Article

Identification of a deep-branching thermophilic clade sheds light on early bacterial evolution

Hao Leng[1,2,5], Yinzhao Wang [1,2,5], Weishu Zhao [1,2], Stefan M. Sievert [3] & Xiang Xiao [1,2,4] ✉

It has been proposed that early bacteria, or even the last universal common ancestor of all cells, were thermophilic. However, research on the origin and evolution of thermophily is hampered by the difficulties associated with the isolation of deep-branching thermophilic microorganisms in pure culture. Here, we isolate a deep-branching thermophilic bacterium from a deep-sea hydrothermal vent, using a two-step cultivation strategy ("Subtraction-Sub-optimal", StS) designed to isolate rare organisms. The bacterium, which we name *Zhurongbacter thermophilus* 3DAC, is a sulfur-reducing heterotroph that is phylogenetically related to Coprothermobacterota and other thermophilic bacterial groups, forming a clade that seems to represent a major, early-diverging bacterial lineage. The ancestor of this clade might be a thermophilic, strictly anaerobic, motile, hydrogen-dependent, and mixotrophic bacterium. Thus, our study provides insights into the early evolution of thermophilic bacteria.

Organisms with optimum growth temperatures higher than 45 °C are designated as thermophiles[1,2]. These thermophiles are ubiquitously distributed in natural or artificial high-temperature environments, such as terrestrial volcanic sites, hot springs, submarine hydrothermal systems, deep subseafloor, oil reservoirs, solar-heated surface soils and industrial incubators[3–8]. Some thermophiles can also live in alkaline, acidic, or high salinity conditions, highlighting their ability to tolerate multiple stressors[9,10]. These so-called "extreme environments" are generally considered analogs of conditions existing on early Earth when life originated in the late Hadean or early Archaean eon[11,12], and multiple experimental studies[13,14] and geological evidence[15] supported the early origin of thermophiles.

Thermophilic bacteria can be found in a wide range of bacterial phyla, yet only a few phyla contain almost exclusively thermophilic bacteria, namely, Aquificota, Dictyoglomota, Coprothermobacterota, Thermotogota, and Thermodesulfobiota. While it has been suggested that thermophily might represent an

ancestral feature of bacteria or even life[16–20], other hypotheses have put forward a non-thermophilic origin of bacteria and placed bacterial lineages such as Chloroflexota or the Candidate Phyla Radiation (CPR) near the root of the bacterial phylogenomic tree[21–23]. A recent study using an outgroup-free gene tree-species tree reconciliation approach rejected previously predicted bacterial roots and placed a new robust root between two major bacterial branches, i.e., Gracilicutes and Terrabacteria, even though with uncertainties regarding the phylogenetic position of the Fusobacteriota lineage[24]. Nevertheless, it remains unclear when the thermophilic bacteria originated and how their metabolisms evolved.

The predicted metabolic potential of the last bacterial common ancestor (LBCA) or early diverging bacterial lineage can shed light on the early Earth environments and early ecosystem[20,25–28]. For thermophilic bacteria, several previous studies focusing on different lineages such as Thermotogota[29], Aquificota[30], and Coprothermobacterota[31]

[1]State Key Laboratory of Microbial Metabolism, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China. [2]International Center for Deep Life Investigation (IC-DLI), Shanghai Jiao Tong University, Shanghai, China. [3]Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA, USA. [4]Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai, Guangdong, China. [5]These authors contributed equally: Hao Leng, Yinzhao Wang. ✉e-mail: zjxiao2018@sjtu.edu.cn

revealed a complex evolutionary history with a high proportion of genes horizontally transferred from other prokaryotes. However, there is still a lack of an evolutionary study based on comprehensive comparative genomic analyses of major thermophilic bacterial clades. Even though cultivation-free metagenomic analyses have greatly expanded our knowledge of bacterial diversity[32,33], research on thermophilic bacteria still requires isolated strains to provide direct evidence of their physiology, such as optimal growth temperature to confirm their thermophilic features. Yet, it is extremely difficult to obtain pure cultured novel thermophilic bacteria, especially for deep-branching clades.

Here, we report the discovery and isolation of a novel thermophilic, piezophilic, sulfur-reducing bacterial strain, *Zhurongbacter thermophilus* 3DAC. Phylogenomic studies of the strain 3DAC indicate that it is closely related to major thermophilic bacterial groups, such as Coprothermobacterota, Dictyoglomota, Caldisericota, and Thermotogota. Phylogenetic ancestral analyses indicate that these closely clustered thermophilic bacterial lineages share one last common ancestor with the potential to metabolize hydrogen, sulfur compounds, and with mixotrophic carbon catabolism.

Results and discussion

Discovery and isolation of a previously unknown thermophilic and piezophilic bacterium

Isolation with a novel "Subtraction-Suboptimal" strategy. To increase the probability to cultivate novel thermophilic bacteria from the rare biosphere, we designed a two-stage cultivation strategy, named "Subtraction-Suboptimal" (StS) strategy (Fig. 1c). The concept of this method is to reduce the complexity of the community and minimize the abundance of dominant heterotrophic organisms by using first an autotrophic culture medium and subculturing for several times (denoted "Subtraction"), followed by switching to a rich heterotrophic medium at a suboptimal temperature to recover the target heterotrophic thermophilic bacteria present in low abundance and to avoid competition from usually dominant fast-growing hyperthermophilic heterotrophic archaea (denoted "Suboptimal").

In this case, the original sample, collected from an active black smoker chimney emitting hydrothermal fluid with a temperature of 231 °C (Fig. 1a, b; ref. 34), was firstly cultivated in a chemolithoautotrophic medium (SME medium, see "Methods") at 50 °C, then consecutively transferred for five times under autotrophic condition,
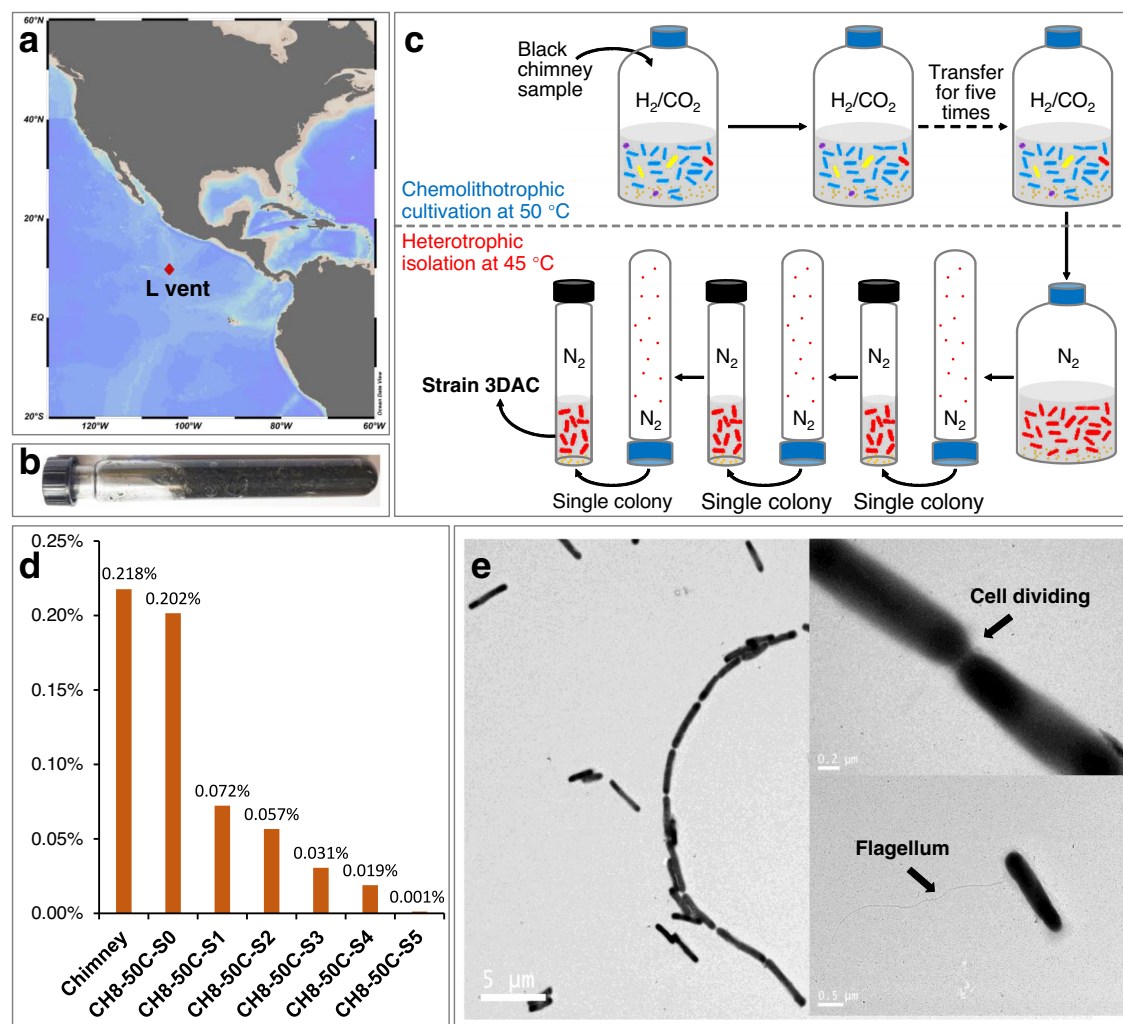
Fig. 1 | Strain isolation procedure. a Location of L-vent at the 9°N vent field on the East Pacific Rise. The map was generated by Ocean Data View (Schlitzer, Reiner, Ocean Data View, https://odv.awi.de, 2021). **b** The black L-vent chimney sample collected in an anaerobic tube. **c** Schematic diagram of the "Subtraction-Suboptimal" (StS) strategy used to isolate *Zhurongbacter thermophilus* strain 3DAC in this study. **d** The track of proportion changes of 3DAC closely related sequences in the chimney sample, original enrichment, and different transfers. Labels: Chimney, the chimney sample; CH8-50C-S0, original enrichment; CH8-50C-S1 to -S5, the first to fifth subcultures transferred from the original enrichment. All samples indicate that strain 3DAC represents a rare species. **e** Transmission electron micrograph of 3DAC, in which 3DAC cells form long filaments with chains of up to 11 cells or more; the micrograph showing a connection between two 3DAC cells; and a long lateral flagellum of 3DAC. The TEM experiment was repeated four times, and long filamentous patterns were observed each time. Source data are provided as a Source Data file.

reducing the microbial community complexity and the ratio of heterotrophs. The conditions of the chemoautotrophic "Subtraction" stage mimicked the natural environment in situ and established a community dominated by autotrophs (mainly *Hydrogenimonas* within Campylobacterota and *Desulfurobacterium* within Aquificota)[35]. It performed similarly to the traditional enrichment method by decreasing the proportion of heterotrophs, but kept the co-cultured heterotrophs with autotrophs. In general, while the community is simplified through the "Subtraction" stage, it still maintains a certain degree of diversity, providing more opportunities for the isolation of rare species supported by autotrophs.

