frontiers in Microbiology

ORIGINAL RESEARCH
published: 17 January 2019
doi: 10.3389/fmicb.2018.03353

Structural and Phylogenetic Diversity of Anaerobic Carbon-Monoxide Dehydrogenases

*Masao Inoue[1], Issei Nakamoto[1], Kimiho Omae[1], Tatsuki Oguro[1], Hiroyuki Ogata[2], Takashi Yoshida[1] and Yoshihiko Sako[1]**

[1] Graduate School of Agriculture, Kyoto University, Kyoto, Japan, [2] Institute for Chemical Research, Kyoto University, Kyoto, Japan

Anaerobic Ni-containing carbon-monoxide dehydrogenases (Ni-CODHs) catalyze the reversible conversion between carbon monoxide and carbon dioxide as multi-enzyme complexes responsible for carbon fixation and energy conservation in anaerobic microbes. However, few biochemically characterized model enzymes exist, with most Ni-CODHs remaining functionally unknown. Here, we performed phylogenetic and structure-based Ni-CODH classification using an expanded dataset comprised of 1942 non-redundant Ni-CODHs from 1375 Ni-CODH-encoding genomes across 36 phyla. Ni-CODHs were divided into seven clades, including a novel clade. Further classification into 24 structural groups based on sequence analysis combined with structural prediction revealed diverse structural motifs for metal cluster formation and catalysis, including novel structural motifs potentially capable of forming metal clusters or binding metal ions, indicating Ni-CODH diversity and plasticity. Phylogenetic analysis illustrated that the metal clusters responsible for intermolecular electron transfer were drastically altered during evolution. Additionally, we identified novel putative Ni-CODH-associated proteins from genomic contexts other than the Wood–Ljungdahl pathway and energy converting hydrogenase system proteins. Network analysis among the structural groups of Ni-CODHs, their associated proteins and taxonomies revealed previously unrecognized gene clusters for Ni-CODHs, including uncharacterized structural groups with putative metal transporters, oxidoreductases, or transcription factors. These results suggested diversification of Ni-CODH structures adapting to their associated proteins across microbial genomes.

Keywords: carbon-monoxide dehydrogenase, genomic context, structural prediction, functional prediction, molecular evolution, horizontal gene transfer

OPEN ACCESS

Edited by:
Frank T. Robb, University of Maryland, Baltimore, United States

Reviewed by:
Michael Köpke, LanzaTech, United States
Stephen Techtmann, Michigan Technological University, United States
Juan M. Gonzalez, Spanish National Research Council (CSIC), Spain

***Correspondence:**
Yoshihiko Sako
sako@kais.kyoto-u.ac.jp

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

Received: *05 October 2018*
Accepted: *31 December 2018*
Published: *17 January 2019*

Citation:
Inoue M, Nakamoto I, Omae K, Oguro T, Ogata H, Yoshida T and Sako Y (2019) Structural and Phylogenetic Diversity of Anaerobic Carbon-Monoxide Dehydrogenases. Front. Microbiol. 9:3353. doi: 10.3389/fmicb.2018.03353

INTRODUCTION

The reversible conversion between carbon monoxide (CO) and carbon dioxide (CO_2) catalyzed by CO dehydrogenases (CODHs) constitutes a key reaction in carbon fixation and energy conservation (King and Weber, 2007; Oelgeschläger and Rother, 2008; Sokolova et al., 2009; Nitschke and Russell, 2013; Can et al., 2014). The reaction ($CO_2 + 2H^+ + 2e^- <=> CO + H_2O$) serves as the initial step of carbon fixation via the carbonyl branch of the Wood-Ljungdahl pathway (WLP) in methanogens and acetogens (Nitschke and Russell, 2013; Schuchmann and Müller, 2014). This step is catalyzed by an Ni-containing CODH (Ni-CODH) subunit of the CODH/acetyl-CoA synthase (ACS) complex,

which is also called the acetyl-CoA decarbonylase/synthase complex (hereafter referred to as ACS), with the CO intermediate being finally converted to acetyl-CoA by this complex (Doukov et al., 2002; Gong et al., 2008; Can et al., 2014). Additionally, organisms termed carboxydotrophs utilize CO as an energy source through this reaction due to its low redox potential (King and Weber, 2007; Oelgeschläger and Rother, 2008; Sokolova et al., 2009). Aerobic carboxydotrophs, such as *Oligotropha carboxidovorans*, use oxygen as a terminal electron acceptor from CO oxidation, with Mo- and Cu-containing CODHs belonging to the xanthine oxidase family (King and Weber, 2007; Hille et al., 2015). By contrast, anaerobic carboxydotrophs use various types of terminal electron acceptors, such as CO_2, proton, sulfate, and ferric iron with Ni-CODHs (Oelgeschläger and Rother, 2008; Sokolova et al., 2009). For example, methanogens and acetogens use the reductive potential from CO for carbon fixation via methyl branch of WLPs (Oelgeschläger and Rother, 2008; Nitschke and Russell, 2013). In addition, hydrogenogenic carboxydotrophs, such as *Carboxydothermus hydrogenoformans* and *Rhodospirillum rubrum*, couple CO oxidation with proton reduction to produce hydrogen, in which the proton motive force might be generated with residual energy via the CODH/energy converting hydrogenase (ECH) complex (Fox et al., 1996; Soboh et al., 2002).

Ni-CODHs are divided into CooS type, which are more frequent in bacteria, and Cdh type, almost all of which are found in archaea (Techtmann et al., 2012). Structural analyses of CooS-type Ni-CODHs, such as *C. hydrogenoformans* CooSII, *R. rubrum* CooS, and *Moorella thermoacetica* AcsA, showed that the homodimeric Ni-CODH contains five metal clusters, with each subunit containing a catalytic C-cluster comprising Ni, Fe, and S along with a [4Fe-4S] cubane-type cluster (B-cluster), and a subunit interface of the dimer that accommodates an additional [4Fe-4S] cluster (D-cluster) (Dobbek et al., 2001; Drennan et al., 2001; Doukov et al., 2002). The acid–base catalysts His and Lys are located around the C-cluster (Jeoung and Dobbek, 2007; Fesseler et al., 2015). Mutations in the catalytic C-cluster and the acid–base catalysts have shown decreased activities in CO oxidation (Kim et al., 2004; Jeon et al., 2005). Intramolecular electron transfer is conducted among these three types of metal clusters, whereas intermolecular electron transfer occurs between the solvent-accessible D-cluster and ferredoxin-like proteins, such as CooF (Dobbek et al., 2001; Singer et al., 2006). Moreover, structural analysis of the *Methanosarcina barkeri* ACS α subunit showed that Cdh-type Ni-CODHs are characterized by an additional two [4Fe-4S] cubane-type clusters (E- and F-clusters) (Gong et al., 2008). These analyses suggest that the E- and F-clusters conduct intermolecular electron transfer to ferredoxin-like proteins, whereas the D-cluster conducts electron transfer to another possible electron carrier, flavin adenine dinucleotide (Gong et al., 2008).

Ni-CODHs belong to the hybrid-cluster protein (HCP) family (as *Prismane* in the Pfam database) comprising Ni-CODHs and HCPs (Finn et al., 2016). Unlike Ni-CODHs, HCP is a monomeric protein that has two metal clusters: a catalytic hybrid cluster comprising Fe, S, and O; and a [4Fe-4S] cluster, corresponding to the C- and B-clusters of Ni-CODHs (Aragão et al., 2008). Biochemical and genetic studies showed that HCPs exhibit no CODH activity, but retain the nitric oxide reductase, hydroxylamine reductase, and peroxidase activities necessary to protect cells against nitrosative and oxidative stresses (Wolfe et al., 2002; Almeida et al., 2006; Yurkiw et al., 2012; Wang et al., 2016). By contrast, Ni-CODH C-cluster variants exhibit decreased CODH activity but increased hydroxylamine reductase activity, implying catalytic similarity between these two enzymes (Heo et al., 2002; Inoue et al., 2013).

The functions of Ni-CODHs have been predicted from their genomic contexts (Matson et al., 2011; Techtmann et al., 2012). The component genes for the CODH/ACS complex are often located in gene clusters along with genes encoding their accessory factors, such as the corrinoid iron-sulfur protein and methyltransferase. Additionally, the CODH/ACS gene clusters are widely distributed in bacteria and archaea (Shin et al., 2016; Adam et al., 2018). The component genes for the CODH/ECH complex in hydrogenogenic carboxidotrophs also form gene clusters along with the genes for CooF, a CO-sensing transcriptional activator CooA, and Ni-insertion accessory proteins (Wu et al., 2005; Sant'Anna et al., 2015; Omae et al., 2017). Accordingly, Techtmann et al. (2012) classified Ni-CODHs into four functional categories: ACS, ECH, CooF, and others. Recently, a novel Ni-CODH-gene cluster containing flavin adenine dinucleotide-dependent NAD(P) oxidoreductase (FNOR) and CooF was also reported and is possibly involved in CO oxidation (Whitham et al., 2015; Geelhoed et al., 2016). Ni-CODHs are also phylogenetically classified into six distinct clades from A to F (Techtmann et al., 2012). Clade A comprises Cdh-type Ni-CODHs, and the others include CooS-type Ni-CODHs. Few Ni-CODHs in clades A, E, and F have been biochemically characterized, with Ni-CODH functions (e.g., ACS, ECH, or CooF) not necessarily related to their phylogenetic clades. Therefore, the majority of Ni-CODHs remain functionally uncharacterized (Techtmann et al., 2012). Additionally, structure-based classification of Ni-CODHs has not yet been addressed.

Current progress in next-generation sequencing techniques has expanded the available information associated with microbial genomes. Here, we expanded the datasets of Ni-CODHs and analyzed their structural variations and genomic contexts in order to reveal previously undescribed structural and functional features. Phylogenetic analysis combined with structural prediction revealed the structural plasticity of the metal clusters and active sites of Ni-CODHs. Furthermore, our genomic context analysis revealed not only novel putative Ni-CODH-associated proteins but also relationships among Ni-CODH structural features, genomic contexts, and taxonomies. Our findings suggest that Ni-CODHs are more diverse in their structures and functions than previously reported.

