



OPEN ACCESS

EDITED BY

Philippe M. Oger,
UMR5240 Microbiologie, Adaptation et
Pathogenie (MAP), France

REVIEWED BY

Christopher L. Hemme,
University of Rhode Island, United States
Jia Wang,
The University of Tennessee, United States

*CORRESPONDENCE

Robert M. Kelly
✉ rmkelly@ncsu.edu

RECEIVED 26 April 2023

ACCEPTED 20 July 2023

PUBLISHED 03 August 2023

CITATION

Bing RG, Willard DJ, Crosby JR,
Adams MWW and Kelly RM (2023) Whither the
genus *Caldicellulosiruptor* and the order
Thermoanaerobacterales: phylogeny,
taxonomy, ecology, and phenotype.
Front. Microbiol. 14:1212538.
doi: 10.3389/fmicb.2023.1212538

COPYRIGHT

© 2023 Bing, Willard, Crosby, Adams and Kelly.
This is an open-access article distributed under
the terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Whither the genus *Caldicellulosiruptor* and the order Thermoanaerobacterales: phylogeny, taxonomy, ecology, and phenotype

Ryan G. Bing¹, Daniel J. Willard¹, James R. Crosby¹,
Michael W. W. Adams² and Robert M. Kelly^{1*}

¹Department of Chemical and Biomolecular Engineering, North Carolina State University, Raleigh, NC, United States, ²Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA, United States

The order Thermoanaerobacterales currently consists of fermentative anaerobic bacteria, including the genus *Caldicellulosiruptor*. *Caldicellulosiruptor* are represented by thirteen species; all, but one, have closed genome sequences. Interest in these extreme thermophiles has been motivated not only by their high optimal growth temperatures ($\geq 70^{\circ}\text{C}$), but also by their ability to hydrolyze polysaccharides including, for some species, both xylan and microcrystalline cellulose. *Caldicellulosiruptor* species have been isolated from geographically diverse thermal terrestrial environments located in New Zealand, China, Russia, Iceland and North America. Evidence of their presence in other terrestrial locations is apparent from metagenomic signatures, including volcanic ash in permafrost. Here, phylogeny and taxonomy of the genus *Caldicellulosiruptor* was re-examined in light of new genome sequences. Based on genome analysis of 15 strains, a new order, Caldicellulosiruptorales, is proposed containing the family Caldicellulosiruptoraceae, consisting of two genera, *Caldicellulosiruptor* and *Anaerocellum*. Furthermore, the order Thermoanaerobacterales also was re-assessed, using 91 genome-sequenced strains, and should now include the family Thermoanaerobacteraceae containing the genera *Thermoanaerobacter*, *Thermoanaerobacterium*, *Caldanaerobacter*, the family Caldanaerobiaceae containing the genus *Caldanaerobius*, and the family Calorimonaceae containing the genus *Calorimonas*. A main outcome of ANI/AAI analysis indicates the need to reclassify several previously designated species in the Thermoanaerobacterales and Caldicellulosiruptorales by condensing them into strains of single species. Comparative genomics of carbohydrate-active enzyme inventories suggested differentiating phenotypic features, even among strains of the same species, reflecting available nutrients and ecological roles in their native biotopes.

KEYWORDS

Caldicellulosiruptor, Thermoanaerobacterales, bacteria, thermophiles, phylogeny, ecology, fermentative anaerobes

1. Introduction

Fermentative anaerobes are biotechnologically relevant microorganisms; the push for renewable energy has identified them as microbial platforms because of their ability to convert plant polysaccharides, including in some cases microcrystalline cellulose, into industrially important products, such as acetone, butanol and ethanol (Bing et al., 2022). In fact, bioprocesses, used to produce industrial solvents prior to the emergence of petrochemicals, have received renewed attention in recent years with the development of molecular genetic tools to facilitate metabolic engineering of fermentative anaerobes (Lee et al., 2012; Charubin et al., 2018).

The taxonomy of fermentative anaerobes (Wagner and Wiegel, 2008), a taxonomically heterogeneous group, is in flux, and not without some controversy (Robitzski, 2022). For example, re-naming of the phylum 'Firmicutes' to 'Bacillota', which includes fermentative anaerobes, was proposed to reflect the fact that it also contains aerobic bacteria from the genus *Bacillus* (Oren and Garrity, 2021). The wide availability of genome sequences has displaced 16S rRNA gene sequence as the primary standard for differentiating one microbe from another. This may ultimately lead to purely alphanumeric designations for microorganisms, replacing the binomial nomenclature devised by Linnaeus centuries ago that still serves as the basis for the International Code of Nomenclature of Prokaryotes (Akst, 2021). For now, genus-species designations remain the taxonomic standard, such classifications can now be assessed using whole genome sequence comparisons. Absent from this characterization, however, is phenotypic information that is central to discerning features that ultimately define the microbiology. Phenotype can map to small sections of the genome not apparent from phylogenetic analysis or taxonomy.

Within the phylum Bacillota (formerly Firmicutes) are many thermophilic fermentative anaerobes currently assigned to the order Thermoanaerobacterales. These bacteria are sources of thermostable enzymes (i.e., glycoside hydrolases) capable of hydrolyzing a wide range of plant polysaccharides (Blumer-Schuette et al., 2008, 2014; Conway et al., 2016, 2018) present in lignocellulosic biomass (Zeldes et al., 2015; Straub et al., 2017, 2019, 2020). This biotechnological perspective serves as motivation to re-visit the taxonomy and phylogeny of fermentative anaerobes currently assigned to the order Thermoanaerobacterales to assess whether current classifications are consistent with emerging genome sequence information. Of particular interest here is the placement of those extremely thermophilic Thermoanaerobacterales ($T_{opt} > 70^{\circ}\text{C}$) that are currently all assigned to the genus *Caldicellulosiruptor* (Blumer-Schuette et al., 2010; Blumer-Schuette, 2020).

Table 1 lists the 13 species (and two strains) currently assigned to the genus *Caldicellulosiruptor* for which genome sequences are available, in addition to information about their genomes and isolation sites. Genomes average 2.71 ± 0.18 Mb with $2,604 \pm 181$ ORFs and G+C content of $35.6 \pm 0.5\%$. Note that it was recently proposed that *Caldicellulosiruptor acetigenus* is actually composed of three previously described species (*C. acetigenus*, *C. kristjanssonii*, *C. lactoaceticus*; Habib et al., 2021) and should be re-classified into a single species and associated strains. Members of *Caldicellulosiruptor* are globally distributed (Blumer-Schuette, 2020). *Caldicellulosiruptor saccharolyticus* (f. *Caldocellum saccharolyticum*; Luthi et al., 1990; Gibbs et al., 1992), the first named species discovered, was isolated

from a piece of wood downstream from a 78°C pool in Taupo, New Zealand, as were many other thermophilic isolates from similar sites in New Zealand that were not designated with taxonomic classifications at that time (Patel et al., 1986). From that group, *Caldicellulosiruptor danielii* and *Caldicellulosiruptor morgani* were eventually named when their genome sequences were reported (Lee et al., 2018). *Caldicellulosiruptor* species have since been isolated from globally diverse thermal features (e.g., Yellowstone National Park, United States; Hveragerði, Iceland; Nagano Prefecture, Japan; Kamchatka, Russia; Changbai, China) and, based on metagenomic data (Blumer-Schuette, 2020), yet uncharacterized species likely inhabit neutral, terrestrial, thermal features wherever they occur on earth. To date, *Caldicellulosiruptor bescii* (Kataeva et al., 2009; Yang et al., 2010; f. *Anaerocellum thermophilum*; Svetlichnyi et al., 1990) has been the focus of most studies of these bacteria, given the availability of molecular genetic tools (Chung et al., 2014; Lipscomb et al., 2016) and biotechnological potential (Lee et al., 2020; Bing et al., 2022) of this species. Not only has *C. bescii* been engineered to produce industrial chemicals (Williams-Rhaesa et al., 2018; Straub et al., 2020), but can also degrade and ferment transgenic lignocellulose (Straub et al., 2019). With the recent availability of additional complete genome sequences (Kataeva et al., 2009; Elkins et al., 2010; Lee et al., 2015; Mendoza and Blumer-Schuette, 2019; Habib et al., 2021; Bing et al., 2023b), it is essential that we re-visit the classification of the *Caldicellulosiruptor* and its taxonomic placement within the order Thermoanaerobacterales to give an updated phylogenetic perspective on these biotechnologically-important microorganisms.

2. Materials and methods

Unless noted otherwise, all computer software parameters used were default parameters.

2.1. Strain and genome sequence information

Genome assemblies for currently classified Thermoanaerobacterales (based on NCBI taxonomy, March 2023) were obtained from the National Center for Biotechnology Information (NCBI). The list of species strains and accession numbers are provided in Supplementary Table S1. 16S ribosomal RNA gene sequences were retrieved based on NCBI annotations for the most complete 16S rRNA sequence found in these assemblies.

Isolation location coordinates for the *Caldicellulosiruptor* were obtained primarily through isolation literature and NCBI biosample information. When needed, locations were estimated as 'best possible' from information contained in the literature (e.g., town names or geographical features). Supplementary Tables S2, S3 contain detailed information on the locations and their accuracy. Geographical distances between locations were calculated via a webserver employing the formula developed by Vincenty (1975).¹

¹ <https://stevemorse.org/nearest/distancebatch.html>

TABLE 1 Isolation and genome sequence information of species currently assigned to genus *Caldicellulosiraptor*.

Species	Isolation site	Isolation environment	Genome (Mb)	ORFs	G+C (%)	Ref
<i>saccharolyticus</i>	Taupo, New Zealand	Wood in hot spring pool (48°C)	2.97	2,924	35.2	van de Werken et al. (2008)
<i>changbaiensis</i>	Changbai Mountains, China	Hot spring sediment (83°C, pH 7)	2.91	2,833	35.1	Mendoza and Blumer-Schuette (2019)
sp. str. F32	Qingdao, China	Biocompost	2.38*	2,426	35.2	Ying et al. (2013)
<i>naganoensis</i>	Nozawaonsen, Nagano Pref., Japan	Hot spring mud & sediment (75–85°C, pH 9.0)	2.49	2,473	35.3	Lee et al. (2015)
<i>morganii</i>	Rotorua, New Zealand	Hot spring (63°C, pH 8.8)	2.48	2,413	36.5	Lee et al. (2015)
<i>danielii</i>	Waimangu, New Zealand	Hot spring	2.83	2,714	35.8	Lee et al. (2015)
<i>hydrothermalis</i>	Pauzhetka, Kamchatka, Russia	Terrestrial neutral hot spring	2.77	2,679	36.1	Blumer-Schuette et al. (2011)
<i>diazotrophicus</i>	Nakabusa, Nagano Pref., Japan	Hot spring N ₂ -poor biomats (78.3°C, pH 8.5–8.9)	2.60	2,449	34.8	Chen et al. (2021b)
<i>owensensis</i>	Owens Lake, CA, United States	Freshwater pond sediment in Dry Lake Bed (32°C, pH 9.0)	2.43	2,333	35.4	Blumer-Schuette et al. (2011)
<i>obsidiansis</i>	Yellowstone Nat. Park, WY, United States	Hot spring (66°C, pH 5.0)	2.53	2,404	35.2	Elkins et al. (2010)
<i>bescii</i>	Valley of Geysers, Kamchatka, Russia	Hot spring caused waterlogged foot of a geyser (90°C, pH 6.5)	2.93	2,828	35.2	Yang et al. (2010)
<i>kronotskyensis</i>	Valley of Geysers, Kamchatka, Russia	Terrestrial neutral hot spring	2.84	2,623	35.1	Blumer-Schuette et al. (2011)
<i>acetigenus (acetigenus)</i>	Hveragerði-Hengill, Iceland	Hot spring biomats and sediments (55–75°C, pH 8.0–8.7)	2.74	2,643	36.3	Onyenwoke et al. (2006) and Habib et al. (2021)
<i>acetigenus (kristjanssonii)</i>			2.80	2,712	36.0	
<i>acetigenus (lactoaceticus)</i>			2.67	2,601	36.1	

*Genome sequence not closed; 127 contigs.