For the "Suboptimal" stage, we switched the cultivation condition from the autotrophic medium to a heterotrophic medium to provide a strong selection for heterotrophs. However, we decreased the cultivation temperature from 50 to 45 °C. While this resulted in a suboptimal growth temperature for thermophilic bacteria, it prevented the growth of heterotrophic hyperthermophilic archaea (i.e., *Thermococcus*, growth temperature ranges between 50 and 100 °C[36]), which were pervasive in the chimney sample, allowing the recovery of rare heterotrophic thermophilic bacteria. After a four-day incubation, growth was observed, followed by the isolation of single transparent and round colonies by using the rolling tube method[37] containing a solid TRM medium.

During the "Subtraction" stage, some low-abundance heterotrophic species co-cultured with the autotrophs can be retained, these rare species can be hardly isolated by traditional methods when fast-growing heterotrophic microorganisms present at the same time. The successful isolation of the strain 3DAC, which represented only ~0.001% of the total community in the fifth transfer of the chemoautotrophic enrichment (~0.2% in the chimney sample, Fig. 1d), highlights the potential of the "StS" strategy to isolate rare species from environmental samples.

Here we provide a new idea of a two-step strategy to achieve the isolation of rare or unknown species that have a similar ecological niche to a dominant group. The "Subtraction" process makes a dilution-to-extinction within a community to realize the simplification of microbial composition, and the "Suboptimal" process recovers the rare species by the suboptimal condition that is more harmful to the dominant group in a similar ecological niche. This "StS" strategy may provide new opportunities to discover rare taxa hidden behind the common dominant taxa in studied environments.

Morphology and physiology. The cells of strain 3DAC are rod-shaped, 1.3–7.5 μm long (average 3.1 ± 1.3, $n = 25$) and 0.4–0.6 μm wide (average 0.5 ± 0.07, $n = 25$). They have a single lateral flagellum and are highly motile under a light microscope. During growth, the strain 3DAC occurs mostly as single cells, but sometimes forms long chains (Fig. 1e and Supplementary Fig. 1). The strain 3DAC grows only under strictly anaerobic conditions and performs as a piezophilic thermophile with: temperature range of 30–75 °C (optimal 70 °C), NaCl range of 1.0–4.5% (w/v) (optimal 2.5%), pH range of 5.5–8.5 (optimal pH 7.0), and hydrostatic pressure range of 0.1–80 MPa (optimal 20 MPa) (Supplementary Fig. 2a–d).

Carbon source utilization. Separately adding peptone, casamino acids, beef extract, yeast extract, and tryptone at a final concentration of 0.2% (w/v) supported the rapid growth of strain 3DAC (Supplementary Fig. 2e). In a basal TRM medium without a carbon source, no growth was detected upon separately adding only one type of any carbohydrate or organic acid (see "Methods") under the tested conditions. However, strain 3DAC grew well with 20 canonical amino acids simultaneously added to the basal TRM medium. Moreover, "leave-one-out" experiments with 20 amino acids revealed that 3DAC relied on 11 essential amino acids (i.e., L-arginine, L-cysteine, L-histidine, L-leucine, L-lysine, L-methionine, L-phenylalanine, L-proline, L-threonine,

L-tryptophan, and L-tyrosine) (Fig. 2b), no growth was observed as the absence of any one of these amino acids. In the presence of 20 amino acids, D-(+)-glucose and pyruvate can be used as additional carbon sources (Fig. 2b and Supplementary Data 1).
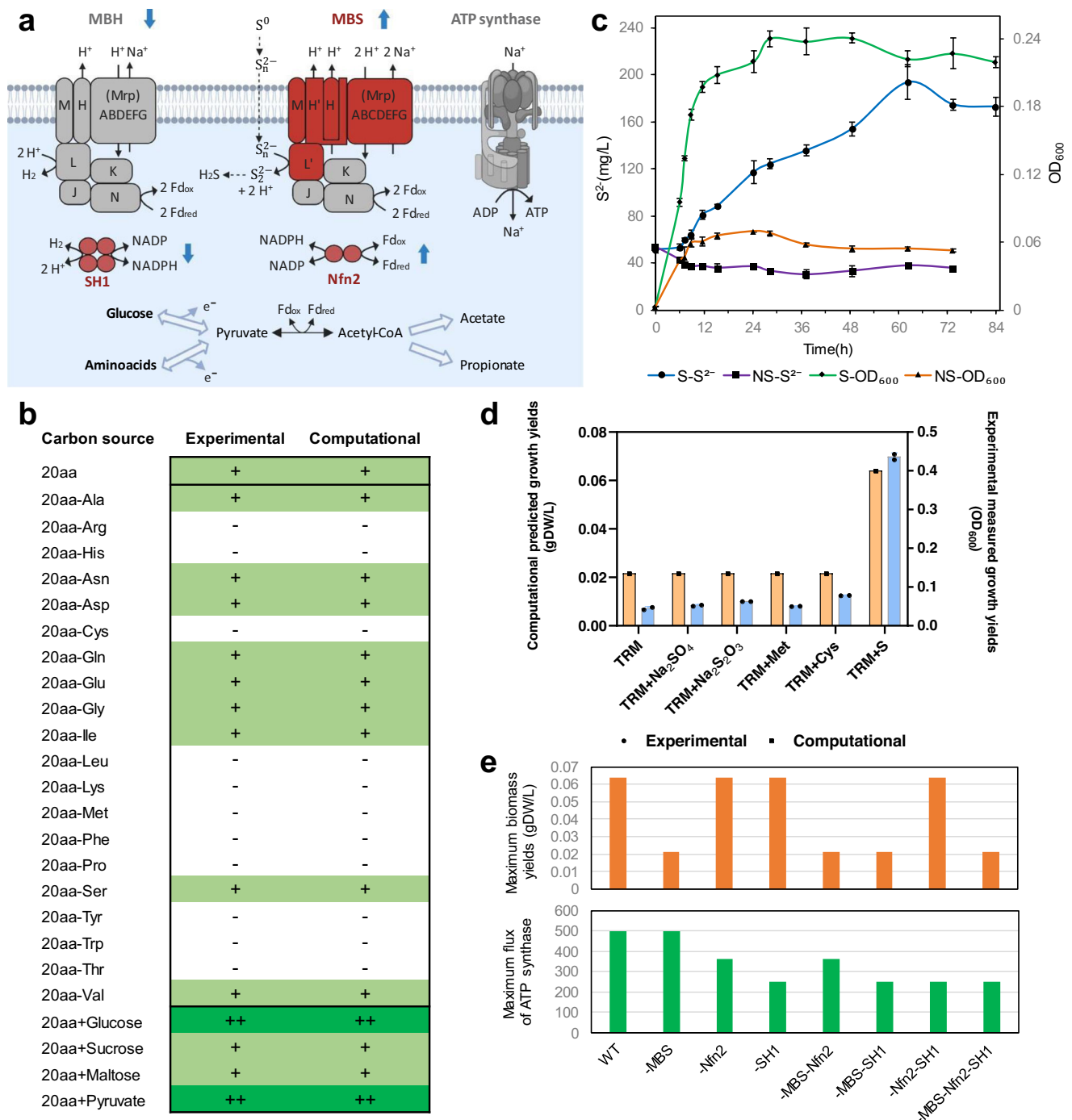
Genomic and phylogenetic analyses indicate 3DAC as a deep-branching, sulfur-reducing heterotrophic bacterium

Genome characteristics. The complete genome of strain 3DAC is a single circular chromosome of 1,588,679 bp with a GC content of 41.15% (Supplementary Fig. 3). The 3DAC genome contains three rRNA gene operons and 46 tRNA genes. A total of 1524 coding DNA sequences (CDSs) with an average length of 980.78 bp, covering 94.09% of the whole genome, can be identified. The strain 3DAC has a complete glycolysis pathway, suggesting a capability of carbohydrate utilization, but many genes related to amino acid synthesis are absent, indicating the requirement of certain amino acids. This phenomenon is confirmed by the results of physiological experiments, the growth of 3DAC depends on the addition of amino acids and can be stimulated by adding carbohydrates (Fig. 2b). The loss of genes related to amino acid synthesis is common among thermophiles. Since the synthesis of amino acids consumes much energy[38], it is more efficient for thermophiles to absorb amino acids directly from the environment, which is more conducive to their adaptation to the hydrothermal environment.

Archaea-derived energetic process in 3DAC. The respiratory energetic process in strain 3DAC contains a Na⁺-driven ATP synthase (Supplementary Fig. 4), a 12-gene cluster coding a membrane-bound hydrogen-producing complex (MBH) and a 13-gene cluster coding the membrane-bound sulfur respiration complex (MBS), which is similar to the energetic process in hyperthermophilic archaea within the order Thermococcales[39–41] (Fig. 2a). MBS enables cells to reduce elemental sulfur (S^0) to hydrogen sulfide (H_2S) in vivo and to use reduced ferredoxin as an electron donor[40]. Genes in MBH of 3DAC, including the key catalytic subunit of hydrogen production (MbhL), are bacterial orthologs, as expected, while most of the MBS-coding genes (10 of 13) of 3DAC are archaeal orthologs encoding the complete membrane arm (MbsA, B, C, D, E, G, H, H' and M subunits) and the key catalytic subunit of sulfur reduction (MbsL) (Fig. 2a and Supplementary Data 1). Compared to the bacterial hydrogen-producing MBH, the archaea-derived MBS enables cells to convert energy more efficiently with a 3.4-fold higher redox potential to reduce elemental sulfur (S^0) to hydrogen sulfide (H_2S) than MBH[39,40]. In addition, two redox-balancing complexes in 3DAC, soluble hydrogenase I (SH1) and NADP-dependent ferredoxin oxidoreductase II (Nfn2), are also archaeal orthologs with a high identity (>60%) as genes reported in hyperthermophilic archaea Thermococcales[42,43].

Consistent with genomic analysis, physiological experiments revealed that the presence of elemental sulfur significantly stimulated the growth of strain 3DAC, compared with the medium without sulfur (Fig. 2c, d). Other sulfur substitutes, such as sodium sulfate, sodium thiosulfate, L-cysteine or L-methionine, did not significantly stimulate growth (Fig. 2d and Supplementary Data 1). During the growth of 3DAC, elemental sulfur is reduced to sulfide (Fig. 2c). This process is proposed to be conducted by MBS with the help of redox balancing via Nfn2 and/or SH1, which is also in agreement with transcriptomic analysis with or without elemental sulfur. We found that most genes (10 of 13) in MBS were significantly upregulated with elemental sulfur in the medium, including the catalytic subunit MbsL. Genes in Nfn2 were also highly expressed with sulfur, indicating the potential coupling of Nfn2 with MBS for redox balancing. In contrast, genes in MBH and SH1 were significantly downregulated in the presence of sulfur, which suggested competition between MBS and MBH and potential coupling between MBH and SH1 (Supplementary Data 1).

