tRNA sequence of hydrogenogenic pCO utilizer MAGs was lower than *C. thermautotrophica* 41 and *C. hydrogenoformans* Z-2901 (Table 3).

Although the species most closely related to both JI_enriched_1_bin9 and JI_enriched_2_bin9 was *G. acetivorans* strain SBH6 (Fig. 2B), differences between the features of these genomes were observed. The genome of *G. acetivorans* strain SBH6 was reported to contain genes related to the CODH/ACS complex, but its CO utilization ability could not be confirmed (Mardanov *et al.*, 2015). The ANI values between

MAGs, JI_enriched_1_bin9 and JI_enriched_2_bin9, and *G. acetivorans* strain SBH6 were 77.44% and 76.85%, respectively. Therefore, it was inferred that JI_enriched_1_bin9 and JI_enriched_2_bin9 are novel pCO utilizer closely related to *G. acetivorans*. Table 4 shows a comparison between *G. acetivorans* strain SBH6 and JI_enriched_1_bin9 and JI_enriched_2_bin9. The genome size and the number of CDSs of the two pCO utilizer MAGs were smaller than those of *G. acetivorans* strain SBH6, and they contained fewer tRNA than *G. acetivorans* strain SBH6 (Table 4).

Because the ANI between JI_enriched_1_bin2 and *Desulfotomaculum kuznetsovii* DSM 6115 was over 96%, it was determined that JI_enriched_1_bin2 belongs to this species (Fig. 2A). The complete genome of *D. kuznetsovii* DSM 6115 has been sequenced and it is able to utilize CO (Parshina *et al.*, 2005). However, JI_enriched_1_bin2 did not include any of the 16S rRNA genes mentioned above. The genome length, G+C content, the number of CDSs and rRNA, tRNA and the number of CRISPR sequences between JI_enriched_1_bin2 and *D. kuznetsovii* DSM 6115 differed. JI_enriched_1_bin2 had a shorter genome but a higher G+C content than *D. kuznetsovii* DSM 6115, while containing fewer CDSs, rRNA, and tRNA (Table 5). JI_enriched_1_bin2 also contained two CRISPR sequences (Table 5). However, it was noted that the completeness value of this MAG was relatively low (Table 2). In conclusion, all the pCO utilizer MAGs contained smaller genomes and fewer CDSs than those of the most closely related species.

Phylogenetic and physiological traits of Ni-CODHs included in pCO utilizer MAGs

Ni-CODHs are phylogenetically divided into 8 clades (A-H) (Inoue *et al.*, 2022). In this study, the predicted physiological functions of Ni-CODHs were mainly of “energy conservation” and “carbon fixation”. Sixteen Ni-CODHs found in five pCO utilizer MAGs were analyzed phylogenetically and were distributed between 4 clades (A, B, D, and F). The phylogeny and physiological function of the 16 Ni-CODHs are discussed as follows.

Within the MAGs most closely related to *T. ferrireducens*, four Ni-CODH contained in JI_enriched_1_bin24 and three Ni-CODH contained in JI_enriched_2_bin23 were placed in Clade F (Fig.

3A). It is likely that one Ni-CODH from both MAGs is classified into Clade B (Fig. 3B). Below, the function of each Ni-CODH and its genomic context were described in detail. CODH-1 and CODH-2 of JI_enriched_1_bin24 and CODH-1 of JI_enriched_2_bin23 were adjacent to genes encoding ferredoxin-like electron transporter protein CooF, the transcriptional regulator protein CooA, the accessory protein CooC and proteins that constitute ECH (Fig. 4A). Therefore, the physiological roles of these three Ni-CODHs were inferred to be energy conservation along with of hydrogen generation. Next, functions of CODH-3 of JI_enriched_1_bin24 and CODH-2 of JI_enriched_2_bin23 were identified as carbon fixation. (Fig. 4B), the genes surrounding these Ni-CODH genes were those for CODH/ACS complex subunits and also accessory protein CooC (Fig. 4B). CODH-4 of JI_enriched_1_bin24 and CODH-3 of JI_enriched_2_bin23 were adjacent to genes for the ferredoxin-like protein CooF and NAD(P)H-nitrite reductase large subunit (Fig. 4C), suggesting possible respiratory roles through $NAD(P)^+$ reduction. CODH-5 in JI_enriched_1_bin24 and CODH-4 in JI_enriched_2_bin23 were adjacent to genes encoding ABC transporter protein (*tauA*, *tauB*, *tauC*) and cysteine synthase A (*cysK*) (Fig. 4D). Although similar Ni-CODH gene clusters were conserved in the genomes of *Calderihabitans maritimus* KKC1 and *Moorella thermoacetica* ATCC 39073, the physiological role of these Ni-CODHs has not been revealed (Omae *et al*., 2017). However, it is suggested that genes encoding the ABC transporter proteins are involved in taurine transport. A previous study reported the ability of microbes to utilize taurine anaerobically (Denger *et al*., 1999). *Desulfonispora thiosulfatigenes* was shown to generate acetic acid, ammonia, and thiosulfate ($S_2O_3^{2-}$) from taurine (Denger *et al*., 1999). Hydrogenogenic pCO utilizers may consume taurine via the same reaction to generate thiosulfate. CO oxidation by Ni-CODH is coupled to the reduction of thiosulfate, hydrogen sulfide is generated by the reduction of thiosulfate and cysteine can be synthesized from hydrogen sulfide by CysK. Together, this could indicate a novel physiological function of Ni-CODH. However, experimental proof will be required to confirm the above hypotheses for $NAD(P)^+$-mediated respiration and taurine transport. Taken together, these multiple Ni-CODH genes and their gene clusters suggested that the

two pCO utilizer MAGs of the genus *Thermolithobacter* were capable of coupling several distinct catalytic reactions with CO utilization such as hydrogenogenic CO oxidation and carbon fixation although, in the genus *Thermoithobacter*, *T. carboxydivorans* was only known to hydrogenogenically grow on CO (Sokolova et al., 2007).

The two MAGs related to *G. acetivorans*, JI_enriched_1_bin9 and JI_enriched_2_bin9, both contained two Ni-CODH genes: one CooS-type CODH gene and one Cdh-type CODH gene (Table 2). The Cdh-type genes were distributed in Clade A, while the CooS-type Ni-CODHs belonged to Clade D (Fig. 3C). To infer their physiological roles, the genomic context of these Ni-CODHs was examined. The Cdh-type Ni-CODH gene of JI_enriched_1_bin9 was adjacent to genes encoding CODH/ACS subunits (beta, gamma, delta, and epsilon) and threonine synthase genes, indicating that its physiological function could be carbon fixation. The remaining CooS-type Ni-CODH gene of JI_enriched_1_bin9 was adjacent to a single, unannotated gene, and the physiological function could not be inferred. The Cdh-type Ni-CODH gene of JI_enriched_2_bin9 was adjacent to genes encoding CODH/ACS subunits (beta and epsilon) and the physiological function is therefore most likely related to carbon fixation. The CooS-type Ni-CODH gene of JI_enriched_2_bin9 was adjacent to a gene for an AMP-binding protein, and thus its physiological function remained unidentified.

JI_enriched_1_bin2 was related to *D. kuznetsovii* and contained three CooS-type Ni-CODH genes (Table 2). Of these, CODH-1 grouped in Clade D, and CODH-2 and CODH-3 belonged to Clade F (Fig. 3A, C). The physiological roles of these three Ni-CODHs were predicted. No genes were found adjacent to CODH-1 and it was therefore functionally unidentified. CODH-2 was adjacent to the gene encoding the CODH/ACS complex subunit beta, while CODH-3 was adjacent to genes for CODH/ACS complex subunits alpha and beta. This likely indicates that the physiological function of these Ni-CODHs could be related to carbon fixation, although the genomic context of these CODH genes was incomplete.

The pCO utilizer MAGs were divided into two domains: Bacteria and Archaea (Fig. 2), however, the

Ni-CODHs of these MAGs were scattered across the four clades in the Ni-CODH tree (Fig. 3). Different phylogenetic traits of pCO utilizer MAGs and Ni-CODHs in the MAGs were observed. JI_enriched_1_bin24 and JI_enriched_2_bin23 were related to *C. thermoautotrophica* 41 based on genome sequences; however, the Ni-CODHs within these MAGs (CODH-5 of JI_enriched_1_bin24 and CODH-4 of JI_enriched_2_bin23) were more closely related to the Ni-CODHs of *Thermanaeromonas toyohensis* and *C. maritimus*. This indicates that, although some MAGs and species were closely related based on their genome sequences, their Ni-CODH sequences were phylogenetically distant. This further supports the previous studies showing complex evolution and punctate distribution of Ni-CODH genes possibly caused by lateral transfers and/or secondary losses (Techtmann *et al.*, 2012; Inoue *et al.*, 2019; Imaura *et al.*, 2023; Suzuki *et al.*, 2023).

