Supplementary Information

for

Branched-chain amino acid specialization drove diversification within *Calditenuaceae* (*Caldarchaeia*) and enables their cultivation

Dengxun Lai[1*], Damon Mosier[2,3], Marike Palmer[1,4], Xavier Mayali[5], Juliet Johnston[5], Walter Saldivar[2], Jonathan K. Covington[1], Jian-Yu Jiao[6], Ranjani Murali[1], Cale O. Seymour[1], Lan Liu[6], Zheng-Shuang Hua[7], Wen-Jun Li[6,8], Peter K. Weber[5], Jennifer Pett-Ridge[5], Daniel R. Colman[9], Eric S. Boyd[9], Takuro Nunoura[10], Jeremy A. Dodsworth[2*] & Brian P. Hedlund[1,11*]

*Correspondence to Dengxun Lai (laid.omics@gmail.com), Jeremy A. Dodsworth (jdodsworth@csusb.edu), and Brian P. Hedlund (brian.hedlund@unlv.edu)

List of Supplementary Notes

Supplementary Note 1. Catalytic types of peptidases and amino acids composition in *Caditenuaceae* 3

Supplementary Note 2. Names proposed for type genomes under SeqCode ...**Error! Bookmark not defined.**3

Supplementary Note 3. Additional interpretations about shared orthogroups, taxonomy, metabolism, and evolution of *Caldienuaceae* 7

List of Supplementary Figures

Supplementary Fig. 1. Overview of sample preparation steps for FISH-nanoSIMS analysis. .. 10

Supplementary Fig. 2. Quantification of ^{13}C and ^{15}N incorporation in single cells from GBS sediment incubated in situ with isotopically labeled substrates, as measured by nanoSIMS. 11

Supplementary Fig. 3. Example of *Cal. ramacidaminiphagus* cells incorporating propionate . 12

Supplementary Fig. 4. Expression of BCAA transporter genes in *Cal. ramacidaminiphagus* cells 13

Supplementary Fig. 5. Boxplots depicting the proportion of amino acids in protein-coding sequences within high-quality genomes included in the study 14

Supplementary Fig. 6. Isoprenoid biosynthesis pathways in *Cal. ramacidaminiphagus* LZ^{Ts} .. 15

Supplementary Fig. 7. TCA cycle in *Cal. ramacidaminiphagus* LZ^{Ts} 16

Supplementary Fig. 8. Similarities of the domains constituting CrOFORs from *Cal. ramacidaminiphagus* LZ^{Ts} and StOFOR2 from *S. tokodaii*. 17

Supplementary Fig. 9. OFOR computed structure model (CSMs)..... 18

Supplementary Fig. 10. Identification of *Calditenuis* 16S rRNA gene sequences..... 19

Supplementary Fig. 11. Robust species phylogeny of *Calditenuaceae*..... 20

Supplementary Fig. 12. Maximum-likelihood tree of *Caldarchaeales* 16S rRNA gene sequences 21

Supplementary Fig. 13. Densitree plot showing genealogical concordance of 34 shared proteins from ar53 marker protein sets 22

Supplementary Fig. 14. Cladograms of BCAA transporter proteins: (a) K01998-LivM, (b) K01997-LivH, (c) K01996-LivF, (d) K01995-LivG 23

Supplementary Fig. 15. Maximum-likelihood phylogenetic tree of PAA transporter proteins: (a) K02028-ABC.PA.A, (b) K02029-ABC.PA.P, (c) K02030-ABC.PA.S. 24

Supplementary Fig. 16. Catalytic types of peptidases in protein-coding sequences in high-quality *Calditenuaceae* genomes 25

Supplementary Fig. 17. Non-metric Multidimensional Scaling (NMDS) analysis of amino acid composition in protein-coding sequences in high-quality *Calditenuaceae* (a) and all the high-quality genomes included in the study (b) 26

Supplementary Fig. 18. Shared orthogroups in the *Calditenuaceae* family 27

Supplementary Fig. 19. Ancestral gene content reconstruction in the *Calditenuaceae* family .. 28

Supplementary Fig. 20. Principal Coordinates Analyses (PCoA) of *Calditenuaceae* genomes 29

Supplementary Fig. 21. Carbon fixation pathways in *Cal. ramacidaminiphagus* LZ^{Ts} 30

Supplementary Fig. 22. Calvin-Benson cycle in *Cal. ramacidaminiphagus* LZ^{Ts}. 31

Supplementary Fig. 23. Structural characterization of the alpha subunit of nitrite oxidoreductase (NxrA). 32

References 33

Supplementary Note 1. Catalytic types of peptidases and amino acids composition in *Calditenuaceae*

Considering the specialization of *Cal. ramacidaminiphagus* for BCAAs, we examined whether the distribution of catalytic peptidase types differs within *Calditenuaceae*. Our analysis revealed notable differences in the catalytic types of peptidases in *Calditenuis* compared to other genera within the family *Calditenuaceae* (Supplementary Fig. 16). Specifically, metallo, cysteine, threonine, and aspartic peptidases were enriched in *Calditenuis*. Since all but one of these peptidases lack signal peptides or membrane anchors, some of them likely mediate the digestion of oligopeptides, releasing BCAAs. Despite this enrichment, the total number of peptidases in *Calditenuaceae* was lower than *Caldarchaeaceae* and *Nitrososphaerales*, potentially reflecting differences in protein degradation strategies. Additionally, although *Calditenuis* did not encode transporters for PAAs and lacked genes encoding substrate-binding and ATP-binding proteins for neutral amino acids, these amino acids could enter the cells via peptide transporters and be liberated by cytoplasmic peptidases (Supplementary Table 14).

Next, we examined amino acid compositions in *Calditenuaceae* (Supplementary Fig. 5 and Supplementary Fig. 17; Supplementary Table 12; Source Data 2). Interestingly, while valine was enriched in *Calditenuis*, isoleucine was depleted. This suggests that *Calditenuis* may have a higher anabolic demand for valine, possibly indicating that other BCAAs might be used catabolically more than valine. To further elucidate this, additional experiments could test the growth of *Calditenuis* on each individual BCAA to better understand its specific requirements and preferences. In addition to BCAAs, other amino acids such as glutamic acid, arginine, glycine, alanine, and proline were enriched in *Calditenuis*.

Supplementary Note 2. Names proposed for type genomes under SeqCode

Description of the class *Caldarchaeia*

Cald.ar.chae'i.a **L. neut. n.** *Caldarchaeum*, archaea from hot environments; -ia, ending to denote a class; **N.L. neut. pl. n.** *Caldarchaeia*, the *Caldarchaeum* class; The type genus is *Caldarchaeum*.

Description of *Calditenuis ramacidaminiphagus*

ram.a.cid.a.mi.ni.pha'gus **L. masc. n.** *ramus*, branch; **N.L. neut. adj.** *acidum aminum*, amino acid; **N.L. masc. suff.** *-phagus*, eater; **N.L. masc. adj.** *ramacidaminiphagus*, a branched-chain amino acid eater.

The nomenclatural type genome is LZ^{Ts}, recovered from Great Boiling Spring, Gerlach, Nevada, USA. The type genome size is 1,506,500 bp (estimated size 1,536,304 bp) with G+C content of 62.0%. CheckM-based quality assessment indicates a completeness of 98.1% and a contamination estimate of 0%. In addition to amino acids, propionate was the second most incorporated substrate.

The type genome is available under GenBank Assembly accession XXXXXXX, BioSample XXXXX, SRA run accession XXXXX.

Description of *Calditenuis aerorheumatis*

Name was originally assigned as *Calditenuis aerorheumensis* in Beam *et al.*, 2016[1], but incorrectly spelled. Oren *et al.*, 2020[2] corrected it to "aerorheumatis".

a.e.ro.rheu'ma.tis. **Gr. masc. n.** *aer*, air; **Gr. neut. n.** *rheuma*, a flow, a current; **N.L. gen. n.** *aerorheumatis*, of an air flow

The species name was proposed based on a medium-quality genome; we propose to assign the genome GC11.bin.20 as the nomenclatural type genome. GC11.bin.20 was recovered from 'Bone Pool' in the Geyser Creek Geyser Basin, Yellowstone National Park, Wyoming, USA. The binned genome size of the type genome is 1,232,858 bp (estimated size 1,333,169 bp) with a G+C content of 61.6%. CheckM-based quality assessment indicates a completeness of 92.5% and a contamination estimate of 0%. The type genome is available through the Integrated Microbial Genome (IMG) database under ID 3300045882_10020.

Description of the genus *Caldaricola*

Cal.da.ri'co.la **L. masc. adj.** *caldarius*, hot; **L. masc. suff.** -*cola*, inhabitant; **N.L. masc. n.** *Caldaricola*, inhabitant of hot environments.

The type species is *Caldarincola sungkaiensis*.

Description of *Caldaricola sungkaiensis*

sung.ka.i.en'sis **N.L. masc. adj.** *sungkaiensis*, of Sungkai.

The nomenclatural type genome is SKYB10. This genome was recovered from the SKY hot spring, a brown biofilm from an alkaline hot spring in Perak, Sungkai, Malaysia. The genome was reported in Liew *et al.*, 2022[3], although that study did not focus on the species and did not provide a name. The binned genome size of the type is 1,166,430 bp (estimated size 1,271,313 bp) with a G+C content of 54.0%. CheckM-based quality assessment indicates a completeness of 91.8% and a contamination estimate of 0%. The type genome is available under GenBank Assembly accession GCA_025059745.1, BioSample PRJNA761511.

Description of the genus *Kaftiomonas*

Kaf.ti.o.mo'nas **Gr. masc. adj.** *kaftós*, very hot or scalding; **L. fem. n.** *monas*, a monad; **N.L. fem. n.** *Kaftiomonas*, a monad from very hot or scalding water.

The type species is *Kaftiomonas caldifontinalis*.

Description of *Kaftiomonas caldifontinalis*

cal.di.fon.ti.na'lis **L. masc. adj.** *caldus*, hot; **L. fem. adj.** *fontinalis*, spring or fountain; **N.L. fem. adj.** *caldifontinalis*, of a hot spring.

The nomenclatural type genome is S04.bin.80, recovered from Sylvan Springs (SYL), Geyser Basin, Yellowstone National Park, Wyoming, USA. The binned genome size of the type is 1,223,960 bp (estimated size 1,235,960 bp) with a G+C content of 48.4%. CheckM-based quality assessment indicates a completeness of 99.03% and a contamination estimate of 0%. The type genome is available through the IMG database under ID 3300045871_10080.

Description of *Kaftiomonas geyserifluvii*

gey.se.ri.flu'vi.i **N.L. neut. n.** *geysirum*, geyser, from Icelandic n. *geysir*; **L. masc. n.** *fluvius*, river or creek; **N.L. gen. n.** *geyserifluvii*, of Geyser Creek.

The nomenclatural type genome is GC01.bin.41, recovered from Unnamed Spring (GCR-01) in the Geyser Creek Geyser Basin, Yellowstone National Park, Wyoming, USA. The binned genome size of the type is 1,449,412 bp (estimated size 1,547,041 bp) with a G+C content of 50.0%. CheckM-based quality assessment indicates a completeness of 93.7% and a contamination estimate of 0%. The type genome is available through the IMG database under ID 3300045887_10041.

Description of *Kaftiomonas yunnanensis*

yu.nan.nen'sis **N.L. fem. adj.** *yunnanensis*, of Yunnan

The nomenclatural type genome is DRTY-6.bins.153.201705, recovered from Diretiyanqu Spring 6 (DRTY-6), Rehai, Yunnan, China. The binned genome size of the type is 1,242,820 bp (estimated size 1,274,047 bp) with a G+C content of 42.6%. CheckM-based quality assessment indicates a completeness of 96.6% and a contamination estimate of 0.97%. The type genome is available under GenBank Assembly accession JBHHLG000000000, BioSample SAMN43776125, SRA run accession SRR30574052.

Description of the genus *Ardentivivens*

Ar.den.ti.vi'vens **L. masc. adj.** *ardens*, burning or scalding; **N.L. pres. part.** *vivens*, living in; **N.L. fem. n.** *Ardentivivens*, organisms living in scalding water. The type species is *Ardentivivens gerlachensis.*

Description of *Ardentivivens gerlachensis*

ger.lach.en'sis **N.L. fem. adj.** *gerlachensis*, referring to Gerlach, NV, USA.

The nomenclatural type genome is GBS85_8, recovered from Great Boiling Spring, Gerlach, Nevada, USA. The metagenome of the binned genome of the type was released in Thomas *et al.*, 2019[4]. The genome size is 1,489,319 bp (estimated size 1,539,129 bp) with a G+C content of 57.0%. CheckM-based quality assessment indicates a completeness of 96.7% and a contamination

estimate of 0%. The type genome is available under GenBank Assembly accession XXXXXXX, BioSample XXXXX, SRA run accession XXXXX.

Description of *Ardentivivens yellowstonensis*

yel.low.ston.en'sis. **N.L. fem. adj.** *yellowstonensis*, referring to Yellowstone.