Furthermore, a genome-scale metabolic model of 3DAC, GEM-i3DAC (Supplementary Data 1), was constructed to simulate the

Carbon source	Experimental	Computational
20aa	+	+
20aa-Ala	+	+
20aa-Arg	-	-
20aa-His	-	-
20aa-Asn	+	+
20aa-Asp	+	+
20aa-Cys	-	-
20aa-Gln	+	+
20aa-Glu	+	+
20aa-Gly	+	+
20aa-Ile	+	+
20aa-Leu	-	-
20aa-Lys	-	-
20aa-Met	-	-
20aa-Phe	-	-
20aa-Pro	-	-
20aa-Ser	+	+
20aa-Tyr	-	-
20aa-Trp	-	-
20aa-Thr	-	-
20aa-Val	+	+
20aa+Glucose	++	++
20aa+Sucrose	+	+
20aa+Maltose	+	+
20aa+Pyruvate	++	++

Fig. 2 | Energy metabolism of strain 3DAC. a Schematic diagram of the energetic process in strain 3DAC. Red shapes represent archaeal orthologs. Gray shapes represent bacterial orthologs. Solid lines represent enzymatic reactions. Dotted lines represent abiotic reactions. Differentially expressed genes were identified using DESeq2 (version 1.12.4). The default DESeq function with betaPrior = FALSE was applied, wherein a negative binomial generalized model with Wald test for significance was performed. The two-sided p-values were corrected for multiple testing using the Benjamini–Hochberg method. Blue arrows indicate significant up- or down-regulation with sulfur in the culture medium based on transcriptomic data [$n = 3$ per treatment, log2(fold change) >1, P value < 0.05] (Supplementary Data 1). This figure was created using BioRender (https://biorender.com/). **b** Growth capabilities on varied carbon sources of strain 3DAC. Computational predictions were simulated by the genome-scale metabolic model of 3DAC (Supplementary Data 1). "20aa": all 20 amino acids supplemented as the sole carbon source. "20aa-X":

removal of individual amino acids X. "20aa+Y": extra carbon source Y supplied with 20 amino acids. "+": similar growth yields compared with the 20aa condition; "++": at least 1.5-fold higher growth yields than 20aa; "-": no growth. **c** Sulfide concentration and OD_{600} during the growth of strain 3DAC. S: medium with elemental sulfur; NS: no elemental sulfur. Error bars represent the standard deviations (SDs) from independent biological triplicates ($n = 3$) during the experiments under 0.1 MPa. Data are presented as average values ± SD. **d** Experimental and computational growth yields of the 3DAC strain by utilizing various sulfur species. In the experimental group, error bars represent the standard deviations (SDs) from independent biological duplicates ($n = 2$) during the experiments under 0.1 MPa. Data are presented as average values ± SD. **e** Simulations of wild-type 3DAC (WT), single deletions (-MBS, -Nfn2, -SH1), double deletions (-MBS-Nfn2, -MBS-SH1 and -Nfn2-SH1) and triple deletions (-MBS-Nfn2-SH1) of archaea-derived complexes in energetic processes at 3DAC. Source data are provided as a Source Data file.

impacts of the respiratory and redox-balancing complexes derived from archaea (i.e., MBS, SH1 and Nfn2) in 3DAC. The model incorporated both annotations from automatic pipelines and manually curated biochemical evidence of (hyper)thermophiles from the literature (Supplementary Data 1). This model was validated to predict growth trends accurately by utilizing varied carbon sources and electron acceptors compared to the experimental observations (Fig. 2b, d). Simulations were further performed on the 3DAC wild-type, single, double, and triple deletions of the archaea-derived complexes MBS, SH1 and Nfn2. Deletion of MBS blocked sulfur consumption and hydrogen sulfide production and theoretically decreased maximum biomass yields to 33% of the WT (Fig. 2e and Supplementary Data 1), which supported the necessity of MBS for utilizing sulfur as an electron acceptor to stimulate 3DAC growth. The single deletion of Nfn2 and SH1 decreased the maximum flux of ATP synthase to 72% and 50%, respectively, which indicated the essential function of Nfn2 and SH1 in ATP production. These results revealed the necessity of these archaea-derived complexes for both growth and ATP production in the 3DAC strain. Overall, strain 3DAC performed as a thermophilic heterotroph combining bacteria-derived hydrogen-producing and archaea-derived sulfur-reducing complexes for respiration. The ability of the thermophilic strain 3DAC to utilize the archaea-derived energy production might indicate its crucial evolutionary place.

Phylogenetic analyses identify 3DAC as a deep-branching lineage.

Phylogenomic analysis based on different concatenated conserved protein sequences[24,44–46] and 16S rRNA gene-based analysis both place the phylogenetic position of strain 3DAC as a sister lineage to Coprothermobacterota, a bacterial phylum proposed by Pavan et al. (2018) [47] (Fig. 3a–c and Supplementary Fig. 5a–c). The most closely related cultured strain of 3DAC based on both 16S rRNA gene phylogenetic and phylogenomic analyses was *Coprothermobacter proteolyticus* DSM5265 but sharing a low 16S rRNA gene identities (81.25–83.00%) and AAI (48–49%) (Supplementary Data 3 and 4). *C. proteolyticus* DSM5265 is an anaerobic thermophilic microbe frequently identified in artificial thermophilic anaerobic systems, and grows better without sodium chloride and is able to reduce thiosulfate to sulfide[31,48–50], while 3DAC was isolated from the natural hydrothermal vent and has a preference to utilize amino acid and elemental sulfur. Phylogenetic analysis of the 16S rRNA genes shows that strain 3DAC, together with other related 16S rRNA genes from enrichment cultures from tubes of *Alvinella pompejana* collected from a deep-sea vent at 13°N on the East Pacific Rise, forms a monophyletic cluster, clearly separating them from Coprothermobacterota (Supplementary Fig. 5c). We also compiled almost complete 16S rRNA gene sequences (longer than 1400 bp, >95% coverage) from Coprothermobacterota and compared the sequence identities with those of the 3DAC clade. The 16S rRNA gene sequence identities ranged from 75.73 to 83.00% (Supplementary Data 5). Since strain 3DAC was isolated from a high-temperature hydrothermal vent environment, we propose 3DAC as *Zhurongbacter thermophilus* 3DAC, named after Zhurong, the fire god in Chinese myth. For higher taxonomic classification, based on the suggested AAI ranges for the whole-genome[51] and median 16S rRNA gene sequence[52] identities to distinguish a new phylum, it is likely that this lineage represents a previously unknown thermophilic bacterial phylum. However, the relative evolutionary divergence (RED) score of strain 3DAC indicates that 3DAC might belong to a class-level lineage within the phylum Coprothermobacterota based on GTDB taxonomy[53]. There is a discordance between traditional classification and GTDB, hence, here we only temporally call the clade represented by 3DAC as a deep-branching lineage Zhurongbacterota until more 3DAC-related species are discovered for more accurate taxonomic assessment, as well as the prokaryotic taxonomic definition of phylum reach a consensus within the scientific community.

Evolutionary history of the 3DAC-related thermophilic clade

Deep-branching thermophilic bacterial phyla cluster closely with 3DAC on phylogenomic trees. Phylogenetic analysis based on different sets of the concatenated conserved protein sequences with different inference models revealed that, in addition to Coprothermobacterota[31,47,50], strain 3DAC is most closely related to other thermophilic bacterial groups, i.e., Caldisericota[54,55], Dictyoglomota[56], Thermotogota[16,57,58], and Thermodesulfobiota[44,47,59–63] (Fig. 3a–c). These results are consistent with previous phylogenetic studies employing different protein makers[55,64], all found that these thermophilic bacterial lineages clustered together. Nevertheless, we noticed that the 16S rRNA gene-based phylogenetic tree places several non-thermophilic bacterial phyla within a clade, which we term "closely clustered thermophilic bacteria" (CCTB) (Supplementary Fig. 5a, b). The phylogenetic conflict between 16S rRNA gene and concatenated conserved protein sequences may be caused by the potential phylogenetic artifacts arising from a compositional bias of 16S rRNA gene, as have shown in previous studies[65–68], which may not accurately reflect the evolutionary history of the thermophilic lineages. On the contrary, concatenated alignments of multiple universally distributed single-copy marker genes have become increasingly important in evolutionary analyses as they provide greater resolution and span a larger portion of the genome than individual marker sequences[69].

We also tested their phylogenetic positions with two outgroup rooting datasets, i.e., using Archaea as an outgroup, and using bacterial root position proposed in ref. 24. The results indicate that these thermophilic groups consistently cluster near the root of the Archaea–Bacteria phylogenomic tree, or display deep-branching monophyly with the sister clade Terrabacteria by using the two rooting methods, respectively (Fig. 3a, b). Physiologically, all the above-mentioned bacterial lineages also share similar metabolic capabilities, for example, nearly all their lineages have an anaerobic lifestyle, are able to use elemental sulfur or sulfur-related compounds and are heterotrophs with the ability to degrade sugars or/and proteins (Supplementary Data 6). Therefore, here we refer to this clade as "closely clustered thermophilic bacteria (CCTB)" to describe these thermophilic bacterial phyla that cluster closely with 3DAC and branch together on bacterial phylogenomic trees. The CCTB contains at least six lineages that have pure cultured strains: *Zhurongbacter*, Caldisericota, Coprothermobacterota, Dictyoglomota, Thermotogota, and Thermodesulfobiota. Some other candidate phyla such as Deinococcota and Synergistota also cluster with the clade (Fig. 3a) and therefore might belong to the CCTB. However, although a number of their members are also thermophilic bacteria, these phyla have evolved multiple non-thermophilic bacteria lineages[70–72]. Besides, some candidate phyla cluster with the CCTB on the 16S rRNA gene tree but display some inconsistency on the phylogenomic tree using different conserved marker protein sequence sets, such as thermophilic Aquificota. We have carefully checked the phylogeny of Aquificota by different methods, and found that most of the genes within Aquificota have undergone horizontal gene transfer (HGT) as previously described[73,74] (Supplementary Figs. 6 and 7), which would influence its phylogenetic inference. In general, further study regarding whether the phyla Deinococcota, Synergistota, Aquificota, and even other closely related clades had a thermophilic origin and branch as sister lineages with CCTB need to be resolved in the future.

