

RESEARCH ARTICLE

Genome-based metabolic and phylogenomic analysis of three *Terrisporobacter* speciesTim Böer¹, Frank R. Bengelsdorf², Mechthild Bömeke¹, Rolf Daniel¹, Anja Poehlein^{1*}

1 Genomic and Applied Microbiology and Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, Georg-August-University Göttingen, Göttingen, Germany, **2** Institute for Molecular Biology and Biotechnology of Prokaryotes, University Ulm, Ulm, Germany

* anja.poehlein@biologie.uni-goettingen.de



Abstract

Acetogenic bacteria are of high interest for biotechnological applications as industrial platform organisms, however, acetogenic strains from the genus *Terrisporobacter* have hitherto been neglected. To date, three published type strains of the genus *Terrisporobacter* are only covered by draft genome sequences, and the genes and pathway responsible for acetogenesis have not been analyzed. Here, we report complete genome sequences of the bacterial type strains *Terrisporobacter petrolearius* JCM 19845^T, *Terrisporobacter mayombeii* DSM 6539^T and *Terrisporobacter glycolicus* DSM 1288^T. Functional annotation, KEGG pathway module reconstructions and screening for virulence factors were performed. Various species-specific vitamin, cofactor and amino acid auxotrophies were identified and a model for acetogenesis of *Terrisporobacter* was constructed. The complete genomes harbored a gene cluster for the reductive proline-dependent branch of the Stickland reaction located on an approximately 21 kb plasmid, which is exclusively found in the *Terrisporobacter* genus. Phylogenomic analysis of available *Terrisporobacter* genomes suggested a reclassification of most isolates as *T. glycolicus* into *T. petrolearius*.

OPEN ACCESS

Citation: Böer T, Bengelsdorf FR, Bömeke M, Daniel R, Poehlein A (2023) Genome-based metabolic and phylogenomic analysis of three *Terrisporobacter* species. PLoS ONE 18(10): e0290128. <https://doi.org/10.1371/journal.pone.0290128>

Editor: Bashir Sajo Mienda, Federal University Dutse, NIGERIA

Received: June 26, 2023

Accepted: September 27, 2023

Published: October 10, 2023

Copyright: © 2023 Böer et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The genome sequence for *T. mayombeii* is available at DDBJ/ENA/GenBank under the accession numbers CP101637-CP101639 and the BioProject/BioSample accession numbers are PRJNA838466 and SAMN28423240. Raw reads can be found in the NCBI sequence read archive under the accession number SRR23391581 for Illumina reads and SRR23390615 for Nanopore reads. The genome sequence for *T. petrolearius* is available at DDBJ/ENA/GenBank under the accession number CP102984-CP102986 and the BioProject/

Introduction

Acetogenic bacteria are of high interest to biotechnology as they enable the conversion of industrial waste gases to valuable biocommodities by synthesis gas fermentation or microbial electrosynthesis (MES). These carbon-recycling technologies can contribute to the current efforts towards achieving climate neutrality. The genus *Terrisporobacter* comprises isolates capable of performing MES and autotrophic growth on H₂ and CO₂ through acetogenesis [1–3]. *Terrisporobacter* consists of mesophilic, Gram-positive and strictly anaerobic members. The cells appear as slightly curved rods, form endospores and are motile by flagella [4]. The four published type strains are *Terrisporobacter mayombeii* DSM 6539^T, *Terrisporobacter glycolicus* DSM 1288^T, *Terrisporobacter petrolearius* JCM 19845^T and *Terrisporobacter hibernicus* NCTC 14625^T. The ability to utilize H₂ and CO₂ via the Wood-Ljungdahl pathway (WLP) has only been shown for *T. mayombeii* (Kane et al. 1991). Acetogenesis has also been described for the isolates *T. glycolicus* DSM 13865 and ATCC 29797, and genes required for the Wood-

BioSample accession numbers are PRJNA839206 and SAMN28513175. Raw reads can be found in the NCBI sequence read archive under the accession number SRR23390574 for Illumina reads and SRR23392100 for Nanopore reads. The genome sequence for *T. glycolicus* is available at DDBJ/ENA/GenBank under the accession number CP117523-CP117524 and the BioProject/BioSample accession numbers are PRJNA929918 and SAMN32972286. Raw reads can be found in the NCBI sequence read archive under the accession number SRR23389261 for Illumina reads and SRR23389232 for Nanopore reads.

Funding: This work was supported by the Open Access Publication Funds of the University of Göttingen. Furthermore, we acknowledge the support of T.B. by the "Deutsche Bundesstiftung Umwelt" (DBU; 20020/640) and partial support of A.P. by the BMBF (ThermoSynCon; 031B0857C). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Ljungdahl pathway are present in all available *Terrisporobacter* draft genomes [2, 5]. So far, no detailed genome-based analysis is available for the genes and pathways involved in acetogenic growth of *Terrisporobacter* strains. The first *Terrisporobacter* species discovered was *T. glycolicus* DSM 1288^T (formerly *Clostridium glycolicum*), which was isolated from mud of a stagnant pond near the Chesapeake and Ohio canal (USA) in 1962 [6]. *T. mayombei* DSM 6539^T (formerly *Clostridium mayombei*) was isolated as the second *Terrisporobacter* type strain from gut contents of the soil-feeding termite *Cubitermes speciosus* in 1991 [3]. The genus *Terrisporobacter* was proposed in 2014 and *T. glycolicus* DSM 1288^T was described as a type species [4]. The third type strain *T. petrolearius* JCM 19845^T was isolated in 2015 using a petroleum reservoir sample from an oilfield in Shengli (China) [7]. *Terrisporobacter othiniensis* (08–306576) and *Terrisporobacter muris* (DSM 29186) were described as different strains. However, both strains appeared to be the same species, but these names have not been published [8, 9]. In 2023, *Terrisporobacter hibernicus* (NCTC 14625^T) was published as a fourth type strain [10]. *Terrisporobacter* members have been found to be strongly associated to pig manure and in particular *T. glycolicus* DSM 1288^T has been discussed as potential emerging anaerobic pathogen [11, 12]. In addition, some members show a positive correlation with the generation of bioelectricity in microbial fuel cells, indicating their electricigenic potential [13]. Correspondingly, *Terrisporobacter* strains have been found at the cathode of MES cells [1]. These insights support the use of strains from the genus *Terrisporobacter* as biocatalysts in biotechnology. The availability of high-quality genome sequences is a prerequisite for further metabolic investigations. Nevertheless, high-quality genome sequences were not available for three published type strains of the *Terrisporobacter* genus. We therefore sequenced and assembled the genomes of *T. petrolearius* JCM 19845^T, *T. mayombei* DSM 6539^T and *T. glycolicus* DSM 1288^T to perform a genome-based metabolic and phylogenomic analysis of these *Terrisporobacter* type strains.

Materials and methods

Cultivation and DNA extraction

Genomic DNA was extracted from cells of a *T. petrolearius* JCM 19845^T culture derived from the Microbe Division of the Japan Collection of Microorganisms (JCM; Tsukuba; Japan). *T. glycolicus* DSM 1288^T and *T. mayombei* DSM 6539^T were grown from lyophilised cultures derived from the German Collection of Microorganisms and Cell Cultures (DSMZ; Braunschweig, Germany) in an overnight culture, incubated at 30°C in DSM 311 medium supplemented with glucose (30 mM). Cell pellets were obtained by centrifugation of 10 ml culture at 13,000 g for 5 min. DNA was isolated by using the instructions for cell samples of the MasterPure™ Complete DNA and RNA purification kit (Epicentre, Madison, WI, USA).

Genome sequencing, assembly and annotation

Illumina sequencing libraries were prepared using the Nextera XT DNA sample preparation kit (Illumina, San Diego, CA, USA) and sequenced with a MiSeq system and v3 chemistry (600 cycles) following the instructions of the manufacturer (Illumina, San Diego, CA, USA). Sequencing kits and program versions for Nanopore sequencing are shown in the [S1 Table](#). The Nanopore sequencing libraries were prepared with 1.5 µg high-molecular-weight DNA using the ligation sequencing kit 1D and the native barcode expansion kit as recommended by the manufacturer (Oxford Nanopore Technologies, Oxford, UK). Nanopore sequencing was conducted for 72 h with the MinION device Mk1B, the SpotON flow cell R9.4.1 and the MinKNOW software following the instructions of the manufacturer (Oxford Nanopore Technologies, Oxford, UK). Demultiplexing and base calling of Nanopore sequencing data were performed with the Albacore/Guppy software. Quality control and adapter trimming of

paired-end Illumina sequences was performed with FastQC (v0.11.9) and Trimmomatic (v0.39; LEADING: 3, TRAILING: 3, SLIDINGWINDOW:4:15, MINLEN:50) [14]. Porechop (v0.2.4) was used for adapter trimming of Nanopore reads, which were subsequently filtered to a minimal read length of 5 kb [15]. Trimmed paired-end Illumina and filtered long Nanopore reads were used as input of Unicycler (v0.5.0) for a *de novo* hybrid genome assembly [16]. The resulting assembly graph was inspected with Bandage (v0.8.1) [17]. Annotation of the resulting genomes was performed with Prokka (v1.14.6) and the locus tag TEMA for *T. mayombei* DSM 6539^T, TEPE for *T. petrolearius* JCM 19845^T and TEGL for *T. glycolicus* DSM 1288^T were chosen [18]. BUSCO scores were obtained using BUSCO (v5.4.4) with the lineage dataset *clostridia_odb10* and the mode for prokaryotic genomes [19].

Genome-based metabolic and phylogenomic analysis

Functional annotation of genes to KEGG categories and the reconstruction of KEGG pathway modules was conducted with BlastKOALA [20]. For metabolites occurring with multiple KEGG pathway modules only the most complete ones were chosen for visualization. Assignment of genes into cluster of orthologous genes (COG) and subsequent functional annotation into COG categories was performed with the eggNOG mapper (v2.1.9) using the eggNOG 5.0 database [21, 22]. Barcharts and heatmaps were plotted in RStudio (v2022.06.0) with the ggplot2 package (v3.4.1) [23, 24]. Pan/core genome analysis was conducted with OrthoVenn2 clustering genes, identifying orthologous clusters and assigning Gene Ontology (GO) annotations [25]. The identification and classification of insertion sequence (IS) elements was performed with the ISESCAN pipeline (v1.7.2.3) [26]. Orthologous genes for the Wood-Ljungdahl pathway and the Stickland reactions were identified by using Proteinortho (v6.0.31) running with default parameters [27]. Furthermore, the Proteinortho output was used for the verification of missing genes in KEGG modules predicted to be incomplete by BlastKOALA. Gene cluster visualizations were prepared with clinker (v1.32) [28]. Virulence factor screening was performed with the VFAnalyzer pipeline using the Virulence Factor Database (VFDB) [29]. The phylogenomic analysis was performed with all *Terrisporobacter* genomes available at NCBI Genbank using pyani (v0.2.12) and the MUMmer alignment option [30]. ANIm percentage identity and alignment length values are shown in S2 Table. The dDDH values for *T. petrolearius* were calculated by the Type (Strain) Genome Server (TYGS) [31].