The nomenclatural type genome is N02.bin.19, recovered from Perpetual Spouter (N-02), Norris Geyser Basin, Yellowstone National Park, Wyoming, USA. The binned genome size of the type is 1,784,888 bp (estimated size 1,863,278 bp) with a G+C content of 54.8%. CheckM-based quality assessment indicates a completeness of 95.8% and a contamination estimate of 0.97%. The type genome is available through the IMG database under ID 3300045867_10019.

Description of the genus *Candentenecus*

Can.dent.en.e'cus **L. pres. part.** *candens*, boiling or red-hot; **N.L. masc. n.** *enecus*, an inhabitant; **N.L. masc. n.** *Candentenecus*, inhabitants of boiling water. The type species is *Candentenecus fervidifontanae.*

Description of *Candentenecus fervidifontanae*

fer.vi.di.fon.ta'na.e **L. masc. adj.** *fervidus*, hot; **L. fem. n.** *fontana*, spring or fountain; **N.L. gen. n.** *fervidifontanae*, of a hot spring.

The nomenclatural type genome is Aig_Gxs02, recovered from Great Boiling Spring, Gerlach, Nevada, USA. The binned genome size of the type is 1,338,385 bp (estimated size 1,351,506 bp) with a G+C content of 52.8%. CheckM-based quality assessment indicates a completeness of 99.0% and a contamination estimate of 0%. The type genome is available under GenBank Assembly accession XXXXXXX, BioSample XXXXX, SRA run accession XXXXX.

Description of *Candentenecus caldifluvii*

cal.di.flu'vi.i **L. masc. adj.** *caldus*, hot; **L. masc. n.** *fluvius*, stream; **N.L. gen. n.** *caldifluvii*, of a hot stream.

The nomenclatural type genome is HR02, recovered from the enrichment culture of samples from a geothermal stream, a subsurface gold mine, Kanagawa, Yokosuka, Japan. The type genome is available under GenBank Assembly accession GCA_002898395.1, BioSample SAMD00093752, SRA run accession DRR101361, DRR101362, DRR101363, DRR101364, DRR101365, DRR101366, DRR101367.

Description of *Candentenecus silaniferventis*

si.la.ni.fer.ven'tis **L. masc. n.** *silanus*, jet of water or fountain; **L. part. adj.** *fervens*, hot; **N.L. gen. n.** *silaniferventis*, of a hot fountain or spring.

The nomenclatural type genome is QQ.bins.91.201705, recovered from Qiaoquan Spring, Rehai, Yunnan, China. The binned genome size of the type is 1,435,363 bp (estimated size 1,463,786 bp) with a G+C content of 52.1%. CheckM-based quality assessment indicates a completeness of 98.1% and a contamination estimate of 0%. The type genome is available under GenBank Assembly accession JBHHMF000000000, BioSample SAMN43776150, SRA run accession SRR30574051.

Supplementary Note 3. Additional interpretations about shared orthogroups, taxonomy, metabolism, and evolution of *Calditenuaceae*

We examined the shared orthogroups in the *Calditenuaceae* high-quality genomes. Pairwise comparisons of shared orthogroups were conducted using the Sørensen index to account for differences in genome size (Supplementary Fig. 18). The lowest Sørensen index observed within *Calditenuis* is 0.835, within *Kaftiomonas* is 0.822, within *Ardentivivens* is 0.803, and within *Candentenecus* is 0.847. In contrast, the highest Sørensen index between different genera is 0.778, except for *Ardentivivens*. Specifically, the highest Sørensen index between *A. gerlachensis* and *Candentenecus* is 0.893, and between *A. yellowstonensis* and *Candentenecus* is 0.822. These patterns can be explained by ancestral gene reconstruction analyses (Supplementary Fig. 19). For instance, the higher Sørensen index between *A. gerlachensis* and *Candentenecus* compared to that between *A. yellowstonensis* and *Candentenecus* is partly due to *A. yellowstonensis* losing 24 orthologous families (Supplementary Fig. 19, node 24), while *Ardentivivens gerlachensis* lost only 4 orthologous families (Supplementary Fig. 19, node 27). Additionally, significant gene gains and losses occurred in *A. yellowstonensis*, particularly in the MGB03.bin.5 genome, which lost 104 orthologous families and gained 35 new ones. This can explain the low Sørensen index observed between MGB03.bin.5 and the other *A. yellowstonensis* genomes (Supplementary Fig. 18 and Supplementary Fig. 19). The low Sørensen index of shared orthogroups among the same species, *A. yellowstonensis*, indicates high genomic diversity within the species. This diversity is likely due to substantial gene gain and loss events, suggesting that the species may be highly adaptable.

Consistent with the taxonomic classifications, principal coordinates analyses (PCoA) based on functional similarity using KO, COG, or arCOG annotations and amino acids composition revealed five genus-level clusters (Supplementary Fig. 20). Previous studies have shown divergent amino acid compositional patterns between thermophiles and mesophiles[5,6,7]. Our study further revealed that within thermophiles, amino acid composition differs at the genus level. *Ardentivivens* and *Candentenecus* were close, however, we separated them into two genera based on the following reasons. First, the AAI values between these two groups are below 65%. Second, GTDB based on relative evolutionary divergence (RED) values separated them into two genera. Third, they are sampled from different continents (Fig. 6). Additionally, although all several genera and all species of *Calditenuaceae* were found only on one continent, the lone exception is genome 146_bin.25 from Hawaii belonging to the same species as the other genomes from Yunnan, China,

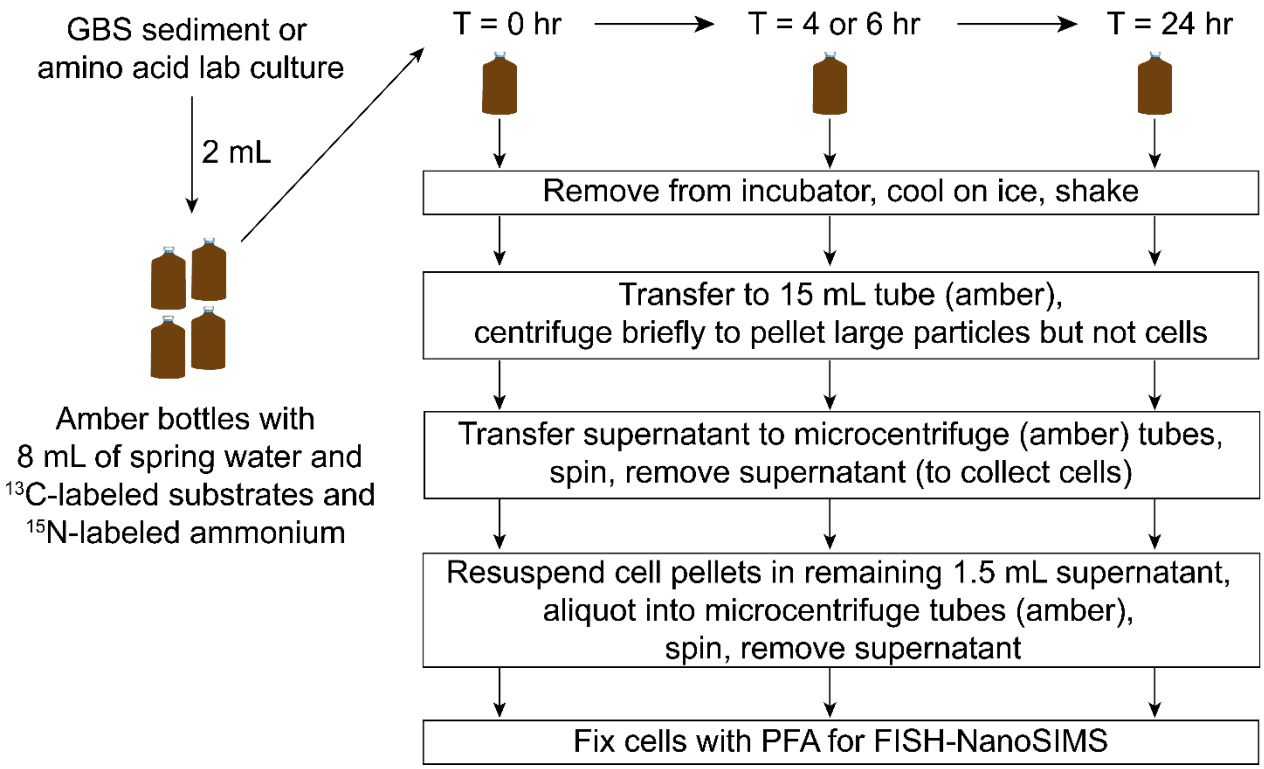
given the approximately 7500 km direct distance between them and the absence of any terrestrial connection. While the mechanisms of microbial dispersal between continents are not well understood, we propose that *C. fumarioli* may have been transported via volcanic ash. Microbial transportation in volcanic ash has been observed[8,9], as the ash can provide protection from UV radiation. Both Hawaii and Yunnan are volcanic regions. It is also possible they could have been transported from other regions that are poorly sampled, such as during the eruption of Krakatoa[10]. We have only one high-quality *C. fumarioli* genome from Hawaii, and obtaining more genomes from that area and the Pacific Ring of Fire could provide further insights.

Several carbon fixation pathways, including the 3-hydroxypropionate/4-hydroxybutyrate (HP/HB)[1], dicarboxylate/4-hydroxybutyrate (DC/HB)[11], Calvin-Benson-Cycle (CBC)[12], and reversed oxidative TCA (roTCA) pathway[13], have been proposed for *Nitrososphaeria*_A. To assess these pathways, we screened genes encoding enzymes involved in those common carbon fixation pathways in *Cal. ramacidaminiphagus* (Supplementary Fig. 21 and Supplementary Fig. 22). The results show that *Cal. ramacidaminiphagus* lacks several key enzymes for the HP/HB, DC/HB, and CBC pathways. Additionally, for the reductive TCA (rTCA) cycle, ATP citrate lyase and citryl-CoA synthase were not annotated. Instead, a gene encoding citrate synthase (K01647), which functions in roTCA, an alternative pathway to generate oxaloacetate, is present and expressed in *Cal. ramacidaminiphagus* (Supplementary Table 9 and Supplementary Table 14). Consistent with this, *Cal. ramacidaminiphagus* showed assimilation activity in bicarbonate incubation in lab cultures (Fig. 2). The low concentration of bicarbonate (0.5 μM) could explain the low atom % excess, as a high concentration of CO_2 is required for the roTCA cycle[14]. The gene encoding citrate synthase is present in all high-quality *Nitrososphaeria*_A genomes (Supplementary Table 14), indicating that the roTCA is a conserved feature within this group.

