MICROBIOLOGY SOCIETY

A comparative genome analysis of the *Bacillota* (*Firmicutes*) class *Dehalobacteriia*

Young C. Song[1], Sophie I. Holland[2]†, Matthew Lee[2], Gao Chen[3], Frank E. Löffler[3,4,5], Michael J. Manefield[2], Philip Hugenholtz[1],* and Ulrike Kappler[6],*

Abstract

Dehalobacterium formicoaceticum is recognized for its ability to anaerobically ferment dichloromethane (DCM), and a catabolic model has recently been proposed. *D. formicoaceticum* is currently the only axenic representative of its class, the *Dehalobacteriia*, according to the Genome Taxonomy Database. However, substantial additional diversity has been revealed in this lineage through culture-independent exploration of anoxic habitats. Here we performed a comparative analysis of 10 members of the *Dehalobacteriia*, representing three orders, and infer that anaerobic DCM degradation appears to be a recently acquired trait only present in some members of the order *Dehalobacteriales*. Inferred traits common to the class include the use of amino acids as carbon and energy sources for growth, energy generation via a remarkable range of putative electron-bifurcating protein complexes and the presence of S-layers. The ability of *D. formicoaceticum* to grow on serine without DCM was experimentally confirmed and a high abundance of the electron-bifurcating protein complexes and S-layer proteins was noted when this organism was grown on DCM. We suggest that members of the *Dehalobacteriia* are low-abundance fermentative scavengers in anoxic habitats.

DATA SUMMARY

The authors confirm all supporting data and protocols have been provided within the article or through Supplementary Data files. Raw peptide data have been deposited in the PRIDE database [1] under accession PXD038410.

INTRODUCTION

Metagenomics and, more recently, metagenome-assembled genomes (MAGs) are revealing the scope and functional potential of the microbial world [2–4]. This is true even for apparently well-studied lineages such as the phylum *Bacillota* (formerly *Firmicutes*). According to the Genome Taxonomy Database (GTDB), a rank-normalized genome-based taxonomy, the *Bacillota* comprises several class-level lineages that are exclusively or predominantly represented by MAGs [2]. A case in point is the class *Dehalobacteriia* that is currently represented by a single axenic culture, *Dehalobacterium formicoaceticum*, so called due to its unusual ability to anaerobically ferment dichloromethane (DCM) to formate and acetate. Note that *D. formicoaceticum* is

Impact Statement

Evolutionary and ecological context is important to understand model microorganisms. Here we show that anaerobic dichloromethane degradation, a key metabolic trait of *Dehalobacterium formicoaceticum*, is a relatively recent acquisition and not representative of the lineage to which this bacterium belongs. By contrast, the potential to use amino acids as carbon and energy sources is widespread in the class *Dehalobacteriia*, a trait we were able to experimentally confirm in *D. formicoaceticum*. We suggest that *Dehalobacteriia* are primarily low-abundance fermentative scavengers in anoxic ecosystems.

classified by LPSN [5] as a member of the family *Peptococcaceae*, which encompasses 25 validly published genera with cultured representatives (https://lpsn.dsmz.de/family/peptococcaceae). However, these genera are phylogenetically diverse and belong to multiple classes according to the rank-normalized GTDB taxonomy [2].

DCM is a chlorinated C1 compound that is widely used as an industrial solvent, resulting in substantial groundwater contamination [6, 7]. It is also produced naturally in environments such as oceans, wetlands and soils and contributes to climate warming [6, 7]. While the phylogenetic distribution and mechanism of aerobic DCM oxidation are well established [8, 9], the same cannot be said for anaerobic DCM metabolism. To date, three genera of anaerobic bacteria are known to degrade DCM: *D. formicoaceticum* [10], '*Ca.* Forminomas warabiya' [11, 12] and '*Ca.* Dichloromethanomonas elyunquensis' [13]. Both *D. formicoaceticum* and '*Ca.* F. warabiya' are members of the *Dehalobacteriia*, while '*Ca.* D. elyunquensis' belongs to a neighbouring *Bacillota* class, the *Desulfitobacteriia*, as defined by GTDB. Recently, a model has been proposed for anaerobic DCM degradation that is based on proteomic and genomic analyses. In this model, the methylene chloride catabolism (Mec) proteins, which include several corrinoid-dependent methyltransferases, convert DCM into Wood–Ljungdahl pathway (WLP) intermediates, such as methylene- or methyl-tetrahydrofolate (THF) [14]. The final metabolic products of anaerobic DCM degradation include formate, acetate and inorganic chloride, as previously described [5–7, 10–14].

It is not known whether anaerobic DCM degradation is common to all members of the *Dehalobacteriia*, and the broader metabolic properties of this class have not been investigated. Here, we performed a phylogenetic and metabolic reconstruction of publicly available *Dehalobacteriia* genomes representing three orders within this class. Anaerobic DCM degradation may be a recently acquired trait only found in some members of the order *Dehalobacteriales*. By contrast, all members of the *Dehalobacteriia* appear to rely on a combination of a glycine cleavage system and the WLP for carbon fixation and energy generation. A number of potential electron-bifurcating complexes were highly conserved and highly expressed in *D. formicoaceticum* growing on DCM, suggesting their metabolic importance. While *D. formicoaceticum* is known as an obligate DCM degrader, we demonstrated its ability to grow on serine in the absence of DCM, an ability that appears to be conserved within the class. The distribution of *Dehalobacteriia* 16S rRNA genes in the environment suggests that members of this lineage are found in anoxic habitats in low abundance, which combined with the metabolic information suggest that *Dehalobacteriia* are fermentative scavengers.

METHODS

Data retrieval

According to GTDB release 06-RS202 (April 2021), there are 17 species representative genomes in the class *Dehalobacteriia*, with an estimated average completeness of 81.8% and contamination of 0.5% (Table S1, available in the online version of this article). For the purpose of this study, nine genomes with an estimated completeness greater than or equal to the average were selected for metabolic inference (Table S1). Despite having an estimated completeness lower than the average (i.e. 78.5%), UBA4997 was also added to this study, as it is one of four genomes in *Dehalobacteriia* that possess 16S small-subunit rRNA (16S SSU rRNA), which was used for relative abundance analysis as described below. Out of these ten sequences, the genomes DUPU01 and RUG14212 were introduced into *Dehalobacteriia* in GTDB 06-RS202, while the remaining eight genomes, including those of the isolate *D. formicoaceticum* [10, 14] and mixed-culture member '*Ca.* F. warabiya' [11, 12], were introduced to this class in the previous GTDB release, 04-RS89 (June 2019). For the purpose of metabolic inference, Prokka [15] was used. Detailed steps of these bioinformatics analyses are described in subsequent sections.

Relative abundance analysis

A set of 11 near full-length 16S SSU rRNA gene sequences obtained from four *Dehalobacteriia* genomes (*D. formicoaceticum*, '*Ca.* F. warabiya', UBA4997 and UBA5757) were checked against a set of commonly used broadly conserved SSU rRNA primers [16] to determine the likelihood of primer bias in rRNA-based environmental surveys. With the exception of 803F, all primers should be able to PCR-amplify *Dehalobacteriia* SSU rRNA genes (Table S2). Using nucleotide BLAST (BLASTN) [17], these sequences were then searched against the database downloaded from archives of the SILVA SSU rRNA gene database (release 138.1, 510508 total sequences [18]) to conservatively (≥ 95% identity) identify 16S rRNA gene sequences belonging to the *Dehalobacteriia*

(currently not annotated in SILVA). An E-value of 1e-3, maximum target hits of 10 and tabular output format (i.e. *-evalue 1e-3 -outfmt 6 -max_target_seqs 10*) were used as parameters for this search operation. Full 16S rRNA gene amplicon datasets associated with the identified *Dehalobacteriia* sequences (Table S3) were obtained using their publication information where available. The sequences that were not published were excluded from subsequent analysis. The relative abundance of *Dehalobacteriia* sequences in each dataset was then determined as the percentage of total sequences in the study.

Metabolic inference of *Dehalobacteriia*

Dehalobacteriia MAG or isolate genomes were annotated using Prokka with *--kingdom Bacteria* and *--addgenes* parameters, where the latter is used to add gene features (e.g. gene names, enzyme commission ID and protein name) for each detected coding region. A list of metabolic pathways inferred to be present in the *Dehalobacteriia* genomes was generated by importing the GenBank (GBK) files generated from Prokka to Pathway Tools software [19]. In addition, sequences of detected proteins were also searched against KO (KEGG Orthology [20]) using BlastKOALA (KEGG Orthology and Links Annotation [21]) to generate a list of KO identifications for each *Dehalobacteriia* genome (Table S4). To determine the presence of membrane-bound redox complexes and cytoplasmic hydrogenases, a subset of proteins identified by Prokka were further searched against the NCBI non-redundant (nr) database using protein blast (BLASTP). The conserved domain database (CDD [22]) associated with the top hit of each query was used to confirm the identification of protein function assignment for *Dehalobacteriia* by comparison with biochemically characterized proteins, such as *Rhodobacter* nitrogen fixation (Rnf) complex [23], *Sporomusa* type ferredoxin-dependent transhydrogenase (Stn [24]) and electron-bifurcating hydrogenase (Hnd [25]). Conservation of the hydrogenase proteins detected from the *Dehalobacteriia* genomes was visualized as a gene-neighbourhood diagram using Clinker [26].

The protein sequences detected from the *Dehalobacteriia* genomes using Prokka were also searched against a customized protein database of the *mec* gene cassette [14] to predict the functional potential of the genomes in relation to anaerobic DCM fermentation. A BLASTP search with the following parameters was used for this operation: an e-value of 1e-3, maximum target hits of 10 and tabular output format.

Phylogenetic inference

Dehalobacteriia MAGs and isolate genomes, along with species-representative genomes in the neighbouring *Bacillota* class-level lineages, as defined in GTDB, were used for phylogenetic inference. The phylogenetic tree was built using IQ-Tree [27] with protein mixture model C60 and 1000 bootstrap cycles as parameter for the maximumum-likelihood (ML) inference.

The GenBank files generated for each of the *Dehalobacteriia* genomes using Prokka were used as input files for Clinker [26] to generate two gene-neighbourhood diagrams, one depicting the conservation of the redox complexes detected in *Dehalobacteriia* and the other showing the conservation of the *mec* cassette present in *D. formicoaceticum*, '*Ca.* F. warabiya' and '*Ca.* Dichloromethanomonas elyunquensis' [14]. A minimum amino acid identity of 30% was used to generate the Clinker plot [26].