Conclusion

In this study, two *Thermolithobacter* pCO utilizer MAGs, two *Geoglobus* pCO utilizer MAGs, and one *Desulfotomaculum* pCO utilizer MAG, all of which were of medium- to high-quality, were constructed. The *Thermolithobacter* MAGs contained Ni-CODH sequences which were suggested to be phylogenetically novel in a previous study (Omae *et al.*, 2021). Moreover, novel Ni-CODH sequences in *Thermolithobacter* and *Geoglobus* pCO utilizer MAGs, which were not clustered with any existing genome-encoded Ni-CODH sequences in the criteria of 90% amino acid sequence identity were discovered. These sequences were not found in the conventional Ni-CODH database (Inoue *et al.*, 2022). These findings were notable since various types of pCO utilizers, especially including hydrogenogenic pCO utilizers, are known to constitute a rare biosphere (Fukuyama *et al.*, 2020) and are difficult to detect even in metagenomic databases (Inoue *et al.*, 2022). Nevertheless, the high-quality or medium-quality MAGs from the pCO utilizers were obtained due to the integration of the traditional enrichment cultures and the recently developed next-generation sequencing techniques as well as the metagenome assembling (binning)

methods. In future, the integrated strategy used in this study would not only contribute to gaining insight into the diversity of the rare biosphere, i.e., CO utilizers, without isolation of them but also provide a genetic basis for establishing the conditions and compositions of media suitable for isolating the CO utilizers from various environments.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP16H06381 (to Y.S.), the Institute for Fermentation, Osaka Grant Number L-2021-1-002 (to T.Y.) and by JST SPRING, Grant Number JPMJSP2110. Computation time was provided by Super Computer System, Institute for Chemical Research, Kyoto University.

References

1. Camacho, Christiam, George Coulouris, Vahram Avagyan, Ning Ma, Jason Papadopoulos, Kevin Bealer, and Thomas L. Madden. 2009. "BLAST+: Architecture and Applications." *BMC Bioinformatics* 10 (1): 421. https://doi.org/10.1186/1471-2105-10-421.
2. Capella-Gutiérrez, Salvador, José M. Silla-Martínez, and Toni Gabaldón. 2009. "TrimAl: A Tool for Automated Alignment Trimming in Large-Scale Phylogenetic Analyses." *Bioinformatics* 25 (15): 1972–73. https://doi.org/10.1093/bioinformatics/btp348.
3. Chaumeil, Pierre-Alain, Aaron J Mussig, Philip Hugenholtz, and Donovan H Parks. 2020. "GTDB-Tk: A Toolkit to Classify Genomes with the Genome Taxonomy Database." *Bioinformatics* 36 (6): 1925–27. https://doi.org/10.1093/bioinformatics/btz848.
4. Chen, Shifu, Yanqing Zhou, Yaru Chen, and Jia Gu. 2018. "Fastp: An Ultra-Fast All-in-One FASTQ Preprocessor." *Bioinformatics* 34 (17): i884–90. https://doi.org/10.1093/bioinformatics/bty560.
5. Dobbek, Holger, Vitali Svetlitchnyi, Lothar Gremer, Robert Huber, and Ortwin Meyer. 2001. "Crystal Structure of a Carbon Monoxide Dehydrogenase Reveals a [Ni-4Fe-5S] Cluster." *Science* 293 (5533): 1281–85. https://doi.org/10.1126/science.1061500.
6. Eddy, Sean R. 2011. "Accelerated Profile HMM Searches." *PLOS Computational Biology* 7 (10): e1002195. https://doi.org/10.1371/journal.pcbi.1002195.
7. Edgar, Robert C. 2010. "Search and Clustering Orders of Magnitude Faster than BLAST." *Bioinformatics* 26 (19): 2460–61. https://doi.org/10.1093/bioinformatics/btq461.

8. Fox, J D, Y He, D Shelver, G P Roberts, and P W Ludden. 1996. “Characterization of the Region Encoding the CO-Induced Hydrogenase of Rhodospirillum Rubrum.” *Journal of Bacteriology* 178 (21): 6200–6208. https://doi.org/10.1128/jb.178.21.6200-6208.1996.

9. Fukuyama, Yuto, Masao Inoue, Kimiho Omae, Takashi Yoshida, and Yoshihiko Sako. 2020. “Chapter Three - Anaerobic and Hydrogenogenic Carbon Monoxide-Oxidizing Prokaryotes: Versatile Microbial Conversion of a Toxic Gas into an Available Energy.” In *Advances in Applied Microbiology*, edited by Geoffrey Michael Gadd and Sima Sariaslani, 110:99–148. Academic Press. https://doi.org/10.1016/bs.aambs.2019.12.001.

10. Harris, Charles R., K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, et al. 2020. “Array Programming with NumPy.” *Nature* 585 (7825): 357–62. https://doi.org/10.1038/s41586-020-2649-2.

11. Huerta-Cepas, Jaime, Kristoffer Forslund, Luis Pedro Coelho, Damian Szklarczyk, Lars Juhl Jensen, Christian von Mering, and Peer Bork. 2017. “Fast Genome-Wide Functional Annotation through Orthology Assignment by EggNOG-Mapper.” *Molecular Biology and Evolution* 34 (8): 2115–22. https://doi.org/10.1093/molbev/msx148.

12. Hyatt, Doug, Gwo-Liang Chen, Philip F. LoCascio, Miriam L. Land, Frank W. Larimer, and Loren J. Hauser. 2010. “Prodigal: Prokaryotic Gene Recognition and Translation Initiation Site Identification.” *BMC Bioinformatics* 11 (1): 119. https://doi.org/10.1186/1471-2105-11-119.

13. Imaura, Yoshinari, Shunsuke Okamoto, Taiki Hino, Yusuke Ogami, Yuka Adachi Katayama, Ayumi Tanimura, Masao Inoue, Ryoma Kamikawa, Takashi Yoshida, and Yoshihiko Sako. 2023. “Isolation, Genomic Sequence and Physiological Characterization of Parageobacillus Sp. G301, an Isolate Capable of Both Hydrogenogenic and Aerobic Carbon Monoxide Oxidation.” *Applied and Environmental Microbiology* 0 (0): e00185-23. https://doi.org/10.1128/aem.00185-23.

14. Inoue, Masao, Issei Nakamoto, Kimiho Omae, Tatsuki Oguro, Hiroyuki Ogata, Takashi Yoshida, and Yoshihiko Sako. 2019. “Structural and Phylogenetic Diversity of Anaerobic Carbon-Monoxide Dehydrogenases.” *Frontiers in Microbiology* 9. https://www.frontiersin.org/articles/10.3389/fmicb.2018.03353.

15. Inoue, Masao, Kimiho Omae, Issei Nakamoto, Ryoma Kamikawa, Takashi Yoshida, and Yoshihiko Sako. 2022. “Biome-Specific Distribution of Ni-Containing Carbon Monoxide Dehydrogenases.” *Extremophiles* 26 (1): 9. https://doi.org/10.1007/s00792-022-01259-y.

16. Jain, Chirag, Luis M. Rodriguez-R, Adam M. Phillippy, Konstantinos T. Konstantinidis, and Srinivas Aluru. 2018. “High Throughput ANI Analysis of 90K Prokaryotic Genomes Reveals Clear Species Boundaries.” *Nature Communications* 9 (1): 5114. https://doi.org/10.1038/s41467-018-07641-9.

17. Kang, Dongwan D., Feng Li, Edward Kirton, Ashleigh Thomas, Rob Egan, Hong An, and Zhong Wang. 2019. “MetaBAT 2: An Adaptive Binning Algorithm for Robust and Efficient Genome Reconstruction from Metagenome Assemblies.” *PeerJ* 7 (July): e7359.

https://doi.org/10.7717/peerj.7359.

18. Katoh, Kazutaka, and Daron M. Standley. 2013. “MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability.” *Molecular Biology and Evolution* 30 (4): 772–80. https://doi.org/10.1093/molbev/mst010.

19. King, Gary M., and Carolyn F. Weber. 2007. “Distribution, Diversity and Ecology of Aerobic CO-Oxidizing Bacteria.” *Nature Reviews Microbiology* 5 (2): 107–18. https://doi.org/10.1038/nrmicro1595.