2.2. Phylogenetic tree inference: 16S ribosomal RNA and genome taxonomy database bac120

16S rRNA gene sequences were aligned using Clustal Omega (v. 1.2.4; Sievers et al., 2011) using flags “--outfmt fa --distmat-out --full --full-iter --percent-id --guidetree-out.” The percent identities output was used for 16S rRNA identity matrices (Supplementary Table S8), which were color coded based on proposed 16S rRNA taxonomic rank “Minimum Sequence ID” values (Yarza et al., 2014). The 16S rRNA alignment was inputted into FastTree (v. 2.1.11; Price et al., 2010) to generate a distance tree using flags “-nt -gtr -gamma.” The 16S rRNA tree was mid-point rooted with Dendroscope 3 (Huson and Scornavacca, 2012).

The bac120 gene marker set from the Genome Taxonomy Database (GTDB) was also used to infer a phylogenetic tree from the same set of Thermoanaerobacterales (based on NCBI taxonomy, March 2023) using GTDB-Tk v2.1.0 (Parks et al., 2018, 2020, 2021; Chaumeil et al., 2019, 2022). The phylogenetic tree was inferred using the “de_novo_wf” workflow with 200 aa per gene marker and using the phylum Thermodesulfobiota as the out-group for tree rooting. The GTDB reference database was excluded to allow for a custom taxonomic classification. Both the 16S rRNA and bac120 trees were visualized and formatted using the Interactive Tree of Life (iTOL; Letunic and Bork, 2006). A color-blind friendly divergent color palette was applied to the different genera (determined by Chroma.js Color

Palette Helper)² (colors: #1d1b73, #243694, #2b53af, #2d74be, #009ba7, #b1ab14, #ba971c, #c5801d, #d36219, #f00000). The same genera colors were applied to the 16S rRNA tree.

2.3. Average identity calculations and pangenome analysis

Average nucleotide identities (ANI) were generated with pyani (v. 0.2.12; Pritchard et al., 2016) average_nucleotide_identity.py BLAST+ method (ANiB; flag “-m ANiB”). Final bidirectional ANiB values were generated by averaging the ANiB values representing the two orientations of the same strain pairs (Supplementary Table S6). Average amino acid identities (AAI) were calculated with GET_HOMOLOGUES (v22082022; Contreras-Moreira and Vinuesa, 2013; Vinuesa and Contreras-Moreira, 2015) with flags “-A -t 0 -M (Supplementary Table S7).” A composite ANiB/AAI matrix of the 91 genomes was calculated and heat-mapped based on the min/max values (40.4–100%) using the Inferno colormap. This same colormap scale was maintained for all ANiB/AAI figures.

Core and pangenome analysis were also completed using GET_HOMOLOGUES. For this, the flags “-A -t 0 -c -z -P -M” were used.

² <https://vis4.net/palettes>

Then compare_clusters.pl. was used to generate the core, soft core, and pangenomes. Finally, parse_pangenome_matrix.pl. with “-f core_Tettelin” was used to generate pan/core genome graphs (Supplementary Data S2).

2.4. Analysis of carbohydrate active enzymes (CAZymes)

For annotation of carbohydrate active enzymes, dbCAN (v 3.0.7; Zhang et al., 2018; Zheng et al., 2023) was locally run with the run_dbcan command using protein.faa files of translated nucleotide coding sequences as input and flags “--signalP=true --gram P” to use SignalP v 4.1 (Nielsen, 2017; Supplementary Data S2). The dbCAN output was then curated where all glycosyl transferases (GTs) were removed, and a consensus was generated between the three prediction algorithms (HMMER, eCAMI, and DIAMOND), where annotations that had agreement between at least 2 prediction algorithms was used as the consensus. Any proteins that had positive hits on only a single algorithm were subjected to a BLAST search on the NCBI webserver against the NCBI non-redundant protein sequences database to obtain an annotation. Any coding sequence deemed to not be involved in carbohydrate catabolism, based on the BLAST search annotation, were removed from the analysis, along with any coding sequences corresponding to proteins under 100 amino acids (Supplementary Table S9). For the Caldicellulosiruptorales, remaining proteins representing the “CAZysome” of each strain were input into GET_HOMOLOGUES for Pan/Core/Soft-core analysis, as described above for whole genomes. The resulting Core/Soft-cores of the Caldicellulosiruptorales were then analyzed for CAZyme annotation and predicted function (Supplementary Table S4).

2.5. Substrate use evaluation of *Caldicellulosiruptor diazotrophicus*

Three strains of Caldicellulosiruptorales were obtained from DSMZ-German Collection of Microorganisms and Cell Cultures GmbH: DSM 6725 (*C. bescii*), DSM 18901 (*C. hydrothermalis*), and DSM 112098 (*C. diazotrophicus*). Strains were adapted to substrates and grown in 125 mL serum bottles, as described previously (Bing et al., 2023a) where modified D671 media with cellobiose (Biosynth-Carbosynth, OC04040) was used for routine growth. All cultures were grown at 75°C with 150 rpm shaking in a New Brunswick Innova 42 incubator shaker. Adaption to beechwood xylan (Biosynth-Carbosynth, YX45751), wet-milled corn fiber (WMCF, provided by Novozymes A/S), or Avicel PH-101 (Millipore-Sigma) was done by passages of 5×10^8 cells (final 1×10^7 cells/ml starting cell density) from 5 g/L cellobiose cultures to 0.5 g/L cellobiose and 4.5 g/L carbohydrate equivalent of substrate (xylan, WMCF, or Avicel). Cell growth was monitored by epifluorescence microscopy, as previously described (Bing et al., 2023a). Once cells reached 5×10^8 cells/ml or after 3 days, 5×10^8 cells were passaged to 5 g/L carbohydrate equivalent of substrate alone. Culture growth was monitored for 7 days. *C. bescii* (DSM 6725) and *C. hydrothermalis* (DSM 18901) were included as controls for expected growth phenotypes on the substrates. All strains were expected to grow on cellobiose, xylan, and WMCF, but only *C. bescii* grows on microcrystalline cellulose (Avicel). The phenotype observed

is robust growth of *C. bescii* on all substrate adaptations and final passages; *C. hydrothermalis* had robust growth on cultures containing cellobiose, xylan, and WMCF, weak growth on the adaption passage to Avicel (containing 0.5 g/L cellobiose) and no growth on Avicel alone.

3. Results

3.1. Proposed changes to the classification of the Thermoanaerobacterales and the genus *Caldicellulosiruptor*

In order to assess the existing taxonomic classification for the genus *Caldicellulosiruptor*, phylogenetic trees were inferred for a set of selected genomes within the order Thermoanaerobacterales (according to NCBI Taxonomy as of March 2023) using both 16S ribosomal RNA gene sequences to represent the existing taxonomic structure and the bac120 gene markers from the Genome Taxonomy Database (GTDB) to assess reclassification (Figure 1). The larger bac120 gene marker set allowed a multiple sequence alignment of ~24,000 amino acids, enabling more robust placement of the selected genomes within the tree and highlights the several areas needed for reclassification. Note that recent publications have already moved *Thermodesulfobium* to its own phylum Thermodesulfobiota (Frolov et al., 2023) and the genera *Calderihabitans*, *Desulfotibacter*, *Moorella*, and *Zhaonella* to the novel order Moorellales (Lv et al., 2020); however, these changes were not yet reflected in the NCBI taxonomy at the time of this analysis.

The 16S ribosomal RNA gene sequence tree includes multiple nodes at the genus level and higher with bootstrap values ≤ 0.50 , indicating low confidence some of the taxonomic arrangements in this tree. Meanwhile, the bac120 gene marker tree has ≥ 0.50 bootstrap values (and ≥ 0.98 with the exception of the *Calorimonas adulescens* node) for all nodes above the species level, implying strong phylogenetic relationships above species-level classification. Based on the bac120 gene marker phylogeny, the Order Thermoanaerobacterales should consist of the Family Thermoanaerobacteraceae containing the genera *Thermoanaerobacter*, *Thermoanaerobacterium*, and *Caldanaerobacter*, the Family Caldanaerobieaceae containing the genus *Caldanaerobius*, and the Family Calorimonaceae containing the genus *Calorimonas*. The current genus *Caldicellulosiruptor* is sufficiently divergent from the order Thermoanaerobacterales that it should be long to a separate Order, Caldicellulosiruptorales, and Family, Caldicellulosiruptoraceae. Within the Family, the species *C. danielli*, *C. hydrothermalis*, *C. diazotrophicus*, *C. owensensis*, *C. obsidiansis*, *C. bescii*, *C. kronotskyensis*, *C. acetigenus*, *C. lactoaceticus*, and *C. kristjanssonii* should be moved to a new genus; we propose this genus be named “*Anaerocellum*” in view of the fact that the type strain *C. bescii* was originally named *Anaerocellum thermophilum* (Svetlichnyi et al., 1990; Kataeva et al., 2009; Yang et al., 2010). The species *C. changbaiensis*, *C. sp. F32*, *C. saccharolyticus*, *C. naganoensis*, and *C. morganii* should remain in the genus *Caldicellulosiruptor*, recognizing that the type strain from this group was *C. saccharolyticus* (Rainey et al., 1994).

Together, the orders Thermoanaerobacterales and Caldicellulosiruptorales likely comprise their own Class (Thermoanaerobacteria), but without comprehensive evaluation of the Class Clostridia, this could not be fully established here

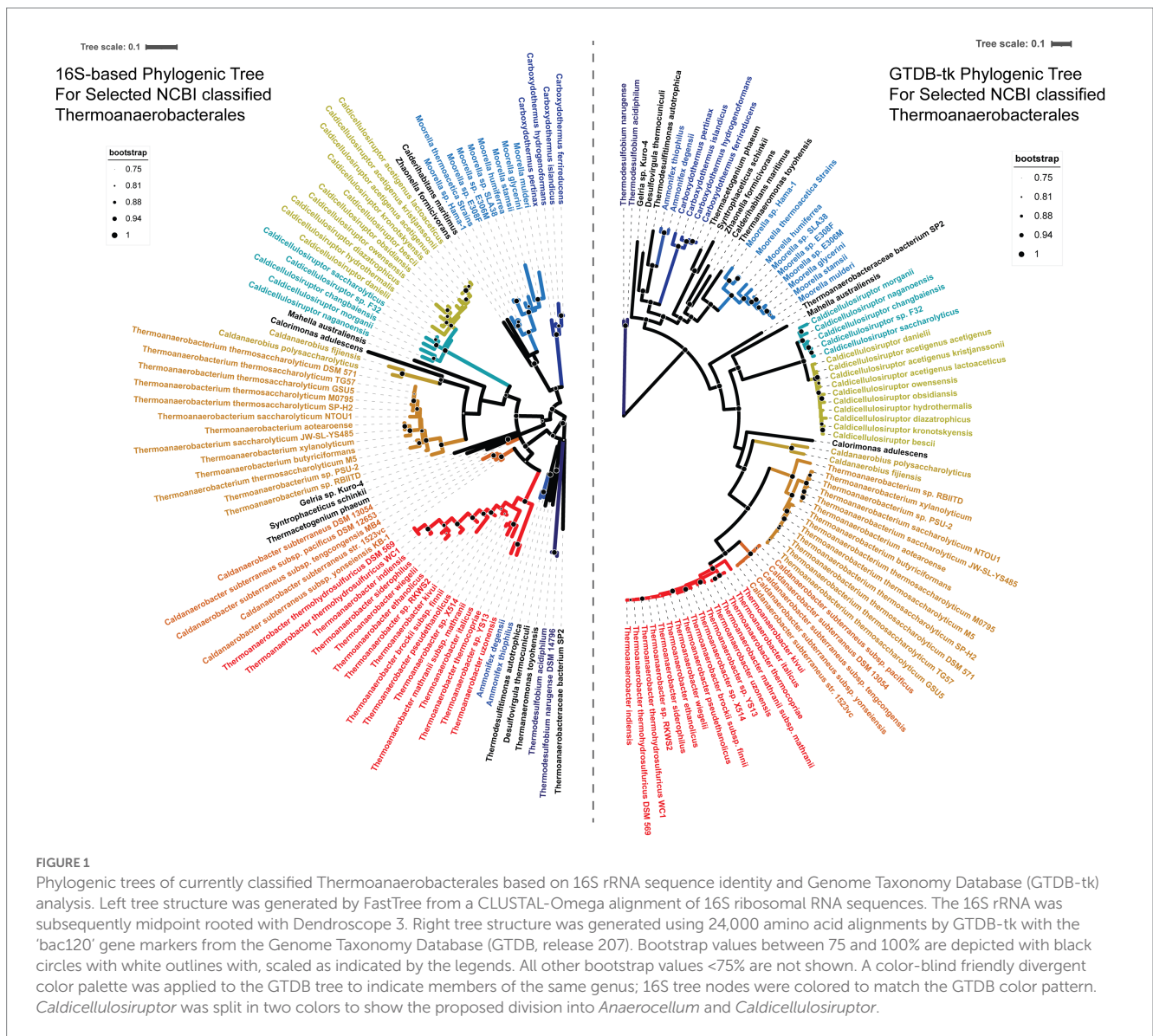


FIGURE 1
Phylogenetic trees of currently classified Thermoanaerobacterales based on 16S rRNA sequence identity and Genome Taxonomy Database (GTDB-tk) analysis. Left tree structure was generated by FastTree from a CLUSTAL-Omega alignment of 16S ribosomal RNA sequences. The 16S rRNA was subsequently midpoint rooted with Dendroscope 3. Right tree structure was generated using 24,000 amino acid alignments by GTDB-tk with the 'bac120' gene markers from the Genome Taxonomy Database (GTDB, release 207). Bootstrap values between 75 and 100% are depicted with black circles with white outlines with, scaled as indicated by the legends. All other bootstrap values <75% are not shown. A color-blind friendly divergent color palette was applied to the GTDB tree to indicate members of the same genus; 16S tree nodes were colored to match the GTDB color pattern. *Caldicellulosiruptor* was split in two colors to show the proposed division into *Anaerocellum* and *Caldicellulosiruptor*.

