Molecular evolution of the Wood-Ljungdahl pathway and the reductive glycine pathway in Thermodesulfobacteriota

Tomoyuki Wakashima[1,2], Keitaro Kume[3,4], Yoko Chiba[1,5*]

[1]RIKEN Center for Sustainable Resource Science, Saitama, Japan

[2]Graduare School of Science and Technology, University of Tsukuba, Ibaraki, Japan

[3]Institute of Medicine, University of Tsukuba, Ibaraki, Japan

[4]Center for Cyber Medicine Research, University of Tsukuba, Ibaraki, Japan

[5]Institute of Life and Environmental Sciences, University of Tsukuba, Ibaraki, Japan

*** Correspondence:**
Yoko Chiba yoko.chiba.ey@a.riken.jp

Keywords: Carbon fixation$_1$, Molecular evolution$_2$, Carbon metabolism$_3$, Bacterial evolution$_4$, Thermodesulfobacteriota$_5$, Wood-Ljungdahl pathway$_6$, Reductive glycine pathway$_7$.

Abstract

Carbon fixation is a fundamental metabolic process that sustains ecosystems, yet its origins and evolutionary history remain largely unresolved. In this study, we focused on the Wood-Ljungdahl (WL) pathway, which is considered one of the most ancient carbon fixation pathways and the reductive glycine (rGly) pathway, which shares several reactions with the WL pathway. The evolutionary scenario of the two carbon fixation pathways was inferred in the phylum Thermodesulfobacteriota, which includes microorganisms that operate either the WL pathway or the rGly pathway for autotrophic growth. The timing of gene gain and loss events was inferred by gene presence/absence analyses for both pathways, together with phylogenetic analyses of their key enzymes. Our results suggested that the common ancestor of Thermodesulfobacteriota possessed all genes encoding key enzymes of both pathways; formate dehydrogenase, the carbon monoxide dehydrogenase/acetyl-CoA synthase complex and the glycine cleavage system. Furthermore, analyses of complete gene sets for the WL and rGly pathway, together with downstream genes required for amino acid biosynthesis, supported the possibility that the common ancestor of this phylum had been capable of autotrophic growth through these carbon fixation pathways. Then, multiple lineages have lost the WL and rGly pathway genes independently during subsequent evolution. Gene replacements also occurred in the glycine cleavage system by regaining genes by horizontal gene transfer. These results suggest that carbon fixation pathways in extant organisms in the phylum Thermodesulfobacteriota arose through a combination of vertical inheritance, gene loss, and horizontal gene transfer.

1 Introduction

Elucidating the origins and early evolution of carbon fixation pathways is essential to understand the origin of life. Whether the first organisms on Earth were autotrophs or heterotrophs remains unresolved (Bada and Lazcano, 2002; Kitadai and Maruyama, 2018; Wächtershäuser, 1988; Kitadai et al., 2021), however, there is no doubt that the emergence of autotrophs was crucial for establishing ecosystem on Earth. This is because organic compounds produced solely through abiotic processes are insufficient to sustain a substantial biomass (Martin and Russell, 2007; Nunoura et al., 2018; Weiss et al., 2016; Berg et al., 2010; Dick, 2019). There is growing interest in how early life fixed carbon dioxide (CO_2) and how it evolved to acquire the complex and highly refined metabolic systems observed today (Fuchs, 2011; Braakman and Smith, 2012; Moody et al., 2024).

To date, seven carbon fixation pathways have been identified in extant organisms in nature (Berg et al., 2010; Fuchs, 2011; Sánchez-Andrea et al., 2020). Among these pathways, the Wood-Ljungdahl (WL) pathway, which is also called the reductive acetyl-CoA pathway, is considered to be one of the oldest ones and existed in the last universal common ancestor (LUCA), owing to its broad conservation across diverse lineages including bacteria and archaea (Martin and Russell, 2007; Weiss et al., 2016; Berg et al., 2010; Braakman and Smith, 2012; Schuchmann and Müller, 2014; Moody et al., 2024; Nitschke and Russell, 2013; Peretó et al., 1999). The old origin of the WL pathway is also supported by its structural simplicity and high energy efficiency (Martin and Russell, 2007; Berg et al., 2010; Fuchs, 2011; Schuchmann and Müller, 2014).

The WL pathway incorporates two molecules of CO_2 through the methyl and the carbonyl branches to synthesize one molecule of acetyl-CoA (Fig. 1, highlighted in blue) (Schuchmann and Müller, 2014). In the methyl branch of the bacterial WL pathway, CO_2 is converted to formate by formate dehydrogenase (FDH) and subsequently reduced to a methyl group ($-CH_3$) bound to tetrahydrofolate (THF) to produce CH_3-THF. In the carbonyl branch, CO_2 is reduced to CO by carbon monoxide dehydrogenase (CODH) and then combines with the methyl group of the CH_3-THF and CoA to become acetyl-CoA by acetyl-CoA synthase (ACS). CODH and ACS involved in the WL pathway form a complex composed of five subunits; AcsA, AcsB, AcsC, and AcsD, are conserved between bacteria and archaea and AcsE and CdhB are specific to bacteria and archaea, respectively (Supplemental Table. S1). In both bacteria and archaea, the genes for these five subunits form a cluster (Adam et al., 2018). Both genetic and biochemical studies have confirmed that this pathway actually functions for carbon fixation in a wide range of taxa, including methanogens, acetogens, and sulfate-reducing bacteria, spanning both archaeal and bacterial domains (Schuchmann and Müller, 2014; Adam et al., 2018; Borrel et al., 2016; Pierce et al., 2008).

The reductive glycine (rGly) pathway (Fig. 1, highlighted in yellow) is the most recently identified CO_2 fixation pathway that supports the autotrophic growth of naturally existing organisms (Sánchez-Andrea et al., 2020; Song et al., 2020). This pathway utilizes the reverse reaction of the glycine cleavage system (GCS), which cleave glycine to methylene tetrahydrofolate (CH_2-THF), CO_2, and ammonia (Fig. 1, highlighted in yellow) (Kikuchi et al., 2008; Kikuchi, 1973; Kim et al., 2017). GCS is composed of five proteins named GCST, GCSPα, GCSPβ, GCSL, and GCSH, and their genes are known to form a cluster (Okamura-Ikeda et al., 1993; Cao et al., 2018; Tezuka and Ohnishi, 2014; Lorio et al., 2010). GCSP catalyzes the decarboxylation of glycine and the product, aminomethyl group ($-CH_2-NH_2$), is transferred to a carrier protein, GCSH. The aminomethyl group reacts with tetrahydrofolate (THF) to produce CH_2-THF and ammonia, which is catalyzed by GCST. The released GCSH is then re-oxidizes by GCSL, a FAD-dependent oxidoreductase (Kikuchi et al., 2008; Kikuchi, 1973). Because the GCS reaction is thermodynamically near equilibrium (Kikuchi et al.,

2008), glycine synthesis through the reverse GCS reaction to supplement organic compound has been known for a long time in some heterotrophs (Schneeberger, 1999; Dürre and Andreesen, 1982; Waber and Wood, 1979; Fuchs, 1986). On the other hand, biochemically demonstrated CO_2 fixation to support the autotrophic growth via the reverse GCS reaction (i.e., the rGly pathway) is limited to two species; *Desulfovibrio desulfuricans* belonging to the phylum Thermodesulfobacteriota (TDB) (Sánchez-Andrea et al., 2020) and *Clostridium drakei* belonging to the phylum Bacillota (Song et al., 2020). Notably, the pathway to provide CH_2-THF from CO_2 is shared between the rGly and WL pathways (Fig. 1, S1) (Sánchez-Andrea et al., 2020), suggesting the profound functional and evolutionary relationships between them (Song et al., 2020; Cotton et al., 2018). In fact, *Clostridium drakei* has been biochemically proven to operate both the rGly and WL pathways for CO_2 fixation (Song et al., 2020). Furthermore, last bacterial common ancestor (LBCA) is suggested to have retained some part of the genes for these metabolic pathways. Ancestral metabolic reconstruction based on the gene phylogenies across major bacterial lineages identified methyl-branch enzyme genes shared between the WL and rGly pathways in the LBCA, although it did not support the presence of FDH, which catalyzes the first reaction in these pathways (Coleman et al., 2021).

FDH working for the WL or rGly pathways forms a hetromeric protein complex (Maia et al., 2015). For example, the FDH in *Moorella thermoacetica*, a model bacterium for the WL pathway, has an $\alpha_2\beta_2$ structure (Yamamoto et al., 1983). In contrast, the FDH in *D. desulfuricans*, which operates the rGly pathway, consists of three subunits; α, β, and γ (Maia et al., 2016). The FDHα subunit in both species contains a metal that binds to CO_2 and belong to the MopB superfamily, however, they are catabolized into distinct groups; the FDHα subunits of *M. thermoacetica* and *D. desulfuricans* are categorized into cytoplasmic FDH and FdhG, respectively, based on the distinct clade on their phylogenetic tree (Wells et al., 2023). In contrast, FDHβ exhibit no homology between in *M. thermoacetica* and in *D. desulfuricans*, and therefore expected to have different evolutionary origins. The phylogenetic diversity of FDH makes it difficult to estimate the existence/absent of the gene in ancestral organisms based on phylogenetic analysis, because the estimation is influenced by the query, taxon sampling, and root positions (Coleman et al., 2021; Powell and Battistuzzi, 2022; Superson et al., 2019).

Here, we estimated the evolutionary scenario of the genes related to CO_2 fixation pathways in a thermophilic and sulfate-reducing bacterial phylum Thermodesulfobacteriota (TDB). The phylum TDB, formerly classified as Deltaproteobacteria, belongs to the Gracilicutes within the bacterial domain (Coleman et al., 2021) and is phylogenetically related to the phyla Myxococcota, Bdellovibrionota, and Proteobacteria (Fig. S2) (Waite et al., 2020). TDB includes *Desulfovibrio desulfuricans* and *Thermodesulfatator indicus* growing chemolithoautotrophically using the rGly and WL pathways, respectively (Sánchez-Andrea et al., 2020; Chiba et al., 2025). Therefore, this phylum is suitable to estimate the presence or absence of the gene set for the rGly and WL pathways in the common ancestor of this phylum and how it evolved to the extant organisms which have either pathway. To access these questions, we conducted presence/absence analyses for both pathway genes to identify the gene sets in extant organisms and performed phylogenetic analyses to estimate how the genes have been inherited.

2 Materials and Methods

2.1 Gene presence/absence analysis

To examine the gene presence/absence, we selected at least one representative species from each of the 42 families within to the phylum TDB based on the classification proposed by Waite et al. (Waite et al., 2020). Priority was given to those whose genomes are registered in RefSeq.

Proteomes for each representative species were predicted from NCBI RefSeq, and the presence or absence of enzymes involved in the WL and rGly, and subsequent amino acid synthesis pathways was estimated by homology searches using BLAST (Altschul et al., 1990). As queries, we used 39 amino acid sequences of the enzymes from *M. thermoacetica, D. desulfuricans, T. indicus, Escherichia coli, Thermus thermophilus* and *Hydrogenobacter thermophilus* (Supplementary Table 2). In the BLAST searches, we defined the homolog as those with an e-value of <1e-5 and query coverage of 70% or more.