20. Letunic, Ivica, and Peer Bork. 2021. “Interactive Tree Of Life (ITOL) v5: An Online Tool for Phylogenetic Tree Display and Annotation.” *Nucleic Acids Research* 49 (W1): W293–96. https://doi.org/10.1093/nar/gkab301.

21. Li, Dinghua, Ruibang Luo, Chi-Man Liu, Chi-Ming Leung, Hing-Fung Ting, Kunihiko Sadakane, Hiroshi Yamashita, and Tak-Wah Lam. 2016. “MEGAHIT v1.0: A Fast and Scalable Metagenome Assembler Driven by Advanced Methodologies and Community Practices.” *Methods*, Pan-omics analysis of biological data, 102 (June): 3–11. https://doi.org/10.1016/j.ymeth.2016.02.020.

22. Li, Heng. 2013. “Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM.” arXiv. https://doi.org/10.48550/arXiv.1303.3997.

23. Mardanov, Andrey V., Galina B. Slododkina, Alexander I. Slobodkin, Alexey V. Beletsky, Sergey N. Gavrilov, Ilya V. Kublanov, Elizaveta A. Bonch-Osmolovskaya, Konstantin G. Skryabin, and Nikolai V. Ravin. 2015. “The Geoglobus Acetivorans Genome: Fe(III) Reduction, Acetate Utilization, Autotrophic Growth, and Degradation of Aromatic Compounds in a Hyperthermophilic Archaeon.” *Applied and Environmental Microbiology* 81 (3): 1003–12. https://doi.org/10.1128/AEM.02705-14.

24. Matsen, Frederick A., Robin B. Kodner, and E Virginia Armbrust. 2010. “Pplacer: Linear Time Maximum-Likelihood and Bayesian Phylogenetic Placement of Sequences onto a Fixed Reference Tree.” *BMC Bioinformatics* 11 (1): 538. https://doi.org/10.1186/1471-2105-11-538.

25. Merrouch, Mériem, Martino Benvenuti, Marco Lorenzi, Christophe Léger, Vincent Fourmond, and Sébastien Dementin. 2018. “Maturation of the [Ni–4Fe–4S] Active Site of Carbon Monoxide Dehydrogenases.” *JBIC Journal of Biological Inorganic Chemistry* 23 (4): 613–20. https://doi.org/10.1007/s00775-018-1541-0.

26. Nguyen, Lam-Tung, Heiko A. Schmidt, Arndt von Haeseler, and Bui Quang Minh. 2015. “IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies.” *Molecular Biology and Evolution* 32 (1): 268–74. https://doi.org/10.1093/molbev/msu300.

27. Oelgeschläger, Ellen, and Michael Rother. 2008. “Carbon Monoxide-Dependent Energy Metabolism in Anaerobic Bacteria and Archaea.” *Archives of Microbiology* 190 (3): 257–69. https://doi.org/10.1007/s00203-008-0382-6.

28. Omae, Kimiho, Yuto Fukuyama, Hisato Yasuda, Kenta Mise, Takashi Yoshida, and Yoshihiko Sako.

2019. “Diversity and Distribution of Thermophilic Hydrogenogenic Carboxydotrophs Revealed by Microbial Community Analysis in Sediments from Multiple Hydrothermal Environments in Japan.” *Archives of Microbiology* 201 (7): 969–82. https://doi.org/10.1007/s00203-019-01661-9.

29. Omae, Kimiho, Tatsuki Oguro, Masao Inoue, Yuto Fukuyama, Takashi Yoshida, and Yoshihiko Sako. 2021. “Diversity Analysis of Thermophilic Hydrogenogenic Carboxydotrophs by Carbon Monoxide Dehydrogenase Amplicon Sequencing Using New Primers.” *Extremophiles* 25 (1): 61–76. https://doi.org/10.1007/s00792-020-01211-y.

30. Omae, Kimiho, Yasuko Yoneda, Yuto Fukuyama, Takashi Yoshida, and Yoshihiko Sako. 2017. “Genomic Analysis of Calderihabitans Maritimus KKC1, a Thermophilic, Hydrogenogenic, Carboxydotrophic Bacterium Isolated from Marine Sediment.” *Applied and Environmental Microbiology* 83 (15): e00832-17. https://doi.org/10.1128/AEM.00832-17.

31. Omaye, Stanley T. 2002. “Metabolic Modulation of Carbon Monoxide Toxicity.” *Toxicology* 180 (2): 139–50. https://doi.org/10.1016/S0300-483X(02)00387-6.

32. Ondov, Brian D., Todd J. Treangen, Páll Melsted, Adam B. Mallonee, Nicholas H. Bergman, Sergey Koren, and Adam M. Phillippy. 2016. “Mash: Fast Genome and Metagenome Distance Estimation Using MinHash.” *Genome Biology* 17 (1): 132. https://doi.org/10.1186/s13059-016-0997-x.

33. Parks, Donovan H., Maria Chuvochina, Pierre-Alain Chaumeil, Christian Rinke, Aaron J. Mussig, and Philip Hugenholtz. 2020. “A Complete Domain-to-Species Taxonomy for Bacteria and Archaea.” *Nature Biotechnology* 38 (9): 1079–86. https://doi.org/10.1038/s41587-020-0501-8.

34. Parks, Donovan H., Maria Chuvochina, David W. Waite, Christian Rinke, Adam Skarshewski, Pierre-Alain Chaumeil, and Philip Hugenholtz. 2018. “A Standardized Bacterial Taxonomy Based on Genome Phylogeny Substantially Revises the Tree of Life.” *Nature Biotechnology* 36 (10): 996–1004. https://doi.org/10.1038/nbt.4229.

35. Parks, Donovan H., Michael Imelfort, Connor T. Skennerton, Philip Hugenholtz, and Gene W. Tyson. 2015. “CheckM: Assessing the Quality of Microbial Genomes Recovered from Isolates, Single Cells, and Metagenomes.” *Genome Research* 25 (7): 1043–55. https://doi.org/10.1101/gr.186072.114.

36. Parshina, S. N., S. Kijlstra, A. M. Henstra, J. Sipma, C. M. Plugge, and A. J. M. Stams. 2005. “Carbon Monoxide Conversion by Thermophilic Sulfate-Reducing Bacteria in Pure Culture and in Co-Culture with Carboxydothermus Hydrogenoformans.” *Applied Microbiology and Biotechnology* 68 (3): 390–96. https://doi.org/10.1007/s00253-004-1878-x.

37. Price, Morgan N., Paramvir S. Dehal, and Adam P. Arkin. 2010. “FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments.” *PLOS ONE* 5 (3): e9490. https://doi.org/10.1371/journal.pone.0009490.

38. Pritchard, Leighton, Rachel H. Glover, Sonia Humphris, John G. Elphinstone, and Ian K. Toth. 2016. “Genomics and Taxonomy in Diagnostics for Food Security: Soft-Rotting Enterobacterial Plant

