AMERICAN SOCIETY FOR MICROBIOLOGY | mSystems®

RESEARCH ARTICLE
Ecological and Evolutionary Science

Defining Genomic and Predicted Metabolic Features of the *Acetobacterium* Genus

Daniel E. Ross,[a,b] Christopher W. Marshall,[c] Djuna Gulliver,[a] Harold D. May,[d] R. Sean Norman[e]

[a]National Energy Technology Laboratory, Pittsburgh, Pennsylvania, USA
[b]Leidos Research Support Team, Pittsburgh, Pennsylvania, USA
[c]Department of Biological Sciences, Marquette University, Milwaukee, Wisconsin, USA
[d]Department of Microbiology & Immunology, Hollings Marine Laboratory, Medical University of South Carolina, Charleston, South Carolina, USA
[e]Department of Environmental Health Sciences, Arnold School of Public Health, University of South Carolina, Columbia, South Carolina, USA

ABSTRACT Acetogens are anaerobic bacteria capable of fixing CO_2 or CO to produce acetyl coenzyme A (acetyl-CoA) and ultimately acetate using the Wood-Ljungdahl pathway (WLP). *Acetobacterium woodii* is the type strain of the *Acetobacterium* genus and has been critical for understanding the biochemistry and energy conservation in acetogens. Members of the *Acetobacterium* genus have been isolated from a variety of environments or have had genomes recovered from metagenome data, but no systematic investigation has been done on the unique and various metabolisms of the genus. To gain a better appreciation for the metabolic breadth of the genus, we sequenced the genomes of 4 isolates (*A. fimetarium*, *A. malicum*, *A. paludosum*, and *A. tundrae*) and conducted a comparative genome analysis (pan-genome) of 11 different *Acetobacterium* genomes. A unifying feature of the *Acetobacterium* genus is the carbon-fixing WLP. The methyl (cluster II) and carbonyl (cluster III) branches of the Wood-Ljungdahl pathway are highly conserved across all sequenced *Acetobacterium* genomes, but cluster I encoding the formate dehydrogenase is not. In contrast to *A. woodii*, all but four strains encode two distinct Rnf clusters, Rnf being the primary respiratory enzyme complex. Metabolism of fructose, lactate, and H_2:CO_2 was conserved across the genus, but metabolism of ethanol, methanol, caffeate, and 2,3-butanediol varied. Additionally, clade-specific metabolic potential was observed, such as amino acid transport and metabolism in the psychrophilic species, and biofilm formation in the *A. wieringae* clade, which may afford these groups an advantage in low-temperature growth or attachment to solid surfaces, respectively.

IMPORTANCE Acetogens are anaerobic bacteria capable of fixing CO_2 or CO to produce acetyl-CoA and ultimately acetate using the Wood-Ljungdahl pathway (WLP). This autotrophic metabolism plays a major role in the global carbon cycle and, if harnessed, can help reduce greenhouse gas emissions. Overall, the data presented here provide a framework for examining the ecology and evolution of the *Acetobacterium* genus and highlight the potential of these species as a source for production of fuels and chemicals from CO_2 feedstocks.

KEYWORDS *Acetobacterium*, pan-genome, metagenome-assembled genome, acetogens

Acetogens are ubiquitous in nature, phylogenetically diverse, and produce acetyl coenzyme A (acetyl-CoA) from fixation of two molecules of CO_2, producing acetate as the sole product from the reductive acetyl-CoA/Wood-Ljungdahl pathway (WLP) (1). Of the over 22 genera that contain acetogens, *Clostridium* and *Acetobacterium* contain the most acetogenic species (2). *Acetobacterium* spp. are of particular interest because

Citation Ross DE, Marshall CW, Gulliver D, May HD, Norman RS. 2020. Defining genomic and predicted metabolic features of the *Acetobacterium* genus. mSystems 5:e00277-20. https://doi.org/10.1128/mSystems.00277-20.

Editor Nick Bouskill, Lawrence Berkeley National Laboratory

The review history of this article can be read here.

Address correspondence to Daniel E. Ross, daniel.ross@netl.doe.gov.

Received 2 April 2020
Accepted 20 August 2020
Published 15 September 2020

TABLE 1 Genome characteristics of 14 *Acetobacterium* isolates and MAGs

Isolate or MAG (accession no.)	Genome size (bp)	GC%	Type of genome	No. of contigs	Completeness (%)	Contamination (%)	Source	Reference(s)
A. woodii (CP002987.1)	4,044,777	39.3	Isolate	1	100	1.43	Black sediment	3, 22
Acetobacterium sp. KB-1 (CP030040.1)	3,976,829	42.5	MAG	1,006	100	0.71	Anaerobic bioreactor	S. Tang, K. Kirvushin, O. Molenda, and E. A. Edwards, 2018, unpublished
A. bakii (LGYO00000000.1)	4,135,228	41.2	Isolate	103	99.29	2.86	Paper mill wastewater pond	92, 93
A. dehalogenans (AXAC00000000.1)	4,045,770	43.8	Isolate	81	99.29	0	Wastewater sludge	94
Acetobacterium sp. MES1 (MJUY00000000.1)	3,654,423	44.3	MAG	65	99.29	0	Brewery wastewater	35, 95
A. malicum DSM 4132^T (WJBE00000000)	4,086,418	43.7	Isolate	92	99.29	1.43	Freshwater sediment	96
A. paludosum DSM 8237^T (WJBD00000000)	3,696,449	40.1	Isolate	86	99.29	2.14	Bog sediment	92
A. tundrae DSM 9173^T (WJBB00000000)	3,567,344	39.7	Isolate	94	99.29	1.43	Tundra wetland soil	97
A. wieringae (LKEU00000000.1)	3,895,828	44.1	Isolate	62	98.57	0.71	Sewage digester	61
Acetobacterium sp. UBA5834 (DJGY00000000.1)	3,463,311	44.1	MAG	87	98.57	1.43	Bioreactor	98
A. fimetarium DSM 8238^T (WJBC00000000)	3,245,738	44.7	Isolate	99	98.57	0.71	Cattle manure	92
Acetobacterium sp. UBA6819 (DKEP00000000.1)	3,607,361	42.8	MAG	474	94.52	3.83	Wastewater	98
Acetobacterium sp. UBA5558 (DIMU00000000.1)	2,612,337	43.6	MAG	110	85	0.71	Bioreactor	98
Acetobacterium sp. UBA11218 (DNHO00000000.1)	2,224,242	44.8	MAG	1,582	57	0.75	Bioreactor	98

the WLP is the genus' defining feature, and *Acetobacterium woodii* has been extensively studied. The genus *Acetobacterium* contains Gram-positive, non-spore-forming, homoacetogenic bacteria and was first described as a genus by Balch and coworkers, with the type strain *Acetobacterium woodii* WB1 (ATCC 29683) (3). Members of the genus *Acetobacterium* have been found in diverse environments, including sulfate-reducing permeable reactive zones (4), anoxic bottom waters of a volcanic subglacial lake (5, 6), seagrass rhizosphere (7), high-temperature gas-petroleum reservoirs (8), anaerobic granular sludge from fruit-processing wastewater (9), and biocathode communities (10–14). Sulfate-reducing bacteria (SRB) are often found in these environments as well, with several studies suggesting a syntrophic partnership between SRB and acetogens (7, 15). In particular, a combination of *Desulfovibrio*, *Sulfurospirillum*, and *Acetobacterium* was proposed to cooperatively participate in microbially induced corrosion (MIC) of steel (15), has been found in production waters from a biodegraded oil reservoir (16), and can be detected in natural subsurface CO_2 reservoirs (17). Moreover, these three microorganisms were the most abundant members of the biocathode community responsible for electrode-driven production of acetate and hydrogen from CO_2 (18–20).

To date, 14 sequenced *Acetobacterium* genomes are publicly available—one complete and closed genome of *Acetobacterium woodii* (a single contiguous sequence) (21) and 13 draft genomes (including metagenome-assembled genomes) of varying completeness (Table 1). *Acetobacterium woodii* is the type strain of the genus, is the best-characterized strain, and has been used as a model organism for understanding the bioenergetics of acetogens (reference 1 and references therein). Importantly, a genetic system also exists for *A. woodii* (22). While *A. woodii* utilizes the Wood-Ljungdahl pathway (WLP) for autotrophic growth, alternative pathways for carbon utilization are operative, including glucose metabolism, 1,2-propanediol degradation, 2,3-butanediol oxidation, ethanol oxidation, caffeate reduction, ethylene glycol metabolism, and alanine metabolism (23–28).

Outside the type strain, not much is known about the *Acetobacterium* genus. The genus appears to have varied metabolic capabilities, including low-temperature growth (29), debromination of polybrominated diphenyl ethers (PBDEs) (30), hexahydro-1,3,5-trinitro-1,3,5-triazine (RDX) degradation (31, 32), electrode-mediated acetogenesis (33), enhanced iron corrosion (34), and isoprene degradation (35). To provide a more robust genome data set, we sequenced and manually curated the genomes of four *Acetobacterium* strains, including *A. malicum* and three psychrophilic strains (*A. fimetarium*, *A. paludosum*, and *A. tundrae*). We also conducted a comparative genomic study to shed light on the phylogenetic relatedness and functional potential of the *Acetobacterium* genus. Our results suggest that while the major metabolic pathways are well conserved (e.g., the Wood-Ljungdahl pathway and accessory components), certain predicted

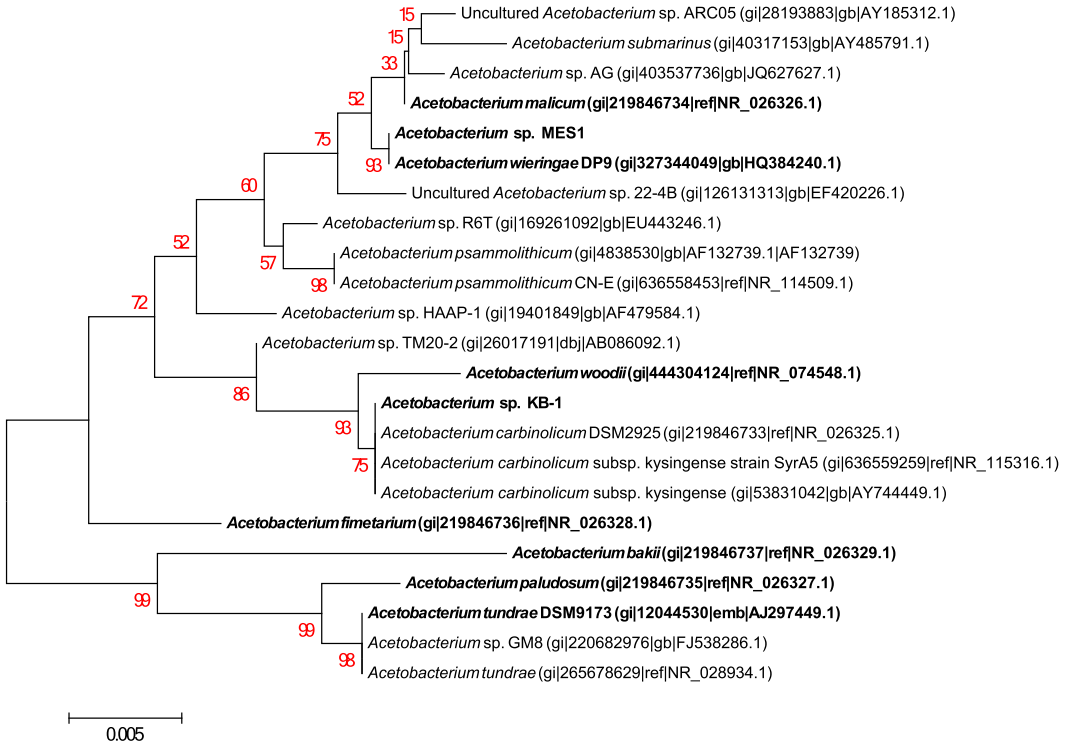

FIG 1 16S rRNA gene phylogeny. Phylogenetic tree of *Acetobacterium* 16S rRNA gene sequences. *Acetobacterium* strains in bold represent strains that have a sequenced genome. Numbers in red indicate percent bootstrap support.

metabolisms and genome features were markedly different than what is known from *Acetobacterium woodii*. Results described here provide insight into the defining common features of the *Acetobacterium* genus as well as ancillary pathways that may operate to provide various strains the ability to survive in diverse environments and potentially be exploited for biotechnological applications, such as the production of fuels and chemicals from CO_2 feedstocks (18, 20).

RESULTS AND DISCUSSION

Phylogeny of the *Acetobacterium* genus. As a first step toward identifying the taxonomic placement of newly sequenced *Acetobacterium* genomes, we collected 16S rRNA gene sequences from the NCBI database, and for metagenome-assembled genomes (MAGs) not in the database, we attempted to extract 16S rRNA gene sequences from metagenome data. Multiple MAGs did not contain 16S rRNA gene sequence, potentially due to metagenome assembly and binning issues. *Acetobacterium dehalogenans* contained a truncated 16S rRNA gene that was less than 400 bp and was not included in the 16S rRNA tree. Therefore, nine 16S rRNA gene sequences of the 13 sequenced *Acetobacterium* genomes were utilized. We reconstructed an updated phylogenetic tree of available 16S rRNA gene sequences (Fig. 1). The *Acetobacterium* genus forms a distinct subgroup with cluster XV of the *Clostridium* subphylum (36). Notably, Willems and Collins (36) found that the psychrophilic strains were not more closely related to one another than were nonpsychrophilic strains—our findings suggest the opposite, which is in agreement with the work of Shin and coworkers (29) and shows distinct clustering of known psychrophiles *A. tundrae*, *A. paludosum*, *A. fimetarium*, and *A. bakii* (Fig. 1). Other *Acetobacterium* strains with sequenced genomes clustered close together, including *A. wieringae* and *Acetobacterium* sp. strain MES1.

Genome characteristics of the *Acetobacterium* genus. To date, there are 14 publicly available sequenced representatives of the *Acetobacterium* genus. Thirteen of the 14 sequenced *Acetobacterium* strains are draft genome assemblies, with the genome of *Acetobacterium woodii* as the only complete and closed genome (contained in

one contiguous sequence). The 13 draft genomes contained a varying number of contigs ranging from 62 (*A. wieringae*) to 1,582 (*Acetobacterium* sp. strain UBA11218). The genome sizes ranged from ~2.2 Mbp (*Acetobacterium* sp. strain UBA11218) to ~4.1 Mbp (*A. bakii*), and the GC% ranged from 39.7 (*A. tundrae*) to 44.8 (*Acetobacterium* sp. strain UBA11218) (Table 1). Each genome, including *A. woodii*, was examined for completeness based upon the *Eubacteriaceae* marker gene set from CheckM (37). *Acetobacterium woodii* and *Acetobacterium* sp. strain KB-1 were 100% complete. The remaining 12 *Acetobacterium* genomes/MAGs were >98% complete with the exception of *Acetobacterium* sp. UBA6819 (94% complete, missing 11 marker genes), *Acetobacterium* sp. strain UBA5558 (85% complete, missing 31 marker genes), and *Acetobacterium* sp. UBA11218 (57% complete, missing 93 marker genes). Furthermore, *Acetobacterium* sp. UBA5558 was only 2.61 Mbp, well below the average genome size (3.69 Mbp), and had a genome coverage of 13×. Likewise, *Acetobacterium* sp. UBA11218 was only 2.22 Mbp with 6.1× genome coverage. The use of nonclosed genomes/MAGs provides access to a larger and potentially more diverse genome set, but gaps between contigs and missing genes present potential errors in analyses. Therefore, the three most incomplete MAGs were excluded from our pan-genome analyses in order to more accurately describe the core components of the genus, while minimizing erroneous results, and all genomes/MAGs, excluding *Acetobacterium* sp. UBA11218 (57% complete), were included in specific pathway identification (e.g., WLP, Rnf, hydrogenase).