(Supplementary Table S5). The placement of *Mahella australiensis* in the bac120 gene marker tree further supports the need for a broader evaluation of the Class Clostridia. The GTDB taxonomy (~5,000 amino acid alignment) keeps the genus *Mahella* within Clostridia, but our ~24,000 amino acid alignment suggests that *Mahella* is more closely related to the Caldicellulosiruptorales, meriting a Family-level classification within the Order Caldicellulosiruptorales.

Reclassification at the species level is further supported by Average Nucleotide Identity (ANI) and Average Amino Acid Identity (AAI) comparisons of genome-sequenced strains currently assigned to the Order Thermoanaerobacterales (Figure 2), including 15 strains within the genus *Caldicellulosiruptor* (Table 1). Color coding in Figure 2B reflects previously proposed taxonomic grouping thresholds for 16S rRNA gene sequence identities (Yarza et al., 2014). Although the 16S rRNA gene sequence analysis is less clear than ANI/AAI, as expected, only a few current Thermoanaerobacterales genera are indicated as having Order level or lower relatedness by 16S rRNA.

To further investigate the proposed re-classification of the Thermoanaerobacterales, a genome-wide assessment was done

(Figure 3). To investigate species-level classification, we assume that ANI values $\geq 94\text{--}96\%$ are indicative of strains of the same species (Richter and Rossello-Mora, 2009; Kim et al., 2014). ANI/AAI results indicate several species from the genera *Thermoanaerobacter*, *Moorella*, *Carboxythermus*, *Thermoanaerobacterium*, *Caldicellulosiruptor*, and the new genus *Anaerocellum* need to be consolidated as strains of previously described species (Table 2). For example, *Thermoanaerobacter ethanolicus* (type strain JW200, the first isolate from this group) should now include an additional 6 strains currently assigned as *T. thermohydrosulfuricus* WC1, *T. thermohydrosulfuricus* DSM569, *T. sp. RKWS2*, *T. siderophilus* SR4, *T. weigelii* RT8.B1, and *T. indiensis* BSB-33, all of which share ANI $\geq 97\%$.

Within the Caldicellulosiruptorales, several species level reclassifications are appropriate. The recently proposed reclassifying of *A. lactoaceticus* and *A. kristjanssonii* as subspecies of *A. acetigenus* (Habib et al., 2021) is supported by ANI values all $\geq 97\%$ (Figure 3). Note that *A. acetigenus* was previously reclassified from *Thermoanaerobium acetigenus* based on 16S rRNA gene sequence

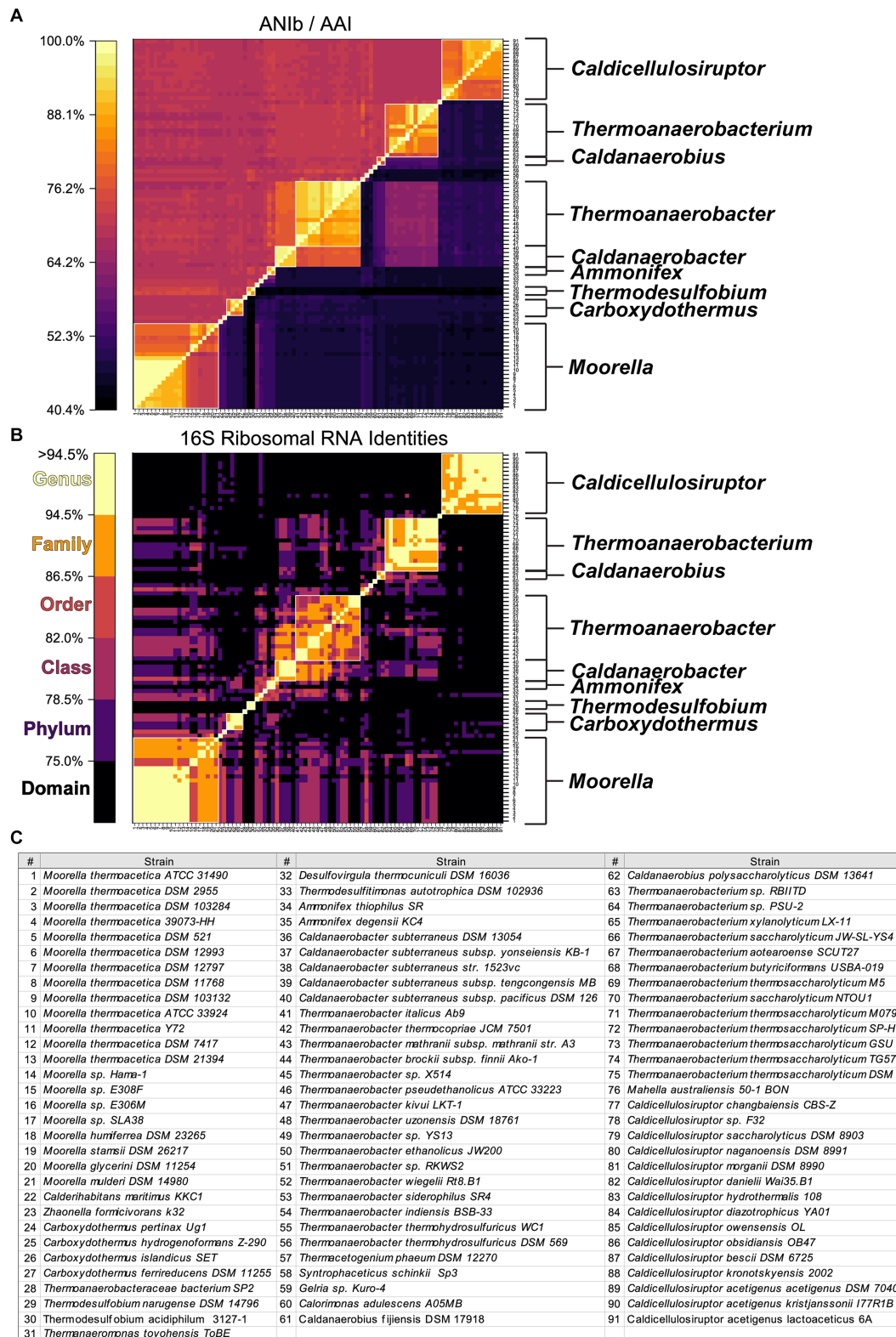


FIGURE 2 Identity matrix assessment of 91 currently classified Thermoanaerobacterales. **(A)** Whole-genome average nucleotide identity (ANIb, upper left triangle) and average amino acid identity (AAI, lower left triangle). **(B)** 16S ribosomal RNA identities from CULSTAL-Omega alignment, color map of taxonomic ranks is based on "Minimum Sequence ID" (Yarza et al., 2014). **(C)** Strain name to number key for (A,B). White outline boxes indicated current NCBI genus classifications, genera with more than 1 strain present are labeled to the left of each matrix in (A,B).

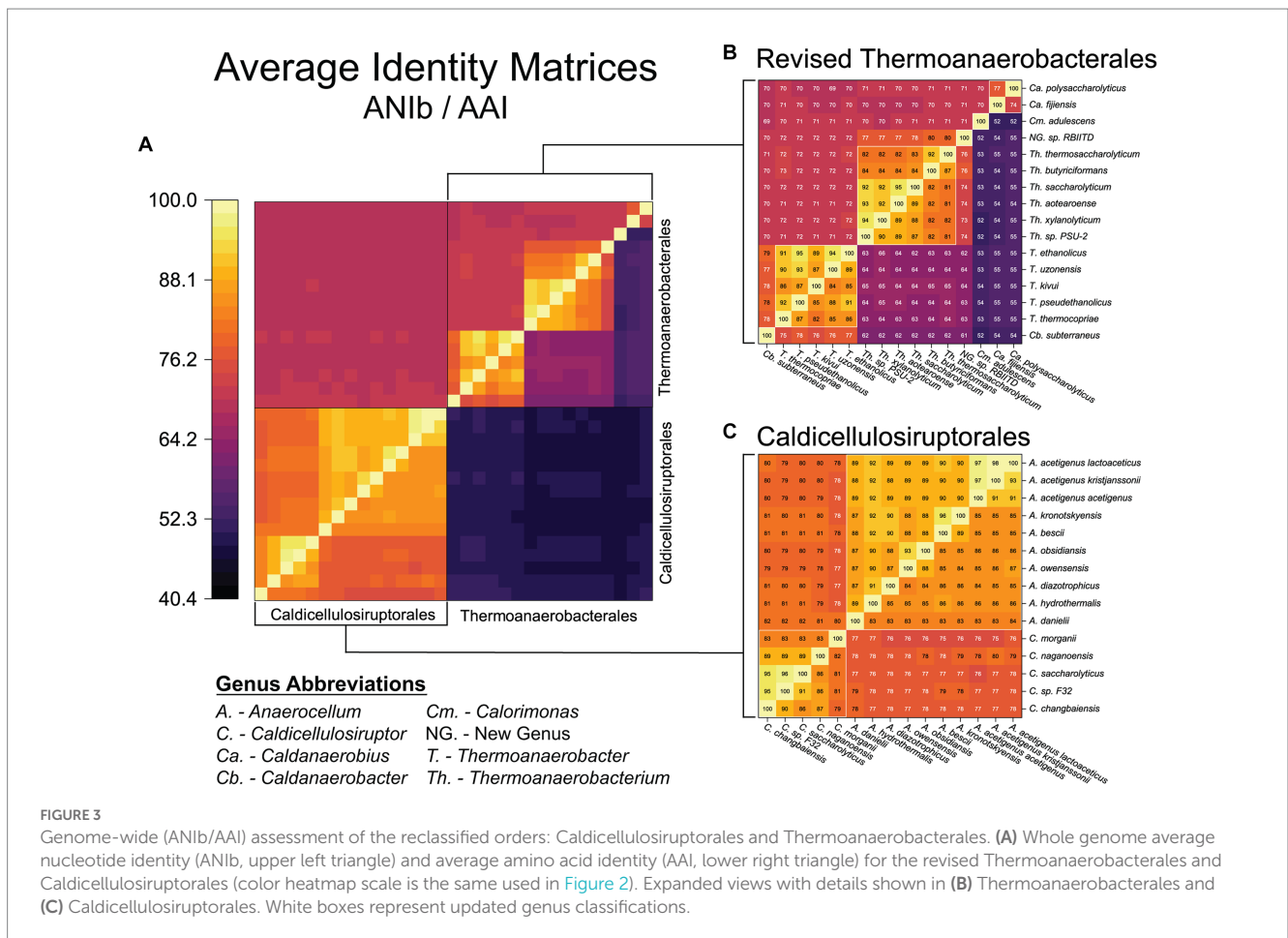


FIGURE 3 Genome-wide (ANiB/AAI) assessment of the reclassified orders: Caldicellulosiruptorales and Thermoanaerobacteriales. **(A)** Whole genome average nucleotide identity (ANiB, upper left triangle) and average amino acid identity (AAI, lower right triangle) for the revised Thermoanaerobacteriales and Caldicellulosiruptorales (color heatmap scale is the same used in Figure 2). Expanded views with details shown in **(B)** Thermoanaerobacteriales and **(C)** Caldicellulosiruptorales. White boxes represent updated genus classifications.

and physiological properties (Onyenwoke et al., 2006). What we now term *A. bescii* and *A. kronotskyensis* should be considered the same species, given that their ANI is 96% (Table 2; Figure 3); as such, *A. kronotskyensis* 2002 is designated as *A. bescii* strain 2002. Two other *kronotskyensis* strains (2006, 2,902) were classified as this species based on 16S rRNA (Miroshnichenko et al., 2008) but have no available genome sequences. These may also be strains of *A. bescii*, but without genome sequences this cannot be verified.