508 Pathogens." *Analytical Methods* 8 (1): 12–24. https://doi.org/10.1039/C5AY02550H.
509 39. Schoelmerich, Marie Charlotte, and Volker Müller. 2019. "Energy Conservation by a Hydrogenase-
510 Dependent Chemiosmotic Mechanism in an Ancient Metabolic Pathway." *Proceedings of the*
511 *National Academy of Sciences* 116 (13): 6329–34. https://doi.org/10.1073/pnas.1818580116.
512 40. Shen, Wei, Shuai Le, Yan Li, and Fuquan Hu. 2016. "SeqKit: A Cross-Platform and Ultrafast Toolkit
513 for FASTA/Q File Manipulation." *PLOS ONE* 11 (10): e0163962.
514 https://doi.org/10.1371/journal.pone.0163962.
515 41. Sokolova, T., J. Hanel, R. U. Onyenwoke, A.-L. Reysenbach, A. Banta, R. Geyer, J. M. González, W.
516 B. Whitman, and J. Wiegel. 2007. "Novel Chemolithotrophic, Thermophilic, Anaerobic Bacteria
517 Thermolithobacter Ferrireducens Gen. Nov., Sp. Nov. and Thermolithobacter Carboxydivorans Sp.
518 Nov." *Extremophiles* 11 (1): 145–57. https://doi.org/10.1007/s00792-006-0022-5.
519 42. Sokolova, Tatyana G., Anne-Meint Henstra, Jan Sipma, Sofiya N. Parshina, Alfons J.M. Stams, and
520 Alexander V. Lebedinsky. 2009. "Diversity and Ecophysiological Features of Thermophilic
521 Carboxydotrophic Anaerobes." *FEMS Microbiology Ecology* 68 (2): 131–41.
522 https://doi.org/10.1111/j.1574-6941.2009.00663.x.
523 43. Sukumaran, Jeet, and Mark Holder. 2010. "DendroPy: A Python Library for Phylogenetic Computing."
524 *Bioinformatics* 26 (12): 1569–71. https://doi.org/10.1093/bioinformatics/btq228.
525 44. Suzuki, Jota, Yoshinari Imaura, Shiho Nishida, Ryoma Kamikawa, and Takashi Yoshida. 2023. "Draft
526 Genome Sequence of Thermolongibacillus Altinsuensis Strain B1-1, a Novel Hydrogenogenic CO
527 Oxidizer Isolated from Sediment from Lake Biwa in Japan." *Microbiology Resource*
528 *Announcements* 0 (0): e00334-23. https://doi.org/10.1128/mra.00334-23.
529 45. Svetlitchnyi, Vitali, Christine Peschel, Georg Acker, and Ortwin Meyer. 2001. "Two Membrane-
530 Associated NiFeS-Carbon Monoxide Dehydrogenases from the Anaerobic Carbon-Monoxide-
531 Utilizing EubacteriumCarboxydothermus Hydrogenoformans." *Journal of Bacteriology* 183 (17):
532 5134–44. https://doi.org/10.1128/JB.183.17.5134-5144.2001.
533 46. Tamura, Koichiro, Glen Stecher, and Sudhir Kumar. 2021. "MEGA11: Molecular Evolutionary
534 Genetics Analysis Version 11." *Molecular Biology and Evolution* 38 (7): 3022–27.
535 https://doi.org/10.1093/molbev/msab120.
536 47. Tanizawa, Yasuhiro, Takatomo Fujisawa, and Yasukazu Nakamura. 2018. "DFAST: A Flexible
537 Prokaryotic Genome Annotation Pipeline for Faster Genome Publication." *Bioinformatics* 34 (6):
538 1037–39. https://doi.org/10.1093/bioinformatics/btx713.
539 48. Techtmann, Stephen, Albert Colman, Alexander Lebedinsky, Tatyana Sokolova, and Frank Robb. 2012.
540 "Evidence for Horizontal Gene Transfer of Anaerobic Carbon Monoxide Dehydrogenases."
541 *Frontiers in Microbiology* 3. https://www.frontiersin.org/articles/10.3389/fmicb.2012.00132.
542 49. Techtmann, Stephen M., Albert S. Colman, and Frank T. Robb. 2009. "'That Which Does Not Kill Us
543 Only Makes Us Stronger': The Role of Carbon Monoxide in Thermophilic Microbial Consortia."

Environmental Microbiology 11 (5): 1027–37. https://doi.org/10.1111/j.1462-2920.2009.01865.x.
50. Wu, Martin, Qinghu Ren, A. Scott Durkin, Sean C. Daugherty, Lauren M. Brinkac, Robert J. Dodson, Ramana Madupu, et al. 2005. "Life in Hot Carbon Monoxide: The Complete Genome Sequence of Carboxydothermus Hydrogenoformans Z-2901." *PLOS Genetics* 1 (5): e65. https://doi.org/10.1371/journal.pgen.0010065.
51. Yoneda, Y., T. Yoshida, T. Daifuku, T. Kitamura, T. Inoue, S. Kano, and Y. Sako. 2013. "Quantitative Detection of Carboxydotrophic Bacteria Carboxydothermus in a Hot Aquatic Environment." *Fundamental and Applied Limnology* 182 (2): 161–70.
52. Yoneda, Yasuko, Sanae I. Kano, Takashi Yoshida, Eitaro Ikeda, Yuto Fukuyama, Kimiho Omae, Shigeko Kimura-Sakai, Takashi Daifuku, Tetsuhiro Watanabe, and Yoshihiko Sako. 2015. "Detection of Anaerobic Carbon Monoxide-Oxidizing Thermophiles in Hydrothermal Environments." *FEMS Microbiology Ecology* 91 (9). https://doi.org/10.1093/femsec/fiv093.

Figures and Tables

Figure legends

Fig. 1. Phylogenetic analysis based on 16S rRNA sequences. (A) Maximum likelihood (ML) phylogenetic tree of hydrogenogenic potential carbon monoxide (pCO) utilizer metagenome assembled genomes (MAGs), JI_enriched_1_bin24 and JI_enriched_2_bin23, and closely related species. (B) ML phylogenetic tree of pCO utilizer MAGs, JI_enriched_1_bin9, and closely related species.

Fig. 2. Genome-based phylogeny of metagenome assembled genomes (MAGs). Purple circles indicate bootstrap scores over 80%. (A) Phylogenetic tree of representatives of Phylum Firmicutes_B bacteria and bacterial MAGs obtained in this study. (B) Phylogenetic tree of representatives of Phylum Halobacteriota archaea and archaeal MAGs obtained in this study.

Fig. 3. Phylogenetic relationships of representative anaerobic Ni-carbon monoxide dehydrogenase (Ni-CODHs) catalytic subunit. Ni-CODH sequences used were those retrieved from public database (MGnify and RefSeq/GenBank) in Inoue *et al.* 2022 and those found only in MAGs reconstructed in this study. Purple squares indicate bootstrap scores over 80%. (A) 424 representative Ni-CODH sequences of Clade F and those found only in MAGs reconstructed in this study. (B) 202 representative Ni-CODH sequences of Clade B and those found only in MAGs reconstructed in this study. (C) 357 representative Ni-CODH sequences of Clade D and those found only in MAGs reconstructed in this study.

Fig. 4. Ni-carbon monoxide dehydrogenase (Ni-CODHs) and surrounding genes (genomic contexts) of JI_enriched_1_bin24 and JI_enriched_2_bin23. Upper figure indicates JI_enriched_1_bin24 and lower figure indicates JI_enriched_2_bin23. Cth_41: *C. thermautotrophica* 41; KKC1: *C. maritimus* KKC1. (A) CODH-1 and CODH-2 of JI_enriched_1_bin24 and CODH-1 of JI_enriched_2_bin23. Ni-CODH genes, energy-converting hydrogenase (ECH) genes, CooA genes, CooC genes and CooF genes are depicted. (B) CODH-3 of JI_enriched_1_bin24 and CODH-2 of JI_enriched_2_bin23. Ni-CODH genes, CooC genes and acethyl-CoA synthase related genes are depicted. (C) CODH-4 of JI_enriched_1_bin24 and CODH-3 of JI_enriched_2_bin23. Ni-CODH genes, CooF genes and NAD(P)H-nitrite reductase large subunit genes are shown. (D) CODH-5 of JI_enriched_1_bin24 and CODH-4 of JI_enriched_2_bin23. Cysteine synthase genes, ABC transporter genes and CODH genes are depicted.

Table 1. Number of paired-end reads and metagenome assembled genomes (MAGs) reconstructed by metagenome sequence from anaerobic enrichment culture.

	JI_enriched_1	JI_enriched_2
Number of		
Raw reads	27,532,830	26,259,718
Q20 pass reads	15,534,972	16,503,906
Number of		
High-quality bins*[1]	9	6
Medium-quality bins*[2]	8	9

*1: completeness>90% and contamination<5%

*2: completeness≥50% and contamination<10%

Table 2. General features of the genomes from metagenome assembled genomes (MAGs) containing Ni-carbon monoxide dehydrogenase (Ni-CODH). *C. hydrogenoformans* Z-2901 CooS II (CooS type) and *Methanosarcina barkeri* ACS complex α subunit (Cdh type) was used as query Ni-CODH sequences (Inoue *et al*., 2019).

Sample ID	Genome Bin ID	Completeness (%)	Contamination (%)	Number of Contigs	Number of CODH genes	
					CooS type	Cdh type
JI_enriched_1	bin2	88.59	6.12	344	3	0
JI_enriched_1	bin9	91.50	0.00	37	1	1
JI_enriched_1	bin24	99.43	1.11	19	5	0
JI_enriched_2	bin9	95.42	1.31	83	1	1
JI_enriched_2	bin23	100.00	0.77	24	4	0

Table 3. Comparison of general features of the genomes from hydrogenogenic pCO utilizer metagenome assembled genomes (MAGs), *C. thermautotrophica* 41^T (Cth_41) and *C. hydrogenoformans* Z-2901 (Chy).