To further examine the relatedness of the *Acetobacterium* genus, we calculated the pairwise average nucleotide identity (ANI) and average amino acid identity (AAI) for each genome and constructed phylogenies based on whole-genome alignments (Fig. 2). The closest species to *Acetobacterium woodii* were *A. malicum*, *A. dehalogenans*, and *Acetobacterium* sp. KB-1, with average nucleotide identity (ANI) and average amino acid identity (AAI) of 80% and 79%, respectively. While these represent distinct species, many of the sequenced *Acetobacterium* genomes were highly similar to one another, potentially representing subspecies or strains. Specifically, *Acetobacterium* sp. strain UBA6819 was closely related to *Acetobacterium* sp. KB-1 (99% ANI); *A. malicum* and *A. dehalogenans* (97% ANI) could be considered different strains of the same species; *Acetobacterium* sp. MES1, *A. wieringae*, *Acetobacterium* sp. UBA5558, and *Acetobacterium* sp. strain UBA5834 are all likely the same species (97 to 100% ANI); and *A. tundrae* and *A. paludosum* (95% ANI) had high sequence similarity across their genomes (Fig. 2A) (38). These findings suggest that, for example, three potential subspecies groups exist (brackets in Fig. 2A). While many of the genomes/MAGs were above the suggested species-level cutoff for ANI of 95% (39–41), we cannot definitively define particular genomes/MAGs as subspecies, and phenotypic characterization should accompany genomic data to confirm these findings.

The *Acetobacterium* pan-genome. Utilizing predicted protein sequences, whole-genome comparisons between *Acetobacterium* genomes were performed. While the presence of a particular pathway does not definitively prescribe function, genetic comparisons reveal evolutionary conservation or divergence and provide valuable information about the functional potential encoded in each genome. Pan-genome analyses utilized gene families or amino acid sequences that clustered together at a defined sequence similarity cutoff. We identified the common set of gene families (core genome), the unique set of gene families (unique genome), and the ancillary set of gene families, which were found in at least two, but not all, genomes (accessory genome). In order to obtain the most accurate pan-genome and core genome estimates, only genomes with a >98% completion were utilized (11 in total).

Whole-genome comparisons validated phylogenetic distances calculated using 16S rRNA genes and ANI/AAI groupings (Fig. 2), though phylogenomic relationships varied from marker gene clustering. For example, *A. fimetarium* was relatively closer to the psychrophilic strains in the pan-genome tree and ANI/AAI groupings, while *A. bakii* was relatively closer to the psychrophilic strains than *A. fimetarium* in the 16S rRNA tree. While 16S rRNA gene-based phylogeny has been extensively used for the *Acetobacte-*

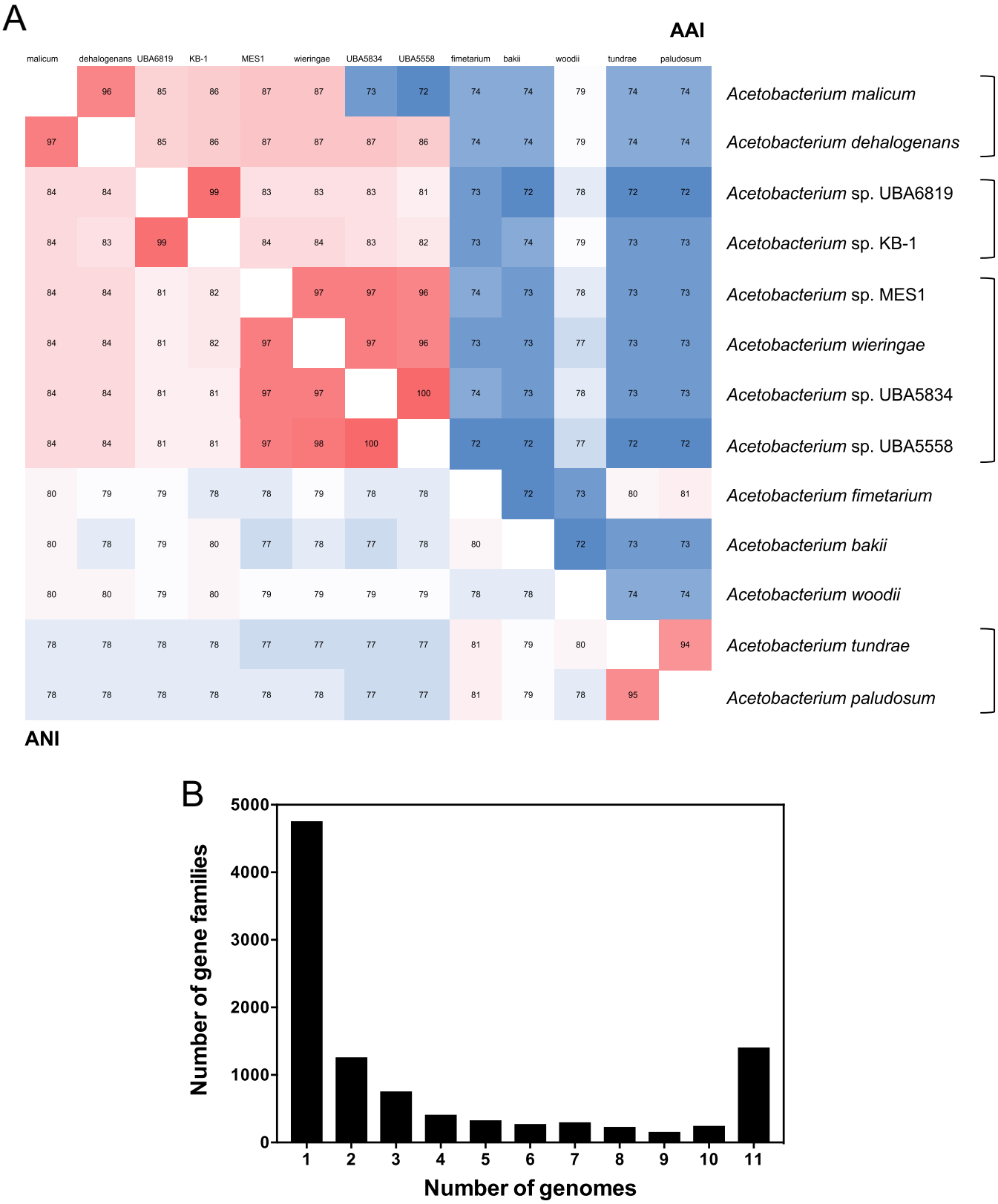

FIG 2 *Acetobacterium* pan-genome. (A) ANI/AAI matrix of 13 sequenced *Acetobacterium* isolate genomes and MAGs. Potential subspecies are denoted in brackets. (B) Pan-genome distribution of gene families found in 1, some (2 to 10), or all genomes (determined using predicted amino acid sequences). (C) Pan-genome phylogeny of 11 *Acetobacterium* strains. Potential subspecies denoted in brackets. Blue bar = presence of Rnf2 complex, green bar = potential for 2,3-butanediol metabolism, purple bar = *A. woodii*-type WLP cluster I, black bar = potential for methanol oxidation, orange bar = potential for caffeate metabolism, red bar = potential for ethanol oxidation. (D) COG classification of predicted protein sequences of the core genome, accessory genome, and unique genome of 11 *Acetobacterium* strains.

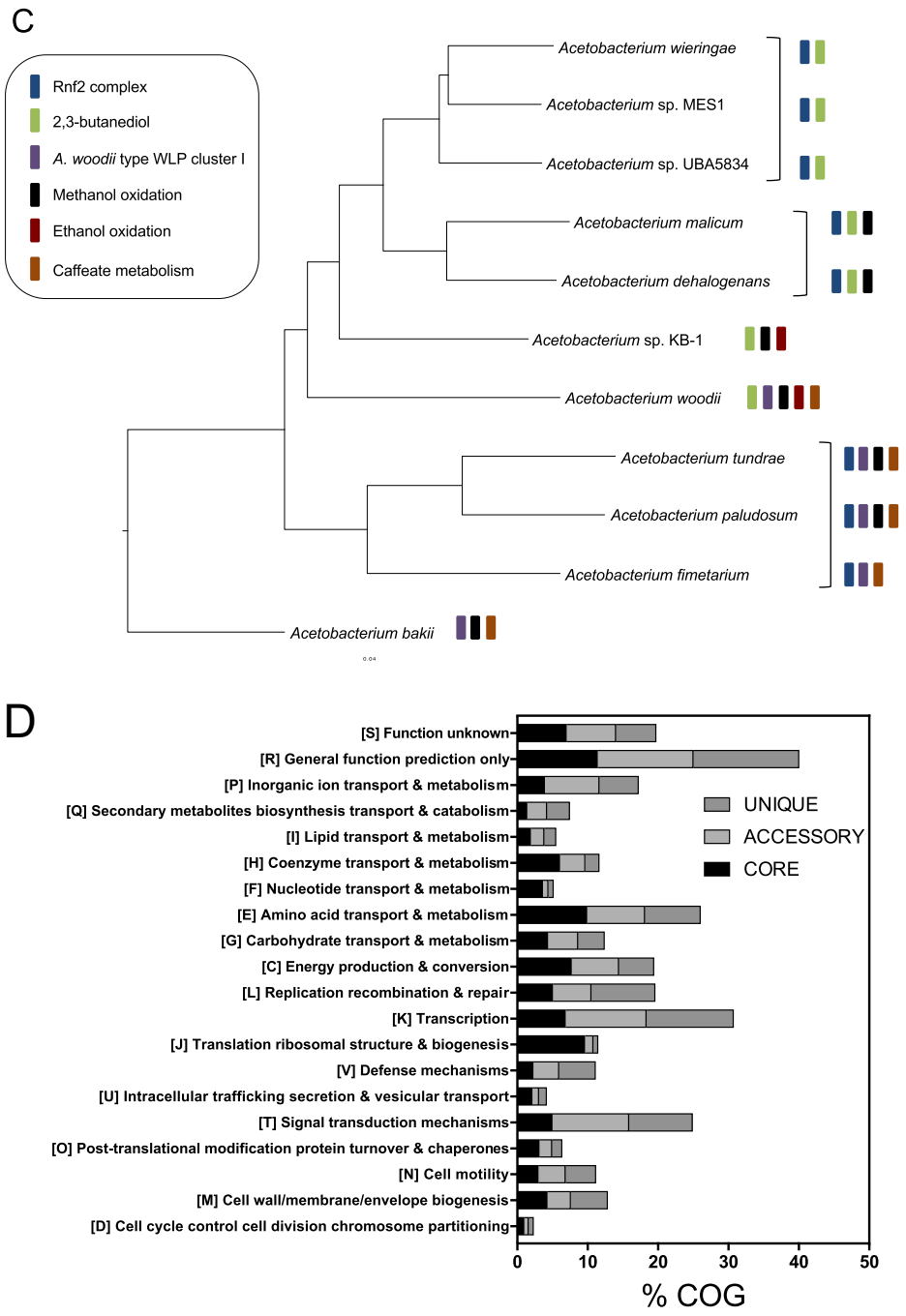

FIG 2 (Continued)

rium genus, improved phylogenetic resolution is achievable with our whole-genome analyses (42, 43).

Inclusion of the 11 most complete *Acetobacterium* genomes revealed a total of 10,126 gene families (40,203 total amino acid sequences). When calculated as the total number of amino acid sequences per category, the percentage of core sequences was 38.5%, while the accessory, unique, and exclusively absent sequences were 43.6%, 11.9%, and 6.1%, respectively. Within the pan-genome, the core genome, accessory genome, and unique genome contained a total of 1,406 (13.9%), 3,964 (39.1%), and 4,756 (47.0%) gene families, respectively (Fig. 2B and Table 2). The pan-genome partitioning of gene families suggests that the unique genome contains the most

TABLE 2 Pan-genome distribution of the 11 most complete *Acetobacterium* genomes

Genome	No. of gene families			
	Core	Accessory	Unique	Exclusively absent[a]
A. bakii	1,406	1,577	811	57
A. dehalogenans	1,406	1,814	358	0
A. fimetarium	1,406	1,070	510	83
Acetobacterium sp. KB-1	1,406	1,631	448	17
A. malicum	1,406	1,915	342	3
Acetobacterium sp. MES1	1,406	1,680	223	3
A. paludosum	1,406	1,497	483	1
A. tundrae	1,406	1,421	486	28
Acetobacterium sp. UBA5834	1,406	1,505	222	26
A. wieringae	1,406	1,860	253	0
A. woodii	1,406	1,548	620	27

[a]Exclusively absent gene families are those found in the denoted strain only.

functional diversity, followed by the accessory genome and then the core genome. Phylogenomic trees utilizing the pan-genome and core genome sequences revealed clustering of the psychrophilic strains *A. tundrae*, *A. paludosum*, and *A. fimetarium* (Fig. 2C). Other clusters included *Acetobacterium wieringae*, *Acetobacterium* sp. MES1, and *Acetobacterium* sp. UBA5834, and *A. malicum* and *A. dehalogenans*, consistent with the ANI and AAI results. Gene families unique to the psychrophilic clade or the *A. wieringae* clade were examined further (see Data Set S1 in the supplemental material).

Protein sequences from core, accessory, and unique gene families were classified with the Clusters of Orthologous Groups (COG) database (44). The core genome had the highest percentage of sequences from amino acid transport & metabolism (category E, 9.9%), translation, ribosomal structure & biogenesis (category J, 9.7%), and energy production and conversion (category C, 7.7%) (Fig. 2D). The highest percentages of sequences in the accessory and unique genomes were from transcription (category K, 11.5% and 12.4%, respectively) and signal transduction mechanisms (category T, 10.9% and 9.0%, respectively). The third most represented categories for the accessory and unique gene families were category E (amino acid transport and metabolism) and category L (replication, recombination, and repair), respectively.

The Wood-Ljungdahl (acetyl-CoA) carbon fixation pathway in *Acetobacterium* spp. The Wood-Ljungdahl pathway (WLP) (or the reductive acetyl-CoA pathway) is thought to be derived from an ancient carbon dioxide-fixing metabolic pathway (45). The WLP is coupled to substrate-level phosphorylation, but the net ATP gain is zero; thus, a chemiosmotic mechanism is employed for energy conservation (46). Generation of a transmembrane gradient is utilized by an F_1F_0 ATP synthase for ATP synthesis and relies upon either the *Rhodobacter* nitrogen fixation (Rnf) complex or the energy-converting ferredoxin-dependent hydrogenase complex (Ech) depending upon the species. *Acetobacterium woodii* has served as a model organism for understanding the mechanisms of energy conservation in Rnf-type acetogens (21, 22, 47–50), and we utilized the *A. woodii* genome as a template to examine the WLP of the other 12 *Acetobacterium* genomes/MAGs.

