Title

Phylogenetic diversity of the carbon monoxide-utilizing prokaryotes and their divergent carbon monoxide metabolisms in the human gut microbiome

Yuka Adachi Katayama[a], Ryoma Kamikawa[a], Takashi Yoshida[a,#]

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

#Address correspondence to Takashi Yoshida: E-mail, yoshida.takashi.7a@kyoto-u.ac.jp; Tel., +81-75-753-6217; Fax, +81-75-753-6226

Keywords

Human gut microbiome; Carbon monoxide-utilizing prokaryotes; Wood-Ljungdahl pathway; Nickel-containing carbon monoxide dehydrogenase; *Blautia*; *Veillonella*

Abstract

Although the production of toxic CO within the human body has been detected, only a few CO-utilizing prokaryotes (CO utilizers) have been reported in the human gut, and their phylogenetic and physiological diversity remains unclear. Here, we unveiled more than thousand representative genomes originating from previously unexplored potential CO-utilizing prokaryotes, which contain CO dehydrogenase (CODH) genes. More than half of CODH-bearing prokaryotes possess genes for the autotrophic Wood–Ljungdahl pathway (WLP). However, 79% of these prokaryotes commonly lack a key gene for WLP, which encodes enzyme that synthesizes formate from CO_2 and reductants such as H_2, suggesting that they share a degenerated WLP. Instead, many were predicted to possess an alternative way of synthesizing formate from pyruvate, which is a product of glycolysis. In addition to degenerated WLP, seven genes neighboring the CODH gene were found, which may reflect diverse utilization of CO in the human gut. Our findings reveal the unique and diverse nature of CO metabolism in the human gut microbiome, suggesting its potential contribution to CO consumption and gut homeostasis.

Impact statement

Carbon monoxide (CO)-utilizing prokaryotes mitigate the toxic impact of CO by

consuming it as energy and/or carbon sources. In addition to various environments, CO is also produced via multiple routes, such as heme degradation, in the human body and accumulates in the gut. Revealing CO-utilizing prokaryotes and their CO metabolisms in the human gut would contribute to gaining insight into how microbial community functions are involved in maintaining human gut homeostasis. Nevertheless, the limited number of CO utilizers in the human gut microbiome have been reported. In our study, a significant proportion of human gut microbial genomes belonging to diverse phyla were revealed to be of potential CO-utilizing prokaryotes. Additionally, the majority of CO-utilizing prokaryote genomes in the human gut have potentially remodeled the Wood–Ljungdahl pathway (WLP), one of the most well-known autotrophic pathways, to the degenerated, heterotrophic form. Moreover, there were seven other genes neighboring CODH in the human gut CO-utilizers, suggesting various CO utilization. Our findings would pave the way for future explorations into microbial metabolic adaptations and their implications for human health.

Data summary

The human gut prokaryote genomes were downloaded from HumGut database (Hiseni et al. 2021; https://arken.nmbu.no/~larssn/humgut/). The accession numbers of

CODH/ACS-bearing genomes from environments without host-association (Inoue *et al*., 2022) are listed in Table S1. Metatranscriptomic datasets were downloaded from the NCBI Sequence Read Archive (SRA) under the Bioproject accession numbers PRJNA354235 and PRJNA707065 and their accession IDs are listed in Table S2.

Introduction

Carbon monoxide (CO) is a colorless, odorless, and tasteless gas that is toxic to many organisms (Ernst and Zibrak 1998). CO has a 200–250 times higher affinity for hemoproteins, such as hemoglobin and cytochrome *c*, than oxygen. Therefore, exposure to >100 ppm of CO inhibits oxygen transport and respiration, resulting in severe effects on human health (Ernst and Zibrak 1998; Alonso *et al.* 2003; Prockop and Chichkova 2007). Nevertheless, the human body produces 16.4–160 μmol, which correspond to 0.5–4.5 ppm, of CO per hour on average, primarily through heme degradation using heme oxygenase-1 (Morse and Choi 2005; Onyiah *et al.* 2014; Hopper *et al.* 2020). From the human body and exogenous sources, CO accumulates in the gastrointestinal tract, potentially influencing the composition of intestinal microbiota (Hopper et al. 2020).

Prokaryotes utilizing CO are called CO utilizers and have been found in various environments, including soil (Suzuki *et al.* 1983), lake (Leigh *et al.* 1981), freshwater lake sediments (Imaura *et al.* 2023), bioreactors (Hattori *et al.* 2000), hot springs (Svetlichny *et al.* 1991), deep hydrothermal vents (Moussard *et al.* 2004), and even gut environment (Trischler *et al.* 2022). All CO utilizers possess CO dehydrogenase (CODH), which catalyzes the reversible conversion of CO_2 and CO ($CO_2 + 2H^+ + 2e^-$

<=> CO + H_2O) (Oelgeschläger and Rother 2008; Sokolova *et al.* 2009). CODHs are physiologically and structurally diverse enzymes that form two major groups, namely nickel- and molybdenum-containing CODHs (Ni-CODH and Mo-CODH); of which, Ni-CODH works under anaerobic conditions (Ragsdale 2004). The catalytic subunit of Mo-CODH (CoxL) is phylogenetically classified into two groups, form I and II, whereas that of Ni-CODH (CooS and CdhA) is phylogenetically and structurally classified into eight distinct clades, A to H, and mini-CODH (Inoue *et al.* 2019, 2022).

Through reactions catalyzed by the abovementioned CODHs, CO utilizers utilize CO as a carbon source and/or as the low redox potential ($E^{0\prime}$= −520 mV) for carbon fixation and/or energy conservation in various CO metabolizing pathways and respiratory chains. For instance, through the reduction and oxidation of the ferredoxin-like protein CooF and flavin adenine dinucleotide-dependent NAD(P) oxidoreductase (FNOR), the reducing power of CO is supplied to NAD^+, thereby generating NADH, which is available for various respiration and biosynthesis reactions (Tian *et al.* 2016; Inoue *et al.* 2019; Slobodkin *et al.* 2019). The reducing power of CO is also supplied to various terminal electron acceptors, including proton, nitrate, and oxygen, in the electron transport chains to generate an electrochemical gradient for ATP synthesis (Oelgeschläger and Rother 2008; Slobodkin *et al.* 2019; Adachi *et al.* 2020).

Among the CO-mediated CO_2 fixation pathways, the Wood-Lijungdahl pathway (WLP; reductive acetyl-CoA pathway) has been best studied. The WLP is recognized as the most ancient autotrophic pathway, which was possessed by the last universal common ancestor (LUCA) (Schuchmann and Müller 2014; Weiss *et al.* 2016; Adam *et al.* 2018). This pathway consists of two branches, carbonyl and methyl branches (Ragsdale and Pierce 2008; Adam *et al.* 2018). In the carbonyl branch, CO_2 is reduced to CO by Ni-CODH and incorporated into the carbonyl moiety of acetyl-CoA. Alternatively, CO can be obtained directly from the surrounding environment and incorporated into the carbonyl branch (Ragsdale 2008). In the methyl branch, formate is synthesized from CO_2 and reductants such as H_2 by formate dehydrogenase (Fdh), which are then reduced stepwise to a methyl moiety (CH_3-) by formyl-tetrahydrofolate synthase (Fhs), methylene-tetrahydrofolate dehydrogenase/cyclohydrolase (FolD), and methylene-tetrahydrofolate reductase (MetF). The generated carbonyl and methyl moieties are incorporated into acetyl-CoA along with coenzyme A (CoA) by acetyl-CoA synthase (ACS), which forms a complex with Ni-CODH (CODH/ACS). Acetyl-CoA is then converted into various compounds such as short-chain fatty acids (SCFA), including acetate, by CO utilizers (Miller and Wolin 1996; Trischler *et al.* 2022). In many of these acetogenic CO utilizers, WLP is linked to energy conservation using Rnf

complexes and/or energy-converting hydrogenase (Ech), where the oxidation of reduced ferredoxin is coupled with the reduction of NAD^+ and protons, respectively (Diender *et al.* 2015; Schoelmerich and Müller 2019).

Although the CO-utilizing abilities of human gut prokaryotes have not yet been fully elucidated, accumulating evidence suggests that CO utilizers are present in the human gut microbiome. Recently, two human gut bacteria, *Blautia luti* and *B. wexlerae*, were found to consume CO via the WLP when formate was added to the medium (Trischler *et al.* 2022). Furthermore, several gastrointestinal prokaryotes such as *Clostridioides* and *Marvinbryantia* possess genes for WLP (Wolin *et al.* 2003; Yao *et al.* 2023). However, the diversity of CO utilizers and their metabolic pathways in the human gut remain poorly understood.

With the advancement of sequencing technologies and bioinformatic tools, extensive databases of human gut prokaryotic genomes have been established. Almeida *et al.* (2021) created a Unified Human Gastrointestinal Genome (UHGG) collection comprising 204,938 non-redundant genomes. Hiseni *et al.* (2021) screened more than 5,700 healthy human gut metagenomes and constructed HumGut, a genome collection containing over 381,000 genomes. Analyzing these databases would deepen our understanding of human intestinal prokaryotes, including CO utilizers and uncultured

prokaryotes, which are dominant in the human gut (Almeida *et al.* 2019, 2021; Omae *et al.* 2019). In this study, we used the HumGut database to investigate CO utilizers and their CO-utilizing pathways in the human gut microbiome.

Methods

Genome sources

Publicly available genome collection, HumGut (Hiseni *et al*. 2021), was used as a source of human gut prokaryote genomes (downloaded from: https://arken.nmbu.no/~larssn/humgut/). In the database, >381,000 genomes detected from 3,534 healthy human gut metagenome samples were clustered into 30,691 representative genomes at 97.5% sequence identity. The previously reported WLP-bearing bacterial genomes without host asscociation (Inoue *et al*. 2022) were utilized and listed in Table S1. Metatranscriptomic datasets derived from 110 different healthy individuals were downloaded from the NCBI Sequence Read Archive (SRA) under the Bioproject accession numbers PRJNA354235 and PRJNA707065 (Table S2).

Detection of *cooS* and *coxL* in human gut prokaryote genomes

CODH-bearing genomes were searched from 30,691 representative human gut

prokaryotic genomes, using *cooS* and *coxL* as genetic markers for Ni- and Mo-CODHs, respectively. The homology search was performed using NCBI blast+ 2.13.0 with cutoffs of an E-value of 10^{-10}, a sequence length of 200 aa, and a sequence identity of 30%, using the following amino acid sequences as queries: clades A–H CODHs and mini CODH from *Methanosarcina barkeri* CODH/ACS α subunit (WP_011305243.1), *Butyrivibrio* sp. CooS (WP_026514536.1), *Clostridium novyi* CooS (WP_039226206.1), *Carboxydothermus hydrogenoformans* CooSV (WP_011342982.1), *Thermococcus onnurineus* CooS (WP_012571978.1), *C. hydrogenoformans* CooSII (WP_011343033.1), *Deltaproteobacteria bacterium* RBG_16_49_23 CooS (OGP75751.1), Candidatus Atribacteria bacterium HGW-Atribacteria-1 CooS (PKP59679.1), and *Thermosinus carboxydivorans* mini CooS (WP_007288589.1), respectively (Inoue *et al*., 2019, 2022; Techtmann *et al*., 2012). The Ni-CODH clades of the detected CooS were determined by their amino acid sequence identity with representative CooS sequences. The detected CooS/CoxL sequences were aligned using MAFFT v7.487 with the E-INS-i application (Katoh and Standley 2013). To examine the motifs of CooS, previously used criteria (Inoue *et al.* 2019) were adopted: no amino acid substitutions in two C-clusters comprising Ni, Fe, and S; a B-cluster comprising cubane-type 4Fe-4S; and a D-cluster comprising an additional 4Fe-4S at the subunit

interface (Dobbek et al., 2001; Doukov *et al*. 2002; Inoue *et al*. 2019). A similar analysis was performed for CoxL with slight modifications. A homology search was performed using *Oligotropha carboxidovorans* CoxL (WP_013913730.1) as the query. The sequences that conserved the Form I CoxL active site (AYRCSFR) (King 2003) alone were selected. The detected Ni-CODHs are listed in Table S3. All genomes containing Ni-CODHs are listed in Table S4.

Phylogenetic tree construction

The taxonomic assignment of 30,691 prokaryotic genomes was performed using GTDB-tk v2.1.1, with the reference data version R207 (Chaumeil *et al.* 2019). The names of the phyla were corrected according to the method described by Oren and Garrity (2021). Genome distances among the *cooS*-containing genomes were calculated using Mashtree v1.2.0 with 1,000 replicates and a minimum depth value of 0 (Ondov *et al.* 2016; Katz *et al*. 2019). The obtained tree was then visualized using the R package, ggtree v3.2.1 (Yu *et al.* 2017). Of the 3,534 human gut metagenome samples (Hiseni *et al.* 2021), the number of metagenome samples, from which the bacterial genomes were detected with 95% or higher identity, was displayed as a heatmap on the outer layer of the phylogenetic tree.

Genome-based prediction of metabolic functions

To characterize the metabolic functions of the CO utilizer in the human gut, three groups of genomes were analyzed using METABOLIC v4.0 (Zhou *et al.* 2022). The first group comprised the CODH-bearing genomes detected in the HumGut database. The second group comprised the non-CODH-bearing genomes of human gut prokaryotes belonging to a family or order with CODH-bearing genomes, that were detected in the HumGut database. The third group comprised the CODH/ACS-bearing bacteria found in environments other than the intestine. The proportions of genomes encoding functions in each genus were calculated for each group. From the HumGut database, we selected Ni-CODH-lacking bacterial genomes belonging to orders or families containing Ni-CODH-bearing bacteria, resulting in 194 genomes (the second group). We also utilized Ni-CODH-bearing bacterial MAG datasets constructed by Inoue et al. (2022). Among the genomes, only the 554 genomes with bacterial CODH/ACS and without “host-associated” tag were used in this analysis. The accession ID of the genomes are listed in Table S1.

Characterization of gene compositions in the *cooS* genomic contexts

Protein-coding genes in the genomes were predicted using Prodigal v2.6.3, with default settings (Hyatt *et al.* 2012). The gene functions were then annotated to the protein coding sequences using eggNOG-mapper v2.1.5 (Cantalapiedra *et al.* 2021). For double-checking, a homology search of the protein-coding sequences against the clusters of orthologous gene (COG) database was performed using DIAMOND BLASTp v 2.0.12 (Buchfink *et al.* 2021). The KEGG orthology (KO) and COGs of 15 genes located in the upstream and downstream of the CooS loci were then extracted and are listed in Tables S5 and S6, respectively. Previously reported CODH-related COGs were also referred: AcsB (COG1614), CooF (COG0437 or COG1142), FNOR (COG1251), ECH (COG3260 and COG3261), and ABC transporter (COG0600, COG1116, and COG0715) (Inoue *et al.* 2019). The gene maps around *cooS* were visualized using the R package gggenomes v0.9.5.9000 (Hackl and Ankenbrand 2022).