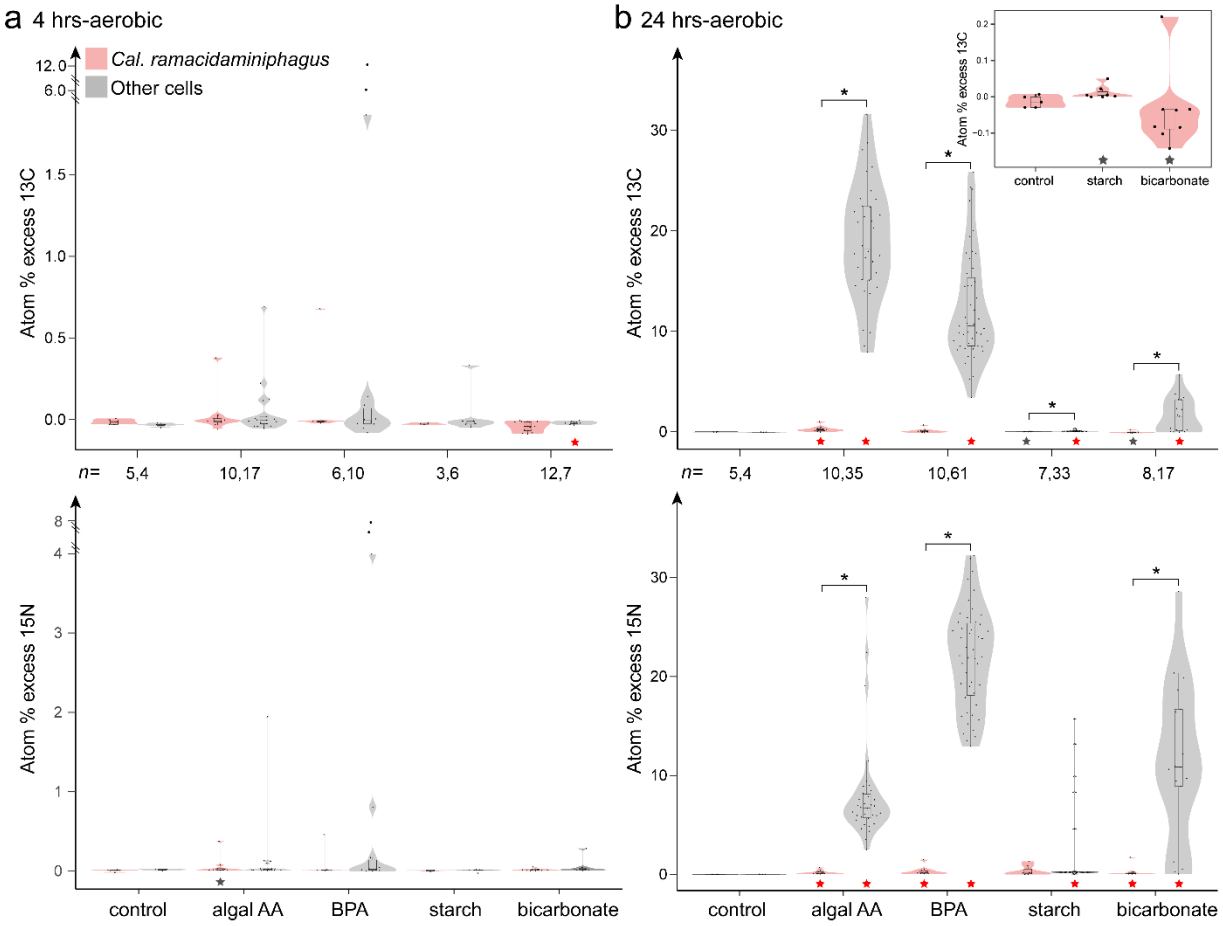
No hydrogenases were detected in *Calditenuis* genomes (Supplementary Table 30), consistent with previous research showing that a *Calditenuis* genome (Aig_JZ_bin_19) from a hot spring in China lacks hydrogenases[13]. This suggests that *Calditenuis* does not rely on hydrogen metabolism for energy conservation and instead depends on alternative redox processes. In contrast, the genera *Ardentivivens* and *Kaftimonas*, as well as the species *Candentenecus fervidifontanae*, possess genes encoding either group 3abd or group 3c Ni-Fe hydrogenases. These hydrogenases may play a role in maintaining redox balance by supplying intracellular reducing equivalents[15]. Aerobic carbon-monoxide dehydrogenases (CoxS, CoxM, and CoxL) were found throughout the family *Calditenuaceae*, except for *Kaftimonas*. Several members of *Calditenuis* possess genes encoding nitric oxide reductase (nNOR and gNOR[16]) and nitrous oxide reductase (NosZ), suggesting that nitric oxide and nitrous oxide could serve as electron acceptors, potentially coupling with the oxidation of BCAAs. *Benthortus lauensis* (family *Wolframiiraptoraceae*) and several members of *Caldarchaeaceae* possess genes encoding nitrite oxidoreductase (NxrA and NxrB), with all NxrA containing conserved residues for Tat signal peptides, Fe-S cluster coordination, molybdenum binding, and nitrite/nitrate interaction sites (Supplementary Fig. 23). According to METABOLIC

results, members of the family *Calditenuaceae* do not appear to be highly involved in sulfur cycling, except for *Kaftimonas*, which possesses genes encoding sulfur dioxygenase (SDO). While KEGG annotations from eggnog-mapper v.2[17] indicate that most members of the class *Caldarchaeia* have sulfide:quinone oxidoreductase (SQR) (Supplementary Table 14), a closer inspection of protein domains revealed that *Cal. ramacidaminiphagus* LZ^{Ts} lacks a flavin adenine dinucleotide (FAD)-binding domain[18], which is essential for its catalytic activity.

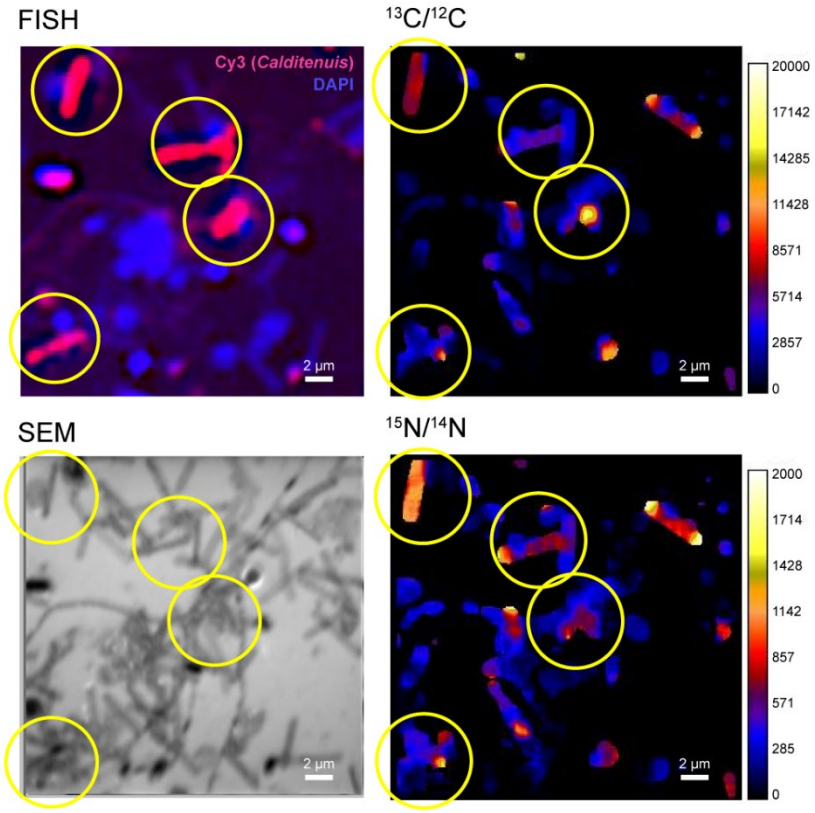
Supplementary Fig. 1. Overview of sample preparation steps for FISH-nanoSIMS analysis. This flowchart outlines the key steps involved in preparing GBS sediment samples and lab culture samples for FISH-nanoSIMS analysis. The process includes 4 and 24 hours for GBS sediment samples, and 6 and 24 hours for lab culture samples.



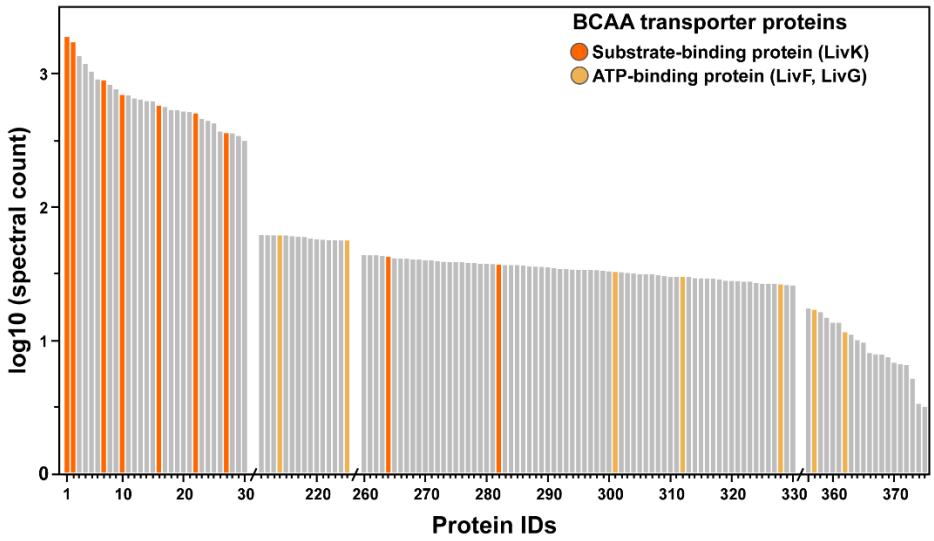
Supplementary Fig. 2. Quantification of 13C and 15N incorporation in single cells from GBS sediment incubated in situ with isotopically labeled substrates, as measured by nanoSIMS. Hybrid violin/box plots showing isotopic enrichments of *Cal. ramacidaminiphagus* (red) and other cells (grey). The violin plot shows the density distribution of all data points, whereas the box plot shows the median (central line) and 25th and 75th percentiles (box edges). Each point reflects the enrichment of cellular ^{13}C-labeled substrates or ^{15}N-ammonium in a single cell. The number of independent cells in each group is denoted by n. Statistical significance was calculated using the Mann-Whitney U-test (*, $p < 0.01$). A five-star symbol indicates significant differences between unlabeled controls and *Cal. ramacidaminiphagus* or other cells (red, $p < 0.01$; grey, $p < 0.05$) (Supplementary table 3). Abbreviations: BPA-butyrate, propionate and acetate; algal AA-algal amino acids.



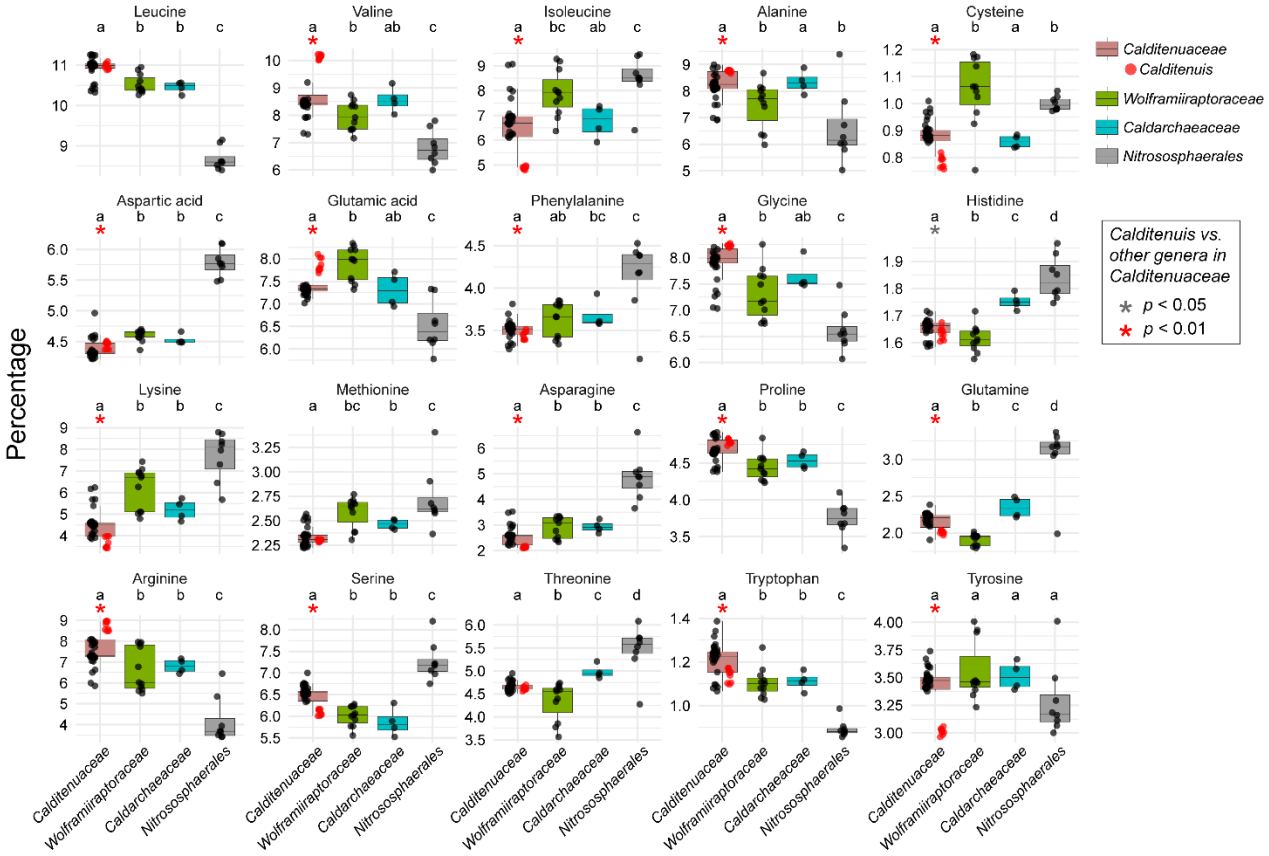
Supplementary Fig. 3. Example of *Cal. ramacidaminiphagus* cells incorporating propionate. CARD-FISH imaging (upper left) of *Cal. ramacidaminiphagus* (Cy3, marked with a yellow circle), scanning electron microscope (SEM) image (lower left), and nanoSIMS ion ratio images reflecting ^{13}C assimilation (upper right) and ^{15}N assimilation (lower right) of representative *Cal. ramacidaminiphagus* cells and other cells for propionate.