Analytical assays

Bacterial cultures and growth conditions

D. formicoaceticum (ATCC 700118) was cultivated at 30 °C in a defined, anoxic, bicarbonate-buffered mineral salt medium containing DCM as the sole energy source, as described by Chen and colleagues [28]. Medium preparation followed the methodology described by Holland and colleagues [12] with modification. In brief, 1 litre of medium contained $CaCl_2.H_2O$ (0.1 g), $MgCl_2.6H_2O$ (0.1 g), NH_4Cl (1.5 g), NaH_2PO_4 (0.6 g), KCl (0.1 g) and 1 ml each of trace element solutions A and B (Tables S5 and S6) as described previously [29]. The medium was purged with N_2 gas for 1 h followed by addition of 2.5 g $NaHCO_3$ and sparging with N_2/CO_2 (80:20) to achieve a pH between 6.8 and 7.0. The pH-adjusted medium (50 ml) was then dispensed into nitrogen-sparged 60 ml serum bottles. The serum bottles were immediately sealed with a Teflon-coated rubber cap and metal crimp, sparged with N_2/CO_2 (80:20) for another 5 min and autoclaved for 20 min at 121 °C.

Before use, the following reagents were aseptically added to each bottle: 500 µl of 5 g l^{-1} yeast extract solution (50 mg l^{-1} final concentration) and 50 µl of 1000× vitamin solution (Table S7 [30]). DCM and/or serine were added to the standard medium as carbon sources with final concentrations of 1 and 10 mM, respectively. Serum bottles were inoculated with 2% (v/v) *D. formicoaceticum* and incubated statically in the dark at 30 °C for up to 31 days [10]. Cultures were resupplied with DCM and/or serine once, to the same concentration, when the initial substrate pulse was depleted.

Determination of DCM concentrations by GC

DCM was quantified on a Shimadzu GC-2010 gas chromatograph equipped with a flame ionzation detector (GC-FID) and a GS-Q column (30 m×0.32 mm; Agilent Technologies). Headspace samples (100 µl) were withdrawn directly from culture flasks using a lockable, gas-tight syringe and manually injected. The oven was initially at 150 °C, then raised by 30 °C min^{-1} to 250 °C. The inlet temperature was 250 °C, split ratio 1:10 and FID temperature 250 °C. The carrier gas was helium (3 ml min^{-1}). A minimum

three-point calibration curve was used. DCM concentrations are reported as the nominal concentration in each serum bottle, calculated from the headspace concentration using Henry's Law dimensionless solubility constant (Hcc=0.107 at 30 °C), as per the OSWER method [31].

Determination of volatile fatty acid concentrations by GC

The volatile fatty acids (VFAs) acetate, formate, butyrate and propionate were quantified as described by Holland *et al.* [12]. VFAs were derivatized to their ethyl esters by combining 500 µl of liquid culture sample, 200 µl of ethanol (100%), and 200 µl of 1 M H_2SO_4 in 10 ml glass vials, which was sealed with Teflon/butyl septa and screw caps, and then incubated at 60 °C for 45 min. Ethyl esters were separated using a Shimadzu GC-2010 GC-FID fitted with a DB-FFAP column (30 m×0.320 mm; Agilent Technologies) and a PAL LHS2-xt-Shim autosampler (Shimadzu). The samples were agitated at 80 °C for 5 min, before 250 µl of sample headspace was injected. The oven was held at a constant temperature of 40 °C for 6 min, the carrier gas was helium (2 ml min^{-1}) and other parameters were as described above for DCM quantification. A six-point calibration curve was used for acetate, butyrate, propionate and formate (0, 0.5, 1, 5, 10 and 15 mM).

Determination of serine concentrations by GC–triple quadrupole MS (GC-TQMS)

Determination of serine in culture samples used the method of Villas-Boas and colleagues [32] with modifications. In a 1.5 ml glass vial, 50 µl of culture sample, 250 µl NaOH (1%), 250 µl ethanol (100%), 5 µl alanine (10 mM; internal standard) and 50 µl pyridine were combined before 20 µl ethyl chloroformate (Ecf) was added to the mixture, followed by shaking for 20 s. A second 20 µl aliquot of Ecf was added and the vial was shaken again for 20 s. DCM (1 ml) was added, and the vial was shaken. After phase separation the upper, aqueous phase was removed, and a small amount of anhydrous Na_2SO_4 was added to remove water. The solvent phase was transferred to a fresh vial for analysis on an Agilent 7890A gas chromatograph system fitted with a 7000A triple quadrupole mass spectrometer (GC-TQMS). The TQMS was operated in the multiple reaction monitoring (MRM) mode. Precursor–product ion transitions (*m/z*) and the collision energy (eV) were 116.2→72.1 (*m/z*), 10 (eV) for alanine and 114.0→86.0 (*m/z*), 10 (eV) for serine [32].

Preparation of protein extracts for proteomic analyses

Cultures were harvested after a 31 day incubation, when approximately 80% of the second substrate pulse had been depleted. The culture suspension (~40 ml) was transferred to an ice-cold plastic Falcon tube and centrifuged at 3270 relative centrifugal force (rcf) for 45 min at 4 °C. The supernatant was removed, leaving 1 ml to suspend the cell pellet. The concentrated cell suspension was transferred to 1.5 ml microfuge tubes and centrifuged at 14000 rcf for 10 min at 4 °C. After removing the supernatant, the pellets were suspended in 100 µl of SDS-MOPS protein extraction buffer [50 mM MOPS, pH 7.9; 150 mM NaCl; 100 µM EDTA; 100 µM $MgCl_2$; 4% (w/v) SDS]. The mixture was transferred to a 2 ml bead-beating tube that contained 0.06 g glass beads (0.5 mm) and a 6.35 mm ceramic sphere. The samples were processed for 5 min at 1800 r.p.m. in a PowerLyser (Qiagen), and followed by centrifugation at 16000 rcf for 10 min at 4 °C. Supernatants were transferred to clean 1.5 ml tubes and again centrifuged at 14000 rcf for 10 min at 4 °C. This supernatant – the final crude extract – was transferred to a 1.5 ml tube and stored at −20 °C until further analysis. Protein quantification was performed using the MicroBCA (ThermoFisherTM) assay as per the manufacturer's instructions. Typical protein concentrations of cell extracts were between 0.6 and 1.7 µg ml^{-1}. Tryptic peptide samples were prepared using a modified version of a filter-assisted sample preparation (FASP) protocol [33]. Detailed steps of the FASP protocol are available in the Supplementary Information.

Liquid chromatography

Peptides were separated by nanoLC using an UltiMate nanoRSLC UPLC and autosampler system (Dionex). Samples (2.5 µl) were concentrated and desalted on a micro C18 precolumn (300 µm×5 mm; Dionex) with H_2O/CH_3CN (98:2, 0.2% trifluoroacetic acid) at 15 µl min^{-1}. After a 4 min wash, the pre-column was switched (Valco 10 port UPLC valve; Valco) into line with a fritless nano column (75 µm×~15 cm) containing C18AQ media (1.9µ, 120 Å; Dr Maisch). Peptides were eluted using a linear gradient of H_2O/CH_3CN (98:2, 0.1% formic acid) to H_2O/CH_3CN (64:36, 0.1% formic acid) at 200 nl min^{-1} over 30 min. High voltage (2000 V) was applied to a low-volume titanium union (Valco) and the tip was positioned ~0.5 cm from the heated capillary (*T*=275 °C) of an Orbitrap Fusion Lumos (Thermo Electron Bremen, Germany) mass spectrometer. Positive ions were generated by electrospray and the Fusion Lumos was operated in data-dependent acquisition mode (DDA).

A survey scan of *m/z* 350–1750 was acquired (resolution=120 000 at *m/z* 200, with an accumulation target value of 400000 ions) and lockmass-enabled (*m/z* 445.12003). Data-dependent tandem MS analysis was performed using a top-speed approach (cycle time of 2 s). MS2 spectra were fragmented by high-energy collision dissociation (HCD; NCE=30) activation mode and the ion-trap was selected as the mass analyser. The intensity threshold for fragmentation was set to 25 000 units. A dynamic exclusion of 20 s was applied with a mass tolerance of 10 p.p.m.
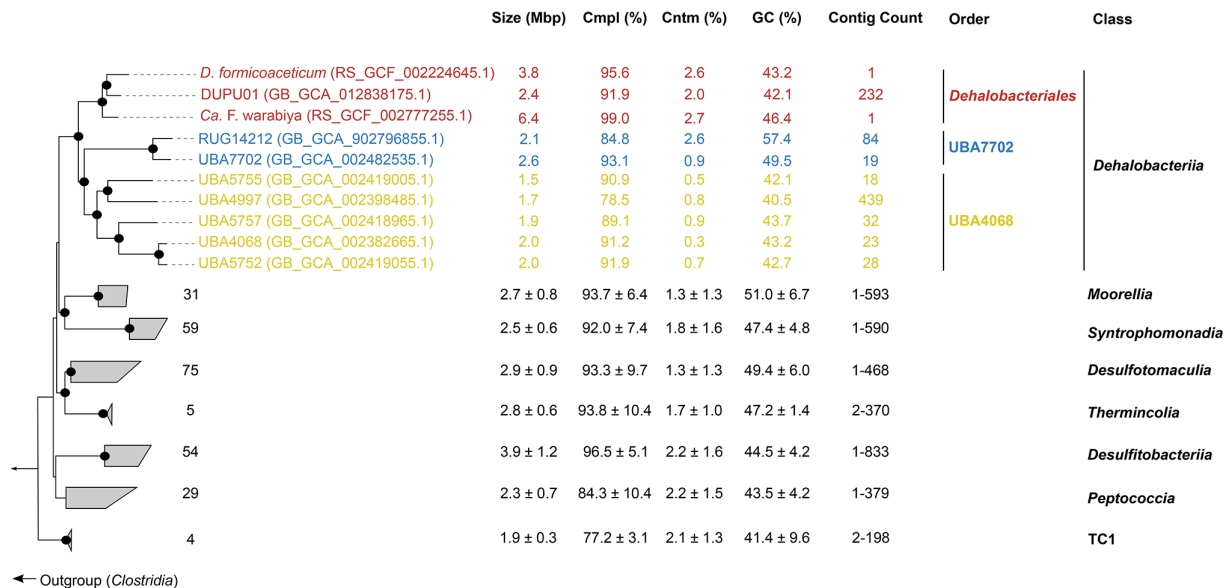
	Size (Mbp)	Cmpl (%)	Cntm (%)	GC (%)	Contig Count	Order	Class
D. formicoaceticum (RS_GCF_002224645.1)	3.8	95.6	2.6	43.2	1	*Dehalobacteriales*	
DUPU01 (GB_GCA_012838175.1)	2.4	91.9	2.0	42.1	232		
Ca. F. warabiya (RS_GCF_002777255.1)	6.4	99.0	2.7	46.4	1		
RUG14212 (GB_GCA_902796855.1)	2.1	84.8	2.6	57.4	84	**UBA7702**	
UBA7702 (GB_GCA_002482535.1)	2.6	93.1	0.9	49.5	19		*Dehalobacteriia*
UBA5755 (GB_GCA_002419005.1)	1.5	90.9	0.5	42.1	18		
UBA4997 (GB_GCA_002398485.1)	1.7	78.5	0.8	40.5	439		
UBA5757 (GB_GCA_002418965.1)	1.9	89.1	0.9	43.7	32	**UBA4068**	
UBA4068 (GB_GCA_002382665.1)	2.0	91.2	0.3	43.2	23		
UBA5752 (GB_GCA_002419055.1)	2.0	91.9	0.7	42.7	28		
31	2.7 ± 0.8	93.7 ± 6.4	1.3 ± 1.3	51.0 ± 6.7	1-593		*Moorellia*
59	2.5 ± 0.6	92.0 ± 7.4	1.8 ± 1.6	47.4 ± 4.8	1-590		*Syntrophomonadia*
75	2.9 ± 0.9	93.3 ± 9.7	1.3 ± 1.3	49.4 ± 6.0	1-468		*Desulfotomaculia*
5	2.8 ± 0.6	93.8 ± 10.4	1.7 ± 1.0	47.2 ± 1.4	2-370		*Thermincolia*
54	3.9 ± 1.2	96.5 ± 5.1	2.2 ± 1.6	44.5 ± 4.2	1-833		*Desulfitobacteriia*
29	2.3 ± 0.7	84.3 ± 10.4	2.2 ± 1.5	43.5 ± 4.2	1-379		*Peptococcia*
4	1.9 ± 0.3	77.2 ± 3.1	2.1 ± 1.3	41.4 ± 9.6	2-198		**TC1**

