PNAS

# Supporting Information for

# Deep-branching acetogens in serpentinized subsurface fluids of Oman

Daniel R. Colman[1], Emily A. Kraus[2], Patrick H. Thieringer[2], Kaitlin Rempfert[3], Alexis S. Templeton[3], John R. Spear[2], and Eric S. Boyd[1*]

[1]*Department of Microbiology and Cell Biology, Montana State University, Bozeman, MT 59717*

[2]*Department of Civil and Environmental Engineering, Colorado School of Mines, Golden, CO 80401*

[3]*Department of Geological Sciences, University of Colorado, Boulder, CO 80309*

Corresponding author: Eric Boyd
Email: *eric.boyd@montana.edu*

**This PDF file includes:**

Supporting text
Tables S1 to S2
Figures S1 to S6
Legends for Datasets S1 to S3
SI References

**Other supporting materials for this manuscript include the following:**

Datasets S1 to S3

**Supplementary Results/Discussion**

**ATP synthases.** Both Type I and Type II 'Acetothermia' MAGs encode ATP synthases allowing for harnessing of chemiosmotic potentials, although they harbor evolutionarily-related, but highly distinct types of ATPase. F-Type and V-Type ATP synthases represent the two predominant means by which life generates ATP through depletion of chemiosmotic potentials (1). The Type II MAGs encode an F-Type ATP synthase complex typical of other 'Acetothermia' and Bacteria (**Supplementary Fig. 4**), while the Type I population MAGs encoded V-type ATP synthases that are generally rare among Bacteria, but prevalent among Archaea. Moreover, the Type II F-Type ATP synthase represents a basal lineage among those synthases, while the Type I V-Type ATP synthase belongs to a derived lineage (**Supplementary Fig. 5**). Thus, it is likely that the F-Type ATP synthases represent the ancestral complex of 'Acetothermia', among which the Type II MAGs represent a very early diverging member. In contrast, the V-Type ATP synthases may have been acquired by the Type I MAGs later in their evolutionary trajectory.

**Carbon fixation/metabolism.** Both Type I and II 'Acetothermia' MAGs encoded ferredoxin-dependent pyruvate synthase complexes (PFOR) (2), in addition to complete glycolytic/gluconeogenic pathways that would allow for the conversion of acetyl-CoA to six-carbon sugars that can be used in biosynthesis (3). Specifically, representative MAGs for both populations encode the key protein involved in gluconeogenesis (an archaeal-like fructose bisphosphate aldolase; FBP). However, slight differences in the enzymes catalyzing the key steps of glycolytic/gluconeogenic pathways were observed when

comparing populations. Specifically, the phosphorylation of pyruvate is likely conducted via two different enzymes (pyruvate-water dikinase (PpsA) and pyruvate orthophosphate dikanase (PpdK)) in the Types I and II populations, respectively (**Fig. 2, Supplementary Table S3**). Further, differences in glucose formation were also observed (e.g., via glucose-6-phosphate isomerase (Pgi1) and glucose/mannose-6-phosphate isomerase (Pgi-Pmi), respectively) (**Fig. 2, Supplementary Table S3**).

Type I MAGs exhibited evidence for additional heterotrophic pathways including a near complete TCA cycle, indicating an ability to harness reducing equivalents from acetyl-CoA to generate NADH that could then be used to generate a chemiosmotic gradient via NADH dehydrogenase (Nuo) (**Fig. 2**). Further, the Type I MAGs encoded several ABC-type transporters for amino acids (i.e., general L-amino acids, branched-chain amino acids) that were not present in the Type II MAGs, suggesting the Type I cells can use these carbon sources for biosynthesis or energy conservation. In contrast to the Type I MAGs, Type II MAGs do not encode a complete TCA cycle but do encode a BmpA-NupABC high-affinity nucleoside transport system that has been shown to operate as nucleoside scavenging systems in *Lactococcus* (4). Type II 'Acetothermia' also encoded partial ABC transporter modules for galactose and maltose, suggesting that they may also be capable of taking up these sugar compounds, when available. Thus, the Type II populations may be able to supplement their autotrophic capacity when exogenous nucleoside or sugar molecules are available.