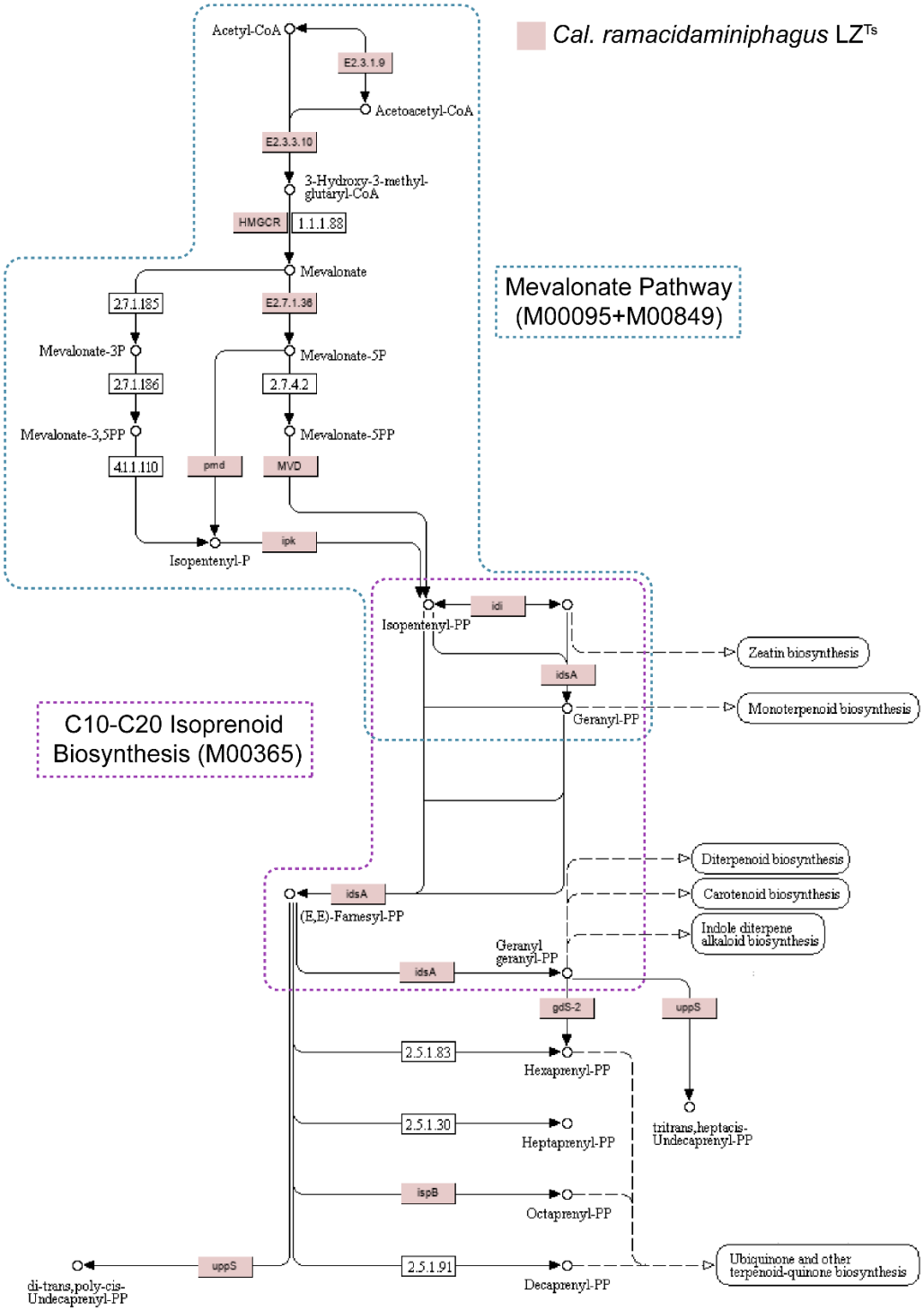
Supplementary Fig. 4. Expression of BCAA transporter genes in *Cal. ramacidaminiphagus* cells. Proteins were identified using metaproteomics from the G-10 culture after three weeks of incubation at 80°C. Proteins are arranged by abundance with peptides matching BCAA transporter proteins highlighted in color (Supplementary table 9).



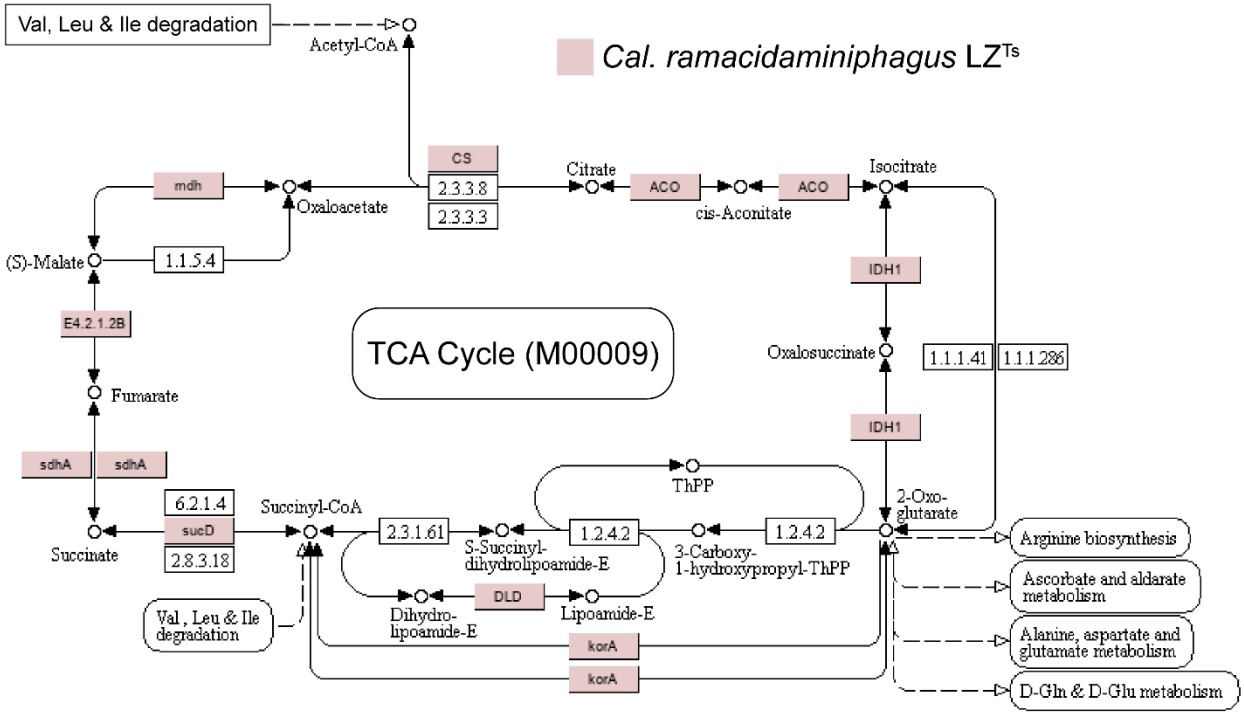
Supplementary Fig. 5. Boxplots depicting the proportion of amino acids in protein-coding sequences within high-quality genomes included in the study. Different groups are color-coded, and significant differences are indicated by different lowercase letters ($p < 0.05$; Supplementary Table 15). Significant differences between *Calditenuis* and other genera in *Calditenuaceae* are marked by asterisks: grey asterisks denote $p < 0.05$, and red asterisks denote $p < 0.01$ (Supplementary Table 12).



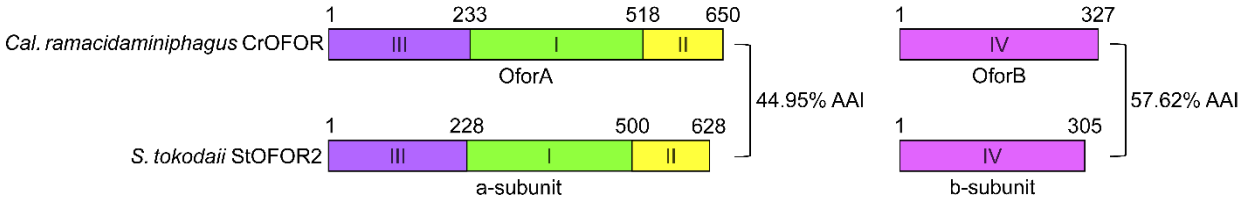
Supplementary Fig. 6. Isoprenoid biosynthesis pathways in *Cal. ramacidaminiphagus* LZ^{Ts}. The dashed green box outlines the mevalonate pathway (KEGG: M00095+M00849), and the dashed purple box outlines the C10-C20 isoprenoid biosynthesis pathway (KEGG: M00365). Genes with a red background indicate their presence of genes in LZ^{Ts}.



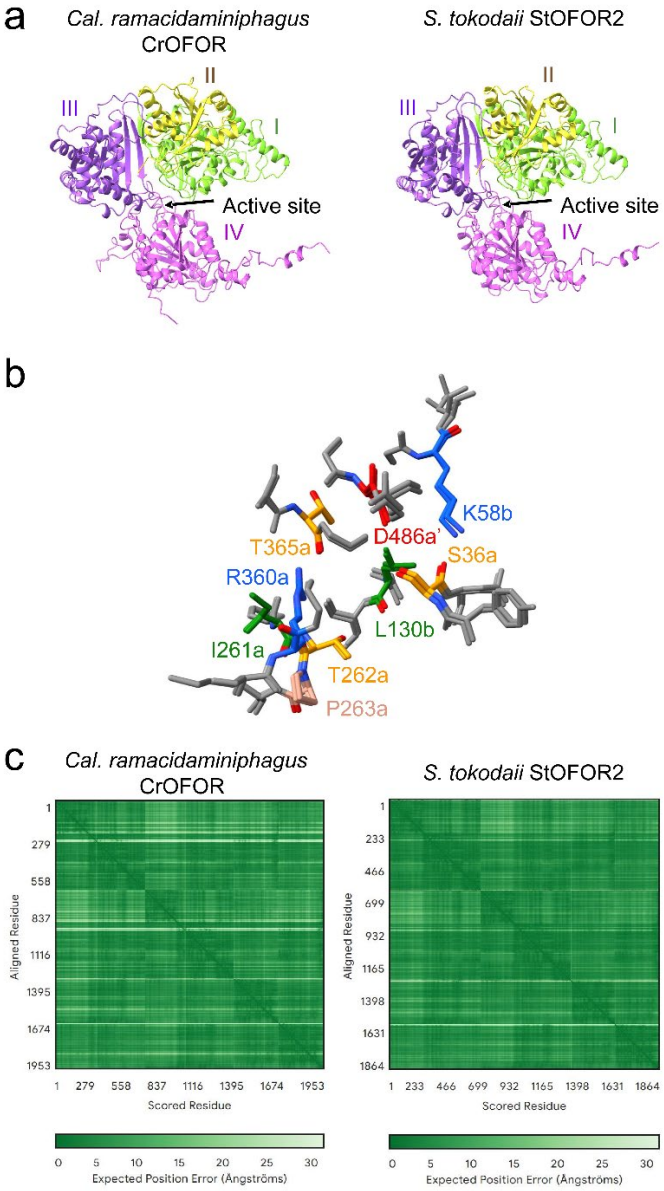
Supplementary Fig. 7. TCA cycle in *Cal. ramacidaminiphagus* LZ^{Ts}. The map shows genes involved in the TCA cycle (KEGG: M00009), with a red background indicating their presence in LZ^{Ts}.



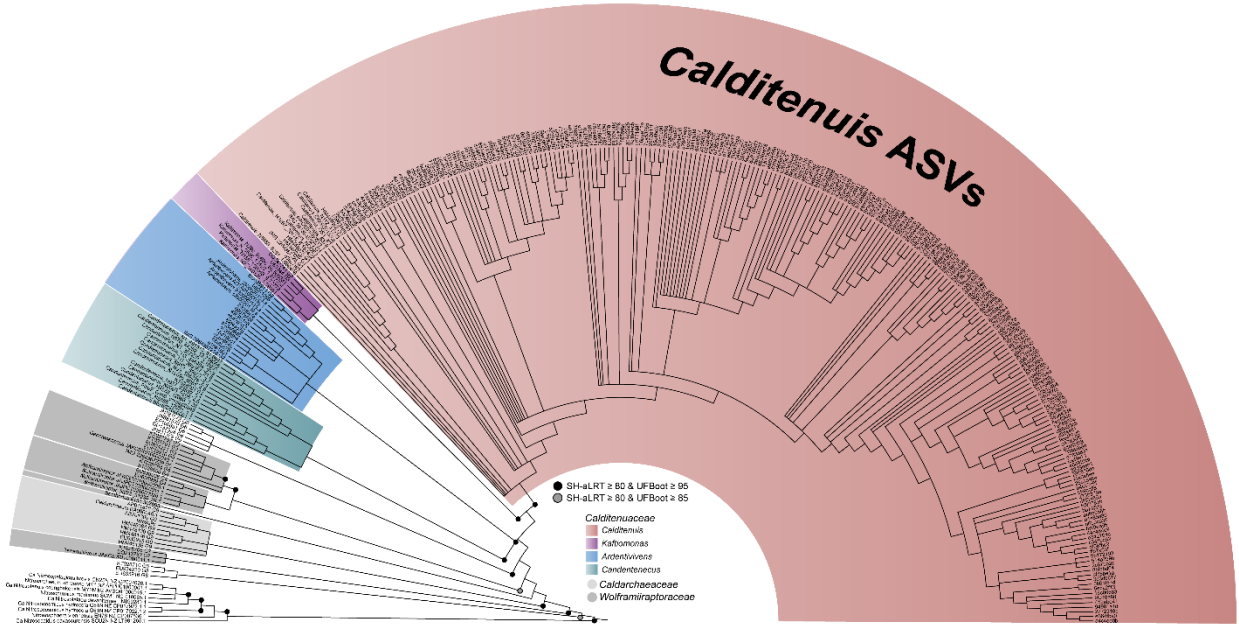
Supplementary Fig. 8. Similarities of the domains constituting CrOFORs from *Cal. ramacidaminiphagus* LZ^{Ts} and StOFOR2 from *S. tokodaii*. Schematics of the amino acid sequences of a- and b-subunits from *Cal. ramacidaminiphagus* CrOFOR and *S. tokodaii* StOFOR2. The start and end residues of Domains I (green), II (yellow), III (purple), and IV (magenta) are shown above each map, along with the AAI of each subunit to the right.