(i) Cluster I of the WLP is divergent across *Acetobacterium* spp. In *Acetobacterium woodii*, the WLP is organized into three separate and distinct gene clusters (Fig. 3) (21). Cluster I contains genes that encode the hydrogen-dependent CO_2 reductase (HDCR), which is the first step of the WLP methyl branch (CO_2 to formate) (51). Specifically, cluster I in *Acetobacterium woodii* encodes two formate dehydrogenase (FDH) isoenzymes (one of which is a selenocysteine-containing version), a formate dehydrogenase accessory protein, a putative FeS-containing electron transfer protein, and an [FeFe]-hydrogenase (FdhF1, HycB1, FdhF2, HycB2, FdhD, HycB3, and HydA2) (Fig. 3) (21). The presence, sequence similarity, and gene arrangement of cluster I were determined for the remaining 12 genomes/MAGs. Only *A. bakii*, *A. fimetarium*, *A. paludosum*, and *A. tundrae* encoded cluster I with high sequence similarity and a gene arrangement identical to *A. woodii* (Fig. 3; Data Set S2). Of these, *A. bakii* and *A. fimetarium* contain

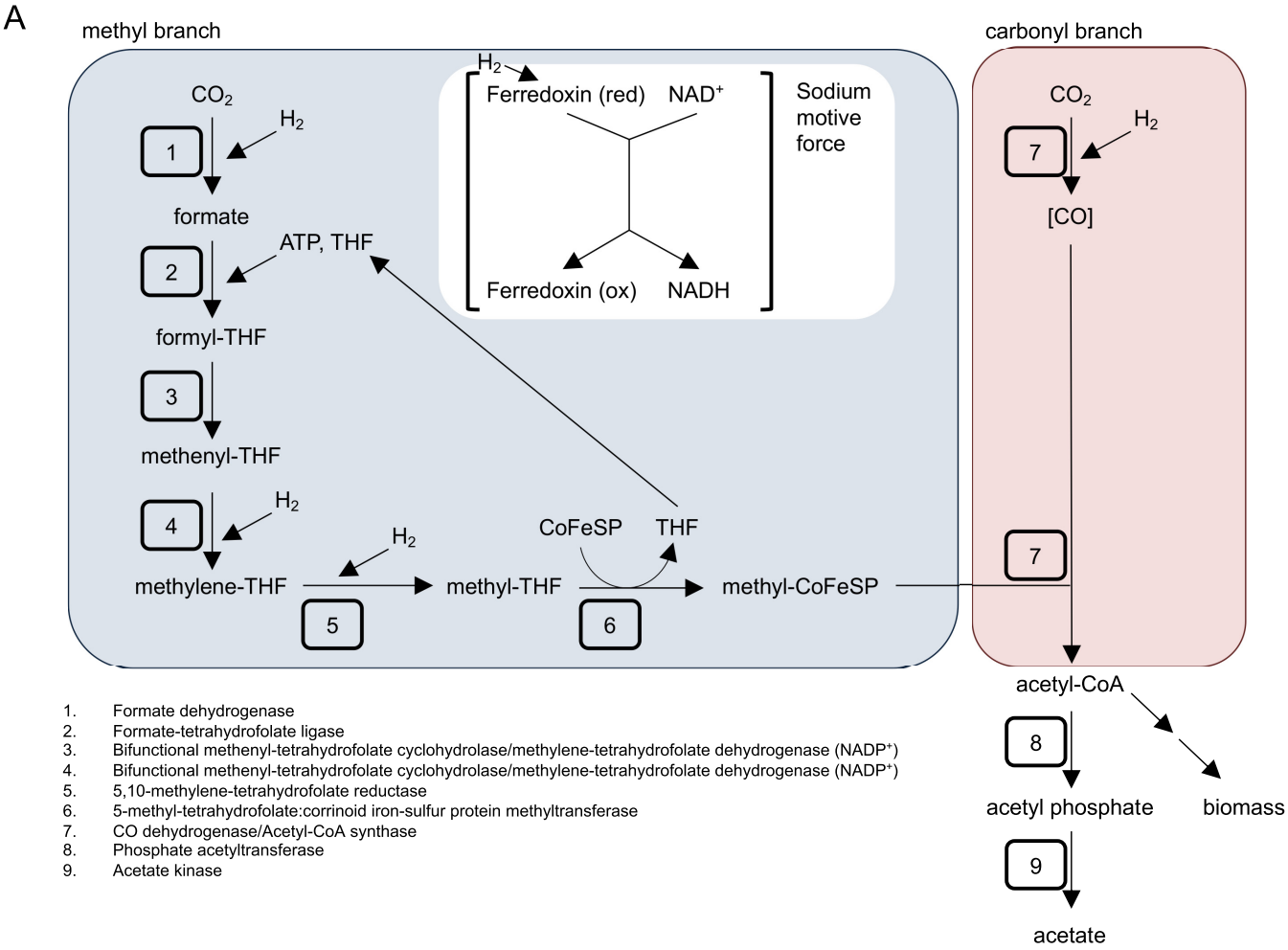

FIG 3 The Wood-Ljungdahl pathway (WLP). (A) Pathway diagram of the methyl and carbonyl branches of the WLP (note: H_2 is not directly used as a reductant). (B) Genetic representation of the WLP cluster I (formate dehydrogenase), cluster II (methyl branch), and cluster III (carbonyl branch). Lines denote contiguous sequences, while the absence of a line represents a discontinuous operon. The numbers represent the protein-encoding gene (peg) numbers generated by RAST. Consecutive numbers are synonymous with contiguous sequences. (C) Gene arrangement of cluster I of the Wood-Ljungdahl pathway. The formate dehydrogenase subunits are denoted in purple. (D) Phylogenetic tree of formate dehydrogenases from *Acetobacterium* spp. and closely related proteins. Four distinct FDH branches are highlighted, corresponding to the four types of FDH cluster I.

slight modifications of the gene cluster compared to *A. woodii*, where *A. bakii* encodes a truncated FdhF1, and *A. fimetarium* encodes only FdhF1 and not FdhF2.

Contrary to recent reports of the conservation of the WLP (38), we found that cluster I was markedly different and more divergent in the remaining genomes, with low sequence identity (<50%) and similarity (Fig. 3 and Data Set S2). Examination of formate dehydrogenase (FDH) protein alignments revealed approximately 220 extra amino acids at the beginning (N-terminal) of the non-*A. woodii*-type FDH that contained multiple conserved cysteine residues, with predicted [2Fe-2S] and [4Fe-4S] motifs (Table S2; Text S2). The genome architecture surrounding the formate dehydrogenase had identical synteny for various groups of strains with four distinct gene patterns, which corresponded to phylogenetic clustering of all FDH proteins from *A. woodii* and non-*A. woodii* FDH clusters (Fig. 3; Table S2). While the *A. woodii* FDH cluster I was quite different from the non-*A. woodii* FDH clusters, the non-*A. woodii* FDH cluster IA, for example, contained proteins with conserved residues for coordination of iron-sulfur clusters and molybdopterin cofactors, with similarities to the *A. woodii* HDCR (Table S2). Based upon the genome architecture surrounding each FDH, and phylogenetic placement of the FDH, we propose four types of FDH cluster I—the *A. woodii* FDH cluster I that contains FdhF1 and FdhF2, and three non-*A. woodii* FDH clusters (IA, IB,

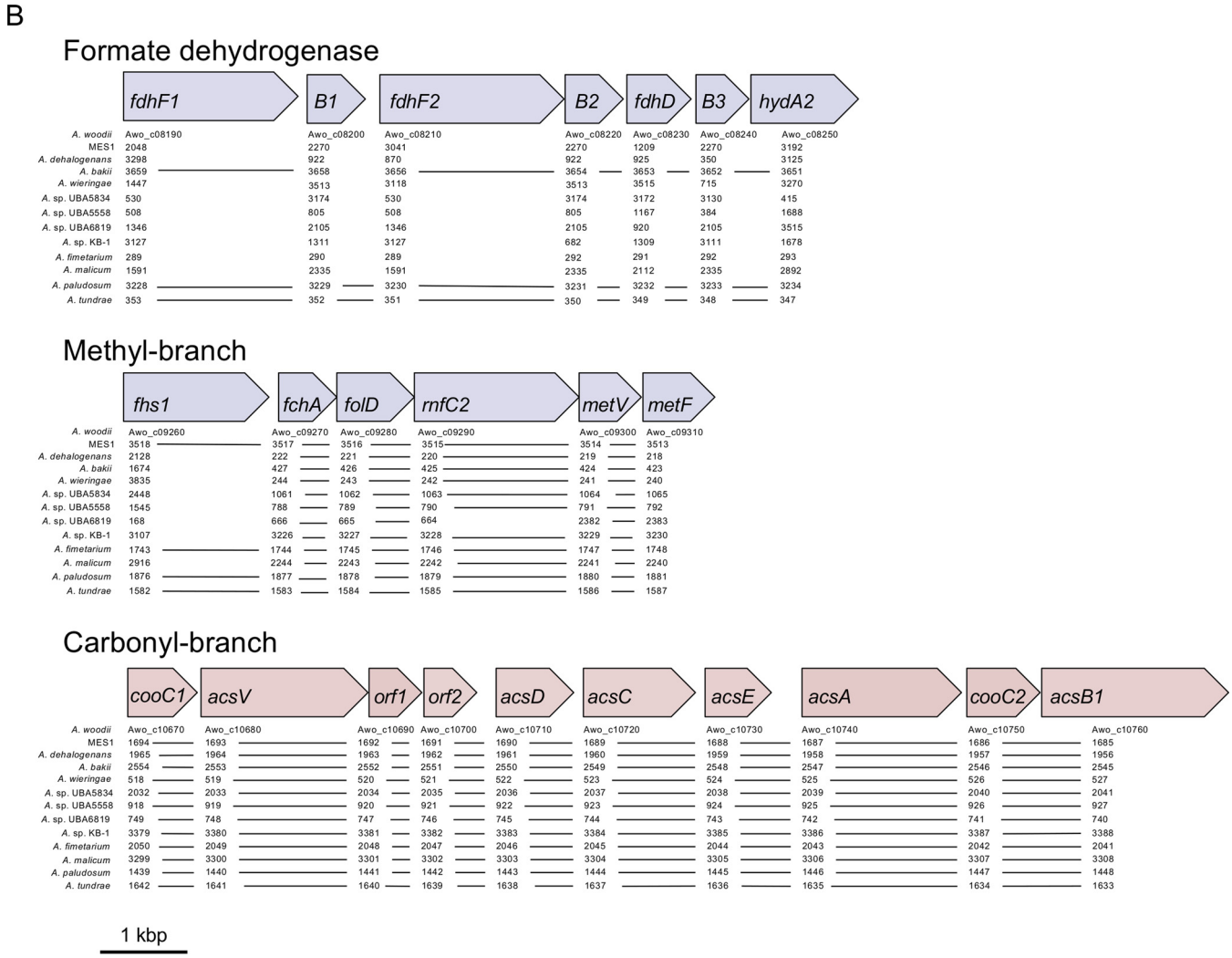

FIG 3 (Continued)

and IC) (Fig. 3). The *A. woodii* HDCR (*A. woodii* FDH cluster I) uses H_2 as an electron donor for CO_2 reduction (52). More work is needed to determine the electron donor for the non-*A. woodii* FDH clusters, but if H_2 is not used directly, other potential alternative electron donors, such as NAD(P)H and reduced ferredoxin, could be utilized in a mechanism similar to *Clostridium autoethanogenum* (53).

The variability in FDH could be a result of the availability and requirement for tungsten. In *Campylobacter jejuni* and *Eubacterium acidaminophilum*, formate dehydrogenase activity is dependent upon tungsten availability, as *tupA* mutants exhibited markedly decreased FDH activity (54, 55). We found that the strains encoding the non-*A. woodii* FDH cluster I also encode a tungsten-specific ABC transporter (*tupABC*). TupA has a high affinity for tungsten and can selectively bind tungsten over molybdate (56). The *tupABC* operon was absent in the strains encoding the *A. woodii* FDH cluster I, though the ability to transport tungsten in these strains may rely upon the *modABC* operon, as all sequenced genomes encode ModA, which is a high-affinity molybdate ABC transporter that binds both molybdate and tungstate (54, 57).

(ii) Conservation of WLP cluster II and cluster III of the *Acetobacterium* genus. Cluster II (methyl branch) encodes enzymes responsible for the conversion of formate to methyltetrahydrofolate (methyl-THF) (Fhs1, FchA, FolD, RnfC2, MetV, and MetF). Unlike cluster I, cluster II is highly conserved across all *Acetobacterium* genomes (Data Set S2). The exception was the dihydrolipoamide dehydrogenase (LpdA1) and glycine

C

A. woodii FDH Cluster I [*A. woodii, A. bakii* (truncated Fdh1), *A. paludosum, A. fimetarium* (only 3-7), *A. tundrae*]

1. Formate dehydrogenase H (EC 1.2.1.2) selenocysteine-containing (Fdh1—AFA47611)
2. Quad-[4Fe-4S] ferredoxin, HycB/HydN/HyfA family (HycB1—AFA47612)
3. Formate dehydrogenase H (EC 1.2.1.2) selenocysteine-containing (Fdh2—AFA47613)
4. Quad-[4Fe-4S] ferredoxin, HycB/HydN/HyfA family (HycB2—AFA37614)
5. Sulfur carrier protein FdhD (FdhD—AFA47615)
6. Quad-[4Fe-4S] ferredoxin, HycB/HydN/HyfA family (HycB3—AFA47616)
7. Periplasmic [FeFe] hydrogenase large subunit (EC 1.12.7.2) (HydA2—AFA47617)

non-*A. woodii* FDH Cluster IA [*A. wieringae, A.* sp. MES1, *A.* sp. UBA5558, *A.* sp. UBA5834, *A. dehalogenans*]

1. NAD-reducing hydrogenase subunit HoxE (EC 1.12.1.2) [2Fe-2S]
2. Electron bifurcating [FeFe] hydrogenase subunit C [2Fe-2S]
3. Electron bifurcating [FeFe] hydrogenase subunit B [2Fe-2S] [4Fe-4S] [Fer4_7]
4. Formate dehydrogenase alpha subunit [4Fe-4S] [molybdopterin cofactor]
5. Formylmethanofuran dehydrogenase associated protein FmdE/Molybdopterin molybdenumtransferase (EC 2.10.1.1)
6. Formate dehydrogenase chain D

non-*A. woodii* FDH Cluster IB [*A.* sp. KB-1, *A.* sp. UBA6819]

1. Hypothetical protein
2. Cadmium efflux system accessory protein
3. Lead, cadmium, zinc and mercury transporting ATPase (EC 3.6.3.3) (EC 3.6.3.5)/Copper-translocating P-type ATPase (EC 3.6.3.4)
4. Formate dehydrogenase alpha subunit
5. Hypothetical protein
5. Uncharacterized corrinoid protein
6. Uncharacterized corrinoid protein

non-*A. woodii* FDH Cluster IC [*A. malicum A. wieringae, A.* sp. MES1, *A.* sp. UBA5558, *A.* sp. UBA5834, *A. dehalogenans, A. bakii* (missing MobA)]

1. Formate dehydrogenase alpha subunit
2. Molybdopterin-guanine dinucleotide biosynthesis protein MobA
3. Hypothetical protein (TrkA-C domain)
4. tRNA-Leu-TAG
5. tRNA-Gly-GCC
6. tRNA-Arg-TCG
7. nucleoid-associated protein
8. type-I 3-dehydroquinate dehydratase