In this study, we defined FDH, the CODH/ACS complex, and the GCS complex as the key enzymes of the WL and rGly pathways. We inferred that a species possessed the WL pathway if its genome encoded both a homolog of FDHα subunit and all five proteins of the CODH/ACS complex; AcsA, AcsB, AcsC, AcsD, and AcsE. We inferred that a species possessed the rGly pathway if its genome encoded homologs of FDHα and all the five GCS complex proteins; GCSPα, GCSPβ, GCST, GCSL, and GCSH.

2.2 Phylogenetic analyses

The target enzymes for phylogenetic analyses are as follows: FDH, the CODH/ACS complex, the GCS complex, pyruvate-ferredoxin oxidoreductase (PFO), phosphoserine phosphatase (PSP), serine hydroxymethyltransferase (SHMT), and the glycine reductase (GR) complex. Finally, a total of 28 enzyme subunits were the subjects of the analyses (Supplementary Table S3).

We collected sequences by MMseqs2 (release 14-7e284) similarity searches (Steinegger and Söding, 2017) against a local NCBI-nr database (ver. 2024/Nov/03). In these searches, we used sequences whose enzymatic functions were biochemically analyzed (Supplementary Table S3). The command line options of MMseqs2 search were "-e 1e-5 -s 7.0 --num-iterations 5". From the result of the similarity search of each query, we first removed bottom 90% of MAG (metagenome-assembled genome)-derived hits by e-value. Then, we selected one representative sequence from each of NCBI taxid set we specified to cover the diversity of the tree of life. This set includes NCBI taxids of 199: 139 from whole organisms (for detail, see the legend to Supplementary Figure 1 in the previous study (Koganemaru et al., submitted)) and 60 related to TDB. These representatives on each query were finally merged. Next, the domain structures of the merged sequences were identified using InterProScan ver. 5.72, and sequences whose domain structures did not match those of the queries were removed (Jones et al., 2014). Sequences with large insertions or deletion were also removed to ensure accurate alignments. Next, in the preliminary phylogeny, we manually excluded sequences that exhibited long branches (>1.5 substitution per site) in the preliminary phylogenetic trees to minimize the effects of long-branch attraction (LBA) artefact (Bergsten, 2005; Graybeal, 1998; Hedtke et al., 2006). Because more stringent filtering criteria were applied in this phylogenetic analysis compared to the preceding gene presence/absence analysis to avoid the LBA artefact, some sequences that had been identified as "present" in the preceding analysis were excluded from the

phylogenetic analysis. The numbers of sequences ultimately used and excluded are listed in Supplementary Table S3.

The curated sequences were aligned using MUSCLE (Edgar, 2004) and subsequently trimmed with G-blocks (Talavera and Castresana, 2007). The following three options were enabled for trimming: I. Allow smaller final blocks, II. Allow gap positions within the final blocks, and III. Allow less strict flanking positions. Phylogenetic trees were then inferred by maximum likelihood method using IQ-TREE ver. 2.3.6 on the final trimmed alignment data (Minh et al., 2020). Model selection was performed using ModelFinder "-m MFP" (Kalyaanamoorthy et al., 2017), which automatically determined the best-fit model for each enzyme. Node support values were evaluated using ultrafast bootstrap analysis "-B 1000" (Hoang et al., 2018). The evolutionary models and the number of selected amino acid sites used to infer each phylogenetic tree were provided in the legends of the respective figures. The workflow for the phylogenetic analysis is illustrated in Supplementary Figure S10.

For FDHα, a single phylogenetic tree was inferred by combining the newly collected homologous sequences of W-FDH (cytoplasmic Fdhs) from *M. thermoacetica* and Mo-FDH (FdhG) from *D. desulfuricans* with known sequences from the MopB superfamily (Wells et al., 2023). Subsequently, these sequences were divided into cytoplasmic Fdhs and FdhG groups, and separate phylogenetic trees were inferred for each.

3 Results

3.1 Presence or absence of carbon fixation pathway genes

The presence or absence of all the 39 genes for enzymes involved in the WL and rGly pathways and subsequent amino acid synthesis was determined in 54 representative species in the phylum TDB, which is classified into clade A to E (Fig. 2, Supplemental Table S4). Literature search revealed at least 15 out of the 54 species can grow chemolithoautotrophically and they were distributed across clades A, B, C and D (Sánchez-Andrea et al., 2020; Pikuta et al., 2003; Krukenberg et al., 2016; Finster et al., 1998; Moussard et al., 2004; Alain et al., 2010; Lai et al., 2016; Kojima et al., 2016; Hamilton-Brehm et al., 2013; Mardanov et al., 2016; Frolova et al., 2018; Slobodkina et al., 2017; Slobodkin et al., 2013; Cravo-Laureau et al., 2004; DeWeerd et al., 1990). In the 54 representative species, 50 and 44 species possessed the homologs of cytoplasmic Fdhs in *M. thermoacetica* and FdhG in *D. desulfuricans*, respectively (Supplemental Table S4-C). As the results, all the analyzed species was found to have at least one FDHα candidate genes with only one exception of an uncultured *Candidatus* Dadabacteria bacterium (Fig.2). Furthermore, 41 species possessed both cytoplasmic Fdhs and FdhG homologs and the 41 species were distributed in all five clades (A-E) of the phylum TDB. This suggests that the common ancestor of the phylum TDB already possessed genes for both cytoplasmic Fdhs and FdhG.

The complete gene set for the CODH/ACS complex was identified in 27 species belonging to clades B, C, D, and E, including *T. indicus*, which actually operates the WL pathway for CO_2 fixation and belongs to clade B (Fig. 2). The complete gene set for the GCS complex was identified in 21 species in clades A, C, D, and E including *D. desulfuricans*, which possesses the rGly pathway and belongs to clade A (Fig. 2). Except for the uncultured *Candidatus* Dadabacteria bacterium, the organisms possessing the gene set for either the CODH/ACS or GCS complex also had FDH genes, suggesting that they may possess the WL or rGly pathway. In particular, species reported to grow

autotrophically have high possibility of fixing CO_2 via either the WL or rGly pathway. All the ten species analyzed in clade A had the gene set for the rGly pathway, whereas none of them had the WL pathway gene set. Conversely, in clade B, 11 out of 13 species conserved the WL pathway gene set, and no species were found to have the rGly pathway gene set (Fig. 2). In contrast to mutually exclusive distribution of the rGly and WL pathways in clade A and B, respectively, both the pathways were estimated to exist in more deeply branching clades C, D, and E compared to clades A and B. Notably, *Desulfoluna butyratoxydans* and *Desulfatirhabdium butyrativorans* in clade C possessed both the gene sets for the WL and rGly pathways. These interspecies distribution of gene sets for the WL and rGly pathways suggested that both the gene sets may have been present in the common ancestor of the phylum TDB.

3.2 Phylogenetic analyses of the enzymes for the WL and rGly pathway

If the common ancestor of the phylum TDB possessed the gene sets for both the WL and rGly pathways and these genes were vertically inherited, the phylogeny of these genes would be consistent with the phylogeny of species and exhibited the monophyly in the phylum TDB. Therefore, we conducted phylogenetic analyses focusing on the enzyme genes constituting these pathways to determine whether each enzyme gene was vertically inherited within this phylum or secondarily acquired through horizontal gene transfer (HGT).

A total of 28 phylogenetic trees of the enzymes related to the WL pathway, rGly pathway and subsequent amino acid synthesis were inferred (Figures 3-5 and Supplementary Figures S3-S9). The following section focuses on FDH, CODH/ACS complex, and GCS complex and provides detailed discussions. These three components play particularly important roles: FDH catalyzes the initial step of both the WL and rGly pathways and CODH/ACS and GCS complexes are directly involved in CO_2 fixation in each pathway.

3.3 Phylogenetic analysis of FDHα

In the phylogenetic tree of cytoplasmic Fdh, the sequences from the phylum TDB did not form a monophyletic group (Fig. S3A). This result did not support the vertical transmission of cytoplasmic Fdh gene from the common ancestor of the phylum TDB. In contrast, FdhG from the phylum TDB formed a monophyletic group (Fig. 3, pink diamond). The branching pattern in this monophyletic group was not fully consistent with the phylogenetic relationships of the host species. For instance, sequences from clade A split into two clades. This phylogenetic incongruence can be explained by gene duplications and subsequent gene losses within the phylum TDB. The presence of sequences derives from the genus *Thermodesulfovibrio*, which belongs to the phylum Nitrospirota, is most probably due to the horizontal gene transfer (HGT) from the phylum TDB to the genus *Thermodesulfovibrio*. Therefore, the phylogenetic tree of FdhG supported the presence of FdhG gene in the common ancestor of the phylum TDB.

Within the monophyletic group of FdhG in TDB shown in pink diamond in Fig 3, only one sequence was included from each clade B species, suggesting that FdhG has been lost or undergone substantial genetic divergence in clade B. Two FdhG candidates in clade B organisms (*Desulfurivibrio alkaliphilus* and *Thermosulfuriphilus ammonigenes*) were sporadically distributed outside this monophyletic clade, suggesting the secondarily acquisition through HGT. Furthermore, several FdhG

sequences from the clade B together with one from the clade E formed a long-branching clade (Fig. 3, blue diamond) whose evolutionary origin remains unclear.

3.4 Phylogenetic analysis of the CODH/ACS complex

In the phylogenetic tree of AcsA, which encodes CODH, the sequences in TDB clustered into two distinct clades supported by bootstrap values (BPs) over 95 each (Fig. 4, highlighted in gray boxes). These two clades were provisionally designated as group 1 and group 2. Group 1 contained AcsA in TDB clades B, C, D, and E, including the one in *T. indicus* possessing the functional WL pathway. In contrast, group 2 contained sequences from clades A, B, C, and E. Eight TDB species encoded two or more AcsA genes within their genomes and each of them were categorized into group 1 and group 2.

The genomic context analysis of group 1 AcsA revealed that 19 out of 25 TDB-derived AcsA including the one in *T. indicus* formed a cluster with other CODH/ACS complex subunit genes (Fig. 4, highlighted with*), supporting its function as a CODH/ACS complex. The other six AcsA without forming the cluster were grouped into one clade (Fig. 4, black diamond). The species with these non-clustered AcsA also retained the previously mentioned AcsA gene that forms the CODH/ACS cluster. Therefore, TDB species with group 1 AcsA may have functional CODH which works in the WL pathway. In contrast, all the 16 AcsA sequences in group 2 derived from the phylum TDB did not form the cluster, suggesting that group 2 AcsA is not involved in acetyl-CoA synthesis (Techtmann et al., 2012). This assumption is supported by group 2 AcsA in *D. desulfuricans*, which has only one AcsA and operates the rGly pathway rather than the WL pathway for CO_2 fixation.

AcsB subunit encoding ACS which catalyzes acetyl-CoA synthesis with CO, a methyl group of CH_3-THF, and CoA. All the analyzed AcsB in the phylum TDB, which distributed in all the clades A to E, was clustered into a single clade with a BP support value of 73 (Fig. S4A). Although the phylogeny of AcsB in the phylum TDB was not fully consistent with the phylogeny of the host species, sequences in the clades B and C species formed distinct monophyletic groups supported by a BP of 100. Sequences from Chloroflexota (*Dehalogenimonas formicexedens*) and Planctomycetota (*Kuenenia* sp. and *Scalindua japonica*) were most probably the results of HGT from the phylum TDB to these organisms outside the phylum TDB. Therefore, the phylogenetic analysis supported the hypothesis that the common ancestor of the phylum TDB already possessed the AcsB gene. Phylogenetic analyses also indicate that the AcsC and AcsD genes, which encode subunits involved in methyltransferase activity, were likely present in the common ancestor of the phylum TDB. This is because AcsC and AcsD in the phylum TDB formed monophyletic clades with BPs over 90, respectively (Figs. S4C, D). Similar to AcsB, the AcsC and AcsD clades also contained sequences from the Chloroflexota (*Dehalococcoides* sp.) suggesting that gene cluster-level HGT of the CODH/ACS complex may have occurred from the phylum TDB to the Chloroflexota.