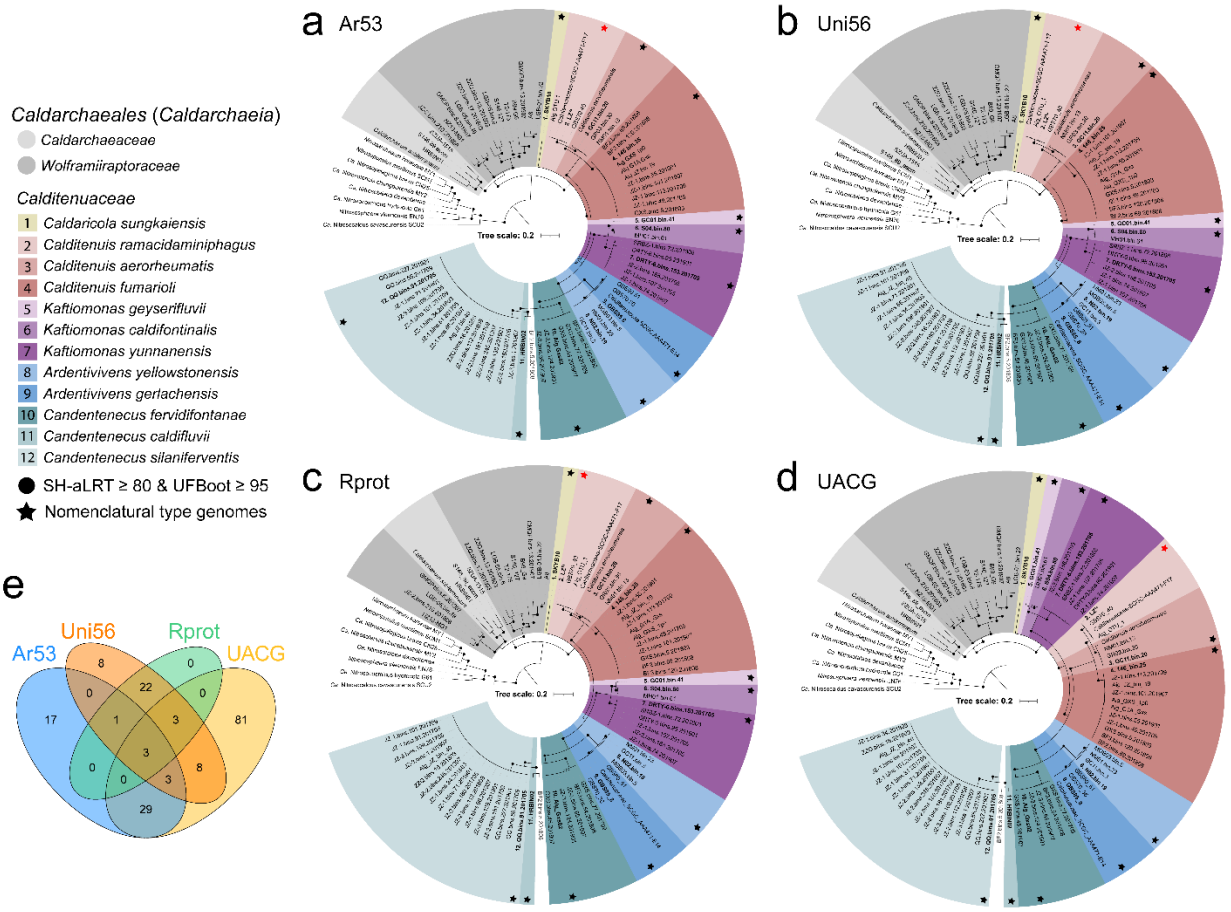
Supplementary Fig. 9. OFOR computed structure model (CSMs). CSMs of *Cal. ramacidaminiphagus* CrOFOR and *S. tokodaii* StOFOR2 generated using AlphaFold 3. Scores for interface predicted template modeling (TM) and predicted TM for CrOFOR were 0.94 and 0.95, respectively, and for StOFOR2 were 0.95 and 0.96, respectively, indicating accurate predictions. **(A)** Heterodimers of the *Cal. ramacidaminiphagus* CrOFOR and *S. tokodaii* StOFOR2 with highlighted Domains I (green), II (yellow), and III (purple) on the alpha subunits, and Domain IV (magenta) on the beta subunits. The active sites are indicated by black arrows. **(B)** Active sites of *Cal. ramacidaminiphagus* CrOFOR and *S. tokodaii* StOFOR2 overlaid with key residues of *Cal. ramacidaminiphagus* CrOFOR colored and labeled. Adjacent amino acids (gray) are shown for structural context. **(C)** PAE plots of homodimeric *Cal. ramacidaminiphagus* CrOFOR and *S. tokodaii* StOFOR2.



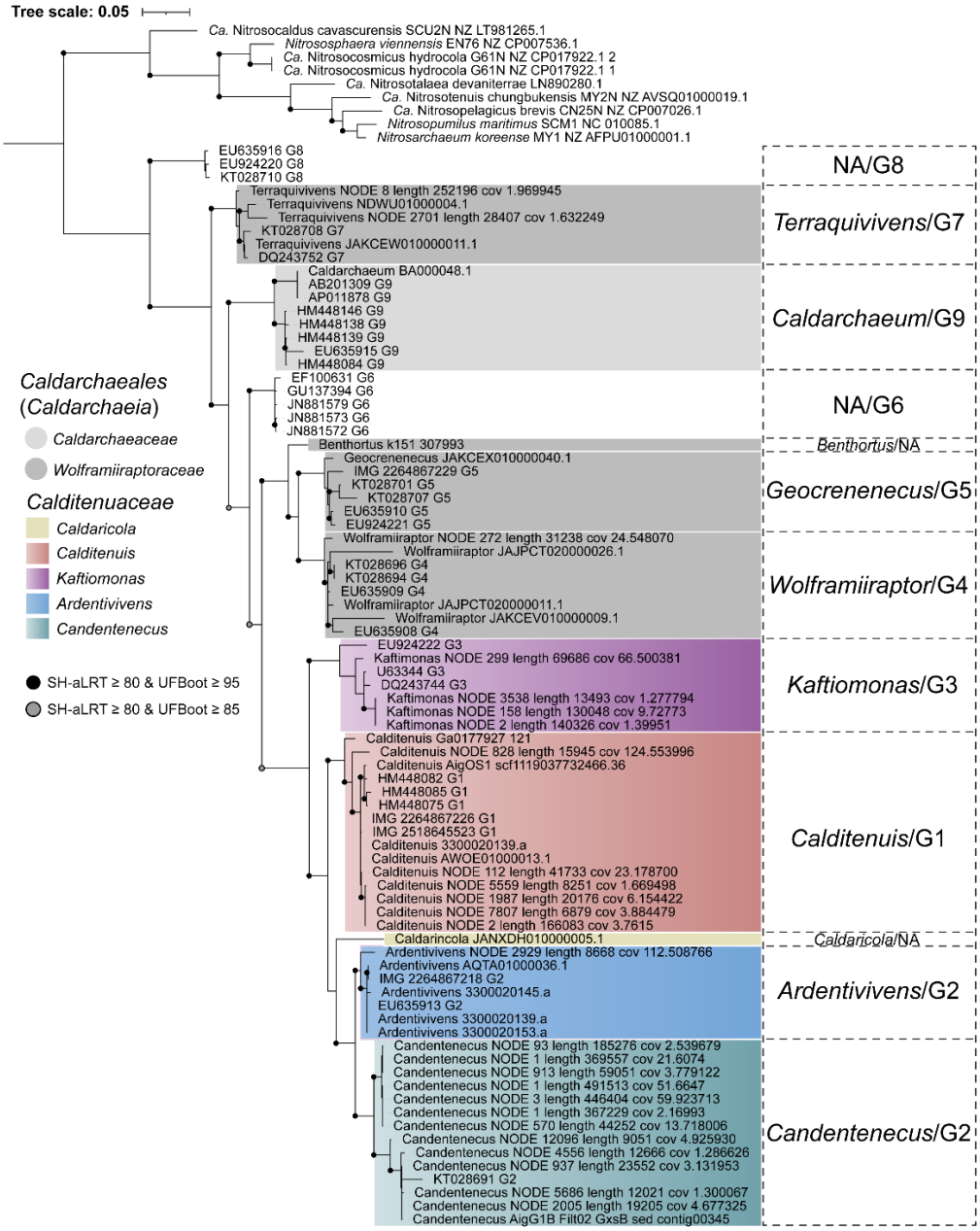
Supplementary Fig. 10. Identification of *Calditenuis* 16S rRNA gene sequences. Cladogram inferred from Maximum-likelihood-based phylogenetic tree of *Caldarchaeales* includes 16S rRNA gene sequences from Supplementary Fig. 12 and sequences from cultures amended with different types of amino acids, identified as *Caldarchaeales* by QIIME with Naive Bayes classifiers trained on the SILVA138 database. Sequences were aligned and trimmed, and phylogenetic reconstruction was performed using IQ-TREE with the best-fit model and 1000 pseudoreplicates. *Calditenuis* sequences are highlighted with a red background. Dotted nodes with darker colors indicate SH-aLRT support ≥80% and UFboot support ≥95%, while lighter colors indicate SH-aLRT support ≥80% and UFboot support ≥85%.



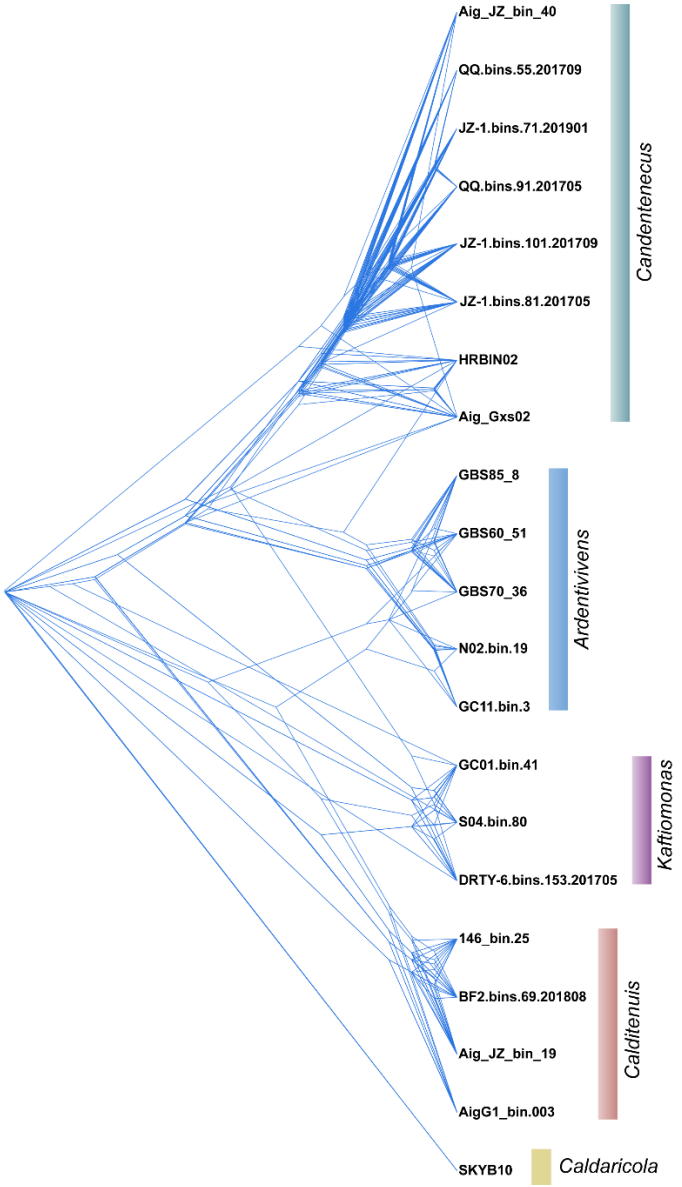
Supplementary Fig. 11. Robust species phylogeny of *Calditenuaceae*. Maximum-likelihood phylogenomic trees were inferred from four conserved marker gene/protein sets: 53 archaeal-specific marker proteins (ar53; **a**), 56 universal marker proteins (uni56; **b**), ribosomal proteins (rprot; **c**), and up-to-date archaeal core genes (UACG; **d**). The optimal substitution model was applied to each partition of the marker sequences (Supplementary Tables 26-29). Supported branches are indicated by filled circles at the nodes at the species level and above. Different species were visually differentiated by distinct colors, with species within the same genus represented in varying shades of the same color. Taxon names in bold and marked with a star represent the proposed nomenclatural types. Taxon names in gray indicate species that lack a high-quality representative genome. **e,** Venn diagram shows the overlap among the marker gene/protein sets of ar53, uni56, rprot, and UACG.