Results and discussion

Genome assembly and annotation

Details for the genome sequencing, assembly and annotation of the three *Terrisporobacter* type strains are summarized in Table 1. Each genome harbors a circular chromosome ranging from 4.03–4.09 Mb, and an additional circular plasmid ranging from 21.29–21.89 Kb (pTM1, pTP1 and pTG1). An additional circular plasmid was obtained for *T. mayombei* DSM 6539^T (pTM2, 8.20 Kb) and *T. petrolearius* JCM 19845^T (pTP2, 6.74 Kb). The GC content of the chromosomes varied between 28.7–28.9%. Prokka annotation resulted in 4,165, 4,208 and 4,038 potential genes for the genomes of *T. mayombei*, *T. petrolearius* and *T. glycolicus*, respectively. All three genomes contained 12 genes coding for 5S, 16S and 23S rRNA. One CRISPR repeat region was detected in the genomes of *T. petrolearius* and *T. glycolicus* with 67 and 28 spacers, respectively. The genome of *T. mayombei* contained two CRISPR repeat regions with 12 and 25 spacers. BUSCO scored a 100% completeness for all three genomes with 98.8% single copy (S) and 1.2% duplicated (D) BUSCO marker genes. No fragmented (F) or missing (M) BUSCO marker genes were identified.

Table 1. Genome features of the three *Terrisporobacter* type strains *T. petrolearius* JCM 19845^T, *T. mayombe* DSM 6539^T, and *T. glycolicus* DSM 1288^T.

Feature	Genome		
	<i>T. mayombe</i> DSM 6539 ^T	<i>T. petrolearius</i> JCM 19845 ^T	<i>T. glycolicus</i> DSM 1288 ^T
Chromosome size (bp)	4,034,168	4,096,312	4,039,277
Number of Illumina reads (250 bp)	1,495,093	1,218,094	4,189,121
Number of Nanopore reads/mean length (bp)	207,411/11,100	295,055/6,700	113,279/20,700
Chromosome mean coverage (Illumina/ Nanopore)	116/557	122/465	444/552
Plasmid size (bp)	pTM1: 21,885; pTM2: 8,218	pTP1: 21,288; pTP2: 6,738	pTG1: 21,319
Plasmid mean coverage (Illumina/Nanopore)	pTM1: 417/2,314 pTM2: 1,008/3,389	pTP1: 557/2,416 pTP2: 428/2,161	pTG1: 1,526/1,573
GC content (%)	28.8	28.9	28.7
Genes	4,165	4,203	4,038
CDS	3,915	3,932	3,826
Functional proteins	2,033	2,112	2,114
Hypothetical proteins	1,882	1,820	1,712
rRNAs (5S; 16S; 23S)	12; 12; 12	12; 12; 12	12; 12; 12
tRNAs	94	93	93
tmRNA	1	1	1
ncRNAs	119	141	82
CRISPR repeats	2	1	1
BUSCO scores	C:100.0% [S:98.8%, D:1.2%] F:0.0%, M:0.0% n:247	C:100.0% [S:98.8%, D:1.2%] F:0.0%, M:0.0% n:247	C:100.0% [S:98.8%, D:1.2%] F:0.0%, M:0.0% n:247

<https://doi.org/10.1371/journal.pone.0290128.t001>

Functional annotation and pan/core genome analysis

The functional annotation into COG categories of the *Terrisporobacter* type strain genomes is visualized in Fig 1A. The eggNOG-mapper assigned categories to 86.9% of coding sequences for *T. mayombe*, 88.1% of coding sequences for *T. petrolearius* and 89% for *T. glycolicus*. The genes of the three *Terrisporobacter* type strain genomes distributed uniformly across the different COG categories, with the exception of the category replication, recombination and repair (L). With 258, the *T. petrolearius* contained significantly more genes than the other two type strains of *T. mayombe* (161) and *T. glycolicus* (150). The pan genome of the three genomes comprised 3,548 gene clusters of which 2,930 gene clusters formed the core genome (Fig 1B). The genome of *T. mayombe* and *T. glycolicus* each harbored 15 exclusive gene clusters, while the genome of *T. petrolearius* harbored 38 exclusive gene clusters. *T. mayombe* contained four exclusive gene clusters, including gene clusters for acyltransferase activity (GO:0016746), phosphorelay sensor kinase activity (GO:0000155), DNA binding (GO:0003677) and metal ion binding (GO:0046872). Furthermore, *T. mayombe* possessed two exclusive gene clusters with GO annotations for the regulation of DNA-templated transcription (GO:0006355) and cobalamin biosynthesis (GO:0009236). *T. petrolearius* contained ten exclusive gene clusters with GO annotations of molecular function, comprising gene clusters for ATP-binding (GO:0005524), phosphorelay response regulator activity (GO:0000156), metal ion binding (GO:0046872), spermidine-importing ATPase activity (GO:0015595), enterobactin transmembrane transporter activity (GO:0042931), transmembrane transporter activity (GO:0022857), methionyl-tRNA formyltransferase activity (GO:0004479), sequence-specific DNA binding (GO:0043565) and two gene clusters encoding ATP hydrolysis activity (GO:0016887). Four

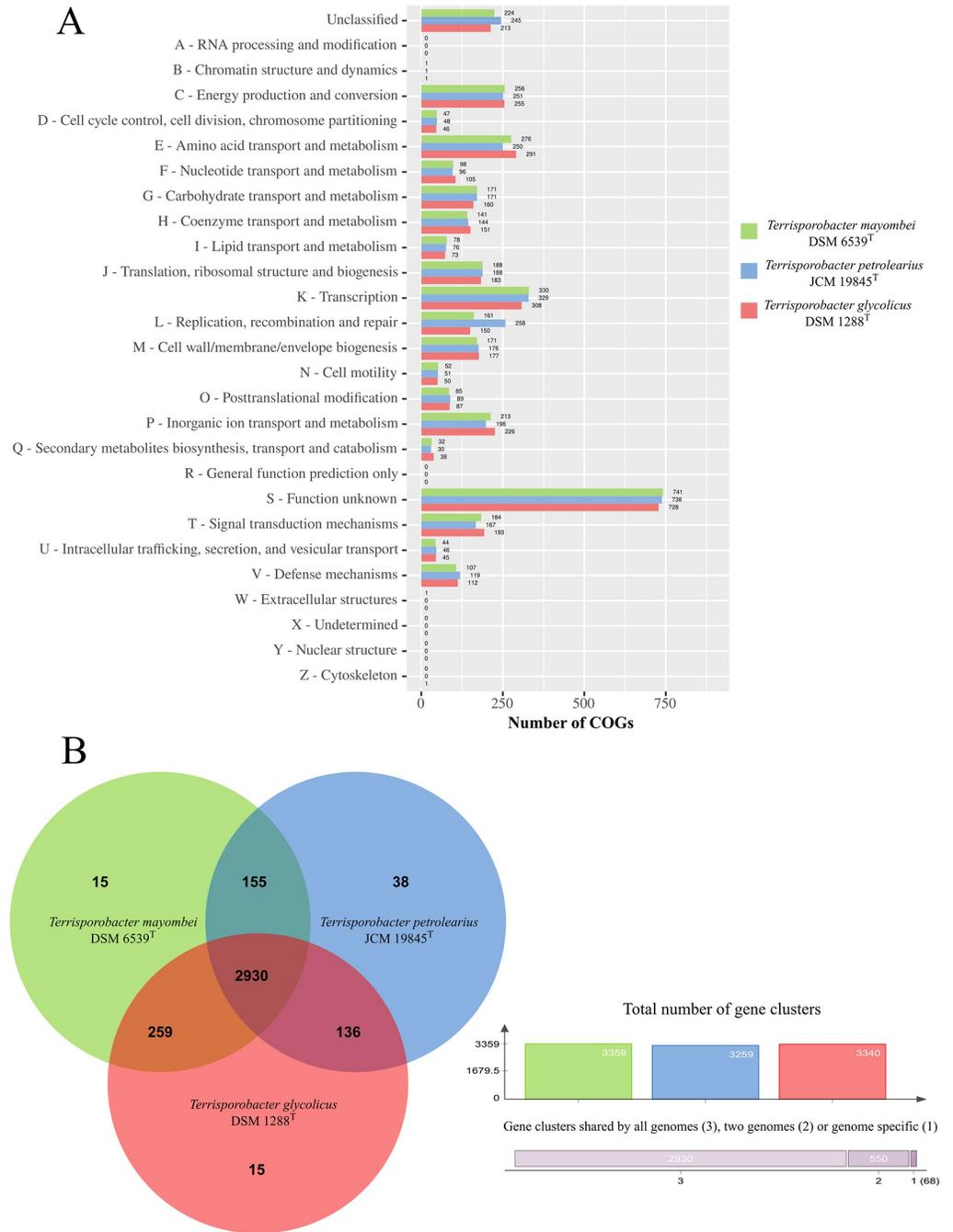


Fig 1. Functional annotation and pan/core genome analysis. **A** Functional annotation of the *Terrisporobacter* type strain genomes into COG categories. The y-axis shows the different COG categories and the x-axis the number of genes assigned to each category by eggNOG-mapper. The bar color indicates the number of genes for the three *Terrisporobacter* type strains *T. mayombei* (green), *T. petrolearius* (blue) and *T. glycolicus* (red). **B** Pan/core genome analysis of three *Terrisporobacter* type strains. The Venn diagram visualizes the exclusive and shared gene clusters for *T. mayombei* (green), *T. petrolearius* (blue) and *T. glycolicus* (red) created with OrthoVenn2. The graph visualizes the total number of gene clusters in the three genomes using the same color code.

<https://doi.org/10.1371/journal.pone.0290128.g001>

exclusive gene clusters with over 80 predicted proteins in total were annotated for transposition (GO:0032196) and DNA-mediated transposition (GO:0006313). In addition, exclusive gene clusters were identified with GO annotations for osmosensory signaling via phosphorelay pathway (GO:0007234), shikimate metabolic process (GO:0019632), antibiotic biosynthetic process (GO:0017000), regulation of DNA-templated transcription (GO:0006355), DNA integration (GO:0015074) and N-terminal protein amino acid acetylation (GO:0006474). *T. glycolicus* contained two exclusive gene clusters with annotated molecular function, including gene clusters for nickel cation transmembrane transporter activity (GO:0015099) and metal ion binding (GO:0046872). Other assigned GO annotations of exclusive gene clusters comprised regulation of systemic acquired resistance (GO:0010112), transmembrane transport (GO:005505085), ethanolamine catabolic process (GO:00463336), regulation of DNA-templated transcription (GO:0006355), polysaccharide biosynthetic process (GO:0000271), nickel cation transport (GO:0015675) and teichoic acid biosynthetic process (GO:0019350). Remaining exclusive gene clusters for the three *Terrisporobacter* genomes did not receive functional annotations.