D

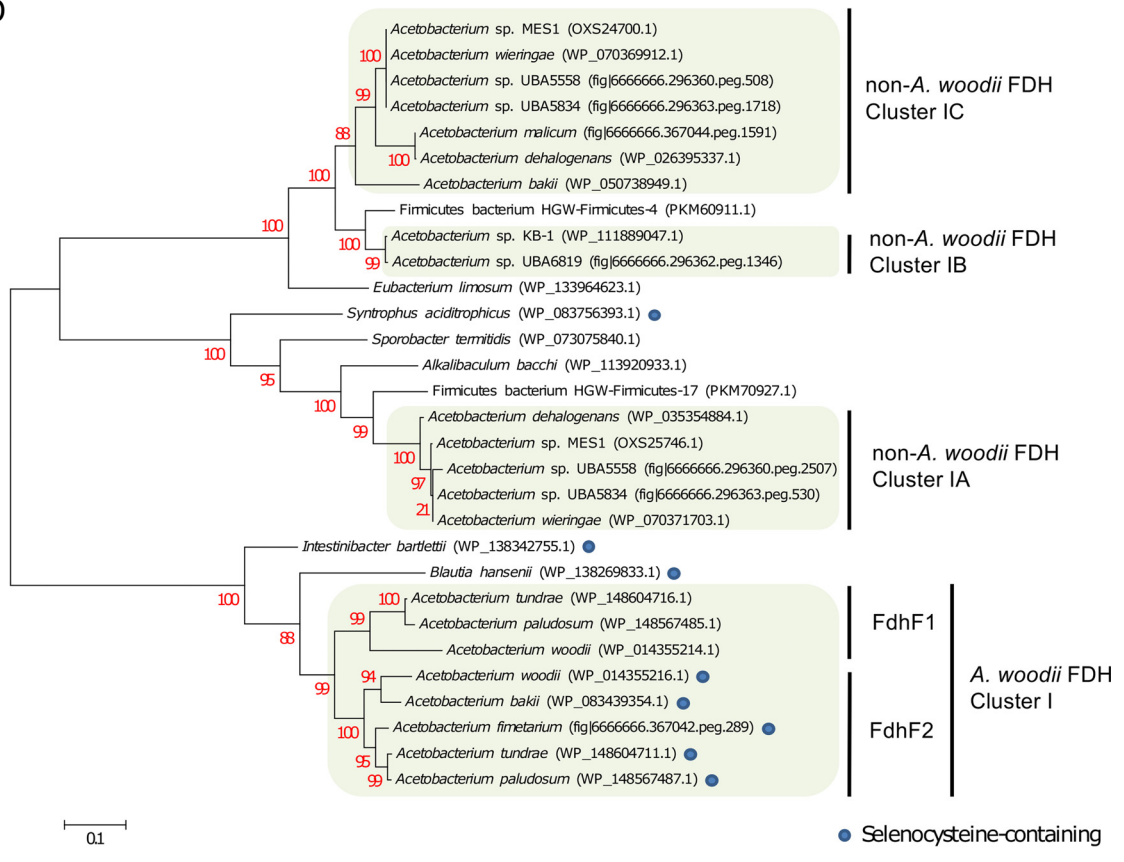

FIG 3 (Continued)

cleavage system H protein (GcvH1), which is found downstream of methylene-tetrahydrofolate reductase (MetF) in *A. woodii* but dispersed throughout the genomes of the other *Acetobacterium* strains (Fig. 3). Other variations in cluster II are found in *A. tundrae* and *A. paludosum*, which contain two formate tetrahydrofolate-ligase (Fhs1) subunits separated by a small hypothetical protein (39 amino acids [aa]). In *A. paludosum* the two copies are identical, while in *A. tundrae* there is one amino acid difference between the two Fhs1 sequences, suggesting a gene duplication event in these two strains.

Cluster III (carbonyl branch) encodes the CO dehydrogenase/acetyl-CoA synthase complex (AcsABCD), methyltransferase (AcsE), and accessory proteins (CooC and AcsV), for conversion of methyl-THF to acetyl-CoA. The carbonyl branch is the most well conserved portion of the WLP in the *Acetobacterium* genus, with high sequence similarity and identical gene arrangement across all genomes—the one exception is a hypothetical protein and an additional AcsB directly downstream of AcsB in *A. bakii*, *A. dehalogenans*, *A. paludosum*, *A. tundrae*, *A. fimetarium*, *A. malicum*, and *Acetobacterium* sp. KB-1 (Fig. 3) (29).

Genes essential for energy conservation. Energy conservation in acetogens requires the generation of a sodium or proton gradient across the cytoplasmic membrane, with an energy-converting ferredoxin:NAD^+ reductase complex (known as the *Rhodobacter* nitrogen fixation [Rnf] complex) or energy-converting ferredoxin-dependent hydrogenase complex (Ech) and a sodium-dependent or proton-dependent ATPase involved (1). In *Acetobacterium woodii*, *Acetobacterium bakii*, and potentially all other members of the genus, a sodium ion gradient is generated via the Rnf and sodium-dependent ATPase.

(i) Conservation of the ferredoxin-NAD^+ oxidoreductase (Rnf complex) and variability in Rnf copy number. The multisubunit integral membrane ferredoxin-NAD^+ oxidoreductase called the *Rhodobacter* nitrogen fixation (Rnf) complex is an energy-coupled transhydrogenase responsible for transfer of electrons from reduced ferredoxin to NAD^+, which in *Acetobacterium woodii* generates a transmembrane sodium ion gradient to drive ATP generation via sodium-dependent ATP synthase (58–60). This process is reversible when the concentration of NADH is greater than that of ferredoxin (22). The Rnf cluster is common across many *Clostridia* including *Clostridium tetani*, *Clostridium kluyveri*, *Clostridium difficile*, *Clostridium phytofermentans*, and *Clostridium botulinum* (47); however, the *Clostridia* type is not sodium dependent and may involve a transmembrane proton gradient (61). While gene arrangement within the Rnf complex differs across prokaryotes, in *Acetobacterium woodii* the genes of the Rnf complex (*rnfCDGEAB*) are polycistronic and cotranscribed (59). All sequenced *Acetobacterium* genomes encode an Rnf complex (Data Set S2). Based upon the presence and conservation of the Rnf complex in sequenced *Acetobacterium* spp., we propose that energy conservation mechanisms are similar across the genus and that the Rnf complex is required.

Microorganisms, such as *Azotobacter vinelandii*, encode two Rnf complexes, one of which is linked to nitrogen fixation and the other of which is expressed independently of nitrogen source (62). *Acetobacterium woodii* encodes only a single Rnf complex (21), but *Acetobacterium wieringae* and *Acetobacterium* sp. MES1 were found to encode two (14, 63, 64). The function of the second Rnf complex in *Acetobacterium* is unknown, but closer examination of the remaining *Acetobacterium* genomes revealed that the majority of *Acetobacterium* genomes encode a second *rnfCDGEAB* cluster (Data Set S2). The genome architecture surrounding the second Rnf complex was also well conserved, with a predicted hydroxymethylpyrimidine transporter, hydroxyethylthiazole kinase, and thiamine-phosphate pyrophosphorylase directly adjacent to the Rnf2 complex (Fig. 4; Data Set S2). Phylogenetic analysis of the two Rnf complexes revealed two distinct clusters: the first Rnf complex clusters with *Eubacterium limosum*, and the second Rnf complex forms a distinct clade with the *Clostridia* Rnf complex (Fig. 4).

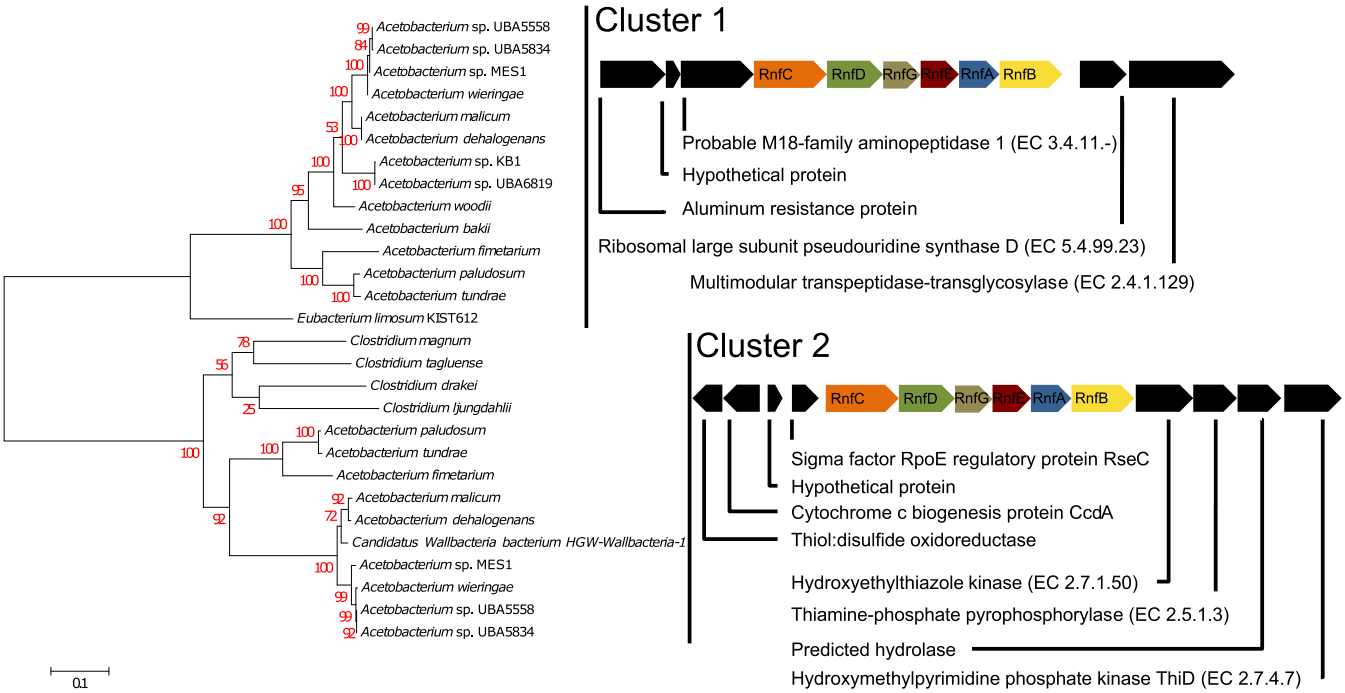

FIG 4 Phylogenetic tree of the RnfCDGEAB protein complex. Predicted protein sequences were concatenated and clustered with MUSCLE (105), and the phylogenomic tree was constructed with the Maximum Likelihood method using the bootstrap method with 1,000 replications for test of phylogeny. The substitution method was Jones-Taylor-Thornton with uniform rates among sites. Clustering and tree construction were performed with MEGA 6.06 (104). Varied genome architecture of predicted protein sequences immediately surrounding each Rnf operon is shown (not drawn to scale).

Sodium pumping by the Rnf complex has been extensively studied in *A. woodii*, and sodium-dependent autotrophic growth has been experimentally verified in *A. bakii*, suggesting a role of the Rnf complex similar to that of *A. woodii* (29, 58, 59). In *A. woodii*, predicted Na^+ translocating amino acid residues were found in the transmembrane subunits RnfA (E88), RnfD (D250 and G300), and RnfE (D129) (59). It is not known if the other *Acetobacterium* spp. utilize a Na^+ gradient or H^+ gradient for energy conservation. To probe this question, we examined the amino acid sequence similarity between the Rnf complexes and found that despite sequence divergence, amino acid residues predicted to be involved in Na^+ binding were conserved across all *Acetobacterium* Rnf complexes. For example, RnfA1 and RnfA2 from *Acetobacterium* sp. MES1 were only 51% identical with 73% positives, yet both retained the conserved glutamic acid (E88). Similarly, RnfE1 and RnfE2 were divergent (53% identity, 69% positives), but aspartic acid at position 129 (D129) was conserved. While experimental validation is needed, we hypothesize that *Acetobacterium* Rnf complexes perform Na^+-specific pumping.

(ii) Conservation of the sodium-dependent ATPase. *Acetobacterium woodii* encodes an integral membrane Na^+ F_1F_0 ATP synthase that generates ATP via a sodium ion gradient and contains both V-type and F-type rotor subunits (65). All sequenced *Acetobacterium* genomes encode the F-type and V-type ATP synthase (Data Set S2), each contained in a separate operon and with amino acid sequence similarity to *Acetobacterium woodii* ranging from 56% (AtpB, *A. fimetarium*) to 98% (AtpE2, *A. bakii*). In particular, the *c*-subunits of the F_1F_0 ATP synthase are responsible for ion translocation across the membrane. Using three *c*-subunits from *Acetobacterium woodii* (Awo_c02160 to -c02180), which is known to use a sodium ion gradient, we queried the remaining 12 *Acetobacterium* genomes for *c*-subunits and, specifically, the Na^+-binding motif. All strains encode at least two subunits with binding motifs specific for Na^+ (Fig. S1). The presence of the F-type and V-type ATP synthase operons, gene synteny, and sequence similarity across all strains, in addition to the presence of Na^+-binding motifs, suggests that all *Acetobacterium* strains utilize this ATP synthase for ATP production via a sodium gradient, similar to what is observed in *A. woodii* and *A. bakii*.

Electron bifurcation. The transmembrane sodium-ion gradient is generated by the Rnf complex, but a number of accompanying enzyme complexes, including an electron-bifurcating hydrogenase, ferredoxins, and electron-transfer flavoproteins (ETF), are also essential for overcoming thermodynamically unfavorable reactions of the WLP (1).

(i) Conservation of the flavin-based electron-bifurcating hydrogenase HydABC. The key hydrogenase responsible for providing reducing equivalents for electron bifurcation in *Acetobacterium woodii* is encoded by *hydA1*, *hydB*, *hydD*, *hydE*, and *hydC* (Awo_c26970-Awo_c27010). This electron-bifurcating hydrogenase, HydABCDE, couples the thermodynamically unfavorable reduction of oxidized ferredoxin with the favorable reduction of NAD^+ (21). With the exception of *Acetobacterium* sp. UBA5558, HydABCDE was well conserved across the *Acetobacterium* genus (>77% identity and 85% coverage) (Data Set S2). A phylogenetic tree of concatenated proteins HydABCDE revealed clustering identical to that of the whole-genome tree, with tight clustering of *Acetobacterium* sp. MES, *Acetobacterium* sp. UBA5834, and *A. wieringae* and clustering of the psychrophilic strains (Fig. S2). This high degree of conservation across the genus highlights the essential function of this enzyme complex in *Acetobacterium* spp.

(ii) Variability in the number of ferredoxins encoded by *Acetobacterium* genomes. Ferredoxins are soluble cellular redox compounds and, in acetogens, play a major role in autotrophic CO_2 fixation (1). Of the electron carriers involved in the WLP, only ferredoxins have been demonstrated to provide the reducing potential required for the reduction of CO_2 to CO in the carbonyl branch. The genome of *Acetobacterium woodii* encodes 17 ferredoxins, the most of any *Acetobacterium* genome to date. The number of ferredoxins encoded by the other *Acetobacterium* genomes ranged from 6 (*Acetobacterium* sp. KB-1 and *Acetobacterium* sp. UBA6819) to 16 (*A. bakii*). Specifically, *A. malicum* and *A. dehalogenans* encode 12 ferredoxins, members of the *A. wieringae* clade encoded 8 to 10, and *Acetobacterium* sp. KB-1 and *Acetobacterium* sp. UBA6189 encoded six. The psychrophilic clade was more varied, as *A. fimetarium* encoded 8, while *A. paludosum* encoded 11, *A. tundrae* encoded 13, and *A. bakii* encoded 16. The overall variability in the number of ferredoxins encoded by each genome may be a casualty of loss of other things in the genome or an indication that it has less or more specific electron transport functions to carry out.