	JI_enriched_1_bin24	JI_enriched_2_bin23	Cth_41	Chy
Total Sequence Length (bp)	1,998,880	2,077,722	2,743,125	2,401,520
G+C content (%)	53.0	52.6	49.09	42.0
Number of				
CDSs	1,991	2,074	2,739	2,439
rRNA	5	5	15	12
tRNA	52	54	73	50
CRISPRs	1	1	1	-
Reference	This study	This study	IMG*1	NCBI*2

*1: https://img.jgi.doe.gov/cgi-bin/m/main.cgi?section=TaxonDetail&page=taxonDetail&taxon_oid=2802429487

*2: https://www.ncbi.nlm.nih.gov/genome/?term=carboxydothermus%20hydrogenoformans

Table 4. General features of the genomes from non-hydrogenogenic pCO utilizer metagenome assembled genomes (MAGs) and closely-related species. Archaeal MAGs (JI_enriched_1_bin9, JI_enriched_2_bin9) and *Geoglobus acetivorans* strain SBH6 (Gac). *C. hydrogenoformans* Z-2901

(Chy) was selected for reference as representative carboxydotroph species.

	JI_enriched_1_bin9	JI_enriched_2_bin9	Gac	Chy
Total Sequence Length (bp)	1,411,080	1,462,938	1,901,110	2,401,520
G+C content (%)	41.5	41.4	46.5	42.0
Number of				
CDSs	1,554	1,576	2,158	2,439
rRNA	3	1	3	12
tRNA	39	42	47	50
CRISPRs	2	4	-	-
Reference	This study	This study	NCBI*1	NCBI*2

*1: https://www.ncbi.nlm.nih.gov/genome/?term=geoglobus%20acetivorans

*2: https://www.ncbi.nlm.nih.gov/genome/?term=carboxydothermus%20hydrogenoformans

Table 5. General features of the genomes from non-hydrogenogenic pCO utilizer metagenome assembled genomes (MAGs) and closely-related species. Bacterial MAG (JI_enriched_1_bin2) and *D. kuznetsovii* DSM 6115 (Dku). *C. hydrogenoformans* Z-2901 (Chy) was selected for reference as representative carboxydotroph species.

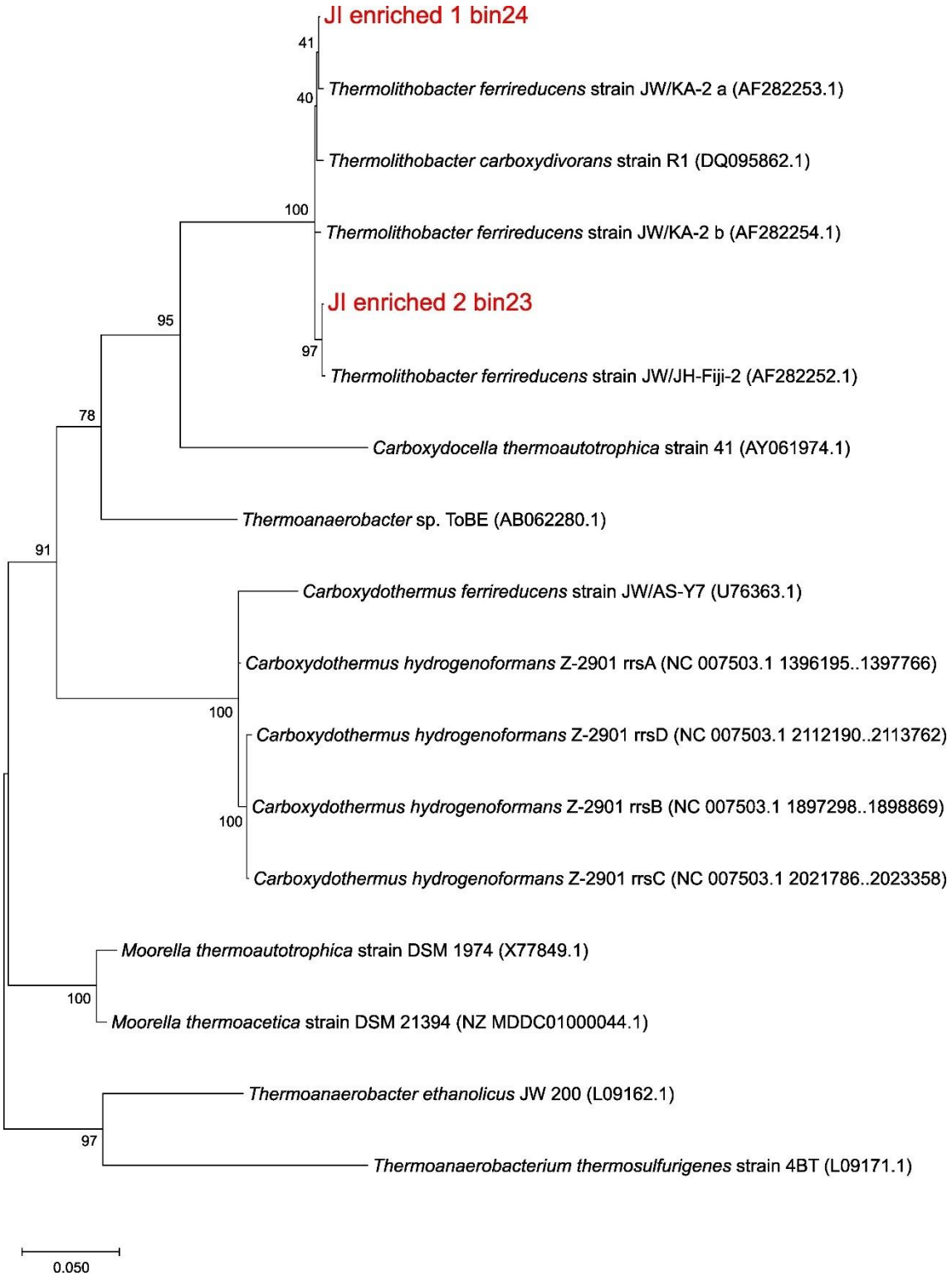
	JI_enriched_1_bin2	Dku	Chy
Total Sequence Length (bp)	2,734,629	3,601,390	2,401,520
G+C content (%)	55.1	54.9	42.0
Number of			
CDSs	2,418	3,398	2,439
rRNA	3	9	12
tRNA	38	46	50
CRISPRs	2	-	-
Reference	This study	NCBI*1	NCBI*2

*1: https://www.ncbi.nlm.nih.gov/genome/3122?genome_assembly_id=172913

*2: https://www.ncbi.nlm.nih.gov/genome/?term=carboxydothermus%20hydrogenoformans

Fig. 1.

(A)


JI enriched 1 bin24
41
Thermolithobacter ferrireducens strain JW/KA-2 a (AF282253.1)
40
Thermolithobacter carboxydivorans strain R1 (DQ095862.1)
100
Thermolithobacter ferrireducens strain JW/KA-2 b (AF282254.1)
JI enriched 2 bin23
95
97
Thermolithobacter ferrireducens strain JW/JH-Fiji-2 (AF282252.1)
78
Carboxydocella thermoautotrophica strain 41 (AY061974.1)
Thermoanaerobacter sp. ToBE (AB062280.1)
91
Carboxydothermus ferrireducens strain JW/AS-Y7 (U76363.1)
Carboxydothermus hydrogenoformans Z-2901 rrsA (NC 007503.1 1396195..1397766)
100
Carboxydothermus hydrogenoformans Z-2901 rrsD (NC 007503.1 2112190..2113762)
100
Carboxydothermus hydrogenoformans Z-2901 rrsB (NC 007503.1 1897298..1898869)
Carboxydothermus hydrogenoformans Z-2901 rrsC (NC 007503.1 2021786..2023358)
Moorella thermoautotrophica strain DSM 1974 (X77849.1)
100
Moorella thermoacetica strain DSM 21394 (NZ MDDC01000044.1)
Thermoanaerobacter ethanolicus JW 200 (L09162.1)
97
Thermoanaerobacterium thermosulfurigenes strain 4BT (L09171.1)
0.050

(B)