AcsE in the phylum TDB were dispersed across the phylogenetic tree and did not form a monophyletic group (Fig. S4D). Therefore, we could not estimate the origin of AcsE in TDB, which encodes a bacterial-specific subunit involved in THF-corrinoid methyltransferase activity, from the phylogenetic analysis alone.

3.5 Phylogenetic analysis of GCS complex

GCSPα in the phylum TDB were divided into two clades provisionally designated as group 1 and 2, supported by a BP of 100 and 94, respectively (Fig. 5A). Group 1 GCSPα was comprised of sequences from clades B, C, D, and E and formed a sister group with GCSPα from the phylum Myxococcota, the closest relative to the phylum TDB. In contrast, group 2 GCSPα included sequences from the species belonging to clade A and the species in the order Geobacterales in clade E. This group 2 also contained sequences from outside the phylum TDB, specifically from Epsilonproteobacteria and the bacterial "PVC group" (Coleman et al., 2021)(Fig. S2). No species possessed both group 1 and group 2 GCSPα genes.

Similar to GCSPα, the phylogenetic analyses of GCSPβ and GCST clearly divided the TDB sequences into two groups. Specifically, group 1 consisted of sequences from the TDB clades B, C, D, and E excluding the order Geobacterales (Fig. 5B, Fig. S8C), while group 2 consisted of sequences from clade A and the order Geobacterales in clade E, along with sequences from Epsilonproteobacteria and the PVC group.

GCSPα, GCSPβ and GCST categorized into group 1 based on the phylogenetic trees suggested to be vertically inherited from the common ancestor of the phylum TDB because the sequences from clades B to E formed a monophyletic group although the sequences in the close relatives in the phylum Myxococcota. In contrast, group 2 GCSPα, GCSPβ, and GCST were nested within the PVC group lineage, suggesting that the genes encoding these three subunits were most likely horizontally transferred from the PVC group to clade A and the order Geobacterales.

In contrast to the case of GCSPα, GCSPβ, and GCST, GCSL and GCSH in the phylum TDB were widely scattered across the phylogenetic trees and did not form monophyletic groups. Therefore, we could not estimate the origin of these proteins.

4 Discussion

In this study, we investigated the origin and evolutionary history of the WL and rGly pathways in the phylum TDB by analyzing gene presence/absence and phylogeny of the corresponding enzymes. As a result, we succeeded in inferring the gene sets of the common ancestor of the TDB phylum, as well as the subsequent processes of gene gain and loss as described below.

The common ancestor of the phylum TDB is suggested to have possessed the genes for CODH/ACS complex. Four out of five CODH/ACS complex subunits, AcsA, B, C, and D, are inferred to have existed in the common ancestor of the TDB phylum based on gene presence/absence and phylogenetic analyses. Although our phylogenetic analysis did not reveal the origin of AcsE, this protein is also expected to have existed in the common ancestor of the phylum TDB. This is because AcsE gene forms a cluster with other CODH/ACS complex genes and is therfore most likely vertically inherited as a cluster.

The common ancestor of the phylum TDB may also have possessed the genes for GCS complex. Both phylogenetic and gene presence/absence analyses supported the vertical inheritance of GCSPα, GCSPβ, and GCST to clade B, C, D and E from the common ancestor of TDB. In contrast, phylogenetic analyses did not clarify the origin of GCSL and GCSH in the phylum TDB although they are widely distributed in all the clades (Supplementary Table S4-D). This could be explained by

rapid sequence evolution caused by low functional constraints or replacement by functionally equivalent proteins because GCSH is a non-catalytic carrier protein and GCSL is an FAD dependent oxidoreductase that re-oxidizes the H-protein (Kikuchi et al., 2008; Kikuchi, 1973). Therefore, we estimated that the common ancestor of TDB may also have possessed the genes for GCSL and GCSH.

The common ancestor of the phylum TDB is also expected to have FdhG, a catalytic unit of FDH, indicating that it may have all the key enzyme genes for the WL and rGly pathways. Furthermore, presence/absence analysis also detected all the genes other than those for the key enzymes for the WL and rGly pathways in the phylum TDB (Supplementary Table S4-D). Namely, enzymes involved in the methyl branch which are shared between the WL and rGly pathways and an enzyme catalyzing the subsequent reaction in the WL pathway (Fig. S1) were identified in all the clades. These findings support the possibility that the common ancestor of the phylum TDB possessed the gene set for the WL and rGly pathways.

Based on the phylogeny of enzymes, the following evolutionary scenario is proposed in the phylum TDB (Fig. 6). The gene set for CODH/ACS complex was lost in the common ancestor of clade A and in clade E after the ancestor of *D. camini* was separated from others (Fig. 6, Blue). GCS complex genes were lost in the common ancestor of clades A-B and in the order Geobacterales within clade E (Fig. 6, Yellow). The CODH/ACS complex and/or GCS complex have also been lost in some species in clades B to E multiple times. Moreover, clade A and the order Geobacterales regained group 2 GCSPα, Pβ, and T from PVC group via HGT (Fig. 6, Pink). These findings indicate that individual lineages have formed unique gene repertoires through gene losses and acquisitions via HGT.

Although the common ancestor of TDB is expected to have possessed all the enzyme genes for the WL and rGly pathways, it remains unsolved whether these genes have worked for CO_2 fixation. To estimate their function for autotrophic growth, we analyzed downstream of the carbon fixation pathways. If the common ancestor of the TDB phylum grew autotrophically via the WL pathway, it would have pyruvate ferredoxin oxidoreductase, phosphoserine phosphatase, and serine hydroxymethyltransferase (Fig. 1), because these enzymes are essential to synthesize serine and glycine from acetyl-CoA through pyruvate. As a result, the three enzymatic genes were detected in all clades of the phylum TDB (Table S4-D). Although phylogenetic analyses of these three enzymes did not support their vertical inheritance from the common ancestor of the phylum TDB (Fig. S4-6) plausibly due to frequent HGT and gene duplication events, these were likely present in the common ancestor of TDB. Therefore, the common ancestor of TDB may have been capable of autotrophic growth via the WL pathway.

We also analyzed the five subunits of the glycine reductase complex (Fig. 1), which catalyzes the conversion of glycine to acetyl-CoA via acetyl phosphate and therefore reported to be essential for autotrophic growth via the rGly pathway (Sánchez-Andrea et al., 2020; Song et al., 2020). Neither gene presence/absence nor phylogenetic analysis supported the presence of the glycine reductase complex in the common ancestor. Still, there remains a possibility that glycine produced via the rGly pathway had been converted to serine by SHMT and then to pyruvate. The common ancestor of the phylum TDB might have grown autotrophically by operating both the WL and rGly pathways like *C. drakei* (Song et al., 2020). Therefore TDB might have been capable of autotrophic growth via the rGly pathway.

To date, no species within the phylum TDB has been reported to operate both the WL and rGly pathways. Our presence/absence analysis demonstrates that two species in clade C (*Desulfoluna*

butyratoxydans and *Desulfatirhabdium butyrativorans*) possess the key gene sets for both the WL and rGly pathways (Fig. 2), however, both species are known as heterotrophs (Suzuki et al., 2008; Balk et al., 2008). Further studies are needed to address the possibility that they grow autotrophically under unknown conditions and/or these metabolic pathways function as auxiliary pathways for organic compound acquisition or as sink for reducing power (Sánchez-Andrea et al., 2020; Schneeberger, 1999).

LBCA is estimated to have a gene set for the rGly pathway except for FDH (Coleman et al., 2021). However, their study used only cytoplasmic Fdh (K05299, K15022) as the query for FDH based on the information of KEGG pathway map, in which only cytoplasmic Fdh is registered (Aug 2025) . Nonetheless, FDH has multiple isomers other than cytoplasmic Fdh, including FdhG that converts CO_2 to formate in the rGly pathway of *D. desulfuricans* (Wells et al., 2023). Therefore, it remains possible that the LBCA did possess FDH that is not classified as cytoplasmic Fdh, like the case of the common ancestor of the phylum TDB which may have possessed FdhG.

In this study, we focused on a specific phylum and estimated the metabolisms of its last common ancestor. This phylum-specific studies ensures appropriate taxon sampling and enables more accurate inference of ancestral traits than the studies focused on deeper lineage, such as LBCA (Powell and Battistuzzi, 2022; Superson et al., 2019). Accumulation of knowledges based on such clade-specific studies across various phyla will lead to a more precise inference of metabolism in ancestral organisms such as LBCA and LUCA.

5 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.

6 Author Contributions

TW, KK and YC conceived and designed the work and wrote the manuscript. TW performed the presence/absence analyses. TW and KK inferred the phylogenetic trees. TW, KK and YC conducted phylogenetic and evolutionary analyses. All authors reviewed and approved the final version of the manuscript.

7 Funding

This work was supported in part by JSPS KAKENHI Grant Number 23K16986 (KK), 19K15745 and 23H04654 (YC). This work was supported by RIKEN Junior Research Associate Program.

8 Acknowledgments

We appreciate the fruitful discussions with Dr. Tetsuo Hashimoto regarding the phylogenetic analyses and evolutionary interpretations. We also thank Mr. Kohei Bamba for his assistance with the single-gene phylogenetic analyses. We are grateful to Dr. Ryuhei Nakamura, Dr. Eric Smith and Dr. Kimiho Omae for their helpful discussions.

9 References

Adam, P. S., Borrel, G., and Gribaldo, S. (2018). Evolutionary history of carbon monoxide dehydrogenase/acetyl-CoA synthase, one of the oldest enzymatic complexes. *Proc. Natl. Acad. Sci. U.S.A.* 115, E1166-E1173. doi: 10.1073/pnas.1716667115

Alain, K., Postec, A., Grinsard, E., Lesongeur, F., Prieur, D., and Godfroy, A. (2010). Thermodesulfatator atlanticus sp. nov., a thermophilic, chemolithoautotrophic, sulfate-reducing bacterium isolated from a Mid-Atlantic Ridge hydrothermal vent. *Int. J. Syst. Evol. Microbiol.* 60, 33-38. doi: 10.1099/ijs.0.009449-0

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403-410. doi: 10.1016/S0022-2836(05)80360-2

Bada, J. L., and Lazcano, A. (2002). Some like it hot, but not the first biomolecules. *Science* 296, 1982-1983. doi: 10.1126/science.1069487

Balk, M., Altinbaş, M., Rijpstra, W. I. C., Damsté, J. S. S., and Stams, A. J. M. (2008). Desulfatirhabdium butyrativorans gen. nov., sp. nov., a butyrate-oxidizing, sulfate-reducing bacterium isolated from an anaerobic bioreactor. *Int. J. Syst. Evol. Microbiol.* 58, 110-115. doi: 10.1099/ijs.0.65396-0

Berg, I. A., Kockelkorn, D., Ramos-Vera, W. H., Say, R. F., Zarzycki, J., Hügler, M., et al. (2010). Autotrophic carbon fixation in archaea. *Nat. Rev. Microbiol.* 8, 447-460. doi: 10.1038/nrmicro2365

Bergsten, J. (2005). A review of long-branch attraction. *Cladistics* 21, 163-193. doi: 10.1111/j.1096-0031.2005.00059.x