Evolution and metabolic potentials of the predicted CCTB common ancestor. Phylogenetic ancestral analyses indicate that the members of the CCTB (Zhurongbacter, Caldisericota, Coprothermobacterota, Dictyoglomota, Thermotogota, and Thermodesulfobiota) might have shared one potentially last common ancestor. The predicted ancestral gene set from six lineages, in general, includes genes coding for thermophilic-related components (putative heat-shock proteins), flagellar that enables its movement, hydrogenase (FeFe group type) as
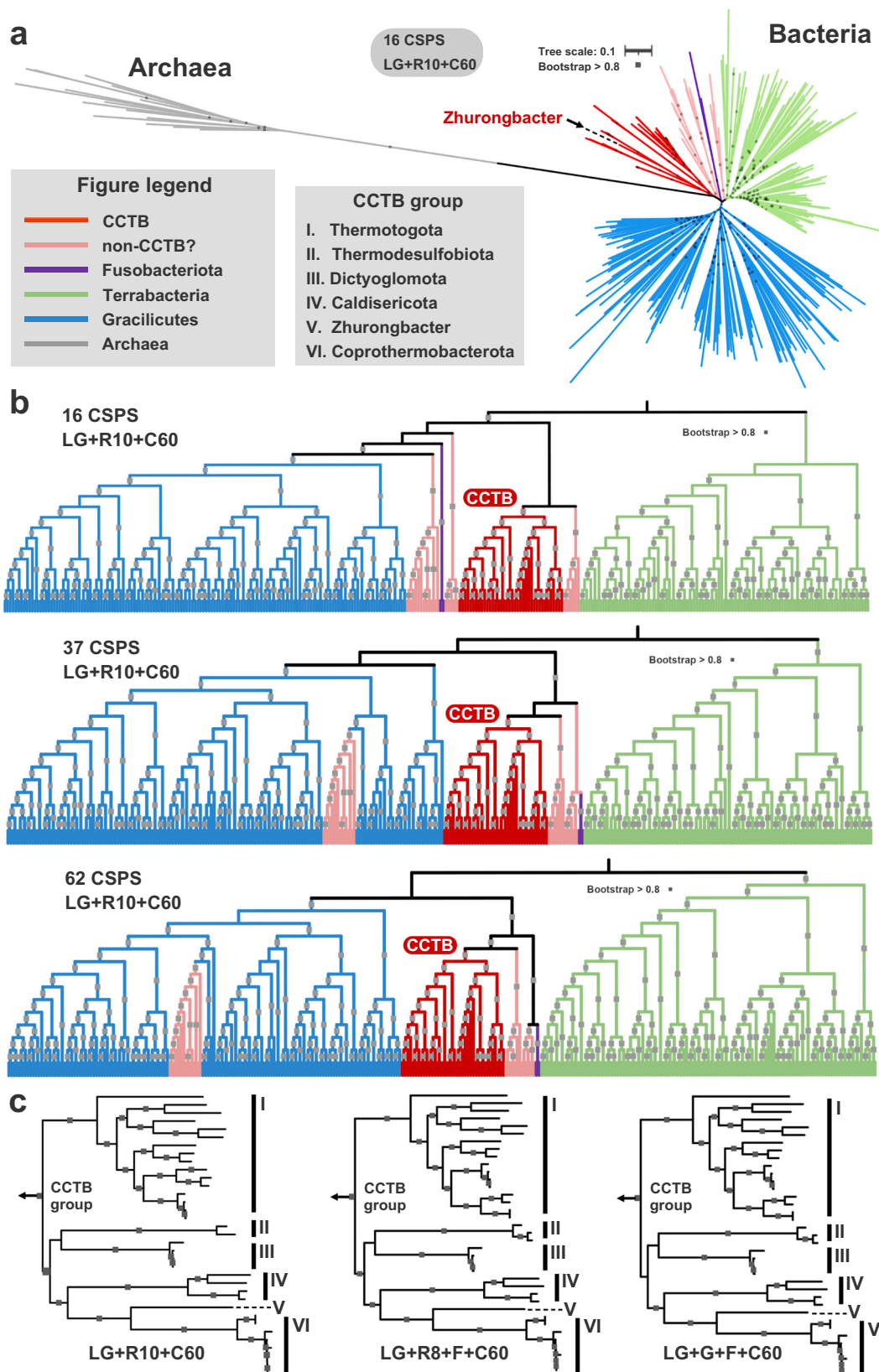
well as saccharide and amino acid utilization pathways (ALE predicted gene copy number >0.3, Supplementary Data 2). Nevertheless, the amino acid biosynthesis pathways are not complete, possibly due to the limitation of ancestral prediction. A genome-scale metabolic model of the CCTB ancestor was further constructed and suggests it is a motile mixotrophic bacterium with carbon dioxide, saccharides and

proteins as substrates (Supplementary Data 2). This predicted ancestral metabolism of the CCTB ancestor also shares many similarities with the recently inferred metabolic capacities of the last universal common ancestor (LUCA) and LBCA[20,24,75]. The CCTB ancestor, LUCA and LBCA are considered strictly anaerobic microorganisms. Both the CCTB ancestor and LUCA are predicted to be thermophilic and share

Fig. 3 | Phylogenetic analyses. a Phylogenomic tree of *Zhurongbacter thermophilus* 3DAC and other major thermophilic bacteria using Archaea as the outgroup with 16 conserved single-copy protein sequences (CSCP) with IQ-Tree using the LG + R10 + C60 model. **b** Phylogenomic tree of *Zhurongbacter thermophilus* 3DAC and other major thermophilic bacteria using 16/37/62 conserved protein sequences. Red color represents major thermophilic bacterial phyla, namely, Zhurongbacterota, Caldisericota, Coprothermobacterota, Dictyoglomota, Thermotogota, and Thermodesulfobiota. The pink color refers to Aquificota, Synergistota and Deinococcota, while the purple color refers to Fusobacteriota. **c** Phylogenomic inferences of the lineages from the CCTB with 37 conserved protein sequences with different models but without the Archaea genome. I-VI represent Thermotogota, Thermodesulfobiota, Dictyoglomota, Caldisericota, Zhurongbacterota, and Coprothermobacterota. The arrows indicate other bacterial branches. In all trees, the gray squares indicate a bootstrap higher than 0.8.
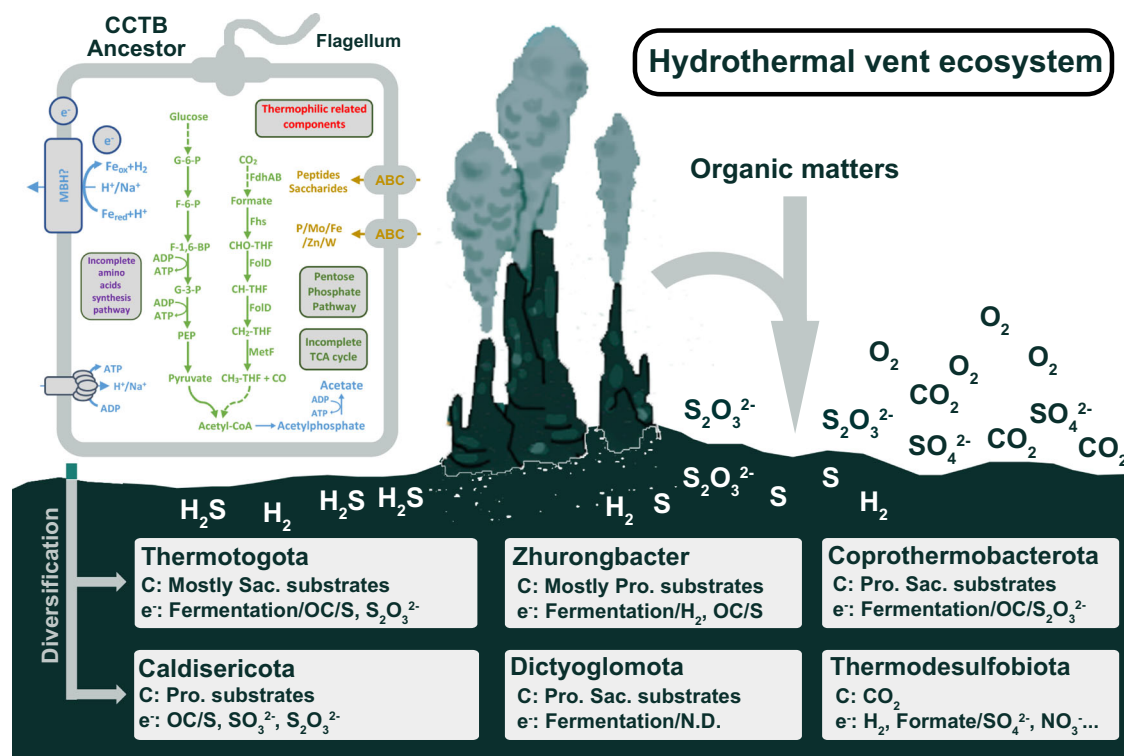
Fig. 4 | Metabolic capability of the CCTB ancestor and its diversification. Illustration of metabolic features of the CCTB ancestor and different bacterial lineages from the CCTB (Supplementary Data 6), the hydrothermal vent here indicates that these bacteria might have originated from similar environments on early Earth but under adaptation radiation during their evolutionary history. Pro proteins, Sac. saccharides, OC organic carbon.

the potential for metabolizing hydrogen, while LBCA and the CCTB ancestor share the predictions for motility and the potential for degrading organic carbon compounds.

Within the CCTB clade, after the diversification of the CCTB ancestor, each lineage adapted to different environments and gradually evolved to form specific features by gene gain and loss (Fig. 4 and Supplementary Fig. 8), especially on sulfur metabolism and carbon utilization pathways. For sulfur respiration, most lineages of the CCTB contained the putative catalytic subunit MbsL-encoding gene of the MBS and/or MBH complex-encoding gene cluster (Supplementary Fig. 9)[40], which is responsible for sulfur and hydrogen metabolisms. As the sulfur respiration capability strongly stimulates the growth of 3DAC in the present study and other strains in the previous studies[76], it indicates that their ancestor might originate from a sulfur compound-rich hydrothermal vent or hot spring environment[77,78]. For carbon assimilation, members of the clades Zhurongbacter, Caldisericota, Coprothermobacterota, Dictyoglomota and Thermotogota are considered as thermophilic heterotrophic bacteria, but they diversified their preferences for carbon source (Supplementary Fig. 8), such as gain and loss of genes for peptide and sugar transport and motility genes (Fig. 4).

As cultivation technologies develop rapidly, more deep-branching thermophilic bacteria clustered with the CCTB will be isolated, and comprehensively characterized through our newly designed cultivation strategy. The discovery of the present novel deep-branching bacterium strain 3DAC with the ability to utilize archaea-derived energy production can be viewed as a piece of the thermophile evolutionary puzzle. With more deep-branching lineages to be discovered in the future, the physiological and evolutionary history of the CCTB will enable us to see through the basic metabolism and niche expansion of early life.

Methods

Sample collection

The original sample was collected with the remotely operated deep-sea research vehicle (ROV) *Jason II* during R/V *Atlantis* research cruise AT26-10 (Dive 761) from a black smoker chimney at L-vent located at the 9°N deep-sea vent field on the East Pacific Rise (9°46.25' N, 104°16.74' W, water depth 2528 m). The temperature of the emitted fluid was 231 °C. More information about the fluid geochemistry and microbial communities in the chimney walls can be found in ref. 34. Pieces of the chimney were stored immediately upon retrieval in sterile anaerobic tubes, and the headspace in the tubes was filled with pure nitrogen gas with slight overpressure to prevent the entry of oxygen. The samples were then stored at 4 °C until enrichment in the laboratory.