← Outgroup (*Clostridia*)

Fig. 1. Phylogenetic relationships between members of the *Dehalobacteriia* and neighbouring *Bacillota* classes as defined by the GTDB. The maximum-likelihood (ML) tree was inferred based on alignment of 120 concatenated marker genes downloaded from GTDB R06-R202. IQ-Tree with protein mixture model C60 and 1000 bootstrap cycles were used as parameters for ML inference. Bootstrap values indicated by black dots on each node of the tree are all ≥90 %. Also shown are the genome size (Mb), estimated genome completion (%), estimated genome contamination (%), GC content (%) and contig count in respective order. The number of representatives in each neighbouring class is shown beside each grey wedge.

Protein identification

The peptide sequence data obtained from the three replicates of the *D. formicoaceticum* sample grown in DCM-containing medium were mapped against the ORFs of the *D. formicoaceticum* genome available in GTDB release R06-202, using ProteinPilot software (version 5.0.2; AB Sciex, 2018). The ProteinPilot results were filtered using a custom R script (R version 3.4.3 [34]) to generate a list of proteins which were identified in at least two replicates of each test set and had an average 'unused' peptide score ≥2 (i.e. 99% confidence of identification).

RESULTS AND DISCUSSION

Environmental distribution of *Dehalobacteriia*

As of GTDB release 06-RS202, three orders are recognized within the class *Dehalobacteriia*: *Dehalobacteriales*, UBA7702 and UBA4068 (Fig. 1). The only characterized members of this class belong to the order *Dehalobacteriales*: *D. formicoaceticum* and '*Ca.* F. warabiya' (Fig. 1; Table S1). Analysis of metadata associated with *Dehalobacteriia* genomes suggests that they are limited to anoxic habitats consistent with the strictly anaerobic lifestyle of the characterized representatives of this class and its neighbouring GTDB-defined classes *Desulfitobacteriia*, *Moorellia*, *Peptococcia*, *Syntrophomonadia* and *Desulfotomaculia* [35–38]. Documented habitats include an Alberta oil sands tailing pond [39], the bovine rumen [40], anaerobic digester sludge [41], an alcohol fermentation pit [42] and a methanol-fed electric cell anode biofilm [43]. To further evaluate the ecological distribution of this group, we explored 16S rRNA gene amplicon data present in the SILVA database. However, SILVA does not currently include *Dehalobacteriia* in its classification scheme. Therefore, we identified 16S rRNA gene sequences putatively belonging to this class by similarity searches using all available 16S rRNA gene sequences derived from genomes classified as *Dehalobacteriia* in GTDB (06-RS202). Only 46 of the 510508 reference sequences in SILVA (v138) had ≥95% similarity to GTDB *Dehalobacteriia* 16S rRNA sequences, all of which were associated with anoxic or low-oxygen habitats, including the murine caecum [44–46], the human digestive tract [47], the insect gut [48], a hydrocarbon-contaminated aquifer [49] and a mud volcano [50]. The predominant inferred metabolic activity of the microbial communities occurring in these habitats is fermentation of metabolites such as VFAs (e.g. propionate, butyrate) and amino acids producing acetate, ethanol, formate, CO_2 or H_2 as end products [51, 52]. In each of these environmental surveys, the *Dehalobacteriia* sequences typically comprised «1% of the total 16S rRNA gene sequences obtained (Table S3), suggesting that this class is not only restricted to anoxic habitats, but also has relatively low environmental abundance.

Anaerobic dichloromethane degradation

Current knowledge of *Dehalobacteriia* metabolism comes from the only pure culture representative, *D. formicoaceticum,* and a highly enriched culture dominated by '*Ca.* F. warabiya'. Both bacteria are notable for their ability to anaerobically degrade DCM.
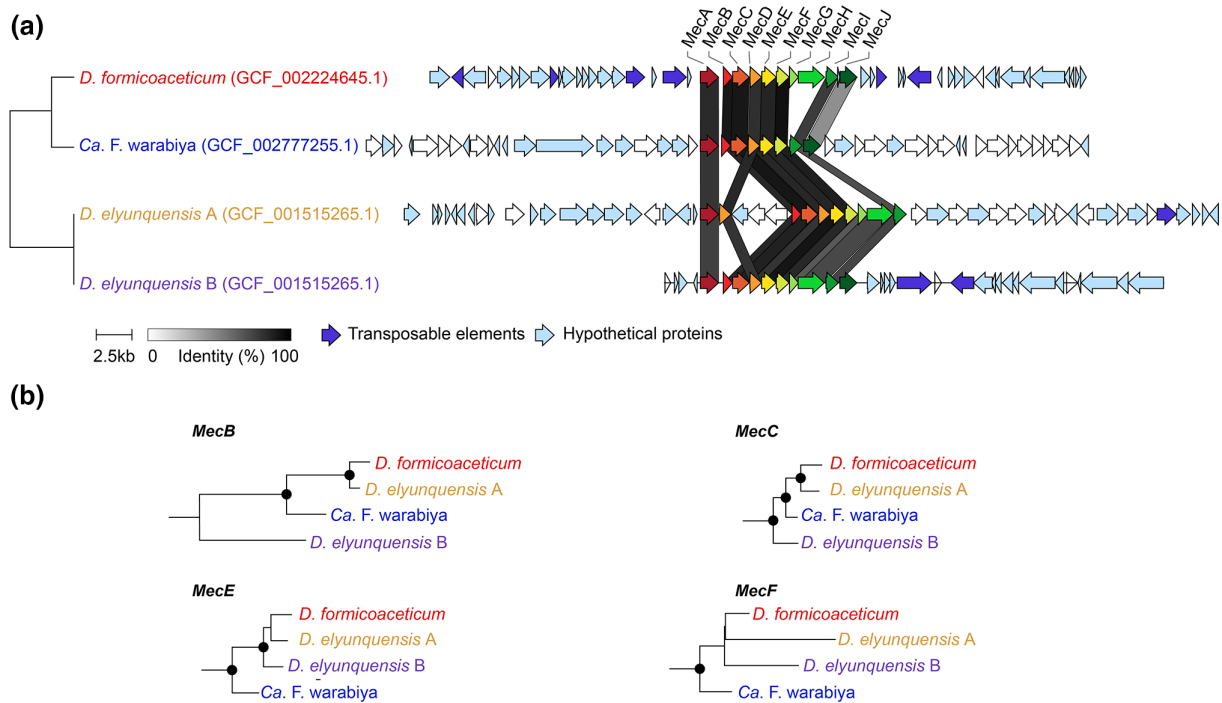
Fig. 2. (a) Gene neighbourhoods of the methylene chloride catabolism (*mec*) cassette and (b) phylogenetic relationships of selected *mec* genes in the *Dehalobacteriia* (*Dehalobacterium formicoaceticum* and '*Ca.* Formimonas warabiya') and *Desulfitobacteriia* ('*Ca.* Dichloromethanomonas elyunquensis'). Filled circles indicate >90% bootstrap support.

Recently, a pathway for anaerobic DCM degradation has been proposed based on comparative genomics and shotgun proteomics that identified a cassette comprising 8–10 genes called the methylene chloride catabolism (*mec*) cassette [14]. This cassette is present in two of the three members of the order *Dehalobacteriales* (*D. formicoaceticum* and '*Ca.* F warabiya'; Fig. 2), and in '*Ca.* Dichloromethanomonas elyunquensis', a member of the neighbouring GTDB-defined class *Desulfitobacteriia*. None of the *Dehalobacteriia* MAGs included in this study were found to harbour close homologues of the *mec* cassette (>30% amino acid identity), suggesting that anaerobic DCM degradation is not ancestral to the *Dehalobacteriia*. Twenty-nine additional genomes classified as *Dehalobacteriia* in the latest GTDB release (08-RS214), not included in the present analysis, also lack close homologues of the *mec* cassette (data not shown). Phylogenetic and gene neighbourhood analysis of the *mec* cassette in *D. formicoaceticum*, '*Ca.* F. warabiya' and '*Ca.* D. elyunquensis' support this inference as there is no conservation of flanking regions and individual *mec* genes have different evolutionary histories (Fig. 2b) suggesting that the *mec* cassettes were present in the *Dehalobacteriia* no earlier than the common ancestor of the order *Dehalobacteriales* and possibly acquired later than that. Furthermore, the flanking regions include transposable elements and many small hypothetical genes, suggesting that the *mec* cassettes are located in mobile accessory elements (Fig. 2a). The absence of an *mec* cassette in DUPU01, the third member of the order *Dehalobacteriales*, may therefore be correct although it is also possible that it was missed as the genome is only in draft (232 contigs, 92% estimated completeness; Fig. 1). Anaerobic DCM degradation via the *mec* cassette is predicted to produce methylene-THF, which is an intermediate of the WLP and connects DCM degradation directly to a dominant carbon turnover pathway in these bacteria.