**Supplementary Table S1**. Sample information for 2015 water samples including select geochemical measurements and host-rock setting from Rempfert et al. 2017.

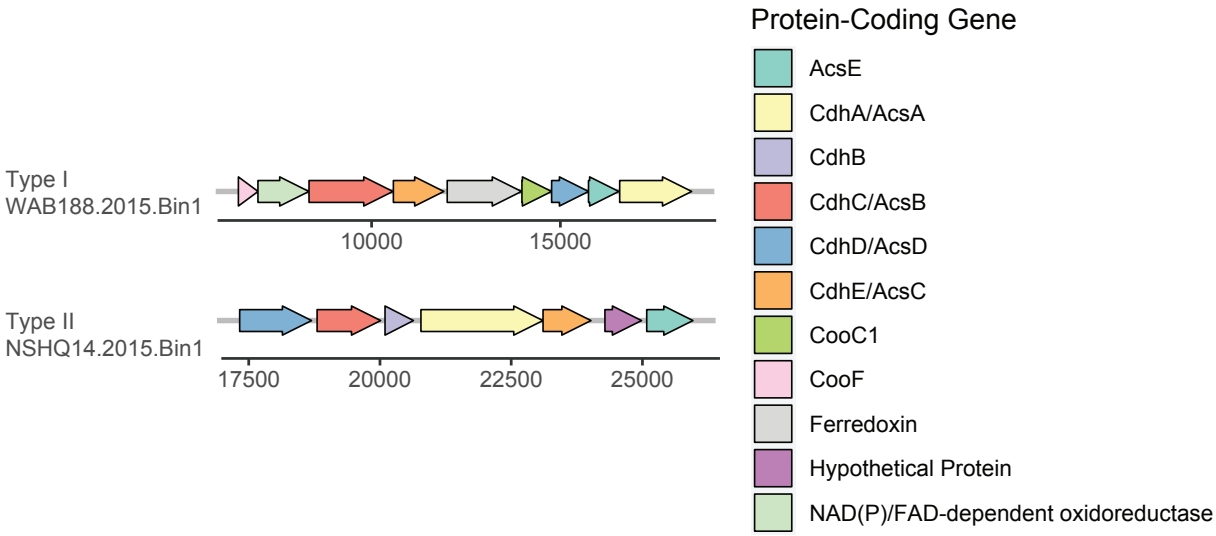
| Sample | WAB103 | NSHQ3B | WAB55 | WAB188 | WAB56 | NSHQ04 | WAB71 | NSHQ14 |
|---|---|---|---|---|---|---|---|---|
| Lithology / Classification | Gabbro | Alkaline Peridotite | Contact | Contact | Hyperalk. Peridotite | Hyperalk. Peridotite / Contact | Hyperalk. Peridotite | Hyperalk. Peridotite |
| pH | 8.2 | 8.4 | 9.3 | 8.7 | 10.6 | 10.6 | 11.0 | 11.4 |
| Temperature | | | | | | | | |
| Conductivity (µS/cm) | 1,095 | 297 | 585 | 427 | 354 | 2,386 | 1,109 | 3,199 |
| Eh (mV) | 130 | 158 | 78 | -220 | -251 | -103 | -126 | -307 |
| DIC (mM) | 2.3 | 2.5 | 2.6 | 2.3 | 0.22 | 0.18[a] | 0.21 | 0.19 |
| $H_2$ (mM) | Bd | Bd | Bd | Bd | Bd | 0.18 | Bd | 2.9 |
| $CH_4$ (mM) | Bd | Bd | Bd | 0.04 | 0.01 | 2.3 | 0.02 | 0.12 |
| Acetate (µM) | 0.43 | 0.47 | 2.0 | 3.8 | Nm | 1.4[a] | 0.63 | 1.2 |
| Formate (µM) | 1.1 | 1.2 | 1.4 | 1.0 | 1.5 | 2.3[a] | 1.5 | 1.7 |
| $SO_4^{2-}$ (mM) | 1.7 | 0.28 | 3.9 | 0.59 | 0.02 | 0.36 | 0.03 | 0.15 |
| $NO_3^-$ (mM) | 0.36 | 0.11 | 0.15 | 0.10 | 0.03 | 0.02 | Bd | 0.02 |
| $NO_2^-$ (mM) | 0.001 | 0.004 | 0.003 | Bd | 0.001 | Bd | Bd | 0.0004 |
| Dissolved $O_2$ (mM) | 0.04 | 0.17 | 0.09 | 0.004 | 0.02 | 0.00 | 0.01 | 0.01 |