MATERIALS AND METHODS

Construction of Ni-CODH Datasets

The amino acid sequences corresponding to Ni-CODHs were obtained from the National Center for Biotechnology

Information (NCBI) non-redundant protein sequence database (NCBI Resource Coordinators, 2018) (as of January 4, 2018) through a BLASTp search (Camacho et al., 2009) using *C. hydrogenoformans* CooSII (WP_011343033) and the *M. barkeri* ACS α subunit (WP_011305243) as queries. Overlapping hits were excluded to construct the non-redundant dataset. Low-scoring and short-length hits (bit score <200; amino acid length <550), including HCPs and partial fragments, were excluded from the dataset (i.e., the non-redundant potential Ni-CODH dataset). Additionally, a multiple sequence alignment of the non-redundant Ni-CODH sequences was prepared using the MAFFT version 7.310 program with the E-INS-i method (Katoh and Standley, 2013). According to the alignment, sequences having no deletions in the D-, B-, and C-clusters and the catalytic residues were used for subsequent steps. To exclude Ni-CODH sequences without genomic information, 2665 Ni-CODH-encoding genomes were curated from ∼130,000 prokaryotic genomes in the NCBI assembly database in GFF3 format (NCBI Resource Coordinators, 2018) (as of January 4, 2018). These contained both genomes from cultured organisms and metagenome-assembled genomes. Moreover, genomes of same species by the NCBI taxonomy were excluded from the dataset unless these genomes encoded Ni-CODHs with different NCBI protein accession numbers in order to construct an unbiased dataset. We selected the genomes in the following order of priority: reference genomes, representative genomes, complete genomes, and draft genomes. For example, there were 1025 Ni-CODH-encoding genomes in *Clostridioides difficile*. In this case, we chose 44 genomes as representatives of this taxon. In the final dataset, the number of non-redundant Ni-CODH protein sequences was 1942 (**Supplementary Table 1**). The numbers of Ni-CODH-encoding genomes and Ni-CODH genes were 1375 and 2241, respectively (**Supplementary Table 2**). It should be noted that CODH genes annotated as pseudogenes were not included in the dataset although it has been reported that a split CODH gene functions *in vivo* by possible read-through event (Liew et al., 2016). The workflow of data retrieval is summarized in **Supplementary Figure 1**.

Phylogenetic and Structural Classification of Ni-CODHs

The 1942 Ni-CODH protein sequences obtained were realigned with the MAFFT version 7.390 program using the E-INS-i method (Katoh and Standley, 2013). The alignment was subsequently trimmed using the trimAl version 1.4.1 program with a gap-threshold value of 0.9 (Capella-Gutiérrez et al., 2009). A phylogenetic tree was then constructed using the FastTree version 2.1.10 program (Price et al., 2010) with an approximate-maximum-likelihood method using the WAG model. Robustness of the topology of the phylogenetic trees was evaluated by local bootstrap values based on 1000 re-samples. The six major clades (clades A–F) and the novel clade G of the tree were assigned as previously reported (Techtmann et al., 2012; **Figure 1** and **Supplementary Table 1**). The tree was drawn using iTOL version 4.2.4 software (Letunic and Bork, 2016).

To further classify Ni-CODH protein sequences in each phylogenetic clade into structural groups, we analyzed the variation patterns of the residues comprising the D-, B-, and C-clusters and the acid–base catalysts using multiple sequence alignment (**Supplementary Table 1**). Additionally, novel motifs were manually searched and classified into different structural groups (see Results).

Structural Prediction Analysis

Structural prediction of a representative protein in each structural group with unknown three-dimensional structure was performed by the homology-modeling server SWISS-MODEL (Biasini et al., 2014). The 3D structures of *M. barkeri* ACS α subunit (PDB ID: 3CF4) and *C. hydrogenoformans* CooSII (PDB ID, 4UDY) were used as templates for structural groups in clade A and in clades B to G, respectively. All models were predicted as homodimers, with the results summarized in **Supplementary Table 3**. All of the molecular graphics were generated using PyMOL (Schrödinger, New York, NY, United States).

Analysis of Ni-CODH-Genomic Contexts

We searched continuous same-stranded genes [i.e., "directons" (Ermolaeva et al., 2001)], including the gene for Ni-CODH. First, Ni-CODH-encoding directons were determined according to the following criteria: the intergenic region was within 300 bp, and the number of protein-coding genes was within 15 genes upstream and downstream of the gene locus for Ni-CODH. Next, all proteins encoded in the Ni-CODH-encoding genomes were annotated with Clusters of Orthologous Groups of proteins (COGs) (Galperin et al., 2015) through RPS-BLAST searches (e-value $< 10^{-6}$) using the NCBI Conserved Domain Database (Marchler-Bauer et al., 2017). Except for the COGs of Ni-CODHs (COG1151 and COG1152), significantly enriched COGs within the Ni-CODH-encoding directons were extracted using Fisher's exact test ($p < 0.05$), followed by the Benjamini–Hochberg false-discovery rate control (Benjamini and Hochberg, 1995) (false-discovery rate <0.05) (**Supplementary Table 4**). Furthermore, the genomic contexts of genes for Ni-CODHs were curated by serially adding flanking genes annotated with the significantly enriched COGs to the Ni-CODH-encoding directon, regardless of their directions (**Supplementary Table 2**).

The presence or absence of the significantly enriched COGs, structural groups of Ni-CODHs, and taxonomies within each genomic context of the Ni-CODH gene were assigned to a binary matrix. A similarity scores based on the Simpson coefficient for all combinations between COGs (≥25 Ni-CODH-containing genomic loci within each COG), structural groups, and taxonomies were calculated from the binary matrix using R version 3.4.2 and its package "proxy" version 0.4–19 (**Supplementary Table 5**). Network analysis was performed to predict the relationships among Ni-CODH-related COGs, structural groups, and taxonomies from the similarity scores (the Simpson coefficient ≥0.4) using the R package "reshape2" version 1.4.3 and visualized using Cytoscape version 3.6.0 (Shannon et al., 2003).

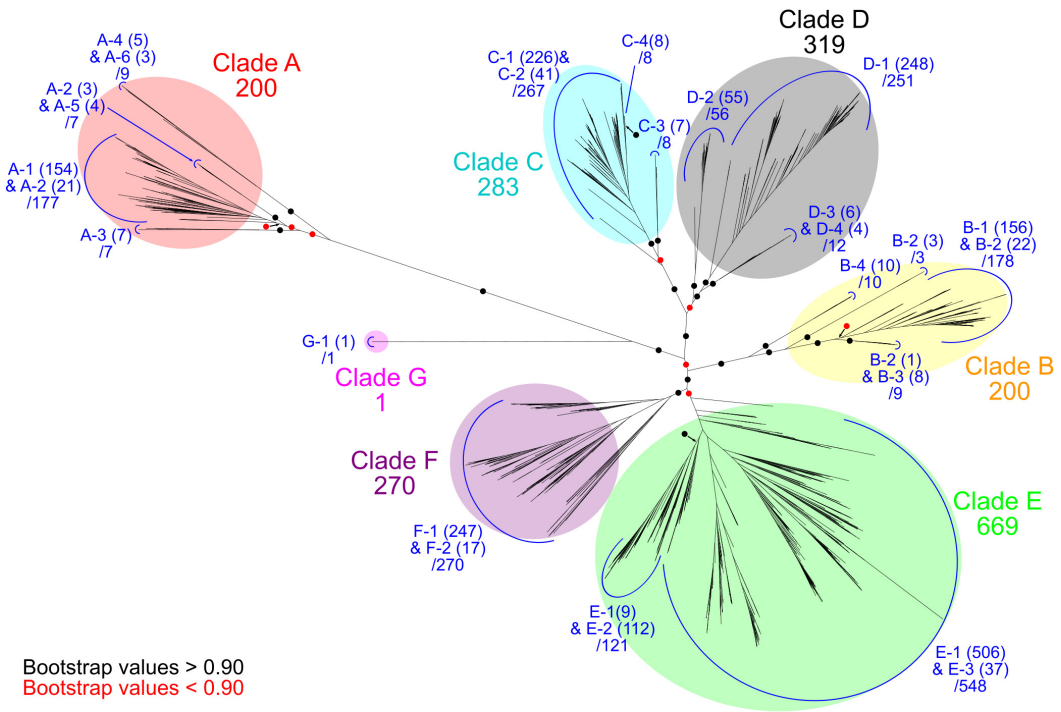

FIGURE 1 | Phylogenetic relationships of Ni-CODHs. An unrooted phylogenetic tree was constructed using an alignment of 1942 Ni-CODHs. Major clades A through G are highlighted by different colors as follows: *red*, clade A; *yellow*, clade B; *cyan*, clade C; *gray*, clade D; *green*, clade E; *purple*, clade F; and *magenta*, clade G. The number of proteins in each clade is designated. Subclades are assigned by the structural groups and colored in *blue* (related to **Figure 2**). The number of proteins in each structural group is shown in parentheses, and the total number of proteins in each subclade is designated. The branches separating the major clades or subclades are encircled and colored according to their bootstrap values, with *black* and *red circles* indicating >0.90 and <0.90 support, respectively.

RESULTS

An Expanded Dataset of Ni-CODHs

We retrieved 2241 genes for Ni-CODH (1942 non-redundant protein sequences) from 1375 genomes including genomes from cultured organisms and metagenome-assembled genomes across 36 phyla and unclassified bacteria (**Table 1** and **Supplementary Tables 1, 2**). Overall, two percent of total prokaryotic genomes encoded at least one Ni-CODH gene and 40 percent of the Ni-CODH-encoding genomes possessed more than one Ni-CODH gene. Our genome dataset was ∼eightfold larger than that from a previous study by Techtmann et al. (2012), which presented a dataset of 292 Ni-CODHs and 180 genomes across eight phyla. The newly identified 28 phyla containing Ni-CODHs were scattered around the tree of life (Hug et al., 2016). A phylogenetic tree of Ni-CODHs was constructed using 1942 non-redundant protein sequences (**Figure 1**). All Ni-CODHs were divided into the previously described six major clades (A–F), except for a novel clade G comprising a single protein (**Figure 1** and **Supplementary Table 1**; Techtmann et al., 2012).