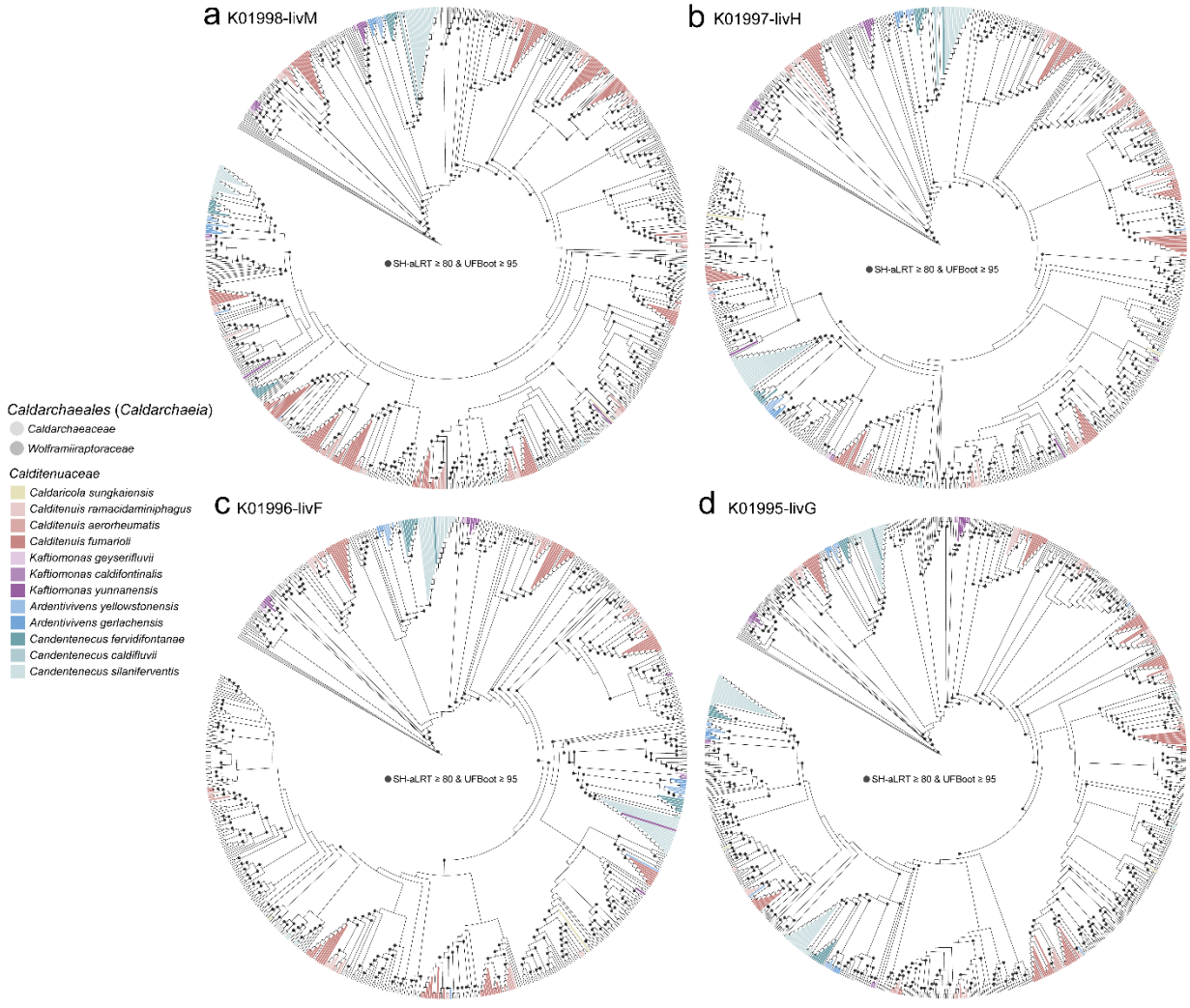
Supplementary Fig. 12. Maximum-likelihood tree of *Caldarchaeales* 16S rRNA gene sequences. 16S rRNA gene sequences from genomes with lengths above 940 bp, along with representative sequences from Hedlund *et al*. (2015), were included for phylogenetic reconstruction. The sequences were aligned and trimmed, and phylogenetic reconstruction was performed using IQ-TREE with the best-fit model and 1000 pseudoreplicates. Different genus-level groups were labeled based on this study and Hedlund *et al*. (2015). "NA" represents cases where no 16S rRNA genes belonging to the corresponding genera were found in the genomes included in the study. Dotted nodes with darker colors indicate SH-aLRT support ≥80% and UFboot support ≥95%, while lighter colors indicate SH-aLRT support ≥80% and UFboot support ≥85%.



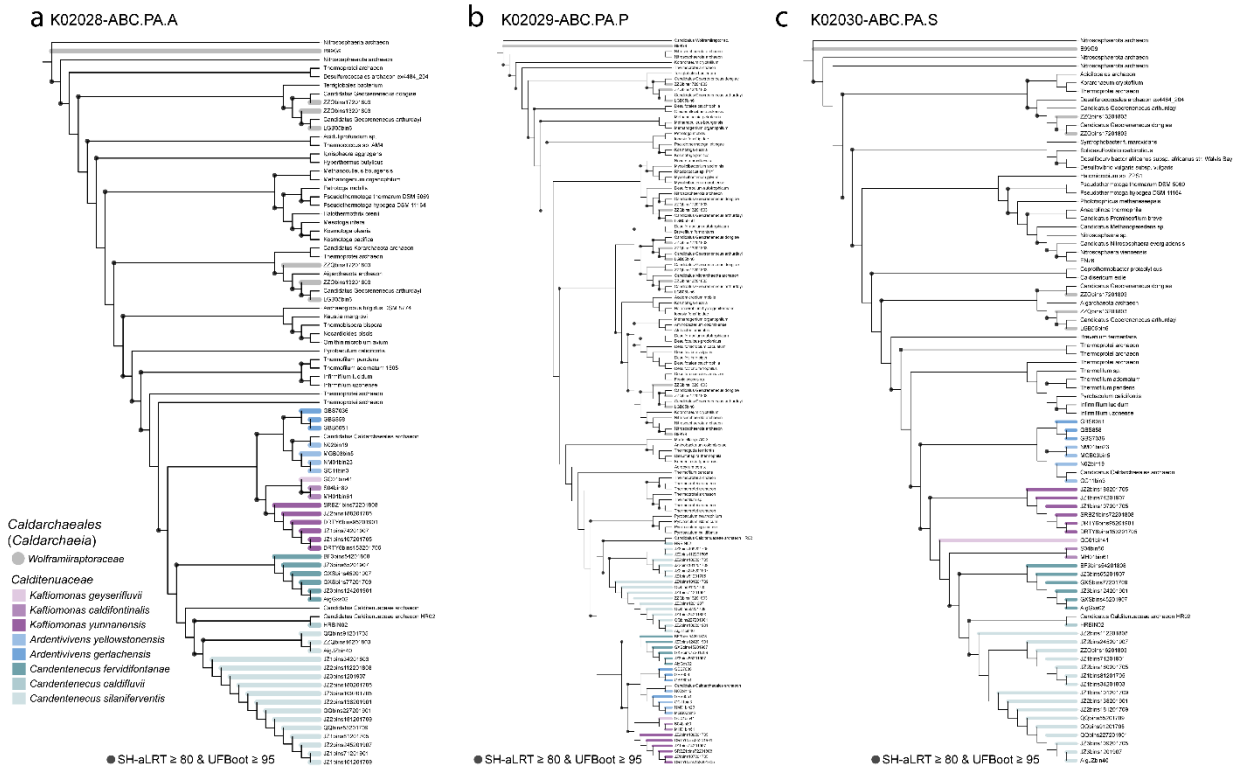
Supplementary Fig. 13. Densitree plot showing genealogical concordance of 34 shared proteins from ar53 marker protein sets. Code developed in-house was used to filter *Calditenuaceae* genomes based on two rules: 1) each species must have at least one representative genome, and 2) as many marker proteins as possible should be retained in the analysis. Ultimately, each species had at least one genome with all 34 marker protein sequences. Single protein trees were constructed using IQ-TREE based on the best-fit model inferred from 1000 replicates. Different colors represent different genus-level groups in the *Calditenuaceae* family.



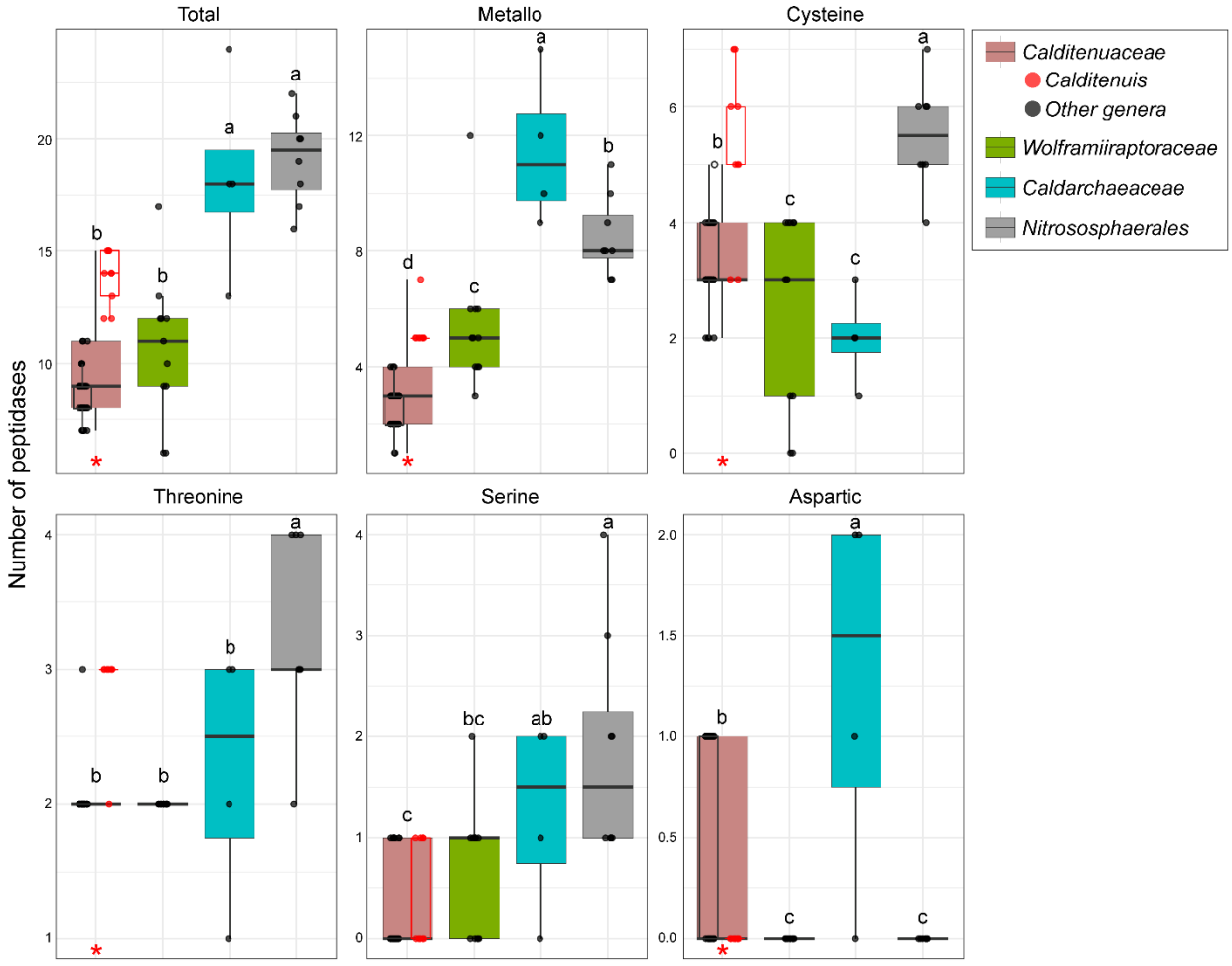
Supplementary Fig. 14. Cladograms of BCAA transporter proteins: (a) K01998-LivM, (b) K01997-LivH, (c) K01996-LivF, (d) K01995-LivG. For each BCAA transporter protein category, sequences annotated as the BCAA transporter protein, along with the top five best-matched sequences from the NCBI NR database and the top five best-matched sequences from the UniProt database, were retrieved, clustered, aligned, and trimmed. A maximum-likelihood (ML) phylogeny was constructed using IQ-TREE based on the best-fit model with UFBoot and SH-aLRT support values inferred from 1000 replicates each and the cladograms were inferred from the ML phylogenies. The taxa was color-coded following the scheme used in Figure 6.