[a] Data are reported from the well in 2016 because these data are not available for 2015.
Abbreviations: Bd indicates below detection.
Abbreviations: Hyperalk: hyperalkaline; Bd, below detection.

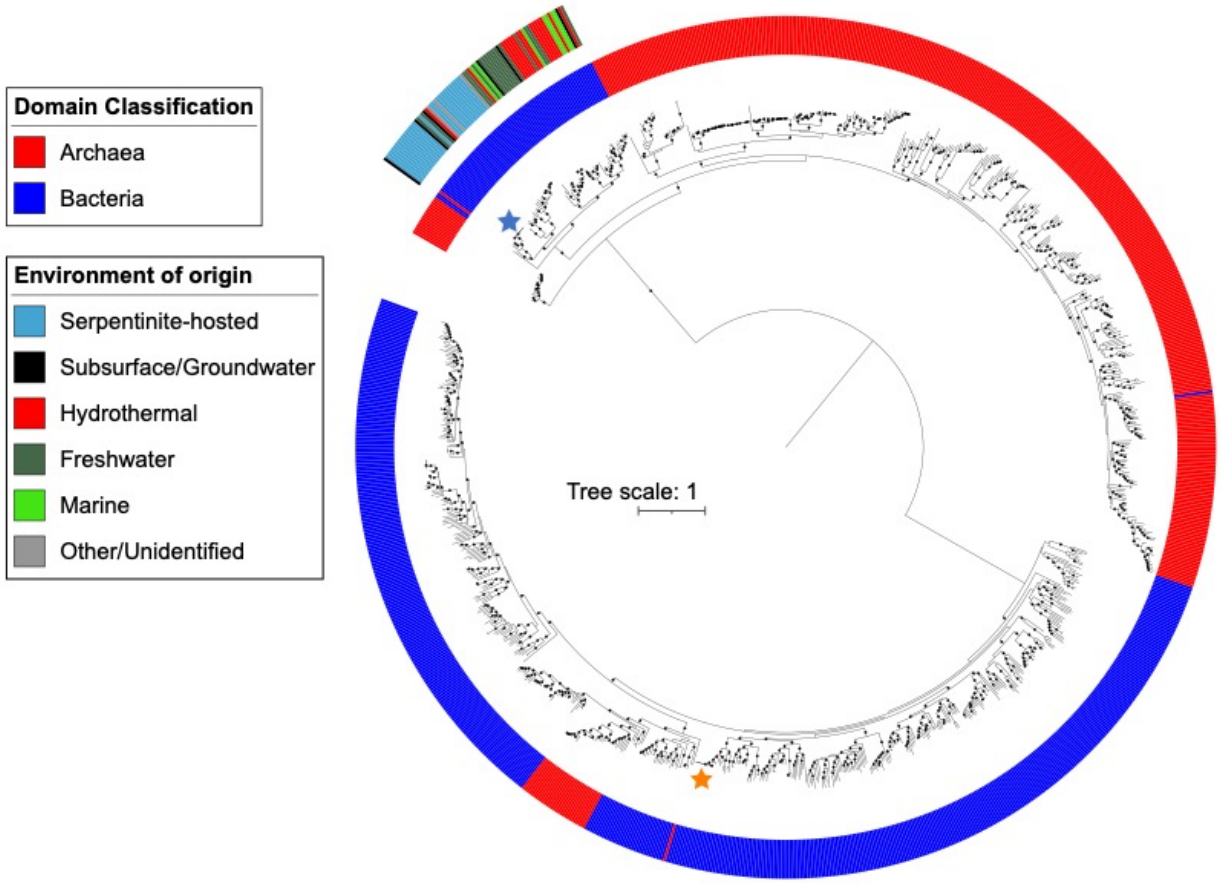
**Supplementary Table S2**. Sample information for 2017 water samples including select geochemical measurements and host-rock setting from Fones et al. 2019.