Structural and Phylogenetic Diversity of Ni-CODHs

According to multiple sequence alignment analysis, we classified each phylogenetic clade into structural groups based on the variation patterns of the catalytically important residues (**Figure 1**, **Table 2**, and **Supplementary Table 1**). The multiple sequence alignment of the representative sequence in each structural group is shown in **Figure 2**. The representative sequences were chosen by the following order of priority: 3D structures are known; genes or proteins are previously described; and RefSeq sequences are available. The sequence motifs of the D-clusters were classified into three types as follows: type I, -Cys-X-X-Cys-Cys-; type II, -Cys-$(X)_{7-13}$-Cys-Cys-; and type III, -Cys-X-X-Cys-$(X)_5$-Cys- (**Figure 2**, **Table 2**, and **Supplementary Table 1**). Additionally, we found that there were D-cluster regions with only one or no Cys residues (see below), as previously reported (Lindahl, 2002). The four Cys residues in the B-cluster and the C-terminal three Cys residues in the C-clusters were highly conserved in all Ni-CODHs with a few exceptions, whereas the other residues in the C-cluster, the acid–base catalysts, and the E- and F-clusters were variable (**Figure 2**, **Table 2**, and **Supplementary Table 1**). Moreover, we found novel Cys motifs predicted to form metal clusters at the N-terminus and insertions, as well as His-rich extensions predicted to bind metal ions at N- or C-termini. Ni-CODHs fused to other proteins were also manually searched and classified (**Supplementary Table 1**). Based on these variations in primary structures, 1921 of the 1942 Ni-CODHs were classified into 24 structural groups, with these groups containing two or fewer proteins grouped into

TABLE 1 | The distribution of Ni-CODHs in various phyla.

Phylum	No. of genomes
Firmicutes	748 (104)[a]
Proteobacteria	232 (30)
Euryarchaeota	199 (36)
Chloroflexi	34
Planctomycetes	28 (2)
Nitrospirae	21
Candidatus Bathyarchaeota	13
Candidatus Omnitrophica	12
Thermodesulfobacteria	9
Aquificae	7 (4)
Lentisphaerae	6
Nitrospinae	6
Spirochaetes	6
Actinobacteria	5
Bacteroidetes	5
Elusimicrobia	5
Armatimonadetes	4
Synergistetes	4
Candidatus Kryptonia	3
Chlorobi	3 (2)
Unclassified bacteria	3
Acidobacteria	2
Candidate division WOR-3	2
Candidatus Aminicenantes	2
Candidatus Desantisbacteria	2
Candidatus Thorarchaeota	2
Crenarchaeota	2
Caldiserica	1
Candidatus Fraserbacteria	1
Candidatus Geothermarchaeota	1
Candidatus Korarchaeota	1 (1)
Candidatus Lokiarchaeota	1
Candidatus Pacearchaeota	1
Candidatus Raymondbacteria	1
Chrysiogenetes	1 (1)
Deferribacteres	1
Thaumarchaeota	1
Total	1375 (180)

[a] *The numbers in parentheses are from a previous study (Techtmann et al., 2012).*

"others," except for structural group G-1 (**Figure 2**, **Table 2**, and **Supplementary Table 1**). The phylogenetic tree showed that some of these structural groups formed subclades, whereas others did not, implying the evolutionary plasticity of Ni-CODHs (**Figure 1**).

Clade A Ni-CODHs were divided into six structural groups (A-1–6) (**Figure 2**, **Table 2**, and **Supplementary Table 1**). Structural group A-1 was typical of clade A Ni-CODHs, such as the *M. barkeri* ACS α subunit characterized by complete active-site motifs, the type I D-cluster, and the E- and F- clusters (Gong et al., 2008). Conversely, structural groups A-2 through A-6 had no Cys motifs in the region of the D-cluster. Among these, structural group A-3 showed no Cys signatures in the region of E- and F-clusters, whereas structural group A-5 showed an additional two Cys × 4 signatures between the E-and F-clusters. Structural groups A-4 and A-6 featured incomplete active-site signatures in both the C-cluster and acid–base catalysts. Additionally, structural group A-6 had an incomplete F-cluster.

Clade B Ni-CODHs harboring type II D-clusters and complete acid–base catalysts were divided into four structural groups (B-1–4) based on variations in the second Cys residue of the C-cluster and the presence or absence of His-rich extensions (**Figure 2**, **Table 2**, and **Supplementary Table 1**). Structural group B-1 had an Asp or Glu residue in place of the second Cys residue of the C-cluster and the N-terminal His-rich region. In structural groups B-2 and B-3, the second Cys residue of the C-cluster was replaced by an Asp residue, and only structural group B-3 had the C-terminal His-rich region. These His-rich regions containing from 4 to 48 His residues were characterized by the repeat of the -His-X-His- motif and similar to the C-terminal extension of CooJ and the N-terminal extension of HypB, which constitute Ni-chaperones associated with Ni-CODHs and hydrogenases, respectively (Watt and Ludden, 1998; Olson and Maier, 2000).

In all four structural groups (C-1–4) of clade C Ni-CODHs, the residues for the C-cluster were completely conserved, and the acid–base catalysts His and Lys were replaced by Tyr and Gln, respectively (**Figure 2**, **Table 2**, and **Supplementary Table 1**). Structural groups C-1, C-2, and C-4 had type II D-clusters, whereas structural group C-3 had only one Cys residue in the region of the D-cluster. We identified novel Cys motifs in structural groups C-1 and C-4, with structural group C-1 having an additional Cys motif containing two to four Cys residues at the N-terminal region, whereas structural group C-4 showed a Cys × 4 signature between the Tyr residue and the His residue in the C-cluster.

Clade D Ni-CODHs were divided into four structural groups (D-1–4) (**Figure 2**, **Table 2**, and **Supplementary Table 1**). Structural groups D-1 and D-2 had the type III D-cluster, whereas structural groups D-3 and D-4 had the type II D-cluster. In structural group D-1, the first Cys residue of the C-cluster was replaced by a Glu residue. Additionally, structural groups D-1, D-3, and D-4 had a Tyr residue at the position of the acid–base catalyst His, whereas structural group D-2 had an Asn residue. Structural group D-3 had an additional Cys motif containing three Cys residues at the N-terminal region.

All structural groups of clades E and F Ni-CODHs (E-1–3 and F-1–2, respectively) displayed conservation of the complete active-site motifs (**Figure 2**, **Table 2**, and **Supplementary Table 1**). Structural groups E-1, E-3, F-1, and F-2 had the type II D-cluster, whereas structural group E-2 had the type III D-cluster. Structural groups E-3 and F-2 had an additional Cys motif containing two or three Cys residues at the N-terminal region. Most of the well-characterized Ni-CODHs were classified into the structural groups of clades E and F: structural group E-1 contained *Thermococcus onnuriensis* CooSI, whereas structural group E-2 contained *D. vulgaris* CooS. Structural group F-1 contained *C. hydrogenoformans* CooSII and *M. thermoacetica* AcsA, whereas structural group F-2 contained *R. rubrum* CooS (Drennan et al., 2001; Doukov et al., 2002; Jeoung and Dobbek, 2007; Hadj-Saïd et al., 2015; Schut et al., 2016).

TABLE 2 | The comparison of structural signatures among structural groups of Ni-CODHs.

Structural group	D-cluster	C-cluster						Acid-base		E- & F-clusters	Novel structural motifs
		His	Cys1	Cys2	Cys3	Cys4	Cys5	His	Lys		
A-1	type I	H	C	C	C	C	C	H	K	+	
A-2	–	H	C	C	C	C	C	H	K	+	
A-3	–	H	C	C	C	C	C	H	K	–	
A-3	–	H	C	C	C	C	C	H	K	–	
A-4	–	D	P	C	C	C	C	Q	R	+	
A-5	–	H	C	C	C	C	C	H	K	+	Two Cys × 4 motifs between E- and F-clusters
A-6	–	N	S	C	C	C	C	H	L	+[b]	
B-1	type II	H	C	D/E	C	C	C	H	K	–	His-rich region at the N-terminus
B-2	type II	H	C	D	C	C	C	H	K	–	
B-3	type II	H	C	D	C	C	C	H	K	–	His-rich region at the C-terminus
B-4	type II	H	C	C	C	C	C	H	K	–	
C-1	type II	H	C	C	C	C	C	Y	Q	–	A Cys × 2 to 4 motif at the N-terminus
C-2	type II	H	C	C	C	C	C	Y	Q	–	
C-3	–[a]	H	C	C	C	C	C	Y	Q	–	
C-4	type II	H	C	C	C	C	C	Y	Q	–	A Cys × 4 motif between the Tyr and His residues
D-1	type III	H	E	C	C	C	C	Y	K	–	
D-2	type III	H	C	C	C	C	C	N	K	–	
D-3	type II	H	C	C	C	C	C	Y	K	–	A Cys × 3 motif at the N-terminus
D-4	type II	H	C	C	C	C	C	Y	K	–	
E-1	type II	H	C	C	C	C	C	H	K	–	
E-2	type III	H	C	C	C	C	C	H	K	–	
E-3	type II	H	C	C	C	C	C	H	K	–	A Cys × 2 to 3 motif at the N-terminus
F-1	type II	H	C	C	C	C	C	H	K	–	
F-2	type II	H	C	C	C	C	C	H	K	–	A Cys × 2 to 3 motif at the N-terminus
G-1	type III	H	T	C	C	C	C	Y	K	–	

[a]*One Cys residue was substituted to other amino acids.*[b]*One Cys residue in the F-cluster was substituted to other amino acids.*

The novel clade G (namely structural group G-1) comprised only a single protein and was the most deeply branched among CooS-type Ni-CODHs (**Figure 1**). Structural group G-1 had the type III D-cluster, with the acid–base catalyst His and the first Cys residue replaced by Tyr and Thr, respectively (**Figure 2**, **Table 2**, and **Supplementary Table 1**). It should be noted that the sequence was derived from the high quality metagenomic assembly (completeness >80% by CheckM) (Parks et al., 2015; Anantharaman et al., 2016).