Genomic-based predictions of the degenerated WLP

In the present study, the gene composition of the CODH/ACS-bearing genomes was analyzed. Genomes that possessed both CooS and AcsB (K14138) were termed CODH/ACS-bearing genomes. All the KOs annotated to bacterial genomes in the human gut and other environments were listed, and the presence of each KO was

222 checked in each genome to calculate the proportion of genomes carrying a given gene in
223 each genome group. When two or more KOs were annotated to a single gene, all
224 annotated KOs were considered for calculating the proportion. Since K00656 (PflD)
225 contains members of PFL-like proteins without PFL activity, proteins annotated as
226 K00656 were aligned using MAFFT v7.487 with the E-INS-i application, and only the
227 sequences that conserved the Cys-Cys active sites (Sawers and Watson 1998) were
228 regarded as potential PFL-coding sequences. The proportions of the genomes carrying
229 the given genes are listed in Table S7. Especially, the proportion of genomes carrying
230 the following genes were extracted from the obtained results: CO metabolism-related
231 genes (*cooC*, K07321; *cooF*, K00196), ACS subunits (*acsB–E*, K14138, K00197,
232 K00194, K15023; *cdhC*, K00193), WLP (*fhs*, K01938; *folD*, K01491; *metF*, K00297;
233 *metV*, K25007), Rnf subunits (*rnfA–G*, KO03612–17; *rnfC2*, K25008), *ech* (K15830,
234 K15832), Hdr (*hdrA–C*, K03388–90), Mvh (*mvhD*, K14127), and PFOR (*por*, K03737;
235 *porA–D*, K00169–72). In addition, formate-related genes were searched in the KO
236 database and KEGG reaction database by inserting the keyword, "formate."
237 Hydrogenases were omitted from the results and other genes were checked as the
238 formate-related proteins: formate transporters (FocB, K03459; FocA, K06212; OxlT,
239 K08177; YfdC, K21990; FdhC, and K21993), formate dehydrogenase catalytic subunits

(FdhA, K05299; FdhF, K22015; and FdoG, K00123) and the other subunits of formate dehydrogenase (K00122–27, K22515, and K22516), PFL (K04069 and K00656), and other 55 genes. Genes that were not present in any genome were removed from the figure.

Taxonomic and functional profiling of metatranscriptomes

Publicly available metatranscriptome datasets of healthy human were retrieved from the SRA database. Among the 110 datasets, 96 datasets were sourced from BioProject PRJNA354235, which includes samples from healthy adult men from United States aged 40–75 years old (Abu-Ali *et al*. 2018), while 14 datasets were sourced from BioProject PRJNA707065, which includes samples from mothers and their six-month-old infants from New Zealand or United Kingdom under the NiPPeR Study (https://www.nipperstudy.com/). Sequence reads were processed using KneadData v0.12.0 (http://huttenhower.sph.harvard.edu/kneaddata). In KneadData, the quality control pipeline was set as follows: FastQC v0.12.1 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was employed before and after quality controls, adaptor trimming was performed with default settings, reads were

filtered by Trimmomatic v0.35 (Bolger, Lohse and Usadel 2014) with a sliding window sizes of four and a minimum quality score of 20, Tandem Repeats Finder (TRF) v4.09 (Benson 1999) was employed with default parameters, and human host RNA (Hg38) and rRNA sequences (Silva v128) were removed using Bowtie2 v2.5.2 (Langmead and Salzberg 2012). Trimmed non-human reads were subsequently mapped to (i) the 1,380 of human gut non-redundant CODH mRNA sequences identified in this study, (ii) the 30,691 human gut prokaryotic genomes in the HumGut database, and (iii) Fdh genes of *B. schinkii* (JANSWJ010000003.1:269667-271859) and *B. hydrogenotrophica* (NZ_CYXL01000001.1:c262854-260023) using CoverM v0.6.1 (https://github.com/wwood/CoverM).

Results

CODH was possessed by 4.2% of human gut prokaryotic genomes

In the present study, we analyzed human gut prokaryotic genomes from the HumGut database (Hiseni *et al.*, 2021). The genomes in the HumGut database were clustered into 30,691 genomes with 97.5% sequence identity (Hiseni *et al.*, 2021). Prior to exploring CO utilizers in the human gut, 30,691 representative genomes were taxonomically classified using GTDB-tk (Chaumeil *et al*. 2019), and 98% of the gut genomes belonged to the phyla Bacillota (19,877 genomes), Bacteroidota (4,011), Actinobacteriota (3,676), Pseudomonadota (1,905), Campylobacterota (227), and Thermodesulfobacteriota (259) (Fig. 1a). The dominant orders in the Bacillota phylum were Oscillospirales (8,642), Lachnospirales (3,608), Christensenellales (1,476), Lactobacillales (1,423), and Veillonellales (1,279) (Fig. 1a).

To identify CO utilizers in the human gut, genes for the catalytic subunit of Ni-CODH clades A–G (*cdhA* and *cooS*) and mini CODH were used as markers for CO utilizers. As a result of homology search, 2,505 amino acid sequences encoded by 2,202 genomes were detected as candidate Ni-CODHs. Of these, Ni-CODHs were further filtered using the following criterion: no amino acid substitutions or deletions in the conserved motifs of the CODH of the five metal clusters and two acid-base catalysts

(Inoue *et al*. 2019). Consequently, 1,380 amino acid sequences in 1,302 genomes were confirmed to conserve Ni-CODH motifs. Most of the genomes contained one Ni-CODH gene (1,231/1,302 CODH-bearing genomes), whereas other genomes contained two or more genes (Fig. 1b). The detected 1,380 Ni-CODHs belonged to clades B (493), C (179), D (14), E (692) and F (1) and mini CODH (1) (Fig. 1c). Although a similar survey was conducted for Mo-CODH using CoxL as a marker, Mo-CODH, which possesses a conserved active site for form I CoxL, was not found in any human gut genome.

Among the 1,302 CODH-bearing genomes, the phylum Bacillota comprised 1, 223 (94%). The remaining 79 genomes were identified as Actinobacteriota, Bacteroidota, Campylobacterota, Thermodesulfobacteriota, Pseudomonadota, Synergistota, and Verrucomicrobiota (Fig. 1d, e). The 1,223 Bacillota- and 79 other phyla-derived genomes belonged to 248 species, 82 genera, and 8 phyla.

Among the 1,223 Bacillota-derived genomes, 379, 341, and 411 belonged to Lachnospirales, Veillonel l ales, and Oscillospirales, respectively; thus, a large proportion of the CODH-bearing genomes were occupied by these three orders (Fig. 1d). CODH-bearing genomes accounted for 10.5%, 26.7%, and 4.8% of the total human gut microbial genomes derived from the orders Lachnospirales, Veillonelalles, and

Oscillospirales, respectively. Within the order Lachnospirales, 38 genera possessed CODH, including *Blautia* (154/291; 154 Ni-CODH-bearing genomes within 291 of the total genomes of the genus in the human gut microbiome), *Fusicatenibacter* (67/215), *Ruminococcus* (29/1,194), *Enterocloster* (21/56), *Anaerobutyricum* (12/30), and UMGS1375 (10/11). Of these, *Blautia* was previously identified as possessing functional WLP (Trischler *et al*. 2022) and was the only taxon to possess three CODHs (HumGut ID: 13241, 3275, 5110, 12728, 20698, 11367, and 8982) (Fig. 1b). In *Blautia*, 24 of 154 CODH-bearing genomes contained two or more CODH genes. The Veillonelalles with Ni-CODH genes consisted of three genera: *Veillonella* (323/443), *Megasphaera* (12/96), and F0422 (7/12). *Veillonella* is a ubiquitous bacterium found in the human body, such as in the oral cavity and gut (Rogosa 1964). The order Oscillospirales contains 13 genera that encode Ni-CODHs in their genomes, including *Faecousia* (282/446), CAG-170 (82/179), and CAG-353 (20/60) (Table S4). Although Oscillospirales occupies a higher proportion of the human gut CODH-bearing genome, most of these bacteria are uncultured; therefore, their physiological characteristics are largely unknown. CODH-bearing genomes were also identified in the Bacillota orders Peptostreptococcales (41/239) and Selenomonadales (27/135) as well as in the Actinobacteriota order Coriobacteriales (23/3,050) (Table S4).

To assess the prevalence of potential CO utilizers in humans, we checked for the presence of CODH-bearing genomes among the available 3,534 datasets of healthy human gut metagenomes (Hiseni *et al*., 2021) (Fig. 1e). The CODH-bearing genomes derived from the Lachnospirales genera *Fusicatenibacter* and *Blautia* were present in 48% and 26% of the metagenome datasets, respectively. The Veillonelalles genera *Veillonella* was detected in 8.3% of the metagenome datasets (Fig. 1e, Table S4). The Oscillospirales genera *Faecousia* and CAG-170 were detected in 13% and 18% of the metagenome datasets, respectively. The above CODH-bearing genomes were appeared more prevalent than or as prevalent as the two prominent CODH-lacking Bacteroidales genera, *Phocaeicola* and *Bacteroides* (Hiseni *et al.* 2021), which were detected in the 18% and 31% metagenome datasets on average, respectively. In contrast, some CODH-bearing genomes were present in only a limited number of the human gut metagenome datasets (Fig. 1e, Table S4). For instance, the CODH-bearing genomes from the genus *Eubacterium* in the order Eubacteriales were found in only 0.2% of metagenomes on average.

Majority of CODH-bearing bacteria were acetate-producing prokaryotes in the human gut

As most of the potential CO utilizers detected in this study have been uncultured and thereby uncharacterized, we reconstructed the genome-based metabolic functions of the potential CO utilizers bearing Ni-CODH genes in the human gut microbiome using METABOLIC v4.0 (Zhou *et al.* 2022). Twenty-three metabolic functions were detected in the potential human gut CO utilizers, and the acetate production function was conserved among them (Fig. S1). In particular, the analysis revealed that 1,162/1,302 Ni-CODH-bearing human gut genomes, which correspond to 73/82 genera, contained genes for acetate production (*acdA*, *ack*, and *pta*) for fermentation, suggesting that most potential CO utilizers produce acetate in the human gut (Fig. S1). Exceptions were the Selenomonadales genera (*Centipeda*, *Mitsuokella*, and *Selenomonas*), whose all 27 genomes lacked the abovementioned genes for acetate production.

To evaluate whether these functions were specific to potential CO utilizers, we compared the metabolic functions of 1,302 Ni-CODH-bearing human gut microbial genomes to those of 194 Ni-CODH-lacking genomes that were closely related to the Ni-CODH-bearing ones. All functions other than CO metabolism were observed in both types of genomes; thus, we could not find any specific metabolic functions in the potential CO utilizers (Fig. S1). This suggests that CO metabolism may be engaged in accessory functions that support the existing functions conserved in both Ni-CODH-

bearing and Ni-CODH-lacking human gut microbiomes.

Eight *cooS* genomic contexts were identified in the human gut prokaryote genomes

Genes that encode proteins involved in CO metabolism are often located close to *cooS* in the genome, enabling the prediction of their physiological roles based on the genomic context (Inoue *et al*. 2019; Matson *et al.* 2011; Techtmann *et al*. 2012). To gain insight into the physiological roles of CO metabolism in the human gut microbiome, we analyzed 15 genes located in the upstream and downstream regions of *cooS* in the 1,302 human gut microbial genomes (Fig. 2, Table S5, S6). Totally, 1,150 KOs and 1,285 COGs were annotated for genes within the 1,380 *cooS* contexts in the 1,302 genomes (Tables S5 and S6). Genomic contexts were manually classified into eight types: WLP, PEPCK, FNOR, ABC transporter, Fe-only hydrogenase, MFS transporter, uncharacterized dehydrogenase, and Cysteine synthase. Below, we describe the estimated physiological functions and detailed phylogenetic distributions of genomic context types (Fig. 2).

The WLP type, characterized by the presence of the gene for ACS β subunit (AcsB; K14138, COG1614), was the most prevalent in the human gut microbial genomes, being detected from 593 of the 1,380 *cooS* contexts. The WLP context was

observed in Oscillospirales (8/13 genera, that is, eight CODH-bearing genera within the total genera of the order; 323/411 CODH-bearing genomes, that is, 323 CODH genomes within the total genomes of the order), Lachnospirales (18/38 genera, 198/379 genomes), Peptostreptococcales (4/9 genera, 13/41 genomes), and Eubacterales *Eubacterium* (3/3 genomes). All genera conserved the consecutive gene arrangement of *cooS-cooC-acsB*, except for the Peptostreptococcales genera, which had the gene arrangement of *cooS-cooC-fhs* and *acsB* at 10 genes downstream of *cooS* (Tables S5 and S6). *acsCDE* was found in the *cooS* contexts of the Lachnospirales and Peptostreptococcus genera but not in seven out of eight Oscillospirales genera (Fig. 2, Table S5, S6).

The second type is PEPCK, which has been unpreceded in *cooS* genomic contexts. PEPCK (*pckA*; K01610, COG1866) is a phosphoenolpyruvate carboxykinase that catalyzes the conversion of phosphoenolpyruvate (PEP), CO_2, and ADP to oxaloacetate (OAA) and ATP, and vice versa. The PEPCK type was found among the genomes of Veillonelales genus *Veillonella* (189/342 genomes). Gene composition varied among PEPCK-type genomic contexts. Of the 189 PEPCK-type genomic contexts of *Veillonella*, 144 included a putative *oxyR*, which encodes a transcriptional regulator of LysR family (K04761, COG0583) and responds to oxidative stress (Sen and Imlay 2021) in the upstream region adjacent to *cooS*. Genes for nitrate reduction (*narGHIJK*)

were also observed in 111 *Veillonella* PEPCK-type genomic contexts, of which 78 genomes also included *oxyR* (Fig. 2).

The third is FNOR type that has been predicted to be associated with CO utilization in the previous studies (Geelhoed *et al*. 2016; Inoue *et al*. 2019, 2022). FNOR (COG1251) is often present in the CO utilizers and is involved in energy conservation through CO oxidation associated with the reduction of $NAD(P)^+$ (Geelhoed *et al*. 2016). In human gut microbial genomes, FNOR genes were found in 138 *cooS* contexts of bacteria, including Lachnospirales (15/38 genera, 92/379 genomes) and Peptostreptococcales (5/9 genera, 18/41 genomes) (Fig. 2).

The following types have been reported previously; however, their physiological functions in CO metabolism are unknown: ABC transporters (K02049–51, COG0715-COG0600-COG1116) were found in various bacteria, such as Lachnospirales (9/38 genera, 40/379 genomes), Campylobacterales *Campylobacter* (37/40 genomes), and Selenomonadales *Selenomonas* (4/4 genomes) (Fig. 2). “Fe-only hydrogenase” (COG4624) was encoded in Lachnospirales (9/38 genera, 43/379 genomes), Oscillospirales (2/13 genera, 6/411 genomes), and Peptostreptococcus (2/9 genera, 4/56 genomes) (Fig. 2, Table S6). The COG4624 gene was located downstream region adjacent to *cooS* in 24 of the 57 *cooS* genomic contexts with COG4624, including

413 contexts from Lachnospirales *Anaerobutyricum*. In 17 of these, *pflAB* (K04069 and
414 K00656) was found within the genomic context of *cooS* (Fig. 2). In addition to the Fe-
415 hydrogenase-type context, COG4624 was also found in the WLP-type context in 27
416 genomes.