Borrel, G., Adam, P. S., and Gribaldo, S. (2016). Methanogenesis and the Wood-Ljungdahl pathway: an ancient, versatile, and fragile association. *Genome Biol. Evol.* 8, 1706-1711. doi: 10.1093/gbe/evw114

Braakman, R., and Smith, E. (2012). The emergence and early evolution of biological carbon-fixation. *PLoS Comput. Biol.* 8:e1002455. doi: 10.1371/journal.pcbi.1002455

Cao, X., Hong, Y., Zhu, L., Hu, Y., and Cronan, J. E. (2018). Development and retention of a primordial moonlighting pathway of protein modification in the absence of selection presents a puzzle. *Proc. Natl. Acad. Sci. U.S.A.* 115, 647-655. doi: 10.1073/pnas.1718653115

Chiba, Y., Sumida, T., Kameya, M., Fukuyama, Y., Wakashima, T., Shimamura, S., et al. (2025). Impact of acetate on CO_2 fixation pathways in thermophilic and hydrogenotrophic bacteria. *bioRxiv*. doi: 10.1101/2025.03.31.646503

Coleman, G. A., Davín, A. A., Mahendrarajah, T. A., Szánthó, L. L., Spang, A., Hugenholtz, P., et al. (2021). A rooted phylogeny resolves early bacterial evolution. *Science* 372:abe0511. doi: 10.1126/science.abe0511

Cotton, C. A., Edlich-Muth, C., and Bar-Even, A. (2018). Reinforcing carbon fixation: CO_2 reduction replacing and supporting carboxylation. *Curr. Opin. Biotechnol.* 49, 49-56. doi: 10.1016/j.copbio.2017.07.014

Cravo-Laureau, C., Matheron, R., Cayol, J. L., Joulian, C., and Hirschler-Réa, A. (2004). Desulfatibacillum aliphaticivorans gen. nov., sp. nov., an n-alkane- and n-alkene-degrading, sulfate-reducing bacterium. *Int. J. Syst. Evol. Microbiol.* 54, 77-83. doi: 10.1099/ijs.0.02717-0

DeWeerd, K. A., Mandelco, L., Tanner, R. S., Woese, C. R., and Suflita, J. M. (1990). Desulfomonile tiedjei gen. nov. and sp. nov., a novel anaerobic, dehalogenating, sulfate-reducing bacterium. *Arch. Microbiol.* 154, 23-30. doi: 10.1007/BF00249173

Dick, G. J. (2019). The microbiomes of deep-sea hydrothermal vents: distributed globally, shaped locally. *Nat. Rev. Microbiol.* 17, 271-283. doi: 10.1038/s41579-019-0160-2

Dürre, P., and Andreesen, J. R. (1982). Pathway of carbon dioxide reduction to acetate without a net energy requirement in Clostridium purinolyticum. *FEMS Microbiol. Lett.* 15, 51-56. doi: 10.1111/j.1574-6968.1982.tb00036.x

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792-1797. doi: 10.1093/nar/gkh340

Finster, K., Liesack, W., and Thamdrup, B. (1998). Elemental sulfur and thiosulfate disproportionation by Desulfocapsa sulfoexigens sp. nov., a new anaerobic bacterium isolated from marine surface sediment. *Appl. Environ. Microbiol.* 64, 119-125. doi: 10.1128/AEM.64.1.119-125.1998

Frolova, A. A., Slobodkina, G. B., Baslerov, R. V., Novikov, A. A., Bonch-Osmolovskaya, E. A., and Slobodkin, A. I. (2018). Thermosulfurimonas marina sp. nov., an autotrophic sulfur-disproportionating and nitrate-reducing bacterium isolated from a shallow-sea hydrothermal vent. *Microbiology* 87, 502-507. doi: 10.1134/S0026261718040082

Fuchs, G. (1986). CO_2 fixation in acetogenic bacteria: variations on a theme. *FEMS Microbiol. Lett.* 39, 181-213. doi: 10.1111/j.1574-6968.1986.tb01859.x

Fuchs, G. (2011). Alternative pathways of carbon dioxide fixation: insights into the early evolution of life? *Annu. Rev. Microbiol.* 65, 631-658. doi: 10.1146/annurev-micro-090110-102801

Graybeal, A. (1998). Is it better to add taxa or characters to a difficult phylogenetic problem? *Syst. Biol.* 47, 9-17. doi: 10.1080/106351598260996

Hamilton-Brehm, S. D., Gibson, R. A., Green, S. J., Hopmans, E. C., Schouten, S., van der Meer, M. T. J., et al. (2013). Thermodesulfobacterium geofontis sp. nov., a hyperthermophilic, sulfate-reducing bacterium isolated from Obsidian Pool, Yellowstone National Park. *Extremophiles* 17, 251-263. doi: 10.1007/s00792-013-0512-1

Hedtke, S. M., Townsend, T. M., and Hillis, D. M. (2006). Resolution of phylogenetic conflict in large data sets by increased taxon sampling. *Syst. Biol.* 55, 522-529. doi: 10.1080/10635150600697358

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* 35, 518-522. doi: 10.1093/molbev/msx281

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236-1240. doi: 10.1093/bioinformatics/btu031

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587-589. doi: 10.1038/nmeth.4285

Kikuchi, G. (1973). The glycine cleavage system: composition, reaction mechanism, and physiological significance. *Mol. Cell. Biochem.* 1, 169-187. doi: 10.1007/BF01659328

Kikuchi, G., Motokawa, Y., Yoshida, T., and Hiraga, K. (2008). Glycine cleavage system: reaction mechanism, physiological significance, and hyperglycinemia. *Proc. Jpn. Acad. Ser. B* 84, 246-263. doi: 10.2183/pjab.84.246

Kim, K., Chiba, Y., Kobayashi, A., Arai, H., and Ishii, M. (2017). Phosphoserine phosphatase is required for serine and one-carbon unit synthesis in Hydrogenobacter thermophilus. *J. Bacteriol.* 199:e00409-17. doi: 10.1128/JB.00409-17

Kitadai, N., and Maruyama, S. (2018). Origins of building blocks of life: a review. *Geosci. Front.* 9, 1117-1153. doi: 10.1016/j.gsf.2017.07.007

Kitadai, N., Nakamura, R., Yamamoto, M., Okada, S., Takahagi, W., Nakano, Y., et al. (2021). Thioester synthesis through geoelectrochemical CO_2 fixation on Ni sulfides. *Commun. Chem.* 4:54. doi: 10.1038/s42004-021-00475-5

Kojima, H., Umezawa, K., and Fukui, M. (2016). Caldimicrobium thiodismutans sp. nov., a sulfur-disproportionating bacterium isolated from a hot spring, and emended description of the genus Caldimicrobium. *Int. J. Syst. Evol. Microbiol.* 66, 1828-1831. doi: 10.1099/ijsem.0.000947

Krukenberg, V., Harding, K., Richter, M., Glöckner, F. O., Gruber-Vodicka, H. R., Adam, B., et al. (2016). Candidatus Desulfofervidus auxilii, a hydrogenotrophic sulfate-reducing bacterium involved in the thermophilic anaerobic oxidation of methane. *Environ. Microbiol.* 18, 3073-3091. doi: 10.1111/1462-2920.13283

Lai, Q., Cao, J., Dupont, S., Shao, Z., Jebbar, M., and Alain, K. (2016). Thermodesulfatator autotrophicus sp. nov., a thermophilic sulfate-reducing bacterium from the Indian Ocean. *Int. J. Syst. Evol. Microbiol.* 66, 3978-3982. doi: 10.1099/ijsem.0.001297

Lorio, J. C., Kim, W. S., Krishnan, A. H., and Krishnan, H. B. (2010). Disruption of the glycine cleavage system enables Sinorhizobium fredii USDA257 to form nitrogen-fixing nodules on agronomically improved North American soybean cultivars. *Appl. Environ. Microbiol.* 76, 4185-4193. doi: 10.1128/AEM.00437-10

Maia, L. B., Fonseca, L., Moura, I., and Moura, J. J. G. (2016). Reduction of carbon dioxide by a molybdenum-containing formate dehydrogenase: a kinetic and mechanistic study. *J. Am. Chem. Soc.* 138, 8834-8846. doi: 10.1021/jacs.6b03941

Maia, L. B., Moura, J. J. G., and Moura, I. (2015). Molybdenum and tungsten-dependent formate dehydrogenases. *J. Biol. Inorg. Chem.* 20, 287-309. doi: 10.1007/s00775-014-1218-2

Mardanov, A. V., Beletsky, A. V., Kadnikov, V. V., Slobodkin, A. I., and Ravin, N. V. (2016). Genome analysis of Thermosulfurimonas dismutans, the first thermophilic sulfur-disproportionating bacterium of the phylum Thermodesulfobacteria. *Front. Microbiol.* 7:950. doi: 10.3389/fmicb.2016.00950

Martin, W., and Russell, M. J. (2007). On the origin of biochemistry at an alkaline hydrothermal vent. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 362, 1887-1926. doi: 10.1098/rstb.2006.1881

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., Lanfear, R., and Teeling, E. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530-1534. doi: 10.1093/molbev/msaa015

Moody, E. R. R., Álvarez-Carretero, S., Mahendrarajah, T. A., Clark, J. W., Betts, H. C., Dombrowski, N., et al. (2024). The nature of the last universal common ancestor and its impact on the early Earth system. *Nat. Ecol. Evol.* 8, 1654-1666. doi: 10.1038/s41559-024-02461-1

Moussard, H., L'Haridon, S., Tindall, B. J., Banta, A., Schumann, P., Stackebrandt, E., Reysenbach, A. L., and Jeanthon, C. (2004). Thermodesulfatator indicus gen. nov., sp. nov., a novel thermophilic chemolithoautotrophic sulfate-reducing bacterium isolated from the Central Indian Ridge. *Int. J. Syst. Evol. Microbiol.* 54, 227-233. doi: 10.1099/ijs.0.02669-0

Nitschke, W., and Russell, M. J. (2013). Beating the acetyl coenzyme A-pathway to the origin of life. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 368:20120258. doi: 10.1098/rstb.2012.0258

Nunoura, T., Chikaraishi, Y., Izaki, R., Suwa, T., Sato, T., Harada, T., et al. (2018). A primordial and reversible TCA cycle in a facultatively chemolithoautotrophic thermophile. *Science* 359, 559-563. doi: 10.1126/science.aao3407

Okamura-Ikeda, K., Ohmura, Y., Fujiwara, K., and Motokawa, Y. (1993). Cloning and nucleotide sequence of the gcv operon encoding the Escherichia coli glycine-cleavage system. *Eur. J. Biochem.* 216, 539-548. doi: 10.1111/j.1432-1033.1993.tb18172.x

Peretó, J. G., Velasco, A. M., Becerra, A., and Lazcano, A. (1999). Comparative biochemistry of CO_2 fixation and the evolution of autotrophy. *Int. Microbiol.* 2, 3-10. doi: 10.2436/im.v2i1.95886

Pierce, E., Xie, G., Barabote, R. D., Saunders, E., Han, C. S., Detter, J. C., et al. (2008). The complete genome sequence of Moorella thermoacetica (f. Clostridium thermoaceticum). *Environ. Microbiol.* 10, 2550-2573. doi: 10.1111/j.1462-2920.2008.01679.x