Some of these atypical structural motifs emerged across clades (**Figure 2**, **Table 2**, and **Supplementary Table 1**). The type III D-cluster was found in structural groups D-1, D-2, E-2, and G-1. The Tyr residue replacing to the acid–base catalyst His was conserved through structural groups of clades C, D, and G except for structural group D-2. The N-terminal additional Cys motifs were found in structural groups C-1, D-3, E-3, and F-2.

Structural Prediction Analysis Reveals the Plasticity of the D-Clusters and Active Sites

Structural information of Ni-CODHs is limited to structural groups A-1, F-1, and F-2 Ni-CODHs. To obtain structural information regarding variations in the structural motifs, we performed structural prediction by homology modeling (**Supplementary Table 3**). First, we compared the conformation of three types of D-clusters. Structural analyses of the *M. barkeri* ACS α subunit (structural group A-1; PDB ID: 3CF4) containing a type I D-cluster and *C. hydrogenoformans*

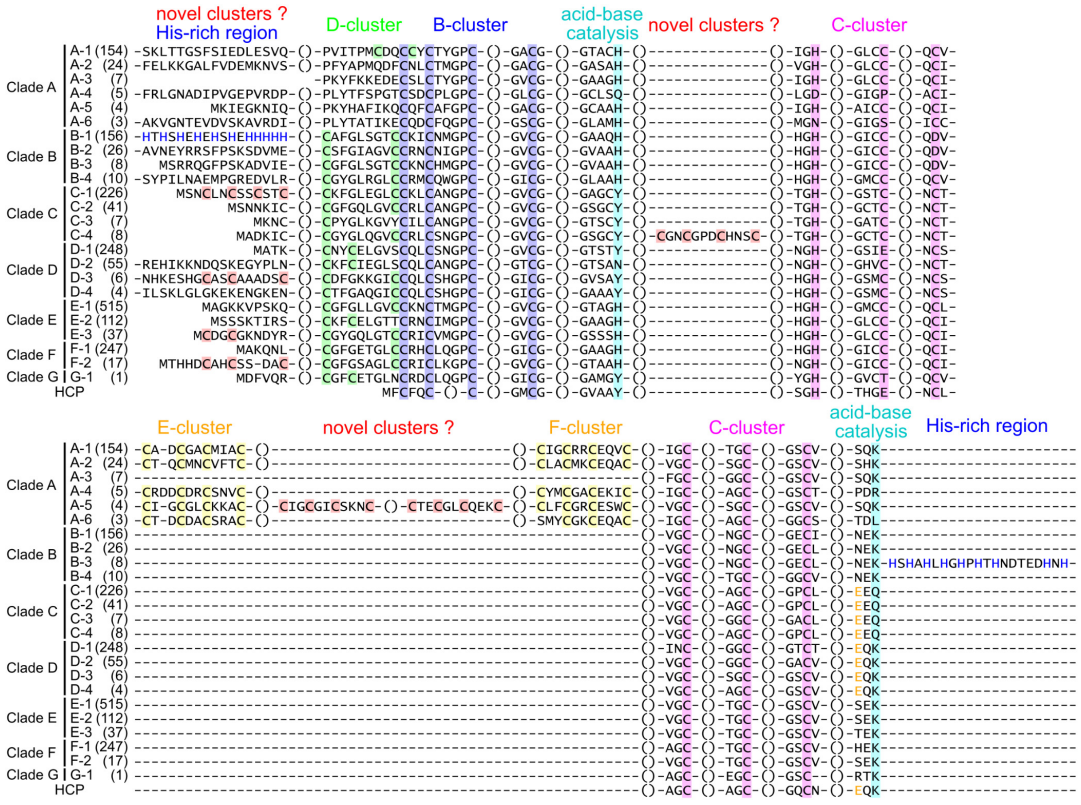

FIGURE 2 | Structural groups of Ni-CODHs. The representative protein sequences of each structural group of Ni-CODHs and HCPs are represented by multiple sequence alignments. The alignment of the N- and C-terminal extensions and the insertions was manually modified for representation. The number of proteins in each structural group is shown in parentheses. The Cys residues of the metal clusters and catalytic residues are highlighted by different color backgrounds as follows: *green*, D-cluster; *blue*, B-cluster; *magenta*, C-cluster; *yellow*, E- and F-clusters; *cyan*, acid-base catalysts; and *red*, putative novel clusters. The His residues in the His-rich regions of structural groups B-1 and B-4 are shown in *blue letters*, and the Glu residues at the equivalent position of Glu494 in the catalytic hybrid cluster of *D. vulgaris* HCP are shown in *yellow letters* (Aragão et al., 2008). The NCBI accession numbers of the representative sequences are as follows: A-1, WP_011305243; A-2, WP_010878596; A-3, OGW06734; A-4, OIP92259; A-5, ODS42986; A-6, OIP30420; B-1, WP_026514536; B-2, WP_015485077; B-3, WP_012645460; B-4, WP_011393470; C-1, WP_039226206; C-2, WP_013237576; C-3, WP_010870233; C-4, WP_044921150; D-1, WP_011342982; D-2, WP_015926279; D-3, WP_079933214; D-4, WP_096205957; E-1, WP_012571978; E-2, WP_010939375; E-3, WP_088535808; F-1, WP_011343033; F-2, WP_011389181; G-1, OGP75751; and HCP, WP_010939296.

CooSII (structural group F-1; PDB ID: 4UDX) containing a type II D-cluster demonstrated that these two D-clusters exhibit different conformations, albeit complete formation of iron-sulfur clusters (**Figures 3A,B**; Gong et al., 2008; Fesseler et al., 2015). The predicted structure of *D. vulgaris* CooS (structural group E-2) with the type III D-cluster showed that the two Cys residues from each subunit faced each other, and that the distance between these four Cys residues was between 4 and 6 Å (**Figure 3C**). These data were consistent with a recent study of the metal-cluster formation of *D. vulgaris* CooS (Hadj-Saïd et al., 2015) and suggested that the Cys motif of the type III D-cluster could form an iron-sulfur cluster in a manner different from other types of D-clusters.

Next, we focused on the structural variations of the active sites. Notably, the sequence alignment of Ni-CODHs with HCPs showed that the eight residues comprising the catalytic site of structural group D-1 were identical to those of HCPs (**Figure 2**; Aragão et al., 2008). Moreover, one additional residue forming the catalytic metal cluster of HCPs [Glu (two residues before the acid–base catalyst Lys)] was conserved in structural group D-1 Ni-CODHs. Structural prediction of *C. hydrogenoformans* CooSV showed that conformation of the active-site residues of structural group D-1 Ni-CODH resembled that of *D. vulgaris* HCP (PDB ID: 1W9M) rather than *C. hydrogenoformans* CooSII (**Figures 4A–C**). The root mean square deviation of these nine residues between *C. hydrogenoformans* CooSV and *D. vulgaris* HCP was 1.2 Å, although the phylogenetic tree of the HCP family showed structural group D-1 Ni-CODHs as being distant from HCPs (**Figure 4D**).

Novel Ni-CODH-Associated Proteins Predicted According to Genomic Contexts

To investigate how such structural features of Ni-CODHs are related to their biological function and to predict the functional properties of each Ni-CODH, we focused on

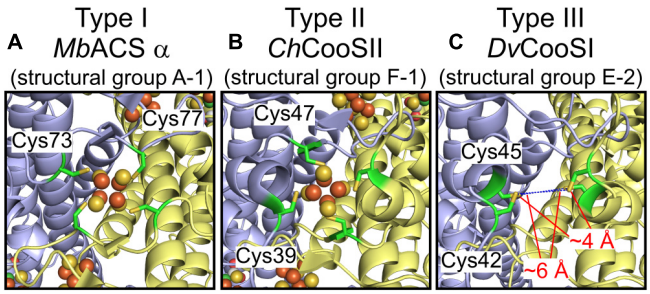

FIGURE 3 | Structural plasticity of the D-clusters. **(A)** The structure of the type I D-cluster of the *M. barkeri* ACS α subunit (structural group A-1; PDB ID: 3CF4). **(B)** The structure of the type II D-cluster of *Carboxydothermus hydrogenoformans* CooSII (structural group F-1; PDB ID: 4UDX). **(C)** The predicted structure of *D. vulgaris* CooS (structural group E-2; NCBI protein accession number: WP_010939375) around the type III D-cluster. All structures are shown in cartoon representation, and each subunit is shown in different colors. The four Cys residues forming the D-cluster are represented as *green sticks*. The Fe and S atoms of the D-clusters are colored in *brown* and *yellow*, respectively. The distances between the Cys residues of the type III D-cluster are indicated.