| Sample | WAB104 | WAB105 | WAB55 | WAB188 | NSHQ04 | WAB71 | NSHQ14B (50 m) | NSHQ14C (85 m) |
|---|---|---|---|---|---|---|---|---|
| Lithology / Classification | Alkaline peridotite | Alkaline Peridotite | Contact | Contact | Hyperalk. Peridotite / Contact | Hyperalk. Peridotite | Hyperalk. Peridotite | Hyperalk. Peridotite |
| pH | 8.5 | 8.3[a] | 9.2 | 7.6[a] | 10.6[b] | 11.0 | 11.1 | 11.3 |
| Temperature | | | | | | | | |
| Conductivity (µS/cm) | 493 | 448 | 1,171 | 926 | 1,249 | 1,803 | 493 | 493 |
| ORP (mV) | 180[a] | 178[a] | 110[a] | 214 | -342[b] | -86[a] | -415 | -253 |
| DIC (mM) | 3.5 | 3.5 | 2.9 | 3.0 | 0.04 | 0.12 | 0.05 | 0.13 |
| $H_2$ (µM) | Bd | Bd | Bd | 0.92 | Bd | 0.51 | 21 | 164 |
| $CH_4$ (µM) | 0.02 | 0.02 | 0.10 | 1.69 | 155 | 12.6 | 34.6 | 12.6 |
| $SO_4^{2-}$ (µM) | 477 | 292 | 875 | 1,130 | 683 | 42 | 131 | 2 |
| $NO_3^-$ (µM) | 123 | 135 | 143 | 118 | 3.0 | 2.5 | Bd | Bd |
| $NO_2^-$ (µM) | Bd | Bd | 8.0 | 6.0 | 3.0 | 2.5 | Bd | Bd |

[a] Data are reported from the well in 2016 because these data are not available for 2017.
[b] Data are reported from the well in 2015 because these data are not available for 2016 or 2017.
Abbreviations: Hyperalk: hyperalkaline; Bd, below detection.