417 "MFS transporter," "functionally uncharacterized dehydrogenase," and "Cysteine
418 synthase" were relatively minor types, as they occupied fewer than 2% of the CODH-
419 bearing genomes and were observed in specific taxa. The MFS transporter is
420 responsible for importing or exporting a wide range of substrates across the membrane
421 using a substrate concentration gradient. The MFS transporters encoded in this context
422 were annotated as COG2223 (NarK, nitrate transporter) and K08177 (OxlT,
423 oxalate/formate antiporter) in Selenomonadales *Mitsuokella* (16/20 genomes) (Fig. 2).
424 In the functionally uncharacterized dehydrogenase context type, genes for K00004
425 (butanediol dehydrogenase) and COG1063 (threonine dehydrogenase or related Zn-
426 dependent dehydrogenase) were found in the downstream region adjacent to *cooS* in
427 Clostridiales *Clostridium* (11/15 CODH-bearing genomes), whereas genes for
428 hydrogenase nickel incorporation protein (*hypAB*, K04651, and K04652), which may be
429 involved in maturation of CooS, were present upstream of *cooS*. Cysteine synthase
430 (CysK; K01738, COG0031) was mainly found in Veillonellales *Megasphaera* (11/12

genomes).

Fdh-lacking WLP is common in the human gut microbiome

As the WLP-type genomic context was the most prevalent among the 1,302 potential CO utilizers in the human gut, we further investigated whether human gut prokaryotes bearing the WLP context were capable of acetyl-CoA synthesis through the WLP. Since WLP could be functional even when *cooS* and *acsB* are not located in the same genomic context (Gencic and Grahame 2020), we here referred to the 667 genomes bearing both *cooS* and *acsB* as CODH/ACS-bearing genomes regardless of whether to retain them as the WLP type context. The presence of other WLP genes such as *cooCF*, *fdh*, *fhs*, *folD*, *metF*, and *acsCDE* was surveyed in all the genomes containing genes for CODH/ACS (667 genomes) (Fig. 3).

More than 88% of CODH/ACS-bearing genomes retained almost the full set of WLP genes, including *acsCD*, *cooC*, *fhs*, *folD*, and *metF* (Fig. 3, Fig. S2). Interestingly, *acsE*, a bacteria-specific gene for the methyltransferase subunits of ACS, was not detected in 78% of the Oscillospirales genomes (7/9 genera) (Fig. 3), whereas *acsE* was conserved in other CODH/ACS-bearing genomes. Although a nearly full set of WLP

genes was detected, we found that the Fdh genes for formate synthesis (*fdhA*, *fdhF*, and *fdoG*) were not detected in 95%, 96%, and 79% of the CODH/ACS-bearing genomes, respectively (Fig. 3), and any other types of the Fdh genes were observed in 79% of the genomes. A previous study reported that two *Blautia* species are capable of CO metabolism in the presence of external formate even though the bacteria lack Fdh gene/activity (Trischler *et al*. 2022). In this study, the presence of Fdh genes were also checked in the 554 CODH/ACS-bearing genomes found in other environments (Inoue *et al*., 2022) (Fig. S2). It revealed that Fdh-lacking WLP is also present in the genomes of other environments. However, compared to human gut microbial genomes with WLP genes, Fdh genes (*fdhA*, *fdhF*, and *fdoG*) were missing in a smaller proportion of bacteria of other environments (28%, 84%, and 17%, respectively) (Fig. S2), and none of the Fdh genes was observed in 14% of the genomes. Thus, the lack of formate synthesis might be a feature highly enriched in the human gut potential CO utilizer. Similarly, *cooF* was not detected in 91% of the CODH/ACS-bearing genomes (accounting for 26/42 genera), whereas *cooF* was absent in 53% of the bacteria isolated from other environments (Fig. 3, Fig. S2).

The above analyses indicate that there may be CO_2/H_2-independent sources of formate for human gut CODH/ACS-bearing bacteria. To estimate how formate is

467 obtained, we surveyed the genes involved in alternative formate synthesis or import by
468 CODH/ACS-bearing bacteria (Fig. 3). The Lachnospirales (16/21 genera, 145/239
469 genomes), a few Oscillospirales genera (2/9 genera), all Peptostreptococcus genera (5/5
470 genera, 22/26 genomes), and Verrucomicrobiales genera Akkermansia (1 genome)
471 contained *pflB* (Fig. 3). These genera are likely capable of producing formate from
472 pyruvate. Three Oscillospirales genera that lacked both *fdh* and *pfl* possessed a gene for
473 formate transporters, *fdhC* (56/291 genomes) (Fig. 3). Two Lachnospirales
474 (*Merdiplasma* and *Fusicatenibacter*) and two Oscillospirales genera (CAG-170 and
475 *Faecousia*) were not found to possess these genes. These genera possess genes encoding
476 proteins involved in formate-producing/utilizing reactions such as GTP degradation
477 (GCH1, *ribBA*) and thiamine metabolism (*thiC*) (Fig. 3).

478 To gain deeper insights into the functions specifically encoded in the human gut
479 microbial genomes with WLP, we compared all the genes in the genomes of human gut
480 bacteria and other environmental bacteria that contain WLP genes (Fig. S1, Table S7).
481 We observed that genes involved in carbohydrate and saccharide degradation were more
482 prevalent in the human gut microbiome containing WLP genes than in bacteria of other
483 environments (Fig. S1, Table S7). Therefore, one of the possible explanations is that
484 potential human gut CO utilizers with WLP are capable of utilizing more diverse

carbohydrates and/or polysaccharides than the CO utilizers of other environments, producing more pyruvate through glycolysis. For pyruvate metabolism, 53% of the human gut CODH/ACS-bearing genomes possessed *pflB*. PFOR (*por*) is also involved in WLP by supplying reduced ferredoxin from pyruvate oxidation (Ragsdale 2004) and is present in 90% of the human gut CODH/ACS-bearing genomes. These genes were less frequently observed in CODH/ACS-bearing bacteria in other environments: *por* and *pflB* were present in 71% and 19% of the bacteria in other environments, respectively (Fig. S2).

Not only genes for carbohydrate/saccharide degradation (above), but also the genes for Rnf complex (*rnfABCDEG*) were detected in a high proportion of WLP-bearing CO utilizers in the human gut (>73%). Group B FeFe hydrogenase, which is involved in ferredoxin-coupled H_2 production (Benoit *et al.* 2020), was relatively enriched in human gut WLP-bearing CO utilizers (64%) (Fig. S2). Since some FeFe hydrogenases are known as electron-bifurcating enzyme that provide Fd to Fd-dependent Fdh of CO-utilizing acetogen (Schuchmann and Müller 2012), the co-occurrence of FeFe hydrogenase and Fdh was checked. As a result, group A13, A4, C1, and C3 FeFe hydrogenases were co-occurring with Fdh genes (Fig. S3). Especially, the group A4 FeFe hydrogenase and *fdhA* were statistically co-occurring, with an adjusted

odds ratio of 2,376 and a 95% confidence interval ranging from 136 to 41,655 (Fig. S3). Furthermore, *fdoG* was co-occurring with group C3 FeFe hydrogenases, with an odds ratio of 59.4 and a 95% confidence interval ranging from 25.6 to 138 (Fig. S3). However, the most prevalent type of hydrogenase, FeFe groupB hydrogenase gene, showed no association with Fdh genes (Fig. S3).

CODH transcripts were detected from 97.3% of metatranscriptomic samples

To gain insights into whether active CO-utilizing bacteria are present in the human gut microbiome, metatrascriptome samples from 110 healthy human individuals were analyzed. Following quality control, 7.8 ± 1.2 million reads were obtained per sample on average (Fig. 4a). Filtered reads were then mapped to 1,380 CODHs to identify CODH transcripts in the human gut microbiome. It was found that CODHs were transcribed in 107 out of 110 metatranscriptome samples, with an average of 1,127 reads mapped to the CODH sequences (Fig. 4b). The number of detected CODHs correlated with the total number of reads (r^2 = 75.7), with nearly full-length CODH mRNAs (cover length >1500 nt) detected in samples with more than 28 million filtered reads (Fig. 4b). The host bacteria of the CODH transcripts were assigned to the 13

521 orders from 7 phyla (Fig. 4c). None of the reads from the CODH-bearing genomes of
522 the phyla Campylobacterota, Desulfobacterota, and Proteobacteria were detected in the
523 healthy human metatranscriptome. At the order level, there were Lachnospirales
524 (average 2.9 $\times 10^5$ RPKM of CODH transcripts across 110 datasets), Oscillospirales (1.7
525 $\times 10^5$ RPKM), Peptostreptococcales (2.6 $\times 10^4$ RPKM), Veillonellales (1.8 $\times 10^4$ RPKM),
526 Synergistales (1.8 $\times 10^3$ RPKM), and other 8 orders (55 to 815 RPKM). At the genera
527 level, CODHs from 66 genera were detected (Fig. S4). In Lachnospirales, the genera
528 *Blautia*, *Choladocola*, *Fusicatenibacter* and UMGS1375 accounted for 37.9%, 19%,
529 12%, and 7.9% of CODH transcripts from Lachnospirales, respectively. In
530 Oscillospirales, the genera *Faecousia* and CAG-170 accounted for 65% and 26% of the
531 transcripts, respectively. In Peptostreptococcales, 58% of CODH transcripts were from
532 the genus *Intestinibacter*. In Veillonellales, 99.7% of CODH transcripts were derived
533 from the genus *Veillonella*. Regarding the function of CODHs, 6 out of 8 types (WLP,
534 PEPCK, FNOR, ABC transporter, Fe-hydrogenase, MFS transporter) were identified in
535 the human gut metatranscriptomes (Fig. 4d). The most abundant type was WLP,
536 accounting for 66% relative abundance on average, while average abundances of other
537 types ranged from 0.02 to 5.0% (Fig. 4d).

538 To determine if the entire WLP is transcribed in the human gut microbiome,

transcripts of WLP-related genes were investigated. Among the most prevalent WLP-bearing species, genomes with completeness greater than 94% were used as the references for read mapping. As a result, transcripts of all WLP genes (*cooS*, *acsB*, *fhs*, *folD*, *metF*) from six species were detected in the human gut metatranscriptome, clearly indicating that WLP is actively transcribed in the human gut (Fig. 5). Transcripts associated with WLP (*rnf*, *pfor*, *pfl*) were also identified in the samples (Fig. 5). Metatranscriptome reads were mapped to *fdhF* from *B. schinkii* and *B. hydrogenotrophica*, resulting in the detection of *B. hydrogenotrophica fdhF* in 4 out of 110 samples, although the cover length was shorter than 500 nt in all four samples. Since PFOR can supply reduced ferredoxin from pyruvate and PFL can supply formate from pyruvate, our interest lay in the expression balance of PFOR and PFL to predict the metabolism of Fdh-lacking bacteria. Both transcripts for PFOR and PFL were identified (Fig. 5 a-d). PFOR transcripts were generally detected at a higher level than PFL, with an average 19.7-fold difference, while PFL transcripts were observed at higher levels than PFOR in three samples (Fig. 5 a-d).

Discussion

Given the continuous production of CO through heme degradation in the human body (Hopper et al. 2020), it was predicted that a significant number of CO-utilizing prokaryotes exist in the human gastrointestinal tract, contributing to the consumption of the accumulating harmful CO. Previous studies indeed demonstrated the rapid consumption of CO by fresh human fecal samples (Levine *et al.* 1982) and by two *Blautia* strains isolated from human feces in the presence of formate (Trischler *et al.* 2022). However, the mechanism through which taxonomically and phylogenetically diverse human gut prokaryotes become potential CO utilizers in the human gut microbiome remains unclear. In the present study, we analyzed CODH-bearing genomes derived from potential CO utilizers in the human gut microbiome and uncovered various Ni-CODH-bearing genomes belonging to 248 species and 82 genera across 8 phyla (Fig. 1). Furthermore, CODH transcripts derived from 66 genera, 5phyla were identified in 110 of metatranscriptome samples obtained from healthy human with various nationality and age (ranging from 6 months to 75 years old). This suggests that the CO utilizing pathways are active and may constitute an important component of the human intestinal environment. To date, only two human gut prokaryotes have been experimentally verified as CO utilizers (Trischler *et al.* 2022) and 280 *Blautia* genomes and 43 acetogenic species were checked to identify the presence of *cooS* (Trischler *et al.*

575 2022; Yao *et al.* 2023). Therefore, our findings extensively expand the catalogue of
576 potential human gut CO utilizers.

577 Our findings are not limited to phylogenetic diversity but include the functional
578 diversity of CO utilizers. The previously identified CO-mediated metabolic pathway in
579 human gut-derived prokaryotes is the WLP (Trischler et al. 2022). However, as their
580 CODH genomic contexts can be divided into eight distinct types presumably involved
581 in distinct physiological roles, the diverse taxa of the human gut microbiome are
582 potentially involved in CO utilization through various pathways.

583 We observed that WLP is most prevalent in the genomes and
584 metatranscriptome of human gut microbes, such as those of Oscillospirales,
585 Lachnospirales, and Peptostreptococcales. Almost all genes responsible for carbonyl
586 and methyl branches, including a gene for utilizing exogenous CO, were identified in
587 the genomes of the human gut microbiome. However, any types of Fdh genes were not
588 detected in large number of the WLP-bearing human gut microbial genomes (Fig. 3).
589 Previous cultivation experiments have suggested that human gut CO utilizers *Blautia*
590 *luti* and *B. wexlerae* in the order Lachnospirales lack Fdh activity but possess functional
591 WLP in formate-added medium (Trischler *et al.* 2022). These bacteria therefore utilize
592 extracellular formate. Furthermore, these bacteria possess PFL genes in their genomes

593 which synthesizes formate and acetyl-CoA from pyruvate (Fig. 6). It suggests that
594 glycolysis would become one of the primary sources of carbon and reducing power for
595 the Fdh-lacking WLP in *Blautia luti* and *B. wexlerae* (Fig. 6), meaning that their WLP
596 may be modified to a heterotrophic form that obtains carbon from organic compounds
597 instead of CO_2. The metabolic pathways of this "heterotrophic WLP" should be tested
598 experimentally. However, the coexistence of PFL and PFOR in *B. luti* and *B. wexlerae*
599 suggests that carbon and electron fluxes may be balanced depending on the available
600 carbohydrates and electron donors in the environment (Fig. 6). Additionally, group B
601 FeFe hydrogenase possessed by these *Blautia* strains may be utilized to dispose of the
602 excess amount of Fd_{red} produced via glycolysis using PFOR as evident from the
603 cultivate experiments of *B. coccoides* demonstrating hydrogen production in the
604 presence of glucose (Liu *et al*. 2015). We further predicted that this surplus Fd_{red} arises
605 due to the saccharide-rich environment of the human gut (Pokusaeva, Fitzgerald and
606 van Sinderen 2011), and part of Fd_{red} is uptaken by Fdh-lacking WLP to produce acetate
607 as well as the hydrogenase.