Note that the inter-genus ANI of the Caldicellulosiruptorales is ≤82%, indicative of the divergence between the genera *Anaerocellum* and *Caldicellulosiruptor*, established by the GTDB-tk analysis. The *Caldicellulosiruptor* genus share an ANiB >83% (89% if *C. morganii* is excluded; Figure 3). Note that *C. morganii* is the most divergent of the Caldicellulosiruptorales by ANI/AAI, although the bac120 (GTDB-tk) phylogenetic tree places it squarely within the *Caldicellulosiruptor* genus (Figure 1). A case can be made that *C. saccharolyticus*, *C. sp. F32*, and *C. changbaiensis* CBZ are the same species (ANI's ≥ 95% and agreement with bac120 phylogenetic inference; Table 2); in fact, *C. saccharolyticus* and *C. sp. F32* have been designated as such by the China General Microbiological Culture Collection (CGMCC). However, the *C. sp. F32* genome is currently in 127 contigs, thus lowering the confidence in its taxonomic placement compared to other *Caldicellulosiruptor* strains.

The core genomes of various taxonomic levels were evaluated in the process of whole-genome analysis. The core and pan genomes of the Thermoanaerobacteraceae are 306 / 6,915 for the family, 1,308 /

3,510 for *Caldanaerobacter*, 815 / 4,087 for *Thermoanaerobacter*, and 1,374 / 4,419 for *Thermoanaerobacterium*. Within the Caldicellulosiruptoraceae, the core and pan genomes are: 1,248 / 3,833 for the family, 1,496 / 3,527 for *Anaerocellum*, and 1,367 / 3,027 for *Caldicellulosiruptor* (Supplementary Figure S1). The pangenome of the Caldicellulosiruptoraceae remains open, implying more genetic diversity within the Family remains to be discovered.

The core and pan genomes of the Thermoanaerobacteraceae, respectively are: 306/6,915 for the family, 1,308/3,510 for *Caldanaerobacter*, 815/4,087 for *Thermoanaerobacter*, and 1,374/4,419 for *Thermoanaerobacterium* (Supplementary Figure S1).

3.2. Global distribution of the order Caldicellulosiruptorales and relationship to taxonomy

The wide global distribution of the Caldicellulosiruptorales is evident from the isolation sites of currently named species as well as from signatures detected in community analyses (Figure 4; Blumer-Schuette, 2020). Given the closely related microbiological features of members of the Caldicellulosiruptorales, it is interesting to consider how these species became globally distributed and the relationships between geography, physiochemical features of isolation sites, and strain relatedness. The fact that all known members of the Caldicellulosiruptorales

TABLE 2 Proposed species-level reclassification of species currently assigned to order Thermoanaerobacterales.

Proposed species name	Strain (T type)	Heterotypic synonyms based on NCBI taxonomy (March 2023)	Justification/notes
<i>Thermoanaerobacter ethanolicus</i>	JW200 ^T	<i>Thermoanaerobacter ethanolicus</i>	ANI/AAI (97–99/90–94%) for all strains general agreement with GTDB-tk. <i>T. weigelli</i> listed as separate species in ~5,000aa MSA GTDB taxonomy, but has >97% ANI with all strains and agrees with ~20,000aa MSA GTDB-tk analysis.
	SR4	<i>Thermoanaerobacter siderophilus</i>	
	BSB-33	<i>Thermoanaerobacter indiensis</i>	
	WC1	<i>Thermoanaerobacter thermohydrosulfuricus</i>	
	L77-66	<i>Thermoanaerobacter thermohydrosulfuricus</i>	
	RT8.B1	<i>Thermoanaerobacter weigelli</i>	
	RKWS2	<i>Thermoanaerobacter</i> sp. RKWS2	
<i>Thermoanaerobacter pseudethanolicus</i>	39E ^T	<i>Thermoanaerobacter pseudethanolicus</i>	ANI/AAI (98–100/94–96%). GTDB-tk agreement.
	AKo-1	<i>Thermoanaerobacter brockii</i> subsp. <i>finnii</i>	
	X514	<i>Thermoanaerobacter</i> sp. X514	
<i>Thermoanaerobacter uzonensis</i>	JW/IW-010 ^T	<i>Thermoanaerobacter uzonensis</i>	ANI/AAI (97/91%). GTDB-tk agreement.
	YS13	<i>Thermoanaerobacter</i> sp. YS13	
<i>Thermoanaerobacter thermocopriae</i>	JT3-3 ^T	<i>Thermoanaerobacter thermocopriae</i>	ANI/AAI (98–99/93%). GTDB-tk agreement.
	Ab9	<i>Thermoanaerobacter italicus</i>	
	A3	<i>Thermoanaerobacter mathranii</i> subsp. <i>mathranii</i>	
New species	AMP ^T	<i>Moorella thermoacetica</i>	GTDB-tk agreement. ANI/AAI (93/85%) support split from other <i>M. thermoacetica</i> as a new species.
<i>Moorella</i> sp. Hama-1	Hama-1 ^T	<i>Moorella</i> sp. Hama-1	ANI/AAI & GTDB-tk agreement
<i>Moorella</i> sp. E308F	E308F ^T	<i>Moorella</i> sp. E308F	ANI/AAI (99/90%), GTDB-tk agreement.
	E306M	<i>Moorella</i> sp. E306M	
<i>Moorella</i> sp. SLA38	SLA38 ^T	<i>Moorella</i> sp. SLA38	ANI/AAI, GTDB-tk agreement.
<i>Carboxydotherrmus hydrogenoformans</i>	Z-2901 ^T	<i>Carboxydotherrmus hydrogenoformans</i>	ANI/AAI (98/91%). GTDB-tk agreement.
	JW/AS-Y7	<i>Carboxydotherrmus ferrireducens</i>	
<i>Thermoanaerobacterium aotearoense</i>	SCUT27 ^T	<i>Thermoanaerobacterium aotearoense</i>	ANI/AAI (100/94%), GTDB-tk agreement. <i>T. saccharolyticum</i> NT0U1 and JW-SL-YS485 are separated
	JW-SL-YS485	<i>Thermoanaerobacterium saccharolyticum</i>	
<i>Thermoanaerobacterium</i> sp. PSU-2	PSU-2 ^T	<i>Thermoanaerobacterium</i> sp. PSU-2	ANI/AAI and GTDB-tk agreement
New genus sp. RBIITD	RBIITD ^T	<i>Thermoanaerobacterium</i> sp. RBIITD	ANI/AAI/GTDB-tk indicated as species of a new genus.
<i>Bioaibacter acetigenes</i>	SP2	Thermoanaerobacteraceae bacterium SP2	~5,000aa MSA GTDB taxonomy placed this unclassified Thermoanaerobacterales as <i>B. acetigenes</i>
<i>Syntrophomonas</i> sp. UBA4844	Kuro-4	<i>Gelria</i> sp. Kuro-4	<i>Gelria</i> sp. Kuro-4 is likely a member of <i>Syntrophomonas</i> based on ~5,000aa MSA GTDB taxonomy.
<i>Caldicellulosiruptor saccharolyticus</i>	Tp 8 T.6.3.3.1 ^T	<i>Caldicellulosiruptor saccharolyticus</i>	ANI/AAI: <i>Csac</i> /CF32 (96/91%), <i>Csac</i> / <i>Ccha</i> : (95/86%), <i>Ccha</i> /CF32: (95/90%) GTDB-tk agreement
	CBZ	<i>Caldicellulosiruptor changbaiensis</i>	
	F32	<i>Caldicellulosiruptor</i> sp. F32	
<i>Anaerocellum besicii</i>	Z-1320 ^T	<i>Anaerocellum besicii</i>	<i>Abes</i> / <i>Akro</i> ANI/AAI (96/89%). GTDB-tk agreement.
	2002	<i>Anaerocellum kronotskyensis</i> 2002	

grow best at temperatures above 70°C differentiates them from almost all other characterized bacteria, which are mostly mesophilic or moderately thermophilic ($T_{opt} \leq 65^\circ\text{C}$). Community analyses of terrestrial hot springs indicate that, above 65°C, Caldicellulosiruptorales dominate (Vishnivetskaya et al., 2015;

Lee et al., 2018). Based on the global presence of Caldicellulosiruptorales in low-salinity terrestrial thermal sites (>65°C) and the relatedness of strains isolated in various regions, additional areas likely to harbor novel strains of Caldicellulosiruptorales can be identified. These areas are detailed