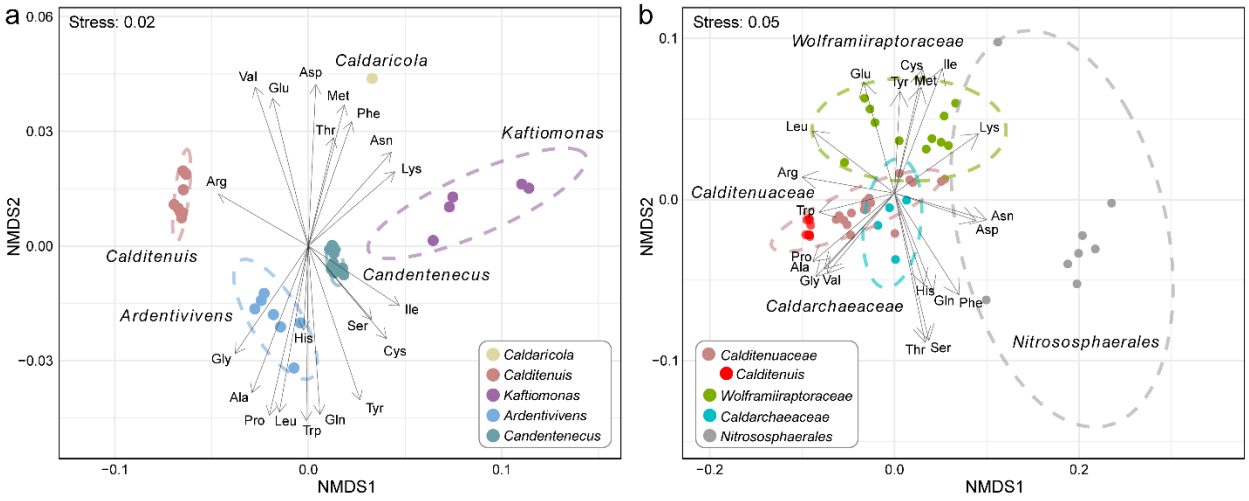
Supplementary Fig. 15. Maximum-likelihood phylogenetic tree of PAA transporter proteins: (a) K02028-ABC.PA.A, (b) K02029-ABC.PA.P, (c) K02030-ABC.PA.S. For each PAA transporter protein category, sequences annotated as the PAA transporter protein, along with the top five best-matched sequences from the NCBI NR database and the top five best-matched sequences from the UniProt database, were retrieved, clustered, aligned, and trimmed. Phylogeny was constructed using IQ-TREE based on the best-fit model with UFBoot and SH-aLRT support values inferred from 1000 replicates each. The taxa was color-coded following the scheme used in Figure 6.



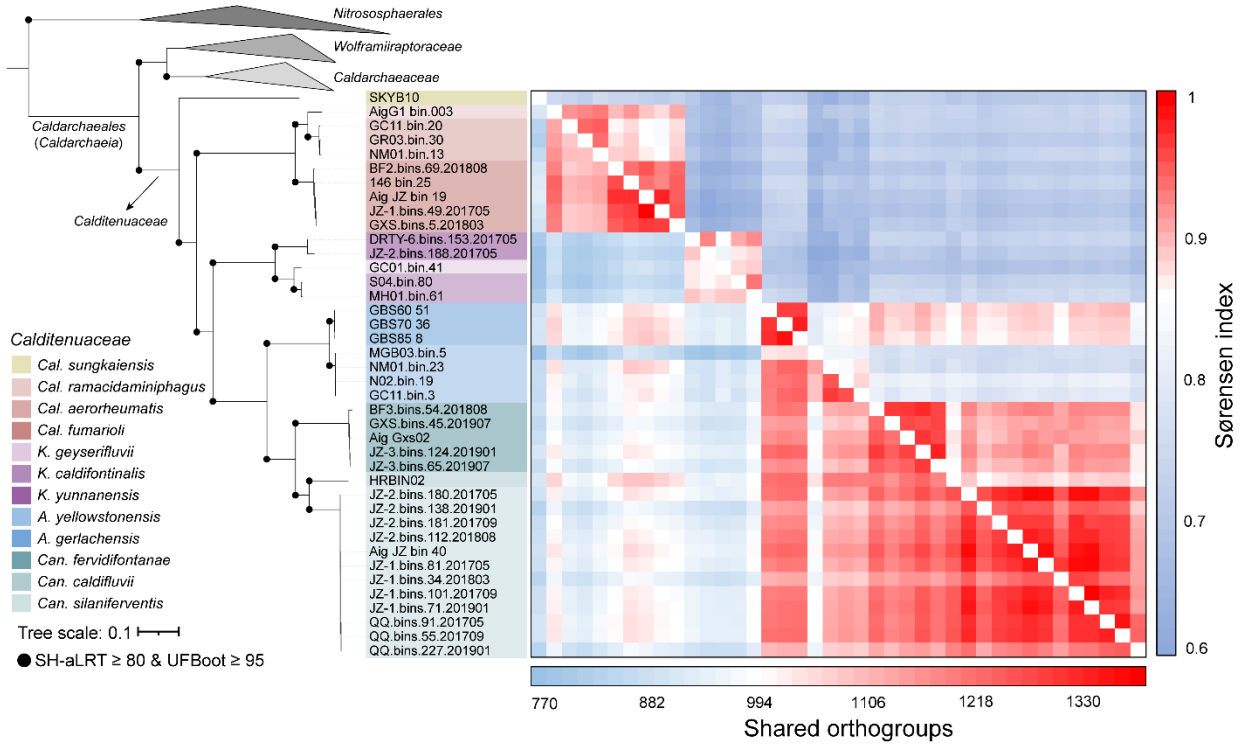
Supplementary Fig. 16. Catalytic types of peptidases in protein-coding sequences in high-quality *Calditenuaceae* genomes. Peptidases are divided into seven classes based on the mechanism of catalysis: metallo, cysteine, serine, aspartic, threonine, glutamic, and asparagine peptidases. Different colors represent different species in the *Calditenuaceae* family.



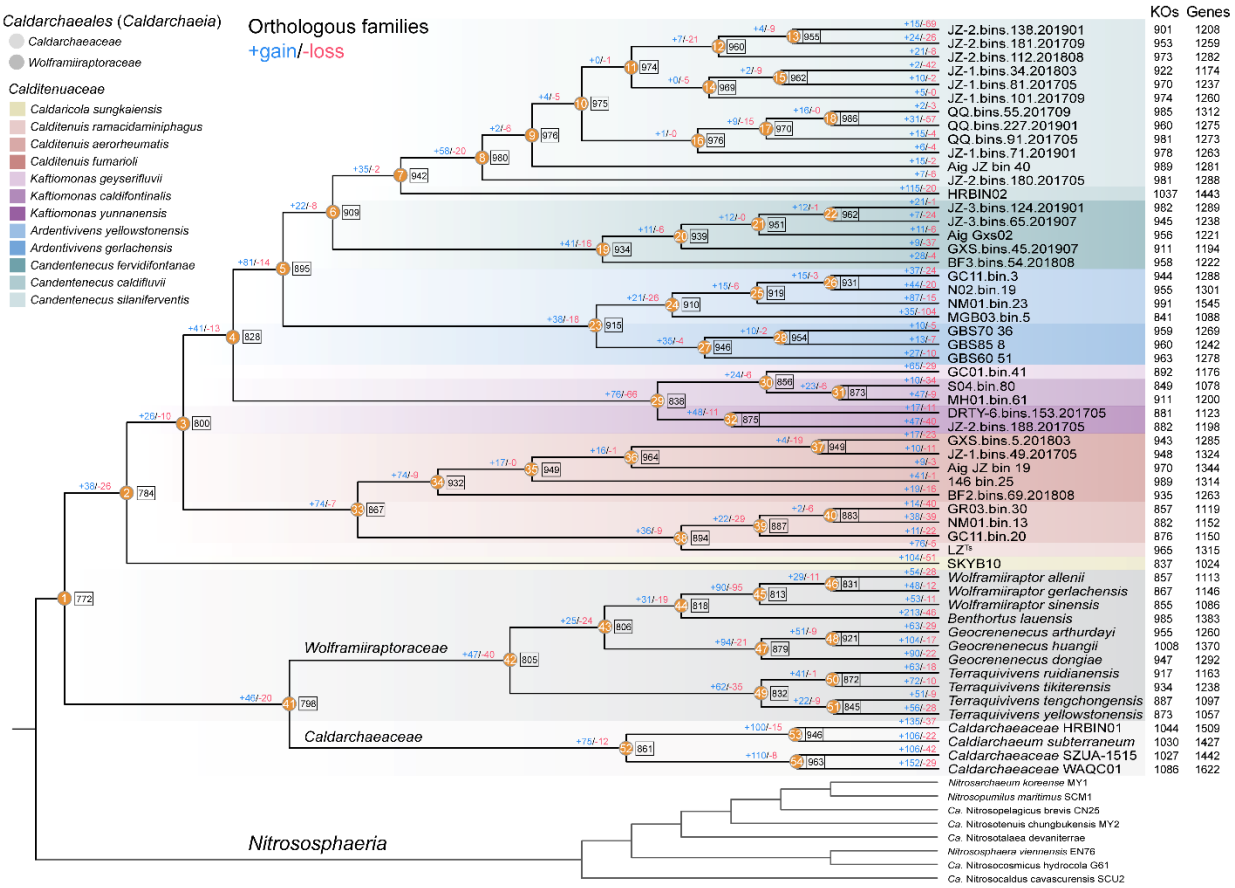
Supplementary Fig. 17. Non-metric Multidimensional Scaling (NMDS) analysis of amino acid composition in protein-coding sequences in high-quality *Calditenuaceae* (a) and all the high-quality genomes included in the study (b). The NMDS analysis utilized Bray-Curtis distances to assess amino acid composition differences. Colored dots represent distinct groups, and ellipses denote 95% confidence intervals.



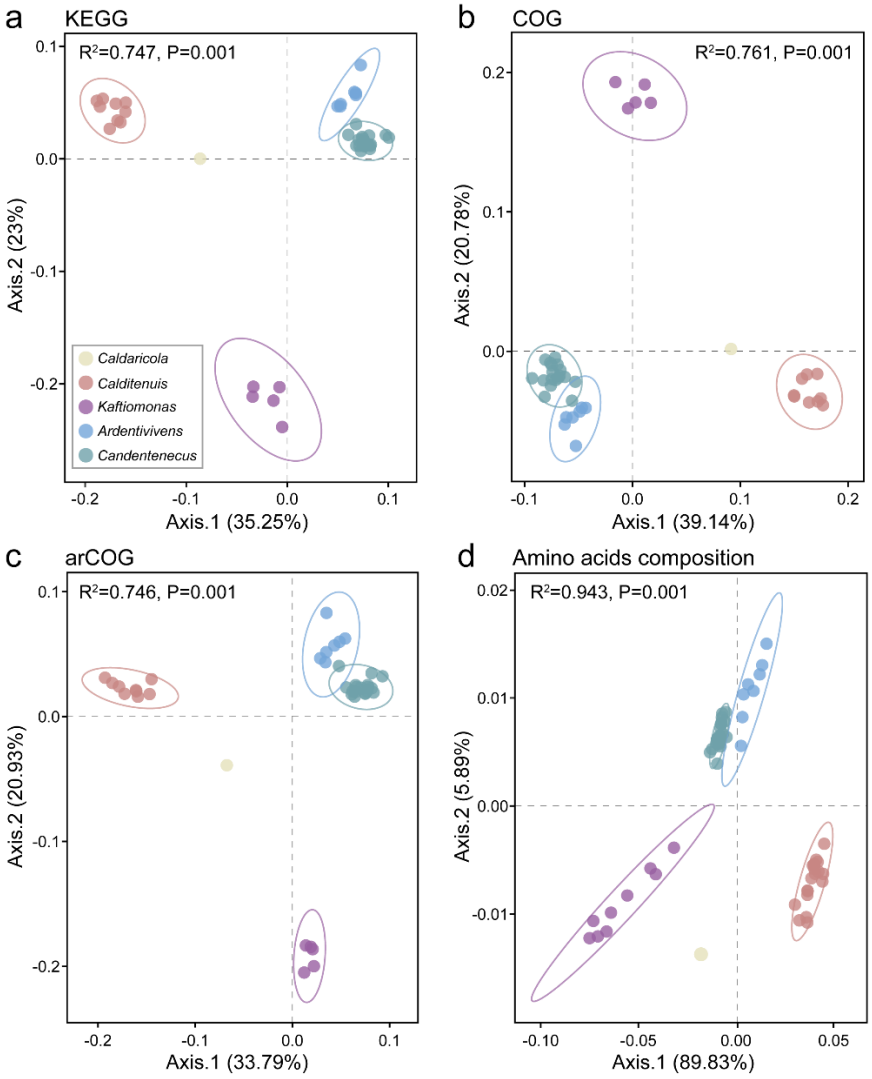
Supplementary Fig. 18. Shared orthogroups in the *Calditenuaceae* family. Heatmap showing shared orthogroups (lower left of the square) and Sørensen dissimilarity index (upper right of the square) between genomes in the high-quality *Calditenuaceae* genomes. The Sørensen index is used to account for differences in genome sizes, providing a more accurate measure of similarity by giving more weight to shared orthogroups.