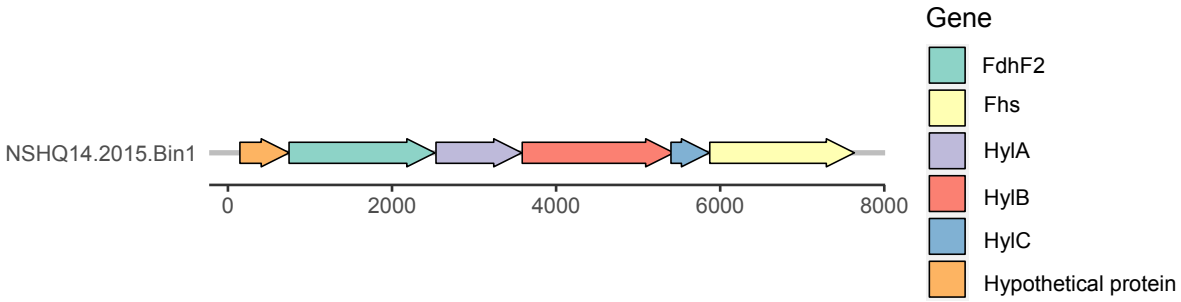



**Supplementary Fig. 1. Genomic co-localization of carbon monoxide dehydrogenase (Cdh)/Acetyl-CoA synthase (Acs) genes in Type I and II 'Acetothermia' metagenome assembled genomes (MAGs).** The Cdh or Acs designation is given for archaeal-like or bacterial-like complexes, as described in the primary text. Other abbreviations include: carbon monoxide dehydrogenase subunits C1 and F (CooC1, CooF, respectively). Genes are colored according to the protein homolog category indicated to the right and gene arrows are scaled to their length on the DNA contig sequence. The position of genes in contigs is given by the number below the gene graphic. The 2015 WAB188 Type I MAG and the 2015 NSHQ14 Type II MAGs were used for reference, but gene synteny was identical among other Type I and Type II MAGs analyzed in this study.




**Supplementary Fig. 2.** Maximum Likelihood (ML) phylogeny of the catalytic subunit the carbon monoxide/acetyl-CoA synthase (CODH/ACS) complex (CdhA) encoded by the Types I and II 'Acetothermia' in context of the 500 closest references from the NCBI nr database of bacterial-like (Type I) or archaeal-like (Type II) homologs (alignment length of 900 amino acid positions). The Types I and II 'Acetothermia' CdhA are indicated by orange and blue stars, respectively. The first annotation ring shows whether the CdhA derive from Archaea (red) or Bacteria (blue), while the second ring indicates the environment of origin for the CdhA homolog (as indicated by the legend on the left). Environmental metadata from the NCBI records were used to assess broad environmental classifications, and only homologs not deposited within the last year were used for the analyses. Black circles show >90% bootstrap support (out of 1,000 bootstraps). Branch

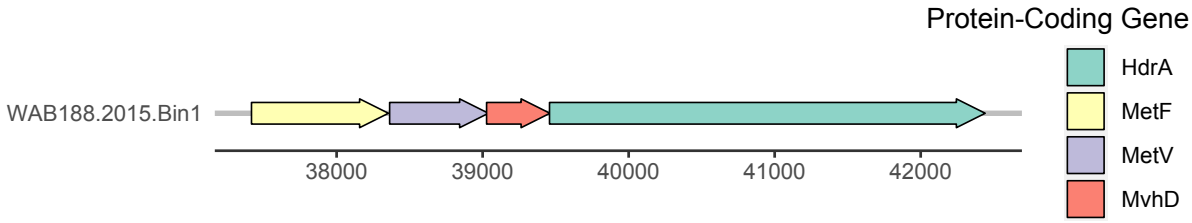
length is scaled based on the expected number of substitutions per site legend within the tree.