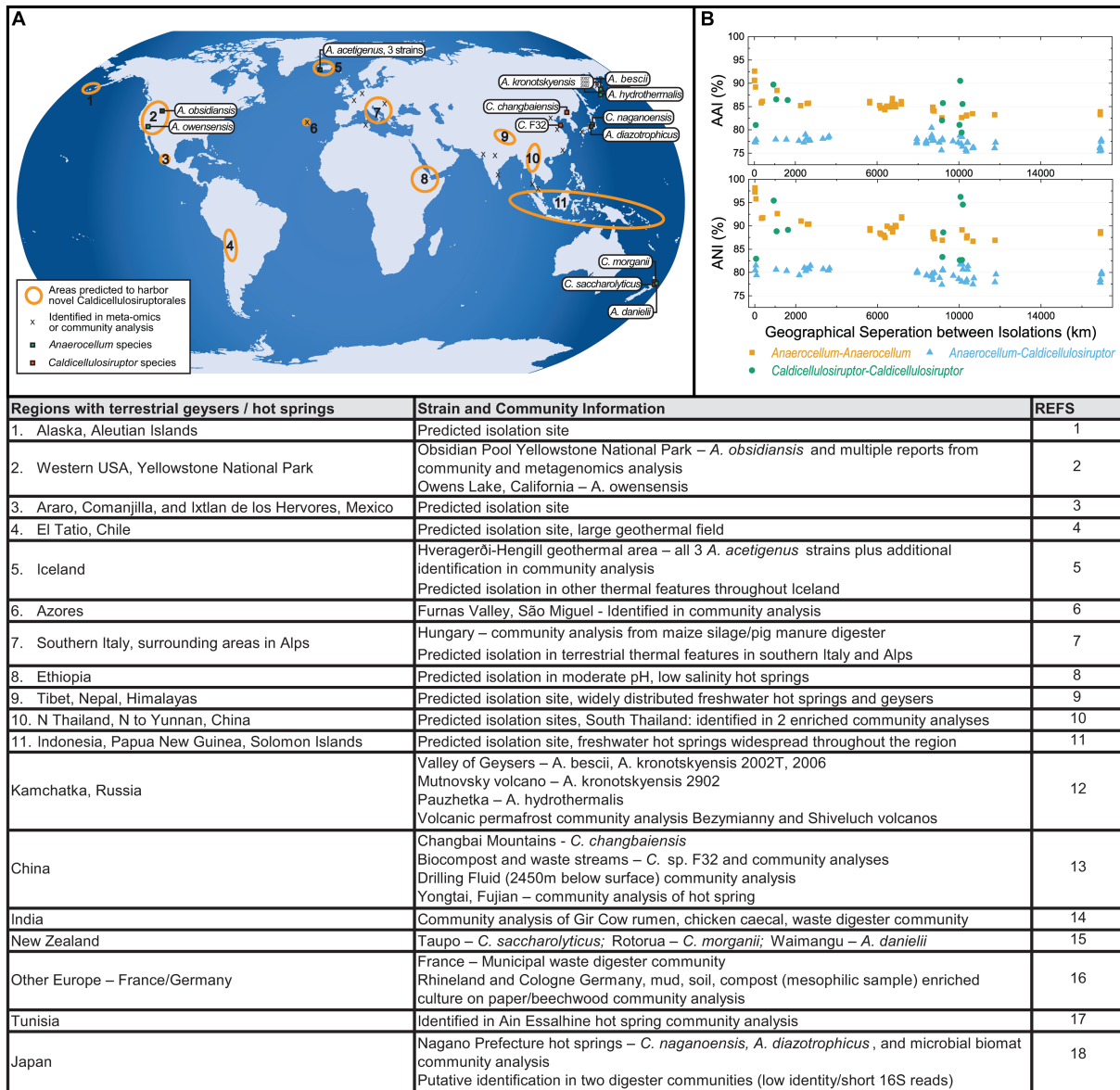


FIGURE 4 Biogeographic distribution of current and putative isolation sites for Caldicellulosiruptorales. **(A)** Strains with genome sequences are labeled, and isolation sites marked. Presence of Caldicellulosiruptorales in multi-omics and community analysis are marked with "X." Circled areas are predicted to harbor Caldicellulosiruptorales. Listed in the table below are details of these areas and isolation sites. **(B)** Whole genome average amino acid identity (AAI) and average nucleotide identity (ANI) for strain pairs with genome sequences plotted against the distance separating their isolation sites; color coded based on pairings within and outside the two proposed genera (*Caldicellulosiruptor* and *Anaerocellum*). Reference list: 1 - (Alaska Department of Natural Resources, 1984); 2 - (Berry et al., 1980; Elkins et al., 2010; Hamilton-Brehm et al., 2010; Vishnivetskaya et al., 2015; Lee et al., 2018); 3, 8, 9, 11 - (Johnson, 2010); 4 - (Fernandez-Turiel et al., 2005); 5 - (Nielsen, 1993; Mladenovska et al., 1995; Sonne-Hansen and Ahring, 1997; Bredholt et al., 1999; Onyenwoke et al., 2006; Orlygsson et al., 2010); 6 - (Sahm et al., 2013); 7 - (Johnson, 2010; Wirth et al., 2012); 10 - (Johnson, 2010; Hniman et al., 2011); 12 - (Rainey et al., 1993; Miroschnichenko et al., 2008; Sahm et al., 2013; Vishnivetskaya et al., 2015, 2022); 13 - (Zhang et al., 2007, 2016; Qiu et al., 2011; Ying et al., 2013; Bing et al., 2015; Liu et al., 2015); 14 - (Pandit et al., 2016, 2018a,b); 15 - (Sissons et al., 1987; Rainey et al., 1993, 1994; Lee et al., 2015, 2018); 16 - (Svetlitchnyi et al., 2013; Lu et al., 2014); 17 - (Sayeh et al., 2010); 18 - (Taya et al., 1988; Rainey et al., 1993; Narihiro et al., 2009; Lee et al., 2015, 2018; Cheng et al., 2018; Nishihara et al., 2018; Chen et al., 2021a,b).

in Figure 4, both circled in the map (Figure 4A) with details provided below the map. There is a notable absence of isolated Caldicellulosiruptorales from Africa and South America, although there are thermal features on these continents hospitable to Caldicellulosiruptorales. Presence of widespread thermal features and high similarity of strains currently isolated from North America and Iceland may indicate that additional more divergent

Caldicellulosiruptorales exist in these areas, which would reflect the diversity seen in China, Kamchatka, Japan, and New Zealand.

It is also interesting to consider how the globally distributed Caldicellulosiruptorales compare with respect to geographic separation of isolation sites and overall genetic relatedness. Figure 4B shows how geographical distance between isolation sites for Caldicellulosiruptorales (Supplementary Table S2) relates to AAI and

ANI. While it is not surprising that species and strains isolated from immediate proximity to each other can have high AAI/ANI values, it is interesting that some species from isolation sites separated by thousands of kilometers are also closely related. Further, several strains isolated from relatively close sites have lower AAI/ANI values, indicating close proximity of isolation does not always infer high strain similarity even within the same genus. Thus, from available genomic data, there is no correlation between geographical spacing of isolation sites and ANI/AAI for the Caldicellulosiruptorales, at least at for distance >50 km. Presumably, at some smaller distance (i.e., <50 km), there could be a correlation, where multiple strains from a single geothermal feature, or very near-by thermal features, could have high relatedness. The Icelandic strains (*A. acetigenus*) may hint at this possibility due to their high relatedness and close proximity of isolation. However, at the same time, species isolated <100 km from each other have much lower relatedness, like those from New Zealand (*C. morgani*, *C. saccharolyticus*, *A. danielii*) or Japan (*C. naganensis*, *A. diazotrophicus*).

3.3. Carbohydrate utilization in the Caldicellulosiruptorales and Thermoanaerobacterales

Comparing and contrasting the underlying metabolism, physiology, and ecology of microorganisms goes beyond 16S rRNA, GTDB-tk, and ANI/AAI analyses. Interest in the Caldicellulosiruptorales was initially driven by their ability to degrade lignocellulosic biomass and the inventory of carbohydrate active enzymes (CAZymes) supporting this characteristic (Blumer-Schuette et al., 2014). Specifically, within the group of CAZymes are intracellular, extracellular, and surface-layer Glycoside Hydrolases (GHs) that process carbohydrates; many of these enzymes have multiple domains and associated carbohydrate binding modules (CBMs; Blumer-Schuette et al., 2010; Conway et al., 2016, 2017, 2018; Crosby et al., 2022; Laemthong et al., 2022). Figure 5 summarizes GH inventories separated by GH family (Drula et al., 2022), for the 15 genome-sequenced Caldicellulosiruptorales and 37 sequenced strains within the revised Thermoanaerobacterales. It is evident that the GH inventory varies widely across genera and species, and even across strains within a species. To this point, *A. acetigenus* str. *Acetigenus* has 15 or more total CAZymes, and 20 or more GH containing CAZymes, than either of the other two strains of this species.

A wide variety of GH family domains are found in all Caldicellulosiruptorales genomes; collectively these enable degradation of carbohydrates found in lignocellulosic biomass to recruit them as carbon and energy sources. Enzyme functions likely associated with these domains include xylan hydrolysis, exo- and endo-acting cleavage of β -glycans, hydrolysis of α -glucans (such as starch, pullulan), pectin degradation, hydrolysis of galactomannans, mono-, di-, and oligosaccharide phosphorylation, and hydrolysis of peptidoglycan or chitin. Domains related to cleavage of terminal glucose residues from β -glycosides (GH1) and α -glucans (GH15), as well as GH51 (likely endo acting β -glucanase or α -L-arabinofuranosidase), are found in all but one species. Domains related to phosphorylases (GH65), uncapping of glucuronic acid decorated xylooligosaccharides (GH67), and hydrolysis of unsaturated glucuronyl/galacturonyl linkages (GH105), and GH4 (possible α -glucosidase, galactosidase, glucuronidase, or galacturonase) are

present in all but two species. GH9 and GH48 domains (and CBM3) are central to enzymes that catalyze microcrystalline cellulose hydrolysis, notably present in the multi-domain cellulase, CelA (Conway et al., 2017) one or both of these GH domains are present in all five *Caldicellulosiruptor* species and six of ten *Anaerocellum* species. Strains lacking these cellulases are unable to grow on microcrystalline cellulose (Avicel). Of note here, *A. diazotrophicus* (which lacks these enzymes) was reported in its isolation paper to grow on Avicel (Chen et al., 2021b). However, as part of this work, *A. diazotrophicus* was readily cultured on cellobiose, beechwood xylan, and corn fiber, but no growth was observed on Avicel at 75°C. This result reinforces that Caldicellulosiruptorales lacking GH9 and GH48 domains cannot grow on microcrystalline cellulose. Both strongly cellulolytic and non-cellulolytic (or weakly cellulolytic) species are found in most Caldicellulosiruptorales isolation biotopes (Vishnivetskaya et al., 2015; Lee et al., 2018). This reflects the cooperative nature of microbial communities. For example, not all species in a biotope need to be cellulolytic to utilize lignocellulose, but instead some species can degrade cellulose through extracellular GHs thereby enabling scavenging of the resulting oligosaccharides by non-cellulolytic species. Interestingly, common to all members of the Caldicellulosiruptorales are peptidoglycan lyases / chitinase (GH23) and peptidoglycan/ chitin binding domains (CBM50). This implicates degradation of bacterial or fungal cell wall remnants to support acquisition of substrates in otherwise nutritionally sparse biotopes.

In the Thermoanaerobacterales genomes, highly represented GH families (GH1-5, GH13, GH15, GH23, GH31, GH36, GH65, GH130) mirror what is seen in the Caldicellulosiruptorales, with two exceptions. GH18 domains (chitinases, lysozymes) are common to the Thermoanaerobacterales, but absent in the Caldicellulosiruptorales. None of the genome sequenced Thermoanaerobacterales are known to degrade crystalline cellulose, and accordingly, none contain genes encoding GH9, GH48, and CBM3 domains, which are important for microcrystalline cellulose degradation. It is also interesting that many Thermoanaerobacterales lack GH10/11 domains that relate to xylan hydrolysis (*Thermoanaerobacterium*, *Caldanaerobius polysaccharolyticus*, and some *Thermoanaerobacter* are exceptions). Among the Thermoanaerobacterales, *Thermoanaerobacterium* have the most diversity in CAZymes, reflecting their hemicellulolytic activity. Similar trends are seen with other CAZyme domains (carbohydrate binding modules, polysaccharide lysases, and carbohydrate esterases; Supplementary Figure S2). Based on CAZyme inventory, Caldicellulosiruptorales have a broader carbohydrate appetite than the Thermoanaerobacterales, especially with respect to crystalline cellulose.

4. Discussion

Clearly, keeping taxonomy up to date in the face of expanding genome sequence databases is challenging. Fortunately, bioinformatic analyses, using tools such as GTDB-tk, ANI, and AAI, can be used to update taxonomy to more accurately reflect phylogeny, although heuristics and criteria for classification have not been universally established. Many fermentative anaerobes were isolated decades ago, prior to the use of whole genome sequence data, such that their taxonomy required updating. While classifications with 16S rRNA sequencing greatly improved taxonomy beyond phenotypic

here through analysis of CAZymes, but extends to other genes with high phenotypic influence, such as toxin-antitoxin biofilm regulation (Lewis et al., 2023)). We show here that when carbohydrate “appetite” is mapped to CAZyme inventory, the Caldicellulosiruptorales differ from the Thermoanaerobacterales in that many species of the former, but not the latter, can degrade microcrystalline cellulose. Most geographical regions harbor both cellulolytic and weak/non-cellulolytic Caldicellulosiruptorales, which could suggest a possible existence of Caldicellulosiruptorales communities, where the non-cellulolytic species benefit from cellulases secreted by the cellulolytic species. In extremely thermophilic terrestrial environments, these non-cellulolytic Caldicellulosiruptorales potentially fill the environmental niche occupied at lower temperatures by other hemicellulolytic and oligosaccharide-consuming organisms, such as *Thermoanaerobacterium* and *Thermoanaerobacter* species. The latter two genera are often found along with cellulolytic Clostridia (such as *Thermoclostridium stercorarium* and *Acetivibrio thermocellus*), forming moderately thermophilic plant biomass degrading communities with diversity spanning multiple Orders (Liu et al., 2008; Bing et al., 2023a). Further, metagenomic analyses to date indicate that microbial biodiversity in terrestrial hot springs decreases with increasing temperature (Lee et al., 2018), such that the gene pool becomes more limited. In line with this, Caldicellulosiruptorales seem to be the only plant biomass degraders in their higher temperature ($\geq 70^{\circ}\text{C}$) environments (based on available isolation and metagenomics reports for these thermal features). Caldicellulosiruptorales seem to have evolved to occupy multiple niches normally filled by multiple Orders of bacteria at lower temperatures. Due to the more limited microbial diversity in these extreme temperatures, Caldicellulosiruptorales species have evolved to harbor a broad array of CAZymes, allowing them to degrade and consume a wide range of polysaccharides, as the presence of other polysaccharide degrading microorganisms is less likely.