"Subtraction-Suboptimal" strategy-guided isolation

The outer layer of the black chimney sample (0.25 g) was first enriched anaerobically in an SME medium[79], which contained the following

components (L^{-1}): NaCl, 20.0 g; $MgSO_4 \cdot 7 H_2O$, 3.5 g; $MgCl_2 \cdot 6 H_2O$, 2.75 g; KCl, 0.325 g; KNO_3, 2.0 g; MES, 4.27 g; NaBr, 50.0 mg; H_3BO_3, 15.0 mg; $SrCl_2 \cdot 6 H_2O$, 7.5 mg; $(NH_4)_2SO_4$, 10.0 mg; KI, 0.05 mg; $Na_2WO_2 \cdot 2H_2O$, 0.1 mg; $CaCl_2 \cdot 2H_2O$, 0.75 g; KH_2PO_4, 0.5 g; $NiCl_2 \cdot 6H_2O$, 2.0 mg; and resazurin, 1.0 mg. The pH of the medium was adjusted to 6.0 using 1 M NaOH. After autoclaving, 1 ml of trace element mixture, 1 ml of vitamin mixture, 1 ml of thiamine solution, and 1 ml of vitamin B12 solution[80] were added to the medium (L^{-1}). Then, 50 ml aliquots of the medium were separated into 150-ml glass serum bottles and supplemented with 0.5 g of elemental sulfur and 0.5 ml of $Na_2S \cdot 9H_2O$ (10%, w/v; pH 7.0), and the tightly stoppered bottles were pressurized with hydrogen/carbon dioxide (80:20; 101 kPa). The cultures were incubated at 50 °C. We obtained the original enrichment in this chemolithotrophic medium, subcultured (1% inoculation) the cells in this medium for five generations, and then transferred the enrichment of the fifth generation (1% inoculation) to a heterotrophic TRM medium[81] with the following composition (L^{-1}): 1 g of yeast extract, 4 g of tryptone, 23 g of NaCl, 5 g of $MgCl_2 \cdot 6H_2O$, 0.7 g of KCl, 0.5 g of $(NH_4)_2SO_4$, 3.3 g of PIPES disodium salt, 1 ml of NaBr (5%, w/v), 1 ml of $SrCl_2 \cdot 6H_2O$ (1%, w/v), 1 ml of KH_2PO_4 (5%, w/v), 1 ml of K_2HPO_4 (5%, w/v), 1 ml of $CaCl_2 \cdot 2 H_2O$ (2%, w/v), and 1 ml of Na_2WO_4 (10 mM). The medium was adjusted to pH 7.0, autoclaved, separated into anaerobic tubes (5 ml), and supplemented with 0.05 g of elemental sulfur and 0.05 ml of $Na_2S \cdot 9 H_2O$ (10%, w/v; pH 7.0), and the headspace of the tubes was pressurized with pure nitrogen gas. Heterotrophic enrichment was performed at 45 °C. Colonies were obtained on solid TRM medium (1.5% Gelrite) using the rolling tube method, which was repeated three times.

Epifluorescence microscopy and transmission electron microscopy

Fluorescent staining by SYBR Green I (Invitrogen) was carried out on cells filtered onto 0.2-μm polycarbonate GTBP membranes (Millipore) following established protocols[82]. Images were captured with a fluorescence microscope (Nikon Eclipse 90i).

For transmission electron microscopy, cells were harvested at $3000 \times g$ for 5 min, fixed in 4% formaldehyde for 2 h at room temperature, and washed with 1× PBS (pH 7.4). Fixed cells were negatively stained with 2% (w/v) phosphotungstic acid (pH 6.5) for observation with a transmission electron microscope (FEI Tecnai G2 Spirit Biotwin) at an acceleration voltage of 120 kV.

Growth characteristics

The effects of the temperature, pH, and NaCl concentration on the growth of strain 3DAC were determined in an anaerobic TRM medium with a headspace gas composed of pure nitrogen gas at atmospheric pressure supplemented with 1% (w/v) elemental sulfur. To evaluate the optimum temperature for growth, strain 3DAC was cultivated at 25–80 °C with 5 °C intervals. The ranges of pH and NaCl concentrations for growth were determined at 45 °C. The optimal pH for growth was determined by varying the pH in the culture medium between 4.0 and 9.0 using the following buffers at a concentration of 10 mM: acetate at pH 4.0, 4.5, and 5.0; MES at pH 5.5 and 6.0; PIPES at pH 6.5 and 7.0; HEPES at pH 7.5 and 8.0; and TAPS at pH 8.5 and 9.0. The pH was adjusted at room temperature before autoclaving. To test the effect of the NaCl concentration on growth, NaCl was added to the medium at final concentrations within the range of 0.5–5.0% (w/v). All of these tests were performed in triplicate in the dark without shaking. Growth was determined by monitoring the OD_{600} with a quartz colorimeter (50 mm optical path) on a spectrophotometer (HACH DR5000). The specific growth rate (μ; h^{-1}) was calculated according to the equation $\mu = \ln[OD_{600}(t_2)/OD_{600}(t_1)]/(t_2 - t_1)$, where $OD_{600}(t_1)$ and $OD_{600}(t_2)$ are the optical densities of liquid cultures measured at 600 nm at time points t_2 and t_1, and t_1 and t_2 are the time points (in hours) of the exponential growth phases[83]. Growth rates were calculated using linear regression analysis of 4 to 6 points along the linear portions of the growth curves.

Determination of growth pressure

The pressure range (0.1, 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100 MPa) was tested with syringes loaded with TRM (2.5% NaCl, pH 7.0) supplemented with 1% (w/v) elemental sulfur. These syringes were placed into high-pressure and high-temperature reactors and incubated at 70 °C. Growth yields and growth rates were measured as previously described. All of these tests were performed in triplicate.

Essential amino acid tests

Mother liquors of 20 kinds of amino acids (L-alanine, L-arginine, L-asparagine, L-aspartic acid, L-cysteine, L-glutamine, L-glutamic acid, glycine, L-histidine, L-isoleucine, L-leucine, L-lysine, L-methionine, L-phenylalanine, L-proline, L-serine, L-threonine, L-tryptophan, L-tyrosine, and L-valine) were prepared. Then, each combination of 19 amino acid mother liquors was added to basal TRM medium (2.5% NaCl, no yeast extract or tryptone, supplemented with 1 ml of trace element solution[84], 10 ml of vitamin solution[85] and 1% (w/v) elemental sulfur) to maintain each amino acid at a final concentration of 0.1 g/L. Basal TRM medium containing all 20 amino acids and no amino acids was set as the positive and negative controls, respectively. The pH of all the tested media was adjusted to 7.0 at room temperature. For all tests, positive cultures were transferred (1% inoculum) into the tested medium at least five times for confirmation of growth. All tests were performed in duplicate at atmospheric pressure and with incubation at 70 °C.

Determination of growth requirements

Various single carbon sources were added to a basal TRM medium (2.5% NaCl, containing only 0.002% yeast extract, without any other organic carbon sources) supplemented with 1% (w/v) elemental sulfur. The following single carbon sources, all at a final concentration of 0.2% (w/v), were tested: peptone, casamino acids, beef extract, yeast extract, tryptone, sodium acetate, sodium formate, sodium citrate, sodium pyruvate, sodium oxalate, D-fructose, D-(+)-glucose, D-(+)-xylose, D-(+)-maltose, sucrose, starch, glycogen, N-acetyl-D-glucosamine, D-mannitol, L-alanine, L-arginine, L-asparagine, L-aspartic acid, L-cysteine, L-glutamine, L-glutamic acid, glycine, L-histidine, L-isoleucine, L-leucine, L-lysine, L-methionine, L-phenylalanine, L-proline, L-serine, L-threonine, L-tryptophan, L-tyrosine, and L-valine. Growth was recorded after 24 and 48 h. Negative (uninoculated basal TRM) and positive (inoculated TRM containing yeast extract and peptone) controls were prepared for each substrate. For all tests, positive cultures were transferred (1% inoculum) into the tested medium for confirmation of growth.

To test carbon source utilization with the addition of 20 kinds of amino acids, various carbon sources at a final concentration of 0.2% (w/v) were added into a basal TRM medium (2.5% NaCl, no yeast extract or tryptone) and supplemented with 20 kinds of amino acids (each at a final concentration of 0.1 g/L) and 1% (w/v) elemental sulfur. The carbon sources used were as follows: D-(+)-glucose, D-fructose, D-(+)-maltose, sucrose, starch, sodium acetate, sodium formate, and sodium pyruvate. Basal TRM medium supplemented with 20 kinds of amino acids and 1% (w/v) elemental sulfur was used as a control.

To examine the ability of the isolate to grow in the absence of elemental sulfur, cells were cultured in TRM medium (2.5% NaCl, sodium sulfate removed) without sulfur. Other sulfur substitutes, such as sodium sulfate (20 mM), sodium sulfite (20 mM), sodium thiosulfate (20 mM), L-cysteine (20 mM) and L-methionine (20 mM), were also tested.

The pH of all the tested media was adjusted to 7.0 at room temperature. All tests were performed in duplicate at atmospheric

pressure and with incubation at 70 °C. In cultures containing insoluble substrates, growth was monitored by microscopic examination.

The sulfide concentration was measured by the methylene blue method[86] using a sulfide measurement kit (HACH) on a spectro-photometer (HACH DR5000).

Genome sequencing and analysis

Genomic DNA was extracted by a modified SDS-based DNA extraction method[87] and sequenced on the PacBio RSII SMRT and Illumina HiSeq X Ten sequencing platforms (BGI, China). The genome was assembled into one contig from the PacBio reads (3668 Mbp, 2308-fold coverage) by Falcon v0.3.0[88] and Celera Assembler v8.3[89]. Then, the assembled genome was corrected with Illumina HiSeq data (736 Mbp, 463-fold coverage) using GATK v1.6–13[90]. Protein-coding genes were predicted by Glimmer v3.02[91], and rRNA and tRNA annotations were performed with RNAmmer v1.2[92] and tRNAscan-SE v1.3.1[93], respectively. Predicted gene functions were annotated with databases including NCBI-nr (updated on Oct 10, 2017), KEGG (BlastKOALA, v2.2, updated on May 15, 2019)[94], and eggNOG-Mapper v2.0.1b-2-g816e190[95]. Average amino acid identity (AAI) values were calculated by CompareM v.0.0.23 (https://github.com/dparks1134/CompareM).

Amplicon sequencing and diversity analyses

For the chimney sample and chemoautotrophic enrichment, the bacterial V4 region of the 16S rRNA gene was amplified by B515F (5′-XXXXXXXXXGTGCCAGCMGCCGCGGTAA-3′) with eight-nucleotide key tags for each sample and B806R (5′-GGACTACHVGGGTWTCTAAT-3′)[96]; the X region represents the various key tags for each sample. The PCR program was 3 min at 95 °C, followed by 35 cycles of 94 °C for 40 s, 56 °C for 1 min, and 72 °C for 1 min. The final extension step was 72 °C for 7 min. The 50 μL amplification mixture contained 1 μL of each forward and reverse primer, 1 μL template DNA, 5 μL 10× ExTaq buffer, 0.5 μL ExTaq polymerase (TaKaRa, Japan), 4 μL of 2.5 mM dNTP mix and 37.5 μL ddwater. PCR products were purified with an E.Z.N.A. Gel Extraction Kit (OmegaBio-Tek, USA) and sequenced on the MiSeq platform (Illumina, USA) according to the manufacturer's instructions.

Data analysis of the 16S rRNA MiSeq sequences was performed using the QIIME version 1.9.1 software pipeline[97] and the QIIME-compatible version of the SILVA-132 database for template-based alignment and taxonomic assignment. Assembled reads that passed the chimera check were clustered into de novo operational taxonomic units (OTUs) at a cutoff of 97% sequence similarity.

RNA sequencing and quantitative real-time PCR

Cells of strain 3DAC cultured with and without sulfur were harvested at the log phase by centrifugation. Three replicates under each condition were prepared for transcriptome analysis. Total RNA was extracted using a Total RNA Extractor (TRIzol) kit (Sangon Biotech, China) according to the manufacturer's protocol. For RNA sequencing library preparation, genomic DNA was removed with DNase I, and ribosomal RNA (rRNA) was removed with a Ribo-off rRNA Depletion Kit (bacteria) (Vazyme Biotech Co., Ltd, China). A complementary DNA (cDNA) library was generated by a VAHTS™ Stranded mRNA-seq Library Prep Kit (Vazyme Biotech Co., Ltd, China) following the manufacturer's instructions. Sequencing was performed on the Illumina HiSeq 2500 (Illumina, USA) platform at Sangon Biotech (Shanghai Co., Ltd., China).