608 Although their CO-utilizing ability has not been confirmed, several human
609 gastrointestinal acetogens, including species from Lachnospirales (*B. producta* and
610 *Marvinbryantia formatexigens*) and Clostridiales (*C. bovifaecis* and *Clostridioides*

611 *difficile*), lack *fdh* but retain WLP genes (Wolin et al. 2003; Yao *et al*. 2020, 2023). This
612 study indicates that the Fdh-lacking WLP are not limited to *Blautia*, *Marvinbryantia*,
613 and *Clostridium* but more prevalently utilized in the human gut CO utilizers represented
614 by the 526 genomes from 32 genera. While a few routes were predicted for formate
615 acquisition, the "heterotrophic WLP" system may represent one favorable pathway, as it
616 is possessed by 53% of WLP-bearing human gut bacteria. Importantly, many of the
617 potential CO utilizers with such “heterotrophic WLP” also retain genes to degrade
618 carbohydrates and polysaccharides to greater extent than the CO utilizers of other
619 environments (Fig. S1, Table S7). The presence of these genes, coupled with the
620 availability of carbohydrates in the human intestinal tract (Pokusaeva, Fitzgerald and
621 van Sinderen 2011), suggests an enhanced provision of substrates for glycolysis, and
622 thereby pyruvate for formate synthesis. This alternative pathway for formate synthesis
623 may be one reason why *fdh* was positively dropped from the genomes of acetogens in
624 the intestinal environment (Yao *et al*., 2023). In addition, the previous findings by
625 Levine *et al.* (1982), which observed higher CO consumption activity of feces samples
626 in the presence of glucose (0.7 mL/h, g feces) than in the absence of glucose (0.2 mL/h,
627 g feces) (Levine *et al.* 1982), might at least partially be caused by the heterotrophic
628 WLP-bearing human gut CO utilizers. Regarding the other routes for formate

acquisition, 141 genomes possessed Fdh genes. These Fdh genes were found to co-occur with putative electron-bifurcating group A13, A4, C1, and C3 hydrogenases (Fig. S3), a condition necessary for utilizing CO to prevent toxicity to the hydrogenase module of H_2-dependent CO_2 reductase (HDCR) (Dietrich and Müller 2023). There were also bacteria lacking both Fdh and PFL. These bacteria may import formate from extracellular fractions by using the formate transporters or by cross-feeding (Gomez *et al*. 2019).

A novel CODH type, which has not been previously reported, was identified in 190 genomes of the genus *Veillonella* (Fig. 2). *Veillonella* mainly harbored clade B CODH, a phylogenetically separate variant from the predominant WLP-associated clade E CODH (Fig. 1). The *cooS* genome of *Veillonella* contained genes encoding PEPCK and NarKGHJI (Fig. 2), which may be involved in CO metabolism. For instance, a biochemical link between CO oxidation and nitrate reduction may be established as observed in *Deferribacter autotrophicus* (Slobodkin *et al.* 2019). However, as the reconstructed pathways described above are based only on metagenome-assembled genomes and metatranscriptomes, further experimental verifications are necessary.

This study revealed the phylogenetic diversity of CO utilizers and their divergent CO-utilizing pathways in the human gut microbiome. Because trace amounts of CO are

constantly produced in the human body, CO utilizers may constantly consume CO in the human gut microbiota, avoiding the accumulation of high concentrations of toxic CO in the gut. Therefore, shedding light on CO utilizers would help elucidate the microbial mechanisms underpinning intestinal homeostasis in humans.

Funding information

This work was supported by the Institute for Fermentation, Osaka (IFO): IFO research grant L-2021-1-002 and G-2024-1-011.

Acknowledgments

Computation time was provided by the Super Computer System, Institute for Chemical Research, Kyoto University.

Author contributions

Yuka Katayama: Conceptualization, Methodology, Software, Validation, Formal analysis, Writing - Original Draft, Writing - Review & Editing, Visualization. **Ryoma Kamikawa**: Conceptualization, Writing - Original Draft, Writing - Review & Editing, Supervision. **Takashi Yoshida**: Conceptualization, Writing - Review & Editing,

Supervision, Funding acquisition.

Conflicts of interest

The authors declare that there are no conflicts of interest.

Figure legends

Fig. 1 The CODH-bearing prokaryotic genomes in the human gut microbiome

(a) Taxonomic classification of publicly available 30,691 human gut prokaryotic genomes. The genomes were the representatives of more than 0.38 million genomes, clustered with 97.5% nucleotide sequence identity (Hiseni *et al*., 2021). The numbers in parentheses show those of genomes. (b) The number of genomes that possess 1–3 CODHs. No genome possessed four or more CODHs in the used dataset. (c) The number of genomes that possess a CODH belonging to any of the six clades. (d) Taxonomic classification of the CODH-bearing genomes. (e) Phylogenetic tree of the CODH-bearing genomes. The tip color represents the genus level classifications. The heat maps around the phylogenetic tree exhibits order level taxonomic classifications, CODH clades B, C, D, E, and F and mini CODH from the inner layer, respectively. The outermost layer shows the prevalence of each genome reported in the previous study (Hiseni *et al.* 2021). The number indicates the metagenome samples where each genome was analyzed in 3,534 metagenome datasets.

Fig. 2 The genomic context of CODH (*cooS*) in the CODH-bearing genomes in the human gut microbiome

The heatmap on the left shows the number of genes detected within the 15 genes upstream and downstream of *cooS* in the genomes of each genus. Genera with more than nine CODH-bearing genomes were selected (red font), and the gene map around *cooS* of the genera is shown on the right. Each box represents protein-coding genes and their coded strands. The gray box represents the others and the hypothetical protein. *cooS*, carbon monoxide dehydrogenase catalytic subunit gene; *cooC*, carbon monoxide dehydrogenase maturation gene; *cooF*, ferredoxin-like protein-coding gene; *acs*, acetyl-CoA synthase gene; *fhs*, formyltetrahydrofolate synthase gene; *folD*, methylene-tetrahydrofolate dehydrogenase/cyclohydrolase gene; *metF*, methylene tetrahydrofolate reductase gene; *metK*, S-adenosylmethionine synthetase gene; *fnor*, flavin adenine dinucleotide-dependent NAD(P) oxidoreductase gene; *pckA*, phosphoenolpyruvate carboxykinase gene. *pfl*, pyruvate: formate lyase gene; *cys*, cysteine synthase gene; *nar*, nitrate reductase gene; MFS transporter, major facilitator superfamily transporter; ABC transporter, ATP-binding cassette transporters.

Fig. 3 The presence rates of the WLP genes and other related genes in the human gut genomes.

Genomes possessing *cooS* and *acsB* were analyzed to reveal the presence of WLP-related and other genes. The presence rates are displayed for each genus on a heatmap. *cooS*, carbon monoxide dehydrogenase catalytic subunit gene; *cooC*, carbon monoxide dehydrogenase maturation gene; *cooF*, ferredoxin-like protein-coding gene; *acsBCDE*, acetyl-CoA synthase gene; *fdhA*/*fdhF*/*fdoG*, formate dehydrogenase catalytic subunit gene; *fhs*, formyltetrahydrofolate synthase gene; *folD*, methylene-tetrahydrofolate dehydrogenase/cyclohydrolase gene; *metF*, methylene tetrahydrofolate reductase gene; *rnf* ferredoxin:NAD^+ oxidoreductase gene; *hycE*, energy-converting hydrogenase catalytic subunit gene; *hdr-mvh*, methylenetetrahydrofolate reductase gene; *por*/*porABCD*, pyruvate: ferredoxin oxidoreductase *pfl*, pyruvate: formate lyase gene. *focA*/*oxlT*/*fdhC*, genes are involved in formate transport. Other genes, genes involved in formate consumption and metabolism. The right bar represents the number of genomes in each genus.

Fig. 4 Prevalence and diversity of CODH transcripts in the healthy human gut

metatranscriptomic datasets.

(a) Composition of sequence reads in the 110 of publicly available metatranscriptomes in the health human stool samples. (b) Number of detected CODHs among 1,380 non-redundant CODH sequences. (c) Relative abundances of host bacteria of the detected CODH sequences in the metatranscriptomic datasets. (d) Relative abundances of CODH types in the metatranscriptomic datasets. In the x-axis of the graphs, datasets with a greater number of filtered RNA are aligned from left to right.

Fig. 5 Abundance of WLP-related transcripts in the human gut metatranscriptomes.

The heatmaps illustrate the relative abundances of WLP transcripts derived from seven different CODH-bearing genera that were most prevalently detected in the 110 human gut metatranscriptomes. Genomes with more than 94% completeness were selected as references to map the six metatranscriptomic datasets with the highest numbers of CODH reads for the selected genomes. The seven genomes were derived from (a) *Blautia wexlerae*, (b) *Choladocola* sp003481535, (c) *Fusicatenibacter saccharivorans*, (d) UMGS1375 sp900066615, (e) *Faecousia* sp900546075, and (f) CAG-170 sp003516765.

Fig. 6 Genome-based metabolic reconstructions of the Fdh-lacking WLP in the human gut

A schematic of Fdh (formate dehydrogenase)-lacking WLP commonly identified in the human gut bacteria, including genera *Blautia* and *Fusicatenibacter* from Bacillota order Lachnospirales. The figure shows an example of metabolic pathways. Pyruvate which is mainly obtained via glycolysis is converted to acetyl-CoA and formate by *pflB* (pyruvate: formate lysase) (red lines). Formate is then progressively converted to methyl moiety in WLP (orange lines) using formyl-tetrahydrofolate synthase (Fhs), methylene-tetrahydrofolate dehydrogenase/cyclohydrolase (FolD), and methylene-tetrahydrofolate reductase (MetF). CO and CH_3 are then incorporated to acetyl-CoA with CO dehydrogenase/acetyl-CoA synthase (CODH/ACS). The WLP may contribute to toxic CO utilization, carbon acquisition, sink of electrons, and carbons obtained by glycolysis in the human gut prokaryotes.

Fig. S1 Metabolic function hits detected in the three groups of genomes

The metabolic functions of the three genome groups were analyzed using METABOLIC v4.0. The y-axis represents the genomes grouped by phylogeny (genera and orders) and

origin. The first group consists of CODH-bearing genomes detected in the HumGut database (red). The second group comprises non-CODH-bearing genomes, which belong to the same family as the CODH-bearing genomes in the HumGut database (yellow). The third group comprised CO utilizers found in environments other than that of the intestines (blue). The x-axis represents the functions identified in each genome group. The functions were characterized by the presence of genes in METABOLIC. The following list represents the functions and the detected genes associated with each function, which are ordered from left to right in the columns: thermophilic-specific (reverse gyrase), amino acid utilization (4-aminobutyrate aminotransferase and related aminotransferases, aminotransferase class I and II, phosphoserine aminotransferase, ornithine/acetylornithine aminotransferase, branched-chain amino acid aminotransferase/4-amino-4-deoxychorismate lyase, aspartate/tyrosine/aromatic aminotransferase, histidinol-phosphate/aromatic aminotransferase, serine-pyruvate aminotransferase/archaeal aspartate aminotransferase), ethanol fermentation (aldehyde dehydrogenase, alcohol dehydrogenase), aromatics degradation (*catA*, *ubiX*||*bsdC*, *bcrABCD*), complex carbon degradation (cellobiosidase, cellulase, beta-glucosidase, arabinosidase, beta-glucuronidase, alpha-L-rhamnosidase, mannan endo-1,4-beta-mannosidase, alpha-D-xyloside xylohydrolase, beta-xylosidase, beta-mannosidase, beta-

galactosidase, alpha-amylase, glucoamylase, pullulanase, isoamylase, chitinase, hexosaminidase), fermentation (*porA*, *adh*, *ldh*, *acdA*||*ack*||*pta*, *acs*, *pflD*), C1 metabolism (*mxaF* or *mdh*, *mauAB*, *fdhA*||*fghA*||*frmA*||mycoS_dep_FDH||*fae*, *fdoG*||*fdwB*||*fdoH*||*fdhAB*, *coxS*||*coxM*||*coxL*), methane metabolism (*pmoABC*, *mmoBD*, *mcrABC*), carbon fixation (Form I, Form II, *mcr*||K14469, K14466||K18861, K18861||4hbl, *cdhD*||*cdhE*||*cooS*, *aclAB*), nitrogen cycling (*amoABC*, *anfDKG*||*nifDK*||*vnfDKG*||*nifH*, *nxrAB*, *napAB*||*narGH*, *nrfADH*||*nirBD*, *nirKS*||*octR*, *norBC*, *nosDZ*, *hzoA*||*hzsA*), sulfur cycling (*fccB*||*sqr*, *dsrABD*||*asrABC*, *sdo*||*sor*, *sreABC*||*sor*, *soxBCY*, *aprA*||*sat*, *phsA*), hydrogenase (FeFe-group-a13, FeFe-group-a2, FeFe-group-a4, FeFe-group-b, FeFe-group-c1, FeFe-group-c2, FeFe-group-c3, Fe hydrogenase, Nife-group-1, NiFe-group-2ade, NiFe-group-2bc, NiFe-group-3abd, NiFe-group-3c, NiFe-group-4a-g, NiFe-group-4hi), oxidative phosphorylation (*nuoABC*, *ndhABC*, *sdhCD*, *petAB*||*fbcH*, *atpAB* (V/A-type), *atpAD* (F-type)), oxygen metabolism (*coxAB*, *ccoNOP*, *cyoABCD*, *cydAB*, *qoxAB*), urea utilization (*ureABC*), halogenated compound utilization (E3.8.1.2||*pcpC*||*cprA*||*pceA*), perchlorate reduction (*pcrAB*), chlorite reduction (*cld*), arsenic cycling (*arxA*||*aioA*, *arrA*), selenate reduction (*ygfMK*||*xdhD*), nitrile hydration (*nthAB*), and metal reduction (iron reduction series genes). Only *acdA*||*ack*||*pta* for fermentation and *lacZ* for complex carbon degradation

are highlighted.

Fig. S2 Comparisons of the gene presence rates between human gut and other environmental WLP-containing bacteria.

The presence of WLP-related genes was compared between the genomes 667 of human gut bacteria and 554 bacteria from other environments, all of which possess CODH/ACS. The genes were classified into four categories: (a) WLP-related genes, (b) genes involved in energy conservation, (c) genes involved in pyruvate metabolism, and (d) hydrogenase genes. The presence of these genes was determined by analyzing KO annotations using eggNOG-mapper v2.2.5. The presence of hydrogenase genes was assessed using METABOLIC v4.0. *cooS*, carbon monoxide dehydrogenase catalytic subunit gene; *cooC*, carbon monoxide dehydrogenase maturation gene; *cooF*, ferredoxin-like protein-coding gene; *acsBCDE*, acetyl-CoA synthase gene; *fdhA*/*fdhF*/*fdoG*, formate dehydrogenase catalytic subunit gene; *fhs*, formyltetrahydrofolate synthase gene; *folD*, methylene-tetrahydrofolate dehydrogenase/cyclohydrolase gene; *metF*, methylene tetrahydrofolate reductase gene; *rnf*, ferredoxin:NAD^+ oxidoreductase gene; *hycE*, energy-converting hydrogenase catalytic subunit gene; *hdr-mvh*, methylenetetrahydrofolate reductase gene; *porABCD*,

pyruvate ferredoxin oxidoreductase gene; *pfl*, pyruvate: formate lyase gene; *focA*/*oxlT*/*fdhC*, genes involved in formate transport. Other genes, genes involved in formate consumption and metabolism.