C1 metabolism in *Dehalobacteriia* mostly relies on an incomplete WLP and the glycine synthase complex

In contrast to the *mec* genes involved in DCM degradation, elements of the WLP were commonly found in the *Dehalobacteriia* (Fig. 3a). The complete WLP enables fixation of CO_2 and production of acetyl-CoA. With the exception of formate dehydrogenase (FdhF), the methyl-branch of the WLP is conserved in all analysed *Dehalobacteriia*. The complete pathway is present in the orders *Dehalobacteriales* and UBA7702 (Fig. 3a), where the carbonyl branch of the WLP comprises a bifunctional carbon monoxide dehydrogenase/acetyl-CoA synthase (AcsAB) that is related to similar enzyme complexes previously described in *Moorella* and *Clostridium* species [53]. The absence of the *acsAB* genes in the order UBA4068 is unusual, as the WLP has been proposed as the major pathway for CO_2 fixation in related *Bacillota*. However, an absence of several key genes of the WLP has been previously described for the organohalide-respiring bacterial species *Dehalococcoides mccartyi*, a member of the phylum *Chloroflexota* [54]. In *Dehalococcoides mccartyi*, *acsA*, *metF* and *fdhF* were absent, and as a result, no CO_2 fixation occurred. Instead, external formate was assimilated into glycine and methionine via the reactions catalysed by glycine or methionine synthase [54].
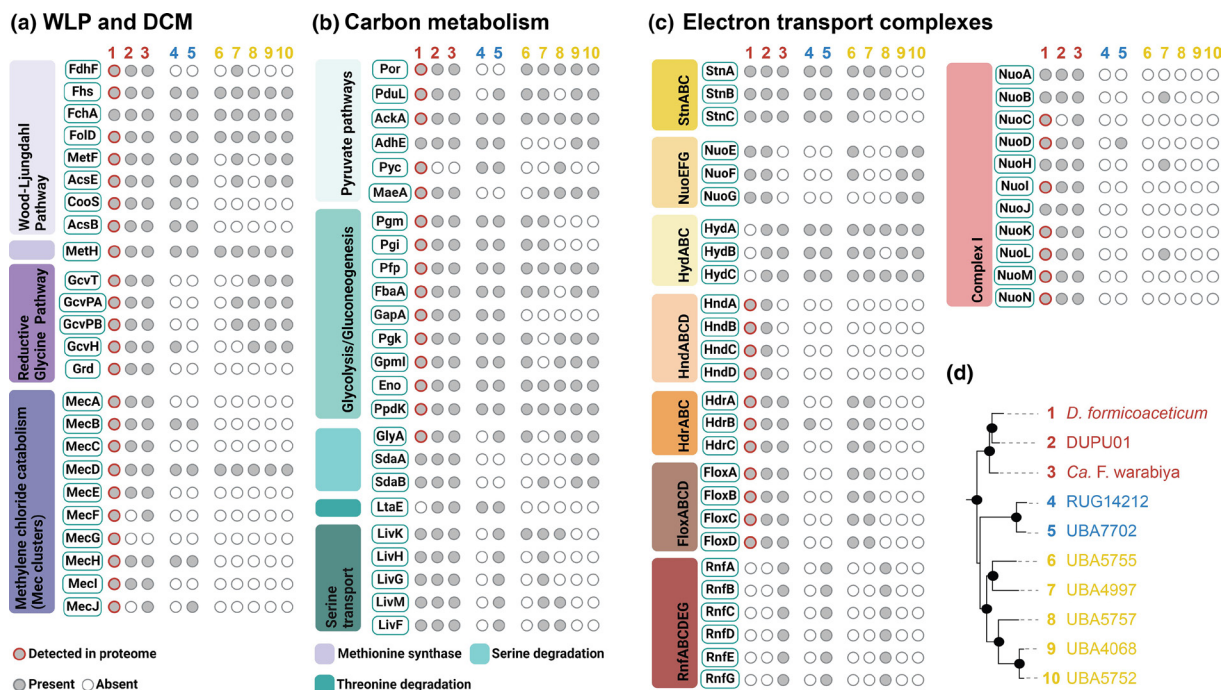
Fig. 3. Distribution of key inferred metabolic traits in the *Dehalobacteriia*. (a) Carbon fixation and DCM degradation-related functions, (b) central carbon metabolism, (c) potential electron bifurcating complexes and (d) legend showing phylogenetic relationships of *Dehalobacteriia* genomes colour-coded by order (red, *Dehalobacteriales*; blue, UBA7702; yellow, UBA4068). The figure was generated using BioRender (https://biorender.com).

Unlike *Dehalococcoides mccartyi*, *fdhF* and *metF* genes are present in most *Dehalobacteriia*, but homologues of *acsAB* are absent in the order UBA4068. Based on the genes that are present, we propose that members of UBA4068 fix CO_2 via the reversible reaction of the FdhF [55], with 5,10-methylene tetrahydrofolate being used as a substrate for the glycine cleavage system (GCS). The four enzymes of the GCS, i.e. the aminomethyltransferase GcvT (EC 2.1.2.10; K00605), glycine cleavage protein H (GcvH; K02437), and glycine dehydrogenase subunits 1 and 2 (GcvPA and GcvPB; EC 1.4.4.2; K00282 and K00283), catalyse the formation of glycine that can subsequently be converted to serine and then pyruvate via the reactions of GlyA and SdaAB that are conserved in several *Dehalobacteriia* (Fig. 3a). In *D. formicoaceticum* cultures, ^{13}C-glycine was detected as a metabolite during growth with ^{13}C-DCM, which supported the presence and function of the GCS [56]. Pyruvate is a central metabolite that can then be used in a variety of metabolic pathways to synthesize new cellular components, but it can also support substrate-level phosphorylation. In this case, pyruvate would be converted to acetyl-CoA and acetyl-phosphate via the reactions catalysed by pyruvate:ferredoxin oxidoreductase (Por; EC 1.2.7.1; K03737) and PduL (EC 2.3.1.222; K015024). Acetyl-phosphate can then be converted to acetate via acetate kinase (AckA), leading to the production of ATP. In the *Dehalobacteriales*, a glycine reductase (Grd, EC 1.21.4.2; K10670-72; K21576-77) is also present that allows direct production of acetyl-phosphate from glycine (Fig. 3a). C1 metabolism in the *Dehalobacteriales* is thus similar to what has been proposed for *Clostridium drakei* [57], while the other two orders of the *Dehalobacteriia* have a more limited carbon fixation repertoire, where carbon assimilation requires formation of serine as an intermediate.

Interestingly, with the exception of the order UBA7702, all examined *Dehalobacteriia* genomes contain operons that encode homologues of the so-called 'aerobic', molybdenum-containing CoxLMS carbon monoxide dehydrogenase that converts CO to CO_2, but is not known to catalyse the reverse reaction (Fig. 4 [57]). In bacteria with a WLP, CO is formed as a by-product of the acetyl-synthase reaction and has been shown to cause growth inhibition in *Dehalococcoides mccartyi*, which only contains a partial WLP [54]. We propose that in the *Dehalobacteriia*, the reaction of 'aerobic' CODH could prevent endogenous CO poisoning.

Pyruvate metabolism

As set out above, pyruvate is a key intermediate in all members of the *Dehalobacteriia*. Once formed, pyruvate can be used by *Dehalobacteriia* for gluconeogenesis (Figs 3b and 4) and to generate intermediates for the pentose phosphate pathway and TCA cycle (e.g. oxaloacetate and malate). These intermediates probably play a role in the biosynthesis of amino acids, as the TCA cycle is incomplete in all members of the *Dehalobacteriia* (Fig. S1). Pyruvate also enables the formation of various fermentative end products. All members of the *Dehalobacteriia* should be able to produce acetate via the reaction of Por described above, while
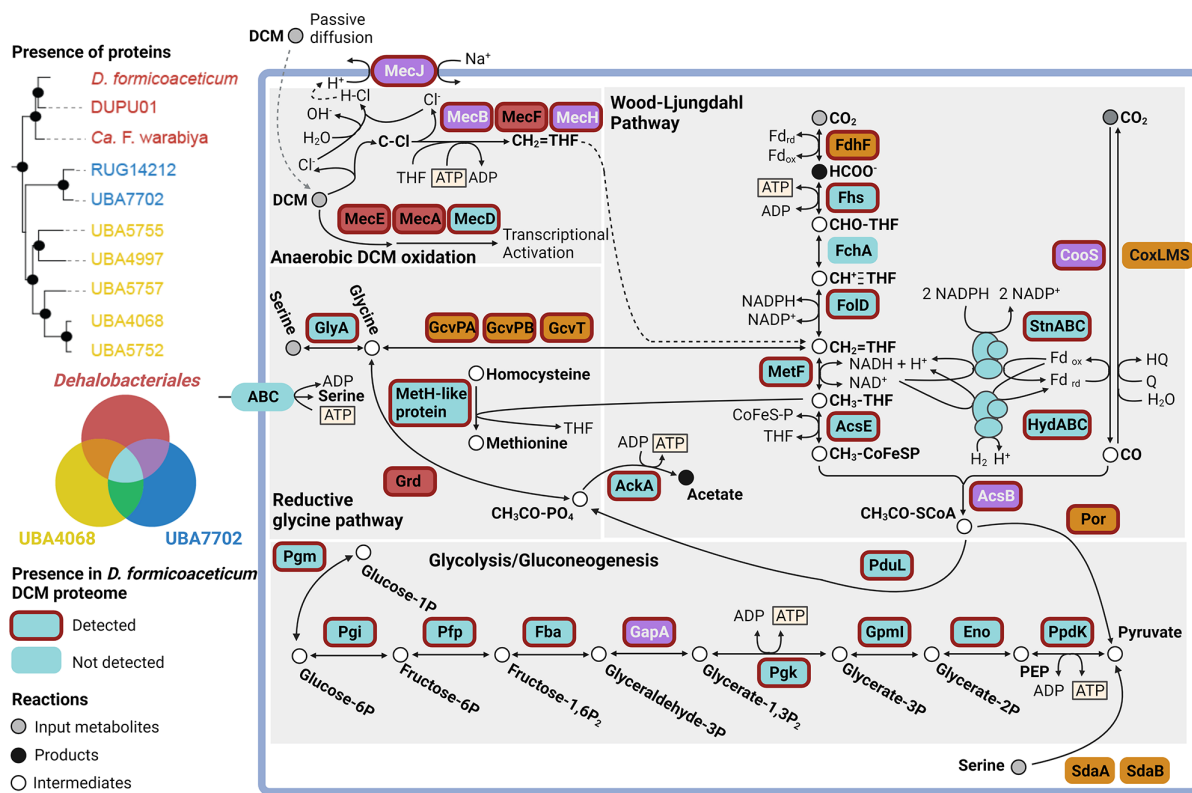
Fig. 4. Overview of key metabolic features detected in members of the *Dehalobacteriia*. Protein names are shown in coloured boxes that indicate the orders of the *Dehalobabteriia* in which the proteins are detected (see Venn diagram on the left-hand side). Proteins detected in the *D. formicoaceticum* proteome are indicated by a red outline.

ethanol production is possible via the acetaldehyde alcohol dehydrogenase (AdhE; EC 1.2.1.10; K04072) that is found in members of all three orders of the *Dehalobacteriia*.