KEGG pathway analysis assigned 49.9% of genes for *T. mayombeii*, 51.3% of genes for *T. petrolearius* and 52.4% of genes for *T. glycolicus* into functional categories (Fig 2). Between the analyzed three *Terrisporobacter* type strains genes distributed uniformly to the different functional categories, with the exception of the category unclassified: genetic information processing. In this category, 113 genes were assigned from the genome of *T. petrolearius*, while only 32 genes derived from the genome of *T. mayombeii* and 28 genes from the genome of *T. glycolicus*.

The functional annotations into COG and KEGG categories match with the described *Terrisporobacter* substrate spectrum specializing on the fermentation of carbohydrates and amino acids [3, 6, 7]. In both functional annotation approaches the gene distribution to the different categories was very similar between the three *Terrisporobacter* type strain genomes with the only exception of the increased gene number in *T. petrolearius* assigned to the categories of unclassified: genetic information processing (KEGG) or replication, recombination and repair (COG). This difference was also identified by pan/core genome analysis, with the four exclusive gene clusters encoded by *T. petrolearius* containing over 80 predicted proteins for transposition. In order to investigate the difference in transposase abundance the presence of IS elements in the genomes were analyzed with the ISESCAN pipeline (Fig 3). In total 94 IS elements were identified for *T. petrolearius*, while only four and six IS elements were detected in the *T. mayombeii* and *T. glycolicus* genomes, respectively. All three genomes harbored IS elements from the IS21 family. One element of family IS66 was present in the genomes of *T. petrolearius* and *T. glycolicus*. The family IS607 was only detected with one element in *T. glycolicus*. The families IS3, IS110 and IS256 were only recorded in *T. petrolearius* and could be identified as the major source of the detected difference in transposase abundance between the three analyzed *Terrisporobacter* genomes. The families IS3 and IS256 contain transposases with the common acidic triad of Asp, Asp and Glu at the active site (DDE transposases). The IS110 family characteristically comprise of IS elements with transposases containing an Asp, Glu, Asp, Asp motif at the active site (DEDD transposase). These IS elements differ from classical DDE transposases as they do not contain terminal inverted repeats and upon integration do not produce flanking target DNA repeats [32]. The identified differences indicate an increased genome plasticity of *T. petrolearius* in comparison to the other two type strains [33].

For a more detailed metabolic investigation of the three *Terrisporobacter* type strain genomes, we analyzed and compared regarding the pathway modules of different KEGG categories (Fig 4). Within the category carbohydrate metabolism, the modules for glycolysis, pyruvate oxidation, pentose phosphate pathway, phosphoribosyl diphosphate (PRPP) biosynthesis,

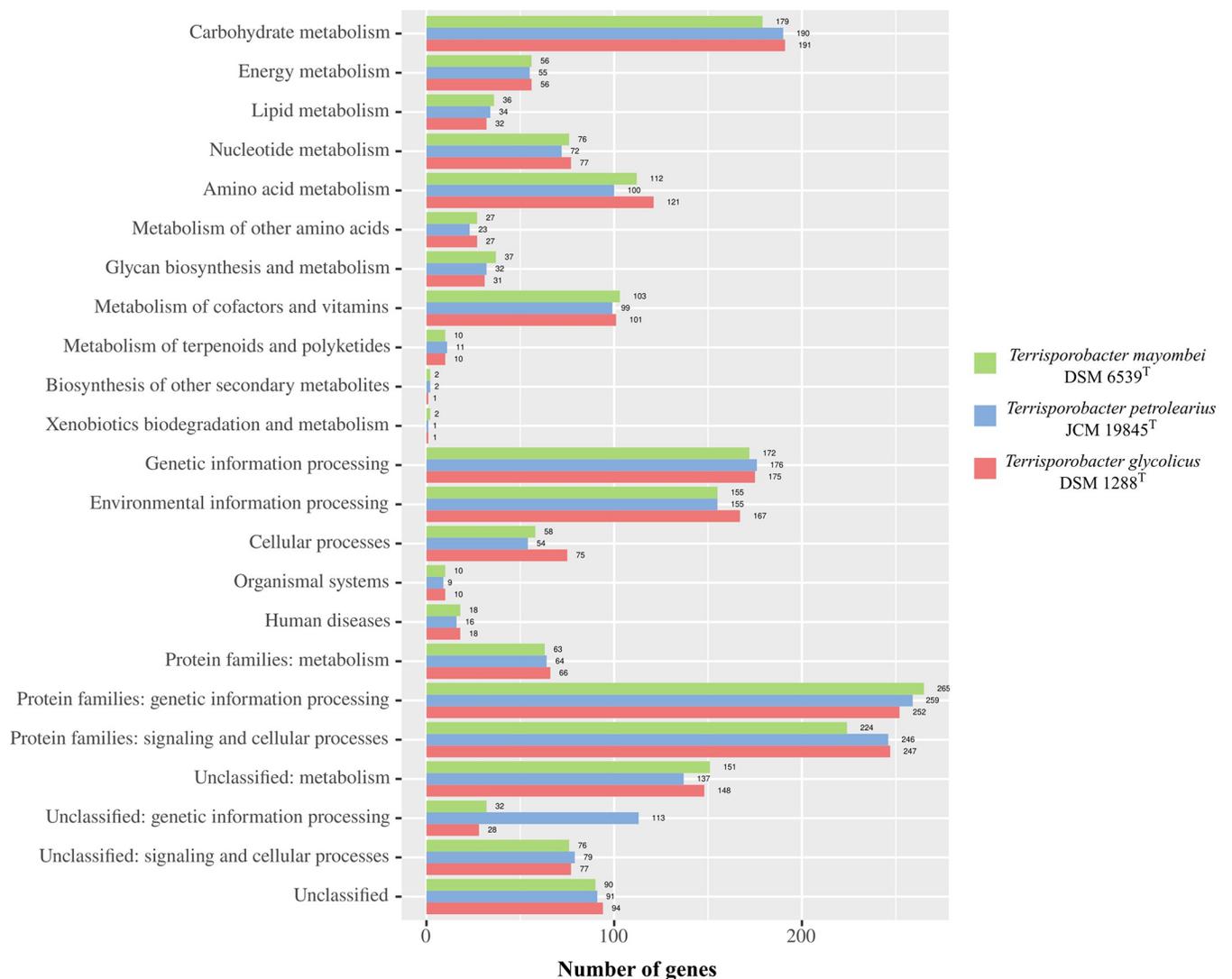


Fig 2. Functional annotation of genes from the *Terrisporobacter* type strain genomes into KEGG categories. The y-axis shows the different functional KEGG categories and the x-axis the number of genes assigned to each category by BlastKOALA. The bar color indicates the number of genes assigned for the three *Terrisporobacter* type strains *T. mayombei* (green), *T. petrolearius* (blue) and *T. glycolicus* (red).

<https://doi.org/10.1371/journal.pone.0290128.g002>

glycogen biosynthesis, nucleotide sugar biosynthesis from glucose and UDP-N-acetyl-D-glucosamine biosynthesis were complete in all three genomes. The modules for gluconeogenesis, citrate cycle, semi-phosphorylative Entner-Doudoroff pathway, D-galacturonate degradation, galactose degradation, glycogen degradation, nucleotide sugar biosynthesis from galactose and propanoyl-CoA metabolism were incomplete in every genome. The module for gluconeogenesis was identified as incomplete due to the lack of a phosphoenolpyruvate (PEP) carboxykinase (both ATP- and GTP-dependent) catalyzing the conversion of oxaloacetate into PEP. Nevertheless, gluconeogenesis remains functional as acetyl-CoA is carboxylated to pyruvate by a pyruvate ferredoxin oxidoreductase. This reaction links the Wood-Ljungdahl pathway to gluconeogenesis and the incomplete citrate cycle, thereby allowing the generation of building blocks for biosynthesis during autotrophic growth [34]. With an incomplete traditional, semi- and non-phosphorylative Entner-Doudoroff pathway, hexoses are metabolized only via glycolysis. Fermentation of pentoses such as xylose can occur via the non-oxidative pentose

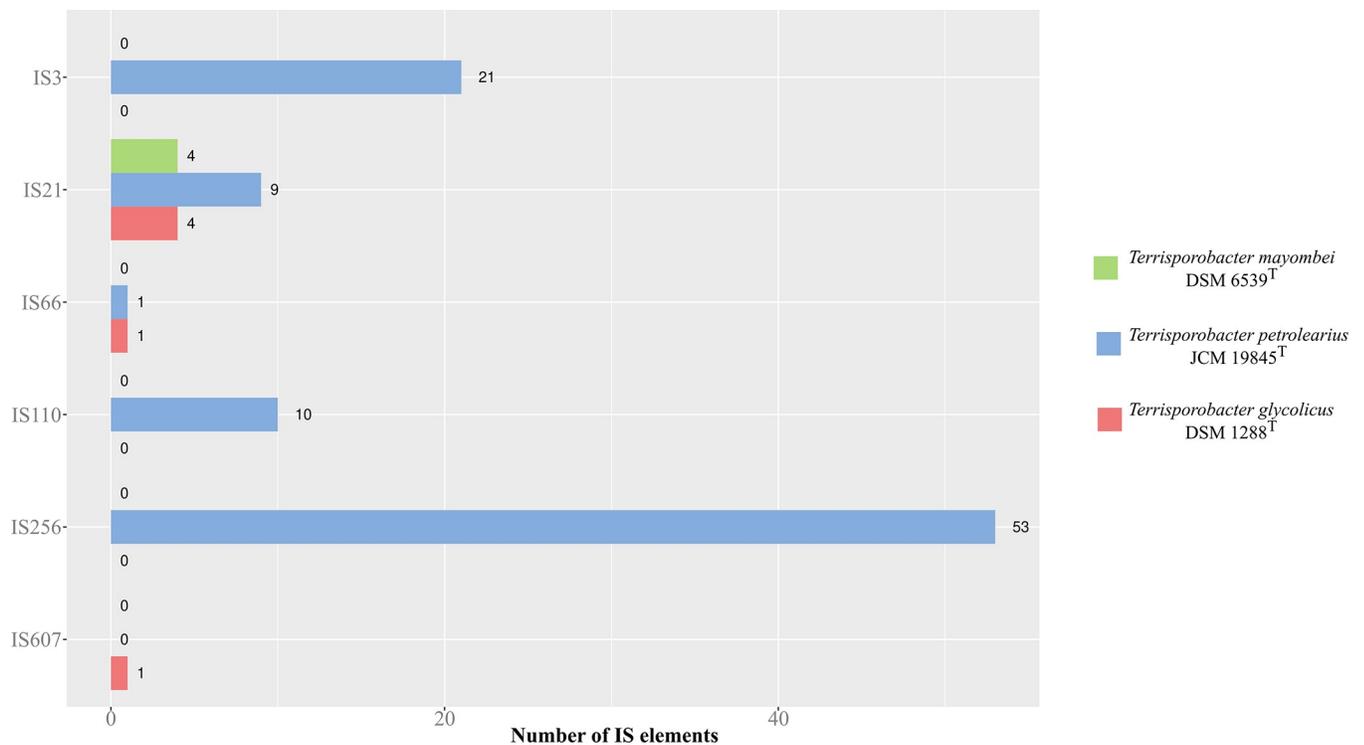


Fig 3. Distribution of IS elements and classification into IS families. The y-axis shows the different IS families and the x-axis the number of IS elements assigned to each IS family by ISESCAN. The bar color indicates the number of genes assigned for the three *Terrisporobacter* type strains *T. mayombeii* (green), *T. petrolearius* (blue) and *T. glycolicus* (red).