Specialization of Caldicellulosiruptorales to their localized environment might drive the gain and loss of CAZymes. How the Caldicellulosiruptorales acquire new CAZymes is not at all understood. The majority of their CAZymes have low homology (via NCBI BLAST search) outside of the Caldicellulosiruptorales, and homologous CAZymes within Caldicellulosiruptorales have similar average percent amino acid identities to that of the whole-genome AAI (Supplementary Figure S3), suggesting vertical gene transfer and evolution is significant for these CAZymes. Caldicellulosiruptoraceae all contain a plethora of transposases encoded in their genomes; this could also potentially aid in the movement of genes, including CAZymes, within their own genomes or between other Caldicellulosiruptorales. Possible evidence for this is found in species with fragmented glucan degradation loci (GDL), including *A. acetigenus* (all three strains) and *C. morgani* (Lee et al., 2018), where transposases are annotated immediately adjacent to GDL fragments. A more detailed evaluation of the role of transposases in the Caldicellulosiruptorales and their communities is obviously needed.

Caldicellulosiruptorales clearly have evolved to fit their environmental niche, where they excel at scavenging a variety of carbohydrates available in their environment. Indeed, CAZyme inventory of the Caldicellulosiruptorales reflect the types of biomasses common to their locales including (hemi)celluloses, pectin, and starches found in woods and grasses, as well as their fruits and seeds. Less

obvious sources of carbohydrate could include chitin or peptidoglycan in fungal/bacterial cell walls, or lichenin ($\beta 1,3$; $\beta 1,4$ glucan) and galactomannan rich organisms like mosses and lichens, including Icelandic moss [which occurs outside of Iceland in most regions containing Caldicellulosiruptorales (Table 1; Ingólfssdóttir, 2000)]. CAZyme inventory of Caldicellulosiruptorales varies by species; some like *A. kronotskyensis* have broad appetites for available carbohydrates. *A. kronotskyensis* contains 44 out of the 51 GH domains present in the Caldicellulosiruptorales pangenome and those missing GH domains are relatively rare across the family. Other Caldicellulosiruptorales that lack cellulases, or have fewer CAZymes, either live in environments with more limited sources of carbohydrates, or have evolved to live in communities with other strains, occupying narrower ecological roles. The latter can be seen for *A. acetigenus* strains, which vary widely in CAZyme inventory despite highly localized geographic proximity and high genome-wide identity. It is possible that these strains may exist in communities together where only a subpopulation degrade cellulose (such as *A. acetigenus lactoaceticus*). However, in order to prove this, a more detailed evaluation of Caldicellulosiruptorales communities is needed. In context of major research interests in the Caldicellulosiruptorales related to lignocellulose degradation, novel CAZymes likely still exist in environments not yet sampled, harbored in undiscovered strains. Consideration of polysaccharides local to isolation environments could aid in finding CAZymes or strains capable of degrading specific substrates.

Beyond CAZymes, the most differentiating feature between the Caldicellulosiruptorales and Thermoanaerobacterales is thermophilicity, given that the differences in optimal growth temperatures exceed 10°C . This is reflected by the AAI of homologous proteins, where increased protein thermostability likely maps to changes in amino acid sequence to maintain function at higher temperatures. As many aspects of fermentative anaerobic metabolism should be highly conserved, many proteins must evolve to become more (or less) thermostable. As such, the revised taxonomy containing two Orders (Caldicellulosiruptorales and Thermoanaerobacterales) reflects this phylogenetic divergence. While the revisions here improve the taxonomy of the Thermoanaerobacterales and Caldicellulosiruptorales, many unresolved taxonomic issues still remain for strains no longer a part of the Thermoanaerobacterales, as well as Class and Phyla classifications for the Clostridia. In any event, bioinformatic tools are useful for taxonomy and phylogeny, but phenotype and ecology must still lean heavily on microbiological details.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary material.

Author contributions

RB and RK conceived the study. RB, DW, JC, MA, and RK carried out the experiments and analysis. RB, MA, and RK wrote the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the US Department of Energy BER Awards DE-SC0019391 and DE-SC0022192. JC acknowledges support from a US DoEd GAANN Fellowship (P200A160061). RB and DW acknowledge support from an NIH Biotechnology Traineeship (NIH T32 GM008776-16).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Akst, J. (2021). "Researchers propose automating the naming of novel microbes" in *The Scientist*. Available at: <https://www.the-scientist.com/notebook/researchers-propose-automating-the-naming-of-novel-microbes--68411>
- Alaska Department of Natural Resources (1984). *Geothermal Resources of Alaska 1983*. Fairbanks, Alaska, USA: U.S. Government Printing Office.
- Berry, G. W., Grim, P. J., and Ikelman, J. A. (1980). "Thermal springs list for the United States" in *National Oceanic and Atmospheric Administration key to Geophysical Records*. Boulder Colorado: United State Department of Commerce, National Oceanic and Atmospheric Administration, National Geophysical and Solar-terrestrial Data Center.
- Bing, R. G., Carey, M. J., Laemthong, T., Willard, D. J., Crosby, J. R., Sulis, D. B., et al. (2023a). Fermentative conversion of unpretreated plant biomass: a thermophilic threshold for indigenous microbial growth. *Bioresour. Technol.* 367:128275. doi: 10.1016/j.biortech.2022.128275
- Bing, R. G., Straub, C. T., Sulis, D. B., Wang, J. P., Adams, M. W., and Kelly, R. M. (2022). Plant biomass fermentation by the extreme thermophile *Caldicellulosiruptor bescii* for co-production of green hydrogen and acetone: techno-economic analysis. *Bioresour. Technol.* 348:126780. doi: 10.1016/j.biortech.2022.126780
- Bing, W., Wang, H., Zheng, B., Zhang, F., Zhu, G., Feng, Y., et al. (2015). *Caldicellulosiruptor changbaiensis* sp. nov., a cellulolytic and hydrogen-producing bacterium from a hot spring. *Int. J. Syst. Evol. Microbiol.* 65, 293–297. doi: 10.1099/ijs.0.065441-0
- Bing, R. G., Willard, D. J., Manesh, M. J. H., Laemthong, T., Crosby, J. R., Adams, M. W., et al. (2023b). Complete genome sequences of *Caldicellulosiruptor acetigenus* DSM 7040, *Caldicellulosiruptor morganii* DSM 8990 (RT8.B8), and *Caldicellulosiruptor naganensis* DSM 8991 (NA10). *Microbiol Resour Announc* 12:e0129222. doi: 10.1128/mra.01292-22
- Blumer-Schuette, S. E. (2020). Insights into thermophilic plant biomass hydrolysis from *Caldicellulosiruptor* systems biology. *Microorganisms* 8:385. doi: 10.3390/microorganisms8030385
- Blumer-Schuette, S. E., Brown, S. D., Sander, K. B., Bayer, E. A., Kataeva, I., Zurawski, J. V., et al. (2014). Thermophilic lignocellulose deconstruction. *FEMS Microbiol. Rev.* 38, 393–448. doi: 10.1111/1574-6976.12044
- Blumer-Schuette, S. E., Kataeva, I., Westpheling, J., Adams, M. W., and Kelly, R. M. (2008). Extremely thermophilic microorganisms for biomass conversion: status and prospects. *Curr. Opin. Biotechnol.* 19, 210–217. doi: 10.1016/j.copbio.2008.04.007
- Blumer-Schuette, S. E., Lewis, D. L., and Kelly, R. M. (2010). Phylogenetic, microbiological, and glycoside hydrolase diversities within the extremely thermophilic, plant biomass-degrading genus *Caldicellulosiruptor*. *Appl. Environ. Microbiol.* 76, 8084–8092. doi: 10.1128/AEM.01400-10
- Blumer-Schuette, S. E., Ozdemir, I., Mistry, D., Lucas, S., Lapidus, A., Cheng, J.-F., et al. (2011). Complete genome sequences for the anaerobic, extremely thermophilic plant biomass-degrading bacteria *Caldicellulosiruptor hydrothermalis*, *Caldicellulosiruptor kristjanssonii*, *Caldicellulosiruptor kronotskyensis*, *Caldicellulosiruptor owensensis*, and *Caldicellulosiruptor lactoaceticus*. *J. Bacteriol.* 193, 1483–1484. doi: 10.1128/JB.01515-10
- Bredholt, S., Sonne-Hansen, J., Nielsen, P., Mathrani, I. M., and Ahring, B. K. (1999). *Caldicellulosiruptor kristjanssonii* sp. nov., a cellulolytic, extremely thermophilic, anaerobic bacterium. *Int. J. Syst. Bacteriol.* 49 Pt 3, 991–996. doi: 10.1099/00207713-49-3-991
- Charubin, K., Bennett, R. K., Fast, A. G., and Papoutsakis, E. T. (2018). Engineering clostridium organisms as microbial cell-factories: challenges & opportunities. *Metab. Eng.* 50, 173–191. doi: 10.1016/j.ymben.2018.07.012
- Chaumeil, P. A., Mussig, A. J., Hugenholtz, P., and Parks, D. H. (2019). GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* 36, 1925–1927. doi: 10.1093/bioinformatics/btz848
- Chaumeil, P. A., Mussig, A. J., Hugenholtz, P., and Parks, D. H. (2022). GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics* 38, 5315–5316. doi: 10.1093/bioinformatics/btac672
- Chen, Y., Nishihara, A., and Haruta, S. (2021a). Nitrogen-fixing ability and nitrogen fixation-related genes of thermophilic fermentative bacteria in the genus *Caldicellulosiruptor*. *Microbes Environ.* 36:n/a. doi: 10.1264/jsm.2021018
- Chen, Y., Nishihara, A., Iino, T., Ohkuma, M., and Haruta, S. (2021b). *Caldicellulosiruptor diazotrophicus* sp. nov., a thermophilic, nitrogen-fixing fermentative bacterium isolated from a terrestrial hot spring in Japan. *Int. J. Syst. Evol. Microbiol.* 71:5014. doi: 10.1099/ijsem.0.005014
- Cheng, H., Asakura, Y., Kanda, K., Fukui, R., Kawano, Y., Okugawa, Y., et al. (2018). Dynamic bacterial community changes in the autothermal thermophilic aerobic digestion process with cell lysis activities, shaking and temperature increase. *J. Biosci. Bioeng.* 126, 196–204. doi: 10.1016/j.jbiosc.2018.02.012
- Chung, D., Cha, M., Guss, A. M., and Westpheling, J. (2014). Direct conversion of plant biomass to ethanol by engineered *Caldicellulosiruptor bescii*. *Proc. Natl. Acad. Sci. U. S. A.* 111, 8931–8936. doi: 10.1073/pnas.1402210111
- Contreras-Moreira, B., and Vinuesa, P. (2013). GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl. Environ. Microbiol.* 79, 7696–7701. doi: 10.1128/AEM.02411-13
- Conway, J. M., Crosby, J. R., Mckinley, B. S., Seals, N. L., Adams, M. W., and Kelly, R. M. (2018). Parsing *in vivo* and *in vitro* contributions to microcrystalline cellulose hydrolysis by multidomain glycoside hydrolases in the *Caldicellulosiruptor bescii* secretome. *Biotechnol. Bioeng.* 115, 2426–2440. doi: 10.1002/bit.26773
- Conway, J. M., Mckinley, B. S., Seals, N. L., Hernandez, D., Khatibi, P. A., Poudel, S., et al. (2017). Functional analysis of the glucan degradation locus in *Caldicellulosiruptor bescii* reveals essential roles of component glycoside hydrolases in plant biomass deconstruction. *Appl. Environ. Microbiol.* 83, e01828–e01817. doi: 10.1128/AEM.01828-17
- Conway, J. M., Pierce, W. S., Le, J. H., Harper, G. W., Wright, J. H., Tucker, A. L., et al. (2016). Multidomain, surface layer-associated glycoside hydrolases contribute to plant polysaccharide degradation by *Caldicellulosiruptor* species. *J. Biol. Chem.* 291, 6732–6747. doi: 10.1074/jbc.M115.707810
- Crosby, J. R., Laemthong, T., Bing, R. G., Zhang, K., Tanwee, T. N. N., Lipscomb, G. L., et al. (2022). Biochemical and regulatory analyses of xylanolytic regulons in *Caldicellulosiruptor bescii* reveal genus-wide features of hemicellulose utilization. *Appl. Environ. Microbiol.* 88:e0130222. doi: 10.1128/aem.01302-22
- Drula, E., Garron, M. L., Dogan, S., Lombard, V., Henrissat, B., and Terrapon, N. (2022). The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Res.* 50, D571–D577. doi: 10.1093/nar/gkab1045
- Elkins, J. G., Lochner, A., Hamilton-Brehm, S. D., Davenport, K. W., Podar, M., Brown, S. D., et al. (2010). Complete genome sequence of the cellulolytic thermophile *Caldicellulosiruptor obsidiansis* OB47T. *J. Bacteriol.* 192, 6099–6100. doi: 10.1128/JB.00950-10
- Fernandez-Turiel, J. L., Garcia-Valles, M., Gimeno-Torrente, D., Saavedra, J., and Martinez-Manent, S. (2005). The hot spring and geyser sinters of El Tatio, northern Chile. *Sediment. Geol.* 180, 125–147. doi: 10.1016/j.sedgeo.2005.07.005
- Frolov, E. N., Lebedinsky, A. V., Elcheninov, A. G., and Kublanov, I. V. (2023). Taxonomic proposal for a deep branching bacterial phylogenetic lineage: transfer of the family *Thermodesulfobiaceae* to *Thermodesulfobiales* ord. nov., *Thermodesulfobii* classis nov. and *Thermodesulfobiota* phyl. nov. *Syst. Appl. Microbiol.* 46:126388. doi: 10.1016/j.syapm.2022.126388