Energy conservation via membrane-bound redox complexes

Only members of the order *Dehalobacteriales* have the potential for energy conservation via standard oxidative phosphorylation as they encode the ubiquitous, membrane-bound ATP synthase and 11 Nuo subunits necessary to form a functional NADH dehydrogenase/complex I [58] (Fig. 5). However, all *Dehalobacteriia* genomes encode putative electron-bifurcating protein complexes known to mediate redox processes in related bacteria [24, 25].

When investigating the presence of Complex I homologues, we discovered genes encoding NuoE- and NuoF-like proteins that are homologues of the HydA/HndA and HydB/HndC proteins that occur in electron-bifurcating hydrogenases [24]. Further investigation showed that at least one putative electron-bifurcating hydrogenase is found in all *Dehalobacteriia* genomes. A hydrogenase resembling HydABC from *Thermotoga maritima* [59] is most commonly found in members of the group, although in UBA5757, the gene cluster appears to be incomplete and therefore potentially non-functional (Fig. 5). In *D. formicoaceticum*, the HydABC complex appears to have been replaced by an Hnd-like hydrogenase complex (HndABCD; EC 1.12.1.3; K17992, K18330-2) that has been characterized as catalysing the reduction of NAD^+ and ferredoxin in *Desulfovibrio fructosovorans* [24, 25]. Interestingly, the DUPU01 genome contains copies of both gene clusters, suggesting functional redundancy (Fig. 5). The presence of conserved HydABC/HndABCD gene clusters in *Dehalobacteriia* indicates that this lineage may use hydrogenase-driven reduction of NAD^+ and ferredoxin to supply reducing equivalents for CO_2 fixation via the WLP or conversion of acetyl-CoA to pyruvate catalysed by Por.

In addition to the two hydrogenase-like complexes, we identified a gene cluster that appears to encode an StnABC transhydrogenase-like complex [24] and is present in seven of the ten analysed *Dehalobacteriia* genomes, only being absent in some members of the order UBA4068 (Figs 3c and 4). In this complex, the NuoE- and NuoF-like subunits occur together with a third subunit that contains a GltD domain (Fig. 5), suggesting that it may be able to interconvert the oxidized and reduced forms of NAD^+, $NAD(P)^+$ and ferredoxin as described for *Sporomusa* [24]. The genes encoding this complex occur together with a gene encoding a cytoplasmic, Mo-containing formate dehydrogenase; however, the functional significance of this arrangement is unclear at
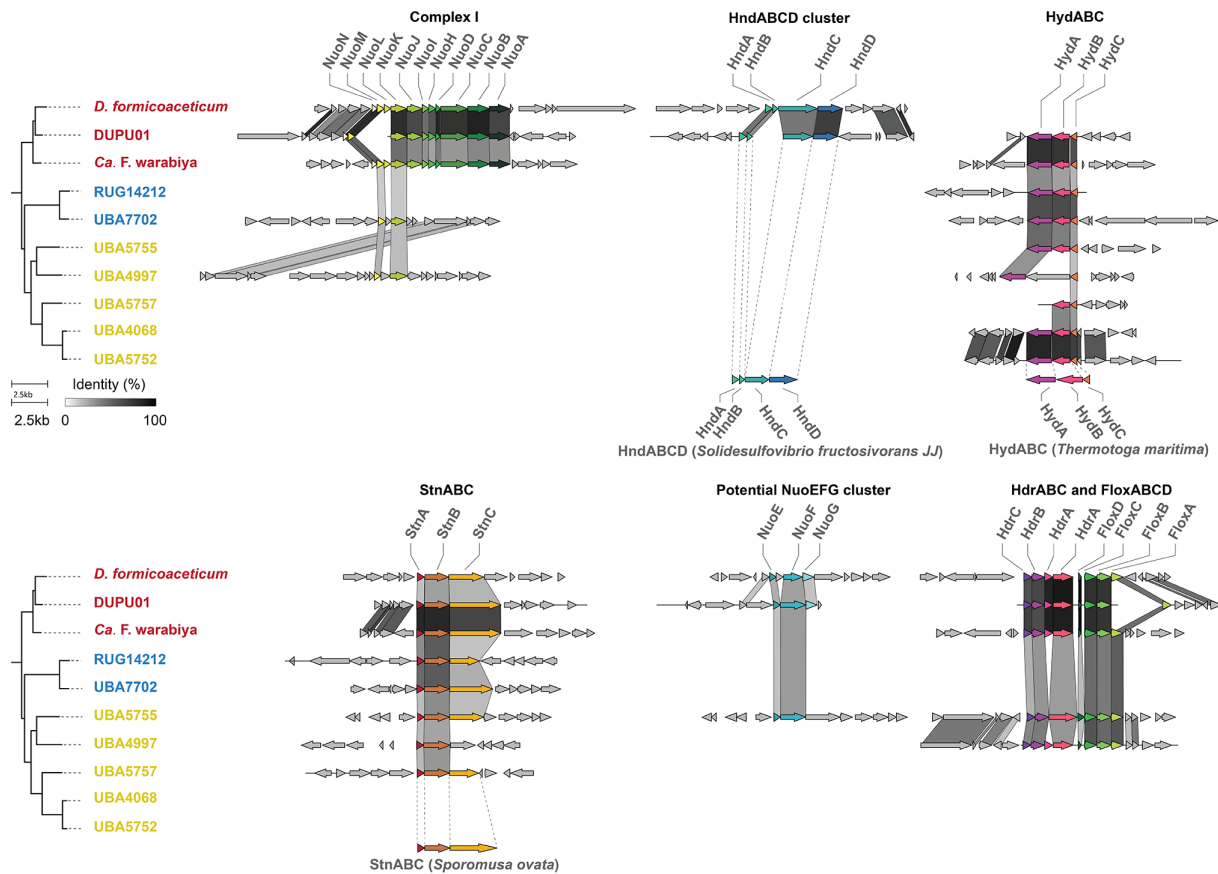
Fig. 5. Gene neighbourhood diagrams of redox complexes identified in *Dehalobacteriia* genomes. The greyscale bands indicate amino acid identities between the corresponding pair of genes (a minimum amino acid identity of 30% was used to connect genes between genomes).

present. A third gene cluster is present in members of the orders *Dehalobacteriales* and UB4068 and consists of the NuoE- and NuoF-like subunits as well as a subunit with homology to NuoG and other iron–sulphur cluster proteins (Fig. 5). We were unable to identify characterized, close homologues of this gene cluster, but it is possible that this complex may function as a hydrogenase, similar to HndABC. Despite the presence of multiple hydrogenases and the potential in some orders to fix CO_2 via the formate dehydrogenase reaction, the ability to grow as hydrogen-dependent chemolithotrophs has been excluded for *D. formicoaceticum* and '*Ca*. F. warabiya' [9, 10].

Two additional redox complexes that have been demonstrated to be involved in redox reactions in sulphate-reducing bacteria [60] are present in the *Dehalobacteriales* and some members of the order UBA4068. These gene clusters encode an HdrABC heterodisulphide reductase-like complex, which is associated with genes encoding a Flox-type NADH dehydrogenase that has been proposed to deliver electrons to the Hdr system (Fig. 5). The role of the HdrABC and Flox complexes in *Dehalobacteriia* is currently unknown but could involve providing reducing equivalents to the WLP.

Amino acid metabolism

As already indicated, CO_2 fixation in *Dehalobacteriia* via the WLP appears to be connected to the synthesis of glycine/ serine and possibly methionine as end products (Fig. 4). The inferred reliance of *Dehalobacteriia* on glycine and serine as intermediates in carbon assimilation suggests that they may be able to use these amino acids as carbon sources. Externally supplied glycine could be converted to serine using serine/glycine hydroxymethyltransferase (GlyA; EC 2.1.2.1; K00600) that is conserved in most members of this class, and both externally supplied and endogenously produced serine is predicted to be assimilated as pyruvate after conversion by L-serine dehydratase (SdaAB; EC 4.3.1.17; K01752) as described above. In the *Dehalobacteriales*, glycine could alternatively be converted to acetyl-phosphate by glycine reductase. With the exception of *D. formicoaceticum,* members of the *Dehalobacteriales* and UBA7702 orders may also be able to convert threonine to glycine for assimilation using threonine aldolase (LtaE; K01620; EC 4.1.2.48). This is an interesting discovery, as growth on amino

acids alone has so far not been demonstrated for the well-characterized isolated representative of this group, *D. formicoaceticum*, and warrants further investigation.

Experimental verification of metabolic inferences

To verify some of our metabolic inferences for the *Dehalobacteriia*, we generated proteome data (*n*=3 biological replicates) for *D. formicoaceticum* grown with DCM as the only energy source. An average of 1294±43 proteins (false discovery rate of 1%) were detected in the replicate samples, representing 1340 unique proteins. Of these, 1012 proteins (Table S8) or 28% of the theoretical *D. formicoaceticum* proteome were detected in all three DCM-culture replicates and used for further analyses. Of these 1012 proteins, 702 could be assigned a KEGG ID, collectively representing 22 functional protein categories (Tables S9 and S10). The five most abundant categories were 'Genetic information processing' (16.4% of the KEGG-annotated proteins, 115 proteins), followed by 'Amino acid metabolism' (9.6%, 97 proteins), 'Energy metabolism' (8.9%, 90 proteins), 'Metabolism of cofactors and vitamins (8.3%, 84 proteins)' and 'Carbohydrate metabolism' (7.9%, 80 proteins). These categories are all associated with basic cellular function, which is expected.