(iii) Variability in electron transfer flavoprotein (ETF) copy number. Electron transfer flavoprotein (EtfAB) is the electron bifurcation module in *Acetobacterium woodii* (66). *A. woodii* encodes two separate EtfAB modules. One is adjacent to CarABC and part of the caffeate reduction operon (CarDE), while the other is directly upstream of a lactate dehydrogenase (LctD), a potential L-lactate permease (LctE), and a potential lactate racemase (LctF) and involved in lactate metabolism (67). Examination of the remaining 12 genomes revealed that, with the exception of *Acetobacterium* sp. UBA5558, all genomes encode at least one copy of EtfAB. Specifically, *A. tundrae* encodes four copies of EtfAB; *A. bakii* and *A. paludosum* encode three copies; *A. woodii*, *A. malicum*, *A. fimetarium*, and *A. dehalogenans* each encode two copies; and the remaining strains have a single copy. The gene architecture downstream of the single EtfAB module encoded by all strains (except *Acetobacterium* sp. UBA5558) was identical and included machinery for lactate metabolism. The four EtfAB modules of *A. tundrae* were not well conserved, with amino acid identities of EtfA ranging from 39% to 55% and of EtfB ranging from 43 to 48%, suggesting they are not duplications but potentially four different gene transfer events. Interestingly, one of the EtfAB modules found only in *A. tundrae*, *A. paludosum*, and *A. bakii* was adjacent to a butyryl-CoA dehydrogenase (acyl-CoA dehydrogenase), an L-carnitine dehydratase (CoA transferase), and a butyryl-CoA dehydrogenase (acyl-CoA dehydrogenase), suggesting that other electron acceptors besides caffeate can be utilized—potentially crotonyl-CoA (68).

Other genome features: alternative electron donors/acceptors utilized by the *Acetobacterium* genus. *Acetobacterium* species can use a wide range of substrates for carbon and energy, indicating a generalist lifestyle for the genus despite the perception

TABLE 3 Growth phenotypes and predicted metabolisms of *Acetobacterium* spp.[a]

	A. woodii	*A. bakii*	*A. fimetarium*	*A. paludosum*	*A. tundrae*	*A. wieringae*	*A. malicum*	*A. dehalogenans*		*A.* sp. MES1	*A.* sp. UBA5558	*A.* sp. UBA5834	*A.* sp. UBA6819	*A.* sp. KB-1	Pathway/enzyme(s)
	Cultured isolates									MAGs					
H_2/CO_2	+	+	+	+	+	+	+	+							Wood-Ljungdahl Pathway (WLP)
CO	+	+	+	+	+	ND	ND	+							WLP (CO-dehydrogenase)
Formate	+	+	+	+	+	+	+	+							hydrogen-dependent CO_2 reductase (HDCR)/formate dehydrogenase
Fructose	+	+	+	+	+	+	+	+							Glycolysis
Glucose	+	+	–	+	+	–	–	+							Glycolysis
Methanol	+	+	–	+	+	–	–	+							MtaRC1WXYBC2A
Ethanol	+	–	–	–	–	+	–	+							bifunctional acetaldehyde-CoA/alcohol dehydrogenase
1,2-propanediol	+	ND	ND	ND	ND	+	+	ND							PduABCDEGHKLJNOPST
2,3-Butanediol	+	–	+	–	–	–	–	ND							TPP-dependent acetoin dehydrogenase
Lactate	+	+	+	+	+	+	+	+							lactate dehydrogenase/lactate permease/lactate racemase (LctABCDEF)
Glycine betaine	+	+	+	+	+	+	+	ND							MttA2/OpuD2/MttB10/MttC6/OpuD3
Alanine	+														alanine dehydrogenase
Caffeate	+														CarABCDE

[a]Growth phenotypes were compiled from experimental data with observed growth (+), no growth (−), or not determined (ND) (reference 109 and references therein). Metabolisms were predicted as present (dark blue) or absent (yellow). Light blue represents pathways that are complete but in which one or two predicted proteins have low sequence identity. Dark boxes represent inconsistencies between experimentally validated phenotypes and predicted metabolisms from genome data. Growth phenotypes are unavailable for the five metagenome-assembled genomes on the right (*Acetobacterium* sp. MES1, *Acetobacterium* sp. UBA5558, *Acetobacterium* sp. UBA5834, *Acetobacterium* sp. UBA6819, and *Acetobacterium* sp. KB-1).

that acetogens are specialists (Table 3). This could provide an advantage in anoxic environments where competition for limited substrates is high (69, 70). As an example, encoded in the genomes of most *Acetobacterium* species are carbon utilization pathways for 1,2-propanediol, 2,3-butanediol, ethanol, lactate, alanine, methanol, glucose, fructose, and glycine betaine, and utilization of various electron acceptors, including caffeate (Data Set S2). Combining data of growth phenotypes of *Acetobacterium* isolates with *in silico* genomic evidence of metabolic pathways, we hypothesize predicted metabolisms of the *Acetobacterium* genus, in particular strains that have not been isolated (MAGs) or isolates that have yet to be tested for a particular phenotype (Table 3; Data Set S3).

(i) 1,2-Propanediol. Acetogens metabolize alcohols as an alternative to autotrophic acetogenic growth, including *Acetobacterium carbinolicum*, *Acetobacterium woodii*, and *Acetobacterium wieringae* (71). *Acetobacterium woodii* can utilize 1,2-propanediol as the sole carbon and energy source for growth (27). In anaerobic environments, formation of 1,2-propanediol results from the degradation of fucose and rhamnose, constituents of bacterial exopolysaccharides and plant cell walls (72). The 1,2-propanediol degradation pathway is encoded by the *pduABCDEGHKL* gene cluster, which in *A. woodii* (Awo_c25930 to Awo_c25740) contains 20 genes with similarity to the *pdu* cluster of *Salmonella enterica* (27). The presence, homology, and gene arrangement of the *pdu* gene cluster in each of the other 12 *Acetobacterium* genomes suggest that 1,2-propanediol degradation is conserved across the *Acetobacterium* genus (Data Set S2). Furthermore, all strains contain a histidine kinase and response regulator upstream of the 1,2-propanediol cluster, suggesting that the regulatory and expression mechanisms of this pathway are conserved. In *A. woodii*, it was proposed to sense alcohols (chain length >2) or aldehyde intermediates—a mechanism we hypothesize is employed by all *Acetobacterium* species.

(ii) 2,3-Butanediol. The 2,3-butanediol oxidation pathway is encoded by the *acoR-ABCL* operon in *Acetobacterium woodii*, and unlike 1,2-propanediol degradation, the WLP accepts reducing equivalents from 2,3-butanediol oxidation to generate acetate (26). AcoR is a putative transcriptional activator, AcoA is a thiamine PP_i (TPP)-dependent acetoin dehydrogenase (alpha subunit), AcoB is a TPP-dependent acetoin dehydroge-

nase (beta subunit), AcoC is a dihydrolipoamide acetyltransferase, and AcoL is a dihydrolipoamide dehydrogenase (26, 73, 74). Amino acid sequence homology and operon structure suggest that 2,3-butanediol oxidation is a common metabolism across the *Acetobacterium* genus, with the exception of *A. fimetarium*, *A. paludosum*, and *A. tundrae* (Data Set S2). *A. bakii* encodes this cluster but shows low percent identity (<50%) and positives (<66%) for the AcoR subunit, AcoC subunit, and AcoL subunit. The genome architecture surrounding the 2,3-butanediol operon showed slight variations in some strains, as *A. bakii* has an additional 2,3-butanediol dehydrogenase (WP_050738536.1) and a small hypothetical protein (WP_050738535.1) between AcoR and AcoA (Data Set S2). Likewise, *A. malicum*, *A. dehalogenans*, *Acetobacterium* sp. MES1, *A. wieringae*, *Acetobacterium* sp. UBA5558, and *Acetobacterium* sp. UBA5834 encode a small hypothetical protein between *acoR* and *acoA*.

(iii) Ethanol. In acetogens, the oxidation of primary aliphatic alcohols, such as ethanol, is coupled to the reduction of CO_2 (75). The ethanol oxidation pathway has been elucidated in *A. woodii* and contains a bifunctional acetaldehyde-CoA/alcohol dehydrogenase (AdhE) that is upregulated during growth on ethanol (23), suggesting this was the primary enzyme responsible for ethanol oxidation. Examination of the other sequenced *Acetobacterium* genomes revealed *Acetobacterium* sp. KB-1 and *Acetobacterium* sp. UBA6819 were the only strains to encode a complete AdhE with 87% sequence identity to *A. woodii* (Data Set S2). Despite the lack of AdhE in the other species of *Acetobacterium*, species like *A. wieringae* are capable of growth on ethanol (Table 3) (64, 71, 75–77). It is likely that *A. wieringae* and the other *Acetobacterium* strains that do not encode the *A. woodii*-type bifunctional acetaldehyde-CoA/alcohol dehydrogenase employ an alternative pathway for ethanol oxidation. Biochemical analyses are needed to confirm this phenotype, but one hypothesis is that the conversion of ethanol to acetate proceeds via acetaldehyde by the activities of alcohol dehydrogenase (converts ethanol to acetaldehyde) and aldehyde:ferredoxin oxidoreductase (converts acetaldehyde directly to acetate), similar to *Thermacetogenium phaeum* (78). The genome of *Acetobacterium* sp. MES1 encodes an alcohol dehydrogenase (OXS24683.1) upstream of a tungsten-containing aldehyde:ferredoxin oxidoreductase (AFO) (OXS24682.1), both of which are also found in the *A. wieringae* genome. *A. woodii* does encode a similar alcohol dehydrogenase (Adh3; alcohol dehydrogenase, iron-type [AFA47416]) but does not encode a similar tungsten-containing AFO.

(iv) Lactate. Utilization of lactate as a carbon and energy source has been observed in many *Acetobacterium* strains, including *A. woodii*, *A. wieringae*, *A. dehalogenans*, *A. carbinolicum*, *A. malicum*, *A. fimetarium*, *A. bakii*, *A. paludosum*, and *A. tundrae* (71, 79, 80). The protein machinery responsible for lactate metabolism is encoded by the *lctABCDEF* operon (67, 81). The operon includes a transcriptional regulator (LctA), an electron transfer flavoprotein beta (LctB) and alpha (LctC), lactate dehydrogenase (LctD), a potential L-lactate permease (LctE), and a potential lactate racemase (LctF). With the exception of *Acetobacterium* sp. UBA5558, all sequenced *Acetobacterium* strains encode the lactate operon (Data Set S2).

(v) Alanine. *A. woodii* encodes an alanine degradation pathway for utilization of alanine as a sole carbon and energy source (24). The alanine degradation pathway consists of a pyruvate:ferredoxin (flavodoxin) oxidoreductase (PFO) (AWO_RS12520), a sodium:alanine symporter family protein (AWO_RS12525), alanine dehydrogenase (AWO_RS12530), and the Lrp/AsaC family transcriptional regulator (AWO_RS12535). Examination of this operon revealed it was well conserved across the *Acetobacterium* genus, with the exception of *A. fimetarium*, *A. paludosum*, and *A. tundrae*, which lacked a sodium:alanine symporter, alanine dehydrogenase, and the transcriptional regulator, the details of which are discussed below (Data Set S2).

(vi) Caffeate. Some acetogenic bacteria utilize phenyl acrylates as alternative electron acceptors (82). One such phenyl acrylate, caffeate, is produced during lignin degradation and may be an available substrate in environments containing vegetation (83). Recently, it was observed that *A. woodii* has the ability to couple caffeate reduction

with ATP synthesis (25). The caffeate reduction operon in *A. woodii* is encoded by *carA2*, *carB2*, *carC*, *carD*, and *carE* (Awo_c15700 to Awo_c15740). Specifically, CarA is a hydrocaffeyl-CoA:caffeate CoA transferase (25), CarB is an ATP-dependent acyl-CoA synthetase (84), CarC is a caffeyl-CoA reductase, and CarDE is an electron transfer protein (49). Examination of the other 12 *Acetobacterium* genomes revealed BLAST hits with low similarity (<53% identity) with the exception of three psychrophilic strains (*A. paludosum*, *A. fimetarium*, and *A. tundrae*) and CarC, CarD, and CarE from *Acetobacterium bakii* (74%, 73%, and 73% amino acid sequence identity and 88%, 85%, and 86% amino acid sequence coverage, respectively) (Data Set S2). The lack of similar and congruous sequences suggests that only *A. woodii*, *A. fimetarium*, *A. paludosum*, and *A. tundrae* are capable of caffeate reduction via this pathway.

Potential adaptations for enhanced surface colonization. Many of the *Acetobacterium* genomes/MAGs encoded portions of the Widespread Colonization Island (WCI), which mediates nonspecific adherence to surfaces and biofilm formation (85). The products of the WCI are responsible for assembly and secretion of bundled pili (86) and may be important for colonization of diverse environments (87). The general structure of the WCI genomic region is similar for all sequenced *Acetobacterium* strains with the exception of *A. woodii* and includes 1 to 4 small hypothetical proteins (~56 aa), followed by TadZ, Von Willebrand factor type A, RcpC/CpaB, TadZ/CpaE, TadA/VirB11/CpaF, TadB, and TadC (Data Set S2). The small hypothetical proteins from *Acetobacterium* sp. MES1 show significant similarity (100% coverage, 98% identity) with the Flp/Fap pilin component from *A. wieringae* (OFV69504.1 to OFV69507.1). Likewise, immediately upstream of the Flp/Fap pilin components is a hypothetical protein with high similarity (100% coverage, 79% identity) to a prepilin peptidase from *A. dehalogenans* (WP_026393143.1). The presence of this potentially biofilm-enhancing WCI found in many of the species of *Acetobacterium* but not in *A. woodii* highlights the importance of expanding genome-informed biochemical analyses to species and strains beyond the type strain of a genus. One intriguing hypothesis resulting from the discovery of this WCI is that it may enhance surface attachment on insoluble electron donors like metallic iron or electrodes. A recent study suggests *A. malicum* and an isolate most closely related to *A. wieringae* more effectively extract electrons from solid Fe(0) coupons than does *A. woodii* (34). Additionally, *Acetobacterium* sp. MES1 is capable of colonizing cathodes in microbial electrosynthesis systems (14, 33), but *A. woodii* has failed in all such attempts (88, 89).

Closer examination of the strains that were phylogenetically most closely related to *Acetobacterium wieringae*, which include *Acetobacterium* sp. MES1 (capable of electroacetogenesis), revealed 42 proteins unique to this clade (Data Set S1). Of these unique protein sequences, several were of potential relevance for possible surface colonization and electron transport, including cardiolipin synthase for the production of membrane phospholipids, the global regulator diguanylate cyclase (Fig. S3), and methylenetetrahydrofolate reductase (methylene-THF reductase) (Data Set S1).