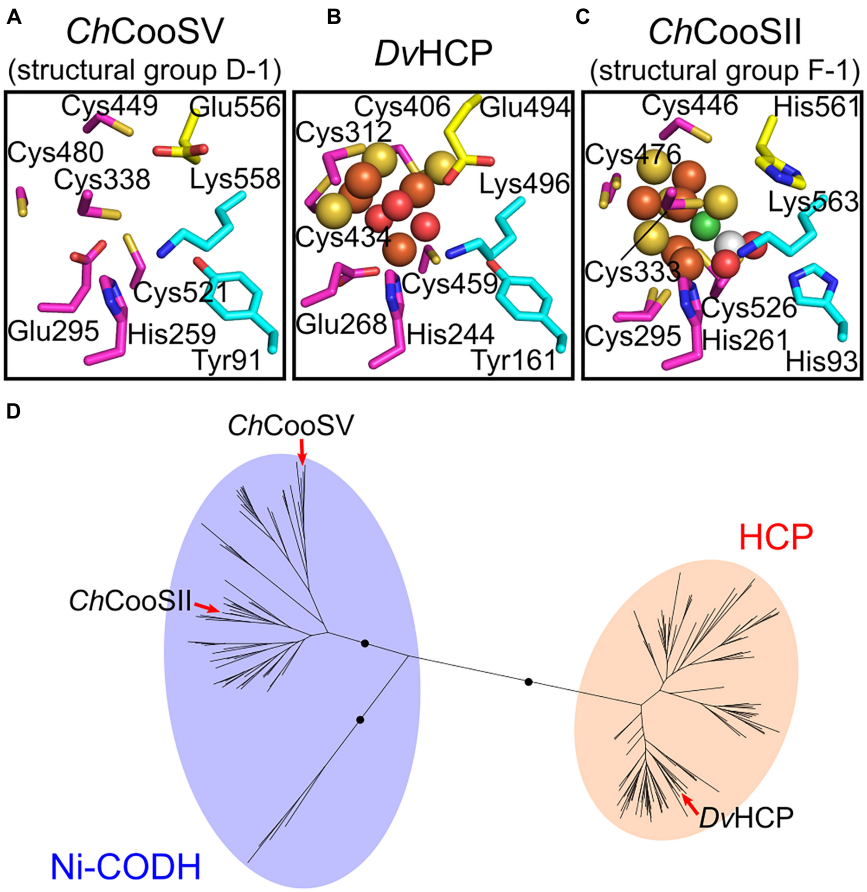

FIGURE 4 | Structural and phylogenetic relationships of the catalytic sites between structural group D-1 Ni-CODHs and HCPs. **(A)** The predicted structure of the catalytic site of *C. hydrogenoformans* CooSV (structural group D-2; NCBI protein accession number: WP_011342982). **(B)** The structure of the catalytic site of *D. vulgaris* HCP (PDB ID: 1W9M). **(C)** The structure of the catalytic site of *C. hydrogenoformans* CooSII (structural group F-1; PDB ID: 4UDX). The residues forming the catalytic sites are represented in *stick forms* and colored as follows (related to **Figure 2**): *magenta*, the C-clusters and hybrid cluster; *cyan*, acid–base catalysts; and *yellow*, the Glu residue unique to the hybrid cluster of HCPs and the corresponding residues of *C. hydrogenoformans* CooSV and CooSII. The Ni, Fe, S, and O atoms of the metal clusters are colored in *green*, *brown*, *yellow*, and *red*, respectively. The C and O atoms of CO_2 are colored in *white* and *red*, respectively. **(D)** The phylogenetic tree of the HCP family from the Pfam seed dataset (PF03063) (Finn et al., 2016). The branches of *C. hydrogenoformans* CooSII and CooSV and *D. vulgaris* HCP are indicated by *red arrows*.

genomic contexts for Ni-CODHs. We classified all proteins encoded in the Ni-CODH-encoding genomes into COGs, followed by enrichment analysis to curate significantly abundant COGs in the Ni-CODH-encoding directons (**Supplementary Table 4**). The significantly enriched COGs (false-discovery rate <0.05) included a series of previously described Ni-CODH-associated proteins, such as CooF (COG0437 and COG1142), CooC (COG3640), CooA (COG0664), FNOR (COG1251), and components of WLP and ECH (**Supplementary Table 4**). Additionally, our analysis identified novel putative Ni-CODH-associated proteins, including various types of oxidoreductases, ABC transporters, and transcription factors, as detailed in subsequent sections. Based on these data, we determined the genomic contexts of genes for Ni-CODH by adding flanking genes annotated with these COGs to the Ni-CODH-encoding directon (**Supplementary Table 2**).

Relationships Between Structural Groups and Genomic Contexts for Ni-CODHs

To elucidate the structural and functional relationships of Ni-CODHs, we performed comprehensive network analysis using similarity scores among structural groups of Ni-CODHs, Ni-CODH-associated proteins, and taxonomies of Ni-CODH-encoding genomes based on the genomic contexts of 2457 genes for Ni-CODH (**Figure 5**, **Supplementary Figures 2, 3**, and **Supplementary Table 5**). The network precisely presented functional gene clusters encoding Ni-CODHs, such as WLP, ECH, and FNOR (Techtmann et al., 2012; Geelhoed et al., 2016; Shin et al., 2016).

Structural groups A-1, A-2, A-4, A-5, A-6, E-1, E-3, and F-1 Ni-CODHs were associated with the largest group comprising proteins involved in both methyl and carbonyl branches of WLP, including the ACS catalytic β subunit (COG1614) and other subunits (**Figure 5** and **Supplementary Figures 2, 3**). This group also contained various types of oxidoreductases, such as glycine-cleavage system components (COG0509 and COG1249), heterodisulfide reductase-like proteins (COG1148, COG1150, COG1908, and COG2048), and NADH:ubiquinone oxidoreductase-like proteins (COG1034, COG1894, and COG1905). Metallochaperones were also found in the WLP group, including CooC and the recently characterized Ni-insertion accessory protein CooT (COG1532) (Timm et al., 2017). Structural group E-2 Ni-CODHs were also associated with CooC, which was in turn related to the ECH group, as well as the WLP group.

Structural groups F-1 and F-2 Ni-CODHs were associated with ECH-related proteins, including [NiFe] hydrogenase catalytic subunits (COG3261 and COG3260), accessory subunits (COG0650, COG0651, and COG1143), and their maturation protease (COG0680) (**Figure 5** and **Supplementary Figures 2, 3**). The metallochaperones HypA (COG0375) and HypB (COG0378), CooA, and CooF (COG1142) were also classified into the ECH group.

Structural groups C-1 and C-2 Ni-CODHs were grouped with FNOR, CooF (COG0437), an NsrR/IscR-like transcriptional regulator (COG1959), and functionally unknown proteins (**Figure 5** and **Supplementary Figures 2, 3**). CooX (COG1143), a ferredoxin-like accessory protein of the ECH complex (Soboh et al., 2002), was clustered into the FNOR group, as well as the ECH group, suggesting potential involvement in Ni-CODH/FNOR function. Notably, two orthologous groups of CooF (COG0437 and COG1142) were mainly categorized into different functional groups: COG0437 was associated with FNOR, whereas COG1142 was associated with ECH. Structural groups D-3 and G-1 Ni-CODHs were also related to FNOR and COG1142, suggesting the functional plasticity of these CooF-like proteins.

Furthermore, the network also revealed novel putative gene clusters for Ni-CODH (**Figure 5** and **Supplementary Figures 2, 3**). All structural groups of clade B Ni-CODHs were grouped with the NitT/TauT family ABC transporter components (COG0600, COG0715, and COG1116). Structural group D-1 Ni-CODHs were associated with a functionally unknown zinc-binding dehydrogenase family protein (COG1063) along with the ECH-related metallochaperones HypA and HypB (Riveros-Rosas et al., 2003; Watanabe et al., 2015). Structural group C-4 was associated with a [FeFe] hydrogenase catalytic subunit (COG4624). A histidine kinase (COG0642) and a response regulator (COG0745), as a two-component system, were not related to any structural groups of Ni-CODHs, suggesting co-existence with this system and various structural groups of Ni-CODHs. The genomic context analysis suggested that these proteins likely exhibit Ni-CODH-associated functions; however, the functional relationships between these proteins and Ni-CODHs remain to be elucidated.

Ni-CODH Genomic Contexts Are Scattered Across Microbial Genomes

Our network analysis also revealed taxonomic profiles of Ni-CODHs (**Figure 5**). Regardless of their phyla or domains, the family level taxa were scattered across the network and associated with various types of Ni-CODHs and their related proteins. A previous study speculated that clade A Ni-CODHs are specific to archaea with one exception, whereas clades B and F are specific to bacteria (Techtmann et al., 2012). In the present study, however, all clades of Ni-CODHs were found in both domains (**Figure 5** and **Supplementary Tables 2, 5**). Among these, structural group A-1 Ni-CODHs were mainly found in archaea, whereas structural groups B-1, C-2, and F-1 Ni-CODHs were mainly found in bacteria, with a few exceptions. Structural groups A-2, A-5, C-3, and D-4 were found only in archaea, whereas structural groups A-3, A-4, A-6, B-2, B-3, B-4, C-1, C-4, D-2, D-3, E-3, F-2, and G-1 were found only in bacteria. Some of these domain-specific structural groups of Ni-CODHs, such as A-3, A-4, A-5, A-6, C-3, D-2, D-3, and D-4, were deeply branched in each clade (**Figure 1** and **Supplementary Table 2**).

DISCUSSION

In this study, we performed a comprehensive analysis of an expanded Ni-CODH dataset focusing on their structures and