All enzymes needed for both the methyl- and carbonyl- branches of the WLP, the degradation of DCM via Mec proteins and enzymes involved in the reductive glycine pathway (i.e. GCS and glycine reductase) were identified and most were present in the top 10% of detected proteins in all three biological replicates (5/8 WLP, 5/6 Mec proteins and 6/13 proteins of the reductive glycine pathway) (Table S8). The presence of Mec, WLP and reductive glycine pathway proteins in the proteomics data is consistent with results described by Murdoch *et al.* [12], confirming the importance of C1-assimilation pathways in DCM metabolism of *D. formicoaceticum*. Furthermore, the presence of serine hydroxymethyltransferase (GlyA) and glycine reductase (Grd) suggests that *D. formicoaceticum* may be able to assimilate serine into biomass as predicted from metabolic reconstruction (Fig. 4). Our prediction that methionine is a product of the WLP was supported by the observation that a methionine synthase (MetH)-related protein was also highly abundant (Table S8). Production of acetate, a known end product of *D. formicoaceticum* metabolism, is probably catalysed by AckA and PduL that were also present in the top 10% of detected proteins.

In addition to C1 metabolism-related proteins, several proteins involved in amino acid transport and metabolism were highly abundant. This included an NADP$^+$-specific glutamate dehydrogenase (GdhA; EC 1.4.1.4; K00262), a branched-chain amino acid aminotransferase (IlvE; EC 2.6.1.42; K00826), as well as tryptophan synthase (EC 4.2.1.20; K01695 and K01696), diaminopimelate decarboxylase (LysA; EC 4.1.1.20; K01586) and ʟʟ-diaminopimelate aminotransferase (K10206; EC 2.6.1.83). A substrate-binding protein of an ATP-binding cassette (ABC) transporter with specificity for leucine, isoleucine and valine (LivK; K01999) was also detected (Table S4). This suggests that amino acids play an important role in *D. formicoaceticum* metabolism even during growth on DCM-containing minimal medium. In this medium the small amount of yeast extract added would have provided some limited access to an external supply of amino acids. Unexpectedly, the top 10% of proteins also contained a taurine:pyruvate aminotransferase and a sulfoacetaldehyde-acetyltransferase (Table S8) that allow the conversion of taurine and pyruvate to acetyl-phosphate and sulphite via the intermediate sulfoacetaldehyde, indicating that *D. formicoaceticum* may be able to use taurine as a carbon source.
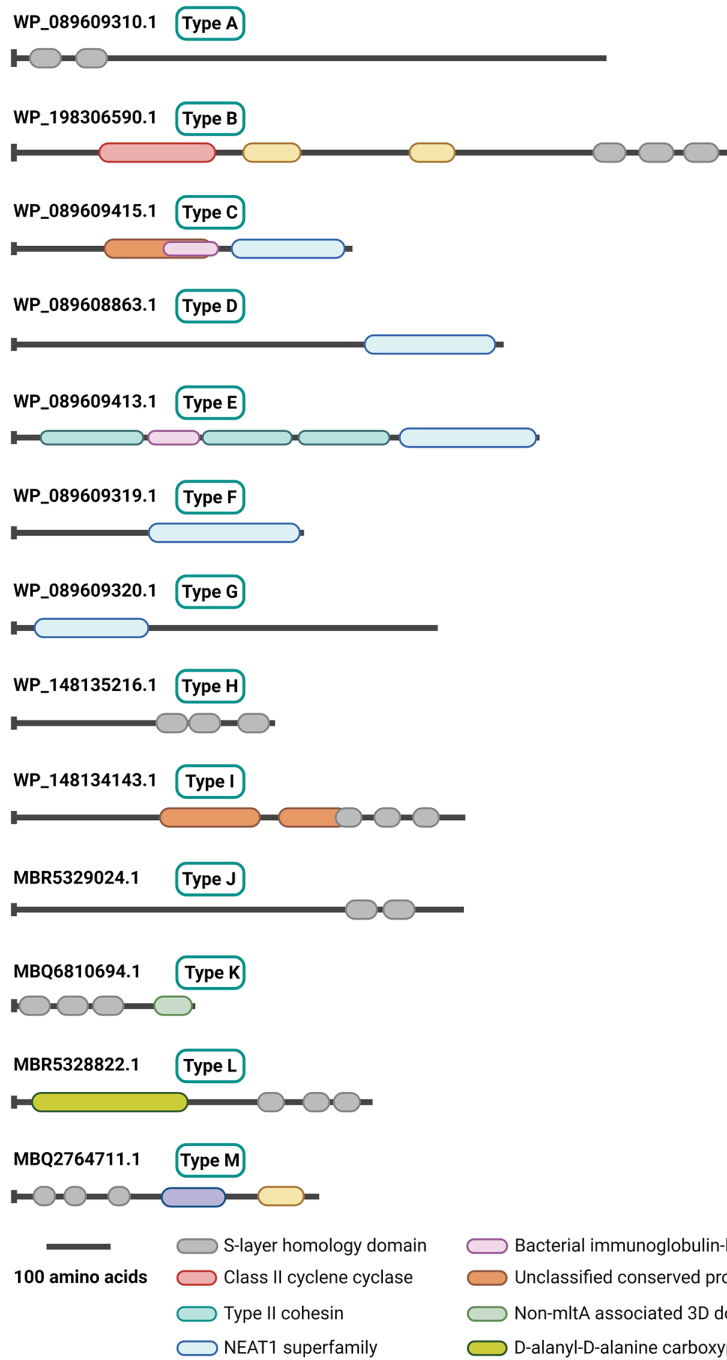
Energy conservation via redox complexes

Proteins of several of the putative electron-bifurcating complexes described above were also abundant in the proteome. All four subunits of the putative cytoplasmic [FeFe] hydrogenase Hnd (HndABCD) were present in the top 10% of detected proteins (Table S11). Similarly, all subunits of the protein complex that resembles the *Sporomusa* StnABC transhydrogenase were present in the top 10% of expressed proteins (Tables S8 and S11). Other electron-bifurcating complexes that do not use hydrogen as an electron donor, including HdrABC heterodisulphide reductase/FloxABCD were also highly abundant (Tables S8 and S11). Overall, the abundance of proteins from electron-bifurcating complexes was similar to core WLP proteins, suggesting that they play key roles in energy conservation of *D. formicoaceticum* and that hydrogenase activity may be linked to DCM degradation.
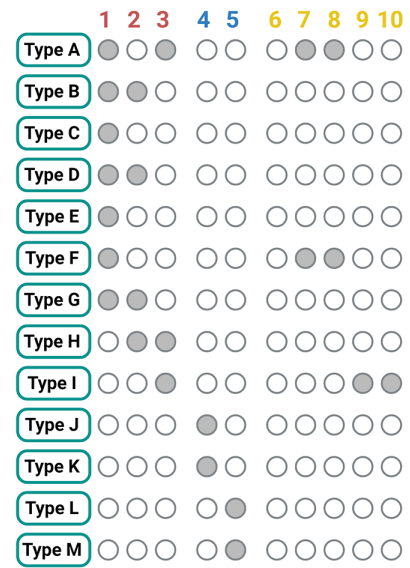
S-layer proteins are abundant in *D. formicoaceticum* during growth on DCM

In addition to proteins involved in energy and carbon metabolism, seven surface layer (S-layer) proteins (SLPs; WP_089609413.1, WP_089608863.1, WP_089609310.1, WP_198306590.1, WP_089609319.1 and WP_089609320.1) were present in the top 10% of detected proteins (Table S8), demonstrating that, as in many other bacteria, SLPs are highly abundant in *D. formicoaceticum*. SLPs typically contain one or more S-layer homology domain (pfam00395) at either the N- or C-terminus of the protein (Fig. 6) with a number of other domains known to occur that determine SLP function [61]. Firmicute SLPs are best characterized in *Bacillus anthracis* and *Clostridioides difficile*, where they facilitate functions such as growth phase adaptation, and interactions with host cells and other surfaces [62]. However, *D. formicoaceticum* SLPs had low similarity to *B. anthracis* and *C. difficile* SLPs (<35% amino acid identity), and this was limited to the S-layer homology or NEAr Transporter (NEAT; cd06920) domains representing between 16 and 34% of the protein. NEAT domains have been suggested to play a role in iron transport [63]. The only exception was SLP WP_089609319.1, for which an alignment of the protein sequence to its closest homologue in *B. anthracis* (WP_140420396.1) covered 68% of the sequence length with an amino acid identity of 30%.
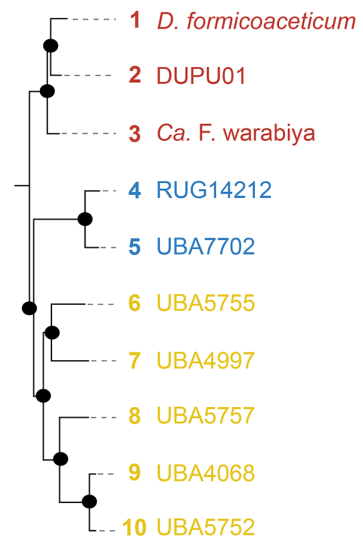
Fig. 6. (a) S-layer proteins identified in *Dehalobacteriia* genomes. (b) Identification of S-layer proteins in *Dehalobacteriia* genomes based on BLASTP homology search against *D. formicoaceticum* proteins. Filled and open circles indicate presence and absence of proteins respectively. (c) Legend showing phylogenetic relationships of *Dehalobacteriia* genomes colour-coded by order (Fig. 1).