Potential adaptations to a psychrophilic lifestyle. Psychrophilic microorganisms have adapted to survive and grow in cold environments by varying membrane fluidity, optimizing transcription and translation (e.g., overexpression of RNA helicases, post-transcriptional regulation of RNA), and expressing cold shock proteins and cold-adapted enzymes (90). *Acetobacterium bakii* employs posttranscriptional regulation for cold adaptation, and at low temperatures a lipid biosynthesis pathway (ABAKI_c35860 to -c35970), cold shock protein CspL (ABAKI_c09820 and ABAKI_c26430), and a Dead-box helicase (ABAKI_c00160 to -c00180) were upregulated (29). Examination of these pathways across the *Acetobacterium* genus revealed the lipid biosynthesis pathway was conserved, but for the psychrophilic strains the genome architecture surrounding this cluster was different. Furthermore, two of the four psychrophilic strains (*A. tundrae* and *A. paludosum*) encoded twice as many cold shock proteins (4 total) as the other *Acetobacterium* strains (1 to 2 total).

Other minor variations were observed in the psychrophilic strains, in particular the sodium-dependent ATPase and 1,2-propanediol operon. The subunits of the sodium-dependent ATPase involved that constitute the membrane-bound motor were less well conserved than the catalytic subunits, perhaps indicative of the varying membranes across *Acetobacterium* (Data Set S2). Furthermore, *A. tundrae*, *A. paludosum*, and *A. fimetarium* were the only strains that do not encode the *c1* subunit. Minor variability in the 1,2-propanediol operon structures of three psychrophilic strains (*A. fimetarium*, *A. paludosum*, and *A. tundrae*) was evident, as a heme-binding protein (pfam03928) located between *pduO* and *pduP* was missing.

One adaptation to survival in cold environments that psychrophiles employ is to increase protein flexibility and stability (90). As a result, psychrophiles often contain a higher proportion of hydrophobic amino acid residues, such as alanine and glycine (91, 92). The psychrophilic strains are the only sequenced *Acetobacterium* strains to lack an alanine degradation pathway and glycine cleavage system. One possible explanation for the loss of these pathways in the psychrophilic *Acetobacterium* strains is an increased utilization of alanine and glycine in cold-adapted protein synthesis (91). Interestingly, the alanine degradation pathway encodes alanine dehydrogenase, which may also play an important role in NH_4^+ assimilation (93). Glutamate dehydrogenase, another enzyme shown to be important for nitrogen assimilation, was also absent from the psychrophilic strains (Data Set S2). Thus, the psychrophilic strains may have adapted to obtain nitrogen from alternative sources, using an alternative pathway.

To further determine what functional genome attributes may separate the psychrophilic *Acetobacterium* species from their nonpsychrophilic counterparts, we identified gene families unique to the psychrophilic species (Data Set S1). Our analyses revealed 40 annotated gene families found only in the psychrophilic clade and included a calcium-translocating P-type ATPase, sodium:proton antiporter, cupin, ribulose 1,5-bisphosphate carboxylase, and short-chain dehydrogenase (SDR) family oxidoreductase, among others (Data Set S1). The presence of a unique calcium-translocating P-type ATPase and sodium:proton antiporter suggests modified abilities to transport calcium and sodium. The 1,5-bisphosphate carboxylase (RuBisCO) and cupin may be involved in the methionine salvage pathway, which has been observed in other bacteria (94, 95). SDR family oxidoreductases have a wide range of activities, including metabolism of amino acids, cofactors, and carbohydrates, and may play a role in redox sensing (96). More work is needed to determine the exact function each protein plays in the psychrophilic strains, but these findings can be used as a framework for targeted functional analyses in future studies to determine specific roles these unique proteins play in the metabolisms of psychrophilic *Acetobacterium* strains and their persistence in cold environments.

Conclusions. Acetogens are a phylogenetically diverse group of microorganisms capable of converting CO_2 into acetate. *Acetobacterium woodii* has been used as a model organism to study the WLP and accessory components required for energy conservation in Rnf-type acetogens. We sequenced four *Acetobacterium* isolates from the culture collection (ATCC) (*A. fimetarium*, *A. malicum*, *A. paludosum*, and *A. tundrae*) and performed pan-genome analysis on the 11 most complete genomes/MAGs. We examined the functional potential of the available *Acetobacterium* genomes to shed light on the diverse genome attributes, gene arrangement and architecture, and potential metabolic capabilities of the *Acetobacterium* genus. Using the type strain of the genus (*A. woodii*) as a framework for pathway identification and comparison, we found the common and conserved pathways included the WLP and accessory components (electron-bifurcating hydrogenase, ATP synthase, Rnf complex, ferredoxin, and electron transfer flavoprotein [ETF]), glycolysis/gluconeogenesis, and 1,2-propanediol. Figure 5 presents a metabolic overview of the *Acetobacterium* genus. Interestingly, the hydrogen-dependent CO_2 reductase (HDCR) is unevenly distributed among the genus. Based upon presence/absence, gene arrangement, and sequence similarity of many pathways, we hypothesize that divergent metabolisms found in a subset of genomes

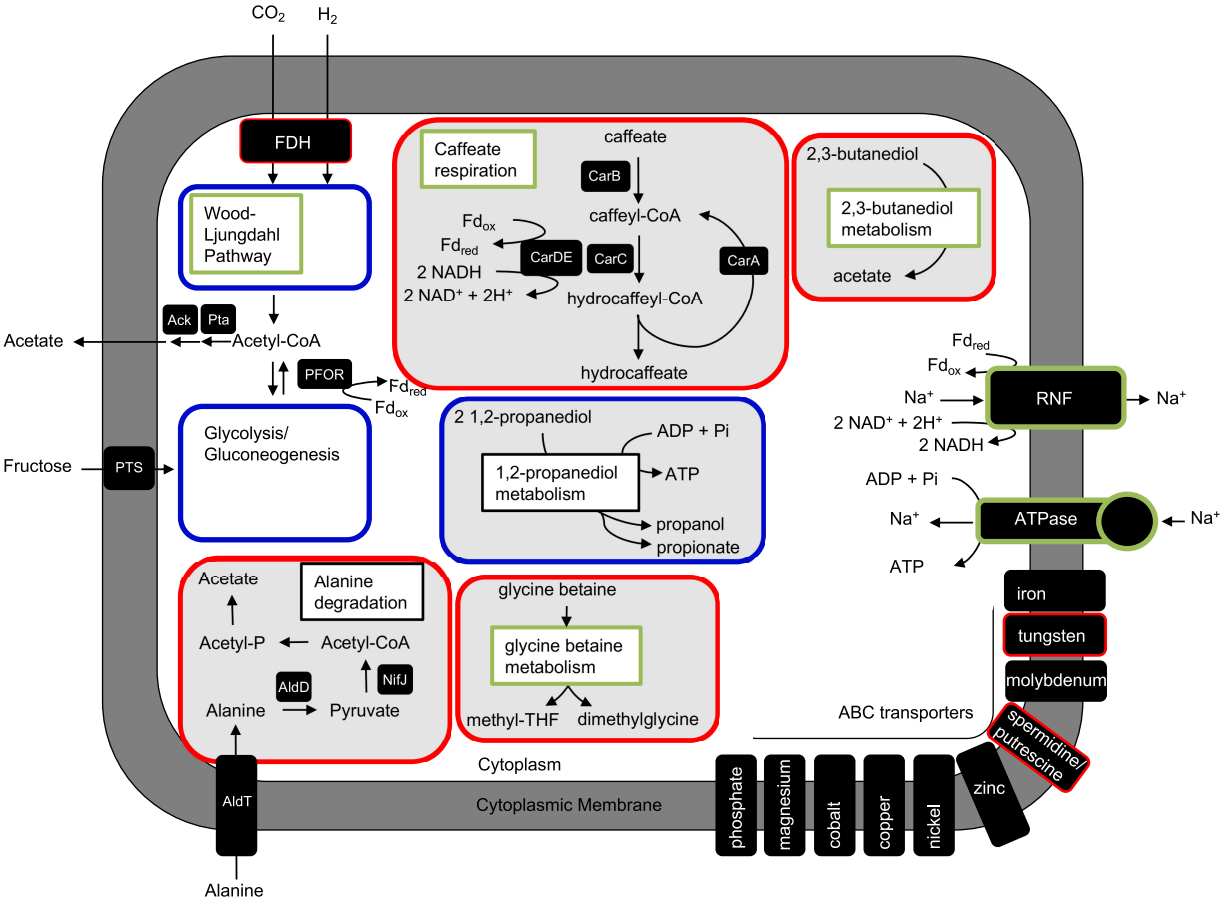

FIG 5 Metabolic overview of the *Acetobacterium* genus. Major metabolic pathways have been biochemically and genetically verified in *Acetobacterium woodii*, and predicted metabolisms are based upon the presence of each operon and similarity of predicted amino acid sequences. Pathways and accessory components framed in blue are encoded by all sequenced *Acetobacterium* strains. Pathways denoted in red are found in only a subset of *Acetobacterium* strains. Metabolic pathways denoted in green (e.g., glycine betaine, caffeate, and 2,3-butanediol) are linked to the Wood-Ljungdahl pathway.

included caffeate reduction, 2,3-butanediol oxidation, ethanol oxidation, alanine metabolism, glycine cleavage system, and methanol oxidation. Notably, the psychrophilic strains encode unique amino acid transport and utilization, and ion transport, which may have evolved for survival in low-temperature environments. Members of the *A. wieringae* clade encode over 40 unique predicted protein sequences, six of which were annotated as diguanylate cyclases (DGCs), enzymes that catalyze the production of the secondary messenger cyclic-di-GMP known to induce biofilm formation (97) and increase tolerance to reactive oxygen species (98). These unique DGCs may play a role in the environments which these strains inhabit, potentially aiding in attachment to various surfaces, including carbon-based electrodes. Overall, the comparative genomic analysis performed on the *Acetobacterium* genus provides a framework to understand the conserved metabolic processes across the genus, as well as identifying divergent genomic features (e.g., surface attachment) that can be exploited for targeted biotechnological applications using *Acetobacterium* strains, such as CO_2-based hydrogen storage (19), microbial electrosynthesis, and conversion of CO or syngas to acetate or other commodity chemicals.

MATERIALS AND METHODS

DNA extraction, DNA sequencing, and genome assembly of four *Acetobacterium* isolates. For *A. fimetarium*, *A. malicum*, *A. paludosum*, and *A. tundrae*, freeze-dried cells were obtained from ATCC and reconstituted on minimal freshwater medium growing autotrophically on H_2:CO_2 (12). Chromosomal DNA was extracted using the AllPrep DNA/RNA minikit (Qiagen) according to the manufacturer's protocol. Extracted DNA was processed for Illumina sequencing using the Nextera XT protocol (Illumina).

Samples were barcoded to enable multiplex sequencing of four samples using a single MiSeq V3 kit (2 × 301). Raw paired-end sequences were quality trimmed with CLC Genomics workbench (Qiagen) with a quality score cutoff of Q30, resulting in a total of 297, 911, 471, and 548 million bp for *A. fimetarium*, *A. malicum*, *A. paludosum*, and *A. tundrae*, respectively. Trimmed paired-end reads were assembled with SPAdes (v. 3.7.0) using the –careful flag to reduce mismatches and short indels (99). Assembly of trimmed paired-end reads resulted in four 'high-quality' draft genomes, a designation which is based upon minimum mandatory genome reporting standards of >90% completion, <5% contamination, the presence of 23S, 16S, and 5S rRNA genes, and at least 18 tRNAs (100) (Table 1; also see Table S1 in the supplemental material). The SPAdes assembled contigs were assessed for quality using Quast (101) and CheckM (v. 1.0.7) (37).

Genome completeness. Each *Acetobacterium* genome or metagenome-assembled genome (MAG) was assessed for completeness and contamination (Table 1) using the available *Eubacteriaceae* marker gene set from CheckM (version 1.0.13) (37).

Manual genome curation and pathway identification. The *Acetobacterium woodii* genome (https://www.ncbi.nlm.nih.gov/nuccore/379009891?report=genbank) was utilized for pathway identification, gene synteny, and sequence similarity. Fasta files of all genomes were downloaded from the NCBI genome database (August 2018) and uploaded to RAST (102, 103) for annotation. The RASTtk pipeline was used to analyze each genome. The BLAST algorithm in RAST was used to find specific gene sequences and to assess gene synteny of operons (e.g., *rnfCDGEAB*). Sequence similarity of predicted protein sequences is presented in Data Set S2. To determine what cutoffs should be utilized in assessing pathway conservation, we analyzed RecA across all sequenced *Acetobacterium* strains and found the sequence identity ranged from 100% to 86%. Pathways were categorized as highly conserved (high sequence similarity [86 to 100%] and near-identical gene arrangement), conserved (high to medium sequence similarity [50 to 86%] with variation in gene arrangement), or divergent (low sequence similarity [0 to 50%] with variation in gene arrangement). Thus, well-conserved sequences should share at least 86% identity. RAST protein-encoding gene (peg) identifiers reveal the synteny of genes that encode each protein in an operon, and where available, NCBI accession numbers are provided.

Concatenated and single protein trees. Predicted protein sequences encoded by the *hydABDEC* operon, Rnf operon, or individual diguanylate cyclases (DGCs) were identified in each genome. Proteins encoded by each operon were manually concatenated into a single contiguous amino acid sequence. Concatenated protein sequences were aligned in MEGA6.06 (104) using MUSCLE (105) with the following parameters: gap open penalty (−2.9), gap extend penalty (−0.01), hydrophobicity multiplier (1.2), and UPGMB clustering method with a minimum diagonal length (lambda) of 24. A protein tree was constructed using the Maximum Likelihood method with the following parameters: test of phylogeny = bootstrap method, number of bootstrap replications = 1,000, substitutions type = amino acid, model/method = Jones-Taylor-Thornton (JTT) model, rates among sites = uniform rates, gap/missing data treatment = partial deletion, ML heuristic method = nearest-neighbor-interchange (NNI), and branch swap filter = very strong.

Pan-genome analysis. Pan-genomics was performed with the Bacterial Pan Genome Analysis Tool (BPGA) version 1.0.0 (106). Initially, default BPGA parameters were utilized, which include protein clustering at 50% sequence identity cutoff with USEARCH (107). Further analysis was performed at various clustering cutoffs ranging from 10% to 99% to examine the pan-genome partitioning (e.g., total gene families, core gene families, accessory gene families, and unique genes). Resulting pan-genome and core genome trees were visualized using FigTree (http://tree.bio.ed.ac.uk/software/figtree/). The representative *Acetobacterium* core gene sequences from BPGA were uploaded to BlastKOALA (108) to evaluate KEGG pathway predictions.

Genotype/phenotype analysis. Observed growth phenotypes on various substrates from the work of Simankova and coworkers (109), and references therein, were overlaid with predicted metabolisms from genome data. Predicted metabolisms were determined as described above and also determined with KofamKOALA (110). Amino acid fasta files from RAST were uploaded to the KofamKOALA website (www.genome.jp/tools/kofamkoala) and analyzed using a 0.01 E value threshold. KofamKOALA data were compiled, and all pathway modules were examined (Data Set S3).

Data availability. The sequenced strains were deposited in the Sequence Read Archive (SRA) under the accession numbers WJBE00000000, WJBD00000000, WJBC00000000, and WJBB00000000. KBase narratives for *Acetobacterium* genomes can be found at https://narrative.kbase.us/narrative/ws.53630.obj.1.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

TEXT S1, DOCX file, 0.03 MB.

TEXT S2, DOCX file, 0.1 MB.

FIG S1, PDF file, 0.1 MB.

FIG S2, EPS file, 0.1 MB.

FIG S3, PDF file, 0.1 MB.

TABLE S1, DOCX file, 0.01 MB.

TABLE S2, DOCX file, 0.02 MB.

DATA SET S1, XLSX file, 0.01 MB.

DATA SET S2, XLSX file, 0.1 MB.