<https://doi.org/10.1371/journal.pone.0290128.g003>

phosphate pathway and has been described for the type strains of *T. mayombeii* and *T. glycolicus* [3, 6]. All three analyzed genomes contain an incomplete citrate cycle and are not able to degrade glycogen and galactose. In the category metabolism of cofactors and vitamins every genome contained complete gene sets for the modules of thiamine salvage, NAD biosynthesis, coenzyme A biosynthesis, C1-unit interconversion and the two cobalamin biosynthesis modules. The module for the biosynthesis of pyridoxal-P was only complete in *T. mayombeii* and *T. glycolicus*, while this module was identified to be incomplete in *T. petrolearius* lacking the PdxT subunit of the Pyridoxal 5'-phosphate synthase. Both modules for the biosynthesis of biotin were only complete in *T. mayombeii* and incomplete in the other two type strains. The modules for the biosynthesis of thiamine, riboflavin, pantothenate, pimeloyl-ACP, lipoic acid, tetrahydrofolate, tetrahydrobiopterin, L-threo-tetrahydrobiopterin, molybdenum cofactor, heme and tocopherol/tocotorienol were incomplete in all three *Terrisporobacter* type strains. Siroheme biosynthesis was found to be only complete in *T. mayombeii* and *T. glycolicus*. The genome of *T. petrolearius* is missing genes encoding the sirohydrochlorin ferrochelatase (SirB), the glutamyl-tRNA reductase (HemA) and the glutamate-1-semialdehyde 2,1-aminomutase (HemL). However, all three genomes encode sulfite reductases requiring siroheme as essential prosthetic group [35]. The incomplete module for heme biosynthesis in all three genomes imply the dependence on heme acquisition from the environment or the host during potential pathogenesis [36]. The category amino acid metabolism contained complete gene sets for the modules threonine biosynthesis, cysteine biosynthesis, lysine biosynthesis, polyamine biosynthesis, histidine biosynthesis and histidine degradation in all three genomes. The modules for tyrosine and phenylalanine biosynthesis were only complete for *T. mayombeii* and

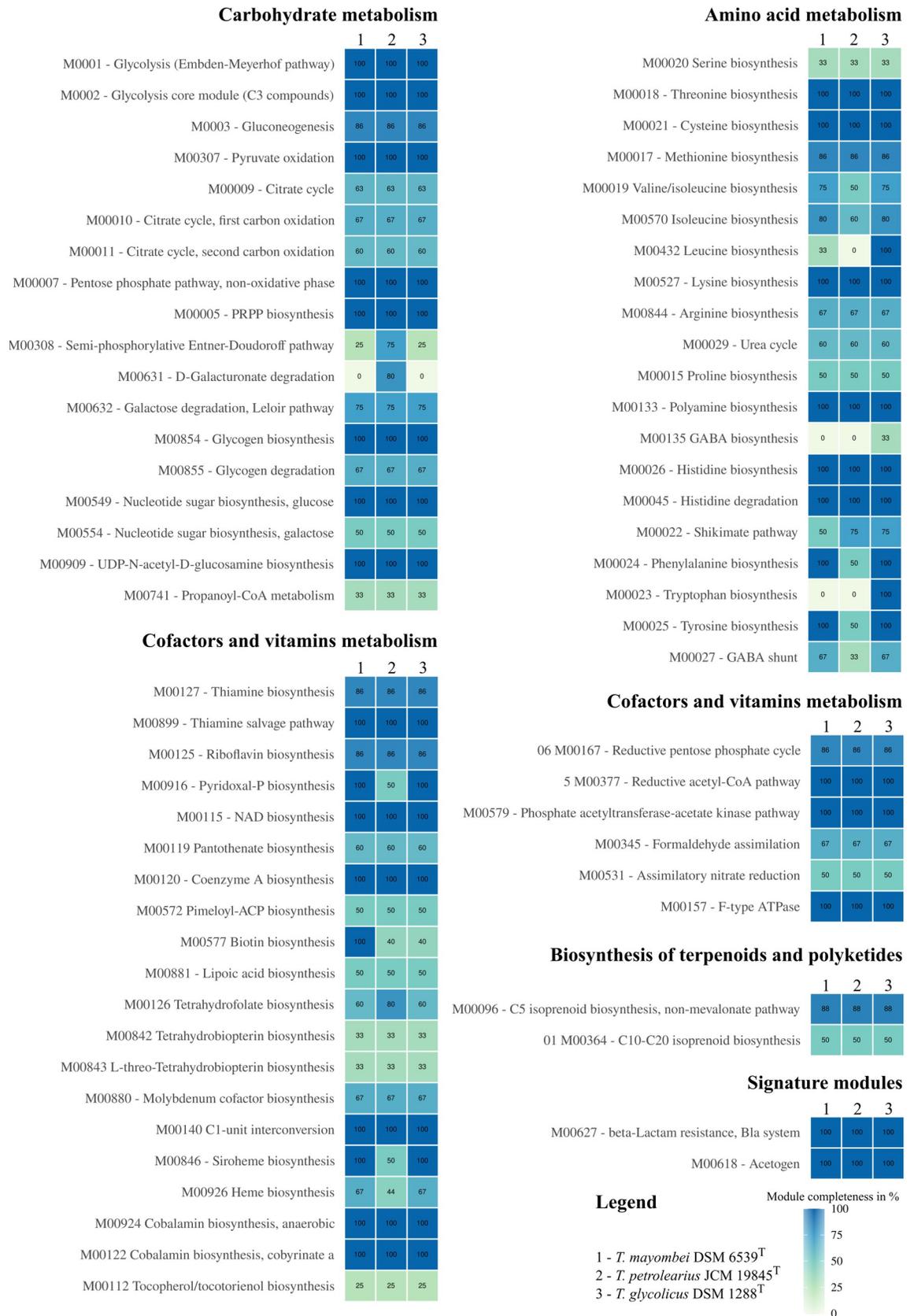


Fig 4. Completeness of KEGG pathway modules for KEGG categories. The completeness of KEGG modules is given in percentage from the total amount of required blocks and the amount of detected blocks in a pathway module for the three *Terrisporobacter* type strains *T. mayombi* (1), *T. petrolearius* (2) and *T. glycolicus* (3). For metabolites with multiple KEGG modules only the most complete modules detected are shown. The legend for the color gradient is shown in the bottom right.

<https://doi.org/10.1371/journal.pone.0290128.g004>

T. glycolicus. Furthermore, the modules for leucine and tryptophan biosynthesis were only complete in *T. glycolicus*. The modules for the biosynthesis of serine, methionine, valine, isoleucine, arginine, proline, gamma-aminobutyric acid GABA and the shikimate pathway were incomplete in all three *Terrisporobacter* type strains. This also applied to the modules for the urea cycle and GABA shunt. The identified incomplete modules in the categories metabolism of cofactors and vitamins and amino acid metabolism explain the observed dependence on the addition of vitamins and amino acids for significant growth of *Terrisporobacter* members [3, 6, 7]. The reconstruction of KEGG pathway modules showed auxotrophies for different vitamins/cofactors and amino acids in all three genomes. Furthermore, several species-specific vitamin and amino acid auxotrophies were identified. In the category energy metabolism complete modules were detected in every *Terrisporobacter* type strain for reductive acetyl-CoA pathway, phosphate acetyltransferase-acetate kinase pathway and F-type ATPase. The modules reductive pentose phosphate cycle, formaldehyde assimilation and assimilatory nitrate reduction were incomplete in all three *Terrisporobacter* strains. No complete modules were present for the category of biosynthesis of terpenoids and polyketides. The signature modules for beta-lactam resistance and acetogen were complete in all three *Terrisporobacter* type strains. The prerequisites for growth on H₂ and CO₂ through acetogenesis are fulfilled for every *Terrisporobacter* type strain analyzed, as shown by the complete modules for the reductive acetyl-CoA pathway (WLP), phosphate acetyltransferase-acetate kinase pathway and the complete signature module for acetogens.

Pathway reconstruction for acetogenesis

The genes necessary to employ the WLP were present in all three type strains of the *Terrisporobacter* genus. The detected genes for the WLP are shown in Fig 5A with the corresponding locus tags for each genome. Overall, the genomic organization of the WLP genes in the *Terrisporobacter* type strains resembled the structure detected in the acetogen *Clostridioides difficile* [37]. The methyl-branch starts with the reversible hydrogen-dependent CO₂ reductase complex (HDCR) using H₂ as a direct electron donor for the reduction of CO₂ to formate by a formate dehydrogenase (FDH). A potential HDCR-complex was encoded by a gene cluster consisting of a selenocysteine-containing formate dehydrogenase H (*fdhF*), two copies of the hydrogenase-4 component A (*hyfA*), a NADP-reducing hydrogenase subunit (*hndD*), a putative formate transporter (*focA*) and a sulfur carrier protein (*fdhD*). In *C. difficile* the HDCR-complex gene cluster did not contain a formate transporter and *fdhD* was located adjacent to the *fdhF* gene, while in *Terrisporobacter* it was located at the end of the gene cluster (Fig 5B). In addition to the HDCR-complex, the oxidation of H₂ in *Terrisporobacter* can be performed by orthologs of the electron-bifurcating hydrogenase HydABC (TEMA_26770–26790, TEPE_39410–39430, TEGE_37760–37780). The activation of formate with tetrahydrofolate (THF) is carried out under consumption of one ATP by the formyl-THF synthetase (FTS) producing N¹⁰-formyl-THF, which is subsequently transformed to N⁵,N¹⁰-methenyl-THF via cyclization and dehydration by the methenyl-THF cyclohydrolase (MTC). The methylene-THF dehydrogenase (MTD) reduces N⁵,N¹⁰-methenyl-THF to N⁵,N¹⁰-methylene-THF, which is then further reduced to N⁵-methyl-THF by the methylene-THF reductase (MTR). Finally, the methyl-group is transferred to a corrinoid-iron-sulfur protein (CoFeSP) by the methyl-