While S-layer homology and NEAT domains were commonly found in the *D. formicoaceticum* SLPs, we also identified other associated domains (Fig. 6). One SLP (WP_089609413.1) has multiple 'Type II cohesion domains (cd08547)' that have been shown to play a role in the attachment of the cellulose-degrading cellulosome to the bacterial cell wall. This SLP also contains a single bacterial IG-like domain (pfam02368), which mediates host interactions in *Escherichia coli* and has also been detected in SLPs from *Bacillus* species [63]. However, most of the *D. formicoaceticum* SLPs, including the most highly abundant one (WP_198306590.1), lack identifiable domains and therefore their function remains unclear. It is interesting to note, however, that
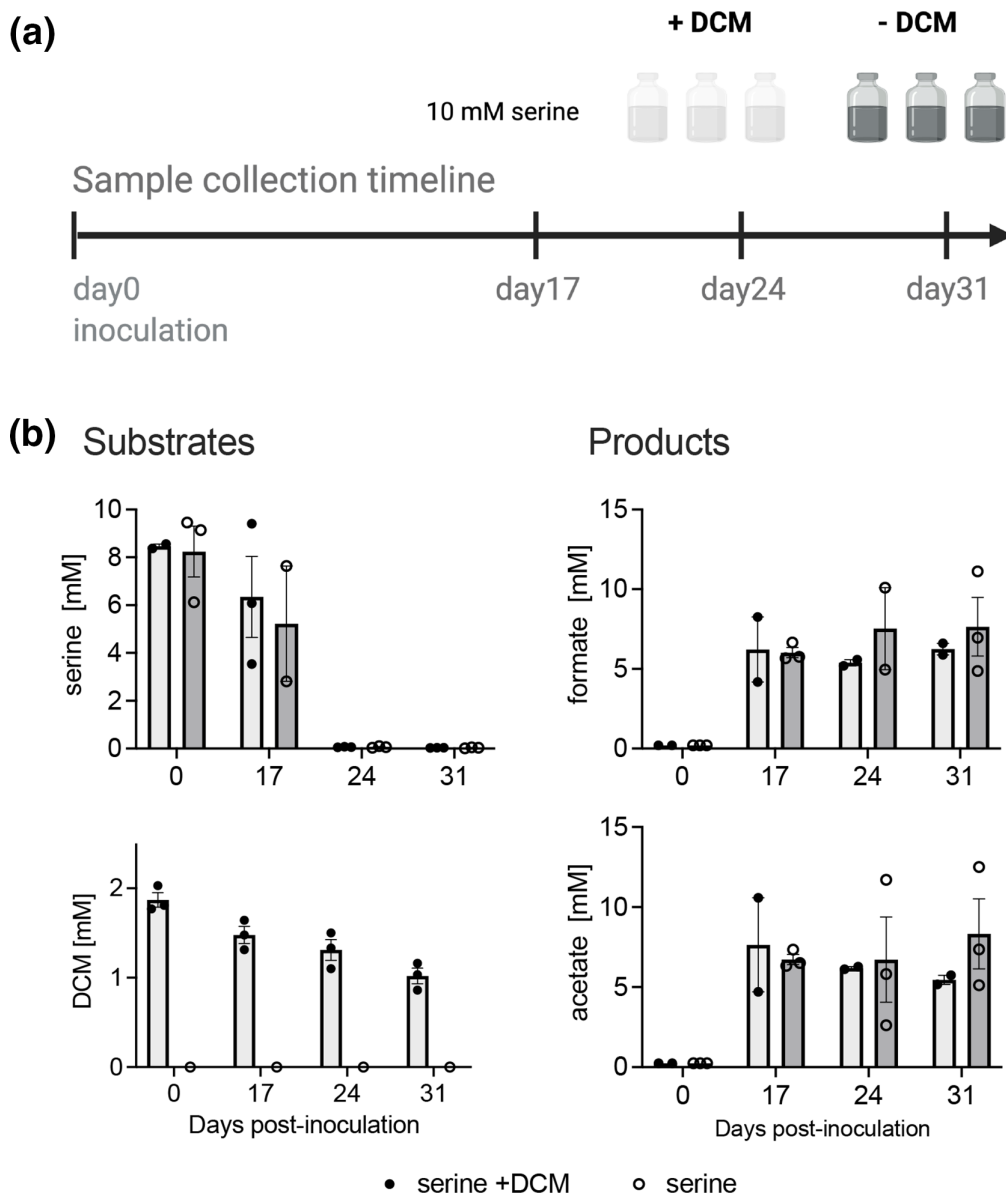
Fig. 7. Serine degradation by *D. formicoaceticum* in the presence or absence of DCM. (a) Experimental set-up. (b) GC-MS measurement of serine, formate, acetate and DCM in both DCM and non-DCM samples over a 31 day cultivation period.

while SLPs from *B. anthracis* appear to always contain three consecutive S-layer homology domains, this is only the case for one of the *D. formicoaceticum* SLPs. In summary, we show that SLPs found in *D. formicoaceticum* are highly expressed during growth on DCM but may also play key roles under other growth conditions. Additionally, they have novel protein domain structures that add to our knowledge of SLPs and warrant biochemical characterization.

Metabolic reconstruction indicated that *Dehalobacteriia* including *D. formicoaceticum*, which is known as an obligate DCM-degrader [10], may be able to use amino acids such as serine as a carbon and energy source. To test this, we grew *D. formicoaceticum* in mineral medium containing 10 mM serine as the carbon source or 10 mM serine in combination with 2 mM DCM for 31 days, using three biological replicate cultures for each condition.

In all cultures, turbidity indicated bacterial growth and on day 17 post-inoculation only 0.06±0.01 and 0.07±0.03 mM serine remained in samples with and without DCM, respectively. In cultures with added DCM, the average DCM concentration decreased from 1.9±0.1 to 1.5±0.1 mM at day 17. Serine-only cultures produced acetate (5.4±3.7 mM) and formate (6.5±4.8 mM) as the dominant end products. These products are consistent with cultures containing DCM only, although differing from the 2:1 molar ratio of formate and acetate previously observed [10]. Cultures containing both serine and DCM also produced these end

products (7.3±0.6 mM acetate, 8.1±0.6 mM formate), which was expected due to the presence of DCM. On day 17, all cultures were supplemented with an additional 10 mM serine and 2 mM DCM that continued to be consumed until day 31 (Fig. 7).

These observations indicate that *D. formicoaceticum* is able to use serine as a growth substrate in the absence of DCM, consistent with the metabolic reconstruction for members of the *Dehalobacteriia* (Fig. 4). Serine metabolism may have proceeded along various routes including assimilation into the WLP using the glycine cleavage system, or the SdaAB enzyme system. We note, however, that in the proteomics data we collected, high levels of expression were only observed for the reductive glycine system, while SdaAB was not present (Table S4). Furthermore, the proteins necessary to convert serine to the observed end products, acetate and formate, are conserved in the orders of *Dehalobacteriia*, and expressed in *D. formicoaceticum* during growth in the presence of DCM alone, which might indicate that they are constitutively produced (Figs 3a and 4; Table S8). Metabolism of serine is a novel expansion of the known substrate range of *D. formicoaceticum*, which has been considered an obligate DCM degrader for over 20 years [10]. '*Ca*. F. warabiya' has also been reported to grow with substrates other than DCM, including choline, glycine betaine and methanol [12]. Notably, DCM consumption proceeded at a slower rate than usual in the DCM and serine cultures, suggesting that serine may be used preferentially, or cause inhibition of DCM metabolism.

CONCLUDING REMARKS

In summary, we show that anaerobic DCM degradation, regarded as a defining trait of *D. formicoaceticum*, is probably only present in two of the ten investigated members of the class *Dehalobacteriia* and, moreover, appears to have been recently acquired via horizontal gene transfer. By contrast, use of serine as a growth substrate, and use of the glycine cleavage pathway and WLP in central carbon metabolism appear to be broadly conserved in this class. A number of predicted electron-bifurcating complexes identified in the genomes of these bacteria appear to support fermentative metabolism and warrant further investigation. The validity of these inferences was confirmed using the only pure culture representative of the *Dehalobacteriia*, *D. formicoaceticum*, which was able to grow on serine in the absence of DCM. Taken together, our data indicate that DCM degradation is a recently acquired and restricted trait in the *Dehalobacteriia* and that amino acid fermentation is a widespread and ancestral trait in this lineage.

Conflicts of interest
The authors have no conflicts of interest to declare.

References

1. Perez-Riverol Y, Bai J, Bandla C, García-Seisdedos D, Hewapathirana S, *et al*. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res* 2022;50:D543–D552.

2. Parks DH, Chuvochina M, Chaumeil P-A, Rinke C, Mussig AJ, *et al*. A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat Biotechnol* 2020;38:1079–1086.

3. Pasolli E, Asnicar F, Manara S, Zolfo M, Karcher N, *et al*. Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. *Cell* 2019;176:649–662.

4. Kyrpides NC, Hugenholtz P, Eisen JA, Woyke T, Göker M, *et al*. Genomic encyclopedia of bacteria and archaea: sequencing a myriad of type strains. *PLOS Biol* 2014;12:e1001920.

5. Parte AC, Sardà Carbasse J, Meier-Kolthoff JP, Reimer LC, Göker M. List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ. *Int J Syst Evol Microbiol* 2020;70:5607–5612.

6. Hossaini R, Chipperfield MP, Montzka SA, Leeson AA, Dhomse SS, *et al*. The increasing threat to stratospheric ozone from dichloromethane. *Nat Commun* 2017;8:15962.

7. Gribble GW. *Naturally Occurring Organohalogen Compounds - A Comprehensive Update*. Vienna: Springer; 2010.

8. Muller EEL, Hourcade E, Louhichi-Jelail Y, Hammann P, Vuilleumier S, *et al*. Functional genomics of dichloromethane utilization in *Methylobacterium extorquens* DM4. *Env Microbiol* 2011;13:2518–2535.

9. Firsova JE, Doronina NV, Trotsenko YuA. Analysis of the key functional genes in new aerobic degraders of dichloromethane. *Microbiology* 2010;79:66–72.

10. Mägli A, Wendt M, Leisinger T. Isolation and characterization of *Dehalobacterium formicoaceticum* gen. nov. sp. nov., a strictly anaerobic bacterium utilizing dichloromethane as source of carbon and energy. *Arch Microbiol* 1996;166:101–108.

11. Holland SI, Ertan H, Montgomery K, Manefield MJ, Lee M. Novel dichloromethane-fermenting bacteria in the *Peptococcaceae* family. *ISME J* 2021;15:1709–1721.

12. Holland SI, Edwards RJ, Ertan H, Wong YK, Russell TL, *et al*. Whole genome sequencing of a novel, dichloromethane-fermenting *Peptococcaceae* from an enrichment culture. *PeerJ* 2019;7:e7775.