Fig. S3 The co-occurrence of genes for Fdh and FeFe hydrogenase in the genomes of CODH/ACS-bearing bacteria in the human gut

(a) The count of genomes containing Fdh genes and FeFe hydrogenase genes, as well as those lacking either gene. These counts were utilized to calculate odds ratio. (b) The odds ratio for the co-presence of Fdh and FeFe hydrogenase genes. The red characters indicate there were significant differences.

Fig. S4 Genera level diversity in the healthy human gut metatranscriptomic datasets.

The genera level classification of the hosts of detected CODH transcripts are displayed. In the x-axis of the graph, datasets with a greater number of filtered RNA are aligned from left to right.

References

Abu-Ali GS, Mehta RS, Lloyd-Price J *et al*. Metatranscriptome of human faecal microbial communities in a cohort of adult men. *Nat Microbiol* 2018;3:356–66.

Adachi Y, Inoue M, Yoshida T *et al*. Genetic engineering of carbon monoxide-dependent hydrogen-producing machinery in *Parageobacillus thermoglucosidasius*. *Microbes Environ* 2020;**35**:ME20101.

Adam PS, Borrel G, Gribaldo S. Evolutionary history of carbon monoxide dehydrogenase/acetyl-CoA synthase, one of the oldest enzymatic complexes. *Proc Natl Acad Sci U S A* 2018;**115**:E1166–73.

Almeida A, Mitchell AL, Boland M *et al.* A new genomic blueprint of the human gut microbiota. *Nature 2019 568:7753* 2019;**568**:499–504.

Almeida A, Nayfach S, Boland M *et al.* A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nat Biotechnol* 2021;**39**:105–14.

Alonso J-R, Cardellach F, López S *et al.* Carbon monoxide specifically inhibits cytochrome c oxidase of human mitochondrial respiratory chain. *Pharmacol Toxicol* 2003;**93**:142–6.

Benoit SL, Maier RJ, Sawers RG *et al*. Molecular hydrogen metabolism: a widespread trait of pathogenic bacteria and protists. *Microbiology and Molecular Biology Reviews* 2020;**84**, DOI: 10.1128/MMBR.00092-19.

Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* 1999;27:573–80.

Buchfink B, Reuter K, Drost H-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods* 2021;**18**:366–8.

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;30:2114–20.

Cantalapiedra CP, Hernández-Plaza A, Letunic I *et al.* eggNOG-mapper v2: Functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol Biol Evol* 2021;**38**:5825–9.

Chaumeil P-A, Mussig AJ, Hugenholtz P *et al*. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 2019, DOI: 10.1093/bioinformatics/btz848.

Diender M, Stams AJM, Sousa DZ. Pathways and bioenergetics of anaerobic carbon monoxide fermentation. *Front Microbiol* 2015;**6**:1275.

Dietrich HM, Müller V. Ferredoxin as a physiological electron donor for carbon dioxide fixation to formate in a bacterial carbon dioxide reductase. *ACS Catal* 2023;**13**:12374–82.

Ernst A, Zibrak JD. Carbon monoxide poisoning. *New England Journal of Medicine*

1998;**339**:1603–8.

Garner RE, Kraemer SA, Onana VE *et al*. A genome catalogue of lake bacterial diversity and its drivers at continental scale. *Nat Microbiol* 2023;8:1920–34.

Geelhoed JS, Henstra AM, Stams AJM. Carboxydotrophic growth of *Geobacter sulfurreducens*. *Appl Microbiol Biotechnol* 2016;**100**:997–1007.

Gencic S, Grahame DA. Diverse energy-conserving pathways in *Clostridium difficile*: Growth in the absence of amino acid stickland acceptors and the role of the Wood-Ljungdahl pathway. *J Bacteriol* 2020;**202**:233–53.

Gomez, J. A.L., Mukhopadhya, I., Duncan, S. H., Louis, P., Shaw, S., Collie□Duguid, E., Crost, E., Juge, N., & Flint, H. J. (2019). Formate cross□feeding and cooperative metabolic interactions revealed by transcriptomics in co□cultures of acetogenic and amylolytic human colonic bacteria. *Environmental Microbiology*, 21(1), 259–271. https://doi.org/10.1111/1462-2920.14454

Gurbich TA, Almeida A, Beracochea M *et al*. MGnify Genomes: A resource for biome-specific microbial genome catalogues. *J Mol Biol* 2023;435:168016.

Hackl T, Ankenbrand MJ. gggenomes: A Grammar of Graphics for Comparative Genomics, R package version 0.9.5.9000. *https://github.com/thackl/gggenomes* 2022.

Hattori S, Kamagata Y, Hanada S *et al*. *Thermacetogenium phaeum* gen. nov., sp. nov., a strictly anaerobic, thermophilic, syntrophic acetate-oxidizing bacterium. *Int J Syst Evol Microbiol* 2000;**50**:1601–9.

Hiseni P, Rudi K, Wilson RC *et al*. HumGut: a comprehensive human gut prokaryotic genomes collection filtered by metagenome data. *Microbiome* 2021;**9**:1–12.

Hopper CP, de La Cruz LK, Lyles K v. *et al.* Role of carbon monoxide in host–gut microbiome communication. *Chem Rev* 2020;**120**:13273–311.

Hyatt D, LoCascio PF, Hauser LJ *et al*. Gene and translation initiation site prediction in metagenomic sequences. *Bioinformatics* 2012;**28**:2223–30.

Imaura Y, Okamoto S, Hino T *et al.* Isolation, Genomic sequence and physiological characterization of *Parageobacillus* sp. G301, an isolate capable of both hydrogenogenic and Aerobic Carbon Monoxide Oxidation. *Appl Environ Microbiol* 2023, DOI: 10.1128/aem.00185-23.

Inoue M, Nakamoto I, Omae K *et al*. Structural and phylogenetic diversity of anaerobic carbon-monoxide dehydrogenases. *Front Microbiol* 2019;**9**:3353.

Inoue M, Omae K, Nakamoto I *et al*. Biome-specific distribution of Ni-containing carbon monoxide dehydrogenases. *Extremophiles* 2022;**26**:1–11.

Katoh K, Standley DM. MAFFT Multiple sequence alignment software version 7:

904 improvements in performance and usability. *Mol Biol Evol* 2013;**30**:772–80.

905 Katz L, Griswold T, Morrison S *et al.* Mashtree: a rapid comparison of whole genome
906 sequence files. *J Open Source Softw* 2019;**4**:1762.

907 King GM. Molecular and culture-based analyses of aerobic carbon monoxide oxidizer
908 diversity. *Appl Environ Microbiol* 2003;**69**:7257–65.

909 Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*
910 2012;9:357–9.

911 Leigh JA, Mayer F, Wolfe RS. *Acetogenium kivui*, a new thermophilic hydrogen-
912 oxidizing acetogenic bacterium. *Arch Microbiol* 1981;**129**:275–80.

913 Levine AS, Bond JH, Prentiss RA *et al.* Metabolism of carbon monoxide by the colonic
914 flora of humans. *Gastroenterology* 1982;**83**:633–7.

915 Liu C, Li J, Zhang Y et al. Influence of glucose fermentation on CO_2 assimilation to
916 acetate in homoacetogen *Blautia coccoides* GA-1. *J Ind Microbiol Biotechnol*
917 2015;**42**:1217–24.

918 Miller TL, Wolin MJ. Pathways of acetate, propionate, and butyrate formation by the
919 human fecal microbial flora. *Appl Environ Microbiol* 1996;**62**:1589–92.

920 Morse D, Choi AMK. Heme Oxygenase-1. *Am J Respir Crit Care Med* 2005;**172**:660–
921 70.

Moussard H, L'Haridon S, Tindall BJ *et al. Thermodesulfatator indicus* gen. nov., sp. nov., a novel thermophilic chemolithoautotrophic sulfate-reducing bacterium isolated from the central Indian ridge. *Int J Syst Evol Microbiol* 2004;**54**:227–33.

Oelgeschläger E, Rother M. Carbon monoxide-dependent energy metabolism in anaerobic bacteria and archaea. *Arch Microbiol* 2008;**190**:257–69.

Omae K, Fukuyama Y, Yasuda H *et al.* Diversity and distribution of thermophilic hydrogenogenic carboxydotrophs revealed by microbial community analysis in sediments from multiple hydrothermal environments in Japan. *Arch Microbiol* 2019;**201**:969–82.

Ondov BD, Treangen TJ, Melsted P *et al.* Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol* 2016;**17**:132.

Onyiah JC, Sheikh SZ, Maharshak N *et al.* Heme oxygenase-1 and carbon monoxide regulate intestinal homeostasis and mucosal immune responses to the enteric microbiota. *Gut Microbes* 2014;**5**:220–4.

Oren A, Garrity GM. Valid publication of the names of forty-two phyla of prokaryotes. *Int J Syst Evol Microbiol* 2021;**71**, DOI: 10.1099/ijsem.0.005056.

Pokusaeva K, Fitzgerald GF, van Sinderen D. Carbohydrate metabolism in Bifidobacteria. Genes Nutr 2011;6:285–306.

Prockop LD, Chichkova RI. Carbon monoxide intoxication: An updated review. *J Neurol Sci* 2007;**262**:122–30.

Ragsdale SW. Life with carbon monoxide. *Crit Rev Biochem Mol Biol* 2004;**39**:165–95.

Ragsdale SW. Enzymology of the Wood-Ljungdahl Pathway of acetogenesis. *Ann N Y Acad Sci* 2008;**1125**:129–36.

Ragsdale SW, Pierce E. Acetogenesis and the Wood–Ljungdahl pathway of CO_2 fixation. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 2008;**1784**:1873–98.

Rogosa M. THE GENUS *VEILLONELLA* I. *J Bacteriol* 1964;**87**:162–70.

Sawers G, Watson G. A glycyl radical solution: oxygen-dependent interconversion of pyruvate formate-lyase. *Mol Microbiol* 1998;**29**:945–54.

Schoelmerich MC, Müller V. Energy conservation by a hydrogenase-dependent chemiosmotic mechanism in an ancient metabolic pathway. *Proc Natl Acad Sci U S A* 2019;**116**:6329–34.

Schuchmann K, Müller V. A bacterial electron-bifurcating hydrogenase. *Journal of Biological Chemistry* 2012;287:31165–71.

Schuchmann K, Müller V. Autotrophy at the thermodynamic limit of life: a model for energy conservation in acetogenic bacteria. 2014, DOI: 10.1038/nrmicro3365.Sen

A, Imlay JA. How microbes defend themselves from incoming hydrogen peroxide. *Front Immunol* 2021;**12**, DOI: 10.3389/fimmu.2021.667343.

Slobodkin A, Slobodkina G, Allioux M *et al*. Genomic insights into the carbon and energy metabolism of a thermophilic deep-sea bacterium deferribacter autotrophicus revealed new metabolic traits in the phylum deferribacteres. *Genes (Basel)* 2019;**10**:849.

Sokolova TG, Henstra A-M, Sipma J *et al*. Diversity and ecophysiological features of thermophilic carboxydotrophic anaerobes. *FEMS Microbiol Ecol* 2009;**68**:131–41.

Suzuki Y, Kishigami T, Inoue K *et al*. *Bacillus thermoglucosidasius* sp. nov., a new species of obligately thermophilic Bacilli. *Syst Appl Microbiol* 1983;**4**:487–95.

Stewart RD, Auffret MD, Warr A *et al*. Compendium of 4,941 rumen metagenome-assembled genomes for rumen microbiome biology and enzyme discovery. *Nat Biotechnol* 2019;37:953–61.

Svetlichny VA, Sokolova TG, Gerhardt M *et al*. *Carboxydothermus hydrogenoformans* gen. nov., sp. nov., a CO-utilizing thermophilic anaerobic bacterium from hydrothermal environments of Kunashir Island. *Syst Appl Microbiol* 1991;**14**:254–60.

Tian L, Lo J, Shao X *et al*. Ferredoxin:NAD $^{+}$ oxidoreductase of

Thermoanaerobacterium saccharolyticum and its role in ethanol formation. *Appl Environ Microbiol* 2016;**82**:7134–41.

Trischler R, Roth J, Sorbara MT *et al.* A functional Wood–Ljungdahl pathway devoid of a formate dehydrogenase in the gut acetogens *Blautia wexlerae*, *Blautia luti* and beyond. *Environ Microbiol* 2022;**24**:3111–23.

Weiss MC, Sousa FL, Mrnjavac N *et al.* The physiology and habitat of the last universal common ancestor. *Nat Microbiol* 2016;**1**:16116.

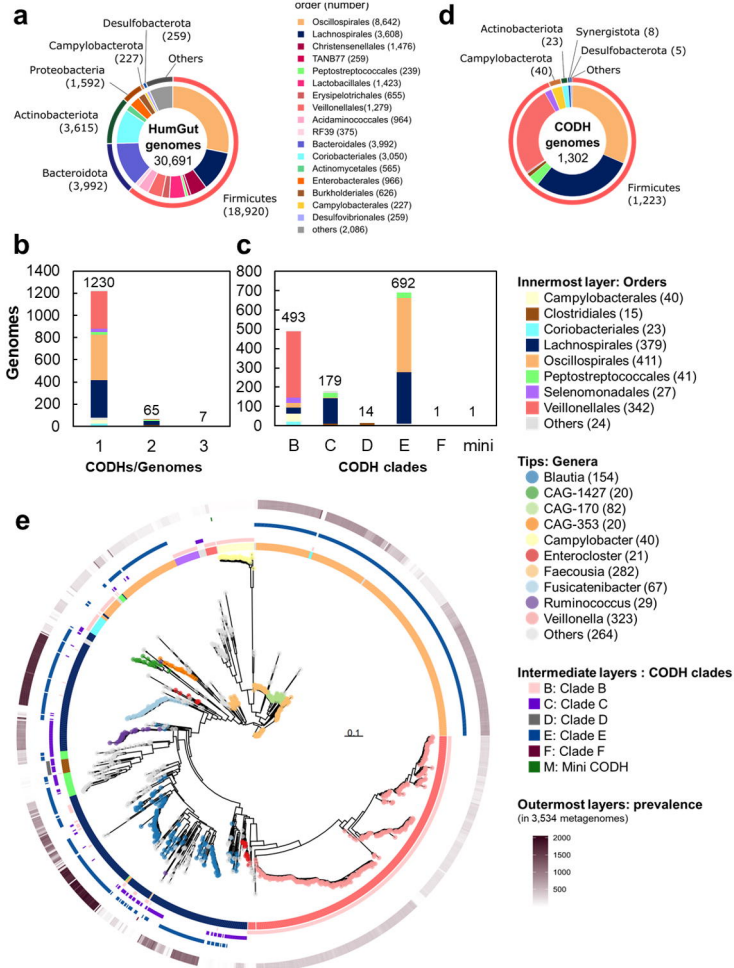
Wolin MJ, Miller TL, Collins MD *et al.* Formate-dependent growth and homoacetogenic fermentation by a bacterium from human feces: Description of *Bryantella formatexigens* gen. nov., sp. nov. *Appl Environ Microbiol* 2003;**69**:6321–6.