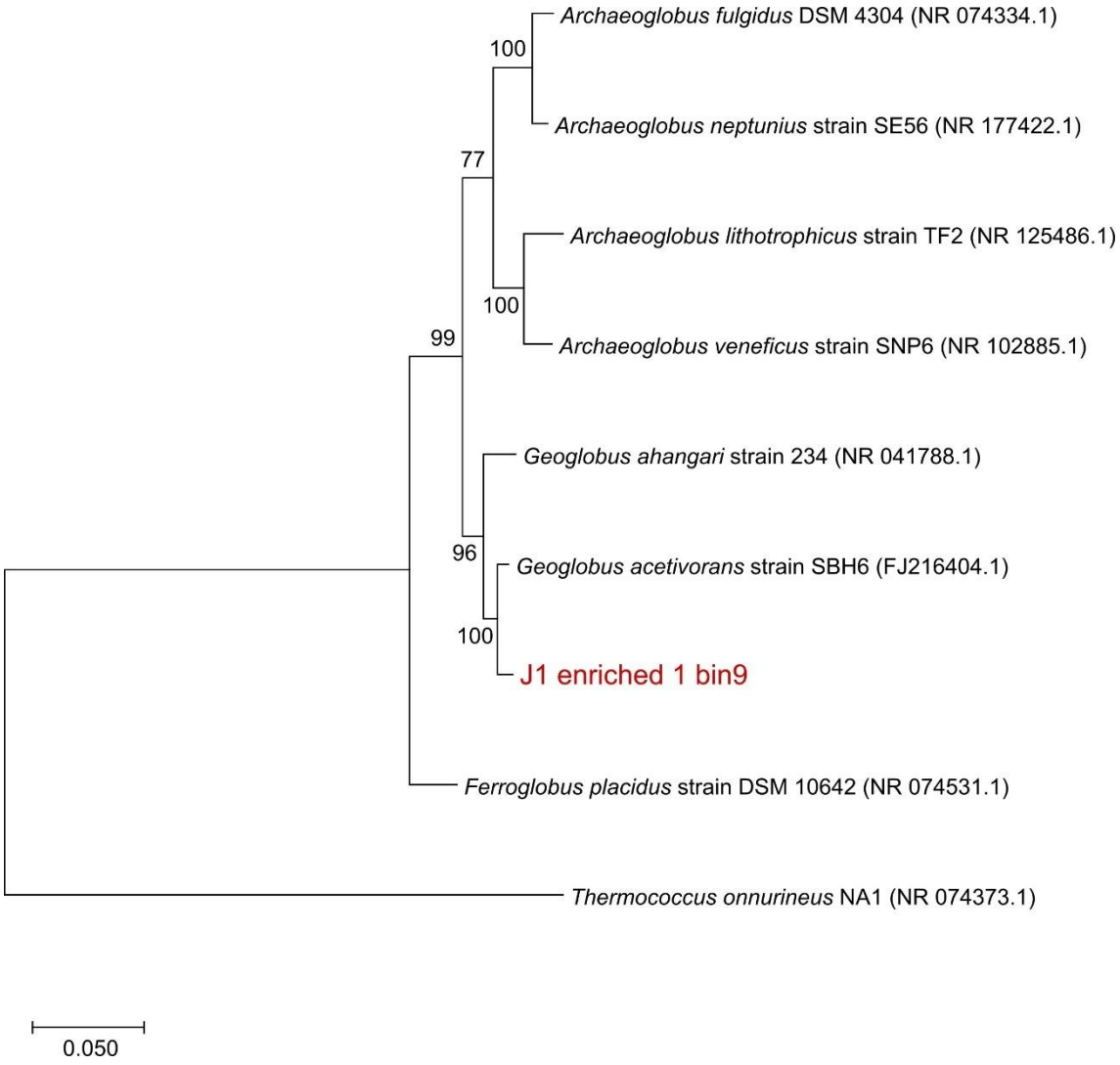
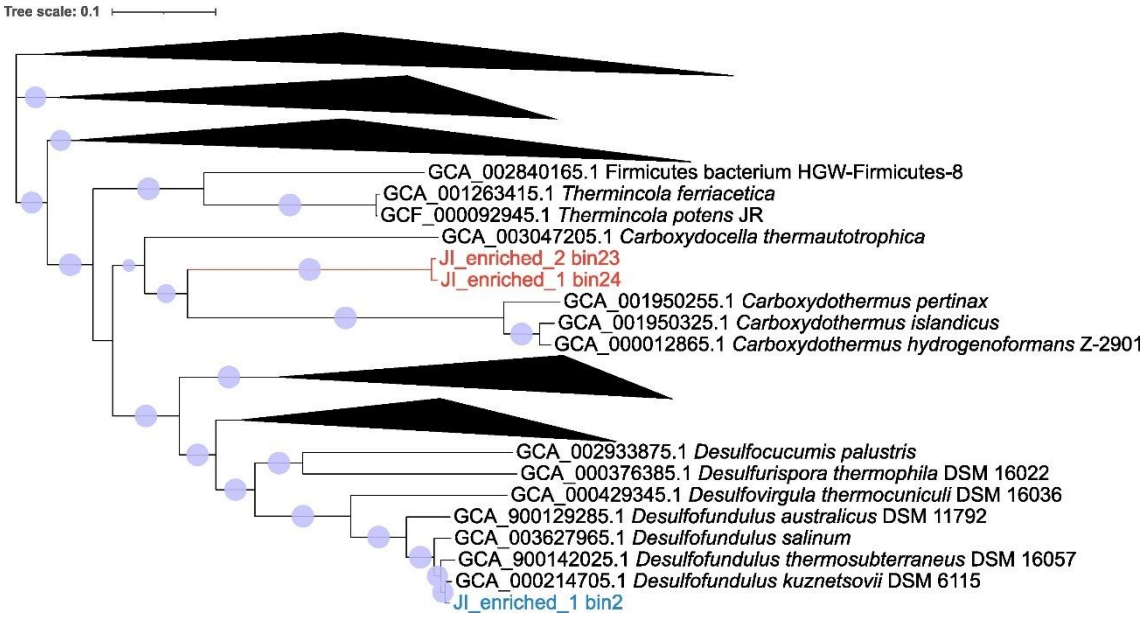

Archaeoglobus fulgidus DSM 4304 (NR 074334.1)
100
Archaeoglobus neptunius strain SE56 (NR 177422.1)
77
Archaeoglobus lithotrophicus strain TF2 (NR 125486.1)
100
99
Archaeoglobus veneficus strain SNP6 (NR 102885.1)
Geoglobus ahangari strain 234 (NR 041788.1)
96
Geoglobus acetivorans strain SBH6 (FJ216404.1)
100
J1 enriched 1 bin9
Ferroglobus placidus strain DSM 10642 (NR 074531.1)
Thermococcus onnurineus NA1 (NR 074373.1)
0.050

Fig. 2.

(A)


Tree scale: 0.1
GCA_002840165.1 Firmicutes bacterium HGW-Firmicutes-8
GCA_001263415.1 Thermincola ferriacetica
GCF_000092945.1 Thermincola potens JR
GCA_003047205.1 Carboxydocella thermautotrophica
JI_enriched_2 bin23
JI_enriched_1 bin24
GCA_001950255.1 Carboxydothermus pertinax
GCA_001950325.1 Carboxydothermus islandicus
GCA_000012865.1 Carboxydothermus hydrogenoformans Z-2901
GCA_002933875.1 Desulfocucumis palustris
GCA_000376385.1 Desulfurispora thermophila DSM 16022
GCA_000429345.1 Desulfovirgula thermocuniculi DSM 16036
GCA_900129285.1 Desulfofundulus australicus DSM 11792
GCA_003627965.1 Desulfofundulus salinum
GCA_900142025.1 Desulfofundulus thermosubterraneus DSM 16057
GCA_000214705.1 Desulfofundulus kuznetsovii DSM 6115
JI_enriched_1 bin2

(B)