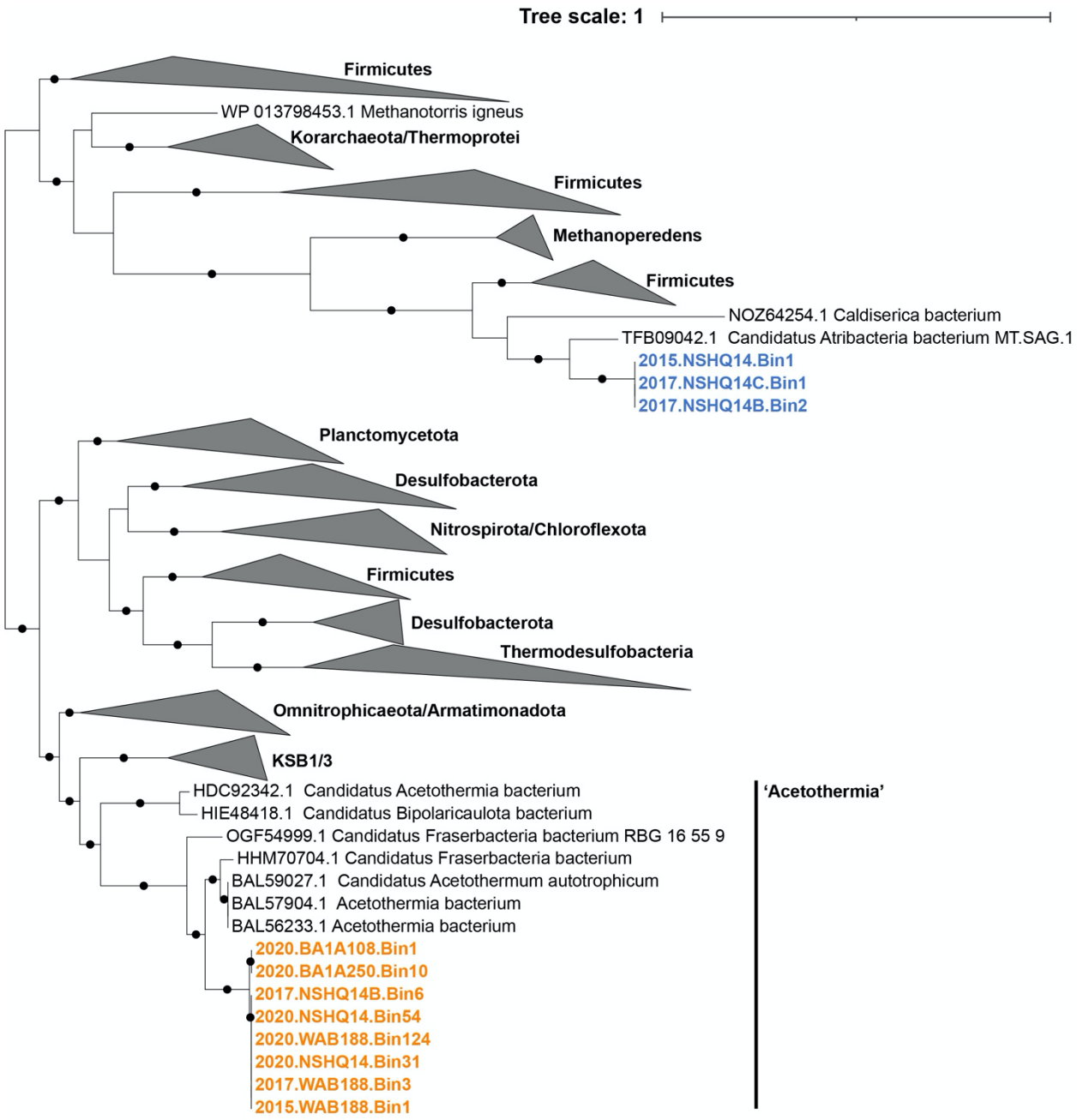

**Supplementary Fig. 3. Genomic co-localization of formate metabolizing genes involved in the western branch of the Wood-Ljungdahl pathway of the Type II 'Acetothermia' metagenome assembled genomes (MAGs). Abbreviations:** formate dehydrogenase (Fdh), formate-tetrahydrofolate—ligase (Fhs), hydrogenase-like protein (Hyl). Genes are colored according to the protein homolog category indicated to the right and gene arrows are scaled to their length on the DNA contig sequence. The position of genes within contigs is given by the number below the gene graphic. The 2015 NSHQ14 Type II MAG is used for reference.