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1212538/full#supplementary-material>

- Gibbs, M. D., Saul, D. J., Luthi, E., and Bergquist, P. L. (1992). The beta-mannanase from "*Caldocellum saccharolyticum*" is part of a multidomain enzyme. *Appl. Environ. Microbiol.* 58, 3864–3867. doi: 10.1128/aem.58.12.3864-3867.1992
- Habib, N., Rao, M. P. N., Xiao, M., Jan, S. A., and Li, W. J. (2021). Genome-based reclassification of *Caldicellulosiruptor lactoaceticus* and *Caldicellulosiruptor kristjanssonii* as later heterotypic synonyms of *Caldicellulosiruptor acetigenus*. *Int. J. Syst. Evol. Microbiol.* 71:4927. doi: 10.1099/ijsem.0.004927
- Hamilton-Brehm, S. D., Mosher, J. J., Vishnivetskaya, T., Podar, M., Carroll, S., Allman, S., et al. (2010). *Caldicellulosiruptor obsidiansis* sp. nov., an anaerobic, extremely thermophilic, cellulolytic bacterium isolated from obsidian Pool, Yellowstone national Park. *Appl. Environ. Microbiol.* 76, 1014–1020. doi: 10.1128/AEM.01903-09
- Hniman, A., Prasertsan, P., and Thong, A. O. (2011). Community analysis of thermophilic hydrogen-producing consortia enriched from Thailand hot spring with mixed xylose and glucose. *Int. J. Hydrog. Energy* 36, 14217–14226. doi: 10.1016/j.ijhydene.2011.05.087
- Huson, D. H., and Scornavacca, C. (2012). Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Syst. Biol.* 61, 1061–1067. doi: 10.1093/sysbio/sys062
- Ingólfssdóttir, K. (2000). "Bioactive compounds from Iceland moss" in *Bioactive Carbohydrate Polymers*. ed. B. S. Paulsen (Dordrecht: Springer Netherlands), 25–36.
- Johnson, W. R. (2010). World Geysir Locations. Available at: <https://www.johnstonsarchive.net/geysers/geysmapw.html> (Accessed April 21, 2023).
- Kataeva, I. A., Yang, S. J., Dam, P., Poole, F. L. 2nd, Yin, Y., Zhou, F., et al. (2009). Genome sequence of the anaerobic, thermophilic, and cellulolytic bacterium "*Anaerocellum thermophilum*" DSM 6725. *J. Bacteriol.* 191, 3760–3761. doi: 10.1128/JB.00256-09
- Kim, M., Oh, H. S., Park, S. C., and Chun, J. (2014). Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int. J. Syst. Evol. Microbiol.* 64, 346–351. doi: 10.1099/ijms.0.059774-0
- Laemthong, T., Bing, R. G., Crosby, J. R., Adams, M. W., and Kelly, R. M. (2022). Engineering *Caldicellulosiruptor bescii* with surface layer homology domain-linked glycoside hydrolases improves plant biomass solubilization. *Appl. Environ. Microbiol.* 88, e01274–e01222. doi: 10.1128/aem.01274-22
- Lee, L. L., Blumer-Schuetz, S. E., Izquierdo, J. A., Zurawski, J. V., Loder, A. J., Conway, J. M., et al. (2018). Genus-wide assessment of lignocellulose utilization in the extremely thermophilic genus *Caldicellulosiruptor* by genomic, pangenomic, and metagenomic analyses. *Appl. Environ. Microbiol.* 84:e02694-17. doi: 10.1128/AEM.02694-17
- Lee, L. L., Crosby, J. R., Rubinstein, G. M., Laemthong, T., Bing, R. G., Straub, C. T., et al. (2020). The biology and biotechnology of the genus *Caldicellulosiruptor*: recent developments in 'Caldi world'. *Extremophiles* 24, 1–15. doi: 10.1007/s00792-019-01116-5
- Lee, L. L., Izquierdo, J. A., Blumer-Schuetz, S. E., Zurawski, J. V., Conway, J. M., Cottingham, R. W., et al. (2015). Complete genome sequences of *Caldicellulosiruptor* sp. strain Rt8.B8, *Caldicellulosiruptor* sp. strain Wai35.B1, and "*Thermoanaerobacter cellulolyticus*". *Genome Announc.* 3:e00440-15. doi: 10.1128/genomeA.00440-15
- Lee, J., Jang, Y. S., Choi, S. J., Im, J. A., Song, H., Cho, J. H., et al. (2012). Metabolic engineering of *Clostridium acetobutylicum* ATCC 824 for isopropanol-butanol-ethanol fermentation. *Appl. Environ. Microbiol.* 78, 1416–1423. doi: 10.1128/AEM.06382-11
- Leticus, I., and Bork, P. (2006). Interactive tree of life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 23, 127–128. doi: 10.1093/bioinformatics/btl529
- Lewis, A. M., Willard, D. J., Manesh, M. J. H., Sivabalasarma, S., Albers, S.-V., and Kelly, R. M. (2023). Stay or go: sulfolobales biofilm dispersal is dependent on a bifunctional VapB antitoxin. *MBio* 14, e00053–e00023. doi: 10.1128/mbio.00053-23
- Lipscomb, G. L., Conway, J. M., Blumer-Schuetz, S. E., Kelly, R. M., and Adams, M. W. W. (2016). A highly thermostable kanamycin resistance marker expands the tool kit for genetic manipulation of *Caldicellulosiruptor bescii*. *Appl. Environ. Microbiol.* 82, 4421–4428. doi: 10.1128/AEM.00570-16
- Liu, Y., Yu, P., Song, X., and Qu, Y. (2008). Hydrogen production from cellulose by co-culture of *Clostridium thermocellum* JN4 and *Thermoanaerobacterium thermosaccharolyticum* GD17. *Int. J. Hydrog. Energy* 33, 2927–2933. doi: 10.1016/j.ijhydene.2008.04.004
- Liu, Z., Zhao, C., Deng, Y., Huang, Y., and Liu, B. (2015). Characterization of a thermostable recombinant β -galactosidase from a thermophilic anaerobic bacterial consortium YTY-70. *Biotechnol. Biotechnol. Equip.* 29, 547–554. doi: 10.1080/13102818.2015.1015244
- Lu, F., Bize, A., Guillot, A., Monnet, V., Madigou, C., Chapleur, O., et al. (2014). Metaproteomics of cellulose methanisation under thermophilic conditions reveals a surprisingly high proteolytic activity. *ISME J.* 8, 88–102. doi: 10.1038/ismej.2013.120
- Luthi, E., Jasmat, N. B., and Bergquist, P. L. (1990). Xylanase from the extremely thermophilic bacterium "*Caldocellum saccharolyticum*": overexpression of the gene in *Escherichia coli* and characterization of the gene product. *Appl. Environ. Microbiol.* 56, 2677–2683. doi: 10.1128/aem.56.9.2677-2683.1990
- Lv, X. M., Yang, M., Dai, L. R., Tu, B., Chang, C., Huang, Y., et al. (2020). *Zhaonella formicivorans* gen. nov., sp. nov., an anaerobic formate-utilizing bacterium isolated from Shengli oilfield, and proposal of four novel families and *Moorellales* ord. nov. in the phylum *Firmicutes*. *Int. J. Syst. Evol. Microbiol.* 70, 3361–3373. doi: 10.1099/ijsem.0.004178
- Mendoza, C., and Blumer-Schuetz, S. E. (2019). Complete genome sequence of *Caldicellulosiruptor changbaiensis* CBS-Z, an extremely thermophilic, cellulolytic bacterium isolated from a hot spring in China. *Microbiol Resour Announc* 8:e00021-19. doi: 10.1128/MRA.00021-19
- Milojevic, T., Cramm, M. A., Hubert, C. R. J., and Westall, F. (2022). "freezing" thermophiles: from one temperature extreme to another. *Microorganisms* 10:2417. doi: 10.3390/microorganisms10122417
- Miroshnichenko, M. L., Kublanov, I. V., Kostrikin, N. A., Tourova, T. P., Kolganova, T. V., Birkeland, N. K., et al. (2008). *Caldicellulosiruptor kronotskyensis* sp. nov., and *Caldicellulosiruptor hydrothermalis* sp. nov., two extremely thermophilic, cellulolytic, anaerobic bacteria from Kamchatka thermal springs. *Int. J. Syst. Evol. Microbiol.* 58, 1492–1496. doi: 10.1099/ijms.0.65236-0
- Mladenovska, Z., Mathrani, I. M., and Ahring, B. K. (1995). Isolation and characterization of *Caldicellulosiruptor lactoaceticus* sp. nov., an extremely thermophilic, cellulolytic, anaerobic bacterium. *Arch. Microbiol.* 163, 223–230. doi: 10.1007/BF00305357
- Narihiro, T., Terada, T., Kikuchi, K., Iguchi, A., Ikeda, M., Yamauchi, T., et al. (2009). Comparative analysis of bacterial and archaeal communities in methanogenic sludge granules from upflow anaerobic sludge blanket reactors treating various food-processing, high-strength organic wastewaters. *Microbes Environ.* 24, 88–96. doi: 10.1264/jmsme.2.ME08561
- Nielsen, P. I. M. M. B. K. A. (1993). *Thermoanaerobium acetigenum* spec. nov., a new anaerobic, extremely thermophilic, xylanolytic non-spore-forming bacterium isolated from an Icelandic hot spring. *Arch. Microbiol.* 159, 460–464. doi: 10.1007/BF00288594
- Nielsen, H. (2017). "Predicting secretory proteins with SignalP" in *Protein Function Prediction: Methods and Protocols*. ed. D. Kihara (New York, NY: Springer New York), 59–73.
- Nishihara, A., Haruta, S., Mcglynn, S. E., Thiel, V., and Matsuura, K. (2018). Nitrogen fixation in thermophilic chemosynthetic microbial communities depending on hydrogen, sulfate, and carbon dioxide. *Microbes Environ.* 33, 10–18. doi: 10.1264/jmsme2.ME17134
- Onyenwoke, R. U., Lee, Y. J., Dabrowski, S., Ahring, B. K., and Wiegell, J. (2006). Reclassification of *Thermoanaerobium acetigenum* as *Caldicellulosiruptor acetigenus* comb. nov. and emendation of the genus description. *Int. J. Syst. Evol. Microbiol.* 56, 1391–1395. doi: 10.1099/ijms.0.63723-0
- Oren, A., and Garrity, G. M. (2021). Valid publication of the names of forty-two phyla of prokaryotes. *Int. J. Syst. Evol. Microbiol.* 71:5056. doi: 10.1099/ijsem.0.005056
- Orlygsson, J., Sigurbjornsdottir, M. A., and Bakken, H. E. (2010). Bioprospecting thermophilic ethanol and hydrogen producing bacteria from hot springs in Iceland. *Icel. Agric. Sci.* 23, 73–85. Available at: https://ias.is/wp-content/uploads/Icelandic_Agricultural_Sciences_23_2010/Bioprospecting-thermophilic-ethanol.pdf
- Pandit, P. D., Gulhane, M. K., Khardenavis, A. A., and Purohit, H. J. (2016). Mining of hemicellulose and lignin degrading genes from differentially enriched methane producing microbial community. *Bioresour. Technol.* 216, 923–930. doi: 10.1016/j.biortech.2016.06.021
- Pandit, R. J., Hinsu, A. T., Patel, S. H., Jakhesara, S. J., Koringa, P. G., Bruno, F., et al. (2018b). Microbiota composition, gene pool and its expression in Gir cattle (*Bos indicus*) rumen under different forage diets using metagenomic and metatranscriptomic approaches. *Syst. Appl. Microbiol.* 41, 374–385. doi: 10.1016/j.syapm.2018.02.002
- Pandit, R. J., Hinsu, A. T., Patel, N. V., Koringa, P. G., Jakhesara, S. J., Thakkar, J. R., et al. (2018a). Microbial diversity and community composition of caecal microbiota in commercial and indigenous Indian chickens determined using 16s rDNA amplicon sequencing. *Microbiome* 6:115. doi: 10.1186/s40168-018-0501-9
- Parks, D. H., Chuvochina, M., Chaumeil, P.-A., Rinke, C., Mussig, A. J., and Hugenholtz, P. (2020). A complete domain-to-species taxonomy for bacteria and archaea. *Nat. Biotechnol.* 38, 1079–1086. doi: 10.1038/s41587-020-0501-8
- Parks, D. H., Chuvochina, M., Rinke, C., Mussig, A. J., Chaumeil, P.-A., and Hugenholtz, P. (2021). GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res.* 50, D785–D794. doi: 10.1093/nar/gkab776
- Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarshewski, A., Chaumeil, P.-A., et al. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* 36, 996–1004. doi: 10.1038/nbt.4229
- Patel, B. K., Morgan, H. W., and Daniel, R. M. (1986). Unusual microorganisms observed in New Zealand hot springs. *Microb. Ecol.* 12, 181–186. doi: 10.1007/BF02011202
- Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490. doi: 10.1371/journal.pone.0009490
- Pritchard, L., Glover, R. H., Humphris, S., Elphinstone, J. G., and Toth, I. K. (2016). Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Anal. Methods* 8, 12–24. doi: 10.1039/C5AY02550H