DATA SET S3, XLSX file, 0.04 MB.

ACKNOWLEDGMENTS

This work was performed in support of the U.S. Department of Energy's Fossil Energy CO_2 Utilization Research Program. The research was executed through the NETL Research and Innovation Center's CO_2 Utilization Technologies FWP. Research performed by Leidos Research Support Team staff was conducted under the RSS contract 89243318CFE000003.

D.E.R. performed analyses, analyzed data, and wrote the manuscript. C.W.M. analyzed data and wrote the manuscript. H.D.M., R.S.N., and D.G. edited the manuscript.

We declare no conflict of interest.

Funding was provided by the U.S. Department of Energy, Advanced Research Project Agency–Energy (award DE-AR0000089).

This work was funded by the Department of Energy, National Energy Technology Laboratory, an agency of the United States Government, through a support contract with Leidos Research Support Team (LRST). Neither the United States Government nor any agency thereof, nor any of their employees, nor LRST, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

REFERENCES

1. Schuchmann K, Müller V. 2016. Energetics and application of heterotrophy in acetogenic bacteria. Appl Environ Microbiol 82:4056–4069. https://doi.org/10.1128/AEM.00882-16.
2. Bengelsdorf FR, Beck MH, Erz C, Hoffmeister S, Karl MM, Riegler P, Wirth S, Poehlein A, Weuster-Botz D, Durre P. 2018. Bacterial anaerobic synthesis gas (syngas) and CO2+H2 fermentation. Adv Appl Microbiol 103:143–221. https://doi.org/10.1016/bs.aambs.2018.01.002.
3. Balch WE, Magrum LJ, Fox GE, Wolfe RS, Woese CR. 1977. An ancient divergence among the bacteria. J Mol Evol 9:305–311. https://doi.org/10.1007/BF01796092.
4. Pereyra LP, Hiibel SR, Pruden A, Reardon KF. 2008. Comparison of microbial community composition and activity in sulfate-reducing batch systems remediating mine drainage. Biotechnol Bioeng 101:702–713. https://doi.org/10.1002/bit.21930.
5. Gaidos E, Marteinsson V, Thorsteinsson T, Johannesson T, Runarsson AR, Stefansson A, Glazer B, Lanoil B, Skidmore M, Han S, Miller M, Rusch A, Foo W. 2009. An oligarchic microbial assemblage in the anoxic bottom waters of a volcanic subglacial lake. ISME J 3:486–497. https://doi.org/10.1038/ismej.2008.124.
6. Marteinsson VT, Runarsson A, Stefansson A, Thorsteinsson T, Johannesson T, Magnusson SH, Reynisson E, Einarsson B, Wade N, Morrison HG, Gaidos E. 2013. Microbial communities in the subglacial waters of the Vatnajokull ice cap, Iceland. ISME J 7:427–437. https://doi.org/10.1038/ismej.2012.97.
7. Küsel K, Pinkart HC, Drake HL, Devereux R. 1999. Acetogenic and sulfate-reducing bacteria inhabiting the rhizoplane and deep cortex cells of the sea grass *Halodule wrightii*. Appl Environ Microbiol 65:5117–5123. https://doi.org/10.1128/AEM.65.11.5117-5123.1999.
8. Shimizu S, Ueno A, Ishijima Y. 2011. Microbial communities associated with acetate-rich gas-petroleum reservoir surface facilities. Biosci Biotechnol Biochem 75:1835–1837. https://doi.org/10.1271/bbb.110243.
9. Sancho Navarro S, Cimpoia R, Bruant G, Guiot SR. 2016. Biomethanation of syngas using anaerobic sludge: shift in the catabolic routes with the CO partial pressure increase. Front Microbiol 7:1188. https://doi.org/10.3389/fmicb.2016.01188.
10. Arends JBA, Patil SA, Roume H, Rabaey K. 2017. Continuous long-term electricity-driven bioproduction of carboxylates and isopropanol from CO2 with a mixed microbial community. J CO2 Util 20:141–149. https://doi.org/10.1016/j.jcou.2017.04.014.
11. LaBelle EV, Marshall CW, Gilbert JA, May HD. 2014. Influence of acidic pH on hydrogen and acetate production by an electrosynthetic microbiome. PLoS One 9:e109935. https://doi.org/10.1371/journal.pone.0109935.
12. Marshall CW, Ross DE, Fichot EB, Norman RS, May HD. 2012. Electrosynthesis of commodity chemicals by an autotrophic microbial community. Appl Environ Microbiol 78:8412–8420. https://doi.org/10.1128/AEM.02401-12.
13. Saheb-Alam S, Singh A, Hermansson M, Persson F, Schnurer A, Wilen BM, Modin O. 2017. Effect of start-up strategies and electrode materials on carbon dioxide reduction on biocathodes. Appl Environ Microbiol 84:e02242-17. https://doi.org/10.1128/AEM.02242-17.
14. Marshall CW, Ross DE, Handley KM, Weisenhorn PB, Edirisinghe JN, Henry CS, Gilbert JA, May HD, Norman RS. 2017. Metabolic reconstruction and modeling microbial electrosynthesis. Sci Rep 7:8391. https://doi.org/10.1038/s41598-017-08877-z.
15. Usher KM, Kaksonen AH, Bouquet D, Cheng KY, Geste Y, Chapman PG, Johnston CD. 2015. The role of bacterial communities and carbon dioxide on the corrosion of steel. Corros Sci 98:354–365. https://doi.org/10.1016/j.corsci.2015.05.043.
16. Grabowski A, Nercessian O, Fayolle F, Blanchet D, Jeanthon C. 2005. Microbial diversity in production waters of a low-temperature biodegraded oil reservoir. FEMS Microbiol Ecol 54:427–443. https://doi.org/10.1016/j.femsec.2005.05.007.
17. Freedman AJE, Tan B, Thompson JR. 2017. Microbial potential for carbon and nutrient cycling in a geogenic supercritical carbon dioxide reservoir. Environ Microbiol 19:2228–2245. https://doi.org/10.1111/1462-2920.13706.
18. Marshall CW, LaBelle EV, May HD. 2013. Production of fuels and chemicals from waste by microbiomes. Curr Opin Biotechnol 24:391–397. https://doi.org/10.1016/j.copbio.2013.03.016.
19. Müller V. 2019. New horizons in acetogenic conversion of one-carbon

substrates and biological hydrogen storage. Trends Biotechnol 37:1344–1354. https://doi.org/10.1016/j.tibtech.2019.05.008.
20. Straub M, Demler M, Weuster-Botz D, Dürre P. 2014. Selective enhancement of autotrophic acetate production with genetically modified *Acetobacterium woodii*. J Biotechnol 178:67–72. https://doi.org/10.1016/j.jbiotec.2014.03.005.
21. Poehlein A, Schmidt S, Kaster AK, Goenrich M, Vollmers J, Thurmer A, Bertsch J, Schuchmann K, Voigt B, Hecker M, Daniel R, Thauer RK, Gottschalk G, Müller V. 2012. An ancient pathway combining carbon dioxide fixation with the generation and utilization of a sodium ion gradient for ATP synthesis. PLoS One 7:e33439. https://doi.org/10.1371/journal.pone.0033439.
22. Westphal L, Wiechmann A, Baker J, Minton NP, Müller V. 2018. The Rnf complex is an energy-coupled transhydrogenase essential to reversibly link cellular NADH and ferredoxin pools in the acetogen *Acetobacterium woodii*. J Bacteriol 200:e00357-18. https://doi.org/10.1128/JB.00357-18.
23. Bertsch J, Siemund AL, Kremp F, Müller V. 2016. A novel route for ethanol oxidation in the acetogenic bacterium *Acetobacterium woodii*: the acetaldehyde/ethanol dehydrogenase pathway. Environ Microbiol 18:2913–2922. https://doi.org/10.1111/1462-2920.13082.
24. Dönig J, Müller V. 2018. Alanine, a novel growth substrate for the acetogenic bacterium *Acetobacterium woodii*. Appl Environ Microbiol 84:e02023-18. https://doi.org/10.1128/AEM.02023-18.
25. Hess V, Gonzalez JM, Parthasarathy A, Buckel W, Müller V. 2013. Caffeate respiration in the acetogenic bacterium *Acetobacterium woodii*: a coenzyme A loop saves energy for caffeate activation. Appl Environ Microbiol 79:1942–1947. https://doi.org/10.1128/AEM.03604-12.
26. Hess V, Oyrik O, Trifunovic D, Müller V. 2015. 2,3-Butanediol metabolism in the acetogen *Acetobacterium woodii*. Appl Environ Microbiol 81:4711–4719. https://doi.org/10.1128/AEM.00960-15.
27. Schuchmann K, Schmidt S, Martinez Lopez A, Kaberline C, Kuhns M, Lorenzen W, Bode HB, Joos F, Müller V. 2015. Nonacetogenic growth of the acetogen *Acetobacterium woodii* on 1,2-propanediol. J Bacteriol 197:382–391. https://doi.org/10.1128/JB.02383-14.
28. Trifunovic D, Schuchmann K, Müller V. 2016. Ethylene glycol metabolism in the acetogen *Acetobacterium woodii*. J Bacteriol 198:1058–1065. https://doi.org/10.1128/JB.00942-15.
29. Shin J, Song Y, Jin S, Lee JK, Kim DR, Kim SC, Cho S, Cho BK. 2018. Genome-scale analysis of *Acetobacterium bakii* reveals the cold adaptation of psychrotolerant acetogens by post-transcriptional regulation. RNA 24:1839–1855. https://doi.org/10.1261/rna.068239.118.
30. Ding C, Chow WL, He J. 2013. Isolation of *Acetobacterium* sp. strain AG, which reductively debrominates octa- and pentabrominated diphenyl ether technical mixtures. Appl Environ Microbiol 79:1110–1117. https://doi.org/10.1128/AEM.02919-12.
31. Adrian NR, Arnett CM. 2004. Anaerobic biodegradation of hexahydro-1,3,5-trinitro-1,3,5-triazine (RDX) by *Acetobacterium malicum* strain HAAP-1 isolated from a methanogenic mixed culture. Curr Microbiol 48:332–340. https://doi.org/10.1007/s00284-003-4156-8.
32. Sherburne LA, Shrout JD, Alvarez PJ. 2005. Hexahydro-1,3,5-trinitro-1,3,5-triazine (RDX) degradation by *Acetobacterium paludosum*. Biodegradation 16:539–547. https://doi.org/10.1007/s10532-004-6945-6.
33. Marshall CW, Ross DE, Fichot EB, Norman RS, May HD. 2013. Long-term operation of microbial electrosynthesis systems improves acetate production by autotrophic microbiomes. Environ Sci Technol 47:6023–6029. https://doi.org/10.1021/es400341b.
34. Philips J, Monballyu E, Georg S, De Paepe K, Prevoteau A, Rabaey K, Arends JBA. 2019. An Acetobacterium strain isolated with metallic iron as electron donor enhances iron corrosion by a similar mechanism as *Sporomusa sphaeroides*. FEMS Microbiol Ecol 95:fiy222. https://doi.org/10.1093/femsec/fiy222.
35. Kronen M, Lee M, Jones ZL, Manefield MJ. 2019. Reductive metabolism of the important atmospheric gas isoprene by homoacetogens. ISME J 13:1168–1182. https://doi.org/10.1038/s41396-018-0338-z.
36. Willems A, Collins MD. 1996. Phylogenetic relationships of the genera *Acetobacterium* and *Eubacterium* sensu stricto and reclassification of *Eubacterium alactolyticum* as *Pseudoramibacter alactolyticus* gen. nov., comb. nov. Int J Syst Bacteriol 46:1083–1087. https://doi.org/10.1099/00207713-46-4-1083.
37. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055. https://doi.org/10.1101/gr.186072.114.
38. Esposito A, Tamburini S, Triboli L, Ambrosino L, Chiusano ML, Jousson O. 2019. Insights into the genome structure of four acetogenic bacteria with specific reference to the Wood-Ljungdahl pathway. Microbiologyopen 8:e938. https://doi.org/10.1002/mbo3.938.
39. Chun J, Oren A, Ventosa A, Christensen H, Arahal DR, da Costa MS, Rooney AP, Yi H, Xu X-W, De Meyer S, Trujillo ME. 2018. Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. Int J Syst Evol Microbiol 68:461–466. https://doi.org/10.1099/ijsem.0.002516.
40. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. 2007. DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol 57:81–91. https://doi.org/10.1099/ijs.0.64483-0.
41. Konstantinidis KT, Tiedje JM. 2005. Towards a genome-based taxonomy for prokaryotes. J Bacteriol 187:6258–6264. https://doi.org/10.1128/JB.187.18.6258-6264.2005.
42. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil PA, Hugenholtz P. 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat Biotechnol 36:996–1004. https://doi.org/10.1038/nbt.4229.
43. Parks DH, Chuvochina M, Chaumeil PA, Rinke C, Mussig AJ, Hugenholtz P. 2020. A complete domain-to-species taxonomy for Bacteria and Archaea. Nat Biotechnol https://doi.org/10.1038/s41587-020-0501-8.
44. Tatusov RL, Galperin MY, Natale DA, Koonin EV. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res 28:33–36. https://doi.org/10.1093/nar/28.1.33.
45. Weiss MC, Sousa FL, Mrnjavac N, Neukirchen S, Roettger M, Nelson-Sathi S, Martin WF. 2016. The physiology and habitat of the last universal common ancestor. Nat Microbiol 1:16116. https://doi.org/10.1038/nmicrobiol.2016.116.
46. Müller V, Frerichs J. 2013. Acetogenic bacteria. eLS https://doi.org/10.1002/9780470015902.a0020086.pub2.
47. Biegel E, Müller V. 2011. A Na+-translocating pyrophosphatase in the acetogenic bacterium *Acetobacterium woodii*. J Biol Chem 286:6080–6084. https://doi.org/10.1074/jbc.M110.192823.
48. Buckel W, Thauer RK. 2013. Energy conservation via electron bifurcating ferredoxin reduction and proton/Na+ translocating ferredoxin oxidation. Biochim Biophys Acta 1827:94–113. https://doi.org/10.1016/j.bbabio.2012.07.002.
49. Imkamp F, Biegel E, Jayamani E, Buckel W, Müller V. 2007. Dissection of the caffeate respiratory chain in the acetogen *Acetobacterium woodii*: identification of an Rnf-type NADH dehydrogenase as a potential coupling site. J Bacteriol 189:8145–8153. https://doi.org/10.1128/JB.01017-07.
50. Schmidt S, Biegel E, Müller V. 2009. The ins and outs of Na+ bioenergetics in *Acetobacterium woodii*. Biochim Biophys Acta 1787:691–696. https://doi.org/10.1016/j.bbabio.2008.12.015.
51. Ceccaldi P, Schuchmann K, Müller V, Elliott SJ. 2017. The hydrogen dependent CO2 reductase: the first completely CO tolerant FeFe-hydrogenase. Energy Environ Sci 10:503–508. https://doi.org/10.1039/C6EE02494G.
52. Schuchmann K, Muller V. 2013. Direct and reversible hydrogenation of CO2 to formate by a bacterial carbon dioxide reductase. Science 342:1382–1385. https://doi.org/10.1126/science.1244758.
53. Wang S, Huang H, Kahnt J, Mueller AP, Kopke M, Thauer RK. 2013. NADP-specific electron-bifurcating [FeFe]-hydrogenase in a functional complex with formate dehydrogenase in Clostridium autoethanogenum grown on CO. J Bacteriol 195:4373–4386. https://doi.org/10.1128/JB.00678-13.
54. Makdessi K, Andreesen JR, Pich A. 2001. Tungstate uptake by a highly specific ABC transporter in *Eubacterium acidaminophilum*. J Biol Chem 276:24557–24564. https://doi.org/10.1074/jbc.M101293200.
55. Smart JP, Cliff MJ, Kelly DJ. 2009. A role for tungsten in the biology of *Campylobacter jejuni*: tungstate stimulates formate dehydrogenase activity and is transported via an ultra-high affinity ABC system distinct from the molybdate transporter. Mol Microbiol 74:742–757. https://doi.org/10.1111/j.1365-2958.2009.06902.x.
56. Andreesen JR, Makdessi K. 2008. Tungsten, the surprisingly positively acting heavy metal element for prokaryotes. Ann N Y Acad Sci 1125:215–229. https://doi.org/10.1196/annals.1419.003.
57. Rech S, Wolin C, Gunsalus RP. 1996. Properties of the periplasmic ModA