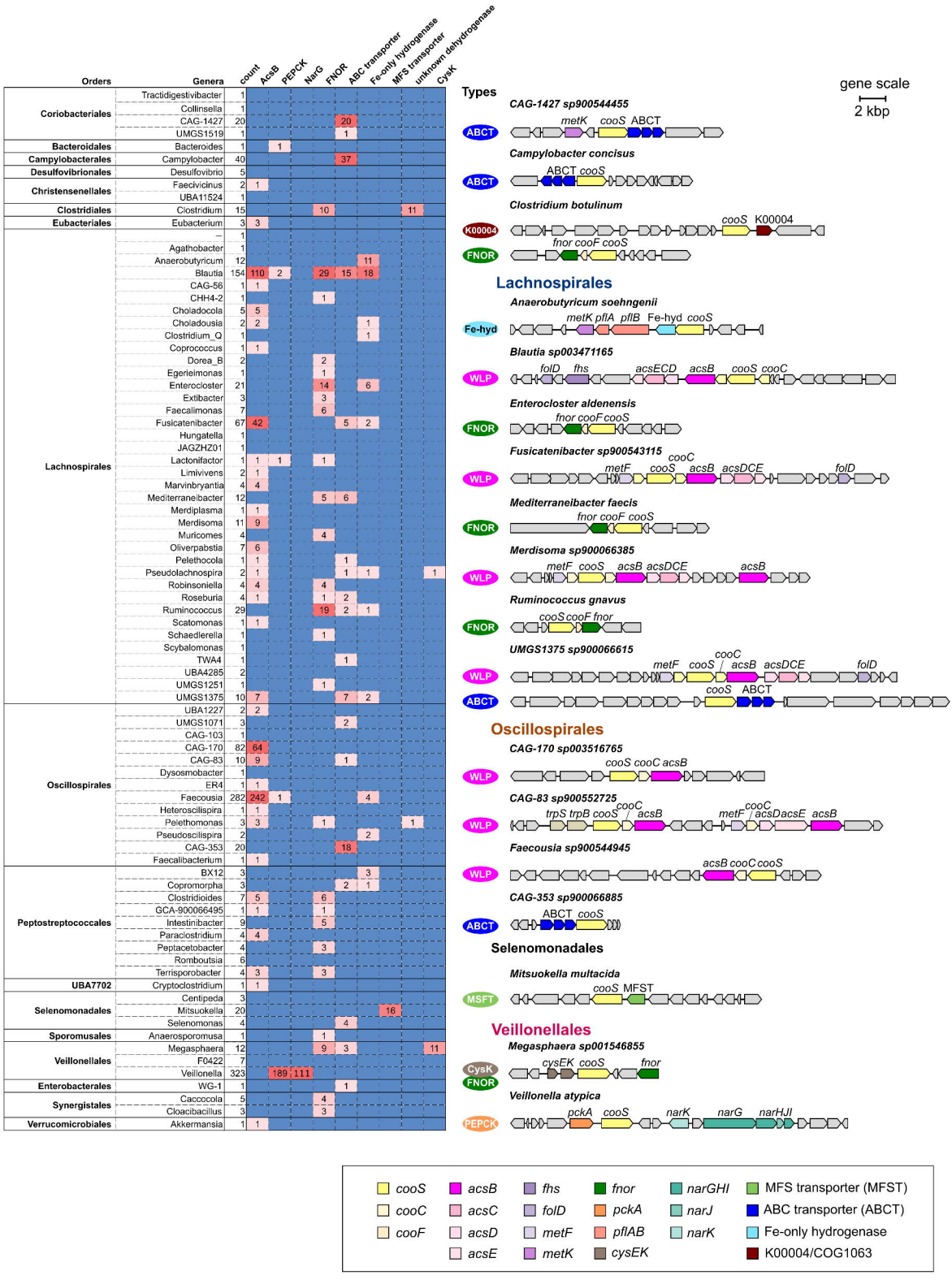
Yao Y, Fu B, Han D et al. Formate-dependent acetogenic utilization of glucose by the fecal acetogen *Clostridium bovifaecis*. *Appl Environ Microbiol* 2020;86, DOI: 10.1128/AEM.01870-20.

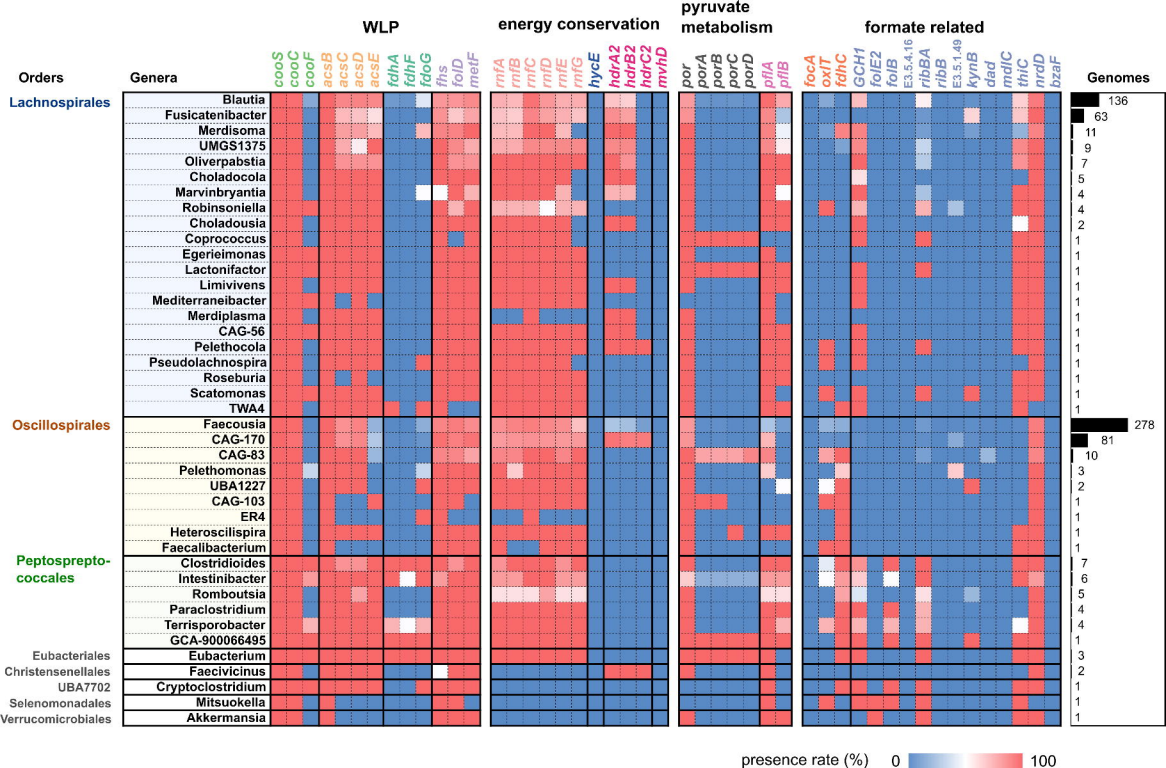
Yao Y, Fu B, Han D *et al.* Reduction, evolutionary pattern and positive selection of genes encoding formate dehydrogenase in Wood–Ljungdahl pathway of gastrointestinal acetogens suggests their adaptation to formate□rich habitats. *Environ Microbiol Rep* 2023;**15**:129–41.

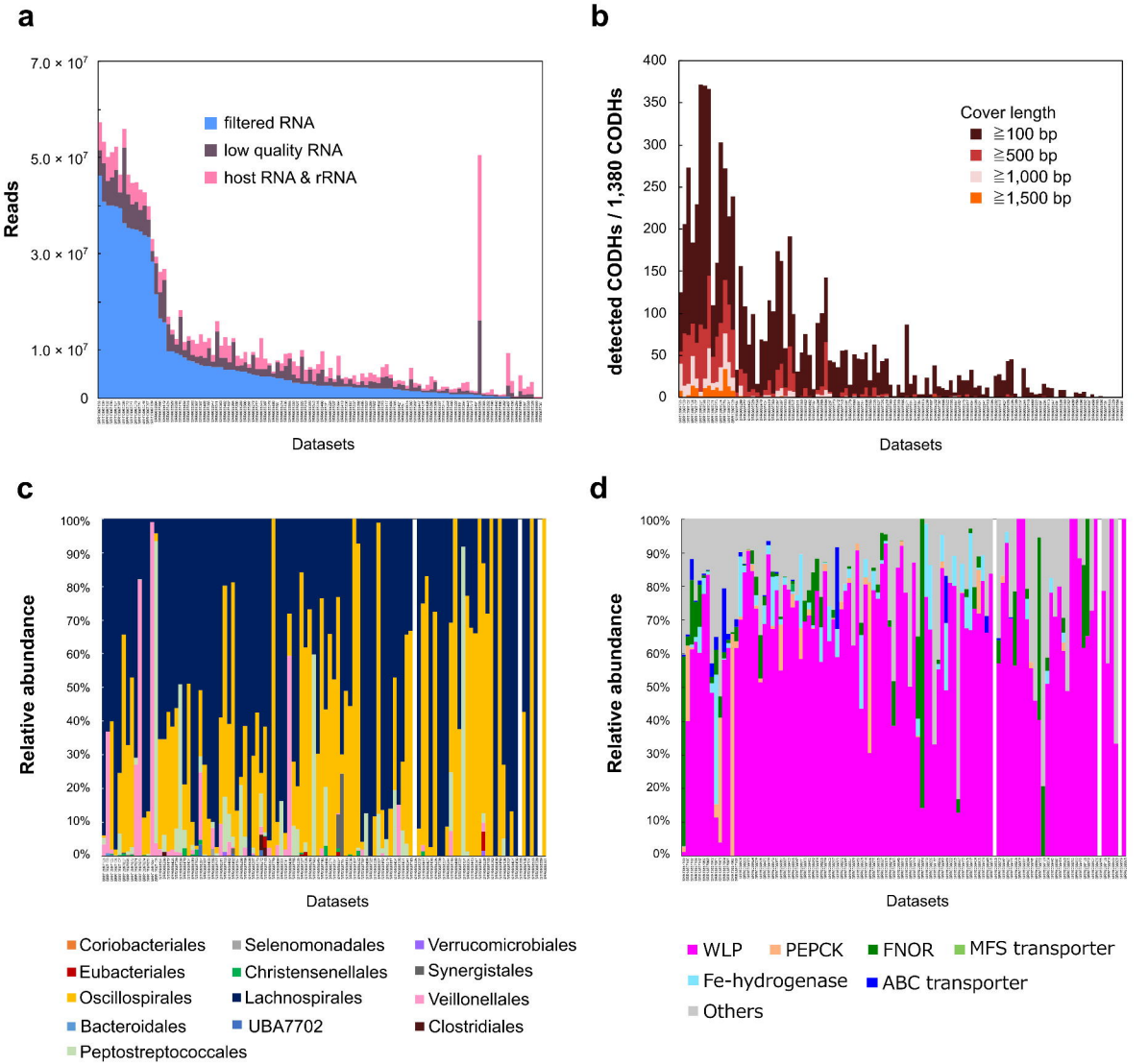
Yu G, Smith DK, Zhu H *et al*. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol* 2017;**8**:28–36.

Zhou Z, Tran PQ, Breister AM *et al*. METABOLIC: high-throughput profiling of microbial genomes for functional traits, metabolism, biogeochemistry, and community-scale functional networks. *Microbiome* 2022;**10**:33.


a
Desulfobacterota (259)
Campylobacterota (227)
Others
Proteobacteria (1,592)
Actinobacteriota (3,615)
HumGut genomes 30,691
Bacteroidota (3,992)
Firmicutes (18,920)
order (number)
Oscillospirales (8,642)
Lachnospirales (3,608)
Christensenellales (1,476)
TANB77 (259)
Peptostreptococcales (239)
Lactobacillales (1,423)
Erysipelotrichales (655)
Veillonellales(1,279)
Acidaminococcales (964)
RF39 (375)
Bacteroidales (3,992)
Coriobacteriales (3,050)
Actinomycetales (565)
Enterobacterales (966)
Burkholderiales (626)
Campylobacterales (227)
Desulfovibrionales (259)
others (2,086)
d
Actinobacteriota (23)
Synergistota (8)
Desulfobacterota (5)
Campylobacterota (40)
Others
CODH genomes 1,302
Firmicutes (1,223)
b
Genomes
1400
1200
1000
800
600
400
200
0
1230
65
7
1
2
3
CODHs/Genomes
c
800
700
600
500
400
300
200
100
0
493
179
14
692
1
1
B
C
D
E
F
mini
CODH clades
Innermost layer: Orders
Campylobacterales (40)
Clostridiales (15)
Coriobacteriales (23)
Lachnospirales (379)
Oscillospirales (411)
Peptostreptococcales (41)
Selenomonadales (27)
Veillonellales (342)
Others (24)
Tips: Genera
Blautia (154)
CAG-1427 (20)
CAG-170 (82)
CAG-353 (20)
Campylobacter (40)
Enterocloster (21)
Faecousia (282)
Fusicatenibacter (67)
Ruminococcus (29)
Veillonella (323)
Others (264)
e
0.1
Intermediate layers : CODH clades
B: Clade B
C: Clade C
D: Clade D
E: Clade E
F: Clade F
M: Mini CODH
Outermost layers: prevalence
(in 3,534 metagenomes)
2000
1500
1000
500