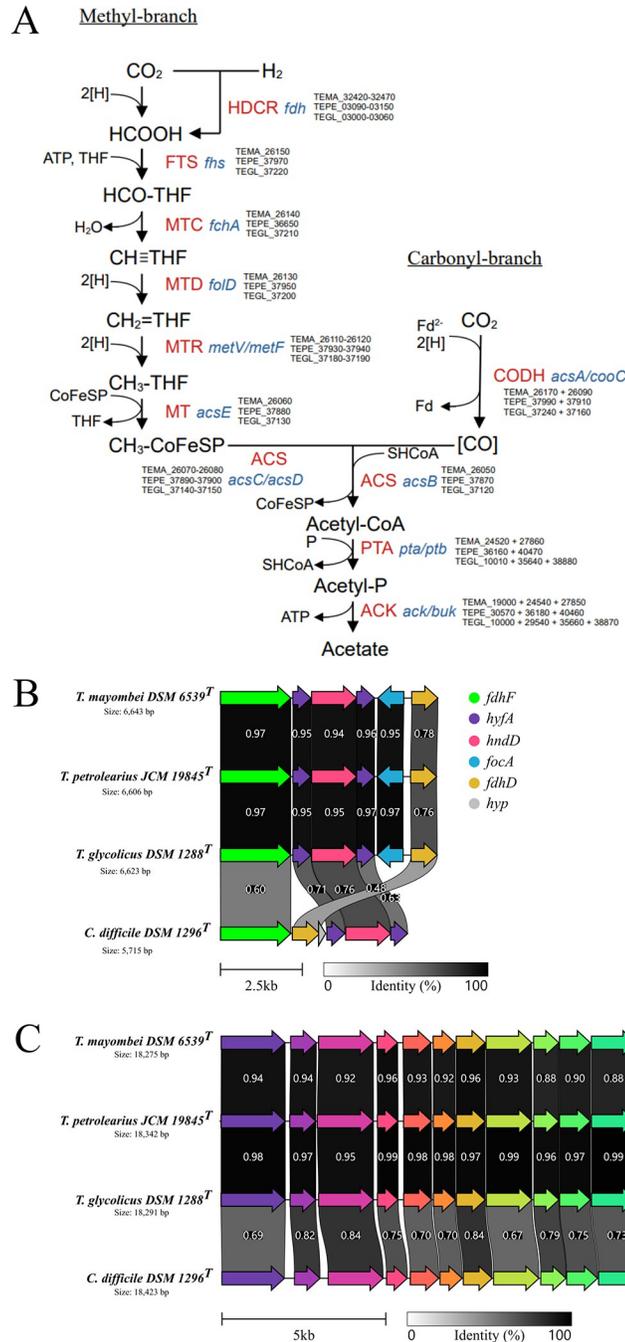


Fig 5. Reconstruction of the Wood-Ljungdahl pathway in *Terrisporobacter*. **A:** Genome-based reconstruction of the Wood-Ljungdahl pathway for the three type strains of the *Terrisporobacter* genus. Genome locus tags (black) of the catalyzing enzymes (red) and the required genes (blue) are shown for the three type strains of the *Terrisporobacter* genus (TEMA, *T. mayombei* DSM 6539^T; TEPE, *T. petrolearius* JCM 19845^T; TEGL, *T. glycolicus* DSM 1288^T). **B:** Genomic organization of the HDCR-complex gene cluster in the three *Terrisporobacter* type strains in comparison to the type strain of *C. difficile* DSM 1296^T. The following gene abbreviations were used: *fdhF*, formate dehydrogenase H; *hyfA*, hydrogenase-4 component A; *hndD*, NADP-reducing hydrogenase; *focA*, putative formate transporter; *fdhD*, sulfur carrier protein. **C:** Genomic organization of the Wood-Ljungdahl gene cluster in the three *Terrisporobacter* type strains in comparison to the type strain of *C. difficile* DSM 1296^T. The following gene abbreviations were used: *acsA*, anaerobic carbon-monoxide dehydrogenase catalytic subunit; *acsF*, carbon monoxide dehydrogenase accessory protein; *fhs*, formyl THF synthetase; *fchA*, methenyl THF cyclohydrolase; *folD*, bifunctional cyclohydrolase/dehydrogenase; *metV*, methylene THF reductase C-terminal catalytic subunit; *metF*, methylene THF reductase large subunit; *lpdA*, dihydrolipoyl dehydrogenase; *cooC*, carbon monoxide dehydrogenase accessory protein; *acsD*, CoFeSP

small subunit; *acsC*, CoFeSP large subunit; *acsE*, methyl THF CoFeSP methyltransferase; *acsB*, carbon monoxide dehydrogenase/acetyl-CoA synthase subunit beta; *gcvH*, glycine cleavage system H protein; *acsV*, corrinoid activation/regeneration protein. The following enzyme abbreviations were used: HDCR, hydrogen-dependent CO₂ reductase; FTS, formyl-THF synthetase; MTC, methenyl-THF cyclohydrolase; MTD, methylene-THF dehydrogenase; MTR, methylene-THF reductase; MT, methyl-transferase; CODH, carbon monoxide dehydrogenase; ACS, acetyl-CoA synthetase; PTA, phosphotransacetylase, ACK, acetate kinase. The legend for size and sequence identity of translated genes are shown at the bottom of **B** and **C** in kilobases (kb) and from 0% (white) to 100% (black).

<https://doi.org/10.1371/journal.pone.0290128.g005>

transferase (MT). In the carbonyl-branch the carbon monoxide dehydrogenase (CODH) reduces CO₂ to a carbonyl-group using reduced ferredoxin. This carbonyl group is fused with the methyl-group from the methyl-branch and the cofactor coenzyme A (CoA), which is catalyzed by the acetyl-CoA synthetase (ACS) and releases CoFeSP. Acetyl-CoA is converted to acetyl phosphate (Acetyl-P) by a phosphotransacetylase (PTA) or alternatively by a phosphotransbutyrylase (PTB). The transfer of the phosphate group to ADP forming ATP is catalyzed by an acetate or alternatively butyrate kinase (ACK/BUK) and leads to the formation of the final product acetate. The genus *Terrisporobacter* is reported to produce only acetate as a fermentation product of the WLP (homoacetogenesis) [2]. The majority of genes involved in the WLP were organized as a gene cluster of approximately 18 kb. These gene clusters showed the same organization as in *C. difficile* or acetogenic bacteria belonging to the genus *Clostridium* (Fig 5C) [38]. Energy conservation is accomplished with the RNF-complex (*rnfABCDEG*; TEMA_15820–15870, TEPE_33980–33930, TEGL_15820–15870) by coupling the electron transfer from reduced ferredoxin to NAD with the translocation of protons or sodium ions across the cytoplasmic membrane. A membrane-bound ATP synthase can then use the created ion gradient to produce ATP [39]. The redox pool of NADH and reduced ferredoxin is linked to the cellular redox pool (NADPH) with a *Sporomusa* type Nfn (STN) electron-bifurcating transhydrogenase (*stnABC*, TEMA_09310–09330, TEPE_23120–23140, TEGL_20000–20020) [40].

Terrisporobacter plasmids

All three *Terrisporobacter* type strain genomes harbored a circular plasmid with an approximate size of 21 kb (pTM1, pTP1 and pTG1). *T. petrolearius* and *T. mayombeii* contained an additional smaller circular plasmid of 6.7 kb (pTP2) and 8.2 kb (pTM2) in size. The annotation of the smaller plasmids pTP2 and pTM2 yielded only genes encoding hypothetical proteins. Hence, inferring possible functional roles of these plasmids was not possible. A circular plasmid of similar size (21.7 kb) as the 21 kb plasmids was also detected in the complete genome of *T. hibernicus* NCTC 14625^T [10]. These plasmids appeared to be a shared trait in the *Terrisporobacter* genus, as they were present with an almost identical genetic organization in the genomes of every *Terrisporobacter* type strain (Fig 6). All three larger plasmids contained a full set of genes required for the proline-dependent reductive reactions of the Stickland reaction. The Stickland reaction describes the coupled oxidation and reduction of amino acids for the conservation of energy in anaerobic bacteria [41]. The plasmids encoded the proprotein *prdA*, which is most likely posttranslationally cleaved resulting in one small protein containing the pyruvoyl-group for binding of D-proline and one larger scaffold protein. The proline reductase was completed by the gene products of *prdB* and *prdC*. As in *C. difficile*, only *prdB* encodes a selenocysteine residue (Sec), while the *prdC* gene did not encode a Sec as typically present in other amino acid-fermenting bacteria such as *Acetoanaerobium sticklandii* or *Clostridium botulinum* [42]. The use of a selenoprotein requires the cotranslational insertion of Sec guided by a complex machinery [43, 44]. The Sec insertion complex was encoded by the *Terrisporobacter* plasmids by the three genes *selA*, *selB*, *selD* and the selenium-binding tRNA gene *selC*. In addition to genes for proline reduction, the plasmids contained genes for the partitioning proteins

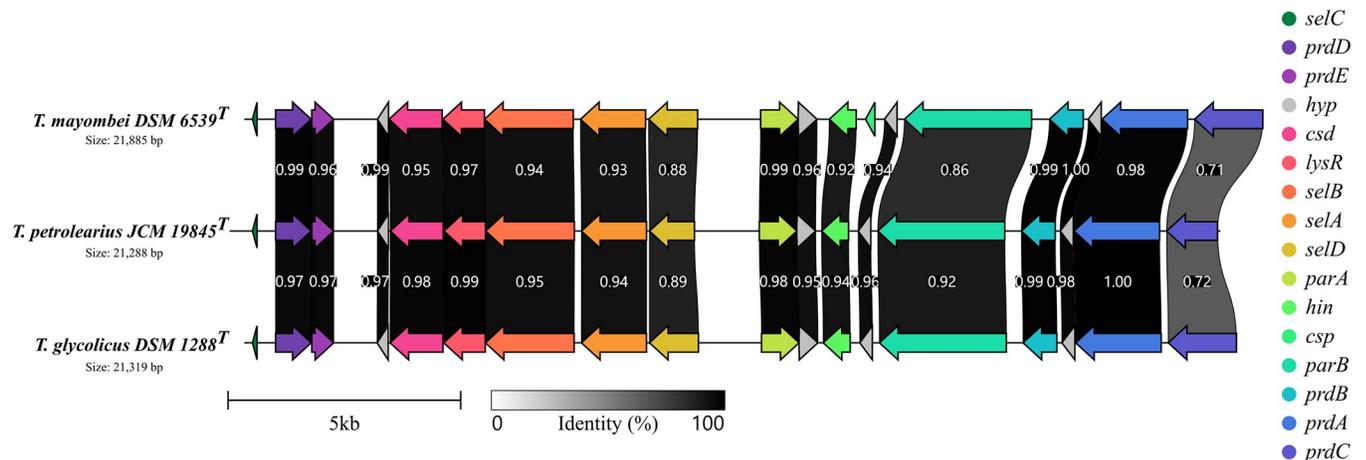


Fig 6. Comparison of the circular plasmids in the three *Terrisporobacter* type strains. The plasmids encode a proline-dependent gene cluster for the reductive branch of the Stickland reaction. The legend for the size and the sequence identity of translated genes is shown at the bottom in kilobases (kb) and from 0% (white) to 100% (black). The following genes are encoded: *selC*, tRNA-Sec; *prdD*, proline reductase cluster protein PrdD; *prdE*, proline reductase cluster protein PrdE; *csd*, putative cysteine desulfurase; *lysR*, lysR family transcriptional regulator; *selB*, selenocysteine-specific elongation factor; *selA*, L-seryl-tRNA (Sec) selenium transferase; *selD*, selenide water dikinase; *parA*, parA family protein; *hin*, DNA-invertase; *csp*, cold shock protein; *parB*, DNA primase; *prdB*, D-proline reductase subunit gamma; *prdA*, D-proline reductase proprotein PrdA; *prdC*, proline reductase-associated electron transfer protein; *hyp*, hypothetical protein.