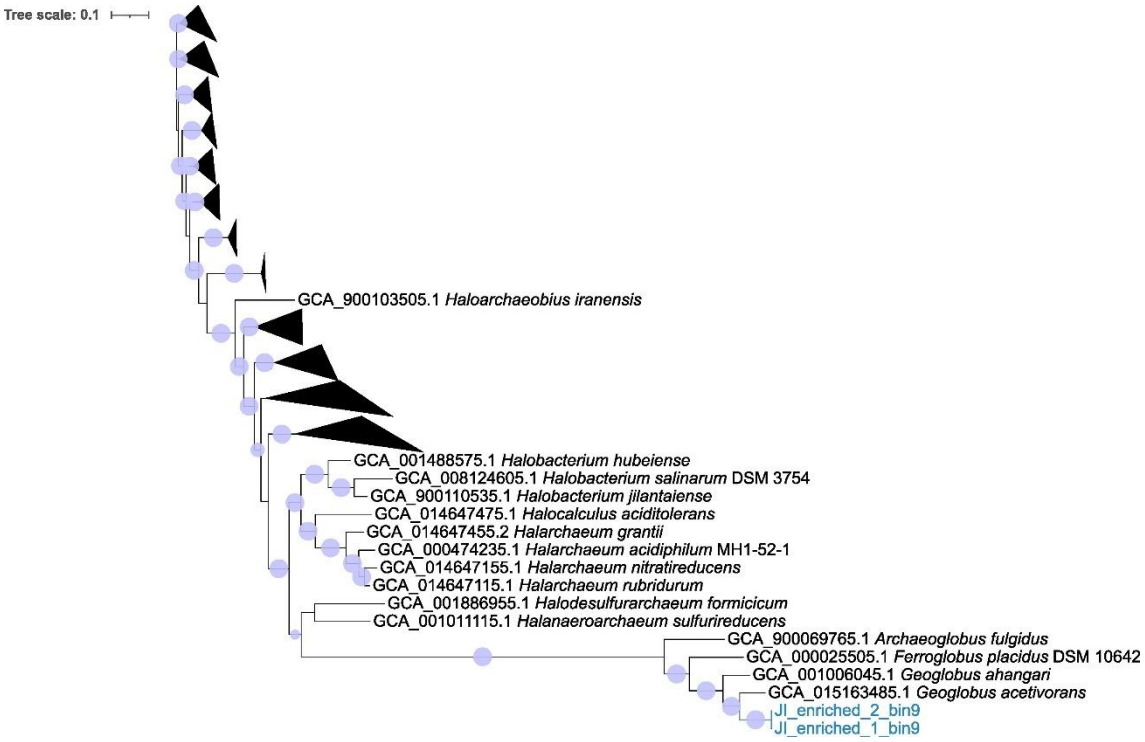
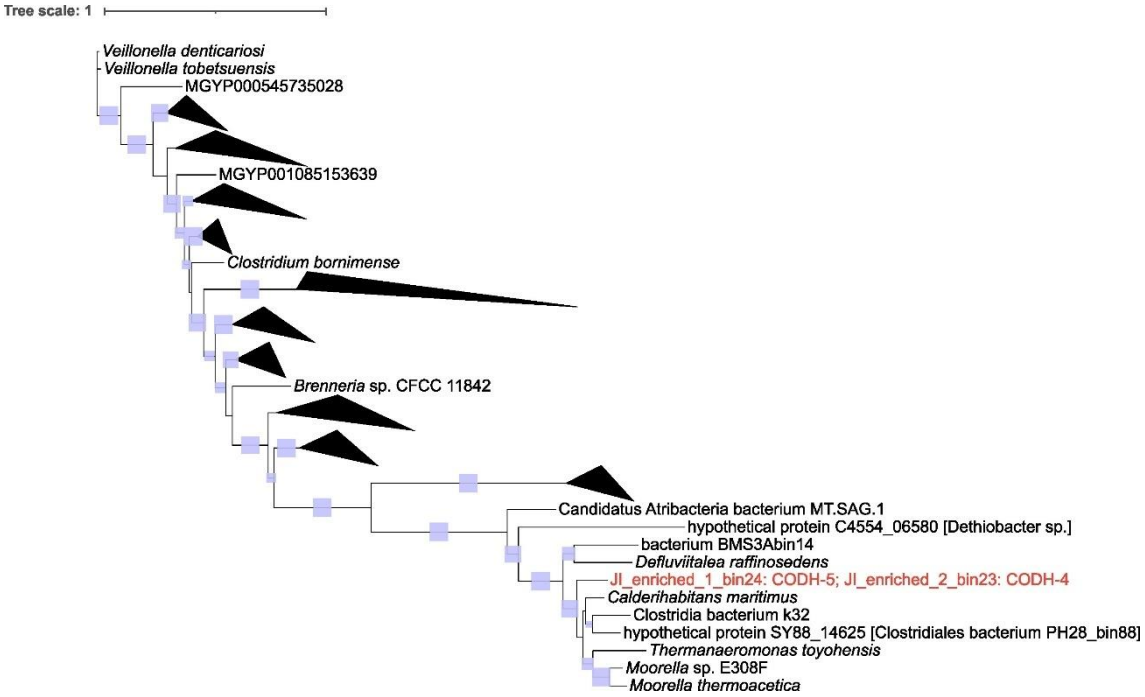

Tree scale: 0.1
GCA_900103505.1 Haloarchaeobius iranensis
GCA_001488575.1 Halobacterium hubeiense
GCA_008124605.1 Halobacterium salinarum DSM 3754
GCA_900110535.1 Halobacterium jilantaiense
GCA_014647475.1 Halocalculus aciditolerans
GCA_014647455.2 Halarchaeum grantii
GCA_000474235.1 Halarchaeum acidiphilum MH1-52-1
GCA_014647155.1 Halarchaeum nitratireducens
GCA_014647115.1 Halarchaeum rubridurum
GCA_001886955.1 Halodesulfurarchaeum formicicum
GCA_001011115.1 Halanaeroarchaeum sulfurireducens
GCA_900069765.1 Archaeoglobus fulgidus
GCA_000025505.1 Ferroglobus placidus DSM 10642
GCA_001006045.1 Geoglobus ahangari
GCA_015163485.1 Geoglobus acetivorans
JI_enriched_2_bin9
JI_enriched_1_bin9

Fig. 3.

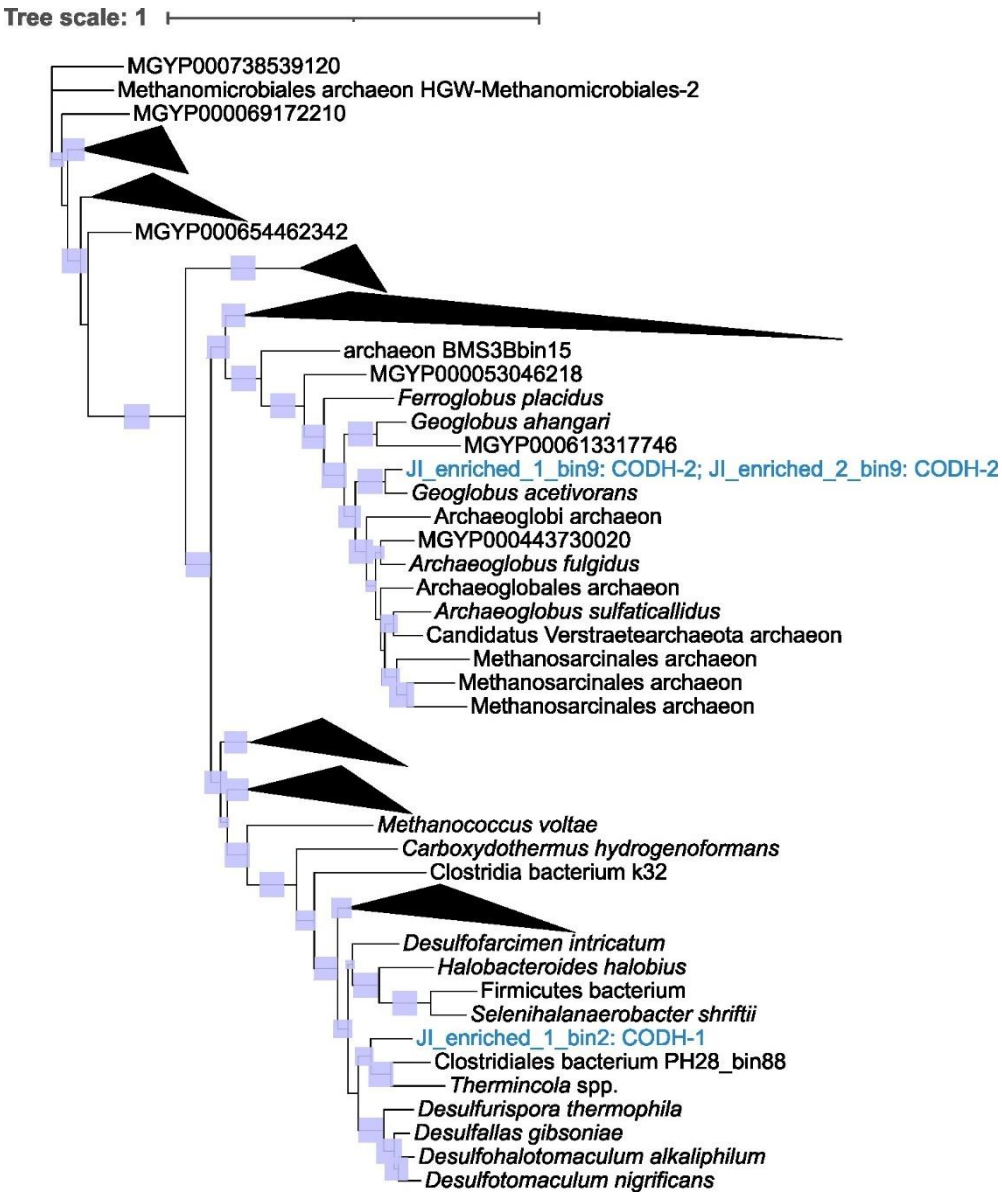
(A)