Pikuta, E. V., Hoover, R. B., Bej, A. K., Marsic, D., Whitman, W. B., Cleland, D., and Krader, P. (2003). Desulfonatronum thiodismutans sp. nov., a novel alkaliphilic, sulfate-reducing bacterium capable of lithoautotrophic growth. *Int. J. Syst. Evol. Microbiol.* 53, 1327-1332. doi: 10.1099/ijs.0.02598-0

Powell, C. L. E., and Battistuzzi, F. U. (2022). Testing phylogenetic stability with variable taxon sampling. *Methods Mol. Biol.* 2569, 167-188. doi: 10.1007/978-1-0716-2691-7_8

Sánchez-Andrea, I., Guedes, I. A., Hornung, B., Boeren, S., Lawson, C. E., Sousa, D. Z., Bar-Even, A., Claassens, N. J., and Stams, A. J. M. (2020). The reductive glycine pathway allows autotrophic growth of Desulfovibrio desulfuricans. *Nat. Commun.* 11:5090. doi: 10.1038/s41467-020-18906-7

Schneeberger, A. (1999). Net synthesis of acetate from CO_2 by Eubacterium acidaminophilum through the glycine reductase pathway. *FEMS Microbiol. Lett.* 177, 1-6. doi: 10.1016/S0378-1097(99)00248-7

Schuchmann, K., and Müller, V. (2014). Autotrophy at the thermodynamic limit of life: a model for energy conservation in acetogenic bacteria. *Nat. Rev. Microbiol.* 12, 809-821. doi: 10.1038/nrmicro3365

Slobodkin, A. I., Reysenbach, A. L., Slobodkina, G. B., Kolganova, T. V., Kostrikina, N. A., and Bonch-Osmolovskaya, E. A. (2013). Dissulfuribacter thermophilus gen. nov., sp. nov., a thermophilic, autotrophic, sulfur-disproportionating, deeply branching deltaproteobacterium from a deep-sea hydrothermal vent. *Int. J. Syst. Evol. Microbiol.* 63, 1967-1971. doi: 10.1099/ijs.0.046938-0

Slobodkina, G. B., Reysenbach, A. L., Kolganova, T. V., Novikov, A. A., Bonch-Osmolovskaya, E. A., and Slobodkin, A. I. (2017). Thermosulfuriphilus ammonigenes gen. nov., sp. nov., a thermophilic, chemolithoautotrophic bacterium capable of respiratory ammonification of nitrate with elemental sulfur. *Int. J. Syst. Evol. Microbiol.* 67, 3474-3479. doi: 10.1099/ijsem.0.002142

Song, Y., Lee, J. S., Shin, J., Lee, M., Jin, S., Kang, S., et al. (2020). Functional cooperation of the glycine synthase-reductase and Wood-Ljungdahl pathways for autotrophic growth of Clostridium drakei. *Proc. Natl. Acad. Sci. U.S.A.* 117, 7516-7523. doi: 10.1073/pnas.1912289117

Steinegger, M., and Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* 35, 1026-1028. doi: 10.1038/nbt.3988

Superson, A. A., Phelan, D., Dekovich, A., and Battistuzzi, F. U. (2019). Choice of species affects phylogenetic stability of deep nodes: an empirical example in Terrabacteria. *Bioinformatics* 35, 3608-3616. doi: 10.1093/bioinformatics/btz121

Suzuki, D., Ueki, A., Amaishi, A., and Ueki, K. (2008). Desulfoluna butyratoxydans gen. nov., sp. nov., a novel Gram-negative, butyrate-oxidizing, sulfate-reducing bacterium isolated from an estuarine sediment in Japan. *Int. J. Syst. Evol. Microbiol.* 58, 826-832. doi: 10.1099/ijs.0.65306-0

Talavera, G., and Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* 56, 564-577. doi: 10.1080/10635150701472164

Techtmann, S. M., Lebedinsky, A. V., Colman, A. S., Sokolova, T. G., Woyke, T., Goodwin, L., and Robb, F. T. (2012). Evidence for horizontal gene transfer of anaerobic carbon monoxide dehydrogenases. *Front. Microbiol.* 3:132. doi: 10.3389/fmicb.2012.00132

Tezuka, T., and Ohnishi, Y. (2014). Two glycine riboswitches activate the glycine cleavage system essential for glycine detoxification in Streptomyces griseus. *J. Bacteriol.* 196, 1369-1376. doi: 10.1128/JB.01480-13

Waber, L. J., and Wood, H. G. (1979). Mechanism of acetate synthesis from CO_2 by Clostridium acidiurici. *J. Bacteriol.* 140, 468-478. doi: 10.1128/jb.140.2.468-478.1979

Wächtershäuser, G. (1988). Before enzymes and templates: theory of surface metabolism. *Microbiol. Rev.* 52, 452-484. doi: 10.1128/mr.52.4.452-484.1988

Waite, D. W., Chuvochina, M., Pelikan, C., Parks, D. H., Yilmaz, P., Wagner, M., et al. (2020). Proposal to reclassify the proteobacterial classes Deltaproteobacteria and Oligoflexia, and the phylum Thermodesulfobacteria into four phyla reflecting major functional capabilities. *Int. J. Syst. Evol. Microbiol.* 70, 5972-6016. doi: 10.1099/ijsem.0.004213

Weiss, M. C., Sousa, F. L., Mrnjavac, N., Neukirchen, S., Roettger, M., Nelson-Sathi, S., and Martin, W. F. (2016). The physiology and habitat of the last universal common ancestor. *Nat. Microbiol.* 1:16116. doi: 10.1038/nmicrobiol.2016.116

Wells, M., Kim, M., Akob, D. M., Basu, P., and Stolz, J. F. (2023). Impact of the dimethyl sulfoxide reductase superfamily on the evolution of biogeochemical cycles. *Microbiol. Spectr.* 11:e04145-22. doi: 10.1128/spectrum.04145-22

Yamamoto, I., Saiki, T., Liu, S. M., and Ljungdahl, L. G. (1983). Purification and properties of NADP-dependent formate dehydrogenase from Clostridium thermoaceticum, a tungsten-selenium-iron protein. *J. Biol. Chem.* 258, 1826-1832. doi: 10.1016/S0021-9258(18)33062-X

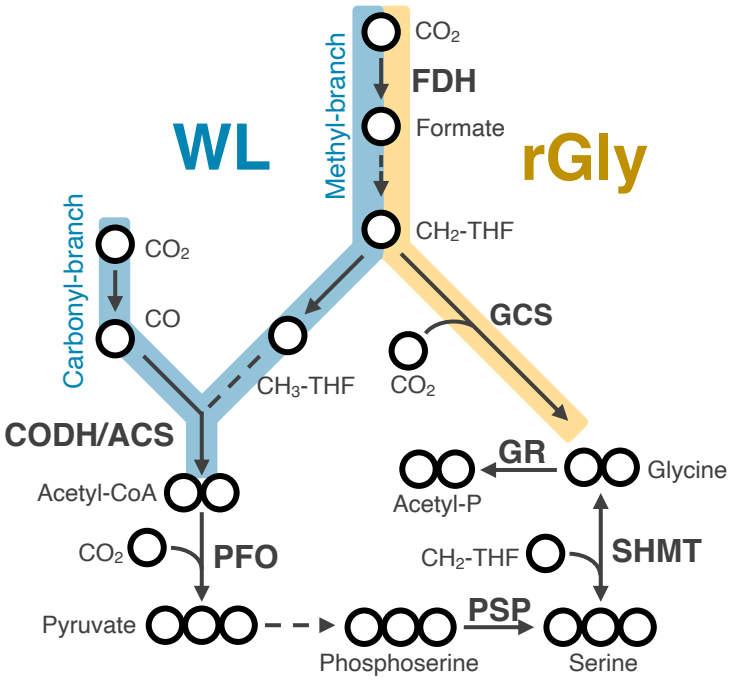


Fig. 1
Metabolic map of the Wood–Ljungdahl (WL) and reductive glycine (rGly) pathways.
Blue and yellow arrows indicate reactions in the WL and rGly pathways, respectively. Arrows represent enzymatic reactions and dashed arrows indicate reactions catalyzed by multiple enzymes. Bold letters show enzyme names. Circles represent number of carbons. THF: tetrahydrofolate; FDH: formate dehydrogenase; CODH/ACS: carbon monoxide dehydrogenase/acetyl-CoA synthase; GCS: glycine cleavage system; PFO: pyruvate-ferredoxin oxidoreductase; SHMT: serine hydroxymethyltransferase; PSP: phosphoserine phosphatase; GR: glycine reductase.

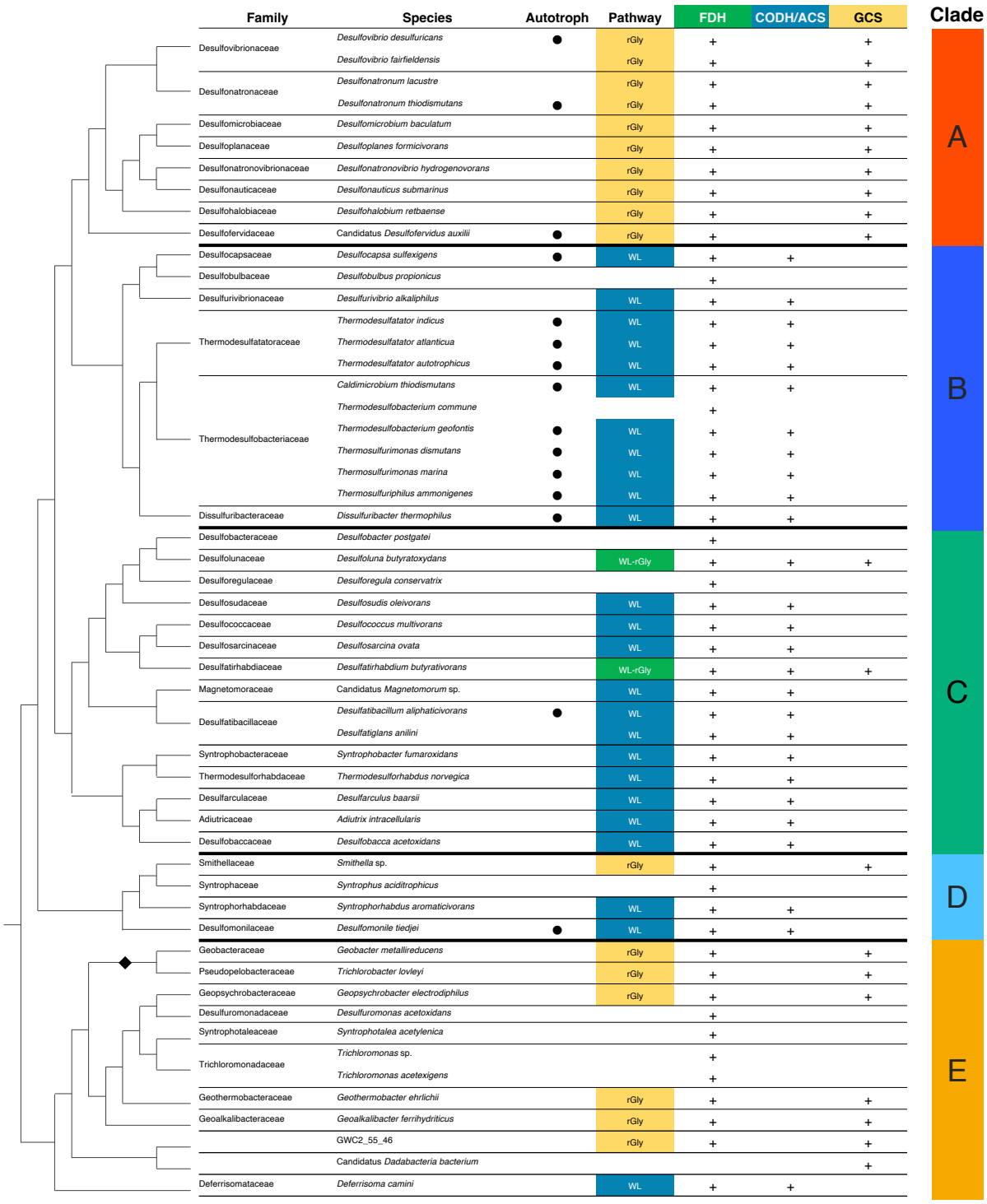


Fig. 2

Presence or absence of enzymes for the WL and rGly pathways in Thermodesulfobacteriota species.