<https://doi.org/10.1371/journal.pone.0290128.g006>

ParA and ParB, a DNA-invertase (Hin), four hypothetical proteins and a putative cysteine desulfurase (Csd). The plasmid of *T. mayombei* additionally encoded a protein annotated as cold shock protein (Csp). Genes for the reduction of glycine/sarcosine/betaine and the oxidative branch of the Stickland reaction were all encoded chromosomally. This also applied to the proline racemase (PrdF) required for the transformation of amino acids from the L to the D stereoisomer (TEMA_04260 + 37640/TEPE_08380/TEGL_07940) and the membrane-bound protein PrdG (TEMA_37630 + 37650/TEPE_08370 + 08390/TEGL_07930 + 07950). Orthologs of the putative DNA-binding protein PrdX and the sigma-54-dependent transcriptional activator protein PrdR were not identified.

Virulence factor screening

Multiple potential virulence factors were detected in the genomes of the three *Terrisporobacter* type strains comprising the VF classes adherence, regulation, toxin, antiphagocytosis, cell surface components, copper uptake, enzyme, glycosylation systems, immune evasion, iron uptake, others, secretion system, serum resistance and immune evasion and stress adaptation (Table 2). Six virulence factors for cell adherence were detected. Genes encoding a fibronectin-binding protein (*fbpA/fbp68*) and a heat shock protein (*groEL*) were found in every *Terrisporobacter* type strain. These proteins with adhesive function are characteristically present in pathogenic Gram-positive bacteria such as *C. difficile* or *Clostridium perfringens* [45–47].

Furthermore, the *Terrisporobacter* type strain genomes encoded proteins for cell adherence, which are absent in genomes of other pathogenic members of the genus *Clostridium*. These include homologs of the adhesion protein Lap from the genus *Listeria*, the polar flagella from the genus *Aeromonas* and type IV pili from the genus *Neisseria* [48–50]. The genes *cheA/cheY*, *lisR/lisK* and *virR/virS* from the VF class regulation are virulence factors in *Listeria monocytogenes* and were also detected in the *Terrisporobacter* type strains [51–53]. Genes for the production of the toxins hemolysin and cytolyisin involved in host cell lysis were detected in every *Terrisporobacter* strain. Furthermore, several virulence factors were detected associated to

Table 2. Virulence factor screen for three *Terrisporobacter* type strains with VFAnalyzer using the VFDB.

VF class	Virulence factors	Related genes	<i>T. mayombe</i> DSM 6539 ^T	<i>T. petrolearius</i> JCM 19845 ^T	<i>T. glycolicus</i> DSM 1288 ^T	
Adherence	CD0873	<i>CD0873</i>	-	TEPE_05460	-	
	Fibronectin-binding protein	<i>fbpA/fbp68</i>	TEMA_19280	TEPE_30830	TEGL_29820	
	GroEL	<i>groEL</i>	TEMA_27370	TEPE_39970	TEGL_38350	
	Listeria adhesion protein (<i>Listeria</i>)	<i>lap</i>	TEMA_20190; TEMA_20240	TEPE_31700; TEPE_31750	TEGL_30950; TEGL_31000	
	Polar flagella (<i>Aeromonas</i>)		<i>fljP</i>	TEMA_40940	TEPE_11950	TEGL_11730
			<i>flmH</i>	-	TEPE_06230	TEGL_05600
Type IV pili (<i>Neisseria</i>)		<i>pilT</i>	TEMA_31630	TEPE_02190	TEGL_02200	
Regulation	CheA/CheY (<i>Listeria</i>)	<i>cheY</i>	TEMA_41120	TEPE_12130	TEGL_11910	
	LisR/LisK (<i>Listeria</i>)	<i>lisR</i>	TEMA_11910	-	-	
	VirR/VirS (<i>Listeria</i>)	<i>virR</i>	-	TEPE_20480	-	
	Sigma A (<i>Mycobacterium</i>)	<i>sigA/rpoV</i>	TEMA_16220	-	TEGL_26790	
Toxin	Hemolysin	Undetermined	TEMA_35520	TEPE_06520	TEGL_05930	
		Undetermined	TEMA_18780	TEPE_30320	TEGL_29320	
	Cytolysin (<i>Enterococcus</i>)	<i>cylR2</i>	TEMA_35430	TEPE_25820	TEGL_05860	
Antiphagocytosis	Capsular polysaccharide (<i>Vibrio</i>)	<i>rmlB</i>	-	TEPE_11780	-	
	Capsule (<i>Enterococcus</i>)		TEMA_02390	TEPE_14200	TEGL_14000	
Cell surface components	Trehalose-recycling ABC transporter (<i>Mycobacterium</i>)	<i>sugC</i>	TEMA_29560	TEPE_00240	TEGL_00240	
Copper uptake	Copper exporter (<i>Mycobacterium</i>)	<i>ctpV</i>	TEMA_18330	-	-	
Enzyme	Streptococcal enolase (<i>Streptococcus</i>)	<i>eno</i>	TEMA_33900	TEPE_04590	TEGL_04390	
Glycosylation system	O-linked flagellar glycosylation (<i>Campylobacter</i>)	<i>pseB</i>	-	TEPE_09350; TEPE_35440	-	
Immune evasion	Capsule (<i>Acinetobacter</i>)		TEMA_39640	TEPE_35360	-	
	Capsule (<i>Staphylococcus</i>)	<i>capG</i>	TEMA_39660	-	-	
	Capsule (<i>Streptococcus</i>)			TEMA_39140	TEPE_10060	-
			<i>rfaA-1</i>	TEMA_40760	TEPE_11770	-
			<i>cps2K</i>	-	-	TEGL_10420
			<i>cps4I</i>	-	-	TEGL_06050; TEGL_34830
	Exopolysaccharide (<i>Haemophilus</i>)	<i>galE</i>	TEMA_00320	-	-	
	Polyglutamic acid capsule (<i>Bacillus</i>)		<i>capB</i>	TEMA_21920	TEPE_33320	TEGL_32800
			<i>capC</i>	TEMA_21910	TEPE_33310	TEGL_32790
	Polysaccharide capsule (<i>Bacillus</i>)			-	-	TEGL_10330; TEGL_34750
			<i>galE3</i>	TEMA_39740	TEPE_10800	TEGL_10450; TEGL_12550
			<i>epsE</i>	-	TEPE_10660	-
			<i>ugd</i>	-	TEPE_10670	-
Iron uptake	Heme biosynthesis (<i>Haemophilus</i>)	<i>hemB</i>	TEMA_36710	TEPE_07530	TEGL_07060	
		<i>hemL</i>	TEMA_36520	-	-	
	Periplasmic binding protein-dependent ABC transport systems (<i>Vibrio</i>)	<i>vctC</i>	TEMA_37360	TEPE_08140	-	
Others	O-antigen (<i>Yersinia</i>)	<i>ddhA</i>	-	-	TEGL_34860	
Secretion system	Type III secretion system (<i>Chlamydia</i>)	<i>cdsN</i>	TEMA_40860	TEPE_11870	TEGL_11650	
Serum resistance and immune evasion	LPS (<i>Francisella</i>)		TEMA_40770	-	-	
			TEMA_39610	-	-	
		<i>wbtI</i>	TEMA_40740	TEPE_11750	-	
Stress adaptation	Catalase-peroxidase (<i>Mycobacterium</i>)	<i>katG</i>	TEMA_06520	TEPE_26360	TEGL_17140	

<https://doi.org/10.1371/journal.pone.0290128.t002>

capsule formation involved in antiphagocytosis and immune evasion. For the verified *Terrisporobacter* pathogen *T. glycolicus* the *ddhA* gene was detected encoding an O-antigen essential for virulence in *Yersinia enterocolitica* [54]. Homologs of iron uptake proteins have been identified with similarity to proteins from the genus *Haemophilus* (HemB/L) and *Vibrio* (VctC), which matched with the incomplete heme biosynthesis encoded by all three genomes. Overall, the abundance and diversity of virulence factors detected in the three *Terrisporobacter* type strains underline the pathogenic potential harbored by this bacterial genus.

Phylogenomic analysis

The complete genome sequences were compared to available genome sequences of the *Terrisporobacter* genus by average nucleotide identity analysis based on a MUMmer alignment (ANIm) (Fig 7). The genomes of the type strains *T. glycolicus* DSM 1288^T and *T. mayombe* DSM 6539^T showed the highest identity to each other (93.5%). This value was below the species delineation threshold of 95% and supported the classification as separate *Terrisporobacter* species [55]. Using the same threshold, *T. glycolicus* DSM 1288^T and *T. petrolearius* JCM 19845^T also clustered as two distinct species with an identity of 93.1%. However, the genome *T. petrolearius* JCM 19845^T showed the highest identity to the genome sequences of non-type *T. glycolicus* strains MGYG-HGUT-00005 (96.3%), KPPR-9 (96.4%), WW3900 (96.3%), UBA8115 (95.8%), FS03 (96.3%), *T. petrolearius* UN03-225 (96.4%) and *T. hibernicus* NCTC 14625^T (96.4%). These identities were above the species identity threshold of 95%, suggesting that these genomes belonged to the same species, namely *T. petrolearius*. Mitchell et al. reported a 67.4% d4 dDDH value of the complete *T. hibernicus* type strain genome to the draft genome of *T. petrolearius* JCM 19845^T calculated by the Type Strain Genome Server (TYGS) [10, 31]. TYGS submission of the complete genome of *T. petrolearius* JCM 19845^T yielded a d4 dDDH value of 69.2% to the complete *T. hibernicus* genome. However, TYGS recommends the use of the dDDH values d0 and d6 for the comparison of complete genomes, as they also take the genome length into consideration in their calculations [56]. The dDDH values for d0 and d6 between the two complete genomes of *T. petrolearius* JCM 19845^T and *T. hibernicus* were 72.2% and 74.1%, respectively. These dDDH values were above the generally applied species threshold of 70% and thereby further supported the classification of the isolate *T. hibernicus* as member of the *T. petrolearius* species. Our phylogenomic analysis based on complete type strain genomes suggested the existence of subspecies clusters within the *T. petrolearius* species rather than two distinct species. With 99.6% identity between the genome sequences of *T. "othiniensis"* 08–306576 and *T. "muris"* DSM 29186 both names describe isolates of the same species as already described by Afriza et al. [9]. Both isolates were not published under the ICNP as *T. "othiniensis"* was not deposited in a culture collection and *T. "muris"* is only deposited in one culture collection. Hence, no type strain and valid name description are currently available for these *Terrisporobacter* strains. In addition, the genome sequence of the isolate *T. mayombe* MSK4.1 showed an identity of 99.7% to the genomes of *T. "othiniensis"* and *T. "muris"* and hence represents another isolate of this species. The genome sequence of *T. mayombe* ban5-GNNATD2-55 isolate from human feces was below the species delineation threshold of 95% and thereby represented another undescribed *Terrisporobacter* species. The ANIm analysis of metagenomic assembled genomes from the *Terrisporobacter* genus indicated the existence of at least three more *Terrisporobacter* species without available isolates.