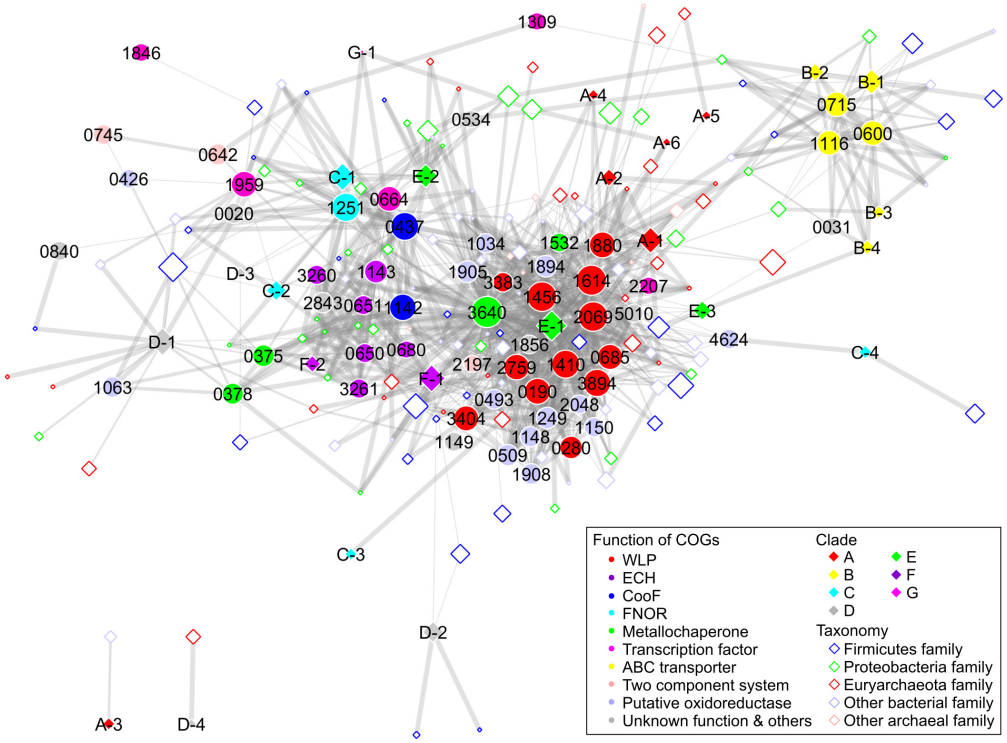

FIGURE 5 | Network analysis of Ni-CODH-related proteins, structural groups of Ni-CODHs, and taxonomies focusing on the genomic contexts. The network is represented by an unweighted force-directed layout and manually modified to avoid the overlapping nodes. The width of the edges is scaled proportionally according to the Simpson coefficient. The size of the nodes is scaled proportionally according to the log of the number of gene loci of Ni-CODHs with each COG, structural group, and taxonomy. The COGs, structural groups of Ni-CODHs, and family level taxonomies are shown by filled circles, filled diamonds, and open diamonds, respectively, and colored according to their functions, clades, and phyla as *inset*, respectively. The taxonomy omitted network is shown in **Supplementary Figure 3** in order to explain the functional groups of the Ni-CODH-related proteins.

genomic contexts in order to classify Ni-CODHs and their related proteins and predict their novel functions. The previous two comprehensive studies for classification of Ni-CODHs have only focused on specific genomic contexts (i.e., CODH/ACS, CODH/ECH, and CODH/CooF) or clades (Techtmann et al., 2012; Adam et al., 2018). Techtmann et al. (2012) focuses on CODH/ACS, CODH/ECH, and CODH/CooF, whereas Adam et al. (2018) focuses on CODH/ACS subunits in more detail. In particular, there are few descriptions in the functions of clades B and D Ni-CODHs and the structural diversity of Ni-CODHs. Our work provided more comprehensive and unbiased descriptions with larger dataset than these previous studies. Our phylogenetic analysis represented not only the previously reported six clades but also the novel clade G (**Figure 1**). Furthermore, our structure-based classification showed that the structural motifs of Ni-CODHs were more diverse and plastic than previously reported (**Table 2** and **Figures 2–4**). Moreover, the motifs consisting of D-clusters were variable, even within the same clade. Additionally, the active-site motifs in well-characterized Ni-CODHs were not conserved in structural groups A-3, A-5, B-1, B-2, and B-3 or in any structural groups in clades C, D, and G.

Our structure-based classification also identified novel structural motifs of Ni-CODHs (**Figure 2** and **Supplementary Table 1**). In particular, structural groups A-5 and C-4 have insertions of novel Cys motifs. The two Cys × 4 motifs in structural group A-5 were tandemly inserted in the same region of the E- and F-clusters, suggesting that additional two [4Fe-4S] clusters could conduct electron transfer. Conversely, one Cys × 4 motif in structural group C-4 were inserted between the Tyr residue and the His residue in the C-cluster predicted to be opposite side of the D-cluster on 3D structure, implying possible functions in protein-protein interaction or non-canonical electron transfer to the [FeFe] hydrogenase catalytic subunit (COG4624). In addition, structural groups C-1, D-3, E-3, and F-2 have another types of additional N-terminal Cys motifs in which the numbers of Cys residues were variable from two to four. The crystal structure of *R. rubrum* CooS indicates that the N-terminal Cys-containing region is disordered (Drennan et al., 2001). We speculated that these Cys-containing extensions might have a metallochaperone-like function. However, the structural and functional properties of these novel Cys motifs remain unknown.

Additionally, structural groups of B-1 and B-3 Ni-CODHs harbor His-rich extensions likely responsible for metal binding similar to the Ni chaperones CooJ and HypB (**Figure 2** and **Supplementary Table 1**; Watt and Ludden, 1998; Olson and Maier, 2000). Our genomic context analysis also identified

NitT/TauT-family ABC transporters as putative clade B Ni-CODH-associated proteins, which consist of three components: an ATP-binding cassette module, a transmembrane module, and a periplasmic solute-binding module (**Figure 5** and **Supplementary Tables 4, 5**). We found that the periplasmic module is homologous to a periplasmic solute metal-binding protein from *Staphylococcus aureus* (PDB ID: 3UN6) (data not shown). The Ni-insertion machinery of Ni-CODHs involves a variety of metallochaperones (Merrouch et al., 2018). CooC is considered a major component in various organisms, whereas CooJ and CooT provide assistance in some organisms, such as *R. rubrum*, suggesting diverse Ni-insertion mechanisms in the maturation of Ni-CODHs. Therefore, we hypothesized that the ABC transporters import metal ions, such as Ni^{2+}, and then the His-rich regions of structural groups B-1 and B-3 Ni-CODHs bind the metal ions and insert them into the catalytic sites of Ni-CODHs.

Our structural prediction analysis suggested that the active-site residues of structural group D-1 Ni-CODHs were similar to those of HCPs, although they were phylogenetically distant from each other (**Figure 4**). A previous study reported that a structural group D-1-mimicking mutant (C295E) of *C. hydrogenoformans* CooSII exhibits not only dramatically decreased Ni-binding and CO-oxidation activities but also increased hydroxylamine reductase activity (Inoue et al., 2013). Therefore, we speculate that structural group D-1 Ni-CODHs likely have similar structural and functional properties as HCPs as a result of convergent evolution. A genetic analysis of *Clostridium autoethanogenum* has also shown that structural group D-1 Ni-CODH is unable to compensate inactivation of AcsA (Liew et al., 2016). However, genetic analysis of *M. acetivorans* suggests potential involvement of structural group D-1 Ni-CODHs in CO oxidation (Rother et al., 2007). Therefore, the biochemical characterization of structural group D-1 Ni-CODHs is required to elucidate their catalytic activity.

Ni-CODH/FNOR represents a recently identified putative functional unit for CO oxidation (Whitham et al., 2015; Geelhoed et al., 2016); however, no biochemical study of this type of FNOR has been reported. Notably, structural groups C-1, C-2, D-3, and G-1 Ni-CODHs were related to FNOR and CooF according to our network analysis (**Figure 5** and **Supplementary Figures 2, 3**). The genes for some structural group D-1, D-2, D-3, E-1, E-2, E-3, and F-1 Ni-CODHs also coexist with FNOR in their genomic contexts (**Supplementary Tables 2, 5**). The deletion mutant of a gene for structural group C-2 Ni-CODH adjacent to FNOR and CooF in *C. autoethanogenum* has exhibited growth deficiency in the presence of CO probably because of the excess reducing equivalent, suggesting its role in CO_2 reduction (Liew et al., 2016). Conversely, transcriptomic analysis of *C. ljungdahlii* showed that expression of the gene cluster for Ni-CODH, comprising structural group C-1 Ni-CODH, FNOR, and CooF, is upregulated upon oxygen exposure, suggesting their use of CO as a source of electrons for oxygen detoxification (Whitham et al., 2015). Similar phenomenon was observed in structural group C-1 Ni-CODH from *C. acetobutylicum* although the genomic context was not conserved (Hillmann et al., 2009). Additionally, a study of *Geobacter sulfurreducens* growth in the presence of CO suggests that a gene cluster encoding structural group E-1 Ni-CODHs, FNOR, and CooF is likely responsible for CO oxidation (Geelhoed et al., 2016). Our analysis suggested that FNOR could be coupled with various structural groups of Ni-CODHs, and that the genes for FNOR are widespread across microbial genomes along with genes encoding Ni-CODHs, such as WLP or ECH.

Additionally, we showed that only structural groups E-1, F-1, and F-2 Ni-CODHs coexisted with the [NiFe] hydrogenase catalytic subunits based on their genomic contexts (**Supplementary Tables 2, 5**). These data imply a requirement of the complete structural motifs for Ni-CODH/ECH coupling. Conversely, all structural groups of Ni-CODHs in clades A, E, and F, except for structural groups A-3 and F-2, coexisted with the ACS β subunit regardless of their complete or incomplete structural motifs. However, it remains necessary to examine whether the structural groups harboring incomplete structural motifs are able to catalyze the CODH/ACS reaction.

Unveiling the specific interactions between Ni-CODH phylogeny and microbial physiology are quite challenging because of possible functional redundancy and horizontal gene transfer (Techtmann et al., 2012; Sant'Anna et al., 2015). Our dataset could provide some interactions between specific structural group and microbial physiology. It has been shown that carboxydotrophic organisms rely on clades A, E, and F Ni-CODHs for CO oxidizing activity (Matschiavelli et al., 2012; Techtmann et al., 2012; Sant'Anna et al., 2015). In our dataset, structural groups B-4 and D-3 Ni-CODHs were also specifically found in carboxydotrophic bacteria in addition to their clade F Ni-CODHs forming gene clusters with ACS or ECH (**Supplementary Table 2**). Structural group B-4 was found in carboxydtrophic genera *Moorella*, *Calderihabitans*, and *Thermanaeromonas* (a candidate carboxydtroph possessing a CODH/ECH gene cluster), whereas structural group D-3 was found in *Carboxydocella* and *Thermanaeromonas*. Interestingly, these structural groups possessed the complete motif for the C-cluster, suggesting possible roles for carboxydotrophic metabolism in these organisms. Conversely, methanogenic archaea depend on clade A Ni-CODHs for their acetoclastic methanogenesis (Gong et al., 2008; Techtmann et al., 2012). However, structural group C-3 was specifically found in methanogenic genera *Methanocaldococcus* and *Methanotorris* in addition to structural group A-1. We also found that structural group D-4 was conserved in anaerobic methane oxidizing genera *Candidatus Methanoperedens* with structural group A-2. These structural groups of Ni-CODHs might have possible roles in their unique metabolisms.