The tree on the left shows a phylogenetic tree of Thermodesulfobacteriota species, modified from Waite et al.. The order Geobacterales in clade E is indicated by a black diamond (◆). Autotrophic species are indicated by black circles (●). The presence or absence of each enzyme is based on predicted proteomes. FDH is considered to be present (+) when a species encodes either the homolog of cytoplasmic Fdh or FdhG. CODH/ACS is considered to be present if all five subunits of the CODH/ACS complex are encoded. GCS is considered present when all five proteins of the glycine cleavage system are encoded. Pathway: When CODH/ACS and GCS are present together with FDH, the species is estimated to have the WL and rGly pathways, respectively. Clade: the phylum Thermodesulfobacteriota were divided into five monophyletic clades in this study.

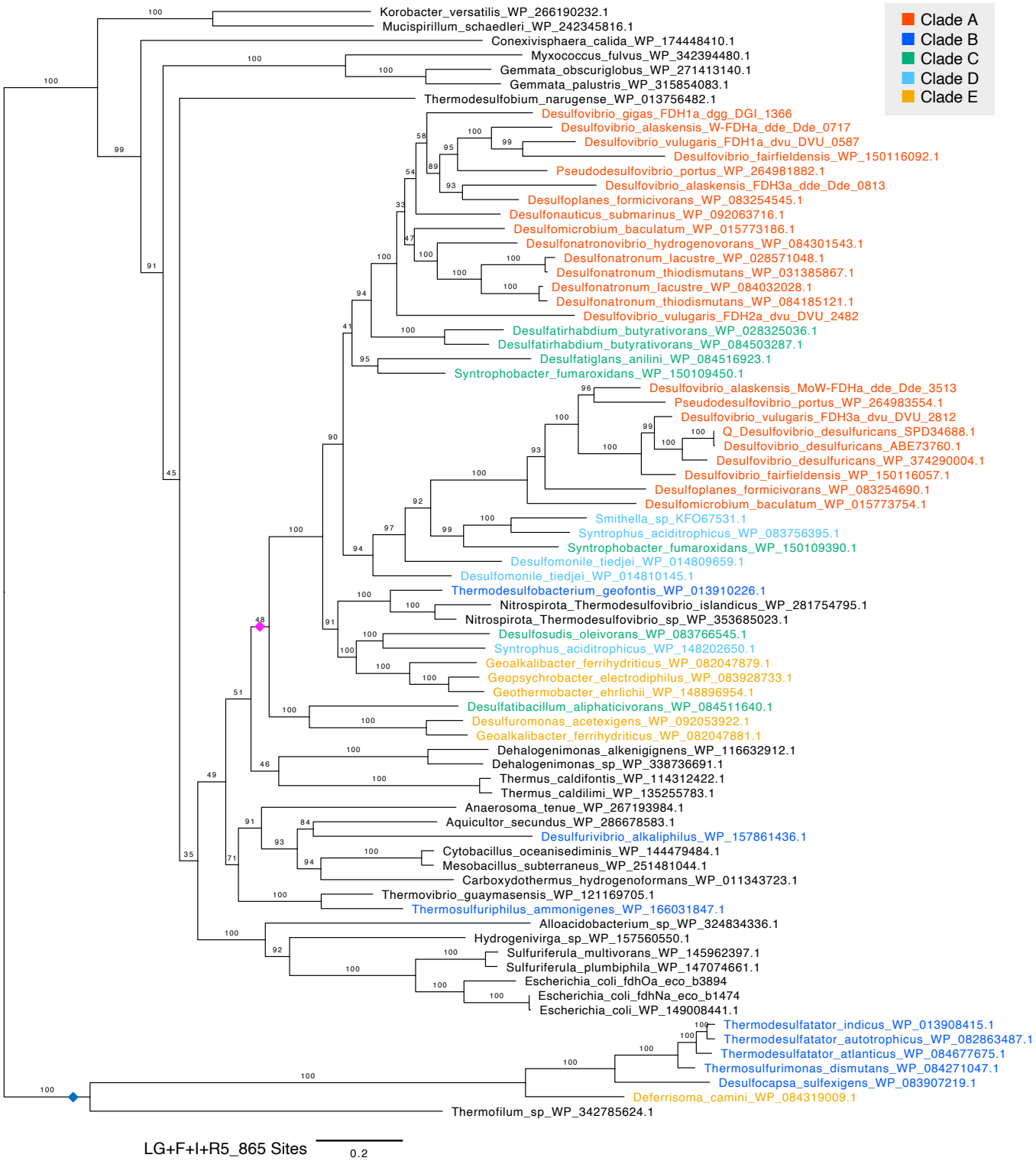


Fig. 3

Maximum likelihood phylogenetic tree of FdhG.

Bootstrap values were calculated using the ultrafast bootstrap method. The evolutionary model, scale bar and number of sites are shown in the figure. Sequence in the phylum TDB are highlighted with colors. A pink diamond indicates a clade composed of sequences from TDB species. A blue diamond indicates clade B and E sequences that were positioned outside the TDB largest clade.

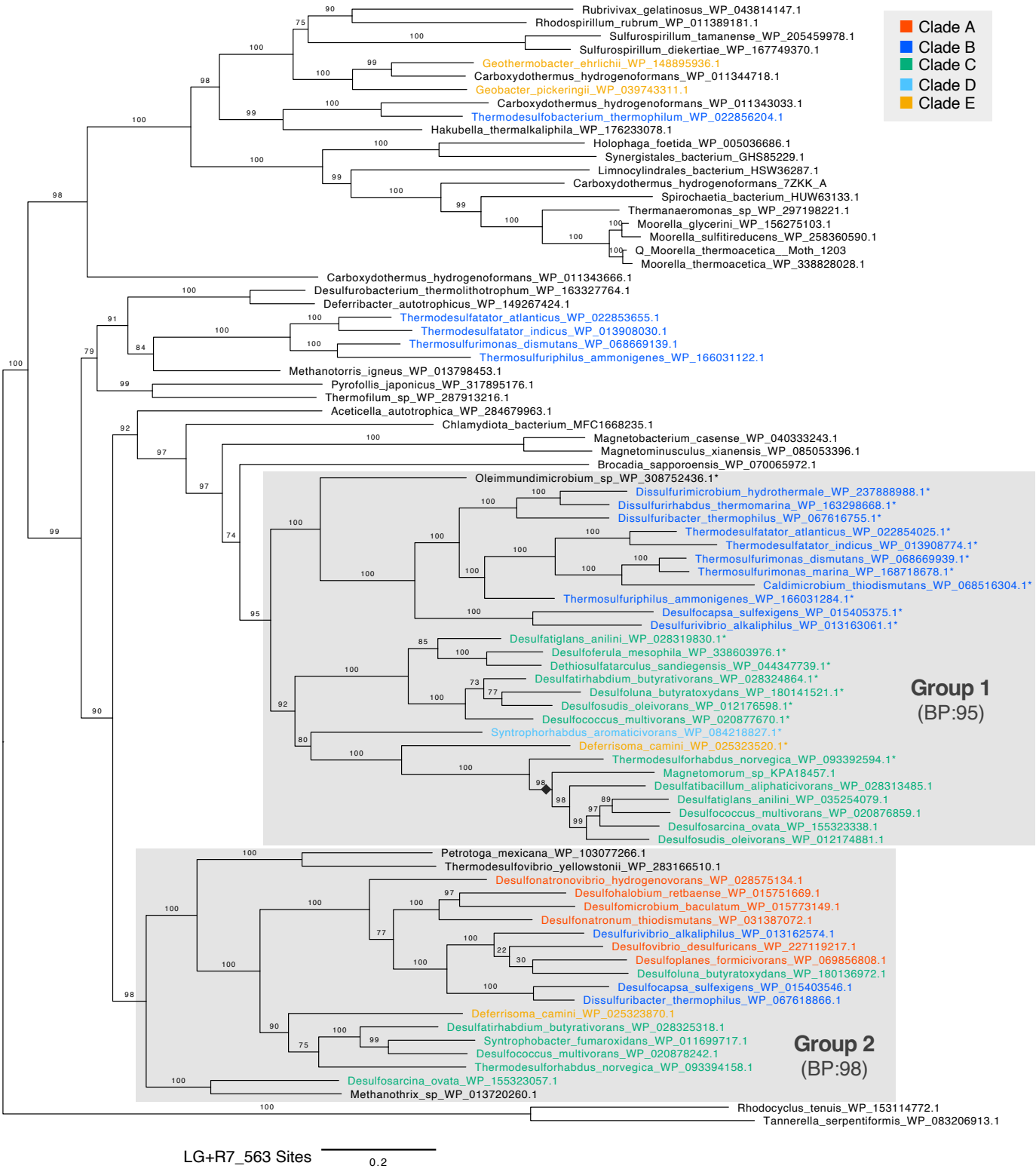
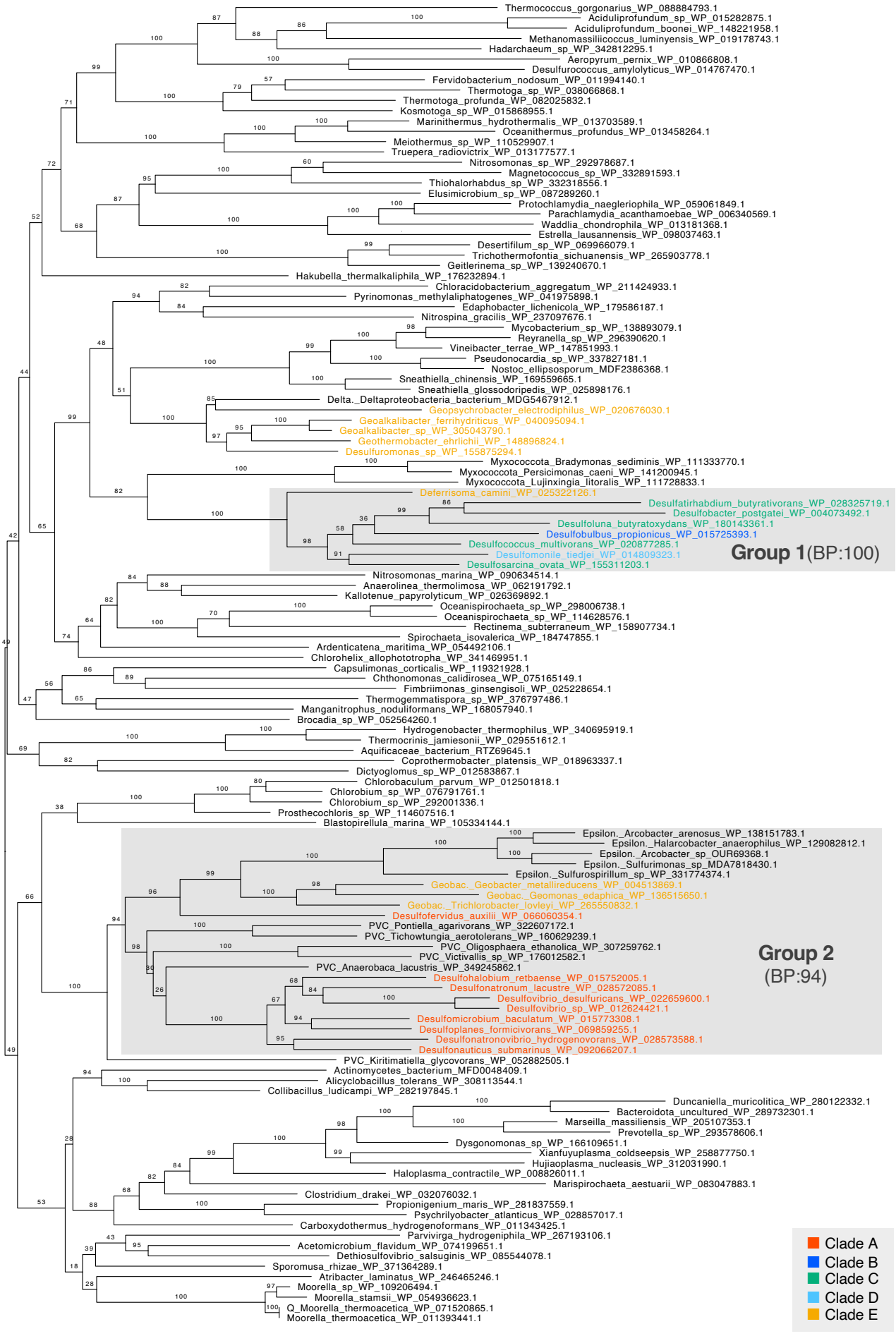