Tree scale: 1
Candiadtus Scalindua japonica
Candidatus Kuenenia stuttgartiensis
Tepidibacillus decaturensis
JI_enriched_1_bin24: CODH-2
Carboxydocella spp.
JI_enriched_1_bin24: CODH-4; JI_enriched_2_bin23: CODH-3
Clostridia bacterium k32
Peptococcaceae bacterium DCMF
Sulfurospirillum sp. SCADC
Geothermobacter ehrlichii
Sulfurimonas sp.
Calderihabitans maritimus
Thermoanaerosceptrum fracticalcis
JI_enriched_1_bin24:CODH-1; JI_enriched_2_bin23: CODH-1
Carboxydocella spp.
JI_enriched_1_bin24: CODH-3; JI_enriched_2_bin23: CODH-2
Firmicutes bacterium HGW-Firmicutes-8
Thermincola potens
Desulfitibacter sp. BRH_c19
Desulfofundulus thermobenzoicus
MGYP000545467787
Desulfotomaculum copahuensis
JI_enriched_1_bin2: CODH-2
Desulfofarcimen acetoxidans
Desulfofarcimen acetoxidans
Desulfofarcimen intricatum
Desulfurispora thermophila
Desulfallas gibsoniae
Desulfallas thermosapovorans DSM 6562
Peptococcaceae bacterium BRH_c8a
Peptococcaceae bacterium BICA1-7
Firmicutes bacterium
Desulfotomaculum copahuensis
Desulfofundulus thermobenzoicus
Desulfovirgula thermocuniculi
JI_enriched_1_bin2: CODH-3

(B)


Tree scale: 1
Veillonella denticariosi
Veillonella tobetsuensis
MGYP000545735028
MGYP001085153639
Clostridium bornimense
Brenneria sp. CFCC 11842
Candidatus Atribacteria bacterium MT.SAG.1
hypothetical protein C4554_06580 [Dethiobacter sp.]
bacterium BMS3Abin14
Defluviitalea raffinosedens
JI_enriched_1_bin24: CODH-5; JI_enriched_2_bin23: CODH-4
Calderihabitans maritimus
Clostridia bacterium k32
hypothetical protein SY88_14625 [Clostridiales bacterium PH28_bin88]
Thermanaeromonas toyohensis
Moorella sp. E308F
Moorella thermoacetica

682 **(C)**



683

684

685

686

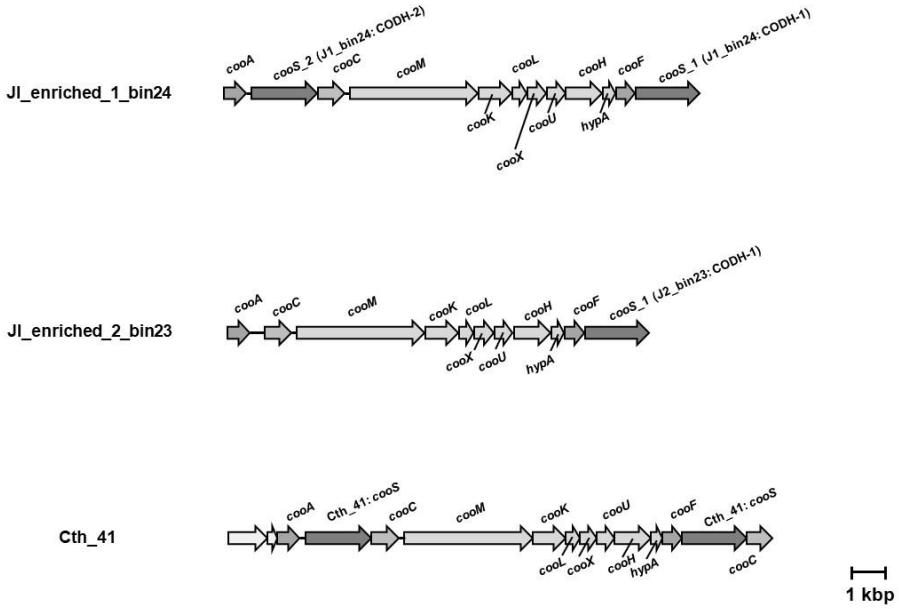
687

688

689

690 **Fig. 4**

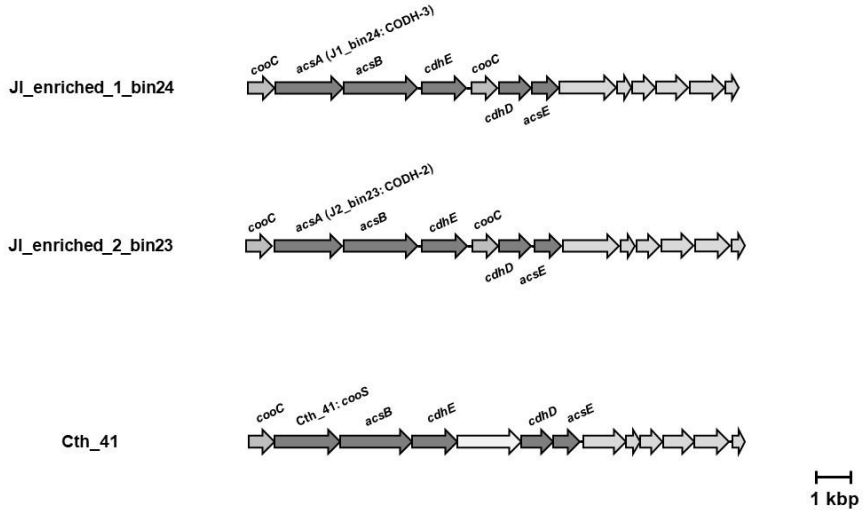
691 **(A)**



692

693

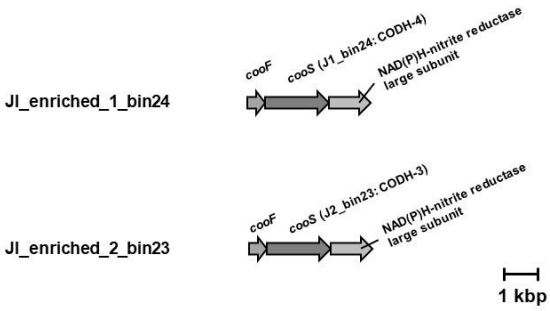
694 **(B)**



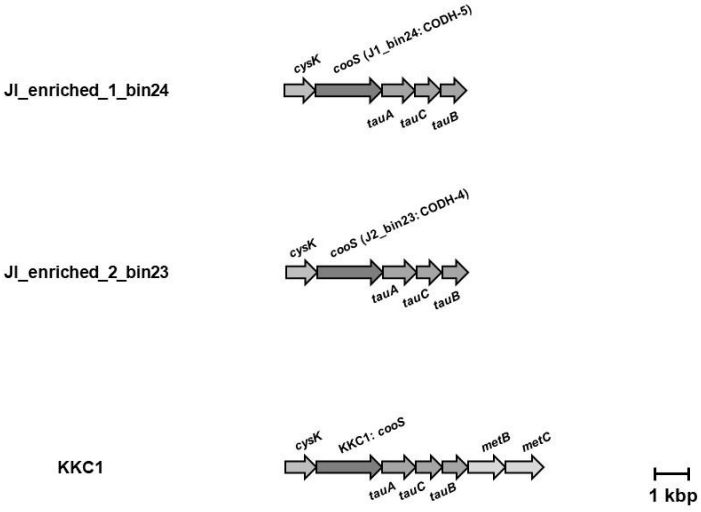
695

696

(C)


cooF
cooS (J1_bin24: CODH-4)
NAD(P)H-nitrite reductase large subunit
JI_enriched_1_bin24
cooF
cooS (J2_bin23: CODH-3)
NAD(P)H-nitrite reductase large subunit
JI_enriched_2_bin23
1 kbp

(D)


JI_enriched_1_bin24
cysK
cooS (J1_bin24: CODH-5)
tauA
tauC
tauB
JI_enriched_2_bin23
cysK
cooS (J2_bin23: CODH-4)
tauA
tauC
tauB
KKC1
cysK
KKC1: cooS
metB
metC
tauA
tauC
tauB
1 kbp