13. Kleindienst S, Chourey K, Chen G, Murdoch RW, Higgins SA, *et al*. Proteogenomics reveals novel reductive dehalogenases and methyltransferases expressed during anaerobic dichloromethane metabolism. *Appl Environ Microbiol* 2019;85:e02768–18.

14. Murdoch RW, Chen G, Kara Murdoch F, Mack EE, Villalobos Solis MI, *et al*. Identification and widespread environmental distribution of a gene cassette implicated in anaerobic dichloromethane degradation. *Global Change Biology* 2022;28:2396–2412.

15. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;30:2068–2069.

16. Engelbrektson A, Kunin V, Wrighton KC, Zvenigorodsky N, Chen F, *et al*. Experimental factors affecting PCR-based estimates of microbial species richness and evenness. *ISME J* 2010;4:642–647.

17. NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 2016;44:D7–D19.

18. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, *et al*. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 2013;41:D590–6.

19. Karp PD, Midford PE, Billington R, Kothari A, Krummenacker M, *et al*. Pathway Tools version 23.0 update: software for pathway/genome informatics and systems biology. *Briefings Bioinfo* 2021;22:109–126.

20. Aoki-Kanehisa KF, Kanehisa M. Gene annotation and pathway mapping in KEGG. In: *Comparative Genomics Methods in Molecular Biology*, vol. 396. Humana Press, 2007.

21. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol* 2016;428:726–731.

22. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, *et al*. CDD: NCBI's conserved domain database. *Nucleic Acids Res* 2015;43:D222–6.

23. Kuhns M, Trifunović D, Huber H, Müller V. The Rnf complex is a Na+ coupled respiratory enzyme in a fermenting bacterium, *Thermotoga maritima*. *Commun Biol* 2020;3:431.

24. Kremp F, Roth J, Müller V. The *Sporomusa* type Nfn is a novel type of electron-bifurcating transhydrogenase that links the redox pools in acetogenic bacteria. *Sci Rep* 2020;10:14872.

25. Kpebe A, Benvenuti M, Guendon C, Rebai A, Fernandez V, *et al*. A new mechanistic model for an O_2-protected electron-bifurcating hydrogenase, Hnd from *Desulfovibrio fructosovorans*. *Biochim Biophys Acta Bioenerg* 2018;1859:1302–1312.

26. Gilchrist CLM, Chooi Y-H. clinker & clustermap.js: automatic generation of gene cluster comparison figures. *Bioinformatics* 2021;37:2473–2475.

27. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 2017;14:587–589.

28. Chen G, Shouakar-Stash O, Phillips E, Justicia-Leon SD, Gilevska T, *et al*. Dual carbon-chlorine isotope analysis indicates distinct anaerobic dichloromethane degradation pathways in two members of *Peptococcaceae*. *Environ Sci Technol* 2018;52:8607–8616.

29. Wolin EA, Wolin MJ, Wolfe RS. Formation of methane by bacterial extracts. *J Biol Chem* 1963;238:2882–2886.

30. Wolin EA, Wolin MJ, Wolfe RS. Formation of methane by bacterial extracts. *J Biol Chem* 1963;238:2882–2886.

31. Staudinger J, Roberts PV. A critical review of Henry's law constants for environmental applications. *Crit Rev Env Sci Tech* 1996;26:205–297.

32. Villas-Bôas SG, Delicado DG, Akesson M, Nielsen J. Simultaneous analysis of amino and nonamino organic acids as methyl chloroformate derivatives using gas chromatography-mass spectrometry. *Anal Biochem* 2003;322:134–138.

33. Wiśniewski JR. Filter aided sample preparation - a tutorial. *Anal Chim Acta* 2019;1090:23–30.

34. R: A language and environment for statistical computing. R Core Team. R Foundation for Statistical Computing. Vienna, Austria; 2017. https://www.R-project.org/ [accessed 1 November 2019].

35. Wang S, Huang H, Kahnt J, Thauer RK. A reversible electron-bifurcating ferredoxin- and NAD-dependent [FeFe]-hydrogenase (HydABC) in *Moorella thermoacetica*. *J Bacteriol* 2013;195:1267–1275.

36. Pierce E, Xie G, Barabote RD, Saunders E, Han CS, *et al*. The complete genome sequence of *Moorella thermoacetica* (f. *Clostridium thermoaceticum*). *Env Microbiol* 2008;10:2550–2573.

37. Villemur R, Lanthier M, Beaudet R, Lépine F. The *Desulfitobacterium* genus. *FEMS Microbiol Rev* 2006;30:706–733.

38. Imachi H, Sekiguchi Y, Kamagata Y, Hanada S, Ohashi A, *et al*. *Pelotomaculum thermopropionicum* gen. nov., sp. nov., an anaerobic, thermophilic, syntrophic propionate-oxidizing bacterium. *Int J Syst Evol Microbiol* 2002;52:1729–1735.

39. An D, Caffrey SM, Soh J, Agrawal A, Brown D, *et al*. Metagenomics of hydrocarbon resource environments indicates aerobic taxa and genes to be unexpectedly common. *Environ Sci Technol* 2013;47:10708–10717.

40. Martínez-Álvaro M, Auffret MD, Stewart RD, Dewhurst RJ, Duthie C-A, *et al*. Identification of complex rumen microbiome interaction within diverse functional niches as mechanisms affecting the variation of methane emissions in bovine. *Front Microbiol* 2020;11:659.

41. Campanaro S, Treu L, Rodriguez-R LM, Kovalovszki A, Ziels RM, *et al*. New insights from the biogas microbiome by comprehensive genome-resolved metagenomics of nearly 1600 species originating from multiple anaerobic digesters. *Biotechnol Biofuels* 2020;13:25.

42. Guo M-Y, Huo D-Q, Ghai R, Rodriguez-Valera F, Shen C-H, *et al*. Metagenomics of ancient fermentation pits used for the production of chinese strong-aroma liquor. *Genome Announc* 2014;2:e01045-14.

43. Yamamuro A, Kouzuma A, Abe T, Watanabe K. Metagenomic analyses reveal the involvement of syntrophic consortia in methanol/electricity conversion in microbial fuel cells. *PLoS One* 2014;9:e98425.

44. Wen L, Ley RE, Volchkov PY, Stranges PB, Avanesyan L, *et al*. Innate immunity and intestinal microbiota in the development of Type 1 diabetes. *Nature* 2008;455:1109–1113.

45. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, *et al*. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* 2006;444:1027–1031.

46. Ley RE, Bäckhed F, Turnbaugh P, Lozupone CA, Knight RD, *et al*. Obesity alters gut microbial ecology. *Proc Natl Acad Sci* 2005;102:11070–11075.

47. Ley RE, Turnbaugh PJ, Klein S, Gordon JI. Microbial ecology: human gut microbes associated with obesity. *Nature* 2006;444:1022–1023.

48. Liu N, Zhang L, Zhou H, Zhang M, Yan X, *et al*. Metagenomic insights into metabolic capacities of the gut microbiota in a fungus-cultivating termite (*Odontotermes yunnanensis*). *PLoS One* 2013;8:e69184.

49. Tischer K, Kleinsteuber S, Schleinitz KM, Fetzer I, Spott O, *et al*. Microbial communities along biogeochemical gradients in a hydrocarbon-contaminated aquifer. *Environ Microbiol* 2013;15:2603–2615.

50. Cheng T-W, Chang Y-H, Tang S-L, Tseng C-H, Chiang P-W, *et al*. Metabolic stratification driven by surface and subsurface interactions in a terrestrial mud volcano. *ISME J* 2012;6:2280–2290.

51. Hubalek V, Buck M, Tan B, Foght J, Wendeberg A, *et al*. Vitamin and amino acid auxotrophy in anaerobic consortia operating under methanogenic conditions. *mSystems* 2017;2:e00038-17.

52. Gieg LM, Fowler SJ, Berdugo-Clavijo C. Syntrophic biodegradation of hydrocarbon contaminants. *Curr Opin Biotechnol* 2014;27:21–29.

53. 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* 2018;115:E1166–E1173.

54. Zhuang W-Q, Yi S, Bill M, Brisson VL, Feng X, *et al*. Incomplete Wood-Ljungdahl pathway facilitates one-carbon metabolism in organohalide-respiring *Dehalococcoides mccartyi*. *Proc Natl Acad Sci* 2014;111:6419–6424.

55. Bassegoda A, Madden C, Wakerley DW, Reisner E, Hirst J. Reversible interconversion of CO2 and formate by a molybdenum-containing formate dehydrogenase. *J Am Chem Soc* 2014;136:15473–15476.

56. Song Y, Lee JS, Shin J, Lee GM, Jin S, *et al*. Functional cooperation of the glycine synthase-reductase and Wood–Ljungdahl pathways for autotrophic growth of *Clostridium drakei*. *Proc Natl Acad Sci* 2020;117:7516–7523.

57. Cordero PRF, Bayly K, Man Leung P, Huang C, Islam ZF, *et al*. Atmospheric carbon monoxide oxidation is a widespread mechanism supporting microbial survival. *ISME J* 2019;13:2868–2881.

58. Moparthi VK, Hägerhäll C. The evolution of respiratory chain complex I from a smaller last common ancestor consisting of 11 protein subunits. *J Mol Evol* 2011;72:484–497.

59. **Schut GJ**, **Adams MWW**. The iron-hydrogenase of *Thermotoga maritima* utilizes ferredoxin and NADH synergistically: a new perspective on anaerobic hydrogen production. *J Bacteriol* 2009;191:4451–4457.

60. **Meyer B**, **Kuehl JV**, **Price MN**, **Ray J**, **Deutschbauer AM**, *et al*. The energy-conserving electron transfer system used by *Desulfovibrio alaskensis* strain G20 during pyruvate fermentation involves reduction of endogenously formed fumarate and cytoplasmic and membrane-bound complexes, Hdr-Flox and Rnf. *Environ Microbiol* 2014;16:3463–3486.

61. **Ravi J**, **Fioravanti A**. S-layers: the proteinaceous multifunctional armors of gram-positive pathogens. *Front Microbiol* 2021;12:663468.

62. **Fagan RP**, **Fairweather NF**. Biogenesis and functions of bacterial S-layers. *Nat Rev Microbiol* 2014;12:211–222.

63. **Tarlovsky Y**, **Fabian M**, **Solomaha E**, **Honsa E**, **Olson JS**, *et al*. A Bacillus anthracis S-layer homology protein that binds heme and mediates heme delivery to IsdC. *J Bacteriol* 2010;192:3503–3511.