The sequencing quality controls of the raw reads produced by Illumina HiSeq 2500 were assessed using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and all samples passed. Low-quality reads (Q value of <20), ambiguous "N" nucleotides, adapter sequences and fragments of <35 bp were removed from raw reads with Trimmomatic[98] for further analysis. rRNA was removed using a sortmerna[99] with default settings. Read mapping of the non-rRNA to the genome of *Zhurongbacter thermophilus* 3DAC was performed using Bowtie 2[100], and the relevant mapping information was summarized. The uniquely mapped reads were collected and analyzed with the DESeq2 package based on the Poisson distribution to identify the differentially expressed genes (DEGs). Estimation of gene expression levels was based on the transcript per million (TPM) values[101].

To validate the RNA-Seq data, the expression levels of ten randomly selected genes were quantified using quantitative real-time PCR (qRT-PCR). After the extraction of RNA, residual DNA was removed by a TURBO DNA-free Kit (Thermo Fisher Scientific, USA), and cDNA synthesis reverse transcription was followed by a RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific, USA). The primers used for quantitative real-time PCR (RT-qPCR) were designed by Primer-BLAST[102] (http://www.ncbi.nlm.nih.gov/tools/primer-blast) and are listed in Supplementary Data 7. Gene expression was quantified in different samples using the 16S rRNA gene as a reference gene. RT-qPCR was performed on a StepOnePlus Real-Time PCR instrument (Applied Biosystems, USA) with PowerUp SYBR Green Master Mix (Applied Biosystems, USA) using the following program: 95 °C for 10 min, followed by 40 cycles of 95 °C for 15 s and 60 °C for 1 min. All qRT-PCR analyses were performed with four technical replicates. The RNA-Seq fold changes were plotted against the qRT-PCR fold changes, and the correlation coefficients (R^2) between these two datasets were calculated (Supplementary Fig. 10).

Phylogenetic analyses

For the 16S rRNA gene tree, the 16S gene sequences were aligned using SINA v1.2.11[103], and the full alignment was stripped of columns containing 70% or more gaps with a Perl script. The maximum-likelihood phylogeny of the 16S rRNA gene tree was inferred by IQ-TREE[104] v2.0.6 with the "-MFP -B 1000" options for best-fit model selection and ultrafast bootstrap. The reference sequences of other bacteria were downloaded from the Silva database, sequences smaller than 1400 bp were filtered out, and OTUs were picked from the remaining sequences at a cutoff of 97%. The above-described representative OTU sequences were used for the tree.

For phylogenomic analysis based on conserved proteins, the representative bacterial dataset used by Coleman et al.,[24] and selected archaeal reference genomes from the different phyla were downloaded from the NCBI prokaryotic genome database (https://www.ncbi.nlm.nih.gov/assembly/). These reference genomes and the 3DAC genome from this study were used to construct a phylogenomic tree based on a concatenated alignment of three different sets of marker genes (Supplementary Data 8–10)[24,45,46]. Specifically, each of the marker protein sequences from the reference genomes and the 3DAC genome was aligned using the MAFFT algorithm v7.313[105] with the parameters --ep 0 --genafpair --maxiterate 1000 and filtered with tri-mAl v1.4.rev2[106] with the parameter -automated1. Then, all marker protein sequences were concatenated into a single alignment, and phylogenetic trees were built using IQ-Tree[104] v2.0.6 with the best-fitting model LG + C60 + R10 with an ultrafast bootstrap value of 1000.

To illustrate the phylogenetic affiliation of the phylum Aquificota, reference genomes were downloaded from the NCBI database, and 5-10 genomes were downloaded for each phylum according to the NCBI taxonomy. Then each of the 37 marker protein sequences from the reference genomes was extracted and aligned using the MAFFT algorithm v7.313 with the parameters --ep 0 --genafpair --maxiterate 1000 and filtered with trimAl v1.4.rev2 with the parameter -automated1. Then 37 marker protein sequences were separately used to build phylogenetic trees using IQ-Tree v2.0.6 with the model LG + C60 + F + G with an ultrafast bootstrap value of 1000 to evaluate the potential HGT events.

For the phylogenetic analysis of the MbsL protein, its sequences were extracted from the reference genomes of Caldisericota, Thermotogota, Thermodesulfobiota, Aquificota, Zhrongbacter, and Thermococcales. These sequences with references were then aligned using the MAFFT algorithm v7.313 with the parameters --ep 0 --genafpair

--maxiterate 1000 and filtered with trimAl v1.4.rev2 with the parameter -automated1. After that, the phylogenetic tree was built using IQ-Tree v2.0.6 by the best-fitting model with an ultrafast bootstrap value of 1000. All trees were visualized by iTOL[107] online software.

Evolutionary analyses of the CCTB

A total of 43 cultured bacterial complete genomes (Supplementary Data 11) from the CCTB, as well as potential candidate phyla were selected for evolutionary analyses. Genomes were downloaded from NCBI genomic database and predicted with Prodigal version 2.6.3[108]. Then all genomes were annotated with eggNOG-Mapper v2.0.1b-2-g816e190[95]. The comparative genomic analysis was performed by OrthoFinder[109] with default parameters. A total of 6,578 orthogroups from 43 representative genomes were obtained from OrthoFinder[109]. The orthogroups that contain four or more sequences were separately aligned using the MAFFT algorithm v7.313[105] with the parameters --ep 0 --genafpair --maxiterate 1000, filtered with trimAl v1.4.rev2[106] with the parameter -automated1 and then trees were constructed using IQ-Tree with the best-fitting model. To address the evolutionary history of the CCTB, ancestral family gene sets were inferred using the program amalgamated likelihood estimation (ALE) version 0.4[110]. Only bacteria with complete genome information were considered in the present study because incomplete genomes would lead to biased analyses in gene gain and loss processes. Here, the phyla Synergistota and Deinococcota were used as the outgroups. The representative genomes from the phyla Aquificota, Thermosulfidibacterota and Thermodesulfobacterota were also considered in the ALE analysis. For the CCTB, cultured bacteria with complete genomes in the phylum Dictyoglomota includes *Dictyoglomus turgidum* DSM6724 and *Dictyoglomus thermophilum* H-6-12; the phylum Thermodesulfobiota includes Thermodesulfobium acidiphilum 3127-1 GCA and Thermodesulfobium narugense DSM14796; the phylum Caldisericota includes *Caldisericum exile* AZM16c01; the phylum Zhurongbacter includes *Zhurongbacter thermophilus* 3DAC; the phylum Coprothermobacterota includes *Coprothermobacter platensis* DSM11748 and *Coprothermobacter proteolyticus* DSM5265; and the phylum Thermotogota includes *Pseudothermotoga hypogea* DSM11164, *Thermotoga neapolitana* DSM4359, *Thermosipho melanesiensis* BI429, *Fervidobacterium pennivorans* DSM9078, *Kosmotoga pacifica* SLHLJ1, *Mesotoga prima* MesG1Ag42, *Defluviitoga tunisiensis* L3, *Petrotoga mobilis* SJ95, *Marinitoga piezophila* KA3, and *Athalassotoga saccharophila* NAS-01. Ancestor metabolic reconstruction was performed with the predicted ancestral gene set (851 sequences) at the node of the CCTB with an average gene copy number larger than 0.3.

Metabolic modeling

A genome-scale metabolic model of *Zhurongbacter thermophilus* 3DAC, GEM-i3DAC, was constructed according to the complete genome of the 3DAC strain. The draft model construction referred to ortholog mapping to public databases (KEGG[111], EggNOG[95], BIGG[112,113], ModelSeed[114], and TCDB[115]) (performed in November 2020) and the published model of the related thermophilic bacterial strain *Thermotoga maritima* MSB8[116,117]. Extensive manual curations were carried out following draft model reconstruction to integrate the latest biochemical evidence of enzymatic functions in (hyper)thermophiles. Overall, literature evidence was assigned to 57 reactions in the 3DAC model. The biomass equations of 3DAC were formulated based on the biosynthesis of macromolecules, including DNA, RNA, and proteins. Their biosynthesis was defined to account for the millimole composition of each building block in assembling one gram of a given component and the associated energy cost. Simulations of carbon source utilization, electron acceptors, and gene deletion were formulated with exchange constraints that represent the corresponding culture conditions used in the experiments.

The KO terms of the predicted ancestral gene set were used to construct the initial genome-scale metabolic model of the last common ancestor of the CCTB. Reactions for each KO term were obtained from the latest KEGG database (performed in March 2022). Compounds were initially introduced from the latest KEGG database, but the physiological charges of each and the modified hydrogen number of charged compounds were added. The biomass objective function was introduced from the *Escherichia coli* core model[118], including DNA, RNA, and proteins that accounted for ~80% of the dry weight of cells[119] (Supplementary Data 2). Metabolic simulations were performed with exchange constraints identified based on the environmental parameters measured in typical hydrothermal vents[120] and the potential nutrients and products revealed by metabolic reconstruction (Supplementary Data 2). The current knowledge of amino acid synthesis pathways in bacteria and archaea based on gene annotation is inadequate, and amino acids that are missing or incomplete in the synthesis pathway may also be nonessential[121,122]. 3DAC have very similar metabolic characteristics to *Thermococcus*, and based on our previous results of essential amino acid testing for *Thermococcus eurythermalis* A501 and the presence/absence of their key enzymes in the complete genome[123], we have delineated a criterion for determining essential amino acids at the genomic level, trying to reduce the bias caused by this knowledge gap. Essential amino acids were identified as those missing 50% or more of the reactions from the whole biosynthesis pathway. Other amino acids were treated as nonessential. To quantitatively simulate the growth yields using the tested sole carbon sources, each carbon source was constrained to 10 mM of total carbon to avoid the influence of different carbon numbers in the formula, with all essential amino acids (1 mM for each).

For both the 3DAC model and the ancestor model, consistency checks of all reactions were computationally conducted using the *formulacheck, chargecheck* and *masscheck* functions in PSAMM version 1.1.1 (https://zhanglab.github.io/psamm/)[124], and the unbalanced equations were manually curated. Gap-filling processes were performed to complete the synthesis of the main biomass components (DNA, RNA, and proteins) and essential cofactors (i.e., NAD$^+$/NADP$^+$ and CoA) by the *fastgapfill* function in PSAMM. Flux variation analysis (FVA) to optimize biomass production was performed using the *fva* function in PSAMM with IBM ILOG CPLEX Optimizer version 12.7.1.0.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The genome sequence of *Zhurongbacter thermophilus* 3DAC was reconstructed and deposited under accession number CP046447. Amplicon sequencing data have been deposited at SRA under accession numbers SRR14072822, SRR14072823, SRR14072817, SRR14072825, SRR14072824, SRR14072798, and SRR14072797, and transcriptome sequence data have been deposited at SRA under accession numbers SRR14072890 and SRR14072891. Source data are provided with this paper.