Fig. 4

Maximum likelihood phylogenetic tree of AcsA (CODH).

Bootstrap values were calculated using the ultrafast bootstrap method. The evolutionary model, scale bar and number of sites are shown in the figure. Sequence in the phylum TDB are highlighted with colors. Asterisks (*) indicate sequences encoded in a cluster with other CODH/ACS complex genes. A black diamond (◆) indicates a clade within group 1 composed of sequences that do not form a gene cluster.

A


Thermococcus_gorgonarius_WP_088884793.1
Aciduliprofundum_sp_WP_015282875.1
Aciduliprofundum_boonei_WP_148221958.1
Methanomassiliicoccus_luminyensis_WP_019178743.1
Hadarchaeum_sp_WP_342812295.1
Aeropyrum_pernix_WP_010866808.1
Desulfurococcus_amylolyticus_WP_014767470.1
Fervidobacterium_nodosum_WP_011994140.1
Thermotoga_sp_WP_038066868.1
Thermotoga_profunda_WP_082025832.1
Kosmotoga_sp_WP_015868955.1
Marinithermus_hydrothermalis_WP_013703589.1
Oceanithermus_profundus_WP_013458264.1
Meiothermus_sp_WP_110529907.1
Truepera_radiovictrix_WP_013177577.1
Nitrosomonas_sp_WP_292978687.1
Magnetococcus_sp_WP_332891593.1
Thiohalorhabdus_sp_WP_332318556.1
Elusimicrobium_sp_WP_087289260.1
Protochlamydia_naegleriophila_WP_059061849.1
Parachlamydia_acanthamoebae_WP_006340569.1
Waddlia_chondrophila_WP_013181368.1
Estrella_lausannensis_WP_098037463.1
Desertifilum_sp_WP_069966079.1
Trichothermofontia_sichuanensis_WP_265903778.1
Geitlerinema_sp_WP_139240670.1
Hakubella_thermalkaliphila_WP_176232894.1
Chloracidobacterium_aggregatum_WP_211424933.1
Pyrinomonas_methylaliphatogenes_WP_041975898.1
Edaphobacter_lichenicola_WP_179586187.1
Nitrospina_gracilis_WP_237097676.1
Mycobacterium_sp_WP_138893079.1
Reyranella_sp_WP_296390620.1
Vineibacter_terrae_WP_147851993.1
Pseudonocardia_sp_WP_337827181.1
Nostoc_ellipsosporum_MDF2386368.1
Sneathiella_chinensis_WP_169559665.1
Sneathiella_glossodoripedis_WP_025898176.1
Delta._Deltaproteobacteria_bacterium_MDG5467912.1
Geopsychrobacter_electrodiphilus_WP_020676030.1
Geoalkalibacter_ferrihydriticus_WP_040095094.1
Geoalkalibacter_sp_WP_305043790.1
Geothermobacter_ehrlichii_WP_148896824.1
Desulfuromonas_sp_WP_155875294.1
Myxococcota_Bradymonas_sediminis_WP_111333770.1
Myxococcota_Persicimonas_caeni_WP_141200945.1
Myxococcota_Lujinxingia_litoralis_WP_111728833.1
Deferrisoma_camini_WP_025322126.1
Desulfatirhabdium_butyrativorans_WP_028325719.1
Desulfobacter_postgatei_WP_004073492.1
Desulfoluna_butyratoxydans_WP_180143361.1
Desulfobulbus_propionicus_WP_015725393.1
Desulfococcus_multivorans_WP_020877285.1
Desulfomonile_tiedjei_WP_014809323.1
Desulfosarcina_ovata_WP_155311203.1
Group 1(BP:100)
Nitrosomonas_marina_WP_090634514.1
Anaerolinea_thermolimosa_WP_062191792.1
Kallotenue_papyrolyticum_WP_026369892.1
Oceanispirochaeta_sp_WP_298006738.1
Oceanispirochaeta_sp_WP_114628576.1
Rectinema_subterraneum_WP_158907734.1
Spirochaeta_isovalerica_WP_184747855.1
Ardenticatena_maritima_WP_054492106.1
Chlorohelix_allophototropha_WP_341469951.1
Capsulimonas_corticalis_WP_119321928.1
Chthonomonas_calidirosea_WP_075165149.1
Fimbriimonas_ginsengisoli_WP_025228654.1
Thermogemmatispora_sp_WP_376797486.1
Manganitrophus_noduliformans_WP_168057940.1
Brocadia_sp_WP_052564260.1
Hydrogenobacter_thermophilus_WP_340695919.1
Thermocrinis_jamiesonii_WP_029551612.1
Aquificaceae_bacterium_RTZ69645.1
Coprothermobacter_platensis_WP_018963337.1
Dictyoglomus_sp_WP_012583867.1
Chlorobaculum_parvum_WP_012501818.1
Chlorobium_sp_WP_076791761.1
Chlorobium_sp_WP_292001336.1
Prosthecochloris_sp_WP_114607516.1
Blastopirellula_marina_WP_105334144.1
Epsilon._Arcobacter_arenosus_WP_138151783.1
Epsilon._Halarcobacter_anaerophilus_WP_129082812.1
Epsilon._Arcobacter_sp_OUR69368.1
Epsilon._Sulfurimonas_sp_MDA7818430.1
Epsilon._Sulfurospirillum_sp_WP_331774374.1
Geobac._Geobacter_metallireducens_WP_004513869.1
Geobac._Geomonas_edaphica_WP_136515650.1
Geobac._Trichlorobacter_lovleyi_WP_265550832.1
Desulfofervidus_auxilii_WP_066060354.1
PVC_Pontiella_agarivorans_WP_322607172.1
PVC_Tichowtungia_aerotolerans_WP_160629239.1
PVC_Oligosphaera_ethanolica_WP_307259762.1
PVC_Victivallis_sp_WP_176012582.1
PVC_Anaerobaca_lacustris_WP_349245862.1
Desulfohalobium_retbaense_WP_015752005.1
Desulfonatronum_lacustre_WP_028572085.1
Desulfovibrio_desulfuricans_WP_022659600.1
Desulfovibrio_sp_WP_012624421.1
Desulfomicrobium_baculatum_WP_015773308.1
Desulfoplanes_formicivorans_WP_069859255.1
Desulfonatronovibrio_hydrogenovorans_WP_028573588.1
Desulfonauticus_submarinus_WP_092066207.1
Group 2
(BP:94)
PVC_Kiritimatiella_glycovorans_WP_052882505.1
Actinomycetes_bacterium_MFD0048409.1
Alicyclobacillus_tolerans_WP_308113544.1
Collibacillus_ludicampi_WP_282197845.1
Duncaniella_muricolitica_WP_280122332.1
Bacteroidota_uncultured_WP_289732301.1
Marseilla_massiliensis_WP_205107353.1
Prevotella_sp_WP_293578606.1
Dysgonomonas_sp_WP_166109651.1
Xianfuyuplasma_coldseepsis_WP_258877750.1
Hujiaoplasma_nucleasis_WP_312031990.1
Haloplasma_contractile_WP_008826011.1
Marispirochaeta_aestuarii_WP_083047883.1
Clostridium_drakei_WP_032076032.1
Propionigenium_maris_WP_281837559.1
Psychrilyobacter_atlanticus_WP_028857017.1
Carboxydothermus_hydrogenoformans_WP_011343425.1
Parvivirga_hydrogeniphila_WP_267193106.1
Acetomicrobium_flavidum_WP_074199651.1
Dethiosulfovibrio_salsuginis_WP_085544078.1
Sporomusa_rhizae_WP_371364289.1
Atribacter_laminatus_WP_246465246.1
Moorella_sp_WP_109206494.1
Moorella_stamsii_WP_054936623.1
Q_Moorella_thermoacetica_WP_071520865.1
Moorella_thermoacetica_WP_011393441.1
Clade A
Clade B
Clade C
Clade D
Clade E
Q.pfam+I+R7_344 Sites
0.2