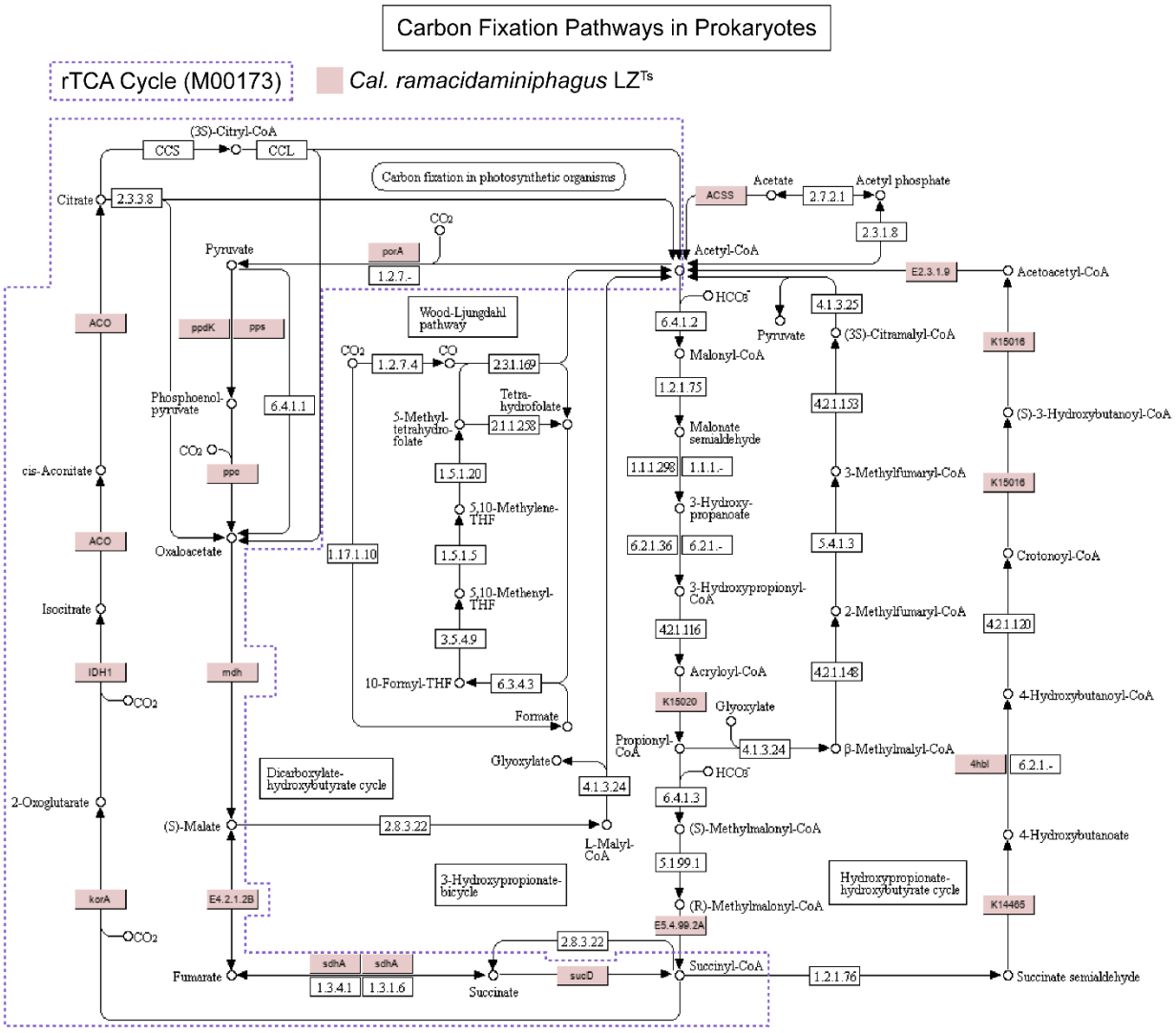
Supplementary Fig. 19. Ancestral gene content reconstruction in the *Calditenuaceae* family. Count software was used to capture the dynamics of the evolution of orthologous families. The tree topology on the left is derived from the tree established using 53 conserved archaeal marker sequences. The numbers of gain and loss events are marked on each lineage of the tree. "+" symbols represent gain events (blue), and "-" symbols represent loss events (red). The color scheme follows that of Figure 6 to differentiate taxa.



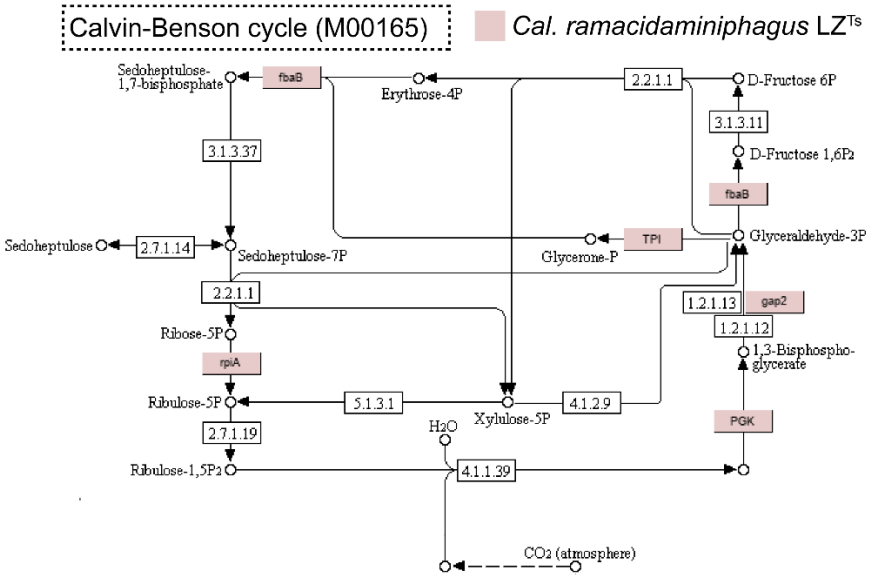
Supplementary Fig. 20. Principal Coordinates Analyses (PCoA) of *Calditenuaceae* genomes. PCoA plots depict distances between *Calditenuaceae* genomes using Bray-Curtis dissimilarity based on functional profiling annotated by (a) KEGG, (b) COG, (c) arCOG, and (d) amino acid composition. Different colors represent genus-level groups.



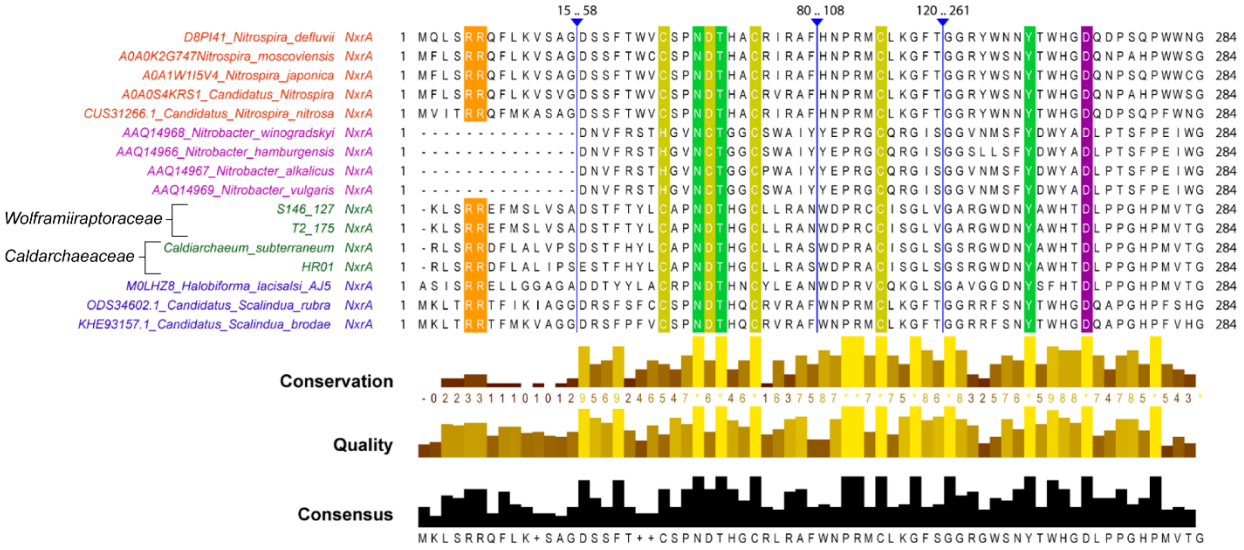
Supplementary Fig. 21. Carbon fixation pathways in *Cal. ramacidaminiphagus* LZ^{Ts}. The map illustrates genes involved in the carbon fixation pathways (KEGG: map00720), with a red background highlighting their presence in LZ^{Ts}.



Supplementary Fig. 22. Calvin-Benson cycle in *Cal. ramacidaminiphagus* LZ^{Ts}. The map illustrates genes involved in the Calvin-Benson cycle (KEGG: M00165), with a red background highlighting their presence in LZ^{Ts}.



Supplementary Fig. 23. Structural characterization of the alpha subunit of nitrite oxidoreductase (NxrA). Partial sequence alignment showcasing key conserved residues. Highlighted regions indicate Tat signal peptides (orange), Fe-S cluster coordinating residues (yellow), molybdenum-binding residues (purple), and nitrite/nitrate interaction sites (green).



References

1. Beam JP*, et al.* Ecophysiology of an uncultivated lineage of Aigarchaeota from an oxic, hot spring filamentous 'streamer'community. *ISME J* **10**, 210-224 (2016).
2. Oren A, Garrity GM, Parker CT, Chuvochina M, Trujillo ME. Lists of names of prokaryotic Candidatus taxa. *Int J Syst Evol Microbiol* **70**, 3956-4042 (2020).
3. Liew KJ*, et al.* Integrating multi-platform assembly to recover MAGs from hot spring biofilms: insights into microbial diversity, biofilm formation, and carbohydrate degradation. *Environmental Microbiome* **19**, 29 (2024).
4. Thomas SC*, et al.* Position-specific metabolic probing and metagenomics of microbial communities reveal conserved central carbon metabolic network activities at high temperatures. *Front Microbiol* **10**, 1427 (2019).
5. Kreil DP, Ouzounis CA. Identification of thermophilic species by the amino acid compositions deduced from their genomes. *Nucleic Acids Res* **29**, 1608-1615 (2001).
6. Fukuchi S, Nishikawa K. Protein surface amino acid compositions distinctively differ between thermophilic and mesophilic bacteria. *J Mol Biol* **309**, 835-843 (2001).
7. Zhou X-X, Wang Y-B, Pan Y-J, Li W-F. Differences in amino acids composition and coupling patterns between mesophilic and thermophilic proteins. *Amino Acids* **34**, 25-33 (2008).
8. Dawrs SN*, et al.* Hawaiian volcanic ash, an airborne fomite for nontuberculous mycobacteria. *GeoHealth* **8**, e2023GH000889 (2024).
9. Van Eaton AR, Harper MA, Wilson CJ. High-flying diatoms: Widespread dispersal of microorganisms in an explosive volcanic eruption. *Geology* **41**, 1187-1190 (2013).
10. Self S, Rampino MR. The 1883 eruption of Krakatau. *Nature* **294**, 699-704 (1981).
11. Nunoura T*, et al.* Insights into the evolution of Archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. *Nucleic Acids Res* **39**, 3204-3223 (2011).
12. Rinke C*, et al.* Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**, 431-437 (2013).
13. Hua Z-S*, et al.* Genomic inference of the metabolism and evolution of the archaeal phylum Aigarchaeota. *Nat Commun* **9**, 2832 (2018).
14. Steffens L*, et al.* High CO2 levels drive the TCA cycle backwards towards autotrophy. *Nature* **592**, 784-788 (2021).
15. Peters JW*, et al.* [FeFe]-and [NiFe]-hydrogenase diversity, mechanism, and maturation. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* **1853**, 1350-1369 (2015).
16. Murali R*, et al.* Diversity and evolution of nitric oxide reduction in bacteria and archaea. *Proc Natl Acad Sci* **121**, e2316422121 (2024).
17. Huerta-Cepas J, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res* **47**, D309-D314 (2019).
18. Brito JA*, et al.* Structural and functional insights into sulfide: quinone oxidoreductase. *Biochemistry* **48**, 5613-5622 (2009).