Our phylogenetic analysis also provided information clarifying the evolutionary history of Ni-CODHs. In Cdh-type Ni-CODHs (clade A), structural groups A-2 through A-6 not harboring the D-cluster were deeply branched, with structural groups A-2, A-4, A-5, and A-6 being more deeply branched than structural group A-3 not harboring the E- and F-clusters (**Figures 1, 2**). These data imply that the ancestor of Cdh-type Ni-CODHs did not likely harbor the D-cluster but rather the E- and F- clusters. This idea might be supported by a previous structural study of the *M. barkeri* ACS α subunit, which suggested that the E- and F-clusters are responsible for electron transfer

to or from a ferredoxin-like protein rather than the D-cluster (Gong et al., 2008). In fact, structural groups A-2, A-4, A-5, and A-6 formed gene cluster with ACS subunits, suggesting possible involvements for WLP without D-cluster. Therefore, acquisition of the type I D-cluster in structural group A-1 might enable electron transfer to flavin adenine dinucleotide.

Moreover, in the CooS-type Ni-CODHs (clades B–G), structural groups harboring type II D-clusters were deeply branched in clades D and E, whereas structural group C-3 harboring the incomplete D-cluster was deeply branched in clade C, and structural group G-1 had a type III D-cluster. Regarding the catalytic site, structural groups B-4, D-2, D-3, and D-4 harboring the complete motif required to form the C-cluster were deeply branched, implying that mutations in the C-cluster motifs occurred after branching into each clade, and that the ancestor of clades B through F likely harbored the complete motif necessary to form the C-cluster. Conversely, structural group G-1 had a mutation in the second Cys residue of the C-cluster. These structural and evolutionary plasticities in CooS-type Ni-CODHs render it difficult to infer their ancestral form. Therefore, additional sequence information for clade G Ni-CODHs in particular is required to ascertain the ancestor of the CooS-type Ni-CODHs.

Previous studies suggested that the CODH/ACS complex likely existed in the last universal common ancestor (LUCA) (Weiss et al., 2016; Adam et al., 2018). However, it has been difficult to predict the Ni-CODHs that existed in the LUCA based on phylogenetic analysis of Ni-CODHs because of the horizontal gene transfer between domains. In our analysis, some of the domain-specific structural groups of Ni-CODHs were deeply branched in clades A, C, and D (**Figure 1** and **Supplementary Table 2**). These data raise the possibilities that the ancestor of Ni-CODHs in each clade might have existed in the LUCA or might have been transferred between domains early after branching of each clade.

Conversely, we found a possible horizontal gene transfer across the domains in shallow branches of the tree. Only one archaeon *Methanosarcina horonobensis*, was found to possess a clade B Ni-CODH (structural group B-1) exhibiting high similarities (sequence identities >70%) to those from *Caldicoprobacter* and *Ruminiclostridium* genera. Furthermore, the structural group B-1 Ni-CODH from *M. horonobensis* was found in putative gene cluster of ABC transporters like other clade B Ni-CODHs. The three components of ABC transporters from *M. horonobensis* also showed high similarities (sequence identities >60%) to those from *Caldicoprobacter* and *Ruminiclostridium* genera. These results imply that there would be a variety of evolutionary histories of Ni-CODHs.

Overall, our expanded dataset and genomic context analysis revealed novel putative Ni-CODH-associated proteins and relationships among Ni-CODH structural features, genomic contexts, and taxonomies, suggesting unanticipated structural and functional diversity and plasticity. The enhanced understanding of biochemically uncharacterized Ni-CODHs, their associated protein functions, and Ni-CODH evolutionary history will likely facilitate future research in anaerobic microbe metabolism.

AUTHOR CONTRIBUTIONS

MI, TY, and YS conceived and designed the work. MI, IN, and TO constructed the datasets of Ni-CODHs and performed the phylogenetic analyses. MI performed the comparative sequence and structural prediction analyses. MI, IN, KO, and HO performed genomic context analysis. MI, TY, and YS wrote the paper with help from KO and HO.

FUNDING

This work was supported by a Grant-in-Aid for Scientific Research (16H06381 to YS) from the Ministry of Education, Culture, Sports, Science, and Technology (MEXT).

ACKNOWLEDGMENTS

The part of the computational analysis was performed at the Super Computer System, Institute for Chemical Research, Kyoto University.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.03353/full#supplementary-material

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.1716667115

Almeida, C. C., Romão, C. V., Lindley, P. F., Teixeira, M., and Saraiva, L. M. (2006). The role of the hybrid cluster protein in oxidative stress defense. *J. Biol. Chem.* 281, 32445–32450. doi: 10.1074/jbc.M605888200

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

Aragão, D., Mitchell, E. P., Frazão, C. F., Carrondo, M. A., and Lindley, P. F. (2008). Structural and functional relationships in the hybrid cluster protein family: structure of the anaerobically purified hybrid cluster protein from *Desulfovibrio vulgaris* at 1.35 Å resolution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 64, 665–674. doi: 10.1107/S0907444908009165

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.* 57, 289–300. doi: 10.2307/2346101

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., et al. (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* 42, W252–W258. doi: 10.1093/nar/gku340

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. doi: 10.1186/1471-2105-10-421

Can, M., Armstrong, F. A., and Ragsdale, S. W. (2014). Structure, function, and mechanism of the nickel metalloenzymes, CO dehydrogenase, and acetyl-CoA synthase. *Chem. Rev.* 114, 4149–4174. doi: 10.1021/cr400461p

Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348

Dobbek, H., Svetlitchnyi, V., Gremer, L., Huber, R., and Meyer, O. (2001). Crystal structure of a carbon monoxide dehydrogenase reveals a [Ni-4Fe-5S] cluster. *Science* 293, 1281–1285. doi: 10.1126/science.1061500

Doukov, T. I., Iverson, T. M., Seravalli, J., Ragsdale, S. W., and Drennan, C. L. (2002). A Ni-Fe-Cu center in a bifunctional carbon monoxide dehydrogenase/acetyl-CoA synthase. *Science* 298, 567–572. doi: 10.1126/science.1075843

Drennan, C. L., Heo, J., Sintchak, M. D., Schreiter, E., and Ludden, P. W. (2001). Life on carbon monoxide: X-ray structure of *Rhodospirillum rubrum* Ni-Fe-S carbon monoxide dehydrogenase. *Proc. Natl. Acad. Sci. U.S.A.* 98, 11973–11978. doi: 10.1073/pnas.211429998

Ermolaeva, M. D., White, O., and Salzberg, S. L. (2001). Prediction of operons in microbial genomes. *Nucleic Acids Res.* 29, 1216–1221. doi: 10.1093/nar/29.5.1216

Fesseler, J., Jeoung, J. H., and Dobbek, H. (2015). How the [$NiFe_4S_4$] Cluster of CO Dehydrogenase Activates CO_2 and NCO^-. *Angew. Chem. Int. Ed.* 54, 8560–8564. doi: 10.1002/anie.201501778

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., et al. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44, D279–D285. doi: 10.1093/nar/gkv1344

Fox, J. D., He, Y., Shelver, D., Roberts, G. P., and Ludden, P. W. (1996). Characterization of the region encoding the CO-induced hydrogenase of *Rhodospirillum rubrum*. *J. Bacteriol.* 178, 6200–6208. doi: 10.1128/jb.178.21.6200-6208.1996

Galperin, M. Y., Makarova, K. S., Wolf, Y. I., and Koonin, E. V. (2015). Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res.* 43, D261–D269. doi: 10.1093/nar/gku1223

Geelhoed, J. S., Henstra, A. M., and Stams, A. J. M. (2016). Carboxydotrophic growth of *Geobacter sulfurreducens*. *Appl. Microbiol. Biotechnol.* 100, 997–1007. doi: 10.1007/s00253-015-7033-z

Gong, W., Hao, B., Wei, Z., Ferguson, D. J., Tallant, T., Krzycki, J. A., et al. (2008). Structure of the $\alpha_2\varepsilon_2$ Ni-dependent CO dehydrogenase component of the *Methanosarcina barkeri* acetyl-CoA decarbonylase/synthase complex. *Proc. Natl. Acad. Sci. U.S.A.* 105, 9558–9563. doi: 10.1073/pnas.0800415105

Hadj-Saïd, J., Pandelia, M. E., Léger, C., Fourmond, V., and Dementin, S. (2015). The carbon monoxide dehydrogenase from *Desulfovibrio vulgaris*. *Biochim. Biophys. Acta* 1847, 1574–1583. doi: 10.1016/j.bbabio.2015.08.002

Heo, J., Wolfe, M. T., Staples, C. R., and Ludden, P. W. (2002). Converting the NiFeS carbon monoxide dehydrogenase to a hydrogenase and a hydroxylamine reductase. *J. Bacteriol.* 184, 5894–5897. doi: 10.1128/JB.184.21.5894-5897.2002

Hille, R., Dingwall, S., and Wilcoxen, J. (2015). The aerobic CO dehydrogenase from *Oligotropha carboxidovorans*. *J. Biol. Inorg. Chem.* 20, 243–251. doi: 10.1007/s00775-014-1188-4

Hillmann, F., Döring, C., Riebe, O., Ehrenreich, A., Fischer, R. J., and Bahl, H. (2009). The role of PerR in O_2-affected gene expression of *Clostridium acetobutylicum*. *J. Bacteriol.* 191, 6082–6093. doi: 10.1128/JB.00351-09

Hug, L. A., Baker, B. J., Anantharaman, K., Brown, C. T., Probst, A. J., Castelle, C. J., et al. (2016). A new view of the tree of life. *Nat. Microbiol.* 1:16048. doi: 10.1038/nmicrobiol.2016.48

Inoue, T., Takao, K., Yoshida, T., Wada, K., Daifuku, T., Yoneda, Y., et al. (2013). Cysteine 295 indirectly affects Ni coordination of carbon monoxide dehydrogenase-II C-cluster. *Biochem. Biophys. Res. Commun.* 441, 13–17. doi: 10.1016/j.bbrc.2013.09.143