**Supplementary Fig. 4. Phylogenetic Reconstruction of ATP synthase variants in the Types I and II 'Acetothermia' metagenome assembled genomes (MAGs).** A Maximum-Likelihood (ML) phylogenetic reconstruction is shown for the homologous F-Type ATPase B subunits and V-Type ATPase A subunits (n=1,344 alignment positions). The V-Type A subunits encoded by the Type I 'Acetothermia' are shown in orange, and the F-Type B subunit for the Type II 'Acetothermia' are shown in blue. Other clades are either shown in black text or are collapsed as triangles and represent selections from other major archaeal, eukaryal, and bacterial groups, as analyzed in Mulkidjanian et al. 2008

(1). Black circles indicate >90% bootstrap support (out of 1,000 bootstraps). The major separation between F-Type and V-Type ATPases is demarcated on the tree.




**Supplementary Fig. 5. Genomic co-localization of heterodisulfide reductase (Hdr) 'A' subunits and methylene-reductase (Met) 'V' and 'F' in Type I 'Acetothermia' metagenome assembled genomes (MAGs).** Other abbreviations include: methyl-viologen reducing hydrogenase subunit D (MvhD). Genes are colored according to the protein homolog category indicated to the right and gene arrows are scaled to their length on the DNA contig sequence. The position of genes within contigs is given by the number below the gene graphic. The 2015 WAB188 Type I MAG was used for reference.




**Supplementary Fig. 6. Phylogenetic placement of Types I and II 'Acetothermia' bacterial-like carbon monoxide dehydrogenase/acetyl-CoA synthase complexes (Cdh/Acs) (CdhA) proteins among other bacterial lineages.** **a**) Maximum-Likelihood (ML) phylogeny of the catalytic protein subunit of the carbon monoxide

dehydrogenase/acetyl-CoA synthase complexes (Cdh/Acs) encoded by the Type I and II 'Acetothermia' metagenome assembled genomes (MAGs) in context of other bacterial-like CdhA (alignment length of 854 amino acid positions). The Type II 'Acetothermia' bacterial-like CdhA are highlighted in bolded blue text. Note that these subunits are distinct from the archaeal-like CdhA that are components of archaeal-like Cdh complexes encoded by operons in the Type II 'Acetothermia' MAGs (**Fig. 4a**). The Type I 'Acetothermia' are highlighted in bolded orange text, nested within a larger 'Acetothermia' CdhA clade. Black circles show >90% bootstrap support for branches (out of 1,000 bootstraps). Branch length is scaled based on the expected number of substitutions per site legend in the top right. Clades are collapsed as triangles, with the taxonomic groups they correspond to indicated next to the triangles.

**Supplementary Dataset S1 (text file). Reference genomes used to reconstruct the bacterial domain phylogeny.**

**Supplementary Dataset S2 (XLS file). Protein identification in the Types I and II 'Acetothermia' referenced in the main text.** NCBI database identifications are shown for homologs that were identified by KEGG-based or BLAST-based searches for the 2015 and 2017 MAGs. JGI IMG gene accession numbers corresponding to homologs in the 2020 MAGs are also provided.

**Supplementary Dataset S3 (newick file).** Unrooted phylogenetic reconstruction of the bacterial domain based on 722 genomes.

**Supplementary References**


1. A. Y. Mulkidjanian, M. Y. Galperin, K. S. Makarova, Y. I. Wolf, E. V. Koonin, Evolutionary primacy of sodium bioenergetics. *Biol Direct* **3**, 13 (2008).
2. C. Furdui, S. W. Ragsdale, The role of pyruvate ferredoxin oxidoreductase in pyruvate synthesis during autotrophic growth by the Wood-Ljungdahl pathway. *J Biol Chem* **275**, 28494-28499 (2000).
3. J. Shin, Y. Song, Y. Jeong, B. K. Cho, Analysis of the core genome and pan-genome of autotrophic acetogenic bacteria. *Front Microbiol* **7**, 1531 (2016).
4. J. Martinussen, C. Sorensen, C. B. Jendresen, M. Kilstrup, Two nucleoside transporters in Lactococcus lactis with different substrate specificities. *Microbiology (Reading)* **156**, 3148-3157 (2010).