References

1. Lieph, R., Veloso, F. A. & Holmes, D. S. Thermophiles like hot T. *Trends Microbiol.* **14**, 423–426 (2006).
2. Hedlund, B. P. et al. Uncultivated thermophiles: current status and spotlight on 'Aigarchaeota'. *Curr. Opin. Microbiol.* **25**, 136–145 (2015).
3. Kato, S. et al. Isolation and characterization of a thermophilic sulfur- and iron-reducing thaumarchaeote from a terrestrial acidic hot spring. *ISME J.* **13**, 2465–2474 (2019).

4. Green-Saxena, A. et al. Active sulfur cycling by diverse mesophilic and thermophilic microorganisms in terrestrial mud volcanoes of Azerbaijan. *Environ. Microbiol.* **14**, 3271–3286 (2012).

5. Sievert, S. M. & Vetriani, C. Chemoautotrophy at deep-sea vents past, present, and future. *Oceanography* **25**, 218–233 (2012).

6. L'Haridon, S., Reysenbacht, A. L., Glénat, P., Prieur, D. & Jeanthon, C. Hot subterranean biosphere in a continental oil reservoir. *Nature* **377**, 223–224 (1995).

7. Etchebehere, C., Pavan, M. E., Zorzopulos, J., Soubes, M. & Muxi, L. *Coprothermobacter platensis* sp. nov., a new anaerobic proteolytic thermophilic bacterium isolated from an anaerobic mesophilic sludge. *Int. J. Syst. Bacteriol.* **48**(Pt 4), 1297–1304 (1998).

8. Lu, Z. & Lu, Y. *Methanocella conradii* sp. nov., a thermophilic, obligate hydrogenotrophic methanogen, isolated from Chinese rice field soil. *PLoS ONE* **7**, e35279 (2012).

9. Reysenbach, A. L. et al. A ubiquitous thermoacidophilic archaeon from deep-sea hydrothermal vents. *Nature* **442**, 444–447 (2006).

10. Mesbah, N. M., Hedrick, D. B., Peacock, A. D., Rohde, M. & Wiegel, J. *Natranaerobius thermophilus* gen. nov., sp. nov., a halophilic, alkalithermophilic bacterium from soda lakes of the Wadi An Natrun, Egypt, and proposal of Natranaerobiaceae fam. nov. and Natranaerobiales ord. nov. *Int. J. Syst. Evol. Microbiol.* **57**, 2507–2512 (2007).

11. Abramov, O. & Mojzsis, S. J. Microbial habitability of the Hadean Earth during the late heavy bombardment. *Nature* **459**, 419–422 (2009).

12. Sleep, N. H. The Hadean-Archaean environment. *Cold Spring Harb. Perspect. Biol.* **2**, a002527 (2010).

13. Jordan, S. F. et al. Promotion of protocell self-assembly from mixed amphiphiles at the origin of life. *Nat. Ecol. Evol.* **3**, 1705–1714 (2019).

14. Damer, B. & Deamer, D. The hot spring hypothesis for an origin of life. *Astrobiology* **20**, 429–452 (2020).

15. Dodd, M. S. et al. Evidence for early life in Earth's oldest hydrothermal vent precipitates. *Nature* **543**, 60–64 (2017).

16. Pollo, S. M., Zhaxybayeva, O. & Nesbo, C. L. Insights into thermoadaptation and the evolution of mesophily from the bacterial phylum Thermotogae. *Can. J. Microbiol.* **61**, 655–670 (2015).

17. Giulio, M. D. The ancestor of the Bacteria domain was a hyperthermophile. *J. Theor. Biol.* **224**, 277–283 (2003).

18. Giulio, M. D. The universal ancestor was a thermophile or a hyperthermophile: tests and further evidence. *J. Theor. Biol.* **221**, 425–436 (2003).

19. Battistuzzi, F. U. & Hedges, S. B. A major clade of prokaryotes with ancient adaptations to life on land. *Mol. Biol. Evol.* **26**, 335–343 (2009).

20. Weiss, M. C. et al. The physiology and habitat of the last universal common ancestor. *Nat. Microbiol.* **1**, 16116 (2016).

21. Cavalier-Smith, T. Rooting the tree of life by transition analyses. *Biol. Direct* **1**, 19 (2006).

22. Castelle, C. J. & Banfield, J. F. Major new microbial groups expand diversity and alter our understanding of the tree of life. *Cell* **172**, 1181–1197 (2018).

23. Zhu, Q. et al. Phylogenomics of 10,575 genomes reveals evolutionary proximity between domains Bacteria and Archaea. *Nat. Commun.* **10**, 5477 (2019).

24. Coleman, G. A. et al. A rooted phylogeny resolves early bacterial evolution. *Science* **372**, eabe0511 (2021).

25. Martin, W. F. & Sousa, F. L. Early microbial evolution: the age of Anaerobes. *Cold Spring Harbor. Perspect. Biol.* **8**, a018127 (2015).

26. Weiss, M. C., Preiner, M., Xavier, J. C., Zimorski, V. & Martin, W. F. The last universal common ancestor between ancient Earth chemistry and the onset of genetics. *PLoS Genet.* **14**, e1007518 (2018).

27. Sleep, N. H. Geological and geochemical constraints on the origin and evolution of life. *Astrobiology* **18**, 1199–1219 (2018).

28. Martin, W. F., Weiss, M. C., Neukirchen, S., Nelson-Sathi, S. & Sousa, F. L. Physiology, phylogeny, and LUCA. *Microb. Cell* **3**, 582–587 (2016).

29. Nelson, K. E. et al. Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature* **399**, 323–329 (1999).

30. Giovannelli, D. et al. Insight into the evolution of microbial metabolism from the deep-branching bacterium, *Thermovibrio ammonificans*. *eLife* **6**, e18990 (2017).

31. Kunath, B. J. et al. From proteins to polysaccharides: lifestyle and genetic evolution of *Coprothermobacter proteolyticus*. *ISME J.* **13**, 603–617 (2019).

32. Nayfach, S. et al. A genomic catalog of Earth's microbiomes. *Nat. Biotechnol.* **39**, 499–509 (2021).

33. Parks, D. H. et al. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat. Microbiol.* **2**, 1533–1542 (2017).

34. Hou, J. et al. Microbial succession during the transition from active to inactive stages of deep-sea hydrothermal vent sulfide chimneys. *Microbiome* **8**, 102 (2020).

35. Leng, H., Zhao, W. & Xiao, X. Cultivation and metabolic insights of an uncultured clade, Bacteroidetes VC2 .1 Bac22 (Candidatus Sulfidibacteriales ord. nov.), from deep-sea hydrothermal vents. *Environ. Microbiol.* **24**, 2484–2501 (2022).

36. Leigh, J. A., Albers, S. V., Atomi, H. & Allers, T. Model organisms for genetics in the domain Archaea: methanogens, halophiles, Thermococcales and Sulfolobales. *FEMS Microbiol. Rev.* **35**, 577–608 (2011).

37. Hungate, R. E. In *Methods in Microbiology* 3 (eds Norris, J. R. & Ribbons, D. W.) 117–132 (Academic Press, 1969).

38. Lever, M. A. et al. Life under extreme energy limitation: a synthesis of laboratory- and field-based investigations. *FEMS Microbiol. Rev.* **39**, 688–728 (2015).

39. Wu, C.-H., Schut, G. J., Poole, F. L., Haja, D. K. & Adams, M. W. W. Characterization of membrane-bound sulfane reductase: a missing link in the evolution of modern day respiratory complexes. *J. Biol. Chem.* **293**, 16687–16696 (2018).

40. Yu, H. et al. Structure of the respiratory MBS complex reveals iron-sulfur cluster catalyzed sulfane sulfur reduction in ancient life. *Nat. Commun.* **11**, 5953 (2020).

41. Lipscomb, G. L., Schut, G. J., Scott, R. A. & Adams, M. W. W. SurR is a master regulator of the primary electron flow pathways in the order Thermococcales. *Mol. Microbiol.* **104**, 869–881 (2017).

42. Nguyen, D. M. N. et al. Two functionally distinct NADP(+)-dependent ferredoxin oxidoreductases maintain the primary redox balance of *Pyrococcus furiosus*. *J. Biol. Chem.* **292**, 14603–14616 (2017).

43. Wu, C.-H., McTernan, P. M., Walter, M. E. & Adams, M. W. W. Production and application of a soluble hydrogenase from *Pyrococcus furiosus*. *Archaea* **2015**, 912582 (2015).

44. Mukherjee, S. et al. 1,003 reference genomes of bacterial and archaeal isolates expand coverage of the tree of life. *Nat. Biotechnol.* **35**, 676–683 (2017).

45. Hug, L. A. et al. A new view of the tree of life. *Nat. Microbiol.* **1**, 16048 (2016).

46. Wang, Y., Wegener, G., Hou, J., Wang, F. & Xiao, X. Expanding anaerobic alkane metabolism in the domain of Archaea. *Nat. Microbiol.* **4**, 595–602 (2019).

47. Pavan, M. E. et al. Proposal for a new classification of a deep branching bacterial phylogenetic lineage: transfer of *Coprothermobacter proteolyticus* and *Coprothermobacter platensis* to Coprothermobacteraceae fam. nov., within Coprothermobacterales ord. nov., Coprothermobacteria classis nov. and

Coprothermobacterota phyl. nov. and emended description of the family Thermodesulfobiaceae. *Int. J. Syst. Evol. Microbiol.* **68**, 1627–1632 (2018).

48. Kersters, I. et al. Isolation of *Coprothermobacter proteolyticus* from an anaerobic digest and further characterization of the species. *Syst. Appl. Microbiol.* **17**, 289–295 (1994).

49. Ollivier, B. M. et al. Emendation of the genus Thermobacteroides: *Thermobacteroides proteolyticus* sp. nov., a proteolytic acetogen from a methanogenic enrichment. *Int. J. Syst. Bacteriol.* **35**, 425–428 (1985).

50. Gagliano, M. C., Braguglia, C. M., Petruccioli, M. & Rossetti, S. Ecology and biotechnological potential of the thermophilic fermentative *Coprothermobacter* spp. *FEMS Microbiol. Ecol.* **91**, fiv018 (2015).

51. Hugenholtz, P., Skarshewski, A. & Parks, D. H. Genome-based microbial taxonomy coming of age. *Cold Spring Harb. Perspect. Biol.* **8**, a018085 (2016).

52. Yarza, P. et al. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat. Rev. Microbiol.* **12**, 635–645 (2014).

53. Parks, D. H. et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* **36**, 996–1004 (2018).

54. Mori, K., Yamaguchi, K., Sakiyama, Y., Urabe, T. & Suzuki, K.-i. *Caldisericum exile* gen. nov., sp. nov., an anaerobic, thermophilic, filamentous bacterium of a novel bacterial phylum, Caldiserica phyl. nov., originally called the candidate phylum OP5, and description of Caldisericaceae fam. nov., Caldisericales ord. nov. and Caldisericia classis nov. *Int. J. Syst. Evol. Microbiol.* **59**, 2894–2898 (2009).

55. Segata, N., Bornigen, D., Morgan, X. C. & Huttenhower, C. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nat. Commun.* **4**, 2304 (2013).

56. Nishida, H., Beppu, T. & Ueda, K. Whole-genome comparison clarifies close phylogenetic relationships between the phyla Dictyoglomi and Thermotogae. *Genomics* **98**, 370–375 (2011).