molybdate-binding protein of *Escherichia coli*. J Biol Chem 271:2557–2562. https://doi.org/10.1074/jbc.271.5.2557.
58. Biegel E, Müller V. 2010. Bacterial Na+-translocating ferredoxin:NAD+ oxidoreductase. Proc Natl Acad Sci U S A 107:18138–18142. https://doi.org/10.1073/pnas.1010318107.
59. Biegel E, Schmidt S, González JM, Müller V. 2011. Biochemistry, evolution and physiological function of the Rnf complex, a novel ion-motive electron transport complex in prokaryotes. Cell Mol Life Sci 68:613–634. https://doi.org/10.1007/s00018-010-0555-8.
60. Biegel E, Schmidt S, Müller V. 2009. Genetic, immunological and biochemical evidence for a Rnf complex in the acetogen *Acetobacterium woodii*. Environ Microbiol 11:1438–1443. https://doi.org/10.1111/j.1462-2920.2009.01871.x.
61. Hess V, Schuchmann K, Müller V. 2013. The ferredoxin:NAD+ oxidoreductase (Rnf) from the acetogen *Acetobacterium woodii* requires Na+ and is reversibly coupled to the membrane potential. J Biol Chem 288:31496–31502. https://doi.org/10.1074/jbc.M113.510255.
62. Curatti L, Brown CS, Ludden PW, Rubio LM. 2005. Genes required for rapid expression of nitrogenase activity in *Azotobacter vinelandii*. Proc Natl Acad Sci U S A 102:6291–6296. https://doi.org/10.1073/pnas.0501216102.
63. Poehlein A, Bengelsdorf FR, Schiel-Bengelsdorf B, Daniel R, Dürre P. 2016. Genome sequence of the acetogenic bacterium *Acetobacterium wieringae* DSM 1911T. Genome Announc 4:e01430-16. https://doi.org/10.1128/genomeA.01430-16.
64. Arantes AL, Moreira JPC, Diender M, Parshina SN, Stams AJM, Alves MM, Alves JI, Sousa DZ. 2020. Enrichment of anaerobic syngas-converting communities and isolation of a novel carboxydotrophic *Acetobacterium wieringae* strain JM. Front Microbiol 11:58. https://doi.org/10.3389/fmicb.2020.00058.
65. Fritz M, Müller V. 2007. An intermediate step in the evolution of ATPases–the F1F0-ATPase from *Acetobacterium woodii* contains F-type and V-type rotor subunits and is capable of ATP synthesis. FEBS J 274:3421–3428. https://doi.org/10.1111/j.1742-4658.2007.05874.x.
66. Herrmann G, Jayamani E, Mai G, Buckel W. 2008. Energy conservation via electron-transferring flavoprotein in anaerobic bacteria. J Bacteriol 190:784–791. https://doi.org/10.1128/JB.01422-07.
67. Weghoff MC, Bertsch J, Müller V. 2015. A novel mode of lactate metabolism in strictly anaerobic bacteria. Environ Microbiol 17:670–677. https://doi.org/10.1111/1462-2920.12493.
68. Li F, Hinderberger J, Seedorf H, Zhang J, Buckel W, Thauer RK. 2008. Coupled ferredoxin and crotonyl coenzyme A (CoA) reduction with NADH catalyzed by the butyryl-CoA dehydrogenase/Etf complex from Clostridium kluyveri. J Bacteriol 190:843–850. https://doi.org/10.1128/JB.01417-07.
69. Schink B. 1997. Energetics of syntrophic cooperation in methanogenic degradation. Microbiol Mol Biol Rev 61:262–280. https://doi.org/10.1128/.61.2.262-280.1997.
70. Thauer RK, Jungermann K, Decker K. 1977. Energy conservation in chemotrophic anaerobic bacteria. Bacteriol Rev 41:100–180. https://doi.org/10.1128/MMBR.41.1.100-180.1977.
71. Eichler B, Schink B. 1984. Oxidation of primary aliphatic alcohols by *Acetobacterium carbinolicum* sp. nov., a homoacetogenic anaerobe. Arch Microbiol 140:147–152. https://doi.org/10.1007/BF00454917.
72. Obradors N, Badia J, Baldoma L, Aguilar J. 1988. Anaerobic metabolism of the L-rhamnose fermentation product 1,2-propanediol in *Salmonella typhimurium*. J Bacteriol 170:2159–2162. https://doi.org/10.1128/jb.170.5.2159-2162.1988.
73. Oppermann FB, Schmidt B, Steinbuchel A. 1991. Purification and characterization of acetoin:2,6-dichlorophenolindophenol oxidoreductase, dihydrolipoamide dehydrogenase, and dihydrolipoamide acetyltransferase of the *Pelobacter carbinolicus* acetoin dehydrogenase enzyme system. J Bacteriol 173:757–767. https://doi.org/10.1128/jb.173.2.757-767.1991.
74. Oppermann FB, Steinbuchel A. 1994. Identification and molecular characterization of the aco genes encoding the *Pelobacter carbinolicus* acetoin dehydrogenase enzyme system. J Bacteriol 176:469–485. https://doi.org/10.1128/jb.176.2.469-485.1994.
75. Buschhorn H, Dürre P, Gottschalk G. 1989. Production and utilization of ethanol by the homoacetogen *Acetobacterium woodii*. Appl Environ Microbiol 55:1835–1840. https://doi.org/10.1128/AEM.55.7.1835-1840.1989.
76. Braun M, Mayer F, Gottschalk G. 1981. *Clostridium aceticum* (Wieringa), a microorganism producing acetic acid from molecular hydrogen and carbon dioxide. Arch Microbiol 128:288–293. https://doi.org/10.1007/BF00422532.
77. Schink B. 1984. Fermentation of 2,3-butanediol by *Pelobacter carbinolicus* sp. nov. and *Pelobacter propionicus* sp. nov., and evidence for propionate formation from C2 compounds. Arch Microbiol 137:33–41. https://doi.org/10.1007/BF00425804.
78. Keller A, Schink B, Müller N. 2019. Alternative pathways of acetogenic ethanol and methanol degradation in the thermophilic anaerobe *Thermacetogenium phaeum*. Front Microbiol 10:423. https://doi.org/10.3389/fmicb.2019.00423.
79. Kotsyurbenko OR, Simankova MV, Nozhevnikova AN, Zhilina TN, Bolotina NP, Lysenko AM, Osipov GA. 1995. New species of psychrophilic acetogens: *Acetobacterium bakii* sp. nov., *A. paludosum* sp. nov., *A. fimetarium* sp. nov. Arch Microbiol 163:29–34. https://doi.org/10.1007/BF00262200.
80. Tanaka K, Pfennig N. 1988. Fermentation of 2-methoxyethanol by *Acetobacterium malicum* sp. nov. and Pelobacter venetianus. Arch Microbiol 149:181–187. https://doi.org/10.1007/BF00422003.
81. Schoelmerich MC, Katsyv A, Sung W, Mijic V, Wiechmann A, Kottenhahn P, Baker J, Minton NP, Müller V. 2018. Regulation of lactate metabolism in the acetogenic bacterium *Acetobacterium woodii*. Environ Microbiol 20:4587–4595. https://doi.org/10.1111/1462-2920.14412.
82. Dilling S, Imkamp F, Schmidt S, Müller V. 2007. Regulation of caffeate respiration in the acetogenic bacterium *Acetobacterium woodii*. Appl Environ Microbiol 73:3630–3636. https://doi.org/10.1128/AEM.02060-06.
83. Blum U, Wentworth TR, Klein K, Worsham AD, King LD, Gerig TM, Lyu SW. 1991. Phenolic acid content of soils from wheat-no till, wheat-conventional till, and fallow-conventional till soybean cropping systems. J Chem Ecol 17:1045–1068. https://doi.org/10.1007/BF01402933.
84. Hess V, Vitt S, Müller V. 2011. A caffeyl-coenzyme A synthetase initiates caffeate activation prior to caffeate reduction in the acetogenic bacterium *Acetobacterium woodii*. J Bacteriol 193:971–978. https://doi.org/10.1128/JB.01126-10.
85. Planet PJ, Kachlany SC, Fine DH, DeSalle R, Figurski DH. 2003. The widespread colonization island of *Actinobacillus actinomycetemcomitans*. Nat Genet 34:193–198. https://doi.org/10.1038/ng1154.
86. Schreiner HC, Sinatra K, Kaplan JB, Furgang D, Kachlany SC, Planet PJ, Perez BA, Figurski DH, Fine DH. 2003. Tight-adherence genes of *Actinobacillus actinomycetemcomitans* are required for virulence in a rat model. Proc Natl Acad Sci U S A 100:7295–7300. https://doi.org/10.1073/pnas.1237223100.
87. Kachlany SC, Planet PJ, Desalle R, Fine DH, Figurski DH, Kaplan JB. 2001. flp-1, the first representative of a new pilin gene subfamily, is required for non-specific adherence of *Actinobacillus actinomycetemcomitans*. Mol Microbiol 40:542–554. https://doi.org/10.1046/j.1365-2958.2001.02422.x.
88. Deutzmann JS, Spormann AM. 2017. Enhanced microbial electrosynthesis by using defined co-cultures. ISME J 11:704–714. https://doi.org/10.1038/ismej.2016.149.
89. Nevin KP, Woodard TL, Franks AE, Summers ZM, Lovley DR. 2010. Microbial electrosynthesis: feeding microbes electricity to convert carbon dioxide and water to multicarbon extracellular organic compounds. mBio 1:e00103-10. https://doi.org/10.1128/mBio.00103-10.
90. D'Amico S, Collins T, Marx J-C, Feller G, Gerday C. 2006. Psychrophilic microorganisms: challenges for life. EMBO Rep 7:385–389. https://doi.org/10.1038/sj.embor.7400662.
91. Goldstein RA. 2007. Amino-acid interactions in psychrophiles, mesophiles, thermophiles, and hyperthermophiles: insights from the quasi-chemical approximation. Protein Sci 16:1887–1895. https://doi.org/10.1110/ps.072947007.
92. Methe BA, Nelson KE, Deming JW, Momen B, Melamud E, Zhang X, Moult J, Madupu R, Nelson WC, Dodson RJ, Brinkac LM, Daugherty SC, Durkin AS, DeBoy RT, Kolonay JF, Sullivan SA, Zhou L, Davidsen TM, Wu M, Huston AL, Lewis M, Weaver B, Weidman JF, Khouri H, Utterback TR, Feldblyum TV, Fraser CM. 2005. The psychrophilic lifestyle as revealed by the genome sequence of *Colwellia psychrerythraea* 34H through genomic and proteomic analyses. Proc Natl Acad Sci U S A 102:10913–10918. https://doi.org/10.1073/pnas.0504766102.
93. Bellion E, Tan F. 1987. An NAD+-dependent alanine dehydrogenase from a methylotrophic bacterium. Biochem J 244:565–570. https://doi.org/10.1042/bj2440565.
94. Ashida H, Saito Y, Kojima C, Kobayashi K, Ogasawara N, Yokota A. 2003. A functional link between RuBisCO-like protein of Bacillus and

photosynthetic RuBisCO. Science 302:286–290. https://doi.org/10.1126/science.1086997.
95. Badger MR, Bek EJ. 2008. Multiple Rubisco forms in proteobacteria: their functional significance in relation to CO2 acquisition by the CBB cycle. J Exp Bot 59:1525–1541. https://doi.org/10.1093/jxb/erm297.
96. Kavanagh KL, Jörnvall H, Persson B, Oppermann U. 2008. Medium- and short-chain dehydrogenase/reductase gene and protein families: the SDR superfamily: functional and structural diversity within a family of metabolic and regulatory enzymes. Cell Mol Life Sci 65:3895–3906. https://doi.org/10.1007/s00018-008-8588-y.
97. Valentini M, Filloux A. 2016. Biofilms and c-di-GMP signaling: lessons from *Pseudomonas aeruginosa* and other bacteria. J Biol Chem 291:12547–12555. https://doi.org/10.1074/jbc.R115.711507.
98. Fernandez NL, Waters CM. 2019. Cyclic di-GMP increases catalase production and hydrogen peroxide tolerance in *Vibrio cholerae*. Appl Environ Microbiol 85:e01043-19. https://doi.org/10.1128/AEM.01043-19.
99. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
100. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Schulz F, Jarett J, Rivers AR, Eloe-Fadrosh EA, Tringe SG, Ivanova NN, Copeland A, Clum A, Becraft ED, Malmstrom RR, Birren B, Podar M, Bork P, Weinstock GM, Garrity GM, Dodsworth JA, Yooseph S, Sutton G, Glockner FO, Gilbert JA, Nelson WC, Hallam SJ, Jungbluth SP, Ettema TJG, Tighe S, Konstantinidis KT, Liu WT, Baker BJ, Rattei T, Eisen JA, Hedlund B, McMahon KD, Fierer N, Knight R, Finn R, Cochrane G, Karsch-Mizrachi I, Tyson GW, Rinke C, Genome Standards Consortium, Lapidus A, Meyer F, Yilmaz P, Parks DH, Eren AM, Schriml L, Banfield JF, Hugenholtz P, Woyke T. 2018. Corrigendum: minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol 36:660. https://doi.org/10.1038/nbt0718-660a.
101. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086.
102. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9:75. https://doi.org/10.1186/1471-2164-9-75.
103. Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Thomason JA, III, Stevens R, Vonstein V, Wattam AR, Xia F. 2015. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci Rep 5:8365. https://doi.org/10.1038/srep08365.
104. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–2729. https://doi.org/10.1093/molbev/mst197.
105. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. https://doi.org/10.1093/nar/gkh340.
106. Chaudhari NM, Gupta VK, Dutta C. 2016. BPGA—an ultra-fast pan-genome analysis pipeline. Sci Rep 6:24373. https://doi.org/10.1038/srep24373.
107. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460–2461. https://doi.org/10.1093/bioinformatics/btq461.
108. Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol 428:726–731. https://doi.org/10.1016/j.jmb.2015.11.006.
109. Simankova MV, Kotsyurbenko OR, Stackebrandt E, Kostrikina NA, Lysenko AM, Osipov GA, Nozhevnikova AN. 2000. *Acetobacterium tundrae* sp. nov., a new psychrophilic acetogenic bacterium from tundra soil. Arch Microbiol 174:440–447. https://doi.org/10.1007/s002030000229.
110. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, Ogata H. 2020. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. Bioinformatics 36:2251–2252. https://doi.org/10.1093/bioinformatics/btz859.