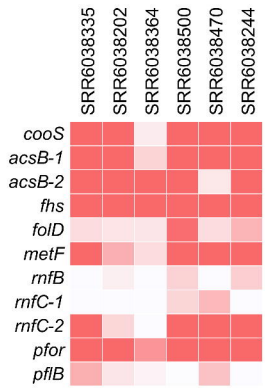

Orders
Genera
count
AcsB
PEPCK
NarG
FNOR
ABC transporter
Fe-only hydrogenase
MFS transporter
unknown dehydrogenase
CysK
Coriobacteriales
Tractidigestivibacter 1
Collinsella 1
CAG-1427 20 20
UMGS1519 1 1
Bacteroidales
Bacteroides 1 1
Campylobacterales
Campylobacter 40 37
Desulfovibrionales
Desulfovibrio 5
Christensenellales
Faecivicinus 2 1
UBA11524 1
Clostridiales
Clostridium 15 10 11
Eubacteriales
Eubacterium 3 3
Lachnospirales
– 1
Agathobacter 1
Anaerobutyricum 12 11
Blautia 154 110 2 29 15 18
CAG-56 1 1
CHH4-2 1 1
Choladocola 5 5
Choladousia 2 2 1
Clostridium_Q 1 1
Coprococcus 1 1
Dorea_B 2 2
Egerieimonas 1 1
Enterocloster 21 14 6
Extibacter 3 3
Faecalimonas 7 6
Fusicatenibacter 67 42 5 2
Hungatella 1
JAGZHZ01 1
Lactonifactor 1 1 1 1
Limivivens 2 1
Marvinbryantia 4 4
Mediterraneibacter 12 5 6
Merdiplasma 1 1
Merdisoma 11 9
Muricomes 4 4
Oliverpabstia 7 6
Pelethocola 1 1 1
Pseudolachnospira 2 1 1 1 1
Robinsoniella 4 4 4
Roseburia 4 1 1 2
Ruminococcus 29 19 2 1
Scatomonas 1 1
Schaedlerella 1 1
Scybalomonas 1
TWA4 1 1
UBA4285 2
UMGS1251 1 1
UMGS1375 10 7 7 2
Oscillospirales
UBA1227 2 2
UMGS1071 3 2
CAG-103 1
CAG-170 82 64
CAG-83 10 9 1
Dysosmobacter 1
ER4 1 1
Faecousia 282 242 1 4
Heteroscilispira 1 1
Pelethomonas 3 3 1 1
Pseudoscilispira 2 2
CAG-353 20 18
Faecalibacterium 1 1
Peptostreptococcales
BX12 3 3
Coprocomorpha 3 2 1
Clostridioides 7 5 6
GCA-900066495 1 1 1
Intestinibacter 9 5
Paraclostridium 4 4
Peptacetobacter 4 3
Romboutsia 6
Terrisporobacter 4 3 3
UBA7702
Cryptoclostridium 1 1
Selenomonadales
Centipeda 3
Mitsuokella 20 16
Selenomonas 4 4
Sporomusales
Anaerosporomusa 1 1
Veillonellales
Megasphaera 12 9 3 11
F0422 7
Veillonella 323 189 111
Enterobacterales
WG-1 1 1
Synergistales
Caccocola 5 4
Cloacibacillus 3 3
Verrucomicrobiales
Akkermansia 1 1
Types
gene scale
2 kbp
CAG-1427 sp900544455
ABCT
metK cooS ABCT
Campylobacter concisus
ABCT
ABCT cooS
Clostridium botulinum
K00004
cooS K00004
FNOR
fnor cooF cooS
Lachnospirales
Anaerobutyricum soehngenii
Fe-hyd
metK pflA pflB Fe-hyd cooS
Blautia sp003471165
WLP
folD fhs acsECD acsB cooS cooC
Enterocloster aldenensis
FNOR
fnor cooF cooS
Fusicatenibacter sp900543115
WLP
metF cooS cooC acsB acsDCE folD
Mediterraneibacter faecis
FNOR
fnor cooF cooS
Merdisoma sp900066385
WLP
metF cooS acsB acsDCE acsB
Ruminococcus gnavus
FNOR
cooS cooF fnor
UMGS1375 sp900066615
WLP
metF cooS cooC acsB acsDCE folD
ABCT
cooS ABCT
Oscillospirales
CAG-170 sp003516765
WLP
cooS cooC acsB
CAG-83 sp900552725
WLP
trpS trpB cooS cooC acsB metF cooC acsDacsE acsB
Faecousia sp900544945
WLP
acsB cooC cooS
CAG-353 sp900066885
ABCT
ABCT cooS
Selenomonadales
Mitsuokella multacida
MSFT
cooS MFST
Veillonellales
Megasphaera sp001546855
CysK
FNOR
cysEK cooS fnor
Veillonella atypica
PEPCK
pckA cooS narK narG narHJI
cooS
cooC
cooF
acsB
acsC
acsD
acsE
fhs
folD
metF
metK
fnor
pckA
pflAB
cysEK
narGHI
narJ
narK
MFS transporter (MFST)
ABC transporter (ABCT)
Fe-only hydrogenase
K00004/COG1063


WLP
energy conservation
pyruvate metabolism
formate related
Orders
Genera
Genomes
cooS
cooC
cooF
acsB
acsC
acsD
acsE
fdhA
fdhF
fdoG
fhs
folD
metF
rnfA
rnfB
rnfC
rnfD
rnfE
rnfG
hycE
hdrA2
hdrB2
hdrC2
mvhD
por
porA
porB
porC
porD
pflA
pflB
focA
oxlT
fdhC
GCH1
folE2
folB
E3.5.4.16
ribBA
ribB
E3.5.1.49
kynB
dad
mdlC
thiC
nrdD
bzaF
Lachnospirales
Blautia
Fusicatenibacter
Merdisoma
UMGS1375
Oliverpabstia
Choladocola
Marvinbryantia
Robinsoniella
Choladousia
Coprococcus
Egerieimonas
Lactonifactor
Limivivens
Mediterraneibacter
Merdiplasma
CAG-56
Pelethocola
Pseudolachnospira
Roseburia
Scatomonas
TWA4
Oscillospirales
Faecousia
CAG-170
CAG-83
Pelethomonas
UBA1227
CAG-103
ER4
Heteroscilispira
Faecalibacterium
Peptosprepto-coccales
Clostridioides
Intestinibacter
Romboutsia
Paraclostridium
Terrisporobacter
GCA-900066495
Eubacteriales
Eubacterium
Christensenellales
Faecivicinus
UBA7702
Cryptoclostridium
Selenomonadales
Mitsuokella
Verrucomicrobiales
Akkermansia
136
63
11
9
7
5
4
4
2
1
1
1
1
1
1
1
1
1
1
1
278
81
10
3
2
1
1
1
1
7
6
5
4
4
1
3
2
1
1
1
presence rate (%)
0
100

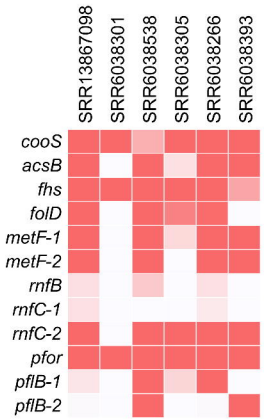

a
Reads
7.0 × 10^7
5.0 × 10^7
3.0 × 10^7
1.0 × 10^7
0
filtered RNA
low quality RNA
host RNA & rRNA
Datasets
b
detected CODHs / 1,380 CODHs
400
350
300
250
200
150
100
50
0
Cover length
≧100 bp
≧500 bp
≧1,000 bp
≧1,500 bp
Datasets
c
Relative abundance
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Datasets
Coriobacteriales
Selenomonadales
Verrucomicrobiales
Eubacteriales
Christensenellales
Synergistales
Oscillospirales
Lachnospirales
Veillonellales
Bacteroidales
UBA7702
Clostridiales
Peptostreptococcales
d
Relative abundance
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Datasets
WLP
PEPCK
FNOR
MFS transporter
Fe-hydrogenase
ABC transporter
Others