Conclusions

Complete genomes for the three *Terrisporobacter* type strains of *T. mayombe*, *T. petrolearius* and *T. glycolicus* were obtained and revealed a *Terrisporobacter* characteristic 21 kb plasmid

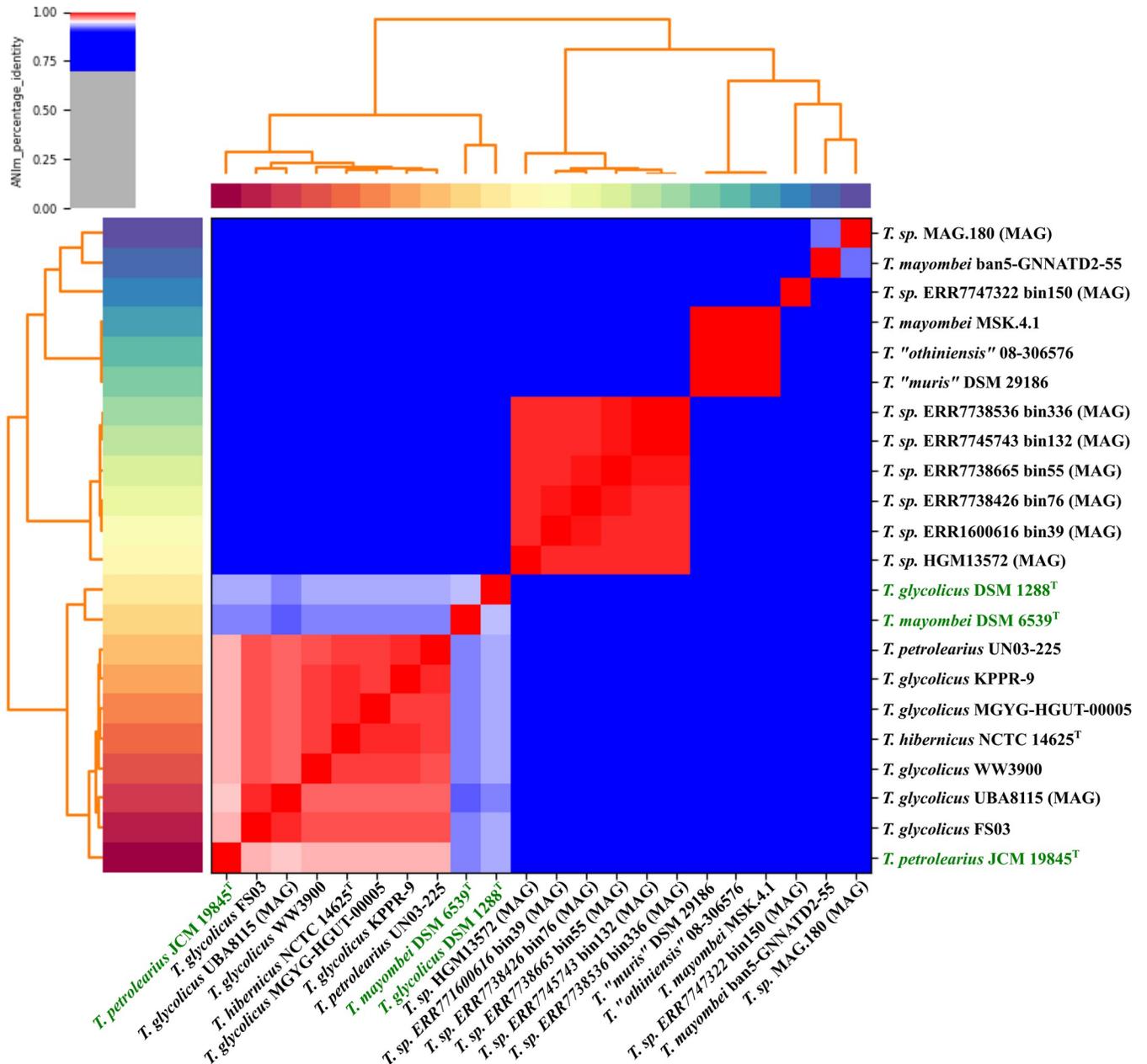


Fig 7. Phylogenomic analysis of *Terrisporobacter* genome sequences. The legend for sequence identities is shown in the top left corner. Sequence identities higher than 95% are shown in red, while lower values are colored grey to blue. The three type strains of the genomes sequenced from the *Terrisporobacter* genus in this study are highlighted in green. Genome sequences derived from metagenomic assemblies are indicated by MAG in parentheses.

<https://doi.org/10.1371/journal.pone.0290128.g007>

encoding a gene cluster for the reductive proline-dependent branch of the Stickland reaction. Functional annotations of the genomes showed species-specific auxotrophies for amino acids and vitamins. The genetic prerequisites for acetogenesis were given in all three *Terrisporobacter* type strains and the WLP could successfully be reconstructed. Our metabolic analysis supports future research for biotechnological applications of acetogenic strains from the genus *Terrisporobacter*. The detection of multiple potential virulence factors also demonstrated a pathogenic potential harbored by the genus *Terrisporobacter*. Phylogenomic analysis of

available *Terrisporobacter* genomes suggested a reclassification of most isolates now termed *T. glycolicus* into *T. petrolearius*.

Supporting information

S1 Table. Kits and program versions for Nanopore sequencing.

(DOCX)

S2 Table. ANIm percentage identity and alignment length values. ANIm percentage identity values (white, in %) and corresponding alignment lengths (grey, in bp) for available *Terrisporobacter* genomes at NCBI Genbank.

(DOCX)

Acknowledgments

We thank Melanie Heinemann for technical assistance and Avril von Hoyningen-Huene for proofreading.

Author Contributions

Conceptualization: Tim Böer.

Data curation: Tim Böer.

Investigation: Tim Böer, Anja Poehlein.

Methodology: Tim Böer, Anja Poehlein.

Visualization: Tim Böer.

Writing – original draft: Tim Böer.

Writing – review & editing: Frank R. Bengelsdorf, Mechthild Bömeke, Rolf Daniel.

References

1. Wang D, Liang Q, Chu N, Zeng RJ, Jiang Y. Deciphering mixotrophic microbial electrosynthesis with shifting product spectrum by genome-centric metagenomics. *Chem Eng J*. 2023; 451:139010.
2. Küsel K, Karnholz A, Trinkwalter T, Devereux R, Acker G, Drake HL. Physiological Ecology of *Clostridium glycolicum* RD-1, an Aerotolerant Acetogen Isolated from Sea Grass Roots. *Appl Environ Microbiol*. 2001; 67(10):4734–41. <https://doi.org/10.1128/AEM.67.10.4734-4741.2001> PMID: 11571179
3. Kane MD, Brauman A, Breznak JA. *Clostridium mayombi* sp. nov., an H₂/CO₂ acetogenic bacterium from the gut of the African soil-feeding termite, *Cubitermes speciosus*. *Arch Microbiol*. 1991; 156(2):99–104.
4. Gerritsen J, Fuentes S, Grievink W, van Niftrik L, Tindall BJ, Timmerman HM, et al. Characterization of *Romboutsia ilealis* gen. nov., sp. nov., isolated from the gastro-intestinal tract of a rat, and proposal for the reclassification of five closely related members of the genus *Clostridium* into the genera *Romboutsia* gen. nov., *Intestinib*. *Int J Syst Evol Microbiol*. 2014 May 1; 64(5):1600–16.
5. Ohwaki K, Hungate RE. Hydrogen utilization by clostridia in sewage sludge. *Appl Environ Microbiol*. 1977; 33(6):1270–4. <https://doi.org/10.1128/aem.33.6.1270-1274.1977> PMID: 879782
6. Gaston LW, Stadtman ER. FERMENTATION OF ETHYLENE GLYCOL BY *CLOSTRIDIUM GLYCOLICUM*, SP. N. *J Bacteriol*. 1963; 85(2):356–62.
7. Deng Y, Guo X, Wang Y, He M, Ma K, Wang H, et al. *Terrisporobacter petrolearius* sp. nov., isolated from an oilfield petroleum reservoir. *Int J Syst Evol Microbiol*. 2015; 65(10):3522–6. <https://doi.org/10.1099/ijsem.0.000450> PMID: 26297478
8. Lund LC, Sydenham TV, Høgh SV, Skov M, Kemp M, Justesen US. Draft Genome Sequence of “*Terrisporobacter othiniensis*” Isolated from a Blood Culture from a Human Patient. *Genome Announc*. 2015; 3(2):e00042–15. <https://doi.org/10.1128/genomeA.00042-15> PMID: 25744994