57. Bhandari, V. & Gupta, R. S. Molecular signatures for the phylum (class) Thermotogae and a proposal for its division into three orders (Thermotogales, Kosmotogales ord. nov. and Petrotogales ord. nov.) containing four families (Thermotogaceae, Fervidobacteriaceae fam. nov., Kosmotogaceae fam. nov. and Petrotogaceae fam. nov.) and a new genus Pseudothermotoga gen. nov. with five new combinations. *Antonie Van. Leeuwenhoek* **105**, 143–168 (2014).

58. Nesbo, C. L. et al. Genomic analysis of the mesophilic Thermotogae genus *Mesotoga* reveals phylogeographic structure and genomic determinants of its distinct metabolism. *Environ. Microbiol.* **21**, 456–470 (2019).

59. Mori, K., Kim, H., Kakegawa, T. & Hanada, S. A novel lineage of sulfate-reducing microorganisms: Thermodesulfobiaceae fam. nov., *Thermodesulfobium narugense*, gen. nov., sp. nov., a new thermophilic isolate from a hot spring. *Extremophiles* **7**, 283–290 (2003).

60. Zhang, W. & Lu, Z. Phylogenomic evaluation of members above the species level within the phylum Firmicutes based on conserved proteins. *Environ. Microbiol. Rep.* **7**, 273–281 (2015).

61. Kunisawa, T. Evolutionary relationships of completely sequenced Clostridia species and close relatives. *Int. J. Syst. Evol. Microbiol.* **65**, 4276–4283 (2015).

62. Ludwig, W., Schleifer, K.-H. & Whitman, W. B. in *Systematic Bacteriology* (eds Vos, P. D. et al.) Ch. Chapter 1, 1–13 (Springer New York, 2009).

63. Frolov, E. N., Lebedinsky, A. V., Elcheninov, A. G. & Kublanov, I. V. Taxonomic proposal for a deep branching bacterial phylogenetic lineage: transfer of the family Thermodesulfobiaceae to Thermodesulfobiales ord. nov., Thermodesulfobiia classis nov. and Thermodesulfobiota phyl. nov. *Syst. Appl. Microbiol.* **46**, 126388 (2023).

64. Nobu, M. K. et al. Unique H2-utilizing lithotrophy in serpentinite-hosted systems. *ISME J.* **17**, 95–104 (2022).

65. Mooers, A. Ø. & Holmes, E. C. The evolution of base composition and phylogenetic inference. *Trends Ecol. Evol.* **15**, 365–369 (2000).

66. Khachane, A. N., Timmis, K. N. & dos Santos, V. A. Uracil content of 16S rRNA of thermophilic and psychrophilic prokaryotes correlates inversely with their optimal growth temperatures. *Nucleic Acids Res.* **33**, 4016–4022 (2005).

67. Gribaldo, S. & Philippe, H. Ancient phylogenetic relationships. *Theor. Popul. Biol.* **61**, 391–408 (2002).

68. Waite, D. W. et al. 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 (2020).

69. Rinke, C. et al. Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**, 431–437 (2013).

70. Rosenberg, E. In *The Prokaryotes: Other Major Lineages of Bacteria and The Archaea* (eds Rosenberg, E. et al.) 613–615 (Springer Berlin Heidelberg, 2014).

71. Jumas-Bilak, E. & Marchandin, H. In *The Prokaryotes* (eds Rosenberg, E. et al.) Ch. Chapter 384, 931–954 (Springer Berlin Heidelberg, 2014).

72. Albuquerque, L. & da Costa, M. S. In *The Prokaryotes* (eds Rosenberg, E. et al.) Ch. Chapter 128, 955–987 (Springer Berlin Heidelberg, 2014).

73. Eveleigh, R. J., Meehan, C. J., Archibald, J. M. & Beiko, R. G. Being *Aquifex aeolicus*: untangling a hyperthermophile's checkered past. *Genome Biol. Evol.* **5**, 2478–2497 (2013).

74. Boussau, B., Guéguen, L. & Gouy, M. Accounting for horizontal gene transfers explains conflicting hypotheses regarding the position of aquificales in the phylogeny of bacteria. *BMC Evol. Biol.* **8**, 272 (2008).

75. Xavier, J. C. et al. The metabolic network of the last bacterial common ancestor. *Commun. Biol.* **4**, 413 (2021).

76. Bertoldo, C. & Antranikian, G. In *The Prokaryotes* (eds Dworkin, M. et al.) Ch. Chapter 5, 69–81 (Springer New York, 2006).

77. Reysenbach, A.-L. & Shock, E. Merging genomes with geochemistry in hydrothermal ecosystems. *Science* **296**, 1077–1082 (2002).

78. Reysenbach, A.-L. & Cady, S. L. Microbiology of ancient and modern hydrothermal systems. *Trends Microbiol.* **9**, 79–86 (2001).

79. Vetriani, C., Speck, M. D., Ellor, S. V., Lutz, R. A. & Starovoytov, V. *Thermovibrio ammonificans* sp. nov., a thermophilic, chemolithotrophic, nitrate-ammonifying bacterium from deep-sea hydrothermal vents. *Int. J. Syst. Evol. Microbiol* **54**, 175–181 (2004).

80. Widdel, F., Boetius, A. & Rabus, R. In *The Prokaryotes* (eds Dworkin, M. et al.) Ch. Chapter 33, 1028–1049 (Springer New York, 2006).

81. Zeng, X. et al. Pyrococcus CH1, an obligate piezophilic hyperthermophile: extending the upper pressure-temperature limits for life. *ISME J.* **3**, 873–876 (2009).

82. Patel, A. et al. Virus and prokaryote enumeration from planktonic aquatic environments by epifluorescence microscopy with SYBR Green I. *Nat. Protoc.* **2**, 269–276 (2007).

83. Peng, P. et al. Concurrent haloalkanoate degradation and chlorate reduction by *Pseudomonas chloritidismutans* AW-1T. *Appl. Environ. Microbiol.* **83**, e00325–00317 (2017).

84. Balch, W. E., Fox, G. E., Magrum, L. J., Woese, C. R. & Wolfe, R. S. Methanogens: reevaluation of a unique biological group. *Microbiol Rev.* **43**, 260–296 (1979).

85. Wolin, E. A., Wolin, M. J. & Wolfe, R. S. Formation of methane by bacterial extracts. *J. Biol. Chem.* **238**, 2882–2886 (1963).

86. Reese, B. K., Finneran, D. W., Mills, H. J., Zhu, M. X. & Morse, J. W. Examination and refinement of the determination of aqueous hydrogen sulfide by the methylene blue method. *Aquat. Geochem.* **17**, 567–582 (2011).

87. Natarajan, V. P., Zhang, X., Morono, Y., Inagaki, F. & Wang, F. A modified SDS-based DNA extraction method for high quality environmental DNA from seafloor environments. *Front. Microbiol.* **7**, 986 (2016).

88. Chin, C. S. et al. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Methods* **13**, 1050–1054 (2016).

89. Halpern, A. L. et al. Consensus generation and variant detection by Celera Assembler. *Bioinformatics* **24**, 1035–1040 (2008).

90. DePristo, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).

91. Delcher, A. L., Bratke, K. A., Powers, E. C. & Salzberg, S. L. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* **23**, 673–679 (2007).

92. Lagesen, K. et al. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* **35**, 3100–3108 (2007).

93. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955–964 (1997).

94. Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* **428**, 726–731 (2016).

95. Huerta-Cepas, J. et al. Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. *Mol. Biol. Evol.* **34**, 2115–2122 (2017).

96. Caporaso, J. G. et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl Acad. Sci. USA* **108**, 4516–4522 (2011).

97. Caporaso, J. G. et al. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).

98. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

99. Kopylova, E., Noe, L. & Touzet, H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**, 3211–3217 (2012).

100. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

101. Wagner, G. P., Kin, K. & Lynch, V. J. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci.* **131**, 281–285 (2012).

102. Ye, J. et al. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinforma.* **13**, 134 (2012).

103. Pruesse, E., Peplies, J. & Glockner, F. O. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **28**, 1823–1829 (2012).

104. Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).

105. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

106. Capella-Gutierrez, S., Silla-Martinez, J. M. & Gabaldon, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).

107. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259 (2019).

108. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinforma.* **11**, 119 (2010).

109. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).

110. Szollosi, G. J., Rosikiewicz, W., Boussau, B., Tannier, E. & Daubin, V. Efficient exploration of the space of reconciled gene trees. *Syst. Biol.* **62**, 901–912 (2013).

111. Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M. & Tanabe, M. KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res.* **49**, D545–D551 (2021).

112. King, Z. A. et al. BiGG Models: a platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res.* **44**, D515–D522 (2016).

113. Schellenberger, J., Park, J. O., Conrad, T. M. & Palsson, B. O. BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. *BMC Bioinforma.* **11**, 213 (2010).

114. Seaver, S. M. D. et al. The ModelSEED Biochemistry Database for the integration of metabolic annotations and the reconstruction, comparison and analysis of metabolic models for plants, fungi and microbes. *Nucleic Acids Res.* **49**, D575–D588 (2021).

115. Saier, M. H. Jr et al. The Transporter Classification Database (TCDB): 2021 update. *Nucleic Acids Res.* **49**, D461–D467 (2020).

116. Zhang, Y. et al. Three-dimensional structural view of the central metabolic network of *Thermotoga maritima*. *Science* **325**, 1544–1549 (2009).

117. Lerman, J. A. et al. In silico method for modelling metabolism and gene product expression at genome scale. *Nat. Commun.* **3**, 929 (2012).

118. Orth, J. D. et al. A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism—2011. *Mol. Syst. Biol.* **7**, 535 (2011).

119. Neidhardt, F. C., Ingraham, J. L. & Schaechter, M. *Physiology of the Bacterial Cell; A Molecular Approach.* (Sinauer Associates, 1990).

120. Martin, W., Baross, J., Kelley, D. & Russell, M. J. Hydrothermal vents and the origin of life. *Nat. Rev. Microbiol.* **6**, 805–814 (2008).

121. Cario, A., Lormieres, F., Xiang, X. & Oger, P. High hydrostatic pressure increases amino acid requirements in the piezo-hyperthermophilic archaeon *Thermococcus barophilus*. *Res. Microbiol.* **166**, 710–716 (2015).

122. Price, M. N. et al. Filling gaps in bacterial amino acid biosynthesis pathways with high-throughput genetics. *PLoS Genet.* **14**, e1007147 (2018).

123. Zhao, W. et al. Proteome-wide 3D structure prediction provides insights into the ancestral metabolism of ancient archaea and bacteria. *Nat. Commun.* **13**, 7861 (2022).

124. Steffensen, J. L., Dufault-Thompson, K. & Zhang, Y. PSAMM: a portable system for the analysis of metabolic models. *PLoS Comput. Biol.* **12**, e1004732 (2016).

Acknowledgements

Author contributions

X.X. designed and supervised the project. X.X. and S.M.S. collected the chimney samples. H.L. carried out the experiments. Y.W.

performed the phylogenetic and evolutionary analyses with input from H.L. W.Z. curated the genome annotations, and performed the metabolic model constructions and simulations. Y.W., H.L., and W.Z. analyzed the data. Y.W., H.L., W.Z., and X.X. wrote the manuscript. S.M.S. revised the manuscript. All authors contributed to the final version of the manuscript and approved the manuscript and supporting information.

Competing interests

The authors declare no competing interests.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.