Lachnospirales

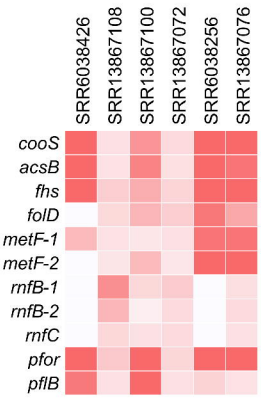
(a) *Blautia wexlerae*



(b) *Choladocola* sp003481535

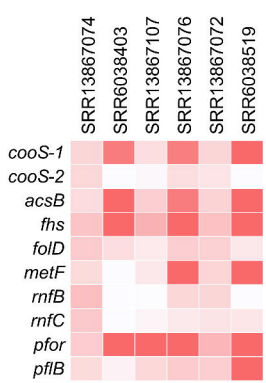


(c) *Fusicatenibacter saccharivorans*

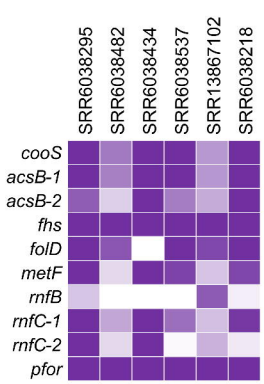


Oscillospirales

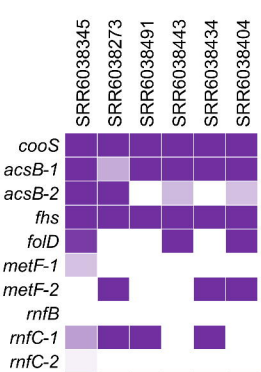
(d) UMGS1375 sp900066615

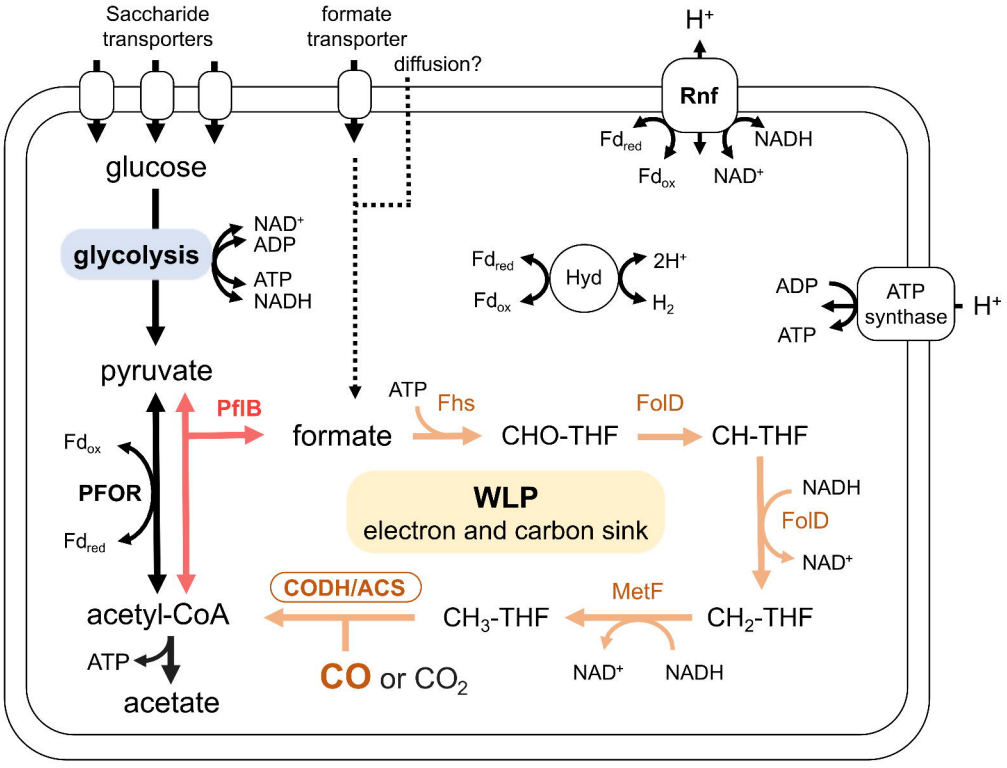


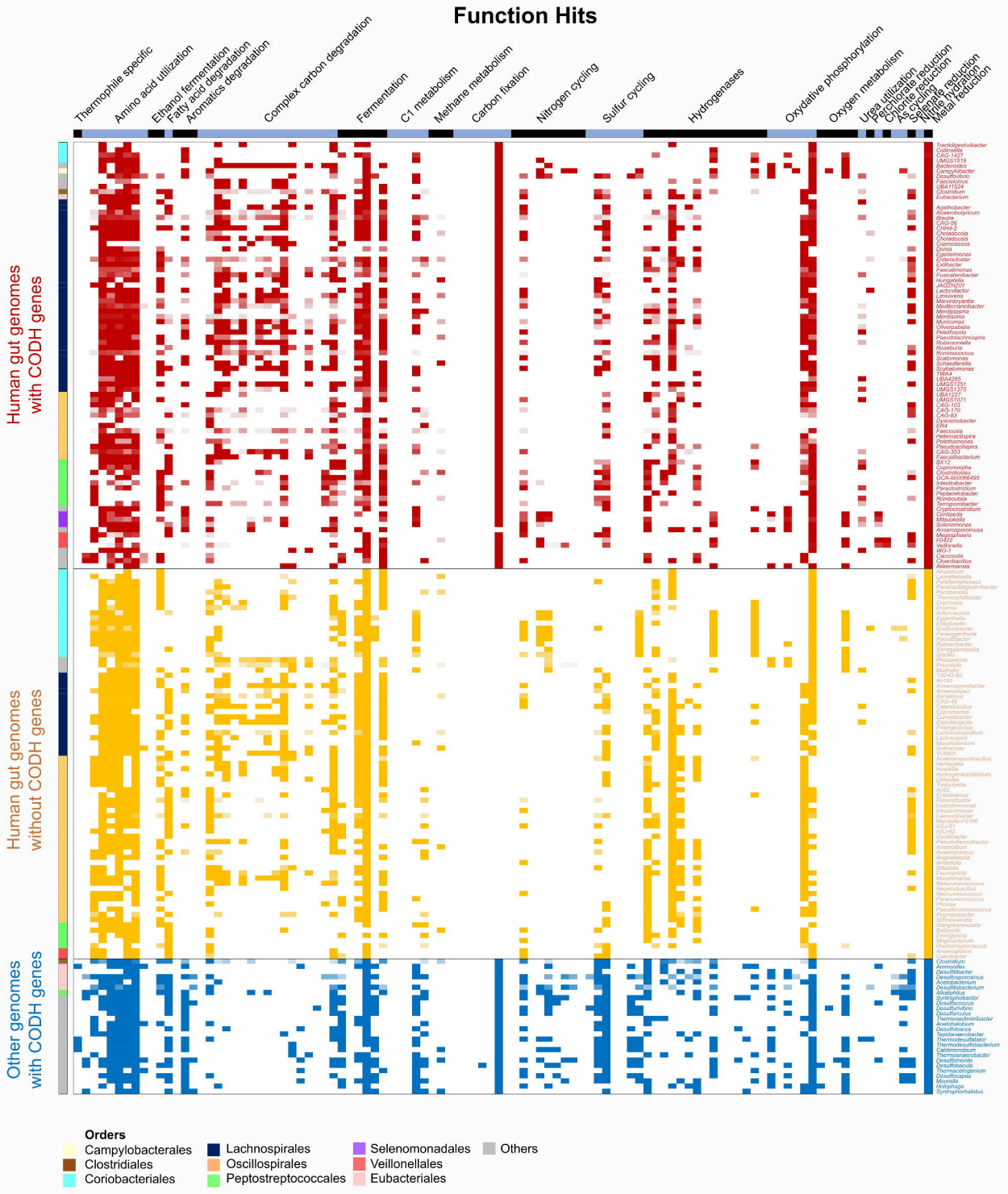
(e) *Faecousia* sp900546075

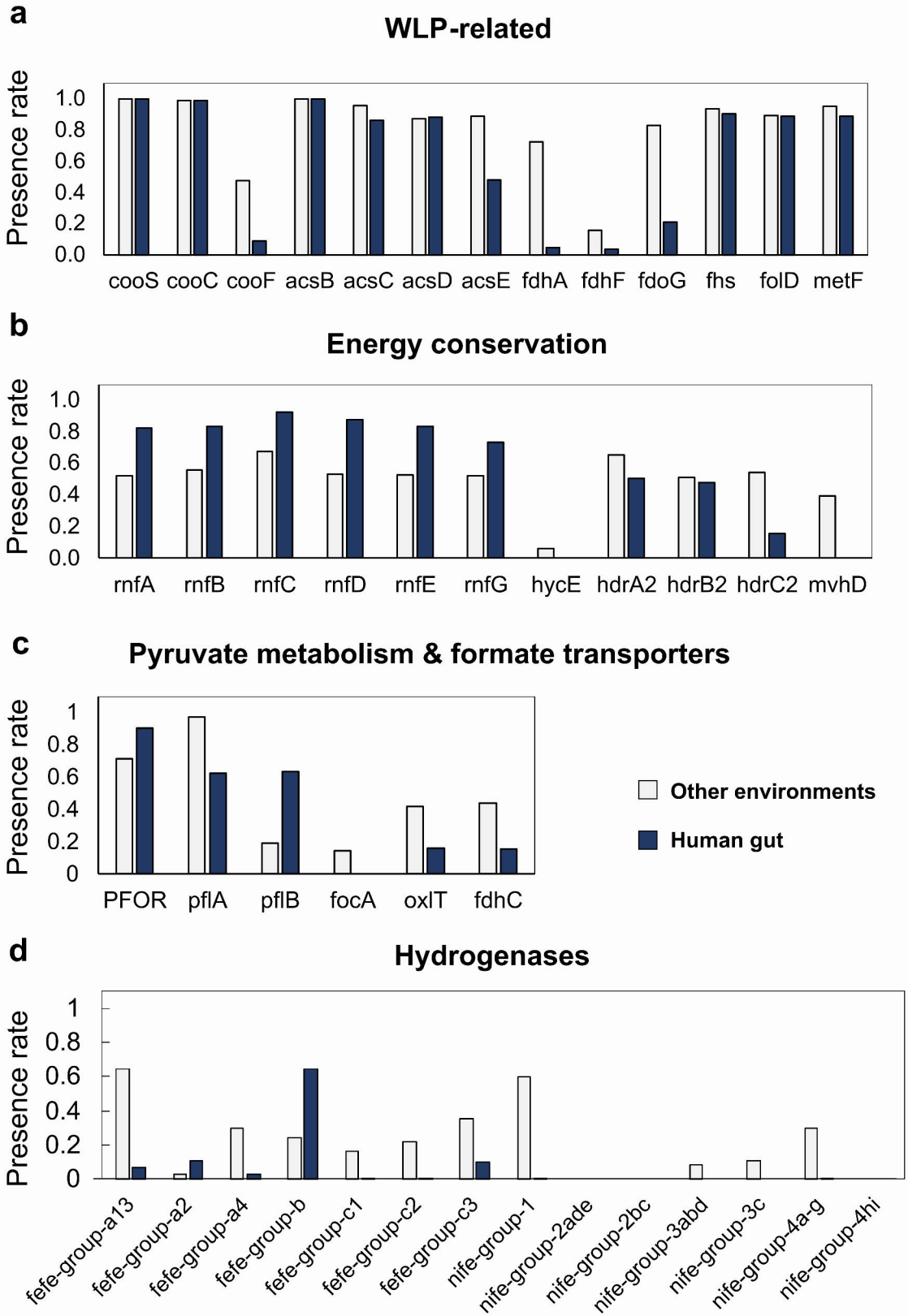


(f) CAG-170 sp003516765




Saccharide transporters
formate transporter
diffusion?
H+
Rnf
Fdred
Fdox
NADH
NAD+
glucose
glycolysis
NAD+
ADP
ATP
NADH
Fdred
Fdox
Hyd
2H+
H2
ADP
ATP
ATP synthase
H+
pyruvate
PflB
formate
ATP
Fhs
CHO-THF
FolD
CH-THF
Fdox
PFOR
Fdred
WLP
electron and carbon sink
NADH
FolD
NAD+
CODH/ACS
acetyl-CoA
CH3-THF
MetF
CH2-THF
ATP
acetate
CO or CO2
NAD+
NADH


Function Hits
Thermophile specific
Amino acid utilization
Ethanol fermentation
Fatty acid degradation
Aromatics degradation
Complex carbon degradation
Fermentation
C1 metabolism
Methane metabolism
Carbon fixation
Nitrogen cycling
Sulfur cycling
Hydrogenases
Oxydative phosphorylation
Oxygen metabolism
Urea utilization
Perchlorate reduction
Chlorite reduction
As. cycling
Selenate reduction
Nitrate reduction
Nitrite reduction
Metal reduction
Human gut genomes with CODH genes
Human gut genomes without CODH genes
Other genomes with CODH genes
Orders
Campylobacterales
Clostridiales
Coriobacteriales
Lachnospirales
Oscillospirales
Peptostreptococcales
Selenomonadales
Veillonellales
Eubacteriales
Others


a
WLP-related
Presence rate
1.0
0.8
0.6
0.4
0.2
0.0
cooS cooC cooF acsB acsC acsD acsE fdhA fdhF fdoG fhs folD metF
b
Energy conservation
Presence rate
1.0
0.8
0.6
0.4
0.2
0.0
rnfA rnfB rnfC rnfD rnfE rnfG hycE hdrA2 hdrB2 hdrC2 mvhD
c
Pyruvate metabolism & formate transporters
Presence rate
1
0.8
0.6
0.4
0.2
0
PFOR pflA pflB focA oxlT fdhC
Other environments
Human gut
d
Hydrogenases
Presence rate
1
0.8
0.6
0.4
0.2
0
fefe-group-a13
fefe-group-a2
fefe-group-a4
fefe-group-b
fefe-group-c1
fefe-group-c2
fefe-group-c3
nife-group-1
nife-group-2ade
nife-group-2bc
nife-group-3abd
nife-group-3c
nife-group-4a-g
nife-group-4hi

a Raw data (genome count)

fdhA vs group A13 FeFe hydrogenase		
	fefe-group-a13 +	fefe-group-a13 -
fdhA+	19	13
fdhA-	26	609

fdhF vs group A13 FeFe hydrogenase		
	fefe-group-a13 +	fefe-group-a13 -
fdhF+	15	9
fdhF-	30	613

fdoG vs group A13 FeFe hydrogenase		
	fefe-group-a13 +	fefe-group-a13 -
fdoG+	23	118
fdoG-	22	504

fdhA vs group A2 FeFe hydrogenase		
	fefe-group-a2 +	fefe-group-a2 -
fdhA+	1	31
fdhA-	69	566

fdhF vs group A2 FeFe hydrogenase		
	fefe-group-a2 +	fefe-group-a2 -
fdhF+	1	23
fdhF-	69	574

fdoG vs group A2 FeFe hydrogenase		
	fefe-group-a2 +	fefe-group-a2 -
fdoG+	9	132
fdoG-	61	465

fdhA vs group A4 FeFe hydrogenase		
	fefe-group-a4 +	fefe-group-a4 -
fdhA+	21	11
fdhA-	0	635

fdhF vs group A4 FeFe hydrogenase		
	fefe-group-a4 +	fefe-group-a4 -
fdhF+	19	5
fdhF-	2	641

fdoG vs group A4 FeFe hydrogenase		
	fefe-group-a4 +	fefe-group-a4 -
fdoG+	21	120
fdoG-	0	526

fdhA vs group B FeFe hydrogenase		
	fefe-group-b +	fefe-group-b -
fdhA+	21	11
fdhA-	407	228

fdhF vs group B FeFe hydrogenase		
	fefe-group-b +	fefe-group-b -
fdhF+	18	6
fdhF-	410	233

fdoG vs group B FeFe hydrogenase		
	fefe-group-b +	fefe-group-b -
fdoG+	66	75
fdoG-	362	164

fdhA vs group C1 FeFe hydrogenase		
	fefe-group-c1 +	fefe-group-c1 -
fdhA+	3	29
fdhA-	2	633

fdhF vs group C1 FeFe hydrogenase		
	fefe-group-c1 +	fefe-group-c1 -
fdhF+	3	21
fdhF-	2	641

fdoG vs group C1 FeFe hydrogenase		
	fefe-group-c1 +	fefe-group-c1 -
fdoG+	5	136
fdoG-	0	526

fdhA vs group C2 FeFe hydrogenase		
	fefe-group-c2 +	fefe-group-c2 -
fdhA+	0	32
fdhA-	2	633

fdhF vs group C2 FeFe hydrogenase		
	fefe-group-c2 +	fefe-group-c2 -
fdhF+	0	24
fdhF-	2	641

fdoG vs group C2 FeFe hydrogenase		
	fefe-group-c2 +	fefe-group-c2 -
fdoG+	0	141
fdoG-	2	524

fdhA vs group C3 FeFe hydrogenase		
	fefe-group-c3 +	fefe-group-c3 -
fdhA+	13	19
fdhA-	53	582

*fdhA = K05299

fdhF vs group C3 FeFe hydrogenase		
	fefe-group-c3 +	fefe-group-c3 -
fdhF+	7	17
fdhF-	59	584

*fdhF = K22015

fdoG vs group C3 FeFe hydrogenase		
	fefe-group-c3 +	fefe-group-c3 -
fdoG+	60	81
fdoG-	6	520

*fdoG = K00123

b Odds ratio

**Red character: significantly different

fdhA vs group A13 FeFe hydrogenase	
Adjusted odds ratio	33.2
Upper confidence interval	73.6
Lower confidence interval	15.0

fdhF vs group A13 FeFe hydrogenase	
Adjusted odds ratio	32.8
Upper confidence interval	79.6
Lower confidence interval	13.5

fdoG vs group A13 FeFe hydrogenase	
Adjusted odds ratio	4.4
Upper confidence interval	8.2
Lower confidence interval	2.4

fdhA vs group A2 FeFe hydrogenase	
Adjusted odds ratio	0.4
Upper confidence interval	2.0
Lower confidence interval	0.1

fdhF vs group A2 FeFe hydrogenase	
Adjusted odds ratio	0.5
Upper confidence interval	2.8
Lower confidence interval	0.1

fdoG vs group A2 FeFe hydrogenase	
Adjusted odds ratio	0.5
Upper confidence interval	1.1
Lower confidence interval	0.3

fdhA vs group A4 FeFe hydrogenase	
Adjusted odds ratio	2376.2
Upper confidence interval	41654.9
Lower confidence interval	135.6

fdhF vs group A4 FeFe hydrogenase	
Adjusted odds ratio	909.8
Upper confidence interval	4335.7
Lower confidence interval	190.9

fdoG vs group A4 FeFe hydrogenase	
Adjusted odds ratio	187.9
Upper confidence interval	3123.4
Lower confidence interval	11.3

fdhA vs group B FeFe hydrogenase	
Adjusted odds ratio	1.0
Upper confidence interval	2.2
Lower confidence interval	0.5

fdhF vs group B FeFe hydrogenase	
Adjusted odds ratio	1.6
Upper confidence interval	4.0
Lower confidence interval	0.7

fdoG vs group B FeFe hydrogenase	
Adjusted odds ratio	0.4
Upper confidence interval	0.6
Lower confidence interval	0.3

fdhA vs group C1 FeFe hydrogenase	
Adjusted odds ratio	30.1
Upper confidence interval	158.8
Lower confidence interval	5.7

fdhF vs group C1 FeFe hydrogenase	
Adjusted odds ratio	41.8
Upper confidence interval	223.9
Lower confidence interval	7.8

fdoG vs group C1 FeFe hydrogenase	
Adjusted odds ratio	42.4
Upper confidence interval	772.0
Lower confidence interval	2.3

fdhA vs group C2 FeFe hydrogenase	
Adjusted odds ratio	3.9
Upper confidence interval	82.9
Lower confidence interval	0.2

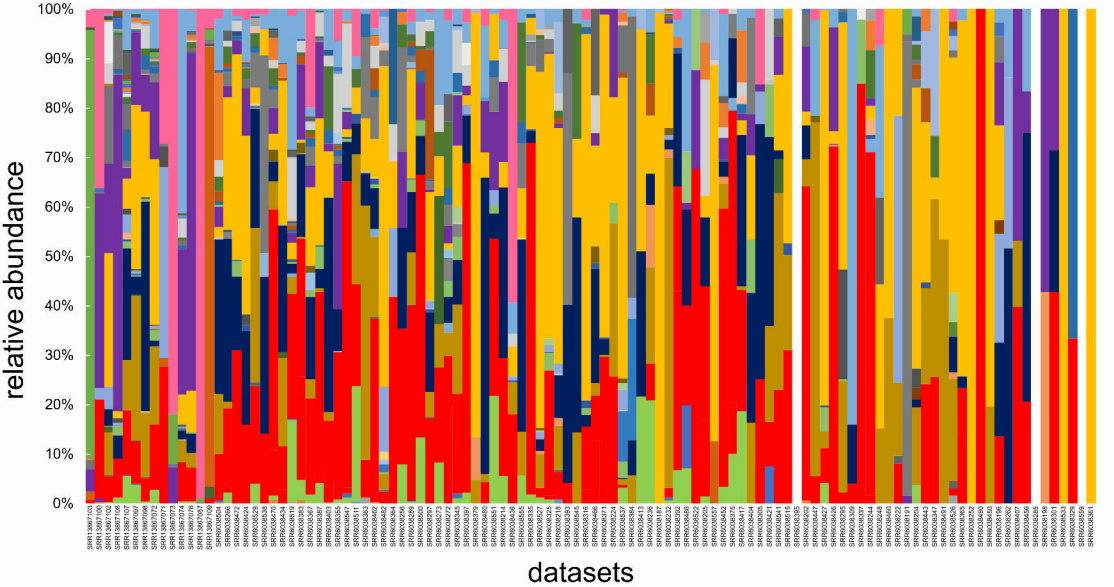
fdhF vs group C2 FeFe hydrogenase	
Adjusted odds ratio	5.2
Upper confidence interval	112.0
Lower confidence interval	0.2

fdoG vs group C2 FeFe hydrogenase	
Adjusted odds ratio	0.7
Upper confidence interval	15.5
Lower confidence interval	0.0

fdhA vs group C3 FeFe hydrogenase	
Adjusted odds ratio	7.5
Upper confidence interval	15.9
Lower confidence interval	3.6

fdhF vs group C3 FeFe hydrogenase	
Adjusted odds ratio	4.2
Upper confidence interval	10.3
Lower confidence interval	1.7

fdoG vs group C3 FeFe hydrogenase	
Adjusted odds ratio	59.4
Upper confidence interval	138.0
Lower confidence interval	25.6



- g__
- g__Agathobacter
- g__Akkermansia
- g__Anaerobutyricum
- g__BX12
- g__Bacteroides
- g__Blautia
- g__CAG-103
- g__CAG-1427
- g__CAG-170
- g__CAG-353
- g__CAG-56
- g__CAG-83
- g__CHH4-2
- g__Caccocola
- g__Choladocola
- g__Choladousia
- g__Cloacibacillus
- g__Clostridioides
- g__Clostridium
- g__Coprococcus
- g__Copromorpha
- g__Cryptoclostridium
- g__Dorea_B
- g__Dysosmobacter
- g__ER4
- g__Enterocloster
- g__Eubacterium
- g__F0422
- g__Faecalibacterium
- g__Faecalimonas
- g__Faecivicinus
- g__Faecousia
- g__Fusicatenibacter
- g__GCA-900066495
- g__Heteroscilispira
- g__Hungatella_A
- g__Intestinibacter
- g__JAGZHZ01
- g__Lactonifactor
- g__Limivivens
- g__Marvinbryantia
- g__Mediterraneibacter
- g__Megasphaera
- g__Merdiplasma
- g__Merdisoma
- g__Mitsuokella
- g__Muricomes
- g__Oliverpabstia
- g__Pelethocola
- g__Pelethomonas
- g__Peptacetobacter
- g__Pseudolachnospira
- g__Pseudoscilispira
- g__Robinsoniella
- g__Romboutsia
- g__Roseburia
- g__Ruminococcus_B
- g__Scatomonas
- g__Schaedlerella
- g__Terrisporobacter
- g__Tractidigestivibacter
- g__UBA11524
- g__UMGS1071
- g__UMGS1375
- g__Veillonella