Jeon, W. B., Singer, S. W., Ludden, P. W., and Rubio, L. M. (2005). New insights into the mechanism of nickel insertion into carbon monoxide dehydrogenase: analysis of *Rhodospirillum rubrum* carbon monoxide dehydrogenase variants with substituted ligands to the [Fe_3S_4] portion of the active-site C-cluster. *J. Biol. Inorg. Chem.* 10, 903–912. doi: 10.1007/s00775-005-0043-z

Jeoung, J. H., and Dobbek, H. (2007). Carbon dioxide activation at the Ni,Fe-cluster of anaerobic carbon monoxide dehydrogenase. *Science* 318, 1461–1464. doi: 10.1126/science.1148481

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Kim, E. J., Feng, J., Bramlett, M. R., and Lindahl, P. A. (2004). Evidence for a proton transfer network and a required persulfide-bond-forming cysteine residue in Ni-containing carbon monoxide dehydrogenases. *Biochemistry* 43, 5728–5734. doi: 10.1021/bi036062u

King, G. M., and Weber, C. F. (2007). Distribution, diversity and ecology of aerobic CO-oxidizing bacteria. *Nat. Rev. Microbiol.* 5, 107–118. doi: 10.1038/nrmicro1595

Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 44, W242–W245. doi: 10.1093/nar/gkw290

Liew, F., Henstra, A. M., Winzer, K., Köpke, M., Simpson, S. D., and Minton, N. P. (2016). Insights into CO_2 fixation pathway of *Clostridium autoethanogenum* by targeted mutagenesis. *mBio* 7:e00427-16. doi: 10.1128/mBio.00427-16

Lindahl, P. A. (2002). The Ni-containing carbon monoxide dehydrogenase family: light at the end of the tunnel? *Biochemistry* 41, 2097–2105. doi: 10.1021/bi015932+

Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C. J., Lu, S., et al. (2017). CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* 45, D200–D203. doi: 10.1093/nar/gkw1129

Matschiavelli, N., Oelgeschläger, E., Cocchiararo, B., Finke, J., and Rother, M. (2012). Function and regulation of isoforms of carbon monoxide dehydrogenase/acetyl coenzyme A synthase in *Methanosarcina acetivorans*. *J. Bacteriol.* 194, 5377–5387. doi: 10.1128/JB.00881-12

Matson, E. G., Gora, K. G., and Leadbetter, J. R. (2011). Anaerobic carbon monoxide dehydrogenase diversity in the homoacetogenic hindgut microbial communities of lower termites and the wood roach. *PLoS One* 6:e19316. doi: 10.1371/journal.pone.0019316

Merrouch, M., Benvenuti, M., Lorenzi, M., Léger, C., Fourmond, V., and Dementin, S. (2018). Maturation of the [Ni–4Fe–4S] active site of carbon monoxide dehydrogenases. *J. Biol. Inorg. Chem.* 23, 613–620. doi: 10.1007/s00775-018-1541-0

NCBI Resource Coordinators (2018). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 46, D8–D13. doi: 10.1093/nar/gkx1095

Nitschke, W., and Russell, M. J. (2013). Beating the acetyl coenzyme A-pathway to the origin of life. *Philos. Trans. R. Soc. B Biol. Sci.* 368:20120258. doi: 10.1098/rstb.2012.0258

Oelgeschläger, E., and Rother, M. (2008). Carbon monoxide-dependent energy metabolism in anaerobic bacteria and archaea. *Arch. Microbiol.* 190, 257–269. doi: 10.1007/s00203-008-0382-6

Olson, J. W., and Maier, R. J. (2000). Dual roles of *Bradyrhizobium japonicum* nickelin protein in nickel storage and GTP-dependent Ni mobilization. *J. Bacteriol.* 182, 1702–1705. doi: 10.1128/JB.182.6.1702-1705.2000

Omae, K., Yoneda, Y., Fukuyama, Y., Yoshida, T., and Sako, Y. (2017). Genomic analysis of *Calderihabitans maritimus* KKC1, a thermophilic, hydrogenogenic, carboxydotrophic bacterium isolated from marine sediment. *Appl. Environ. Microbiol.* 83:e00832-17. doi: 10.1128/AEM.00832-17

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

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

Riveros-Rosas, H., Julián-Sánchez, A., Villalobos-Molina, R., Pardo, J. P., and Piña, E. (2003). Diversity, taxonomy and evolution of medium-chain dehydrogenase/reductase superfamily. *Eur. J. Biochem.* 270, 3309–3334. doi: 10.1046/j.1432-1033.2003.03704.x

Rother, M., Oelgeschläger, E., and Metcalf, W. W. (2007). Genetic and proteomic analyses of CO utilization by *Methanosarcina acetivorans*. *Arch. Microbiol.* 188, 463–472. doi: 10.1007/s00203-007-0266-1

Sant'Anna, F. H., Lebedinsky, A. V., Sokolova, T. G., Robb, F. T., and Gonzalez, J. M. (2015). Analysis of three genomes within the thermophilic bacterial species *Caldanaerobacter subterraneus* with a focus on carbon monoxide dehydrogenase evolution and hydrolase diversity. *BMC Genomics* 16:757. doi: 10.1186/s12864-015-1955-9

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

Schut, G. J., Lipscomb, G. L., Nguyen, D. M. N., Kelly, R. M., and Adams, M. W. W. (2016). Heterologous production of an energy-conserving carbon monoxide dehydrogenase complex in the hyperthermophile *Pyrococcus furiosus*. *Front. Microbiol.* 7:29. doi: 10.3389/fmicb.2016.00029

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

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

Singer, S. W., Hirst, M. B., and Ludden, P. W. (2006). CO-dependent H_2 evolution by *Rhodospirillum rubrum*: role of CODH:CooF complex. *Biochim. Biophys. Acta* 1757, 1582–1591. doi: 10.1016/j.bbabio.2006.10.003

Soboh, B., Linder, D., and Hedderich, R. (2002). Purification and catalytic properties of a CO-oxidizing:H_2-evolving enzyme complex from *Carboxydothermus hydrogenoformans*. *Eur. J. Biochem.* 269, 5712–5721. doi: 10.1046/j.1432-1033.2002.03282.x

Sokolova, T. G., Henstra, A. M., Sipma, J., Parshina, S. N., Stams, A. J. M., and Lebedinsky, A. V. (2009). Diversity and ecophysiological features of thermophilic carboxydotrophic anaerobes. *FEMS Microbiol. Ecol.* 68, 131–141. doi: 10.1111/j.1574-6941.2009.00663.x

Techtmann, S. M., Lebedinsky, A. V., Colman, A. S., Sokolova, T. G., Woyke, T., Goodwin, L., et al. (2012). Evidence for horizontal gene transfer of anaerobic carbon monoxide dehydrogenases. *Front. Microbiol.* 3:132. doi: 10.3389/fmicb.2012.00132

Timm, J., Brochier-Armanet, C., Perard, J., Zambelli, B., Ollagnier-de-Choudens, S., Ciurli, S., et al. (2017). The CO dehydrogenase accessory protein CooT is a novel nickel-binding protein. *Metallomics* 9, 575–583. doi: 10.1039/C7MT00063D

Wang, J., Vine, C. E., Balasiny, B. K., Rizk, J., Bradley, C. L., Tinajero-Trejo, M., et al. (2016). The roles of the hybrid cluster protein, Hcp and its reductase, Hcr, in high affinity nitric oxide reduction that protects anaerobic cultures of *Escherichia coli* against nitrosative stress. *Mol. Microbiol.* 100, 877–892. doi: 10.1111/mmi.13356

Watanabe, S., Kawashima, T., Nishitani, Y., Kanai, T., Wada, T., Inaba, K., et al. (2015). Structural basis of a Ni acquisition cycle for [NiFe] hydrogenase by Ni-metallochaperone HypA and its enhancer. *Proc. Natl. Acad. Sci. U.S.A.* 112, 7701–7706. doi: 10.1073/pnas.1503102112

Watt, R. K., and Ludden, P. W. (1998). The identification, purification, and characterization of CooJ: a nickel-binding protein that is co-regulated with the Ni-containing CO dehydrogenase from *Rhodospirillum rubrum*. *J. Biol. Chem.* 273, 10019–10025. doi: 10.1074/jbc.273.16.10019

Weiss, M. C., Sousa, F. L., Mrnjavac, N., Neukirchen, S., Roettger, M., Nelson-Sathi, S., et al. (2016). The physiology and habitat of the last universal common ancestor. *Nat. Microbiol.* 1:16116. doi: 10.1038/nmicrobiol.2016.116

Whitham, J. M., Tirado-Acevedo, O., Chinn, M. S., Pawlak, J. J., and Grunden, A. M. (2015). Metabolic response of *Clostridium ljungdahlii* to oxygen exposure. *Appl. Environ. Microbiol.* 81, 8379–8391. doi: 10.1128/AEM.02491-15

Wolfe, M. T., Heo, J., Garavelli, J. S., and Ludden, P. W. (2002). Hydroxylamine reductase activity of the hybrid cluster protein from *Escherichia coli*. *J. Bacteriol.* 184, 5898–5902. doi: 10.1128/JB.184.21.5898-5902.2002

Wu, M., Ren, Q., Durkin, A. S., Daugherty, S. C., Brinkac, L. M., Dodson, R. J., et al. (2005). Life in hot carbon monoxide: the complete genome sequence of *Carboxydothermus hydrogenoformans* Z-2901. *PLoS Genet.* 1:e65. doi: 10.1371/journal.pgen.0010065

Yurkiw, M. A., Voordouw, J., and Voordouw, G. (2012). Contribution of rubredoxin:oxygen oxidoreductases and hybrid cluster proteins of *Desulfovibrio vulgaris* Hildenborough to survival under oxygen and nitrite stress. *Environ. Microbiol.* 14, 2711–2725. doi: 10.1111/j.1462-2920.2012.02859.x

Conflict of Interest Statement: 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 © 2019 Inoue, Nakamoto, Omae, Oguro, Ogata, Yoshida and Sako. 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.