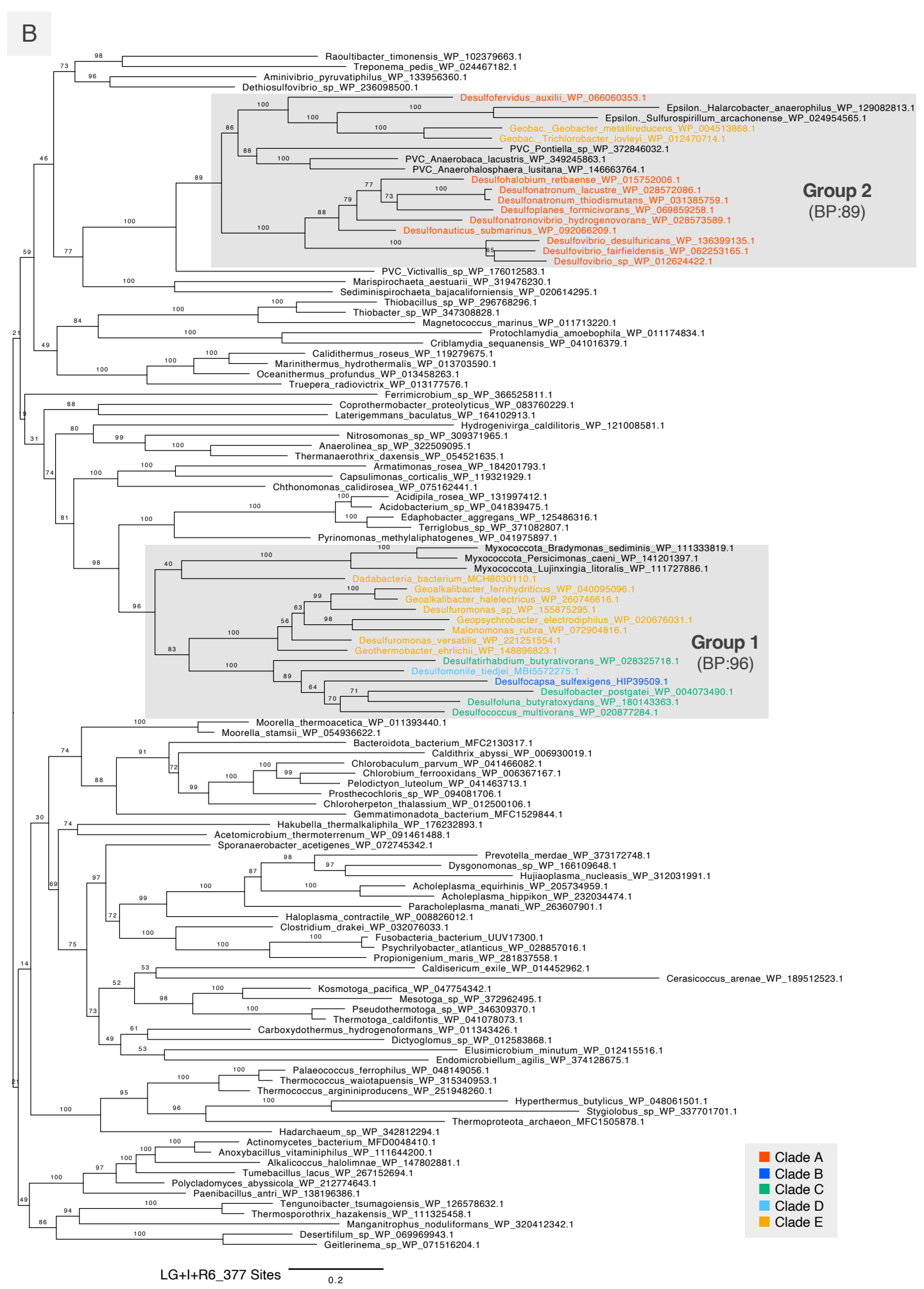

B
Raoultibacter_timonensis_WP_102379663.1
Treponema_pedis_WP_024467182.1
Aminivibrio_pyruvatiphilus_WP_133956360.1
Dethiosulfovibrio_sp_WP_236098500.1
Desulfofervidus_auxilii_WP_066060353.1
Epsilon._Halarcobacter_anaerophilus_WP_129082813.1
Epsilon._Sulfurospirillum_arcachonense_WP_024954565.1
Geobac._Geobacter_metallireducens_WP_004513868.1
Geobac._Trichlorobacter_lovleyi_WP_012470714.1
PVC_Pontiella_sp_WP_372846032.1
PVC_Anaerobaca_lacustris_WP_349245863.1
PVC_Anaerohalosphaera_lusitana_WP_146663764.1
Desulfohalobium_retbaense_WP_015752006.1
Desulfonatronum_lacustre_WP_028572086.1
Desulfonatronum_thiodismutans_WP_031385759.1
Desulfoplanes_formicivorans_WP_069859258.1
Desulfonatronovibrio_hydrogenovorans_WP_028573589.1
Desulfonauticus_submarinus_WP_092066209.1
Desulfovibrio_desulfuricans_WP_136399135.1
Desulfovibrio_fairfieldensis_WP_062253165.1
Desulfovibrio_sp_WP_012624422.1
Group 2
(BP:89)
PVC_Victivallis_sp_WP_176012583.1
Marispirochaeta_aestuarii_WP_319476230.1
Sediminispirochaeta_bajacaliforniensis_WP_020614295.1
Thiobacillus_sp_WP_296768296.1
Thiobacter_sp_WP_347308828.1
Magnetococcus_marinus_WP_011713220.1
Protochlamydia_amoebophila_WP_011174834.1
Criblamydia_sequanensis_WP_041016379.1
Calidithermus_roseus_WP_119279675.1
Marinithermus_hydrothermalis_WP_013703590.1
Oceanithermus_profundus_WP_013458263.1
Truepera_radiovictrix_WP_013177576.1
Ferrimicrobium_sp_WP_366525811.1
Coprothermobacter_proteolyticus_WP_083760229.1
Laterigemmans_baculatus_WP_164102913.1
Hydrogenivirga_caldilitoris_WP_121008581.1
Nitrosomonas_sp_WP_309371965.1
Anaerolinea_sp_WP_322509095.1
Thermanaerothrix_daxensis_WP_054521635.1
Armatimonas_rosea_WP_184201793.1
Capsulimonas_corticalis_WP_119321929.1
Chthonomonas_calidirosea_WP_075162441.1
Acidipila_rosea_WP_131997412.1
Acidobacterium_sp_WP_041839475.1
Edaphobacter_aggregans_WP_125486316.1
Terriglobus_sp_WP_371082807.1
Pyrinomonas_methylaliphatogenes_WP_041975897.1
Myxococcota_Bradymonas_sediminis_WP_111333819.1
Myxococcota_Persicimonas_caeni_WP_141201397.1
Myxococcota_Lujinxingia_litoralis_WP_111727886.1
Dadabacteria_bacterium_MCH8030110.1
Geoalkalibacter_ferrihydriticus_WP_040095096.1
Geoalkalibacter_halelectricus_WP_260746616.1
Desulfuromonas_sp_WP_155875295.1
Geopsychrobacter_electrodiphilus_WP_020676031.1
Malonomonas_rubra_WP_072904816.1
Desulfuromonas_versatilis_WP_221251554.1
Geothermobacter_ehrlichii_WP_148896823.1
Desulfatirhabdium_butyrativorans_WP_028325718.1
Desulfomonile_tiedjei_MBI5572275.1
Desulfocapsa_sulfexigens_HIP39509.1
Desulfobacter_postgatei_WP_004073490.1
Desulfoluna_butyratoxydans_WP_180143363.1
Desulfococcus_multivorans_WP_020877284.1
Group 1
(BP:96)
Moorella_thermoacetica_WP_011393440.1
Moorella_stamsii_WP_054936622.1
Bacteroidota_bacterium_MFC2130317.1
Caldithrix_abyssi_WP_006930019.1
Chlorobaculum_parvum_WP_041466082.1
Chlorobium_ferrooxidans_WP_006367167.1
Pelodictyon_luteolum_WP_041463713.1
Prosthecochloris_sp_WP_094081706.1
Chloroherpeton_thalassium_WP_012500106.1
Gemmatimonadota_bacterium_MFC1529844.1
Hakubella_thermalkaliphila_WP_176232893.1
Acetomicrobium_thermoterrenum_WP_091461488.1
Sporanaerobacter_acetigenes_WP_072745342.1
Prevotella_merdae_WP_373172748.1
Dysgonomonas_sp_WP_166109648.1
Hujiaoplasma_nucleasis_WP_312031991.1
Acholeplasma_equirhinis_WP_205734959.1
Acholeplasma_hippikon_WP_232034474.1
Paracholeplasma_manati_WP_263607901.1
Haloplasma_contractile_WP_008826012.1
Clostridium_drakei_WP_032076033.1
Fusobacteria_bacterium_UUV17300.1
Psychrilyobacter_atlanticus_WP_028857016.1
Propionigenium_maris_WP_281837558.1
Caldisericum_exile_WP_014452962.1
Cerasicoccus_arenae_WP_189512523.1
Kosmotoga_pacifica_WP_047754342.1
Mesotoga_sp_WP_372962495.1
Pseudothermotoga_sp_WP_346309370.1
Thermotoga_caldifontis_WP_041078073.1
Carboxydothermus_hydrogenoformans_WP_011343426.1
Dictyoglomus_sp_WP_012583868.1
Elusimicrobium_minutum_WP_012415516.1
Endomicrobiellum_agilis_WP_374128675.1
Palaeococcus_ferrophilus_WP_048149056.1
Thermococcus_waiotapuensis_WP_315340953.1
Thermococcus_argininiproducens_WP_251948260.1
Hyperthermus_butylicus_WP_048061501.1
Stygiolobus_sp_WP_337701701.1
Thermoproteota_archaeon_MFC1505878.1
Hadarchaeum_sp_WP_342812294.1
Actinomycetes_bacterium_MFD0048410.1
Anoxybacillus_vitaminiphilus_WP_111644200.1
Alkalicoccus_halolimnae_WP_147802881.1
Tumebacillus_lacus_WP_267152694.1
Polycladomyces_abyssicola_WP_212774643.1
Paenibacillus_antri_WP_138196386.1
Tengunoibacter_tsumagoiensis_WP_126578632.1
Thermosporothrix_hazakensis_WP_111325458.1
Manganitrophus_noduliformans_WP_320412342.1
Desertifilum_sp_WP_069969943.1
Geitlerinema_sp_WP_071516204.1
Clade A
Clade B
Clade C
Clade D
Clade E
LG+I+R6_377 Sites
0.2

Fig. 5

Maximum likelihood phylogenetic tree of GCSP α-subunit (A) and β-subunit (B).

Bootstrap values were calculated using the ultrafast bootstrap method. The evolutionary model, scale bar and number of sites are shown in the figure. Sequence in the phylum TDB are highlighted with colors.

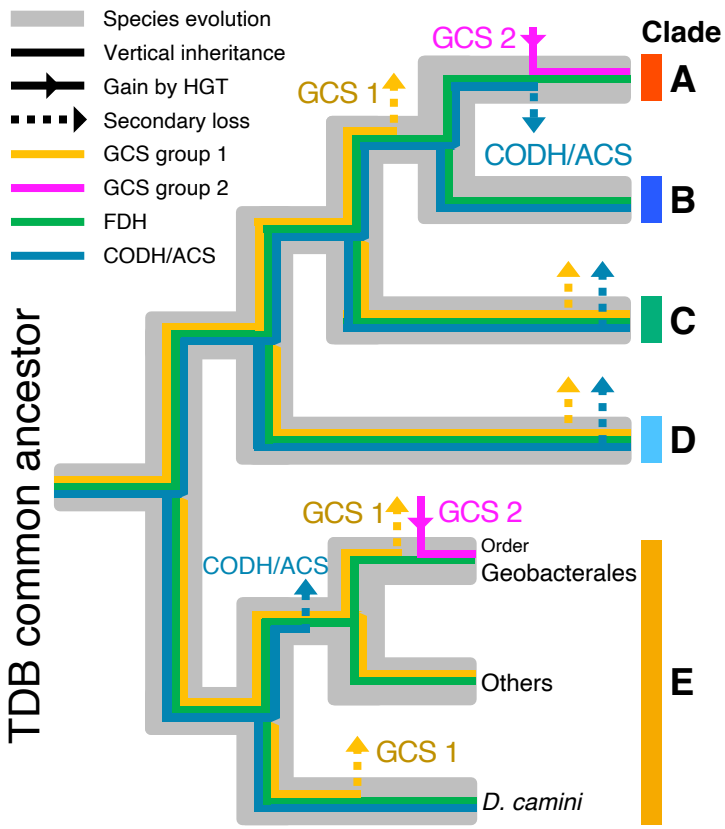


Fig. 6
Gain and loss scenarios of gene sets for the WL and rGly pathways in the phylum Thermodesulfobacteriota.
The phylogenetic tree of the phylum TDB is shown gray. Presence of absence of FDH, CODH/ACS complex, and GCS complex are shown with color bars explained in the figure.