- Qiu, C., Wen, J., and Jia, X. (2011). Extreme-thermophilic biohydrogen production from lignocellulosic bioethanol distillery wastewater with community analysis of hydrogen-producing microflora. *Int. J. Hydrog. Energy* 36, 8243–8251. doi: 10.1016/j.ijhydene.2011.04.089
- Rainey, F. A., Donnison, A. M., Janssen, P. H., Saul, D., Rodrigo, A., Bergquist, P. L., et al. (1994). Description of *Caldicellulosiruptor saccharolyticus* gen. nov., sp. nov.: an obligately anaerobic, extremely thermophilic, cellulolytic bacterium. *FEMS Microbiol. Lett.* 120, 263–266. doi: 10.1111/j.1574-6968.1994.tb07043.x
- Rainey, F. A., Janssen, P. H., Morgan, H. W., and Stackebrandt, E. (1993). A biphasic approach to the determination of the phenotypic and genotypic diversity of some anaerobic, cellulolytic, thermophilic, rod-shaped bacteria. *Antonie Van Leeuwenhoek* 64, 341–355. doi: 10.1007/BF00873092
- Richter, M., and Rossello-Mora, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *PNAS* 106, 19126–19131. doi: 10.1073/pnas.0906412106
- Robitzski, D. (2022). “Newly renamed prokaryote phyla cause uproar” in *The Scientist*. Available at: <https://www.the-scientist.com/news-opinion/newly-renamed-prokaryote-phyla-cause-uproar-69578>
- Sahm, K., John, P., Nacke, H., Wemheuer, B., Grote, R., Daniel, R., et al. (2013). High abundance of heterotrophic prokaryotes in hydrothermal springs of the Azores as revealed by a network of 16S rRNA gene-based methods. *Extremophiles* 17, 649–662. doi: 10.1007/s00792-013-0548-2
- Šantl-Temkiv, T., Amato, P., Casamayor, E. O., Lee, P. K. H., and Pointing, S. B. (2022). Microbial ecology of the atmosphere. *FEMS Microbiol. Rev.* 46:fuac009. doi: 10.1093/femsre/fuac009
- Sayeh, R., Birrien, J. L., Alain, K., Barbier, G., Hamdi, M., and Prieur, D. (2010). Microbial diversity in Tunisian geothermal springs as detected by molecular and culture-based approaches. *Extremophiles* 14, 501–514. doi: 10.1007/s00792-010-0327-2
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal omega. *Mol. Syst. Biol.* 7:539. doi: 10.1038/msb.2011.75
- Sissons, C. H., Sharrock, K. R., Daniel, R. M., and Morgan, H. W. (1987). Isolation of cellulolytic anaerobic extreme thermophiles from New Zealand thermal sites. *Appl. Environ. Microbiol.* 53, 832–838. doi: 10.1128/aem.53.4.832-838.1987
- Sonne-Hansen, J., and Ahring, B. K. (1997). Anaerobic microbiology of an alkaline Icelandic hot spring. *Microb. Ecol.* 23, 31–38. doi: 10.1111/j.1574-6941.1997.tb00388.x
- Straub, C. T., Bing, R. G., Otten, J. K., Keller, L. M., Zeldes, B. M., Adams, M. W. W., et al. (2020). Metabolically engineered *Caldicellulosiruptor bescii* as a platform for producing acetone and hydrogen from lignocellulose. *Biotechnol. Bioeng.* 117, 3799–3808. doi: 10.1002/bit.27529
- Straub, C. T., Khatibi, P. A., Wang, J. P., Conway, J. M., Williams-Rhaesa, A. M., Peszlen, I. M., et al. (2019). Quantitative fermentation of unpretreated transgenic poplar by *Caldicellulosiruptor bescii*. *Nat. Commun.* 10:3548. doi: 10.1038/s41467-019-11376-6
- Straub, C. T., Zeldes, B. M., Schut, G. J., Adams, M. W., and Kelly, R. M. (2017). Extremely thermophilic energy metabolisms: biotechnological prospects. *Curr. Opin. Biotechnol.* 45, 104–112. doi: 10.1016/j.copbio.2017.02.016
- Svetlichnyi, V. A., Svetlichnaya, T. P., Chernykh, N. A., and Zavarzin, G. A. (1990). *Anaerocellum thermophilum* gen. Nov., sp. nov., an extremely thermophilic cellulolytic eubacterium isolated from hot-springs in the valley of geysers. *Mikrobiologiya* 59, 598–604.
- Svetlichnyi, V. A., Kensch, O., Falkenhan, D. A., Korseska, S. G., Lippert, N., Prinz, M., et al. (2013). Single-step ethanol production from lignocellulose using novel extremely thermophilic bacteria. *Biotechnol. Biofuels* 6:31. doi: 10.1186/1754-6834-6-31
- Taya, M., Hinoki, H., Yagi, T., and Kobayashi, T. (1988). Isolation and characterization of an extremely thermophilic, cellulolytic, anaerobic bacterium. *Appl. Microbiol. Biotechnol.* 29, 474–479. doi: 10.1007/BF00269071
- Van De Werken, H. J. G., Verhaart, M. R. A., Vanfossen, A. L., Willquist, K., Lewis, D. L., Nichols, J. D., et al. (2008). Hydrogenomics of the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus*. *Appl. Environ. Microbiol.* 74, 6720–6729. doi: 10.1128/AEM.00968-08
- Vincenty, T. (1975). Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations. *Surv. Rev.* 23, 88–93. doi: 10.1179/sre.1975.23.176.88
- Vinuesa, P., and Contreras-Moreira, B. (2015). “Robust identification of orthologues and paralogues for microbial pan-genomics using GET_HOMOLOGUES: a case study of pInCA/C plasmids” in *Bacterial Pangonomics: Methods and Protocols*, eds. A. Mengoni, M. Galardini and M. Fondi (New York, NY: Springer New York), 203–232.
- Vishnivetskaya, T. A., Hamilton-Brehm, S. D., Podar, M., Mosher, J. J., Palumbo, A. V., Phelps, T. J., et al. (2015). Community analysis of plant biomass-degrading microorganisms from obsidian Pool, Yellowstone National Park. *Microb. Ecol.* 69, 333–345. doi: 10.1007/s00248-014-0500-8
- Vishnivetskaya, T. A., Mironov, V. A., Abramov, A. A., Shcherbakova, V. A., and Rivkina, E. M. (2022). Biogeochemical characteristics of earth’s volcanic permafrost: an analog of extraterrestrial environments. *Astrobiology* 22, 812–828. doi: 10.1089/ast.2021.0137
- Wagner, I. D., and Wiegel, J. (2008). Diversity of thermophilic anaerobes. *Ann. N. Y. Acad. Sci.* 1125, 1–43. doi: 10.1196/annals.1419.029
- Williams-Rhaesa, A. M., Rubinstein, G. M., Scott, I. M., Lipscomb, G. L., Poole Ii, F. L., Kelly, R. M., et al. (2018). Engineering redox-balanced ethanol production in the cellulolytic and extremely thermophilic bacterium, *Caldicellulosiruptor bescii*. *Metab Eng Commun* 7:e00073. doi: 10.1016/j.mec.2018.e00073
- Wirth, R., Kovacs, E., Maroti, G., Bagi, Z., Rakhely, G., and Kovacs, K. L. (2012). Characterization of a biogas-producing microbial community by short-read next generation DNA sequencing. *Biotechnol. Biofuels* 5:41. doi: 10.1186/1754-6834-5-41
- Yang, S. J., Kataeva, I., Wiegel, J., Yin, Y., Dam, P., Xu, Y., et al. (2010). Classification of ‘*Anaerocellum thermophilum*’ strain DSM 6725 as *Caldicellulosiruptor bescii* sp. nov. *Int. J. Syst. Evol. Microbiol.* 60, 2011–2015. doi: 10.1099/ijs.0.017731-0
- Yarza, P., Yilmaz, P., Pruesse, E., Glockner, F. O., Ludwig, W., Schleifer, K. H., et al. (2014). Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat. Rev. Microbiol.* 12, 635–645. doi: 10.1038/nrmicro3330
- Ying, Y., Meng, D., Chen, X., and Li, F. (2013). An extremely thermophilic anaerobic bacterium *Caldicellulosiruptor* sp. F32 exhibits distinctive properties in growth and xylanases during xylan hydrolysis. *Enzym. Microb. Technol.* 53, 194–199. doi: 10.1016/j.enzmictec.2013.04.004
- Zeldes, B. M., Keller, M. W., Loder, A. J., Straub, C. T., Adams, M. W., and Kelly, R. M. (2015). Extremely thermophilic microorganisms as metabolic engineering platforms for production of fuels and industrial chemicals. *Front. Microbiol.* 6:1209. doi: 10.3389/fmicb.2015.01209
- Zhang, G., Dong, H., Jiang, H., Xu, Z., and Eberl, D. E. (2007). Unique microbial community in drilling fluids from Chinese continental scientific drilling. *Geomicrobiol J.* 23, 499–514. doi: 10.1080/01490450600875860
- Zhang, F., Yang, J.-H., Dai, K., Ding, Z.-W., Wang, L.-G., Li, Q.-R., et al. (2016). Microbial dynamics of the extreme-thermophilic (70 °C) mixed culture for hydrogen production in a chemostat. *Int. J. Hydrog. Energy* 41, 11072–11080. doi: 10.1016/j.ijhydene.2016.04.085
- Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., et al. (2018). dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 46, W95–W101. doi: 10.1093/nar/gky418
- Zheng, J., Ge, Q., Yan, Y., Zhang, X., Huang, L., and Yin, Y. (2023). dbCAN3: automated carbohydrate-active enzyme and substrate annotation. *Nucleic Acids Res.* 51, W115–W121. doi: 10.1093/nar/gkad328