9. Afrizal A, Jennings SAV, Hitch TCA, Riedel T, Basic M, Panyot A, et al. Enhanced cultured diversity of the mouse gut microbiota enables custom-made synthetic communities. *Cell Host Microbe*. 2022; 30(11):1630–45. <https://doi.org/10.1016/j.chom.2022.09.011> PMID: 36208631
10. Mitchell M, Nguyen S V, Connor M, Fairley DJ, Donoghue O, Marshall H, et al. *Terrisporobacter hibernicus* sp. nov., isolated from bovine faeces in Northern Ireland. *Int J Syst Evol Microbiol*. 2023; 73(1):005667. <https://doi.org/10.1099/ijsem.0.005667> PMID: 36748477
11. Chen T, Xie G, Mi J, Wen X, Cao Z, Ma B, et al. Recovery of the Structure and Function of the Pig Manure Bacterial Community after Enrofloxacin Exposure. *Microbiol Spectr*. 2022; 10(3):e02004–21. <https://doi.org/10.1128/spectrum.02004-21> PMID: 35604139
12. Cheng MP, Domingo MC, Lévesque S, Yansouni CP. A case report of a deep surgical site infection with *Terrisporobacter glycolicus*/T. Mayombei and review of the literature. *BMC Infect Dis*. 2016; 16:529. <https://doi.org/10.1186/s12879-016-1865-8> PMID: 27686579
13. Li Y, Li X, Sun Y, Zhao X, Li Y. Cathodic microbial community adaptation to the removal of chlorinated herbicide in soil microbial fuel cells. *Environ Sci Pollut Res*. 2018; 25:16900–12.
14. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014; 30(15):2114–20. <https://doi.org/10.1093/bioinformatics/btu170> PMID: 24695404
15. Wick RR, Judd LM, Gorrie CL, Holt KE. Completing bacterial genome assemblies with multiplex MinION sequencing. *Microb Genomics*. 2017; 3(10):e000132. <https://doi.org/10.1099/mgen.0.000132> PMID: 29177090
16. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol*. 2017; 13(6):e1005595. <https://doi.org/10.1371/journal.pcbi.1005595> PMID: 28594827
17. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics*. 2015; 31(20):3350–2. <https://doi.org/10.1093/bioinformatics/btv383> PMID: 26099265
18. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014; 30(14):2068–9. <https://doi.org/10.1093/bioinformatics/btu153> PMID: 24642063
19. Manni M, Berkeley MR, Seppely M, Simão FA, Zdobnov EM. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Mol Biol Evol*. 2021; 38(10):4647–54. <https://doi.org/10.1093/molbev/msab199> PMID: 34320186
20. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J Mol Biol*. 2016; 428(4):726–31. <https://doi.org/10.1016/j.jmb.2015.11.006> PMID: 26585406
21. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol Biol Evol*. 2021; 38(12):5825–9. <https://doi.org/10.1093/molbev/msab293> PMID: 34597405
22. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res*. 2019; 47(D1):D309–14. <https://doi.org/10.1093/nar/gky1085> PMID: 30418610
23. R Core Team. R: A language and environment for statistical computing. R Found Stat Comput [Internet]. 2018; Available from: <https://www.r-project.org/>
24. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. 2016.
25. Xu L, Dong Z, Fang L, Luo Y, Wei Z, Guo H, et al. OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res*. 2019; 47(W1):W52–8. <https://doi.org/10.1093/nar/gkz333> PMID: 31053848
26. Xie Z, Tang H. ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. *Bioinformatics*. 2017; 33(21):3340–7. <https://doi.org/10.1093/bioinformatics/btx433> PMID: 29077810
27. Lechner M, Findeiß S, Steiner L, Marz M, Stadler PF, Prohaska SJ. Proteinortho: Detection of (Co-) orthologs in large-scale analysis. *BMC Bioinformatics*. 2011; 12:124. <https://doi.org/10.1186/1471-2105-12-124> PMID: 21526987
28. Gilchrist CLM, Chooi YH. clinker & clustermap.js: automatic generation of gene cluster comparison figures. *Bioinformatics*. 2021; 37(16):2473–5. <https://doi.org/10.1093/bioinformatics/btab007> PMID: 33459763
29. Liu B, Zheng D, Jin Q, Chen L, Yang J. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Res*. 2019; 47(D1):D687–92. <https://doi.org/10.1093/nar/gky1080> PMID: 30395255

30. Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Anal Methods*. 2016; 8(1):12–24.
31. Meier-Kolthoff JP, Göker M. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nat Commun*. 2019; 10:2182. <https://doi.org/10.1038/s41467-019-10210-3> PMID: 31097708
32. Siguier P, Goubeyre E, Chandler M. Bacterial insertion sequences: their genomic impact and diversity. *FEMS Microbiol Rev*. 2014; 38(5):865–91. <https://doi.org/10.1111/1574-6976.12067> PMID: 24499397
33. Bennett PM. Genome Plasticity. In: Woodford N, Johnson AP, editors. *Genomics, Proteomics, and Clinical Bacteriology: Methods and Reviews*. Totowa, NJ; 2004. p. 71–113.
34. Furdai C, Ragsdale SW. The Role of Pyruvate Ferredoxin Oxidoreductase in Pyruvate Synthesis during Autotrophic Growth by the Wood-Ljungdahl Pathway. *J Biol Chem*. 2000; 275(37):28494–9. <https://doi.org/10.1074/jbc.M003291200> PMID: 10878009
35. Crane BR, Getzoff ED. The relationship between structure and function for the sulfite reductases. *Curr Opin Struct Biol*. 1996; 6(6):744–56. [https://doi.org/10.1016/s0959-440x\(96\)80003-0](https://doi.org/10.1016/s0959-440x(96)80003-0) PMID: 8994874
36. Choby JE, Skaar EP. Heme Synthesis and Acquisition in Bacterial Pathogens. *J Mol Biol*. 2016; 428(17):3408–28. <https://doi.org/10.1016/j.jmb.2016.03.018> PMID: 27019298
37. Gencic S, Grahame DA. Diverse Energy-Conserving Pathways in *Clostridium difficile*: Growth in the Absence of Amino Acid Stickland Acceptors and the Role of the Wood-Ljungdahl Pathway. *J Bacteriol*. 2020; 202(20):e00233–20. <https://doi.org/10.1128/JB.00233-20> PMID: 32967909
38. Poehlein A, Cebulla M, Ilg MM, Bengelsdorf FR, Schiel-Bengelsdorf B, Whited G, et al. The Complete Genome Sequence of *Clostridium acetivum*: a Missing Link between Fnf- and Cytochrome-Containing Autotrophic Acetogens. *MBio*. 2015; 6(5):e01168–15. <https://doi.org/10.1128/mBio.01168-15> PMID: 26350967
39. Schuchmann K, Müller V. Autotrophy at the thermodynamic limit of life: a model for energy conservation in acetogenic bacteria. *Nat Rev Microbiol*. 2014; 12(12):809–21. <https://doi.org/10.1038/nrmicro3365> PMID: 25383604
40. Kremp F, Roth J, Müller V. The Sporomusa type Nfn is a novel type of electron-bifurcating transhydrogenase that links the redox pools in acetogenic bacteria. *Sci Rep*. 2020; 10:14872. <https://doi.org/10.1038/s41598-020-71038-2> PMID: 32913242
41. Stickland LH. Studies in the metabolism of the strict anaerobes (genus *Clostridium*). *Biochem J*. 1935; 29(4):889–98.
42. Fonknechten N, Chaussonnerie S, Tricot S, Lajus A, Andreesen JR, Perchat N, et al. *Clostridium sticklandii*, a specialist in amino acid degradation: revisiting its metabolism through its genome sequence. *BMC Genomics*. 2010; 11:555. <https://doi.org/10.1186/1471-2164-11-555> PMID: 20937090
43. Gonzalez-Flores JN, Shetty SP, Dubey A, Copeland PR. The molecular biology of selenocysteine. *Biomol Concepts*. 2013; 4(4):349–65. <https://doi.org/10.1515/bmc-2013-0007> PMID: 25436585
44. Zinoni F, Birkmann A, Stadtman TC, Böck A. Nucleotide sequence and expression of the selenocysteine-containing polypeptide of formate dehydrogenase (formate-hydrogen-lyase-linked) from *Escherichia coli*. *Proc Natl Acad Sci*. 1986; 83(13):4650–4. <https://doi.org/10.1073/pnas.83.13.4650> PMID: 2941757
45. Hennequin C, Janoir C, Barc M, Collignon A, Karjalainen T. Identification and characterization of a fibronectin-binding protein from *Clostridium difficile*. *Microbiology*. 2003; 149(10):2779–87.
46. Collignon A, Bourlioux P, Waligora-Dupriet AJ, Karjalainen T, Barc MC, Hennequin C, et al. GroEL (Hsp60) of *Clostridium difficile* is involved in cell adherence. *Microbiology*. 2001; 147(1):87–96. <https://doi.org/10.1099/00221287-147-1-87> PMID: 11160803
47. Mehdizadeh Gohari I, A. Navarro M, Li J, Shrestha A, Uzal F, A. McClane B. Pathogenicity and virulence of *Clostridium perfringens*. *Virulence*. 2021; 12(1):723–53. <https://doi.org/10.1080/21505594.2021.1886777> PMID: 33843463
48. Wampler JL, Kim KP, Jaradat Z, Bhunia AK. Heat Shock Protein 60 Acts as a Receptor for the *Listeria* Adhesion Protein in Caco-2 Cells. *Infect Immun*. 2004; 72(2):931–6. <https://doi.org/10.1128/IAI.72.2.931-936.2004> PMID: 14742538
49. Tomás JM. The Main *Aeromonas* Pathogenic Factors. *ISRN Microbiol*. 2012; 2012:256261. <https://doi.org/10.5402/2012/256261> PMID: 23724321
50. Thanassi DG, Bliska JB, Christie PJ. Surface organelles assembled by secretion systems of Gram-negative bacteria: diversity in structure and function. *FEMS Microbiol Rev*. 2012; 36(6):1046–82. <https://doi.org/10.1111/j.1574-6976.2012.00342.x> PMID: 22545799
51. Dons L, Eriksson E, Jin Y, Rottenberg ME, Kristensson K, Larsen CN, et al. Role of Flagellin and the Two-Component CheA/CheY System of *Listeria monocytogenes* in Host Cell Invasion and Virulence. *Infect Immun*. 2004; 72(6):3237–44. <https://doi.org/10.1128/IAI.72.6.3237-3244.2004> PMID: 15155625

52. Cotter PD, Guinane CM, Hill C. The LisRK Signal Transduction System Determines the Sensitivity of *Listeria monocytogenes* to Nisin and Cephalosporins. *Antimicrob Agents Chemother*. 2002; 46(9):2784–90. <https://doi.org/10.1128/AAC.46.9.2784-2790.2002> PMID: 12183229
53. Mandin P, Fsihi H, Dussurget O, Vergassola M, Milohanic E, Toledo-Arana A, et al. VirR, a response regulator critical for *Listeria monocytogenes* virulence. *Mol Microbiol*. 2005; 57(5):1367–80. <https://doi.org/10.1111/j.1365-2958.2005.04776.x> PMID: 16102006
54. Bengoechea JA, Najdenski H, Skurnik M. Lipopolysaccharide O antigen status of *Yersinia enterocolitica* O:8 is essential for virulence and absence of O antigen affects the expression of other *Yersinia* virulence factors. *Mol Microbiol*. 2004; 52(2):451–69.
55. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol*. 2007; 57(1):81–91. <https://doi.org/10.1099/ijs.0.64483-0> PMID: 17220447
56. Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics*. 2013; 14:60. <https://doi.org/10.1186/1471-2105-14-60> PMID: 23432962